

RAUL ARRIOJA JUAREZ

INGENIERO CIVIL, FACULTAD DE INGENIERIA, UNAM, MEXICO, 1975
MASTER OF SCIENCES, UNIVERSITY OF CALIFORNIA, BERKELEY, USA, 1977

A CORRELAÇÃO ESPÚRIA NA ENGENHARIA HIDRÁULICA

Tese apresentada à Escola Politécnica da Universidade
de São Paulo para a obtenção do título de Doutor em Engenharia

Área de concentração:
Engenharia Hidráulica

Orientador: Prof. Dr. Podalyro Amaral de Souza
Professor Doutor do Departamento de Engenharia Hidráulica
e Sanitária da EPUSP

Coorientador: Dr. Gabriel Echávez Aldape
Profesor Titular C
División de Estudios de Posgrado de la Facultad de Ingeniería (DEPFI)
Universidad Nacional Autónoma de México (UNAM)

SÃO PAULO

2001

À minha mãe
à minha esposa
à minha filha.

AGRADECIMENTOS

O autor agradece a todos os que, direta ou indiretamente, colaboraram para conclusão deste trabalho.

Reconhecimentos especiais são feitos para:

ao Prof. Dr. Podalyro Amaral de Souza, pela orientação, confiança e estímulo dados para o desenvolvimento deste trabalho.

ao Prof. Dr. Gabriel Echávez Aldape, *Profesor Titular C definitivo de tiempo completo* da División de Estudios de Posgrado de la Facultad de Ingeniería (DEPFI) da Universidad Nacional Autónoma de México (UNAM), que na qualidade de coorientador deste trabalho no México, garantiu a orientação técnica ao desenvolvimento deste trabalho.

ao Dipl. Geol. Horst Herbert Blaesig Schlotfeldt, graduado da Ludwig Maximilians Universität München (Institut für Allgemeine und Angewandte Geologie) pelo apoio logístico fornecido.

ao Lic. F. M., *Candidato a M. C., Diplomado de la CEE* Raúl Magaña Zamora, *Profesor Titular C* da ESIME Culhuacan do Instituto Politécnico Nacional (IPN) no México pela ajuda fornecida para conclusão deste trabalho.

ao Dr. Rodolfo Silva Casarin, pesquisador do Instituto de Ingeniería, UNAM, pelas sugestões fornecidas para o dimensionamento do canal reprodutor de ondas descrito no Capítulo 5 desta tese.

ao Ing. Jesús Manuel Acevedo Trejo, *Profesor Titular B* da ESIME Culhuacan do Instituto Politécnico Nacional (IPN) no México.

ao Sr. Alberto Sahonero Coronel e à Arq. Carola Sahonero Gutiérrez pela hospitalidade oferecida nas viagens pela Bolívia, rumo ao Brasil.

Sem estas ajudas não teria sido possível elaborar esta tese.

SUMÁRIO

Lista de figuras	VI
Lista de tabelas	VII
Lista de símbolos	VIII
Resumo	
XI	
Abstract	XII
Resumen	XIII
1 INTRODUÇÃO	1
2 REVISÃO BIBLIOGRÁFICA SOBRE O TEMA DE CORRELAÇÃO ESPÚRIA ORIENTADO À ENGENHARIA HIDRÁULICA	3
2.1 Introdução	3
2.2 Revisão bibliográfica do tema abordado	4
2.3 Conclusões	18
3 A FÓRMULA DE PEARSON PARA CALCULAR O COEFICIENTE DE CORRELAÇÃO ESPÚRIA	20
3.1 Introdução	20
3.2 Exemplo de aplicação da fórmula de Pearson a um caso de estudo	22
3.3 Exemplo numérico simples de aplicação	32
4 DESENVOLVIMENTO DE UM MODELO COMPUTACIONAL PARA O CÁLCULO DO COEFICIENTE DE CORRELAÇÃO ESPÚRIA	34
4.1 Introdução	34

4.2 Descrição do programa ESPURI	34
4.3 Exemplos de organização de arquivos para o programa e dos respectivos arquivos de saída de resultados	39
4.4 Compilação do programa ESPURI	47
4.5 Usos de planilha de cálculo eletrônica para calcular os dados de ingresso para o programa num caso de estudo	48
5 ORIENTAÇÕES PARA APLICAÇÕES NA ENGENHARIA HIDRÁULICA	57
5.1 Introdução	57
5.2 Caso de estudo No. 1. Proposta de equação para uma curva chave de um canal trapezoidal	60
5.3 Caso de estudo No. 2. Geração de onda num canal reprodutor de ondas	76
5.4 Caso de estudo No. 3. Vertedouro triangular	80
CONCLUSÕES	84
SUGESTÕES PARA FUTURAS PESQUISAS	87
REFERÊNCIAS BIBLIOGRÁFICAS	91
ANEXO	93
Introdução	93
Exemplos de aplicação	94

LISTA DE FIGURAS

Figura 1.- Disposição das variáveis y e z no plano cartesiano.	22
Figura 2.- Diagrama de fluxo do programa de computador ESPURI.	49
Figura 3.- Seção transversal de um canal trapezoidal.	60
Figura 4.- Correlação entre y e z	69
Figura 5.- Seção longitudinal de um canal reproduzidor de ondas.	76
Figura 6.- Elevação do quebramar proposto com dimensionamento.	77
Figura 7.- Vista em planta do canal de ondas.	79
Figura 8.- Seção transversal de um vertedouro triangular de crista delgada.	80
Figura 9.- Seção longitudinal de um vertedouro triangular de crista delgada instalado num canal retangular.	82
Figura 10.- Gráfico para obtenção do coeficiente de correlação espúria.	100

LISTA DE TABELAS

Tabela 1.- Modelos algébricos propostos em forma de quociente.	27
Tabela 2.- Diversos modelos algébricos propostos em forma de produto.	30
Tabela 3.- Modelos algébricos mais freqüentes na engenharia hidráulica.	59
Tabela 4.- Planilha de cálculo para o caso de estudo No. 1.	68
Tabela 5.- Planilha de cálculo eletrônica para os parâmetros estatísticos que integram o arquivo de ingresso para o programa ESPURI.	70
Tabela 6.- Cálculo do número de Froude para os valores da Tabela 4.	71
Tabela 7.- Planilha de cálculo eletrônica para os parâmetros estatísticos que integram o arquivo de ingresso para o programa ESPURI para o exercício onde usam-se números aleatórios para todas as variáveis.	72
Tabela 8.- Classificação do escoamento de um vertedouro triangular de crista delgada.	83
Tabela 9.- Classificação do coeficiente de correlação	85
Tabela 10.- Influência do valor relativo do coeficiente de variação sobre o coeficiente ρ .	90
Tabela 11.- Coletânea de modelos cuja solução gráfica obtém-se através da Figura 10.	99

LISTA DE SÍMBOLOS

A = área da seção transversal do canal

b = largura do canal

C = coeficiente de rugosidade de Chezy

coeficientes

$C_i = \frac{\sigma_i}{m_i}$ coeficiente de variação

$\left| \frac{C_i}{C_j} \right|$ = valor absoluto do quociente dos coeficientes de variação C_i, C_j

$C1, P, Q, R, S$ = variáveis de FORTRAN que caracterizam a função y

$C2, T, U, V, W$ = variáveis de FORTRAN que caracterizam a função z

$CR (I,J)$ = variável de FORTRAN para indicar o coeficiente de correlação entre as variáveis com subíndices i e j

d = profundidade do escoamento

d_{cr} = profundidade crítica

D = tamanho de grão

Δ = altura da onda de areia

ΔW = mudança de armazenamento da água por ano

E = valor da energia usada na geração da onda

Φ = ângulo de abertura do vertedouro

F = número de Froude

$$f_a = \frac{\delta}{\delta m_a} f(m_1, m_2, m_3, m_4) \quad a = 1, \dots, 4 \text{ derivada parcial de uma função } f(m_i) \text{ com}$$

respeito a uma variável m_a

g = constante gravitacional

$$g_b = \frac{\delta}{\delta m_b} g(m_5, m_6, m_7, m_8) \quad b = 5, \dots, 8 \text{ derivada parcial de uma função } g(m_i) \text{ com}$$

respeito a uma variável m_b

$$GE = 100 \frac{\rho_{yz}}{r_{yz}} = \text{grau de espureidade da correlação}$$

γ = peso específico da água.

h_5 = altura de onda gerada para uma distância “D” horizontal de 5 vezes a profundidade de escoamento no ponto considerado

H = carga hidráulica do vertedouro

KI = 5 índice dentro do programa ESPURI

KF = 8 índice dentro do programa ESPURI

Λ = comprimento da onda de areia

= cotangente do talude do canal

m_1, m_2, \dots, m_8 = médias

M (I) = variável de FORTRAN para indicar a média i

n = coeficiente de rugosidade de Manning

NI = 1 índice dentro do programa ESPURI

NF = 4 índice dentro do programa ESPURI

p, q, r, s, t, u, v, w = expoentes

P_e = altura efetiva de chuva

P_{e-prom} = altura efetiva média de chuva

= vazão

r_{yz} = coeficiente de correlação de Pearson

$r_{12}, r_{13}, \dots, r_{18}, \dots, r_{78}$ = coeficientes de correlação entre variáveis

R = raio hidráulico da seção transversal do canal

ρ = coeficiente de correlação

ρ_{yz} = coeficiente de correlação espúria

s = declividade do canal

S = declividade da linha de energia

SIGMA (I) = variável de FORTRAN para indicar o desvio padrão i

$\sigma_1, \sigma_2, \dots, \sigma_8$ = desvios padrões

v_0 = velocidade de aproximação do escoamento d'água

v = velocidade do fluxo da água no canal

V = escoamento direto anual

V_{prom} = escoamento direto médio anual

W_i = armazenamento da água por ano

variáveis aleatórias

$y = C_1 X_1^p X_2^q X_3^r X_4^s$ função adimensional

$z = C_2 X_5^t X_6^u X_7^v X_8^w$ função adimensional

RESUMO

A análise de correlação é uma técnica estatística muito difundida que investiga e modela a relação entre duas ou mais variáveis aleatórias ou conjuntos de dados experimentais ordenados por duplas para fazer previsões para uma variável aleatória considerada como dependente em função de uma ou várias variáveis aleatórias consideradas independentes. Na engenharia hidráulica é freqüente a obtenção de parâmetros adimensionais pelo uso do teorema de Buckingham, por meio da formação de produtos “ π ” para obter ditos parâmetros. Na engenharia experimental tenta-se estabelecer correlações entre dois ou mais parâmetros adimensionais para propósitos de interpolações ou extrapolações. Às vezes inclui-se uma variável comum nos dois ou mais parâmetros adimensionais correlacionados com o intuito de melhorar o grau de correlação, o que seria refletido num valor mais alto do coeficiente de correlação. Porém, esta prática pode levar a valores enganosamente altos de dito coeficiente pelo efeito da correlação espúria. Neste trabalho desenvolve-se uma metodologia estatística para calcular o grau de correlação espúria que afeta um coeficiente de correlação dado para estabelecer critérios de aceitação ou rejeição de diversos modelos algébricos de correlação propostos freqüentemente na hidráulica que tenham uma variável comum. Também se fazem recomendações para evitar o problema da correlação espúria. Para facilitar a aplicação desta metodologia, desenvolve-se um programa de computador escrito em FORTRAN.

ABSTRACT

Correlation analysis is a statistical technique of widespread use which investigates and models the relationship among two or more random variables or sets of experimental data ordered by pairs to forecast values for a random variable considered as dependent as a function of one or several random variables deemed as independent. In hydraulic engineering it is frequent to obtain adimensional parameters using the Buckingham theorem, through the formation of “pi” products to obtain these parameters. In experimental engineering it is often intended to establish a correlation among two or more adimensional parameters for purposes of interpolation or extrapolation. Sometimes a common variable is included in the two or more correlated adimensional parameters in order to improve the degree of correlation, which would result in a higher value of the correlation coefficient. However, this practice can lead to deceptively high values of such coefficient due to the effect of spurious correlation. In this thesis a statistical methodology is developed to calculate the degree of spurious correlation which affects a given coefficient of correlation in order to establish acceptance or rejection criteria for several correlation algebraic models of frequent use in hydraulics which have a common variable. Recommendations are also formulated to avoid the problem of spurious correlation. To facilitate the application of this methodology, a computer program written in FORTRAN is developed.

RESUMEN

El análisis de correlación es una técnica estadística de uso muy difundido que investiga y modela la relación entre dos o más variables aleatorias o conjuntos de datos experimentales ordenados por pares para hacer predicciones para una variable aleatoria considerada como dependiente en función de una o varias variables aleatorias consideradas independientes. En ingeniería hidráulica es frecuente la obtención de parámetros adimensionales con el uso del teorema de Buckingham, por medio de la formación de productos “pi” para obtener dichos parámetros. En la ingeniería experimental suele intentarse establecer correlaciones entre dos o más parámetros adimensionales para propósitos de interpolaciones o extrapolaciones. En ocasiones se incluye una variable común en los dos o más parámetros adimensionales correlacionados con objeto de mejorar el grado de correlación, lo que se reflejaría en un valor más alto del coeficiente de correlación. Sin embargo, esta práctica puede conducir a valores engañosamente altos de dicho coeficiente por el efecto de la correlación espuria. En este trabajo se desarrolla una metodología estadística para calcular el grado de correlación espuria que afecta a un coeficiente de correlación dado con objeto de establecer criterios de aceptación o rechazo de diversos modelos algebraicos de correlación de uso frecuente en la hidráulica que tienen una variable en común. Asimismo se hacen recomendaciones para evitar el problema de la correlación espuria. Para facilitar la aplicación de esta metodología, se desarrolla un programa de computadora escrito en FORTRAN.

RAUL ARRIOJA JUAREZ

Ingeniero Civil, Facultad de Ingeniería, UNAM, México, 1975
Master of Sciences, University of California, Berkeley, USA, 1977

A CORRELAÇÃO ESPÚRIA NA ENGENHARIA HIDRÁULICA

*Tese apresentada à Escola Politécnica da
Universidade de São Paulo para a obtenção
do título de Doutor em Engenharia*

SÃO PAULO

2002

CAPITULO 1

INTRODUÇÃO

Na engenharia é comum tentar estabelecer relações algébricas ou geométricas entre duas ou mais variáveis aleatórias ou entre conjuntos de dados experimentais ordenados com o intuito de prever o valor de uma das variáveis aleatórias em função das outras. Concordando com as características do experimento realizado, assinala-se a função de variável dependente a uma das variáveis aleatórias consideradas e às outras se assinala a de variáveis independentes. A variável dependente caracteriza-se por ser a grandeza cujo valor deseja-se conhecer, medindo-a em função das outras variáveis que são mais simples de medir, podem ser medidas com maior exatidão, ou são de ordem secundária. Esta técnica matemática é conhecida genericamente como análise de correlação e é amplamente usada na engenharia, sendo o exemplo mais difundido a correlação dos mínimos quadrados, que permite estabelecer relações algébricas empíricas entre variáveis minimizando o erro existente entre valores observados e valores estimados, seguindo o critério de minimizar a soma dos quadrados destas diferenças. Na análise de correlação define-se o coeficiente de correlação, que é um parâmetro estatístico que tenta quantificar numericamente a porcentagem do grau de correlação estatística entre duas variáveis propostas. Porém, para certas condições experimentais apresenta-se o fenômeno da correlação espúria, ou falsa, que contribui enganosamente a uma sobre estimativa do valor do coeficiente de correlação, o que é inconveniente do ponto de vista experimental. O primeiro objetivo deste trabalho visa o desenvolvimento de uma metodologia para estimar a porcentagem de correlação espúria que pode apresentar-se numa análise de correlação entre variáveis ou grupos de

variáveis, e também desenvolver estratégias para minimizar ou eliminar este tipo de correlação.

Outrossim, cada variável independente que intervém num fenômeno tem um comportamento diferente, com faixas de variação diversas e mesmo de natureza distinta. Assim, freqüentemente a velocidade de um fluido pode-se variar de forma contínua, de modo que cubra a faixa de interesse experimental, enquanto que as variáveis geométricas, por exemplo, o diâmetro de um tubo ou a largura de um canal, variam de forma discreta com uma possível dispersão mínima de cada valor, e a viscosidade do fluido, mesmo que intervenha de forma importante no fenômeno e por isso deve ser considerada, não é modificada nem uma só vez (é considerada constante). Então também é importante, como segundo objetivo deste trabalho, considerar esta situação para futuros trabalhos de pesquisas (vide Sugestões para futuras pesquisas) com o objetivo de analisar como esta diferença de comportamento intervém e influi nos resultados de um experimento.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA SOBRE O TEMA DE CORRELAÇÃO ESPÚRIA ORIENTADO À ENGENHARIA HIDRÁULICA

2.1 Introdução

Na engenharia hidráulica é freqüente a obtenção de parâmetros adimensionais empregando técnicas de análise dimensional baseadas no teorema de Buckingham (formação de produtos “pi” para obter parâmetros adimensionais). Os parâmetros adimensionais (homogêneos) obtidos devem ser tão simples quanto possível e devem incluir o menor número possível de variáveis representadas por grandezas de referência. Por exemplo, na maioria dos problemas de engenharia consideram-se como dimensões de referência a massa, o comprimento e o tempo, e o número de parâmetros adimensionais que se pode calcular com o teorema de Buckingham são três a menos que as grandezas de referência que intervenham no problema (habitualmente três ou duas).

Existem razões para usar a análise dimensional na engenharia, como por exemplo: 1) as dimensões são conceitos inventados pelo homem que não correspondem necessariamente aos fenômenos naturais estudados (por tanto, requer-se de uma ferramenta adicional para o estudo dos fenômenos naturais); 2) as representações funcionais em termos de quantidades adimensionais são a forma mais simples de relações dimensionalmente homogêneas e são invariantes das unidades de medição; 3) a análise dimensional agrupa as variáveis consideradas em parâmetros adimensionais para um fenômeno estudado (Mahmood, 1980). Uma definição de análise dimensional pode ser “um método pelo qual se deduz informação acerca de um fenômeno através de uma única premissa que os fenômenos possam ser descritos por uma equação dimensionalmente correta que inclua certas variáveis” (a

method by which we deduce information about a phenomenon from the single premise that the phenomena can be described by a dimensionally correct equation having certain variables), (Mahmood, 1980).

2.2 Revisão bibliográfica do tema abordado

Na experimentação usam-se também, com muita frequência, inferências baseadas em interpolações ou extrapolações para formular predições, nas quais se estima o valor de uma variável dependente em termos de uma ou várias variáveis independentes, através de uma correlação estatística estabelecida por uma relação bem definida entre parâmetros dependentes e independentes que são parte de grandezas adimensionais que são as que governam a correlação dada. Pode-se definir a correlação como “O processo para estabelecer uma relação entre uma variável e uma ou mais variáveis relacionadas. A correlação é simples se existir somente uma variável independente; e é múltipla se existir mais de uma variável independente.” (*The process of establishing a relation between a variable and one or more related variables. Correlation is simple if there is only one independent variable; multiple if there is more than one independent variable.*), (Langbein, 1960b). Para obter uma correlação entre uma e outra ou mais variáveis, usam-se preferentemente equações matemáticas obtidas através de métodos estatísticos de regressão que são bem conhecidos, como por exemplo, o método dos mínimos quadrados (Miller, 1985). A correlação matemática entre duas variáveis dadas pode exprimir-se por dois tipos de curvas de regressão, que são uma reta (quando a regressão entre as variáveis é linear) ou uma curva de grau n (quando a regressão entre as ditas variáveis é não linear), segundo a tendência geral apreciada nos dados experimentais quando estes são plotados num diagrama de dispersão (Miller, 1985). Também se pode obter uma curva de correlação baseada num método de correlação gráfica, como por exemplo, pelo método proposto por

Searcy (1960), do U. S. Geological Survey, no qual se usa um gráfico logarítmico para plotar as medianas das vazões consideradas para duas estações de medição que se deseja correlacionar, agrupadas segundo faixas verticais e horizontais, que podem variar de 5 a 10 em número, e que contenham pelo menos três pontos, determinando graficamente para cada faixa a mediana do grupo de pontos da seguinte maneira: se o número de pontos numa faixa for par, obtém-se a média dos três pontos médios do grupo (tanto no sentido vertical como horizontal); se o número de pontos da faixa for ímpar, obtém-se a média dos três pontos médios do grupo, dando ao ponto médio um peso dois. Se uma faixa contém menos de três pontos, deve ser descartada (Searcy, 1960). Oferece-se outra exposição bastante ampla do tema da correlação gráfica em Linsley (1975), onde se estuda em primeiro termo a correlação gráfica entre duas variáveis, depois a correlação gráfica entre três variáveis, e finalmente a obtenção da correlação axial entre quatro ou mais variáveis. Para o primeiro caso de estudo, menciona-se que para uma correlação linear, a reta de melhor ajuste deve passar sempre pelo ponto definido pelas médias dos valores observados das duas variáveis consideradas X e Y, mesmo para uma reta obtida por meio da correlação gráfica que para uma reta obtida pelo método dos mínimos quadrados. Para a rápida obtenção de uma reta de melhor ajuste pelo método gráfico, podem dividir-se os dados em dois ou mais grupos de valores para obter suas respectivas médias, e depois obter as médias das médias até definir a reta procurada. Porém, o método gráfico fornece uma reta que minimiza a soma absoluta dos desvios entre valores verdadeiros e valores estimados, enquanto que o método analítico dos mínimos quadrados minimiza a soma dos quadrados dos ditos desvios. Nas duas seções restantes do Apêndice A (Linsley, 1975) descrevem-se métodos parecidos à metodologia empregada para construir nomogramas para obter correlações entre três ou mais variáveis. Os dois tipos de correlação, gráfica ou numérica, visam o mesmo objetivo,

porém, diferem no procedimento empregado. Segundo Searcy, as vantagens do procedimento de correlação gráfica empregado no seu caso de estudo de dois pontos de medição hidrométrica, um deles com um longo registro histórico de vazões e o outro com um registro curto, são que o plotado dos pontos da relação das vazões no papel logarítmico tende a normalizar os dados e a transformar a relação originalmente curvilínea numa relação linear, e que o método analítico dos mínimos quadrados tende a dar um peso excessivo aos pontos extremos ou remotos da relação, contribuindo assim a mover a reta (ou curva) de correlação proposta aproximando-a dos pontos extremos, introduzindo assim um viés indesejável (Searcy, 1960). Obtém-se o mesmo efeito, embora menor, se forem usadas as médias dos grupos de vazões no método gráfico em lugar das medianas.

O critério mais comumente empregado para tentar avaliar o grau de correlação aparente entre variáveis ou funções de predição propostas é o coeficiente de correlação ρ , que está associado com uma representação gráfica da correlação proposta. Geralmente considera-se que se forem obtidos valores altos do coeficiente de correlação, os parâmetros relacionados linearmente na correlação são eficientes para predizer o valor da variável dependente. Porém, este critério estatístico não é suficiente para que a norma de predição proposta seja realmente eficiente, quer dizer, que realmente se baseie em relações causa-efeito bem definidas. Para garantir esta condição, devem empregar-se provas estatísticas de hipótese para calcular intervalos de confiança para os valores estimados da variável dependente. Na prática estas provas ignoram-se, quer por uma enganosa aparência de correlação que surge dos dados experimentais, quer pela impossibilidade de calcular intervalos de confiança para as variáveis independentes consideradas. Isto gera condições que conduzem à correlação espúria. O uso indiscriminado das técnicas de análise

dimensional já mencionadas, baseadas no teorema de Buckingham, para calcular parâmetros adimensionais sem contemplar a razão física que deve ser considerada para estabelecer uma relação válida causa-efeito, conduz a formulações incorretas que causam o problema da correlação espúria. Um valor alto do coeficiente de correlação associado à análise estatística não é suficiente para estabelecer uma relação causa-efeito.

