

02/2
21

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

DIVISION DE ESTUDIOS SUPERIORES DE LA
FACULTAD DE INGENIERIA



Correlación Espuria en la Ingeniería Experimental



TESIS DE MAESTRIA

Que para obtener el grado de
Maestro en Ingeniería (Hidráulica)
p r e s e n t a :
ING. CARLOS ALFONSO BREMAUNTZ MONDE

México, D. F. 1974

**TESIS CON
FALSA DE ORIGEN**

294541



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

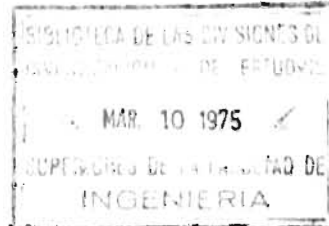
02/2
21

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

DIVISION DE ESTUDIOS SUPERIORES DE LA
FACULTAD DE INGENIERIA



**Correlación Espuria en la Ingeniería
Experimental**



TESIS DE MAESTRIA

Que para obtener el grado de
Maestro en Ingeniería (Hidráulica)

presenta:

ING. CARLOS ALFONSO BREMAUNTZ MONDE

México, D. F.

1974

**TESIS CON
FALLA DE ORIGEN**

2294541

Este trabajo fue dirigido por el Dr. en I. Gabriel
Echávez Aldape a quien agradezco la guía que
me prestó.

I N D I C E

| | Página |
|---|--------|
| C A P I T U L O I | |
| INTRODUCCION | 1 |
| C A P I T U L O II | |
| ANTECEDENTES | 4 |
| C A P I T U L O III | |
| LA CORRELACION REAL ENTRE DOS FUNCIONES | 9 |
| C A P I T U L O IV | |
| FORMULA GENERAL DE LA CORRELACION ESPURIA | 16 |
| C A P I T U L O V | |
| MODELOS DE RELACIONES MAS COMUNES | 20 |
| 5.1) Ejemplos del uso de la gráfica de correlaciones totalmente espurias (Gráfica 1) | 21 |

| | Página |
|---------------------------------------|--------|
| C A P I T U L O VI | |
| GRADO ESPURICO Y OTRAS CORRELACIONES | 26 |
| 6.1) Grado espurio | 26 |
| 6.2) La correlación exagerada | 27 |
| 6.3) La correlación parcial implícita | 27 |
| C A P I T U L O VII | |
| EJEMPLOS NUMERICOS | 32 |
| C A P I T U L O VIII | |
| C O N C L U S I O N E S | 48 |
| B I B L I O G R A F I A | 50 |

C A P I T U L O I

INTRODUCCION

Para resolver problemas complejos en Ingeniería es usual hacer hipótesis simplificadoras que es necesario justificar con métodos experimentales. Para esto se seleccionan las hipótesis que parecen importantes en -- una teoría determinada y se obtienen datos que den información directa --- acerca de la aceptabilidad de esas hipótesis. La decisión acerca del significado de los datos puede conducir a la confirmación, revisión o rechazo de las hipótesis y, con ello, la teoría que originó. Como, en general se tienen conjuntos de datos o resultados, es necesaria la aplicación de técnicas probabilísticas para poder hacer inferencias realmente significativas.

A menudo se desea saber si dos conjuntos de datos están relacionados y conocer el grado de su relación. Este grado de asociación se le conoce como CORRELACION, y matemáticamente se le puede definir de la siguiente manera (1) :

Si existe una relación lineal entre dos variables X y Y, sea.

\bar{X} la media de la variable X

\bar{Y} la media de la variable Y

$x = X' - \bar{X}$ en cada valor X' de la variable X

$y = Y' - \bar{Y}$ en cada valor Y' de la variable Y

la correlación será:

$$r = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} = \frac{S_{xy}}{S_x S_y}$$

donde.

$$S_{xy} = \frac{\sum xy}{N} \text{ covariancia de } X \text{ y } Y.$$

$$S_x = \sqrt{\frac{\sum x^2}{N}} \text{ desviación estandar de } X$$

$$S_y = \sqrt{\frac{\sum y^2}{N}} \text{ desviación estandar de } Y.$$

N número de datos.

Pudiendo aplicar esta definición a una correlación curvilínea o no-lineal. En otras palabras, el coeficiente de correlación mide la bondad de ajuste de la recta o curva de regresión supuesta para el conjunto de datos. En este trabajo el término coeficiente de correlación es utilizado para significar al coeficiente de correlación lineal.

La definición de coeficiente de correlación puede extenderse a relaciones de grupos de variables.

Establecer la existencia de una correlación entre dos variables o entre dos grupos de variables puede ser la meta, o solamente un paso en una investigación, como es el caso cuando se usan medidas de correlación para probar la confiabilidad de las observaciones.

La correlación entre un grupo de variables aleatorias independientes entre sí es nula, pero la correlación entre combinaciones de estas mis

mas variables hechas de tal manera que tengan variables comunes entre --- ellas será una correlación no-nula debida a la repetición de aquellas. Este tipo de correlación se conoce como CORRELACION ESPURIA, y puede tener efectos muy importantes en la Ingeniería Experimental, ya que muy a menudo se repiten variables en los ejes de un diagrama de dispersión o de un plano coordenado.

El objeto de este trabajo es obtener un método para valuar estas correlaciones no-existentes, al mismo tiempo que conocer las correlaciones que se pueden confundir con la espuria. Para esto se presentarán los antecedentes del problema, se definirán las fórmulas generales tanto de la correlación real como de la espuria, se encontrarán los modelos para las relaciones más comunes en la Hidráulica, se definirán otros índices de correlación y se resolverán dos ejemplos numéricos.

Este estudio se encaminará principalmente hacia las relaciones de variables en Hidráulica y Mecánica de Fluidos, aunque se puede aplicar a cualquier campo como: Medicina, Biología, Química o Psicología, siempre y cuando se quieran asociar datos para hacer posteriormente inferencias estadísticas.

C A P I T U L O I I

A N T E C E D E N T E S

En 1897, Karl Pearson escribió un artículo con el título "Una forma de correlación espuria que puede surgir cuando los índices son utilizados en la medición de órganos" (2), en donde, de cuatro variables correlacionadas, Pearson formó las proporciones o índices $\frac{x_1}{x_3}$ y $\frac{x_2}{x_4}$, definiendo para estas proporciones sus medias, desviaciones estándar y coeficientes de correlación en términos de las funciones correspondientes a las variables simples.

Pearson también consideró el caso de correlación espuria entre las siguientes proporciones $\frac{x_1}{x_3}$ y $\frac{x_2}{x_3}$, que surge del hecho de que x_3 es común a las dos proporciones, dándole el nombre de correlación espuria.

Aunque lo que encontró Pearson es válido y aplicable a relaciones en cualquier clase de estudio experimental, inexplicablemente se le ha --- prestado poca atención.

Lowell J. Reed definió a la correlación espuria como sigue (3): --- "Aunque no exista correlación entre dos conjuntos de variables cualesquiera

ra, existirá una correlación entre dos funciones cualesquiera de estas variables, siempre y cuando estas funciones tengan alguna de las dos variables en común. La correlación existente bajo estas condiciones será llamada correlación espuria".

Reed extendió los resultados de Pearson al caso general de dos conjuntos de variables y determinó sus medias, desviaciones estandar, coeficientes de correlación y correlación espuria de las funciones en términos de aquellas variables.

Clayes F. derivó de la fórmula original de Pearson las fórmulas específicas para muchos casos especiales y citó ejemplos en literatura petrográfica de inferencias falsas por las consecuentes correlaciones encontradas (4).

Manuel A. Bonson, en 1965, publicó el artículo "Correlación espuria en Hidráulica e Hidrología" (5), en donde, basándose en las investigaciones de Pearson y Reed, hace un resumen de lo que es la correlación espuria y de los diferentes aspectos relacionados con ésta, llegando a preocuparse por el tamaño conveniente de los coeficientes de variación y por los límites de confianza en tal estimación. Además incluye algunos ejemplos de artículos publicados en donde la correlación es espuria, presentando también algunos modelos de relaciones espurias.

Yalin, M.S., y Kamphuis, J.W., en 1971 publicaron el artículo "El Análisis Dimensional y la correlación Espuria" (6), en donde determinan la efectividad del análisis dimensional en problemas cuando el número de factores involucrados es grande y las teorías conocidas insuficientes; explicando también el peligro que implica el utilizar el análisis dimensional, ya que fácilmente se puede llegar a una correlación falsa. En este artículo existen ciertas contradicciones respecto a lo que es realmente la correlación espuria, y aparentemente Yalin trató de reivindicarse, pues en 1964 escribió un artículo "Propiedades geométricas de las ondas de arena" (7),-

en donde presenta dos gráficas con un alto grado de espureidad.

En el Análisis Dimensional existe un campo potencial para los diferentes casos de correlación espuria. Muchas de las proporciones adimensionales que son graficadas unas con otras, contienen elementos aleatorios comunes, surgiendo así la correlación inexistente y dando origen a que si queremos obtener alguna de las variables no repetidas en alguno de los ejes de una gráfica plana, vamos a obtener un cierto valor alterado debido a la correlación espuria; pues tanto en valores adimensionales como en dimensionales, si se tienen las siguientes proporciones graficadas $\frac{x_1}{x_3}$ contra $\frac{x_2}{x_3}$, en donde todos son números aleatorios, se puede pensar en predecir un valor de x_1 a partir de x_2 o un valor de x_2 a partir de x_1 , aunque x_1 y x_2 son independientes, por lo que esto no es razonable. Ahora que si x_1 y x_2 tienen una cierta correlación, la presencia de la variable repetida x_3 va a provocar una cierta alteración en la correlación real de x_1 y x_2 , por lo que va a existir cierto grado espurio.

