CONTROL ESTADISTICO DE CALIDAD

(del 3 de octubre al 30 de noviembre)

FECHA	HORARIO	T E M A	PROFESOR
octubre 3	18 a 21 h	OBJETIVOS E IMPORTANCIA DEL CONTROL ESTADISTICO DE CALIDAD	Ing. Manuel Marín González
octubre 5,10,17, 19,24 y 26	18 a 21 h	TEORIA DE PROBABILIDADES ESTADISTICA DESCRIPTIVA INFERENCIA ESTADISTICA	Dr. Octavio A. Rascón Chávez
octubre 31, Nov. 7 9 y 14	18 a 21 h	CARTAS DE CONTROL	M. en I. Augusto Villarreal A.
noviembre 16,21 y 23	18 a 21 h	MUESTREO DE INSPECCION	M. en I. Augusto Villarreal A.
noviembre 28 y 30	18 a 21 h	CONFIABILIDAD DE COMPONENTES Y SISTEMAS CLAUSURA	Dr. Octavio A. Rascón Chávez

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

CONTROL ESTADISTICO DE CALIDAD

EL CONTROL DE CALIDAD COMO MECANISMO PARA ADECUAR LA CIENCIA Y LA TECNOLOGIA CESIDADES SOCIO ECONOMICAS

ING. MANUEL MARIN GONZALEZ

OCTUBRE, 1978.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

EL CONTROL DE CALIDAD COMO MECANISMO PARA ADECUAR LA CIENCIA Y LA TECNOLOGIA A LAS NECESIDADES SOCIO ECONOMICAS

Material presentado por el Ing Manuel Marín González, Jefe de la Unidad de Normalización Básica del CONACYT, al Curso de Control Estadístico de Calidad en el Centro de Educación Contínua de la Facultad de Ingeniería de la UNAM, el 3 de octubre de 1978.

El control de calidad no es solamente el mecanismo que permite verificar sistemáticamente que los productos se fabriquen conforme a rigurosas especificaciones previamente establecidas; también es el mecanismo mediante el cual la ciencia y la tecnología se vinculan a las exigencias planteadas por las necesidades socio - económicas, consecuentemente es el mecanismo del desarrollo. Un mecanismo madiante el cual se consiguen objetivos rigurosos previamente especificados, se denomina técnicamente servomecanismo. El control de calidad es el servomecanismo por el cual la ciencia y la tecnología impulsan el desarrollo. El presente trabajo desarrolla este tema en seis apartados. En el primero precisaremos el concepto teórico del servomecanismo, como medición, evaluación y acción apropiada; en el segundo, el punto nodal del servomecanismo, la testificación, como medición del nivel actual de la calidad que descubre las deficiencias que deben ser superadas; en el tercero, las experiencias concretas en el establecimiento de centros de testificación; en el cuarto, la metodología derivada de esa experiencia; en el quinto, el aspecto socio-económico, filosofía, en el -

que se intenta concretar cómo entre lo anterior y lo posterior está la normalización, vale decir, la tendencia objetiva de los hechos, tendencia que puede ser proyectada hacia el futuro con cierta probabilidad, pero también la posibilidad de retroalimentar los errores que se advirtieran respecto de nuevas experiencias; y finalmente, en el sexto, como conclusión hablaremos de la circularidad del control de calidad como el mecanismo que conduce a la definición de las acciones apropiadas de la ciencia y la tecnología para impulsar el desarrollo.

1. <u>El servomecanismo</u>.- El servomecanismo se capta eficazmente en el sencillo esquema adjunto, esquema que representa el más elemental circuito mediante el cual el resultado de la producción es medido y evaluado, de manera que de la evaluación podemos desprender la razón apropiada para corregir, si ello es necesario, la producción de manera de mantenerla dentro de valores previamente establecidos. El servomecanismo es así el mecanismo del control de calidad. Prácticamente desde sus inicios ha sido acreditada una definición del control por la Sociedad Americano de Ingenieros Mecánicos (ASME); medición del valor actual de una variable, o el individual de una serie, y acción apropiada para mantener dicho valor dentro de magnitudes previamente establecidas. Con todo y la sencillez, tanto de la definición del control de calidad, como del servomecanismo, conviene examinar cada uno de sus elementos. Son cuatro:

1.1 <u>Medición</u>.- Desde el pasaje bíblico de la creación, según el cual después de cada una de las grandes acciones sucesivas de Dios, él advierte que lo hecho estaba bien, y en base de eso proseguía la creación. Medir es referir cualquier cosa a otra, tomada como patrón de referencia. Hay patrones para cada parámetros que requerimos medir, patrones de longitud, masa, tiempo, temperatura, corriente eléctrica, intensidad luminosa, substancia. El patrón de referencia también puede ser una norma, vale decir, la definición de la calidad esperdad, referida a varios parámetros y, consecuentemente a varios patrones. Este acto de la medición va desde la simple apreciación visual más o menos esmerada, hasta la medición más rigurosa, efectuada con los instrumentos que ofrecen la mayor confiabilidad. La medición es el elemento esencial de la ciencia. Solo cuando los conceptos son referidos a mediciones rigurosas, son realmente científicos. Es este el sentido de la formulación de Lord Kelvin, la medición es el proceso y el único proceso por el que conocemos.

1.2 Evaluación.- La evaluación podría ser una parte dela medición misma, es le medición misma. Sin embargo dado el desarrollo independiente, aunque interdependiente de la metrología y la normalización vale la pena hacer la definición. Actualmente por metrología se entiende con precisión, la medición propiamente dicha, vale decir la identificación de patrones apropiados y la metrología refiere las magnitudes a esos patrones. La metrología es una ciencia básica que tiene amplísimos desarrollo. Estos se presentan no solamente en relación con la multitud de métodos de medición que cada día van mejorando los establecidos, sino también en virtud de que el patrón va mejorándose en cuanto a la identificación de fenómenos invariantes más accesibles a los que se pueda referir ya a la frecuencia no solamente las unida-- des de longitud y tiempo, sino que se intenta también referir a ella la unidad eléctrica, la luminosa y la de temperatura, incluso no está lejano el día en que igualmente se pueda referir a la frecuencia la ma sa y, en consecuencia, hasta la unidad de substancia. Ello es debido a que la frecuencia es el fenómeno más universal, accesible y manipulable, a más de disponible. En cuanto a la normalización, también es ta tiene amplísimos desarrollos, quizá hasta cada día se acerca más su concepto al de la ciencia misma. La normalización es la definición científica de la calidad. Esta, la calidad, que no es sino la medida en la que la producción alcanza las exigencias de la necesidad, en su de sarrollo da origen al conocimiento cada vez más riguroso. Si se inicia en una discriminación, bueno - malo, posteriormente va refirién dose a características ponderables, más tarde mesurables y finalmen te rigurosamente definidas. Este rigor se logra por supuesto por la metodología científica, si bien sin perder jamás, puesto que en el con cepto de calidad es consustancial la dialéctica, el encuentro entre la intencionalidad de la producción y la exigencia de la necesidad. Ello - determina que en la definición rigurosa de la calidad participen invariablemente productores y usuarios, independientemente de que ellos llegaran a la ciencia. Más tarde se hace necesario reducir la calidad a la medición y ello constituye la normalización básica.

1.3 Error.- La evaluación trae consigo el descubrimiento de una deficiencia o insuficiencia de la producción respecto a sus exigencias, - el servomecanismo lo llama técnicamente error. Por muchas razones a este error se le da la mayor importancia, al grado de que al servomecanismo mismo suele llamársele "error actuated device". La designación no es absurda en manera alguna, aún menos conforme a la moderna filosofía del absurdo, en la que el propio hombre es considerado como un "error actuated device". En efecto, el hombre no acepta errores, reacciona naturalmente, incluso a veces violentamente ante la súbita sensación de hallarse fuera de norma. El principio es la confrontación, si bien la dinámica arranca de la identificación del --- error. También este error es el que impulsa el desarrollo y lo hace en virtud de que al advertir tal error, el impulso natural es el de superarlo. El error conduce a la acción apropiada.

1.4 Acción apropiada.- Esta puede ser desde la simple acción surgida del sentido común, la aplicación de tecnologías del dominio público, la búsqueda de tecnologías apropiadas y aún la investigación de nuevas tecnologías, nuevos principios, nueva ciencia. Luego se ve que el olvido de estos principios son los que han conducido al mundo al evidente desgarramiento que sucede actualmente. No solamente por un lado camina la producción y por otro la ciencia y la tecnología, sino que vivimos el anacronismo de la coexistencia de una ciencia pura y otra - aplicada. Tal vez ello sea solamente un reflejo de la exacerbación de la lucha de clases que los científicos repiten sin siquiera tener conciencia de ello. Pero es el enfrentamiento del hombre a una realidad que lo sobrepasa lo que invariablemente lo impulsa a la investigación en - todos sus aspectos. El servomecanismo, el mecanismo del control de calidad, es el mecanismo que al vincular la ciencia y la tecnología - con la actividad productiva, impulsa su desarrollo.

2. Testificación.- Por muchas razones la testificación es el punto - nodal de todo el control de calidad. Solo en cuanto medimos es que descubrimos nuestras deficiencias, mismas que nos impulsan a la superación de las mismas. La testificación supera todas las subjetividades. Frecuentemente nuestro atraso se debe a que suponemos que por el solo hecho de aplicar una tecnología suficientemente acreditada, con ello tenemos para asegurar la calidad. Sin embargo cuántos trastornos se derivan de esta creencia. Por ejemplo, cada vez se demuestra más claramente que los productos fabricados en México correspondientes a empresas extranjeras ofrecen una calidad diferente. Ello es natural, no solamente en virtud de que además de la tecnología actúan - otros elementos, como son las materias primas y materiales, la mano de obra, etc., sino principalmente por otras dos razones diferentes. Primero, la propia tecnología difícilmente es la misma, más aún cuando en ella incluimos un concepto mucho más amplio que el del -- simple funcionamiento del equipo y, segundo, la falta de exigencias. Si la calidad insistimos, es dialéctica, encuentro entre producción y exigencia de la necesidad, si ésta no existe o existe débilmente, la calidad no existe rigurosamente. Todo ello lo resuelve la testificación.

Esto nos da cuenta de la situación realmente existente. Por supuesto, a la testificación corresponde actualmente también amplísimos desarrollos. No es ya la simple rutina que se práctica en los laboratorios nacionales. La testificación tiene metodologías cada vez más rigurosas, ella misma obedece a normas cada vez más precisas. Principalmente tiene que obedecer a tres elementos esenciales, a la normalización, al muestreo y a la metrología. La testificación tiene que referirse a normas precisas, mismas que tienen que aplicarse conforme a planes de muestreo rigurosos, y los equipos tienen que ser referidos sistemáticamente a patrones que los mantengan en un estado confiable.

3. Experiencias concretas.- CONACYT puede ofrecer una interesante experiencia que confirma la tesis anterior. A la fecha ha establecido toda una serie de centros de investigación y asistencia tecnológica conforme a móviles que van desde la promoción profesional del Consejo, al planteamiento de necesidades concretas y el requerimiento más o menos veleidoso que tiene su origen en móviles hasta políticos. También el control de calidad ha estado en la base de la creación de algunos de estos centros. Concretamente se pueden mencionar el Centro de Investigación y Asistencia Tecnológica del Estado de Guajanuato -- (CIATEG), el Instituto de Investigaciones Siderúrgicas (IMIS), el Centro de Investigación en Ciencias del Mar (CICIMAR), etc. Se trata en todos estos casos del establecimiento de un centro que en su origen no tuvo otro propósito que el de ofrecer una testificación de calidad en la coyuntura materia prima - producto. En el caso del CIATEG, -

entre las industrias del cuero y el calzado; entre el hierro y el acero y la industria metal-mecánica, en el caso del IMIS; y entre los productos pesqueros directamente extraidos del mar y su utilización en las empacadoras y distribuidoras, en el caso de CICIMAR. Se estableció una capacidad técnica de testificación definida conforme a las normas más apropiadas y a la capacidad más conveniente. Lo primero que ocurre es el mejoramiento de la calidad del producto, pero sucesivamente van desprendiéndose una concientización de los abastecedores, una concientización de los fabricantes, usuarios de esos abastecimientos, la introducción de métodos de sentido común en ambos lados, la aplicación de técnicas apropiadas y hasta la investigación de nuevas técnicas y nuevos principios. La experiencia de CONACYT es suficiente para confirmar no solamente el desarrollo hasta estos niveles sino, lo que es más importante, que cuando este ha sido el móvil el centro promovido por CONACYT ha tenido mayor consistencia y desarrollo. Puede decirse incluso que por este procedimiento CONACYT ha descubierto un método apropiado para el establecimiento de un nuevo centro de investigación y asistencia tecnológica.

4. Metodología.- La metodología que CONACYT ha elaborado, actualmente la tiene prácticamente sistematizada. Se parte de la identificación precisa del sector. Ello se hace en función de algunos signos o índices apropiados, por ejemplo el aporte de los diversos sectores productivos al producto industrial regional, el número de establecimientos, el monto de su inversión, el número de trabajadores, etc.,

datos todos ellos que permiten medir de manera apropiada el nivel de productividad. El centro tiene como objetivo fundamental el incremento de esta productividad en función de las deficiencias cuantificables de la calidad. Identificado el sector aún se verifica la trascendencia de su actividad. Ello se hace por la tendencia objetiva que -- ofrezca la actividad productiva del sector identificado. Esta tendencia objetiva puede identificarse en relación con la producción, el consumo, los procesos, las calidades, los precios, las importaciones, las exportaciones, los problemas de transferencia, etc. Se trata de identificar la medida en la que la producción alcanza las exigencias de la necesidad, no solamente en volumen, sino principalmente en calidad y precio. Identificada esta tendencia se plantea enseguida la problemática - que ello determina en la investigación y el cubrimiento que esta hace de la misma. Cubrimiento del que se determina la factibilidad. Tiene que verse sin embargo la adecuación del centro a las necesidades de su implementación y aún la confiabilidad, cosa que se hace en función de todos los factores anteriores. Posteriormente, una vez establecido el centro, se establecen con todo rigor sus metas concretas y el calendario de su cumplimiento con el objeto de seguir después su desarrollo, en la inteligencia de que todos los errores que se identificaran en relación con esas metas y calendarios habrán de corregirse - rigurosamente. La norma invariable del seguimiento consiste en advertir la medida en la que la producción alcanza las exigencias. Es en esta medida en la que se advierte la eficacia del centro. En - -

CONACYT hemos llegado incluso a la necesidad de introducir la circularidad del control a las propias empresas productivas, cosa que estamos haciendo felizmente en un trabajo que incluye no solamente las acciones del propio CONACYT, sino también la participación de los propios industriales y otros elementos activos de la localidad. Nos ayudamos incluso de la asistencia técnica internacional.

5. El aspecto socio-económico, filosofía.- La esencia del control de calidad está en la adecuación de la ciencia y la tecnología a las exigencias socio-económicas. La anacrónica división entre ciencia pura y ciencia aplicada distorsiona aún más que ayuda al desarrollo. Frecuentemente la investigación pura se traduce más en servicio al capital internacional que al propio desarrollo nacional. La esencia del control, la esencia de la normalización esta en la interpolación, extrapolación y rectificación de las hipótesis del desarrollo. El punto de partida es invariablemente el hombre y sus necesidades, es él el que plantea las exigencias del desarrollo y lo hace en todos los órdenes. Toda la producción, pero el desarrollo de ésta demanda el conocimiento cada vez más riguroso, extenso y avanzado. Es por la retroalimentación de la información captada al verificar los resultados actuales de la producción que descubrimos las deficiencias, errores que determinan las acciones apropiadas de la ciencia y la tecnología para la superación.

6. Conclusiones.- Es la circularidad del control de calidad el mecanismo que conduce a la generación de tecnologías apropiadas y a la in

vestigación de nuevos avances en la ciencia para superar los niveles del desarrollo, es pues el servomecanismo por el cual ciencia y tecnología impulsan el desarrollo. Es esa circularidad aún la que determina las características de la capacitación que, de otra manera, están al garete, con sus correspondientes repercusiones no solo económicas, sino aún sociales. La circularidad del control de calidad, la óptima aplicación de las capacidades a las exigencias socio-económicas ha de superar muchas de las contradicciones y desgarramientos de la sociedad contemporánea.

MMG'lmm
2/X/78

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

CONTROL ESTADISTICO DE CALIDAD

TEORIA DE PROBABILIDADES

ESTADISTICA DESCRIPTIVA

INFERENCIA ESTADISTICA

DR. OCTAVIO RASCON CHAVEZ

OCTUBRE, 1978

TEORIA DE PROBABILIDADES

POR DR. OCTAVIO A. RASCON CH.

EXPERIMENTO. PARA FINES DE ESTE CURSO, SE ENTENDERA POR EXPERIMENTO A TODO PROCESO DE OBSERVACION. ASI UN EXPERIMENTO PUEDE SER PLANEADO Y REALIZADO POR EL HOMBRE, O PUEDE SER EFECTUADO POR LA NATURALEZA EN CASO DE UN FENOMENO NATURAL. POR EJEMPLO, EL LANZAR UNA MONEDA O UN DADO Y OBSERVAR LA CARA QUE QUEDA HACIA ARRIBA, ES UN EXPERIMENTO PLANEADO Y REALIZADO POR EL HOMBRE. EL OBSERVAR LA CANTIDAD DE AGUA QUE LLUEVE ANUALMENTE EN UNA CIUDAD, ES UN EXPERIMENTO ASOCIADO A UN FENOMENO NATURAL.

AL RESULTADO DE UN EXPERIMENTO SE LE DENOMINA *DATO* . A UN GRUPO DE DATOS SE LE LLAMA *MUESTRA*.

PROBABILIDAD: ES UNA MEDIDA DE LA CERTIDUMBRE QUE SE LE ASOCIA A LA OCURRENCIA U OBSERVACION DE UN RESULTADO DETERMINADO, AL REALIZARSE EL EXPERIMENTO CORRESPONDIENTE.

LA *TEORIA DE PROBABILIDADES* ES UNA RAMA DE LAS MATEMATICAS APLICADAS QUE TRATA LO CONCERNIENTE A LA ASIGNACION Y MANEJO DE PROBABILIDADES.

ESTADISTICA: ES LA RAMA DE LAS MATEMATICAS QUE SE ENCARGA DE ENSEÑAR LAS REGLAS PARA COLECTAR, ORGANIZAR, PRESENTAR Y PROCESAR LOS DATOS OBTENIDOS AL REALIZAR VARIAS VECES EL EXPERIMENTO ASOCIADO A UN FENOMENO DE INTERES Y PARA INFERIR CONCLUSIONES ACERCA DE ESTE ULTIMO. PROPORCIONA, ADEMAS, LOS METODOS PARA EL DISEÑO DE EXPERIMENTOS Y PARA TOMAR DECISIONES CUANDO APARECEN SITUACIONES DE INCERTIDUMBRE.

ESTADISTICA

* *DESCRIPTIVA.*- TRATA LO CONCERNIENTE A LA OBTENCION, ORGANIZACION, PROCESAMIENTO Y PRESENTACION DE LOS DATOS.
* *INFERENCIAL.*- TRATA LO CONCERNIENTE A LOS METODOS PARA INFERIR CONCLUSIONES ACERCA DEL FENOMENO DEL CUAL PROVIENEN LOS DATOS

ESTADISTICA DESCRIPTIVA

POR DR. OCTAVIO A. RASCON CH.

DATO U OBSERVACION: ES EL RESULTADO DE REALIZAR UN EXPERIMENTO

MUESTRA: ES UNA COLECCION DE DATOS

MUESTREO: PROCESO DE ADQUISICION DE UNA MUESTRA

MUESTREO
- *CON REEMPLAZO*.- CUANDO CADA ELEMENTO OBSERVADO SE REINTEGRA AL LOTE DEL CUAL FUE EXTRAIDO ANTES DE EXTRAER EL SIGUIENTE.
- *SIN REEMPLAZO*.- CUANDO CADA ELEMENTO OBSERVADO NO SE REINTEGRA AL LOTE.

POBLACION: TOTAL DE DATOS QUE SE PUEDEN OBTENER AL REALIZAR UNA SECUENCIA EXHAUSTIVA DE EXPERIMENTOS

POBLACION
- *DISCRETA*.- TIENE UN NUMERO FINITO O UN NUMERO INFINITO NUMERABLE DE DATOS POSIBLES
- *CONTINUA*.- TIENE UN NUMERO INFINITO NO NUMERABLE DE DATOS POSIBLES

EJEMPLOS

1. EXPERIMENTO: LANZAMIENTO DE UNA MONEDA DIEZ VECES

   POBLACION: SUCESION INFINITA NUMERABLE DE "CARAS" Y "CRUCES" (DISCRETA)

   MUESTRA: GRUPO DE 10 OBSERVACIONES

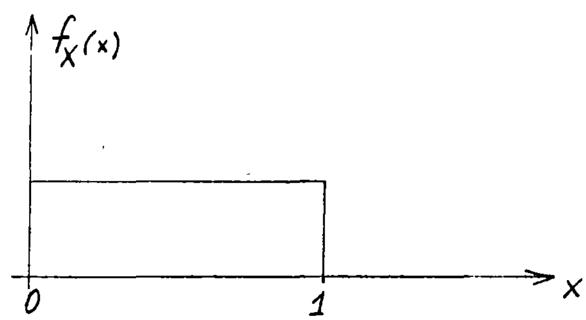
2. EXPERIMENTO: MEDICION DE LA PRECIPITACION PLUVIAL DIARIA EN LA CIUDAD DE MEXICO DURANTE DIEZ AÑOS

   POBLACION: SUCESION INFINITA NO NUMERABLE DE VALORES (CONTINUA)

   MUESTRA: GRUPO DE 3652 OBSERVACIONES (TOMANDO DOS AÑOS BISIESTOS DE 29 DIAS EN FEBRERO)

MUESTRA ALEATORIA: ES UNA MUESTRA OBTENIDA DE TAL MANERA QUE *TODOS* LOS ELEMENTOS DE LA POBLACION TIENEN LA *MISMA* PROBABILIDAD DE SER OBSERVADOS Y, ADEMAS, LA OBSERVACION DE UN ELEMENTO *NO AFECTA* LA PROBABILIDAD DE OBSERVAR CUALQUIER OTRO, ES DECIR, SI SON INDEPENDIENTES.

TABLA DE NUMEROS ALEATORIOS: ES UNA TABLA QUE CONTIENE NUMEROS QUE CONSTITUYEN UNA MUESTRA ALEATORIA OBTENIDA DE UNA DISTRIBUCION DE PROBABILIDADES UNIFORME, QUE GENERALMENTE CORRESPONDE A UNA VARIABLE ALEATORIA QUE PUEDE ASUMIR VALORES ENTRE 0 Y 1, MULTIPLICADOS POR 10^r, EN DONDE r ES EL NUMERO DE DIGITOS QUE SE DESEA TENGAN LOS NUMEROS.



LAS TABLAS QUE SE USEN PARA OBTENER UNA MUESTRA ALEATORIA DEBEN CONTENER NUMEROS CON MAYOR NUMERO DE DIGITOS QUE LOS QUE TIENE EL TOTAL DE ELEMENTOS DE LA POBLACION QUE SE VA A MUESTREAR. POR EJEMPLO, SI SE VA A OBTENER UNA MUESTRA ALEATORIA DE UN LOTE DE LENTES PARA MICROSCOPIO QUE TIENE 10,000 ELEMENTOS, LA TABLA QUE SE USE DEBERA TENER NUMEROS ALEATORIOS CON 5 O MAS DIGITOS.

METODO DE MUESTREO ALEATORIO

1. SE ENUMERAN LOS ELEMENTOS DE LA POBLACION
2. SE FIJA EL CRITERIO DE SELECCION DE LOS NUMEROS ALEATORIOS (POR EJEMPLO, SE DEFINE QUE RENGLONES Y QUE COLUMNAS SE VAN A LEER)
3. SE INDICA QUE DIGITOS SE VAN A ELIMINAR EN CASO DE QUE LOS NUMEROS DE LA TABLA TENGAN MAS DIGITOS QUE LOS NECESARIOS
4. SE LEEN LOS NUMEROS, DE ACUERDO CON LO FIJADO EN LOS PUNTOS 2 Y 3, Y SE EXTRAEN DEL LOTE LOS ELEMENTOS QUE TIENEN LOS NUMEROS LEIDOS. ESTOS CONSTITUYEN LA MUESTRA FISICA CON LA CUAL REALIZAR LOS EXPERIMENTOS. LAS OBSERVACIONES CONSTITUIRAN LA MUESTRA ALEATORIA DESEADA.

NOTA: TODOS LOS NUMEROS QUE SE REPITAN SE CONSIDERAN SOLO UNA VEZ. TAMBIEN SE ELIMINAN LOS NUMEROS MAYORES DEL TAMAÑO DEL LOTE.

EJEMPLO

SE TIENE UN LOTE DE 1,000 TRANSISTORES NUMERADOS DEL UNO AL MIL, CUYA CALIDAD SE VA A VERIFICAR ESTADISTICAMENTE, PARA LO CUAL SE DECIDE TOMAR UNA MUESTRA DE 40 ELEMENTOS Y MEDIR SU AMPLIFICACION, USANDO LA TABLA DE NUMEROS ALEATORIOS ANEXA, CON EL CRITERIO DE TOMAR TODOS LOS RENGLONES IMPARES ELIMINANDO EL ULTIMO DIGITO, LA MUESTRA FISICA SERIAN LOS TRANSISTORES CORRESPONDIENTES A LOS NUMEROS 0415, 0006, 0394, 0998, 0530, 0160, ETC.

TABLA DE NUMEROS ALEATORIOS

Renglón \ Columna	1	2	3	4	5	6	7	8	9	10	11
1	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649
2	18629	81953	05520	91962	04739	13092	37662	94822	94730	06496	35090
3	73115	47498	47498	87637	99016	00060	88824	71013	18735	20286	23153
4	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812
5	30405	03946	23792	14422	15059	45799	22716	19792	09983	74353	68668
6	16631	35006	85900	32388	52390	52390	16815	69298	38732	38480	73817
7	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052
8	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318
9	31624	76384	17403	03941	44167	64486	64758	75366	76554	01601	12614
10	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806

AGRUPAMIENTO DE DATOS

FRECUENCIA DE UN EVENTO.- ES EL NUMERO DE VECES QUE OCURRE EL EVENTO AL OBTENER UNA MUESTRA DE LA POBLACION CORRESPONDIENTE.

FRECUENCIA RELATIVA DE UN EVENTO.- ES EL COCIENTE DE SU FRECUENCIA ENTRE EL TOTAL DE ELEMENTOS (TAMAÑO) DE LA MUESTRA.

FRECUENCIA RELATIVA ACUMULADA.- ES LA ACUMULACION (SUMA) DE LAS FRECUENCIAS RELATIVAS HASTA UN VALOR DADO, PARTIENDO DEL VALOR (O DEL INTERVALO) MAS PEQUEÑO, EN OTRAS PALABRAS, ES LA FRECUENCIA DE VALORES MENORES O IGUALES QUE UN VALOR DADO.

FRECUENCIA COMPLEMENTARIA.- ES LA FRECUENCIA DE VALORES MAYORES QUE UN VALOR DADO = NUMERO DE DATOS - FRECUENCIA ACUMULADA.

DISTRIBUCION DE FRECUENCIAS

CON OBJETO DE FACILITAR LA INTERPRETACION DE LOS DATOS QUE SE TIENEN EN UNA MUESTRA, ES CONVENIENTE AGRUPARLOS POR VALORES O POR INTERVALOS DE VALORES, FORMANDO ASI UNA TABLA DE DISTRIBUCION DE FRECUENCIAS.

PARA FACILITAR EL CALCULO DE LAS FRECUENCIAS ES UTIL ORDENAR LOS DATOS EN FORMA CRECIENTE O DECRECIENTE DE VALORES, FORMANDO ASI UNA *TABLA DE DATOS ORDENADOS*.

EJEMPLO

EN UNA ESCUELA SECUNDARIA SE LES APLICO A 30 PROFESORES UN EXAMEN SOBRE PEDAGOGIA. LAS CALIFICACIONES (DATOS) QUE SE OBTUVIERON FUERON (YA ESTAN ORDENADOS EN FORMA CRECIENTE)

$\underbrace{57,\ 59,}_{A}\ \underbrace{65,\ 67,\ 67,\ 67,\ 69,}_{B}\ \underbrace{72,\ 73,\ 73,\ 77,\ 78,\ 78,}_{C}$

$\underbrace{81,\ 81,\ 83,\ 83,\ 83,\ 84,\ 84,\ 87,\ 88,\ 89,\ 89,}_{D}\ \underbrace{91,\ 91,\ 93,}_{E}$

$\underbrace{95,\ 97,\ 99}_{E}$

AGRUPAMIENTO DE VALORES

CALIFICACION	FRECUENCIA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
57	1	1/30	1/30
59	1	1/30	2/30
65	1	1/30	3/30
67	3	3/30	6/30
69	1	1/30	7/30
72	1	1/30	8/30
73	2	2/30	10/30
77	1	1/30	11/30
78	2	2/30	13/30
81	2	2/30	15/30
83	3	3/30	18/30
84	2	2/30	20/30
87	1	1/30	21/30
88	1	1/30	22/30
89	2	2/30	24/30
91	2	2/30	26/30
93	1	1/30	27/30
95	1	1/30	28/30
97	1	1/30	29/30
99	1	1/30	30/30=1
	$\Sigma=30$	$\Sigma=30/30=1$	

¿CUAL ES LA FRECUENCIA RELATIVA DE VALORES MENORES O IGUALES QUE 93,: 27/30

AGRUPAMIENTO POR INTERVALOS

LIMITES DE CLASES: SON LOS VALORES MINIMO Y MAXIMO DE CADA INTERVALO

MARCAS DE CLASE: SON LOS VALORES MEDIOS DE CADA INTERVALO DE CLASE

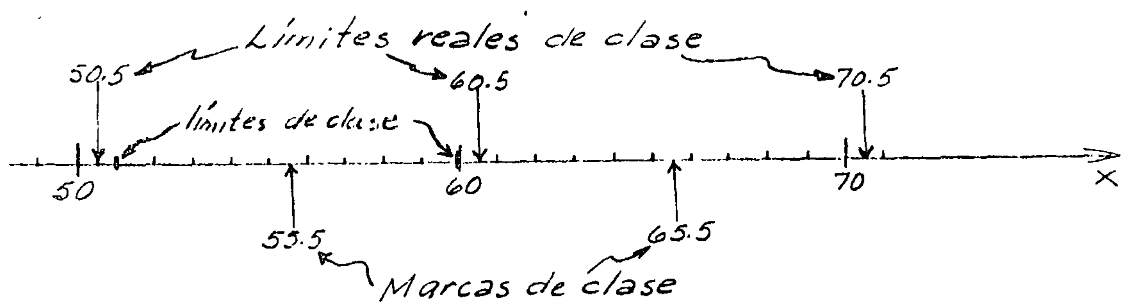
LIMITES REALES DE CLASE: SON LOS VALORES MINIMO Y MAXIMO QUE SON FRONTERA ENTRE LOS INTERVALOS. ESTOS DEBEN TENER UNA CIFRA DECIMAL MAS QUE LOS DATOS.

EVENTO (INTERVALO DE CALIFICACIONES)	ELEMENTOS OBSERVADOS	FRE-CUENCIA	FRECUENCIA RELATIVA
A = {51-60}	57,59	2	2/30
B = {61-70}	65,67,67,67,69	5	5/30
C = {71-80}	72,73,73,77,78,78	6	6/30
D = {81-90}	81,81,83,83,83,84, 84,87,88,89,89	11	11/30
E = {91-100}	91,91,93,95,97,99	6	6/30
		Σ=30	30/30=1

LIMITES INFERIORES DE CLASE | LIMITES SUPERIORES DE CLASE

EVENTO	LIMITES DE CLASE		LIMITES REALES DE CLASE		MARCAS DE CLASE
	INFERIOR	SUPERIOR	INFERIOR	SUPERIOR	
A	51	60	50.5	60.5	55.5
B	61	70	60.5	70.5	65.5
C	71	80	70.5	80.5	75.5
D	81	90	80.5	90.5	85.5
E	91	100	90.5	100.5	95.5

Evento	Elementos corresp. a los intervalos	Frecuencia	Frecuencia relativa	Frecuencia acumulada	Frecuencia relativa acumulada
A: 51-60	59,57	2	2/30=0.067(6.7%)	2	0.067
B: 61-70	67,65,69,67,67	5	5/30=0.166(16.6%)	2+5=7	0.067+0.166=0.233
C: 71-80	72,73,73,77,78,78	6	6/30=0.200(20%)	7+6=13	0.233+0.200=0.433
D: 81-90	83,88,84,89,83,84, 89,87,81,83,81	11	11/30=0.367(36.7%)	13+11=24	0.433+0.367=0.800
E: 91-100	99,91,97,95,91,93	6	6/30=0.200(20%)	24+6=30	0.800+0.200=1.000
		30	1.000		



A = {X: 50.5 < X $\leq$ 60.5}

B = {X: 60.5 < X $\leq$ 70.5}

C = {X: 70.5 < X $\leq$ 80.5}

D = {X: 80.5 < X $\leq$ 90.5}

E = {X: 90.5 < X $\leq$ 100.5}

LIMITES REALES INFERIORES DE CLASE

LIMITES REALES SUPERIORES DE CLASE

A MAYOR NUMERO DE DATOS SE REQUIERE MAYOR NUMERO DE INTERVALOS, PERO SE RECOMIENDA QUE ESTE NUMERO ESTE ENTRE 5 Y 20, SUPONIENDO QUE EN PROMEDIO CAIGAN 5 O MAS ELEMENTOS EN CADA INTERVALO. ASI, SI SE TIENEN 30 DATOS, SE RECOMIENDA USAR 30/5=6 INTERVALOS.

EL PROCESO DE AGRUPAMIENTO SE INDICARA AL MISMO TIEMPO QUE SE REALIZA EL SIGUIENTE EJEMPLO.

EJEMPLO

EN UN ESTUDIO ANTROPOLOGICO SE OBTUVO UNA MUESTRA DE 30 ESTATURAS

DE LOS VARONES ADULTOS RESIDENTES EN UNA REGION. LOS DATOS, ORDENADOS EN FORMA CRECIENTE DE VALORES, FUERON LOS SIGUIENTES: 159,161,163,163,163,167,167,167,167,168,168,168,169,169,170, 171,171,173,174,175,175,175,178,179,181,181,183,184,187,191 CM. OBTENER LA TABLA DE DISTRIBUCION DE FRECUENCIAS.

SOLUCION:

1. DETERMINACION DEL RANGO DE LA MUESTRA

   RANGO = VALOR MAXIMO - VALOR MINIMO = 191-159=32 CM

2. DETERMINACION DEL NUMERO DE INTERVALOS

   NUMERO DE INTERVALOS = $\frac{30}{5} = 6$

3. DETERMINACION DE LOS LIMITES DE CLASE

   ANCHO DE LOS INTERVALOS = $\frac{\text{RANGO}}{\text{NUMERO}} = \frac{32}{6} = 5.3$

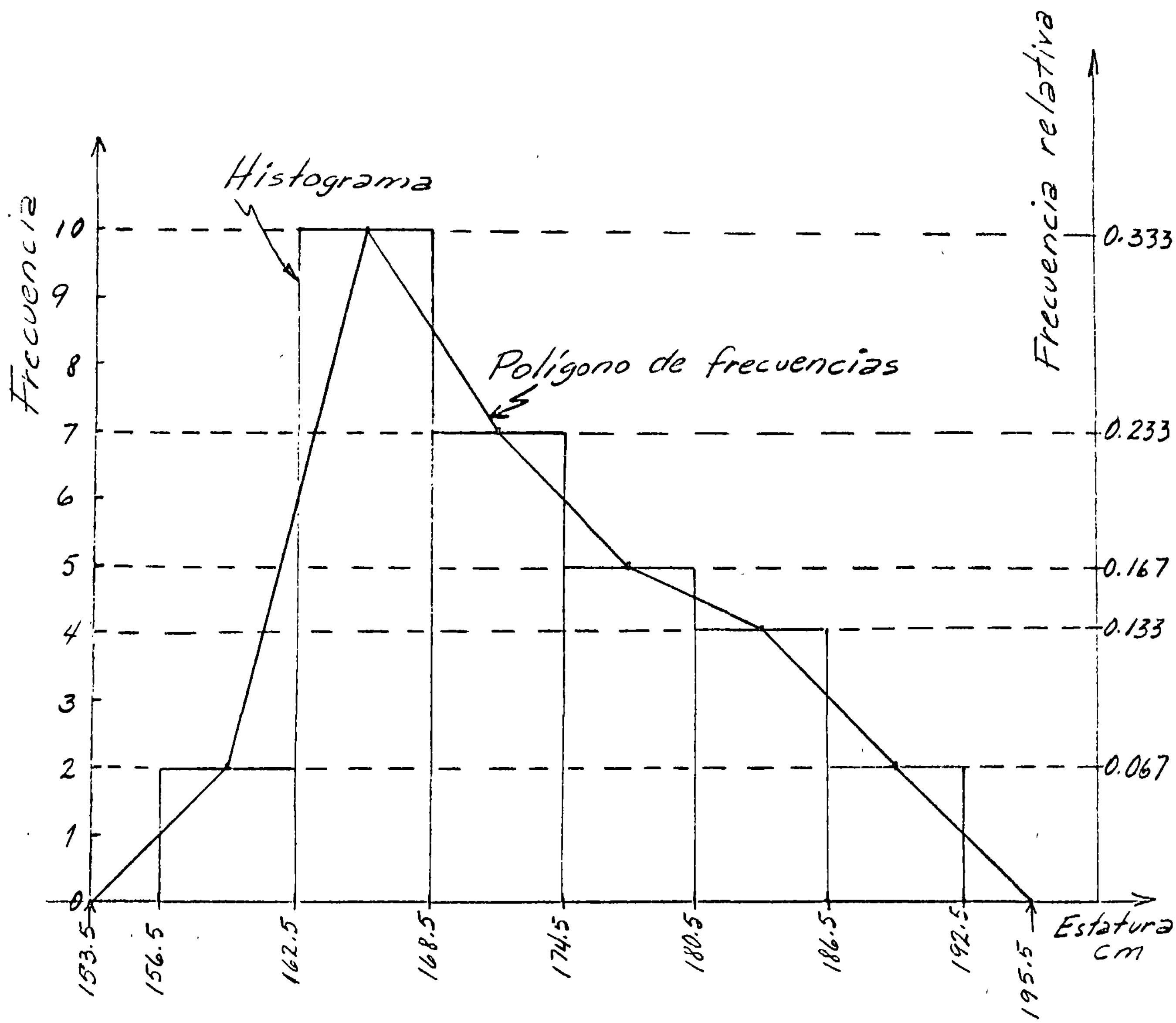
   TOMAREMOS UN ANCHO DE 6 CM, CON LO CUAL EL RANGO DEL AGRUPAMIENTO ES 6 x 6 =36 CM. LA DIFERENCIA DE RANGOS ES 36-32=4, QUE SE REPARTE EN LOS DOS INTERVALOS EQUITATIVAMENTE. POR LO TANTO, LOS INTERVALOS RESULTAN SER:

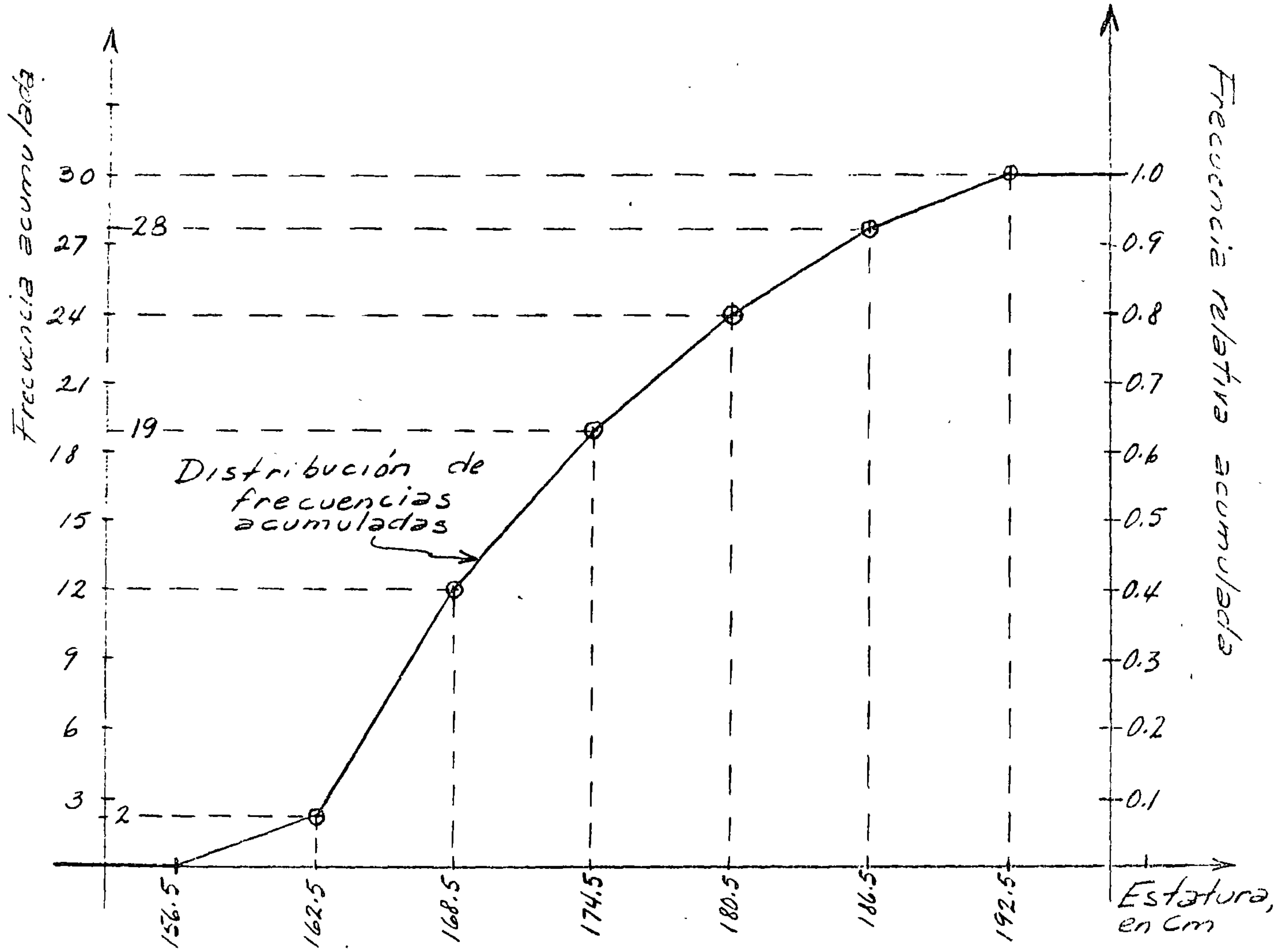
   157-162, 163-168, 169-174, 175-180, 181-186, 187-192

4. INTEGRACION DE LA TABLA:

INTERVALO	LIMITES REALES		FREC.	FREC. REL.	FREC. ACUM.	FREC. REL. ACUM.
	INF.	SUP.				
157-162	156.5	162.5	2	$\frac{2}{30} = 0.067$	2	0.067
163-168	162.5	168.5	10	$\frac{10}{30} = 0.333$	12	0.400
169-174	168.5	174.5	7	$\frac{7}{30} = 0.233$	19	0.633
175-180	174.5	180.5	5	$\frac{5}{30} = 0.167$	24	0.800
181-186	180.5	186.5	4	$\frac{4}{30} = 0.133$	28	0.933
187-192	186.5	192.5	2	$\frac{2}{30} = 0.067$	30	1.000
			$\Sigma = 30$	$\Sigma = 1.000$		

PRESENTACION GRAFICA DE LAS DISTRIBUCIONES DE FRECUENCIAS

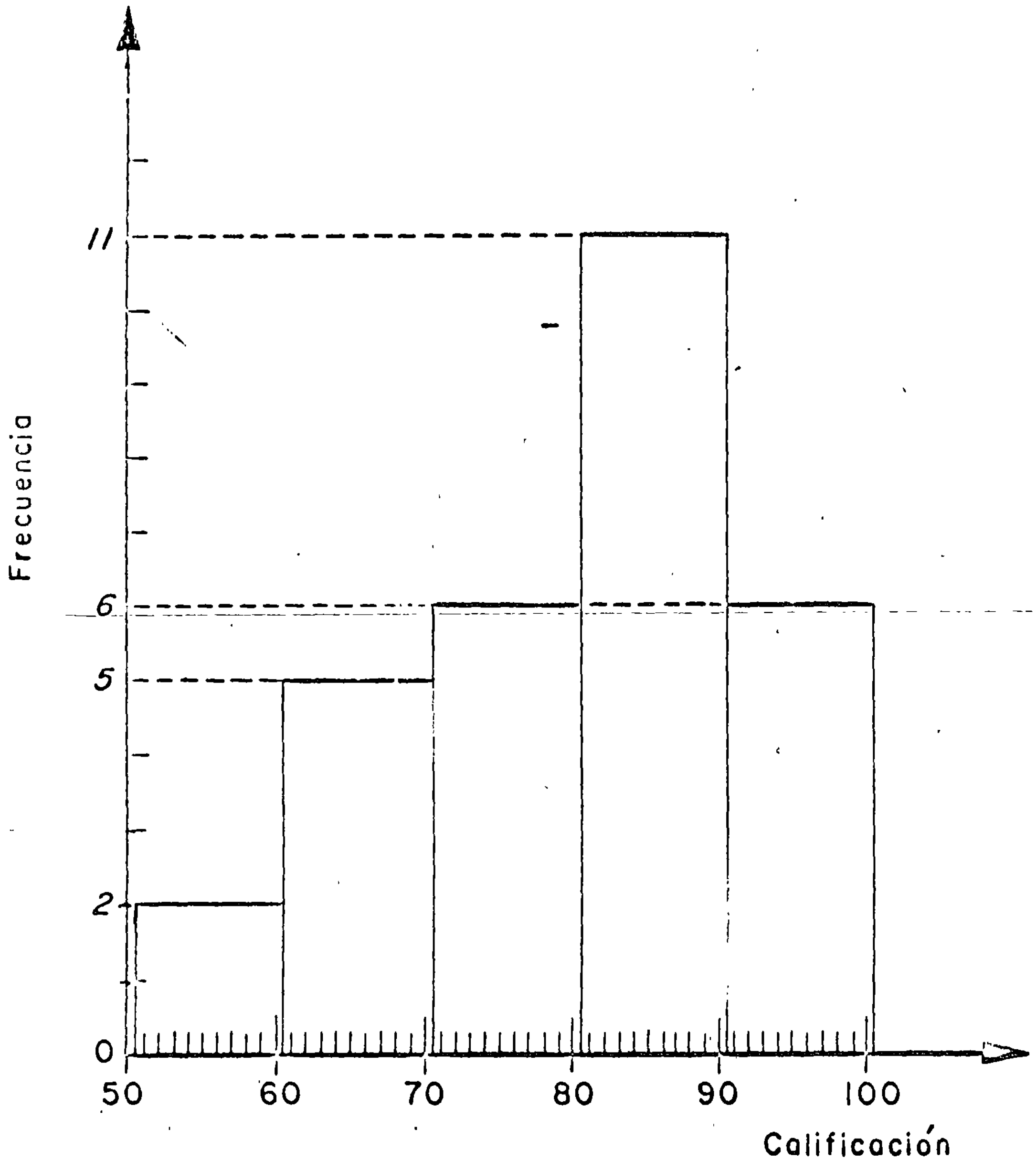




¿CUAL ES LA FRECUENCIA DE VALORES MAYORES QUE 180.5?: 30-24=6

LA FRECUENCIA RELATIVA ACUMULADA COMPLEMENTARIA ES: 1-0.800=0.200 (20%)

HISTOGRAMA DEL PROBLEMA DE LAS CALIFICACIONES EN PEDAGOGIA

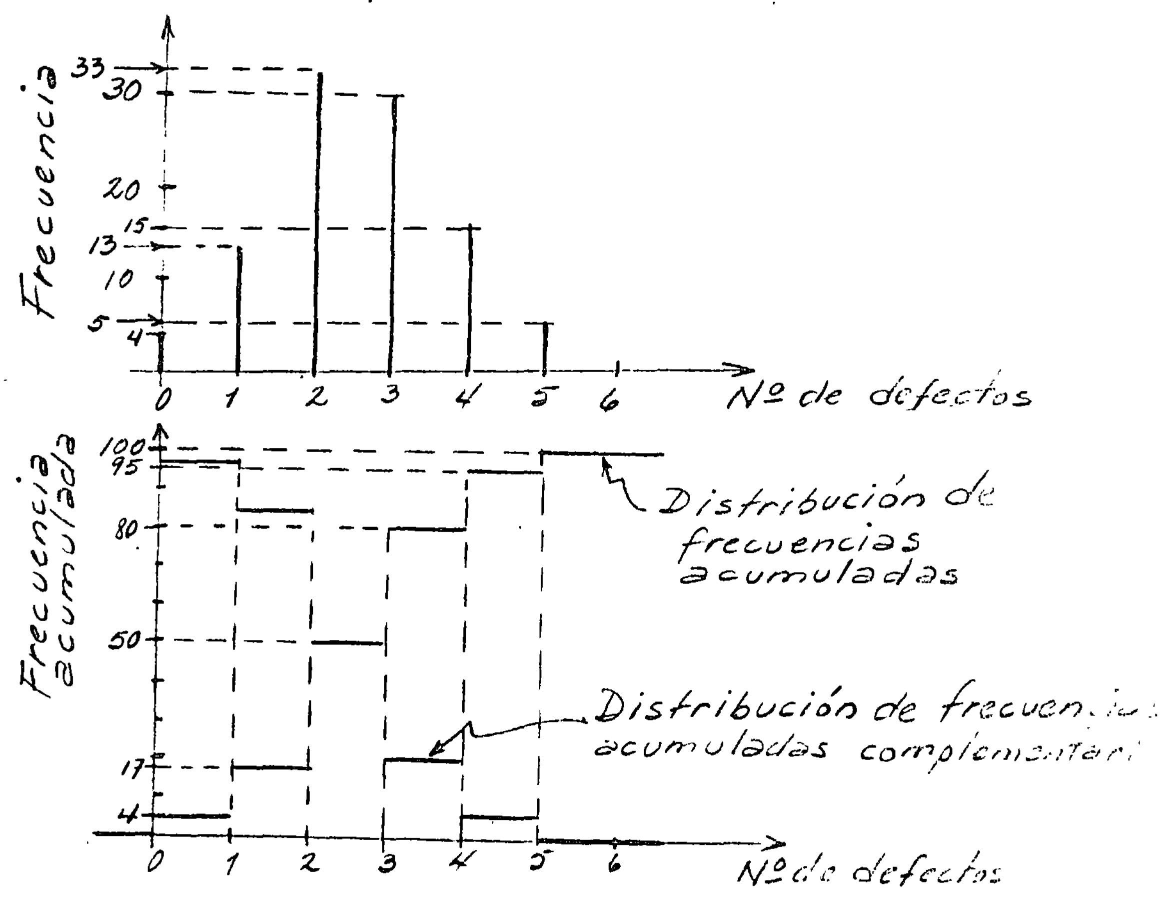


TAREA: DIBUJAR EL POLIGONO DE FRECUENCIAS Y LAS CURVAS DE FRECUENCIAS RELATIVAS ACUMULADAS Y COMPLEMENTARIAS.

EJEMPLO

EN UN ESTUDIO SOBRE LA CALIDAD DE LOS MONOBLOCKS PRODUCIDOS POR UNA FABRICA, SE OBTUVO UNA MUESTRA ALEATORIA DE 100 ELEMENTOS, A LOS CUALES SE LES CONTO EL NUMERO DE DEFECTOS DE FABRICACION. LA DISTRIBUCION DE FRECUENCIAS QUE SE OBTUVO ES LA SIGUIENTE:

NUMERO DE DEFECTOS	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA ACUMULADA COMPLEMENTARIA
0	4	4	96 (100-4)
1	13	17	83 (100-17)
2	33	50	50 (100-50)
3	30	80	20 (100-80)
4	15	95	5 (100-95)
5	5	100	0 (100-100)
	100		

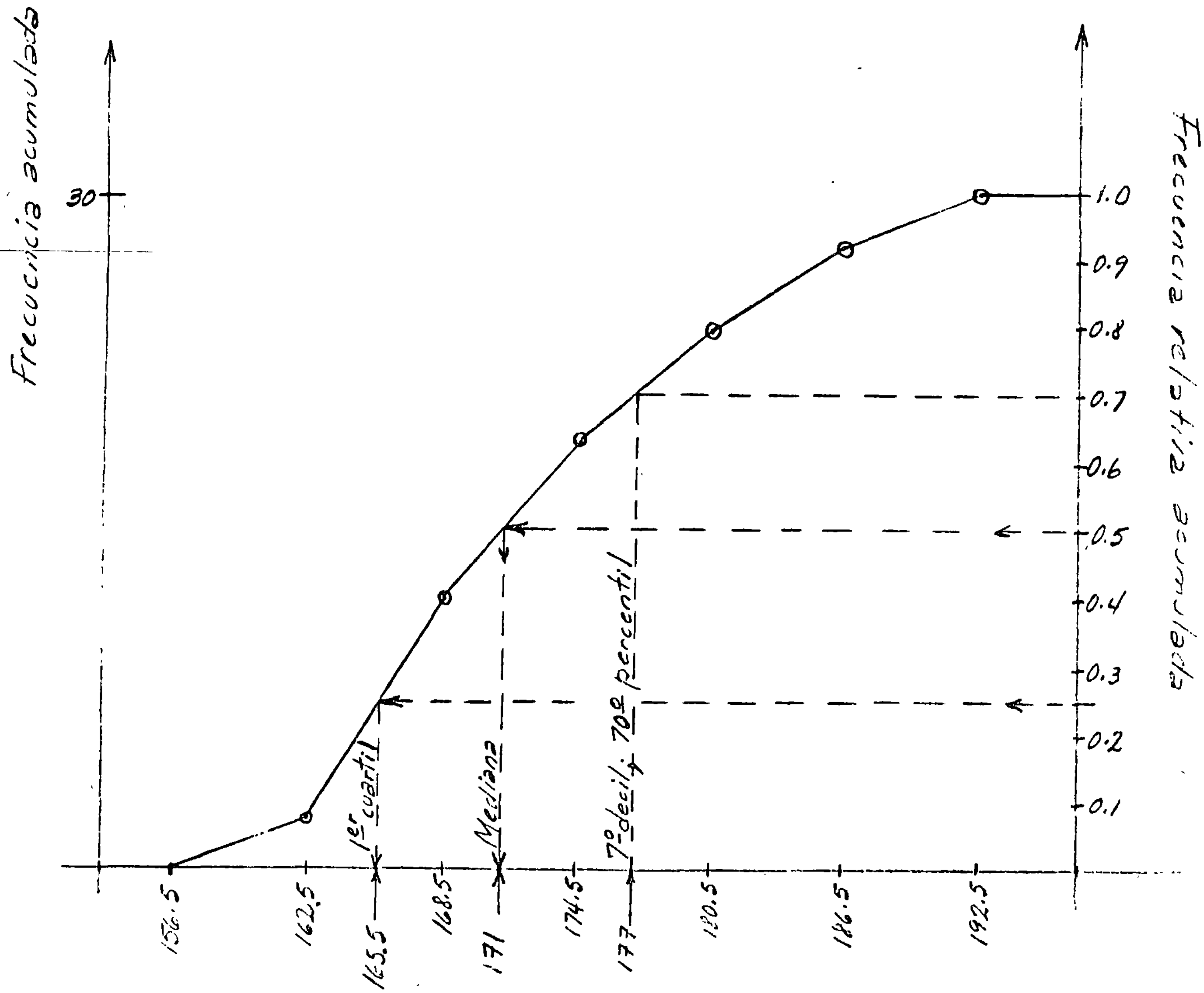


PERCENTILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 1 POR CIENTO.

DECILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 10 POR CIENTO.

CUARTILES:- SON LOS VALORES DE LA VARIABLE CORRESPONDIENTES A FRECUENCIAS RELATIVAS ACUMULADAS QUE SON MULTIPLOS DE 25 POR CIENTO.

MEDIANA:- VALOR DE LA VARIABLE CORRESPONDIENTE A LA FRECUENCIA RELATIVA ACUMULADA DE 50%.



MEDIDAS REPRESENTATIVAS DE LOS DATOS

MEDIDAS DE TENDENCIA CENTRAL

VALOR MEDIO O PROMEDIO ARITMETICO

(PARA LOS DATOS NO AGRUPADOS:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

DONDE x_i SON LOS VALORES DE LOS DATOS Y n ES EL TAMAÑO DE LA MUESTRA.

SI LOS DATOS ESTAN AGRUPADOS Y f_j ES LA FRECUENCIA DEL j-ESIMO INTERVALO Y x_j ES LA MARCA DE CLASE CORRESPONDIENTE, ENTONCES

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{K} f_j x_j \quad ; \quad K = \text{NUMERO DE INTERVALOS}$$

EJEMPLO

SEA EL EJEMPLO ENUNCIADO ANTERIORMENTE DE LOS DEFECTOS EN MONOBLOCKS. SE TENIA:

j	No. DE DEFECTOS x	FRECUENCIA f	fx
1	0	4	4 x 0 = 0
2	1	13	13 x 1 = 13
3	2	33	33 x 2 = 66
4	3	30	30 x 3 = 90
5	4	15	15 x 4 = 60
K=6	5	5	5 x 5 = 25
		Σ=100	$\sum_{j=1}^{6}$ 254

$$\bar{x} = \frac{254}{100}$$

$\bar{x}$ = 2.54 DEFECTOS POR MONOBLOCK

MODO.- ES EL VALOR DE LA VARIABLE QUE APARECE CON MAYOR FRECUENCIA EN UNA MUESTRA. SI LOS DATOS ESTAN AGRUPADOS, EL MODO ES LA MARCA DE CLASE DEL INTERVALO QUE TIENE LA MAYOR FRECUENCIA.

EJEMPLO

EN EL PROBLEMA DE LOS MONOBLOCKS EL MODO ES 2. EN EL PROBLEMA DE LAS ESTATURAS DE LOS VARONES ADULTOS DE UNA CIUDAD EL MODO ES 165.5 CM.

MEDIANA: ES EL VALOR DE LA VARIABLE QUE CORRESPONDE AL 50% DE LA FRECUENCIA RELATIVA ACUMULADA.

SI LOS DATOS ESTAN AGRUPADOS POR INTERVALOS, LA MEDIANA SE PUEDE CALCULAR CON LA FORMULA (ADEMAS DE GRAFICAMENTE, COMO YA SE VIO):

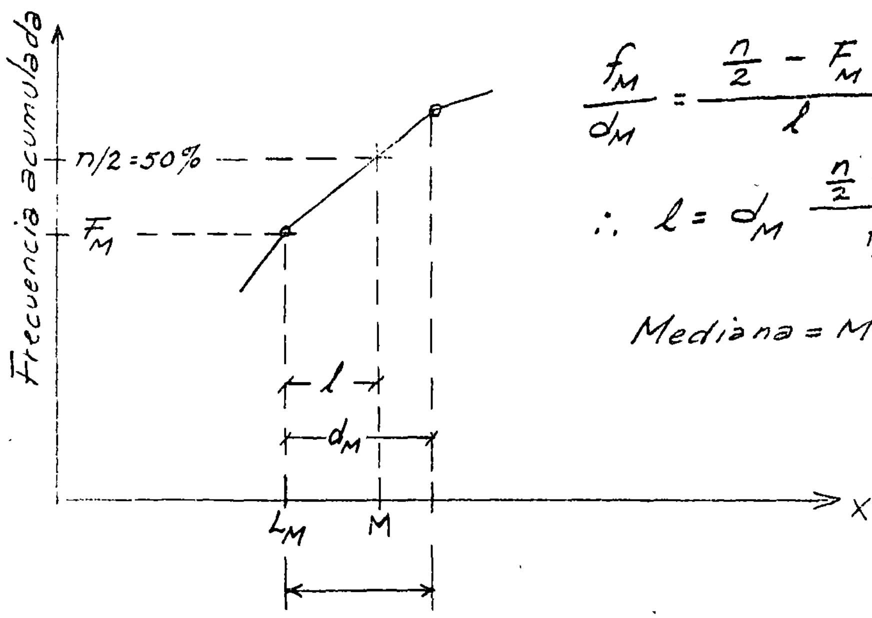
$$\text{MEDIANA} = M = L_M + \frac{\frac{n}{2} - F_M}{f_M} d_M$$

DONDE L_M = LIMITE INFERIOR REAL DEL INTERVALO QUE CONTIENE A LA MEDIANA

f_M Y d_M = RESPECTIVAMENTE, A LA FRECUENCIA Y ANCHO DEL INTERVALO QUE CONTIENE A LA MEDIANA

F_M = FRECUENCIA ACUMULADA HASTA EL INTERVALO QUE CONTIENE A LA MEDIANA EXCLUSIVE

n = TAMAÑO DE LA MUESTRA



Intervalo que contiene a la mediana

$$\frac{f_M}{d_M} = \frac{\frac{n}{2} - F_M}{l}$$

$$\therefore\ l = d_M \frac{\frac{n}{2} - F_M}{f_M}$$

$$\text{Mediana} = M = L_M + l$$

EJEMPLO

EN UN ESTUDIO PARA DETERMINAR LOS TIEMPOS EN QUE UNA MUESTRA ALEATORIA DE INDIVIDUOS REACCIONABA A CIERTOS ESTIMULOS PSICOLOGICOS SE OBTUVO LO SIGUIENTE:

j	MARCA DE CLASE x, EN SEG	LIMITES REALES	FRECUENCIA f	FRECUENCIA ACUMULADA, F	fx, SEG
1	0.10	0.075-0.125	2	2	0.20
2	0.15	0.125-0.175	7	9	1.05
3	0.20	0.175-0.225	14	23	2.80
4	0.25	0.225-0.275	4	27	1.00
K=5	0.30	0.275-0.325	3	30	0.90
			$\Sigma=30$		$\sum_{j=1}^{5} f_i x_i = 5.95$

$$\bar{x} = \frac{5.95}{30} = 0.198 \text{ SEG}$$

MODO = 0.20 SEG

$$d_M = 0.05,\ L_M = 0.20 - \frac{0.05}{2} = 0.175,\ F_M = 9$$

$$n/2 = 30/2 = 15,\ f_M = 14$$

$$\text{MEDIANA} = M = 0.175 + \frac{15 - 9}{14}\, 0.05$$

$$M = 0.175 + \frac{0.30}{14} = 0.175 + 0.021 = 0.196 \text{ SEG}$$

MEDIDAS DE DISPERSION

RANGO = MAXIMO VALOR OBSERVADO - MINIMO VALOR OBSERVADO

VARIANCIA : SI LOS DATOS *NO* ESTAN AGRUPADOS;

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

SI LOS DATOS ESTAN AGRUPADOS:

$$S_X^2 = \frac{1}{n} \sum_{j=1}^{K} f_j (x_j - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{k} f_i \, x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

DONDE LAS x_j SON LOS VALORES DE LAS MARCAS DE CLASE DE LOS INTERVALOS.

DESVIACION ESTANDAR

$$S_X = \sqrt{S_X^2}$$

COEFICIENTE DE VARIACION

$$v_X = S_X / \bar{x}$$

EJEMPLO

EN UN ESTUDIO SOBRE LA TEMPERATURA MAXIMA DIARIA EN UNA CIUDAD SE OBTUVO LO SIGUIENTE DURANTE UNA PRIMAVERA:

j	INTERVALOS DE TEMPERATURA, °F	MARCA DE CLASE, °F	FRECUENCIA f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
1	55 - 63	59	2	118	-21.3	453.7	907.4
2	64 - 72	68	6	408	-12.3	151.3	907.8
3	73 - 81	77	7	539	- 3.3	10.9	76.3
4	82 - 90	86	9	774	5.7	32.5	292.5
5	91 - 99	95	6	570	14.7	216.1	1296.6
			30	2409			3480.6

$$\bar{x} = \frac{2409}{30} = 80.3 \text{ °F}$$

$$S_X^2 = \frac{3480.6}{30} = 116 \text{ °F}^2$$

$$S_X = \sqrt{116} = 10.8 \text{ °F}$$

$$v_X = \frac{10.8}{80.3} = 0.134 \quad (13.4\%)$$

MODO = 86

$d_M=9$, $L_M=72.5$, $f_M=7$, $F_M=8$, $\frac{n}{2} = \frac{30}{2} = 15$

$$\text{MEDIANA} = M = 72.5 + \frac{15-8}{7}\, 9 = 72.5 + 9 = 81.5 \text{ °F}$$

T A R E A

0.78	0.38	0.72	0.65	0.72	0.92	0.78	0.65	0.92	0.78
1.36	1.13	0.65	0.48	0.83	0.48	0.72	0.18	0.65	0.78
0.65	1.00	0.78	0.78	1.03	1.26	0.48	0.18	1.06	0.96
0.65	0.92	0.72	0.78	0.78	0.48	0.28	0.36	0.83	0.48
0.78	0.49	0.36	0.78	0.78	0.83	0.88	0.96	1.03	1.21
0.88	0.57	0.72	1.03	0.92	0.96	0.78	1.09	0.92	1.12
0.65	0.65	0.83	0.72	0.72	0.78	0.72	1.09	0.83	0.83
0.83	1.06	0.57	0.78	1.23	1.09	1.03	0.18	0.65	1.34
0.96	0.65	0.48	1.18	1.12	0.18	0.48	0.72	0.57	0.55
0.96	0.65	0.96	0.51	0.65	1.21	1.48	0.96	0.96	1.40

1. Agrupar datos por intervalos y elaborar tabla con frecuencias, frecuencias relativas, frecuencias acumuladas y frecuencias relativas acumuladas (anotar límites, límites reales y marcas de clase).

2. Dibujar: a) Histograma

   b) Polígono de frecuencias

   c) Curva de frecuencias acumuladas

3. Calcular todas las medidas de tendencia central y de dispersión que se han estudiado

   a. Sin agrupar datos

   b. Con datos agrupados

SIMBOLOS DE DESIGUALDADES:

- $<$ menor que
- $\leq$ menor o igual que
- $>$ mayor que
- $\geq$ mayor o igual que
- $\neq$ diferente de

TEORIA DE CONJUNTOS

UN *CONJUNTO* ES UNA COLECCION BIEN DEFINIDA DE OBJETOS.

NOTACION: LOS CONJUNTOS SE DENOTAN USUALMENTE CON LETRAS MAYUSCULAS, Y SUS ELEMENTOS SE ANOTAN DENTRO DE UN PAR DE LLAVES.

EJEMPLOS

A) EL CONJUNTO DE NUMEROS ANOTADOS EN UN DADO ES

$S = \{1, 2, 3, 4, 5, 6\}$

B) EL CONJUNTO DE LOS NUMEROS ENTEROS MENORES QUE 5 ES

$S = \{-\infty, \ldots, -3, -2, -1, 0, 1, 2, 3, 4\}$

o $S = \{x:$ x ES ENTERO Y $x \leq 4\}$

C) EL CONJUNTO DE LOS NUMEROS ENTEROS POSITIVOS MENORES QUE 5 ES

$E = \{0, 1, 2, 3, 4\}$

$E = \{x:$ ES ENTERO Y $0 \leq x \leq 4\}$

D) EL CONJUNTO DE LOS CONTINENTES ES

C = {ASIA, EUROPA, AMERICA, AFRICA, OCEANIA}

E) EL CONJUNTO DE MARCAS QUE TIENE UNA MONEDA ES

M = {CARA, CRUZ}

F) EL CONJUNTO DE NUMEROS MAYORES DE 5 PERO MENORES O IGUALES QUE 10

$S_1 = \{x: 5 < x \leq 10\}$

CONJUNTOS
- *FINITOS*.- CUANDO TIENEN UN NUMERO FINITO DE ELEMENTOS
- *INFINITOS*.- CUANDO TIENEN UN NUMERO INFINITO DE ELEMENTOS

SUBCONJUNTOS

PARA EXPRESAR QUE UN ELEMENTO *PERTENECE* A UN CONJUNTO SE USA EL SIMBOLO ε. PARA EXPRESAR QUE *NO PERTENECE* SE USA EL SIMBOLO $\not\varepsilon$.

EJEMPLO

SI $S_1 = \{X: 5<X\leq 10\}$, ENTONCES.

$3 \not\varepsilon S_1$; $5 \not\varepsilon S_1$; $8 \varepsilon S_1$; $10 \varepsilon S_1$

PARA EXPRESAR QUE UN CONJUNTO *ESTA CONTENIDO* EN OTRO SE USA EL SIMBOLO C; SI *NO ESTA CONTENIDO* SE USA EL SIMBOLO $\not\subset$.

PARA QUE UN CONJUNTO ESTE CONTENIDO EN OTRO SE REQUIERE QUE *TODOS* SUS ELEMENTOS LO ESTEN, ES DECIR, QUE TODOS SUS ELEMENTOS PERTENEZCAN A AMBOS CONJUNTOS.

EJEMPLO

SEAN E={3,5}; F={3,8}; G={7,9}. $E\not\subset S_1$; $F\not\subset S_1$; $G C S_1$

SI UN CONJUNTO, B, ESTA CONTENIDO EN OTRO, S, SE DICE QUE B ES *SUBCONJUNTO* DE S.

EJEMPLO

$B = \{X: 3\leq X\leq 8\}$ Y $S_1=\{X: 5<X\leq 10\}$

EN ESTE CASO:

$G C S_1 \Rightarrow G$ ES SUBCONJUNTO DE S_1

$B\not\subset S_1 \Rightarrow B$ NO ES SUBCONJUNTO DE S_1

SE DICE QUE DOS CONJUNTOS SON IGUALES CUANDO CONTIENEN LOS MISMOS ELEMENTOS (NO IMPORTA EL ORDEN EN QUE ESTOS SE ESCRIBAN)

EJEMPLO

SEAN $A=\{1,3,5,7\}$, $B=\{7,5,1,3\}$ Y $C=\{7,5,1\}$

EN TAL CASO, $A = B \neq C$

CONJUNTO VACIO

DE LA MISMA MANERA QUE EXISTE EL CERO EN LOS NUMEROS, EN LA TEORIA DE CONJUNTOS EXISTE EL *CONJUNTO VACIO*, EL CUAL *NO* TIENE ELEMENTOS. USUALMENTE SE DENOTA $\emptyset$.

EJEMPLO

¿CUAL ES EL CONJUNTO DE ELEMENTOS, X, TALES QUE $2X=7$ Y X ES ENTERO?

SOLUCION .- ES EL CONJUNTO VACIO, $\emptyset$.

A $\emptyset$ SE LE CONSIDERA COMO SUBCONJUNTO DE CUALQUIER CONJUNTO. ASI, POR EJEM, TODOS LOS SUBCONJUNTOS DEL CONJUNTO

$S =\{2,5,10\}$ SON: $\{2\};\{5\};\{10\};\{2,5\};\{2,10\};\{5,10\};\{2,5,10\}$ Y $\emptyset$.

ESPACIO DE EVENTOS

ASOCIADO A UN EXPERIMENTO SIEMPRE HAY UN CONJUNTO DE RESULTADOS POSIBLES; A DICHO CONJUNTO SE LE LLAMA *ESPACIO DE EVENTOS*.

EJEMPLOS

EL ESPACIO DE EVENTOS ASOCIADO AL EXPERIMENTO DE LANZAR UN DADO Y ANOTAR LA CARA QUE QUEDA HACIA ARRIBA ES

$S =\{1,2,3,4,5,6\}$

EL ESPACIO DE EVENTOS CORRESPONDIENTE AL EXPERIMENTO DE LANZAR DOS DADOS Y ANOTAR LOS NUMEROS QUE QUEDAN HACIA ARRIBA ES

$$S=\left\{\begin{array}{l}(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\\(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\\(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\\(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\\(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\end{array}\right\}$$

SI EN ESTE EXPERIMENTO LA OBSERVACION DE INTERES FUESE LA SUMA DE LOS DOS NUMEROS OBSERVADOS, ENTONCES EL ESPACIO DE EVENTOS DEL EXPERIMENTO SERIA

$$S=\{2,3,4,5,6,7,8,9,10,11,12\}$$

A TODO SUBCONJUNTO DE UN ESPACIO DE EVENTOS SE LE LLAMA *EVENTO*. A LOS EVENTOS QUE TIENEN UN SOLO ELEMENTO DEL ESPACIO SE LES LLAMA *EVENTOS SIMPLES*.

SI AL REALIZAR UN EXPERIMENTO SE OBSERVA UN ELEMENTO DEL EVENTO A, ENTONCES SE DICE QUE *OCURRIO* O *SE VERIFICO* EL EVENTO A. POR EJEMPLO, SI $A=\{2,4\}$ Y AL LANZAR UN DADO SE OBSERVA EL 2 O 4, SE DICE QUE OCURRIO EL EVENTO A; SI SE OBSERVA CUALQUIER OTRO NUMERO, ENTONCES SE DICE QUE *NO* OCURRIO A.

ESPACIOS DE EVENTOS

- *DISCRETOS*.- SI SUS ELEMENTOS PUEDEN NUMERARSE O CONTARSE. TIENEN UN NUMERO FINITO O INFINITO NUMERABLE DE ELEMENTOS.
- *CONTINUOS*.- SI SUS ELEMENTOS *NO* PUEDEN ENUMERARSE. TIENEN UN NUMERO INFINITO NO NUMERABLE DE ELEMENTOS

EJEMPLO

LOS ESPACIOS DE EVENTOS S_1={CARA, CRUZ}; S_2={1,2,3,4,5,6,}; S_3={VERDE, ROJO} SON DISCRETOS. LOS ESPACIOS DE EVENTOS $S_4=\{X: -\infty<X\leq 0\}$; $S_5\{Z: Z\geq 3\}$; $S_6=\{Y: 3\leq Y\leq 80\}$ SON CONTINUOS.

EJEMPLO

¿QUE TIPOS DE ESPACIOS DE EVENTOS CORRESPONDEN A LOS SIGUIENTES EXPERIMENTOS?

A) CONTEO DEL NUMERO DE GRANOS DE UNA MAZORCA DE MAIZ

$S=\{0,1,2,3,\ldots,\infty\}$, ES DISCRETO E INFINITO

B) MEDICION DE LA LONGITUD DE UNA ESPIGA DE TRIGO

$S=\{X: 0<X<\infty\}$, X EN CM, ES CONTINUO E INFINITO

C) MEDICION DEL EFECTO DE UNA VACUNA, EN TERMINOS DE "EXITO" O "FRACASO"

S={EXITO, FRACASO} ES DISCRETO Y FINITO.