Além dos argumentos acima expostos, freqüentemente podem aparecer variáveis independentes repetidas mesmo nos parâmetros dependentes que nos independentes numa relação proposta, cuja presença introduz uma correlação estatística que é fictícia, e que é conhecida como correlação espúria.

O primeiro estudo conhecido na literatura técnica sobre o tema da correlação espúria foi escrito por Pearson em 1896, aplicado ao estudo da biologia. Neste estudo o autor estabelece que se houver duas funções $u = f_1(x, y), v = f_2(z, y)$ que contém três variáveis x, y, z selecionadas aleatoriamente, de modo a não ter correlação entre x, y , entre y, z , ou entre z, x , ainda pode calcular-se uma correlação entre u e v se se formarem as frações

$\frac{x}{y}, \frac{z}{y}$ para o terno de variáveis (Pearson, 1896). Porém, o coeficiente de correlação obtido é

espúrio, desde que foi obtido a partir de variáveis com distribuição aleatória. Pearson oferece um exemplo de sua hipótese ilustrando um caso no qual obtém uma quantidade de ossos de um *ossuarium*, os quais são arranjados por grupos procedentes de esqueletos

individuais. Então se formam os índices $\frac{fémur}{húmero}, \frac{tibia}{húmero}$. Aparentemente, pode se esperar

que exista uma relação orgânica entre estes índices desde que pode calcular-se o coeficiente de correlação entre eles. Porém, devido às características antropométricas do esqueleto

humano, Pearson conclui que o mesmo valor do coeficiente de correlação pode ser obtido se os ossos tivessem sido sorteados em forma completamente aleatória. Por esta razão, Pearson define a correlação orgânica espúria como a correlação que se encontra entre índices quando os valores absolutos dos órgãos têm sido selecionados em forma completamente aleatória: *“I shall accordingly define spurious organic correlation as the correlation which will be found between indices, when the absolute values of the organs have been selected purely at random”*.

No campo da medicina o uso de índices é muito comum. Por exemplo, o índice provavelmente mais utilizado na pesquisa médica seja o índice de massa corporal (*Body Mass Index, (BMI)*), que se obtém como o quociente do peso de uma pessoa entre a estatura elevada ao quadrado, e que se usa como um índice de obesidade (Kronmal, 1993). Outro exemplo é o índice de volume expiratório forçado (*Forced Expiratory Volume, (FEV1)*) dividido entre a estatura da pessoa elevada ao quadrado usado para determinar padrões específicos dependentes da idade para a função pulmonar para ambos os sexos (Kronmal, 1993). As correlações obtidas entre índices sem considerar as verdadeiras relações causa – efeito entre as variáveis estudadas podem conduzir a conclusões erradas e esquisitas. Por exemplo, cita-se o caso de um estudo no qual se obtém uma correlação entre o índice do número de cegonhas por cada 10 000 mulheres e o índice do número de crianças por 10 000 mulheres para uma dada região, para chegar à conclusão absurda de que o número de cegonhas influi na taxa de natalidade (Kronmal, 1993). Porém, através de uma manipulação diferente dos dados, na qual se relaciona linearmente o número de cegonhas como variável independente com a taxa de natalidade, chega-se a uma conclusão diretamente contrária, quer dizer, o número de cegonhas está inversamente relacionado com a taxa de natalidade (Kronmal, 1993). É óbvio que os índices propostos nesta experiência não têm nenhuma

racionalidade do ponto de vista biológico. Para resolver estes paradoxos, recomenda-se evitar o uso de índices e exprimi-los como parte do modelo de regressão linear através de interações (produtos, por exemplo,) (Kronmal, 1993). Porém, nas cartas dirigidas ao autor (Kronmal, 1993) (*Letters to the Editors*), comenta-se que às vezes, no campo da medicina, há necessidade de usar índices por razões clínicas e também, propõe-se o uso de modelos de função exponencial ou alométricos do tipo de $Y = \alpha X^\beta \varepsilon$, que podem linearizar-se através do uso de logaritmos.

No campo da sociologia é freqüente o uso de índices que têm como denominador o tamanho da população para poder comparar quantidades, apresentando-se também o problema da correlação espúria (Kronmal, 1993).

Chayes (1949) escreveu um estudo sobre correlação espúria aplicado à petrografia. Baseando-se na fórmula geral de Reed (1921) para encontrar o coeficiente de correlação entre dois índices em termos dos coeficientes de correlação das quatro variáveis consideradas e seus coeficientes de variação (Pearson, 1896), Chayes calcula formulações para cinco casos de correlação entre índices que se apresentam em seguida:

Caso 1, correlação entre dois índices com denominador comum, $y = \frac{x_1}{x_2}$, $z = \frac{x_3}{x_2}$

Caso 2, correlação de um índice com uma terceira variável, $y = \frac{x_1}{x_2}$, $z = x_3$

Caso 3, correlação de um índice com seu próprio denominador, $y = \frac{x_1}{x_2}$, $z = x_2$

Caso 4, correlação de um índice com seu próprio numerador, $y = \frac{x_1}{x_2}$, $z = x_1$

Caso 5, correlação entre índices quando o denominador de um é o mesmo que o numerador

$$\text{do outro, } y = \frac{x_1}{x_2}, z = \frac{x_2}{x_4}$$

Em seguida apresentam-se aplicações de formulações obtidas a casos de estudo da petrografia. Chayes comenta que se obteve a formulação de Pearson sob o suposto de que as variáveis individuais consideradas (medições absolutas) caracterizaram-se por estar normalmente distribuídas e que para cada variável o coeficiente de variação (desvio padrão dividido pela média) era suficientemente pequeno para ignorar as potências terceira e maiores do coeficiente. Também se comenta que (Reed, 1921) chegou à mesma formulação que Pearson sem especificar a possível distribuição probabilística das variáveis consideradas. Finalmente Chayes conclui que, ao formar dois índices a partir de quatro variáveis, não haverá correlação entre eles se não houver correlação entre as variáveis, e que, em geral, haverá correlação entre os índices se houver correlação entre uma ou mais duplas das variáveis consideradas (medições absolutas) se cada índice inclui uma variável da dupla. Chayes conclui também que a correlação dos índices não é um critério confiável para avaliar as correlações das variáveis consideradas (medições absolutas).

No campo da estatística, o problema da correlação espúria estuda-se sob o nome de multicolinearidade (Montgomery, 1992), a qual se define como uma dependência quase-linear entre as variáveis de regressão consideradas. A multicolinearidade é um problema sério que pode afetar profundamente a utilidade de um modelo de regressão (Montgomery, 1992). A presença da multicolinearidade pode afetar profundamente a precisão de predição dos coeficientes de regressão estimados para um modelo de regressão proposto (por exemplo, o método dos mínimos quadrados) (Montgomery, 1992). Se houver uma forte

multicolinearidade, os coeficientes de regressão estimados tornam-se muito sensíveis aos dados da população amostrada. Também, as extrapolações que se fazem não serão muito confiáveis (Montgomery, 1992). Conforme Montgomery e Mason citam-se quatro fontes primárias de multicolinearidade, que são: 1) o método usado para coletar dados ou técnica de amostragem (por exemplo, se voluntária ou involuntariamente coletam-se amostras somente de um subespaço do espaço das variáveis de regressão, (Mason, 1975); 2) as restrições físicas que limitam ao modelo ou a população estudada (por exemplo, em química analítica, onde a soma de um número de componentes numa solução deve ser constante embora os valores das componentes individuais variem (Mason, 1975); 3) a eleição do modelo; e 4) um modelo sobre-definido, caracterizado pela existência de mais variáveis de regressão que observações (Mason, 1975). Em seguida examinam-se os efeitos da multicolinearidade sobre o modelo de regressão de mínimos quadrados (Montgomery, 1992), algumas técnicas matemático-estatísticas para detectar a multicolinearidade (Montgomery, 1992), e finalmente discutem-se alguns métodos propostos para remediar a multicolinearidade que são: 1) coleta de dados adicionais (recomendado conforme Montgomery como o melhor método para combater a multicolinearidade); 2) uma nova eleição de modelo; 3) a aplicação da técnica de regressão em cristas (Nagib Elian, 1988) (*ridge regression*) em lugar do método de mínimos quadrados; 4) a aplicação do método generalizado de regressão em cristas (*generalized ridge regression*), que é uma extensão do método anterior; 5) a aplicação do método de regressão das componentes principais, 6) a aplicação do método de regressão de raízes latentes, 7) a comparação e avaliação de estimadores com viés. Mason recomenda também fazer uma nova seleção de variáveis para implementar o modelo de regressão.

Na economia, estuda-se a multicolinearidade no campo da econometria na aplicação de modelos de regressão. Na construção de um modelo de regressão múltipla, é preciso estimar os coeficientes individuais de regressão. Se estes coeficientes ou regressores forem ortogonais, podem se fazer inferências estatísticas com relativa facilidade (Montgomery, 1992). Diz-se que os regressores são ortogonais quando não há uma relação linear entre eles. Frequentemente, porém, os regressores não são ortogonais em muitas aplicações. Às vezes esta não ortogonalidade não é séria. Porém, se os regressores estiverem quase perfeitamente relacionados linearmente, as inferências baseadas no modelo correspondente podem estar erradas. Esta dependência linear entre os regressores é uma causa da multicolinearidade. Para obter um bom resultado nas inferências, um modelo linear de regressão deve ter as três seguintes características essenciais: 1) a variável aleatória dependente y deve ter uma distribuição normal de probabilidade; 2) deve ser válida a hipótese de homocedasticidade para o modelo (Nagib Elian, 1988), quer dizer, para todo valor de x , a variância ao redor da reta de regressão deve ser constante ($Var(\varepsilon_i) = constante$). (A condição contrária é denominada heterocedasticidade); 3) o erro deve ser independente para cada valor de x . Uma estratégia para corrigir a multicolinearidade pode ser segundo duas etapas sucessivas: detecção e correção (Farrar, 1967). A multicolinearidade é considerada como uma condição estatística em vez de matemática, e por tanto, o problema é avaliado em termos de sua severidade, mas não em termos de sua existência ou não existência (Farrar, 1967). Propõe-se uma regra empírica para diagnosticar o grau de severidade da multicolinearidade (ou intercorrelação) que consiste em que, dado r_{ij} = correlação de ordem zero entre duas variáveis independentes, e R_y = correlação múltipla entre variáveis dependentes e independentes, a multicolinearidade

não pode ser ignorada se $r_{ij} \geq R_y$ (Farrar, 1967). Considera-se então a multicolinearidade como uma faceta e um sintoma de um planejamento experimental defeituoso, cuja solução requer um aumento dos dados existentes para incluir informação adicional (Farrar, 1967). Com o intuito de diagnosticar a presença da multicolinearidade num conjunto de variáveis X , assim como a localização e o padrão de interdependência entre as variáveis para identificar as variáveis que devem ser eliminadas para melhorar o modelo proposto, que pode ser expresso na forma linear $y = Xb + u$, para o qual, através da regressão de mínimos quadrados obtêm-se os estimadores $b' = (X'X)^{-1}X'y$, propõe-se um roteiro baseado na manipulação da matriz de correlação invertida $(X'X)^{-1}$, (Farrar, 1967). Ainda outro enfoque do problema de multicolinearidade pode ser consultado em Leamer (1973), onde oferece-se uma interpretação Bayesiana, segundo a qual a principal fonte de multicolinearidade radica (tem origem) na base de dados disponíveis. Porém, o enfoque proposto nesta referência é somente matemático, sem se propor um algoritmo para implementar a solução.

Segundo Benjamin (1970), estuda-se a aplicação da probabilidade e estatística a problemas de engenharia civil, fazendo-se uma breve alusão ao problema de correlação espúria, a qual pode conduzir a conclusões erradas acerca de uma relação de causa-efeito entre variáveis quando, de fato, não existe nenhuma. Uma fonte comum de correlação espúria provém de procurar normalizar os dados experimentais ao dividir-lhes por um fator que é propriamente uma variável aleatória. Por exemplo, se se tem duas variáveis aleatórias independentes X e

Y , e se formar os índices $\frac{X}{Z}$ e $\frac{Y}{Z}$ que não são independentes, uma correlação alta entre estes índices não significa que há correlação entre as variáveis X e Y .

No campo da hidráulica e da hidrologia também se encontram casos de existência de correlação espúria (Benson, 1965). Na maioria dos casos, desenvolvem-se índices com variáveis comuns com o intuito de formar quocientes adimensionais que facilitem a construção de relações empíricas que podem ser representadas graficamente para propósitos de predições, interpolações ou extrapolações, ou para tentar relacionar variáveis para a formação de grupos adimensionais que possam conduzir à expressão numa lei física através do teorema de Buckingham. Porém, às vezes não se conhecem as conseqüências do tipo particular de manipulação dos dados. Por exemplo, se as variáveis empregadas estão correlacionadas em algum grau, haverá um grau maior de correlação entre os índices formados com essas variáveis que o grau de correlação entre as variáveis originais (Benson, 1965). Por isto a correlação espúria também pode definir-se como a mudança aparente do valor do coeficiente de correlação entre variáveis que se apresenta quando estas variáveis transformam-se em parâmetros adimensionais que incluem variáveis comuns. A análise dimensional é um campo potencial para a proliferação da correlação espúria. O uso de índices ou grupos adimensionais para construir gráficos não é uma prática errada por ela mesma, e também os coeficientes de correlação calculados entre estes grupos não são errados, desde que a correlação entre estes grupos se interprete somente em termos de grupos (ou somas), mas não em termos de variáveis individuais. O fato de um gráfico adimensional mostrar uma dispersão pequena não é evidencia suficiente de uma boa relação entre elementos individuais dos grupos (por exemplo, o fato de existir uma relação linear entre os grupos adimensionais correlacionados não implica que exista relação linear entre

as variáveis individuais consideradas para os grupos, (Benson, 1965), e também não é uma indicação da confiabilidade de uma estimação de um elemento individual feita através desta relação. Embora os gráficos adimensionais representem relações mais generalizadas que os gráficos dimensionados, e que os primeiros possam alongar-se com maior confiabilidade além do intervalo dos dados, para um gráfico com grupos adimensionais que tenham elementos aleatórios comuns, existem as seguintes possibilidades: 1) uma dispersão pequena aparente dos pontos que definem a relação é causada somente pela presença das variáveis comuns; 2) a forma geométrica da relação está baseada parcial ou totalmente nas características das variáveis comuns; 3) há pouca ou nenhuma confiabilidade na extrapolação da relação além do intervalo dos dados (Benson, 1965). Na seqüência, apresentam-se exemplos de casos de estudo onde se observa o uso de índices ou parâmetros adimensionais com variáveis comuns.

Conforme Hely (1963) tenta-se estabelecer uma relação entre a vazão de uma bacia e diferentes fatores ambientais. Chega-se á conclusão que a precipitação média é a principal influência meteorológica sobre o valor da vazão e que as características geológicas da bacia representam a principal influência geomorfológica. Esta última influência estima-se com o

parâmetro $\frac{Q_{90}}{Q_a}$, onde Q_{90} é a vazão que é igualada ou excedida 90 por cento das vezes e

Q_a é a vazão média na bacia. Logo se constrói uma relação gráfica entre $\frac{Q_{90}}{Q_a}$ para a vazão

diária e $\frac{Q_{90}}{Q_a}$ para a vazão mensal com o intuito de estimar Q_{90} diário. Outros objetivos

procurados com o índice $\frac{Q_{90}}{Q_a}$ são que os valores calculados para diferentes estações hidrométricas sejam mais comparáveis, que o efeito do tamanho da bacia (área) possa ser removido de uma maneira análoga que se usar a grandeza de vazão por unidade de área, e que possa reduzir-se o efeito das diferenças na precipitação média.

Em Yalin (1964) apresenta-se um estudo de transporte de sedimento, no qual estabelece-se uma relação gráfica entre dois índices formados com diversas variáveis aleatórias com o intuito de estabelecer a forma da função estudada:

$$\frac{\Delta}{d_{cr}} \approx \phi \left(\frac{d}{d_{cr}} \right)$$

onde

Δ = altura da onda de areia

d = profundidade do escoamento

d_{cr} = profundidade crítica

Porém, não há certeza da existência de uma relação entre Δ e d , e a forma da função pode depender somente das propriedades de d_{cr} (Benson, 1965).

Com o intuito de estudar a forma e as propriedades das ondas de areia formadas no leito do escoamento, estabelece-se uma relação gráfica experimental de laboratório entre dois índices (Yalin, 1964) que são:

$$\frac{d}{D}, \frac{d}{D} \sqrt{\frac{\Lambda}{D}}$$

onde

D = tamanho de grão

Λ = comprimento da onda de areia

Em Yevdjovich (1963), com o intuito de estudar as relações chuva-escoamento, apresenta-se a seguinte relação gráfica:

$$\frac{W_i}{V_{prom}} , \frac{P_e}{P_{e-prom}}$$

que, devido às definições dos termos envolvidos, equivale a:

$$\frac{\Delta W}{V_{prom}} , \frac{(V + \Delta W)}{V_{prom}}$$

onde

W_i = armazenamento da água por ano

ΔW = variável aleatória que representa a mudança de armazenamento da água por ano

V = variável aleatória que representa o escoamento direto anual

P_e = altura efetiva de chuva

P_{e-prom} = altura efetiva média de chuva

V_{prom} = escoamento direto anual médio

Segundo Yevdjovich (1963) conclui-se que, devido à existência de uma boa correlação gráfica entre os índices considerados, ΔW deve ser uma função dos valores anteriores de

P_{e-prom} ou de V_{prom} . Porém, esta conclusão é contestada pelo fato dos índices considerados

terem um denominador comum (V_{prom}) que é constante, e pelo fato de que ΔW está presente em ambos índices, o que contribui para a bondade do ajuste, mas não pode se ter uma conclusão certa (Benson, 1965).

Em (Mirajgoker, 1963), apresenta-se um estudo do efeito natural da rugosidade em canais revestidos. Nesta experiência se fizeram experimentos com um canal de laboratório de concreto com pedras coladas no fundo para simular a rugosidade segundo um padrão

definido de número de pedras por jarda quadrada. Da experiência obteve-se uma relação

gráfica entre o número de Froude F e a função de resistência proposta $\frac{C}{\sqrt{g}}$, obtida a partir da fórmula de Chezy através de análise dimensional e expressa como equação 7 (Mirajgaoker, 1963). Porém, sabe-se que por definição

$$\frac{C}{\sqrt{g}} = \frac{F}{\sqrt{S}}$$

onde

C = coeficiente de rugosidade de Chezy

F = número de Froude

S = declividade da linha de energia

então, a relação

$$F, \frac{C}{\sqrt{g}}$$

equivale à relação

$$F, \frac{F}{\sqrt{S}}$$

e por tanto, a relação gráfica proposta é espúria, desde que equivale a grafar F contra F (Mirajgaoker, 1963).

A contribuição nova ao campo da hidráulica desta tese é o estabelecimento de uma metodologia estatística e experimental de caráter universal que permita determinar se um tipo de relação algébrica que pretende usar-se para relacionar dois grupos adimensionais propostos é aconselhável ou deve ser substituída por outro modelo algébrico, de acordo

com o valor observado do coeficiente de correlação espúria que se calcula para um modelo algébrico em forma geral.

2.3 Conclusões

Como resultado de uma revisão da bibliografia técnica disponível, obteve-se a fórmula de Pearson (Reed, 1921), a qual permite avaliar os coeficientes de correlação requeridos para obter o grau de correlação espúria presente num problema específico de correlação de variáveis, segundo a metodologia já exposta. Se concluiu que até hoje não tem havido modificações substanciais na conceituação matemática do fenômeno de correlação espúria desde que este foi estudado com mais profundidade na década de sessenta. Também se observou que o tema de correlação espúria tem sido pouco estudado pelos pesquisadores e geralmente seus efeitos são ignorados na engenharia experimental aplicada à hidráulica, quer por desconhecimento, quer por negligência.

CAPÍTULO 3

A FÓRMULA DE PEARSON PARA CALCULAR O COEFICIENTE DE CORRELAÇÃO ESPÚRIA

3.1 Introdução

Em termos simples, a correlação matemática pode ser descrita como o grau de associação de um conjunto de variáveis. A correlação espúria caracteriza-se por um coeficiente que mede o grau de associação de combinações de funções aleatórias independentes que têm variáveis comuns entre si. O fenômeno da correlação espúria foi identificado pela primeira vez por Karl Pearson ao final do século passado (Pearson, 1896). Posteriormente, diversos autores escreveram trabalhos focalizados à experimentação na engenharia hidráulica (ou mesmo para outras disciplinas, (Chayes, 1949) nos que aparecem exemplos da correlação espúria (Benson, 1965; Hely, 1963; Mirajgaoker, 1963; Yalin, 1964; Yevdjovich, 1963). Para atingir os objetivos deste trabalho, é necessário desenvolver um programa de computador que permita calcular o grau de espureidade de uma correlação entre dois grupos de variáveis concordando com o seguinte roteiro (Bremauntz, 1974):

- a) Calcula-se o coeficiente de correlação de Pearson r_{yz} para as duas funções aleatórias em estudo.
- b) Calcula-se a correlação totalmente espúria ρ_{yz} para estas funções. Este coeficiente de correlação espúria obtém-se recalculando o coeficiente de correlação de Pearson para os seguintes valores do coeficiente de correlação entre funções: $r_{ij} = 0$ para $i \neq j$ e $r_{ii} = 1$.

c) Finalmente, calcula-se o grau de espureidade da correlação proposta que se exprime

como segue: $GE = 100 \frac{\rho_{yz}}{r_{yz}}$ que indica a porcentagem de espureidade na correlação.

O coeficiente de correlação de Pearson entre duas funções $y = C_1 X_1^p X_2^q X_3^r X_4^s$ e

$z = C_2 X_5^t X_6^u X_7^v X_8^w$ exprime-se pela seguinte equação (1) (Reed, 1921):

$$r_{yz} = \frac{\sum_{a=1}^4 \sum_{b=5}^8 f_a g_b r_{ab} \sigma_a \sigma_b}{\left(\sum_{a=1}^4 \sum_{c=1}^4 f_a f_c r_{ac} \sigma_a \sigma_c \right)^{1/2} \left(\sum_{b=5}^8 \sum_{d=5}^8 g_b g_d r_{bd} \sigma_b \sigma_d \right)^{1/2}} \quad (1)$$

$$f_a = \frac{\delta}{\delta m_a} f(m_1, m_2, m_3, m_4) \quad a = 1, \dots, 4$$

$$g_b = \frac{\delta}{\delta m_b} g(m_5, m_6, m_7, m_8) \quad b = 5, \dots, 8$$

onde

$X_1, X_2, X_3, \dots, X_8$ variáveis aleatórias das funções

C_1, C_2 coeficientes das funções

p, q, r, s, t, u, v, w expoentes das variáveis

m_1, m_2, \dots, m_8 médias das variáveis

$\sigma_1, \sigma_2, \dots, \sigma_8$ desvios padrões das variáveis

$r_{12}, r_{13}, \dots, r_{18}, \dots, r_{78}$ coeficientes de correlação entre variáveis

3.2 Exemplo de aplicação da fórmula de Pearson a um caso de estudo

Para a aplicação prática da fórmula de Pearson, considera-se convencionalmente que a função y mencionada anteriormente grafa-se sempre no eixo das ordenadas (eixo vertical) do plano cartesiano, enquanto que a função z também mencionada anteriormente grafa-se sempre no eixo das abscissas (eixo horizontal) do plano cartesiano, o que se ilustra na Figura 1:

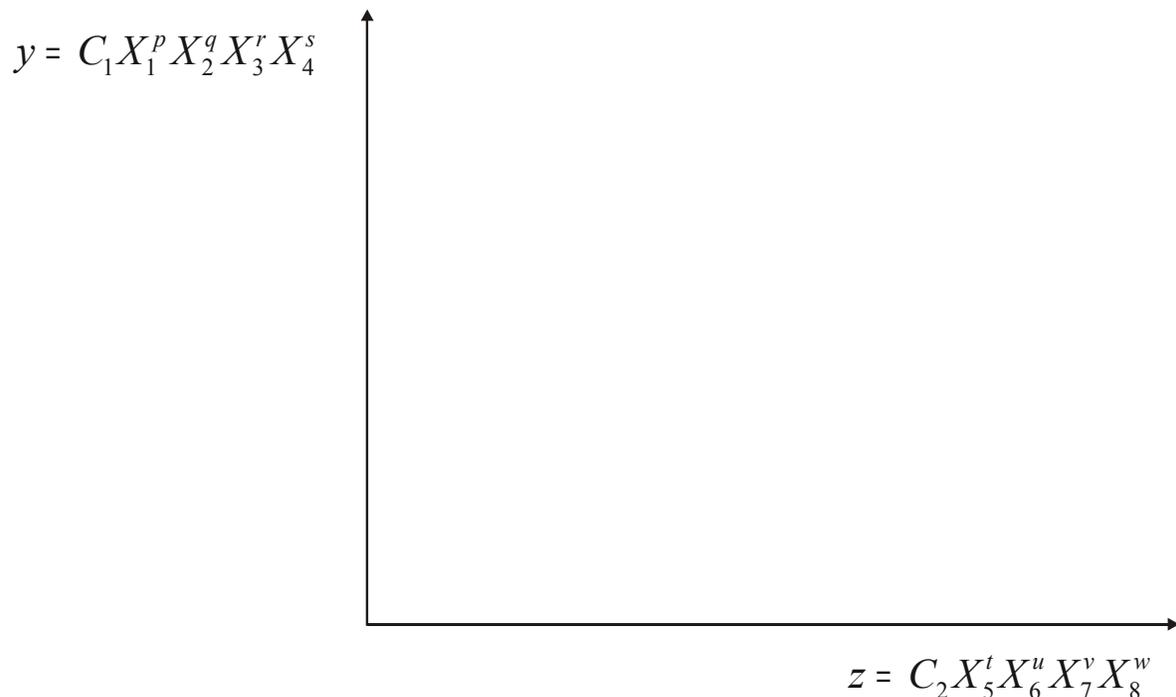


Figura 1.- Disposição das variáveis y e z no plano cartesiano.