Como ejemplos de correlaciones espurias en representaciones adimensionales, se tienen los siguientes:

En un artículo publicado por Hely, "Algunas relaciones entre escurrimiento característico y las circunstancias que rodean a la cuenca del río Delaware" (8), se encuentra una gráfica de $\frac{Q_{90(\text{diario})}}{Q_a}$ contra ---

$\frac{Q_{90(\text{mensual})}}{Q_a}$, donde :

Q_a es el gasto medio anual del escurrimiento en la cuenca, y

Q_{90} es el gasto del 90% de duración diaria o mensual. Declarándose que esta gráfica sirve para formar una base de estimación de curvas mensuales al escurrimiento diario del 90% de duración. Pero como las tres variables involucradas son aleatorias y además se repite la variable Q_a , entonces surge una correlación espuria.

En el artículo publicado por Yalin, M. Selim, anteriormente nombrado (7), se grafica $\frac{\Delta}{d_c}$ contra $\frac{d}{d_c}$ y de las curvas promedio deriva una relación entre Δ y d_c , Δ es la altura de la onda de arena, d_c el tirante crítico y d el tirante. Se dice que esta gráfica está hecha para determinar la forma de la función. Pero como es espuria, la forma de la función va a depender sólo de las propiedades del elemento común d_c . En el mismo artículo está una gráfica de $\frac{d}{D} \sqrt{\frac{A}{D}}$ contra $\frac{d}{D}$, donde A es la amplitud de la onda y D el tamaño del grano. Siendo también esta relación--espuria, ya que todas las variables son aleatorias y están repetidas las variables d y D.

En un artículo publicado por Yevdjovich, "Fluctuaciones de años húmedos y secos, Parte I" (9), se incluye una gráfica de $\frac{W}{V}$ contra $\frac{P_c}{P_c}$, donde W es el escurrimiento, \bar{V} la lluvia media anual efectiva. De la definición de los términos anteriores se deduce que es una gráfica $\frac{\Delta W}{\bar{V}}$ contra $\frac{V + \Delta W}{\bar{V}}$, donde ΔW es el cambio anual del almacenamiento de agua en una cuenca y V es el escurrimiento anual. Siendo ΔW y V variables aleatorias. El denominador común en ambas proporciones es una constante y también tiene algo de contribución a la relación entre las proporciones. La bondad del ajuste es debida, al menos en parte, a la presencia de ΔW en ambos lados.

En otro artículo publicado por Mirajgoker, "Efectos de la rugosidad natural de los canales abiertos" (10), se muestra una gráfica de $\frac{F}{\sqrt{g}}$ contra $\frac{C}{S}$, donde C es el coeficiente de rugosidad de Chezy, F el número de Froude y g la aceleración de la gravedad. Por definición $\frac{C}{\sqrt{g}}$ es igual a $\frac{F}{S}$, donde S es la pendiente de energía, por lo tanto la relación graficada es de F contra $\frac{F}{S}$. En este artículo la gráfica es espuria en el sentido de -

que S es una constante en el conjunto de experimentos y por lo tanto F es trazada contra sí misma. La forma de relación indicada (una recta aproximadamente a 45°) está basada en el hecho de que una variable es graficada contra sí misma.

Ha sido considerado que las gráficas adimensionales representan relaciones más generalizadas que las gráficas de combinaciones dimensionales y que es posible extenderse más allá del rango de datos con mayor confianza; ahora que no todo es ganancia, sino que van a existir también ciertos escollos, tales como la presencia de puntos singulares provocando discontinuidades.

Con lo antes expuesto se ve la importancia de tomar este tipo de correlación, sobre todo en disciplinas como la Hidráulica en donde el análisis dimensional juega un papel tan importante.

+ relación de ... $g =$

C A P I T U L O I I I

La Correlación Real Entre dos Funciones

Antes de derivar la fórmula general de la correlación espuria es -- conveniente presentar las ecuaciones para obtener la correlación real, la media y la desviación estandar entre dos funciones.

Reed desarrolló una ecuación general para la obtención de la correlación entre dos funciones a partir de las medias, desviaciones estandar y coeficientes de correlación de primer orden de las variables. Esta ecuación fue obtenida por aproximaciones de series infinitas y su exactitud es limitada, ya que todos los términos de tercer o más alto grado en los coeficientes de variación, C , fueron despreciados. Por lo tanto en un problema en que la desviación estandar es más grande que la media, la fórmula no dá suficiente aproximación del coeficiente de correlación. En la mayoría de los casos la proporción de la desviación estandar a la media es suficientemente pequeña de tal manera que la aproximación anterior es suficiente.

A continuación se presenta la ecuación de la correlación entre dos funciones a partir de la cual se derivan las fórmulas de la media y desviación estandar de una función. Además dicha ecuación también dá origen a --

la fórmula general de la correlación espuria que se verá en el siguiente--
capítulo.

Sea :

- $x_1, x_2, x_3, x_4, x_{15}, \dots, x_n$ un conjunto de n variables
 $m_1, m_2, m_3, m_4, m_{15}, \dots, m_n$ las medias de las variables
 $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_{15}, \dots, \sigma_n$ las desviaciones estandar
y $r_{12}, r_{13}, r_{17}, \dots, r_{n-1, n}$ los coeficientes de correlación en
las variables tomadas en pares.
- y sea $x_3, x_4, x_8, x_{20}, x_{21}, \dots, x_k$ un conjunto de K variables.
 $m_3, m_4, m_8, m_{20}, m_{21}, \dots, m_k$ las medias de las variables
 $\sigma_3, \sigma_4, \sigma_8, \sigma_{20}, \sigma_{21}, \dots, \sigma_k$ las desviaciones estandar.
y $r_{34}, r_{38}, r_{320}, \dots, r_{k-1, k}$ los coeficientes de correlación en
tre las variables tomadas en pares.

ahora sea $y = f(x_1, x_2, x_3, x_4, x_{15}, \dots, x_n)$

y $Z = F(x_3, x_4, x_8, x_{20}, x_{21}, \dots, x_k)$

representando dos funciones cualesquiera de estos dos conjuntos de varia--
bles.

Entonces la correlación entre estas dos funciones está dada por:

$$r_{yz} = \frac{\sum_{a=n_1}^{n_f} \sum_{b=k_1}^{k_f} f_a F_b r_{ab} \sigma_a \sigma_b}{\sqrt{\left[\sum_{a=n_1}^{n_f} \sum_{c=n_1}^{n_f} f_a f_c r_{ac} \sigma_a \sigma_c \right] \left[\sum_{d=k_1}^{k_f} \sum_{e=k_1}^{k_f} F_d F_e r_{de} \sigma_d \sigma_e \right]}} \quad (1)$$

donde,

- n_1 variable inicial de y
- n_f variable final de y
- k_1 variable inicial de z
- k_f variable final de z

$$f_a = \frac{\partial f(m_1, m_2, m_3, m_4, m_{15}, \dots, m_n)}{\partial m_a} \quad (1.1)$$

a indica el número de los subíndices de las variables de la función y.

$$F_b = \frac{\partial F(m_3, m_4, m_a, m_{20}, m_{21}, \dots, m_k)}{\partial m_b} \quad (1.2)$$

b indica el número de los subíndices de las variables de la función Z.

Derivándose de la ec. (1) las siguientes fórmulas.

a) Para el valor medio de $y = f(x_1, x_2, x_3, x_4, x_{15}, \dots, x_n)$

$$m_y = f + \frac{1}{2} \sum_{a=1}^{n_f} \sum_{c=1}^{n_f} f_{ac} r_{ac} \sigma_a \sigma_c \quad (2)$$

donde $f = f(m_1, m_2, m_3, m_4, m_{15}, \dots, m_n) \quad (2.1)$

$$y \quad f_{ac} = \frac{\partial^2 f(m_1, m_2, m_3, m_4, m_{15}, \dots, m_n)}{\partial m_a \partial m_c} \quad (2.2)$$

b) Para la desviación estandar de $y = f(x_1, x_2, x_3, x_4, x_{15}, \dots, x_n)$

$$\sigma_y = \sqrt{\sum_{a=1}^{n_f} \sum_{c=1}^{n_f} f_a f_c r_{ac} \sigma_a \sigma_c} \quad (3)$$

c) Para el valor medio de $Z = F(x_3, x_4, x_8, x_{20}, x_{21}, \dots, x_k)$

$$m_z = g + \frac{1}{2} \sum_{a=k_1}^{k_f} \sum_{b=k_1}^{k_f} F_{ab} r_{ab} \sigma_a \sigma_b \quad (4)$$

donde:

$$g = F(m_1, m_2, m_3, m_4, m_5, \dots, m_k) \quad (4.1)$$

$$y = F_{ab} = \frac{\partial^2 f(m_1, m_2, m_3, m_4, m_5, \dots, m_k)}{\partial m_a \partial m_b} \quad (4.2)$$

d) Para la desviación estandar de $Z = F(x_1, x_2, x_3, x_4, x_5, \dots, x_k)$

$$\sigma_Z = \sqrt{\sum_{a=k_1}^{k_f} \sum_{b=k_1}^{k_f} F_{ab} F_{ba} \sigma_a \sigma_b} \quad (5)$$

Para ilustrar las Ecs. (1), (2), (3), (4) y (5), se aplicarán al caso de correlación entre proporciones. Esto es:

Sean las variables x_1, x_2, x_3, x_4 y las proporciones $y = \frac{x_1}{x_2}$ y $Z = \frac{x_3}{x_4}$

$y = f(x_1, x_2) = \frac{x_1}{x_2}$, la función es el cociente $f(m_1, m_2) = \frac{m_1}{m_2}$