D) MEDICION DEL NUMERO DE MILIGRAMOS DE UN ANTIBIOTICO CONTENIDO EN UNA CAPSULA

$S=\{Y: 0\leq Y<\infty\}$ Y en mg, ES CONTINUO E INFINITO.

COMPLEMENTO DE UN EVENTO

EL *COMPLEMENTO* DE UN EVENTO A ES OTRO EVENTO QUE CONTIENE TODOS LOS ELEMENTOS DEL ESPACIO DE EVENTOS CORRESPONDIENTE QUE NO ESTAN EN A. USUALMENTE SE DENOTA CON UNA TILDE SOBRE EL SIMBOLO QUE CORRESPONDE AL EVENTO QUE COMPLEMENTA, $\bar{A}$.

EJEMPLOS

SI S={1,2,3,4,5,6} Y A={1,3,5} ENTONCES $\bar{A}$={2,4,6}.

SI $S=\{X: 0\leq X\leq 58\}$ Y $A=\{X: 3<X\leq 17\}$, ENTONCES $\bar{A}=\{X: 0\leq X\leq 3,\ 17<X\leq 58\}$

EVENTOS MUTUAMENTE EXCLUSIVOS

CUANDO DOS O MAS EVENTOS NO PUEDEN OCURRIR SIMULTANEAMENTE AL REALIZAR UNA SOLA VEZ UN EXPERIMENTO, SE DICE QUE ESTOS SON *MUTUAMENTE EXCLUSIVOS*, ES DECIR, DOS EVENTOS SON MUTUAMENTE EXCLUSIVOS CUANDO NO TIENEN NI UN SOLO ELEMENTO EN COMUN.

EJEMPLO

A) CUALQUIER EVENTO Y SU COMPLEMENTO SON MUTUAMENTE EXCLUSIVOS.

B) ¿SON $E=\{Y: 0\leq Y\leq 25\}$ Y $A=\{2,50,100\}$ MUTUAMENTE EXCLUSIVOS?

NO, PORQUE TIENEN EL ELEMENTO 2 EN COMUN.

OPERACIONES CON EVENTOS

UNION

LA *UNION* DE DOS EVENTOS ES OTRO EVENTO CUYOS ELEMENTOS SON TODOS LOS DE AMBOS. LA OPERACION DE UNION SE DENOTA CON EL SIMBOLO $\cup$.

EJEMPLOS

A) SI $A=\{2,4,6\}$ Y $B=\{1,6,12\}$, ENTONCES

$G=A\cup B=\{1,4,6,12,2\}$

B) ¿SON A Y B MUTUAMENTE EXCLUSIVOS? NO PORQUE TIENEN EL 6 EN COMUN.

C) SI $D=\{Y: 0\leq Y\leq 13\}$ Y $E=\{Y: 20\leq Y\leq 50\}$,

ENTONCES

$D\cup E=\{Y: 0\leq Y\leq 13,\ 20\leq Y\leq 50\}$

D) SI $F=\{Y: 8\leq Y\leq 20\}$, ENTONCES

$D\cup F=\{Y; 0\leq Y\leq 20\}$.

E) SI $G=\{Y: 3\leq Y\leq 10\}$, ENTONCES

$D\cup G=\{Y: 0\leq Y\leq 13\}=D$; OBSERVESE QUE EN ESTE CASO $G\subset D$. EN GENERAL, SI $A\subset B$, ENTONCES $A\cup B=B$.

EN GENERAL, LA UNION DE VARIOS EVENTOS ES OTRO EVENTO CUYOS ELEMENTOS SON TODOS LOS DE LOS EVENTOS QUE SE UNEN.

EJEMPLO

$A \cup B \cup F = K = \{1,2,4,6,\ y: 8 \leq y \leq 20\}$

INTERSECCION

LA *INTERSECCION* DE DOS EVENTOS ES EL CONJUNTO DE ELEMENTOS QUE PERTENECEN SIMULTANEAMENTE A AMBOS. PARA DENOTAR LA OPERACION DE INTERSECCION SE USA EL SIMBOLO $\cap$.

EJEMPLOS

A) $A = \{2,3,6\}$ Y $B = \{2,6,10\}$ ENTONCES $A \cap B = C = \{2,6\}$

B) $D = \{Y: 4 \leq Y \leq 5\}$, ENTONCES $A \cap D = \emptyset$.

OBSERVESE QUE EN ESTE EJEMPLO A Y D SON MUTUAMENTE EXCLUSIVOS, YA QUE *NO* TIENEN NINGUN ELEMENTO EN COMUN. SIEMPRE QUE DOS EVENTOS SON MUTUAMENTE EXCLUSIVOS, SU INTERSECCION ES EL CONJUNTO VACIO.

EN GENERAL, LA *INTERSECCION* DE VARIOS EVENTOS ES EL CONJUNTO DE ELEMENTOS QUE TODOS ELLOS TIENEN EN COMUN.

EJEMPLO

SI $A = \{2,3,6,8\}$; $B=\{2,3,10,100\}$; $C=\{Y: 0 \leq Y \leq 5\}$ Y $D=\{Y: 2 \leq Y \leq 4\}$, ENTONCES

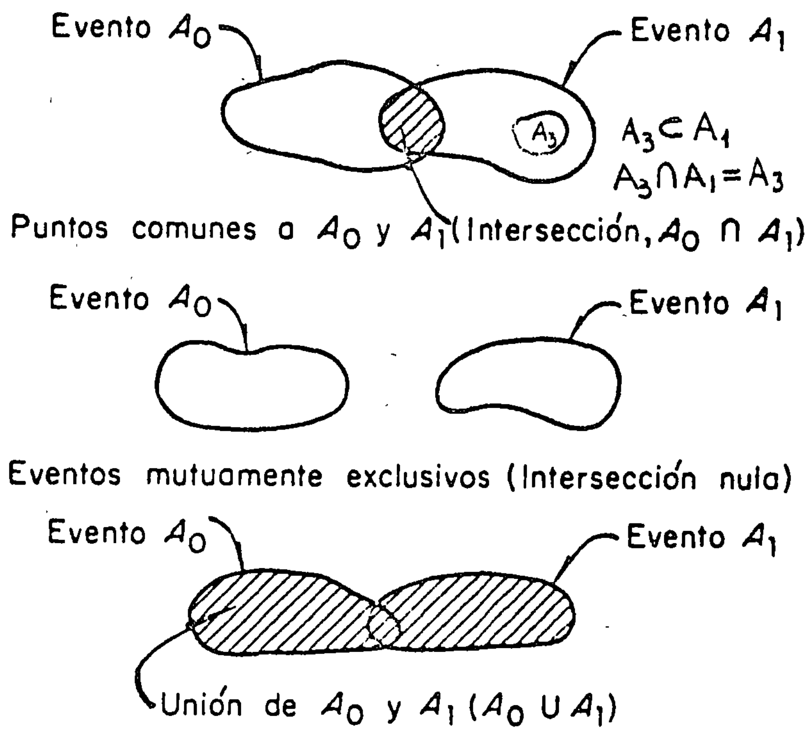
$A \cap B \cap C \cap D = E = \{2,3\}$

$A \cup B \cup C \cup D = F = \{Y: 0 \leq Y \leq 5,\ 6,8,10,100\}$

LA OCURRENCIA DE UN EVENTO "Y" OTRO IMPLICA LA OCURRENCIA DE AMBOS A LA VEZ, ES DECIR, QUE SE VERIFIQUE LA *INTERSECCION*. LA OCURRENCIA DE UN EVENTO "O" ALGUN OTRO, IMPLICA LA OCURRENCIA DE CUALQUIERA DE ELLOS, ES DECIR DE LA *UNION*.

DIAGRAMAS DE VENN

UNA MANERA DE ILUSTRAR GRAFICAMENTE LAS OPERACIONES CON CONJUNTOS ES MEDIANTE LOS DIAGRAMAS DE VENN. EN ESTOS, CADA CONJUNTO SE REPRESENTA POR UNA CURVA CERRADA QUE ENCIERRA LOS ELEMENTOS QUE LE CORRESPONDEN.



Diagramas de Venn (unión e intersección de eventos)

TEORIA DE PROBABILIDADES

AL LANZAR UNA MONEDA NO PODEMOS PREDECIR CON CERTEZA CUAL CARA QUEDARA HACIA ARRIBA. LO UNICO QUE SE PUEDE ASEGURAR, SI LA MONEDA NO ESTA CARGADA, ES QUE AMBAS CARAS TIENEN LA MISMA OPORTUNIDAD DE SALIR, ES DECIR, QUE LOS EVENTOS SIMPLES {CARA} Y {CRUZ} TIENEN LA MISMA *PROBABILIDAD* DE OCURRIR.

COMO YA SE DIJO, LA *PROBABILIDAD* DE QUE OCURRA UN EVENTO ES UNA MEDIDA DEL GRADO DE CONFIANZA QUE SE TIENE DE QUE ESTE OCURRA AL REALIZAR EL EXPERIMENTO CORRESPONDIENTE.

EXISTEN POR LO MENOS TRES MANERAS DE ASIGNARLE UNA PROBABILIDAD A UN EVENTO:

1. EN TERMINOS DE LOS RESULTADOS DE REPETIR VARIAS VECES UN EXPERIMENTO (METODO FRECUENCIAL).

2. APLICANDO LA DEFINICION CLASICA DE PROBABILIDADES.

3. CON BASE EN UN MODELO MATEMATICO (PROBABILISTICO) DEL FENOMENO DE QUE SE TRATE.

METODO FRECUENCIAL

SI N(A) ES EL NUMERO DE VECES QUE SE OBSERVA EL EVENTO A AL REALIZAR N VECES UN EXPERIMENTO, LA FRECUENCIA RELATIVA DE A, DEFINIDA COMO N(A)/N, SE CONSIDERA COMO ESTIMACION DE LA PROBABILIDAD DE A,

$$P(A) = \frac{N(A)}{N}$$

YA QUE, EN EL LIMITE, $P(A) = \lim_{N\to\infty} \frac{N(A)}{N}$

EJEMPLO

DE UNA URNA QUE CONTIENE BOLAS ROJAS, BLANCAS Y AZULES, SE SACO UNA BOLA, SE ANOTO SU COLOR Y SE REGRESO A LA URNA. SI ESTE EXPERIMENTO SE REPITE 20 VECES Y LOS RESULTADOS SON

b,b,a,r,r,r,a,b,r,a,b,b,a,r,b,r,r.a,r,a, DONDE

r = ROJA, b = BLANCA, a = AZUL

¿QUE PROBABILIDADES LE ASIGNARIA A LOS EVENTOS B={b}, A={a}, y R={r}, DE ACUERDO CON EL METODO FRECUENCIAL?

EN ESTA MUESTRA SE TIENE QUE N(B)=6, N(A)=6, N(R)=8, N=20

POR LO QUE $P(B)=\frac{6}{20} = \frac{3}{10}$; $P(A) = \frac{6}{20} = \frac{3}{10}$; $P(R) = \frac{8}{20} = \frac{4}{10}$

NOTESE QUE LOS EVENTOS B, A Y R SON MUTUAMENTE EXCLUSIVOS, YA QUE SON EVENTOS SIMPLES, Y QUE

$$P(B) + P(A) + P(R) = \frac{3}{10} + \frac{3}{10} + \frac{4}{10} = 1 = P(S)$$

EN DONDE $S = \{r,b,a\}$

DEFINICION CLASICA DE PROBABILIDADES

SI N(A) ES EL NUMERO DE MANERAS IGUALMENTE PROBABLES EN QUE PUEDE OCURRIR EL EVENTO A Y N ES EL NUMERO TOTAL DE ELEMENTOS DEL ESPACIO DE EVENTOS CORRESPONDIENTE, ENTONCES LA PROBABILIDAD DE A ES

$$P(A) = \frac{N(A)}{N}$$

EJEMPLOS

A) SI EN UNA URNA SE TIENEN 5 BOLAS BLANCAS Y 15 NEGRAS, Y SE VA A SELECCIONAR UNA AL AZAR, ¿CUAL ES LA PROBABILIDAD DE QUE SEA BLANCA(A={BLANCA})?:

$N = 5+15=20;\ N(A)=5 \Rightarrow P(A)=\frac{5}{20}=\frac{1}{4}$

B) SI SE LANZAN DOS DADOS, ¿CUAL ES LA PROBABILIDAD DE QUE

1. SALGA UN 2 Y UN 5 (EVENTO B)?
2. LA SUMA SEA 7 (EVENTO A)

PARA EL INCISO 1 EL ESPACIO DE EVENTOS ES:

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2.1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

SI EL DADO NO ESTA CARGADO, CADA PAREJA DE NUMEROS ES IGUALMENTE PROBABLE. EN TAL CASO, N=36 y N(B)=2 (APARECE (2,5) O (5,2))

$\Rightarrow P(B)=2/36=1/18.$

PARA EL INCISO 2 EL ESPACIO DE EVENTOS ES

$S_1=\{2,3,4,5,6,7,8,9,10,11,12\}$

PERO NO TODOS LOS ELEMENTOS (EVENTOS SIMPLES) SON IGUALMENTE PROBA-

BLES, YA QUE, POR EJEMPLO, EL 2 SOLO APARECERA SI SE OBSERVA LA PAREJA (1,1), EN CAMBIO EL 3 APARECERA SI OCURREN LAS PAREJAS (1,2) O (2.1), ES DECIR, EL 3 TIENE EL DOBLE DE PROBABILIDAD QUE EL 2. POR ESTO, PARA CALCULAR LA PROBABILIDAD DE QUE LA SUMA SEA 7 ES NECESARIO TRABAJAR CON EL ESPACIO S Y CONTAR LAS MANERAS POSIBLES DE QUE LA SUMA SEA 7, LO CUAL OCURRE SI SE OBSERVA CUALQUIERA DE LAS PAREJAS (6,1), (5,2), (4,3), (3,4), (2,5) o (1,6), ES DECIR, HAY 6 MANERAS IGUALMENTE PROBABLES DE QUE OCURRA EL EVENTO A. POR LO TANTO

$$P(A) = \frac{N(A)}{N} = \frac{6}{36} = \frac{1}{6}$$

PROCEDIENDO DE ESTA MANERA SE PUEDEN CALCULAR LAS PROBABILIDADES DE QUE LA SUMA SEA 2,3,4,ETC. LOS RESULTADOS SON:

$$P(\{2\}) = \frac{1}{36};\ P(\{3\}) = \frac{2}{36};\ P(\{4\}) = \frac{3}{36};\ P(\{5\}) = \frac{4}{36};$$

$$P(\{6\}) = \frac{5}{36};\ P(\{7\}) = \frac{6}{36};\ P(\{8\}) = \frac{5}{36};\ P(\{9\}) = \frac{4}{36};$$

$$P(\{10\}) = \frac{3}{36}\ ;\ P(\{11\}) = \frac{2}{36}\ \text{y}\ P(\{12\}) = \frac{1}{36}$$

(OBSERVESE QUE $\sum_{i=2}^{12} P(\{i\})=1$

ASIGNACION DE PROBABILIDADES MEDIANTE UN MODELO MATEMATICO

MEDIANTE ESTE METODO LAS PROBABILIDADES SE ASIGNAN A PARTIR DE UN MODELO MATEMATICO QUE INVOLUCRE TODOS LOS FACTORES POSIBLES QUE INTERVIENEN EN LA ALEATORIEDAD DEL FENOMENO.

AXIOMAS DE LA TEORIA DE PROBABILIDADES

LAS PROBABILIDADES QUE SE ASIGNAN A LOS DIFERENTES EVENTOS RELACIONADOS CON UN FENOMENO ALEATORIO DEBEN CUMPLIR CON LOS SIGUIENTES TRES AXIOMAS:

AXIOMA 1: LA PROBABILIDAD DE OCURRENCIA DE UN EVENTO A ES UN NUMERO, P(A), QUE SE LE ASIGNA A DICHO EVENTO, CUYO VALOR QUEDA EN EL INTERVALO

$$0 \leq P(A) \leq 1$$

AXIOMA 2: SI S ES UN ESPACIO DE EVENTOS, ENTONCES

$$P(S)=1$$

AXIOMA 3: LA PROBABILIDAD, P(C), DE LA UNION, C, DE DOS EVENTOS MUTUAMENTE EXCLUSIVOS, A Y B, ES IGUAL A LA SUMA DE LAS PROBABILIDADES DE ESTOS, ES DECIR,

$$P(A \cup B) = P(C) = P(A) + P(B)$$

EJEMPLOS

A) EN EL PROBLEMA DEL LANZAMIENTO DE UN DADO QUE NO ESTA CARGADO SE PUEDE ASIGNAR A CADA NUMERO (A CADA EVENTO SIMPLE) UNA PROBABILIDAD DE 1/6, SI $A=\{2,4\}$ Y $B=\{5,6\}$, ENTONCES, PUESTO QUE $A=\{2\} \cup \{4\}$ Y $B=\{5\} \cup \{6\}$, Y QUE LOS EVENTOS ELEMENTALES SON MUTUAMENTE EXCLUSIVOS ENTRE SI, APLICANDO EL AXIOMA 3 SE OBTIENEN:

$$P(A)=P(\{2\}) + P(\{4\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(B)=P(\{5\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

SI $C=A \cup B$, Y DADO QUE A Y B SON EVENTOS MUTUAMENTE EXCLUSIVOS:

$$P(C) = P(A) + P(B) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

ADEMAS, OBSERVESE QUE SE CUMPLE CON LOS AXIOMAS 1 Y 2, YA QUE

$$P(S) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\})$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$$

EJEMPLO

EN EL PROBLEMA DEL LANZAMIENTO DE DOS DADOS, ¿CUAL ES LA PROBABILIDAD QUE AL REALIZAR UNA VEZ EL EXPERIMENTO LA SUMA DE LOS DOS NUMEROS QUE QUEDEN HACIA ARRIBA SEA 7 U 11? ESTO ES EQUIVALENTE A PREGUNTAR POR LA PROBABILIDAD DE QUE OCURRA EL EVENTO $C = \{7\} \cup \{11\}$. PUESTO QUE $\{7\}$ Y $\{11\}$ SON EVENTOS MUTUAMENTE EXCLUSIVOS:

$$P(C) = P(\{7\}) + P(\{11\}) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}$$

EJEMPLO

EN UN LABORATORIO SE PROBARON 100 VIGAS DE CONCRETO REFORZADO NOMINALMENTE IDENTICAS, Y SE ANOTARON LAS CARGAS CON LAS CUALES FALLO CADA UNA. DE ESTA SUCESION DE EXPERIMENTOS SE ASIGNARON, EN TERMINOS DE LAS FRECUENCIAS RELATIVAS CORRESPONDIENTES, LAS SIGUIENTES PROBABILIDADES:

SI $A = \{X: 0 \le X \le 20 \text{ ton}\}$; $P(A) = 0.17$

SI $B = \{X: 20 < X \le 40 \text{ ton}\}$; $P(B) = 0.24$

SI $C = \{X: 40 < X \le 60 \text{ ton}\}$; $P(C) = 0.27$

SI $D = \{X: 60 < X \le 80\}$; $P(D) = 0.13$

SI $E = \{X: 80 < X \le 100\}$; $P(E) = 0.11$

SI $F = \{X: 100 < X\}$; $P(F) = 0.08$

$$\Sigma P(.) = 1.00$$

SI SE REALIZA UNA VEZ MAS EL EXPERIMENTO, CALCULEMOS LAS SIGUIENTES PROBABILIDADES:

A) QUE LA RESISTENCIA SEA MENOR O IGUAL QUE 80 TON. PUESTO QUE $G= \{X: 0 \leq X \leq 80 \text{ ton}\}$ SE TIENE QUE $G= A \cup B \cup C \cup D$, POR LO QUE $P(G)=P(A) + P(B) + P(C) + P(D) = 0.17 + 0.24 + 0.27 + 0.13 = 0.81$

B) LA PROBABILIDAD QUE RESISTA MAS DE 60 TONS. PUESTO QUE $H= \{X: 60<X \leq \infty\}$ O $H=\{X: X>60\}$ SE TIENE QUE $H= D \cup E \cup F$, POR LO QUE $P(H) = P(D) + P(E) + P(F) = 0.13 + 0.11 + 0.08 =0.32$

C) LA PROBABILIDAD QUE RESISTA MAS DE 40 TON, PERO CUANDO MUCHO 100 TON.

PUESTO QUE $I=\{X: 40<X \leq 100\}$ SE TIENE QUE $I = C \cup D \cup E$

POR LO QUE $P(I) = P(C) + P(D) + P(E) = 0.27 + 0.13 + 0.11 = 0.51$

TEOREMAS

DOS TEOREMAS IMPORTANTES QUE SE DEDUCEN A PARTIR DE LOS AXIOMAS SON:

TEOREMA 1.

SI A ES UN EVENTO DEL ESPACIO S, ENTONCES $P(\bar{A})=1-P(A)$

DEMOSTRACION

PUESTO QUE A Y $\bar{A}$ SON MUTUAMENTE EXCLUSIVOS Y ADEMAS $A\cup\bar{A}=S$, ENTONCES, $P(S)=P(A)+P(\bar{A})=1$

$\Rightarrow P(\bar{A})=1-P(A)$

CASO PARTICULAR: PUESTO QUE $P(\bar{S})=1-P(S)=0$ Y $\bar{S}=\emptyset$, SE TIENE QUE $P(\emptyset)=0$

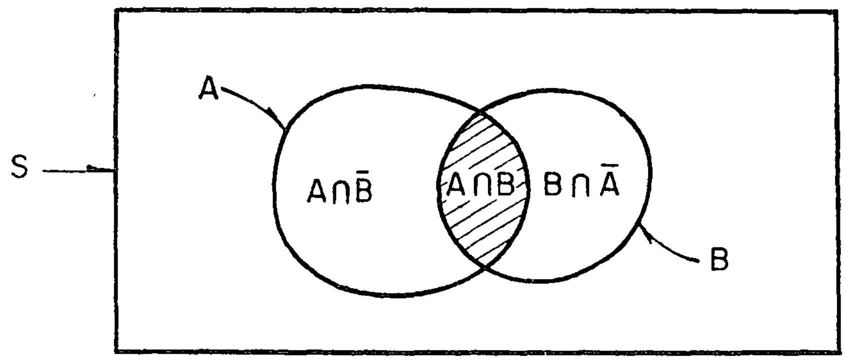
TEOREMA 2.

SI A Y B SON DOS EVENTOS CUALQUIERA DE S, ENTONCES

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

DEMOSTRACION

SEA EL DIAGRAMA DE VENN:



$A \cup B = (A \cap B) \cup (A \cap \bar{B}) \cup (B \cap \bar{A})$. PUESTO QUE $A \cap B$, $A \cap \bar{B}$ Y $B \cap \bar{A}$ SON MUTUAMENTE EXCLUSIVOS, SE TIENE QUE $P(A \cup B) = P(A \cap B) + P(A \cap \bar{B}) + P(B \cap \bar{A})$.

SUMANDO Y RESTANDO $P(A \cap B)$ Y AGRUPANDO TERMINOS SE OBTIENE

$P(A \cup B) = \left[P(A \cap B) + P(A \cap \bar{B})\right] + \left[P(A \cap B) + P(B \cap \bar{A})\right] - P(A \cap B)$

PERO $A = (A \cap \bar{B}) \cup (A \cap B) \Rightarrow P(A \cap B) + P(A \cap \bar{B}) = P(A)$

Y $B = (A \cap B) \cup (B \cap \bar{A}) \Rightarrow P(A \cap B) + P(B \cap \bar{A}) = P(B)$, POR LO QUE

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

EJEMPLO

EN UNA URNA SE TIENEN 28 TIRAS DE PAPEL Y EN CADA UNA SE ENCUENTRA ANOTADA UNA LETRA DISTINTA DEL ALFABETO. CALCULE LA PROBABILIDAD DE QUE AL EXTRAER AL AZAR UNA TIRA:

A) SE OBTENGA UNA VOCAL

B) SE OBTENGA a O z

C) OCURRAN C Y D, DONDE $C=\{x,y,z\}$ y $D=\{b,c,y,z\}$

D) OCURRA C O D

Respuestas

A) $A=\{a,e,i,o,u\} \Rightarrow P(A)=\frac{5}{28}$

B) $B=\{a,z\} \Rightarrow P(B)=\frac{2}{28}$

C) $F=C\cap D=\{y,z\} \Rightarrow P(F)=\frac{2}{28}$

D) $E=C\cup D=\{b,c,x,y,z\} \Rightarrow P(E)=\frac{5}{28}$

o $P(E)=P(C)+P(D)-P(C\cap D)$

$P(C\cap D)=P(F)=\frac{2}{28} \Rightarrow P(E)=\frac{3}{28}+\frac{4}{28}-\frac{2}{28}=\frac{5}{28}$

PROBABILIDAD CONDICIONAL

LA PROBABILIDAD CONDICIONAL, $P(A|B)$, DEL EVENTO A, DADO QUE EL B HA OCURRIDO SE CALCULA CON LA FORMULA

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \quad ; \; P(B)>0 \qquad (1)$$

EVENTOS INDEPENDIENTES

SI DOS EVENTOS, A Y B, SON INDEPENDIENTES, LA PROBABILIDAD DE A NO SE ALTERA SI OCURRE EL EVENTO B; ES DECIR, DOS EVENTOS SON *INDEPENDIENTES* SI

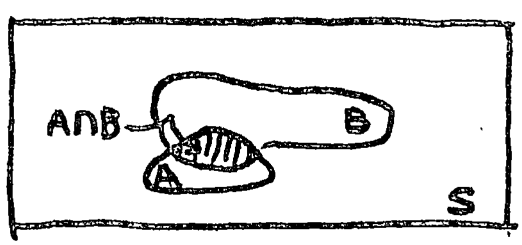
$$P(A|B) = P(A)$$

EN TAL CASO, DE LA ECUACION 1 :

$$P(A\cap B) = P(A) \times P(B) \qquad (1')$$

PUESTO QUE $P(A\cap B) = N(A\cap B)/N(S)$ Y $P(B) = N(B)/N(S)$ LA ECUACION 1 SE PUEDE ESCRIBIR COMO

$$P(A|B) = \frac{\frac{N(A\cap B)}{N(S)}}{\frac{N(B)}{N(S)}} = \frac{N(A\cap B)}{N(B)} \qquad (2)$$



EL TRABAJAR CON LA ECUACION 2 EQUIVALE A EMPLEAR UN ESPACIO DE EVENTOS REDUCIDO DE S A B.

EJEMPLO

EN UNA URNA HAY 10 TRANSISTORES BUENOS Y 10 DEFECTUOSOS. ¿CUAL ES LA PROBABILIDAD DE SACAR UNO BUENO Y UNO DEFECTUOSO (EN CUALQUIER ORDEN) AL REALIZAR DOS EXTRACCIONES AL AZAR, SI HAY REEMPLAZO DEL PRIMER TRANSISTOR OBSERVADO?

LEY GENERAL DE MULTIPLICACION

DE LA ECUACION (1):

$$P(A \cap B) = P(A|B)P(B)$$

ESTA ECUACION SE PUEDE GENERALIZAR A MAS DE DOS EVENTOS ASI:

$$P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1)P(E_2|E_1)\ldots P(E_k|E_1, E_2, \ldots, E_{k-1}) \qquad (3)$$

POR EJEMPLO, SI K=4

$$P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1, E_2) \times$$
$$\times\ P(E_4|E_1, E_2, E_3)$$

EJEMPLO

¿CUAL ES LA PROBABILIDAD DE QUE AL EXTRAER SIN REEMPLAZO CUATRO CARTAS AL AZAR DE UN PAQUETE DE 52, LAS DOS PRIMERAS SEAN DIAMANTES Y LAS DOS ULTIMAS SEAN CORAZONES (EVENTO E)?

SEAN $A = \{$LA 1a. ES DIAMANTE$\}$, $B = \{$LA 2a. ES DIAMANTE$\}$,

$C = \{$LA 3a. ES CORAZON$\}$, $D = \{$LA 4a. ES CORAZON$\}$.

EN TAL CASO

$$E = A \cap B \cap C \cap D = \{(d, d, c, c)\}$$

$P(A) = 13/52$, $P(B|A) = 12/51$, $P(C|A,B) = 13/50$

$P(D|A,B,C) = 12/49$

APLICANDO LA ECUACION 3 SE OBTIENE

$$P(E) = \frac{13}{52}\,\frac{12}{51}\,\frac{13}{50}\,\frac{12}{49} = \frac{78}{20825}$$

SI LOS EVENTOS E_i QUE APARECEN EN LA ECUACION (3) SON INDEPENDIENTES, ENTONCES

$$P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1) \times P(E_2) \times \ldots \times P(E_k)$$

QUE ES LA *LEY GENERAL DE MULTIPLICACION*

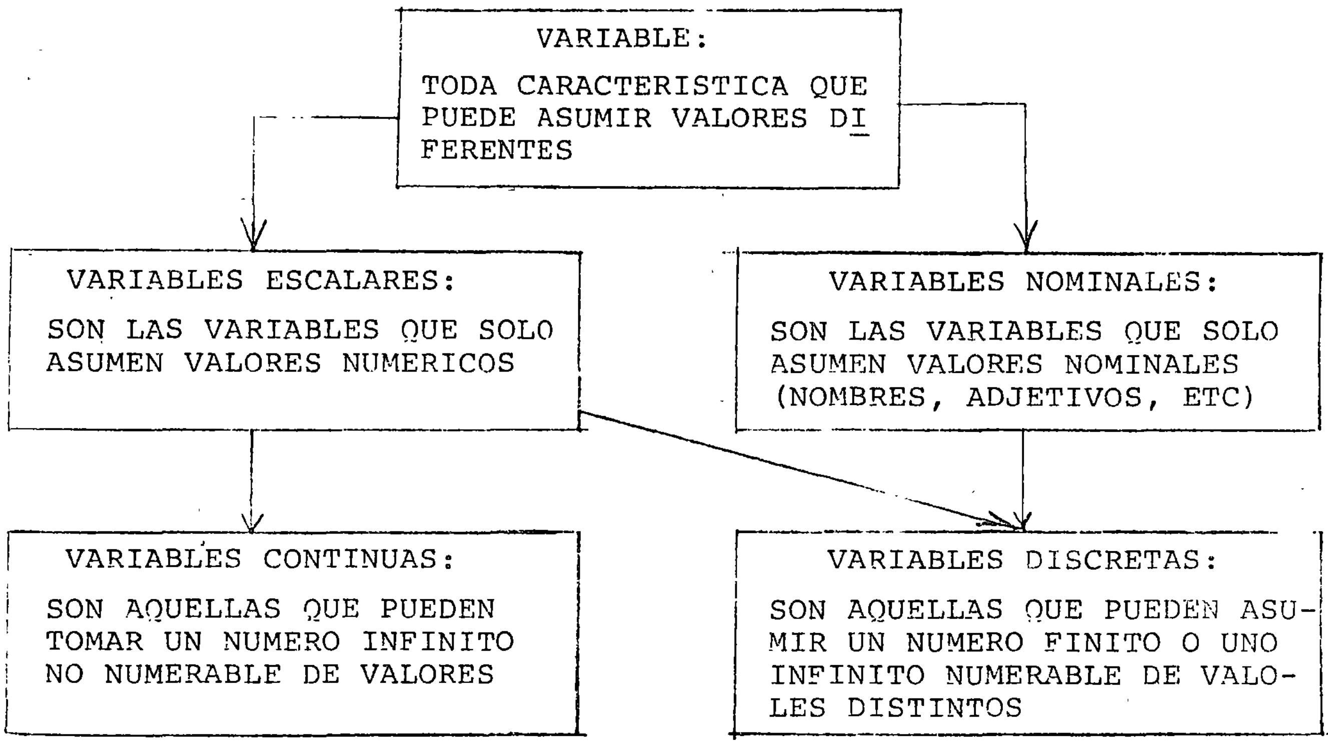
LEY GENERAL DE LA ADICION

SI TODOS LOS EVENTOS E_i SON MUTUAMENTE EXCLUSIVOS ENTRE SI, EL AXIOMA 3 TAMBIEN SE GENERALIZA A:

$$P(E_1 U E_2 U \ldots U E_k) = P(E_1)+P(E_2)+\ldots+P(E_k)$$

VARIABLES ALEATORIAS

CLASIFICACION DE VARIABLES



UNA *VARIABLE ALEATORIA* ES UNA VARIABLE TAL QUE NO PUEDE PREDECIRSE CON CERTEZA EL VALOR QUE ASUMIRA AL REALIZAR UN EXPERIMENTO. POR EJEMPLO, LA RESISTENCIA O CARGA DE FALLA DE UNAS VIGAS ES UNA VARIABLE ALEATORIA, YA QUE ANTES DE ROMPER UNA VIGA TOMADA AL AZAR NO SE PUEDE PRECISAR CUAL SERA SU RESISTENCIA. EN LA SIGUIENTE TABLA SE PRESENTAN LOS RESULTADOS EXPERIMENTALES CON 15 VIGAS DE CONCRETO REFORZADO, OBSERVANDOSE QUE ESTOS VARIAN DE UNAS A OTRAS DE MANERA ALEATORIA.

TABLA 2. PRUEBAS DE VIGAS DE CONCRETO REFORZADO

Número de la viga	Carga de agrietamiento, en kg	Carga de falla, en kg
1	4 700	4 700
2	3 840	4 220
3	3 270	4 360
4	2 310	4 680
5	2 950	4 270
6	4 810	4 810
7	2 720	4 590
8	2 720	4 490
9	4 310	4 310
10	2 950	4 630
11	4 220	4 220
12	2 720	4 340
13	2 720	4 340
14	2 630	4 770
15	2 950	4 630

A TODO EXPERIMENTO SE LE PUEDE ASOCIAR AL MENOS UNA VARIABLE ALEATORIA, DEPENDIENDO ESTA DEL PROBLEMA QUE SE TENGA PLANTEADO. POR EJEMPLO, EN EL CASO DE LA RESISTENCIA DE LAS VIGAS DE VARIABLE ALEATORIA PUEDE SER DIRECTAMENTE DICHA RESISTENCIA, EN CUYO CASO SU ESPACIO DE EVENTOS SERIA

$$S_1 = \{X: 0 \leq X < \infty\}$$

LA VARIABLE TAMBIEN PUDO HABER SIDO UNA CUYO ESPACIO DE EVENTOS FUERA

$$S_2 = \{EXITO, FRACASO\}$$

EN DONDE EL EXITO OCURRIRIA SI LA VIGA RESISTIERA MAS DE CIERTA CANTIDAD, POR EJEMPLO 4600 KG, Y EL FRACASO OCURRIRIA SI RESISTIERA MENOS, ES DECIR:

8

EXITO: SI $X \geq 4600$ KG

FRACASO: SI $X < 4600$ KG

LEYES DE PROBABILIDADES

EL COMPORTAMIENTO DE UNA VARIABLE ALEATORIA SE DESCRIBE MEDIANTE SU *LEY DE PROBABILIDADES*, LA CUAL PUEDE ESPECIFICARSE DE DIFERENTES FORMAS. LA MANERA MAS COMUN DE HACERLO ES MEDIANTE SU *DISTRIBUCION O DENSIDAD DE PROBABILIDADES*.

A FIN DE EVITAR CONFUSION, SE EMPLEARA UNA LETRA MAYUSCULA PARA DENOTAR UNA VARIABLE ALEATORIA, Y LA MINUSCULA CORRESPONDIENTE PARA LOS VALORES QUE PUEDE ASUMIR. SI LA VARIABLE ALEATORIA X ES DISCRETA Y PUEDE ASUMIR LOS VALORES x_i, SU DENSIDAD DE PROBABILIDADES, $f_X(x)$ SERA EL CONJUNTO DE LAS PROBABILIDADES

$$p_X(x_i) = P(X = x_i)$$

LA CUAL SE LEE "PROBABILIDAD DE QUE $X = x_i$". ESTO ES

$$f_X(x) = \{p_X(x_i)\}$$

PARA QUE UNA DENSIDAD DE PROBABILIDADES SATISFAGA LOS TRES AXIOMAS DE LA TEORIA DE PROBABILIDADES, SE DEBEN CUMPLIR LOS SIGUIENTES REQUISITOS

A) $0 \leq p_X(x_i) \leq 1$ PARA TODA x_i

B) $\sum_{i=1}^{n} p_X(x_i) = 1$, DONDE n ES EL NUMERO TOTAL DE VALORES QUE PUEDE ASUMIR X

C) $P(x_m \leq X \leq x_r) = \sum_{i=m}^{i=r} p_X(x_i)$; $m \leq r$

LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS O FUNCION DE DISTRIBUCION

OTRA FORMA DE ESPECIFICAR LA LEY DE PROBABILIDADES DE UNA VARIABLE ALEATORIA ES MEDIANTE LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS, $F_X(x_i)$, QUE SE DEFINE COMO EL CONJUNTO DE LAS SUMAS PARCIALES DE LAS PROBABILIDADES, $P_X(x_i)$, CORRESPONDIENTES A TODOS LOS VALORES DE X MENORES O IGUALES QUE x_i. POR LO TANTO, ESTA FUNCION DA LAS PROBABILIDADES DE QUE LA VARIABLE ALEATORIA TOME VALORES MENORES O IGUALES QUE x_m PARA CUALQUIER m, ES DECIR

$$F_X(x) = \{F_X(x_m)\} \quad ; \quad m = 1,2,\ldots,n$$

EN DONDE

$$F_X(x_m) = \sum_{i=1}^{i=m} P_X(x_i) = P(X \leq x_m)$$

EJEMPLO

SEA X LA VARIABLE ALEATORIA DISCRETA "NUMERO TOTAL DE CARROS QUE SE DETIENEN EN UNA ESQUINA DEBIDO A LA LUZ ROJA DE UN SEMAFORO". SI LAS PROBABILIDADES ASOCIADAS A CADA VALOR, DETERMINADAS POR EL METODO FRECUENCIAL, SON

$$P_X(x) = \begin{cases} 0.1 & \text{SI} \quad x = 0 \\ 0.2 & \text{SI} \quad x = 1 \\ 0.3 & \text{SI} \quad x = 2 \\ 0.2 & \text{SI} \quad x = 3 \\ 0.1 & \text{SI} \quad x = 4 \\ 0.1 & \text{SI} \quad x = 5 \\ 0 & \text{SI} \quad x \geq 6 \end{cases}$$

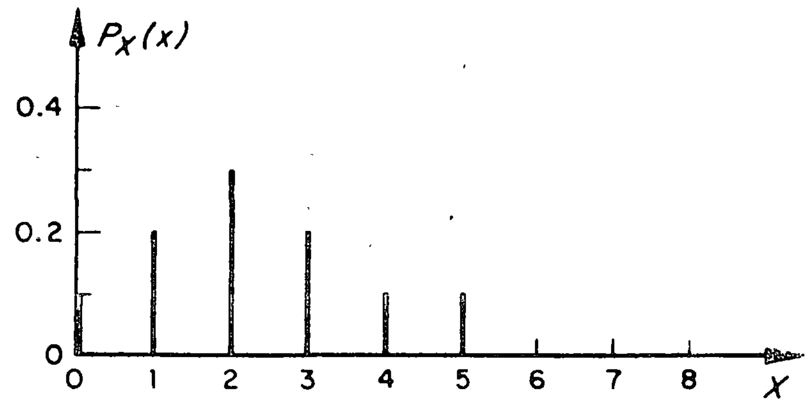
LA DISTRIBUCION DE PROBABILIDADES Y LA DE PROBABILIDADES ACUMULADAS CORRESPONDIENTES SERAN

x	$f_X(x)$	$F_X(x)$
<0	0	0
0	0.1	0.1
1	0.2	0.3
2	0.3	0.6
3	0.2	0.8
4	0.1	0.9
5	0.1	1.0
≥6	0	1.0

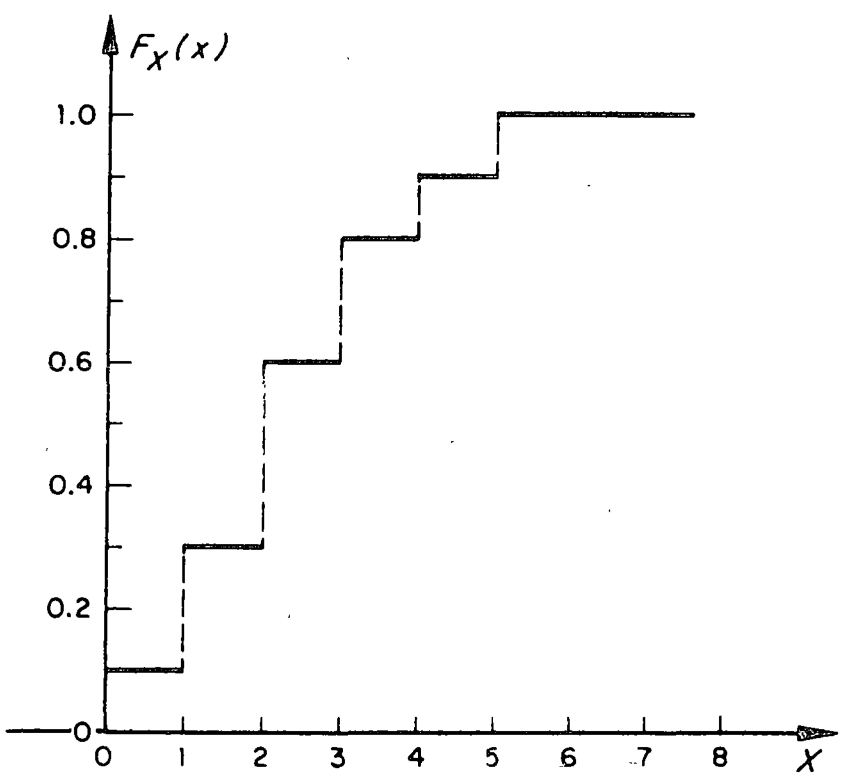
O SEA

$$F_X(x) = \begin{cases} 0, & \text{SI } x < 0 \\ 0.1, & \text{SI } 0 \leq x < 1 \\ 0.3, & \text{SI } 1 \leq x < 2 \\ 0.6, & \text{SI } 2 \leq x < 3 \\ 0.8, & \text{SI } 3 \leq x < 4 \\ 0.9, & \text{SI } 4 \leq x < 5 \\ 1.0, & \text{SI } 5 \leq x \end{cases}$$

LAS GRAFICAS DE ESTAS DISTRIBUCIONES SE PRESENTAN EN LA FIGURA DE LA SIGUIENTE HOJA.



a) Distribución de probabilidades



b) Función de distribución

Ley de probabilidades del ejemplo del tráfico

EJEMPLO

SEA LA VARIABLE ALEATORIA X DEFINIDA POR LA SUMA DE LOS DOS NUMEROS QUE QUEDEN HACIA ARRIBA AL LANZAR DOS DADOS. EN ESTE CASO EL ESPACIO DE EVENTOS ES

$$S = \{2,3,4,5,6,7,8,9,10,11,12\}$$

Y LA DENSIDAD DE PROBABILIDADES ES

$$f_X(x) = \{\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}\}$$

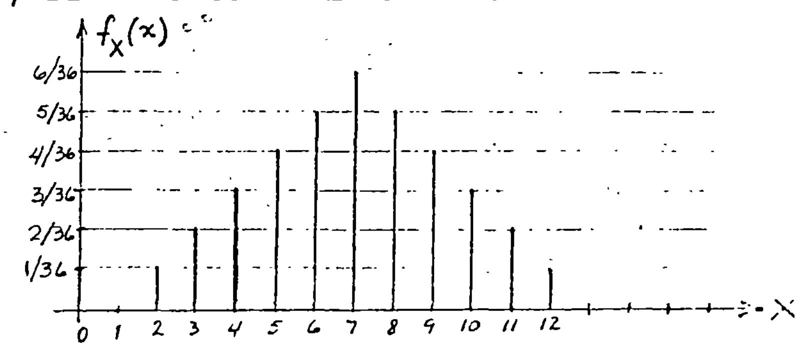
EN ESTE CASO $X_1=2,\ X_2=3,\ldots,X_{11}=12$

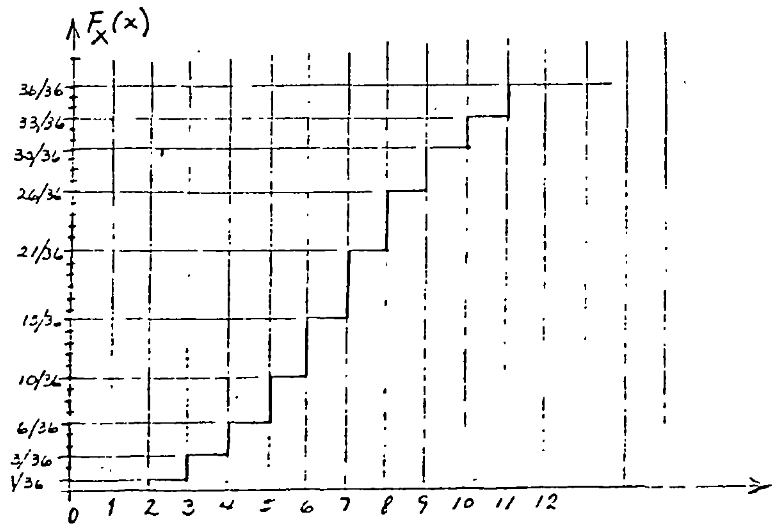
$$Y \quad f_X(2) = \frac{1}{36},\ f_X(3) = \frac{2}{36},\ldots,f_X(12) = \frac{1}{36}$$

ESTAS PROBABILIDADES FUERON CALCULADAS EN UN EJEMPLO PREVIO SOBRE PROBABILIDADES DE EVENTOS .

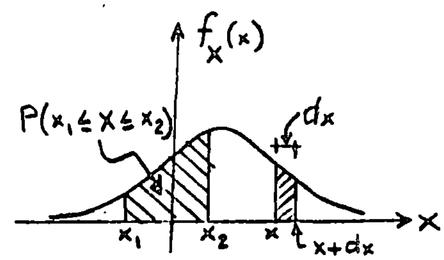
CON ESTAS PROBABILIDADES SE PUEDE OBTENER LA FUNCION DE DISTRIBUCION O DE PROBABILIDADES ACUMULADAS, DE LA SIGUIENTE MANERA:

x	$f_X(x)$	$F_X(x)$
<2	0	0
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36=1
>12	0	1
	$\Sigma=1$	





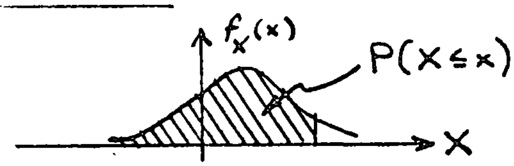
EN EL CASO DE UNA VARIABLE ALEATORIA CONTINUA, X, LA PROBABILIDAD DE QUE ESTA TOME UN VALOR COMPRENDIDO ENTRE x Y x + dx ESTA DADA POR $f_X(x)dx$, DONDE $f_X(x)$ ES LA DENSIDAD DE PROBABILIDADES DE X. POR LO TANTO, LA PROBABILIDAD DE QUE X ASUMA VALORES COMPRENDIDOS EN EL INTERVALO $x_1 \leq X \leq x_2$ ES



$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)\, dx$$

LA INTERPRETACION GRAFICA DE ESTA PROBABILIDAD ES QUE CORRESPONDE AL AREA BAJO LA CURVA DE $f_X(x)$ COMPRENDIDA ENTRE x_1 Y x_2.

PUESTO QUE $F_X(x) = P(X \leq x) = P(-\infty \leq X \leq x)$, Y EN VIRTUD DE LA ECUACION ANTERIOR SE TIENE QUE LA FUNCION DE DISTRIBUCION ES:



$$F_X(x) = \int_{-\infty}^{x} f_X(U)\, dU$$

DONDE U ES SOLO UNA VARIABLE MUDA DE INTEGRACION. EL VALOR DE ESTA INTEGRAL ES IGUAL AL AREA BAJO LA CURVA DE $F_X(x)$ A LA IZQUIERDA DE x. DE ESTA ECUACION SE CONCLUYE QUE

$$\frac{dF_X(x)}{dx} = \frac{d}{dx}\left(\int_{-\infty}^{x} f_X(U)\, dU\right) = f_X(x)$$

ALGUNAS PROPIEDADES DE $F_X(x)$ SON:

$$0 \leq F_X(x) \leq 1$$

$$F_X(-\infty) = 0$$

$$F_X(\infty) = 1$$

$$F_X(x + \varepsilon) \geq F_X(x), \text{ SI } \varepsilon \geq 0$$

$$F_X(x_2) - F_X(x_1) = P(x_1 \leq X \leq x_2)$$

PARA SATISFACER LOS AXIOMAS DE LA TEORIA DE PROBABILIDADES SE NECESITA QUE

$$f_X(x) \geq 0 \text{ PARA TODA } x$$

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

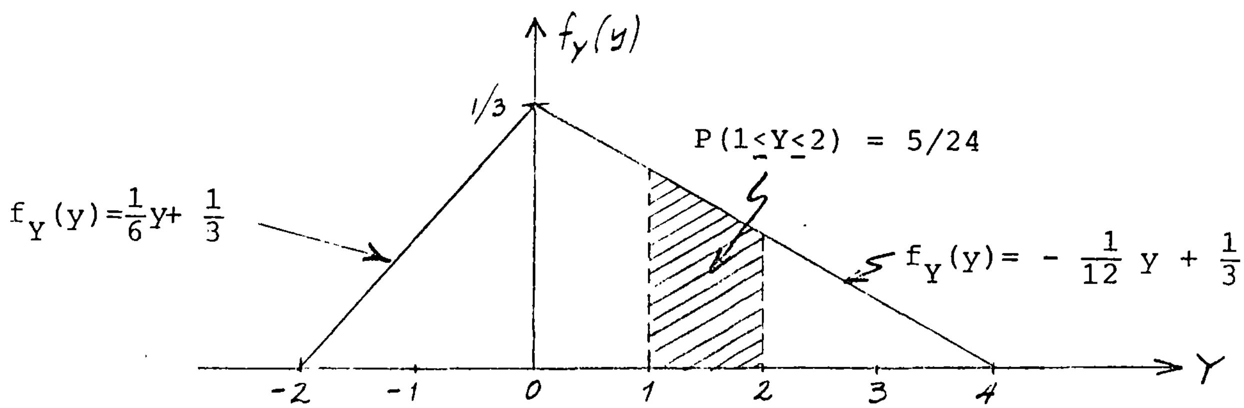
EJEMPLO

SEA UNA VARIABLE ALEATORIA CONTINUA CUYA DENSIDAD DE PROBABILIDADES ES DE FORMA TRIANGULAR DADA POR LAS SIGUIENTES ECUACIONES:

$$f_Y(y) = \frac{1}{6}y + \frac{1}{3}, \text{ SI } -2 \leq Y \leq 0$$

$$f_Y(y) = -\frac{1}{12}y + \frac{1}{3}, \text{ SI } 0 \leq Y \leq 4$$

$$f_Y(y) = 0 \qquad \text{SI } Y \leq 2 \text{ O } Y \geq 4$$



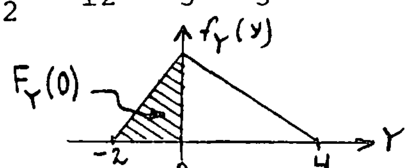
LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS ES, ENTONCES:

SI $-2 \leq Y \leq 0$

$$F_Y(y) = \int_{-\infty}^{Y} f_Y(U)dU = \int_{-2}^{Y} \left(\frac{1}{6}U + \frac{1}{3}\right)dU$$

$$= \left[\frac{U^2}{12} + \frac{U}{3}\right]_{-2}^{Y} = \frac{Y^2}{12} + \frac{Y}{3} + \frac{1}{3}$$

SI $0 \leq Y \leq 4$

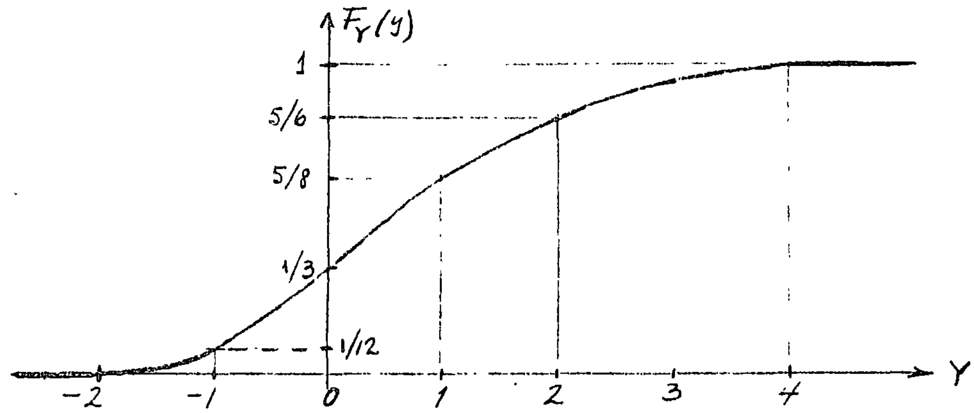


$$F_Y(y) = F_Y(0) + \int_0^Y \left(-\frac{1}{12}U + \frac{1}{3}\right) dU = \frac{1}{3} + \left[\frac{-U^2}{24} + \frac{U}{3}\right]_0^Y =$$

$$= \frac{1}{3} - \frac{Y^2}{24} + \frac{Y}{3} \qquad \text{SI } 0 \leq Y \leq 4$$

$F_Y(y) = 0 \qquad \text{SI} \qquad y \leq -2$

$F_Y(y) = 1 \qquad \text{SI} \qquad y \geq 4$



SI SE DESEA CALCULAR LA PROBABILIDAD DE QUE AL REALIZAR UNA VEZ EL EXPERIMENTO QUE INVOLUCRA A DICHA VARIABLE, EL VALOR QUE SE OBSERVE CAIGA EN EL INTERVALO $1 \leq Y \leq 2$, ENTONCES

$$P[1 \leq Y \leq 2] = \int_1^2 \left(-\frac{1}{12}y + \frac{1}{3}\right) dy = \left[-\frac{y^2}{24} + \frac{y}{3}\right]_1^2 = \frac{5}{24}$$

O

$$P[1 \leq Y \leq 2] = F_Y(2) - F_Y(1) = \frac{5}{6} - \frac{5}{8} = \frac{5}{24}$$

FUNCION DE DISTRIBUCION COMPLEMENTARIA

EL COMPLEMENTO, $G_X(x)$, DE LA DISTRIBUCION DE PROBABILIDADES ACUMULADAS SE UTILIZA CUANDO LAS DECISIONES SE TOMAN CON BASE EN PROBABILIDADES DE QUE SE EXCEDA UN VALOR DADO DE LA VARIABLE. LA *FUNCION DE DISTRIBUCION COMPLEMENTARIA* SE DEFINE COMO

$$G_X(x) = P(X>x) = 1 - F_X(x)$$

EJEMPLO

PARA EL PROBLEMA ANTERIOR DE LA VELOCIDAD MAXIMA ANUAL DEL VIENTO, CALCULEMOS LA PROBABILIDAD DE QUE ESTA SEA MAYOR DE 140 KM/H:

$$G_X(140) = P(X \geq 140) = \int_{140}^{\infty} 0.033 e^{-0.033x} dx = 0.0099$$

O, ALTERNATIVAMENTE

$$P(X \geq 140) = 1-F_X(140) = G_X(140) = 1-(1-e^{-0.033x140}) = e^{-4.62} = 0.0099$$

ESPERANZAS

LA *ESPERANZA* DE UNA FUNCION $g(X)$, DE UNA VARIABLE ALEATORIA DISCRETA, X, ES, POR DEFINICION

$$E(g(X)) = \sum_{i=1}^{i=n} g(x_i)P_X(x_i)$$

O PARA UNA VARIABLE CONTINUA

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx$$

EJEMPLOS

1. SI $g(X)$ = CONSTANTE = c

$$E(c) = c\int_{-\infty}^{\infty} f_X(x)\,dx = c$$

2. SI $g(X) = cx$

$$E[cx] = c\overbrace{\int_{-\infty}^{\infty} x f_X(x)\,dx}^{E[X]} = cE[X].$$

3. SI $g(X) = a + bx$

$$E[a+bx] = a\int_{-\infty}^{\infty} f_X(x)\,dx + b\int_{-\infty}^{\infty} x f_X(x)\,dx = a + bE[X]$$

4. SI $g(X) = g_1(X) + g_2(X)$

$$E[g_1(X) + g_2(X)] = \int_{-\infty}^{\infty} g_1(x)f_X(x)\,dx + \int_{-\infty}^{\infty} g_2(x)f_X(x)\,dx$$

$$= E[g_1(X)] + E[g_2(X)]$$

EJEMPLO

SI X ES UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES EXPONENCIAL, CALCULAR LA ESPERANZA DE LA FUNCION

$$g(X) = X^2$$

EN ESTE CASO SE TIENE QUE

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{SI } 0 \leq X \leq \infty \text{ , Y} \quad f_X(x) = 0 \quad \text{SI } X<0$$

POR LO QUE

$$E(X^2) = E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = \lambda \int_{-\infty}^{\infty} x^2 e^{-\lambda x} dx$$

$$= \lambda \left[\frac{-x^2 e^{-\lambda x}}{\lambda} \right]_0^{\infty} + \frac{2\lambda}{\lambda} \int_0^{\infty} x e^{-\lambda x} dx = \frac{-2}{\lambda^2} \left[e^{-\lambda x}(1+\lambda x) \right]_0^{\infty} = \frac{2}{\lambda^2}$$

EN GENERAL, A LA ESPERANZA DE X^2 SE LE DENOMINA *VALOR MEDIO CUADRATICO.*

Ejemplo. Construcción de la carpeta de una carretera.

Un contratista construirá la carpeta de una carretera en tramos de 50 m; el gobierno aceptará o rechazará cada tramo de acuerdo con una prueba de control de calidad. El contratista tiene la opción de pedir el concreto a una de dos plantas premezcladoras; la planta A cobra 140 pesos/m^3 y la B 160 pesos/m^3, pero, el control de calidad que se lleva en la planta B es mejor, lo cual hace más probable que un tramo dado pase favorablemente la prueba de aceptación. Tomando en cuenta que en cada tramo se usan 100 m^3 de concreto y que la probabilidad de que el proveniente de la planta A no pase la prueba de control es 0.10, y la de B es 0.05, el constructor deberá decidirse por cuál planta usar. El árbol de decisiones de este problema es el mostrado en la fig 6.4, donde $P(\theta_1)$ y $P(\theta_2)$ son las probabilidades de que ocurran θ_1 y θ_2, respectivamente. La utilidad $U_1 = u(a_1, \theta_1)$ es la que corresponde a utilizar la planta A y que la carpeta pase la prueba de control de calidad; en este caso la utilidad (negativa) es el costo del concreto ($14,000.00) más la colocación (supongamos $100,000.00), por lo cual $U_1 =$ -114,000.00. $U_2 = u(a_1, \theta_2)$ es la que corresponde a usar la planta A y que la carpeta no pase la prueba de calidad; en este caso el constructor deberá demoler y reconstruir el tramo con los siguientes costos:

Carpeta demolida	Pérdida de prestigio	$ 5,000.00
	Mano de obra de demolición	15,000.00
	Concreto	14,000.00
	Mano de obra de colocación	100,000.00

Reconstrucción	Mano de obra	$ 100,000.00
	Concreto	14,000.00
	T O T A L	$ 248,000.00

De manera similar se obtienen U_3 y U_4, cuyos valores resultan ser $U_3 = -\ \$116{,}000.00$ y $U_4 = -\ \$252{,}000.00$.

Si la decisión se tomara sin considerar las probabilidades de aceptar la carpeta, el constructor se decidiría por la planta A, ya que la pérdida (utilidad negativa) sería menor. Si sí se toman en cuenta y adoptamos como *criterio de decisión* el escoger la planta que conduzca a una *esperanza de pérdida* menor se tendrá (recuerde que la esperanza de la variable aleatoria X, $E[X]$, es $E[X] = \sum_{i=1}^{m} P[X_i]X_i$, donde las X_i son los valores que puede asumir X, y $P[X_i]$ son las probabilidades correspondientes):

Para la planta A:

$$E[U] = 0.90 \times (-114{,}000) + 0.10 \times (-248{,}000) = -\$127{,}400.$$

Para la planta B:

$$E[U] = 0.95 \times (-116{,}000) + 0.05 \times (-252{,}000) = -\$122{,}800.$$

Comparando ambas cifras se concluye que la decisión de comprar el concreto de la planta B conduce a una pérdida esperada menor que la de la planta A, es decir, se escoge la planta B aunque el precio unitario del concreto sea mayor.

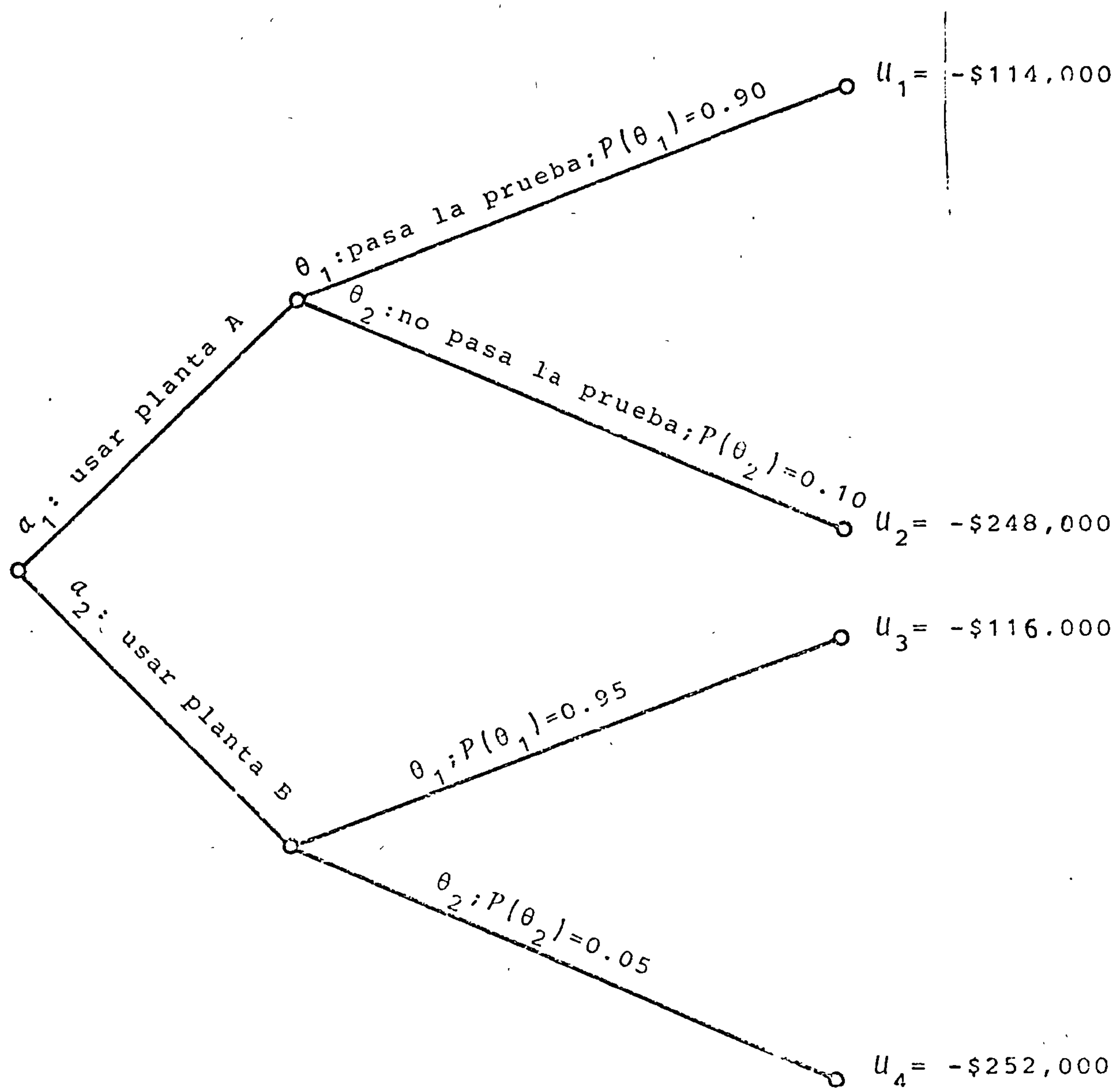


Fig 6.4 Arbol de decisiones del ejemplo 6.2

MEDIDAS DE TENDENCIA CENTRAL

LA *MEDIA O ESPERANZA*, $E[X]$, DE UNA VARIABLE ALEATORIA, X, SE CALCULA CON LAS ECUACIONES ANTERIORES PARA EL CASO EN QUE $g(X)=X$. DE ESTA MANERA, SI LA VARIABLE ES DISCRETA, SU ESPERANZA QUEDA DADA POR

$$E(X) = \sum_{i=1}^{i=n} x_i \, P_X(x_i)$$

DONDE n ES EL TOTAL DE VALORES QUE X PUEDE ASUMIR.

PARA EL CASO DE UNA VARIABLE ALEATORIA CONTINUA, LA MEDIA ES

$$m_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

OTRAS MEDIDAS USUALES DE TENDENCIA CENTRAL DE UNA VARIABLE ALEATORIA SON LA *MEDIANA* Y EL *MODO*. LA PRIMERA SE DEFINE COMO EL VALOR DE LA VARIABLE AL CUAL CORRESPONDE UNA PROBABILIDAD ACUMULADA DE 50%, Y LA SEGUNDA, COMO EL VALOR DE LA VARIABLE AL CUAL CORRESPONDE LA MAYOR PROBABILIDAD.

EJEMPLO

SI LA DENSIDAD DE PROBABILIDADES DE LA VARIABLE ALEATORIA X CORRESPONDE A LOS ERRORES EN UNA NIVELACION, ES LA DE LA SEGUNDA COLUMNA DE LA SIGUIENTE TABLA, LA MEDIA DE DICHA VARIABLE RESULTA SER 4 167 LA MEDIANA 4000 Y EL MODO 4000 MICRAS. LOS CALCULOS CORRESPONDIENTES SE LOCALIZAN EN LA TERCERA COLUMNA.

x_i, EN MICRAS	$P_X(x_i)$	$x_i P_X(x_i)$, EN MICRAS	$F_X(x_i)$
0	6/60	0	6/60
1 000	2/60	2 000/60	8/60
2 000	4/60	8 000/60	12/60
3 000	8/60	24 000/60	20/60
4 000	13/60	52 000/60	$33/60 \simeq 0.5$
5 000	12/60	60 000/60	45/60
6 000	7/60	42 000/60	52/60
7 000	4/60	28 000/60	56/60
8 000	2/60	16 000/60	58/60
9 000	2/60	18 000/60	60/60
TOTAL: E[X] = 250 000/60=4 167 MICRAS			

EJEMPLO

CALCULAR LA ESPERANZA DE UNA VARIABLE ALEATORIA CUYA DENSIDAD DE PROBABILIDADES ES TRIANGULAR DADA POR

$$f_Y(y) = \frac{1}{6}y + \frac{1}{3} \qquad \text{SI } -2 \leq y \leq 0$$

$$f_Y(y) = \frac{-1}{12}y + \frac{1}{3} \qquad \text{SI } 0 \leq y \leq 4$$

$$f_Y(y) = 0 \qquad \text{SI } y \leq -2 \quad \text{O} \quad y \geq 4$$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy = \int_{-2}^{0} y\left(\frac{y}{6} + \frac{1}{3}\right)dy + \int_{0}^{4} y\left(\frac{-y}{12} + \frac{1}{3}\right)dy$$

$$= \left[\frac{y^3}{18} + \frac{y^2}{6}\right]_{-2}^{0} + \left[\frac{-y^3}{36} + \frac{y^2}{6}\right]_{0}^{4} = \frac{2}{3}$$

EJEMPLO

CALCULAR LA ESPERANZA DE UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES EXPONENCIAL

$$f_X(x) = \lambda e^{-\lambda x}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \lambda \int_{0}^{\infty} x e^{-\lambda x} dx = \lambda \left[\frac{-e^{-\lambda x}}{\lambda^2} (1+\lambda x) \right]_0^{\infty} = \frac{1}{\lambda}$$

MEDIDAS DE DISPERSION

UNA MEDIDA MUY COMUN DE LA DISPERSION O VARIABILIDAD DE LOS VALORES QUE PUEDE ASUMIR UNA VARIABLE ALEATORIA ES LA *VARIANCIA*, LA CUAL SE DENOTA COMO $\sigma^2(X)$ O VAR (X), LA CUAL SE DEFINE COMO LA ESPERANZA DE LA FUNCION $g(X) = [X-E(X)]^2$. ASI, PARA UNA VARIABLE ALEATORIA DISCRETA

$$\sigma^2(X) = VAR(X) = \sum_{i=1}^{i=n} (x_i - E(X))^2 \, P_X(x_i)$$

Y PARA UNA CONTINUA

$$\sigma^2(X) = VAR(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)dx$$

DESARROLLANDO EL INTEGRANDO DE ESTA ULTIMA ECUACION:

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x^2-2xE(X) + E^2(X))f_X(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)dx - 2E(X)\int_{-\infty}^{\infty} xf_X(x)dx + E^2(X)\int_{-\infty}^{\infty} f_X(x)dx = E[X^2]-E^2[X]$$

ES DECIR, LA VARIANCIA SE PUEDE CALCULAR COMO LA DIFERENCIA DEL VALOR MEDIO CUADRATICO Y EL CUADRADO DE LA MEDIA DE X.

OTRAS MEDIDAS DE DISPERSION DE LA VARIABLE ALEATORIA X SON LA *DESVIACION ESTANDAR*, $\sigma(X)$, LA CUAL ES IGUAL A LA RAIZ CUADRADA DE LA VARIANCIA, Y EL *COEFICIENTE DE VARIACION* QUE SE DEFINE COMO

$$\nu(X)=\sigma(X)/E(X) \quad \text{SI } E(X)\neq 0$$

EJEMPLO

EN LA SIGUIENTE TABLA SE CALCULA LA VARIANCIA DE LA VARIABLE ALEATORIA CUYA DENSIDAD DE PROBABILIDADES SE PRESENTO EN EL EJEMPLO ANTERIOR ($E(x) = 4167$ micras).

$x_i - E(X)$ EN MICRAS	$(x_i - E(x))^2$	$P_X(x_i)$	$(x_i - E(X))^2 P_X(x_i)$, EN MICRAS
-4 167		6/60	1 740 000
-3 167		2/60	333 000
-2 167		4/60	313 000
-1 167		8/60	181 000
- 167		13/60	6 000
833		12/60	139 000
1 833		7/60	390 000
2 833		4/60	531 000
3 833		2/60	487 000
4 833		2/60	687 000

TOTAL: 4 798 000 MICRAS$^2 = \sigma^2(X)$

LA DESVIACION ESTANDAR Y EL COEFICIENTE DE VARIACION DE ESTA VARIABLE ALEATORIA SON, RESPECTIVAMENTE,

$$\sigma(X) = \sqrt{4\ 798\ 000} = 2\ 200 \text{ MICRAS, Y } \nu(X) = \sigma(X)/E(X) = \frac{2\ 200}{4\ 167} = 0.528$$

EJEMPLO

SI X ES UNA VARIABLE ALEATORIA CON DISTRIBUCION DE PROBABILIDADES EXPONENCIAL, CALCULAR SU VARIANCIA, DESVIACION ESTANDAR Y COEFICIENTE DE VARIACION:

$$\sigma^2(X) = E(X-E[X])^2 = \int_{-\infty}^{\infty}(x-E[X])^2\lambda e^{-\lambda x}dx = \lambda\int_0^{\infty}(x^2-2xE[X]+E^2[X])e^{-\lambda x}dx$$

$$= \lambda\int_0^{\infty}x^2e^{-\lambda x}dx - 2E[X]\lambda\int_0^{\infty}xe^{-\lambda x}dx + E^2[X]\int_0^{\infty}\lambda e^{-\lambda x}dx$$

$$= \frac{2}{\lambda^2} - 2\,\frac{1}{\lambda}\,\frac{1}{\lambda} + \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

YA QUE $E(X) = 1/\lambda$ Y $E[x^2] = 2/\lambda^2$.

USANDO LA FORMULA $\sigma^2(X) = E[X^2] - E^2[X]$, Y TOMANDO EN CUENTA QUE $E[X^2] = 2/\lambda^2$ SE OBTIENE:

$$\sigma^2(X) = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$$

EN CONSECUENCIA, LA DESVIACION ESTANDAR ES

$$\sigma(X) = \sqrt{1/\lambda^2} = 1/\lambda$$

Y EL COEFICIENTE DE VARIACION

$$\nu(X) = \sigma(X)/E(X) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda}} = 1$$

8B

EJEMPLO

SEA Y UNA VARIABLE ALEATORIA CON DENSIDAD DE PROBABILIDADES TRIANGULAR DADA POR

$$f_Y(y) = \frac{1}{6}y + \frac{1}{3} \qquad \text{SI} \quad -2 \leq y \leq 0$$

$$f_Y(y) = \frac{-1}{12}y + \frac{1}{3} \qquad \text{SI} \quad 0 \leq y \leq 4$$

$$f_Y(y) = 0 \qquad \text{SI} \quad y \leq -2 \text{ O } y \geq 4$$

CALCULAR LA VARIANCIA, LA DESVIACION ESTANDAR Y EL COEFICIENTE DE VARIACION.