No desenvolvimento desta tese, consideram-se dois tipos genéricos de relações algébricas entre as duas funções y e z consideradas: relações expressas como quocientes e relações expressas como produtos. Na Tabela 1 (diversos modelos algébricos propostos em forma de quociente) apresenta-se uma amostra de relações expressas como quocientes para as

funções y e z de uso mais freqüente na engenharia, enquanto que na Tabela 2 (Diversos modelos algébricos propostos em forma de produto) apresenta-se uma amostra de relações expressas como produtos para as ditas funções também de uso muito freqüente na engenharia.

Como caso de estudo, considera-se o seguinte modelo algébrico que posteriormente é incluído na Tabela 1 deste Capítulo (modelo número 5 da tabela):

$$y = \frac{x_1}{x_2}$$

$$z = \frac{x_3}{x_2}$$

Calcula-se o numerador para aplicar a equação (1):

$$\text{para } y = \frac{x_1}{x_2} \quad \text{tem-se} \quad f(m_1, m_2) = \frac{m_1}{m_2}$$

$$\text{para } z = \frac{x_3}{x_2} \quad \text{tem-se} \quad g(m_3, m_2) = \frac{m_3}{m_2}$$

então

$$f_1 = \frac{\partial f}{\partial m_1} = \frac{\partial}{\partial m_1} \left(\frac{m_1}{m_2} \right) = \frac{1}{m_2}$$

$$f_2 = \frac{\partial f}{\partial m_2} = \frac{\partial}{\partial m_2} \left(\frac{m_1}{m_2} \right) = -\frac{m_1}{m_2^2}$$

$$g_3 = \frac{\partial g}{\partial m_3} = \frac{\partial}{\partial m_3} \left(\frac{m_3}{m_2} \right) = \frac{1}{m_2}$$

$$g_2 = \frac{\partial g}{\partial m_2} = \frac{\partial}{\partial m_2} \left(\frac{m_3}{m_2} \right) = -\frac{m_3}{m_2^2}$$

substituindo-se os resultados no numerador da equação (1) para $a = 1,2$ $b = 2,3$

$$\sum_{a=1}^4 \sum_{b=5}^8 f_a g_b r_{ab} \sigma_a \sigma_b = f_1 g_2 r_{12} \sigma_1 \sigma_2 + f_1 g_3 r_{13} \sigma_1 \sigma_3 + f_2 g_2 r_{22} \sigma_2 \sigma_2 + f_2 g_3 r_{23} \sigma_2 \sigma_3 =$$

$$- \left(\frac{m_3}{m_2^3} \right) r_{12} \sigma_1 \sigma_2 + \left(\frac{1}{m_2^2} \right) r_{13} \sigma_1 \sigma_3 + \left(\frac{m_1 m_3}{m_2^4} \right) r_{22} \sigma_2^2 - \left(\frac{m_1}{m_2^3} \right) r_{23} \sigma_2 \sigma_3$$

multiplicando-se a equação por $\frac{m_2^2}{m_1 m_3}$ e substituindo os valores de

$$C_1 = \frac{\sigma_1}{m_1} \quad C_2 = \frac{\sigma_2}{m_2} \quad C_3 = \frac{\sigma_3}{m_3} \quad C_4 = \frac{\sigma_4}{m_4}$$

tem-se

$$\begin{aligned} & - \left(\frac{1}{m_2 m_1} \right) r_{12} \sigma_1 \sigma_2 + \left(\frac{1}{m_1 m_3} \right) r_{13} \sigma_1 \sigma_3 + \left(\frac{1}{m_2^2} \right) r_{22} \sigma_2^2 - \left(\frac{1}{m_2 m_3} \right) r_{23} \sigma_2 \sigma_3 \\ & = - r_{12} \left(\frac{\sigma_1}{m_1} \right) \left(\frac{\sigma_2}{m_2} \right) + r_{13} \left(\frac{\sigma_1}{m_1} \right) \left(\frac{\sigma_3}{m_3} \right) + r_{22} \left(\frac{\sigma_2}{m_2} \right)^2 - r_{23} \left(\frac{\sigma_2}{m_2} \right) \left(\frac{\sigma_3}{m_3} \right) \\ & \quad - r_{12} C_1 C_2 + r_{13} C_1 C_3 + C_2^2 r_{22} - r_{23} C_2 C_3 \end{aligned}$$

Considerando-se o primeiro termo do denominador para $a = 1,2$ $c = 1,2$

$$\sum_{a=1}^4 \sum_{c=1}^4 f_a f_c r_{ac} \sigma_a \sigma_c = f_1 f_1 r_{11} \sigma_1 \sigma_1 + f_1 f_2 r_{12} \sigma_1 \sigma_2 + f_2 f_1 r_{21} \sigma_2 \sigma_1 + f_2 f_2 r_{22} \sigma_2 \sigma_2 =$$

Substituindo-se os valores correspondentes das derivadas e multiplicando-se a equação por

$$\frac{m_2^2}{m_1^2}, \text{ tem-se:}$$

$$r_{11} \left(\frac{\sigma_1^2}{m_1^2} \right) - r_{12} \left(\frac{\sigma_1}{m_1} \right) \left(\frac{\sigma_2}{m_2} \right) - r_{21} \left(\frac{\sigma_1}{m_1} \right) \left(\frac{\sigma_2}{m_2} \right) + r_{22} \left(\frac{\sigma_2^2}{m_2^2} \right) =$$

$$r_{11}C_1^2 - r_{12}C_1C_2 - r_{21}C_1C_2 + r_{22}C_2^2$$

Considerando-se o segundo termo do denominador para $b = 2,3$ $d = 2,3$, tem-se:

$$\sum_{b=1}^4 \sum_{d=1}^4 g_b g_d r_{bd} \sigma_b \sigma_d = g_2 g_2 r_{22} \sigma_2 \sigma_2 + g_2 g_3 r_{23} \sigma_2 \sigma_3 + g_3 g_2 r_{32} \sigma_3 \sigma_2 + g_3 g_3 r_{33} \sigma_3 \sigma_3$$

Substituindo-se os valores correspondentes das derivadas e multiplicando-se a equação por $\frac{m_2^2}{m_3^2}$ tem-se:

$$r_{22} \left(\frac{\sigma_2^2}{m_2^2} \right) - r_{23} \left(\frac{\sigma_2}{m_2} \right) \left(\frac{\sigma_3}{m_3} \right) - r_{32} \left(\frac{\sigma_2}{m_2} \right) \left(\frac{\sigma_3}{m_3} \right) + r_{33} \frac{\sigma_3^2}{m_3^2} =$$

$$r_{22}C_2^2 - r_{23}C_2C_3 + r_{33}C_3^2$$

Substituindo os resultados obtidos na equação (1):

$$r_{yz} = \frac{(-r_{12}C_1C_2 + r_{13}C_1C_3 + r_{22}C_2^2 - r_{23}C_2C_3)}{(r_{11}C_1^2 - r_{12}C_1C_2 - r_{21}C_1C_2 + r_{22}C_2^2)^{1/2} (r_{22}C_2^2 - r_{23}C_2C_3 - r_{32}C_3C_2 + r_{33}C_3^2)^{1/2}} \quad (2)$$

Se considerarmos que $r_{11} = r_{22} = r_{33} = 1$ e que $r_{12} = r_{21}$ e $r_{23} = r_{32}$, obtém-se (Chayes, 1949)

$$r_{yz} = \frac{r_{13}C_1C_3 - r_{12}C_1C_2 - r_{23}C_2C_3 + C_2^2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2} (C_2^2 + C_3^2 - 2r_{23}C_2C_3)^{1/2}} \quad (3)$$

A equação (3) exprime o coeficiente de correlação total entre as duas funções consideradas.

Pode-se calcular o coeficiente de correlação espúria considerando-se a condição de $r_{ij} = 0$

para $i \neq j$ e $r_{ii} = 1$. Por tanto, tem-se

$$r_{12} = r_{13} = r_{23} = r_{21} = r_{32} = 0$$

$$r_{22} = r_{11} = r_{33} = 1$$

então (Chayes, 1949)

$$\rho_{yz} = \frac{C_2^2}{(C_1^2 + C_2^2)^{1/2}(C_2^2 + C_3^2)^{1/2}} \quad (3)$$

equação que exprime o coeficiente de correlação espúria. A equação (3) pode ser obtida diretamente da equação (1) utilizando-se o mesmo procedimento matemático empregado para obter a equação (3).

Em continuação apresenta-se uma pequena tabela que ilustra os valores dos coeficientes de correlação para alguns modelos algébricos.

Tabela 1.- Modelos algébricos propostos em forma de quociente.

Modelo	Correlação total r_{yz}	Correlação espúria ρ_{yz} $r_{ij} = 0 \quad i \neq j$ $r_{ii} = 1$
$y = \frac{x_1}{x_2}$ $z = \frac{x_3}{x_4}$	$\frac{r_{13}C_1C_3 - r_{14}C_1C_4 - r_{23}C_2C_3 + r_{24}C_2C_4}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_3^2 + C_4^2 - 2r_{34}C_3C_4)^{1/2}}$	$\rho_{yz} = 0$
$y = \frac{x_1}{x_2}$ $z = x_3$	$\frac{r_{13}C_1 - r_{23}C_2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}}$	$\rho_{yz} = 0$
$y = \frac{x_1}{x_2}$ $z = \frac{x_1}{x_3}$	$\frac{C_1^2 - r_{13}C_1C_3 - r_{12}C_1C_2 + r_{23}C_2C_3}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_1^2 + C_3^2 - 2r_{13}C_1C_3)^{1/2}}$	$\frac{C_1^2}{(C_1^2 + C_2^2)^{1/2}(C_1^2 + C_3^2)^{1/2}}$ $\rho_{yz} > 0$
$y = \frac{x_1}{x_2}$ $z = \frac{x_1}{x_4}$	$\frac{C_1^2 - r_{14}C_1C_4 - r_{12}C_1C_2 + r_{24}C_2C_4}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_1^2 + C_4^2 - 2r_{14}C_1C_4)^{1/2}}$	$\frac{C_1^2}{(C_1^2 + C_2^2)^{1/2}(C_1^2 + C_4^2)^{1/2}}$ $\rho_{yz} > 0$
$y = \frac{x_1}{x_2}$ $z = \frac{x_3}{x_2}$	$\frac{r_{13}C_1C_3 - r_{12}C_1C_2 - r_{23}C_2C_3 + C_2^2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_2^2 + C_3^2 - 2r_{23}C_2C_3)^{1/2}}$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2}(C_2^2 + C_3^2)^{1/2}}$ $\rho_{yz} > 0$
$y = \frac{x_1}{x_2}$ $z = \frac{x_2}{x_4}$	$\frac{r_{12}C_1C_2 - r_{14}C_1C_4 + r_{24}C_2C_4 - C_2^2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_2^2 + C_4^2 - 2r_{24}C_2C_4)^{1/2}}$	$\frac{-C_2^2}{(C_1^2 + C_2^2)^{1/2}(C_2^2 + C_4^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{x_1}{x_2}$ $z = x_1$	$\frac{C_1 - r_{12}C_2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}}$	$\frac{C_1}{(C_1^2 + C_2^2)^{1/2}}$ $\rho_{yz} > 0$

Modelo	Correlação total r_{yz}	Correlação espúria ρ_{yz}
		$r_{ij} = 0 \quad i \neq j$ $r_{ii} = 1$
$y = \frac{x_1}{x_2}$ $z = x_2$	$\frac{r_{12}C_1 - C_2}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}}$	$\frac{-C_2}{(C_1^2 + C_2^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{1}{x_2}$ $z = \frac{x_2}{x_4}$	$\frac{r_{24}C_4 - C_2}{(C_2^2 + C_4^2 - 2r_{24}C_2C_4)^{1/2}}$	$\frac{-C_2}{(C_2^2 + C_4^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{x_5}{x_4}$ $z = \frac{x_6}{x_4}$	$\frac{r_{56}C_5C_6 - r_{54}C_5C_4 - r_{46}C_4C_6 + C_4^2}{(C_5^2 + C_4^2 - 2r_{54}C_5C_4)^{1/2}(C_4^2 + C_6^2 - 2r_{46}C_4C_6)^{1/2}}$	$\frac{C_4^2}{(C_5^2 + C_4^2)^{1/2}(C_4^2 + C_6^2)^{1/2}}$ $\rho_{yz} > 0$
$y = \frac{x_7}{x_8}$ $z = \frac{x_8}{x_9}$	$\frac{r_{78}C_7C_8 - r_{79}C_7C_9 + r_{89}C_8C_9 - C_8^2}{(C_7^2 + C_8^2 - 2r_{78}C_7C_8)^{1/2}(C_8^2 + C_9^2 - 2r_{89}C_8C_9)^{1/2}}$	$\frac{-C_8^2}{(C_7^2 + C_8^2)^{1/2}(C_8^2 + C_9^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{x_{10}}{x_{11}}$ $z = x_{10}$	$\frac{C_{10} - r_{1011}C_{11}}{(C_{10}^2 + C_{11}^2 - 2r_{1011}C_{10}C_{11})^{1/2}}$	$\frac{C_{10}}{(C_{10}^2 + C_{11}^2)^{1/2}}$ $\rho_{yz} > 0$
$y = \frac{x_{12}}{x_{13}}$ $z = x_{13}$	$\frac{r_{1213}C_{12} - C_{13}}{(C_{12}^2 + C_{13}^2 - 2r_{1213}C_{12}C_{13})^{1/2}}$	$\frac{-C_{13}}{(C_{12}^2 + C_{13}^2)^{1/2}}$ $\rho_{yz} < 0$

Modelo	Correlação total r_{yz}	Correlação espúria ρ_{yz}
		$r_{ij} = 0 \quad i \neq j$ $r_{ii} = 1$
$y = \frac{1}{x_{14}}$ $z = \frac{x_{14}}{x_{15}}$	$\frac{r_{1415}C_{15} - C_{14}}{(C_{14}^2 + C_{15}^2 - 2r_{1415}C_{14}C_{15})^{1/2}}$	$\frac{-C_{14}}{(C_{14}^2 + C_{15}^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{x_{16}}{x_{17}}$ $z = x_{17}x_{18}$	$\frac{r_{1617}C_{16}C_{17} + r_{1618}C_{16}C_{18} - C_{17}^2 - r_{1718}C_{17}C_{18}}{(C_{16}^2 + C_{17}^2 - 2r_{1617}C_{16}C_{17})^{1/2}(C_{17}^2 + C_{18}^2 + 2r_{1718}C_{17}C_{18})^{1/2}}$	$\frac{-C_{17}^2}{(C_{16}^2 + C_{17}^2)^{1/2}(C_{17}^2 + C_{18}^2)^{1/2}}$ $\rho_{yz} < 0$
$y = \frac{x_{19}}{x_{20}^p}$ $z = x_{20}$	$\frac{r_{1920}C_{19}C_{20} - pC_{20}^2}{(C_{19}^2 + p^2C_{20}^2 - 2pr_{1920}C_{19}C_{20})^{1/2}(C_{20}^2)^{1/2}}$	$\frac{-pC_{20}^2}{(C_{19}^2 + p^2C_{20}^2)^{1/2}(C_{20}^2)^{1/2}}$ $\rho_{yz} < 0$

Tabela 2.- Diversos modelos algébricos propostos em forma de produto.

Modelo	Correlação total r_{yz}	Correlação espúria ρ_{yz} $r_{ij} = 0 \quad i \neq j$ $r_{ii} = 1$
$y = x_1 x_2$ $z = x_3 x_4$	$\frac{r_{13}C_1C_3 + r_{14}C_1C_4 + r_{23}C_2C_3 + r_{24}C_2C_4}{(C_1^2 + C_2^2 + 2r_{12}C_1C_2)^{1/2}(C_3^2 + C_4^2 + 2r_{34}C_3C_4)^{1/2}}$	$\rho_{yz} = 0$
$y = x_1 x_2$ $z = x_3 x_2$	$\frac{r_{13}C_1C_3 + r_{12}C_1C_2 + r_{23}C_2C_3 + C_2^2}{(C_1^2 + C_2^2 + 2r_{12}C_1C_2)^{1/2}(C_2^2 + C_3^2 + 2r_{23}C_2C_3)^{1/2}}$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2}(C_2^2 + C_3^2)^{1/2}}$ $\rho_{yz} > 0$
$y = x_1 x_2$ $z = x_3$	$\frac{r_{13}C_1 + r_{23}C_2}{(C_1^2 + C_2^2 + 2r_{12}C_1C_2)^{1/2}}$	$\frac{r_{23}C_2}{(C_1^2 + C_2^2 + 2r_{12}C_1C_2)^{1/2}}$ $\rho_{yz} > 0$
$y = x_1 x_2$ $z = x_2$	$\frac{r_{12}C_1 + C_2}{(C_1^2 + C_2^2 + 2r_{12}C_1C_2)^{1/2}}$	$\frac{C_2}{(C_1^2 + C_2^2)^{1/2}}$ $\rho_{yz} > 0$
$y = x_{21} x_{22}$ $z = x_{23} x_{24}$	$\frac{r_{2123}C_{21}C_{23} + r_{2124}C_{21}C_{24} + r_{2223}C_{22}C_{23} + r_{2224}C_{22}C_{24}}{(C_{21}^2 + C_{22}^2 + 2r_{2122}C_{21}C_{22})^{1/2}(C_{23}^2 + C_{24}^2 + 2r_{2324}C_{23}C_{24})^{1/2}}$	$\rho_{yz} = 0$

Modelo	Correlação total r_{yz}	Correlação espúria ρ_{yz} $r_{ij} = 0 \quad i \neq j$ $r_{ii} = 1$
$y = x_{25}x_{26}$ $z = x_{27}x_{26}$	$\frac{r_{2527}C_{25}C_{27} + r_{2526}C_{25}C_{26} + r_{2627}C_{26}C_{27} + C_{26}^2}{(C_{25}^2 + C_{26}^2 + 2r_{2526}C_{25}C_{26})^{1/2}(C_{27}^2 + C_{26}^2 + 2r_{2726}C_{27}C_{26})^{1/2}}$	$\frac{C_{26}^2}{(C_{25}^2 + C_{26}^2)^{1/2}(C_{27}^2 + C_{26}^2)^{1/2}}$ $\rho_{yz} > 0$
$y = x_{28}x_{29}$ $z = x_{30}$	$\frac{r_{2830}C_{28} + r_{2930}C_{29}}{(C_{28}^2 + C_{29}^2 + 2r_{2829}C_{28}C_{29})^{1/2}}$	$\rho_{yz} = 0$
$y = x_{31}x_{32}$ $z = x_{32}$	$\frac{r_{3132}C_{31} + C_{32}}{(C_{31}^2 + C_{32}^2 + 2r_{3132}C_{31}C_{32})^{1/2}}$	$\frac{C_{32}}{(C_{31}^2 + C_{32}^2)^{1/2}}$ $\rho_{yz} > 0$

A técnica de correlação espúria tem sido ilustrada com exemplos simples, que contemplam variáveis com expoente um. Embora seja possível calcular matematicamente o coeficiente de correlação espúria com a equação (1) derivando as funções envolvidas segundo as regras do cálculo, é evidente que este labor pode complicar-se muito para funções muito envolvidas. No desenvolvimento deste trabalho, a equação (1) será resolvida numericamente por meio dum programa de computador para calcular as derivadas numericamente.

Por exemplo, para $f(m) = C_1 m_1^p m_2^q m_3^r m_4^s$

$$\text{tem-se } \frac{\partial f(m)}{\partial m_1} = C_1 p m_1^{p-1} m_2^q m_3^r m_4^s$$

3.3 Exemplo numérico simples de aplicação

Sejam as funções

$$y = \frac{x_1}{x_2} \quad z = \frac{x_3}{x_4}$$

com os seguintes dados:

Médias	Desvios padrão	Coefficientes de variação
$m_1 = 10$	$\sigma_1 = 2,0$	$C_1 = 0,20$
$m_2 = 12$	$\sigma_2 = 3,6$	$C_2 = 0,30$
$m_3 = 11$	$\sigma_3 = 4,4$	$C_3 = 0,40$
$m_4 = 13$	$\sigma_4 = 6,5$	$C_4 = 0,50$

Coefficientes de correlação		
$r_{12} = 0,90$	$r_{13} = 0,91$	$r_{14} = 0,92$
$r_{23} = 0,93$	$r_{24} = 0,94$	$r_{34} = 0,95$

Calcular a correlação total r_{yz} , a correlação espúria ρ_{yz} e o grau de espureidade $GE = \frac{\rho_{yz}}{r_{yz}}$.

Usam-se então as fórmulas correspondentes da tabela 1 (modelo número 1 da tabela). Desta tabela, pode-se dizer antecipadamente que o coeficiente de correlação espúria ρ_{yz} é zero pelo fato de não ter uma variável comum nas duas funções.

Considerando então a fórmula correspondente,

$$r_{yz} = \frac{r_{13}C_1C_3 - r_{14}C_1C_4 - r_{23}C_2C_3 + r_{24}C_2C_4}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2}(C_3^2 + C_4^2 - 2r_{34}C_3C_4)^{1/2}}$$

Substituindo-se, obtém-se:

$$r_{yz} = \frac{[0,91(0,20)(0,40) - 0,92(0,20)(0,50) - 0,93(0,30)(0,40) + 0,94(0,30)(0,50)]}{[(0,20^2 + 0,30^2 - 2(0,90)(0,20)(0,30))^{1/2}(0,40^2 + 0,50^2 - 2(0,95)(0,40)(0,50))^{1/2}}$$

$$r_{yz} = \frac{0,0102}{(0,1483)(0,1732)} = \frac{0,0102}{0,0257} = 0,3971$$

Então,

$$GE = \frac{0}{0,3971} = 0$$

Como $GE = 0$, significa que não há correlação espúria entre as funções consideradas. Se for obtido um valor $G = 100\%$, isso significa que a correlação entre funções proposta reduz-se somente a uma autocorrelação linear da variável repetida, o que pode ser expresso apenas como uma reta desenhada a 45° no plano cartesiano. O grau de espureidade cresce proporcionalmente com o valor absoluto de GE , o qual varia de 0 a 100%. Não existe um valor limite de GE que possa ser definido analiticamente para rejeitar uma correlação proposta. O pesquisador deverá decidir se aceita ou rejeita a dita correlação, baseado numa avaliação criteriosa que inclua todos os fatores a serem considerados.