$Z = F(x_3, x_4) = \frac{x_3}{x_4}$, la función es el cociente $F(m_3, m_4) = \frac{m_3}{m_4}$

De la Ec. (1.1) para $a = 1$ y 2

$$f_1 = \frac{1}{m_2}, \quad f_2 = -\frac{m_1}{m_2^2}$$

De la Ec. (1.2) para $b = 3$ y 4.

$$F_3 = \frac{1}{m_4}, \quad F_4 = -\frac{m_3}{m_4^2}$$

Ahora sustituyendo los valores en la Ec. (1);

Para el Numerador de la Ec. (1), queda:

$$\frac{1}{m_2} \frac{1}{m_3} r_{13} \sigma_1 \sigma_3 - \frac{1}{m_2} \frac{m_2}{m_4^2} r_{14} \sigma_1 \sigma_4 - \frac{m_1}{m_2^2} \frac{1}{m_4} r_{23} \sigma_2 \sigma_3 +$$

$$+ \frac{m_1}{m_2^2} \frac{m_3}{m_4^2} r_{24} \sigma_2 \sigma_4$$

multiplicando todo por $\frac{m_2 m_4}{m_1 m_3}$ y sustituyendo los siguientes valores:

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

$$\text{queda: } r_{13} C_1 C_3 - r_{14} C_1 C_4 - r_{23} C_2 C_3 + r_{24} C_2 C_4$$

Ahora para el 1er término del denominador de la Ec. (1).

$$\sqrt{\frac{1}{m_2} \frac{1}{m_2} r_{11} \sigma_1^2 - \frac{1}{m_2} \frac{m_1}{m_2^2} r_{12} \sigma_1 \sigma_2 - \frac{1}{m_2} \frac{m_1}{m_2^2} r_{21} \sigma_2 \sigma_1 + \frac{m_1}{m_2^2} \frac{m_1}{m_2^2} r_{22} \sigma_2^2}$$

multiplicando todo por $\sqrt{\frac{m_2^2}{m_1^2}} = \frac{m_2}{m_1}$ y sustituyendo los coeficientes de variación C respectivos, queda.

$$\sqrt{r_{11} C_1^2 - 2 r_{12} C_1 C_2 + r_{22} C_2^2}$$

pero como $r_{11} = r_{22} = 1$

$$\sqrt{C_1^2 + C_2^2 - 2 r_{12} C_1 C_2}$$

Para el segundo término del denominador de la EC. (1) sería el mismo procedimiento, obteniéndose finalmente para la Ec. (1)

$$r_{y'z} = \frac{r_{13} C_1 C_3 - r_{14} C_1 C_4 - r_{23} C_2 C_3 + r_{24} C_2 C_4}{\sqrt{(C_1^2 + C_2^2 - 2 r_{12} C_1 C_2) (C_3^2 + C_4^2 - 2 r_{34} C_3 C_4)}}$$

Siendo esta última la fórmula de Pearson para el coeficiente de correlación entre dos índices o proporciones.

Ahora obteniendo la media de y .

$$\text{De (2.1)} \quad f = f(m_1, m_2) = \frac{m_1}{m_2}$$

$$\text{De (2.2)} \quad f_{11} = 0$$

$$f_{22} = 2 \frac{m_1}{m_2^2}$$

$$f_{12} = f_{21} = - \frac{1}{m_2^2}$$

Sustituyendo estos valores en la Ec. (2)

$$m_y = \frac{m_1}{m_2} [1 + C_2^2 - r_{12} C_1 C_2]$$

de la misma manera con la Ec. (4) se obtiene la media para Z .

$$m_z = \frac{m_3}{m_4} [1 + C_4^2 - r_{34} C_3 C_4]$$

y por último las desviaciones estandar.

Para y , mediante la ecuación (3).

$$s_y = \frac{m_1}{m_2} \sqrt{C_1^2 + C_2^2 - 2 r_{12} C_1 C_2}$$

y para la desviación estandar de Z, mediante la ec. (5) se obtiene.

$$\sigma_z = \frac{m_4}{m_3} \sqrt{C_3' + C_4' - 2 r_{34} C_3 C_4}$$

C A P I T U L O I V

Fórmula General de la Correlación Espuria

Una fórmula general para la correlación espuria puede ser derivada directamente de la ec. (1).

Si se tiene:

$$y = f(x_1, x_2, \dots, x_n) \text{ y } Z = F(x_3, x_4, \dots, x_k)$$

habrá correlación espuria (total) cuando alguna o algunas de las variables de y sean iguales a alguna o algunas variables de Z , y todas las demás correlaciones entre las variables sean nulas.

Sea:

$$\begin{aligned} x_1 &= x_3 \\ x_2 &= x_4 \\ &\vdots \\ x_h &= x_L \text{ donde } h \leq n \text{ y } L \leq k \end{aligned}$$

Entonces, por la definición de correlación espuria $r_{a,b} = 0$ excepto para los casos en donde:

$$a = b \left(\begin{array}{c} < h \\ < L \end{array} \right) \text{ en que } r_{(a,b)} = 1.$$

Así, de la ecuación (1), para $h = L$, se tendrá:

$$\rho_{yz} = \frac{\sum_{e=1}^h f_e F_e \sigma_e^2}{\sqrt{\left[\sum_{a=n_1}^n f_a^2 \sigma_a^2 \right] \left[\sum_{b=x_1}^k F_b^2 \sigma_b^2 \right]}} \quad (6)$$

donde ρ_{yz} representa a la correlación espuria.

En la ec. (1) están incluidas todas las formas de correlación espuria para dos funciones.

Por ejemplo: cuando $y = \frac{x_1}{x_2}$ y $Z = \frac{x_3}{x_4}$ habrá 3 casos diferentes de correlación espuria.

- a) Cuando las proporciones tienen el mismo denominador $x_2 = x_4$
- b) Cuando el denominador de una es el mismo numerador de la otra.

$$x_1 = x_4 \quad \text{ó} \quad x_2 = x_3$$

- c) Cuando las proporciones tienen numeradores comunes $x_1 = x_3$

También cuando las funciones consideradas son productos pueden tener un factor común. $y = x_1 x_2$ y $Z = x_3 x_2$

Como ejemplo ilustrativo del uso de la ec. (6), se calculará la correlación espuria de las proporciones $y = \frac{x_1}{x_2}$ y $Z = \frac{x_3}{x_2}$

Para este caso;

$$y = f(x_1, x_2) = \frac{x_1}{x_2} \quad \text{la función es el cociente } f(m_1, m_2) = \frac{m_1}{m_2}$$

$$Z = f(x_2, x_3) = \frac{x_3}{x_2} \text{ la función es el cociente } f(m_2, m_3) = \frac{m_3}{m_2}$$

De la ec. (1.1)

$$f_1 = \frac{\partial f(m_1, m_2)}{\partial m_1}$$

Para $a = 1$ $f_1 = \frac{1}{m_2}$

Para $a = 2$ $f_2 = -\frac{m_1}{m_2^2}$

De la Ec. (1.2) $F_b = \frac{\partial F(m_2, m_3)}{\partial m_3}$

Para $b = 2$ $F_2 = -\frac{m_3}{m_2^2}$

Para $b = 3$ $F_3 = \frac{1}{m_2}$

El numerador de la ec. (6) será :

$$\sum_{e=2}^4 f_e F_e \sigma_e^2 = f_2 F_2 \sigma_2^2 = \frac{m_1 m_3}{m_2^4} \sigma_2^2$$

multiplicando todo por $\frac{m_2^4}{m_1 m_3}$, y sustituyendo los coeficientes C, queda-
 C_2^2 .

Ahora para el primer término del denominador de la ec. (6)

$$\sqrt{\frac{1}{m_2^2} \sigma_1^2 + \frac{m_1^2}{m_2^4} \sigma_2^2}, \text{ multiplicando todo por } \sqrt{\frac{m_2^2}{m_1^2}}$$

y sustituyendo las Ces. queda.

$$\sqrt{C_1^2 + C_2^2}$$

y para el segundo término del denominador $\sqrt{\frac{1}{m_2^2} \sigma_3^2 + \frac{m_3^2}{m_2^4} \sigma_2^2}$,

multiplicando por $\sqrt{\frac{m_2^2}{m_3^2}}$ y sustituyendo las Ces. queda igual a -

$$\sqrt{C_3^2 + C_2^2}$$

Por lo tanto la ecuación final será igual a :

$$r_{yz} = \frac{C_2^2}{\sqrt{(C_1^2 + C_2^2)(C_3^2 + C_2^2)}}$$

Otra manera de obtener la ecuación anterior es directamente de la ecuación (1), calculando el valor de la correlación real para $y = \frac{x_1}{x_2}$ y $Z = \frac{x_3}{x_2}$, para luego sustituir los valores de : $r_{12} = r_{13} = r_{23} = 0$

Particularmente este último método es el más conveniente, ya que se obtiene también la correlación real y la correlación espuria de una sola vez. En el siguiente capítulo se utiliza este último método para calcular los valores de las correlaciones espurias de algunos de los modelos más comunes que se presentan en la Hidráulica.

C A P I T U L O V

Modelos de Relaciones más Comunes

De la tabla I, en la primera columna, se enlistan los modelos de relaciones más comunes en la Hidráulica, siendo y la ordenada y Z la abscisa de una gráfica plana; en la segunda columna se encuentran las correlaciones reales (Pearson) y en la tercera columna las correlaciones totalmente espurias.