CALCULAREMOS PRIMERO EL VALOR MEDIO CUADRATICO PARA LUEGO APLICAR LA ECUACION $\sigma^2(Y) = E(Y^2) - E^2(Y)$

$$E[Y^2] = \int_{-2}^{0} y^2\left(\frac{1}{6}y + \frac{1}{3}\right)dy + \int_{0}^{4} y^2\left(-\frac{y}{12} + \frac{1}{3}\right)dy = \left[\frac{y^4}{24} + \frac{y^3}{9}\right]_{-2}^{0} + \left[\frac{-y^4}{48} + \frac{y^3}{9}\right]_{0}^{4} = 2$$

$$\sigma^2(Y) = 2 - (2/3)^2 = 14/9$$

$$\sigma(Y) \doteq 1.25 \quad \left(\sqrt{14/9}\right)$$

$$\nu(Y) = 1.25/(2/3) = 1{,}88$$

DISTRIBUCIONES PARTICULARES

VARIABLES ALEATORIAS DISCRETAS

DISTRIBUCION BINOMIAL O DE BERNOULLI

LA DISTRIBUCION BINOMIAL O DE BERNOULLI SE EMPLEA COMO DENSIDAD DE PROBABILIDADES DE VARIABLES ALEATORIAS DISCRETAS ASOCIADOS A EXPERIMENTOS EN LOS QUE SOLO HAY (O SOLO IMPORTAN) DOS RESULTADOS POSIBLES, UNO DE LOS CUALES USUALMENTE SE DENOMINA "EXITO" Y, EL OTRO, "FRACASO". ($S = \{$éxito, fracaso$\}$)

SEAN p= PROBABILIDAD DE OBSERVAR "EXITO" AL REALIZAR UNA VEZ EL EXPERIMENTO

q= PROBABILIDAD DE "FRACASO" = 1-p

X= VARIABLE ALEATORIA "NUMERO DE EXITOS OBSERVADOS AL REPETIR n VECES EL EXPERIMENTO "CON REEMPLAZO"

LA DISTRIBUCION DE PROBABILIDADES BINOMIAL ES

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad : \quad x = 0, 1, \ldots, n$$

SE PUEDE DEMOSTRAR QUE LOS PARAMETROS DE ESTA DISTRIBUCION SON

$$E(X) = np \ , \quad \sigma^2(X) = npq$$

SI n=2, ENTONCES X PUEDE ASUMIR LOS VALORES 0, 1 y 2, ES DECIR $S=\{0,1,2\}$. EL ESPACIO DE EVENTOS DEL EXPERIMENTO ES

$$S_1 = \{\underbrace{(\text{FRACASO, FRACASO})}_{X=0}, \underbrace{(\text{EXITO, FRACASO})}_{X=1}, \underbrace{(\text{FRACASO,EXITO})}_{X=1},$$

$$\underbrace{(\text{EXITO, EXITO})}_{X=2}\}$$

$$f_X(x) = \{P(0), P(1), P(2)\}$$

OBSERVESE QUE $x=0$ OCURRE DE UNA MANERA, $x=1$, DE DOS, Y $x=2$, DE UNA. ESTOS RESULTADOS SE PUEDEN OBTENER PERMUTANDO DOS GRUPOS, UNO CON x Y EL OTRO CON $n-x$ ELEMENTOS:

$x=0$: $\quad {}_2P_{0,2} = \frac{2!}{0!x2!} = 1$; $P(\{0\}) = q \times q = q^2 = p^0q^2$

$x=1$: $\quad {}_2P_{1,1} = \frac{2!}{1!x1!} = 2$; $P(\{1\}) = 2pq$

$x=2$: $\quad {}_2P_{2,0} = \frac{2!}{2!x0!} = 1$; $P(\{2\}) = pxp = p^2 = p^2q^0$

$$\sum_{i=0}^{2} P(\{i\}) = q^2+2pq+p^2 = (p+q)^2 = 1$$

(OBSERVESE QUE LOS ELEMENTOS DE S_1 NO SON IGUALMENTE PROBABLES, A MENOS QUE $p = q = 1/2$.)

SI $n = 3$, $S = \{0,1,2,3\}$, e = EXITO Y f = FRACASO, ENTONCES

$S_1 = \{(f,f,f),\ (e,f,f),\ (f,e,f),\ (f,f,e),\ (e,e,f),\ (e,f,e),\ (f,e,e),\ (e,e,e)\}$

$x = 0$: ${}_3P_{0,3} = \frac{3!}{0!\times 3!} = 1$; $P(\{0\}) = 1\ p^0 q^3 = q^3$

$x = 1$: ${}_3P_{1,2} = \frac{3!}{1!\times 2!} = 3$; $P(\{1\}) = 3\ pq^2$

$x = 2$: ${}_3P_{2,1} = \frac{3!}{2!\times 1!} = 3$; $P(\{2\}) = 3p^2q$

$x = 3$: ${}_3P_{3,0} = \frac{3!}{3!\times 0!} = 1$; $\underline{P(\{3\}) = 1p^3q^0 = p^3}$

$$\sum_{i=0}^{3} P(\{i\}) = (p+q)^3 = 1$$

PASANDO AL CASO GENERAL DE CUALQUIER VALOR DE n, LA PROBABILIDAD DE QUE OCURRAN x EXITOS Y $n-x$ FRACASOS *EN UN ORDEN DETERMINADO* ES

$$P(X=x) = p^x q^{n-x}$$

EN VIRTUD DE LA LEY GENERAL DE MULTIPLICACION.

UN ORDEN POSIBLE SERIA, POR EJEMPLO,

$$\underbrace{\text{EXITO, EXITO,} \ldots, \text{EXITO}}_{x},\ \underbrace{\text{FRACASO,} \ldots, \text{FRACASO}}_{n-x}$$

AHORA BIEN, LOS x EXITOS PUEDEN OCURRIR EN ${}_nP_{x,n-x}$ ORDENES DISTINTOS. CADA UNO CON PROBABILIDAD $p^x q^{n-x}$. POR LO TANTO, EN VIRTUD DE LA LEY GENERAL DE ADICION, LA DISTRIBUCION DE PROBABILIDADES DE X RESULTA SER

$$f_X(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x} \ ; \ x = 0,1,\ldots,n$$

LA CUAL SE CONOCE CON EL NOMBRE DE *BINOMIAL O DE BERNOULLI*.

9

LA ESPERANZA DE LA DISTRIBUCION DE BERNOULLI ES

$$E(X) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= np \underbrace{\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}}_{(p+q)^{n-1}} = np(p+q)^{n-1} = np$$

LA VARIANCIA DE LA DISTRIBUCION BINOMIAL ES

$$\sigma^2(X) = E[(X-E(X))^2] = E[(X-np)^2]$$

PERO $E[(X-np)^2] = E(X^2-2npX + n^2p^2) = E[X + X(X-1) - 2npX + n^2p^2]$

$$=E[(1-2np)X] + E[X(X-1)] + E(n^2p^2)$$

$$=(1-2np)np + n^2p^2 + \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p^x q^{n-x} x(x-1)$$

$$=np-n^2p^2 + \sum_{x=2}^{n} n(n-1) \frac{(n-2)!}{(x-2)!(n-x)!} p^2 p^{x-2} q^{n-x}$$

$$=np-n^2p^2 + n(n-1)p^2 \underbrace{\sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x}}_{(p+q)^{n-2}}$$

$$=np-n^2p^2 + n(n-1)p^2(p+q)^{n-2} = np-np^2=np(1-p)=npq$$

EN RESUMEN, PARA LA DISTRIBUCION BINOMIAL,

$$E(X) = np \quad ; \quad \sigma^2(X) = npq \; ; \; \sigma(X) = \sqrt{npq}$$

TABLA 1

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!\,(n-k)!}\, p^k\, q^{n-k}$$

				p		
n	x	0.10	0.20	0.30	0.40	0.50
2	0	.8100	.6400	.4900	.3600	.2500
	1	.9900	.9600	.9100	.8400	.7500
3	0	.7290	.5120	.3430	2160	.1250
	1	.9720	.8960	.7840	.6480	.5000
	2	.9990	.9920	.9730	.9360	.8750
4	0	.6561	.4096	.2401	.1296	.0625
	1	.9477	.8192	.6517	.4752	.3125
	2	.9963	.9728	9163	.8208	.6875
	3	.9999	.9984	.9919	.9744	.9375
5	0	.5905	.3277	.1681	.0778	.0312
	1	.9185	.7373	.5282	.3370	.1875
	2	.9914	.9421	.8369	.6826	.5000
	3	.9995	.9933	.9692	.9130	.8125
	4	1 0000	.9997	.9976	.9898	.9688
6	0	.5314	.2621	.1176	.0467	.0156
	1	.8857	.6554	.4202	.2333	.1094

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
6	2	.9842	.9011	.7443	.5443	.3438
	3	.9987	.9830	.9295	.8208	.6562
	4	.9999	.9984	.9891	.9590	.8906
	5	1.0000	.9999	.9993	.9959	.9844
7	0	.4783	.2097	.0824	.0280	.0078
	1	.8503	.5767	.3294	.1586	.0625
	2	.9743	.8520	.6471	.4199	.2266
	3	.9973	.9667	.8740	.7102	.5000
	4	.9998	.9953	.9712	.9037	.7734
	5	1.0000	.9996	.9962	.9812	.9375
	6	1.0000	1.0000	.9998	.9984	.9922
8	0	.4305	.1678	.0576	.0168	.0039
	1	.8131	.5033	.2553	.1064	.0352
	2	.9619	.7969	.5518	.3154	.1445
	3	.9950	.9437	.8059	.5941	.3633
	4	.9996	.9896	.9420	.8263	.6367
	5	1.0000	.9988	.9887	.9502	.8555
	6	1.0000	.9999	.9987	.9915	.9648
	7	1.0000	1.0000	.9999	.9993	.9961
9	0	.3874	.1342	.0404	.0101	.0020
	1	.7748	.4362	.1960	.0705	.0195
	2	.9470	.7382	.4628	.2318	.0898
	3	.9917	.9144	.7297	.4826	.2539
	4	.9991	.9804	.9012	.7334	.5000
	5	.9999	.9969	.9747	.9006	.7461
	6	1.0000	.9997	.9957	.9750	.9102
	7	1.0000	1.0000	.9996	.9962	.9805
	8	1.0000	1.0000	1.0000	.9997	.9980
10	0	.3487	.1074	.0282	.0060	.0010
	1	.7361	.3758	.1493	.0464	.0107
	2	.9298	.6778	.3828	.1673	.0547
	3	.9872	.8791	.6496	.3823	.1719
	4	.9984	.9672	.8497	.6331	.3770

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
10	5	.9999	.9936	.9527	.8338	.6230
	6	1.0000	.9991	.9894	.9452	.8281
	7	1.0000	.9999	.9984	.9877	.9453
	8	1.0000	1.0000	.9999	.9983	.9893
	9	1.0000	1.0000	1.0000	.9999	.9990
11	0	.3138	.0859	.0198	.0036	.0005
	1	.6974	.3221	.1130	.0302	.0059
	2	.9104	.6174	.3127	.1189	.0327
	3	.9815	.8389	.5696	.2963	.1133
	4	.9972	.9496	.7897	.5328	.2744
	5	.9997	.9883	.9218	.7535	.5000
	6	1.0000	.9980	.9784	.9006	.7256
	7	1.0000	.9998	.9957	.9707	.8867
	8	1.0000	1.0000	.9994	.9941	.9673
	9	1.0000	1.0000	1.0000	.9993	.9941
	10	1.0000	1.0000	1.0000	1.0000	.9995
12	0	.2824	.0687	.0138	.0022	.0002
	1	.6590	.2749	.0850	.0196	.0032
	2	.8891	.5583	.2528	.0834	.0193
	3	.9744	.7946	.4925	.2253	.0730
	4	.9957	.9274	.7237	.4382	.1938
	5	.9995	.9806	.8822	.6652	.3872
	6	.9999	.9961	.9614	.8418	.6128
	7	1.0000	.9994	.9905	.9427	.8062
	8	1.0000	.9999	.9983	.9847	.9270
	9	1.0000	1.0000	.9998	.9972	.9807
	10	1.0000	1.0000	1.0000	.9997	.9968
	11	1.0000	1.0000	1.0000	1.0000	.9998
13	0	.2542	.0550	.0097	.0013	.0001
	1	.6213	.2336	.0637	.0126	.0017
	2	.8661	.5017	.2025	.0579	.0112
	3	.9658	.7473	.4206	.1686	.0461
	4	.9935	.9009	.6543	.3530	.1334
	5	.9991	.9700	.8346	.5744	.2905
	6	.9999	.9930	.9376	.7712	.5000
	7	1.0000	.9988	.9818	9023	.7095

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	0.10	0.20	*p* 0.30	0.40	0.50
13	8	1.0000	.9998	.9960	.9679	.8666
	9	1.0000	1.0000	.9993	.9922	.9539
	10	1.0000	1.0000	.9999	.9987	.9888
	11	1.0000	1.0000	1.0000	.9999	.9983
	12	1.0000	1.0000	1.0000	1.0000	.9999
14	0	.2288	.0440	.0068	.0008	.0001
	1	.5846	.1979	.0475	.0081	.0009
	2	.8416	.4481	.1608	.0398	.0065
	3	.9559	.6982	.3552	.1243	.0287
	4	.9908	.8702	.5842	.2793	.0898
	5	.9985	.9561	.7805	.4859	.2120
	6	.9998	.9884	.9067	.6925	.3953
	7	1.0000	.9976	.9685	.8499	.6047
	8	1.0000	.9996	.9917	.9417	.7880
	9	1.0000	1.0000	.9983	.9825	.9102
	10	1.0000	1.0000	.9998	.9961	.9713
	11	1.0000	1.0000	1.0000	.9994	.9935
	12	1.0000	1.0000	1.0000	.9999	.9991
	13	1.0000	1.0000	1.0000	1.0000	.9999
15	0	.2059	.0352	.0047	.0005	.0000
	1	.5490	.1671	.0353	.0052	.0005
	2	.8159	.3980	.1268	.0271	.0037
	3	.9444	.6482	.2969	.0905	.0176
	4	.9873	.8358	.5155	.2173	.0592
	5	.9978	.9389	.7216	.4032	.1509
	6	.9997	.9819	.8689	.6098	.3036
	7	1.0000	.9958	.9500	.7869	.5000
	8	1.0000	.9992	.9848	.9050	.6964
	9	1.0000	.9999	.9963	.9662	.8491
	10	1.0000	1.0000	.9993	.9907	.9408
	11	1.0000	1.0000	.9999	.9981	.9824
	12	1.0000	1.0000	1.0000	.9997	.9963
	13	1.0000	1.0000	1.0000	1.0000	.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
18	0	.1501	.0180	.0016	.0001	.0000
	1	.4503	.0991	.0142	.0013	.0001
	2	.7338	.2713	.0600	.0082	.0007
	3	.9018	.5010	.1646	.0328	.0038
	4	.9718	.7164	.3327	.0942	.0154
	5	.9936	.8671	.5344	.2088	.0481
	6	.9988	.9487	.7217	.3743	.1189
	7	.9998	.9837	.8593	.5634	.2403
	8	1.0000	.9957	.9404	.7368	.4073
	9	1.0000	.9991	.9790	.8653	.5927
	10	1.0000	.9998	.9939	.9424	.7597
	11	1.0000	1.0000	.9986	.9797	.8811
	12	1.0000	1.0000	.9997	.9942	.9519
	13	1.0000	1.0000	1.0000	.9987	.9846
	14	1.0000	1.0000	1.0000	.9998	.9962
	15	1.0000	1.0000	1.0000	1.0000	.9993
	16	1.0000	1.0000	1.0000	1.0000	.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000
19	0	.1351	.0144	.0011	.0001	.0000
	1	.4203	.0829	.0104	.0008	.0000
	2	.7054	.2369	.0462	.0055	.0004
	3	.8850	.4551	.1332	.0230	.0022
	4	.9648	.6733	.2822	.0696	.0096
	5	.9914	.8369	.4739	.1629	.0318
	6	.9983	.9324	.6655	.3081	.0835
	7	.9997	.9767	.8180	.4878	.1796
	8	1.0000	.9933	.9161	.6675	.3238
	9	1.0000	.9984	.9674	.8139	.5000
	10	1.0000	.9997	.9895	.9115	.6762
	11	1.0000	1.0000	.9972	.9648	.8204
	12	1.0000	1.0000	.9994	.9884	.9165
	13	1.0000	1.0000	.9999	.9969	.9682
	14	1.0000	1.0000	1.0000	.9994	.9904
	15	1.0000	1.0000	1.0000	.9999	.9978
	16	1.0000	1.0000	1.0000	1.0000	.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	p 0.10	0.20	0.30	0.40	0.50
16	0	.1353	.0281	.0033	.0003	.0000
	1	.5147	.1407	.0261	.0033	.0003
	2	.7892	.3518	.0994	.0183	.0021
	3	.9316	.5981	.2459	.0651	.0106
	4	.9830	.7982	.4499	.1666	.0384
	5	.9967	.9183	.6598	.3288	.1051
	6	.9995	.9733	.8247	.5272	.2272
	7	.9999	.9930	.9256	.7161	.4018
	8	1.0000	.9985	.9743	.8577	.5982
	9	1.0000	.9998	.9929	.9417	.7728
	10	1.0000	1.0000	.9984	.9809	.8949
	11	1.0000	1.0000	.9997	.9951	.9616
	12	1.0000	1.0000	1.0000	.9991	.9894
	13	1.0000	1.0000	1.0000	.9999	.9979
	14	1.0000	1.0000	1.0000	1.0000	.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	.1668	.0225	.0023	.0002	.0000
	1	.4818	.1182	.0193	.0021	.0001
	2	.7618	.3096	.0774	.0123	.0012
	3	.9174	.5489	.2019	.0464	.0064
	4	.9779	.7582	.3887	.1260	.0245
	5	.9953	.8943	.5968	.2639	.0717
	6	.9992	.9623	.7752	.4478	.1662
	7	.9999	.9891	.8954	.6405	.3145
	8	1.0000	.9974	.9597	.8011	.5000
	9	1.0000	.9995	.9873	.9081	.6855
	10	1.0000	.9999	.9968	.9652	.8338
	11	1.0000	1.0000	.9993	.9894	.9283
	12	1.0000	1.0000	.9999	.9975	.9755
	13	1.0000	1.0000	1.0000	.9995	.9936
	14	1.0000	1.0000	1.0000	.9999	.9988
	15	1.0000	1.0000	1.0000	1.0000	.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000

TABLA 1 (continuación)

FUNCION DE DISTRIBUCION BINOMIAL

$$F_X(x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

n	x	0.10	0.20	P 0.30	0.40	0.50
20	0	.1216	.0115	.0008	.0000	.0000
	1	.3917	.0692	.0076	.0005	.0000
	2	.6769	.2061	.0355	.0036	.0002
	3	.8670	.4114	.1071	.0160	.0013
	4	.9568	.6296	.2375	.0510	.0059
	5	.9887	.8042	.4164	.1256	.0207
	6	.9976	.9133	.6080	.2500	.0577
	7	.9996	.9679	.7723	.4159	.1316
	8	.9999	.9900	.8867	.5956	.2517
	9	1.0000	.9974	.9520	.7553	.4119
	10	1.0000	.9994	.9829	.8725	.5881
	11	1.0000	.9999	.9949	.9435	.7483
	12	1.0000	1.0000	.9987	.9790	.8684
	13	1.0000	1.0000	.9997	.9935	.9423
	14	1.0000	1.0000	1.0000	.9984	.9793
	15	1.0000	1.0000	1.0000	.9997	.9941
	16	1.0000	1.0000	1.0000	1.0000	.9987
	17	1.0000	1.0000	1.0000	1.0000	.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000

EJEMPLO

SI SE LANZA AL AIRE SEIS VECES UNA MONEDA HOMOGENEA,

A) ¿CUAL ES LA PROBABILIDAD DE OBTENER DOS "CARAS"?

B) ¿CUAL ES LA PROBABILIDAD DE OBTENER POR LO MENOS CUATRO "CARAS" $(X\geq 4)$?

C) ¿CUANTO VALEN LA ESPERANZA Y LA DESVIACION ESTANDAR?

SOLUCION

A) PUESTO QUE LA MONEDA ES HOMOGENEA SE TIENE $p=1/2$ Y $q=1-1/2=1/2$, DONDE p ES LA PROBABILIDAD DE OBSERVAR "CARA" (CARA = EXITO) EN UN LANZAMIENTO. POR TANTO

$$P[X=2]=f_X(2)=\frac{6!}{2!(6-2)!}(\tfrac{1}{2})^2(1/2)^{6-2}=\frac{6!}{2!\ 4!}(1/2)^6=\frac{15}{64}=0.2344$$

B) PARA QUE SE CUMPLA $X\geq 4$ EN SEIS LANZAMIENTOS, SE NECESITA QUE SE OBSERVEN 4,5 o 6 CARAS. PUESTO QUE ESTOS TRES EVENTOS SON MUTUAMENTE EXCLUSIVOS, SE TIENE

$$P[X\geq 4]=f_X(4)+f_X(5)+f_X(6)$$

CALCULANDO LOS TRES SUMANDOS COMO EN LA PREGUNTA ANTERIOR, RESULTA

$$P[X\geq 4]=\frac{6!}{4!\ 2!}(1/2)^4(1/2)^{6-4}+\frac{6!}{5!\ 1!}(1/2)^5(1/2)^{6-5}+\frac{6!}{6!\ 0!}(1/2)^6(1/2)^{6-6}$$

$$=\frac{15}{64}+\frac{6}{64}+\frac{1}{64}=\frac{11}{32}=0.3438$$

C) $E[X]=np=6(1/2)=3$

$$\sigma^2[X]=npq=6(1/2)(1/2)=3/2\ ,\quad \sigma(X)=\sqrt{3/2}=1.22$$

DISTRIBUCION HIPERGEOMETRICA

CUANDO SE TIENE UNA VARIABLE ALEATORIA DISCRETA CUYO ESPACIO DE EVENTOS TIENE SOLO DOS ELEMENTOS, DIGAMOS S={EXITO, FRACASO}, Y SE REALIZA UN MUESTREO *SIN REEMPLAZO*, ENTONCES LOS RESULTADOS DE CADA EXPERIMENTO *NO SON INDEPENDIENTES NI LA PROBABILIDAD DE EXITO PERMANECE CONSTANTE*, COMO EN LA DISTRIBUCION BINOMIAL, POR LO QUE ESTA ULTIMA NO ES APLICABLE.

SEA X LA VARIABLE ALEATORIA NUMERO DE EXITOS OBSERVADOS AL REPETIR n VECES EL EXPERIMENTO CONSISTENTE EN EXTRAER, SIN REEMPLAZO, ELEMENTOS DE UN LOTE QUE TIENE N OBJETOS DE LOS CUALES M SON "EXITOS". EL NUMERO DE ELEMENTOS QUE TIENE EL ESPACIO DE EVENTOS DEL EXPERIMENTO ES

$$N(S) = {}_{N}C_{n}$$

EL NUMERO, N({X=x}), DE MANERAS POSIBLES E IGUALMENTE PROBABLES DE OBTENER x EXITOS ES

EN EL LOTE

$$\overbrace{\underbrace{\odot\odot\odot\cdots\odot\odot}_{M\ \text{EXITOS}}\;\Big|\;\underbrace{\odot\odot\odot\cdots\odot\odot}_{(N-M)\ \text{FRACASOS}}}^{N\ \text{ELEMENTOS}}$$

EN LA MUESTRA

$$\overbrace{\underbrace{\odot\odot\odot\cdot\cdot\odot\odot}_{}\;\Big|\;\underbrace{\odot\odot\cdot\cdot\odot}_{}}^{n\ \text{ELEMENTOS}}$$

x EXITOS QUE SE PUEDEN ELEGIR DE ${}_{M}C_{x}$ MANERAS	(n-x) FRACASOS QUE SE PUEDEN ELEGIR DE ${}_{N-M}C_{n-x}$ MANERAS

CADA ELECCION POSIBLE DE x EXITOS SE COMBINA CON CADA ELECCION POSIBLE DE (n-x) FRACASOS; POR LO TANTO, EL NUMERO TOTAL DE MANERAS DE OBTENER x EXITOS EN n EXTRACCIONES SIN REEMPLAZO ES

$$N(\{X=x\}) = ({}_{M}C_{x})({}_{N-M}C_{n-x})$$

POR LO TANTO

$$P(\{X=x\}) = f_X(x) = \frac{({}_{M}C_{x})({}_{N-M}C_{n-x})}{{}_{N}C_{n}} = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \quad x=0,1,\ldots,n$$

EN DONDE $\binom{M}{x} = \frac{M!}{x!(M-x)!}$, $\binom{N-M}{n-x} = \frac{(N-M)!}{(n-x)!(N-M-n+x)!}$

Y $\binom{N}{n} = \frac{N!}{n!(N-n)!}$

QUE SE CONOCE COMO *DISTRIBUCION HIPERGEOMETRICA*. LA MEDIA Y LA VARIANCIA DE ESTA DISTRIBUCION SON

$$E(X) = \sum_{x=0}^{n} x\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = nM/N$$

Y

$$\sigma^2(X) = \sum_{x=0}^{n} \left(x-\frac{nM}{N}\right)^2 \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{Mn(N-M)(N-n)}{N^2(N-1)}$$

RESPECTIVAMENTE.

EJEMPLO

EN UN PROBLEMA DE CONTROL ESTADISTICO DE CALIDAD, SE TIENE UN LOTE DE 100 TRANSFORMADORES DE CORRIENTE ELECTRICA, DE LOS CUALES 40 SON DEFECTUOSOS (NO CUMPLEN LAS NORMAS DE FABRICACION). ¿CUAL ES LA PROBABILIDAD DE OBTENER UNO DEFECTUOSO DE TRES SELECCIONADOS AL AZAR SIN REEMPLAZO?

$$P[X=1] = \frac{\binom{40}{1}\binom{100-40}{3-1}}{\binom{100}{3}} = \frac{\binom{40}{1}\cdot\binom{60}{2}}{\binom{100}{3}}$$

$$= \frac{\frac{40!}{39! \times 1!} \times \frac{60!}{58! \times 2!}}{\frac{100!}{97! \times 3!}} = 0.438$$

APROXIMACION DE LA DISTRIBUCION HEPERGEOMETRICA MEDIANTE LA BINOMIAL

CUANDO *N ES GRANDE* Y *n PEQUEÑO* $(N \geq 10n)$, LA DISTRIBUCION BINOMIAL SE PUEDE USAR *COMO APROXIMACION* DE LA HIPERGEOMETRICA. DE ESTA APROXIMACION SE HECHA MANO CUANDO LOS CALCULOS CON ESTA ULTIMA RESULTAN TEDIOSOS.

EN EL CASO DEL EJEMPLO ANTERIOR, SI SE USA LA DENSIDAD BINOMIAL SE OBTIENE, CON $p=40/100 = 0.40$ Y $n=3$

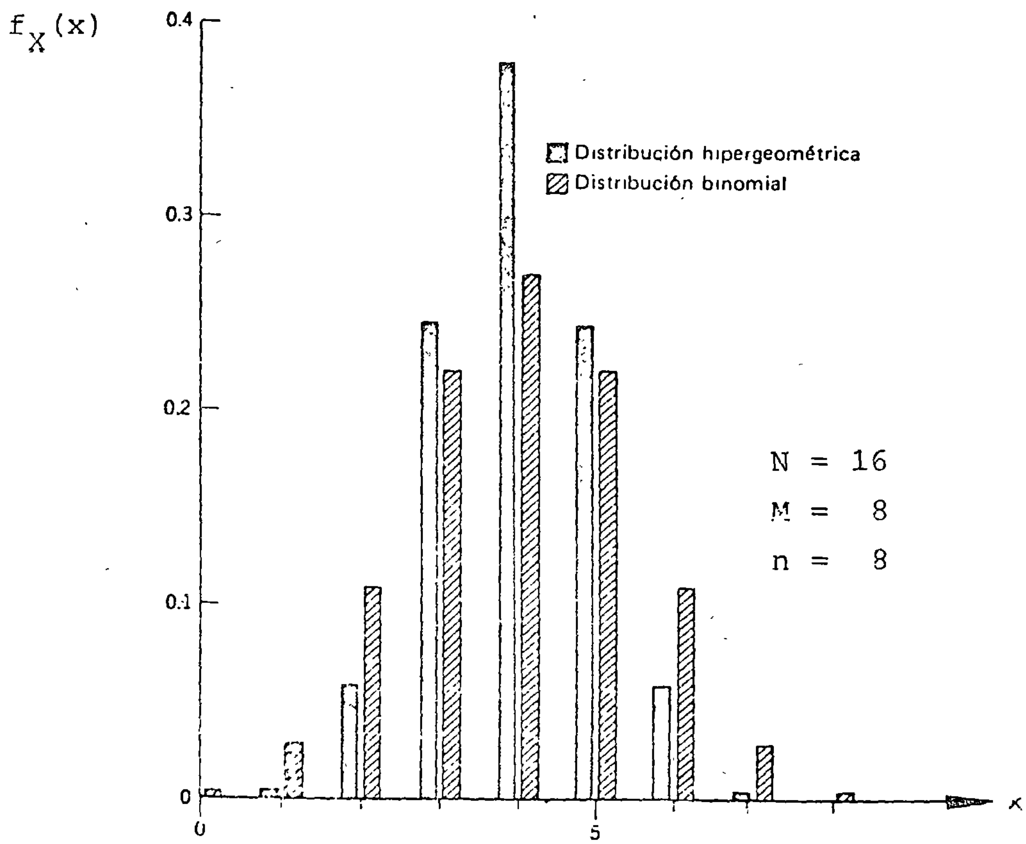
$$P[X=1] = \frac{3!}{1!\ 2!}(0.40)^1(0.60)^2 = 0.432$$

FORMULA DE STIRLING

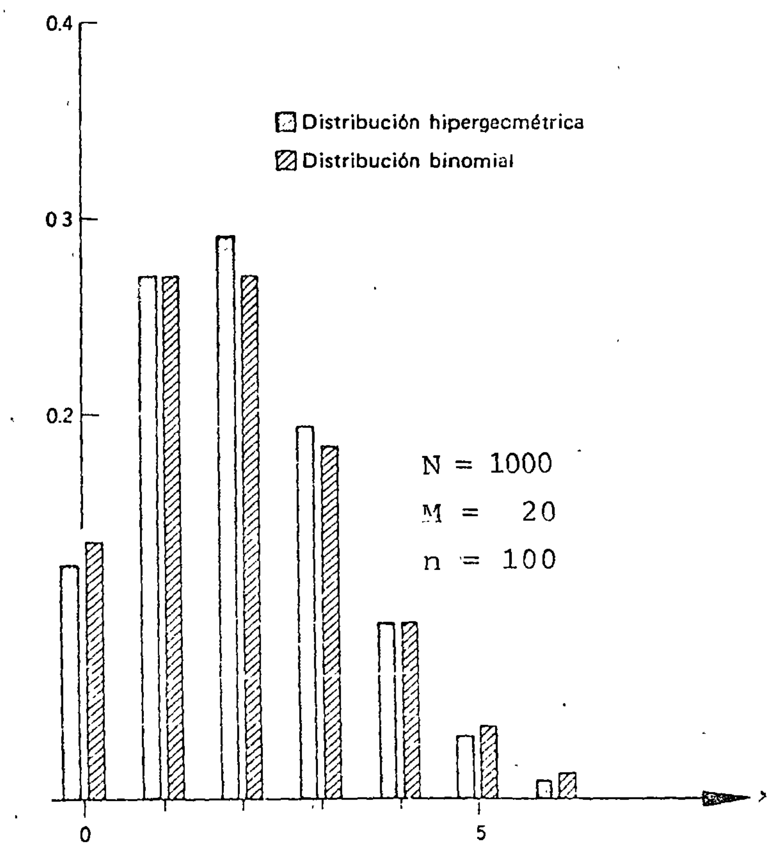
$$N! \simeq \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n \quad ; \quad (e = 2.718...)$$

Error = 2% SI n = 4

Error = 0.8% SI n = 10



COMPARACION DE LAS DISTRIBUCIONES HIPERGEOMETRICA Y BINOMIAL



10

DISTRIBUCION DE POISSON

UNA DISTRIBUCION DE PROBABILIDADES PARA UNA VARIABLE ALEATORIA DISCRETA, X, DE LA FORMA

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \; ; \; x = 0, 1, 2, ...$$

SE LLAMA *DISTRIBUCION DE POISSON*; EN ESTA ECUACION λ ES UNA CONSTANTE. SE PUEDE DEMOSTRAR QUE LA MEDIA Y LA VARIANCIA PARA ESTA DISTRIBUCION QUEDAN DADAS POR

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \lambda$$

$$\sigma^2(X) = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{\lambda^x e^{-\lambda}}{x!} = \lambda$$

ES POSIBLE DEMOSTRAR QUE LA DISTRIBUCION DE POISSON PUEDE EMPLEARSE COMO UNA PROXIMACION DE LA DE BERNOULLI, TOMANDO $\lambda = np$ CUANDO n ES GRANDE Y p PEQUEÑA, PERO DE TAL MANERA QUE $npq > 1$. AL RESPECTO, SI $n=20$ Y $p=0.05$, ENTONCES EL ERROR QUE SE TIENE AL USAR DICHA APROXIMACION ES MENOR DE 3 POR CIENTO PARA VALORES DE X MENORES DE 3; PARA $X=4$ Y $X=5$ LOS ERRORES RESPECTIVOS SON 15 Y 41 POR CIENTO, DEBIDO A QUE NO SE CUMPLE CON LA CONDICION DE QUE npq SEA MAYOR DE UNO, YA QUE $npq = 20 \times 0.05 \times 0.95 = 0.95$.

TABLA 2

FUNCION DE DISTRIBUCION DE POISSON

$$F_X(x) = \sum_{k=0}^{x} \frac{e^{-\lambda} \lambda^k}{k!}$$

X	λ .50	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
0	.607	.368	.135	.050	.018	.007	.002	.001	.000	.000
1	.910	.736	.406	.199	.092	.040	.017	.007	.003	.001
2	.986	.920	.677	.423	.238	.125	.062	.030	.014	.006
3	.998	.981	.857	.647	.433	.265	.151	.082	.042	.021
4	1.000	.996	.947	.815	.629	.440	.285	.173	.100	.055
5	1.000	.999	.983	.961	.785	.616	.446	.301	.191	.116
6	1.000	1.000	.995	.966	.889	.762	.606	.450	.313	.207
7	1.000	1.000	.999	.988	.949	.867	.744	.599	.453	.324
8	1.000	1.000	1.000	.996	.979	.932	.847	.729	.593	.456
9	1.000	1.000	1.000	.999	.992	.968	.916	.830	.717	.587
10	1.000	1.000	1.000	1.000	.997	.986	.957	.901	.816	.706
11	1.000	1.000	1.000	1.000	.999	.995	.980	.947	.888	.803
12	1.000	1.000	1.000	1.000	1.000	.998	.991	.973	.936	.876
13	1.000	1.000	1.000	1.000	1.000	.999	.996	.987	.966	.926
14	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.983	.959
15	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998	.992	.978
16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.996	.989
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.995
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLA 2 (continuación)

FUNCION DE DISTRIBUCION DE POISSON

$$F_X(x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

x	λ = 10.0	11.0	12.0	13.0	14.0	15.0
2	.003	.001	.001	.000	.000	.000
3	.010	.005	.002	.001	.000	.000
4	.029	.015	.008	.004	.002	.001
5	.067	.038	.020	.011	.006	.003
6	.130	.079	.046	.026	.014	.008
7	.220	.143	.090	.054	.032	.018
8	.333	.232	.155	.100	.062	.037
9	.458	.341	.242	.166	.109	.070
10	.583	.460	.347	.252	.176	.118
11	.697	.579	.462	.353	.260	.185
12	.792	.689	.576	.463	.358	.268
13	.864	.781	.682	.573	.464	.363
14	.917	.854	.772	.675	.570	.466
15	.951	.907	.844	.764	.669	.568
16	.973	.944	.899	.835	.756	.664
17	.986	.968	.937	.890	.827	.749
18	.993	.982	.963	.930	.883	.819
19	.997	.991	.979	.957	.923	.875
20	.998	.995	.988	.975	.952	.917
21	.999	.998	.994	.986	.971	.947
22	1.000	.999	.997	.992	.983	.967
23	1.000	1.000	.999	.996	.991	.981
24	1.000	1.000	.999	.998	.995	.989
25	1.000	1.000	1.000	.999	.997	.994
26	1.000	1.000	1.000	1.000	.999	.997
27	1.000	1.000	1.000	1.000	.999	.998
28	1.000	1.000	1.000	1.000	1.000	.999
29	1.000	1.000	1.000	1.000	1.000	1.000

EJEMPLO

SI LA PROBABILIDAD DE QUE FALLE UNA VARILLA DE ACERO AL APLICARLE UNA DETERMINADA FUERZA DE TENSION ES DE 0.001, ¿CUAL ES LA PROBABILIDAD DE QUE DE 2000 VARILLAS PROBADAS FALLEN A) TRES, B) MAS DE DOS?

CON $\lambda = 2000 \times 0.001 = 2$ Y CONSIDERANDO QUE $npq = 1.9 > 1$, SE PUEDE USAR LA DISTRIBUCION DE POISSON COMO APROXIMACION DE LA BINOMIAL:

$$a) \qquad P[X = 3] = \frac{\lambda^3 \; e^{-\lambda}}{3!}$$

$$P[X = 3] = \frac{2^3 \; e^{-2}}{3!} = 0.18$$

EN ESTE CASO LA DISTRIBUCION BINOMIAL DA COMO RESULTADO

$$P[X=3] = \frac{2000!}{3! \times 1997!} (0.001)^3 (0.999)^{1997} = 0.184$$

$$b) \qquad P[X > 2] = 1 - P[X \leqslant 2] = 1 - F_X(2) = 1 - \Big\{ P[X = 0] +$$

$$+ P[X = 1] + P[X = 2] \Big\} = 1 - \frac{2^0 \; e^{-2}}{0!} - \frac{2^1 \; e^{-2}}{1!} - \frac{2^2 \; e^{-2}}{2!} =$$

$$= 1 - \frac{1}{e^2} - \frac{2}{e^2} - \frac{2}{e^2} = 1 - \frac{5}{e^2} = 0.323$$

EJEMPLO

UNA COMPAÑIA ASEGURADORA DESPUES DE MUCHOS AÑOS DE EXPERIENCIA HA ESTIMADO QUE EL 0.004% DE LA POBLACION FALLECE ANUALMENTE POR ACCIDENTE AUTOMOVILISTICO. SI ESTA COMPAÑIA TIENE 40,000 ASEGURADOS, ¿CUAL ES LA PROBABILIDAD DE QUE 2 DE ELLOS MUERAN EN UN AÑO POR ESE TIPO DE ACCIDENTE?

SEA X EL NUMERO DE PERSONAS QUE MUEREN ANUALMENTE DE ENTRE LOS ASEGURADOS, POR ACCIDENTE. LA MEDIA DE X ES

$$E[X] = 0.00004 \times 40{,}000 = 1.6 = \lambda$$

ADEMAS, TOMANDO EN CUENTA QUE npq>1, SE PUEDE USAR SIN GRAN ERROR LA DISTRIBUCION DE POISSON:

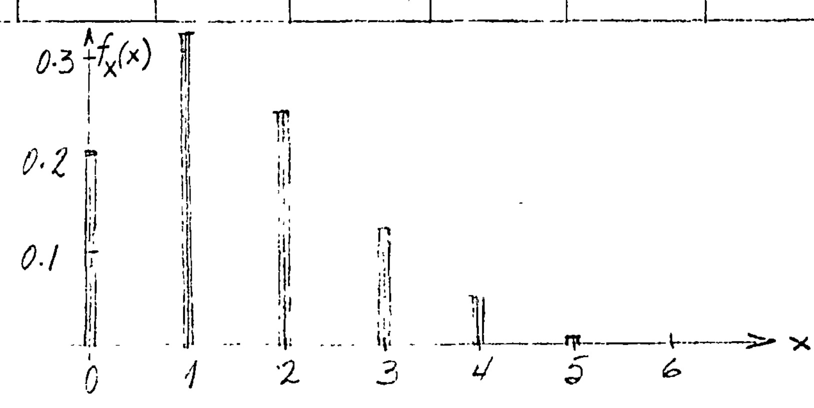
$$P[X=x] = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(1.6)^x e^{-1.6}}{x!} \quad ; \; x=0,\ 1,\ 2,\ldots$$

POR LO QUE

$$P[X=2] = \frac{(1.6)^2 e^{-1.6}}{2!} = \frac{0.2019 \times 2.56}{2} = 0.26$$

LA DISTRIBUCION DE PROBABILIDADES PARA ESTA VARIABLE ALEATORIA ES:

x	0	1	2	3	4	5	6	...
$f_X(x)$	0.202	0.323	0.258	0.138	0.055	0.018	0.005	...



EJEMPLO

EN LA AMPLIACION DEL CARRIL PARA DAR VUELTA A LA IZQUIERDA EN UNA AVENIDA, SOLO HAY CAPACIDAD PARA 3 AUTOS COMO MAXIMO ESPERANDO LA FLECHA LUMINOSA DEL SEMAFORO. EN UN ESTUDIO ESTADISTICO DEL TRANSITO EN ESE LUGAR SE ENCONTRO QUE EN CADA CICLO DE LUCES DEL SEMAFORO HAY EN PROMEDIO 6 AUTOS QUE VAN A DAR VUELTA. ¿CUAL ES LA PROBABILIDAD DE QUE EN UN CICLO DEL SEMAFORO, TOMADO AL AZAR, SE CONGESTIONE EL TRANSITO POR EXCEDERSE LA CAPACIDAD DEL CARRIL?

$$P[X>3] = ?$$

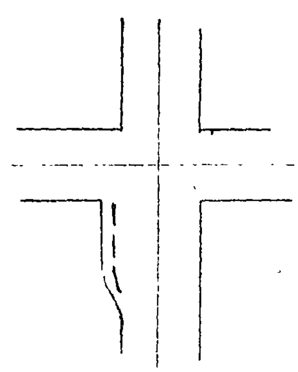
SI $A=\{X>3\}$, $\bar{A}=\{X\leq 3\}$

$P(\bar{A}) = 1-P(A)$ O $P(A)= 1-P(\bar{A})$. CON $\lambda=6$,

$$P(\bar{A}) = P[X\leq 3] = \sum_{x=0}^{x=3} f_X(x) = \sum_{x=0}^{x=3} \frac{e^{-6}6^x}{x!}$$

$$P(\bar{A}) = e^{-6}(1 + 6 + \frac{6^2}{2} + \frac{6^3}{6}) = 61e^{-6} = 0.152$$

$$P[A] = P[X>3] = 1-0.152 = 0.848$$



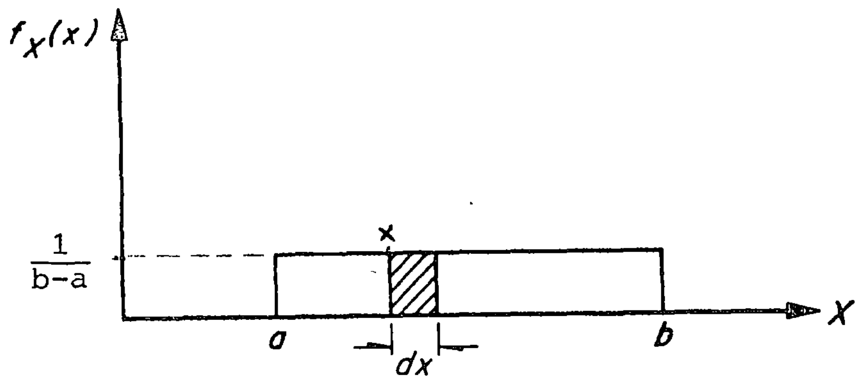
VARIABLES ALEATORIAS CONTINUAS

DISTRIBUCION UNIFORME

SE DICE QUE UNA VARIABLE ALEATORIA CONTINUA, X, TIENE *DISTRIBUCION UNIFORME* ENTRE X = a Y X = b(b>a) SI

$$f_X(x) = \text{CONSTANTE} = \frac{1}{b-a} \quad , \qquad a \leq X \leq b$$

LO QUE SIGNIFICA QUE LA PROBABILIDAD DE OBTENER UN VALOR ENTRE x Y x + dx ES LA MISMA PARA CUALQUIER x COMPRENDIDA ENTRE a Y b. LA GRAFICA DE DICHA DISTRIBUCION ES



Distribución uniforme de una variable aleatoria continua

LA ESPERANZA Y LA VARIANCIA DE LA DISTRIBUCION UNIFORME SE CALCULAN DE LA SIGUIENTE MANERA:

$$E[X] = \int_a^b x \frac{1}{b-a}\, dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = (b+a)/2$$

$$\sigma^2(X) = \int_a^b (x-E[X])^2 \frac{1}{b-a}\, dx = \int_a^b \frac{x^2}{b-a}\, dx + \int_a^b \frac{(E[X])^2}{b-a}\, dx -$$

$$- \int_a^b \frac{2xE[X]}{a-b}\, dx$$

$$= \left[\frac{x^3}{3(b-a)}\right]_a^b + \left[\frac{(E[X])^2}{b-a}\; x\right]_a^b - \left[\frac{2\;E[X]}{b-a}\;\frac{x^2}{2}\right]_a^b =$$

$$= \frac{b^3-a^3}{3(b-a)} + (E[X])^2 - E[X]\,(b+a) = \frac{(b-a)^2}{12}$$

DISTRIBUCION NORMAL

UNA DE LAS DISTRIBUCIONES DE VARIABLES ALEATORIAS CONTINUAS MAS UTIL ES LA *DISTRIBUCION NORMAL* O DE *GAUSS*, DEFINIDA POR LA ECUACION

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

DONDE μ ES LA MEDIA Y σ LA DESVIACION ESTANDAR DE X.

SI SE HACE LA TRANSFORMACION

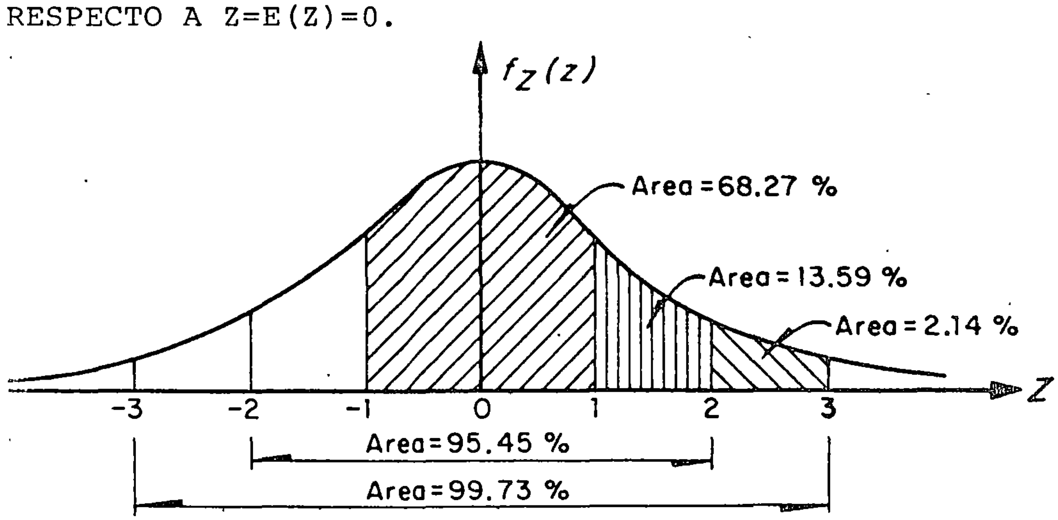
$$Z = (X-\mu)/\sigma$$

ENTONCES LA ECUACION ANTERIOR SE REDUCE A LA LLAMADA *FORMA ESTANDAR*, CUYA ECUACION ES

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

EN ESTE CASO LA VARIABLE ALEATORIA Z TIENE DISTRIBUCION NORMAL CON *MEDIA IGUAL A CERO Y VARIANCIA IGUAL A UNO.*

EXISTEN TABLAS PARA CALCULAR LAS PROBABILIDADES DE UNA VARIABLE ASOCIADA A UNA DISTRIBUCION NORMAL ESTANDAR. EN LA SIGUIENTE FIGURA SE MUESTRA LA FORMA DE CAMPANA DE ESTA DISTRIBUCION, OBSERVANDOSE LA SIMETRIA RESPECTO A Z=E(Z)=0.



Distribución normal de una variable aleatoria continua

LA UTILIDAD DE LA DISTRIBUCION NORMAL ESTANDAR RADICA EN QUE

$$P\left[x_1 \leq X \leq x_2\right] = \int_{x_1}^{x_2} f_X(x)\,dx = P\left[z_1 \leq Z \leq z_2\right] = \int_{z_1}^{z_2} f_Z(z)\,dz$$

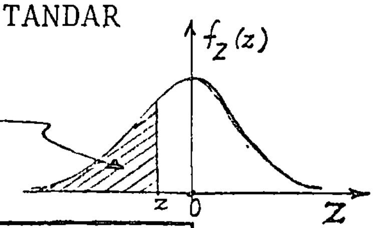
DONDE

$$z_1 = \frac{z_{1-\mu}}{\sigma} \quad Y \quad z_2 = \frac{x_{2-\mu}}{\sigma}$$

TABLA 3

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

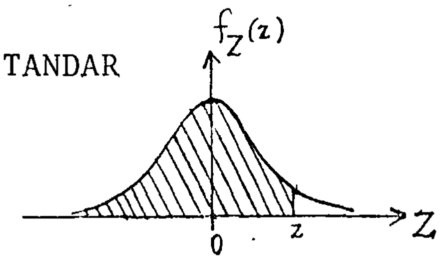


z	0	1	2	3	4	5	6	7	8	9
−3.	.0013									
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	0154	.0150	.0146	.0143
−2.0	.0227	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1921	.1894	.1867
−.7	.2420	.2389	.2358	.2326	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

TABLA 3 (continuación)

FUNCION DE DISTRIBUCION NORMAL ESTANDAR

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$



Z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8079	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	9987									

EJEMPLO

COMO RESULTADO DE UNA LARGA SERIE DE EXPERIMENTOS PROBANDO A COMPRESION SIMPLE CILINDROS DE CONCRETO, SE HA ESTIMADO QUE LA ESPERANZA DE LA RESISTENCIA ES DE 240 KG/CM^2 Y LA DESVIACION ESTANDAR DE 30 KG/CM^2.

A) ¿CUAL ES LA PROBABILIDAD DE QUE OTRO CILINDRO TOMADO AL AZAR RESISTA MENOS DE 240 KG/CM^2?

B. ¿CUAL ES LA PROBABILIDAD DE QUE RESISTA MAS DE 330 KG/CM^2?

C) ¿CUAL ES LA PROBABILIDAD DE QUE SU RESISTENCIA ESTE EN EL INTERVALO DE 210 A 240 KG/CM^2?

SUPONGASE QUE LA DISTRIBUCION DE PROBABILIDADES ES NORMAL.

SOLUCION

A) PARA EMPLEAR LAS TABLAS DE DISTRIBUCION NORMAL ES NECESARIO ESTANDARIZAR LA VARIABLE X, EMPLEANDO $\mu=240$ Y $\sigma=30$, CON $x_1=240$:

$$z_1 = \frac{240 - 240}{30} = 0$$

RECURRIENDO A LA TABLA DE LA DISTRIBUCION NORMAL SE OBTIENE

$$P[X \leq 240] = P[Z \leq 0] = 0.5$$

O SEA, LA PROBABILIDAD QUE CORRESPONDE AL AREA SOMBREADA DE LA SIGUIENTE FIGURA:

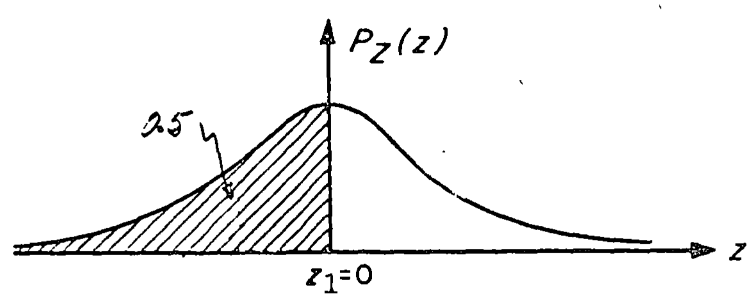


Fig 16. Distribución normal correspondiente al inciso c del ejemplo

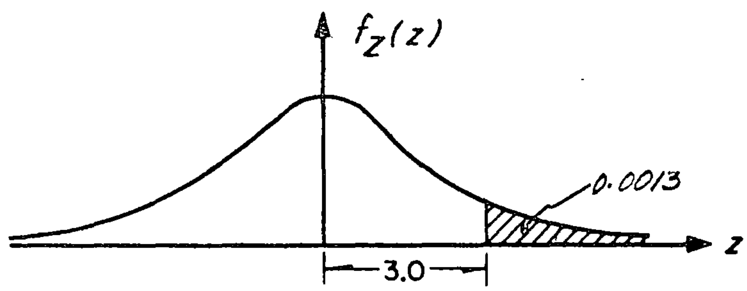
B) EL VALOR ESTANDARIZADO DE LA VARIABLE, PARA x_1=330 KG/CM^2, ES

$$z_1 = \frac{330 - 240}{30} = 3$$

POR LO QUE

$$P[X \geq 330] = P[Z \geq 3] = 1-0.9987 = 0.0013$$

QUE ES EL AREA SOMBREADA DE LA SIGUIENTE FIGURA:



Distribución normal correspondiente al inciso b del ejemplo

C) LOS VALORES ESTANDARIZADOS DE LA VARIABLE, PARA x_1=210 Y x_2=240 SON:

$$z_1 = \frac{210 - 240}{30} = -1$$

$$z_2 = \frac{240 - 240}{30} = 0$$

POR LO QUE

$$P[210 \leq X \leq 240] = P[-1 \leq Z \leq 0] = 0.3413$$

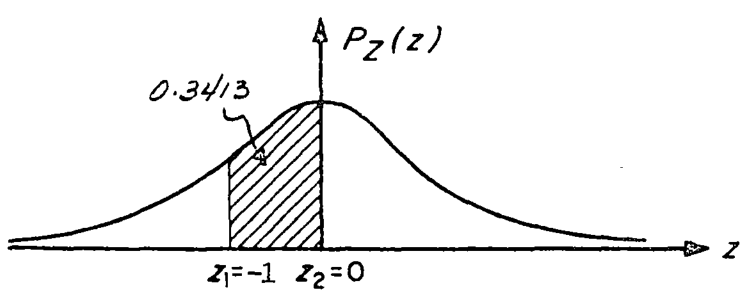


Fig 16. Distribución normal correspondiente al inciso c del ejemplo

TEOREMA CENTRAL DEL LIMITE

IDENTICAS

SEAN LAS VARIABLES ALEATORIAS $X_1, X_2, \ldots, X_k$, CON DENSIDADES DE PROBABILIDADES (ARBITRARIAS), CUYA SUMA SE DENOTARA COMO W, ES DECIR

$$W = X_1 + X_2 + \ldots + X_k$$

ES POSIBLE DEMOSTRAR EL TEOREMA DENOMINADO *TEOREMA CENTRAL DEL LIMITE*, CUYO ENUNCIADO INDICA QUE CONFORME AUMENTA EL NUMERO DE VARIABLES INVOLUCRADAS EN LA SUMA ANTERIOR (AL AUMENTAR k), LA DENSIDAD DE PROBABILIDADES DE W TIENDE A SER LA DISTRIBUCION NORMAL. ADEMAS SE PUEDE DEMOSTRAR QUE SI TODAS LAS VARIABLES $X_1, X_2, \ldots, X_k$ TIENEN DISTRIBUCION NORMAL, ENTONCES, RIGUROSAMENTE, W TAMBIEN LA TIENE, INDEPENDIENTEMENTE DEL NUMERO DE VARIABLES QUE APAREZCAN EN LA SUMA.

A PARTIR DEL TEOREMA DEL LIMITE CENTRAL SE DEMUESTRA QUE LA DISTRIBUCION DE BERNOULLI SE PUEDE APROXIMAR MEDIANTE LA NORMAL CUANDO EL NUMERO DE REPETICIONES DEL EXPERIMENTO ES GRANDE (30 O MAS), CON LO CUAL SE LOGRA UN AHORRO CONSIDERABLE DE LABOR NUMERICA EN LA SOLUCION DE ALGUNOS PROBLEMAS. PARA MEJORAR ESTA APROXIMACION, CONVIENE EFECTUAR UNA CORRECCION POR CONTINUIDAD, LA CUAL SE JUSTIFICA POR USAR UNA DISTRIBUCION CONTINUA EN VEZ DE UNA DISCRETA, SUMANDO O RESTANDO, SEGUN SEA EL CASO, 0.5 AL VALOR DE X QUE SE USE. POR EJEMPLO, SI SE DESEA CUANTIFICAR LA PROBABILIDAD DE QUE DE 2000 ENSAYES SE LOGREN DE 3 A 6 EXITOS, LOS LIMITES REALES QUE SE DEBEN USAR AL APLICAR LA DISTRIBUCION CONTINUA SON $x_1=2.5$ Y $x_2=6.5$.

EJEMPLO

SI LA PROBABILIDAD DE QUE FALLE UNA VARILLA DE ACERO AL APLICARLE CIERTA CARGA ES DE 0.001, DETERMINAR LA PROBABILIDAD DE QUE EN 2000 VARILLAS PROBADAS FALLEN MAS DE DOS.

USANDO LA DISTRIBUCION DE BERNOULLI SE OBTIENE

$$P[X > 2] = 1 - P[X \leq 2] = 1 - (P[X = 0] + P[X = 1] + P[X = 2]) =$$

$$= 1 - (\frac{2\ 000!}{2\ 000!\ 0!}(0.001)^0(0.999)^{2\ 000} + \frac{2\ 000!}{1\ 999!\ 1!}(0.001)^1(0.999)^{1\ 999} +$$

$$+ \frac{2\ 000!}{1\ 998!\ 2!}(0.001)^2(0.999)^{1\ 998}) = 0.3255$$

LOS CALCULOS NECESARIOS PARA OBTENER LA SOLUCION SON BASTANTE MAS TEDIOSOS QUE LOS QUE DEBEN EFECTUARSE APROVECHANDO QUE EL NUMERO DE REPETICIONES DEL EXPERIMENTO ES GRANDE, A FIN DE UTILIZAR LA DISTRIBUCION NORMAL. EN ESTAS CIRCUNSTANCIAS, LA PROBABILIDAD DE QUE $X \leq 2$ EN EL CASO DISCRETO, EQUIVALE A LA DE QUE $X \leq 2.5$ EN EL CONTINUO; ASI

$$\mu = np = 2\ 000 \times 0.001 = 2$$

$$\sigma = \sqrt{npq} = \sqrt{2\ 000 \times 0.001 \times 0.999} = 1.41$$

$$P[X \leq 2.5] = P[Z \leq \frac{2.5 - 2}{1.41}] = P[Z \leq 0.355] = 0.6387$$

DE DONDE

$$P[x > 2.5] = 1 - P[X \leq 2.5] = 1 - 0.6387 = 0.3613$$

INFERENCIA ESTADISTICA

Por: M en I Augusto Villarreal Aranda*

1. Introducción

La parte de la estadística que proporciona las reglas para inferir ciertas características de una población a partir de muestras extraídas de ella, junto con indicaciones probabilísticas de la veracidad de tales inferencias, se llama *inferencia estadística*.

En la inferencia estadística se estudian las relaciones existentes entre una población, las muestras obtenidas de ella, y las técnicas para estimar parámetros, tales como la media y la varianza, o bien para determinar si las diferencias entre dos muestras son debidas al azar, etc.

2. Distribuciones muestrales

Si se consideran todas las muestras posibles de tamaño

* *Secretario Académico*, División de Estudios Superiores, Facultad de Ingeniería, UNAM y *Profesor investigador*, Instituto de Ingeniería, UNAM

n que pueden extraerse de una población, y para cada una se calcula el valor del promedio aritmético, este seguramente variará de una muestra a otra, ya que depende de los valores de los datos que se hayan obtenido en cada muestra. Por lo tanto, el promedio aritmético es en sí una variable aleatoria, como también lo son, por la misma razón, el rango y la variancia de la muestra.

A todo elemento que es función de los valores de los datos que se tienen en una muestra se le denomina *estadística;* toda estadística es, entonces, una variable aleatoria cuya distribución de probabilidades se conoce como *distribución muestral.* Si, por ejemplo, la estadística considerada es la variancia de la muestra, su densidad de probabilidades se llama *distribución muestral de la variancia.*

En forma similar se pueden obtener las distribuciones muestrales de la desviación estándar, del rango, etc., cada una de las cuales tendrá sus propios parámetros, lo que permite hablar de la media y la desviación estándar de la variancia, etc.

3. Muestreo con y sin remplazo

Cuando se efectúa un muestreo en una población de tal manera que cada elemento de la misma se pueda escoger más de una vez, se dice que el muestreo es *con remplazo*; en caso contrario, el muestreo es *sin remplazo*. Si de una urna se quiere extraer una muestra de bolas de colores, se puede proceder de dos maneras: se saca al azar una bola, se anota su color y se regresa a la urna antes de obtener otra, y así sucesivamente; en este caso el muestreo es *con remplazo*. La segunda forma consiste en extraer al azar todas las bolas que constituyen la muestra sin regresarlas a la urna, siendo entonces un muestreo *sin remplazo*.

4. Distribucion muestral del promedio aritmético

Supóngase que se extraen sin remplazo todas las muestras posibles de tamaño n de una población finita de tamaño $N_p > n$. Si la media y la desviación estándar de la distribución muestral del promedio aritmético se denotan con $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$, y la media y la desviación estándar de la población con μ y σ, respectivamente, entonces es posible demostrar que se cumplen las siguientes ecuaciones

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Además, si la población es infinita (o el muestreo es con remplazo), los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

puesto que

$$\lim_{N_p \to \infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{\sigma}{\sqrt{n}}$$

Para valores grandes de $n (n \geq 30)$ se demuestra, empleando el teorema del límite central, que la distribución muestral del promedio aritmético es aproximadamente una distribución normal con media $\mu_{\bar{X}}$ y desviación estándar $\sigma_{\bar{X}}$, independientemente de cuál sea la densidad de probabilidades de X, la variable aleatoria asociada a la población. Si esta variable tiene distribución normal, la distribución muestral del promedio aritmético también es normal, aun para valores pequeños de n $(n < 30)$.

Ejemplo 4.1

Supóngase que se tiene una población finita formada por los datos 1,2,3,4,5. Se desea conocer la media y la desviación estándar de la distribución muestral del promedio aritmético, considerando las muestras de tamaño 3 obtenidas sin remplazo.

Primer procedimiento.

Siendo la población finita y el muestreo sin remplazo, es posible obtener la distribución muestral correspondiente para calcular después sus parámetros, considerando que el número total de muestras distintas de tamaño 3 que pueden obtenerse a partir de una población de 5 elementos es

$$\frac{5!}{3!(5-3)!} = 10$$

Dichas muestras son las siguientes, junto con sus promedios aritméticos correspondientes:

	$\bar{X}_i$		$\bar{X}_i$
1, 2, 3	6/3	3, 4, 5	12/3
1, 2, 4	7/3	3, 4, 1	8/3
1, 2, 5	8/3	4, 5, 1	10/3
2, 3, 4	9/3	4, 5, 2	11/3
2, 3, 5	10/3	5, 1, 3	9/3

Para calcular la media y la desviación estándar, se emplea la siguiente tabla

$\bar{X}_i$	6/3	7/3	8/3	8/3	9/3	9/3	10/3	10/3	11/3	12/3
$\bar{X}_i^2$	36/9	49/9	64/9	64/9	81/9	81/9	100/9	100/9	121/9	144/9

$$\sum_{i=1}^{10} \bar{X}_i = 90/3 \qquad\qquad \sum_{i=1}^{10} \bar{X}_i^2 = 840/9$$

$$\mu_{\bar{X}} = \bar{\bar{X}} = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i = \frac{1}{10} \cdot \frac{90}{3} = 3$$

$$\sigma_{\bar{X}}^2 = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i^2 - \bar{\bar{X}}^2 = \frac{1}{10} \cdot \frac{840}{9} - (3)^2 =$$

$$= 9.333 - 9.000 = 0.333 \Rightarrow \sigma_{\bar{X}} = \sqrt{0.333} = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Segundo procedimiento.

Por tratarse de una población finita, se verifica que

$$\mu_{\bar{X}} = \mu \qquad \text{y} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en donde $N_p = 5$, $n = 3$ y $\mu = 3$.

El valor de σ^2 de la población es

$$\sigma^2 = \frac{1+4+9+16+25}{5} - (3)^2 = \frac{55}{5} - 9 = 11-9 = 2$$

Por lo tanto, $\sigma = \sqrt{2} = 1.4145$ y

$$\sigma_{\bar{X}} = \frac{1.4145}{\sqrt{3}} \sqrt{\frac{5-3}{5-1}} = (0.8164)(0.7071) = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Comparando los resultados, se puede observar que ambos procedimientos conducen a la obtención de los mismos valores de $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$ para la distribución muestral del promedio aritmético.

Ejemplo 4.2

En una bodega se tienen cinco mil varillas de acero; el valor medio del peso, X, de cada varilla es de 5.02 kg, y la desviación estándar 0.3 kg. Hallar la probabilidad de que una muestra de cien varillas, escogida al azar, tenga un peso total

a. entre 496 y 500 kg

b. de más de 510 kg.

Para la distribución muestral del promedio, se tiene que $\mu_{\bar{X}} = \mu = 5.02$ kg y, por tratarse de una población finita,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \frac{0.30}{\sqrt{100}} \sqrt{\frac{5000 - 100}{5000 - 1}} = 0.027$$

a. El peso total de la muestra estará entre 496 y 500 kg si el peso promedio de las cien varillas se encuentra entre 4.96 y 5.00 kg. Puesto que la muestra es mayor de 30 elementos se puede considerar como aproximadamente normal a la distribución muestral, y los valores estándar correspondientes a $\bar{X} = 4.96$ y a $\bar{X} = 5.00$ se obtienen mediante la transformación

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

es decir,

$$Z_1 = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$Z_2 = \frac{5.00 - 5.02}{0.027} = -0.74$$

En la fig 4.1 se puede apreciar que

$$P[496 \leqslant X \leqslant 500] = P[-2.22 \leqslant Z \leqslant -0.74] =$$

$$= P[-2.22 \leqslant Z \leqslant 0] - P[-0.74 \leqslant Z \leqslant 0]$$

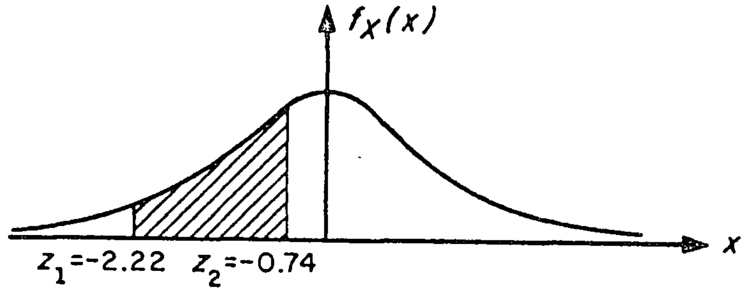


Fig 4.1 Distribución normal correspondiente al ejemplo

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y Z queda finalmente

$$P[496 \leq X \leq 500] = 0.4868 - 0.2704 = 0.2164$$

b. El peso total de la muestra excederá de 510 kg si el peso promedio de las cien varillas pasa de 5.10 kg.

Estandarizando dicho valor, queda

$$Z_3 = \frac{5.10 - 5.02}{0.027} = 2.96$$

Calculando el área bajo la curva normal a la derecha de este valor (fig 4.2), se tiene que

$$P[X \geq 510] = P[Z \geq 2.96] = P[Z \geq 0] - P[0 \leq Z \leq 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

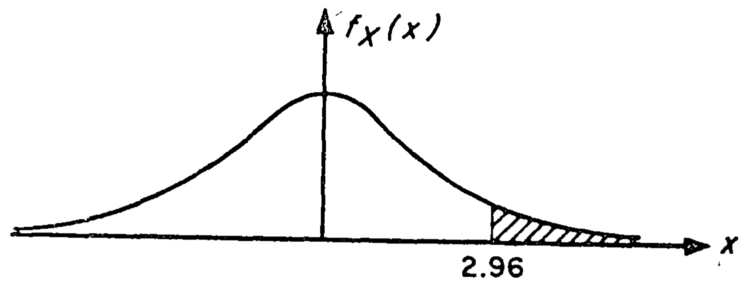


Fig 4.2 Distribución normal correspondiente al ejemplo

5. Distribución muestral de diferencias de promedios aritméticos

Con frecuencia se presenta el caso en el que se tienen datos de dos poblaciones con variables aleatorias asociadas X y Y, respectivamente, surgiendo la duda de si estas se pueden considerar como una sola, es decir, si X = Y. Para probar estadísticamente esta hipótesis (como se verá más adelante), es necesario obtener las distribuciones muestrales de la diferencia de los promedios y de las variancias de las muestras de ambas variables.

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de muestras aleatorias de tamaño n_X y n_Y de dos poblaciones con características X y Y, respectivamente. Se puede demostrar que la distribución muestral de la diferencia de los promedios correspondientes a poblaciones infinitas con medias μ_X y μ_Y y desviaciones estándar σ_X y σ_Y, tiene los siguientes parámetros:

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X} + \frac{\sigma^2_Y}{n_Y}}$$

si las muestras son independientes.

Esta distribución también es aplicable a poblaciones finitas si el muestreo es con remplazo. Para el caso de poblaciones finitas en las cuales el muestreo se hace sin remplazo, los parámetros de la distribución muestral de la diferencia de los promedios aritméticos son

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} = \mu_X - \mu_Y$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}} = \sqrt{\frac{\sigma^2_X}{n_X}\,\frac{N_X - n_X}{N_X - 1} + \frac{\sigma^2_X}{n_Y}\,\frac{N_Y - n_Y}{N_Y - 1}}$$

suponiendo que las muestras sean independientes.

Ejemplo 5.1

Considérese que de una población X se obtienen tres muestras posibles, cuyos correspondientes promedios aritméticos son 3, 7 y 8. De otra población Y se extraen dos muestras posibles, con promedios 2 y 4, respectivamente. Se deben obtener los parámetros de la distribución muestral de las diferencias de los promedios aritméticos.

Primer procedimiento

Todas las posibles diferencias de promedios aritméticos de X con los de Y serían

$$\begin{matrix} 3-2 & 7-2 & 8-2 \\ 3-4 & 7-4 & 8-4 \end{matrix} \Longrightarrow \begin{matrix} 1 & 5 & 6 \\ -1 & 3 & 4 \end{matrix}$$

Es decir,

$$\mu_{\bar{X}-\bar{Y}} = \frac{-1+1+3+4+5+6}{6} = \frac{18}{6} = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{(-1-3)^2 + (1-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (6-3)^2}{6} =$$

$$= \frac{34}{6} = \frac{17}{3}$$

Segundo procedimiento

Se sabe que

$$\mu_{\bar{X}-\bar{Y}} = \mu_{\bar{X}} - \mu_{\bar{Y}} \quad ; \quad \sigma^2_{\bar{X}-\bar{Y}} = \sigma^2_{\bar{X}} + \sigma^2_{\bar{Y}}$$

Por ello,

$$\mu_{\bar{X}} = \frac{3+7+8}{3} = \frac{18}{3} = 6$$

$$\mu_{\bar{Y}} = \frac{2+4}{2} = \frac{6}{2} = 3$$

$$\sigma^2_{\bar{X}} = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

$$\sigma^2_{\bar{Y}} = \frac{(2-3)^2 + (4-3)^2}{2} = \frac{2}{2} = 1$$

$$\mu_{\bar{X}-\bar{Y}} = 6 - 3 = 3$$

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{14}{3} + 1 = \frac{17}{3}$$

Se observa que ambos procedimientos conducen a los mismos resultados.

Ejemplo 5.2

Las varillas de acero que fabrica una compañía A tienen un peso medio de 6.5 kg y una desviación estándar de 0.4, en tanto que las producidas por una empresa B tienen un peso medio de 6.3 kg y una desviación estándar de 0.3 kg. Si se toman muestras aleatorias de 100 varillas de cada fábrica, ¿cuál es la probabilidad de que las de la compañía A tengan un peso promedio de por lo menos

a. 0.35 kg

b. 0.10 kg

mayor que el de la compañía B?

Se puede suponer en este caso que las distribuciones muestrales involucradas son normales, en virtud de que el tamaño de ambas muestras es mayor de 30 elementos. También se puede suponer que ambas poblaciones son infinitas, y siendo $\bar{X}_A$ y $\bar{X}_B$ los pesos promedios de las muestras de las fábricas A y B, respectivamente, entonces

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 6.5 - 6.3 = 0.20 \text{ kg}$$

$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} = 0.05 \text{ kg}$$

La variable estandarizada de la diferencia de los promedios es

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_{\bar{X}_A - \bar{X}_B}}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0.20}{0.05}$$

a. Estandarizando la diferencia de 0.35 kg se llega a

$$Z_1 = \frac{0.35 - 0.20}{0.05} = \frac{0.15}{0.05} = 3$$

La probabilidad deseada es el área bajo la curva normal a la derecha de $Z = 3$, es decir

$$P\left[\bar{X}_A \geqslant \bar{X}_B + 0.35\right] = P\left[Z \geqslant 3\right] = 0.500 - 0.4987 = 0.0013$$

b. Al estandarizar la diferencia de 0.10 kg, la variable Z resulta

$$Z_2 = \frac{0.10 - 0.20}{0.05} = \frac{-0.1}{0.05} = -2$$

La probabilidad requerida es el área bajo la curva normal a la derecha de $Z = -2$, es decir

$$P\left[\bar{X}_A \geqslant \bar{X}_B + 0.10\right] = P\left[Z \geqslant -2\right] = 0.5 + 0.4772 = 0.9772$$

6. Teoría estadística de la estimación

En la práctica profesional a menudo resulta necesario inferir información acerca de una población mediante el uso de muestras extraídas de ella; una parte básica de dicha inferencia consiste en *estimar* los valores de los parámetros de la población (media, variancia, etc.) a partir de las estadísticas correspondientes de la muestra, como se explica a continuación.

7. Estimadores puntuales. Clasificación

Si un estimador de un parámetro de la población consiste en un solo valor de una estadística, se le conoce como *estimador puntual* del parámetro.

Cuando la media de la distribución muestral de una estadística es igual al parámetro que se está estimando de la población, entonces la estadística se conoce como *estimador insesgado* del parámetro; si no sucede así, entonces se denomina *estimador sesgado*. Ambos estimadores son puntuales, y sus valores correspondientes se llaman estimaciones insesgadas o sesgadas, respectivamente. Dicho de otra manera, si S es una estadística cuya distribución muestral tiene media μ_S, y el parámetro correspondiente de la población es θ, se dice que S es un estimador insesgado de θ si

$$\mu_S = \theta$$

Por otra parte, si la estadística S_n de la muestra tiende a ser igual al parámetro θ de la población a medida que se hace más grande el tamaño de la muestra, entonces la estadística recibe el nombre de *estimador consistente* del parámetro.

Empleando símbolos, si

$$\lim_{n \to \infty} S_n = \theta$$

resulta que la estadística S_n es un estimador consistente. Por ejemplo, el promedio aritmético es un estimador insesgado y consistente de la media, y la variancia de la muestra es un estimador sesgado y consistente de la variancia de la población.

Si las distribuciones muestrales de varias estadísticas tienen el mismo valor de la media, se dice que la estadística que cuenta con la menor variancia es un *estimador eficiente* de dicha media, en tanto que las estadísticas restantes se conocen como *estimadores ineficientes* del parámetro.

Por ejemplo, las distribuciones muestrales del promedio aritmético y de la mediana cuentan con medias que son, en ambos casos, iguales a la media de la población. Sin embargo, la variancia de la distribución muestral del promedio aritmético es menor que la de la distribución de la mediana, por lo que el promedio aritmético obtenido de una muestra aleatoria proporciona un estimador eficiente de la media de la población, en tanto que la mediana obtenida de la muestra proporciona un estimador ineficiente de dicho parámetro.

8. Estimación de intervalos de confianza para los parámetros de una población

La estimación de un parámetro de una población mediante un par de números entre los cuales se encuentra, con cierta probabilidad, el valor de dicho parámetro, se llama estimación del intervalo del mismo.