CAPÍTULO 4

DESENVOLVIMENTO DE UM MODELO COMPUTACIONAL PARA O CÁLCULO DO COEFICIENTE DE CORRELAÇÃO ESPÚRIA

4.1 Introdução

Para facilitar a aplicação da metodologia estatística apresentada no Capítulo 3 a casos de estudo de correlação e poder calcular facilmente os coeficientes de correlação total e espúria requeridos, escreveu-se um programa de computador em FORTRAN, nomeado ESPURI, cuja estrutura se descreve na seqüência. O programa tem duas subrotinas de cálculo e pode imprimir resultados ou apresentá-los na tela no caso de usar um computador pessoal. Descrevem-se, também, os formatos dos arquivos de ingresso de dados e de resultados gerados pelo programa.

4.2 Descrição do programa ESPURI

4.2.1 Arquivo de entrada de dados

Os dados de entrada para o programa são lidos através de um arquivo de dados. O arquivo de dados deve constar de um título, os coeficientes e expoentes das funções y , z que se deseja correlacionar, as correspondentes médias e desvios padrões das variáveis aleatórias consideradas e os coeficientes de correlação calculados pelo usuário que devem ordenar-se numa matriz triangular superior de ordem de 7×7 . O programa admite que cada função tenha até quatro variáveis, que seria o caso mais geral. Se for desejável empregar funções que tenham menos variáveis, então os expoentes associados com essas variáveis se fazem iguais a zero. Por exemplo, se as variáveis X_3 , X_4 , X_7 , X_8 não intervêm nas funções, então os expoentes associados $r = s = v = w = 0$. Portanto para este caso particular e por condições de programação, deve-se fazer $m_3 = m_4 = m_7 = m_8 = 1$ e $\sigma_3, \sigma_4 = \sigma_7 = \sigma_8 = 0$. Os

coeficientes de correlação associados com variáveis não definidas são necessariamente iguais a zero. Utiliza-se no programa a fórmula do coeficiente de correlação de Pearson (equação 1 do Capítulo 3). A subrotina SOL1 (RYZ) calcula o coeficiente r_{yz} e a subrotina SOL2 (ESP) calcula a correlação totalmente espúria ρ_{yz} para a qual se considera $r_{ij} = 0$ para $i \neq j$ e $r_{ij} = 1$ se $i = j$. Finalmente, tanto os dados como os resultados são impressos ou são mostrados na tela do computador no caso de usar computador pessoal, incluindo a

porcentagem de espureidade $GE = 100 \left(\frac{\rho_{yz}}{r_{yz}} \right)$. A ordem de ingresso dos dados para o

programa é como segue, considerando as funções $y = C_1 X_1^p X_2^q X_3^r X_4^s$ e

$$z = C_2 X_5^t X_6^u X_7^v X_8^w :$$

VARIÁVEL	CONCEITO
TÍTULO	Título da rodada
	PARA A FUNÇÃO $y = C_1 X_1^p X_2^q X_3^r X_4^s$
C_1	Constante
p	Primeiro expoente
q	Segundo expoente
r	Terceiro expoente
s	Quarto expoente
	PARA A FUNÇÃO $z = C_2 X_5^t X_6^u X_7^v X_8^w$
C_2	Constante
t	Primeiro expoente

u Segundo expoente

v Terceiro expoente

w Quarto expoente

DADOS ESTATÍSTICOS DAS FUNÇÕES

m_i Médias > 0 , $i = 1 \dots 8$

σ_i Desvios padrões, $i = 1 \dots 8$

MATRIZ TRIANGULAR DE COEFICIENTES DE CORRELAÇÃO

$r_{12}, r_{13}, r_{14}, \dots, r_{18}$

$r_{23}, r_{24}, \dots, r_{28}$

r_{34}, \dots, r_{38}

.....

r_{78}

Os dados de entrada para o programa devem ordenar-se num arquivo segundo o seguinte formato para FORTRAN, no qual se incluem já as variáveis que serão utilizadas no programa. Se usar o processador de palavra Word com ambiente Windows, o nome do arquivo deverá ter a extensão TXT e o arquivo deverá guardar-se sob a opção SOLO TEXTO.

4.2.2 Formato para o arquivo de entrada:

TITULO DA RODADA

C1, P, Q, R, S

C2, T, U, V, W

M (1), SIGMA (1)

M (2), SIGMA (2)

M (3), SIGMA (3)

M (4), SIGMA (4)

M (5), SIGMA (5)

M (6), SIGMA (6)

M (7), SIGMA (7)

M (8), SIGMA (8)

CR (1,2), CR (1,3), CR (1,4), CR (1,5), CR (1,6), CR (1,7), CR (1,8)

CR (2,3), CR (2,4), CR (2,5), CR (2,6), CR (2,7), CR (2,8)

CR (3,4), CR (3,5), CR (3,6), CR (3,7), CR (3,8)

CR (4,5), CR (4,6), CR (4,7), CR (4,8)

CR (5,6), CR (5,7), CR (5,8)

CR (6,7), CR (6,8)

CR (7,8)

4.2.3 Arquivo de saída de resultados

Como resultado da rodada do programa, gerar-se-á um arquivo de resultados no qual se incluem os dados originais, os coeficientes de correlação calculados pelo programa ordenados numa matriz quadrada de ordem 8×8 , e mostram-se os coeficientes calculados r_{yz} , ρ_{yz} e GE (grau de espureidade), que são os resultados do programa. O arquivo de resultados apresentará a informação da seguinte maneira:

a) Título da rodada (até 80 caracteres).

- b) NI = 1, NF = 4, KI = 5, KF = 8 que são índices que indicam que se manipulam 4 variáveis para a primeira função e outras 4 variáveis para a segunda função.
- c) Médias e desvios padrões das variáveis X_i , ($i = 1, 8$). No programa se estipulará que, se $m_i = 1$ e $\sigma_i = 0$, isto significa que a variável X_i não intervém na função considerada.
- d) Correlações r_{ij} ($i = 1, 8, j = 1, 8$) entre as variáveis X_i .
- e) Correlação final de Pearson, utilizando a equação proposta.
- f) Coeficiente de correlação espúria, calculada com a equação proposta e considerando que $r_{ij} = 0$ para $i \neq j$ e $r_{ij} = 1$ para $i = j$.
- g) Grau de espureidade $GE = 100 \left(\frac{\rho_{yz}}{r_{yz}} \right)$.

Na seqüência apresenta-se o formato para o arquivo de saída que usa o programa:

TITULO DA RODADA

NI, NF, KI, KF

M (1), SIGMA (1)

M (2), SIGMA (2)

M (3), SIGMA (3)

M (4), SIGMA (4)

M (5), SIGMA (5)

M (6), SIGMA (6)

M (7), SIGMA (7)

M (8), SIGMA (8)

CR (1,1), CR (1,2), CR (1,3), CR (1,4), CR (1,5), CR (1,6), CR (1,7), CR (1,8)

CR (2,1), CR (2,2), CR (2,3), CR (2,4), CR (2,5), CR (2,6), CR (2,7), CR (2,8)

CR (3,1), CR (3,2), CR (3,3), CR (3,4), CR (3,5), CR (3,6), CR (3,7), CR (3,8)

CR (4,1), CR (4,2), CR (4,3), CR (4,4), CR (4,5), CR (4,6), CR (4,7), CR (4,8)

CR (5,1), CR (5,2), CR (5,3), CR (5,4), CR (5,5), CR (5,6), CR (5,7), CR (5,8)

CR (6,1), CR (6,2), CR (6,3), CR (6,4), CR (6,5), CR (6,6), CR (6,7), CR (6,8)

CR (7,1), CR (7,2), CR (7,3), CR (7,4), CR (7,5), CR (7,6), CR (7,7), CR (7,8)

CORRELACION FINAL Ryz =

CORRELACION ESPURIA =

GRADO ESPURIO =

4.3 Exemplos de organização de arquivos para o programa e dos respectivos arquivos de saída de resultados

Em continuação apresentam três exemplos de como organizar um arquivo de dados para o programa com as respectivas saídas que o programa deve fornecer:

Exemplo No. 1. - Considerem-se as seguintes funções cujo grau de correlação espúria será calculado com o programa:

$$y = \frac{x_1}{x_2}, z = \frac{x_5}{x_6}$$

Estas funções podem exprimir-se da seguinte maneira (capítulo 3):

$$y = C_1 X_1^p X_2^q X_3^r X_4^s$$

$$z = C_2 X_5^t X_6^u X_7^v X_8^w$$

Faz-se a substituição correspondente das constantes, índices e variáveis concordando com a ordem das equações originais,

$$y = X_1^1 X_2^1 X_3^0 X_4^0$$

$$z = X_5^1 X_6^{-1} X_7^0 X_8^0$$

$$C_1 = 1, p = 1, q = 1, r = 0, s = 0$$

$$C_2 = 1, t = 1, u = -1, v = 0, w = 0$$

Propõe-se, então, o seguinte conjunto de valores para caracterizar estas funções, concordando com a ordem de ingresso dos dados exposta acima para o programa:

Título da rodada

$$C_1 = 1, p = 1, q = -1, r = 0, s = 0$$

$$C_2 = 1, t = 1, u = -1, v = 0, w = 0$$

$$m_1 = 10, \sigma_1 = 2$$

$$m_2 = 12, \sigma_2 = 3.6$$

$$m_3 = 1, \sigma_3 = 0$$

$$m_4 = 1, \sigma_4 = 0$$

$$m_5 = 11, \sigma_5 = 4.4$$

$$m_6 = 13, \sigma_6 = 6.5$$

$$m_7 = 1, \sigma_7 = 0$$

$$m_8 = 1, \sigma_8 = 0$$

$$r_{12} = 0.90, r_{13} = 0, r_{14} = 0, r_{15} = 0.91, r_{16} = 0.92, r_{17} = 0, r_{18} = 0$$

$$r_{23} = 0, r_{24} = 0, r_{25} = 0.93, r_{26} = 0.94, r_{27} = 0, r_{28} = 0$$

$$r_{34} = 0, r_{35} = 0, r_{36} = 0, r_{37} = 0, r_{38} = 0$$

$$r_{45} = 0, r_{46} = 0, r_{47} = 0, r_{48} = 0$$

$$r_{56} = 0.95, r_{57} = 0, r_{58} = 0$$

$$r_{67} = 0, r_{68} = 0$$

$$r_{78} = 0$$

Ordenando os dados segundo o formato requerido em FORTRAN para ingresso ao

programa, obtém-se o seguinte arquivo, nomeado neste caso particular DADOS1.TXT :

```

Y = X1/X2      Z = X5/X6
1, 1, -1, 0, 0
1, 1, -1, 0, 0
10, 2.0
12, 3.6
1, 0
1, 0
11, 4.4
13, 6.5
1, 0
1, 0
0.90, 0, 0, 0.91, 0.92, 0, 0
      0, 0, 0.93, 0.94, 0, 0

```

```

0, 0, 0, 0, 0
  0, 0, 0, 0
    0.95, 0, 0
      0, 0
        0

```

Após rodar o programa ESPURI para o arquivo DADOS1.TXT, obtém-se o seguinte resultado, gerando-se o arquivo DADOS1.OUT, o qual é impresso como saída do programa:

Y = X1/X2 Z = X5/X6

NI,NF,KI,KF 1 4 5 8

MEDIAS DESVIACIONES ESTANDAR

```

10.000      2.000
12.000      3.600
 1.000      .000
 1.000      .000
11.000      4.400
13.000      6.500
 1.000      .000
 1.000      .000

```

```

CORRELACIONES CR(1,J) = 1.000  .900  .000  .000  .910  .920  .000  .000
CORRELACIONES CR(2,J) =  .900  1.000  .000  .000  .930  .940  .000  .000
CORRELACIONES CR(3,J) =  .000  .000  1.000  .000  .000  .000  .000  .000
CORRELACIONES CR(4,J) =  .000  .000  .000  1.000  .000  .000  .000  .000
CORRELACIONES CR(5,J) =  .910  .930  .000  .000  1.000  .950  .000  .000
CORRELACIONES CR(6,J) =  .920  .940  .000  .000  .950  1.000  .000  .000
CORRELACIONES CR(7,J) =  .000  .000  .000  .000  .000  .000  1.000  .000
CORRELACIONES CR(8,J) =  .000  .000  .000  .000  .000  .000  .000  1.000

```

CORRELACION FINAL Ryz = .3970

CORRELACION ESPURIA = .0000

GRADO ESPURIO = .00 %

Exemplo No. 2.- Considere-se o seguinte exemplo, que é parecido ao caso de estudo No. 1 considerado no Capítulo 6 como exercício para ser resolvido na planilha eletrônica proposta:

$$y = \frac{x_1 x_4}{x_2}, z = C \frac{x_5^2 x_4^3}{x_2^5}$$

Estas funções podem exprimir-se da seguinte maneira (Capítulo 3):

Considerando o modelo algébrico original proposto para as funções y e z ,

$$y = C_1 X_1^p X_2^q X_3^r X_4^s$$

$$z = C_2 X_5^t X_6^u X_7^v X_8^w$$

Faz-se a substituição correspondente das constantes, índices e variáveis concordando com a ordem das equações originais,

$$y = X_1^1 X_2^{-1} X_3^0 X_4^1$$

$$z = C_2 X_5^2 X_6^{-5} X_7^3 X_8^0$$

$$C_1 = 1, p = 1, q = -1, r = 0, s = 1$$

$$C_2, t = 2, u = -5, v = 3, w = 0$$

com a observação de que $X_6 = X_2, X_7 = X_4, C_2 = C; X_3, X_8$, não se definem; observe-se que o uso do programa requer a definição das variáveis X_5, X_6, X_7, X_8 , que são variáveis mudas (*dummy variables*): observe-se também a peculiaridade de que as variáveis

X_1, X_2, X_3, X_4 não podem aparecer na função z se antes não aparecem na função \mathcal{Y} .

Observe-se que $C_2 = 1/g = 1/9,81 = 0,1019$.

Propõe-se, então, o seguinte conjunto de valores para caracterizar estas funções, concordando com a ordem de ingresso dos dados exposta acima para o programa:

Título da rodada

$$C_1 = 1, p = 1, q = -1, r = 0, s = 1$$

$$C_2 = 0.1019, t = 2, u = -5, v = 3, w = 0$$

$$m_1 = 1, \sigma_1 = 0.10$$

$$m_2 = 3, \sigma_2 = 0.01$$

$$m_3 = 4, \sigma_3 = 0.5$$

$$m_4 = 0.5, \sigma_4 = 0.01$$

$$m_5 = 4, \sigma_5 = 0.5$$

$$m_6 = 3, \sigma_6 = 0.01$$

$$m_7 = 0.5, \sigma_7 = 0.01$$

$$m_8 = 1, \sigma_8 = 0$$

$$r_{12} = 0, r_{13} = 0, r_{14} = 0, r_{15} = 0, r_{16} = 0, r_{17} = 0, r_{18} = 0$$

$$r_{23} = 0, r_{24} = 0, r_{25} = 0, r_{26} = 1, r_{27} = 0, r_{28} = 0$$

$$r_{34} = 0, r_{35} = 0, r_{36} = 0, r_{37} = 0, r_{38} = 0$$

$$r_{45} = 0, r_{46} = 0, r_{47} = 1, r_{48} = 0$$

$$r_{56} = 0, r_{57} = 0, r_{58} = 0$$

$$r_{67} = 0, r_{68} = 0$$

$$r_{78} = 0$$

Ordenando os dados segundo o formato requerido em FORTRAN para ingresso ao

programa, obtém-se o seguinte arquivo, nomeado neste caso particular DADOS2.TXT :

$$Y = X1 \ X4/X2 \quad Z = C \ X5^{**2} \ X4^{**3} / \ X2^{**5}$$

$$1, \ 1, \ -1, \ 0, \ 1$$

$$0.1019, \ 2, \ -5, \ 3, \ 0$$

$$1, \ 0.10$$

$$3, \ 0.01$$

$$4, \ 0.5$$

$$.5, \ 0.01$$

$$4, \ 0.5$$

$$3, \ 0.01$$

$$.5, \ 0.01$$

$$1, \ 0$$

$$0, \ 0, \ 0, \ 0, \ 0, \ 0, \ 0$$

$$0, \ 0, \ 0, \ 1, \ 0, \ 0$$

$$0, \ 0, \ 0, \ 0, \ 0$$

```

0, 0, 1, 0
  0, 0, 0
    0, 0
      0

```

Após rodar o programa ESPURI para o arquivo DADOS2.TXT, obtém-se o seguinte resultado, gerando-se o arquivo DADOS2.OUT, o qual é impresso como saída do programa:

Y = X1 X4/X2 Z = C X5**2 X4**3/ X2**5

NI,NF,KI,KF 1 4 5 8

MEDIAS DESVIACIONES ESTANDAR

```

1.000           .100
3.000           .010
4.000           .500
.500           .010
4.000           .500
3.000           .010
.500           .010
1.000           .000

```

CORRELACIONES CR(1,J) = 1.000 .000 .000 .000 .000 .000 .000 .000

CORRELACIONES CR(2,J) = .000 1.000 .000 .000 .000 1.000 .000 .000

CORRELACIONES CR(3,J) = .000 .000 1.000 .000 1.000 .000 .000 .000

CORRELACIONES CR(4,J) = .000 .000 .000 1.000 .000 .000 1.000 .000

CORRELACIONES CR(5,J) = .000 .000 1.000 .000 1.000 .000 .000 .000

CORRELACIONES CR(6,J) = .000 1.000 .000 .000 .000 1.000 .000 .000

CORRELACIONES CR(7,J) = .000 .000 .000 1.000 .000 .000 1.000 .000

CORRELACIONES CR(8,J) = .000 .000 .000 .000 .000 .000 .000 1.000

CORRELACION FINAL Ryz = .0478

$$\text{CORRELACION ESPURIA} = .0478$$

$$\text{GRADO ESPURIO} = 100.00 \%$$

Este resultado indica que a correlação proposta entre funções é irrelevante e que deve ser descartada.

Exemplo No. 3. - Considere-se o seguinte exemplo, que também é considerado no Capítulo 5 como exercício proposto para ser resolvido na planilha eletrônica EXCEL:

$$y = \frac{x_1}{x_2}, z = \frac{\left(\frac{x_5}{x_6}\right)^{1/3}}{x_2}$$

Estas funções podem exprimir-se da seguinte maneira (Capítulo 3):

$$y = C_1 X_1^p X_2^q X_3^r X_4^s$$

$$z = C_2 X_5^t X_6^u X_7^v X_8^w$$

Faz-se a substituição correspondente das constantes, índices e variáveis concordando com a ordem das equações originais,

$$y = X_1^1 X_2^{-1} X_3^0 X_4^0$$

$$z = X_5^{1/3} X_6^{-1/3} X_7^{-1} X_8^0$$

$$C_1 = 1, p = 1, q = -1, r = 0, s = 0$$

$$C_2 = 1, t = 1/3, u = -1/3, v = -1, w = 0$$

com a observação que $X_7 = X_2$ e que X_3, X_4, X_8 não se definem.

Propõe-se, então, o seguinte conjunto de valores para caracterizar estas funções, concordando com a ordem de ingresso dos dados exposta acima para o programa:

Título da rodada

$$C_1 = 1, p = 1, q = -1, r = 0, s = 0$$

$$C_2 = 1, t = 0.3333, u = -0.3333, v = -1, w = 0$$

$$m_1 = 4.857, \sigma_1 = 3.112$$

$$m_2 = 31.5, \sigma_2 = 3.969$$

$$m_3 = 1, \sigma_3 = 0$$

$$m_4 = 1, \sigma_4 = 0$$

$$m_5 = 2.078, \sigma_5 = 2.852$$

$$m_6 = 1000, \sigma_6 = 0$$

$$m_7 = 31.5, \sigma_7 = 3.969$$

$$m_8 = 1, \sigma_8 = 0$$

$$r_{12} = -0.2367, r_{13} = 0, r_{14} = 0, r_{15} = 0.8332, r_{16} = 0, r_{17} = -0.2367, r_{18} = 0$$

$$r_{23} = 0, r_{24} = 0, r_{25} = 0.7149, r_{26} = 0, r_{27} = 1.0, r_{28} = 0$$

$$r_{34} = 0, r_{35} = 0, r_{36} = 0, r_{37} = 0, r_{38} = 0$$

$$r_{45} = 0, r_{46} = 0, r_{47} = 0, r_{48} = 0$$

$$r_{56} = 0, r_{57} = 0.7149, r_{58} = 0$$

$$r_{67} = 0, r_{68} = 0$$

$$r_{78} = 0$$

Ordenando os dados segundo o formato requerido em FORTRAN para ingresso ao

programa, obtém-se o seguinte arquivo, nomeado neste caso particular DADOS3.TXT :

```

Y = X1 / X2      Z = (X5/X6)**(1/3)/X2
1, 1, -1, 0, 0
1, 0.3333, -0.3333, -1, 0
4.857, 3.112
31.5, 3.969
1, 0
1, 0
2.078, 2.852
1000, 0
31.5, 3.969
1, 0
-0.2367,      0,      0,      0.8322,      0,      -0.2367,      0
              0,      0,      0.7149,      0,              1.0,      0
              0,              0,      0,              0,      0
              0,      0,              0,      0,      0
              0,              0.7149,      0
              0,      0
              0

```

Após rodar o programa ESPURI para o arquivo DADOS3.TXT, obtém-se o seguinte resultado, gerando-se o arquivo DADOS3.OUT, o qual é impresso como saída do programa:

$$Y = X1 / X2 \quad Z = (X5/X6)**(1/3)/X2$$

NI,NF,KI,KF	1	4	5	8
MEDIAS	DESVIACIONES ESTANDAR			
4.857	3.112			
31.500	3.969			
1.000	.000			
1.000	.000			
2.078	2.852			
1000.000	.000			
31.500	3.969			
1.000	.000			

CORRELACIONES CR(1,J) =	1.000	-.237	.000	.000	.832	.000	-.237	.000
CORRELACIONES CR(2,J) =	-.237	1.000	.000	.000	.715	.000	1.000	.000
CORRELACIONES CR(3,J) =	.000	.000	1.000	.000	.000	.000	.000	.000
CORRELACIONES CR(4,J) =	.000	.000	.000	1.000	.000	.000	.000	.000
CORRELACIONES CR(5,J) =	.832	.715	.000	.000	1.000	.000	.715	.000
CORRELACIONES CR(6,J) =	.000	.000	.000	.000	.000	1.000	.000	.000
CORRELACIONES CR(7,J) =	-.237	1.000	.000	.000	.715	.000	1.000	.000
CORRELACIONES CR(8,J) =	.000	.000	.000	.000	.000	.000	.000	1.000

CORRELACION FINAL Ryz = .9230

CORRELACION ESPURIA = .0512

GRADO ESPURIO = 5.55 %

Este resultado indica que a correlação proposta entre funções pode implicar uma relação causa-efeito válida entre as variáveis propostas, mesmo que estejam repetidas.