Para obtener las medias y desviaciones estándar de las ordenadas y de las abscisas se procede de la siguiente manera:

Si es un cociente, por ejemplo.

$$L = \frac{x_2}{x_1}, \text{ siendo } L = y \text{ ó } Z, \text{ entonces.}$$

$$m_L = \frac{m_S}{m_I} [1 + C_I^2 - r_{SI} C_S C_I]$$

$$\sigma_L = \frac{m_S}{m_I} [C_S^2 + C_I^2 - 2 r_{SI} C_S C_I]^{1/2}$$

Si es un cociente con el denominador elevado a la potencia P ---
 (Modelo 10), por ejemplo:

$$L = \frac{x_s}{x_i^p}, \text{ entonces,}$$

$$m_L = \frac{m_s}{m_i^p} \left[i - p r_{si} C_s C_i + \frac{1}{2} p (p + 1) C_i^2 \right]$$

$$\sigma_L = \left(\frac{m_s^2}{m_i^{2p}} \right)^{1/2} \left(C_s^2 + m_s^{2p-2} C_i^2 - 2 m_s^{p-1} r_{si} C_s C_i \right)^{1/2}$$

Y si es un producto, por ejemplo:

$$L = x_s x_i, \text{ entonces:}$$

$$m_L = m_s m_i (1 + r_{si} C_s C_i)$$

$$\sigma_L = m_s m_i (C_s^2 + C_i^2 + 2 r_{si} C_s C_i)^{1/2}$$

Las ecuaciones correspondientes a las correlaciones totalmente espurias están graficadas para los diferentes valores de las relaciones entre los coeficientes de variación en la Gráfica 1, en donde el signo de la ordenada $\pm \rho_{y_i}$ o sea la correlación totalmente espuria va a estar determinado por el signo proporcionado en la tercera columna de la Tabla 1.

El superíndice $\underline{0}$ de algunos valores de las abscisas, significa -- que esas proporciones van a estar correlacionadas con la curva $\underline{0}$, para -- así poder obtener el valor de la correlación totalmente espuria.

5.1.- EJEMPLOS DEL USO DE LA GRAFICA DE CORRELACIONES TOTALMENTE ESPURIAS (GRAFICA 1)

Ejemplo 1.- Sea una relación $y = \frac{x_1}{x_2}$ contra $Z = \frac{x_1}{x_3}$, en don-

de los coeficientes de variación son:

$$C_1 = 0.2, \quad C_2 = 0.3 \quad \text{y} \quad C_3 = 0.4$$

Se ve en la tabla I que es una relación modelo tipo (3) y en la gráfica I se indica que para el modelo (3) hay que obtener las relaciones $\left| \frac{C_3}{C_1} \right| = 2$ y $\left| \frac{C_2}{C_1} \right| = 1.5$, por lo que la correlación espuria será $\rho_{y,z} = + 0.248$, donde el signo positivo se obtuvo de la tabla I -- columna 3.

Ejemplo 2.- Sea una relación $y = \frac{x_7}{x_8}$ contra $Z = \frac{x_5}{x_3}$, en donde

los coeficientes de variación son:

$$C_7 = 1.6, \quad C_8 = C_3 = - 0.8$$

Se ve en la tabla I que es una relación modelo tipo (5) y en la gráfica I se indica que para el modelo (5) hay que calcular las relaciones $\left| \frac{C_8}{C_3} \right| = 1$ y $\left| \frac{C_7}{C_3} \right| = 2$, por lo que la correlación espuria será $\rho_{y,z} = - 0.32$, donde el signo negativo se obtuvo de la tabla I columna 3.

Ejemplo 3.- Sea una relación $y = \frac{x_{14}}{x_{12}}$ contra $Z = x_{13}$, en donde los coeficientes de variación son: $C_{14} = 0.9$ y $C_{13} = 0.3$.

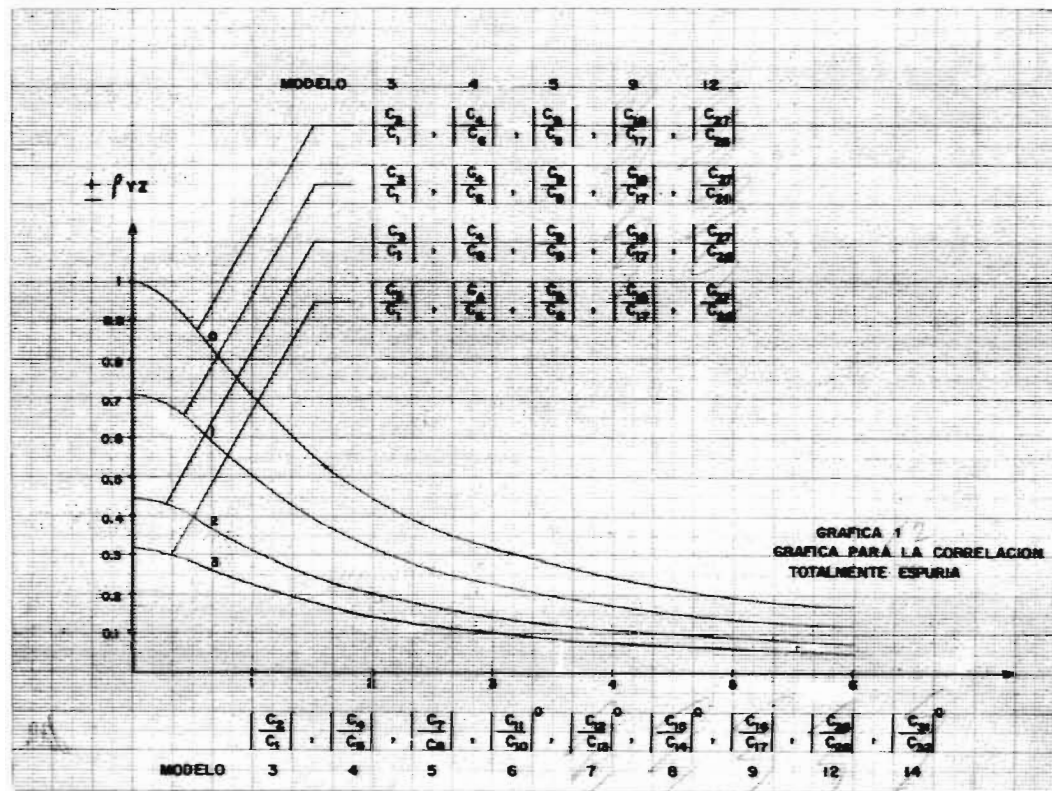
Se ve en la tabla I que es una relación modelo tipo (7) y en la gráfica I se indica que para el modelo (7) solo hay que obtener la relación $\left| \frac{C_{14}}{C_{13}} \right| = 3$, ya que tiene el superíndice 0 y por lo tanto esta proporción va a estar correlacionada con la curva 0, para así calcular el valor de $\rho_{y,z} = - 0.32$, donde el signo negativo se obtiene de la tabla I columna 3.

| MODELO | Correlación Real (Pearson) r_{yz} | Correlación Totalmente Espuria. $r_{i,j} = 0, (i \neq j), r_{yz}$ |
|--|---|--|
| (1) $y = \frac{x_1}{x_2}$ $z = \frac{x_3}{x_4}$ | $\frac{r_{13} C_1 C_3 - r_{14} C_1 C_4 - r_{23} C_2 C_3 + r_{24} C_2 C_4}{(C_1^2 + C_2^2 - 2r_{12} C_1 C_2)^{1/2} (C_3^2 + C_4^2 - 2r_{34} C_3 C_4)^{1/2}}$ | 0 |
| (2) $y = \frac{x_1}{x_2}$ $z = x_3$ | $\frac{r_{13} C_1 - r_{23} C_2}{(C_1^2 + C_2^2 - 2r_{12} C_1 C_2)^{1/2}}$ | 0 |
| (3) $y = \frac{x_1}{x_2}$ $z = \frac{x_3}{x_4}$ | $\frac{C_1^2 - r_{13} C_1 C_3 - r_{14} C_1 C_4 + r_{23} C_2 C_3}{(C_1^2 + C_2^2 - 2r_{12} C_1 C_2)^{1/2} (C_3^2 + C_4^2 - 2r_{34} C_3 C_4)^{1/2}}$ | $\frac{C_1^2}{(C_1^2 + C_2^2)^{1/2} (C_3^2 + C_4^2)^{1/2}} = + r_{yz}$ |
| (4) $y = \frac{x_5}{x_6}$ $z = \frac{x_7}{x_8}$ | $\frac{r_{56} C_5 C_6 - r_{57} C_5 C_7 - r_{68} C_6 C_8 + C_6^2}{(C_5^2 + C_6^2 - 2r_{56} C_5 C_6)^{1/2} (C_7^2 + C_8^2 - 2r_{78} C_7 C_8)^{1/2}}$ | $\frac{C_6^2}{(C_5^2 + C_6^2)^{1/2} (C_7^2 + C_8^2)^{1/2}} = + r_{yz}$ |
| (5) $y = \frac{x_7}{x_8}$ $z = \frac{x_9}{x_{10}}$ | $\frac{r_{78} C_7 C_8 - r_{79} C_7 C_9 + r_{80} C_8 C_{10} - C_8^2}{(C_7^2 + C_8^2 - 2r_{78} C_7 C_8)^{1/2} (C_9^2 + C_{10}^2 - 2r_{90} C_9 C_{10})^{1/2}}$ | $\frac{-C_8^2}{(C_7^2 + C_8^2)^{1/2} (C_9^2 + C_{10}^2)^{1/2}} = - r_{yz}$ |
| (6) $y = \frac{x_{10}}{x_{11}}$ $z = x_{12}$ | $\frac{C_{10} - r_{10,11} C_{11}}{(C_{10}^2 + C_{11}^2 - 2r_{10,11} C_{10} C_{11})^{1/2}}$ | $\frac{C_{10}}{(C_{10}^2 + C_{11}^2)^{1/2}} = + r_{yz}$ |