Sea S una estadística obtenida de una muestra de tamaño n para estimar el valor del parámetro θ, y sea σ_S la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1 - \alpha$, de que el valor de θ se localice en el intervalo de $S - z_c\,\sigma_S$ a $S + z_c\,\sigma_S$, donde z_c es una constante, se escribe en la forma

$$P[S - z_c\,\sigma_S \leqslant \theta \leqslant S + z_c\,\sigma_S] = 1 - \alpha$$

Si se fija el valor de $1 - \alpha$, se puede obtener el valor de z_c necesario para que se satisfaga la ecuación anterior, con lo cual queda definido el *intervalo de confianza* del parámetro θ, $(S - z_c\,\sigma_S,\ S + z_c\,\sigma_S)$, correspondiente al nivel de *confianza* $1 - \alpha$.

La constante z_c que fija el intervalo de confianza se conoce como *valor crítico*. Si la distribución de S es normal, el valor de z_c correspondiente a uno de α se obtiene de la tabla de áreas bajo la curva normal o de la tabla 8.1 siguiente.

TABLA 8.1 VALORES DE z_c PARA DISTINTOS NIVELES DE CONFIANZA

Nivel de confianza, en porcentaje	z_c
99.73	3.00
99.00	2.58
98.00	2.33
96.00	2.05
95.45	2.00
95.00	1.96
90.00	1.64
80.00	1.28
68.27	1.00
50.00	0.674

Ejemplo 8.1

Sea el promedio aritmético $\bar{X}$ una estadística con distribución normal. Las probabilidades o niveles de confianza de que $\mu_{\bar{X}}$ (o μ de la población) se encuentre localizada entre los límites $\bar{X} \pm \sigma_{\bar{X}}$, $\bar{X} \pm 2\ \sigma_{\bar{X}}$ y $\bar{X} \pm 3\ \sigma_{\bar{X}}$ son 68.26, 95.44 y 99.73%, respectivamente, obteniéndose dichos valores de la tabla de áreas bajo la curva normal. Lo anterior significa que el intervalo $\bar{X} \pm 3\ \sigma_{\bar{X}}$ contendrá a $\mu_{\bar{X}}$ en el 99.73 por ciento de las muestras de tamaño n, por lo que los intervalos de confianza de 68.26, 95.44 y 99.73 por ciento para estimar a μ son $(\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}})$ $(\bar{X} - 2\ \sigma_{\bar{X}}, \bar{X} + 2\ \sigma_{\bar{X}})$ y $(\bar{X} - 3\ \sigma_{\bar{X}}, \bar{X} + 3\ \sigma_{\bar{X}})$, lo cual se aprecia en la *fig* 8.1 siguiente.

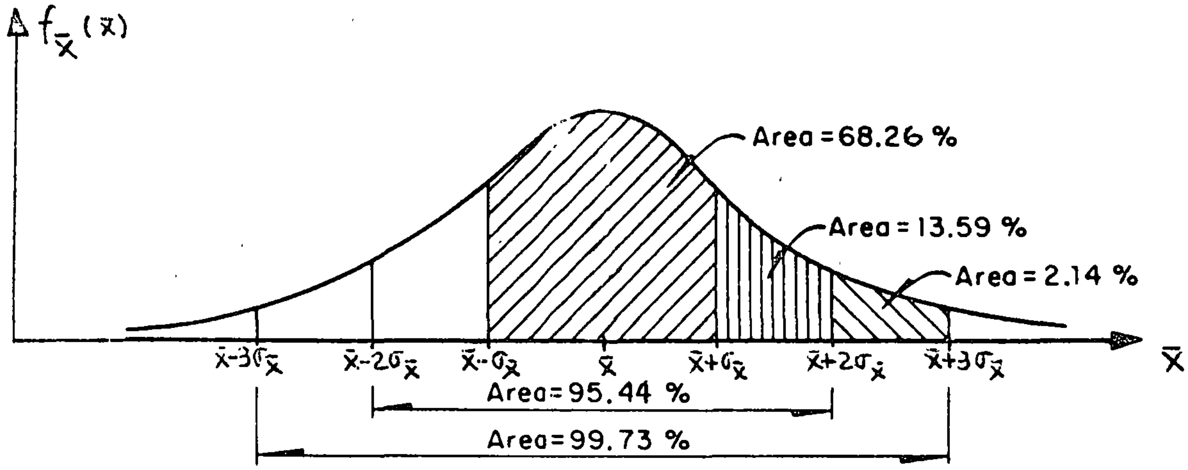


Fig 8.1

9. Estimación de intervalos de confianza para la media

Los límites de confianza para la media de una población con variable aleatoria X asociada están dados por

$$\bar{X} \pm z_c\, \sigma_{\bar{X}}$$

en donde z_c depende del nivel de confianza deseado. Si $\bar{X}$ tiene distribución normal, z_c puede obtenerse en forma directa de la tabla 8.1. Por ejemplo, los límites de confianza de 95 y 99 por ciento para estimar la media, μ, de la población son $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\ \sigma_{\bar{X}}$, respectivamente. Al obtener estos límites hay que usar el valor calculado de $\bar{X}$ para la muestra correspondiente.

Entonces, los límites de confianza para la media de la población quedan dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}}$$

en caso de que el muestreo se haga a partir de una población infinita o de que se efectúe con remplazo a partir de una población finita, o por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

si el muestreo es sin remplazo a partir de una población finita de tamaño N_p.

Ejemplo 9.1

Las mediciones de los diámetros de una muestra aleatoria de 100 tubos de albañal mostraron una media de 32 cm y una desviación estándar de 2 cm. Obténganse los límites de confianza de

a. 95 por ciento

b. 97 por ciento

para el diámetro medio de todos los tubos.

a. De la tabla 8.1, los límites de confianza del 95 por ciento son

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 32 \pm 1.96(2/\sqrt{100}) = 32 \pm 0.392 \text{ cm}$$

o sea 31.608 y 32.392, en donde se ha empleado el valor de S_X para estimar el de σ de la población, puesto que la muestra es suficientemente grande (mayor de 30 elementos). Esto significa que con una probabilidad de 95 por ciento, el valor de μ_X se encuentra entre 31.608 y 32.392 cm.

b. Si $Z = z_c$ es tal que el área bajo la curva normal a la derecha de z_c es el 1.5 por ciento del área total, entonces el área entre 0 y z_c es $0.5 - 0.015 = 0.485$, por lo que de la tabla de áreas bajo la curva normal se obtiene $z_c = 2.17$. Por lo tanto, los límites de confianza del 97 por ciento son:

$$\bar{X} \pm 2.17\sigma/\sqrt{n} = 32 \pm 2.17(2/\sqrt{100}) = 32 \pm 0.434 \text{ cm}$$

y el intervalo de confianza respectivo es (31.566 cm, 32.434 cm).

Ejemplo 9.2

Una muestra aleatoria de 50 calificaciones de cierto examen de admisión tiene un promedio aritmético de 72 puntos, con desviación estándar igual a 10. Si el examen se aplicó a 1018 personas, obtener

a. El intervalo de confianza del 95% para la media del total de calificaciones.

b. El tamaño de muestra necesario para que el error en la estimación de la media no exceda de 2 puntos, considerando el mismo nivel de confianza.

c. El nivel de confianza para el cual la media de la población sea 72 ± 1 puntos.

a. Si se estima a σ de la población con S_X de la muestra y se considera que la población es finita, los límites de confianza son, puesto que $\bar{X} = 72$, $Z_c = 1.96$, $S_X = 10$, $N_p = 1018$ y $n = 50$,

$$72 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm 1.96\ (1.4142)\ (0.9755)$$

$$72 \pm 2.704$$

y el intervalo de confianza respectivo es

$$(69.296,\quad 74.704)$$

b. Puesto que el error en la estimación de la media es, para población finita,

$$\text{Error en la estimación} = Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en este caso se tendría

$$Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} < 2$$

o sea, para un nivel de confianza de 95%,

$$1.96 \frac{10}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

$$\frac{19.6}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

Elevando al cuadrado la desigualdad, queda

$$\frac{384.16}{n} \quad \frac{1018 - n}{1017} < 4$$

o sea

$$87.85 < n$$

Por lo cual, se requieren al menos 88 elementos en la muestra para que el error en la estimación no exceda de 2 puntos, para $1 - \alpha = 0.95$.

c. Los límites de confianza son, en este caso

$$72 \pm Z_c \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm Z_c \ (1.4142)(0.9755)$$

o sea

$$72 \pm 1.3795 \ Z_c$$

Puesto que se desea que el valor de la media sea 72 ± 1 puntos, se verifica que

$$1 = 1.3795 \ Z_c$$

Es decir

$$Z_c = \frac{1}{1.3795} = 0.725$$

El área bajo la curva normal estándar entre 0 y $Z_c = 0.725$ es, por interpolación lineal, igual a 0.2657. Por lo tanto, el nivel de confianza es igual al doble del área anterior, es decir, 2(0.2657) = 0.5314 (o 53.14%), tal como se muestra en la *fig* 9.1.

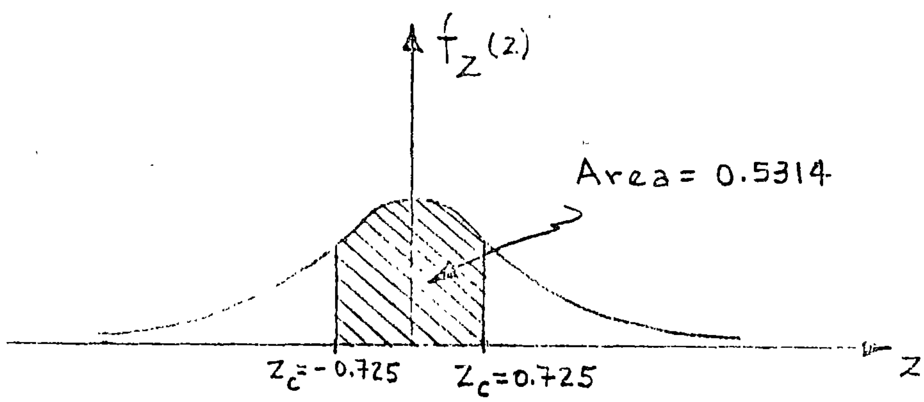


Fig 9.1

10. Intervalos de confianza para diferencias de medias

Los límites de confianza para la diferencia de las medias cuando las poblaciones X y Y son infinitas, o cuando el muestreo se realiza con remplazo de poblaciones finitas, se encuentran dados por

$$\bar{X} - \bar{Y} \pm Z_c\, \sigma_{\bar{X} - \bar{Y}} = \bar{X} - \bar{Y} \pm z_c \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

en donde $\bar{X}$, n_X y $\bar{Y}$, n_Y son los respectivos promedios aritméticos y tamaños de las dos muestras extraídas de las poblaciones, y σ_X y σ_Y las desviaciones estándar de estas últimas.

En el caso de que las poblaciones X y Y sean finitas y el muestreo sin remplazo, los límites de confianza son

$$\bar{X} - \bar{Y} \pm Z_c \sigma_{\bar{X}-\bar{Y}} = \bar{X} - \bar{Y} \pm Z_c \sqrt{\frac{\sigma_X^2}{n_X} \frac{N_X - n_X}{N_X - 1} + \frac{\sigma_Y^2}{n_Y} \frac{N_Y - n_Y}{N_Y - 1}}$$

en donde N_X y N_Y son los tamaños de las poblaciones X y Y, respectivamente.

Las dos ecuaciones anteriores son válidas únicamente si las muestras aleatorias seleccionadas son independientes.

Ejemplo 10.1

Para el ejemplo de las varillas tratado anteriormente (5.2), encontrar el intervalo de confianza del 95.45% para las diferencias de las medias de las poblaciones.

Siendo $\bar{X}_A = \mu_A = 6.5$ kg, $\sigma_A = 0.4$ kg, $\bar{X}_B = \mu_B = 6.3$ kg, $\sigma_B = 0.3$ kg y $n_A = n_B = 100$, los límites de confianza para la diferencia de las medias son, empleando la tabla 8.1

$$\bar{X}_A - \bar{X}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 6.5 - 6.3 \pm 2 \sqrt{\frac{(0.4)^2}{100} + \frac{(0.3)^2}{100}} =$$

$$= 0.2 \pm 0.1$$

Por lo tanto, el intervalo de confianza respectivo es (0.1, 0.3).

Ejemplo 10.2

Se tienen en una bodega 3000 focos de marca X, y 5000 de marca Y. Se extrae una muestra aleatoria de 150 focos de la marca X, y se obtiene una duración promedio de 1400 horas, con desviación estándar igual a 120 horas. Otra muestra aleatoria de 200 focos de la marca Y tuvo una duración promedio de 1200 horas, con desviación estándar igual a 80 horas. Obtener intervalos de confianza de

a. 95%

b. 99%

para la diferencia de los tiempos medios de duración de los focos de ambas marcas.

a. Puesto que se trata de poblaciones finitas y $\bar{X} = 1400$ h, $S_X = 120$ h, $N_X = 3000$, $n_X = 150$, $\bar{Y} = 1200$ h, $S_Y = 80$ h, $N_Y = 5000$ y $n_Y = 200$, se obtiene, estimando a σ_X y σ_Y con S_X y S_Y, respectivamente

$$1400 - 1200 \pm 1.96 \sqrt{\frac{(120)^2}{150} \quad \frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200} \quad \frac{5000 - 200}{5000 - 1}}$$

$$200 \pm 1.96\ (11.04)$$

$$200 \pm 21.638$$

o sea, (178.362, 221.638), puesto que de la tabla 8.1, para un nivel de confianza de 95%, $Z_c = 1.96$.

b. En este caso, al emplear la tabla 8.1 se obtiene

$Z_c = 2.58$ para un nivel de confianza de 99%, por lo cual los límites son

$$1400 - 1200 \pm 2.58 \sqrt{\frac{(120)^2}{150}\,\frac{3000 - 150}{3000 - 1} + \frac{(80)^2}{200}\,\frac{5000 - 2000}{5000 - 1}}$$

$$200 \pm 2.58\ (11.04)$$

$$200 \pm 28.483$$

y el intervalo de confianza es

$$(171.517,\ 228.483)$$

11. Pruebas de hipótesis

Supóngase que una empresa armadora de automóviles está en la disyuntiva de emplear una nueva marca de bujías en sus unidades o la que regularmente utiliza, y que su departamento de control de calidad debe decidir, con base en la información de las muestras de las dos marcas distintas. Las decisiones de este tipo, es decir, que se basan en estudios estadísticos, reciben el nombre de *decisiones estadísticas*, y a los procedimientos que permiten decidir si se acepta o rechaza una hipótesis se les llama *pruebas de hipótesis*, *pruebas de significancia* o *reglas de decisión*.

Al tomar decisiones estadísticas, es necesario postular las diversas alternativas o cursos de acción que pueden adoptarse.

En el caso particular de una *prueba de hipótesis* solamente se tienen dos cursos de acción posibles, los que se denotarán como H_0 y H_1. A la acción H_0 se le llama *hipótesis nula*, y a la H_1, *hipótesis alternativa*. Por ejemplo, si la hipótesis nula establece que $\mu_1 = \mu_2$, la hipótesis alternativa puede ser una de las siguientes:

$$\mu_1 > \mu_2, \mu_1 < \mu_2 \text{ o } \mu_1 \neq \mu_2$$

Al realizar una prueba de hipótesis, se prueba siempre la verdad de la hipótesis nula H_0, aun cuando de antemano se desee rechazarla.

12. Errores de los tipos I y II. Nivel de significancia

En muchas ocasiones se presenta el caso de que se rechaza una hipótesis nula cuando en realidad debería ser aceptada; cuando esto sucede se dice que se ha cometido un *error de tipo I*. En otras ocasiones se acepta una hipótesis nula siendo en realidad falsa; en este caso se dice que se ha cometido un *error de tipo II*.

Al probar una hipótesis nula, a la máxima probabilidad con la que se está dispuesto a cometer un error del tipo I se le llama *nivel de significancia*, α, de la prueba, el cual dentro de la práctica se acostumbra establecer de 5 por ciento (0.05) o 10 por ciento (0.1). El complemento del nivel de significancia, $1 - \alpha$, se conoce como *nivel de confianza*.

Si, por ejemplo, al realizar una prueba de hipótesis se escoge un nivel de significancia de 10 por ciento, significa que existen 10 posibilidades en 100 de que se rechace ésta cuando debería ser aceptada; es decir, que se rechaza a un nivel de significancia del 10 por ciento, y que la probabilidad de que la decisión haya sido errónea es de 0.1.

13. Comportamiento de los errores tipos I y II

Supóngase que se trata de probar la hipótesis nula de que la media, μ_S, de la distribución muestral de la estadística S es μ_1, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

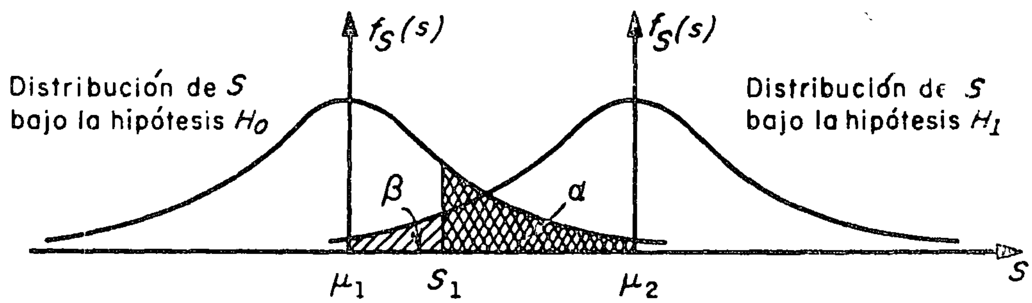
$$H_1 : \mu_S = \mu_2$$

En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar H_0 es la siguiente:

Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico S_1, rechácese H_0; en caso contrario, acéptese.

Es evidente que si H_0 es verdadera, entonces α (área con rayado doble) es la probabilidad de que $S > S_1$, o sea la de rechazar a H_0 siendo verdadera (error tipo I). Por otro lado, si H_1 es verdadera, entonces β (área con rayado sencillo) es la probabilidad de que $S < S_1$, o sea la de aceptar H_0 siendo falsa (error tipo II).

Obsérvese que si se aumenta el valor de S_1 se reduce la probabilidad α, pero se incrementa la β; lo contrario sucede si se disminuye el valor de S_1.



$$P[S>S_1] = \alpha \text{ (error tipo I)}$$
$$P[S<S_1] = \beta \text{ (error tipo II)}$$

Fig 13.1 *Probabilidades de los errores tipos I y II en pruebas de hipótesis.*

En realidad, la única forma posible en la cual se pueden minimizar simultáneamente los errores de tipos I y II es aumentando el tamaño de la muestra, para hacer más "picudas" las distribuciones muestrales de la estadística bajo las hipótesis H_0 y H_1.

Al observar la fig 13.2 siguiente, es posible concluir que el tamaño de los errores I y II es menor para un tamaño de muestra igual a 100 que para un tamaño igual a 50, considerando la misma regla de decisión anterior.

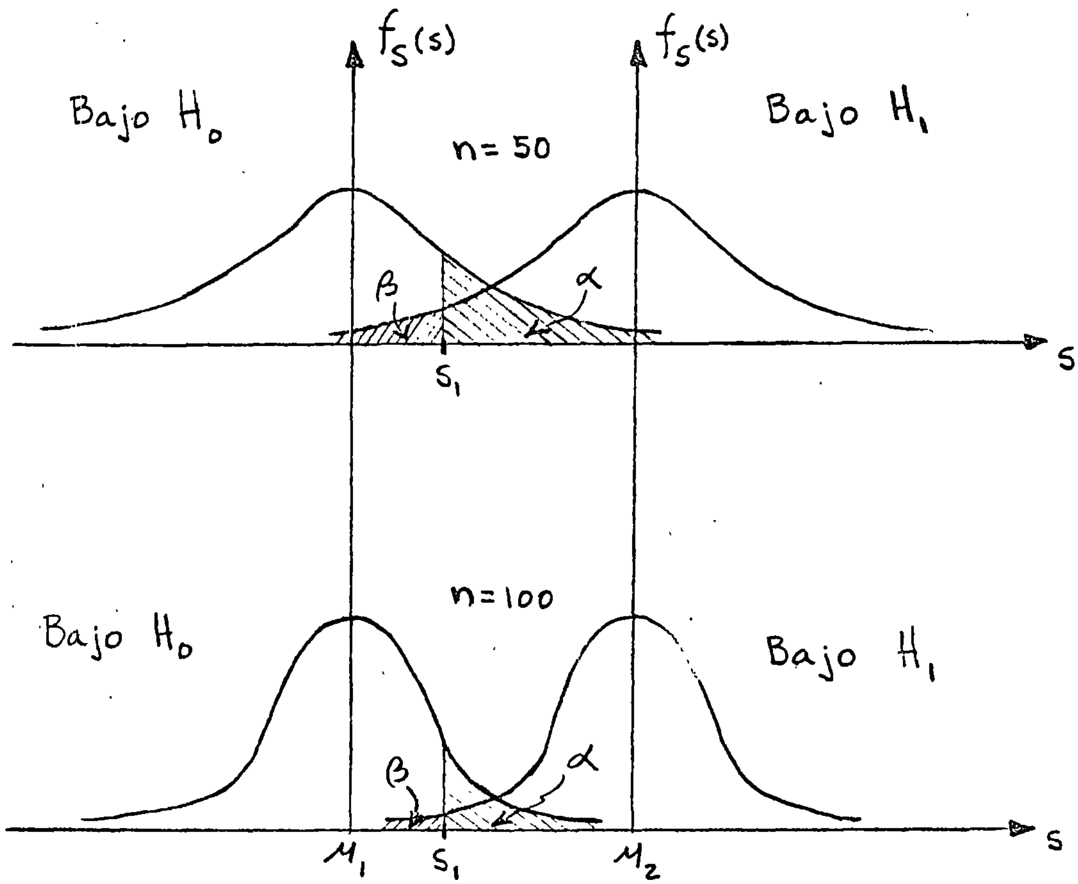


Fig 13.2

Sin embargo, esta técnica de reducción simultánea de ámbos tipos de errores no siempre puede ponerse en práctica, debido a razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo o de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del α por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia α.*

Al conjunto de los valores de la estadística en el que se rechaza la hipótesis nula se le denomina *región crítica, de rechazo, o de significancia.* Por el contrario, al conjunto de los valores de la estadística en que se acepta la hipótesis, se le llama *región de aceptación.*

Considérese que la distribución muestral de la estadística S es normal con desviación estándar σ_S, que la variable Z resulta de estandarizar a S, que la hipótesis nula, H_0, es que la media de S vale μ_S, y que la hipótesis alternativa H_1 es que dicha media es diferente de μ_S, es decir, que

$$Z = \frac{S - \mu_S}{\sigma_S}$$

H_0: media de la distribución muestral de $S = \mu_S$

H_1: media de la distribución muestral de $S \neq \mu_S$

Si se adopta la regla de decisión de aceptar la hipótesis H_0, si el valor de Z cae dentro del intervalo central que encierra al 99 por ciento del área de la distribución de probabilidades, entonces H_0 se aceptará en el caso en que

$$-2.58 \leq Z \leq 2.58$$

empleando la tabla de áreas bajo la curva normal estándar. Pero si el valor estandarizado de la estadística se encuentra fuera de dicho intervalo, se concluye que el evento puede ocurrir con probabilidad de 0.01 si la hipótesis H_0 es verdadera (área rayada total de la fig 14.1). En tal caso, el valor Z de la variable estándar difiere *significativamente* del que se podría esperar de acuerdo con la hipótesis nula, lo cual inclina a rechazarla a un nivel de confianza del 99 por ciento.

De lo anterior de deduce que el área total rayada de la fig 14.1 es el nivel de significancia α de la prueba, y representa la probabilidad de cometer un error del tipo I. Por ello, la región de aceptación de H_0 es $-2.58 \leq Z \leq 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

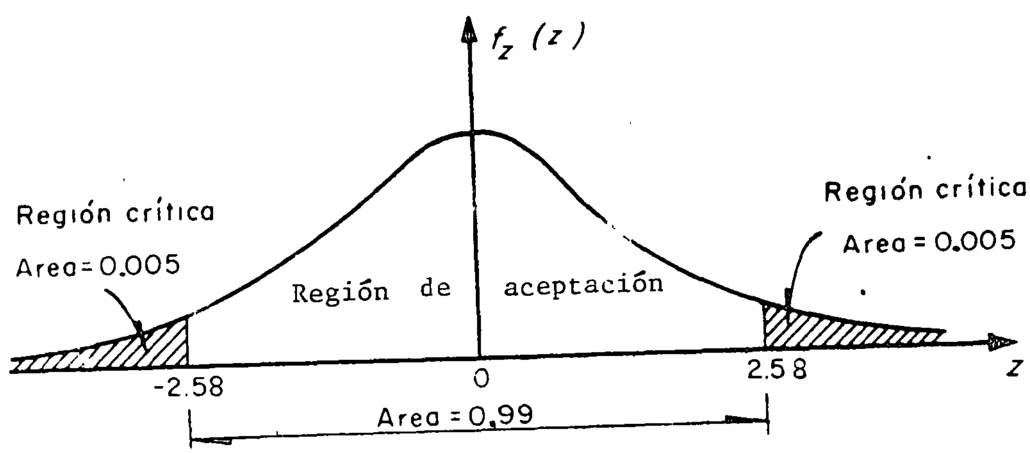


Fig 14.1 Región de significancia

En la tabla 14.1 se presentan los valores de la variable estandarizada, Z, que limitan las regiones de aceptación y de rechazo para el caso en el que la estadística involucrada en la prueba tenga distribución muestral normal. Cuando en alguna prueba de hipótesis se consideren niveles de significancia diferentes a los que aparecen en la tabla mencionada, resulta necesario emplear la de áreas bajo la curva normal estándar.

TABLA 14.1 VALORES CRITICOS DE z

Nivel de significancia, α	Valores de z para pruebas de una cola	Valores de z para pruebas de dos colas
0.1	−1.281 o 1.281	−1.645 y 1.645
0.05	−1.645 o 1.645	−1.960 y 1.960
0.01	−2.326 o 2.326	−2.575 y 2.575
0.005	−2.575 o 2.575	−2.810 y 2.810

15. Pruebas de una y de dos colas

En la prueba de hipótesis del ejemplo anterior, la región de rechazo de la hipótesis nula quedó en ambos extremos (colas) de la distribución muestral de la estadística involucrada en la prueba; a las pruebas de este tipo se les denomina *pruebas de dos colas*. Cuando la región de rechazo se encuentra solamente en un extremo de la distribución muestral en cuestión, se les llama *pruebas de una cola.*

Las pruebas de dos colas se presentan cuando en la hipótesis alternativa aparece el signo $\neq$ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S \neq \mu_1$$

en donde μ_S es la media de la estadística S, y μ_1 es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

y

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

las pruebas resultan de una cola.

16. Pruebas de hipótesis para la media

Para el caso de una población infinita (o finita en que se muestree con remplazo), cuya desviación estándar σ se conoce o se puede estimar adecuadamente, si se tiene que la estadística S obtenida de la muestra es el promedio aritmético, entonces la media de su distribución muestral es $\mu_S = \mu_{\bar{X}} = \mu$, y su desviación estándar es $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ son, respectivamente, la media y la desviación estándar de la variable aleatoria X asociada a la población, y n es el tamaño de la muestra. En tal caso, si $\bar{X}$ tiene distribución normal, la variable estandarizada correspondiente será

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Para el caso de muestreo sin remplazo de población finita, se tiene que $\sigma_S = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$, en donde N_p es el tamaño de la población, por lo que la variable estandarizada será

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}}$$

En los dos casos anteriores, el valor de Z correspondiente al de $\bar{X}$ de la muestra es el que se debe comparar con el valor crítico correspondiente al nivel de significancia fijado, para así aceptar o no la hipótesis nula (prueba de una cola). Si se trata de una prueba de dos colas, el valor de Z se debe comparar con los dos valores críticos que corresponden al valor de α seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de α.

Ejemplo 16.1

Se sabe que el promedio de calificaciones de una muestra aleatoria de tamaño 100 de los estudiantes de tercer año de ingeniería civil es de 7.6, con una desviación estándar de 0.2. Si μ denota la media de la población de esas calificaciones, X, y si se supone que $\bar{X}$ tiene distribución normal, probar la hipótesis

$\mu = 7.65$ en contra de la hipótesis alternativa $\mu \neq 7.65$, usando un nivel de significancia de

a. 0.05

b. 0.01

Para la solución se deben considerar las hipótesis

$$H_0 : \mu = 7.65$$

$$H_1 : \mu \neq 7.65$$

Puesto que $\mu \neq 7.65$ incluye valores menores y mayores de 7.65, se trata de una prueba de dos colas.

La estadística bajo consideración es el promedio aritmético, $\bar{X}$, de la muestra, que se supone extraída de una población infinita. La distribución muestral de $\bar{X}$ tiene media $\mu_{\bar{X}} = \mu$, y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ denotan, respectivamente, la media y la desviación estándar de la población de calificaciones.

Bajo la hipótesis H_0 (considerándola verdadera), se tiene que

$$\mu_{\bar{X}} = 7.65 = \mu$$

y utilizando la desviación estándar de la muestra como una estimación de σ, lo cual se supone razonable por tratarse de una muestra grande,

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.2/\sqrt{100} = 0.2/10 = 0.02$$

a. Para la prueba de dos colas a un nivel de significancia de 0.05 se establece la siguiente regla de decisión

> *Aceptar* H_0 *si el valor* Z *correspondiente al valor del promedio de la muestra se encuentra dentro del intervalo de* -1.96 *a* 1.96 *(tabla 14.1). En caso contrario, rechazar* H_0.

Puesto que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.6 - 7.65}{0.02} = -2.5$$

se encuentra fuera del rango de -1.96 a 1.96, se rechaza la hipótesis H_0 a un nivel de significancia de 0.05.

b. Si el nivel de significancia es 0.01, el intervalo de -1.96 a 1.96 de la regla de decisión del inciso α se remplaza por el de -2.58 a 2.58 tabla (14.1). Entonces, puesto que el valor muestral $Z = -2.5$ se encuentra dentro de este intervalo, se acepta la hipótesis H_0 a un nivel de significancia de 0.01.

Ejemplo 16.2

La resistencia media a la ruptura de cables de acero fabricados por la empresa X es de 905 kg. Una empresa consultora sugiere a X que cambie su proceso de manufactura, con lo cual incrementará la resistencia de sus cables. Se prueba el nuevo proceso, y se extrae una muestra aleatoria de 50 cables, obteniéndose para ellos una resistencia promedio de 926 kg, con desviación estándar igual a 42 kg. ¿Se puede considerar que el nuevo proceso realmente incrementa la resistencia, con un nivel de confianza de 99%?

En este caso, se debe plantear una prueba de hipótesis de una cola, para la cual

$$H_0 : \mu = 905 \text{ kg}$$

$$H_1 : \mu > 905 \text{ kg}$$

Puesto que el tamaño de la muestra es suficientemente grande, se puede aproximar la distribución muestral de la resistencia promedio mediante una normal, y estimar el valor de σ de la población mediante S_X de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a H_0, se tiene que

$$\mu_{\bar{X}} = \mu = 905 \text{ kg}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

Para la prueba de una cola a un nivel de significancia de $\alpha = 1 - (1 - \alpha) = 1 - 0.99 = 0.01$, la regla de decisión es

Aceptar H_0 si el valor estandarizado de $\bar{X}$ de la muestra es menor o igual a $Z_c = 2.326$ (tabla 14.1); en caso contrario, rechazar H_0.

En virtud de que

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{926 - 905}{5.94} = 3.535$$

es mayor de 2.326, se rechaza H_0 a un nivel de significancia de 1%, concluyéndose que en realidad el nuevo proceso sí incrementa la resistencia de los cables.

17. Pruebas de diferencias de medias

Sean $\bar{X}$ y $\bar{Y}$ los promedios aritméticos obtenidos de dos muestras de tamaños n_X y n_Y, extraídas respectivamente de dos poblaciones con medias μ_X y μ_Y, y desviaciones estándar σ_X y σ_Y. Se trata de probar la hipótesis nula, H_0, de que no existe diferencia entre las medias, es decir, que $\mu_X = \mu_Y$. Si n_X y n_Y son suficientemente grandes (>30), la distribución muestral de las diferencias de los promedios es aproximadamente normal. Dicha distribución muestral es rigurosamente normal si las variables aleatorias X y Y asociadas a la población tienen distribución normal, aunque n_X y n_Y sean menores de 30. Para esta distribución muestral, la variable estandarizada Z, que se compara con los valores críticos correspondientes, se encuentra dada por

$$Z = \frac{X - Y - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{X - Y - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}$$

con la cual se puede probar la hipótesis nula H_0 en contra de otras hipótesis alternativas, H_1, a un nivel apropiado de significancia.

Ejemplo 17.1

En el laboratorio de pruebas de una empresa fabricante de aparatos electrónicos se ensayaron dos marcas de transistores, *A* y *B*, de características similares, con objeto de comprobar su ganancia de voltaje. Se tomaron muestras aleatorias de 100 transistores de cada marca, arrojando una ganancia promedio de 31 decibeles, con desviación estándar de 0.3 decibeles para la marca *A*, y 30.9 decibeles de ganancia promedio, con desviación estándar de 0.4 decibeles para la otra. ¿Existe una diferencia significativa entre las ganancias en voltaje de los transistores a un nivel de significancia de

a. 0.05

b. 0.01?

Si μ_A y μ_B son las medias respectivas de las dos poblaciones infinitas a las que corresponden las muestras, la prueba de hipótesis adopta la forma siguiente:

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

Entonces, el valor de Z es, bajo la hipótesis H_0:

$$Z = \frac{\bar{x}_A - \bar{x}_B}{\sigma_{\bar{x}_A - \bar{x}_B}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{31 - 30.9}{\sqrt{\frac{(0.3)^2}{100} + \frac{(0.4)^2}{100}}} = 2$$

a. Puesto que se trata de una prueba de dos colas a un nivel de significancia de 0.05, la diferencia es significativa si el valor de Z se encuentra fuera del intervalo de -1.96 a 1.96. Como este es el caso, puede concluirse que efectivamente existe diferencia significativa en la ganancia en voltaje de los transistores.

b. Si la prueba es a un nivel de significancia de 0.01, la diferencia es significativa si Z se encuentra fuera del rango de -2.58 a 2.58. Partiendo del hecho de que Z = 2, la diferencia entre las ganancias es producto del azar, y se acepta la hipótesis de que ambos tipos de transistores tienen igual ganancia media en voltaje a un nivel de confianza de 99 por ciento.

Ejemplo 17.2

La estatura promedio de 50 estudiantes varones tomados al azar que participan en actividades deportivas es de 173 cm, con desviación estándar de 6.3 cm. Otra muestra aleatoria de 50 estudiantes varones que no participan en ese tipo de actividades tiene promedio de estatura igual a 171 cm, con desviación estándar igual a 7.1 cm. Probar la hipótesis de que los estudiantes varones que practican deportes son más altos que los que no lo hacen, a un nivel de significancia de 0.05.

Se debe decidir entre las hipótesis

$$H_0 : \mu_X = \mu_Y$$

$$H_1 \quad \mu_X > \mu_Y$$

siendo X la variable aleatoria asociada a la población infinita de estaturas de alumnos que practican deportes, y Y la asociada a la de estudiantes que no lo hacen, que también es infinita.

Bajo la hipótesis H_0 , se tiene que

$$\mu_{\bar{X}-\bar{Y}} = 0$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} = \sqrt{\frac{(6.3)^2}{50} + \frac{(7.1)^2}{50}} = 1.3424$$

Entonces, el valor de Z es

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{173 - 171}{1.3424} = \frac{2}{1.3424} = 1.489$$

Puesto que se trata de una prueba de hipótesis de una cola, a un nivel $\alpha = 0.05$, se rechazaría H_0 si el valor de Z muestral fuera mayor del valor crítico para dicho nivel, el cual es $Z_c = 1.645$. Puesto que $Z < Z_c$, en este caso se concluye que la diferencia en las estaturas de ambos grupos de estudiantes se debe únicamente al azar.

3.4 Muestras pequeñas

Como ya se indicó, para muestras grandes ($n \geqslant 30$) las distribuciones muestrales de muchas estadísticas son aproximadamente normales, siendo tanto mejor la aproximación cuanto mayor es el tamaño de n. Sin embargo, cuando se trata de muestras en las que $n < 30$, llamadas *muestras pequeñas*, la aproximación no es suficientemente buena, por lo que resulta necesario introducir una teoría apropiada para su estudio.

Al estudio de las distribuciones muestrales de las estadísticas para muestras pequeñas se le llama *teoría estadística de las muestras pequeñas.* Existen al respecto tres distribuciones importantes: *Ji cuadrada, F* y *t de Student.*

3.4.1 Distribución *Ji* cuadrada (χ^2)

Hasta ahora solo se ha tratado la distribución muestral de la media. En esta sección se verá lo concerniente a la distribución muestral de la variancia, S_X^2, para muestras aleatorias extraídas de poblaciones normales. Puesto que S_X no puede ser negativa, es de esperarse que su distribución muestral no sea una curva normal, ya que esta tiene ordenadas mayores de cero en el lado de las abscisas negativas. De hecho, la estadística S_X^2 se puede estudiar si se consideran muestras aleatorias de tamaño n extraídas de una población normal con desviación estándar σ_X y si para cada muestra se calcula el valor de la estadística.

$$\chi^2 = \frac{n\ S_X^2}{\sigma^2} \tag{3.14}$$

donde S_X^2 es la variancia de la muestra.

El número de grados de libertad, ν, de una estadística se define como

$$\nu = n - k$$

siendo n el tamaño de la muestra y k el número de parámetros de la población que deben estimarse a partir de ella.

La distribución muestral de la estadística χ^2 está dada por la ecuación

$$f\ (\chi^2) = U\ \chi^{\nu-2}\ e^{-\frac{1}{2}\chi^2}$$

en la que U es una constante que hace que el área total bajo la curva resulte igual a uno, y $\nu = n - 1$ es el número de grados de libertad. Esta distribución se llama *Ji cuadrada*, misma que se presenta en la fig 21 para distintos valores de ν.

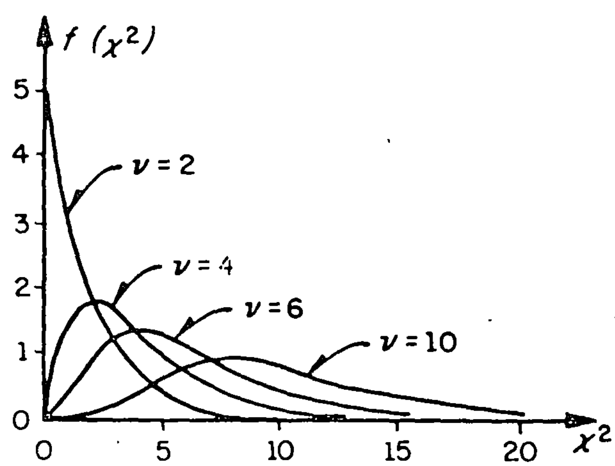
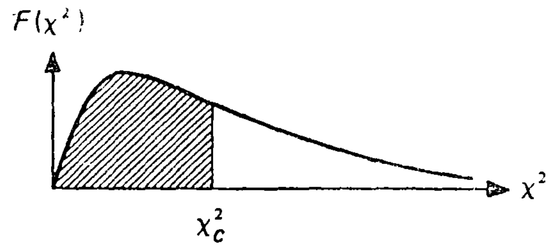


Fig 21. Distribución Ji cuadrada para distintos valores de ν

45

TABLA 8. VALORES CRITICOS χ^2_c



ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.016	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.76	5.39	3.36	1.92	1.06	.711	.483	.297	.207
5	16.7	15.2	12.8	11.15	9.2	6.63	4.35	2.67	1.61	1.15	.831	.554	.413
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.18	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.35	7.57	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.2	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.2	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.7	30.6	27.5	25.1	22.3	18.2	14.3	11.0	8.55	7.26	6.25	5.22	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.73	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.45	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.8	35.6	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.02
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.5	13.15	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.5	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.7	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	43.0	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.12	82.4	77.9	74.2	70.1	67.3

No obstante que la distribución *Ji* cuadrada solo se ha presentado en el estudio de las muestras pequeñas, cabe aclarar que es válida para aquellas mayores de 30 si la variable aleatoria involucrada tiene distribución normal.

3.4.1.1 Intervalo de confianza para la variancia

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la variancia de la población en términos de la variancia de una muestra extraída de ella, a un nivel de confianza dado $1 - \alpha$, si se hace uso de los valores críticos χ_c^2 de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística χ^2, estaría dado por

$$\chi_{c'}^2 < \frac{n\, S_X^2}{\sigma^2} < \chi_c^2$$

donde $\chi_{c'}^2$ y χ_c^2 son los valores críticos para los cuales el $(1 - \alpha)/2$ por ciento del área se encuentra en los extremos izquierdo y derecho de la distribución, respectivamente.

Con base en lo anterior, se concluye que

$$\frac{n\, S_X^2}{\chi_c^2} < \sigma^2 < \frac{n\, S_X^2}{\chi_{c'}^2}$$

es un intervalo de confianza para estimar a σ^2 a un nivel de confianza $1 - \alpha$.

3.4.1.2 Prueba de hipótesis para la variancia

La prueba de hipótesis para la variancia de una población normal se efectúa calculando el valor de la estadística χ^2 y estableciendo las hipótesis H_0 y H_1 apropiadas, es decir, se adoptan reglas de decisión similares a las usadas para la estadística Z.

Ejemplo

La variancia del tiempo de elaboración de cierto producto es igual a 40 min; sin embargo, su proceso de manufactura se modifica y se toma una muestra de veinte tiempos, para la cual la variancia resulta ser igual a 62 min. ¿Es significativo el aumento del tiempo de elaboración a un nivel de significancia de

a) 0.05

b) 0.01?

Se debe decidir de entre las hipótesis

$$H_0 : \sigma^2 = 40 \text{ min}$$

$$H_1 : \sigma^2 > 40 \text{ min}$$

Suponiendo que la hipótesis nula es correcta, el valor de la estadística χ^2 para la muestra considerada es

$$\chi^2 = \frac{n\,S_X^2}{\sigma^2} = \frac{(20)\,(62)}{40} = 31$$

a) Como se trata de una prueba de una cola, la hipótesis H_0 se rechazaría si el valor de la estadística χ^2 fuera mayor que el de χ^2 para un nivel de significancia igual a 0.05, el cual, para $\nu = 20 - 1 = 19$ grados de libertad resulta ser 30.1 (tabla 8). Como $31 > 30.1$, H_0 se rechaza a un nivel de significancia de 0.05.

b) En este caso, el valor de χ^2 para un nivel de significancia de 0.01 y 19 grados de libertad es igual a 36.2. Puesto que $31 < 36.2$, se acepta H_0 a un nivel de significancia de 0.01.

3.4.2 Distribución F

Al efectuar la prueba de hipótesis de igualdad de medias para muestras pequeñas, en la siguiente sección se supondrá que las variancias de las poblaciones a las que corresponden tales muestras son iguales. Por lo tanto, es necesario probar antes si tal suposición es correcta. Para ello, debe considerarse que si S_X^2, n_X y S_Y^2, n_Y son respectivamente la variancia y el tamaño de dos muestras extraídas de poblaciones normales que tienen igual variancia, entonces

$$F = \frac{S_X^2}{S_Y^2} \tag{3.15}$$

TABLA 9. VALORES F_c PARA α = 0.01

v_2 = Grados de libertad del denominador	v_1 = Grados de libertad del numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4.052	5 000	5.403	5.625	5.764	5.859	5.928	5.982	6.023	6.056	6.106	6.157	6.209	6.235	6.261	6.287	6.313	6.339	6 366
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50	99.50	99.50	99.50	99.50
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	26.90	26.70	26.60	26.50	26.40	26.30	26.20	26.10
4	21.20	18.00	16.70	16.00	15.50	15.50	15.00	14 80	14.70	14.50	14.40	14.20	14.00	13.90	13.80	13.70	13.70	13.60	13.50
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.70	10.90	9.77	9.15	8.75	8 47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.87
7	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.30	8.65	7.59	7.01	6.63	6.37	6.17	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.83
9	10.60	8.02	6.99	6.42	6.06	5.81	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9 66	7.22	6.22	5.68	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.03	3.93	3.86	3.78	3 69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.95	2.87
16	8.53	6.23	5.29	4.77	4.43	4.20	4.03	3.89	3.78	3.69	3.55	3.40	3.26	3.18	3.10	3.02	2.93	2.84	2.76
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3 16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.03	5.79	4.87	4.36	4.04	3.81	3.64	3.50	3.41	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.83	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.46	3.30	3.17	3.07	2.98	2.84	2.71	2.54	2.47	2.39	2.29	2.20	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.14	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.65	4.79	3.95	3 48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1 95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1 32	1.00

resulta ser el valor de una variable aleatoria (estadística) que tiene distribución F, con parámetros $\nu_X = n_X - 1$ y $\nu_Y = n_Y - 1$. Esta distribución (fig 22) cuenta con dos parámetros, ν_X y ν_Y, que son los grados de libertad que corresponden a la variancia del numerador y del denominador de la ec 3.15, respectivamente. Cuando se hace referencia a una distribución F en particular, siempre se dan primero los grados de libertad para la variancia del numerador; es decir, $F(\nu_X, \nu_Y)$. En la tabla 9 se presentan los valores críticos F_c para distintos valores de ν_X y ν_Y y un nivel de significancia de 0.01. Cuando los grados de libertad ν_X o ν_Y no se encuentren en dicha tabla, el valor de F se puede obtener mediante interpolación lineal. Si se desea probar la hipótesis a otros niveles de significancia, es factible emplear las tablas de la distribución F (refs 9 y 11).

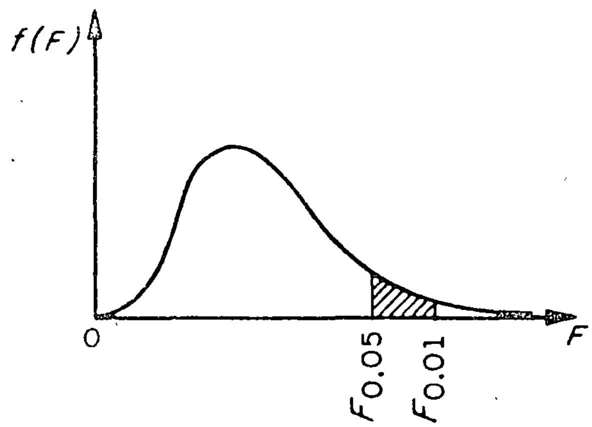


Fig 22. Distribución F.

De acuerdo con lo anterior, se puede probar la hipótesis nula

$$H_0: \ \sigma_X^2 = \sigma_Y^2$$

en contra de alguna hipótesis alternativa adecuada haciendo uso del hecho de que el cociente S_x^2/S_Y^2 es una estadística que tiene distribución F.

Ejemplo

Una empresa manufacturera de cartón prensado va a decidir acerca del empleo de una prensadora A o una B a fin de obtener un grosor determinado en su producto. El problema estriba en que ambas prensadoras proporcionan grosores muy similares, es decir, que la variancia de los grosores para las dos máquinas es la misma. Para decidir acertadamente, se toma una muestra aleatoria de 31 cartones prensados por la máquina A y otra de 41 por la B. Como las variancias del grosor para los cartones de las muestras resultan ser de 12 y de 5 micras, respectivamente, se establecen las hipótesis

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 > \sigma_B^2$$

con objeto de probarlas a un nivel de significancia de 0.01.

El valor de la estadística F resulta

$$F = \frac{S_A^2}{S_B^2} = \frac{12}{5} = 2.4$$

Puesto que $\nu_A = 31 - 1 = 30$ y $\nu_B = 41 - 1 = 40$, en la tabla 9 se puede ver que para un nivel de significancia de 0.01 el valor, F_c, de $F(30, 40)$ es 2.11. De acuerdo con estos valores, la hipótesis H_0 se rechazaría si el valor de F fuera mayor que F_c (30, 40).

Puesto que lo anterior resulta ser cierto, se rechaza H_0, concluyéndose que la prensadora B sería la mejor elección.

3.4.3 Distribución t de Student

Si se consideran muestras de tamaño n extraídas de una población normal con media μ y variancia desconocida, para cada muestra se puede calcular la estadística T definida mediante la fórmula

$$T = \frac{\overline{X} - \mu}{S_X} \sqrt{n - 1} \tag{3.16}$$

donde X es el promedio y S_X la desviación estándar de la muestra.

La distribución muestral de T (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{(1 + \frac{t^2}{\nu})(\nu + 1)/2}$$

ojo: va de exponente

en la que U es una constante que hace que el área bajo la curva sea igual a uno, y $\nu = n - 1$ es el número de grados de libertad.

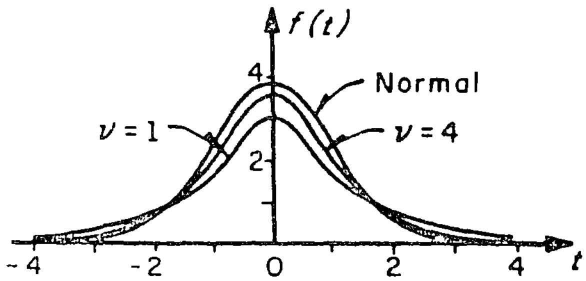


Fig 23. Distribución t de Student para distintos valores de ν

En la fig 23 se aprecia que conforme ν (o n, el tamaño de la muestra) aumenta, la distribución de $f(t)$ se aproxima a la distribución normal.

3.4.3.1 Límites e intervalos de confianza

De manera similar a como se hizo con la distribución normal, es posible estimar los límites de confianza de la media, μ, de una población mediante los *valores críticos*, t_c, de la distribución t, que dependen del tamaño de la muestra y del nivel de confianza deseado, encontrándose dichos valores en la tabla 10.

Así pues,

$$- t_c \leqslant \frac{\overline{X} - \mu}{S_X} \sqrt{n - 1} \leqslant t_c$$

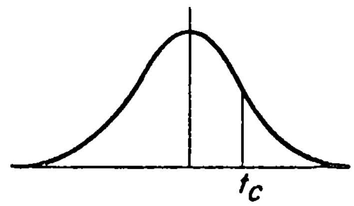
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\overline{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \overline{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\overline{X} \pm t_c \frac{\sigma_X}{\sqrt{n-1}}$$

TABLA 10. VALORES t_c PARA LA DISTRIBUCION t DE STUDENT



ν	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.07	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.275	.138
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.04	3.36	2.58	2.02	1.48	.920	.727	.560	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.91	1.43	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.36	.871	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.693	.537	.258	.128
15	2.95	2.61	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.128
19	2.87	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.256	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.05	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.71	1.31	.855	.683	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.30	.853	.683	.530	.256	.127
40	2.70	2.43	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.528	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

3.4.3.2 Pruebas de hipótesis

La prueba de hipótesis para la media de una población se puede efectuar con muestras pequeñas en forma análoga a la de muestras de tamaño mayor de 30 si en lugar de utilizar a la estadística Z se emplea la T. Entonces, si se consideran dos muestras aleatorias cuyos tamaños, desviaciones estándar y promedios son n_X, S_X, $\overline{X}$ y n_Y, S_Y, $\overline{Y}$, respectivamente, extraídas de poblaciones normales de igual variancia ($\sigma_X^2 = \sigma_Y^2$), se puede probar la hipótesis, H_0, de que las muestras provienen de una misma población, es decir, de que también sus medias son iguales, utilizando la estadística T definida por

$$T = \frac{\overline{X} - \overline{Y}}{\epsilon \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \tag{3.17}$$

donde

$$\epsilon = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \tag{3.18}$$

cuya distribución es la t de Student, con $\nu = n_X + n_Y - 2$ grados de libertad.

Ejemplo

Conforme al plan de desarrollo agrícola de una región, se probó un nuevo fertilizante para maíz. Para ello se escogieron 24 ha de terreno, aplicándose dicho producto a la mitad de ellas. El promedio de producción de maíz en la zona que se usó fertilizante fue de 5.3 ton, con una desviación estándar de 0.40 ton, en tanto que en la otra zona el promedio fue de 5.0 ton, con desviación estándar de 0.36 ton

De acuerdo con los resultados, ¿se puede concluir que existe un aumento significativo en la producción de maíz al usar fertilizante, si se utiliza un nivel de significancia de

a) 0.01

b) 0.05?

Solución

Para probar la hipótesis de igualdad de medias es indispensable saber primero si las muestras provienen de dos poblaciones normales de igual variancia. En ese caso, si σ_X^2 y σ_Y^2 denotan a las variancias de la producción de maíz en la zona tratada y en la no tratada, respectivamente, se debe probar la hipótesis nula H_0: $\sigma_X^2 = \sigma_Y^2$ en contra de la hipótesis alternativa H_1: $\sigma_X^2 > \sigma_Y^2$ a los dos niveles de significancia establecidos.

El valor de la estadística F es, de la ec 3.15,

$$F = \frac{S_X^2}{S_Y^2} = \frac{(0.40)^2}{(0.36)^2} = 1.27$$

y el valor crítico de F (11, 11), obtenido de la tabla 9 mediante interpolación lineal, resulta 4.47. Por lo tanto, como $1.27 < 4.47$, se acepta la hipótesis nula a un nivel de significancia de 0.01.

El valor crítico de F (11, 11) a un nivel de significancia de 0.05 (ref 9) es 2.82, de ahí que como $1.27 < 2.82$, también se acepta la hipótesis H_0.

Con base en lo anterior, se debe decidir entre las hipótesis

H_0: $\mu_X = \mu_Y$ (la diferencia en los promedios se debe al azar)

H_1: $\mu_X > \mu_Y$ (el fertilizante mejora la producción)

Bajo la hipótesis H_0, se tiene que

$$\epsilon = \sqrt{\frac{n_X\, S_X^2 + n_Y\, S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$t = \frac{5.3 - 5.0}{0.397\sqrt{\frac{1}{12} + \frac{1}{12}}} = 1.85$$

a) Puesto que se trata de una prueba de una cola a un nivel de significancia de 0.01, se rechaza la hipótesis H_0 si t es mayor que el valor crítico, t_c, correspondiente a dicho nivel, el cual para $\nu = n_X + n_Y - 2 = 12 + 12 - 2 = 22$ grados de libertad, se obtiene de la tabla 8 como $t_c = 2.51$. Como $t < t_c$, la hipótesis H_0 no se puede rechazar a un nivel de significancia de 0.01.

b) Si el nivel de significancia de la prueba es de 0.05, se rechaza H_0 si t es mayor que el valor t_c respectivo que para 22 grados de libertad es $t_c = 1.72$, por lo que de acuerdo con lo anterior, H_0 se rechaza a un nivel de significancia de 0.05.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

CONTROL ESTADISTICO DE CALIDAD

CARTAS DE CONTROL

M. en I. AUGUSTO VILLARREAL A.

OCTUBRE, 1978

CARTAS DE CONTROL

Por: M en I Augusto Villarreal A. *

INTRODUCCION

Aunque existe la tendencia generalizada a pensar que el Control de Calidad es de desarrollo reciente, realmente no existe nada nuevo en la idea básica de elaborar un producto caracterizado por un alto grado de uniformidad.

Durante siglos, hábiles artesanos han procurado elaborar productos que se distingan por su superior calidad, y una vez que han logrado obtener un cierto estándar de calidad óptimo, eliminar dentro de lo posible la variación entre productos que nominalmente deben resultar iguales.

La idea de que la Estadística puede resultar un instrumento muy útil para asegurar un estándar adecuado de calidad para los productos manufacturados, se remonta no más allá del advenimiento de la producción masiva, y el uso extendido de los métodos estadísticos para resolver problemas de control de calidad es aún más reciente.

Muchos problemas que aparecen durante la elaboración de un producto son susceptibles de ser resueltos empleando tratamientos estadísticos, por lo que al hablar de control estadístico de calidad, nos estaremos refiriendo esencialmente a las dos técnicas especiales que se discutirán en esta parte del curso: uso de las Cartas de Control y muestreo de aceptación.

* Profesor Investigador, División de Estudios Superiores e Instituto de Ingeniería, UNAM

Conviene mencionar que la palabra calidad, al ser empleada de aquí en adelante, se referirá a alguna propiedad medible o contable de algún producto, tal como el diámetro de un balín de acero, la resistencia de una viga de concreto, el número de defectos en una pieza de tela, la eficacia de cierta droga, etc.

IDEAS SOBRE CARTAS DE CONTROL

A muchos individuos les puede sorprender el hecho de que dos artículos aparentemente idénticos, elaborados bajo condiciones cuidadosamente controladas, de las mismas materias primas, y por una misma máquina con diferencia de pocos segundos, puedan, sin embargo, diferir en muchos aspectos.

En efecto, cualquier proceso de manufactura, aun siendo muy bueno, se encuentra caracterizado por una cierta cantidad de variación que es de naturaleza aleatoria, y que no puede ser eliminada en forma completa.

Cuando la variabilidad presente en un proceso de producción se limita a variación aleatoria se dice que el proceso se encuentra en un estado de control estadístico.

Tal estado se puede alcanzar cuando se eliminan aquellos problemas causados por otro tipo de variación, llamada variación sistemática, que es de naturaleza más bien determinística, y que se puede achacar, por ejemplo, a operadores mal entrenados, materia prima de baja calidad, máquinas en mal estado, etc.

Ya que los procesos de manufactura se encuentran rara vez libres de estos problemas, conviene contar con algún método sistemático para detectar desviaciones serias de un estado de control estadístico cuando ocurren, o inclusive antes de que ocurran, tales desviaciones.

Ese método sistemático de detección se puede tener mediante el empleo de las llamadas Cartas de Control.

TIPOS DE CARTAS DE CONTROL

En lo que sigue distinguiremos entre las cartas de control para mediciones o variables ($\bar{X}$, R, σ) y las cartas de control para atributos (p, c), dependiendo de que las observaciones que estemos analizando sean mediciones o datos contados o calculados, respectivamente.

Un ejemplo del primer caso sería la longitud de las varillas de acero de una muestra. Como ejemplo del segundo caso tendríamos el número de focos defectuosos en una muestra de tamaño dado.

CONFIGURACION DE LAS CARTAS DE CONTROL

En cualquiera de los casos mencionados, una carta de control consiste de una Línea Central, correspondiente a la calidad promedio a la que el proceso debe funcionar, y dos líneas que corresponden al Límite Superior de Control (LSC) y al Límite Inferior de Control (LIC), respectivamente, tal como se muestra en la Fig 1.

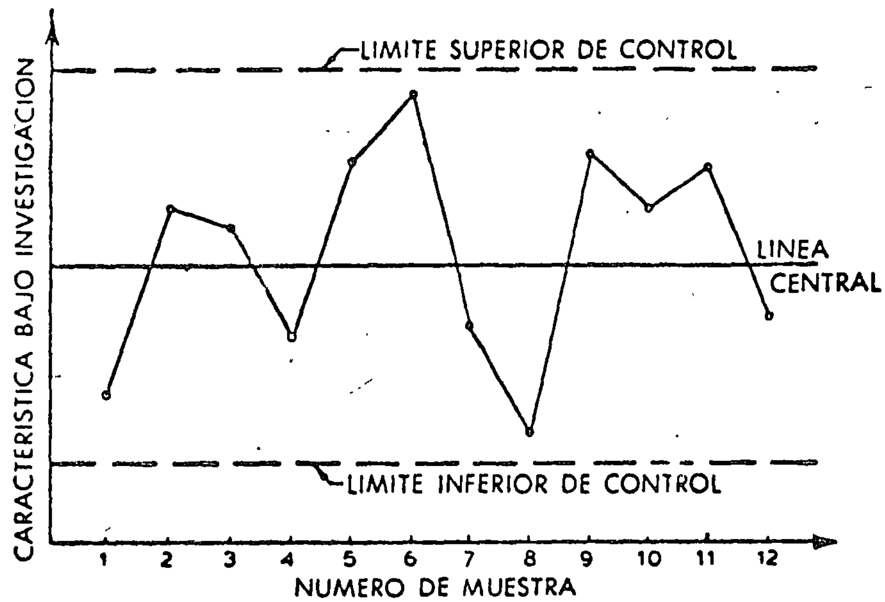


Fig 1. Aspecto general de una carta de control

Estos límites se escogen en forma tal que los valores que se encuentren dentro de ellos se puedan atribuir al azar, en tanto que los valores que caigan fuera de ellos se puedan considerar como indicaciones de falta de control.

No obstante la idea anterior, conviene mencionar que en la Fig 2 que se presenta a continuación se pueden considerar otras posibles situaciones de "falta de control" que ameritan investigarse:

1. Cuando dos de tres puntos sucesivos caen en la zona A.
2. Cuando cuatro de cinco puntos sucesivos caen en la zona B o más allá.
3. Cuando ocho puntos sucesivos caen en la zona C o más allá.

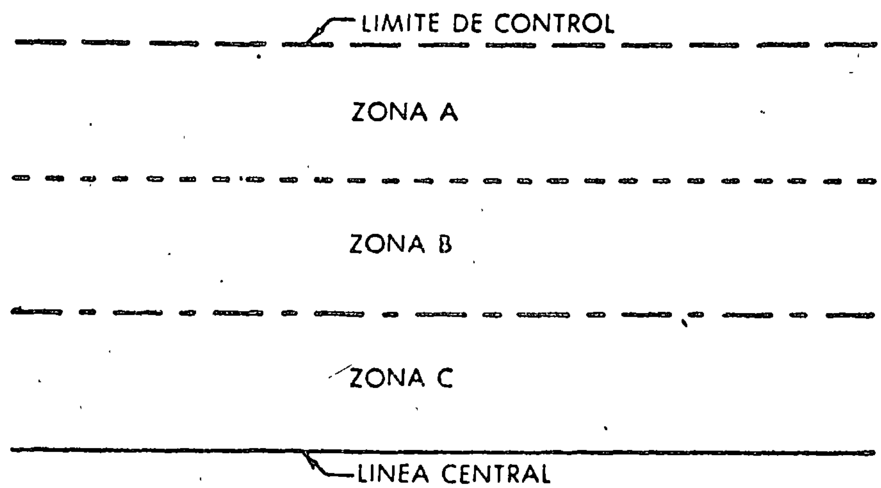


Fig 2 Diagrama que define las zonas A, B y C usadas en el análisis de Cartas de Control.

Debe hacerse notar que cada una de las zonas A, B y C constituye la tercera parte del área entre la línea central y un límite de control, y que las pruebas mencionadas se aplican a ambas mitades de la carta de control, pero se aplican <u>separadamente para cada mitad</u>, y nunca a las dos mitades en combinación.

EXPLICACION DEL EMPLEO DE LAS CARTAS DE CONTROL

Si se grafican en una carta los resultados obtenidos a partir de muestras tomadas periódicamente a intervalos frecuentes, es posible verificar por medio de ella si el proceso se encuentra bajo control, o si se encuentra presente en el proceso la variación sistemática del tipo descrito anteriormente.

Cuando un punto graficado cae fuera de los límites de control, es necesario encontrar el problema que causó tal evento dentro del proceso. Pero aun si los puntos caen dentro de los límites mencionados, alguna tendencia, o cierto patrón de los mismos, puede indicar que se debe llevar a cabo alguna acción para prevenir y así evitar algún problema serio.

La habilidad para "leer" las cartas de control y para determinar a partir de ellas cuál acción correctiva debe llevarse a cabo, - se obtiene a partir de la experiencia y del juicio altamente desarrollado. Un practicante del control estadístico de la calidad debe no sólo comprender los fundamentos estadísticos de la - materia, sino también encontrarse identificado plenamente con los procesos que desea controlar.

CARTAS DE CONTROL PARA MEDICIONES (VARIABLES)

Cuando se requiere establecer control estadístico de la calidad de algún producto en términos de mediciones o variables, es costumbre ejercer tal control sobre la calidad media del proceso, - al igual que sobre su variabilidad.

La primera meta se logra al graficar los promedios de muestras - extraídas periódicamente en la llamada carta de control para los promedios, o simplemente carta $\bar{X}$. La variabilidad se puede controlar de igual forma si se grafican los rangos o las desviaciones estándar de las muestras en las llamadas cartas R o cartas σ, respectivamente, dependiendo de cuál estadística se emplee para estimar la desviación estándar de la población.

Si se conocen la media μ y la desviación estándar σ de la población (proceso) y es razonable suponer las mediciones obtenidas como muestras extraídas de una población normal, se puede asegurar que con probabilidad 1 - α el promedio aritmético de una muestra aleatoria de tamaño n se encontrará entre

$$\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

ó

$$\mu - z_{\alpha/2}\,\sigma_{\bar{X}} \quad \text{y} \quad \mu + z_{\alpha/2}\,\sigma_{\bar{X}}$$

puesto que $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ para el caso de la distribución muestral del promedio aritmético, cuando se muestrea de una población infinita. La suposición de que la extracción de muestras aleatorias se hace de una población infinita es válida en el caso presente, puesto que, por ejemplo, la producción de cierto producto en una fábrica tiende a infinito conforme pasa el tiempo.

Los dos límites anteriores ($\mu \pm z_{\alpha/2}\sigma_{\bar{X}}$) proporcionan entonces límites inferiores y superiores de control y, bajo las suposiciones anteriores, permiten al practicante del control de calidad determinar si se debe o no llevar a cabo algún ajuste en el proceso, al graficar los promedios aritméticos obtenidos de muestras de tamaño n en una carta como la que se muestra en la Fig 1.

Conviene establecer en este momento que al emplear una carta de control para los promedios, lo que se hace realmente es probar hipótesis nulas de que a un cierto nivel de confianza 1-α el valor de la media de la distribución muestral de los promedios sea igual al valor de la calidad nominal del proceso, o al de la calidad media calculada para el mismo, μ_o. Para estas pruebas secuenciales de hipótesis, se emplean como estadísticas de prueba los valores de los promedios aritméticos obtenidos de muestras aleatorias extraídas de la población (o proceso). Es decir, se realizan pruebas de hipótesis para las cuales

$$H_0 : \mu = \mu_o$$

$$H_1 : \mu \neq \mu_o$$

(Prueba de dos colas; cada prueba se realiza con el valor $\bar{X}_i$ de la muestra i)

en donde μ es la media de la distribución muestral del promedio aritmético, μ_o la calidad nominal o calidad media calculada del proceso, y $\bar{X}_i$ (i=1,2,3,...) el valor del promedio aritmético obtenido de la iésima muestra aleatoria. La forma secuencial de estas pruebas de hipótesis se muestra en la Fig 3 que se presenta a continuación.

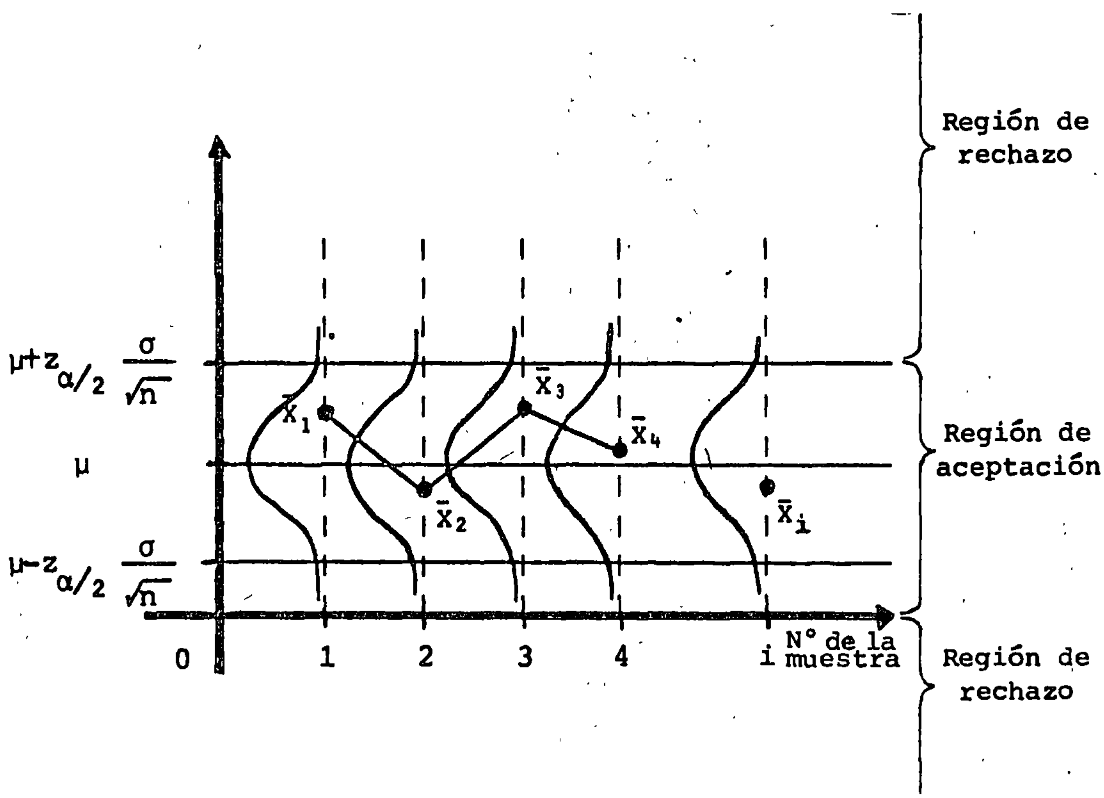


Fig 3 Pruebas de hipótesis que se realizan al emplear una carta de control para los promedios

Si se consideran problemas prácticos, los valores de μ y σ del proceso se desconocen, y es entonces conveniente estimar sus valores a partir de muestras tomadas mientras el proceso se encuentre "bajo control", tal como se explica más adelante. En la práctica es entonces difícil llegar a establecer límites de control del tipo $\mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ al desconocerse μ y σ, independientemente de que en muchos casos es demasiado arriesgado considerar a las mediciones como muestras aleatorias extraídas de una población normal.

En lugar de lo anterior, en el control de calidad industrial se emplean comúnmente los límites de control de "tres desviaciones estándar" o de "tres sigmas", que se obtienen al sustituir a $z_{\alpha/2}$ por un 3 al calcular los límites de control.

Conforme a lo anterior, con los límites de control

$$\mu \pm 3\sigma_{\bar{X}} \qquad \text{ó} \qquad \mu \pm 3 \frac{\sigma}{\sqrt{n}}$$

se puede confiar en que en el 99.73% de los casos el proceso no será declarado "fuera de control", cuando de hecho se encuentra "bajo control".

En otras palabras, estos límites de control permiten considerar que la probabilidad máxima de rechazar la hipótesis

$$H_0 : \Theta = \theta_o$$

cuando debería de ser aceptada (probabilidad de cometer un error de tipo I) es de 0.27%, siendo θ_o un valor de calidad fijo del proceso, y θ el del parámetro correspondiente de la distribución muestral de la estadística bajo consideración.

ELABORACION DE LA CARTA DE CONTROL PARA LOS PROMEDIOS ($\bar{X}$)

a. Caso en que se conocen la media μ y la desviación estándar σ de la población.

Línea central ———————— μ

Límites de control ———————— $\mu \pm 3\sigma_{\bar{x}}$ ó $\mu \pm 3\frac{\sigma}{\sqrt{n}}$

ó $\mu \pm A\sigma$, siendo $A = \frac{3}{\sqrt{n}}$

en donde los valores de A se obtienen de la tabla I, en función de n, el tamaño de la muestra.

<u>Ejemplo</u>: Sea el proceso de elaboración de varillas de acero para las cuales se sabe que el diámetro medio es de 2.5 cm, con una desviación estándar de 0.01 cm. Se desea efectuar control del diámetro de las mismas, para lo cual se extraen periódicamente muestras de cinco varillas. Se pide establecer la línea central y los límites de control para una carta $\bar{X}$.

<u>Solución</u>. Siendo $\mu = 2.5$ cm, $\sigma = 0.01$ y $n = 5$, se tiene que:

Línea central = μ=2.5

Límites de control:

$$2.5 \pm 3\frac{\sigma}{\sqrt{n}} = 2.5 \pm \frac{3(0.01)}{\sqrt{5}} = 2.5 \pm 0.0134 \Longrightarrow 2.5134, \quad 2.4866$$

o, de la tabla I

$$2.5 \pm A\sigma = 2.5 \pm 1.342(0.01) = 2.5 \pm 0.01342 \Longrightarrow 2.51342, \quad 2.48658$$

b. Caso en que se desconocen μ y σ.

Para este caso, que es el más común, es necesario estimar, como se dijo anteriormente, tales parámetros con base en muestras preliminares. Para el caso, normalmente se acostumbra emplear un mínimo de 20 a 25 muestras de 4 ó 5 elementos, obtenidas consecutivamente cuando el proceso está "bajo control".

Sin embargo, como veremos más adelante, se pueden emplear procedimientos estadísticos más formales para determinar el número de muestras (y de elementos en las mismas) más adecuado para las cartas $\bar{X}$.

Entonces, si se utilizan κ muestras preliminares, cada una de tamaño n, se puede estimar con adecuada precisión el valor de μ mediante

$$\bar{\bar{X}} = \frac{1}{\kappa}\sum_{i=1}^{\kappa} \bar{X}_i$$

siendo $\bar{\bar{X}}$ un estimador insesgado y consistente de μ, donde $\bar{X}_i$ denota al promedio aritmético de la iésima muestra, y $\bar{\bar{X}}$ es el promedio de los promedios de las muestras.

El valor de σ de la población puede ser estimado a partir de las desviaciones estándar o de los rangos de las muestras. Si el tamaño de las mismas es pequeño, usualmente el rango proporciona un estimador eficiente de σ, además de que el proceso de cálculo del mismo es bastante más simple que el de la desviación estándar para las muestras.

Sin embargo, es conveniente, cuando se requiere bastante precisión en el cálculo de los límites de control, estimar a σ mediante las desviaciones estándar de las muestras. Tal es el caso, por ejemplo, de muestras de productos que son caros, y que deben destruirse al momento de tomar las mediciones.

b.1 Estimando a σ mediante los rangos de las muestras

Hay que obtener primero el valor $\bar{R}$, que es el rango promedio de los rangos de las κ muestras, es decir,

$$\bar{R} = \frac{1}{\kappa} \sum_{i=1}^{k} R_i$$

Puesto que la estadística $\bar{R}$ siempre estima por encima de su valor real a la desviación estándar de la población, se obtiene un estimador sesgado. Debido a ello, es indispensable afectar el valor de $\bar{R}$ en forma tal de obtener un estimador insesgado de σ, para lo cual se hace

$$\text{Estimador insesgado de } \sigma = \frac{\bar{R}}{d_2}$$

El factor d_2 en la expresión anterior se obtiene experimentalmente al identificar el valor de la media en las distribuciones muestrales del cociente R/σ para distintos valores de n, considerando una población en la cual el valor de σ es conocido. Por ejemplo, para muestras de tamaño cinco ($n=5$), se ha obtenido experimentalmente el valor $d_2=2.326$, tal como se muestra en la Fig 4.