4.4 Compilação do programa ESPURI

O programa ESPURI esta escrito em linguagem de alto nível FORTRAN 77 e compila-se como segue:

- a) O programa é editado sob o nome de ESPURI.FOR
- b) O programa é compilado com a ordem FOR1
- c) Se não aparecer erro, usa-se a ordem PAS2 para criar o arquivo ESPURI.OBJ
- d) O programa enlase-se com a ordem LINK para obter o programa executável ESPURI.EXE
- e) Usa-se o programa ESPURI.EXE para rodar os exemplos.

Na Figura 1 apresenta-se o diagrama de fluxo do programa ESPURI

4.5 Uso de planilha de cálculo eletrônica para calcular os dados de ingresso para o programa num caso de estudo

No Capítulo 5 usa-se uma planilha de cálculo eletrônica para gerar os dados de entrada para o programa ESPURI usando números aleatórios calculados pela planilha, com o intuito de calcular os valores das variáveis envolvidas nas funções y e z , colocá-las em forma “adimensional” e grafá-las, para observar se as funções seguem alguma lei. Depois, usa-se o programa ESPURI para calcular o coeficiente de correlação espúria.

Em continuação apresentam-se o diagrama de fluxo para o programa ESPURI assim como o listado do programa escrito em FORTRAN 77:

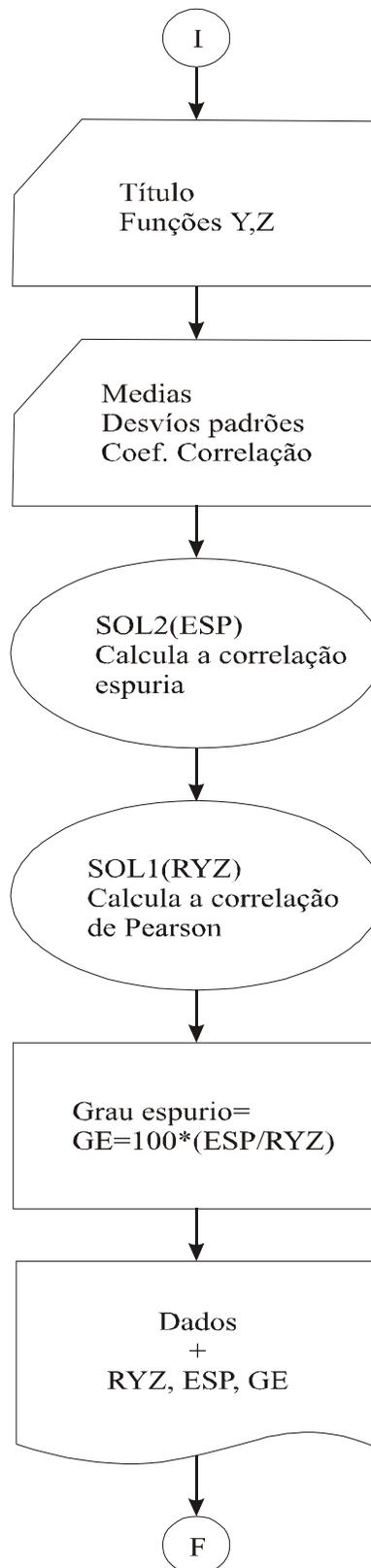


Figura 2.- Diagrama de fluxo do programa de computador ESPURI

```

$STORAGE:2
$LARGE
PROGRAM ESPURI
C#####
C
C   PROGRAMA PARA CALCULAR LA CORRELACION DE PEARSON,
C   CORRELACION ESPURIA Y GRADO ESPURIO,
C   APLICANDO FORMULA GENERAL PARA DOS FUNCIONES
C
C   Y = C1  X1**P  X2**Q  X3**R  X4**S
C   Z = C2  X5**T  X6**U  X7**V  X8**W
C
C#####
C   CHARACTER ARCH*12, TITULO (80)*1
C   REAL M
C   COMMON F(20), G(20)
C   COMMON SIGMA(20), M(20), CR(10,10)
C   COMMON NI,NF,KI,KF
C   COMMON C1, C2, P, Q, R, S, T, U, V, W, X
C -----
C   WRITE (*,'(A\)' ) ' ARCHIVO DE ENTRADA: '
C   READ (*,11) ARCH
11   FORMAT (A12)
C   OPEN (1,FILE=ARCH,STATUS='OLD')
C   WRITE (*,'(A\)' ) ' ARCHIVO DE SALIDA '
C   READ (*,11) ARCH
C   OPEN (2,FILE=ARCH,STATUS= 'NEW')
C -----
C
C   LECTURA DE DATOS
C
C   TITULO, CONSTANTES DE FUNCIONES, MEDIAS Y DESVIACIONES
C   ESTANDAR Y COEFICIENTES DE CORRELACION
C
C   READ (1,101) TITULO
C   RANGO DE LA FUNCION Y DE (1..4), RANGO DE LA FUNCION Z DE
C   (5..8)
C   NI=1
C   NF=4
C   KI=5
C   KF=8
C   READ (1,*) C1, P, Q, R, S
C   READ (1,*) C2, T, U, V, W
C   LAS MEDIAS DEBEN TENER UN VALOR DIFERENTE DE CERO
C   DO 5 I=NI,KF
C   READ (1,*) M(I), SIGMA(I)
5   CONTINUE

```

```

C INICIALIZA MATRIZ DE COEFICIENTES DE CORRELACION
  DO 13 I=1,10
  DO 14 J=1,10
  IF (I .EQ. J) THEN
  CR(I,J)=1.0
  ELSE
  CR(I,J)=0.0
  ENDIF
14     CONTINUE
13     CONTINUE
C
C CALCULO DE CORRELACION ESPURIA
C
  CALL SOL2 (ESP)
  DO 15 I=NI,KF-1
  READ (1,*) (CR(I,J), J=I+1,KF)
15     CONTINUE
C INICIALIZA FUNCIONES
  DO 18 I=1,20
  F(I) = 0.0
  G(I) = 0.0
18     CONTINUE
C
C CALCULO DE CORRELACION PEARSON
C
  CALL SOL1 (RYZ)
C
C GRADO ESPURIO EN PORCENTAJE
C
  IF (RYZ .EQ. 0.0) THEN
  GE = 0.0
  ELSE
  GE = 100*ESP/RYZ
  ENDIF
C
C IMPRESION DE RESULTADOS
C
  WRITE (2,110) TITULO
  WRITE (2,111) NI, NF, KI, KF
  WRITE (2,114)
  WRITE (2,120) C1, P, Q, R, S
  WRITE (2, 121) C2, T, U, V, W
  WRITE (2, 114)
  WRITE (2, 112)
  DO 66 I=NI, KF
  WRITE (2,113) M(I), SIGMA(I)
66     CONTINUE

```

```

        WRITE (2,114)
        DO 30 I=NI, KF
30      WRITE (2,116) I, (CR (I,J), J=NI,KF)
        WRITE (2,114)
        WRITE (2,117) RYZ
        IF (ABS(RYZ) .GT. 1.0) THEN
        WRITE (2,115)
        ELSE
        ENDIF
        WRITE (2,114)
        WRITE (2,118) ESP
        IF (ABS(ESP) .GT. 1.0 ) THEN
        WRITE (2,115)
        ELSE
        ENDIF
        WRITE (2,114)
        WRITE (2,119) GE
C
C  FORMATOS
C
101      FORMAT (80A1)
110      FORMAT (1X, 80A1,/)
111      FORMAT (3X, 'NI,NF,KI,KF' 4I10,/)
112      FORMAT (3X, 'MEDIAS      DESVIACIONES ESTANDAR',/)
113      FORMAT (3X, 2F10.3)
114      FORMAT (3X, /)
115      FORMAT (10X, '#### ERROR, CORRELACION FUERA DE RANGO
###')
116      FORMAT (3X, ' CORRELACIONES CR(' ,I1, ',J) = ',8F6.3,/)
117      FORMAT (3X, ' CORRELACION FINAL Ryz = ' , F10.4)
118      FORMAT (3X, ' CORRELACION ESPURIA = ' , F10.4)
119      FORMAT (3X, ' GRADO ESPURIO = ' , F7.2, ' %')
120      FORMAT (3X, ' FUNCION Y =', F5.2, '* X1**', F5.2, '
*X2**', F5.2, '*X3**',F5.2, '*X4**', F5.2,/)
121      FORMAT (3X, ' FUNCION Z =', F5.2, '* X5**', F5.2, '
*X6**', F5.2, '*X7**',F5.2, '*X8**', F5.2,/)
        END

```

```

      SUBROUTINE SOL1(RYZ)
C SUBROUTINA QUE EFECTUA LA CORRELACION DE PEARSON
C
      REAL M
      COMMON F(20), G(20)
      COMMON SIGMA (20), M(20), CR(10,10)
      COMMON NI, NF, KI, KF
      COMMON C1, C2, P, Q, R, S, T, U, V, W
C
C CALCULO DE LAS FUNCIONES F ( ) Y G ( )
C
      F(1) = C1*P*(M(1)**(P-1))* (M(2)**Q)* (M(3)**R)* (M(4)**S)
      F(2) = C1*(M(1)**P)*Q*(M(2)**(Q-1))* (M(3)**R)* (M(4)**S)
      F(3) = C1*(M(1)**P)* (M(2)**Q)*R*(M(3)**(R-1))* (M(4)**S)
      F(4) = C1*(M(1)**P)* (M(2)**Q)* (M(3)**R)*S*(M(4)**(S-1))
      G(5) = C2*T*(M(5)**(T-1))* (M(6)**U)* (M(7)**V)* (M(8)**W)
      G(6) = C2*(M(5)**T)*U*(M(6)**(U-1))* (M(7)**V)* (M(8)**W)
      G(7) = C2*(M(5)**T)* (M(6)**U)*V*(M(7)**(V-1))* (M(8)**W)
      G(8) = C2*(M(5)**T)* (M(6)**U)* (M(7)**V)*W*(M(8)**(W-1))
C
COMPLEMENTO DE CORRELACIONES
C
      DO 21 I=1,10
      DO 22 J=1,10
      IF (CR(I,J) .NE. 0.0) CR(J,I) = CR(I,J)
22 CONTINUE
21 CONTINUE
C
CALCULO DEL NUMERADOR
C
      AC = 0.0
      DO 10 I=NI,NF
      DO 11 J=KI,KF
      AC = AC + (F(I)*G(J)*CR(I,J)*SIGMA(I)*SIGMA(J))
11 CONTINUE
10 CONTINUE
      ARR = AC
C
CALCULO DEL PRIMER DENOMINADOR
C
      AC = 0.0
      DO 15 I=NI,NF
      DO 20 J=NI,NF
      AC = AC + (F(I)*F(J)*CR(I,J)*SIGMA(I)*SIGMA(J))
20 CONTINUE
15 CONTINUE
      AB1 = SQRT (AC)

```

```

C
CALCULO DEL SEGUNDO DENOMINADOR
C
  AC = 0.0
  DO 25 I=KI, KF
  DO 30 J=KI, KF
  AC = AC + (G(I)*G(J)*CR(I, J)*SIGMA(I)*SIGMA(J))
30  CONTINUE
25  CONTINUE
  AB2 = SQRT (AC)
C
CORRELACION PEARSON
C
  RYZ = ARR/(AB1*AB2)
  RETURN
  END

```

```

SUBROUTINE SOL2 (ESP)
C SUBROUTINA QUE EFECTUA LA CORRELACION ESPURIA
  REAL M
  COMMON F(20), G(20)
  COMMON SIGMA(20), M(20), CR(10,10)
  COMMON NI, NF, KI, KF
  COMMON C1, C2, P, Q, R, S, T, U, V, W
C
CALCULO DE LAS FUNCIONES F( ) Y G( )
C

```

```

  F(1) = C1*P*(M(1)**(P-1))* (M(2)**Q)* (M(3)**R)* (M(4)**S)
  F(2) = C1*(M(1)**P)*Q*(M(2)**(Q-1))* (M(3)**R)* (M(4)**S)
  F(3) = C1*(M(1)**P)* (M(2)**Q)*R*(M(3)**(R-1))* (M(4)**S)
  F(4) = C1*(M(1)**P)* (M(2)**Q)* (M(3)**R)*S*(M(4)**(S-1))
  G(5) = C2*T*(M(5)**(T-1))* (M(6)**U)* (M(7)**V)* (M(8)**W)
  G(6) = C2*(M(5)**T)*U*(M(6)**(U-1))* (M(7)**V)* (M(8)**W)
  G(7) = C2*(M(5)**T)* (M(6)**U)*V*(M(7)**(V-1))* (M(8)**W)
  G(8) = C2*(M(5)**T)* (M(6)**U)* (M(7)**V)*W*(M(8)**(W-1))

```

```

C
COMPLEMENTO DE CORRELACIONES
C
      DO 21 I=1,10
      DO 22 J=1,10
      IF (CR(I,J) .NE. 0.0) CR(J,I) = CR(I,J)
22    CONTINUE
21    CONTINUE
C
C CALCULO DE CORRELACIONES = 1.0 POR ESPUREIDAD
C
      DO 33 I=NI,KF
      DO 34 J=NI,KF
      IF (M(I).EQ.M(J) .AND. SIGMA(I).EQ.SIGMA(J) .AND.
SIGMA(I).NE.0.0) CR(I,J)=1.0
      ELSE
      ENDIF
34    CONTINUE
33    CONTINUE
C
CALCULO DEL NUMERADOR
C
      AC = 0.0
      DO 10 I=NI,NF
      DO 11 J=KI,KF
      AC = AC + (F(I)*G(J)*CR(I,J)*SIGMA(I)*SIGMA(J))
11    CONTINUE
10    CONTINUE
      ARR = AC
C
CALCULO DEL PRIMER DENOMINADOR
C
      AC = 0.0
      DO 15 I = NI,NF
      DO 20 J = NI,NF
      AC = AC + (F(I)*F(J)*CR(I,J)*SIGMA(I)*SIGMA(J))
20    CONTINUE
15    CONTINUE
      AB1 = SQRT (AC)
C
CALCULO DEL SEGUNDO DENOMINADOR
C
      AC = 0.0
      DO 25 I=KI,KF
      DO 30 J=KI,KF
      AC = AC + (G(I)*G(J)*CR(I,J)*SIGMA(I)*SIGMA(J))
30    CONTINUE

```

```
25  CONTINUE
    AB2 = SQRT (AC)
C
CORRELACION ESPURIA
C
    ESP = ARR / (AB1 * AB2)
    RETURN
    END
```

CAPÍTULO 5

ORIENTAÇÕES PARA APLICAÇÕES NA ENGENHARIA HIDRÁULICA

5.1 Introdução

Neste capítulo apresentam-se três casos de estudo da hidráulica que ilustram detalhadamente o uso da metodologia exposta nos capítulos 2 e 3 e o uso do programa desenvolvido no Capítulo 4 para avaliar o grau de correlação espúria presente quando se tenta correlacionar dois parâmetros adimensionais que têm uma ou mais variáveis repetidas. Usando as técnicas habituais da estatística, é possível construir uma relação entre as duas funções estudadas que possa ser linearizada por métodos matemáticos e por tanto se obterá um valor alto do coeficiente de correlação (aproximado de 1), o que pode conduzir a pensar erradamente que a dita relação é representativa de uma relação causa – efeito que é válida do ponto de vista físico, explicando assim a fenomenologia do processo estudado. O intuito deste capítulo é mostrar como avaliar esta situação para poder decidir objetivamente se esta relação proposta deve descartar-se ou não. No desenvolvimento destes exemplos, para o caso de estudo No. 1, propõe-se uma equação para uma curva chave que relaciona a profundidade de escoamento d com a vazão Q de um canal trapezoidal. Nesta experiência tenta-se pesquisar se, para uma relação matemática funcional com variáveis repetidas, representada por uma curva de correlação obtida por alguma metodologia analítica, pode-se apresentar o efeito de autocorrelação das variáveis repetidas (correlação de uma variável com ela mesma), o que conduziria a um alto valor do coeficiente de correlação espúria que é mascarado por um valor também alto do coeficiente de correlação total. Para atingir este objetivo, usam-se dados sintéticos gerados por meio de números aleatórios (Helmlinger, 1989), num primeiro caso segundo um roteiro racional baseado nas equações da hidráulica

superficial que governam este tipo de escoamento, e num segundo caso usando números aleatórios indiscriminadamente, com o intuito de desvendar o efeito da natureza dos dados utilizados na equação matemática que governa a curva chave. Do experimento deverá concluir-se se existe provavelmente uma verdadeira relação causa – efeito entre as variáveis relacionadas, que deverá ser conferida com pesquisas de laboratório. Para facilitar o estudo da aplicação da metodologia desenvolvida neste trabalho a casos freqüentes da hidráulica, apresentam-se na Tabela 1 (Benson, 1965; Helmlinger, 1989) as relações algébricas de variáveis mais freqüentemente encontradas no campo da hidráulica. Os casos de estudo 2 e 3 (geração de ondas num canal reprodutor de ondas e cálculo da vazão em um vertedouro triangular de crista delgada) apresentam-se como temas possíveis de pesquisa, sem serem desenvolvidos.

Tabela 3.- Modelos algébricos mais freqüentes na engenharia hidráulica.

Modelo	Z	Y
1	x_1	x_2
2	$\frac{x_3}{x_4}$	$\frac{x_1}{x_2}$
3	x_3	$\frac{x_1}{x_2}$
4	$\frac{x_1}{x_4}$	$\frac{x_1}{x_2}$
5	$\frac{x_3}{x_2}$	$\frac{x_1}{x_2}$
6	$\frac{x_2}{x_4}$	$\frac{x_1}{x_2}$
7	x_1	$\frac{x_1}{x_2}$
8	x_2	$\frac{x_1}{x_2}$
9	$\frac{x_2}{x_4}$	$\frac{1}{x_2}$
10	$x_3 x_4$	$x_1 x_2$
11	$x_3 x_2$	$x_1 x_2$
12	x_3	$x_1 x_2$
13	x_2	$x_1 x_2$

5.2 Caso de estudo No. 1. Proposta de equação para uma curva chave de um canal trapezoidal.

Por exemplo, tenta-se construir uma equação para relacionar a profundidade de escoamento d com a vazão Q para um canal trapezoidal (Figura 1), que na hidráulica é conhecida como curva chave. Tenta-se então estudar o efeito da correlação espúria sobre esta relação tentativa entre dois grupos adimensionais propostos para estudar o fluxo de água num canal que são:

$$\frac{md}{b}, \frac{Q^2 m^3}{gb^5}$$

onde

m = cotangente do talude do canal

d = profundidade do escoamento

b = largura do canal

Q = vazão

g = constante gravitacional

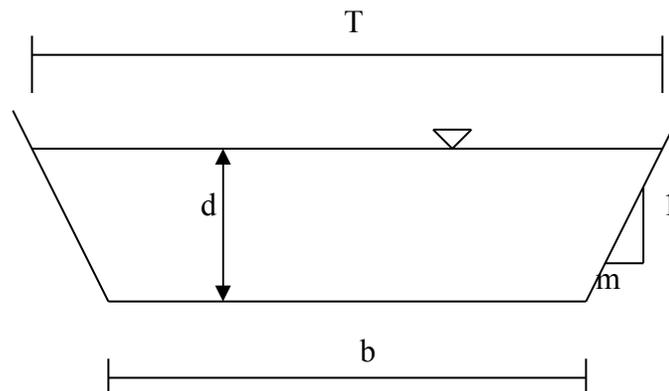


Figura 3.- Seção transversal de um canal trapezoidal.

Os dois números adimensionais propostos podem se relacionar pela seguinte relação funcional:

$$\frac{md}{b} = f\left(\frac{Q^2 m^3}{gb^5}\right)$$

Com referência a Figura 1 do Capítulo 3, identificam-se as variáveis $y = C_1 X_1^p X_2^q X_3^r X_4^s$ e $z = C_2 X_5^t X_6^u X_7^v X_8^w$ para este caso de estudo como:

$$y = \frac{md}{b} \quad (1)$$

$$z = \frac{Q^2 m^3}{gb^5} \quad (2)$$

obtendo-se os seguintes valores para os coeficientes, as variáveis e os expoentes:

para a variável y ,

$$C_1 = 1, X_1 = m, X_2 = b, X_3 = \text{indefinida}, X_4 = d, p = 1, q = -1, r = 0, s = 1$$

para a variável z ,

$$C_2 = \frac{1}{g} = 0.1019, X_5 = Q, X_6 = X_2 = b, X_7 = X_1 = m, X_8 = \text{indefinida}, t = 2, u = -5, v = 3, w = 0$$

então,

$$y = \frac{X_1 X_4}{X_2} \quad (3)$$

$$z = C_2 \frac{X_5^2 X_1^3}{X_2^5} \quad (4)$$

Porém, para usar o programa do Capítulo 4, as funções y e z devem ser expressas da seguinte maneira,

$$y = \frac{X_1 X_4}{X_2}$$

$$z = C_2 \frac{X_5^2 X_7^3}{X_6^5}$$

Com o intuito de decidir se esta relação proposta pode ser válida como representativa do fenômeno físico estudado, propõe-se um exercício matemático que consiste em usar números aleatórios para gerar um conjunto (base) de dados obtidos das funções y e z (equações 3 e 4). Porém, o programa ESPURI vai mostrar até que grau esta relação proposta arbitrariamente é uma autocorrelação entre uma ou várias variáveis repetidas, o que a pode tornar redundante, ou se ela pode ser aceita. Para realizar esta simulação matemática, devem ser seguidas as diretrizes do método científico, cuja primeira premissa é que deve existir uma igualdade de condições para obter uma igualdade de resultados (princípio de *certaribus paribus*), com o intuito de que o experimento esteja controlado. Quer dizer, os números aleatórios não podem ser usados indiscriminadamente para todas as variáveis contempladas nas equações 1 e 2, desde que isto causaria a existência de distribuições de probabilidade diferentes para as variáveis, incluindo a grande dificuldade de ter que pesquisar a natureza verdadeira destas distribuições, coisa que nem sempre poderá ser feita. Desde a colocação inicial do problema aceita-se que as variáveis m e b representam a geometria do canal considerado, que não pode mudar dentro do experimento considerado, aceita-se que a variável g depende do entorno físico do problema, e que sua variação com a altitude sobre o nível medio do mar é desprezível, considerando-se então como uma constante, e aceita-se que existe uma relação funcional unívoca diretamente

proporcional, bem definida entre as variáveis d e Q , ($Q = f(d)$), que não pode ser ignorada sem cometer um grande erro. Esta relação funcional conduz à conclusão que as variáveis d e Q têm a mesma distribuição de probabilidade. Então, a única variável cujo valor pode ser simulado com números aleatórios é d , (a profundidade do canal), sendo o valor de Q fixado através de uma equação algébrica. O uso indiscriminado de números aleatórios para todas as variáveis conduziria a um experimento que não pode ser controlado, desvirtuando-se assim o verdadeiro significado do método científico. O experimento matemático proposto vai ser conduzido mudando o valor de uma variável (d), com a conseqüente mudança proporcional da variável Q , e conservando constantes as outras variáveis (princípio de *mutatis mutandis*).

Por meio da equação de continuidade, a relação $Q = f(d)$ pode exprimir-se como $Q = Av$, onde $Av = f(d)$, sendo A = área da seção transversal do canal e v = velocidade do escoamento d'água no canal, comumente calculada com a equação de Manning. Neste experimento considera-se um canal trapezoidal, por tanto,

$$A = (b + md)d$$

$$v = \frac{1}{n} R^{\frac{2}{3}} s^{\frac{1}{2}}, \text{ com } R = \frac{(b + md)d}{b + 2d\sqrt{1 + m^2}}$$

onde

n = coeficiente de rugosidade de Manning

R = raio hidráulico da seção transversal do canal

s = declividade do canal

Para operar o experimento matemático, usa-se a planilha de cálculo eletrônica apresentada na Tabela 4.