T A B L A 1

| MODELO | Correlación Real (Pearson) r_{yz} | Correlación Totalmente Espuria $r_{ij} = 0, (i \neq j) \quad \rho_{yz}$ |
|--|--|---|
| (7) $y = \frac{x_{12}}{x_{13}}$ $z = x_{13}$ | $\frac{r_{12,13} C_{12} - C_{13}}{(C_{12}^2 + C_{13}^2 - 2r_{12,13}C_{12}C_{13})^{1/2}}$ | $\frac{-C_{13}}{(C_{12}^2 + C_{13}^2)^{1/2}} = -\rho_{yz}$ |
| (8) $y = \frac{1}{x_{14}}$ $z = \frac{x_{14}}{x_{15}}$ | $\frac{r_{14,15} C_{14} - C_{15}}{(C_{14}^2 + C_{15}^2 - 2r_{14,15}C_{14}C_{15})^{1/2}}$ | $\frac{-C_{15}}{(C_{14}^2 + C_{15}^2)^{1/2}} = -\rho_{yz}$ |
| (9) $y = \frac{x_{16}}{x_{17}}$ $z = x_{17}, x_{18}$ | $\frac{r_{16,17}C_{16}C_{17} + r_{16,18}C_{16}C_{18} - C_{17}^2 - r_{17,18}C_{17}C_{18}}{(C_{16}^2 + C_{17}^2 - 2r_{16,17}C_{16}C_{17})^{1/2}(C_{17}^2 + C_{18}^2 + 2r_{17,18}C_{17}C_{18})^{1/2}}$ | $\frac{-C_{17}^2}{(C_{16}^2 + C_{17}^2)^{1/2}(C_{17}^2 + C_{18}^2)^{1/2}} = -\rho_{yz}$ |
| (10) $y = \frac{x_{19}}{x_{20}}$ $z = x_{20}$ | $\frac{r_{19,20} C_{19}C_{20} - p C_{20}^2}{(C_{19}^2 + p^2 C_{20}^2 - 2p r_{19,20}C_{19}C_{20})^{1/2} (C_{20}^2)^{1/2}}$ | $\frac{-p C_{20}^2}{(C_{19}^2 + p^2 C_{20}^2)^{1/2} (C_{20}^2)^{1/2}} = -\rho_{yz}$ |
| (11) $y = x_{21}x_{22}$ $z = x_{23}x_{24}$ | $\frac{r_{21,22}C_{21}C_{22} + r_{21,24}C_{21}C_{24} + r_{22,23}C_{22}C_{23} + r_{22,24}C_{22}C_{24}}{(C_{21}^2 + C_{22}^2 - 2r_{21,22}C_{21}C_{22})^{1/2}(C_{23}^2 + C_{24}^2 + 2r_{23,24}C_{23}C_{24})^{1/2}}$ | 0 |
| (12) $y = x_{25}x_{26}$ $z = x_{27}x_{28}$ | $\frac{r_{25,26}C_{25}C_{26} + r_{25,28}C_{25}C_{28} + r_{26,27}C_{26}C_{27} + C_{26}^2}{(C_{25}^2 + C_{26}^2 + 2r_{25,26}C_{25}C_{26})^{1/2}(C_{27}^2 + C_{28}^2 + 2r_{27,28}C_{27}C_{28})^{1/2}}$ | $\frac{C_{26}^2}{(C_{25}^2 + C_{26}^2)^{1/2}(C_{27}^2 + C_{28}^2)^{1/2}} = +\rho_{yz}$ |

| MODELO | Correlación Real (Pearson) r_{yz} | Correlación Totalmente Espuria $r_{ij} = 0 (i \neq j)$ r'_{yz} |
|---|--|---|
| (13) $y = x_{28} x_{29}$ $Z = x_{30}$ | $\frac{r_{28,30} C_{28} + r_{29,30} C_{29}}{(C_{28}^2 + C_{29}^2 + 2r_{28,29} C_{28} C_{29})^{1/2}}$ | 0 |
| (14) $y = x_{31} x_{32}$ $Z = x_{33}$ | $\frac{r_{31,32} C_{31} + C_{32}}{(C_{31}^2 + C_{32}^2 + 2r_{31,32} C_{31} C_{32})^{1/2}}$ | $\frac{C_{32}}{(C_{31}^2 + C_{32}^2)^{1/2}} = + r'_{32}$ |



C A P I T U L O VI

Grado Espurio y Otras Correlaciones

En este capítulo se verá lo que es el grado de espureidad o de falsedad, además de otros tipos de correlaciones que se podrían fácilmente -- confundir con la correlación espuria con el objeto de tener una visión más amplia del tema.

6.1.- GRADO ESPURIO.

El grado espurio de una correlación se obtiene como el porcentaje -- de la correlación espuria a partir de la correlación real y es un índice -- que mide la falsedad de una relación.

Así por ejemplo para una correlación no-espuria o verdadera.

Si r_{y2} (real) = 0.90 y por lo tanto ρ_{y2} (espuria) = 0.0 el grado espurio va a ser igual a 0 %.

Luego, para una correlación totalmente espuria.

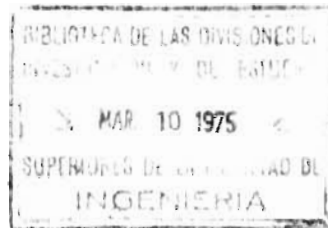
Si r_{y2} (real) = 0.90 y por lo tanto ρ_{y2} (espuria) = 0.90 el grado espurio correspondiente es del 100 %. Y para una correlación par-

cialmente espuria.

$$\text{Si } r_{12} \text{ (real) } = 0.90$$

$$\text{y } r_{12} \text{ (espuria) } = 0.45$$

entonces el grado espurio es del 50 %.



6.2.- LA CORRELACION EXAGERADA .

En aplicaciones prácticas, una correlación es a menudo exagerada -- por el trazo de cantidades comunes a lo largo de ambos ejes en una gráfica creyendo que así se va a aumentar la correlación y obteniendo dicha correlación como la de dos variables únicas X y Y, sin tomar en cuenta la intercorrelación de las variables que las constituyen y sin hacer caso de la correlación espuria debido a la presencia de las variables repetidas. Calculando así una correlación mayor que la real o sea exagerada.

Si se procede de la manera anterior en realidad se tendrá como resultado una correlación mal obtenida, debido a no hacer uso de la fórmula de correlación real (Pearson) para grupos de variables (columna 2, tabla 1)

6.3.- LA CORRELACION PARCIAL IMPLICITA .

Cuando se observa una correlación entre dos variables siempre hay la posibilidad de que esta correlación sea debida a la asociación entre cada una de las dos variables y una tercera. Por ejemplo, entre un grupo de niños de una escuela es posible encontrar una alta correlación entre el peso de los niños y la estatura. Esta correlación puede no reflejar una relación genuina o directa entre estas dos variables, sino más bien resulta del hecho de que tanto el peso como la estatura están asociados con una tercera variable, la edad.

En el diseño experimental se tiene la opción de introducir controles experimentales para eliminar la influencia de la tercera variable o de usar métodos estadísticos para eliminar su influencia. Así para el ejem--

pló anterior un control experimental sería escoger niños de la misma edad.

En el caso de que los controles experimentales no sean factibles, - entonces pueden aplicarse controles estadísticos.

Un control estadístico sería la correlación parcial (11). En esta correlación los efectos de variación por una tercera variable sobre la relación entre las otras dos son eliminados. En otras palabras, la correlación entre X_1 y X_2 se encuentra al tener la tercera variable X_3 como un valor constante. Así la correlación parcial es :

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}} \quad (7)$$

Si se tratara de más variables, entonces sería:

$$r_{12.345} = \frac{r_{12.34} - r_{13.34} r_{23.34}}{\sqrt{1 - r_{13.34}^2} \sqrt{1 - r_{23.34}^2}} \quad (8)$$

dejando constantes a las variables X_3 , X_4 y X_5 .

Las ecuaciones (7) y (8) son válidas para cuando se tienen distribuciones normales entre las variables. En el caso de que la suposición de normalidad no sea realista o que la muestra sea pequeña, se hará uso de la estadística no paramétrica (12). Así en la estadística de orden o no paramétrica existe un coeficiente de correlación parcial que se denomina de "rango de Kendall", que se define como:

$$z_{12.3} = \frac{z_{12} - z_{13} z_{23}}{\sqrt{1 - z_{13}^2} \sqrt{1 - z_{23}^2}} \quad (9)$$

donde los valores de Z_{xy} son los llamados coeficientes de correlación simple de Kendall y se definen como:

$$Z_{xy} = \frac{S}{\frac{1}{2} N (N - 1)} \quad (10)$$

Siendo N el número de datos ordenados según X y Y , y S se obtendría a partir de los rangos de X o de Y , así por ejemplo.

Si se tienen las siguientes observaciones de lluvia y escurrimiento en una cuenca.

| | | | | |
|----------|-----|-----|-----|-----|
| Tormenta | a | b | c | d |
| Lluvia | 130 | 150 | 110 | 90 |
| Esc. | 600 | 200 | 800 | 400 |

Se pregunta si es posible a partir de la lluvia calcular el escurrimiento.

Se podrá responder afirmativamente a esta pregunta solo si el grado de correlación entre las variables es aceptable. Para esto se utiliza la estadística no paramétrica ya que el tamaño de la muestra es demasiado pequeña, por lo que se usará el coeficiente de correlación simple de Kendall.