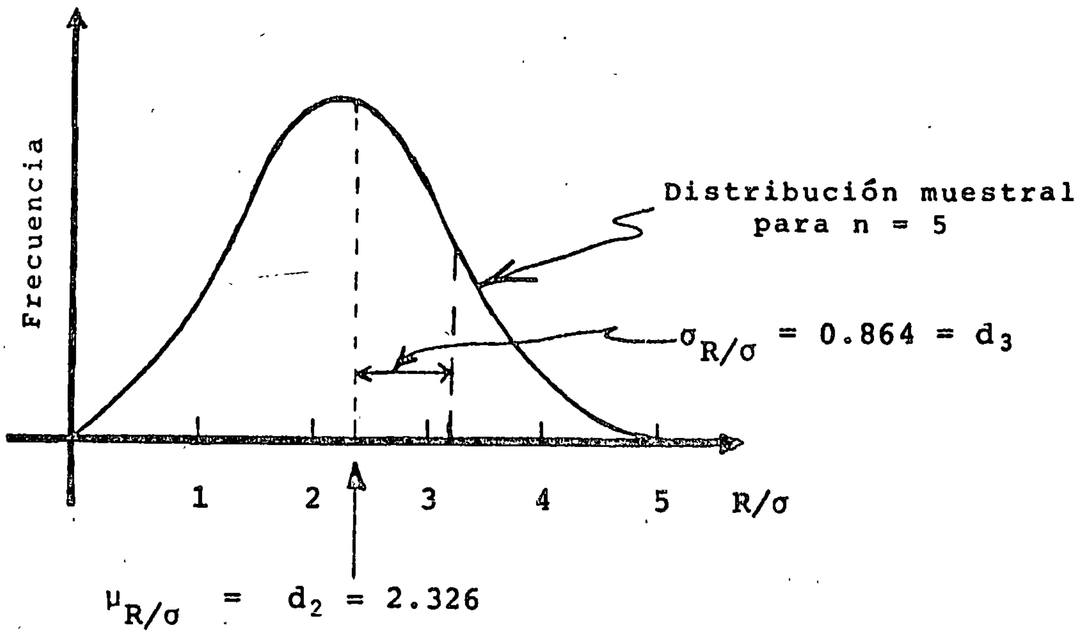


Fig 4. Distribución muestral de R/σ para n=5, suponiendo σ conocida.

En la tabla I se presentan los valores del factor d_2 para distintos tamaños de muestra, observándose que conforme se incrementa el valor de n aumenta el de ese factor, lo cual permite concluir que el rango estima mejor a la desviación estándar cuando las muestras son pequeñas.

De acuerdo con lo anterior, se pueden emplear las siguientes expresiones en la elaboración de la carta de control para los promedios:

Línea Central —— $\bar{\bar{X}}$

Límites de Control —— $\bar{\bar{X}} \pm 3\dfrac{\sigma}{\sqrt{n}}$ ó $\bar{\bar{X}} \pm \dfrac{3\bar{R}}{d_2\sqrt{n}}$

Para abreviar el cálculo de los límites de control a partir de los rangos de las muestras, se ofrece en la tabla I el factor

$$A_2 = \frac{3}{d_2\sqrt{n}}$$

cuyo empleo permite establecer los límites de control como

$$\bar{\bar{X}} \pm A_2 \bar{R}$$

b.2 Estimando a σ mediante las desviaciones estándar de las muestras

Se debe obtener primero el valor de $\bar{\sigma}$, que es el promedio de las desviaciones estándar de las muestras, es decir

$$\bar{\sigma} = \frac{1}{K} \sum_{i=1}^{k} S_i$$

En donde S_i denota la desviación estándar de la iésima muestra. No siendo tampoco $\bar{\sigma}$ un estimador insesgado de la desviación estándar de la población, ya que siempre la estima por abajo de su valor real, hay que afectar dicho valor por un cierto factor para hacerlo insesgado, es decir

$$\text{Estimador insesgado de } \sigma = \frac{\bar{\sigma}}{c_2}$$

Los valores de c_2 se reportan en la tabla I en función del tamaño de la muestra, y se obtienen mediante un procedimiento similar al explicado para el factor d_2.

Con base en lo anterior, los parámetros de la carta de control para los promedios son los siguientes:

Línea Central —— $\bar{\bar{X}}$

Límites de Control —— $\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}}$ ó $\bar{\bar{X}} \pm \frac{3\bar{\sigma}}{c_2\sqrt{n}}$

De nuevo, para abreviar el cálculo de los límites de control para la carta $\bar{X}$, obtenidos ahora a partir de las desviaciones estándar de las muestras, se puede emplear el factor dado en la tabla I

$$A_1 = \frac{3}{c_2\sqrt{n}}$$

con el cual los límites de control quedan como

$$\bar{\bar{X}} \pm A_1 \bar{\sigma}$$

NUMERO MINIMO DE MUESTRAS REQUERIDO PARA LA ELABORACION DE CARTAS $\bar{X}$

En este momento conviene establecer el número mínimo de muestras preliminares, m, así como el tamaño de las mismas, n, que es necesario considerar para estimar adecuadamente los parámetros de una carta de control para los promedios.

El asegurar que un mínimo de 20 0 25 muestras con 4 o 5 elementos cada una son necesarias para obtener los valores de $\bar{\bar{X}}$, $\bar{R}$ o $\bar{\sigma}$, frecuentemente choca con el argumento de que por razones de costo, tiempo, etc., se debe emplear un número menor de ellas. Por ello, se han preparado tablas como las II y IIIque se presentan al final, que permiten obtener una solución cuantitativa para este problema.

Cuando se emplea el rango $\bar{R}$ como estimador de σ para la elaboración de una carta $\bar{X}$, y como se verá más adelante, para una carta R, la tabla II permite determinar el número mínimo, m, de muestras de tamaño n que se deben emplear para tener poco más de un 98% de nivel de confianza de que los promedios aritméticos obtenidos de las muestras se encuentren dentro de los límites de control que se calculen para la carta $\bar{X}$, suponiendo únicamente la presencia de variación aleatoria.

De la misma manera, se establecen en la tabla III los valores óptimos de m y n, cuando se emplean las desviaciones estándar de las muestras para obtener el estimador $\bar{\sigma}$ de la desviación estándar de la población.

<u>Ejemplo:</u> Sea una fábrica que produce varillas de acero, en la cual se desea ejercer control sobre el peso de las mismas. Para ello, se seleccionan veinte muestras aleatorias de cinco varillas cada una, obteniéndose los valores que se reportan en la tabla siguiente:

Número de la muestra	Valores individuales del peso, Kg					Promedio Aritmético $\bar{X}$	Rango R	Desviación estándar S_x
	X_1	X_2	X_3	X_4	X_5			
1	11.1	9.4	11.2	10.4	10.1	10.44	1.8	0.6651
2	9.6	10.8	10.1	10.8	11.0	10.46	1.4	0.5276
3	9.7	10.0	10.0	9.8	10.4	9.98	0.7	0.2400
4	10.1	8.4	10.2	9.4	11.0	9.82	2.6	0.8727
5	12.4	10.0	10.7	10.1	11.3	10.90	2.4	0.8832
6	10.1	10.2	10.2	11.2	10.1	10.36	1.1	0.4224
7	11.0	11.5	11.8	11.0	11.3	11.32	0.8	0.3059
8	11.2	10.0	10.9	11.2	11.0	10.86	1.2	0.4454
9	10.6	10.4	10.5	10.5	10.9	10.58	0.5	0.1720
10	8.3	10.2	9.8	9.5	9.8	9.52	1.9	0.6493
11	10.6	9.9	10.7	10.2	11.4	10.56	1.5	0.5083
12	10.8	10.2	10.5	8.4	9.9	9.96	2.4	0.8357
13	10.7	10.7	10.8	8.6	11.4	10.44	2.8	0.9562
14	11.3	11.4	10.4	10.6	11.1	10.96	1.0	0.3929
15	11.4	11.2	11.4	10.1	11.6	11.14	1.5	0.5352
16	10.1	10.1	9.7	9.8	10.5	10.04	0.8	0.2800
17	10.7	12.8	11.2	11.2	11.3	11.44	2.1	0.7116
18	11.9	11.9	11.6	12.4	11.4	11.84	1.0	0.3382
19	10.8	12.1	11.8	9.4	11.6	11.14	2.7	0.9708
20	12.4	11.1	10.8	11.0	11.9	11.44	1.6	0.6086
SUMA						213.20	31.80	11.3211

Solución: Puesto que se desconoce la media del proceso, esta se - puede estimar en forma insesgada mediante

$$\bar{\bar{X}} = \frac{1}{20} \sum_{i=1}^{20} \bar{X}_i$$

Los valores de los promedios aritméticos $\bar{X}_i$ (i=1,2,...20) de las muestras se reportan en la tabla anterior, por lo cual la línea central es

$$\bar{\bar{X}} = \frac{1}{20} (213.20) = 10.66$$

Se obtendrán ahora los límites inferior y superior de - control estimando primero a σ mediante los rangos de las muestras, y después mediante las desviaciones estándar - correspondientes.

a. Estimando a σ mediante los rangos de las muestras

El valor de $\bar{R}$ es

$$\bar{R} = \frac{1}{20} \sum_{i=1}^{20} R_i$$

Los valores R_i para i=1,2,...,20 se encuentran en - la tabla inicial, por lo que

$$\bar{R} = \frac{1}{20} (31.80) = 1.59$$

Los límites de control para la carta de los promedios son

$$\bar{\bar{X}} \pm A_2 \bar{R}$$

Y, de la tabla I, para n=5, se obtiene $A_2 = 0.577$, - quedando

$$10.66 \pm \underbrace{0.577\ (1.59)}_{0.92}$$

O sea

Línea Central —— 10.66

Límites de Control —— 10.66±0.92 $\Rightarrow$ 11.58, 9.74

b. Estimando a σ mediante las desviaciones estándar de las muestras

El valor de $\bar{\sigma}$ es

$$\bar{\sigma} = \frac{1}{20}\ (11.3211) = 0.57$$

Los límites de control son ahora

$$\bar{\bar{X}} \pm A_1\ \bar{\sigma}$$

De la tabla I, para n=5, se obtiene

A_1 = 1.596, quedando

$$10.66 \pm \underbrace{1.596(0.57)}_{0.91}$$

O sea

Línea Central —— 10.66

Límites de Control —— 10.66±0.91 $\Rightarrow$ 11.57, 9.75

En la Fig 5 que se presenta a continuación se muestra la carta de control obtenida empleando ambos procedimientos.

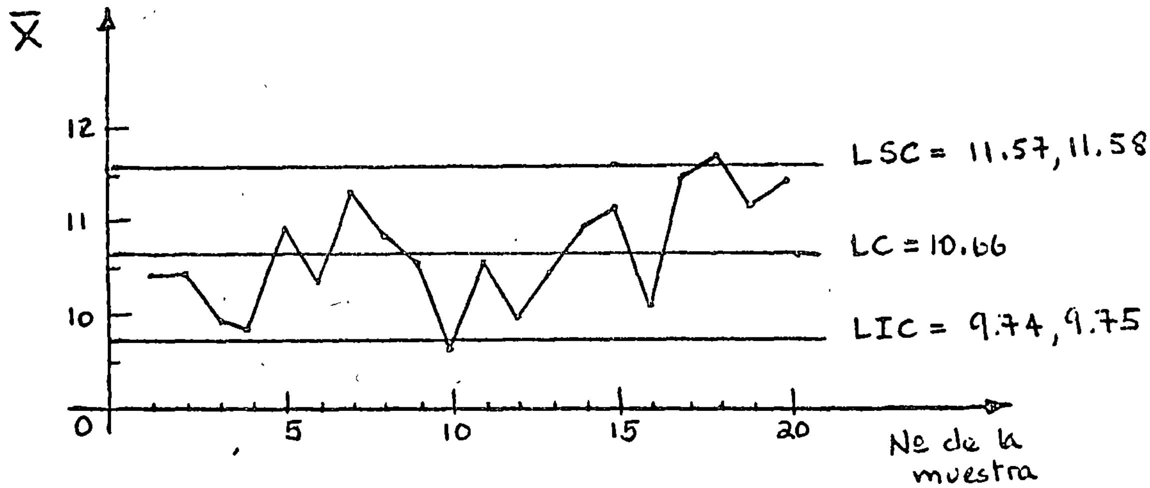


Fig 5 Carta de control $\bar{X}$ obtenida para el ejemplo de las varillas de acero

CARTAS PARA CONTROLAR LA VARIABLILIDAD DE UN PROCESO

Al controlar estadísticamente un proceso puede no ser suficiente - fijar la atención en su "calidad media", sino también en la variabilidad del mismo. Aun cuando es razonable suponer que un incremento en las fluctuaciones de los valores de los promedios aritméticos graficados en una carta $\bar{X}$ se relaciona con un incremento en la variabilidad del proceso, es posible determinar con mayor objetividad y precisión los cambios que experimenta ésta mediante el empleo de las llamadas cartas R y σ, que se elaboran a partir de los rangos y las desviaciones estándar de las muestras, respectivamente.

Conviene mencionar que aun cuando cualquiera de las dos cartas mencionadas permite ejercer control estadístico sobre la variabilidad de un proceso, usualmente se prefiere la carta para los rangos, R, ya que su elaboración es más sencilla que la de σ, que corresponde a las desviaciones estándar. Por otra parte, la carta R conduce a resultados altamente confiables, a la vez que muestra con claridad ciertas tendencias de los valores de las muestras que deben investigarse.

IMPORTANCIA DEL CONTROL DE LA VARIABILIDAD DE UN PROCESO

La importancia del control sobre la variabilidad de un proceso mediante el empleo de las cartas para los rangos o las desviaciones estándar, se hace evidente al considerar que un cambio brusco en aquella característica es de consecuencias más serias que un cambio similar en la "calidad media". Si el proceso experimenta un cambio en ésta última, normalmente se puede regresar al punto de partida efectuando ajustes simples en los dispositivos de producción (por ejemplo, recalibración de herramientas de corte, dosificadoras, etc).

Sin embargo, si el proceso sufre un cambio brusco en su variabilidad, para regresar al punto de partida son necesarios ajustes más costosos y tardados, tales como reparaciones mayores en los dispositivos de producción, o inclusive la compra de un nuevo dispositivo de procesamiento.

Los cambios efectivos en la variabilidad de un proceso afectan necesariamente el desempeño de una carta $\bar{X}$, ya que, como se recordará, los límites de control para la carta de los promedios se establecen a partir de los valores $\bar{R}$ o $\bar{\sigma}$ que se suponen, después de ser afectados por los factores de corrección correspondientes, como buenos estimadores de la desviación estándar del proceso. Si los valores del rango y la desviación estándar de las muestras aumentan, se hace evidente que la carta $\bar{X}$ no operará correctamente.

En contraste con lo anterior, los cambios significativos que se verifican en la carta $\bar{X}$ no necesariamente provocan efectos similares en las cartas R y σ, ya que en la elaboración de ellas no intervienen los promedios aritméticos de las muestras, tal como se verá a continuación.

Por lo anteriormente expuesto, es conveniente ejercer, cuando así sea posible, control simultáneo sobre la "calidad media" y la variabilidad de un proceso.

ELABORACION DE LAS CARTA DE CONTROL PARA LOS RANGOS (CARTA R)

Al igual que para la carta $\bar{X}$, se pueden considerar dos casos distintos en la elaboración de la carta para los rangos: cuando se conoce la desviación estándar σ del proceso y cuando ésto no sucede. En cualquiera de los casos anteriores, se debe observar siempre que el procedimiento de obtención de la línea central y de los límites de control para la carta R, se basa en la distribución muestral de los rangos de muestras aleatorias de tamaño n, extraídas de una población normal.

a. Caso en el que se conoce la desviación estándar σ de la - - Población

De acuerdo con lo anterior, es fácil comprender que los pará metros de la carta de control para los rangos son

Línea Central —— μ_R

Límites de Control —— $\mu_R \pm 3\sigma_R$

Sin embargo, normalmente no conocen los valores de la media y la desviación estándar de la distribución muestral de los rangos. En esta situación, la lógica indica que para estimar el valor de μ_R se debe emplear el de $\bar{R}$, el promedio de los rangos de muestras preliminares. Sin embargo, si se recuerda que

$$\sigma = \frac{\bar{R}}{d_2}$$

entonces

$$\bar{R} = d_2 \ \sigma$$

Y, puesto que se conoce el valor de σ, se puede escribir

Línea Central —— $\bar{R}$ o $d_2\sigma$

quedando finalmente

Línea Central —— $d_2\sigma$

en donde los valores de d_2 se presentan en la tabla I.

Por lo que respecta a σ_R , si se observa nuevamente la Fig 4 se puede ver que la desviación estándar de la distribución mues ral de la estadística R/σ, para el caso de muestras de tamaño 5 es, en forma experimental

$$\sigma_{R/\sigma} = d_3 = 0.864$$

Lo anterior permite considerar que si σ es conocida (y por tan to constante) es válido escribir

$$\sigma_{R/\sigma} = \frac{\sigma_R}{\sigma} d_3$$

o sea

$$\sigma_R = \sigma_{R/\sigma}\ \sigma = d_3\sigma = 0.864\ \sigma$$

En el caso en que n sea diferente de cinco, los valores del factor d_3 se pueden obtener de la tabla I.

Empleando el valor de σ_R así obtenido, los límites de control son, en general, los siguientes

$$d_2\sigma \pm 3d_3\sigma$$

o sea

$$d_2\sigma - 3d_3\sigma \Rightarrow (d_2 - 3d_3)\sigma \Rightarrow D_1\ \sigma$$

$$d_2\sigma \quad 3d_3\sigma \Rightarrow (d_2 + 3d_3)\sigma \Rightarrow D_2\ \sigma$$

en donde

$$D_1 = d_2 - 3d_3 \quad y \quad D_2 = d_2 + 3d_3$$

Los valores de D_1 y D_2 se reportan también en la tabla I en función de n, el tamaño de la muestra.

Conforme a lo anterior, los parámetros de la carta de control para los rangos, cuando σ es conocida, son

Línea Central —— $d_2\ \sigma$

Límite Inferior de Control —— $D_1\ \sigma$

Límite Superior de Control —— $D_2\ \sigma$

b. Caso en el que se desconoce la desviación estándar σ de la población

En este caso en necesario estimar a μ_R de la distribución muestral de los rangos mediante $\bar{R}$, empleando un número adecuado de muestras preliminares, normalmente el mismo que se emplea para la elaboración de una carta $\bar{X}$. Al respecto, conviene recordar que la carta R (o la σ) generalmente se construye <u>después</u> de la carta $\bar{X}$, y que, por lo tanto, se emplean para su elaboración las mismas muestras aleatorias. De acuerdo con ésto, la línea central resulta ser

$$\text{Línea Central} \text{ —— } \bar{R}$$

En este caso se requieren límites de control del tipo

$$\bar{R} \pm 3\sigma_R$$

Puesto que ahora se desconocen σ_R y σ, se pueden hacer, para el límite inferior de control

$$\bar{R} - 3\sigma_R = \bar{R} - \frac{3\,\bar{R}\,\sigma_R}{\bar{R}} = \left(1 - 3\,\frac{\sigma_R}{\bar{R}}\right)\bar{R}$$

$$= \left(1 - 3\,\frac{\frac{\sigma_R}{\sigma}}{\frac{\bar{R}}{\sigma}}\right)\bar{R} = \left(1 - 3\,\frac{d_3}{d_2}\right)\bar{R}$$

$$= \left(\frac{d_2 - 3d_3}{d_2}\right)\bar{R} = \left(\frac{D_1}{d_2}\right)\bar{R}$$

Para el límite superior de control se obtiene

$$\bar{R} + 3\sigma_R = \bar{R}\left(\frac{D_2}{d_2}\right)$$

En la tabla I se presentan los valores de

$$D_3 = \frac{D_1}{d_2} \qquad y \qquad D_4 = \frac{D_2}{d_2}$$

en función de n .

Finalmente, los parámetros de la carta R cuando se desconoce el valor de σ de la población son los siguientes:

Línea Central —— $\bar{\bar{R}}$

Límite Inferior de Control —— $D_3\bar{\bar{R}}$

Límite Superior de Control —— $D_4\bar{\bar{R}}$

ELABORACION DE LA CARTA DE CONTROL PARA LAS DESVIACIONES ESTANDAR (CARTA σ)

En la elaboración de la carta para las desviaciones estándar también se deben considerar los dos casos posibles: cuando se conoce la - desviación estándar de la población y cuando ésto no es así. De - igual manera, el procedimiento para obtener los paráme‿tros de la carta se fundamenta en la distribución muestral de las desviaciones estándar de muestras aleatrias de tamaño n, extraídas de una población normal.

a. Caso en el que se conoce la desviación estándar σ de la población

Con base en la distribución muestral de las desviaciones estándar de las muestras, se pueden establecer los parámetros de la carta σ, a saber

Línea Central —— μ_{s_x}

Límites de Control —— $\mu_{s_x} \pm 3\sigma_{s_x}$

Al desconocerse, como ocurre normalmente, los valores de μ_{s_x} y σ_{s_x} de la distribución muestral, se debe estimar primero μ_{s_x} a partir de $\bar{\sigma}$, el promedio de las desviaciones estándar de las muestras preliminares. Sin embargo, no es necesario realizar en este caso ese cálculo si se recuerda que

$$\sigma = \frac{\bar{\sigma}}{c_2}$$

o sea

$$\bar{\sigma} = c_2\ \sigma$$

Y, en virtud de que el valor de σ es conocido, se llega a

Línea Central —— $\bar{\sigma}$ o $c_2\sigma$

quedando finalmente

Línea Central —— $c_2\sigma$

en donde los valores de c_2 se pueden obtener de la tabla I.

Bajo la suposición de que la población de la cual se extraen las muestras aleatorias se encuentra distribuida en forma normal (o aproximadamente normal), se puede demostrar que la desviación estándar de la distribución muestral de las desviaciones estándar es

$$\sigma_{s_x} = \frac{\sigma}{\sqrt{2n}}$$

en donde n denota al tamaño de las muestras. Empleando el va lor de σ_{S_X} anterior, los límites de control se pueden establecer como

$$\mu_{S_X} \pm 3\sigma_{S_X} = c_2\sigma \pm 3\frac{\sigma}{\sqrt{2n}}$$

o sea

$$c_2\sigma - 3\frac{\sigma}{\sqrt{2n}} = \left(c_2 - \frac{3}{\sqrt{2n}}\right)\sigma = B_1\sigma$$

$$c_2\sigma + 3\frac{\sigma}{\sqrt{2n}} = \left(c_2 + \frac{3}{\sqrt{2n}}\right)\sigma = B_2\sigma$$

en donde

$$B_1 = c_2 - \frac{3}{\sqrt{2n}}$$

$$B_2 = c_2 + \frac{3}{\sqrt{2n}}$$

Los valores de B_1 y B_2 se proporcionan en la tabla I, en función del valor de n. Entonces, los parámetros de la carta σ son, finalmente

Línea Central —— $c_2\sigma$

Límite Inferior de Control —— $B_1\sigma$

Límite Superior de Control —— $B_2\sigma$

b. Caso en el que se desconoce la desviación estándar σ de la población

En este caso es necesario estimar a μ_{S_X} mediante $\bar{\sigma}$, empleando un número suficiente de muestras aleatorias preliminares.

De acuerdo con lo anterior, la línea central de la carta σ es

Línea Central —— $\bar{\sigma}$

Los límites de control serán entonces del tipo

$$\bar{\sigma} \pm 3\sigma_{s_X}$$

Puesto que ahora se desconoce el valor de σ, pero se sabe que

$$\sigma = \frac{\bar{\sigma}}{c_2}$$

el límite inferior de control resulta ser

$$\bar{\sigma} - 3\sigma_{s_X} = \bar{\sigma} - 3\,\frac{\sigma}{\sqrt{2n}} = \bar{\sigma} - 3\,\frac{\bar{\sigma}}{c_2\sqrt{2n}}$$

$$= \left(1 - \frac{3}{c_2\sqrt{2n}}\right)\bar{\sigma}$$

Para el límite superior de control se obtiene

$$\bar{\sigma} + 3\sigma_{s_X} = \left(1 + \frac{3}{c_2\sqrt{2n}}\right)\bar{\sigma}$$

En la tabla I se presentan los valores de

$$B_3 = 1 - \frac{3}{c_2\sqrt{2n}} \quad \text{y} \quad B_4 = 1 + \frac{3}{c_2\sqrt{2n}}$$

en función del valor de n.

Finalmente, los parámetros de la carta σ, cuando no se conoce la desviación estándar de la población, quedan como

Línea Central —— $\bar{\sigma}$

Límite Inferior de Control —— $B_3\bar{\sigma}$

Límite Superior de Control —— $B_4\bar{\sigma}$

Ejemplo: Sea el proceso de elaboración de varillas de acero mencionado en la página 10 de estos apuntes. En él se informa que el diámetro medio de las varillas es igual a 2.5 cm, con desviación estándar de 0.01 cm. En este caso se pide establecer los parámetros de las cartas de control R y σ, considerando que se extraen periódicamente muestras de cinco varillas.

Solución:

a. Carta R

Puesto que se conoce el valor de la desviación estándar de la población, y en virtud de que n=5, se obtiene, empleando la tabla I

$$\text{LC} \text{ —— } d_2\sigma = 2.326(0.01) = 0.02326$$

$$\text{LIC} \text{ —— } D_1\sigma = 0(0.01) = 0.0000$$

$$\text{LSC} \text{ —— } D_2\sigma = 4.918(0.01) = 0.04918$$

b. Carta σ

En este caso, puesto que σ=0.01 y n=5, se obtiene, con el uso de la tabla I

$$\text{LC} \text{ —— } c_2\sigma = 0.8407(0.01) = 0.008407$$

$$\text{LIC} \text{ —— } B_1\sigma = 0(0.01) = 0.00000$$

$$\text{LSC} \text{ —— } B_2\sigma = 1.756(0.01) = 0.01756$$

<u>Ejemplo</u>: Con el fin de investigar la variabilidad en el proceso de producción de varillas de acero mencionado en la página 16, se desea elaborar las cartas de control R y σ correspondientes, considerando la información contenida en la tabla de la misma página.

<u>Solución</u>:

En este caso se desconoce la desviación estándar de la población, por lo cual es indispensable emplear los valores de $\bar{R}$ y $\bar{\sigma}$, considerando que el tamaño de la muestra es 5.

a. Carta R

El valor de $\bar{R}$, obtenido durante el proceso de elaboración de la carta $\bar{X}$ correspondiente, es $\bar{R} = 1.59$. Considerando este valor, y empleando la tabla I, los parámetros de la carta de control R resultan

$$\text{LC} \text{ —— } \bar{R} = 1.590$$

$$\text{LIC} \text{ —— } D_3\bar{R} = 0(1.59) = 0.000$$

$$\text{LSC} \text{ —— } D_4\bar{R} = 2.115(1.59) = 3.363$$

En la Fig 6 se presenta la carta R para este problema.

b. Carta σ

Considerando que al calcular para este problema los parámetros de la carta $\bar{X}$ se obtuvo $\bar{\sigma} = 0.57$, la carta σ queda definida con

$$\text{LC} \text{ —— } \bar{\sigma} = 0.57$$

$$\text{LIC} \text{ —— } B_3\bar{\sigma} = 0(0.57) = 0.00$$

$$\text{LSC} \text{ —— } B_4\bar{\sigma} = 2.089(0.57) = 1.19$$

En la Fig 7 se muestra la carta de control σ correspondiente.

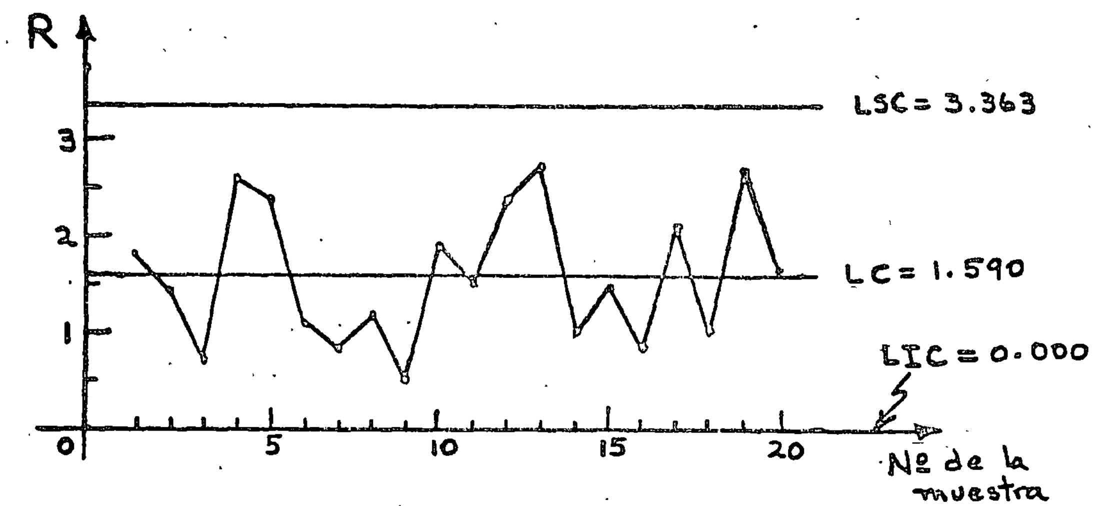


Fig 6 Carta de control R obtenida para el ejemplo de las varillas de acero

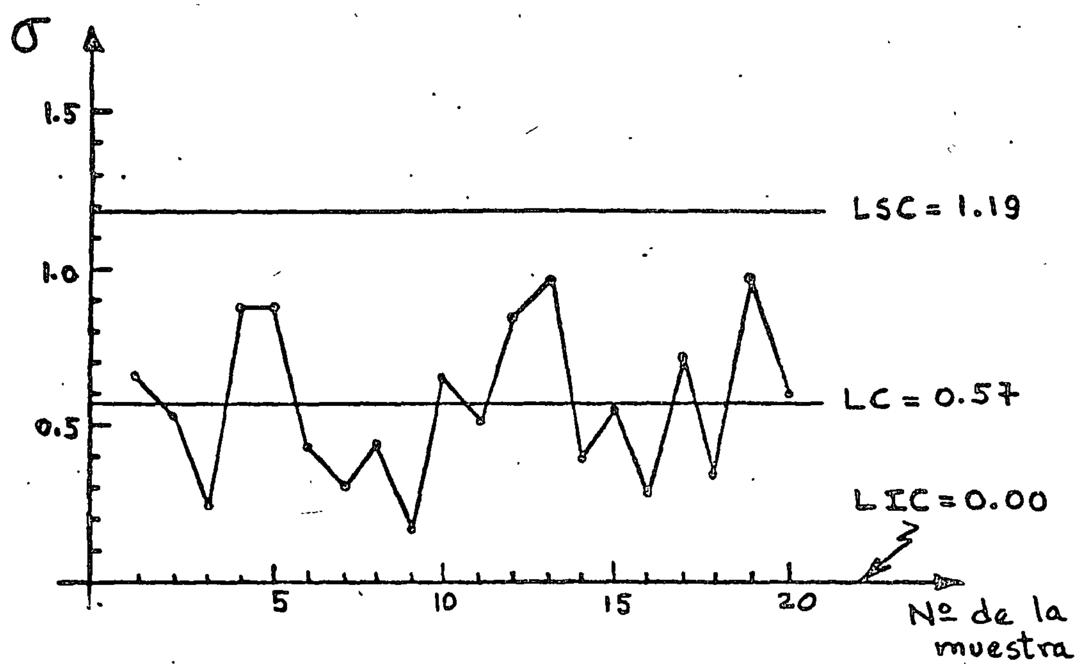


Fig 7 Carta de control σ obtenida para el ejemplo de las varillas de acero

CARTAS DE CONTROL PARA MEDICIONES (ELEMENTOS INDIVIDUALES)

Se han establecido las cartas $\bar{X}$, R y σ considerando que existe la posibilidad de conocer la media μ y/o la desviación estándar σ de la población (proceso), o bien, cuando estos parámetros se desconocen, que es posible obtener un número adecuado de muestras aleatorias de ella, cuyos tamaños sean cuando menos igual a dos, con el fin de estimar con buena precisión los valores de dichos parámetros.

Sin embargo, en muchas ocasiones no se conocen los parámetros del proceso, y únicamente es posible contar con muestras de tamaño uno, es decir, muestras con un solo elemento. Cuando ésto sucede, la técnica para calcular los límites de control en las cartas para mediciones se fundamenta en el empleo de los llamados rangos móviles, que se explican a continuación.

Si, por ejemplo, se cuenta con el conjunto de datos X_i $(i=1,2,...,n)$ registrados en orden, se definen los rangos móviles de orden dos como

$$\left|X_i - X_{i+1}\right| \qquad ; \qquad 1 \leq i \leq n-1$$

es decir

$$\left|X_1 - X_2\right| , \left|X_2 - X_3\right| , \ldots , \left|X_{n-1} - X_n\right|$$

Si se trata de rangos móviles de orden tres, éstos se definen como

$$\left|X_i - X_{i+2}\right| \qquad ; \qquad 1 \leq i \leq n-2$$

es decir

$$\left|X_1 - X_3\right| , \left|X_2 - X_4\right| , \ldots , \left|X_{n-2} - X_n\right|$$

La obtención de los rangos móviles de orden superior al tres se hace siguiendo las ideas anteriores.

En forma numérica, si se tienen los datos registrados en orden 4, 6, 4, 3 y 7, los rangos móviles de orden dos son

$$|4 - 6| = 2\ , \quad |6 - 4| = 2\ , \quad |4 - 3| = 1\ , \quad |3 - 7| = 4$$

y los de orden tres son

$$|4 - 4| = 0\ , \quad |6 - 3| = 3\ , \quad |4 - 7| = 3$$

El empleo de los rangos móviles para la obtención de los límites de control es importante en este caso, debido a que, si se trata de rangos móviles de orden dos, se puede considerar que el valor de cualquiera de ellos debe obtenerse a partir de los valores de dos elementos individuales registrados en orden. Dicho de otra manera, un rango móvil de orden dos debe provenir de una muestra "ficticia" de tamaño dos. En la misma forma, un rango móvil de orden tres tiene que obtenerse a partir de tres elementos individuales, lo cual permite "crear" muestras de tamaño tres.

De acuerdo con lo anterior, es factible establecer los límites de control para las cartas de control, en el caso de elementos individuales, empleando los factores de la tabla I, que se encuentran tabulados a partir de muestras de tamaño dos.

a. Elaboración de la carta X (elementos individuales)

En este caso, la línea central está dada por

$$\bar{X} = \frac{1}{K} \sum_{i=1}^{K} X_i$$

en donde X_i $(i=1,2,...,K)$ denota a los valores de los datos individuales.

Los límites de control requeridos son

$$\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}}$$

Puesto que el tamaño real de la muestra es uno, la expresión - anterior se puede escribir

$$\bar{X} \pm 3\frac{\sigma}{\sqrt{1}} = \bar{X} \pm 3\sigma$$

Debido a que el valor de σ se desconoce, pero es posible obtener el de $\bar{R}$ (promedio de los rangos móviles), la última expresión puede transformarse algebraicamente de la siguiente manera:

$$\bar{X} \pm 3\sigma = \bar{X} \pm \frac{3\sigma\,\bar{R}}{\bar{R}} = \bar{X} \pm \frac{3\bar{R}}{\frac{\bar{R}}{\sigma}}$$

$$\bar{X} \pm \frac{3\,\bar{R}}{d_2} = \bar{X} \pm E_2\bar{R}$$

en donde

$$E_2 = \frac{3}{d_2}$$

Los valores de E_2 se pueden obtener de la tabla I en función de n, que representa ahora el tamaño "ficticio" de la muestra, o el orden de los rangos móviles.

De acuerdo con lo anterior, los parámetros de la carta de control X para elementos individuales son

Línea Central —— $\bar{X}$

Límite Inferior de Control —— $\bar{X} - E_2\bar{R}$

Límite Superior de Control —— $\bar{X} + E_2\bar{R}$

b. Elaboración de la carta R* (rangos móviles)

En este caso, la línea central está dada por el valor del promedio de los rangos móviles, es decir

$$\bar{R} = \frac{1}{K} \sum_{i=1}^{K} R_1$$

En donde R_i $(i=1,2,...,K)$ denota a los valores de los rangos móviles, obtenidos a partir de los datos individuales registrados en orden.

Los límites de control se obtienen considerando que se desconoce el valor de la desviación estándar de la población, en la forma ya explicada para la carta R.

De acuerdo con lo anterior, los parámetros de la carta de control R* para los rangos móviles son

Línea Central —— $\bar{R}$

Límite Inferior de Control —— $D_3\bar{R}$

Límite Superior de Control —— $D_4\bar{R}$

en donde los valores de D_3 y D_4 se obtienen de la tabla I en función de n, el tamaño "ficticio" de la muestra, u orden de los rangos móviles.

Ejemplo: Considérese un proceso de destilación y mezclado de alcohol, para el cual se desea ejercer control sobre el porcentaje de metanol existente. Se extraen 26 lotes sucesivos de alcohol, y se obtiene el porcentaje de metanol correspondiente para cada uno de ellos. Los valores se presentan en la tabla siguiente, y se pide construir cartas X y R* considerando rangos móviles de orden dos.

Lote	Porcentaje de metanol, X	Rango móvil,R	Lote	Porcentaje de metanol, X	Rango móvil, R
1	4.6	---	14	5.5	0.1
2	4.7	0.1	15	5.2	0.3
3	4.3	0.4	16	4.6	0.6
4	4.7	0.4	17	5.5	0.9
5	4.7	0	18	5.6	0.1
6	4.6	0.1	19	5.2	0.4
7	4.8	0.2	20	4.9	0.3
8	4.8	0	21	4.9	0
9	5.2	0.4	22	5.3	0.4
10	5.0	0.2	23	5.0	0.3
11	5.2	0.2	24	4.3	0.7
12	5.0	0.2	25	4.5	0.2
13	5.6	0.6	26	4.4	0.1
			SUMA	128.1	7.2

<u>Solución</u>: El valor del promedio de los rangos móviles de orden dos es

$$\bar{R} = \frac{1}{25} \sum_{i=1}^{25} R_i = \frac{1}{25} (7.2) = 0.288$$

a. Carta X

La línea central de esta carta es $\bar{X}$, cuyo valor es

$$\bar{X} = \frac{1}{26} \sum_{i=1}^{26} X_i = \frac{1}{26} (128.1) = 4.927$$

De la tabla I se obtiene $E_2 = 2.66$ para n=2, - siendo los límites de control

$$\bar{\bar{X}} \pm E_2\bar{R} = 4.927 \pm 2.66(0.288)$$

$$= 4.927 \pm 0.7661$$

Finalmente, los parámetros de la carta X quedan como

LC —— 4.927

LIC—— 4.927 - 0.7661 = 4.161

LSC—— 4.927 + 0.7661 = 5.693

En la Fig 8 se presenta la gráfica correspondiente.

b. Carta R*

La línea central para esta carta es $\bar{R} = 0.288$, y los límites de control se obtienen empleando la tabla I considerando que n=2. De ahi que

LC —— 0.288

LIC—— $D_3\bar{R} = 0(0.288) = 0.000$

LSC—— $D_4\bar{R} = 3.267(0.288) = 0.941$

La Fig 9 muestra la carta R* para este problema.

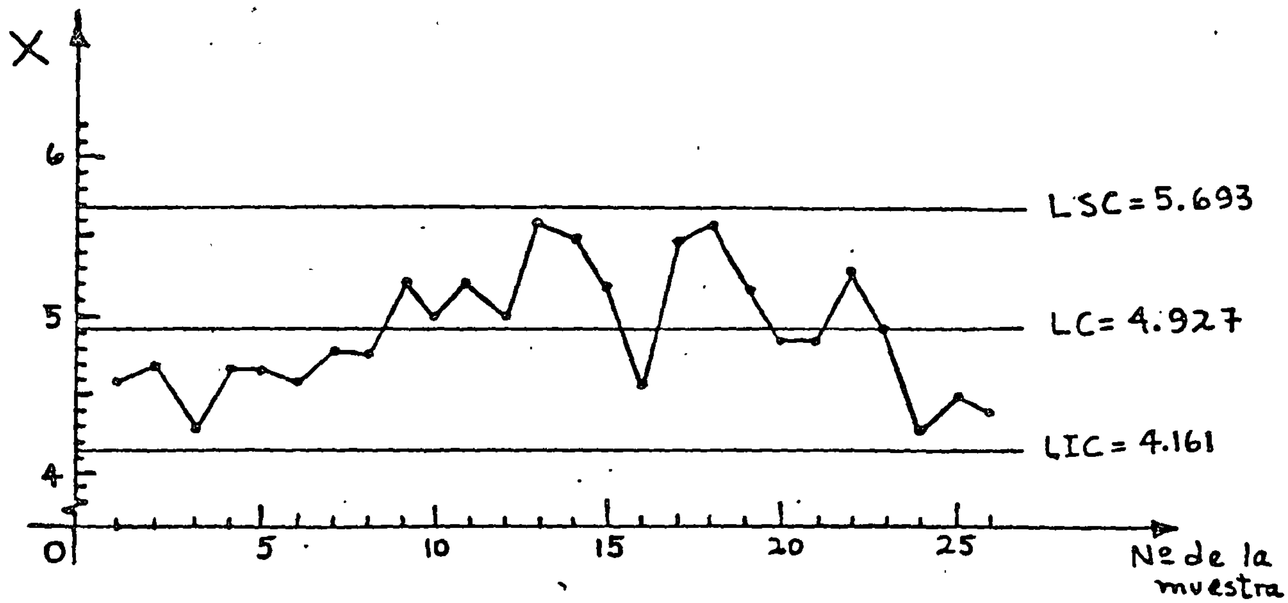


Fig 8 Carta de control X obtenida para el ejemplo de los lotes de alcohol

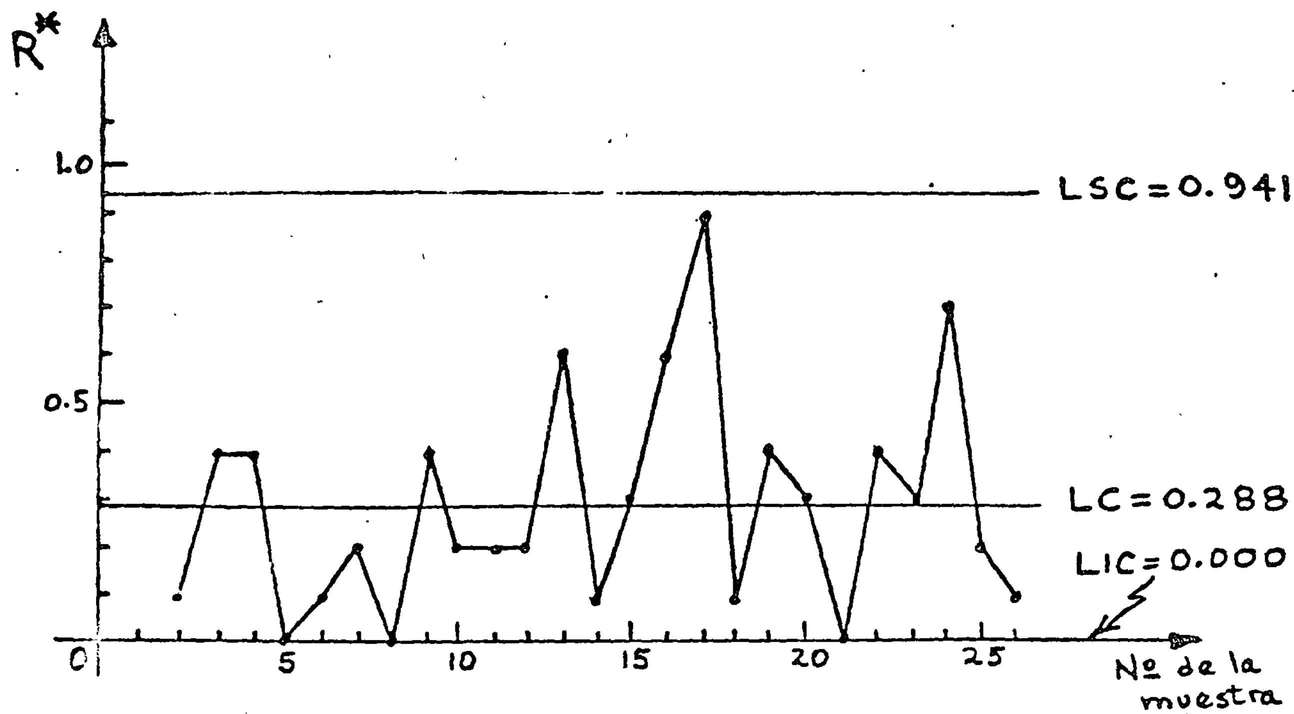


Fig 9 Carta de ccntrol R* obtenida para el ejemplo de los lotes de alcohol

CARTAS DE CONTROL PARA ATRIBUTOS

El término atributo, tal como se emplea en el control de calidad, indica la propiedad que tiene un producto de ser bueno o malo, es decir, permite reconocer si la característica de calidad del mismo se encuentra dentro de ciertos requerimientos específicos o no.

Aunque generalmente se puede obtener información más completa de las mediciones hechas a productos terminados, a menudo consume menos tiempo y dinero el comparar la calidad de un producto en contra de ciertas especificaciones mínimas, sobre la base, por ejemplo, de considerar que sirve o no, o que es bueno o malo.

Por ejemplo, al ejercer control sobre el diámetro de un balín de acero, es más simple y rápido el determinar si éste pasa por un agujero hecho en una placa de acero templado con el diámetro adecuado, que realizar la medición del diámetro con un micrómetro.

Se establecerán ahora los dos tipos fundamentales de cartas de control que se utilizan en conexión con el muestreo por atributos: la carta para la proporción de elementos defectuosos, o carta p, y la carta para el número de defectos, o carta c.

Considérese por ejemplo una muestra de 50 fusibles en la cual se encontró, después de probar todos ellos, que contiene dos elementos defectuosos. En este caso, la proporción de fusibles defectuosos en la muestra es de 2/50 = 0.04.

Por otra parte, debe observarse que si se prueba una sola unidad producida, esta puede tener varios defectos pero, sin embargo, puede o no ser una unidad defectuosa. Tal es el caso, por ejemplo, de rollos (unidades) de tela de determinada longitud, que pueden tener cierto número de imperfecciones pero no necesariamente ser considerados como defectuosos. No obtante, en muchas aplicaciones prácticas una unidad producida se considera defectuosa si tiene - cuando menos un defecto.

La distribución de la proporción y del número de elementos defectuosos en un proceso es obviamente binomial, en tanto que la del número de defectos es de Poisson. Sin embargo, para la elaboración de la carta p se aprovecha la propiedad que tiene la distribución muestral de las proporciones de ser aproximada mediante una distribución normal cuando el tamaño de la muestra es grande, y la proporción de elementos defectuosos no se acerca a cero o a uno.

ELABORACION DE LAS CARTAS DE CONTROL p Y np PARA LA PROPORCION DE - DEFECTUOSOS Y EL NUMERO DE DEFECTUOSOS

Los límites de control que se requieren en este caso son

$$\mu_p \pm 3\sigma_p$$

en donde μ_p es la media de la distribución muestral de las proporciones, y σ_p la desviación estándar correspondiente. Como μ_p de - esta distribución es igual al parámetro P de la población, la estadística p de la muestra estima en forma insesgada a este último.

Si no se conoce el valor de P de la población, lo cual en la práctica es frecuente, se debe disponer de K muestras de tamaño n constante para obtener el valor del estimador insesgado

$$\bar{p} = \frac{1}{K} \sum_{i=1}^{K} p_i$$

en donde p_i $(i=1,2,...,K)$ denota el valor de la proporción en la muestra i. Empleando el valor así obtenido, la línea central es

Línea Central —— $\bar{p}$

En textos de estadística se demuestra que la desviación estándar de la distribución muestral de las proporciones es

$$\sigma_p = \sqrt{\frac{P\ (1-P)}{n}}$$

por lo cual los límites de control son

$$\bar{p} \pm 3\sigma_p = \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Finalmente, los parámetros de la carta de control p quedan como

Línea Central —— $\bar{p}$

Límite Inferior de Control —— $\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

Límite Superior de Control —— $\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

A partir de los parámetros anteriores se pueden derivar los de la llamada carta np, o sea, para el número de defectuosos. Para ello, es necesario multiplicar dichos parámetros por n para así obtener, en el caso de los límites de control

$$n\left(\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right) = n\bar{p} \pm 3n\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= n\bar{p} \pm 3\sqrt{\frac{n^2\bar{p}\ (1-\bar{p})}{n}}$$

$$= n\bar{p} \pm \sqrt{3\ n\ \bar{p}\ (1-\bar{p})}$$

y los parámetros resultan ahora

Línea Central —— $n\bar{p}$

Límite Inferior de Control —— $n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$

Límite Superior de Control —— $n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$

Ejemplo: Para un proceso de elaboración de fusibles se desea ejercer control sobre la proporción de elementos defectuosos, así como sobre el número de ellos. Para ello, se seleccionan 40 muestras aleatorias de 50 fusibles cada una, y se obtienen los valores reportados en la tabla siguiente.

Se desea construir las cartas p y np correspondientes.

Número de la muestra	Número de fusibles defectuosos	Proporción de defectuosos, p	Número de la muestra	Número de fusibles defectuosos	Proporción de defectuosos, p
1	2	0.04	21	1	0.02
2	1	0.02	22	1	0.02
3	2	0.04	23	4	0.08
4	0	0.00	24	2	0.04
5	2	0.04	25	2	0.04
6	3	0.06	26	4	0.08
7	4	0.08	27	1	0.02
8	2	0.04	28	3	0.06
9	0	0.00	29	3	0.06
10	3	0.06	30	2	0.04
11	0	0.00	31	3	0.06
12	1	0.02	32	6	0.12
13	2	0.04	33	2	0.04
14	2	0.04	34	3	0.06
15	3	0.06	35	2	0.04
16	5	0.10	36	3	0.06
17	1	0.02	37	1	0.02
18	2	0.04	38	0	0.00
19	3	0.06	30	2	0.04
20	1	0.02	40	0	0.00
			S U M A		1.68

Solución: El valor de $\overline{p}$ es

$$\overline{p} = \frac{1}{40} \sum_{i=1}^{40} p_i = \frac{1}{40} (1.68) = 0.042$$

a. Carta p

Los límites de control son, para n=50

$$0.042 \pm 3\sqrt{\frac{(0.042)\ (1-0.042)}{50}} = 0.042 \pm 0.0851$$

por lo cual

$$\text{LC} \longrightarrow 0.0420$$

$$\text{LIC} \longrightarrow 0.042 - 0.0851 = -0.0431 \Rightarrow 0.0000$$

$$\text{LSC} \longrightarrow 0.042 + 0.0851 = 0.1271$$

En este caso, y como se verá a continuación para la carta np, la expresión para el cálculo del límite inferior de control conduce a un valor negativo del mismo. Puesto que no tiene sentido físico hablar de una proporción menor de cero o de un número de defectuosos negativo, en forma arbitraria se asigna a ese límite el valor cero.

En la Fig 10 se presenta la carta de control p correspondiente.

b. Carta np

Puesto que $n\overline{p} = 50(0.042)=2.1$, los límites de control son ahora

$$2.1 \pm 3\sqrt{50(0.042)\ (1-0.042)} = 2.1 \pm 4.255$$

o sea

LC —— 2.1

LIC —— 2.1-4.255 = -2.155 $\Rightarrow$ 0.000

LSC —— 2.1+4.255 = 6.355

En la Fig 10 se presenta la carta np para este problema.

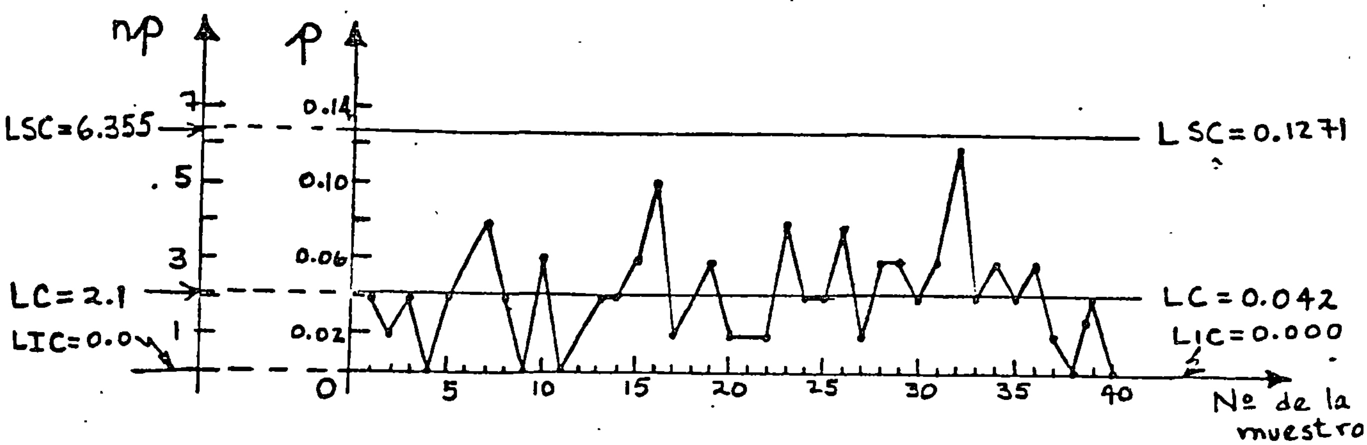


Fig 10 Cartas de control p y np obtenidas para el ejemplo de los fusibles

ELABORACION DE LA CARTA DE CONTROL c PARA EL NUMERO DE DEFECTOS

Existen ocasiones en las que es necesario controlar el número de defectos por unidad en un proceso. Por ejemplo, en la producción de alfombras es importante controlar el número de defectos por metro cuadrado; en la elaboración de papel se requiere controlar el número de defectos por rollo, etc. En estos casos, la variable aleatoria c asociada al número de defectos por unidad tiene una distribución de Poisson.

De lo anterior se desprende que la línea central de la carta de con trol para el número de defectos es el parámetro λ de la distribución de Poisson correspondiente, cuyo valor usualmente se desconoce. En tal situación, se acostumbra estimar en forma insesgada el valor de λ a partir de un mínimo de 20 valores de c, observados previamente en igual número de unidades producidas. De acuerdo con ésto, el valor de

$$\bar{c} = \frac{1}{K} \sum_{i=1}^{K} c_i$$

en donde c_i $(i=1,2,\ldots,K)$ representa el número de defectos observados en la unidad i, se puede emplear como estimador de λ.

Los límites de control requeridos ahora son del tipo

$$\bar{c} \pm 3\sigma_c$$

Puesto que en este caso se observa el número de defectos por unidad, se puede suponer que el tamaño de la muestra es unitario. Por tal motivo, se puede considerar que la desviación estándar de la distribución muestral del número de defectos c es igual a la desviación estándar de la distribución de Poisson y, puesto que $\bar{c}$ estima el valor de λ

$$\sigma_c = \sqrt{\lambda} = \sqrt{\bar{c}}$$

De acuerdo con lo anterior, los parámetros de la carta de control c son

Línea Central —— $\bar{c}$

Límite Inferior de Control —— $\bar{c} - 3\sqrt{\bar{c}}$

Límite Superior de Control —— $\bar{c} + 3\sqrt{\bar{c}}$

Ejemplo: Considérese el proceso de soldadura de dos placas de acero en una fábrica. Diariamente se alcanzan a soldar 8 juntas, y en cada una de ellas se observa el número de defectos existente. Con la información correspondiente a tres días de labor que se presenta en la tabla siguiente, se desea elaborar una carta de control para el número de defectos por junta soldada

Número de la junta soldada	Fecha	Número de defectos
1	Julio 18	2
2		4
3		7
4		3
5		1
6		4
7		8
8		9
9	Julio 19	5
10		3
11		7
12		11
13		6
14		4
15		9
16		9
17	Julio 20	6
18		4
19		3
20		9
21		7
22		4
23		7
24		12
SUMA................		144

Solución: Empleando los valores reportados en la tabla anterior, el valor de $\bar{c}$ resulta

$$\bar{c} = \frac{1}{24} \sum_{i=1}^{24} c_i = \frac{1}{24} (144) = 6$$

Siendo $\bar{c} = 6$, los límites de control quedan como

$$6 \pm 3\sqrt{6} = 6 \pm 7.35$$

Finalmente, los parámetros de la carta c son

LC —— 6

LIC —— 6 - 7.35 = -1.35 ⇒ 0.00

LSC —— 6 + 7.35 = 13.35

Puesto que el número de defectos no puede ser negativo, se fija el valor del límite inferior de control igual a cero.

En la Fig 11 se presenta la carta de control c que corresponde al ejemplo.

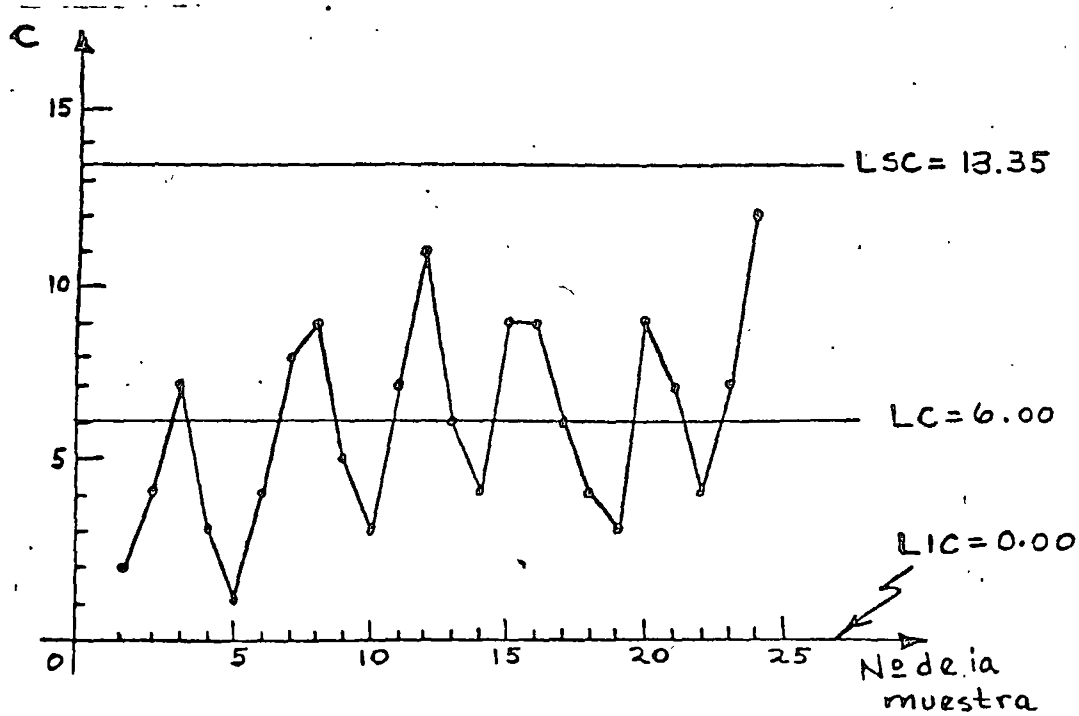


Fig 11 Carta de control c obtenida para el ejemplo de las juntas soldadas

B I B L I O G R A F I A

1. Hansen, B., "Quality Control: Theory and Applications", Prentice Hall, Inc. (1964)

2. Grant, E.L., "Statistical Quality Control", Mc Graw-Hill Book Co. (1971)

3. Ostle, B. "Estadística aplicada", Limusa-Wiley (1973)

4. Miller, I. y Freund, J., "Probability and Statistics - for Engineers", Prentice Hall, Inc. (1965)

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

CONTROL ESTADISTICO DE CALIDAD

MUESTREO DE INSPECCION

M. en I. AUGUSTO VILLARREAL A.

24 OCTUBRE, 1978

MUESTREO DE INSPECCION

Por: M en I Augusto Villarreal Aranda*

1. Introducción

El muestreo de inspección (o de aceptación) se define como el conjunto de todas las acciones que realiza el receptor de producto "terminado" para asegurar la calidad de éste, después de recibirlo del productor.

Este tipo de muestreo puede ser aplicado por un consumidor a los productos que recibe de un vendedor, por un departamento de inspección de producto terminado a los productos recibidos de los departamentos de producción, etc, es decir, se aplica en aquellas ocasiones en que un número grande de unidades producidas se presenta para inspección en forma de lotes, y en donde la forma

* *Secretario Académico*, División de Estudios Superiores, Facultad de Ingeniería, UNAM y *Profesor investigador*, Instituto de Ingeniería, UNAM

lógica de realizar esa tarea es mediante el empleo de la técnica que usa atributos (sirve, no sirve, o pasa, no pasa), con el fin de evitar la tan costosa y tardada inspección al 100%.

Generalmente, con la inspección de lote por lote del producto terminado, existe el acuerdo entre productor y receptor en que

a. los lotes aceptados por el plan de muestreo que se emplee serán aceptados por el receptor como buenos a excepción de aquellas unidades detectadas como defectuosas en todos los lotes durante el proceso de muestreo, las cuales serán reemplazadas por unidades buenas por el productor.

b. los lotes rechazados por el plan de muestreo le serán devueltos al productor para su rectificación.

Sin embargo, existen algunas variantes sobre el acuerdo mencionado. Por ejemplo, algunos receptores de producto terminado emplean la opción de inspeccionar al 100% los lotes rechazados para eliminar los elementos defectuosos, y trasladar el costo de esa operación al productor. Lo anterior se realiza con frecuencia cuando el receptor tiene urgencia de emplear las unidades que recibe del productor. En última instancia el objetivo que se persigue es responsabilizar al productor por la deficiente calidad de un producto terminado.

Para determinar la calidad de un lote, es factible seleccionar una, dos o múltiples muestras aleatorias del mismo, lo cual conduce a considerar planes de muestreo simples, dobles, o múltiples para aceptarlo o rechazarlo. La explicación de cómo y cuando se emplean estos tipos de muestreo se discutirá en esta parte del curso.

2. Plan de muestreo simple

Como se dijo anteriormente, el muestreo de aceptación se aplica a las producciones en masa cuando un *productor* abastece de lotes de artículos a un *receptor*. En situaciones como ésta, se debe decidir individualmente sobre la aceptación o rechazo de cada lote.

En este caso particular, la decisión que se toma se basa en el resultado que se obtiene al inspeccionar una muestra de tamaño "n" que se toma de un lote de "N" artículos, de la cual se determina el número de defectuosos, "X", esto es, de artículos que no cumplen las especificaciones nominales (tamaño, color, resistencia, etc.)

Si el número "X" de artículos defectuosos en la muestra es menor o igual que un número especificado "c" *menor* que "n", se acepta el lote; si el número de defectuosos es mayor que "c", se rechaza. A "c" se le llama el número tolerable de artículos defectuosos o *número de aceptación*. Por lo tanto, las alternativas son

$X \leq c$ se acepta el lote

$X > c$ se rechaza el lote

Resulta evidente que el productor y el receptor deben quedar de acuerdo en cierto *plan de muestreo*, es decir, en cierto tamaño n de muestra y cierto número de aceptación c. Puesto que en este caso el acuerdo se basa en la extracción de una muestra aleatoria única del lote de N artículos, el plan de muestreo a emplearse se denomina *plan de muestreo simple.*

2.1 Probabilidad de aceptación de un lote

Supóngase que si $X \leq c$ se acepta un lote, es decir, ocurre el evento A = {el número de artículos defectuosos en la muestra extraída del lote es menor o igual que el número de aceptación}. En este caso, la probabilidad de dicho evento no depende únicamente del tamaño n de la muestra y del número de aceptación c, sino también del número total de artículos defectuosos que se encuentran en el lote,"M". Si se supone además que el muestreo se realiza sin remplazo, la probabilidad de dicho evento es hipergeométrica, es decir

$$P\ (A) = P\ \{X \leq c\}\ \sum_{X=0}^{c} \frac{C_X^M\ C_{n-X}^{N-M}}{C_n^N} \qquad (2.1)$$

Si no hay artículos defectuosos en el lote, entonces $M = 0$, y el único valor posible que puede asumir X es también 0, por lo cual

$$P\ (A) = P\ \{X \leq c\} = \frac{C_o^o\ C_n^N}{C_n^N} = 1$$

Es decir, la probabilidad de aceptar un lote en el cual no hay elementos defectuosos es igual a la unidad.

Si todos los artículos en un lote son defectuosos, entonces M = N, y el valor de X debe ser igual a n, por lo que

$$P\ (A) = P\ \{X \leq c\} = P\ (\emptyset) = 0$$

en virtud de que la condición inicial es que c < n. Lo anterior indica que la probabilidad de aceptar un lote en el cual todos los artículos son defectuosos es nula.

Conviene hacer notar también que si se mantienen fijos el tamaño de la muestra y el número de aceptación al incrementarse el valor de M, el número de artículos defectuosos en un lote, decrece la probabilidad P (A) de aceptación de este último.

Ejemplo 2.1

Considérese un plan de muestreo simple para el cual N = 10, c = 0 y n = 5. Obténganse los valores de P (A) cuando

a. M = 1

b. M = 3

Solución

a. En este caso, la probabilidad de aceptación es

$$P\ (A) = P\ \{X = 0\} = \frac{C_0^1 \quad C_{5-0}^{10-1}}{C_5^{10}} =$$

$$= \frac{\dfrac{1!}{0!(1-0)!} \quad \dfrac{9!}{5!(9-5)}}{\dfrac{10!}{5!(10-5)!}} = \frac{\dfrac{9x8x7x6}{4x3x2x1}}{\dfrac{10x9x8x7x6}{5x4x3x2x1}} = 0.5$$

b. Para este caso, se obtiene

$$P\ (A) = P\ \{X \leq 0\} = P\ \{X = 0\} = \frac{C_0^3 \quad C_{5-0}^{10-3}}{C_5^{10}} =$$

$$= \frac{\dfrac{3!}{0!(3-0)!} \quad \dfrac{7!}{5!(7-5)!}}{\dfrac{10!}{5!(10-5)!}} = \frac{\dfrac{7x6}{2x1}}{\dfrac{10x9x8x7x6}{5x4x3x2x1}} = 0.0833$$

Lo anterior indica que un plan de muestreo simple para el cual se mantenga fijo el tamaño de la muestra, aun cuando se incremente el número de elementos defectuosos en los lotes, o el número total de elementos en estos últimos, proporciona buena protección en contra de la aceptación errónea de lotes malos.

2.2 Curva característica de operación

Dentro de un plan de muestreo simple, al considerar un número fijo de aceptación, c, y cuando se obtiene una muestra aleatoria de n artículos de un lote para saber si éste se acepta o no, es evidente que se desconoce el número total de artículos defectuosos, M, dentro del mismo. Para que este número se pudiera conocer en forma precisa, se requeriría haber realizado una inspección al 100% en el lote, pero entonces no tendría caso el considerar un plan de muestreo simple.

Por lo anterior, para realizar el cálculo de la probabilidad de aceptación de un lote determinado cuando se desconoce el valor de M, se debe introducir una modificación dentro de la fórmula 2.1. Para ello, considérese que si se divide el número de elementos defectuosos entre el total de elementos para un lote determinado, se obtiene la *fracción de defectuosos*

$$p = \frac{M}{N} \tag{2.2}$$

en el lote. Si p se multiplica por 100, se obtiene el *porcentaje de elementos defectuosos* en dicho lote.

Puesto que M puede tomar dentro de un lote de tamaño N cualquiera de los N + 1 valores 0,1,2,3,...,N-1,N, p puede asumir entonces los N + 1 valores, 1/N, 2/N, 3/N,..., $^{N-1}/N$, 1. Por lo tanto, la probabilidad de aceptación P (A) únicamente se puede definir para los valores mencionados de p.

Si en la ec 2.2 se despeja el valor de M, se obtiene

$$M + Np$$

en forma tal que la ec 2.1 se puede escribir como

$$P\ (A;\ p) = P\ \{X \leq c\} = \sum_{X=0}^{c} \frac{C_X^{Np}\ \ C_{n-X}^{N-Np}}{C_n^N} \qquad (2.3)$$

siendo las probabilidades así obtenidas hipergeométricas.

Si se mantienen fijos los valores de n y c, se pueden graficar las probabilidades de aceptación de un lote en función de los valores de la fracción de elementos defectuosos en el mismo, es decir, de los valores de p. Dicha gráfica contendrá N + 1 puntos, a través de los cuales se puede dibujar la llamada *curva característica de operación* (o curva CO) de un plan de muestreo simple.

Ejemplo 2.2

La fábrica Z elabora cartuchos de dinamita, y los empaca en cajas de 20 unidades. El comprador W acepta cada caja únicamente si al extraer una muestra de dos cartuchos encuentra que ambos son buenos. Elaborar la curva característica de operación correspondiente.

Solución

En este caso, se tiene que N = 20, n = 2 y c = 0. Por lo tanto, las probabilidades de aceptación son, empleando la ec 2.3

$$P\ (A;p) = P\ \{X \leq 0\} = \frac{C_0^{20p}\ \ C_{2-0}^{20-20p}}{C_2^{20}}$$

$$= \frac{\frac{20p!}{0!(20p-0)!} \quad \frac{(20-20p)!}{2!(20-20p-2)!}}{\frac{20!}{2!(20-2)!}} =$$

$$= \frac{\frac{20p!}{0!20p!} \quad \frac{(20-20p)!}{2x1x(18-20p)!}}{\frac{20!}{2x1x18!}} = \frac{18!(20-20p)!}{20!(18-20p)!} =$$

$$= \frac{(20 - 20p)\ (19 - 20p)}{380}$$

Si se le asignan a p los 21 valores 0, 1/20, 2/20, 3/20, ...,19/20,1, se obtienen los correspondientes de P (A; p). Por ejemplo, para p = 10/20 = 0.5, la probabilidad de aceptación es

$$P\ (A;\ 0.5) = \frac{[20 - 20(10/20)]\ [19 - 20\ (10/20)]}{380}$$

$$= \frac{(20 - 10)\ (19 - 10)}{380} = \frac{(10)(9)}{380} = \frac{90}{380} = 0.237$$

Siguiendo el procedimiento anterior, se obtienen los puntos siguientes:

p	P (A; p)
0/20 = 0.00	1.000
1/20 = 0.05	0.900
2/20 = 0.10	0.805
3/20 = 0.15	0.716
4/20 = 0.20	0.632
5/20 = 0.25	0.553
6/20 = 0.30	0.479
7/20 = 0.35	0.411
8/20 = 0.40	0.347
9/20 = 0.45	0.289
10/20 = 0.50	0.237
11/20 = 0.55	0.189
12/20 = 0.60	0.147
13/20 = 0.65	0.111
14/20 = 0.70	0.079
15/20 = 0.75	0.053
16/20 = 0.80	0.032
17/20 = 0.85	0.016
18/20 = 0.90	0.005
19/20 = 0.95	0.000
20/20 = 1.00	0.000

La curva característica de operación correspondientes es la que se hace pasar por los puntos anteriores, y se presenta en la Fig 2.1.

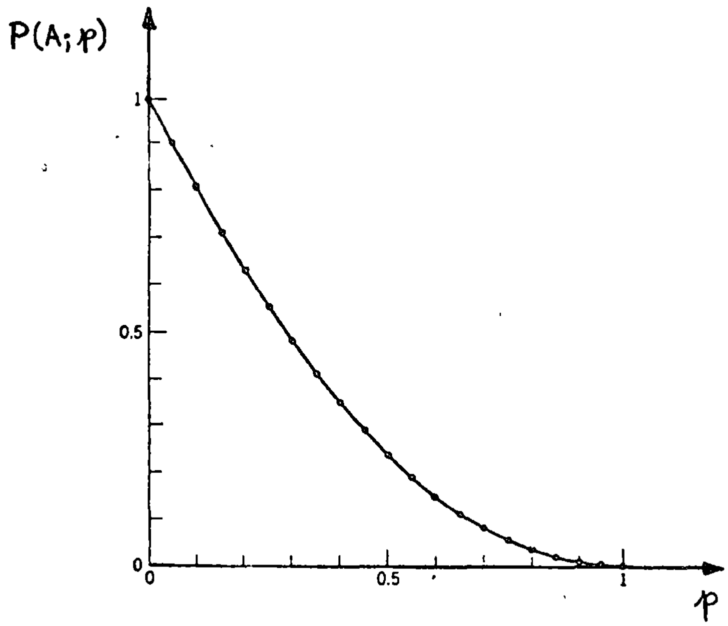


Fig 2.1 Curva CO para un plan de muestreo simple, con N = 20, n = 2 y c = 0.

En la Fig 2.1 se puede observar que a medida que se hace más grande la fracción de defectuosos en el lote (o el número de artículos defectuosos), la probabilidad de aceptación del mismo se va haciendo cada vez menor. Los casos extremos se dan en $p = 0$, en que la aceptación del lote es un evento seguro, y en $p = 1$, cuando es imposible aceptarlo.

2.3 Empleo de la aproximación binomial para construir la curva CO

En la mayor parte de los casos prácticos, el porcentaje de artículos defectuosos en un lote será pequeño (menor del 10%), en tanto que el tamaño del mismo será muy grande (1000 elementos, 10000 elementos, etc), y el de la muestra usualmente será varias veces menor, de tal manera que es posible aproximar las probabilidades dadas por la distribución hipergeométrica (ecs 2.1 y 2.3) empleando la distribución binomial. En particular, la aproximación es buena cuando $N \leq 10\ n$. En estos casos, se puede escribir

$$P\ (A;\ p) = P\ \{X \leq c\} \doteq \sum_{X=0}^{c} C_X^n\ p^X\ (1-p)^{n-X} \qquad (2.4)$$

Se debe observar que siempre se define a p como en la ec 2.2, y que serán mejor aproximadas por la ecuación anterior aquellas probabilidades de aceptación para las cuales el valor de p sea pequeño.

Ejemplo 2.3

En el caso del ejemplo 2.2 anterior, aproxímense las probabilidades de aceptación hipergeométricas para los distintos valores de p mediante la distribución binomial.

Solución

En este caso sí es posible realizar la aproximación pedida, ya que se verifica la condición $N \geq 10n$, porque siendo $N = 20$ y $n = 2$, se tiene que $20 \geq 10(2)$. Por ejemplo, para $p = 0.2$, la aproximación binomial dada por la ec 2.4 conduce al valor

$$P\ (A\ ;\ 0.2) = P\ \{X \leq 0\} = C_0^2\ \ (0.2)^0\ (1-0.2)^{2-0}$$

$$= \frac{2!}{0!\,(2-0)!}\ (0.8)^2 = 0.640$$

en contra del valor exacto 0.632 obtenido mediante la ec 2.3.