Porém, é necessário definir numericamente os valores das variáveis que caracterizam o canal trapezoidal e que intervêm no experimento matemático:

Então se propõe,

$b = \text{largura do canal} = 1,5 \text{ m}$

Para a declividade do talude do canal, escolhe-se um valor de 0,5: 1, segundo Chow (1959), que trata das declividades adequadas de taludes de canal para diferentes materiais, portanto $m = 0,5$ para argila rígida ou terra com revestimento de concreto.

Para o coeficiente de rugosidade n de Manning, escolhe-se um valor de $n = 0,013$ segundo Chow (1959), para canais revestidos de concreto com acabamento reguado.

Para a declividade do canal que se aplica na equação de Manning, escolhe-se um valor de $s = 0,003$.

Na Tabela 4 mostram-se os resultados da experiência, os mesmos que são apresentados em forma gráfica na Figura 4.

Deste caso de estudo poderá concluir-se se a lei matemática da correlação proposta pode considerar-se válida ou se ela deve-se somente ao efeito da correlação espúria procedente das variáveis repetidas nos grupos adimensionais. Existe a possibilidade de o efeito da correlação espúria se amplificar pelo uso de variáveis elevadas a potências grandes, o que poderia observar-se experimentalmente modificando, o valor do expoente para a variável

X_2 na função $z = C_2 \frac{X_5^2 X_1^3}{X_2^5}$, por exemplo.

Para usar o programa ESPURI, precisa-se formar o arquivo de entrada. Os dados para este arquivo obtêm-se da planilha de cálculo eletrônica apresentada na Tabela 5. Forma-se o arquivo de entrada concordando com a ordem estabelecida no Capítulo 4. Para avaliar os parâmetros estatísticos que vão formar o arquivo, lembre-se que, em geral, $r_{ij} = r_{ji}$, que em particular para este caso, $X_2 = X_6, X_1 = X_7$, que X_3, X_8 não se definem e, portanto a média m_i e o desvio padrão σ_i de estas variáveis são iguais a 1 e 0, respectivamente. Todos os coeficientes de correlação associados às variáveis não definidas são iguais a zero. Para as variáveis $X_1 = m = 0,5$ e $X_2 = b = 1,5$, mudam-se levemente os valores pra evitar a divisão entre zero que resultaria do desvio padrão igual a zero para valores constantes. Vide a explicação na página 70. O ordenamento é como segue:

Título da rodada

$$C_1 = 1, p = 1, q = -1, r = 0, s = 1$$

$$C_2 = 0.1019, t = 2, u = -5, v = 3, w = 0$$

$$m_1 = 0.50085714, \sigma_1 = 0.00278003$$

$$m_2 = 1.50028571, \sigma_2 = 0.00268594$$

$$m_3 = 1, \sigma_3 = 0$$

$$m_4 = 0.83354624, \sigma_4 = 0.53764314$$

$$m_5 = 5.11216128, \sigma_5 = 5.04950558$$

$$m_6 = 1.50028571, \sigma_6 = 0.00268594$$

$$m_7 = 0.50085714, \sigma_7 = 0.00278003$$

$$m_8 = 1, \sigma_8 = 0$$

$$r_{12} = -0.12818301, r_{13} = 0, r_{14} = 0.1655682, r_{15} = 0.20507744, r_{16} = -0.12818301, r_{17} = 1, r_{18} = 0$$

$$r_{23} = 0, r_{24} = -0.09654106, r_{25} = -0.13004267, r_{26} = 1, r_{27} = -0.12818301, r_{28} = 0$$

$$r_{34} = 0, r_{35} = 0, r_{36} = 0, r_{37} = 0, r_{38} = 0$$

$$r_{45} = 0.97886916, r_{46} = -0.09654106, r_{47} = 0.1655682, r_{48} = 0$$

$$r_{56} = -0.13004267, r_{57} = 0.20507744, r_{58} = 0$$

$$r_{67} = -0.12818301, r_{68} = 0$$

$$r_{78} = 0$$

Ordenando os dados segundo o formato requerido em FORTRAN para ingresso ao programa, obtém-se o seguinte arquivo, nomeado neste caso particular DADOS4.TXT :

```

Y = X1 X4/X2  Z = C X5**2 X1**3/ X2**5 (Z=C2 X5**2 X7**3/X6**5)
1, 1, -1, 0, 1
0.1019, 2, -5, 3, 0
0.50085714, 0.00278003
1.50028571, 0.00268594
1, 0
0.83354624, 0.53764314
5.11216128, 5.04950558
1.50028571, 0.00268594
0.50085714, 0.00278003
1, 0
-0.12818301, 0, 0.1655682, 0.20507744, -0.12818301, 1, 0
      0, -0.09654106, -0.13004267, 1, -0.12818301, 0
            0, 0, 0, 0, 0
                  0.97886916, -0.09654106, 0.1655682, 0
                          -0.13004267, 0.20507744, 0
                                  -0.12818301, 0
                                          0

```

Após rodar o programa ESPURI, obtém-se o arquivo DADOS4.OUT, que é impresso ou apresentado na tela do computador como resultado:

```

Y = X1 X4/X2  Z = C X5**2 X1**3/ X2**5 (Z=C2 X5**2 X7**3/X6**5)

      NI,NF,KI,KF          1          4          5          8

MEDIAS          DESVIACIONES ESTANDAR

      .501          .003
      1.500          .003
      1.000          .000
      .834          .538
      5.112          5.050
      1.500          .003
      .501          .003
      1.000          .000

CORRELACIONES CR(1,J) =  1.000  -.128  .000  .166  .205  -.128  1.000  .000
CORRELACIONES CR(2,J) = -.128  1.000  .000  -.097  -.130  1.000  -.128  .000
CORRELACIONES CR(3,J) =  .000  .000  1.000  .000  .000  .000  .000  .000

```

CORRELACIONES CR(4,J) = .166 -.097 .000 1.000 .979 -.097 .166 .000
CORRELACIONES CR(5,J) = .205 -.130 .000 .979 1.000 -.130 .205 .000
CORRELACIONES CR(6,J) = -.128 1.000 .000 -.097 -.130 1.000 -.128 .000
CORRELACIONES CR(7,J) = 1.000 -.128 .000 .166 .205 -.128 1.000 .000
CORRELACIONES CR(8,J) = .000 .000 .000 .000 .000 .000 .000 1.000

CORRELACION FINAL Ryz = .9789

CORRELACION ESPURIA = .0001

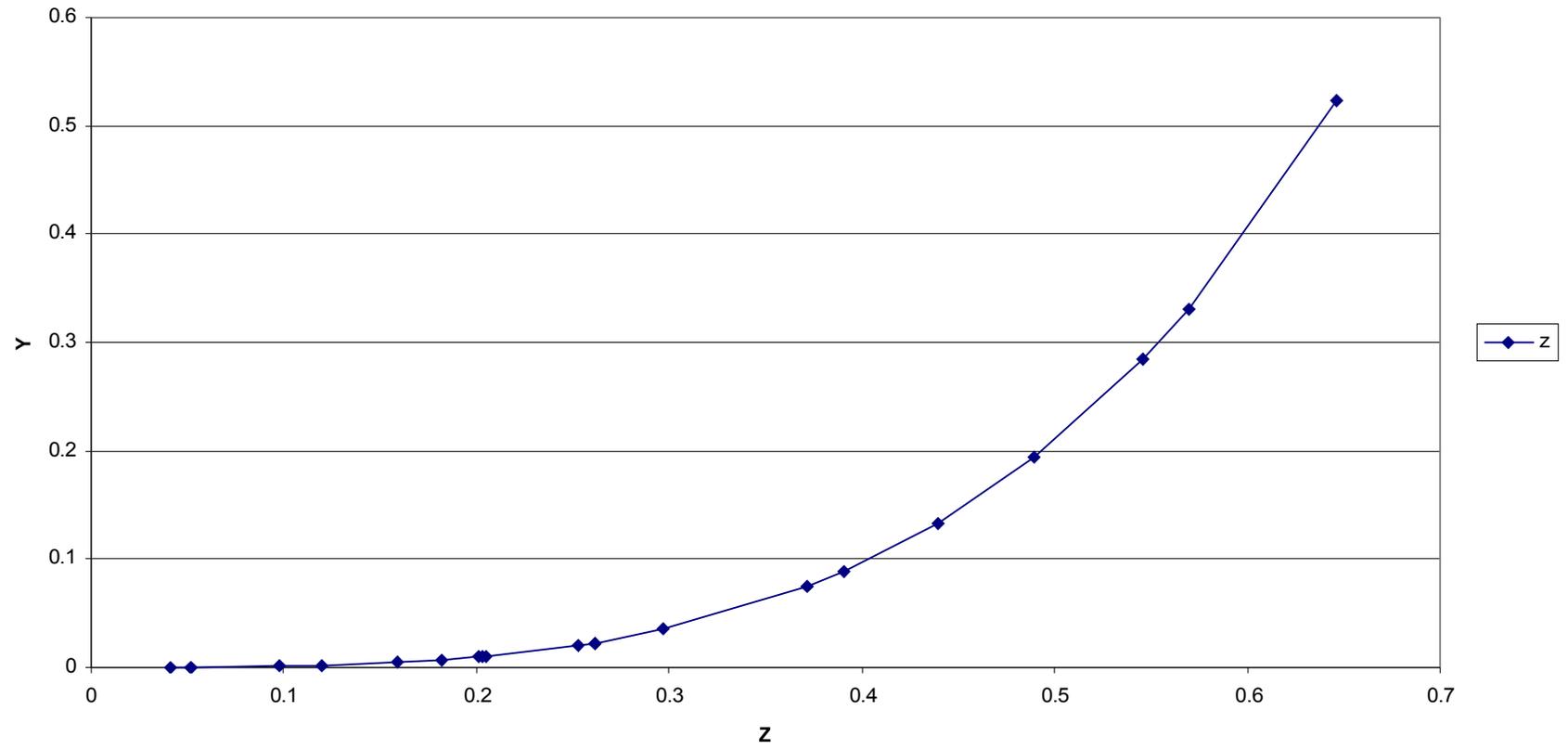
GRADO ESPURIO = .01 %

m	D (m)	b (m)	g (m/s ²)	n	A (m ²)	s	R (m)	v (m/s)	Q (m ³ /s)	y	z	Valores ordenados	
												y	z
0.5	0.122601234	1.5	9.81	0.013	0.191417382	0.003	0.107892768	0.954872747	0.182779241	0.040867078	5.60581E-05	0.04086708	5.6058E-05
0.5	0.154332547	1.5	9.81	0.013	0.243408089	0.003	0.131921491	1.091844967	0.265763896	0.051444182	0.000118516	0.05144418	0.00011852
0.5	0.783803246	1.5	9.81	0.013	1.482878633	0.003	0.455900391	2.495727999	3.700861723	0.261267749	0.022982138	0.05147979	0.00011878
0.5	1.937924972	1.5	9.81	0.013	4.784664056	0.003	0.820228315	3.691819393	17.66411555	0.645974991	0.523562466	0.0973583	0.00093264
0.5	1.171725412	1.5	9.81	0.013	2.444058338	0.003	0.59320974	2.974560389	7.269999121	0.390575137	0.088685665	0.11936648	0.00179923
0.5	0.154439372	1.5	9.81	0.013	0.243584818	0.003	0.132000186	1.092279131	0.266062613	0.051479791	0.000118782	0.15942185	0.00458046
0.5	0.784006902	1.5	9.81	0.013	1.483343764	0.003	0.455979552	2.496016891	3.702451092	0.261335634	0.023001882	0.18221808	0.00706477
0.5	0.603030379	1.5	9.81	0.013	1.086368388	0.003	0.381393742	2.215812375	2.407188517	0.201010126	0.009723104	0.20101013	0.0097231
0.5	0.616045248	1.5	9.81	0.013	1.113823745	0.003	0.3870778	2.237773456	2.492485211	0.205348416	0.010424372	0.20283723	0.01001425
0.5	0.608511699	1.5	9.81	0.013	1.097910791	0.003	0.383794509	2.225101262	2.442962687	0.202837233	0.010014249	0.20534842	0.01042437
0.5	1.468504822	1.5	9.81	0.013	3.281010439	0.003	0.685876306	3.276783955	10.75116236	0.489501607	0.193952576	0.2525858	0.02055562
0.5	0.478265538	1.5	9.81	0.013	0.83176727	0.003	0.323716113	1.986370026	1.652197573	0.159421846	0.004580456	0.26126775	0.02298214
0.5	1.637313472	1.5	9.81	0.013	3.796367912	0.003	0.735567104	3.433197657	13.03368142	0.545771157	0.285048601	0.26133563	0.02300188
0.5	1.317301157	1.5	9.81	0.013	2.843592904	0.003	0.639645703	3.127833304	8.894284588	0.439100386	0.13274151	0.2967225	0.03506496
0.5	0.358099445	1.5	9.81	0.013	0.601266773	0.003	0.261336856	1.722203165	1.03550354	0.119366482	0.001799235	0.37130874	0.0746243
0.5	0.757757402	1.5	9.81	0.013	1.423734243	0.003	0.445697331	2.458351412	3.500039087	0.252585801	0.020555616	0.39057514	0.08868567
0.5	0.890167509	1.5	9.81	0.013	1.73145036	0.003	0.496050059	2.640184511	4.571348421	0.296722503	0.035064961	0.43910039	0.13274151
0.5	0.292074892	1.5	9.81	0.013	0.480766209	0.003	0.223290308	1.550709773	0.745528859	0.097358297	0.000932639	0.48950161	0.19395258
0.5	1.113926221	1.5	9.81	0.013	2.291305146	0.003	0.574144702	2.910481542	6.668801332	0.37130874	0.0746243	0.54577116	0.2850486
0.5	0.546654238	1.5	9.81	0.013	0.969396785	0.003	0.356087438	2.116679206	2.051902017	0.182218079	0.007064769	0.56932845	0.33153148
0.5	1.707985338	1.5	9.81	0.013	4.020584963	0.003	0.755866792	3.49607536	14.05626802	0.569328446	0.331531478	0.64597499	0.52356247

Tabela No. 4.- Planilha de cálculo para o caso de estudo No. 1

Figura 4

Correlação entre Y e Z



	X1=m	X2=b	X5=Q	X4=d
	0.501	1.501	0.18277924	0.12260123
	0.5	1.5	0.2657639	0.15433255
	0.502	1.502	3.70086172	0.78380325
	0.5	1.5	17.6641156	1.93792497
	0.503	1.503	7.26999912	1.17172541
	0.5	1.5	0.26606261	0.15443937
	0.504	1.504	3.70245109	0.7840069
	0.5	1.5	2.40718852	0.60303038
	0.499	1.505	2.49248521	0.61604525
	0.5	1.5	2.44296269	0.6085117
	0.498	1.506	10.7511624	1.46850482
	0.5	1.5	1.65219757	0.47826554
	0.505	1.499	13.0336814	1.63731347
	0.5	1.5	8.89428459	1.31730116
	0.506	1.498	1.03550354	0.35809945
	0.5	1.5	3.50003909	0.7577574
	0.497	1.497	4.57134842	0.89016751
	0.5	1.5	0.74552886	0.29207489
	0.496	1.496	6.66880133	1.11392622
	0.5	1.5	2.05190202	0.54665424
	0.507	1.495	14.056268	1.70798534
media	0.50085714	1.50028571	5.11216128	0.83354624
desv. Padrao	0.00278003	0.00268594	5.04950558	0.53764314
coef. Cor.r11	1		0.97886916	0.97886916
coef. Cor.r12	-0.12818301			
coef. Cor.r15	0.20507744			
coef. Cor.r14	0.1655682			
coef. Cor.r22		1		
coef. Cor.r25		-0.13004267		
coef. Cor.r24		-0.09654106		
coef. Cor.r55			1	
coef. Cor.r54			0.97886916	
coef. Cor.r44				1

Tabela 5.- Planilha de cálculo eletrônica para os parâmetros estatísticos que integram o arquivo de ingresso para o programa ESPURI.

Observe-se que, para as variáveis $X_1 = m = 0,5$, $X_2 = b = 1,5$, mudam-se levemente os valores que deveriam ser constantes, com o intuito de evitar a divisão entre zero que resultaria do desvio padrão igual a zero para os dados originais. Para este caso a folha eletrônica retorna o erro #1DIV/0!, decorrente da formulação usada pelo EXCEL, segundo a qual o coeficiente de correlação entre duas variáveis calcula-se como

$$\rho_{xy} = \frac{Cov(X, Y)}{\sigma_x \sigma_y}, Cov(X, Y) = \frac{1}{n} \sum_{j=1}^n (x_j - \mu_x)(y_j - \mu_y),$$

onde μ_i = média dos valores do conjunto i ,

e

$$-1 \leq \rho_{xy} \leq 1.$$

Como informação adicional, em seqüência fornece-se uma tabela onde se calcula o número de Froude para os valores da Tabela No. 4; observando-se que o número de Froude é menor que 1, o que indica um escoamento sub-crítico. Usa-se a fórmula para canal retangular como uma aproximação só, desde que o erro incorrido por não usar a fórmula correta pra canal trapezoidal é desprezível desde que d não se aproxime do valor da largura do canal.

Tabela 6.- Cálculo do número de Froude para os valores da Tabela 4.

V	D	$F = \frac{v}{\sqrt{gd}}$
0.95487275	0.12260123	0.87069003
1.09184497	0.15433255	0.88735563
2.495728	0.78380325	0.90003463
3.69181939	1.93792497	0.84671541
2.97456039	1.17172541	0.87735501
1.09227913	0.15443937	0.88740142
2.49601689	0.7840069	0.90002189
2.21581238	0.60303038	0.91102204
2.23777346	0.61604525	0.91028064
2.22510126	0.6085117	0.91071148
3.27678396	1.46850482	0.86332758
1.98637003	0.47826554	0.91704597
3.43319766	1.63731347	0.85664007
3.1278333	1.31730116	0.87009477
1.72220317	0.35809945	0.91885727
2.45835141	0.7577574	0.90166324
2.64018451	0.89016751	0.89343694
1.55070977	0.29207489	0.91611271
2.91048154	1.11392622	0.8804448
2.11667921	0.54665424	0.91403785
3.49607536	1.70798534	0.85409113

Para efeito de comparação, repete-se o exercício anterior para o caso no qual as variáveis são todas representadas por números aleatórios, obtendo-se os resultados constantes da

Tabela 7.

Tabela 7.- Planilha de cálculo eletrônica para os parâmetros estatísticos que integram o arquivo de ingresso para o programa ESPURI para o exercício onde usam-se números aleatórios para todas as variáveis.

X1 = m	X4 = d	X2 = b	X5 = Q	$y = \frac{md}{b}$	$z = \frac{Q^2 m^3}{gb^5}$
	(m)	(m)	(m ³ /s)		
0.73361432	0.48275877	0.05222658	0.7848432	6.78119686	63803.1472
0.03759199	0.38516737	0.12860559	0.15635334	0.11258615	0.00376298
0.40226467	0.94006565	0.24680351	0.75744082	1.53221162	4.15726263
0.85268395	0.59602688	0.06855846	0.037577	7.4129807	58.9161333
0.42043845	0.79133739	0.5239859	0.25508907	0.63495729	0.01248027
0.33732313	0.93583523	0.27682864	0.44110471	1.14034033	0.46827139
0.06176617	0.04909059	0.94543987	0.30072833	0.00320712	2.8758E-06
0.81688282	0.33001988	0.37630709	0.66427272	0.71640311	3.24930971
0.81748661	0.3365787	0.92324776	0.03367633	0.29802247	9.4152E-05
0.7473388	0.75070269	0.42762674	0.94652654	1.31196016	2.66578373
0.82564742	0.86242185	0.93351262	0.04849228	0.76277102	0.00019031
0.89462618	0.05630691	0.23875431	0.45399568	0.21098523	19.3911303
0.33502837	0.20147168	0.01742892	0.90293875	3.87279942	1943297.9
0.52490831	0.86471355	0.38079365	0.97279189	1.19197188	1.74249968
0.33582774	0.56309883	0.9857992	0.00046505	0.19182832	8.9688E-10
0.37551443	0.40494122	0.74524922	0.77857226	0.20404083	0.01423315
0.13858546	0.13901999	0.30396406	0.56944851	0.06338299	0.0339064
0.25913149	0.52669063	0.44234735	0.76510939	0.30854062	0.06130861
0.14172856	0.83995974	0.50468112	0.35060963	0.23588417	0.0010896
0.12756351	0.45595268	0.68462362	0.54854518	0.08495606	0.00042333
0.89985972	0.48602703	0.47356083	0.12255101	0.92354797	0.0468393
0.48027677	0.5237232	0.46096881	0.47100627		
0.30241132	0.28401037	0.3066874	0.32938789		
1					coef. Cor.r11
		-0.08184906			coef. Cor.r12
			-0.11091312		coef. Cor.r15
	0.07296142				coef. Cor.r14
		1			coef. Cor.r22
			-0.41595549		coef. Cor.r25
	0.0020379				coef. Cor.r24
			1		coef. Cor.r55
	0.03004759				coef. Cor.r54
	1				coef. Cor.r44

Para usar o programa ESPURI, é necessário formar o arquivo de entrada, como nos casos anteriores. Os dados para este arquivo obtêm-se da planilha de cálculo eletrônica apresentada na Tabela 7. Forma-se o arquivo de entrada concordando com a ordem estabelecida no Capítulo 4. Para avaliar os parâmetros estatísticos que vão formar o arquivo, lembre-se que, em geral, $r_{ij} = r_{ji}$, que em particular para este caso, $X_2 = X_6, X_1 = X_7$, que X_3, X_8 não se definem, e por tanto a média m_i e o desvio padrão σ_i de estas variáveis são iguais a 1 e 0, respectivamente. Todos os coeficientes de correlação associados às variáveis não definidas são iguais a zero. O ordenamento é como segue:

Título da rodada

$$C_1 = 1, p = 1, q = -1, r = 0, s = 1$$

$$C_2 = 0.1019, t = 2, u = -5, v = 3, w = 0$$

$$m_1 = 0.48027677, \sigma_1 = 0.30241132$$

$$m_2 = 0.46096881, \sigma_2 = 0.3066874$$

$$m_3 = 1, \sigma_3 = 0$$

$$m_4 = 0.5237232, \sigma_4 = 0.28401037$$

$$m_5 = 0.47100627, \sigma_5 = 0.32938789$$

$$m_6 = 0.46096881, \sigma_6 = 0.3066874$$

$$m_7 = 0.48027677, \sigma_7 = 0.30241132$$

$$m_8 = 1, \sigma_8 = 0$$

$$r_{12} = -0.08184906, r_{13} = 0, r_{14} = 0.07296142, r_{15} = -0.11091312, r_{16} = -0.08184906, r_{17} = 1, r_{18} = 0$$

$$r_{23} = 0, r_{24} = 0.0020379, r_{25} = -0.41595549, r_{26} = 1, r_{27} = -0.08184906, r_{28} = 0$$

$$r_{34} = 0, r_{35} = 0, r_{36} = 0, r_{37} = 0, r_{38} = 0$$

$$r_{45} = 0.03004759, r_{46} = 0.0020379, r_{47} = 0.07296142, r_{48} = 0$$

$$r_{56} = -0.41595549, r_{57} = -0.11091312, r_{58} = 0$$

$$r_{67} = -0.08184906, r_{68} = 0$$

$$r_{78} = 0$$

Ordenando os dados segundo o formato requerido em FORTRAN para ingresso ao programa, obtém-se o seguinte arquivo, nomeado neste caso particular DADOS5.TXT :

```

Y = X1 X4/X2  Z = C X5**2 X1**3/ X2**5 (Z=C2 X5**2 X7**3/X6**5)
1, 1, -1, 0, 1
0.1019, 2, -5, 3, 0
0.48027677, 0.30241132
0.46096881, 0.3066874
1, 0
0.5237232, 0.28401037
0.47100627, 0.32938789
0.46096881, 0.3066874
0.48027677, 0.30241132
1, 0
-0.08184906, 0, 0.07296142, -0.11091312, -0.08184906, 1, 0
      0, 0.0020379, -0.41575549, 1, -0.08184906, 0
            0, 0, 0, 0, 0
                  0.03004759, 0.0020379, 0.07296142, 0
                          -0.41595549, -0.11091312, 0
                                  -0.08184906, 0
                                          0

```

Após rodar o programa ESPURI, obtém-se o arquivo DADOS5.OUT, que é impresso ou apresentado na tela do computador como resultado:

```

Y=X1 X4/X2  Z = C X5**2 X1**3/X2**5 (Z=C2 X5**2 X7**3/X6**5)

```

```

NI,NF,KI,KF          1          4          5          8

```

```

MEDIAS          DESVIACIONES ESTANDAR

```

```

      .480          .302
      .461          .307
1.000          .000
      .524          .284
      .471          .329
      .461          .307
      .480          .302
1.000          .000

```

CORRELACIONES CR (1, J) = 1.000 -.082 .000 .073 -.111 -.082 1.000 .000
 CORRELACIONES CR (2, J) = -.082 1.000 .000 .002 -.416 1.000 -.082 .000
 CORRELACIONES CR (3, J) = .000 .000 1.000 .000 .000 .000 .000 .000
 CORRELACIONES CR (4, J) = .073 .002 .000 1.000 .030 .002 .073 .000
 CORRELACIONES CR (5, J) = -.111 -.416 .000 .030 1.000 -.416 -.111 .000
 CORRELACIONES CR (6, J) = -.082 1.000 .000 .002 -.416 1.000 -.082 .000
 CORRELACIONES CR (7, J) = 1.000 -.082 .000 .073 -.111 -.082 1.000 .000
 CORRELACIONES CR (8, J) = .000 .000 .000 .000 .000 .000 .000 1.000

CORRELACION FINAL Ryz = .7942

CORRELACION ESPURIA = .7848

GRADO ESPURIO = 98.81 %

O exame dos resultados confere a hipótese inicial de que o uso indiscriminado de números aleatórios pode conduzir a um alto grau de espureidade. Na segunda rodada, pode-se observar a obtenção de uma porcentagem de espureidade de 98.81 %, bem maior que a porcentagem de espureidade de 0.01 % obtida para dados que, mesmo baseados em números aleatórios para a grandeza da profundidade (d), são aplicados com um critério racional como já foi exposto anteriormente. O resultado da primeira rodada sugere que os

números adimensionais $y = \frac{md}{b}$ e $z = \frac{Q^2 m^3}{gb^5}$ propostos para formar a curva chave podem

ter uma relação funcional válida que reflita a existência de verdadeiras relações causa-efeito

que descrevem o fenômeno físico do escoamento num canal aberto. Estas relações causa-efeito devem ser demonstradas através de pesquisas de laboratório. O resultado da segunda rodada sugere que a qualidade dos dados experimentais é muito importante.