Entonces, sean las variables X (Lluvia) y Y (escurrimiento). Se calculan los rangos para los elementos de cada una de las variables, poniéndoles el número de orden de menor a mayor que les corresponda, quedando.

| | | | | |
|--------|---|---|---|---|
| ENSAYO | a | b | c | d |
| X | 3 | 4 | 2 | 1 |
| Y | 3 | 1 | 4 | 2 |

Luego se ordenan los ensayos para que los rangos de X aparezcan en orden natural.

| ENSAYO | d | c | a | b |
|--------|---|---|---|---|
| X | 1 | 2 | 3 | 4 |
| Y | 2 | 4 | 3 | 1 |

Para determinar el valor de S se toma el primer número de la izquierda de Y y se cuentan el número de columnas a la derecha que tienen cifras mayores a tal número, luego se resta el número de columnas a la derecha que tienen cifras menores. Haciendo esto con todas y cada una de las columnas o rangos y sumando los resultados.

$$\text{Así } S = (+2 - 1) + (0 - 2) + (0 - 1) = -2$$

por lo tanto:

$$z_{xy} = \frac{S}{\frac{1}{2} N (N - 1)} = \frac{-2}{\frac{1}{2} 4 (4 - 1)} = -0.33, \text{ que co--}$$

rresponde a un valor pequeño de correlación entre X y Y, por lo que no es posible calcular el escurrimiento a partir de la lluvia sin tener un error de un 67 %.

En el caso de que haya columnas repetidas se dice que existe una "liga" entre X y Y, pudiendo tener varias "ligas". Estas repeticiones de columnas van a afectar el coeficiente de correlación dado por la ec. (10), por lo que para este caso la correlación será:

$$z_{xy} = \frac{S}{\sqrt{\frac{1}{2} N (N - 1) - T_x} \sqrt{\frac{1}{2} N (N - 1) - T_y}} \quad (11)$$

donde:

$$T_x = \frac{1}{2} \sum t(t-1) \text{ siendo } t \text{ el número de observaciones ligadas a cada grupo de ligas de la variable X.}$$

$$T_y = \frac{1}{2} \sum t(t-1) \text{ siendo } t \text{ el número de observaciones ligadas a cada grupo de ligas de la variable Y.}$$

Las ecuaciones (10) y (11) se utilizarán en la ec. (9) según sea el caso, tantas veces como términos Z_{xy} sea necesario calcular. En el último ejemplo del siguiente capítulo se verá como se obtiene el coeficiente de correlación parcial de rango de Kendall mediante el uso de las ecuaciones - (9), (10) y (11).

C A P I T U L O VII

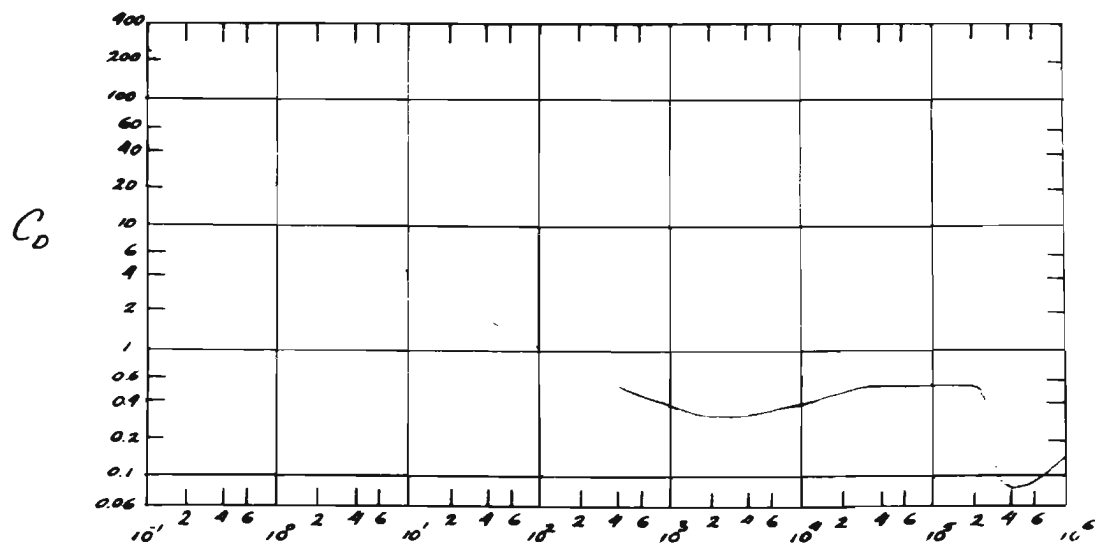
Ejemplos Numéricos

Con el objeto de ilustrar lo antes expuesto se presentan dos ejemplos. El primer ejemplo es una aplicación para la obtención de la correlación real y espuria, al mismo tiempo que muestra como se analizan las relaciones adimensionales. En el segundo ejemplo se obtiene: la correlación real, la correlación espuria, la correlación exagerada y la correlación parcial, a partir de datos experimentales.

Ejemplo 1.- En este ejemplo se estudia la variación bien conocida de la fuerza de arrastre actuando en una esfera en un fluido infinito (13).

Experimentalmente se obtuvo la gráfica 2, después de haber analizado las variables que intervenían en el problema y de haber realizado un Análisis Dimensional (14), llegando a la conclusión de que se debería de construir tal gráfica relacionando al coeficiente de arrastre C_D contra el número de Reynolds R , siendo,

$$C_D = \frac{D}{\frac{\pi}{2} a^2 V_1^2} \quad \text{y} \quad R = \frac{V_1 2 a \rho}{\mu}$$



$$R = \frac{V_0 2 a \rho}{\mu}$$

Gráfica 2.- Coeficiente de arrastre para esferas en función del número de Reynolds.

donde:

- D Fuerza de arrastre
- ρ Densidad del fluido
- a Radio de la esfera
- V_0 Velocidad del flujo
- μ Viscosidad dinámica

Si se hace.

$$x_1 = D$$

$$x_2 = \frac{\pi \rho}{2}$$

$$x_3 = a$$

$$x_4 = V_0$$

$$x_5 = 2 \rho$$

$$x_6 = \mu$$

Queda una relación de $y = \frac{x_1}{x_2 x_3^2 x_4^2}$ contra $Z = \frac{x_3 x_4 x_5}{x_6}$,

a la que se le tratará de calcular su correlación real y espuria.

Sea. $y = f(x_1, x_2, x_3, x_4)$,

la función es $f(m_1, m_2, m_3, m_4) = \frac{m_1}{m_2 m_3^2 m_4^2}$

y $Z = f(x_3, x_4, x_5, x_6)$,

la función es $F(m_3, m_4, m_5, m_6) = \frac{m_3 m_4 m_5}{m_6}$

De la ec. (1.1).

Para $a = 1$, $f_1 = \frac{1}{m_2 m_3^2 m_4^2}$

$$a = 2, \quad f_2 = -\frac{m_1}{m_1' m_3' m_4'}$$

$$a = 3, \quad f_3 = -\frac{2m_1}{m_2 m_3^2 m_4^2}$$

$$a = 4, \quad f_4 = -\frac{2m_1}{m_2 m_3^2 m_4^3}$$

De la ec. (1.2)

$$\text{Para } b = 3, \quad F_3 = \frac{m_a m_b}{m_c}$$

$$b = 4, \quad F_4 = \frac{m_a m_c}{m_b}$$

$$b = 5, \quad F_5 = \frac{m_j m_a}{m_b}$$

$$b = 6, \quad F_6 = -\frac{m_j m_a m_i}{m_b^2}$$

Ahora el numerador de la ec. (1) es.

$$\sum_{a=1}^4 \sum_{b=3}^6 f_a F_b r_{ab} \sigma_a \sigma_b$$

$$= f_1 F_3 r_{13} \sigma_1 \sigma_3 + f_1 F_4 r_{14} \sigma_1 \sigma_4 + f_1 F_5 r_{15} \sigma_1 \sigma_5 + f_1 F_6 r_{16} \sigma_1 \sigma_6 +$$

$$+ f_2 F_3 r_{23} \sigma_2 \sigma_3 + f_2 F_4 r_{24} \sigma_2 \sigma_4 + f_2 F_5 r_{25} \sigma_2 \sigma_5 + f_2 F_6 r_{26} \sigma_2 \sigma_6 +$$

$$+ f_3 F_3 r_{33}^2 \sigma_3 \sigma_3 + f_3 F_4 r_{34} \sigma_3 \sigma_4 + f_3 F_5 r_{35} \sigma_3 \sigma_5 + f_3 F_6 r_{36} \sigma_3 \sigma_6 +$$

$$+ f_4 F_3 r_{43} \sigma_4 \sigma_3 + f_4 F_4 r_{44} \sigma_4 \sigma_4 + f_4 F_5 r_{45} \sigma_4 \sigma_5 + f_4 F_6 r_{46} \sigma_4 \sigma_6$$

Sustituyendo los valores anteriores y tomando en cuenta que.

$$r_{11} = r_{22} = 1. \text{ Luego, multiplicando todo por } \frac{m_2 m_3 m_4 m_i}{m_1 n}$$

y sustituyendo $C_i = \frac{\sigma_i}{m_i}$, donde $1 \leq i \leq 6$, queda.

$$\begin{aligned}
 &= r_{12} C_1 C_2 + r_{14} C_1 C_4 + r_{15} C_1 C_5 - r_{16} C_1 C_6 - \\
 &- r_{23} C_2 C_3 - r_{24} C_2 C_4 - r_{25} C_2 C_5 + r_{26} C_2 C_6 - \\
 &- 2 C_3^2 - 2 r_{34} C_3 C_4 - 2 r_{35} C_3 C_5 + 2 r_{36} C_3 C_6 - \\
 &- 2 C_4^2 - 2 r_{45} C_4 C_5 + 2 r_{46} C_4 C_6.
 \end{aligned}$$