Procediendo en forma similar se calculan los restantes valores de P (A; p), los cuales se presentan de 0.1 en 0.1 en la tabla siguiente, junto con los anteriormente obtenidos en el ejemplo 2.2 para fines de comparación.

p	Hipergeométrica P (A; p)	Binomial P (A; p)
0.00	1.000	1.000
0.10	0.805	0.810
0.20	0.632	0.640
0.30	0.479	0.490
0.40	0.347	0.360
0.50	0.237	0.250
0.60	0.147	0.160
0.70	0.079	0.090
0.80	0.032	0.040
0.90	0.005	0.010
1.00	0.000	0.000

En la tabla se puede observar que las probabilidades de aceptación se aproximan bastante más a las exactas cuando el valor de p se encuentra en la vecindad de $p = 0.10$.

2.4 Empleo de la aproximación de Poisson para construir la curva CO

Como ya se vio, la distribución hipergeométrica se puede aproximar adecuadamente mediante la binomial cuando $N \geq 10$ y $p \leq 0.1$. A su vez, la distribución binomial puede aproximarse suficientemente bien mediante la de Poisson cuando se cumple lo anterior y np es menor de 15, lo cual evita en ocasiones la gran cantidad de labor numérica que se requiere para calcular las probabilidades de aceptación mediante las distribuciones hipergeométrica y binomial.

Entonces, si se hace $\lambda = np$ para la distribución de Poisson, se puede escribir

$$P\ (A;\ p) = P\ \{X \leq c\} \doteq e^{-np} \sum_{X=0}^{c} \frac{(np)^X}{X!}$$

La aproximación anterior es muy útil cuando los lotes son grandes, ya que como se puede apreciar, la ec 2.4 no requiere del manejo de dicho dato para el cálculo de las probabilidades de aceptación que se emplean para construir la curva CO.

Ejemplo 2.4

Obténganse los valores de P (A; p) para p = 0, 0.1, 0.2, 0.3, 0.5 y 1.0 en el caso del plan de muestreo simple del ejemplo 2.2, aproximando mediante la distribución de Poisson.

Solución

Se sabe que n = 2 y c = 0, por lo que

$$np = 2(0) = 0 \;; \qquad P\,(A;\,0) = \frac{e^{-0}\quad 0^{0}}{0!} = 1$$

$$np = 2(0.1) = 0.2; \quad P\,(A;\,0.1) = \frac{e^{-0.2}\quad 0.2^{0}}{0!} = 0.818$$

$$np = 2(0.2) = 0.4; \quad P\,(A;\,0.2) = \frac{e^{-0.4}\quad 0.4^{0}}{0!} = 0.670$$

$$np = 2(0.3) = 0.6; \quad P\,(A;\,0.3) = \frac{e^{-0.6}\quad 0.6^{0}}{0!} = 0.549$$

$$np = 2(0.5) = 1.0; \quad P\,(A;\,0.5) = \frac{e^{-1.0}\quad 1.0^{0}}{0!} = 0.367$$

$$np = 2(1.0) = 2.0; \quad P\,(A;\,1.0) = \frac{e^{-2.0}\quad 2.0^{0}}{0!} = 0.135$$

En la siguiente tabla se comparan los valores hipergeométricos exactos con los obtenidos mediante las aproximaciones binomial y de Poisson.

p	P (A;p) Hipergeométrica	P (A; p) Binomial	P (A;p) Poisson
0	1.000	1.000	1.000
0.1	0.805	0.810	0.818
0.2	0.632	0.640	0.670
0.3	0.479	0.490	0.549
0.5	0.237	0.250	0.367
1.0	0.000	0.000	0.135

Como se puede observar en la tabla anterior, las probabilidades de aceptación calculadas con la fórmula de Poisson difieren bastante de las exactas y de las binomiales cuando p no se encuentra cercano al valor 0.1. Sin embargo, hay que considerar que en el problema anterior los tamaños del lote y la muestra son bastante pequeños, por lo que la aproximación de Poisson no puede ser muy buena.

De hecho, la forma práctica para construir las curvas CO se fundamenta en el método aproximado de Poisson, considerando que los lotes que entrega el productor son muy grandes, y haciendo uso de la tabla 2.1 que se presenta adelante,en la cual se proporcionan, en función del número de aceptación c y del valor $\lambda = np$, las probabilidades de aceptación

$$P\ (A;\ p) = P\ \{X \leq c\} = e^{-np} \sum_{X=0}^{c} \frac{(np)^X}{X!}$$

multiplicadas por mil.

TABLA 2.1

TERMINOS ACUMULATIVOS DE LA APROXIMACION DE POISSON A BINOMIAL

np \ c	0	1	2	3	4	5	6	7	8	c / np
0.02	980	1000	....	....	....	....	....	....	....	0.02
0.04	961	999	1000	....	....	....	....	....	....	0.04
0.06	942	998	1000	....	....	....	....	....	....	0.06
0.08	923	997	1000	....	....	....	....	....	....	0.08
0.10	905	995	1000	....	....	....	....	....	....	0.10
0.15	861	990	999	1000	....	....	....	....	....	0.15
0.20	819	982	999	1000	....	....	....	....	....	0.20
0.25	779	974	998	1000	....	....	....	....	....	0.25
0.30	741	963	996	1000	....	....	....	....	....	0.30
0.35	705	951	994	1000	....	....	....	....	....	0.35
0.40	670	938	992	999	1000	....	....	....	....	0.40
0.45	638	925	989	999	1000	....	....	....	....	0.45
0.50	607	910	986	998	1000	....	....	....	....	0.50
0.55	577	894	982	998	1000	....	....	....	....	0.55
0.60	549	878	977	997	1000	....	....	....	....	0.60
0.65	522	861	972	996	999	1000	....	....	....	0.65
0.70	497	844	966	994	999	1000	....	....	....	0.70
0.75	472	827	959	993	999	1000	....	....	....	0.75
0.80	449	809	953	991	999	1000	....	....	....	0.80
0.90	407	772	937	987	998	1000	....	....	....	0.90
1.00	368	736	920	981	996	999	1000	....	....	1.00
1.10	333	699	900	974	995	999	1000	....	....	1.10
1.20	301	663	879	966	992	998	1000	....	....	1.20
1.30	273	627	857	957	[illegible]	998	1000	....	....	1.30
1.40	247	592	833	946	986	997	999	1000	....	1.40
1.50	223	558	809	934	981	996	999	1000	....	1.50
1.60	202	525	783	921	976	994	999	1000	....	1.60
1.70	183	493	757	907	970	992	998	1000	....	1.70
1.80	165	463	731	891	964	990	997	999	1000	1.80
1.90	150	434	704	875	956	987	997	999	1000	1.90
2.00	135	406	677	857	947	983	995	999	1000	2.00
2.10	122	380	650	839	938	980	994	999	1000	2.10
2.20	110	354	622	819	927	974	993	998	1000	2.20
2.30	100	331	596	799	916	970	991	997	999	2.30
2.40	091	308	570	779	904	964	988	997	999	2.40
2.50	082	287	544	758	891	958	986	996	999	2.50
2.60	074	267	518	736	877	951	983	995	999	2.60
2.70	067	249	494	714	863	943	979	993	998	2.70
2.80	061	231	469	692	848	935	976	992	998	2.80
2.90	055	215	446	670	832	926	971	990	997	2.90
3.00	050	199	423	647	815	916	966	988	996	3.00
3.10	045	185	401	625	798	906	961	986	995	3.10
3.20	041	171	380	603	781	895	955	983	994	3.20
3.30	037	159	359	580	763	883	949	980	993	3.30
3.40	033	147	340	558	744	871	942	977	992	3.40
3.50	030	136	321	537	725	858	935	973	991	3.50
3.60	027	126	303	515	706	844	927	969	988	3.60
3.70	025	116	285	494	687	830	918	965	986	3.70
3.80	022	107	269	473	668	816	909	960	984	3.80
3.90	020	099	253	453	648	801	899	955	981	3.90
4.00	018	092	238	433	629	785	889	949	979	4.00
4.10	017	085	224	414	609	769	879	943	976	4.10
4.20	015	078	210	395	590	753	867	936	972	4.20
4.30	014	072	197	377	570	737	856	929	968	4.30
4.40	012	066	185	359	551	720	844	921	964	4.40
4.50	011	061	174	342	532	703	831	913	960	4.50
4.60	010	056	163	326	513	686	818	905	955	4.60
4.70	009	052	152	310	495	668	805	896	950	4.70
4.80	008	048	143	294	476	651	791	887	944	4.80
4.90	007	044	133	279	458	634	777	877	938	4.90

np \ c	0	1	2	3	4	5	6	7	8	c / np
5.00	007	040	125	265	440	616	762	867	932	5.00
5.10	006	037	116	251	423	598	747	856	925	5.10
5.20	006	034	109	238	406	581	732	845	918	5.20
5.30	005	031	102	225	390	563	717	833	911	5.30
5.40	004	029	095	213	373	546	702	822	903	5.40
5.50	004	027	088	202	358	529	686	809	894	5.50
5.60	004	024	082	191	342	512	670	797	886	5.60
5.70	003	022	077	180	327	495	654	784	877	5.70
5.80	003	021	072	170	313	478	638	771	867	5.80
5.90	003	019	067	160	299	462	622	758	857	5.90
6.00	002	017	062	151	285	446	606	744	847	6.00
6.10	002	016	058	143	272	430	590	730	837	6.10
6.20	002	015	054	134	259	414	574	716	826	6.20
6.30	002	013	050	126	247	399	558	702	815	6.30
6.40	002	012	046	119	235	384	542	687	803	6.40
6.50	002	011	043	112	224	369	527	673	792	6.50
6.60	001	010	040	105	213	355	511	658	780	6.60
6.70	001	009	037	099	202	341	495	643	767	6.70
6.80	001	009	034	093	192	327	480	628	755	6.80
6.90	001	008	032	087	182	314	465	614	742	6.90
7.00	001	007	030	082	173	301	450	599	729	7.00
7.20	001	006	025	072	156	276	420	569	703	7.20
7.40	001	005	022	063	110	253	392	539	676	7.40
7.60	000	004	019	055	125	231	365	510	648	7.60
7.80	...	004	016	048	112	210	338	481	620	7.80
8.00	...	003	014	042	100	191	313	453	593	8.00
8.20	...	003	012	037	089	174	290	425	565	8.20
8.40	...	002	010	032	[illegible]	157	267	399	537	8.40
8.60	...	002	009	028	070	142	246	373	509	8.60
8.80	...	001	007	024	062	123	226	348	482	8.80
9.00	...	001	006	021	055	116	207	324	456	9.00
9.20	...	001	005	018	049	104	189	301	430	9.20
9.40	...	001	005	016	043	093	173	279	404	9.40
9.60	...	001	004	014	[illegible]	084	157	258	380	9.60
9.80	...	001	003	012	[illegible]	075	143	239	356	9.80
10.00	...	001	003	010	029	067	130	220	333	10.00
10.20	...	000	002	009	[illegible]	060	118	203	311	10.20
10.40	...	...	002	008	023	053	107	186	290	10.40
10.60	...	...	002	007	020	048	097	171	269	10.60
10.80	...	...	001	006	017	042	087	157	250	10.80
11.00	...	...	001	005	015	037	079	143	232	11.00
11.20	...	...	001	004	013	033	071	131	215	11.20
11.40	...	...	001	004	012	029	064	119	198	11.40
11.60	...	...	001	003	010	026	057	108	183	11.60
11.80	...	...	001	003	009	023	051	099	169	11.80
12.00	...	...	001	002	008	020	046	089	155	12.00
12.20	...	...	000	002	007	018	041	081	143	12.20
12.40	...	...	...	002	006	016	037	073	131	12.40
12.60	...	...	...	001	005	014	033	066	120	12.60
12.80	...	...	...	001	004	012	029	060	109	12.80
13.00	...	...	...	001	004	011	026	054	100	13.00
13.20	...	...	...	001	003	009	023	049	091	13.20
13.40	...	...	...	001	003	008	020	044	083	13.40
13.60	...	...	...	001	002	007	017	039	075	13.60
13.80	...	...	...	001	002	006	016	035	068	13.80
14.00	...	...	...	000	002	006	014	032	062	14.00
14.20	...	...	...	...	002	005	013	028	056	14.20
14.40	...	...	...	...	001	004	011	025	051	14.40
14.60	...	...	...	...	001	004	010	023	046	14.60
14.80	...	...	...	...	001	003	009	020	042	14.80
15.00	...	...	...	...	001	003	008	018	037	15.00

A continuación se presenta un ejemplo práctico de construcción de una curva CO mediante el método descrito, haciendo uso de la tabla 2.1.

Ejemplo 2.5

Supóngase que un receptor de producto terminado adopta el plan de muestreo simple siguiente:

a. Recibe lotes de ciertos artículos con 1000 unidades c/u.

b. Extrae de cada lote una muestra aleatoria de 20 artículos.

c. Si la muestra extraída contiene dos o más artículos defectuosos, rechaza el lote. De no ser así, lo acepta.

Constrúyase la curva CO correspondiente.

Solución

Puesto que el tamaño de los lotes es grande, se pueden aproximar adecuadamente las probabilidades de aceptación mediante la distribución de Poisson. Para ello, se considera en la práctica que con los valores

$$P\ (A;\ p) = 0.98,\ 0.95,\ 0.70,\ 0.50,\ 0.20,\ 0.10,\ 0.05,\ 0.02$$

se puede definir suficientemente bien la curva CO.

Para construir la curva del plan de muestreo simple indicado, considérese que c = 1 y n = 20. En la columna para la cual c = 1 en la tabla 2.1, se puede ver que el valor más cercano a 980 (0.98 de probabilidad) es 982. Para dicho valor, el correspondiente de np es 0.2, siendo por lo tanto $p = \frac{np}{n} = \frac{0.2}{20} = 0.01.$

El valor más cercano a 950 (0.95 de probabilidad) es en la tabla el 951. Para este valor, np = 0.35 y $p = \frac{0.35}{20} = 0.0175.$

Siguiendo el procedimiento anterior, se llega a

P (A;p)	np	p
1.000	0.00	0.000
0.982	0.20	0.010
0.951	0.35	0.0175
0.699	1.10	0.055
0.493	1.70	0.085
0.199	3.00	0.150
0.099	3.90	0.195
0.052	4.70	0.235
0.021	5.80	0.290
0.000	20.00	1.000

En la Fig 2.2 siguiente se presenta la curva caracterís tica de operación correspondiente al problema.

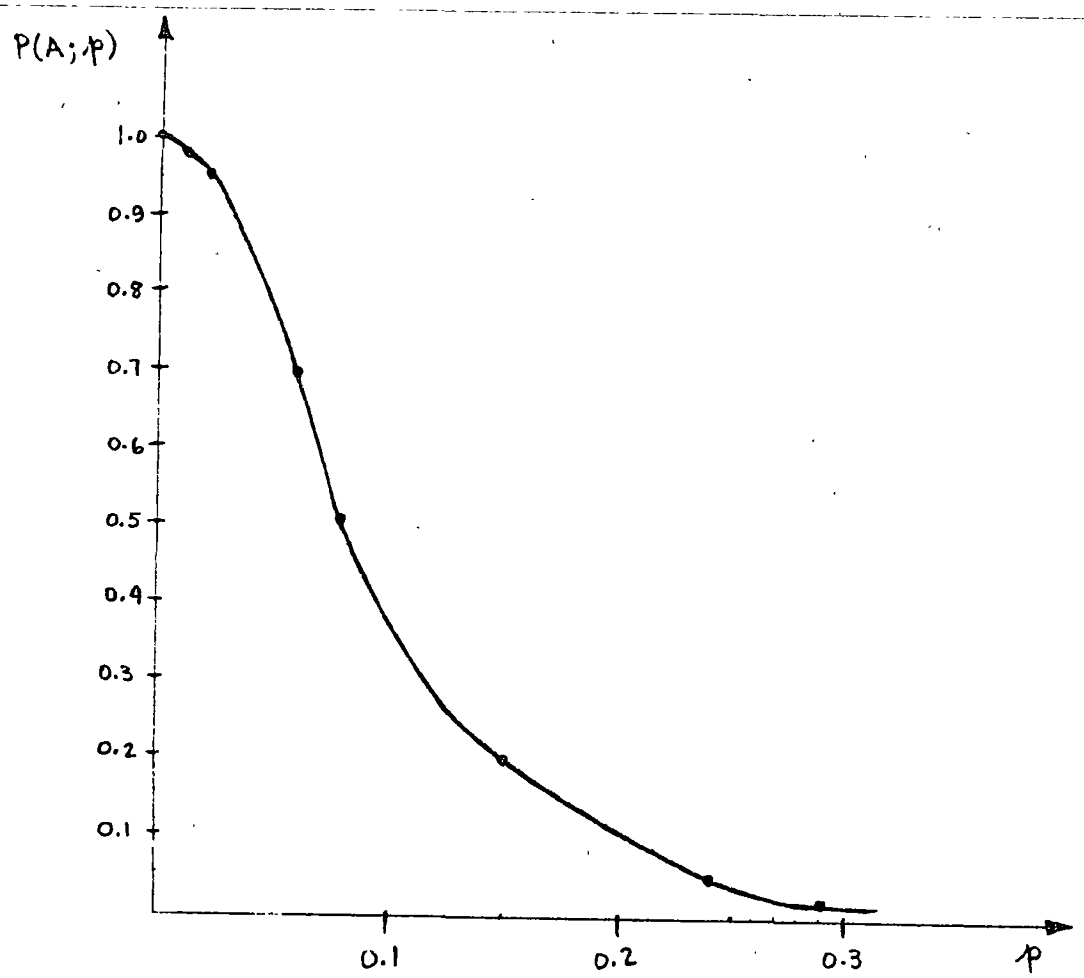


Fig 2.2 Curva característica de operación para plan de muestreo simple con lote grande, c = 1 y n = 20.

2.5 Riesgos en el muestreo de aceptación

Al realizarse los muestreos de aceptación, el productor y el receptor de lotes de artículos tienen intereses distintos al definir un plan de muestreo. El productor puede pedir que la probabilidad, α, de rechazar un lote "bueno" o "aceptable" sea pequeña. Por su parte, el receptor puede exigir que la probabilidad de aceptar un lote "malo" o "no aceptable" sea una cantidad pequeña β.

Para cumplir con ambos compromisos, supóngase que productor y receptor deciden que un lote para el cual p es menor o igual que ciero número p_o es un *lote aceptable*, en tanto que un lote para el que p es mayor o igual que cierto número p_1 $(p_1 > p_o)$ es un lote *no aceptable*, es decir

Si $p \leq p_o$ lote aceptable

Si $p \geq p_1$ lote no aceptable

De acuerdo con lo anterior, α es la probabilidad de rechazar un lote con $p \leq p_o$ y se llama *riesgo del productor*, correspondiendo al error de tipo I que se comete al probar una hipótesis estadística. Por otra parte, β es la probabilidad de aceptar un lote con $p \geq p_1$, se llama *riesgo del receptor*, y corresponde al error de tipo II que se comete al realizar una prueba de hipótesis.

A p_o se le acostumbra llamar *nivel de calidad aceptable* (NCA), y a p_1 *nivel de calidad rechazable* (NCR), o *porcentaje de defectuosos tolerable en un lote* (PDTL). A un lote con $p_o < p < p_1$ se le llama *lote indiferente*.

En la práctica es usual que el acuerdo entre productor y receptor establezca lo siguiente

$$\alpha = \text{Riesgo del productor} \approx 1 - P(A;\, p)_{0.95} = 0.05$$

$$\beta = \text{Riesgo del receptor} \approx P\ (A;\, p)_{0.10} = 0.10$$

Ejemplo 2.6

Para un plan de muestreo simple en el que n = 300 y c = 5, obténganse los valores de p_o y p_1.

Solución

Empleando la tabla 2.1, y considerando los valores P (A; p) que definen adecuadamente a la curva CO, se obtiene:

P (A;p)	np	p
1.000	0.00	0.0000
0.980	2.10	0.0070
0.951	2.60	0.0087
0.703	4.50	0.0150
0.495	5.70	0.0190
0.210	7.80	0.0260
0.104	9.20	0.0307
0.048	10.60	0.0353
0.020	12.00	0.0400
0.000	300.00	1.0000

De acuerdo con la tabla, se tiene que

$$\alpha = 1- P(A;\ p)_{0.951} = 0.0499 \ ; \ p_o = 0.0087$$

$$\beta = P(A;\ p)_{0.104} = 0.104 \ ; \ p_1 = 0.0307$$

En la Fig 2.3 que se presenta a continuación, se muestra la curva CO del plan simple en cuestión, así como los valores del NCA y del NCR.

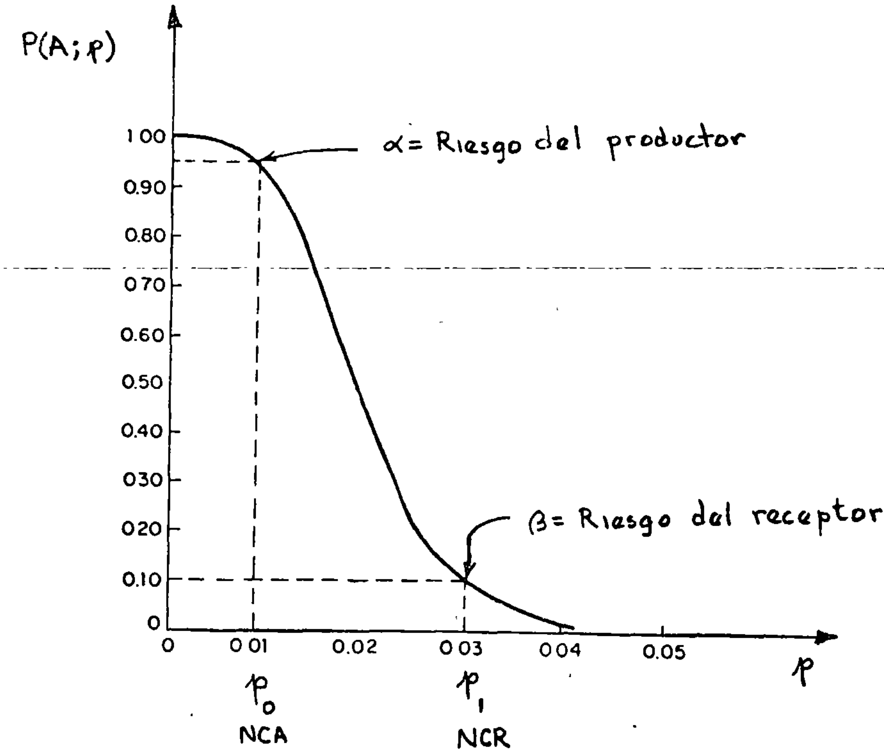


Fig 2.3 Curva CO para plan de muestreo simple con n = 300 y c = 5.

2.6 Cálculo de n y c a partir de p_o, p_1, α y β.

Al observar la Fig 2.3 se puede concluir que los puntos $(p_o, 1-\alpha)$ y (p_1, β) se localizan en la curva CO. Tomando ello en cuenta, existe un método iterativo aproximado para determinar los valores de n y c, considerando conocidos los de p_o, p_1, α y β, de manera que la curva CO pase muy cerca de los puntos mencionados. Dicho procedimiento se expondrá en el ejemplo que sigue, haciendo uso de la tabla 2.1.

Ejemplo 2.7

Para cierto plan de muestreo simple, se fijan los riesgos siguientes:

a. Productor: Aquellos lotes que contengan un 1% de artículos defectuosos se rechazarán en el 5% de los casos.

b. Receptor: Los lotes que contengan un 6% de artículos defectuosos se aceptarán en el 10% del total de casos.

¿Cuáles son los valores del tamaño de la muestra y del número de aceptación que se deben emplear para dicho plan?

Solución

De acuerdo con los datos del problema, se desprende que

$$\alpha = 0.05 \quad ; \quad p_o = 0.01$$

$$\beta = 0.10 \quad ; \quad p_1 = 0.06$$

a. Se considera $c = 0$, con lo cual, de la tabla 2.1,

np_o (para $\alpha = 0.05$ o $P(A;\ 0.01) = 0.95$) $\doteq 0.05$

np_1 (para $\beta = 0.10$) $= 2.30$

Entonces

$$n_\alpha = \frac{np_o}{p_o} = \frac{0.05}{0.01} = 5$$

$$n_\beta = \frac{np_1}{p_1} = \frac{2.30}{0.06} = 38$$

Obviamente, se debe verificar que $n_\alpha = n_\beta$; no siendo este el caso, se hace ahora c = 1.

b. Se considera c = 1, obteniéndose ahora de la tabla 2.1 lo siguiente

$$np_o \text{ (para } \alpha = 0.05) \doteq 0.35$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 3.90$$

Por lo tanto

$$n_\alpha = \frac{0.35}{0.01} = 35$$

$$n_\beta = \frac{3.90}{0.06} = 65$$

Tampoco se verifica que $n_\alpha = n_\beta$; por lo tanto, se hace c = 2.

c. Se considera $c = 2$, y

$$np_o \text{ (para } \alpha = 0.05) \doteq 0.82$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 5.32$$

Ahora, se tiene que

$$n_\alpha = \frac{0.82}{0.01} = 82$$

$$n_\beta = \frac{5.30}{0.06} = 88$$

Ahora n_α y n_β se parecen bastante, pero aún no son iguales. Por lo tanto, se hace $c = 3$ para saber si la diferencia se hace más pequeña.

d. Se considera $c = 3$, y se obtiene

$$np_o \text{ (para } \alpha = 0.05) \doteq 1.37$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 6.68$$

Luego

$$n_\alpha = \frac{1.37}{0.01} = 137$$

$$n_\beta = \frac{6.68}{0.06} = 112$$

Se observa que ahora la diferencia se hace más grande, por lo que el valor real de n se debe encontrar entre 82 y 88 elementos para c = 2. Con el fin de ajustar adecuadamente el valor de n, se puede hacer

$$n = \frac{n_\alpha + n_\beta}{2} = \frac{82 + 88}{2} = 85$$

Por lo tanto, el plan de muestreo simple es el siguiente

$$\alpha = 0.05 \quad ; \quad \beta = 0.10$$

$$p_o = 0.01 \quad ; \quad p_1 = 0.06$$

$$n = 85 \quad ; \quad c = 2$$

cuya curva CO se muestra en la Fig 2.4.

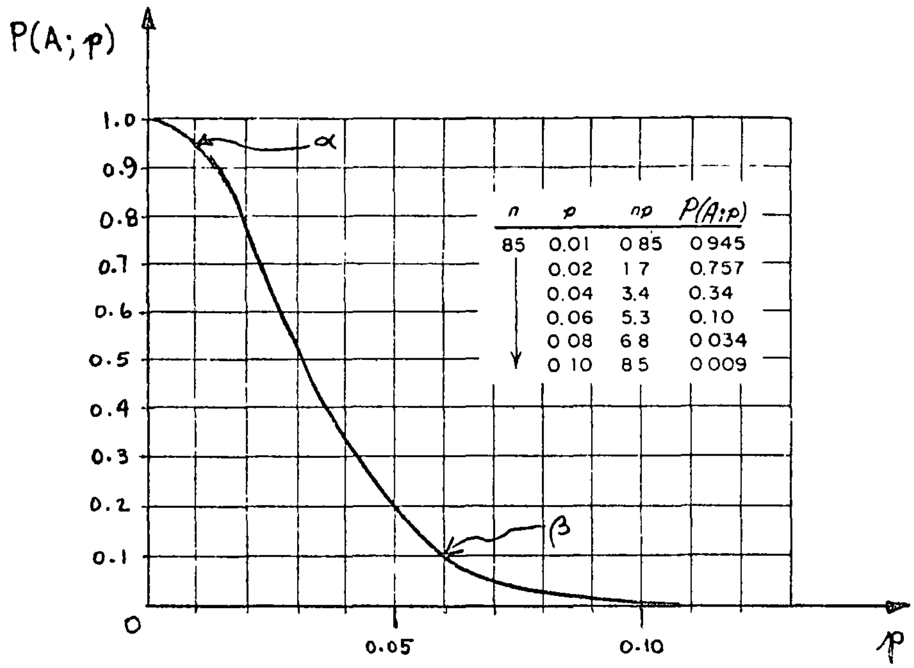


Fig 2.4 Curva CO ajustada para α, β, p_o y p_1 conocidos.

2.7 Comentarios sobre la curva CO

Al comparar las curvas CO de las Figs 2.3 y 2.4, se puede observar que, no obstante el número más grande de artículos defectuosos que permite en la muestra el plan de muestreo asociado a la curva CO de la Fig 2.3, se trata de un mejor plan de aceptación de lotes, en el sentido de que proporciona riesgos más favorables al receptor.

En efecto, ambos planes consideran $\alpha = 0.05$, $\beta = 0.10$ y $p_o \approx 0.01$, pero el plan de la Fig 2.4 aceptará lotes con 6% de defectuosos ($p_1 \approx 0.06$) en el 10% del total de casos, en tanto que el de la Fig 2.3 aceptará lotes con 3% de defectuosos ($p_1 \approx 0.03$)

en el mismo número de casos.

En muchas ocasiones no se comprende con claridad el porqué de un número de aceptación mayor de cero en los planes de muestreo. Si se observa la Fig 2.5, se puede apreciar que las curvas CO (a), (b) y (c) corresponden a planes de muestreo que evitan los artículos defectuosos en la muestra ($c = 0$), pero que tienen riesgos de productor y receptor distintos. Los planes de las curvas CO (d) y (e) consideran 4 y 7 defectuosos en la muestra, respectivamente.

Se observa que las curvas CO con $c = 0$ se caracterizan por patrones cóncavos, en tanto que aquellas con $c \neq 0$ semejan curvas S invertidas.

Los planes de muestreo con $c = 0$ usualmente penalizan más al productor. Asimismo, aquellos planes en que c es mayor de cero proporcionan riesgos más favorables al productor o al receptor, y en muchos casos a ambos.

Se puede afirmar que el riesgo para el receptor se hace más pequeño conforme se incrementa el tamaño de la muestra, en tanto que el riesgo para el productor decrece conforme se permiten uno o más artículos defectuosos en la misma. Esto se puede aclarar si se observan los riesgos en las curvas (c) y (d) de la Fig 2.5.

Las curvas (d) y (e) consideran esencialmente el mismo riesgo para el productor (NCA $\approx$ 0.01 en $\alpha = 0.05$), pero la (e) considera un tamaño de muestra mayor, por lo que el receptor corre un riesgo menor. La curva (f) corresponde a la curva ideal CO, ya que ese plan de muestreo acepta todos los lotes con uno por ciento o menos de artículos defectuosos, y rechaza todos los lotes que contengan más del 1% de defectuosos. Dicha curva obviamente no se puede obtener con las técnicas usuales de muestreo de aceptación.

Lo anterior indica que un plan de muestreo simple será más efectivo en tanto su curva CO correspondiente se asemeje más a la curva ideal de operación.

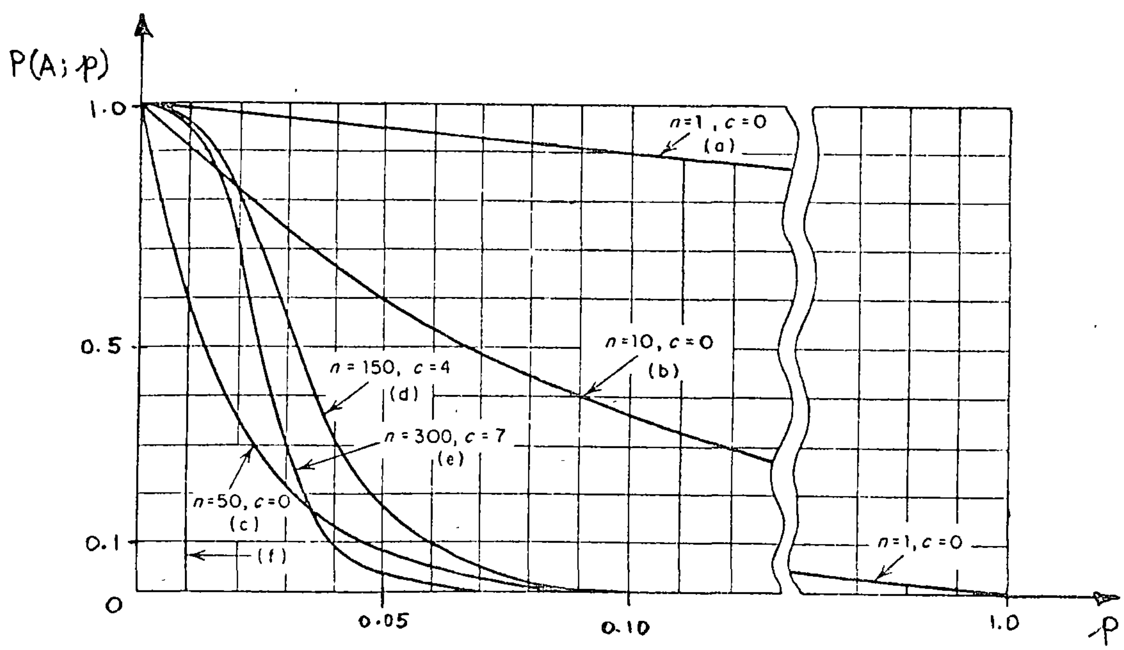


Fig 2.5 Distintos planes de muestreo con c = 0 y c ≠ 0.

3. Plan de muestreo doble

Un plan de muestreo simple requiere que se tome una decisión sobre la aceptación o rechazo de un lote tomando como base la evidencia de *una* muestra extraída del mismo.

Sin embargo, un *plan de muestreo doble* implica la posibilidad de posponer la decisión sobre la aceptación o rechazo de un lote hasta que *una segunda* muestra haya sido extraída. Dicho lote podrá ser aceptado inmediatamente si la primera muestra es muy buena, o rechazado enseguida si la primera muestra es bastante mala. Si la primera muestra no es ni muy buena ni muy mala, la decisión se basa en la evidencia de la primera y segunda muestras combinadas.

En general, los planes de muestreo doble conducen a menos inspección total que los planes sencillos, y también proporcionan la ventaja sicológica que conlleva la idea de dar una segunda oportunidad a los lotes dudosos.

3.1 Símbolos en el muestreo doble

Los siguientes son los símbolos empleados en conexión con el muestreo doble:

N = tamaño del lote

n_1 = tamaño de la primera muestra

c_1 = número de aceptación para la primera muestra

n_2 = tamaño de la segunda muestra

$n_1 + n_2$ = tamaño de la muestra combinada

c_2 = número de aceptación para la muestra combinada

3.2 Interpretación del plan de muestreo doble

Considérese un plan de muestreo doble para el cual se fijan los valores de N, n_1, c_1, n_2 y c_2 ($c_2 > c_1$). La interpretación del proceso que se realiza con dicho plan es la siguiente:

a. Se inspecciona una primera muestra de tamaño n_1 extraída del lote de tamaño N.

b. Se acepta el lote si la muestra anterior contiene c_1 o menos artículos defectuosos.

c. Se rechaza el lote si el número de defectuosos en la muestra excede el valor c_2.

d. Si la primera muestra contiene $c_1 + 1$, $c_1 + 2$, ... o c_2 artículos defectuosos, se extrae e inspecciona una segunda con n_2 elementos.

e. Se acepta el lote sobre la base de la muestra combinada con $n_1 + n_2$ elementos si dicha muestra contiene c_2 artículos defectuosos o menos.

f. Se rechaza el lote si la muestra combinada contiene más de c_2 defectuosos.

3.2 Curva CO de un plan de muestreo doble

De acuerdo con lo que se ha explicado, existen cuatro posiblilidades de que se acepte o se rechace un lote sometido para muestreo doble. Dichas posibilidades son

a. Aceptación después de la primera muestra.

b. Rechazo después de la primera muestra.

c. Aceptación después de la segunda muestra.

d. Rechazo después de la segunda muestra

Tomando como base lo anterior, se explicará a través del ejemplo siguiente la forma como se construye la curva CO para el plan de muestreo doble.

Ejemplo 3.1

Considérese el plan de muestreo doble para el cual el tamaño del lote es muy grande, $n_1 = 50$, $c_1 = 1$, $n_2 = 100$ y $c_2 = 3$.

Constrúyase la curva CO correspondiente.

Solución

Para determinar los puntos de la curva CO, es necesario calcular las probabilidades de que si se toma una segunda muestra el lote sea aceptado, para distintos valores de p. Para ilustrar lo anterior considérese inicialmente el valor $p = 0.02$.

Entonces, un lote puede ser aceptado según el plan anterior en cualquiera de las formas siguientes:

a. un defectuoso o menos en la primera muestra

b. dos defectuosos en la primera muestra, seguido de cero o un defectuoso en la segunda muestra

c. tres defectuosos en la primera muestra, seguidos de cero defectuosos en la segunda muestra.

La probabilidad de aceptar un lote es entonces igual a la suma de las probabilidades de estos diferentes modos por los cuales puede ser aceptado.

Inicialmente, se deben calcular las probabilidades de tener uno o menos, dos o menos y tres o menos defectuosos en la primera muestra. Lo anterior equivale a considerar un plan de muestreo simple para el cual $n_1 = 50$ y $c = 1, 2, 3$. A continuación se deben calcular las probabilidades de tener exactamente dos y tres defectuosos en la primera muestra.

Entonces, con $n_1p = 50(0.02) = 1.00$, se obtiene, empleando la tabla 2.1 y siendo X el número de elementos defectuosos

$$P\{X \leq 1\}_1 = 0.736 \quad ; \quad c = 1 \quad , \quad n_1p = 1.00$$

$$P\{X \leq 2\}_1 = 0.920 \quad ; \quad c = 2 \quad , \quad n_1p = 1.00$$

$$P\{X \leq 3\}_1 = 0.981 \quad ; \quad c = 3 \quad , \quad n_1p = 1.00$$

y

$$P\{X = 2\}_1 = P\{X \leq 2\}_1 - P\{X \leq 1\}_1 = 0.920-0.736 = 0.184$$

$$P\{X = 3\}_1 = P\{X \leq 3\}_1 - P\{X \leq 2\}_1 = 0.981-0.920 = 0.061$$

El subíndice fuera de la llave indica que la probabilidad del evento se calcula con base en la primera muestra.

Ahora bien, si en la primera muestra hay dos defectuosos, los cálculos relacionados con la segunda muestra deberán basarse en $n_2p = 100(0.02) = 2$. El tomar la segunda muestra e inspeccionarla equivale, para efectos de los cálculos, a considerar un nuevo plan de muestreo simple para el resto del lote con número de aceptación igual a 1, ya que este elemento, sumado a los dos defectuosos considerados, permite la aceptación del lote.

Por lo tanto,

$$P\ \{X \leq 1\}_2 = 0.406 \quad ; \quad c = 1, \quad n_2p = 2$$

Si en la primera muestra hay tres defectuosos, los cálculos para la segunda muestra se deben basar en $n_2p = 100(0.02)$ y un número de aceptación igual a cero, es decir

$$P\ \{X \leq 0\}_2 = 0.135 \quad ; \quad c = 0\ , \ n_2p = 2$$

La probabilidad de aceptación es, empleando el concepto de independencia de eventos, la suma de las probabilidades siguientes:

P {un defectuoso o menos en la primera muestra} $= P\ \{X \leq 1\}_1 = 0.736$

+ P {dos defectuosos en la primera muestra, seguidos de cero o un defectuoso en la segunda} $= P\ \{X = 2\}_1\ P\{X \leq 1\}_2 = (0.184)(0.406) =$

$= 0.075$

+ P {tres defectuosos en la primera muestra, seguidos de cero defectuosos en la segunda} $= P\ \{X = 3\}_1\ P\{X \leq 0\}_2 = (0.061)(0.135) =$

$= 0.008$

Entonces,

$$P\{A;\ 0.02\} = 0.736 + 0.075 + 0.008 = 0.819$$

es decir, el punto (0.02, 0.819) se encuentra sobre la curva CO del plan de muestreo doble.

En la forma descrita anteriormente, se pueden calcular también los puntos restantes para definir la curva CO, quedando finalmente

P (A; p)	p
0.98	0.012
0.95	0.015
0.82	0.020
0.70	0.027
0.50	0.037
0.20	0.063
0.10	0.080
0.05	0.100
0.02	0.136

La gráfica de la curva CO correspondiente al plan de muestreo doble propuesto se presenta en la Fig 3.1

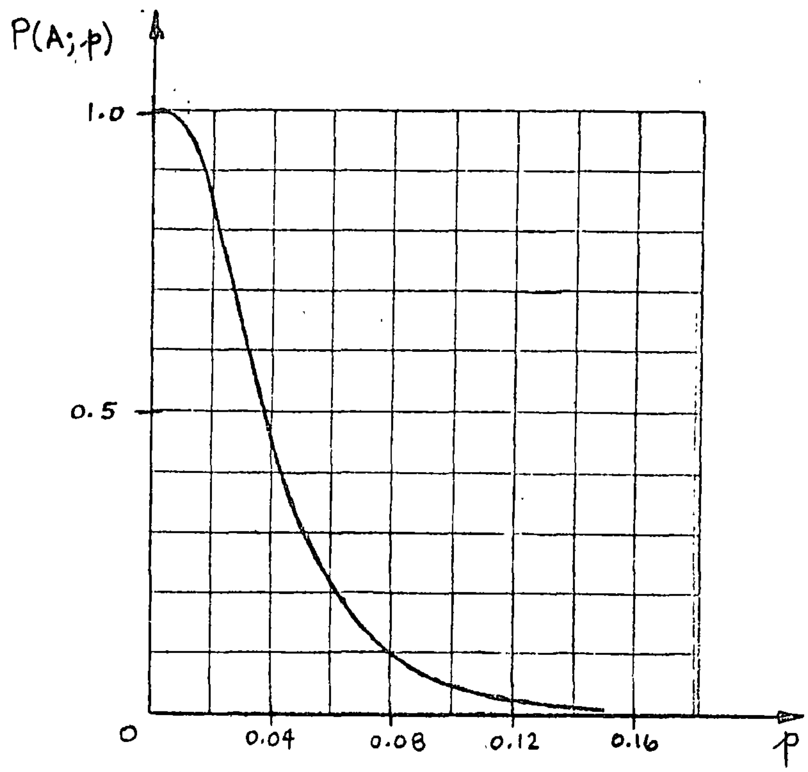


Fig 3.1 Curva CO para plan de muestreo doble con $n_1 = 50$, $c_1 = 1$, $n_2=100$, $c_2 = 3$.

4. Plan de muestreo múltiple

De la misma manera que los planes de muestreo doble pueden diferir la decisión sobre la aceptación o rechazo de un lote hasta que haya sido tomada una segunda muestra, otros planes pueden permitir la extracción de cierto número de muestras antes de que una decisión sea tomada.

Los planes de *muestreo múltiple* son usados cuando se permite la extracción de tres o más muestras de un tamaño preestablecido, y cuando la decisión sobre la aceptación o rechazo de un lote se debe tomar después de la séptima muestra extraída, considerando que no es permitida la aceptación de ese lote con la evidencia obtenida de la primera muestra.

4.1 Interpretación de un plan de muestreo múltiple

Considérese el siguiente plan de muestreo múltiple

Número de la muestra	Tamaño de la muestra individual	Tamaño de la muestra combinada	Número de aceptacion, c	Número de rechazo, r
1	20	20	-	2
2	20	40	0	3
3	20	60	1	3
4	20	80	2	4
5	20	100	2	4
6	20	120	2	4
7	20	140	3	4

La forma de interpretar el plan anterior es la siguiente:

a. Se extrae e inspecciona una muestra de 20 elementos. Si dos o más son defectuosos, se rechaza el lote; si hay uno o cero defectuosos, se extrae e inspecciona una segunda muestra de 20 elementos. (La aceptación del lote no se permite con la primera muestra.)

b. Si en la muestra combinada (20 + 20 = 40) no hay ningún defectuoso, se acepta el lote; si 3 o más artículos son defectuosos se rechaza. De encontrarse uno o dos defectuosos, se toma una tercera muestra de 20 elementos.

c. Si en la muestra combinada (40 + 20 = 60) hay un defectuoso, se acepta el lote; si 3 o más artículos son defectuosos, se rechaza. De encontrarse dos defectuosos, se toma una cuarta muestra de 20 elementos.

d. Si en la muestra combinada (60 + 20 = 80) hay dos defectuosos, se acepta el lote; si 4 o más son defectuosos, se rechaza. De encontrarse tres defectuosos, se toma una quinta muestra de 20 elementos.

.

.

.

g. Si en la muestra combinada (120 + 20 = 140) hay tres defectuosos, se acepta el lote. Si hay cuatro defectuosos o más, se rechaza.

4.2 Curva CO de un plan de muestreo múltiple

La curva característica de operación de un plan de muestreo múltiple se puede obtener siguiendo un procedimiento semejante al empleado en el caso del muestreo doble, haciendo uso de probabilidades condicionales y suponiendo la descomposición del plan múltiple en varios planes sencillos. Desde luego, el cálculo de las probabilidades de aceptación es bastante más complejo, pero el razonamiento es básicamente el mismo.

A continuación, se describirá mediante un ejemplo el procedimiento para la construcción de la curva Co.

Ejemplo 4.1

Considérese el plan de muestreo múltiple descrito anteriormente, y constrúyase la curva CO correspondiente, suponiendo un lote de tamaño grande.

Solución

Los siguientes cálculos corresponden a un solo punto de la curva, para el cual p = 0.02. Cada una de las muestras contiene 20 artículos, por lo que para cada una de llas se tendrá $np = 20(0.02) = 0.4$. Entrando con este valor a la tabla 2.1, y considerando que X denota el número de artículos defectuosos, se obtienen, también para cada muestra, las probabilidades incondicionales siguientes:

$$P_o = P\{X = 0\} = P\{X \leq 0\} = 0.670$$

$$P_1 = P\{X = 1\} = P\{X \leq 1\} - P\{X \leq 0\} = 0.938 - 0.670 = 0.268$$

$$P_2 = P\{X = 2\} = P\{X \leq 2\} - P\{X \leq 1\} = 0.992 - 0.938 = 0.054$$

Tomando en cuenta que A = aceptación, R = rechazo y CM = continúa muestreo, se hace enseguida el análisis muestra por muestra para obtener la probabilidad P (A; 0.02).

a. Muestra 1 (M1)

número de aceptación = c = no hay

número de rechazo = r = 2

0 def M1 $\Rightarrow$ $P_0 = 0.670$ $\Rightarrow$ CM (0 def)

1 def M1 $\Rightarrow$ $P_1 = 0.268$ $\Rightarrow$ CM (1 def)

2 def M1 $\Rightarrow$ $\Rightarrow$ R (2 def)

Probabilidad de aceptación = 0.000

b. Muestra 2 (M2)

c = 0

r = 3

0 def M1, 0 def M2 $\Rightarrow$ $P_{00} = (0.670)(0.670) = 0.449$ $\Rightarrow$ A (0 def)

0 def M1, 1 def M2 $\Rightarrow$ $P_{01} = (0.670)(0.268) = 0.1795$ $\Rightarrow$ CM (1 def)

0 def M1, 2 def M2 $\Rightarrow$ $P_{02} = (0.670)(0.054) = 0.0362$ $\Rightarrow$ CM (2 def)

0 def M1, 3 def M2 $\Rightarrow$ $\Rightarrow$ R (3 def)

1 def M1, 0 def M2 $\Rightarrow$ $P_{10} = (0.268)(0.670) = 0.1795$ $\Rightarrow$ CM (1 def)

1 def M1, 1 def M2 $\Rightarrow$ $P_{11} = (0.268)(0.268) = 0.0718$ $\Rightarrow$ CM (2 def)

1 def M1, 2 def M2 $\Rightarrow$ $\Rightarrow$ R (3 def)

Probabilidad de aceptación = 0.449

Nuevos valores:

P_1 = P {un defectuoso en M2} = 0.1795 + 0.1795 = 0.359

P_2 = P {dos defectuosos en M2} = 0.0362 + 0.0718 = 0.108

c. Muestra 3 (M3)

c = 1

r = 3

1 def M2, 0 def M3 $\Rightarrow$ P_{10} = (0.359)(0.670) = 0.2405 $\Rightarrow$ A (1 def)

1 def M2, 1 def M3 $\Rightarrow$ P_{11} = (0.359)(0.268) = 0.0962 $\Rightarrow$ CM (2 def)

1 def M2, 2 def M3 $\Rightarrow$ $\Rightarrow$ R (3 def)

2 def M2, 0 def M3 $\Rightarrow$ P_{20} = (0.108)(0.670) = 0.0723 $\Rightarrow$ CM (2 def)

2 def M2, 1 def M3 $\Rightarrow$ $\Rightarrow$ R (3 def)

Probabilidad de aceptación = 0.2405

Nuevo valor:

P_2 = P {dos defectuosos en M3} = 0.0962 + 0.0723 = 0.1685

d. Muestra 4 (M4)

c = 2

r = 4

2 def M3, 0 def M4 $\Rightarrow$ P_{20} = (0.1685)(0.670) = 0.1129 $\Rightarrow$ A (2 def)

2 def M3, 1 def M4 $\Rightarrow$ P_{21} = (0.1685)(0.268) = 0.0451 $\Rightarrow$ CM (3 def)

2 def M3, 2 def M4 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.1129

Nuevo Valor:

P_3 = P {3 defectuosos en M4} = 0.0451

e. Muestra 5 (M5)

c = 2

r = 4

3 def M4, 0 def M5 $\Rightarrow$ P_{30} = (0.0451)(0.670) = 0.0302 $\Rightarrow$ CM (3 def)

3 def M4, 1 def M5 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.000

Nuevo valor:

P_3 = P {3 defectuosos en M5} = 0.0302

f. Muestra 6 (M6)

c = 2

r = 4

3 def M5, 0 def M6 $\Rightarrow$ P_{30} = (0.0302)(0.670) = 0.0202 $\Rightarrow$ CM (3 def)

3 def M5, 1 def M6 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0,000

Nuevo valor

P_3 = P {tres defectuosos en M6} = 0.0202

g. Muestra 7 (M7)

c = 3

r = 4

3 def M6, 0 def M7 $\Rightarrow$ P_{30} = (0.0202)(0.670) = 0.0135 $\Rightarrow$ A (3 def)

3 def M6, 1 def M7 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.0135

De acuerdo con lo anterior, la probabilidad de aceptación de un lote, sujeto al plan de muestreo múltiple propuesto con p = 0.02, es

$$P\ (A;\ 0.02) = 0.449 + 0.2405 + 0.1129 + 0.0135 = 0.8159$$

Siguiendo el método descrito, se pueden calcular los valores de las probabilidades de aceptación para distintos valores de p, con los cuales se definen los puntos necesarios para construir la curva característica de operación correspondiente, que se presenta en la Fig 4.1.

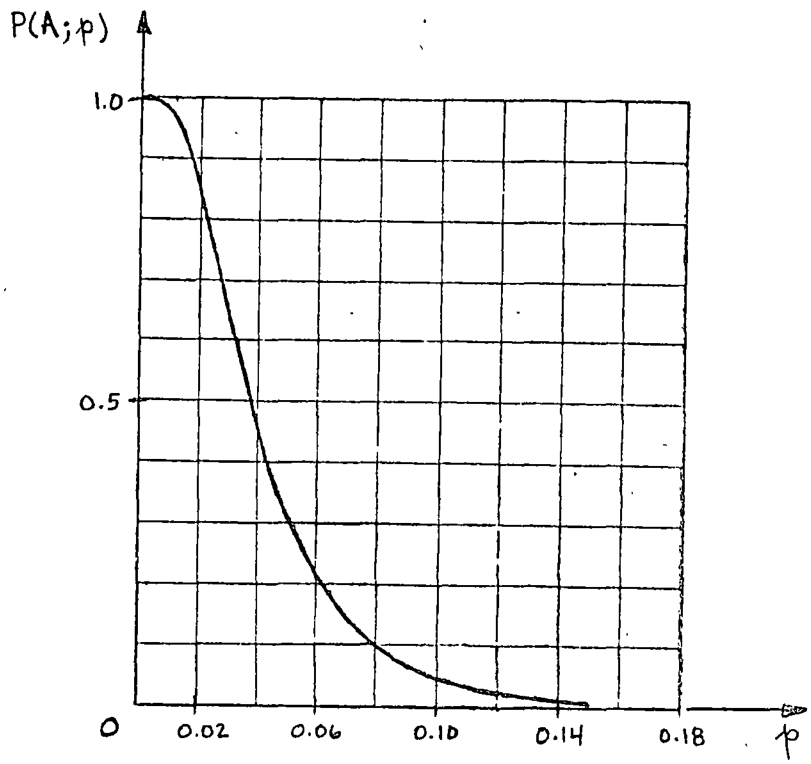


Fig 4.1 Curva CO para un plan de muestreo múltiple

5. Ventajas y desventajas de los planes de muestreo simples, dobles y múltiples

En general, los tres esquemas de muestreo de aceptación que se han presentado se pueden ajustar para proporcionar a lotes con valores de p determinados prácticamente la misma probabilidad de ser aceptados; es decir, si se desea, se puede lograr que las curvas características de operación para los planes simples, dobles y múltiples sean muy parecidas.

No obstante lo anterior, puede suceder que un plan de muestreo de aceptación que ha dado buen resultado para un productor o producto, resulte no tan efectivo para otros. La efectividad de los distintos planes de muestreo expuestos se puede juzgar si se analizan las ventajas y desventajas de cada uno de ellos, en términos de cuatro factores importantes: El número medio de artículos inspeccionados, el costo de administración del plan, la aceptación por parte del productor y la información sobre calidad de los lotes obtenida a largo plazo. En la tabla 5.1 se compara la efectividad de los tres planes estudiados.

Los factores mencionados en la tabla 5.1 deben ser considerados al seleccionar un plan de muestreo. Por ejemplo, en aquellos casos en que el costo de inspección de cada artículo es elevado, la reducción en el número de artículos inspeccionados puede justificar el empleo del muestreo múltiple no obstante su gran complejidad y elevado costo de administración.

Por otro lado, el muestreo simple puede ser el adecuado si el costo de entrenamiento de personal es muy apreciable. Finalmente, si el problema es de acuerdo entre receptor y productor del plan a emplear, posiblemente la solución sea el muestreo doble, ya que es sicológicamente bien aceptado por ambas partes.

TABLA 5.1

COMPARACION ENTRE LOS PLANES DE
MUESTREO SIMPLE, DOBLE Y MULTIPLE

Factor	Plan simple (PS)	Plan doble (PD)	Plan múltiple (PM)
a. Número medio de artículos inspeccionados	El más grande de todos	De 5 a 40% menos que en PS	Aproximadamente 25% menos que en PD
b. Costos de administración (entrenamiento, registros, personal, etc.)	El más bajo de todos	Mayor que el de PS	El más alto de todos
c. Aceptación por parte del productor	Regular	Adecuada	Poca
d. Información a largo plazo sobre calidad de los lotes	La mayor	Menos que en PS	La menor

Ejemplo 3.1 (con $p = 0.02$)

a. Muestra 1 (M1)

$c = 1$

$r = 4$

$np = 50(0.02) = 1.0$; $P_0 = 0.368$; $P_1 = 0.368$; $P_2 = 0.184$; $P_3 = 0.061$

0 def M1 $\Rightarrow P_0 = 0.368$ $\Rightarrow$ A (0 def)

1 def M1 $\Rightarrow P_1 = 0.368$ $\Rightarrow$ A (1 def)

2 def M1 $\Rightarrow P_2 = 0.184$ $\Rightarrow$ CM (2 def)

3 def M1 $\Rightarrow P_3 = 0.061$ $\Rightarrow$ CM (3 def)

4 def M1 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.736

b. Muestra 2 (M2)

$c = 3$

$r = 4$

$np= 100(0.02) = 2$; $P_0 = 0.135$ $P_1 = 0.271$; $P_2 = 0.271$; $P_3 = 0.180$

2 def M1, 0 def M2 $\Rightarrow P_{20} = (0.184)(0.135) = 0.0248 \Rightarrow$ A (2 def)

2 def M1, 1 def M2 $\Rightarrow P_{21} = (0.184)(0.271) = 0.0498 \Rightarrow$ A (3 def)

2 def M1, 2 def M2 $\Rightarrow$ $\Rightarrow$ R (4 def)

3 def M1, 0 def M2 $\Rightarrow P_{30} = (0.061)(0.135) = 0.0082 \Rightarrow$ A (3 def)

3 def M1, 1 def M2 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.0828

$$P\,(A;\ 0.02) = 0.736 + 0.0828 = 0.8188 \approx 0.819$$

CONTROL ESTADISTICO DE CALIDAD

MUESTREO E INSPECCION

TABLAS MIL-STD-105 D

M. en I. AUGUSTO VILLARREAL

TABLA K. LETRAS CLAVE PARA EL TAMAÑO DE MUESTRAS —MIL-STD-105 (ESTÁNDAR ABC)

Tamaño del lote o partida	Niveles de inspección generales		
	I	II	III
2-8	A	A	B
9-15	A	B	C
16-25	B	C	D
26-50	C	D	E
51-90	C	E	F
91-150	D	F	G
151-280	E	G	H
281-500	F	H	J
501-1 200	G	J	K
1 201-3 200	H	K	L
3 201-10 000	J	L	M
10 001-35 000	K	M	N
35 001-150 000	L	N	P
150 001-500 000	M	P	Q
500 001 y más	N	Q	R

Letra clave para el tamaño de la muestra	Tamaño de la muestra	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1 000
		Niveles de calidad aceptable (inspección normal)																									
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 3
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑
D	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	↑
E	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	↑	↑	↑
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑
G	32	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑
H	50	↓	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑
J	80	↓	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑
K	125	↓	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
L	200	↓	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
M	315	↓	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
N	500	↓	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
P	800	↓	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Q	1 250	0 1	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
R	2 000	↑	↑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑

↓ = Usese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.
↑ = Usese el primer plan de muestreo arriba de la flecha.
Ac = Número de aceptación.
Re = Número de rechazo.

TABLA M. TABLA MAESTRA PARA INSPECCION CERRADA (MUESTREO SENCILLO)—MIL-STD-105D (ESTÁNDAR ABC)

Letra clave para el tamaño de la muestra	Tamaño de la muestra	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1 000
		Niveles de calidad aceptable (inspección cerrada)																									
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28
B	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42
C	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	↑
D	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	↑	↑
E	13	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	↑	↑	↑
F	20	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑
G	32	↓	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑
H	50	↓	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑
J	80	↓	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑
K	125	↓	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
L	200	↓	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
M	315	↓	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
N	500	↓	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
P	800	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Q	1 250	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
R	2 000	0 1	↑	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
S	3 150			1 2																							

↓ = Usese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.
↑ = Usese el primer plan de muestreo arriba de la flecha.
Ac = Número de aceptación.
Re = Número de rechazo.

TABLA N. TABLA MAESTRA PARA INSPECCION REDUCIDA (MUESTREO SENCILLO)—MIL-STD-105D (ESTÁNDAR ABC)

Letra clave para el tamaño de la Muestra	Tamaño de la muestra	Niveles de calidad aceptable (inspección normal)																									
		0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1 000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2														↓	0 1		↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	2													↓	0 1	↑	↓	0 2	1 3	2 4	3 5	5 6	7 8	10 11	14 15	21 22	30 31
C	2												↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑
D	3											↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑	
E	5										↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑		
F	8									↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑	↑	↑			
G	13								↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑						
H	20							↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑							
J	32						↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑								
K	50					↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑									
L	80				↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑										
M	125			↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑											
N	200		↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑												
P	315	↓	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑													
Q	500	0 1	↑	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑														
R	800	↑		0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑															

↓ = úsese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.

↑ = úsese el primer plan de muestreo arriba de la flecha.

Ac = número de aceptación.

Re = número de rechazo.

† Si el número de aceptación ha sido excedido, y el número de rechazo no ha sido alcanzado, acéptese el lote, pero reinstálese la inspección normal.

TABLA O. TABLA MAESTRA PARA INSPECCION NORMAL (MUESTREO DOBLE)—MIL-STD-105D (ESTÁNDAR ABC)

Letra clave para el tamaño de la muestra	Muestra	Tamaño de la muestra	Tamaño acumulativo de la muestra	Niveles de calidad aceptable (inspección normal)																									
				0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1 000
				Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A																	↓	†		↓	†	†	†	†	†	†	†	†	†
B	Primera	2	2													↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31
	Segunda	2	4																	1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57
C	Primera	3	3												↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑
	Segunda	3	6																1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57	
D	Primera	5	5											↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑	
	Segunda	5	10															1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57		
E	Primera	8	8										↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑		
	Segunda	8	16														1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57			
F	Primera	13	13									↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑			
	Segunda	13	26													1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27						
G	Primera	20	20								↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑						
	Segunda	20	40												1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27							
H	Primera	32	32							↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑							
	Segunda	32	64											1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27								
J	Primera	50	50						↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑								
	Segunda	50	100										1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27									
K	Primera	80	80					↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑									
	Segunda	80	160									1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27										
L	Primera	125	125				↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑										
	Segunda	125	250								1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27											
M	Primera	200	200			↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑											
	Segunda	200	400							1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27												
N	Primera	315	315		↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑												
	Segunda	315	630						1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27													
P	Primera	500	500	↓	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑													
	Segunda	500	1 000					1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27														
Q	Primera	800	800	†	↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑														
	Segunda	800	1 600				1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27															
	Primera	1 250	1 250	↑		0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑															
	Segunda	1 250	2 500			1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27																

↓ = Usese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.

↑ = Usese el primer plan de muestreo arriba de la flecha.

Ac = Número de aceptación.

Re = Número de rechazo.

† Usese el plan de muestreo sencillo correspondiente (o alternativamente, úsese el plan de muestreo doble hacia abajo, cuando esté disponible).

TABLA R. TABLA MAESTRA PARA INSPECCION NORMAL (MUESTREO MULTIPLE)—MIL-STD-105D (ESTÁNDAR ABC)

Letra clave para el tamaño de la muestra	Muestra	Tamaño de la muestra	Tamaño acumulativo de la muestra	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
				Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↓	↓	†	†	†	†	†	†	†	†	†
B				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	:	:	:	:	:	:	:	:	:	:
C				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	:	:	:	:	:	:	:	:	:	:	:
D	Primera	2	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	4 12	6 16	↑	↑
	Segunda	2	4															‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14	11 19	17 27		
	Tercera	2	6															0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19	19 27	29 39		
	Cuarta	2	8															0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25	27 34	40 49		
	Quinta	2	10															1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29	36 40	53 58		
	Sexta	2	12															1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33	45 47	65 68		
	Séptima	2	14															2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38	53 54	77 78		
E	Primera	3	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	4 12	6 16	↑	↑	↑
	Segunda	3	6														‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14	11 19	17 27			
	Tercera	3	9														0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19	19 27	29 39			
	Cuarta	3	12														0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25	27 34	40 49			
	Quinta	3	15														1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29	36 40	53 58			
	Sexta	3	18														1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33	45 47	65 68			
	Séptima	3	21														2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38	53 54	77 78			
F	Primera	5	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑
	Segunda	5	10													‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14						
	Tercera	5	15													0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19						
	Cuarta	5	20													0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25						
	Quinta	5	25													1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29						
	Sexta	5	30													1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33						
	Séptima	5	35													2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38						
G	Primera	8	8	↓	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑
	Segunda	8	16												‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14							
	Tercera	8	24												0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19							
	Cuarta	8	32												0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25							
	Quinta	8	40												1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29							
	Sexta	8	48												1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33							
	Séptima	8	56												2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38							
H	Primera	13	13	↓	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	13	26											‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14								
	Tercera	13	39											0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19								
	Cuarta	13	52											0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25								
	Quinta	13	65											1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29								
	Sexta	13	78											1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33								
	Séptima	13	91											2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38								
J	Primera	20	20	↓	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	20	40										‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14									
	Tercera	20	60										0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19									
	Cuarta	20	80										0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25									
	Quinta	20	100										1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29									
	Sexta	20	120										1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33									
	Séptima	20	140										2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38									

↓ = Usese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.
↑ = Usese el primer plan de muestreo arriba de la flecha.
Ac = Número de aceptación.
Re = Número de rechazo.
† = Usese el plan de muestreo sencillo correspondiente (o alternativamente, úsese el plan de muestreo múltiple abajo, cuando esté disponible).
: = Usese el plan de muestreo doble correspondiente (o alternativamente, úsese el plan de muestreo sencillo abajo, cuando esté disponible).
‡ = La aceptación no se permite a este tamaño de muestra.

TABLA R. TABLA MAESTRA PARA INSPECCION NORMAL (MUESTREO MULTIPLE)—MIL-STD-105D (ESTÁNDAR ABC) (*Continúa*)

Letra clave para el tamaño de la muestra	Muestra	Tamaño de la muestra	Tamaño acumulativo de la muestra	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
				Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
K	Primera	32	32	↓	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	32	64									‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14										
	Tercera	32	96									0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19										
	Cuarta	32	128									0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25										
	Quinta	32	160									1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29										
	Sexta	32	192									1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33										
	Séptima	32	224									2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38										
L	Primera	50	50	↓	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	50	100								‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14											
	Tercera	50	150								0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19											
	Cuarta	50	200								0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25											
	Quinta	50	250								1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29											
	Sexta	50	300								1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33											
	Séptima	50	350								2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38											
M	Primera	80	80	↓	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	80	160							‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14												
	Tercera	80	240							0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19												
	Cuarta	80	320							0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25												
	Quinta	80	400							1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29												
	Sexta	80	480							1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33												
	Séptima	80	560							2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38												
N	Primera	125	125	↓	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	125	250						‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14													
	Tercera	125	375						0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19													
	Cuarta	125	500						0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25													
	Quinta	125	625						1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29													
	Sexta	125	750						1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33													
	Séptima	125	875						2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38													
P	Primera	200	200	↓	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	200	400					‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14														
	Tercera	200	600					0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19														
	Cuarta	200	800					0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25														
	Quinta	200	1 000					1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29														
	Sexta	200	1 200					1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33														
	Séptima	200	1 400					2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38														
Q	Primera	315	315	†	↑	↓	‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	315	630				‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14															
	Tercera	315	945				0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19															
	Cuarta	315	1 260				0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25															
	Quinta	315	1 575				1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29															
	Sexta	315	1 890				1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33															
	Séptima	315	2 205				2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38															
R	Primera	500	500	↑		‡ 2	‡ 2	‡ 3	‡ 4	0 4	0 5	1 7	2 9	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	Segunda	500	1 000			‡ 2	0 3	0 3	1 5	1 6	3 8	4 10	7 14																
	Tercera	500	1 500			0 2	0 3	1 4	2 6	3 8	6 10	8 13	13 19																
	Cuarta	500	2 000			0 3	1 4	2 5	3 7	5 10	8 13	12 17	19 25																
	Quinta	500	2 500			1 3	2 4	3 6	5 8	7 11	11 15	17 20	25 29																
	Sexta	500	3 000			1 3	3 5	4 6	7 9	10 12	14 17	21 23	31 33																
	Séptima	500	3 500			2 3	4 5	6 7	9 10	13 14	18 19	25 26	37 38																

↓ = úsese el primer plan de muestreo abajo de la flecha. Si el tamaño de la muestra es igual o mayor al tamaño del lote o corrida, hágase la inspección del 100%.
↑ = úsese el primer plan de muestreo arriba de la flecha.
Ac = número de aceptación.
Re = número de rechazo.
† = Usese el plan de muestreo sencillo correspondiente (o alternativamente, úsese el plan de muestreo múltiple abajo, cuando esté disponible)
‡ = La aceptación no se permite con este tamaño de muestra.

TABLA I

Número de observaciones en la muestra n	Carta para promedios			Carta para desviaciones estándar						Carta para rangos							Carta X
	Factores para límites de control			Factores para línea central		Factores para límites de control				Factores para línea central		Factores para límites de control					Factor para límites de control
	A	A_1	A_2	c_2	$1/c_2$	B_1	B_2	B_3	B_4	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4	E_2
2....	2.121	3.760	1.880	0.5642	1.7725	0	1.843	0	3.267	1.128	0.8865	0.853	0	3.686	0	3.267	2.660
3....	1.732	2.394	1.023	0.7236	1.3820	0	1.858	0	2.568	1.693	0.5907	0.888	0	4.358	0	2.575	1.772
4....	1.500	1.880	0.729	0.7979	1.2533	0	1.808	0	2.266	2.059	0.4857	0.880	0	4.698	0	2.282	1.457
5....	1.342	1.596	0.577	0.8407	1.1894	0	1.756	0	2.089	2.326	0.4299	0.864	0	4.918	0	2.115	1.290
6....	1.225	1.410	0.483	0.8686	1.1512	0.026	1.711	0.030	1.970	2.534	0.3946	0.848	0	5.078	0	2.004	1.184
7....	1.134	1.277	0.419	0.8882	1.1259	0.105	1.672	0.118	1.882	2.704	0.3698	0.833	0.205	5.203	0.076	1.924	1.109
8....	1.061	1.175	0.373	0.9027	1.1078	0.167	1.638	0.185	1.815	2.847	0.3512	0.820	0.387	5.307	0.136	1.864	1.054
9....	1.000	1.094	0.337	0.9139	1.0942	0.219	1.609	0.239	1.761	2.970	0.3367	0.808	0.546	5.394	0.184	1.816	1.010
10....	0.949	1.028	0.308	0.9227	1.0837	0.262	1.584	0.284	1.716	3.078	0.3249	0.797	0.687	5.469	0.223	1.777	0.975
11....	0.905	0.973	0.285	0.9300	1.0753	0.299	1.561	0.321	1.679	3.173	0.3152	0.787	0.812	5.534	0.256	1.744	0.946
12....	0.866	0.925	0.266	0.9359	1.0684	0.331	1.541	0.354	1.646	3.258	0.3069	0.778	0.924	5.592	0.284	1.716	0.921
13....	0.832	0.884	0.249	0.9410	1.0627	0.359	1.523	0.382	1.618	3.336	0.2998	0.770	1.026	5.646	0.308	1.692	0.899
14....	0.802	0.848	0.235	0.9453	1.0579	0.384	1.507	0.406	1.594	3.407	0.2935	0.762	1.121	5.693	0.329	1.671	0.881
15....	0.775	0.816	0.223	0.9490	1.0537	0.406	1.492	0.428	1.572	3.472	0.2880	0.755	1.207	5.737	0.348	1.652	0.864
16....	0.750	0.788	0.212	0.9523	1.0501	0.427	1.478	0.448	1.552	3.532	0.2831	0.749	1.285	5.779	0.364	1.636	0.849
17....	0.728	0.762	0.203	0.9551	1.0470	0.445	1.465	0.466	1.534	3.588	0.2787	0.743	1.359	5.817	0.379	1.621	0.836
18....	0.707	0.738	0.194	0.9576	1.0442	0.461	1.454	0.482	1.518	3.640	0.2747	0.738	1.426	5.854	0.392	1.608	0.824
19....	0.688	0.717	0.187	0.9599	1.0418	0.477	1.443	0.497	1.503	3.689	0.2711	0.733	1.490	5.888	0.404	1.596	0.813
20....	0.671	0.697	0.180	0.9619	1.0396	0.491	1.433	0.510	1.490	3.735	0.2677	0.729	1.548	5.922	0.414	1.586	0.803
21....	0.655	0.679	0.173	0.9638	1.0376	0.504	1.424	0.523	1.477	3.778	0.2647	0.724	1.606	5.950	0.425	1.575	0.794
22....	0.640	0.662	0.167	0.9655	1.0358	0.516	1.415	0.534	1.466	3.819	0.2618	0.720	1.659	5.979	0.434	1.566	0.785
23....	0.626	0.647	0.162	0.9670	1.0342	0.527	1.407	0.545	1.455	3.858	0.2592	0.716	1.710	6.006	0.443	1.557	0.778
24....	0.612	0.632	0.157	0.9684	1.0327	0.538	1.399	0.555	1.445	3.895	0.2567	0.712	1.759	6.031	0.452	1.548	0.770
25....	0.600	0.619	0.153	0.9696	1.0313	0.548	1.392	0.565	1.435	3.931	0.2544	0.709	1.804	6.058	0.459	1.541	0.763
Más de 25....	$\frac{3}{\sqrt{n}}$	$\frac{3}{\sqrt{n}}$	......	......	......	*	**	*	**	......	......	......	......	......	......	......	$\frac{3}{d_2}$

* $1 - \frac{3}{\sqrt{2n}}$ ** $1 + \frac{3}{\sqrt{2n}}$

TABLA II

Número mínimo m de muestras de tamaño n requerido para elaborar una carta $\bar{X}$ con una confianza de 98%, cuando se emplean los rangos.

n	m
2	15
3	9
4	7
5	6
6	5
7	5
8	4
9	4
10	4
12	4
14	4
16	3
18	3
20	3

TABLA III

Número mínimo m de muestras de tamaño n requerido para elaborar una carta $\bar{\bar{X}}$ con una confianza de 98%, cuando se emplean las desviaciones estándar.

n	m
2	16
3	9
4	7
5	6
6	5
7	5
8	4
9	4
10	4
12	4
14	3
16	3
18	3
20	3

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

CONTROL ESTADISTIO DE CALIDAD

CONFIABILIDAD DE COMPONENTES Y SISTEMAS

DR. OCTAVIO A. RASCON CHAVEZ

24 OCTUBRE, 1978.

CAPITULO 3

INTRODUCCION A PROCESOS ESTOCASTICOS

por Octavio A. Rascón Ch.*

3.1 *Introducción*

A menudo los especialistas en Investigación de Operaciones, Econometría, Teoría de mantenimiento, etc., se enfrentan a problemas en los cuales se realizan observaciones de cierto fenómeno durante un lapso de tiempo. Cuando la entidad matemática (ya sea variable escalar o vectorial, función, etc.) que caracteriza a ese fenómeno se comporta en forma aleatoria, se dice que la entidad está sujeta (o sigue) a un *proceso estocástico.*

Como ejemplos de cantidades escalares que siguen procesos estocásticos se pueden citar los siguientes:

- El nivel de inventario de cierto producto en una empresa
- El número de vehículos que circulan durante el día por una avenida
- Las variaciones temporales en la calidad de un producto elaborado
- El flujo de agua en un río
- La demanda diaria de agua potable en una ciudad
- El comportamiento de partículas sujetas a impactos que ocurren al azar
- Las aceleraciones del terreno durante un sismo
- El número de personas que esperan servicio en una línea de espera, etc.

3.2 *Definición de un proceso estocástico*

Un proceso estocástico se define como un conjunto, $\{X(t), t \varepsilon T\}$ de variables aleatorias $X(t)$, donde el símbolo ε es el empleado en teoría de conjuntos para indicar *pertenencia* ($t \varepsilon T$ se lee "t está en T") y T es el conjunto de los valores que puede tomar el parámetro determinístico t. Este parámetro suele representar tiempo, distancia, área, volumen, etc.

Cada función $X(t)$ constituye una *realización* o una *función muestra* del proceso; el conjunto de todas las funciones muestra posibles constituye el proceso completo (figs 3.1a y 3.1b).

---

* Jefe de la División de Estudios Superiores de la Facultad de Ingeniería, UNAM.