5.3 Caso de estudo No. 2. Geração de ondas num canal reproduzidor de ondas.

Outro exemplo que pode ser tratado seria estudar a correlação entre dois grupos adimensionais num caso de estudo de geração de ondas para um projeto de hidráulica marítima, que são:

$$\frac{h_5}{d} \left(\frac{E}{\gamma} \right)^{1/3} \frac{1}{d}$$

onde

d = profundidade de escoamento (m)

h_5 = altura de onda gerada para uma distância “D” horizontal de 5 vezes a profundidade de escoamento no ponto considerado (m)

E = valor da energia usada na geração da onda $\left(\frac{\text{kgfm}}{\text{m}} \right)$

γ = peso específico da água, tomado como $\gamma = 1000 \frac{\text{kgf}}{\text{m}^3}$

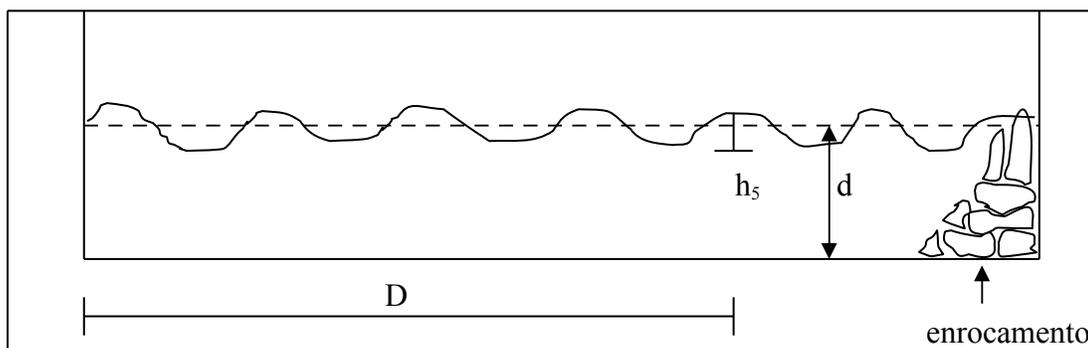


Figura 5.- Seção longitudinal de um canal reprodutor de ondas.

O caso seria completamente similar ao caso de estudo proposto anteriormente, e cita-se somente como possível tema de pesquisa. Na Figura 5 ilustra-se o canal de ondas que seria necessário para este estudo de hidráulica marítima. Na seqüência apresentam-se recomendações de desenho para a construção do canal de ondas.

Considere-se o desenho dum quebramar que deverá funcionar para uma profundidade que varia de 5 a 25 m, cuja largura deve ser de 10 m e cuja base deverá ter uma largura de 120 m (vide Figura 6).

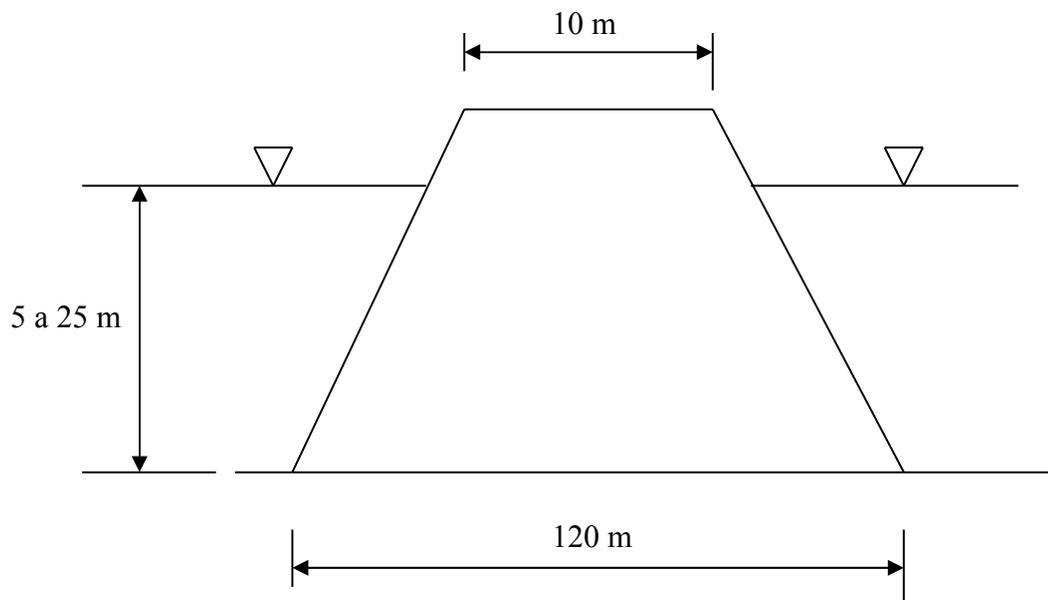


Figura 6.- Elevação do quebramar proposto com dimensionamento.

Em continuação apresentam-se os delineamentos propostos para construir o canal de ondas necessário para este caso de estudo. Aconselha-se, devido às características do quebra-mar, que seu modelo se construa com uma escala de 1 : 25. Concordando com esta escala, o dimensionamento do canal de ondas deve ser como segue:

profundidade: 1m

largura: 0,60 m

comprimento: 30 a 40 m

sobreelevação da parede do canal: 1 m

dimensionamento do batedor (articulado na base para movimento circular ou com capacidade de movimento na direção horizontal para agir como pistão) para produzir as ondas: 1,40 m × 0,60 m

O equipamento necessário para o funcionamento do canal deve ser:

1 ou mais pistões para movimentar o batedor com velocidade de 1 m/s

1 servo-controle para acoplar o motor com o batedor

sensores de níveis

sensores de pressão

sensores de velocidade de água no canal

1 motor de 10 a 40 CV de potência

1 carro viajante para acomodar os sensores

painéis de cristal temperado de 2 m de altura × 2 a 3 m de comprimento para instalar-se numa parede do canal para fotografar as ondas geradas

Aclara-se que este projeto é válido para estudar a parte central do quebra-mar. Se for necessário estudar a ponta do quebra-mar, deve planejar-se um modelo onde se possam gerar ondas multidirecionadas.

Apresenta-se, abaixo, uma planta tentativa do canal de ondas incluindo o modelo do quebra-mar

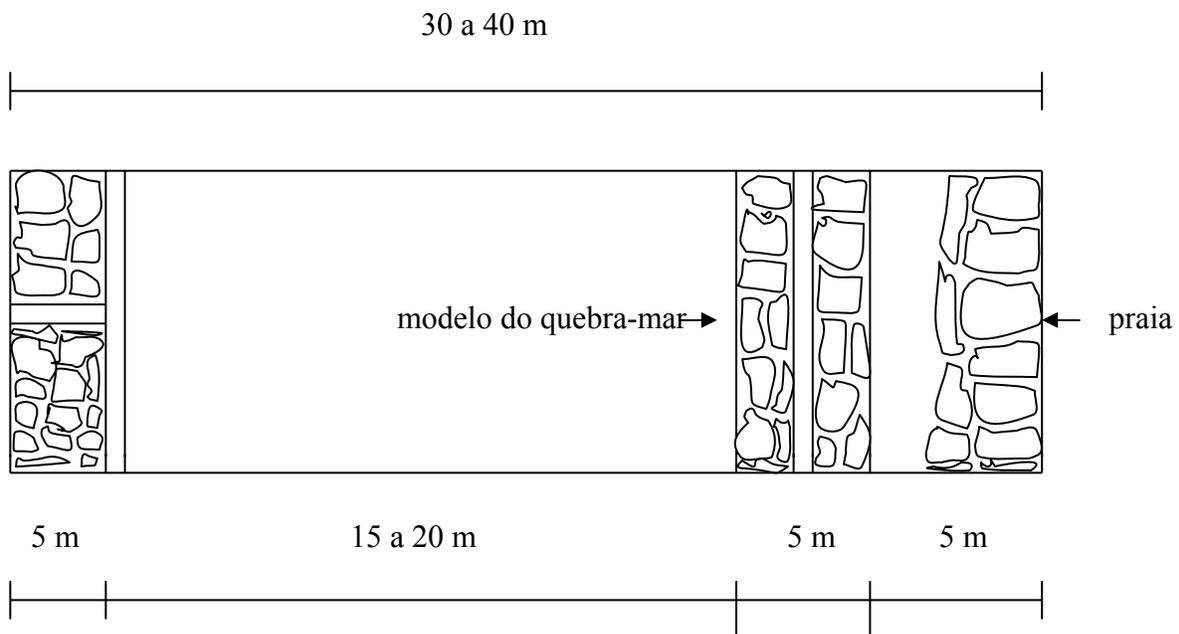


Figura 7.- Vista em planta do canal de ondas.

5.4 Caso de estudo No. 3. Vertedouro triangular

Considere-se um vertedouro triangular de crista delgada (Figura 8) cuja vazão estima-se em função dos parâmetros:

$$Q = f(H, g, v_0, \Phi)$$

onde

Q = vazão

H = carga hidráulica do vertedouro

g = constante gravitacional

v_0 = velocidade de aproximação do escoamento de água

Φ = ângulo de abertura do vertedouro

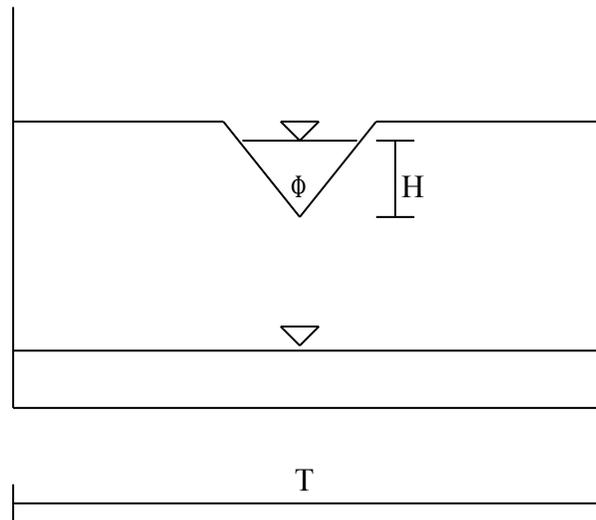


Figura 8.- Seção transversal de um vertedouro triangular de crista delgada.

Para estabelecer uma correlação entre variáveis, propõem-se os seguintes grupos adimensionais:

$$\frac{Q}{v_0 H^2} = f(F, \Phi)$$

onde

$$F = \text{número de Froude} = \frac{v_0}{\sqrt{gH}}$$

Usando a notação proposta no Capítulo 2,

$$y = C_1 X_1^p X_2^q X_3^r X_4^s$$

$$z = C_2 X_5^t X_6^u X_7^v X_8^w$$

esta relação pode exprimir-se como

$$\frac{X_1}{X_2 X_3^2} = f\left(\frac{X_2}{X_5^{\frac{1}{2}} X_3^{\frac{1}{2}}}\right)$$

onde

$$X_1 = Q$$

$$X_2 = v_0$$

$$X_3 = H$$

$$X_5 = g$$

e os expoentes são:

$$p = 1, q = -1, r = -2, t = -\frac{1}{2}$$

sendo os expoentes restantes iguais a zero.

Para realizar o experimento matemático, usam-se números aleatórios para estimar H e depois se estima Q com uma equação apropriada.

Desde que $Q = f(H)$ podem usar-se as equações apropriadas segundo Chow (1959) para uma seção triangular.

Depois se constrói um gráfico de y contra z para observar a tendência da correlação.

Para a construção de o vertedouro triangular de crista delgada, podem-se seguir as seguintes recomendações (French, 1985):

Referindo-se à Figura 9:

1. $H/p \leq 1,2$
2. $H/T \leq 0,4$
3. $0,049 \text{ m} \leq H \leq 0,61 \text{ m}$
4. $p \geq 0,09 \text{ m}$
5. $T \geq 0,61 \text{ m}$
6. O nível de água a jusante do vertedouro deve situar-se embaixo do seu vértice.

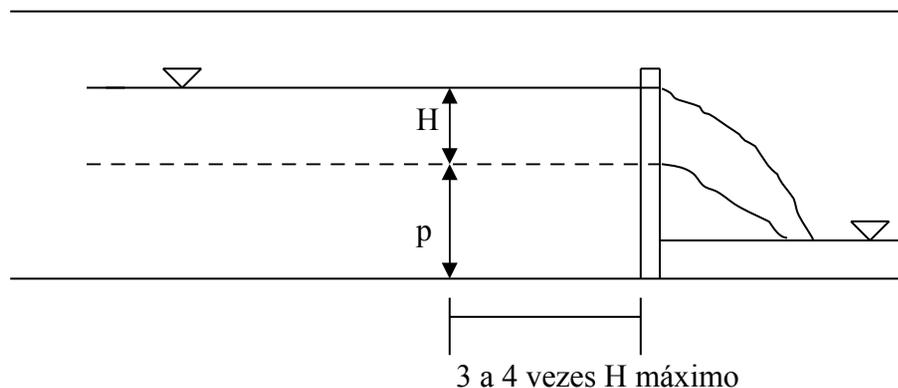


Figura 9.- Seção longitudinal de um vertedouro triangular de crista delgada instalado num canal retangular.

Para a aferição de o vertedouro triangular, deve se considerar que o escoamento no vertedouro seja parcial ou totalmente contraído para a apropriada determinação dos

coeficientes de contração na equação do vertedouro, $Q = C \tan \frac{\phi}{2} H^{\frac{5}{2}}$. Pode se usar a

seguinte tabela para esta determinação (French, 1985):

Tabela 8.- Classificação do escoamento de um vertedouro triangular de crista delgada.

contração parcial	contração total
$H/p \leq 1,2$	$H/p \leq 0,4$
$H/T \leq 0,4$	$H/T \leq 0,2$
$0,049 \text{ m} < H \leq 0,61 \text{ m}$	$0,049 \text{ m} < H \leq 0,38 \text{ m}$
$p \geq 0,09 \text{ m}$	$p \geq 0,46 \text{ m}$
$T \geq 0,61 \text{ m}$	$T \geq 0,91 \text{ m}$

CONCLUSÕES

Neste trabalho apresentaram-se exemplos do uso de índices $\left(\frac{X_i}{X_j}\right)$ para estandarizar dados de qualquer natureza com o intuito de fazer comparações entre grandezas que não poderiam comparar-se em forma absoluta ou com o intuito de estabelecer relações matemáticas que permitam fazer inferências ou extrapolações para avaliar grandezas ou variáveis que sejam de interesse no campo da hidráulica. Embora o uso de índices seja válido do ponto de vista matemático e mesmo atraente como método de pesquisa, este uso deve ser considerado com cautela, desde que exista a possibilidade da incidência do fenômeno da correlação espúria, que pode ser associada a vários fatores, como por exemplo, a mesma estrutura matemática da relação proposta, que em alguns casos pode ser, no fundo, uma autocorrelação de uma variável consigo mesma, o que torna a relação proposta irrelevante, ou, por exemplo, o uso de dados que propiciem o aumento desmedido do coeficiente de correlação espúria, desde que não são dados congruentes com o comportamento físico do fenômeno estudado, ou são dados muito escassos, o que requererá uma coleta adicional de dados para aumentar a base de dados que vai se utilizar, ou a existência de pontos extremos (*outliers*) que influem notoriamente na tendência observada no resto dos dados, ou ao fato de que a relação proposta, quer inclua números adimensionais ou não, tente relacionar variáveis que não têm relação nenhuma do ponto de vista experimental. Antes de tentar trabalhar com uma relação entre variáveis, mesmo que estas variáveis se repitam ou não, resulta de importância fundamental descobrir a existência de relações causa – efeito verdadeiras e demonstráveis entre as variáveis consideradas. Estas relações são tão sutis, que pode acontecer, por exemplo, que uma relação proposta entre duas variáveis não descreva adequadamente o

fenômeno estudado devido a que uma terceira variável, obviamente não considerada na relação, tem sido ignorada completamente, embora esta variável seja relevante para o fenômeno estudado. Isto significa que o pesquisador deve ter um conhecimento completo e profundo do fenômeno estudado antes de tentar propor qualquer relação, o que leva à conclusão que a pesquisa experimental torna-se fundamental para entender adequadamente um processo.

Outrossim, a aplicação mecânica das técnicas de correlação linear ou não linear, sem refletir sobre o verdadeiro significado das variáveis que tentam-se correlacionar e as interações entre elas, conduz à obtenção de valores altos dos diversos coeficientes de correlação que podem calcular-se com as ferramentas habituais da matemática e da estatística, promovendo assim a enganosa idéia de que um valor alto do coeficiente de correlação calculado é uma garantia de validação da lei física representada pela equação matemática considerada. Existe na literatura técnica uma tendência a tentar classificar sistematicamente os resultados de uma análise de correlação em forma indiscriminada. Por exemplo, pode-se pensar numa classificação arbitrária do valor do coeficiente de correlação como segue:

Tabela 9.- Classificação do coeficiente de correlação.

Valor absoluto do coeficiente de correlação	Classificação
0,90 – 1,00	Correlação muito alta
0,70 – 0,89	Correlação alta
0,40 – 0,69	Correlação moderada
0,20 – 0,39	Correlação baixa
0,00 – 0,19	Correlação muito baixa ou ausente

Embora esta classificação possa ser de utilidade, é irrelevante se for desligada das considerações de causalidade acima mencionadas. O programa desenvolvido no Capítulo 4

deste trabalho permite calcular o valor do coeficiente de correlação espúria isoladamente do valor do coeficiente de correlação total, assim como o grau de espureidade como porcentagem. Este programa tenta ser um instrumento útil para decidir se uma relação proposta na qual existem variáveis repetidas deverá ser descartada ou pode ser aceita como válida. Porém, somente o pesquisador pode decidir, baseando seu juízo objetivamente nas evidências acumuladas decorrentes do conhecimento que tenha do fenômeno estudado, assim como em sua experiência, tanto experimental como científica e tecnológica.

SUGESTÕES PARA FUTURAS PESQUISAS

Neste capítulo apresentam-se idéias para futuras pesquisas, que requereriam projetos independentes para serem realizadas ou que não foram realizadas por estarem fora do escopo do tema da presente tese.

O primeiro tema proposto para futuras pesquisas é a influência que diversas variáveis comumente consideradas nos fenômenos da Hidráulica podem ter na quantificação do grau de correlação espúria segundo sua natureza contínua ou discreta assim como suas propriedades estatísticas, por exemplo, o grau de dispersão que a variável considerada pode ter medida pelo desvio padrão σ dos dados considerados. Exemplos de variáveis contínuas típicas no estudo da Hidráulica são a velocidade de um fluido, a profundidade de escoamento, a pressão manométrica ou hidrostática, as forças estáticas ou dinâmicas geradas nos fenômenos estudados, a força de cisalhamento gerada numa fronteira física pelo escoamento, a vazão de um escoamento, o coeficiente de infiltração considerado num escoamento em meio poroso, a temperatura do fluido para estudos da convecção de calor, etc. Exemplos de variáveis discretas são as variáveis geométricas que descrevem os condutos para o escoamento de um fluido tais como o diâmetro de um tubo, a largura de um canal, o talude de um canal, a declividade de um canal ou de um tubo, o ângulo de abertura de um vertedouro triangular, o comprimento de um canal e mesmo diversas grandezas lineares que descrevem a geometria de um conduto; variáveis que descrevem a natureza física dos condutos, como o coeficiente de rugosidade de um canal ou de um tubo, a altura das asperezas de uma parede de um conduto, a porosidade de um meio poroso, etc.; variáveis que descrevem as propriedades do fluido considerado, como a viscosidade dinâmica ou cinemática, o peso específico e a densidade do fluido, a tensão superficial de

um fluido, etc.; variáveis que caracterizam o entorno físico do fenômeno estudado, tais como a constante gravitacional, o albedo, a intensidade de radiação solar, etc. As variáveis contínuas caracterizam-se pela sua grande variabilidade nos valores adotados durante o fenômeno estudado, o que pode conduzir a uma grande dispersão, enquanto que as variáveis discretas caracterizam-se por ser valores fixos que não mudam ou mudam muito pouco ou a longo prazo, tendo então uma dispersão muito pequena ou nula. Desde uma perspectiva da estatística, onde as variáveis freqüentemente são consideradas variáveis aleatórias com uma distribuição associada de densidade de probabilidade e suscetíveis de serem caracterizadas por parâmetros estatísticos, pode-se estudar a influência do comportamento destas variáveis na avaliação do grau de correlação total ou espúria.

O segundo tema proposto para futuras pesquisas é o estudo do efeito dos valores dos expoentes das variáveis consideradas nos parâmetros adimensionais que se tenta correlacionar, $y = C_1 X_1^p X_2^q X_3^r X_4^s$ e $z = C_2 X_5^t X_6^u X_7^v X_8^w$. Se consideramos o caso de estudo No. 1 do Cap. 5, pode-se observar que a lei matemática de correlação obtida na Figura 1 deve-se ao efeito da correlação espúria procedente das variáveis repetidas nos grupos adimensionais. Também conclui-se que existe a possibilidade de que o efeito da correlação espúria se amplificar pelo uso de variáveis elevadas a potências grandes. Esta hipótese

ilustra-se no caso de estudo No. 1, onde se tem a variável X_2 na função $z = C_2 \frac{X_3^2 X_1^3}{X_2^5}$ elevada à quinta potência. Então, deve pesquisarse a possibilidade de que se o valor de um expoente numa variável repetida for muito grande, amplifica-se o efeito da correlação espúria. Propõe-se realizar experimentos onde os expoentes considerados para pesquisa mudem dentro de um intervalo significativo de valores, tentando apreciar não somente o

efeito numérico derivado da mudança do valor do expoente, mas também as possíveis relações causa – efeito que possam ser introduzidas no experimento, desde que as grandezas consideradas freqüentemente representam fenômenos físicos.