Ahora para el primer término del denominador de la ec. (1).

$$\begin{aligned}
 &\left(\sum_{i=1}^4 f_i f_c r_{ic} \sigma_i \sigma_c \right)^{1/2} = \\
 &= (f_1 f_1 r_{11} \sigma_1 \sigma_1 + f_1 f_2 r_{12} \sigma_1 \sigma_2 + f_1 f_3 r_{13} \sigma_1 \sigma_3 + f_1 f_4 r_{14} \sigma_1 \sigma_4 + \\
 &+ f_2 f_1 r_{21} \sigma_2 \sigma_1 + f_2 f_2 r_{22} \sigma_2 \sigma_2 + f_2 f_3 r_{23} \sigma_2 \sigma_3 + f_2 f_4 r_{24} \sigma_2 \sigma_4 + \\
 &+ f_3 f_1 r_{31} \sigma_3 \sigma_1 + f_3 f_2 r_{32} \sigma_3 \sigma_2 + f_3 f_3 r_{33} \sigma_3 \sigma_3 + f_3 f_4 r_{34} \sigma_3 \sigma_4 + \\
 &+ f_4 f_1 r_{41} \sigma_4 \sigma_1 + f_4 f_2 r_{42} \sigma_4 \sigma_2 + f_4 f_3 r_{43} \sigma_4 \sigma_3 + f_4 f_4 r_{44} \sigma_4 \sigma_4)^{1/2}
 \end{aligned}$$

Sustituyendo los valores de las fes; multiplicando todo por.

$$\left(\frac{m_2^2 m_3^4 m_4^4}{m_1^2} \right)^{1/2} \text{ y sustituyendo las Ces. queda.}$$

Como $r_{11} = r_{22} = r_{33} = r_{44} = 1$

$$\begin{aligned}
 &= (C_1^2 - 2r_{12} C_1 C_2 - 4r_{13} C_1 C_3 - 4r_{14} C_1 C_4 + C_2^2 + 4r_{23} C_2 C_3 + \\
 &+ 4r_{24} C_2 C_4 + 4 C_3^2 + 8 r_{34} C_3 C_4 + 4 C_4^2)^{1/2}.
 \end{aligned}$$

y para el segundo término del denominador de la ec. (1)

$$\left(\sum_{i=3}^6 F_i F_c r_{ic} \sigma_i \sigma_c \right)^{1/2} =$$

$$\begin{aligned} & (F_1 F_3 r_{13} \sigma_3 \sigma_3 + F_4 F_4 r_{44} \sigma_4 \sigma_4 + F_5 F_5 r_{55} \sigma_5 \sigma_5 + F_6 F_6 r_{66} \sigma_6 \sigma_6 + \\ & + F_1 F_4 r_{14} \sigma_3 \sigma_4 + F_4 F_1 r_{41} \sigma_4 \sigma_3 + F_5 F_4 r_{54} \sigma_5 \sigma_4 + F_4 F_5 r_{45} \sigma_4 \sigma_5 + \\ & + F_3 F_5 r_{35} \sigma_3 \sigma_5 + F_5 F_3 r_{53} \sigma_5 \sigma_3 + F_6 F_5 r_{65} \sigma_6 \sigma_5 + \\ & + F_5 F_6 r_{56} \sigma_5 \sigma_6 + F_6 F_6 r_{66} \sigma_6 \sigma_6)^{1/2} \end{aligned}$$

Sustituyendo los valores anteriores; multiplicando todo por.

$$\left(\frac{m_6^2}{m_3^2 m_4^2 m_5^2} \right)^{1/2} \text{ y sustituyendo las Ces., queda.}$$

$$\text{Ya que } r_{33} = r_{44} = r_{55} = r_{66} = 1$$

$$\begin{aligned} & = C_3^2 + 2 r_{34} C_3 C_4 + 2 r_{35} C_3 C_5 - 2 r_{36} C_3 C_6 + C_4^2 + \\ & + 2 r_{45} C_4 C_5 + C_5^2 - 2 r_{46} C_4 C_6 - 2 r_{56} C_5 C_6 + C_6^2 \end{aligned}$$

Ahora sustituyendo todos los valores anteriores en la ec. (i).

$$\begin{aligned} r_{yz} &= \frac{r_{13} C_1 C_3 + r_{14} C_1 C_4 + r_{15} C_1 C_5 - r_{16} C_1 C_6 - r_{23} C_2 C_3 - r_{24} C_2 C_4 -}{(C_1^2 - 2r_{17} C_1 C_7 - 4r_{13} C_1 C_3 - 4r_{14} C_1 C_4 + C_7^2 + 4r_{23} C_2 C_3 + 4r_{24} C_2 C_4 +} \\ & \dots \frac{-r_{25} C_2 C_5 + r_{26} C_2 C_6 - 2C_3^2 - 4r_{34} C_3 C_4 - 2C_4^2 - 2r_{45} C_4 C_5 +}{+4C_3^2 + 8r_{34} C_3 C_4 + 4C_4^2)^{1/2}} (C_1^2 + 2r_{34} C_3 C_4 + 2r_{35} C_3 C_5 - 2r_{36} C_3 C_6 + \\ & \dots \frac{+ 2 r_{45} C_4 C_5}{+ C_4^2 + 2 r_{45} C_4 C_5 + C_5^2 - 2 r_{46} C_4 C_6 - 2 r_{56} C_5 C_6 })^{1/2} \end{aligned}$$

Siendo la correlación real más general; ahora que si solo van a variar x_1, x_3, x_4 y x_6 , dejando como constantes a x_2 y x_5 , entonces la correlación real va a ser (teniéndose $C_2 = C_5 = 0$).

$$r_{yz} = \frac{r_{13} C_1 C_3 + r_{14} C_1 C_4 - r_{16} C_1 C_6 - 2 C_3^2 - 4r_{34} C_3 C_4 - 2 C_4^2 +}{(C_1^2 - 4r_{13} C_1 C_3 - 4r_{14} C_1 C_4 + 4 C_3^2 + 8 r_{34} C_3 C_4 + 4 C_4^2)^{1/2}}$$

$$- \frac{+ 2 r_{4c} C_4 C_c}{(C_3^2 + 2r_{34}C_3C_4 - 2r_{36}C_3C_6 + C_4^2 - 2r_{4c}C_4C_c + C_c^2)^{1/2}}$$

obteniendo su correlación totalmente espuria a partir de.

$$r_{13} = r_{14} = r_{16} = r_{34} = r_{36} = r_{46} = 0, \text{ por lo que.}$$

$$\rho_{yZ} = \frac{- 2 (C_3^2 + C_4^2)}{(C_1^2 + 4C_3^2 + 4C_4^2)^{1/2} (C_3^2 + C_4^2 + C_6^2)^{1/2}}$$

Siendo esta última la correlación totalmente espuria, por lo que si se quisiera obtener la fuerza de arrastre $D(x_1)$ o partir de la viscosidad dinámica $\mu(x_6)$ o viceversa, se obtendría una cantidad con una cierta alteración debido a la repetición de las variables x_2, x_3, x_4 y x_c (repetiendo ρ en x_2), en ambos ejes.

Por lo tanto, la presencia de las variables repetidas va a tener como consecuencia la aparición de la correlación espuria, total si no hay intercorrelación entre las variables, y parcial si hay una cierta intercorrelación entre las mismas, por lo que va a dar origen a un grado de espureidad entre y y Z .

Ejemplo 2.- Este ejemplo se basa en el artículo "Derrumbes en embalses" (15), que presenta un estudio realizado en el Instituto de Ingeniería, que trata sobre los efectos hidráulicos producidos por derrumbes de grandes masas al caer en un cuerpo de agua. En él se logran determinar, en forma aproximada, las características de la ondulación producida en función de las características del derrumbe, cuando la ondulación se propaga en una sola dirección (fenómeno bidimensional).

En la tabla 2 se presenta un registro de ensayos, donde la notación es la siguiente :

TABLE 2

ENSAYES BIDIMENSIONALES. CAIDAS EN BLOQUE.