Los procesos estocásticos se pueden clasificar según la naturaleza del conjunto T. Cuando T es un conjunto discontinuo o discreto, se dice que el proceso es de *parámetro discreto*, si T es un conjunto continuo, entonces el proceso es de *parámetro continuo*.

Un proceso de parámetro discreto sería, por ejemplo, la demanda semanal de llantas para automóvil en una fábrica; en tal caso la unidad de tiempo que se usa es de una semana, y el espacio T puede ser el conjunto de todas las semanas del año.

$$T = \{1,2,\ldots,i,\ldots, 52\}$$

en donde el 1 representa la semana número 1, el 2, la número 2, etc. Aquí la demanda en la semana i de cada año es una variable aleatoria; el registro de un año completo es una realización o función muestra del proceso; el grupo de los registros disponibles de años pasados es la muestra del proceso, y el conjunto de los registros de todos los años pasados y futuros constituye el proceso estocástico en cuestión.

Un ejemplo de proceso de parámetro continuo lo constituyen las aceleraciones instantáneas del terreno durante un sismo; cada acelerograma (gráfica aceleración vs tiempo) representa una muestra o realización del proceso. En este caso, la aceleración en el instante t_i es una variable aleatoria. Así, si los elementos de la fig 3.1a representan acelerogramas, se observa que en el tiempo t_i las aceleraciones $X_1(t_i)$, $X_2(t_i)$, $X_3(t_i)$, etc., varían de una muestra a otra, ocurriendo dicha variación al azar.

...

Otra forma de clasificar los procesos estocásticos se basa en la naturaleza del *espacio muestral* de las variables aleatorias que lo constituyen. Así, si dichas variables son continuas, el proceso se denomina *continuo*, y es *discreto* o *discontinuo* si las variables aleatorias son discretas. Por ejemplo, el caso del problema de la demanda de llantas que se mencionó constituye un proceso discreto porque $X(t)$ sólo puede tomar algunos de los valores discretos 0, 1, 2,.... El caso de las aceleraciones del terreno durante un sismo constituye un proceso continuo porque $X(t)$ puede tomar valores del conjunto continuo $\{-\infty, \infty\}$.

3.3 *Descripción de la ley de probabilidades de un proceso estocástico*

Puesto que las ordenadas en cada instante de un proceso estocástico son variables aleatorias, para describir la ley de probabilidades del proceso es necesario establecer todas las *funciones de distribución conjuntas de probabilidades*.

La función de distribución conjunta de probabilidades $F(z_1, z_2,...,z_n)$, de n variables aleatorias escalares, Z_1, Z_2,..., Z_n, se define como la probabilidad, $P[Z_1 \leq z_1, Z_2 \leq z_2,...,Z_n \leq z_n]$, de que simultáneamente Z_1 asuma un valor menor o igual que z_1, que Z_2 asuma un valor menor o igual que z_2, etc., es decir,

$$F(z_1, z_2,..., z_n) = P[Z_1 \leq z_1, Z_2 \leq z_2,...,Z_n \leq z_n]$$

en donde z_1, z_2,...,z_n son valores numéricos. Por lo tanto, la función de distribución conjunta de orden n del proceso estocástico $X(t)$ es la que corresponde a las n variables aleatorias $X(t_1)$, $X(t_2)$,...,$X(t_n)$, para todo t_1, t_2,...,t_n, es decir,

$$F(x_1, x_2,...,x_n; t_1, t_2,...,t_n) = P[X(t_1) \leq x_1, X(t_2) \leq x_2,...,X(t_n) \leq x_n]$$

(El miembro derecho de la ecuación anterior se lee "probabilidad de que simultáneamente $X(t_1) \leq x_1$, $X(t_2) \leq x_2$,..., $X(t_n) \leq x_n$".)

En consecuencia, para contar con la ley de probabilidades del proceso estocástico $X(t)$ es necesario obtener las

...

funciones de distribución conjuntas para todo entero n. Debido a que el conocer todas estas funciones suele ser muy complicado en las aplicaciones prácticas, a menudo el que analiza el proceso se conforma con tener esa función para $n = 2$ como máximo, es decir, con establecer las funciones de distribución de una y dos variables aleatorias.

$$F(x_1;t_1) = P[X(t_1) \leq x_1]$$

$$F(x_1, x_2; t_1, t_2) = P[X(t_1) < x_1, X(t_2) < x_2]$$

Conocidas estas funciones de distribución es posible conocer las densidades de probabilidades correspondientes, mediante las conocidas relaciones de la teoría de probabilidades.

$$f(x_1, t_1) = \frac{dF(x_1, t_1)}{dx_1} \tag{3.1}$$

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \ \partial x_2} \tag{3.2}$$

en donde ∂ es el símbolo de derivación parcial.

Cuando un proceso tiene densidades de probabilidades normales (gausianas), se dice que el proceso es *normal* a *gausiano*.

Ejemplo 3.1

Consideremos el proceso estocástico de parámetro continuo

$$X(t) = A \cos wt + B \text{ sen } wt$$

donde A y B son variables aleatorias independientes, ambas con distribución normal de media cero y variancia σ^2, y ω es una constante positiva. Este es un proceso continuo porque las variables aleatorias A y B son continuas. Calculemos, por ejemplo, la probabilidad de que la integral de 0 a L del cuadrado del proceso sea mayor que c, o sea,

$$P[\int_0^L X^2(t)\,dt > c] = ?$$

donde c es una constante, y $L = 2\pi/\omega$ es el periodo del proceso.

...

El valor de la integral es

$$\int_0^L X^2(t)\,dt = \int_0^L (A^2 \cos^2 \omega t + 2AB \cos \omega t \operatorname{sen} \omega t + B^2 \operatorname{sen}^2 \omega t)\,dt$$

$$= L(A^2+B^2)/2$$

Por lo tanto, la probabilidad de que el miembro izquierdo de la ecuación anterior exceda a c es igual a la probabilidad de que el miembro derecho sea mayor que c, es decir,

$$P[\int_0^L X^2(t)\,dt > c] = P[L(A^2+B^2)/2 > c] = P[(A^2+B^2) > c\omega/\pi]$$

Se puede demostrar (ref 3.1) que como A y B son variables aleatorias independientes con distribución normal, entonces $(A^2+B^2)/\sigma^2 = Z$ tiene distribución ji-cuadrada (X^2) con dos grados de libertad. De aquí se deduce que si hacemos $A^2+B^2=Y$, entonces $Y=Z\sigma^2$, $dz=dy/\sigma^2$ y $P[A^2+B^2>c\omega/\pi] = \int_{c\omega/\pi}^{\infty} \frac{1}{2\sigma^2} e^{-y/2\sigma^2} dy = e^{-c\omega/2\pi\sigma^2}$

Puesto que

$$X(t) = A \cos \omega t + B \operatorname{sen} \omega t = D \cos (\omega t - \theta)$$

donde $D=\sqrt{A^2+B^2}$ y $\theta = \tan^{-1}(B/A)$, se concluye que lo que varía de una función muestra a otra del proceso, conservándose en cada caso la onda armónica de frecuencia ω, es la amplitud, D, y el ángulo de fase, θ, de la onda.

Ejemplo 3.2

Sea el proceso estocástico de parámetro continuo

$$X(t) = R + St$$

donde R y S son dos variables aleatorias independientes con densidades de probabilidades $f_R(r)$ y $f_S(s)$, respectivamente. Cada pareja de s y r (valores que asumen S y R, respectivamente)

...

conduce a una recta que constituye una realización del proceso, por lo que éste está constituído por una familia de líneas rectas con pendientes s, y de ordenadas al origen r, aleatorias.

Para obtener la función de distribución, $F(x;t)$, de primer orden del proceso $X(t)$, hagamos

$$X = R + S_1 \tag{3.3}$$

donde $S_1 = tS$. En tal caso,

$$F_X(x) = P[X \leq x] = P[R + S_1 \leq x] = \iint_{D_X} f_{RS_1}(r,s_1)\,dr\,ds_1 \tag{3.4}$$

donde $f_{RS_1}(r,s_1)$ es la densidad de probabilidades conjunta de las variables aleatorias R y S_1, y la región de integración D_X es tal que $r + s_1 \leq x$; esta región queda representada por la zona sombreada de la fig 3.2, como puede verificarse fácilmente. Por lo tanto, la ec 2.4 queda en la forma

$$F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x-r} f_{RS_1}(r,s_1)\,dr\,ds_1 \tag{3.5}$$

Derivando la ec 3.5 respecto a x obtenemos la densidad de probabilidades, $f_X(x)$, de primer orden de X (ec 3.1)

$$f_X(x) = \int_{-\infty}^{\infty} f_{RS_1}(r, x-r)\,dr \tag{3.6}$$

Puesto que en este caso las variables aleatorias R y S_1 son independientes, se tiene que

$$f_{RS_1}(r, x-r) = f_R(r)\,f_{S_1}(x-r)$$

por lo que

$$f_X(x) = \int_{-\infty}^{\infty} f_R(r)\,f_{S_1}(x-r)\,dr \tag{3.7}$$

Para aplicar la ec 3.7 al problema que estamos resolviendo, es necesario obtener primero la densidad de probabilidades,

...

$f_{S_1}(s_1)$, de la variable aleatoria $S_1 = tS$. Para lograr esto, hagamos lo siguiente:

Si $t > 0$,

$$F_{S_1}(s_1) = P[S_1 \leq s_1] = P[S \leq s_1/t] = \int_{-\infty}^{S_1/t} f_S(s)\,ds$$

La densidad de probabilidades de S_1 se obtiene derivando la ecuación anterior respecto a s_1, es decir,

$$f_{S_1}(s_1) = \frac{d}{ds_1} F_{S_1}(s_1) = \frac{1}{t} f_S(s_1/t) \qquad (3.8)$$

Si $t < 0$, entonces $ts \leq s_1$ para $s > s_1/t$, por lo que

$$F_{S_1}(s_1) = P[S_1 \leq s_1] = P[S > s_1/t] = 1 - \int_{-\infty}^{s_1/t} f_S(s)\,ds$$

y

$$f_{S_1}(s_1) = \frac{d}{ds_1} F_{S_1}(s_1) = 0 - \frac{1}{t} f(s_1/t) = \frac{1}{|t|} f_S(s_1/t) \qquad (3.9)$$

Las ecs 3.8 y 3.9 se pueden combinar para obtener una fórmula válida para todo valor de t, dando como resultado

$$f_{S_1}(s_1) = \frac{1}{|t|} f_S(s_1/t) \qquad (3.10)$$

donde $|t|$ denota el valor absoluto de t.

Sustituyendo la ec 3.10 en la ec 3.7 se obtiene

$$f_X(x) = \frac{1}{|t|} \int_{-\infty}^{\infty} f_R(r) f_S\left(\frac{x-r}{t}\right) dr \qquad (3.11)$$

Consideremos el caso particular en que R y S tienen distribuciones exponenciales con parámetros α y β ($\alpha \neq \beta$), respectivamente, es decir,

$$f_R(r) = \alpha\, e^{-\alpha r} \quad ; r > 0$$

$$f_S(s) = \beta\, e^{-\beta s} \quad ; s > 0$$

En estas circunstancias, el límite inferior de la integral de la ec 3.11 es cero, ya que $f_R(r)$ es cero para valores negativos

...

de r, y el superior es tal que $(x - r)/t \geq 0$, lo cual ocurre para $t > 0$ cuando $r \leq x$, por lo que la ec 2.11 queda en la forma:

$$f_X(x) = \frac{1}{|t|} \int_0^x \alpha\, e^{-\alpha r}\, \beta\, e^{-\beta(x-r)/t} dr$$

$$= \frac{1}{|t|} \int_0^x \alpha\, \beta\, e^{-\beta x/t}\, e^{-(\beta/t-\alpha) r} dr$$

$$= \frac{\alpha\beta}{|t|(\beta/t-\alpha)}\, e^{-\beta x/t} \left[e^{(\beta/t\, -\alpha) r}\right]_0^x$$

$$= \frac{\alpha\beta}{\beta - \alpha t}\, e^{-\beta x/t}\, (e^{(\beta/t\, -\alpha) x} - 1)$$

$$= \frac{\alpha\beta}{\beta - \alpha t}(e^{-\alpha x} - e^{-\beta x/t}) \tag{3.12}$$

El mismo resultado se obtiene para $t < 0$, ya que los límites de integración de la ecuación anterior se intercambian, apareciendo un signo menos fuera de la integral al colocarlos en la posición que tienen para $t > 0$.

3.4 *Media, autocorreación y autocovariancia de un proceso estocástico*

Es bien sabido que el conocer la media y la variancia de una variable aleatoria no es suficiente para determinar su comportamiento probabilístico, sino que es indispensable se sepa cuál es su densidad de probabilidades. Pero, aún cuando ésta no se conozca, los parámetros anteriores son de gran utilidad porque resumen ciertas características de tendencia central y de dispersión de los valores que asume la variable, e incluso permiten estimar, aunque sea burdamente, algunas probabilidades mediante la desigualdad de Chebyshev (ref 3.3).

...

Una utilidad semejante a la que tienen la media y la variancia de una variable aleatoria, la tienen las funciones de *valor medio* (*media*), *autocorrelación* y *autocovariancia* de un proceso estocástico.

La *media*, $E[X(t)] = \eta(t)$, de un proceso continuo $X(t)$ se define como la esperanza de la variable aleatoria $X(t)$ en el instante t, es decir,

$$E[X(t)] = \eta(t) = \int_{-\infty}^{\infty} x f(x;t)\,dx \tag{3.13}$$

donde $E[X(t)]$ se lee "esperanza de $X(t)$". Esta función representa el valor medio del proceso en el instante t. Si el proceso es discreto, la integral de la ec 3.13 se cambia por una suma que abarque todos los valores que puede asumir $X(t)$, es decir,

$$\eta(t) = \sum_{i=-\infty}^{\infty} x_i f(x_i;t) \tag{3.14}$$

Este comentario se aplica también a las definiciones que siguen.

Para plantear las definiciones de las funciones de autocorrelación y de autocovariancia de un proceso estocástico, es necesario indicar que la *correlación*, $R(Z,Y)$, de dos variables aleatorias, Z y Y, se define como la esperanza de su producto. Así, si Z y Y son continuas, se tiene que

$$R(Z,\ Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z\ y\ f(z,y)\ dz\ dy \tag{3.15}$$

en donde $f(z,y)$ es la densidad conjunta de probabilidades de Z y Y (ref 3.1). Por otra parte, la *covariancia*, $C(Z,\ Y)$, de Z y Y se define como la esperanza del producto $\{z-E[Z]\}\{y-E[Y]\}$, es decir,

...

$$C(Z,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{z-E[Z]\}\{y-E[Y]\}f(z,y)dz\ dy \qquad (3.16)$$

Volviendo al proceso estocástico $X(t)$, su *función de autocorrelación*, $R(t_1,t_2)$, se define como la esperanza del producto de las variables aleatorias $X(t_1)$ y $X(t_2)$ para todo t_1 y t_2 (es la correlación de $X(t_1)$ y $X(t_2)$) que, de acuerdo con la ec 3.15, resulta ser

$$R(t_1,t_2) = E[X(t_1)\ X(t_2)] = \int_{-\infty}^{\infty} x_1\ x_2\ f(x_1,x_2;t_1,t_2)dx_1\ dx_2 \quad (3.17)$$

la cual, en general, es una función de t_1 y t_2. Si $t_1 = t_2 = t$, entonces $E[X^2(t)]$ es la función del *valor medio cuadrático del proceso*, o sea, del valor medio del proceso elevado al cuadrado.

La *función de autocovariancia*, $C(t_1\ t_2)$, del proceso $X(t)$ es la covariancia de las variables aleatorias $X(t_1)$ y $X(t_2)$ para todo t_1 y t_2, es decir (ec 3.16)

$$C(t_1,t_2) = E\{[X(t_1) - \eta(t_1)][X(t_2) - \eta(t_2)]\} \quad (3.18)$$

La ec 3.18 se puede escribir como

$$C(t_1,t_2)=E[X(t_1)\ X(t_2)]-E[X(t_1)\eta(t_2)]-E[\eta(t_1)X(t_2)]+E[\eta(t_1)\eta(t_2)]$$

Considerando ahora que la esperanza de una constante es la propia constante, y que la del producto de una constante por una variable aleatoria es igual a la constante multiplicada por la esperanza de la variable, la ecuación anterior se puede reescribir como

$$C(t_1,t_2)=E[X(t_1)X(t_2)] - \eta(t_2)E[X(t_1)] - E[X(t_2)]\eta(t_1)+\eta(t_1)\eta(t_2)$$

$$= R(t_1,t_2) - \eta(t_2)\eta(t_1) - \eta(t_2)\eta(t_1) + \eta(t_1)\eta(t_2) \quad (3.19)$$

Reagrupando términos se llega a

$$C(t_1,t_2) = R(t_1,t_2) - \eta(t_1)\eta(t_2) \qquad (3.20)$$

...

$\sigma^2(R)=1/\alpha^2$, $\sigma^2(S)=1/\beta^2$, $E[R^2] = 2/\alpha^2$ y $E[S^2]=2/\beta^2$, como puede verse en la tabla 2.1. Además, considerando que S y R son variables aleatorias independientes, se tiene que $E[SR]= E[S]E[R]= 1/(\alpha\beta)$, y $\mathrm{Cov}(S,R)=0$. Sustituyendo estos valores en las ecs 3.21, 3.22 y 3.23 se obtiene finalmente

$$\eta(t) = 1/\alpha^2 + t_1 t_2(1/\beta^2)$$
$$R(t_1,t_2) = 2/\alpha^2 + t_1 t_2(2/\beta^2) + (t_1+t_2)[1/(\alpha\beta)]$$
$$C(t_1,t_2) = 1/\alpha^2 + t_1 t_2(1/\beta^2)$$

Ejemplo 3.4. Paseo casual. El camino o paseo casual es un proceso estocástico discreto de parámetro discreto, en el cual los cambios en los valores del proceso ocurren cada T segundos. Dichos cambios obedecen a un mecanismo aleatorio, de tal manera que en cada instante el proceso se incrementa o decrece en una unidad (el proceso da un "paso" a la derecha o a la izquierda, motivo por el cual se denomina *paseo casual*).

Para deducir la distribución de probabilidades de este proceso, supongamos que el mecanismo aleatorio que marca los cambios de valores es el resultado del lanzamiento de una moneda, de manera que si cae cara el proceso da un paso de amplitud s a la derecha, y si cae sol lo da a la izquierda. Con base en esto, es evidente que cada función muestra del proceso dependerá de la secuencia de caras, C, y soles, S, que haya salido; en la figura 3.3 se muestra el principio de una de estas funciones, correspondiente a la secuencia *CCSSSCCS*... (se supone que en t=0 el proceso parte del origen). Obsérvese que las discontinuidades del proceso ocurren en los tiempos

...

$t_i = n_i T$, en donde n_i es el número del lanzamiento de la moneda; además, los valores que puede asumir el paseo casual son... -3, -2, -1, 0, 1, 2, 3,..., es decir, todos los números enteros positivos y negativos.

Supongamos que en los primeros n lanzamientos ocurren k caras, con lo cual en el lapso $t=nT$ se dan k pasos a la derecha y $n - k$ pasos a la izquierda. Por consiguiente, si $X(0) = 0$ (parte del origen) el proceso se encuentra en la posición

$$X(nT) = ks - (n-k)s = (2k - n)s$$

donde s es el tamaño de cada paso.

Si hacemos $2k-n = r$, vemos que $X(nT)$ es una variable aleatoria que toma los valores rs, en donde $r=n,n-2,\ldots,-n$. Claramente $X(nT) = rs$ ocurre cuando aparecen k caras en n lanzamientos, cuya probabilidad queda dada por la distribución binomial. Puesto que

$$k = \frac{r + n}{2}$$

tendremos

$$P[X(nT)=rs] = P\left[k = \frac{r + n}{2}\right] = \binom{n}{\frac{r+n}{2}} p^k q^{n-k} = \frac{n!}{\left(\frac{r+n}{2}\right)!\left(n-\frac{r+n}{2}\right)!} p^k q^{n-k}$$

en donde p es la probabilidad de que la moneda caiga de cara, y $q=1-p$ es la de que caiga de sol. En lo que sigue desarrollaremos el caso en que la moneda es homogénea (no está cargada), es decir, $p=q=1/2$, con lo cual la ecuación anterior queda en la forma

...

$$P[X(nT) = rs] = \frac{n!}{(\frac{r+n}{2})!\,(n-\frac{r+n}{2})!}\,\frac{1}{2^n} \qquad (3.24)$$

Para obtener la esperanza de este proceso aplicamos la ec 3.14, es decir,

$$E[X(nT)] = \sum_{r=-n}^{n} rs\,\frac{n!}{(\frac{r+n}{2})!\,(n-\frac{r+n}{2})!}\,\frac{1}{2^n} \qquad (3.25)$$

lo cual se puede demostrar que vale cero. De igual manera, la función del valor medio cuadrático del proceso es

$$E[X^2(nT)] = \sum_{r=-n}^{n} (rs)^2\,\frac{n!}{(\frac{r+n}{2})!\,(n-\frac{r+n}{2})!}\,\frac{1}{2^n} \qquad (3.26)$$

lo cual da $E[X^2(nT)]=ns^2$. Estos resultados se pueden obtener también si usamos las variables aleatorias independientes X_i que puede tomar los valores s y $-s$ con probabilidades respectivas de 1/2. Evidentemente,

$$E[X_i] = 0 \qquad E[X_i^2] = s^2$$

Puesto que

$$X(nT) = X_1 + X_2 + \ldots + X_n$$

la esperanza de $X(nT)$ es igual a la suma de las esperanzas de las X_i, lo cual da cero; además, puesto que las variables son independientes, la esperanza del cuadrado de $X(nT)$ es igual a la suma de las esperanzas de los cuadrados de las X_i, lo cual es:

$$E[X^2(nT)] = ns^2$$

Para obtener la densidad de probabilidades de segundo orden, $f(x_1,x_2;t_1,t_2)$, con $t_1<t_2$, usaremos la relación (ref 3.1)

...

$$f(x_1,x_2;t_1,t_2) = f(x_2;t_2|x_1;t_1)f(x_1;t_1)$$

en donde $f(x_2;t_2|x_1;t_1)$ es la densidad de probabilidades condicional de que $X_2=x_2$ en $t=t_2$ dado que en $t=t_1$ se tuvo $X_1=x_1$ (ésta nos permite obtener las probabilidades de que el proceso esté en la posición $x_2=r_2 s$ en el tiempo $t_2=n_2 T$, cuando se sabe que en el tiempo $t_1= n_1 T$ estuvo en $x_1= r_1 s)$ y $f(x_1;t_1) = P[X(n_1T)=x_1]$.

Para pasar de $x_1 = r_1 s$ a $x_2= r_2 s$ es necesario dar por lo menos (r_2-r_1) pasos en (n_2-n_1) unidades T de tiempo. Por consiguiente, $f(x_2;t_2|x_1;t_1)=0$ si $(n_2-n_1)<(r_2-r_1)$. Para el caso $(n_2-n_1)\geq(r_2-r_1)$ la densidad condicional se deduce de manera similar a como se hizo para la densidad de primer orden, pero ahora partiendo de $x_1=r_1 s$ en el tiempo $t_1=n_1T$, en vez de partir de $x_1=0$ en el tiempo cero. El resultado es la siguiente distribución binomial

$$f(x_2;t_2|x_1;t_1)=P[X(n_2T)=r_2 s|X(n_1T)= r_1 s] = \binom{n_2-n_1}{\frac{(r_2-r_1)+(n_2-n_1)}{2}}\frac{1}{2^{n_2-n_1}}$$

$$= \frac{(n_2-n_1)!}{\left(\frac{(r_2-r_1)+(n_2-n_1)}{2}\right)!\left[(n_2-n_1)-\frac{(r_2-r_1)+(n_2-n_1)}{2}\right]!}\frac{1}{2^{n_2-n_1}}$$

Tomando esto en cuenta, la distribución conjunta ~~qu~~eda en la forma

$$f[X(n_1T),X(n_2T);\ n_1T,\ n_2T]=\binom{n_2-n_1}{\frac{(r_2-r_1)+(n_2-n_1)}{2}}\binom{n_1}{\frac{r_1+n_1}{2}}\frac{1}{2^{n_2}} \qquad (3.27)$$

Existen algunas variantes del paseo casual, de las cuales las más comunes son:

...

i) La opción de avanzar a la izquierda se cambia por la de permanecer en el mismo sitio; es decir, en este paseo en cada unidad de tiempo o se avanza a la derecha o no se avanza.

ii) Se tienen las tres opciones: avanzar a la derecha con probabilidad p, a la izquierda con probabilidad q, o permanecer en el sitio con probabilidad $v=1-p-q$.

El estudio de los caminos casuales ha encontrado aplicación en muchos problemas de física, ingeniería, inventarios (ref 3.9), etc. Como ejemplo puede citarse la ref 3.10 en la cual se usó un paseo casual para deducir la densidad de probabilidades del número de repeticiones de carga que hay que aplicarle a un objeto para que se rompa (problema de confiabilidad de fatiga de materiales).

3.5 *Procesos estocásticos estacionarios*

Como puede observarse en las ecs 3.1 y 3.2, las densidades de probabilidades de un proceso estocástico son, en general, funciones del tiempo y, por lo tanto, también lo son la esperanza y la autocorrelación del mismo (ecs 3.13 y 3.17). Hay, sin embargo, una clase especial de procesos llamados *procesos estrictamente estacionarios*, tales que sus densidades de probabilidades de cualquier orden son invariantes si se traslada el origen de la escala del tiempo. En particular, la densidad de probabilidades de primer orden no resulta ser función del tiempo y, por consiguiente, la esperanza del proceso es una constante, es decir,

...

$$f(x_1;t) = f(x_1) \tag{3.28}$$

y

$$E[X(t)] = \eta \tag{3.29}$$

Además, en este caso la densidad de probabilidades de segundo orden y la función de autocorrelación no son funciones de los tiempos t_1 y t_2 por separado, sino sólo de su diferencia $\tau=t_2-t_1$, es decir,

$$f(x_1,x_2;t_1,t_2) = f(x_1,x_2;\tau) \tag{3.30}$$

y

$$R(t_1,t_2) = R(\tau) \tag{3.31}$$

De la ec 3.19 se deduce que la covariancia de un proceso estrictamente estacionario también depende sólo de τ, ya que en tal caso queda

$$C(t_1,t_2) = R(\tau)-\eta^2=C(\tau) \tag{3.32}$$

En la mayoría de los procesos estocásticos que se presentan en los problemas prácticos no se cuenta con todas las distribuciones de probabilidades, por lo cual resulta imposible verificar si tales procesos son estrictamente estacionarios. Por tal motivo, se ha convenido en definir un *proceso estacionario en el sentido amplio*, como aquel cuya esperanza es constante y cuya función de autocorrelación es función sólo de la diferencia de tiempos $\tau=t_2-t_1$, sin importar lo que suceda con las distribuciones de probabilidades de cualquier orden.

De las dos definiciones de estacionaridad que se han

...

presentado, se puede concluir que un proceso estrictamente estacionario también es estacionario en el sentido amplio, pero si lo es en el sentido amplio no necesariamente es estrictamente estacionario.

Por otra parte, si un proceso no es estacionario en el sentido amplio, tampoco es estrictamente estacionario, ya que la ec 3.28 no se cumpliría.

Ejemplo 3.5

¿Es el proceso $X(t)=R+St$ visto en el ejemplo 3.3

a) estacionario en el sentido amplio

b) estrictamente estacionario?

a) La respuesta a este problema es inmediata, ya que en la ec 3.21 se observa que la esperanza del proceso sí es función del tiempo y, en la ec 3.22, que la autocorrelación es función de t_1 *y* t_2, y no de $t_2 - t_1$. Por consiguiente, el proceso, *no* es estacionario en el sentido amplio.

b) Puesto que el proceso no es estacionario en el sentido amplio, tampoco es estrictamente estacionario. Esto se puede verificar al observar que la densidad de probabilidades dada en la ec 3.12 es función del tiempo, t.

Ejemplo 3.6

¿Es estacionario, en el sentido amplio, el camino casual estudiado en el ejemplo 3.4?

La respuesta es *no*, ya que aun cuando la esperanza vale cero (es constante), la función de valor medio cuadrático

...

es $n s^2$ (no es constante por depender de n) y la autocorrelación depende de n_2 y n_1 por separado y no de $n_2 - n_1$, puesto que la densidad de probabilidades de segundo orden depende de n_2 y n_1 (ec 3.27).

Demostraremos esto último para el caso del paseo casual en que con probabilidad p se avanza un paso y con probabilidad 1-p se permanece en el sitio en cada T segundos. La densidad de probabilidades de segundo orden correspondiente es (ref 6.13).

$$f[X(n_1T),\ X(n_2T); n_1T, n_2T] = \binom{n_2-n_1}{r_2-r_1}\binom{n_1}{r_1} p^{r_2}(1-p)^{n_2-r_2}$$

Por lo tanto

$$E[X(n_1T)X(n_2T)] = s^2 \sum_{r_1=0}^{n_1} \sum_{r_2=r_1}^{r_1+n_2-n_1} r_2 r_1 \binom{n_2-n_1}{r_2-r_1}\binom{n_1}{r_1} p^{r_2}(1-p)^{n_2-r_2}$$

Haciendo la transformación $m = r_2 - r_1$, multiplicando y dividiendo por $(1-p)^{n_1}$, y reagrupando términos se obtiene:

$$E[X(n_1T)X(n_2T)] = s^2 \sum_{r_1=0}^{n_1} r_1 \binom{n_1}{r_1} p^{r_1}(1-p)^{n_1-r_1} \sum_{m=0}^{n_2-n_1} (m+r_1)\binom{n_2-n_1}{m} p^m (1-p)^{n_2-n_1-m}$$

Dividiendo la suma del lado derecho en dos partes se obtiene

$$\sum_{m=0}^{n_2-n_1} m \binom{n_2-n_1}{m} p^m (1-p)^{n_2-n_1-m} = (n_2-n_1)p$$

y

$$r_1 \sum_{m=0}^{n_2-n_1} \binom{n_2-n_1}{m} p^m (1-p)^{n_2-n_1-m} = r_1$$

ya que la suma se realiza con todos los términos de la distribusión binomial. Por lo tanto

... ...

$$E[X(n_1T)X(n_2T)] = s^2 \sum_{r_1=0}^{n_1} r_1 [r_1 + p(n_2 - n_1)] \binom{n_1}{r_1} p^{r_1} (1-p)^{n_1 - r_1}$$

$$= s^2\{n_1 p[1-p(1-n_1)] + p(n_2-n_1)n_1 p\} = s^2 n_1 p[1-p(1-n_2)]$$

que es una función de n_1 y n_2, y no de su diferencia $n_2 - n_1$.

Ejemplo 3.7

Demostrar que el proceso del ejemplo 3.1, $X(t) = D \cos (\omega t - \phi)$, donde D y ω son constantes, y ϕ es una variable aleatoria con densidad de probabilidades uniforme entre $-\pi$ y π, es estacionario en el sentido amplio.

Para resolver este problema veamos primero si $E[X(t)]$ = constante.

$$E[X(t)] = E[D \cos (\omega t - \phi)]$$

$$= D\ E[\cos \omega t \cos \phi + \operatorname{sen} \omega t \operatorname{sen} \phi]$$

$$= D \cos \omega t\ E[\cos \phi] + D \operatorname{sen} \omega t\ E[\operatorname{sen} \phi]$$

Pero, considerando que la densidad de probabilidades de ϕ es uniforme entre $-\pi$ y π, es decir,

$$f(\phi) = \frac{1}{2\pi} \qquad \text{si } -\pi \leq \phi \leq \pi$$

se obtiene

$$E[\cos \phi] = \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos \phi\ d\phi = \frac{1}{2\pi}\Big[\operatorname{sen} \phi\Big]_{-\pi}^{\pi} = \frac{1}{2\pi}(0-0) = 0$$

y

$$E[\operatorname{sen} \phi] = \int_{-\pi}^{\pi} \frac{1}{2\pi} \operatorname{sen} \phi\ d\phi = \frac{-1}{2\pi}\Big[\cos \phi\Big]_{-\pi}^{\pi} = \frac{-1}{2\pi}[-1-(-1)] = 0$$

Por lo tanto, $E[X(t)]=0$, que es una constante.

Veamos ahora si la autocorrelación es función $\tau = t_2 - t_1$

$$R(t_1,t_2) = E[D\cos(\omega t_1-\phi)\; D\cos(\omega t_2-\phi)] =$$

$$= D^2 E[\cos(\omega t_1-\phi)\cos(\omega t_2-\phi)]$$

Desarrollando los cosenos de la ecuación anterior, efectuando las multiplicaciones y agrupando términos se obtiene

$$R(t_2,t_1) = D^2\;\{\cos\omega t_1\cos\omega t_2 E[\cos^2\phi]+(\cos\omega t_1\operatorname{sen}\omega t_2 +$$

$$+\operatorname{sen}\omega t_1\cos\omega t_2)E[\operatorname{sen}\phi\cos\phi] + \operatorname{sen}\omega t_1\operatorname{sen}\omega t_2 E[\operatorname{sen}^2\phi] \quad (3.33)$$

Pero

$$E[\cos^2\phi] = \int_{-\pi}^{\pi}\frac{1}{2\pi}\cos^2\phi\; d\phi = 1/2$$

$$E[\operatorname{sen}\phi\cos\phi] = \int_{-\pi}^{\pi}\frac{1}{2\pi}\operatorname{sen}\phi\cos\phi\; d\phi = 0$$

$$E[\operatorname{sen}^2\phi] = \int_{-\pi}^{\pi}\frac{1}{2\pi}\operatorname{sen}^2\phi\; d\phi = 1/2$$

por lo que la autocorrelación (ec 3.33) queda en la forma

$$R(t_1,t_2) = D^2\frac{1}{2}(\cos\omega t_1\cos\omega t_2 + \operatorname{sen}\omega t_1\operatorname{sen}\omega t_2)$$

$$= \frac{D^2}{2}\cos(\omega t_2-\omega t_1) = \frac{D^2}{2}\cos\omega(t_2-t_1) = \frac{D^2}{2}\cos\omega\tau$$

que es una función de τ y, en consecuencia, el proceso estudiado sí es estacionario en el sentido amplio.

3.6 *Proceso simple de Poisson*

Uno de los procesos estocásticos que más se emplean para resolver problemas de Investigación de Operaciones, Ingeniería de Sistemas y de Física, es el denominado *proceso simple de Poisson*. Se ha empleado, por ejemplo, para describir la ocurrencia de tormentas, de inundaciones y de flujo de vehículos,

...

en la teoría de espera (colas), etc (refs 3.4 - 3.7).

El proceso simple de Poisson es un proceso estocástico en el que se cuenta el número de ocurrencias de algún evento específico (por ejemplo, el número de vehículos que pasan por cierto punto de una carretera); por este motivo en algunos textos se le denomina *proceso de conteo de Poisson.*

Para estudiar el proceso simple de Poisson, calculemos la densidad de probabilidades del número de ocurrencias de cierto evento en un lapso $t_2 - t_1 = t_a$, cuando el tiempo de ocurrencia de dicho evento tiene densidad de probabilidades uniforme en el intervalo de 0 a L (figs 3.4*a* y 3.4*b*). En este caso, la probabilidad, p, de que ocurra el evento una vez en el período de t_1 a t_2 es

$$p = \frac{t_2 - t_1}{L} = \frac{t_a}{L}$$

Por consiguiente, la probabilidad de que el evento ocurra k veces en el lapso t_a, si sabemos que de 0 a L ocurre n veces es (distribución binomial)

$$P[k \text{ en } t_a] = \binom{n}{k} p^k q^{n-k}$$

donde $q=1-p$

De la teoría de probabilidades sabemos (ref 3.2) que si $n/p \to \infty$, $n/L \to \lambda$ y p es pequeña, entonces la distribución binomial tiende a la de Poisson (ver tabla 2.1) con parámetro $D = np = nt_a/L =$

De la definición de λ se concluye que ésta representa al número medio de ocurrencias por unidad de tiempo, motivo por el cual se le suele llamar *intensidad del proceso*.

Se puede demostrar (ref 3.2) que si t_a y t_b son dos intervalos de tiempo que no se traslapan, entonces los eventos $\{k_a$ ocurrencias en $t_a\}$ y $\{k_b$ ocurrencias en $t_b\}$ son independientes, es decir, el número de ocurrencias del evento en t_a es independiente del número en t_b.

Sea $X(t)$ un proceso estocástico con el cual contamos el número de ocurrencias de un evento. Si hacemos $t_1=0$ y en ese momento iniciamos el conteo, entonces $t_a=t_2-t_1=t_2=t$ y $X(0)=0$; en tal caso la ec 3.34 queda en la forma

$$P[X(t)=k] = P[k \text{ en } t] = e^{-\lambda t}\,\frac{(\lambda t)^k}{k!} \tag{3.35}$$

y constituye la densidad de probabilidades del proceso de Poisson. Cada función muestra de este proceso tiene la forma de una escalera con escalones de altura unitaria localizados en los tiempos t_i en que ocurre cada vez el evento (fig 3.5).

Existen algunas variantes del proceso de Poisson que se estudian en las refs 3.8 y 3.12, de los cuales el más común es el que permite que el número de ocurrencias por unidad de tiempo, λ, no sea una constante, sino una función del tiempo, $\lambda(t)$. En este caso el proceso, $Y(t)$, se denomina *proceso generalizado de Poisson*, y se puede demostrar que su densidad de probabilidades es

$$P[Y(t)=k] = e^{-\mu}\,\frac{(\mu)^k}{k!} \tag{3.36}$$

...

en donde

$$\mu = \int_0^{t} \lambda(t)\ dt$$

Obsérvese que el proceso simple es un caso particular de éste con $\lambda(\tau)=\lambda$, por lo cual $\mu=\lambda\tau$. Tomando en cuenta que la suma algebraica de dos variables aleatorias con distribuciones de Poisson con parámetros respectivos ν y Ω también tiene ese tipo de distribución con parámetro $\nu+\Omega$, se tiene que la diferencia, $X(t_b)-X(t_a)$, del mismo proceso en los tiempos t_b y t_a, con $t_b>t_a$, también es un proceso de Poisson con parámetro $\lambda(t_b-t_a)$, o sea

$$P[X(t_b) - X(t_a) = k] = e^{-\lambda(t_b-t_a)}\frac{[\lambda(t_b-t_a)]^k}{k!} \qquad (3.37)$$

Esta diferencia representa el incremento que tuvo el proceso al pasar del tiempo t_a al t_b.

Puesto que la ec 3.37 es una distribución de Poisson, la media o esperanza y la variancia son

$$E[X(t_b) - X(t_a)] = \sigma^2[X(t_b) - X(t_a)] = \lambda(t_b-t_a) \qquad (3.38)$$

Considerando que la variancia de una variable aleatoria, Z, es

$$\sigma^2(Z) = E[Z^2]-E^2[Z]$$

se tiene que la función del valor medio cuadrático de la diferencia del proceso en cuestión es

$$E\{[X(t_b)-X(t_a)]^2\} = \lambda^2(t_b-t_a)^2+\lambda(t_b-t_a) \qquad (3.39)$$

De las ecs 3.38 y 3.39 es evidente que si $t_a=0$ y $t_b= t$, entonces

$$\eta(t) = E[X(t)] = \lambda t \qquad (3.40)$$

$$E[X^2(t)] = \lambda^2 t + \lambda t = \lambda t(1+\lambda) \qquad (3.41)$$

...

constituyen, respectivamente, la media y la función del valor medio cuadrático del proceso simple de Poisson, $X(t)$.

Para obtener la autocorrelación de $X(t)$, consideremos las diferencias del proceso correspondientes a los intervalos de tiempo que se traslapan, de t_a a t_b, y de t_c a t_d (ver fig 3.6). Para calcular la esperanza $E\{[X(t_b)-X(t_a)][X(t_d)-X(t_c)]\}$, dividamos cada intervalo en dos partes que no se traslapen, ya que de esta manera los incrementos del proceso en esos intervalos son independientes; así,

$$t_b-t_a=(t_c-t_a)+(t_b-t_c)$$

y

$$t_d-t_c=(t_b-t_c)+(t_d-t_b)$$

Por consiguiente,

$$X(t_b)-X(t_a)=[X(t_c)-X(t_a)]+[X(t_b)-X(t_c)]$$

y

$$X(t_d)-X(t_c)=[X(t_b)-X(t_c)]+[X(t_d)-X(t_b)]$$

Tomando en cuenta que la esperanza del producto de los miembros izquierdos de las dos últimas ecuaciones es igual a la esperanza de los productos de los miembros derechos, y que cada uno de los incrementos del proceso encerrados en paréntesis rectangulares es independiente de los demás, en cuyo caso la esperanza del producto es igual al producto de las esperanzas, se llega a que la autocorrelación del incremento del proceso en ambos intervalos es

$$R[X(t_b)-X(t_a),X(t_d)-X(t_c)]=E\{[X(t_b)-X(t_a)][X(t_d)-X(t_c)]\}$$
$$=\lambda^2(t_d-t_c)(t_b-t_a)+\lambda(t_b-t_c) \quad (3.42)$$

...

A partir de esta ecuación, haciendo $t_a=t_c=0$, $t_d=t_2$ y $t_b=t_1$ se obtiene la función de autocorrelación del proceso simple de Poisson, $X(t)$, la cual vale

$$R(t_1,t_2)=\lambda^2 t_1 t_2+\lambda t_1, \text{ si } t_2>t_1 \tag{3.43}$$

De acuerdo con 3.40 y 3.43, la autocovarianza de $X(t)$ vale (ec 3.20)

$$C(t_1,t_2)=\lambda^2 t_1 t_2+ \lambda t_1-(\lambda t_1)(\lambda t_2)=\lambda t_1 \tag{3.44}$$

3.7 *Proceso estocástico de renovación*

El proceso estocástico de renovación trata esencialmente el problema de reposición de objetos o componentes que fallan, tales como focos, máquinas de construcción o industriales, equipo para transporte, etc. Cuando un objeto falla se sustituye con otro; al fallar éste se repone de nuevo, y así sucesivamente. En lo que sigue calcularemos las medias y las variancias del número de reposiciones que se efectúan por unidad de tiempo.

Un *proceso de renovación* es una secuencia X_i de variables aleatorias X_i independientes e idénticamente distribuidas. Aquí X_i denota la duración del objeto introducido en el i-ésimo remplazo.

Si efectuamos la suma

$$S_n = X_1+X_2+\ldots+X_n$$

entonces S_n constituye el tiempo en el cual se efectúa el n-ésimo remplazo del objeto en cuestión.

...

Un problema que puede tratarse como un proceso de renovación, aunque no haya objetos físicos que renovar, es el de flujo de tráfico (ref 3.3): Sea una central telefónica o de correos, la cual recibe órdenes (llamadas) en los instantes aleatorios $\tau_1, \tau_2, \ldots$, en donde $0<\tau_1<\tau_2<\ldots$. Si es razonable suponer (como a menudo sucede) que los tiempos sucesivos de interarrivo de órdenes

$$X_1=\tau_1,\ X_2=\tau_2-\tau_1, \ldots, X_n=\tau_n-\tau_{n-1}$$

son variables aleatorias independientes y con igual densidad de probabilidades, entonces el proceso X_i es un proceso de renovación, ya que constituye los tiempos en que se "renueva" una orden o llamada. Para este caso se puede demostrar (ref 3.3) que si las X_i tienen distribución exponencial con parámetro ν, es decir, si

$$f(x_i) = \nu e^{-\nu x_i}$$

entonces el número de renovaciones hechas hasta el instante t es un proceso de Poisson simple, con media νt.

Consideraremos sólo el caso en que los tiempos a los cuales se asigna la reposición son discretos, por ejemplo, semanas, meses o años. Así, si las unidades de tiempo son semanas, las reposiciones que se hagan de lunes a domingo de la primer semana se anotan en la semana número 1, y así sucesivamente. Estudiaremos dos casos.

a) cuando los objetos que fallan se reponen con elementos nuevos

b) cuando se reemplazan con objetos usados.

3.7.1 *Reemplazo con elementos nuevos*

Consideraremos que la población original de objetos nuevos (en t=0) obtiene la primera renovación en $t=X_1$; los objetos ya renovados obtienen su segunda renovación en el tiempo $t=X_1+X_2$, etc. Sea p_n la probabilidad de que un objeto recién instalado falle después de n unidades de tiempo, y sea p_0=0. Supondremos que todos los objetos operan ininterrumpidamente, y una vez que alguno falla se repone de inmediato con otro nuevo, manteniendo el número total de objetos en uso.

Las reposiciones en cualquier instante son renovaciones ya sea de los objetivos originales o de otros que se repusieron previamente. Sea $g_{i+1}(n)$ el *número esperado* de repuestos de la (i+1)-ésima generación realizados en el tiempo n. La relación que éste guarda con el número de repuestos de la i-ésima generación, $g_i(.)$,

$$g_{i+1}(n)=\sum_{k=0}^{n} g_i(n-k)p_k \qquad (3.45)$$

ya que la renovación de (i+1)-éxima generación en el tiempo t está constituida por los reemplazos que se hagan antes del tiempo n de elementos de la i-ésima generación pero que fallan en el tiempo n; además, el número esperado de reemplazos de la generación i realizados en el tiempo $n-k$ (para $0\leq k\leq n$) es $g_i(n-k)p_k$, por lo que la i-ésima generación se obtiene sumando para toda $k\leq n$, dando como resultado la ec 3.45.

Sea el número esperado total de renovaciones, U_n, la suma de los números esperados de cada generación, es decir,

...

$$u_n = \sum_{i=1}^{n} g_i(n) \tag{3.46}$$

Escribiendo la ec 3.45 explícitamente para $i=1,2,3,\ldots$ y sumando los resultados se obtiene

$$g_2(n) = \sum_{k=0}^{n} g_1(n-k)p_k$$

$$g_3(n) = \sum_{k=0}^{n} g_2(n-k)p_k$$

$$\vdots$$

$$\sum_{i=2}^{n} g_i(n) = \sum_{i=1}^{n} \sum_{k=0}^{n} g_i(n-k)p_k$$

$$= \sum_{k=0}^{n} \Big[\sum_{i=1}^{n} g_i(n-k)p_k \Big] \tag{3.47}$$

Considerando la ec 3.46, la ec 3.47 queda en la forma

$$u_n - g_1(n) = \sum_{k=0}^{n} u_{n-k}p_k \tag{3.48}$$

donde $g_1(n)$ es el número esperado de reposiciones en el tiempo n de la población original.

3.7.2. *Reemplazo con objetos usados*

Si la población original no está constituida totalmente por objetos nuevos, sino que C_k de ellos ya han operado k unidades de tiempo, entonces el total de objetos, N, es

$$N = \sum_{k=0}^{\infty} C_k \tag{3.49}$$

Las probabilidades p_k que se han empleado hasta ahora corresponden a objetos nuevos. Para los objetos usados que pudiera tener la población original, la probabilidad de falla en el tiempo n de un objeto que ya había durado k unidades de tiempo en operación, antes de ser instalado, queda condicionada por la probabilidad, r_k, de que la duración sea mayor de k unidades, la cual es,

$$r_k = p_{k+1} + p_{k+2} + \ldots = 1 - \sum_{n=0}^{k} p_k \tag{3.50}$$

Puesto que la probabilidad condicional de que el objeto falle en $n+k$ unidades de tiempo, dado que debe durar más de k unidades es $p_{n+k} = p_n r_k$, entonces la probabilidad, p_n, de falla de un objeto en el tiempo n será

$$p_n = p_{n+k}/r_k \tag{3.51}$$

y el número esperado de objetos que requerirán renovación en el tiempo n y que ya habían operado k horas es $C_k p_{n+k}/r_k$. Sumando sobre todos los valores posibles de k obtenemos el número esperado de renovaciones de la población original, es decir,

$$g_1(n) = \sum_{k=0}^{\infty} C_k p_{n+k}/r_k \tag{3.52}$$

Sustituyendo la ec 3.48 obtenemos el número medio (esperado) total de renovaciones en el tiempo n, o sea,

$$U_n = \sum_{k=0}^{n} U_{n-k} p_k + \sum_{k=0}^{\infty} C_k p_{n+k}/r_k \tag{3.53}$$

Ejemplo 3.5

Supongamos que en un proceso de renovación discreto la probabilidad de que un objeto dure una unidad de tiempo es $p=0.75$, y la probabilidad que dure k unidades está dada por la distribución bonomial, siendo la duración máxima de cinco unidades de tiempo, es decir,

...

$$p_k = \frac{5!}{(5-k)!}\, p^k q^{5-k}$$

en donde $q=1-p$. Supongamos también que la población original tiene 100 objetos, de los cuales 10 son nuevos, 15 ya habían operado durante una unidad de tiempo, y 75 durante tres unidades de tiempo. Calcular el número esperado total de renovaciones después de seis unidades de tiempo.

Aplicando la distribución binomial para $k=1,\ldots,5$, obtenemos $p_1=0.01$, $p_2=0.09$, $p_3=0.26$, $p_4=0.40$ y $p_5=0.24$; además $p_0=0$, como se había indicado, y $p_k=0$ para $k>5$. Usando estos valores en la ec 3.50 obtenemos

$$\begin{aligned}
r_0 &= 0.01 + 0.09 + 0.26 + 0.40 + 0.24 = 1.00\\
r_1 &= 0.09 + 0.26 + 0.40 + 0.24 = 0.99\\
r_2 &= 0.26 + 0.40 + 0.24 = 0.90\\
r_3 &= 0.40 + 0.24 = 0.64\\
r_4 &= 0.24 = 0.24\\
r_5 &= = 0.00
\end{aligned}$$

De la información que se nos dio acerca de la población original deducimos que $C_0=10$, puesto que diez objetos eran nuevos, $C_1=15$, puesto que 15 objetos tenían una unidad de uso y, análogamente, $C_2=0$, $C_3=75$ y $C_4=C_5=0$. Aplicando la ec 3.52 se llega a

$$g_1(1) = 10 \times 0.01 + 15 \times 0.09/0.99 + 75 \times 0.40/0.64 = 48.34$$

De manera similar se obtiene, $g_1(2)=32.97$, $g_1(3)=8.66$, $g_1(4)=7.63$; $g_1(5)=2.4$ y $g_1(n)=0$ para $n>5$. Por consiguiente, mediante la ec 3.53 obtenemos $U_0=0$

...

$$u_1 = u_{1-0}p_0 + u_{1-1}p_1 + 48.34$$
$$= u_1 \times 0 + 0 \times 0.01 + 48.34 = 48.34$$

$$u_2 = u_{2-0}p_0 + u_{2-1}p_1 + u_{2-2}p_2 + 32.97$$
$$= 0 + 48.34 \times 0.01 + 0 + 32.97 = 33.45$$

$$u_3 = u_{3-0}p_0 + u_{3-1}p_1 + u_{3-2}p_2 + u_{3-3}p_3 + 8.66$$
$$= 0 + 33.45 \times 0.01 + 48.34 \times 0.09 + 0 + 8.66 = 13.35$$

$$u_4 = u_{4-0}p_0 + u_{4-1}p_1 + u_{4-2}p_2 + u_{4-3}p_3 + u_{4-4}p_4 + 7.63$$
$$= 0 + 13.35 \times 0.01 + 33.46 \times 0.09 + 48.34 \times 0.26 + 7.63 = 23.34$$

$$u_5 = u_{5-0}p_0 + u_{5-1}p_1 + u_{5-2}p_2 + u_{5-3}p_3 + u_{5-4}p_4 + u_{5-5}p_5 + 2.4$$
$$= 0 + 23.34 \times 0.01 + 13.35 \times 0.09 + 33.46 \times 0.26 + 48.34 \times 0.40$$
$$+ 2.4 = 31.87$$

$$u_6 = u_{6-0}p_0 + u_{6-1}p_1 + u_{6-2}p_2 + u_{6-3}p_3 + u_{6-4}p_4 + u_{6-5}p_5 + u_{6-6}p_6 + 0$$
$$= 0 + 31.87 \times 0.01 + 23.34 \times 0.09 + 13.35 \times 0.26 + 33.46 \times 0.40 + 48.34 \times 0.24$$
$$= 30.88$$

La suma de las renovaciones promedio después de seis unidades de tiempo es

$$\sum_{n=1}^{6} u_n = 181.24$$

Considerando que $\Sigma p_k = 1$, es posible demostrar (ref 3.7) que cuando n tiene a infinito u_n tiende a $N/(\sum_{k=0}^{\infty} kp_k)$, es decir, que conforme n crece el número esperado de reposiciones en el tiempo n tiende a ser igual al número de objetos en la población original dividido entre el promedio de la duración de cada objeto, lo cual, intuitivamente, es razonable.

...

El lector interesado en profundicar más en este proceso de renovación, puede acudir a las refs 3.3 y 3.7.

3.8 *Confiabilidad*

Uno de los atributos más importantes que debe poseer un sistema es una *confiabilidad* adecuada. Antes de establecer la confiabilidad de un sistema, es necesario analizar si los costos y tiempos de producción, las condiciones de operación y las políticas de mantenimiento que deben observarse durante un periodo de operación considerando, justifican el nivel de confiabilidad deseado. Lo anterior se debe a que, en general, un incremento en la confiabilidad lleva aparejado un crecimiento en los costos de producción y de operación de los sistemas.

La *confiabilidad* de un sistema se define como la *probabilidad* de que éste realice satisfactoriamente ciertas funciones específicas durante un lapso prescrito, bajo ciertas condiciones ambientales.

De la definición anterior se desprende que es necesario conocer:

a) El tiempo de operación (vida útil)

b) El medio ambiente de operación (temperatura, humedad, fuerzas que deberá soportar, etc).

c) ¿Qué es lo que constituye una operación satisfactoria o cuándo se considera que un sistema ha fallado? (por ejemplo, ¿qué nivel de daño se considera como "falla" en un edificio que se ve sujeto a un sismo: colapso total, agrietamiento de muros o agrietamiento de los recubrimientos?).

...

Para calcular la probabilidad de que un sistema no falle (su confiabilidad) es necesario conocer las densidades de probabilidades o la densidad de probabilidades conjunta de las variables aleatorias que se considere podrían influir en la sobrevivencia o falla del sistema. Así, para obtener la confiabilidad de un edificio que se vea sujeto a sismos, es necesario conocer las densidades de probabilidades de su resistencia y de las fuerzas dinámicas que provocarían los temblores que pudieran actuar sobre el edificio. En este ejemplo, la aleatoriedad de la resistencia se debe a variaciones al azar en las resistencias de los materiales de construcción, en las dimensiones de las vigas, losas y columnas, etc; la aleatoriedad de las fuerzas sísmicas se debe a la ocurrencia al azar de los temblores en sitios localizados a diferentes distancias del edificio y con diferentes magnitudes, a las características del subsuelo, etc.

Tanto en el ejemplo del sistema estructural que acabamos de describir, como en los sistemas electrónicos, industriales, etc., son varios los factores que pueden influir en las densidades de probabilidades de las variables aleatorias que intervengan en el problema. Estos factores se pueden dividir en tres grupos, a saber, *factores iniciales*, *factores pre-operacionales y factores de operación*, como puede verse en la fig 3.7.

Si $X(t)$ es un proceso estocástico que define el comportamiento o respuesta del sistema, entonces la densidad de probabilidades de primer orden, obtenida con base en la información, Y, disponible al respecto (ya sea experimental o subje-

...

tiva), será $f|x(t)|y|$. Si se divide el eje del tiempo en unidades discretas (como se hizo en el camino casual), el proceso $X(t)$ será de parámetro discreto; en caso contrario será de parámetro continuo.

Por simplicidad, para el caso discreto escribiremos

$$f[x(t_n)|y] = f[x(t_n)]$$

En caso de que el criterio de falla del sistema sea "ocurre la falla si la respuesta del sistema excede al valor A", entonces la probabilidad de que esto ocurra en el tiempo t_n, si lo único que sabemos es que hasta el tiempo inmediato anterior, t_{n-1}, no ha fallado, será

$$P[X(t_n)>A|X(t_{n-1})<A] = \int_A^\infty f[x(t_n)]dx \qquad (3.54)$$

si el proceso es de parámetro discreto. En caso de que el proceso sea estacionario, su densidad de probabilidades de primer orden no dependerá de t, en cuyo caso se usará en la ecuación anterior $f(x)$ en vez de $f[x(t_n)]$.

Un métodomás usual para determinar la confiabilidad de un sistema consiste en obtener, experimental o subjetivamente, la densidad de probabilidades del *tiempo de falla* del sistema. En forma experimental esto se puede lograr observando varios sistemas idénticos, expuestos a las mismas condiciones ambientales, y anotar los tiempos en que falla cada uno; luego, mediante procedimientos estadísticos, se les ajusta alguna densidad de probabilidades a los tiempos de falla. En forma subjetiva, la densidad se puede asignar con base en el conocimiento de otros sistemas similares, con base en algún modelo matemático (como en la ref 3.10), etc.

En muchas ocasiones, como se verá más adelante, la densidad de probabilidades del tiempo de falla del sistema se deduce de las densidades de sus respectivos componentes que, en general, son más fáciles de determinar.

3.8.1 *Fuerza de mortandad*

En estudios realizados sobre la confiabilidad de los componentes de un sistema se ha encontrado que está caracterizada por tres etapas. La primera es un lapso inicial de *alta mortandad o alta intensidad de falla* (el número medio de fallas por unidad de tiempo es grande), la cual se debe principalmente a un control de calidad deficiente durante la elaboración del componente. Este periodo inicial está limitado por el tiempo T_c indicado en la fig 3.8. Lo que es costumbre hacer en sistemas muy importantes (como los de defensa militar de un país), para eliminar los componentes defectuosos, es usarlos durante un tiempo T_c en algún dispositivo ajeno al sistema, y luego instalar en dicho sistema los que no hayan fallado. La segunda etapa es un periodo de *intensidad de falla constante*, en la que las fallas ocurren al azar y con menor frecuencia que en las etapas 1 y 3; el límite superior de esta parte está indicado con T_D en la fig 3.8. En T_D se inicia la tercera etapa que es de *alta mortandad o alta intensidad de falla*, lo cual equivale a fallas muy frecuentes debidas al deterioro de los componentes. Cuando en un sistema son indeseables este tipo de fallas, estas se pueden reducir siguiendo políticas de mantenimiento preventivo. Estas políticas incluyen pruebas pe-

...

riódicas e inspección de ciertas componentes, y su reemplazo cuando las pruebas indiquen que éstos están próximas a fallar.

La región de intensidad de falla constante corresponde al lapso de operación normal del componente respectivo, en el cual las fallas ocurren al azar y son *independientes*, por lo cual puede considerarse razonablemente que éstas pueden representarse mediante un proceso de Poisson con intensidad λ (el valor de λ varía de un tipo de componente a otro, y se determina experimentalmente). Con esto en mente, la probabilidad de que no ocurra ninguna falla en el periodo de 0 a t se obtiene mediante la ec 3.35, sustituyendo en ella a k=0. El resultado es

$$P[X(t)=0] = e^{-\lambda t}\,\frac{(\lambda t)^0}{0!} = e^{-\lambda t} = P[T\leq t] \tag{3.55}$$

en donde T es la variable aleatoria *tiempo de falla*.

Si analizamos la densidad de probabilidades *exponencial*

$$f(t) = \lambda e^{-\lambda t};\ 0\leq t<\infty \tag{3.56}$$

vemos que su función de distribución (o distribución de probabilidades acumuladas) es

$$P[T\leq t] = F(t) = \int_0^t \lambda e^{-\lambda t} dt = 1-e^{-\lambda t}$$

por lo cual la probabilidad de que la variable T sea mayor que t es

$$P[T>t] = 1-F(t) = e^{-\lambda t} \tag{3.57}$$

Comparando las ecs 3.55 y 3.57 vemos que éstas tienen como factor común a $e^{-\lambda t}$, por lo que deducimos que la variable aleatoria, T, "tiempo de falla", tiene densidad de probabilidades

...

exponencial, y que la probabilidad de que no falle el componente queda dada por la ec 3.57. Es interesante mencionar que experimentalmente se ha confirmado que este tipo de distribución se ha ajustado razonablemente bien a los tiempos de falla de diversos componentes de circuitos electrónicos, tales como transistores, bulbos, etc.

La *fuerza de mortandad o intensidad de falla*, $\beta(\tau)$, de un sistema se define como sigue: $\beta(\tau)d(\tau)$ es la probabilidad de que un sistema falle en el intervalo de τ a $\tau+d\tau$, suponiendo que no ha fallado hasta el tiempo τ.

Supongamos que un sistema ha operado durante un tiempo τ sin ninguna falla. Si la variable aleatoria T es el tiempo de falla del sistema, y $f(t)$ y $F(t)$ son la densidad de probabilidades y la función de distribución de T, respectivamente, entonces la función de distribución de T será

$$F(t|t>\tau) = \frac{P[T< t,\ T>\tau]}{P[T>\tau]};\ t\geq\tau \qquad (3.58)$$

en donde el miembro izquierdo representa la probabilidad de que el sistema sobreviva hasta t dado que sobrebibió hasta τ, y el numerador del miembro derecho es la probabilidad de que simultáneamente ocurran los eventos $\{T<t\}$ y $\{T>\tau\}$, la cual es igual a $F(t)-F(\tau)$. Además, es evidente que el denominador de la ec 3.58 es $P[T>\tau]=1-F(\tau)$. Por consiguiente, la ec 3.58 queda en la forma

$$F(t|t>\tau) = \frac{F(t)-F(\tau)}{1-F(\tau)};\ t\geq\tau \qquad (3.59)$$

La densidad de probabilidades correspondiente se obtiene

...

derivando la ec 3.59 con respecto a t; el resultado es

$$f(t|t>\tau) = \frac{f(t)}{1-F(\tau)} \quad \text{si} \quad t \geq \tau \tag{3.60}$$

$$0 \quad \text{si} \quad t<\tau$$

De la definición de fuerza de mortandad vemos que

$$\beta(\tau)\,d\tau = f(\tau|t>\tau)\,d\tau$$

por lo que

$$\beta(\tau) = \frac{f(\tau)}{1-F(\tau)} = \frac{\frac{dF(\tau)}{d\tau}}{1-F(\tau)} \tag{3.61}$$

Mediante la ec 3.61 se puede obtener la fuerza de mortandad cuando se conoce la densidad de probabilidades de falla del sistema. Por el contrario, si lo que se conoce es $\beta(\tau)$, la densidad de probabilidades se puede obtener resolviendo la ecuación diferencial de la ec 3.61. Integrando ambos miembros de dicha ecuación se obtiene:

$$\int_0^{\tau} \beta(t)\,dt = -\ell n\left[1-F(\tau)\right]$$

(la constante de integración es cero porque $F(0)=0$, debido a que el sistema empezó a operar en $\tau=0$). En la ecuación anterior ℓn denota logaritmo natural (con base e). El antilogaritmo de la ecuación anterior resulta ser

$$F(t) = 1 - e^{-\int_0^{\tau}\beta(t)\,dt} \tag{3.62}$$

o, derivado respecto a τ,

$$f(\tau) = \beta(\tau)\,e^{-\int_0^{\tau}\beta(t)\,dt} \tag{3.63}$$

...

Ejemplo

Sea un sistema en el que la fuerza de mortandad del sistema es la constante λ. Aplicando la ec 3.63, la densidad de probabilidades correspondiente resulta ser

$$f(\tau) = \lambda e^{-\int_0^\tau \lambda dt} = \lambda e^{-\lambda\tau} ; \tau \geq 0$$

que es la distribución exponencial. En conclusión, si la densidad de probabilidades de falla de un sistema es exponencial, entonces su fuerza de mortandad es constante e igual al parámetro λ de la misma.

Ejemplo

Si un sistema tiene una densidad de probabilidades gama, es decir, si

$$f(t) = \begin{cases} c^2 t e^{-ct} & \text{si} \quad t > 0 \\ 0 & \text{si} \quad t < 0 \end{cases}$$

donde c es una constante (el parámetro de la distribución) entonces

$$F(t) = 1 - cte^{-ct} - e^{-ct} \qquad \text{si} \quad t > 0$$

La fuerza de mortandad respectiva se obtiene mediante la ec 3.61, lo cual resulta en

$$\beta(\tau) = \frac{c^2 \tau e^{-c\tau}}{c\tau e^{-c\tau} + e^{-c\tau}} = \frac{c^2\tau}{1 + c\tau}$$

Ejemplo

Otra función de intensidad de falla que se ha empleado en algunos estudios de confiabilidad (ref 3.11) tiene como ecuación a

...

$$\beta(\tau) = \alpha\delta e^{\delta-1}$$

donde α y δ son constantes positivas. Obsérvese que si $\delta>1$ entonces $\beta(\tau)$ crece con τ; si $\delta<1$, decrece, y si $\delta=1$, entonces $\beta(\tau)=\lambda$; este último coincide con el ya estudiado de la distribución de Poisson. Usando esta expresión para $\beta(\tau)$ en la ec 3.63 se obtiene

$$f(\tau) = \alpha\delta\tau^{\lambda-1}e^{-\alpha\tau^{\delta}}$$

que es la distribución de Weibull.

3.8.2. *Pruebas de duración*

Para determinar la densidad de probabilidades de la duración de un componente de un sistema, es necesario extraer una muestra aleatoria de *n* componentes idénticas, probarlas poniéndolas a funcionar en las condiciones ambientales con que lo harán en el sistema y anotando los tiempos de duración de cada componente. En muchas ocasiones la duración de los componentes es larga, por lo que es necesario diseñar pruebas aceleradas, de tal manera que las condiciones ambientales sean más severas que las normales y se ocasionen las fallas más rápidamente. Este tipo de pruebas se pueden usar también para comparar dos tipos diferentes de componentes, para que rápidamente se pueda inferir cuál es más confiable. Es evidente que los resultados de las pruebas aceleradas deben estar correlacionadas con los de las pruebas normales para poder inferir las duraciones en operación normal a partir de las correspondientes a las pruebas aceleradas (estas correlaciones deben determinarse durante la etapa del diseño de la prueba acelerada).

Si existe evidencia experimental o subjetiva suficiente que favorez a la *distribución exponencial* con parámetro λ, como densidad de probabilidades de los tiempos de falla de los componentes, entonces

$$f(t) = \lambda e^{-\lambda t} \; ; \; t>0, \; \lambda>0$$

y el tiempo medio de falla será

$$E|T| = \mu = 1/\lambda \tag{3.64}$$

Supongamos que se ponen a prueba n componentes y que la prueba se da por terminada cuando hayan fallado r de ellos ($r \leq n$), y que los tiempos de falla son $t_1 \leq t_2 \leq \ldots \leq t_r$. Nuestro interés radica en la estimación estadística, $\hat{\mu}$, del tiempo medio de falla, μ.

Usando la teoría desarrollada en la ref 3.11, se puede demostrar que el estimador insesgado del tiempo medio de vida de los componentes es

$$\hat{\mu} = T_r/r \tag{3.65}$$

donde T_r es la duración acumulada de los componentes hasta el tiempo t_r, es decir

$$T_r = \sum_{i=1}^{r} t_i + (n-r)t_r \tag{3.66}$$

para el caso en que los elementos que han fallado *no se reemplazan* por uno nuevo, o

$$T_r = nt_r \tag{3.67}$$

si se han reemplazado. Observe que si la prueba es sin reemplazo, de las ecs 3.65 y 3.66, se obtiene que $\hat{\mu}$ es el promedio de los tiempos observados de falla.

Parà hacer inferencias acerca de μ, usamos el hecho de que $2T_r/\mu$ es una variable que tiene distribución ji-cuadrada con $2r$ grados de libertad (ref 3.11), independientementede que la prueba haya sido con o sin reemplazo. Por consiguiente, el intervalo de confianza para μ correspondiente a un nivel de significancia, α(el nivel de confianza es $1-\alpha$) es

$$\frac{2T_r}{\chi^2_{\alpha/2}} \leq \mu \leq \frac{2T_r}{\chi^2_{1-\alpha/2}} \tag{3.68}$$

en donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son las abscisas de la distribución ji-cuadrada con $2r$ grados de libertad, para los cuales queda un área bajo la curva igual a $\alpha/2$ a su derecha o a su izquierda, respectivamente (fig 3.9).

Para probar hipótesis acerca de μ usamos la misma distribución χ^2 con $2r$ grados de libertad. Si la hipótesis nula (por probar) es que $\mu=\mu_0$(μ_0 es un valor específico), y la hipótesis alternativa es $\mu>\mu_0$, aceptaremos la hipótesis nula con un nivel de confianza $1-\alpha$ si

$$\frac{2T_r}{\mu_0} < \chi^2_{\alpha} \tag{3.69}$$

y en caso contrario la rechazaremos. Si las hipótesis alternativas son $\mu<\mu_0$ o $\mu\neq\mu_0$, aceptaremos la hipótesis nula si

$$\frac{2T_r}{\mu_0} > \chi^2_{1-\alpha/2} \tag{3.70}$$

o si

$$\chi^2_{1-\alpha/2} \leq \frac{2T_r}{\mu_0} \leq \chi^2_{\alpha/2} \tag{3.71}$$

respectivamente.

...

Otra opción que se puede tener al realizar la prueba de duración es la de interrumpir la prueba al concluir un tiempo acumulado de fallas fijo, T, predeterminado, y considerar los k componentes que han fallado hasta ese instante como el valor observado de una variable aleatoria. Ya sea que la prueba se realice con o sin reemplazo, un intervalo de confianza aproximado para el tiempo medio de vida de los componentes es

$$\frac{2T}{\chi^2_{1-\alpha/2}} \leq \mu \leq \frac{2T}{\chi^2_{\alpha/2}} \qquad (3.72)$$

donde $\chi^2_{1-\alpha/2}$ es la abscisa de la distribución ji-cuadrada con $2k + 2$ grados de libertad, para la cual queda un área bajo la curva igual a $\alpha/2$ a su derecha, y $\chi^2_{\alpha/2}$ es la abscisa de la distribución ji-cuadrada con $2k$ grados de libertad, para la cual queda un área bajo la curva igual a $\alpha/2$ a su izquierda. Las pruebas de hipótesis son semejantes a las del caso anterior, pero ahora usando una distribución ji-cuadrada con $2k$ grados de libertad.

Ejemplo

Supongamos que se han puesto a prueba de duración 50 componentes de un sistema, y que ésta se termina al fallar 10 elementos. Si los tiempos de falla resultan ser 65,110,380,420, 505,580,650,840,910 y 950 horas, entonces n=50, r=10 y de la ec 3.66,

$$T_{10} = (65 + 110 +\ldots+ 950) + (50 - 10)\ 950 = 43,410 \text{ hrs}$$

La estimación del tiempo medio de sobrevivencia de los componentes es (ec 3.65)

...

$$\hat{\mu} = \frac{43{,}410}{10} = 4{,}341 \text{ hs}$$

Por consiguiente, de la ec 3.64 se concluye que la intensidad de falla o mortandad de estos componentes es

$$\lambda = 1/\hat{\mu} = 0.00023 \text{ fallas/hr}$$

o 23 fallas cada cien mil horas. El intervalo de confianza del 90% ($\alpha = 1 - 0.9 = 0.10$) para μ es

$$\frac{2x43{,}410}{\chi^2_{0.05}} \leq \mu \leq \frac{2x43{,}410}{\chi^2_{0.95}}$$

en donde la densidad ji-cuadrada tiene 2 x 10= 20 grados de libertad. De las tablas de esta distribución (en la ref 3.11, por ejemplo) se obtienen $\chi^2_{0.05}=31{,}410$ y $\chi^2_{0.95}=10{,}851$, por lo cual resulta

$$2{,}764 \leq \mu \leq 8{,}001$$

Esto significa que con un 90% de probabilidad, el verdadero valor de μ cae entre 2,764 y 8,001 hs.

Supongamos ahora que deseamos probar la hipótesis nula de qu $\lambda=0.00040$, o sea $\mu_0=2{,}500$ hs, contra la hipótesis alternativa de que $\mu>2{,}500$ hrs, con un nivel de confianza del 95% ($\alpha=1-0.95=0.05$). En este caso $\chi^2_{0.05}=31{,}410$ y $2T_r/\mu_0=2x43{,}410/2{,}500=34.728$, que resulta ser mayor que 31.410, por lo cual rechazamos la hipótesis nula con un nivel de significancia de 5%; esto equivale a decir que el tiempo medio de vida excede de 2,500 hs con un 95% de nivel de confianza.

Para estimar experimentalmente la duración media y la variancia de la duración de componentes que no tienen intensidad

...

de falla constante, como la que corresponde a la distribución de Weibull ya descrita, el lector puede acudir a la ref 3.11, pág 376.

Ejemplo

Supongamos que la duración de un componente de un sistema tiene distribución exponencial con parámetro λ, es decir,

$$f(t) = \lambda e^{-\lambda t}$$

Si al fallar un componente se reemplaza de inmediato por otro nuevo, calculemos la densidad de probabilidades de la duración total.

Sean T_1 y T_2 las variables aleatorias que representan a los tiempos de falla del primero y el segundo componente, respectivamente. Puesto que ambos componentes son idénticos, tienen la misma distribución exponencial. Nos interesa la densidad de probabilidades de $T = T_1+T_2$ la cual puede obtenerse mediante la ec 3.7, ya que T_1 y T_2 son independientes (se supone que el funcionamiento del sistema es el mismo con ambos componentes). En estas circunstancias

$$\begin{aligned} f_T(t) &= \int_0^\infty f_{T_1}(t_1) f_{T_2}(t-t_1)\,dt_1 \\ &= \int_0^\infty \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t-t_1)}\,dt_1 = \lambda^2 e^{-\lambda t} \int_0^t dt_1 \\ &= \lambda^2 t e^{-\lambda t} \end{aligned}$$

Aplicando repetidamente este procedimiento podemos calcular la densidad de probabilidades de la duración de n componentes que se usan sucesivamente, reemplazando cada una a la que ha fallado previamente.

...

3.9 *Sistemas en serie y en paralelo*

Puesto que la confiabilidad se ha definido como una probabilidad, ésta se podrá calcular, para un sistema cualquiera, si se conocen las densidades de probabilidades de falla de cada uno de sus componentes. Estas densidades se pueden obtener mediante experimentos diseñados exprofeso o mediante consideraciones de carácter subjetivo basadas en experiencias previas con componentes semejantes, o en la experiencia del que estudia la confiabilidad del sistema.

Muchos sistemas pueden considerarse con los componentes *en serie o en paralelo*. Se dice que un sistema es *en serie* si sus componentes están conectados entre sí de tal manera que al fallar uno de ellos falla el sistema; en la fig 3.10 se muestra la representación clásica de un sistema de este tipo. Un sistema es en *paralelo* si para que falle éste se necesita que fallen todos sus componentes; en la fig 3.11 se encuentra la representación gráfica de un sistema de este tipo.

Para estimar la confiabilidad de un sistema en serie consideraremos que los componentes del mismo son *independientes*, es decir, que el hecho de que uno falle no influye en la probabilidad de que cualquier otro falle. En otras palabras, la confiabilidad del componente se mantiene inalterada cuando cualquier otro falla.