O terceiro tema proposto é o de estabelecer critérios para a avaliação de gráficos propostos que representem curvas de correlação entre parâmetros dimensionais, com o intuito de analisar o gráfico para desmascarar efeitos causados pela correlação espúria introduzida por variáveis repetidas, assim como propor limites numéricos de confiabilidade baseados na análise estatística da correlação proposta com a metodologia exposta neste trabalho. Uma proposta de limites de confiabilidade é sempre muito subjetiva, desde que um valor de correlação total ou espúria não pode ser qualificado diretamente, mas devem considerar-se as condições sob as quais dito valor foi obtido, e também muito importante: as relações causa – efeito que governam o fenômeno estudado devem estar plenamente identificadas para saber, com certeza, os efeitos tanto físicos como funcionais que serão introduzidos na relação proposta por mudanças nas variáveis ou nos expoentes das variáveis.

O quarto tema proposto é a pesquisa dos intervalos de valores válidos para diferentes grandezas que intervêm na formulação para calcular o coeficiente de correlação espúria. Por exemplo, deve se pesquisar qual deve ser a magnitude dos coeficientes de variação que permita que as aproximações desenvolvidas para os coeficientes de correlação compostos sejam válidas para um caso específico. ¿Quão grandes podem ser os coeficientes de variação sem invalidar as expressões para a média, o desvio padrão, e os coeficientes de correlação para os índices propostos? Outra questão seria pesquisar os limites de confiança para estimar o valor de um elemento individual baseando-se numa relação espúria entre índices propostos. Por exemplo, se não houver correlação entre os elementos básicos, os limites de confiança tem uma largura infinita (Benson, 1965).

Para ilustrar a influência dos valores relativos dos coeficientes de correlação, apresenta-se o seguinte exemplo (Benson, 1965), onde se calcula o valor do coeficiente de correlação entre dois índices, que são

$$\frac{X_1}{X_3}, \frac{X_2}{X_3}$$

Na seguinte tabela, mostra-se a variação do coeficiente de correlação com respectu as variações dos coeficientes de variação:

Tabela 10.- Influência do valor relativo do coeficiente de variação sobre o coeficiente ρ .

Coeficientes de variação C_i	Coeficiente de correlação ρ
$C_1 = C_2 = C_3$	0,50 ($\frac{1}{2}$)
$2C_1 = 2C_2 = C_3$	0.80 ($\frac{4}{5}$)
$3C_1 = 3C_2 = C_3$	0.90 ($\frac{9}{10}$)

REFERÊNCIAS BIBLIOGRÁFICAS

BENJAMIN, J.R. & CORNELL, C.A. *Probability, Statistics and Decision for Civil Engineers*, New York: McGraw-Hill, 1970.

BENSON, M.A. Spurious Correlation in Hydraulics and Hydrology, *JOURNAL of the HYDRAULICS DIVISION, ASCE*, v.91, Ser. HY, n.4, p.35 - 42, Jul. 1965.

BOWMAN, C.C. & HANSEN, V.E. Simplification of dimensional analysis. *JOURNAL of the ENGINEERING MECHANICS DIVISION, ASCE*, v.85, Ser. EM, n.1, p.67 – 73, Jan. 1959.

BREMAUNTZ MONGE, C.A. *Correlación espuria en la ingeniería experimental*. Tesis de Maestría. Facultad de Ingeniería. UNAM. México: UNAM, 1974.

CHAYES, F. On Ratio Correlation in Petrography. *JOURNAL of GEOLOGY*, v.57, n.3, p.239 – 254, May 1949.

CHOW, V. T. *Open-channel hydraulics*. New York, McGraw-Hill, 1959.

ECHÁVEZ G., Bremauntz C. Correlación Espuria en la Ingeniería Hidráulica, *In: CONGRESO NACIONAL de HIDRÁULICA*, 4º, Acapulco, México, 1976. Memorias. Acapulco, 1976.

EZEQUIEL, M., & FOX, K. A., *Methods of Correlation and Regression Analysis*. New York, John Wiley, 1959.

FARRAR, D. E. & GLAUBER, R.R. Multicollinearity in Regression Analysis: The Problem Revisited, *The REVIEW of ECONOMICS and STATISTICS*, v.49, n. 1, p. 92-107 Febr. 1967.

FRENCH, Richard H. *Open - Channel Hydraulics*, New York: McGraw – Hill, 1985.

HELMLINGER, K. R. The use of spurious correlation in visualizing correlation coefficients, *COMPUTERS in EDUCATION, ASEE*, v. 9, n2, p. 8 – 13 Abr/Jun. 1989.

HELY, A. G. & OLMSTED, F. H. Some relations between streamflow characteristics and the environment in the Delaware river basin. *PROFESSIONAL PAPER. GEOLOGICAL SURVEY, U. S. DEPARTMENT of the INTERIOR*, Washington, D. C. n.417-B, 1963.

KRONMAL, Richard A. Spurious Correlation and the Fallacy of the Ratio Standard Revisited, *JOURNAL of the ROYAL STATISTICAL SOCIETY, Royal Statistical Society, UK*, Series A, v.156 n.3, p.379 – 392, 1993.

LANGBEIN, Walter B. Section: Fitting Curves to Cyclic Data. Double-Mass Curves, *WATER SUPPLY PAPER. GEOLOGICAL SURVEY. U.S. DEPARTMENT of the INTERIOR*, Washington D. C., n.1541 – B, p. 59 – 66, 1960.

_____. & **ISERI, Kathleen T.** General Introduction and Hydrologic Definitions. *WATER SUPPLY PAPER, GEOLOGICAL SURVEY. U.S. DEPARTMENT of the INTERIOR*, Washington D. C., n.1541-A, 29 pp, 1960.

LEAMER, E. E. Multicollinearity: A Bayesian Interpretation, *The REVIEW of ECONOMICS and STATISTICS*, v. 55, n.3, p. 371-380 Aug. 1973.

LINSLEY, R. K., KOHLER, Max A., & PAULHUS, Joseph L. H. *Hydrology for Engineers*, 2^a ed. New York, McGraw – Hill, 1975.

MAHMOOD, Khalid & SIDDIQUI, M. M. Spurious Correlation in Dimensional Analysis, *JOURNAL of the ENGINEERING MECHANICS DIVISION*, ASCE, v.106, Ser. EM, n.1, pp. 93 – 109, Febr. 1980.

MASON, R.L., GUNST, R.F., & WEBSTER, J.T. Regression Analysis and Problems of Multicollinearity, *COMMUNICATIONS in STATISTICS*, v.4, n.3, p.277 – 292, 1975.

MEZEI, Louis M. *Practical Spreadsheet Statistics & Curve Fitting for Scientists and Engineers*. New York: Prentice Hall, 1990.

MILLER, Irwin & FREUND, John E. *Probability and Statistics for Engineers*, 3a ed., New Jersey: Prentice Hall, 1985.

MIRAJGAOKER, A.G., & CHARLU, K.L.N. Natural Roughness Effects in Rigid Open Channels. *JOURNAL of the HYDRAULICS DIVISION, ASCE*, v. 89, Ser. HY, n.5. p.29-44 Sept. 1963.

MONTGOMERY, Douglas C., & PECK, Elizabeth A. *Introduction to Linear Regression Analysis*, 2a ed., New York: John Wiley, 1992.

NAGIB ELIAN, Silvia. *Análise de regressão*. São Paulo, IME/USP, 1988.

PEARSON, Karl. Mathematical Contributions to the Theory of Evolution - On a Form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs. *PROCEEDINGS, ROYAL SOCIETY OF LONDON*, v. 60, p.489 - 502, 1896.

REED, L.J. On the correlation between any two functions and its application to the general case of spurious correlation. *MATHEMATICS. JOURNAL of the WASHINGTON ACADEMY of SCIENCES*, v.2, n. 19, p.449 - 455 Nov. 1921.

SEARCY, JAMES K. *Graphical Correlation of Gaging–Station Records*, *WATER SUPPLY PAPER. GEOLOGICAL SURVEY*, Washington DC, n.1541–C, p.67–100, 1960.

YALIN, M. S. Geometrical Properties of Sand Waves, *JOURNAL of the HYDRAULICS DIVISION, ASCE*, v. 90, Ser. HY, n.5, p.105 - 119 Sept. 1964.

_____. & **KAMPHUIS, J. W.** Theory of Dimensions and Spurious Correlation, *JOURNAL of HYDRAULIC RESEARCH, IAHR*, v. 9, n.2, p.248 – 265, 1971.

YEVDJEVICH, F. M. Fluctuations of wet and dry years, Part 1: Research data assembly and mathematical models. *HYDROLOGY PAPERS, COLORADO STATE UNIVERSITY*, Fort Collins, Colorado, 1963.

ANEXO

Introdução

Nesta parte apresenta-se um gráfico para obter o valor do coeficiente de correlação espúria segundo a fórmula de Pearson (Capítulo 3) em caso de não poder calcular este coeficiente analiticamente ou usando o programa de computador (Capítulo 4), porém somente para quatro modelos algébricos escolhidos arbitrariamente. A Figura 10 (Bremauntz, 1974; Echávez, 1976) permite esta solução gráfica. Na seqüência explica-se o uso da Figura 10. Para usar a solução gráfica, precisam ser calculadas duas variáveis indicativas baseadas no coeficiente de variação C de uso comum na Estatística e que define-se como:

$$C = \frac{\sigma}{m}$$

onde

C = coeficiente de variação

m = média

σ = desvio padrão

Para facilitar o uso da Figura 10, que é uma solução gráfica da fórmula de Pearson para quatro modelos algébricos considerados, apresenta-se a Tabela 1, onde se indicam os dois parâmetros requeridos para a estimativa gráfica do valor de ρ_{yz} , o coeficiente de correlação espúria para cada modelo, assim como o uso destes parâmetros na Figura 10. Cada parâmetro calcula-se como o valor absoluto da relação dos coeficientes de variação considerados segundo o modelo algébrico escolhido.

Esta figura consta de quatro curvas que estão rotuladas cada uma com um valor que pode ser 0, 1, 2, ou 3, que corresponde a um dos parâmetros que se calculam para a estimação

gráfica. Valores intermediários entre estes números inteiros interpolam-se visualmente entre as curvas. O outro parâmetro calculado se lê no eixo das abscissas da Figura 10. O coeficiente de correlação espúria ρ_{yz} lê-se no eixo das ordenadas, associando-lhe sinal positivo ou negativo segundo o modelo algébrico escolhido. O coeficiente ρ_{yz} varia de 0 a 1.

Exemplos de aplicação

Em continuação apresentam-se quatro exemplos de aplicação para ilustrar o uso do gráfico 10. Usam-se também as fórmulas analíticas já obtidas para os modelos algébricos escolhidos apresentados no Capítulo 3, com o intuito de verificar os resultados estimados graficamente.

Exemplo 1.- Cálculo do coeficiente de correlação espúria ρ_{yz} para o seguinte modelo algébrico e os correspondentes valores dos coeficientes de variação:

$$y = \frac{x_1}{x_2}$$

$$z = \frac{x_1}{x_3}$$

$$C_1 = 0,8$$

$$C_2 = 2,0$$

$$C_3 = 0,9$$

Verifica-se o resultado com a solução analítica baseada na fórmula de Pearson segundo a seguinte fórmula:

$$\rho_{yz} = \frac{C_1^2}{(C_1^2 + C_2^2)^{1/2} (C_1^2 + C_3^2)^{1/2}}$$

Substituindo os correspondentes valores, obtém-se

$$\rho_{yz} = \frac{0,8^2}{(0,8^2 + 2,0^2)^{1/2}(0,8^2 + 0,9^2)^{1/2}} = \frac{0,64}{2,937} = 0,2467$$

Calculam-se os parâmetros indicativos segundo a Tabela 11:

$$\left| \frac{C_2}{C_1} \right| = \left| \frac{2,0}{0,8} \right| = 2,5$$

$$\left| \frac{C_3}{C_1} \right| = \left| \frac{0,9}{0,8} \right| = 1,125$$

Na Figura 10, o valor de 2,5 interpola-se visualmente entre as curvas rotuladas 2 e 3, e depois lê-se o valor de 1,125 no eixo das abscisas. O ponto de cruzamento das referências corresponde nas ordenadas o valor estimado $\rho_{yz} = +0,2467$.

Exemplo 2.- Cálculo do coeficiente de correlação espúria ρ_{yz} para o seguinte modelo algébrico e os correspondentes valores dos coeficientes de variação:

$$y = \frac{x_5}{x_4}$$

$$z = \frac{x_6}{x_4}$$

$$C_4 = 0,8$$

$$C_5 = 0,5$$

$$C_6 = 0,6$$

Verifica-se o resultado com a solução analítica baseada na fórmula de Pearson segundo a seguinte fórmula:

$$\rho_{yz} = \frac{C_4^2}{(C_5^2 + C_4^2)^{1/2} (C_4^2 + C_6^2)^{1/2}}$$

Substituindo os correspondentes valores, obtém-se

$$\rho_{yz} = \frac{0,8^2}{(0,5^2 + 0,8^2)^{1/2} (0,8^2 + 0,6^2)^{1/2}} = \frac{0,64}{0,9434} = 0,6784$$

Calculam-se os parâmetros indicativos segundo a Tabela 11:

$$\left| \frac{C_5}{C_4} \right| = \left| \frac{0,5}{0,8} \right| = 0,625$$

$$\left| \frac{C_6}{C_4} \right| = \left| \frac{0,6}{0,8} \right| = 0,75$$

Na Figura 10, o valor de 0,625 interpola-se visualmente entre as curvas rotuladas 0 e 1, e depois lê-se o valor de 0,75 no eixo das abscissas. O ponto de cruzamento das referências indica nas ordenadas a estimativa $\rho_{yz} = +0,6784$.

Exemplo 3.- Cálculo do coeficiente de correlação espúria ρ_{yz} para o seguinte modelo algébrico e os correspondentes valores dos coeficientes de variação:

$$y = \frac{x_7}{x_8}$$

$$z = \frac{x_8}{x_9}$$

$$C_7 = 0,5$$

$$C_8 = 0,4$$

$$C_9 = 0,6$$

Verifica-se o resultado com a solução analítica baseada na fórmula de Pearson segundo a seguinte fórmula:

$$\rho_{yz} = \frac{-C_8^2}{(C_7^2 + C_8^2)^{1/2}(C_8^2 + C_9^2)^{1/2}}$$

Substituindo os correspondentes valores, obtém-se

$$\rho_{yz} = \frac{-(0,4)^2}{(0,5^2 + 0,4^2)^{1/2}(0,4^2 + 0,6^2)^{1/2}} = \frac{-0,16}{0,4617} = -0,3465$$

Calculam-se os parâmetros indicativos segundo a Tabela 11:

$$\left| \frac{C_7}{C_8} \right| = \left| \frac{0,5}{0,4} \right| = 1,25$$

$$\left| \frac{C_4}{C_8} \right| = \left| \frac{0,6}{0,4} \right| = 1,50$$

Na Figura 10, o valor de 1,25 interpola-se visualmente entre as curvas rotuladas 1 e 2, e depois lê-se o valor de 1,50 no eixo das abscissas. O ponto de cruzamento das referências indica nas ordenadas a estimativa $\rho_{yz} = -0,3465$.

Exemplo 4.- Cálculo do coeficiente de correlação espúria ρ_{yz} para o seguinte modelo algébrico e os correspondentes valores dos coeficientes de variação:

$$y = \frac{x_{10}}{x_{11}}$$

$$z = x_{10}$$

$$C_{10} = 0,50$$

$$C_{11} = 0,20$$

Verifica-se o resultado com a solução analítica baseada na fórmula de Pearson segundo a seguinte fórmula:

$$\rho_{yz} = \frac{C_{10}}{(C_{10}^2 + C_{11}^2)^{1/2}}$$

Substituindo os correspondentes valores, obtém-se

$$\rho_{yz} = \frac{0,5}{(0,5^2 + 0,2^2)^{1/2}} = \frac{0,50}{0,5385} = 0,9285$$

Calculam-se os parâmetros indicativos segundo a Tabela 11:

$$\left| \frac{C_{11}}{C_{10}} \right| = \left| \frac{0,20}{0,50} \right| = 0,4$$

Na Figura 10, usa-se somente a curva rotulada 0. Lê-se o valor de 0,40 no eixo das abscissas, fazendo uma referência vertical até cruzar a curva 0 e encontra-se o valor ρ_{yz} no eixo das ordenadas, estimando-se $\rho_{yz} = +0,9285$. Atribui-se o sinal positivo a ρ_{yz} desde que C_{10} tenha sinal positivo.

Exemplo 5.- Para ilustrar o uso de coeficientes de variação negativos, apresenta-se o seguinte exemplo:

$$y = \frac{x_1}{x_2}$$

$$z = \frac{x_1}{x_3}$$

$$C_1 = -0,5$$

$$C_2 = -0,4$$

$$C_3 = 1,5$$

Verifica-se o resultado com a solução analítica baseada na fórmula de Pearson segundo a seguinte fórmula:

$$\rho_{yz} = \frac{C_1^2}{(C_1^2 + C_2^2)^{1/2}(C_1^2 + C_3^2)^{1/2}}$$

Substituindo os correspondentes valores, obtém-se

$$\rho_{yz} = \frac{(-0,5)^2}{((-0,5)^2 + (-0,4)^2)^{1/2}((-0,5)^2 + 1,5^2)^{1/2}} = \frac{0,25}{1,0124} = 0,2469$$

Calculam-se os parâmetros indicativos segundo a Tabela 11:

$$\left| \frac{C_2}{C_1} \right| = \left| \frac{-0,4}{-0,5} \right| = 0,8$$

$$\left| \frac{C_3}{C_1} \right| = \left| \frac{1,5}{-0,5} \right| = 3,0$$

Na Figura 10, o valor de 0,8 interpola-se visualmente entre as curvas rotuladas 0 e 1, e depois lê-se o valor de 3,0 no eixo das abscissas. O ponto de cruzamento das referências indica nas ordenadas a estimativa $\rho_{yz} = +0,2469$.

Tabela 11.- Coletânea de modelos cuja solução gráfica obtém-se através da Figura 10.

Modelo	Funções	Parâmetros indicativos	Uso dos parâmetros na figura 10	Sinal do coeficiente de correlação espúria
1	$y = \frac{x_1}{x_2}$ $z = \frac{x_1}{x_3}$	$\left \frac{C_2}{C_1} \right $ $\left \frac{C_3}{C_1} \right $	interpolador valor nas curvas rotuladas 0, 1, 2, 3 ler valor no eixo das abscissas	$\rho_{yz} > 0$

2	$y = \frac{x_5}{x_4}$ $z = \frac{x_6}{x_4}$	$\left \frac{C_5}{C_4} \right $ $\left \frac{C_6}{C_4} \right $	interpolat valor nas curvas rotuladas 0, 1, 2, 3 ler valor no eixo das abscisas	$\rho_{yz} > 0$
3	$y = \frac{x_7}{x_8}$ $z = \frac{x_8}{x_9}$	$\left \frac{C_7}{C_8} \right $ $\left \frac{C_9}{C_8} \right $	interpolat valor nas curvas rotuladas 0, 1, 2, 3 ler valor no eixo das abscisas	$\rho_{yz} < 0$
4	$y = \frac{x_{10}}{x_{11}}$ $z = x_{10}$	$\left \frac{C_{11}}{C_{10}} \right $ 0	ler valor no eixo das abscisas usar curva rotulada 0	ρ_{yz} tem o mesmo sinal que C_{10}

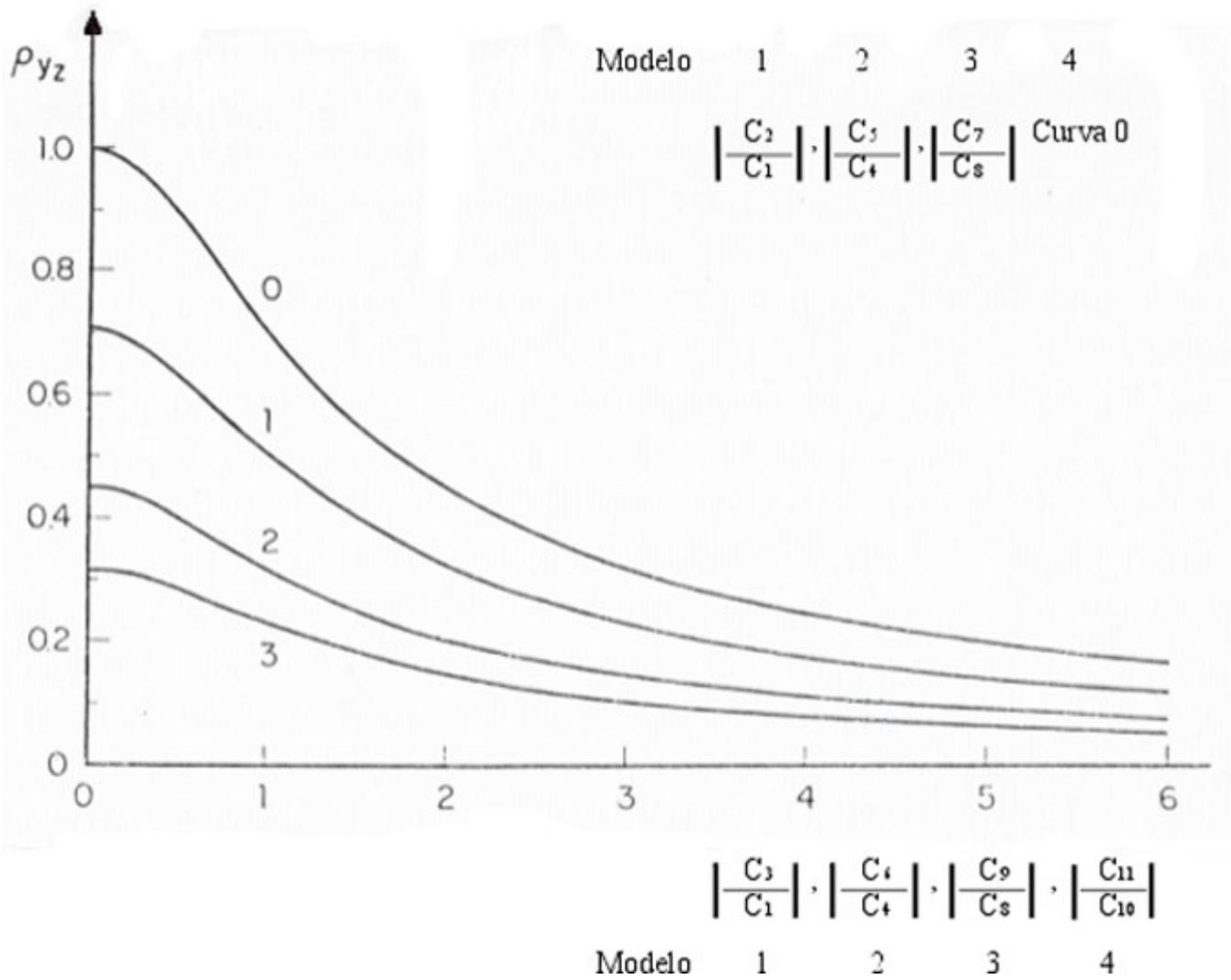


Figura 10.- Gráfico para obtenção do coeficiente de correlação espúria.