Puesto que para que un sistema en serie no falle se requiere que ninguno de sus componentes falle, su confiabilidad será igual al producto de las confiabilidades de cada uno de sus componentes (esto se debe a que el evento "no falla el sistema"

...

es la intersección de los eventos "no falla el componente i" en donde $i=1,2,\ldots,n$, y n es el total de componentes).
En símbolos, la probabilidad de que no falle el sistema antes del tiempo t es

$$P|T \geq t| = R(t) = R_1(t) \times R_2(t) \times \ldots \times R_n(t) = \prod_{i=1}^{n} R_i(t) \qquad (3.73)$$

en donde T es la variable aleatoria "tiempo de falla del sistema", t es un valor que puede asumir T, $R(t)$ es la probabilidad de que no falle el sistema hasta el tiempo t (su confiabilidad hasta t), y $R_i(t)$ es la probabilidad de que la componente i no falle antes de t. De la ec 3.73 se concluye que la confiabilidad de un sistema en serie decrece conforme aumenta el número de sus componentes, ya que se están multiplicando entre sí números menores de uno. Por ejemplo, si $n=4$ y $R_i(t)=0.9$ para toda i (los componentes son idénticos), entonces $R(t)=0.9 \times 0.9 \times 0.9 \times 0.9 = 0.6561$; si $n=5$, $R(t)=0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.59049$

Para que un sistema en paralelo falle es necesario que fallen *todos* sus componentes. Si dichos componentes son independientes, la probabilidad de falla del sistema en algún instante previo a t será

$$P|T \leq t| = F(t) = 1 - R(t) = \prod_{i=1}^{n} |1 - R_i(t)| \qquad (3.74)$$

por lo que la confiabilidad del sistema será $1-P|T \leq t|$, es decir

$$R(t) = 1 - \prod_{i=1}^{n} |1 - R_i(t)| \qquad (3.75)$$

Puesto que todas las probabilidades de falla que aparecen en el miembro derecho de la ec 3.74 son menores que uno, el resultado de aplicarla decrecerá conforme aumenta el número n de componentes, es decir, la probabilidad de supervivencia

...

de un sistema en paralelo aumenta conforme crece el número de sus componentes y, por consiguiente, su confiabilidad (ec 3.75) aumenta.

Por ejemplo, si un sistema en paralelo tiene cuatro componentes (n=4) y si $R_i(t)$=0.9, entonces su probabilidad de falla antes del tiempo t es (ec 3.74)

$$P|T\leq t| = F(t) = 0.1 \times 0.1 \times 0.1 \times 0.1 = 0.0001$$

por lo que su confiabilidad (probabilidad de sobrevivencia) es

$$R(t) = 1 - 0.0001 = 0.9999$$

El hecho de que la confiabilidad de un sistema en paralelo es mayor que la de uno en serie, en igualdad del número de componentes y de sus confiabilidades, hace concluir que una manera de aumentar la confiabilidad de un sistema en serie consiste en ponerle algunos componentes en paralelo a aquellos que tengan baja confiabilidad, con lo cual se forma un sistema *mixto*, como el de la fig 3.12. A los componentes que se agregan con este objeto se les llama *redundantes*, porque no son indispensables para que funcione el sistema. Sin embargo, al añadirle componentes redundantes a un sistema se incrementan su costo, volumen, complejidad, etc., lo que en ocasiones desalienta la utilización de este recurso.

Para calcular la confiabilidad de un sistema mixto primero hay que obtener las confiabilidades de los grupos de componentes que están en paralelo, y luego considerar a dicho grupo como si fuese un elemento conectado en serie con una confiabilidad igual a la del grupo en paralelo. Así, en el caso presentado

...

en la fig 3.12, en que la confiabilidad de cada componente hasta el instante t está anotada abajo de él, el primer grupo de elementos en paralelo tiene una confiabilidad igual a $R_1(t)=$ $1-0.3 \times 0.3 \times 0.3=0.973$; la del segundo grupo es $R_2(t)= 1-$ $0.2 \times 0.2=0.96$ (ver fig 3.13). La confiabilidad del sistema es, entonces,

$$R(t) = 0.99 \times 0.95 \times 0.973 \times 0.96 \times 0.90 = 0.7815$$

Si no hubiese habido componentes redundantes, la confiabilidad hubiera sido

$$R(t) = 0.99 \times 0.95 \times 0.70 \times 0.80 \times 0.90 = 0.4740$$

que es bastante menor que la del sistema que sí los tiene.

3.10 *El modelo exponencial en la confiabilidad de un sistema*

En esta sección emplearemos los resultados obtenidos para calcular las confiabilidades de sistemas en serie y en paralelo, suponiendo que las densidades de probabilidades, $f(t)$, de los tiempos de falla de los componentes son exponenciales, es decir,

$$f_i(t) = \lambda_i e^{-\lambda_i t}$$

en donde λ_i es la intensidad de fallas (número medio de fallas por unidad de tiempo) del i-ésimo componente.

Tomando en cuenta que

$$R(t) = 1-F(t) = 1-\int_0^t f(t)dt$$

obtenemos

$$R_i(t) = 1-\int_0^t \lambda_i e^{-\lambda_i t}dt=e^{-\lambda_i t} \qquad (3.76)$$

...

Si el sistema es en serie, su confiabilidad, de acuerdo con la ec 3.73, será

$$R(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-(\sum_{i=1}^{n} \lambda_i)t} = e^{-\theta t} \tag{3.77}$$

en donde

$$\theta = \sum_{i=1}^{n} \lambda_i \tag{3.78}$$

Puesto que el miembro derecho de la ec 3.77 tiene la misma forma que el de la ec 3.76, deducimos que la densidad de probabilidades del sistema en serie también es exponencial con parámetro θ, es decir, el número medio de fallas del sistema por unidad de tiempo queda dado por la ec 3.78. Además, puesto que el tiempo medio de falla de cada componente es (sec 3.8.2)

$$E_i(t) = \mu_i = 1/\lambda_i$$

tenemos que el tiempo medio falla del sistema, cuando cada componente que falla se reemplaza de inmediato con otro idéntico, es

$$E|T| = \mu = \frac{1}{\theta} = \frac{1}{\sum_{i=1}^{n} \lambda_i} = \frac{1}{\lambda_1+\lambda_2+\ldots+\lambda_n} = \frac{1}{\frac{1}{\mu_1}+\frac{1}{\mu_2}+\ldots+\frac{1}{\mu_n}} \tag{3.79}$$

Para el caso de un sistema en paralelo, si las densidades de probabilidades de falla de las componentes son exponenciales, la confiabilidad es

$$R(t) = 1 - \prod_{i=1}^{n} (1-e^{\lambda_i t}) \tag{3.80}$$

Esta probabilidad no se puede factorizar en tal forma que tenga la apariencia de la ec 3.76, como sucedió con el

...

sistema en serie y, por consiguiente, la distribución de la confiabilidad de un sistema en paralelo no es exponencial. En estas condiciones, la intensidad de falla (fuerza de mortandad) del sistema se tendrá que obtener mediante la ec 3.61, y no resultará ser una constante.

El tiempo medio de falla también es difícil de obtener para el caso general en que las λ_i son diferentes. Si todas las λ_i son iguales a λ, entonces la ec 3.80 resulta en

$$R(t) = 1 - (1 - e^{-\lambda t})^n \qquad (3.81)$$

El desarrollo del bonimio del miembro derecho de la ec 3.81 conduce a

$$R(t) = \binom{n}{1} e^{-\lambda t} - \binom{n}{2} e^{-2\lambda t} + \ldots + (-1)^{n-1} e^{-n\lambda t} \qquad (3.82)$$

en donde $\binom{n}{i}$ denota al número de combinaciones que se pueden formar con n objetos tomando de i en i. Puesto que la densidad de probabilidades es

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}|1 - R(t)| = -\frac{dR(t)}{dt}$$

obtenemos que la densidad de probabilidades del tiempo de falla del sistema en paralelo es

$$f(t) = \lambda\binom{n}{1} e^{-\lambda t} - 2\lambda\binom{n}{2} e^{-2\lambda t} + \ldots + (-1)^{n-1} n\lambda e^{-n\lambda t} \qquad (3.83)$$

La media del tiempo de falla es, entonces

$$E(t) = \mu = \int_0^\infty t f(t)\,dt = \lambda\binom{n}{1}\int_0^\infty t e^{-\lambda t} dt - 2\lambda\binom{n}{2} t e^{-2\lambda t} dt + \ldots$$

$$+ (-1)^{n-1} n\lambda \int^\infty t e^{-n\lambda t} dt$$

$$= \frac{1}{\lambda}\binom{n}{1} - \frac{1}{2\lambda}\binom{n}{2} + \ldots + (-1)^{n-1}\frac{1}{n\lambda}$$

...

Por inducción matemática se puede demostrar que esta ecuación es equivalente a

$$\mu = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right) \tag{3.84}$$

por lo que la fuerza de mortandad resulta ser

$$\theta = \frac{1}{\mu} = \frac{\lambda}{1 + \frac{1}{2} + \ldots + \frac{1}{n}}$$

TABLA 2.1

ALGUNAS DISTRIBUCIONES DE PROBABILIDADES

	Distribución	Densidad de probabilidades, $f_X(x)$	Esperanza	Varianza
DISCRETAS	Binomial ($n=1,2,\ldots;0\leq p\leq 1$; $q=1-p$)	$\binom{n}{x}p^x q^{n-x}$; $x = 0,1,\ldots,n$ 0 ; $x < 0$	np	npq
	Poisson ($\lambda > 0$)	$\frac{e^{-\lambda}\lambda^x}{x!}$; $x = 0,1,\ldots$ 0 ; $x < 0$	λ	λ
	Geométrica ($0\leq p\leq 1$: $q=1-p$)	pq^{x-1} ; $x = 0,1,\ldots$ 0 ; $x < 0$	$1/p$	q/p^2
	Binomial negativa ($r=1,2,\ldots;0\leq p\leq 1$; $q=1-p$)	$\binom{r+x-1}{x}p^r q^x$; $x = 0,1,\ldots$ 0 ; $x < 0$	rq/p	rq/p^2
CONTINUAS	Normal ($-\infty<\eta<\infty$; $0\leq\sigma^2<\infty$)	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\eta)^2/2\sigma^2}$; $-\infty<x<\infty$	η	σ^2
	Exponencial ($\lambda>0$)	$\lambda e^{-\lambda x}$; $x \geq 0$ 0 ; $x < 0$	$1/\lambda$	$1/\lambda^2$
	Uniforme en el intervalo de a a b	$\frac{1}{(b-a)}$; $a\leq x\leq b$ 0 ; $a>x$ o $b<x$	$(a+b)/2$	$(b-a)^2/12$
	Ji-cuadrada (χ^2) con n grados de libertad	$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}x^{(n/2-1)}e^{-x/2}$	n	$2n$

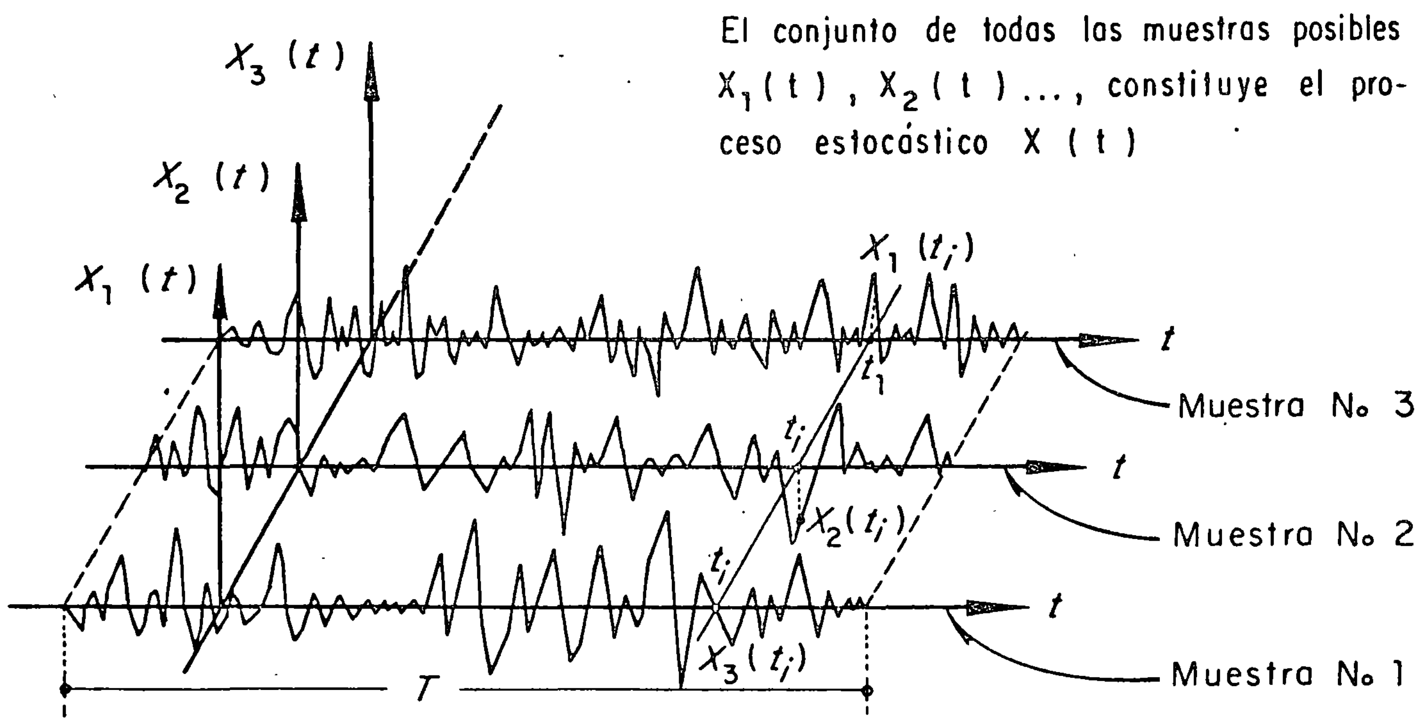


Fig 3.1.a Proceso estocástico de parámetro continuo

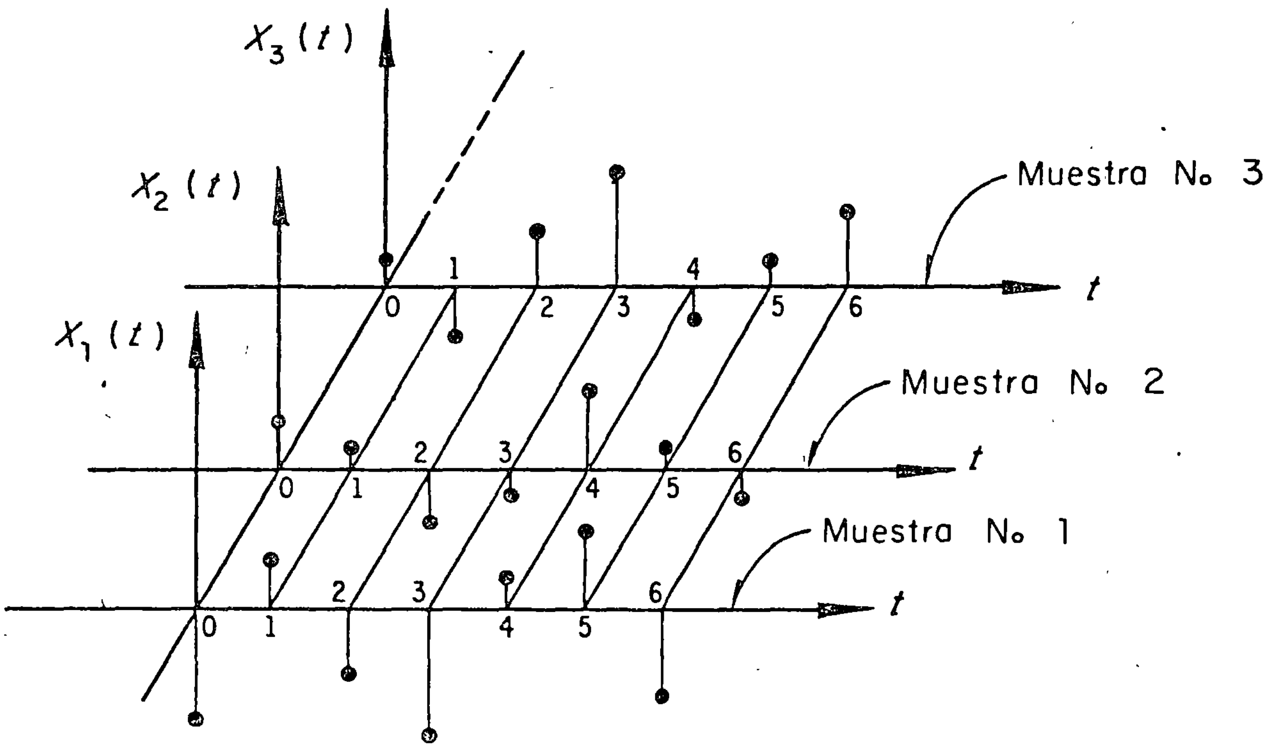


Fig 3.1.b Proceso estocástico de parámetro discreto

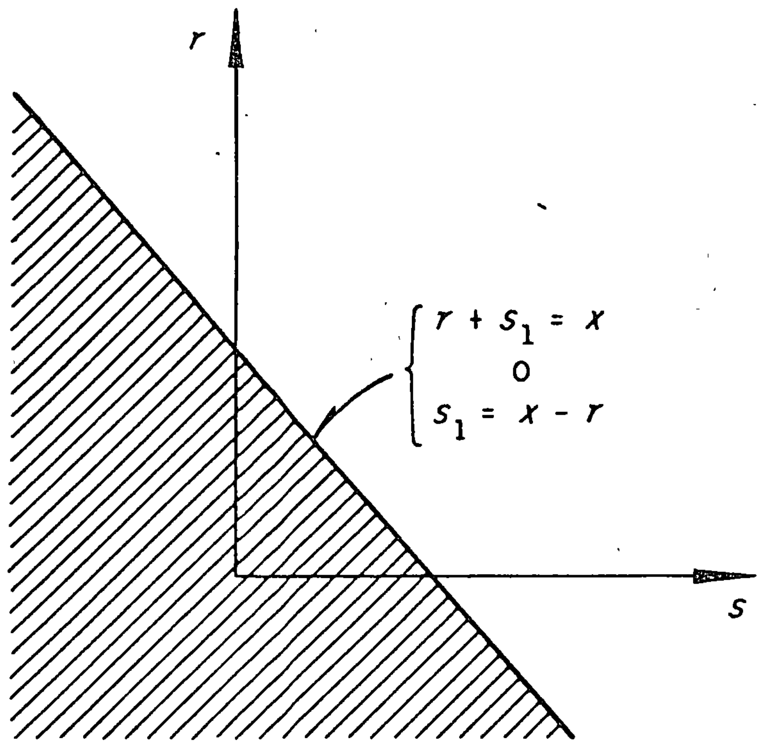


Fig 3.2 Región de integración donde $r + s_1 \leq x$

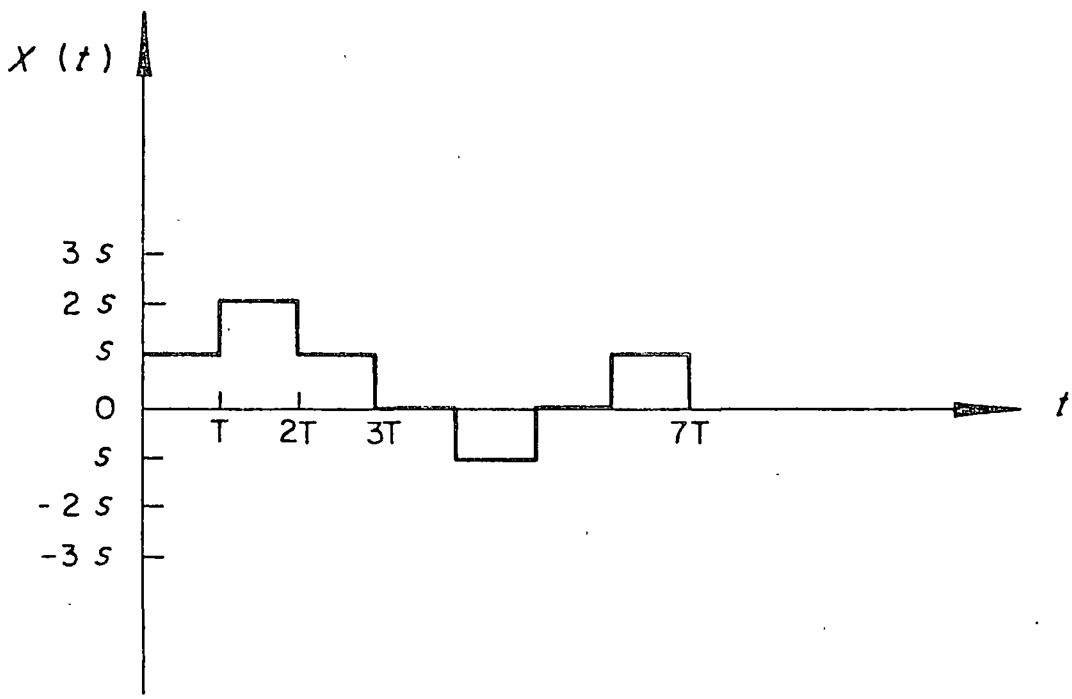


Fig 3.3 Función muestra de un paseo casual

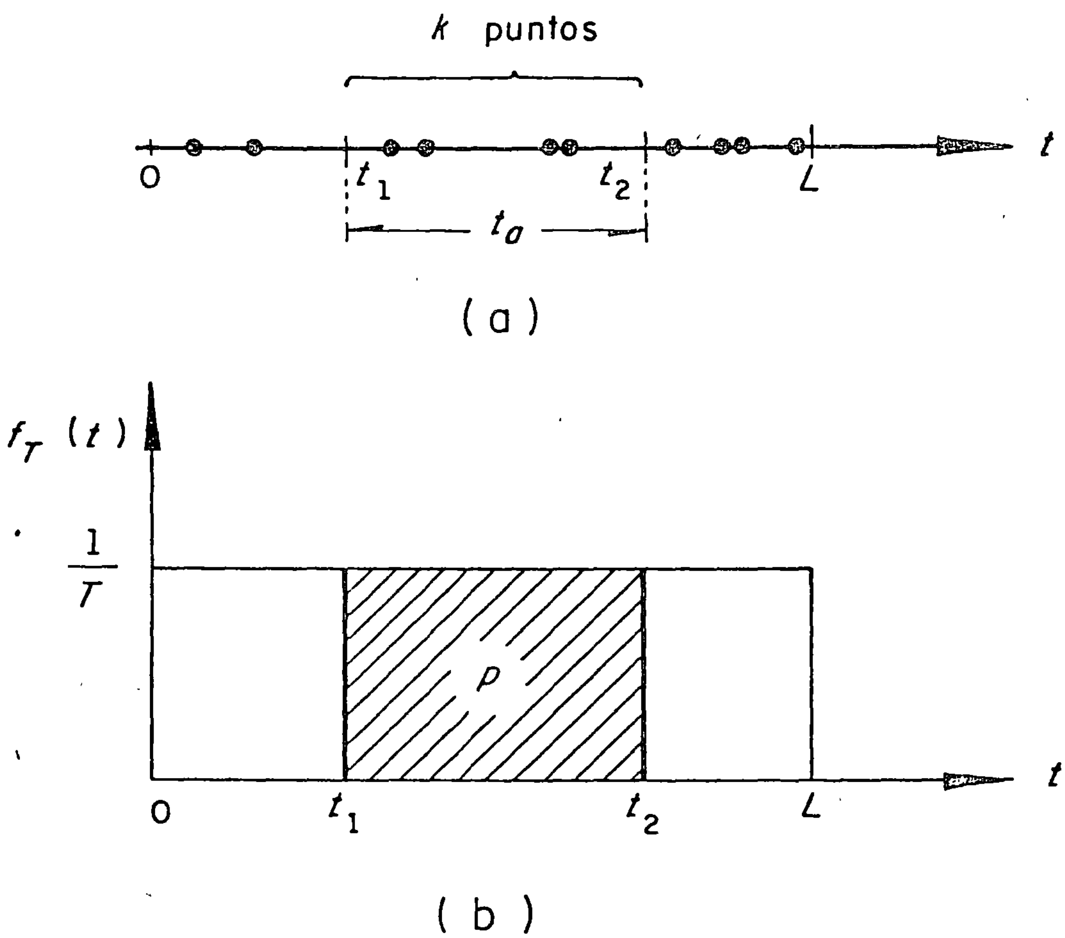


Fig 3.4 Ocurrencia aleatoria de eventos en un lapso de duración L

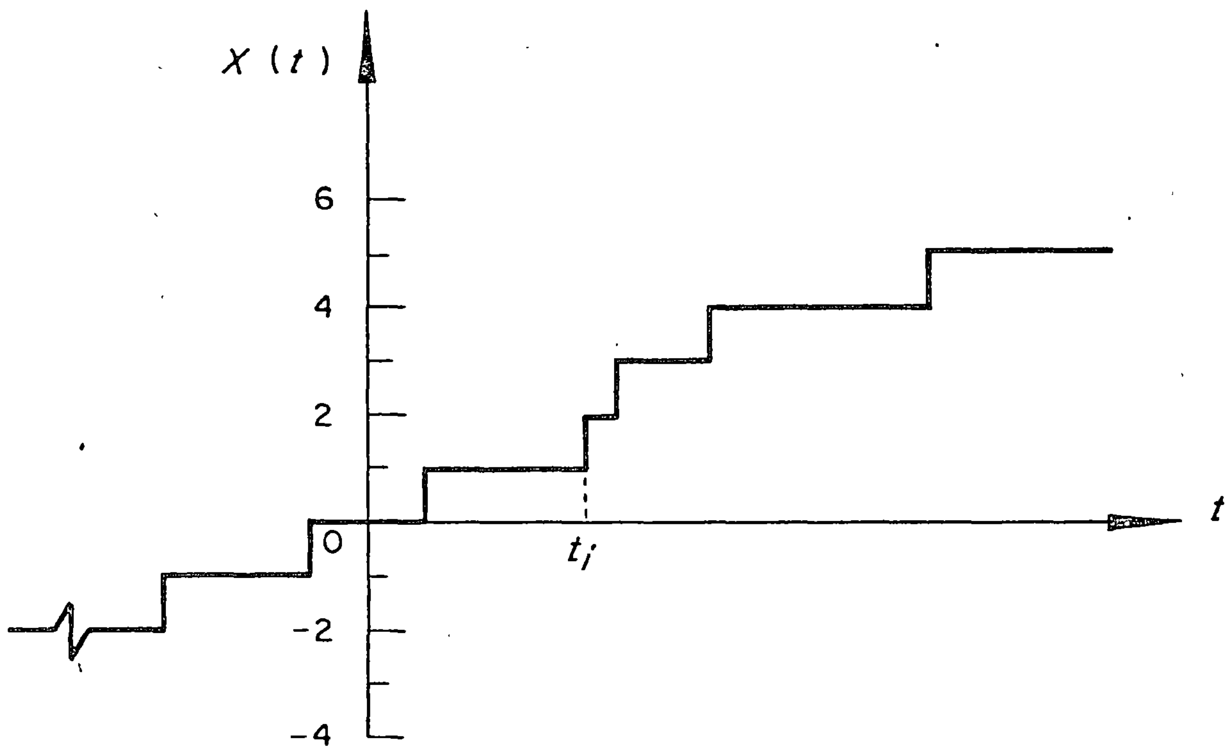


Fig 3.5 Función muestra de un proceso simple de Poisson

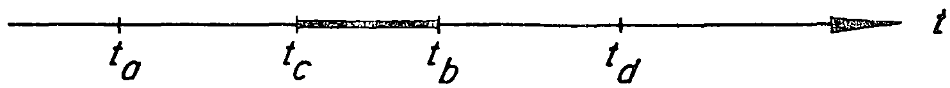


Fig 3.6 Zona de traslape de los intervalos de t_a a t_b, y de t_c a t_d

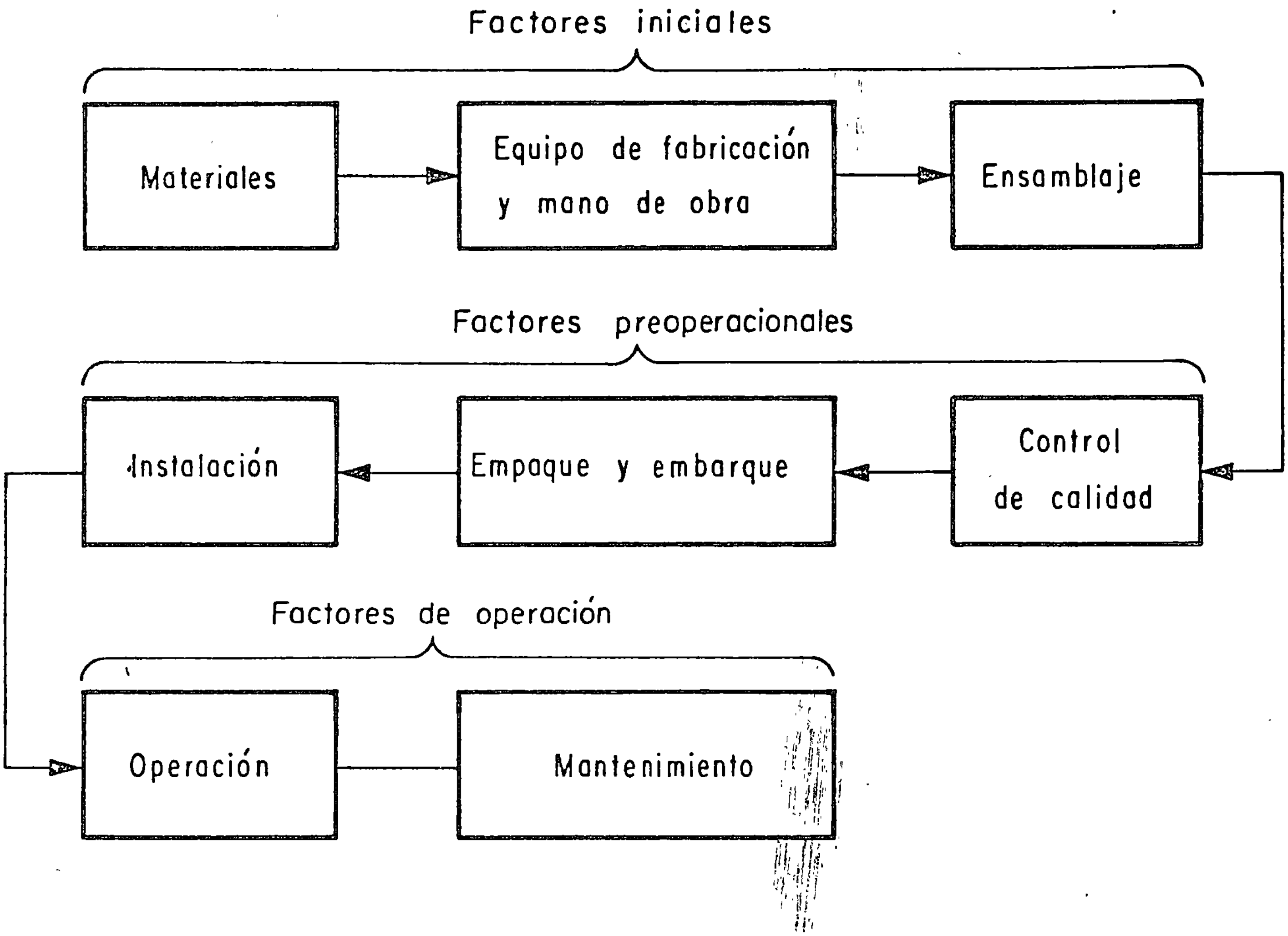


Fig 3.7 Factores que influyen en el comportamiento de un sistema

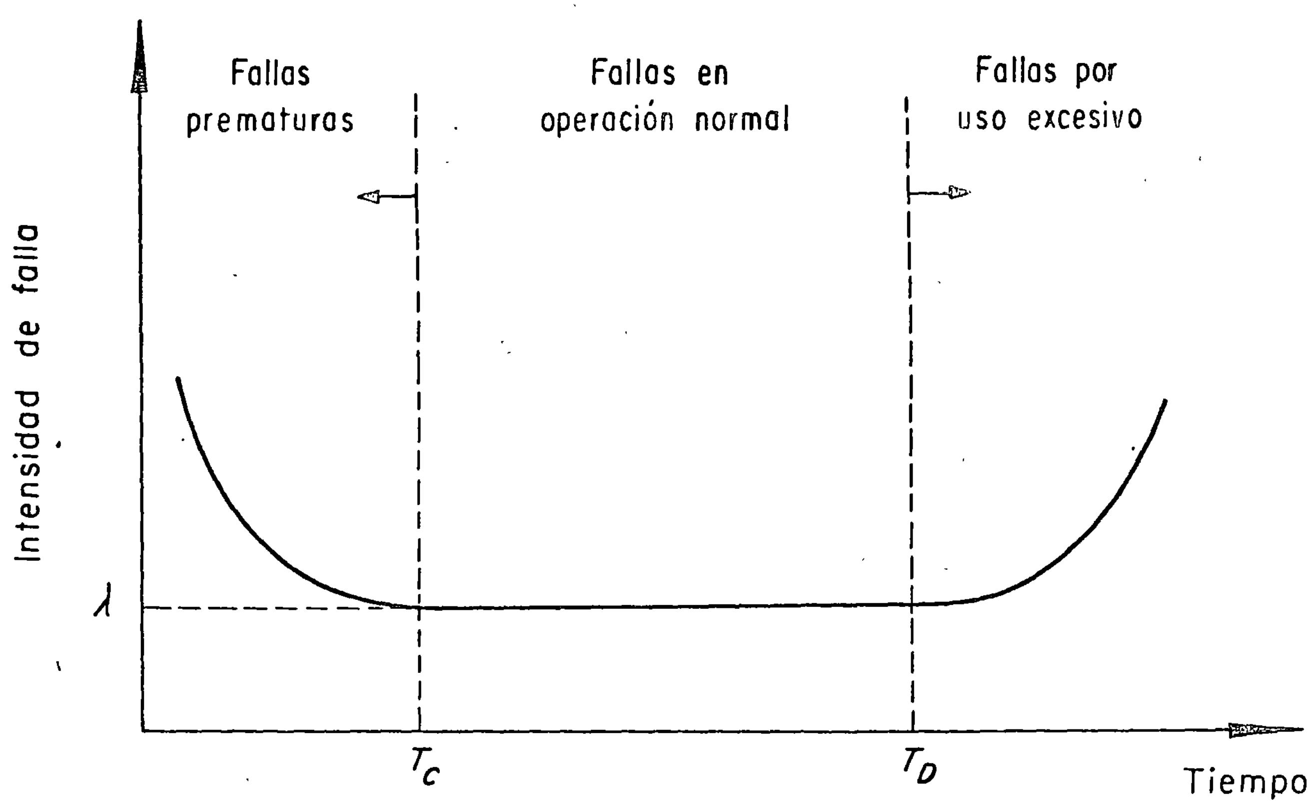


Fig 3.8 Intensidad de falla en función del tiempo de operación

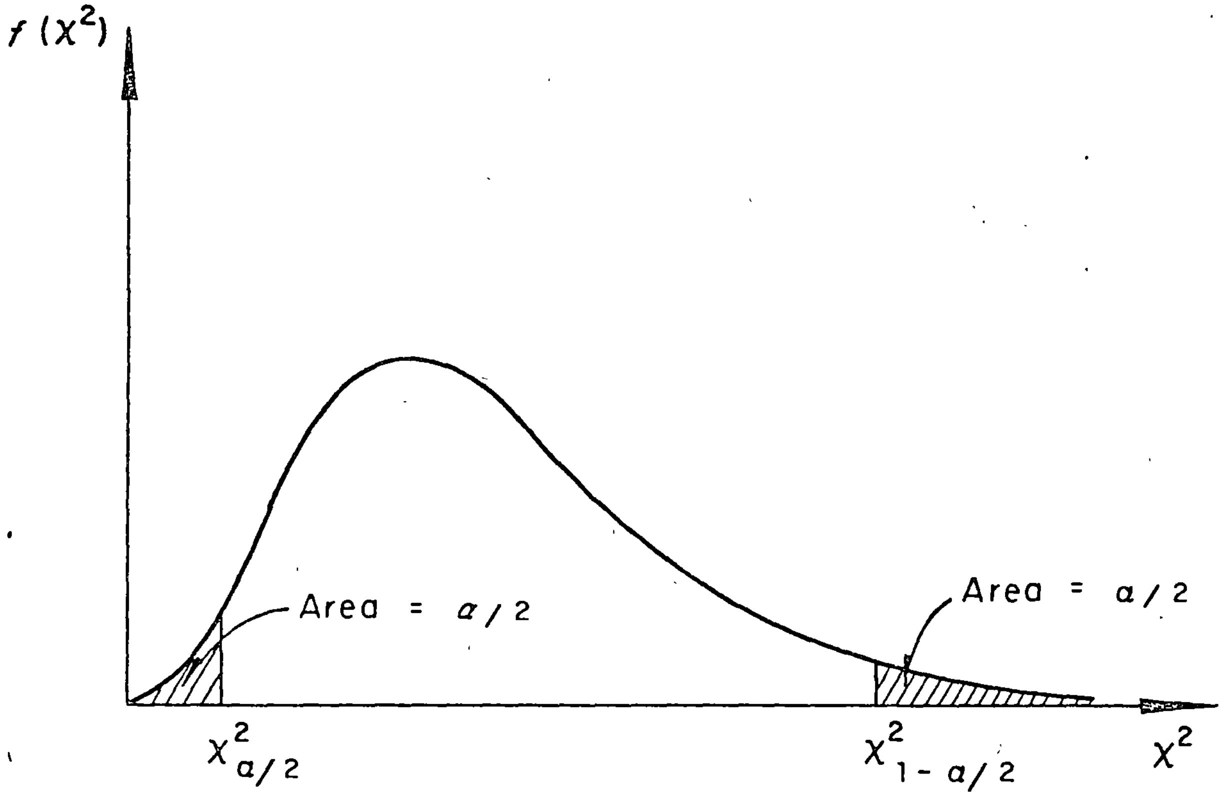


Fig 3.9 Distribución ji - cuadrada

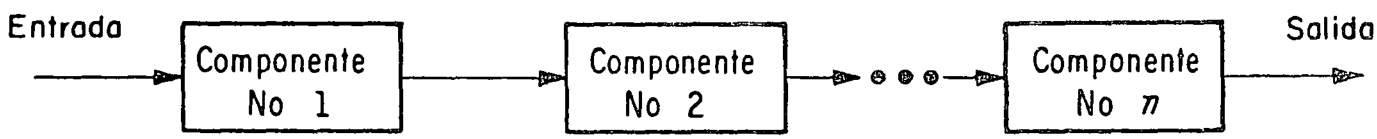


Fig 3.10 Sistema en serie

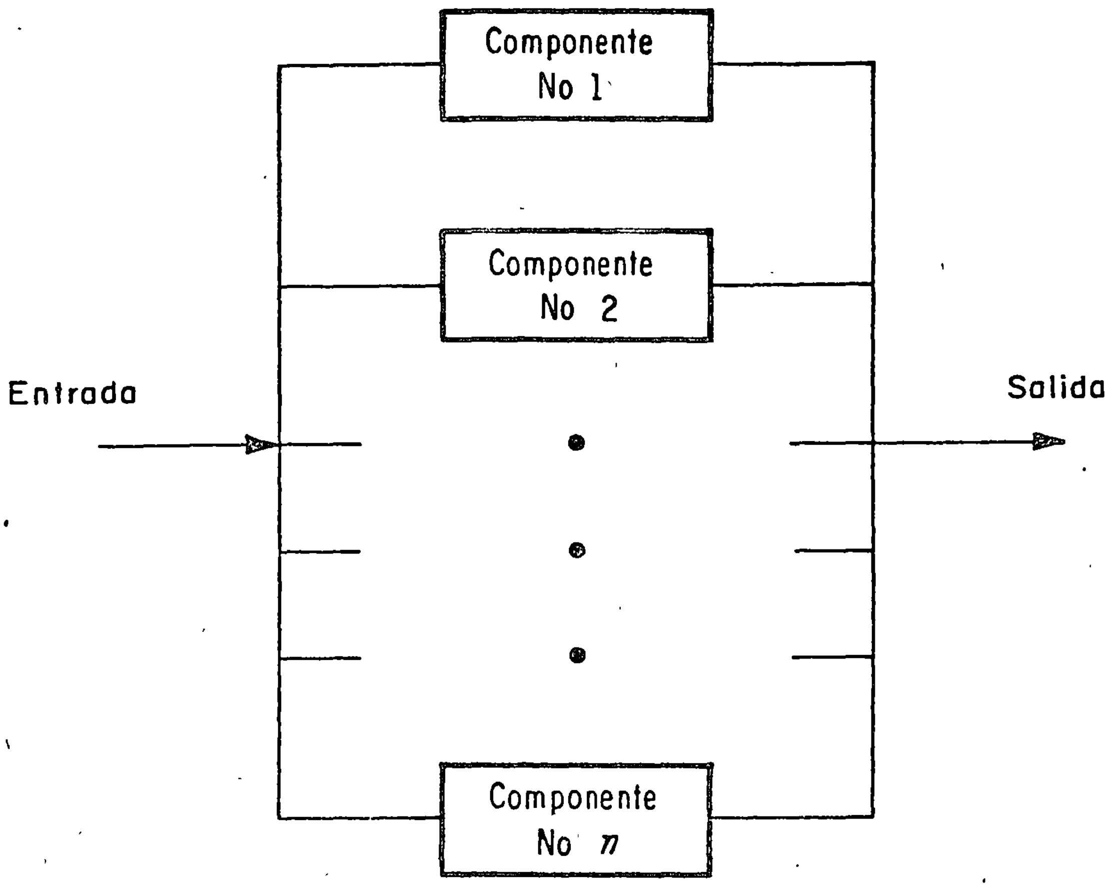


Fig 3.11 Sistema en paralelo

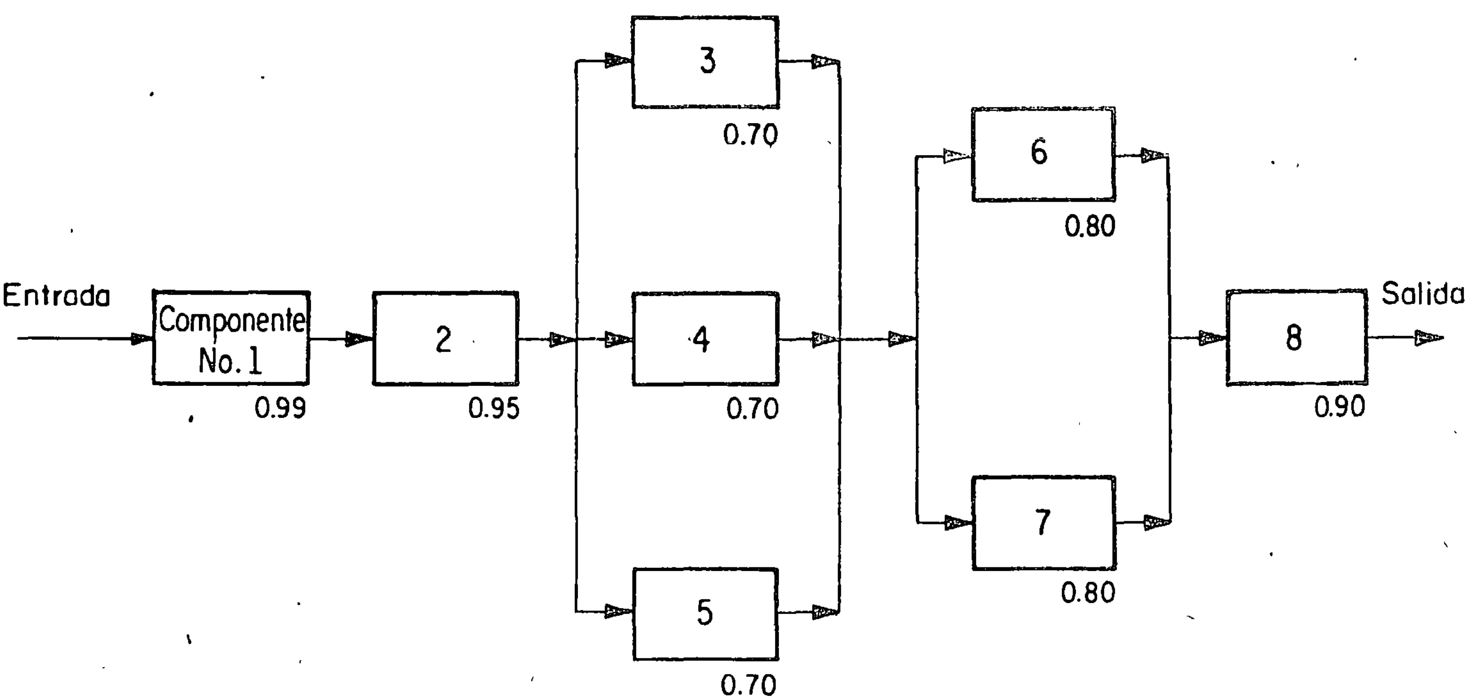


Fig 3.12 Sistema mixto

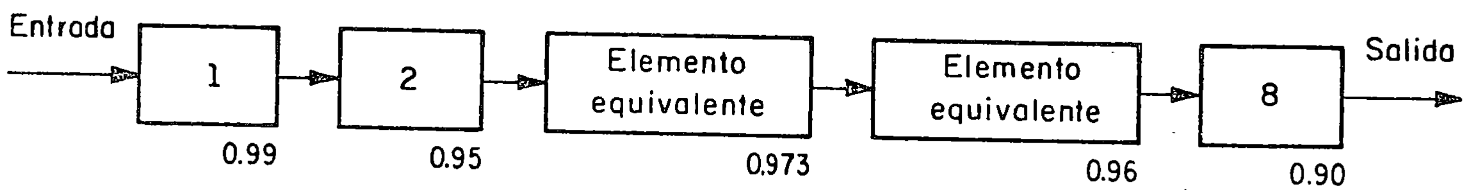


Fig 3.13 Sistema en serie equivalente al de la fig 3.12

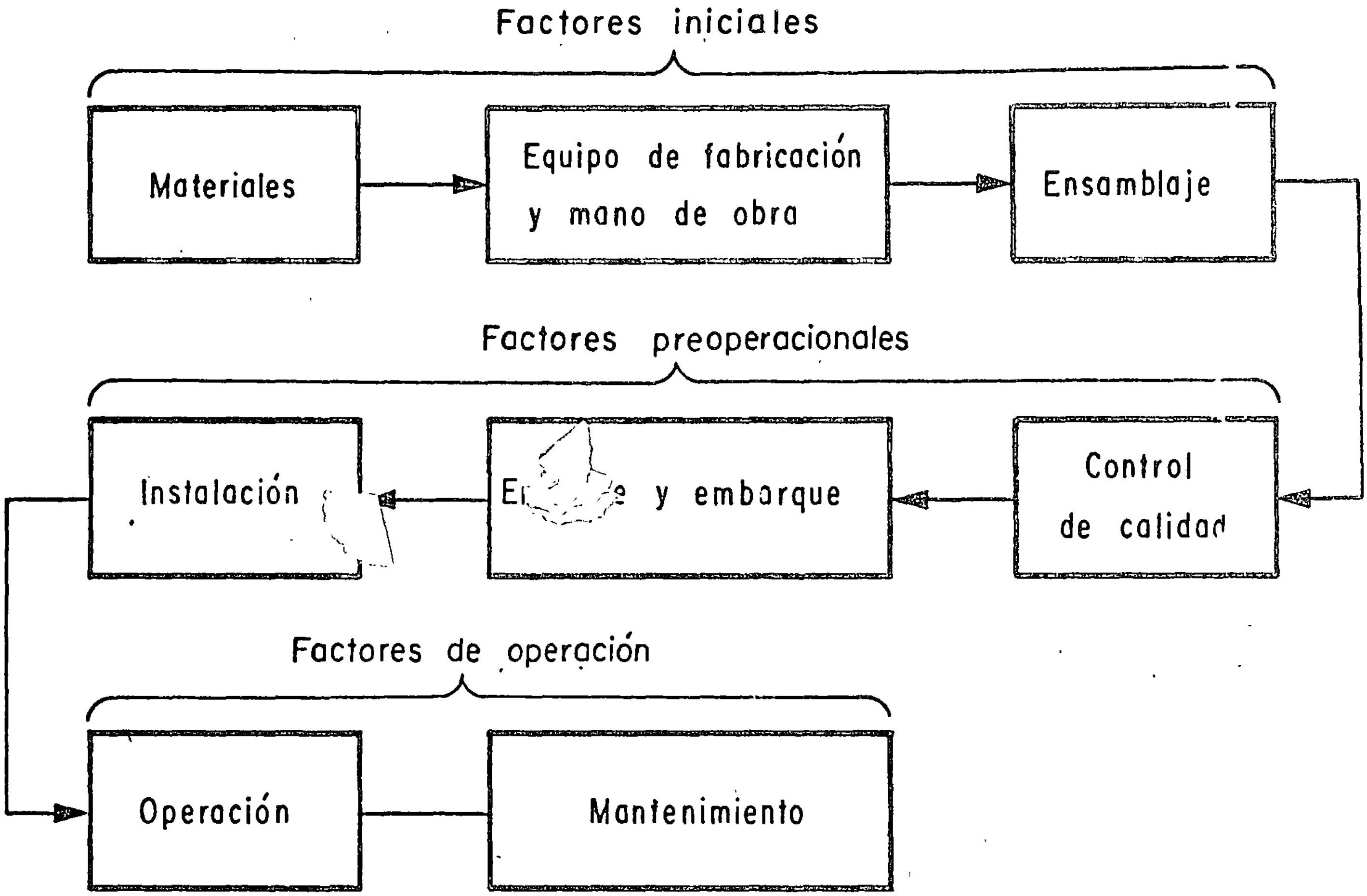


Fig 3.7 Factores que influyen en el comportamiento de un sistema

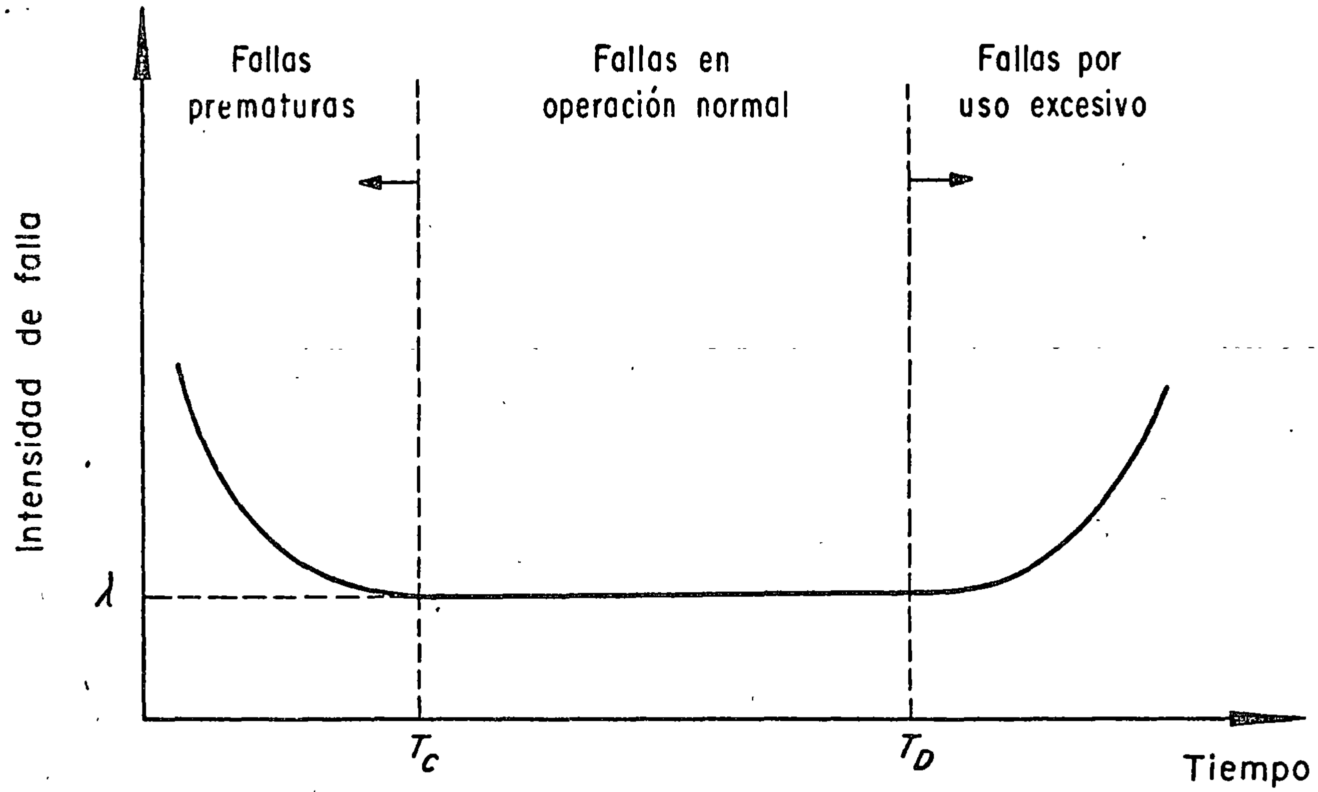


Fig 3.8 Intensidad de falla en función del tiempo de operación

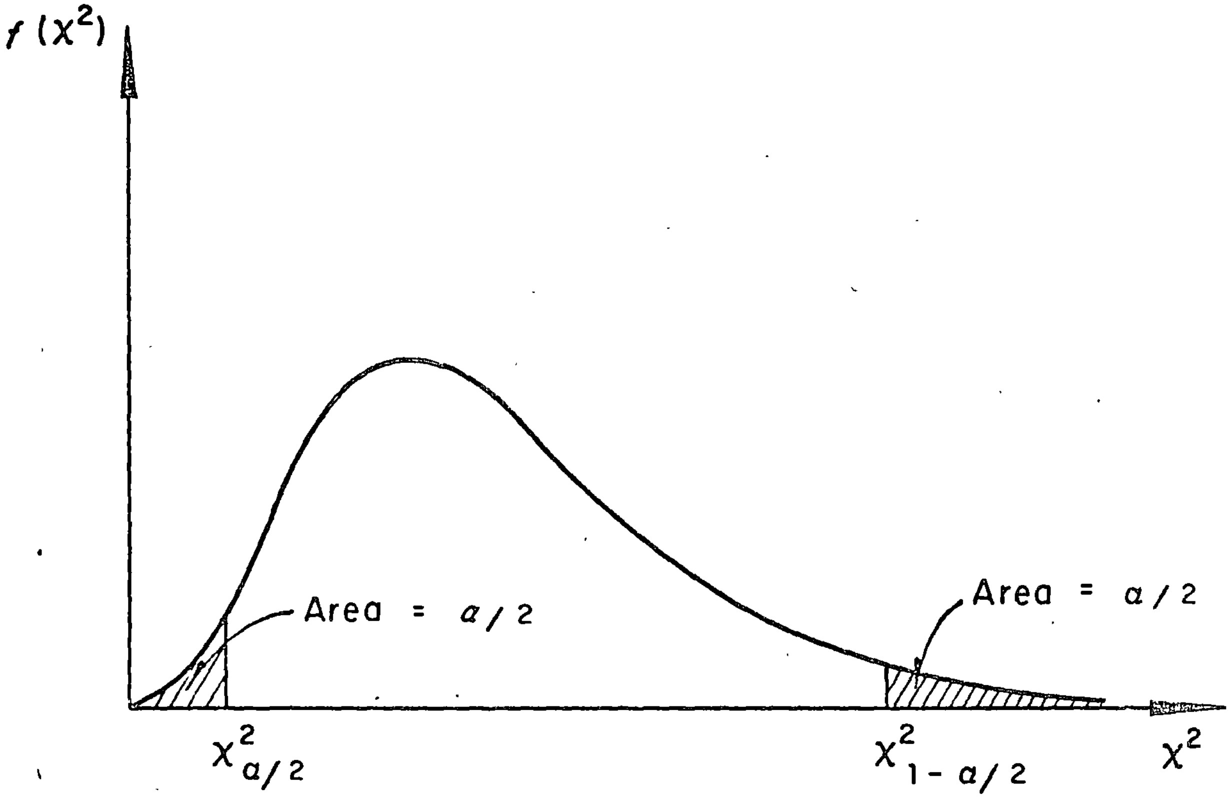


Fig 3.9 Distribución ji - cuadrada

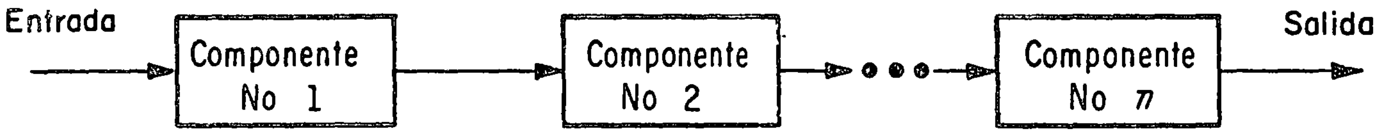


Fig 3.10 Sistema en serie

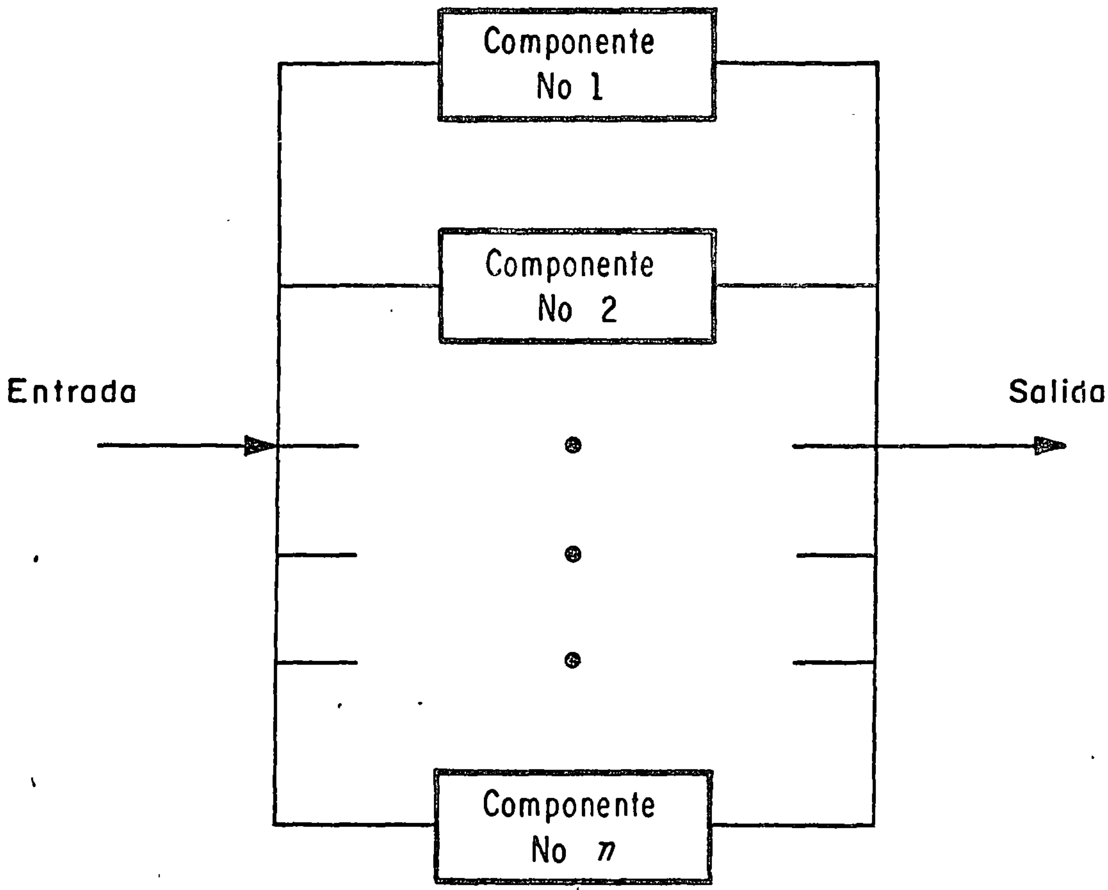


Fig 3.11 Sistema en paralelo

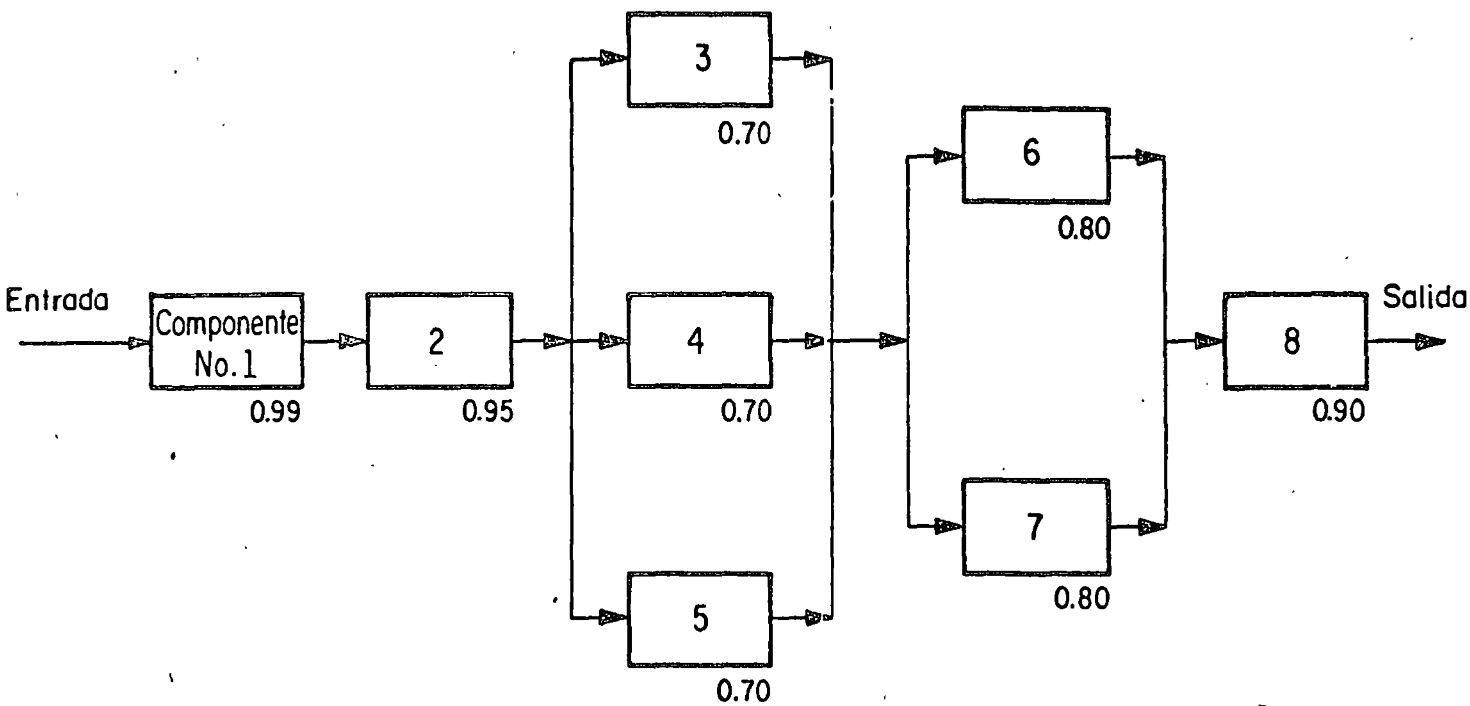


Fig 3.12 Sistema mixto

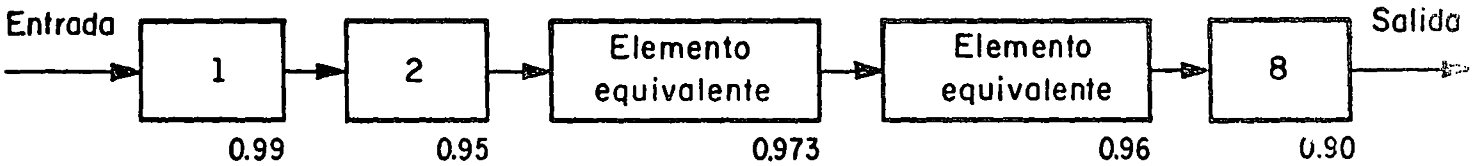
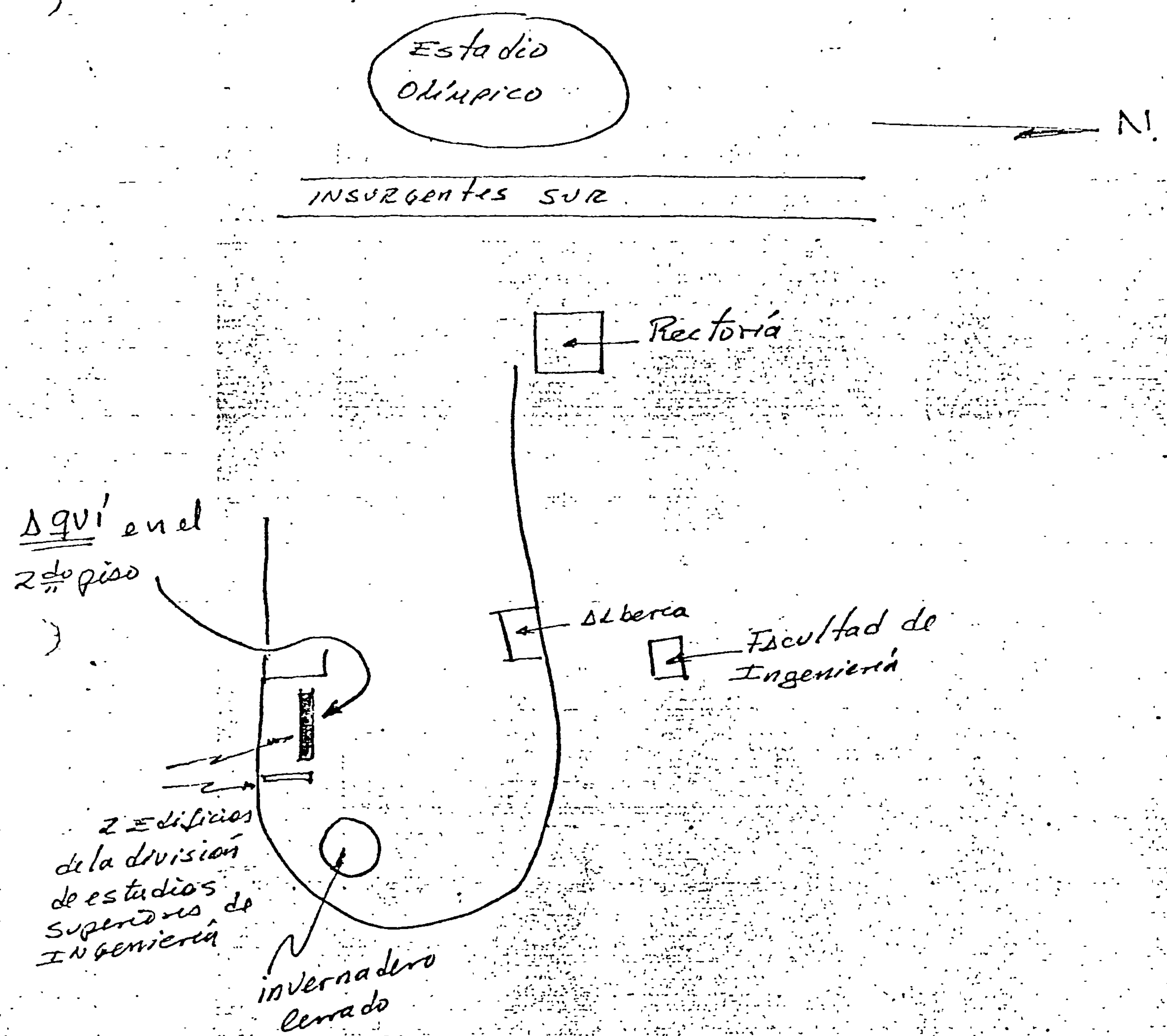


Fig 3.13 Sistema en serie equivalente al de la fig 3.12


Estadio Olímpico
N
INSURGENTES SUR
Rectoría
Aquí en el 2do piso
Alberca
Facultad de Ingeniería
2 Edificios de la división de estudios superiores de Ingeniería
invernadero cerrado

DIRECTORIO DE ASISTENTES AL CURSO: CONTROL ESTADISTICO DE CALIDAD 1978.

1	ALVARO CABALLERO ALMERALLA S.A.R.H.O.P. AV. UNIVERSIDAD Y XOLA MEXICO 12, D.F TEL. 530.40.68	2da. Gustavo E. Campa 59 Vallejo México 14, D.F. Tel. 537.79.69
2	JAIME DE LA CRUZ NOGUEDA S.A.R.H. DIR. GRAL. DE PROTEC. Y ORD. ECOLOGICA REFORMA 107-1° MEXICO 4, D.F. TEL. 566.06.88	Carlos B. Zetina 161-6 Tacubaya México 18, D.F. Tel. 516.39.26
3	PEDRO DIAZ GALVAN OLYMPIA DE MEXICO S.A. ANTIGUA CARR. MEX. PUEBLA KM. 19.5 LOS REYES LA PAZ, EDO. DE MEX. TEL. 585.42.11	Luis de la Rosa 96 Constitución de la Rep. México 14, D.F. Tel. 781.12.97
4	CECILIO JUAREZ HERNANDEZ GENERAL POPO S.A. LAGO ZURICH 245 MEXICO 17, D.F. TEL. 545.70.80	Cerrada de Almacenes 36 Edif. 10-B-4 México 16, D.F. Tel. 561.92.40
5	ROBERTO LANDEROS ORTIZ S.A.H.O.P. XOLA Y AV. UNIVERSIDAD MEXICO 12, D.F. TEL. 519.13.46	Jiquilpan 58 Col. Janitzio México 2, D.F. Tel. 529.41.41
6	LUIS GUILLERMO LIMON LIMON S.A.H.O.P. XOLA Y AVE. UNIVERSIDAD MEXICO 12, D.F. TEL. 538.09.69	Playa Revolcadero 571 México 13, D.F. Tel. 590.38.65
7	RAUL MATA GUZMAN S.A.R.H. REFORMA 107-8° MEXICO, D.F. TEL. 566.06.88 EXT.117	Serapio Rendón 92-42 México 4, D.F.

8	MIGUEL MAYORGA PINI IMMASA PATRIOTISMO 440-3 MEXICO , D.F. TEL. 516.15.53	Séneca 344 México 5, D.F. Tel. 520.71.28
9	MARIA CONCEPCION NEGRETE DE C. GILLETE DE MEXICO S.A. DE C.V. MARIANO ESCOBEDO 155 MEXICO 17, D.F. TEL. 545.69.20 EXT.150	Sn Juan de Ulua 11 Tlalpán México 22, D.F. Tel. 676.50.44 Ext.253
10	RICARDO OLIVERA BUSTAMANTE S.A.H.O.P. XOLA Y AVE. UNIVERSIDAD MEXICO 12, D.F.	
11	MANUEL ORTIZ GARCIA COMISION DEL LAGO DE TEXCOCO S.A.R.H. AV. FCO. I. MADERO Y RIO GUADALUPE MEXICO 14, D.F. TEL. 577.96.11 EXT.31	Bahía de Ballenas 49 México 17, D.F. Tel. 250.39.34
12	HUMBERTO PLANCARTE GONZALEZ S.A.H. O.P. AV. UNIVERSIDAD Y XOLA MEXICO 12, D.F. TEL. 530.46.77	Sur 105 # 1113 Col. Aeronáutica Militar México 9, D.F. Tel. 768.18.69
13	FRANCISCO QUIROS OLAZARRI LAMINADORA MEXICANA DE METALES LAGO ZURICH 168 MEXICO 17, D.F. TEL. 545.71.00	Sur 93 No.10 México 8,D.F. Tel. 552.97.51
14	NOEL RAMOS JIMENEZ COLEGIO DE CIENCIAS Y HUMANIDADES SAN BARTOLO, NAUCALPAN EDO. DE MEX.	Montealegre 19 Bis.1 México 13, D.F. Tel. 672.07.77
15	AGUSTIN LORENZO ROJAS SANTAMARIA S.A.R.H. DIR.GRAL. DE PROTEC. Y ORD.ECO. MEXICO 1, D.F. TEL. 591.14.63	Rastro 423-4 Col. Rosedal México 21, D.F. Tel. 544.79.92

16 PAULINO SANCHEZ PEREZ
CIA. MEX. DE EXPLORACIONES S.A.
RIO BALSAS 101-4
MEXICO 5, D.F.
TEL. 533.62.46

17 JAIME SANCHEZ VILLASEÑOR
DIR. GRAL. DE EST. DEL
TERRITORIO NACIONAL
SAN ANTONIO ABAD 124-3°
MEXICO 8, D.F.
TEL. 578.62.00 EXT.173

Av. Sn. Antonio 167-102
México 18, D.F.
Tel. 515.62.82

18 ALFREDO SALINAS ERUN
INST. MEX. DEL PETROLEO
AV. DE LOS CIEN METROS 152
MEICO 14, D.F.
TEL. 567.66.00

Rincón de los Cedros 36
Bosque Residencial del Sur
México 23, D.F.
Tel. 676.06.47

19 FILIBERTO RAPIA MARTINEZ
DIR. GRAL. DE EST. DEL TER.NAL.
SAN ANTONIO ABAD 124-3°
MEXICO 8, D.F.
TEL. 578.69.00 EXT.173

Navarra 175-1
México 13, D.F.
Tel. 590.42.67

20 ROBERTO TAVERA CARMONA
S.A.H.O.P.
XOLA Y AV. UNIVERSIDAD
MEXICO 12, D.F.
TEL. 530.33.11

Escultores 80
Col. Satélite
Estado de México
Tel. 562.82.62

21 MANUEL TORRES LUGO
INST. MEX. DEL PETROLEO
AV. DE LOS CIEN METROS NO. 152
MEXICO 14, D.F.
TEL. 567.66.00 EXT. 2258

22 AURELIO VARGAS DE LA HOYA
S.A.H.O.P.
XOLA Y AVE. UNIVERSIDAD
MEXICO 12, D.F.

E. Zapata 55
México 13, D.F.
Tel. 539.77.26

23 JOSE LUIS YAÑEZ FERNANDEZ
BANCO DE MEXICO
P. Santos Degollado 10
México 1, D.F.
TEL. 518.05.00

V. P. 498 E-21
Col. Magdalena Mix.
México 2, D.F.
Tel. 768.66.12