| Ensayo | Superficie | α | θ | z | d | V | t_c | E | $V d^2$ | $\sqrt{g/d t_c}$ | $\sqrt{E_c \gamma_c d}$ | h_m | $h_{2.5}$ | h_c |
|--------|------------|----------|----------|-----|-----|-------------------|-------|-------------------|---------|------------------|-------------------------|-------|-----------|-------|
| | | ° | ° | cm | cm | m ³ /m | s | kg/m ² | | | | cm | cm | cm |
| 1 A | Plana | 30 | 60 | 10 | 50 | 0.0490 | 0.48 | 1.011 | 0.196 | 2.14 | 0.200 | 6.0 | 4.0 | 1.5 |
| 2 A | Plana | 30 | 60 | 10 | 50 | 0.0490 | 0.48 | 0.834 | 0.196 | 2.14 | 0.188 | 6.0 | 3.5 | 1.0 |
| 3 A | Plana | 30 | 60 | 10 | 50 | 0.0490 | 0.66 | 0.693 | 0.196 | 2.95 | 0.177 | 4.0 | 3.0 | 2.5 |
| 4 A | Plana | 30 | 60 | 20 | 50 | 0.0909 | 0.48 | 2.326 | 0.358 | 2.14 | 0.265 | 9.0 | 6.0 | 5.0 |
| 5 A | Plana | 30 | 60 | 10 | 25 | 0.0225 | 0.24 | 0.831 | 0.360 | 1.50 | 0.376 | 8.0 | 5.0 | 4.0 |
| 6 A | Plana | 30 | 60 | 10 | 25 | 0.0225 | 0.28 | 0.562 | 0.360 | 1.76 | 0.330 | 8.0 | 5.0 | 4.0 |
| 7 A | Plana | 30 | 60 | 10 | 25 | 0.0225 | 0.52 | 0.430 | 0.360 | 1.27 | 0.302 | 3.5 | 3.0 | 1.0 |
| 8 A | Plana | 30 | 60 | 20 | 25 | 0.0350 | 0.32 | 1.918 | 0.528 | 2.00 | 0.493 | 10.5 | 8.0 | 6.0 |
| 1 B | Plana | 30 | 90 | 10 | 25 | 0.0346 | 0.22 | 1.481 | 0.553 | 1.37 | 0.456 | 15.0 | 9.0 | 7.0 |
| 2 B | Plana | 30 | 90 | 10 | 25 | 0.0346 | 0.18 | 0.970 | 0.553 | 1.12 | 0.396 | 16.0 | 9.0 | 7.0 |
| 3 B | Plana | 30 | 90 | 10 | 25 | 0.0346 | 0.62 | 0.864 | 0.553 | 3.90 | 0.381 | 4.5 | 4.0 | 4.0 |
| 4 B | Plana | 30 | 90 | 20 | 25 | 0.0520 | 0.32 | 4.135 | 0.832 | 2.00 | 0.642 | 16.0 | 12.5 | 11.0 |
| 5 B | Plana | 30 | 90 | 10 | 50 | 0.0780 | 0.36 | 4.212 | 0.312 | 1.60 | 0.323 | 15.0 | 8.0 | 7.0 |
| 6 B | Plana | 30 | 90 | 10 | 50 | 0.0780 | 0.36 | 3.049 | 0.312 | 1.60 | 0.290 | 15.0 | 7.5 | 6.5 |
| 7 B | Plana | 30 | 90 | 10 | 50 | 0.0780 | 0.56 | 2.248 | 0.312 | 2.50 | 0.262 | 8.0 | 7.0 | 6.0 |
| 8 B | Plana | 30 | 90 | 20 | 50 | 0.1390 | 0.54 | 7.189 | 0.558 | 2.40 | 0.386 | 15.0 | 11.0 | 9.5 |
| 1 C | Plana | 60 | 90 | 10 | 50 | 0.0252 | 0.17 | 0.204 | 0.100 | 0.76 | 0.118 | 7.0 | 2.5 | 2.0 |
| 2 C | Plana | 60 | 90 | 10 | 50 | 0.0252 | 0.22 | 0.236 | 0.100 | 0.97 | 0.123 | 7.0 | 2.0 | 2.0 |
| 3 C | Plana | 60 | 90 | 10 | 50 | 0.0252 | 0.30 | 0.251 | 0.100 | 1.34 | 0.126 | 3.5 | 1.5 | 1.5 |
| 4 C | Plana | 60 | 90 | 20 | 50 | 0.0440 | 0.36 | 0.933 | 0.176 | 1.61 | 0.193 | 11.0 | 3.5 | 2.0 |
| 5 C | Plana | 60 | 90 | 10 | 25 | 0.0116 | 0.18 | 0.153 | 0.186 | 1.13 | 0.214 | 6.0 | 1.5 | 1.5 |
| 6 C | Plana | 60 | 90 | 10 | 25 | 0.0116 | 0.18 | 0.165 | 0.186 | 1.13 | 0.220 | 6.0 | 2.0 | 1.5 |
| 7 C | Plana | 60 | 90 | 10 | 25 | 0.0116 | 0.30 | 0.162 | 0.186 | 1.89 | 0.220 | 3.5 | 2.0 | 1.5 |
| 8 C | Plana | 60 | 90 | 20 | 25 | 0.0173 | 0.20 | 0.350 | 0.277 | 1.26 | 0.282 | 10.0 | 4.5 | 3.0 |
| 1 D | Plana | 60 | 90 | 30 | 50 | 0.0620 | 0.48 | 1.098 | 0.248 | 2.14 | 0.207 | 13.0 | 5.0 | 4.0 |
| 2 D | Plana | 60 | 90 | 30 | 50 | 0.0620 | 0.30 | 1.390 | 0.248 | 1.34 | 0.224 | 12.5 | 5.0 | 4.0 |
| 3 D | Plana | 60 | 90 | 30 | 50 | 0.0620 | 0.60 | 1.125 | 0.248 | 2.68 | 0.208 | 6.0 | 5.0 | 4.0 |
| 4 D | Plana | 60 | 90 | 30 | 25 | 0.0180 | 0.15 | 0.717 | 0.288 | 0.94 | 0.357 | 14.0 | 4.5 | 4.0 |
| 5 D | Plana | 60 | 90 | 30 | 25 | 0.0180 | 0.20 | 0.545 | 0.288 | 1.25 | 0.327 | 11.0 | 4.5 | 4.0 |
| 6 D | Plana | 60 | 90 | 30 | 25 | 0.0180 | 0.36 | 0.398 | 0.288 | 2.25 | 0.294 | 5.0 | 4.0 | 3.0 |
| 2 E | Plana | 30 | 60 | 30 | 50 | 0.1057 | 0.48 | 5.426 | 0.422 | 2.13 | 0.351 | 13.0 | 9.0 | 7.0 |
| 3 E | Plana | 30 | 60 | 30 | 50 | 0.1057 | 1.00 | 2.428 | 0.422 | 4.40 | 0.268 | 6.0 | 5.5 | 4.5 |
| 4 E | Plana | 30 | 60 | 30 | 25 | 0.0363 | 0.36 | 1.308 | 0.580 | 2.25 | 0.437 | 12.0 | 11.0 | 9.0 |
| 5 E | Plana | 30 | 60 | 30 | 25 | 0.0363 | 0.62 | 1.052 | 0.580 | 3.90 | 0.408 | 6.0 | 5.0 | 4.0 |
| 1 F | Plana | 30 | 90 | 30 | 25 | 0.0541 | 0.26 | 5.163 | 0.866 | 1.64 | 0.688 | 11.0 | 11.0 | 11.0 |
| 2 F | Plana | 30 | 90 | 30 | 25 | 0.0541 | 0.24 | 4.361 | 0.866 | 1.50 | 0.652 | 10.0 | 9.5 | 9.5 |
| 3 F | Plana | 30 | 90 | 30 | 25 | 0.0541 | 0.58 | 2.092 | 0.866 | 3.65 | 0.510 | 8.0 | 8.0 | 7.0 |
| 4 F | Plana | 30 | 90 | 30 | 50 | 0.1916 | 0.56 | 13.536 | 0.768 | 2.48 | 0.478 | 20.0 | 16.0 | 14.0 |
| 5 F | Plana | 30 | 90 | 30 | 50 | 0.1916 | 0.54 | 16.416 | 0.768 | 2.45 | 0.508 | 20.0 | 16.0 | 14.0 |
| 1 G | Circular | -30 | 60 | 10 | 50 | 0.0836 | 0.54 | 2.142 | 0.334 | 2.38 | 0.257 | 7.0 | 5.0 | 4.0 |
| 2 G | Circular | -30 | 60 | 10 | 50 | 0.0716 | 0.42 | 2.010 | 0.286 | 1.87 | 0.253 | 9.0 | 5.0 | 4.0 |
| 3 G | Circular | -30 | 60 | 10 | 50 | 0.0646 | 0.96 | 0.258 | 0.258 | 4.28 | 0.267 | 4.0 | 3.0 | 3.0 |
| 4 G | Circular | -30 | 60 | 10 | 25 | 0.0179 | 0.28 | 0.298 | 0.287 | 1.76 | 0.267 | 6.0 | 4.0 | 3.0 |
| 5 G | Circular | -30 | 60 | 10 | 25 | 0.0143 | 0.32 | 0.292 | 0.228 | 2.02 | 0.265 | 5.0 | 3.0 | 2.0 |
| 1 I | Circular | -30 | 90 | 10 | 25 | 0.0104 | 0.26 | 0.228 | 0.167 | 1.64 | 0.244 | 4.0 | 2.0 | 1.5 |
| 2 I | Circular | -30 | 90 | 10 | 25 | 0.0123 | 0.26 | 0.280 | 0.197 | 1.64 | 0.234 | 4.0 | 2.0 | 2.0 |
| 3 I | Circular | -30 | 90 | 10 | 25 | 0.0124 | 0.32 | 0.197 | 0.199 | 2.02 | 0.230 | 3.0 | 2.0 | 2.0 |
| 4 I | Circular | -30 | 90 | 10 | 50 | 0.0495 | 0.48 | 1.435 | 0.198 | 2.14 | 0.226 | 7.5 | 4.0 | 4.0 |
| 5 I | Circular | -30 | 90 | 10 | 50 | 0.0425 | 0.40 | 0.930 | 0.170 | 1.79 | 0.195 | 7.0 | 4.0 | 3.0 |
| 6 I | Circular | -30 | 90 | 10 | 50 | 0.0261 | 0.54 | 0.351 | 0.104 | 2.60 | 0.141 | 3.0 | 2.5 | 2.0 |
| 1 N | Circular | 0 | 60 | 10 | 25 | 0.0367 | 0.28 | 1.717 | 0.572 | 1.76 | 0.480 | 10.5 | 6.5 | 5.75 |
| 2 N | Circular | 0 | 60 | 10 | 25 | 0.0350 | 0.26 | 0.997 | 0.560 | 1.63 | 0.400 | 10.5 | 5.5 | 4.5 |
| 3 N | Circular | 0 | 60 | 10 | 25 | 0.0320 | 0.52 | 0.934 | 0.512 | 3.27 | 0.390 | 4.0 | 4.0 | 3.5 |
| 4 N | Circular | 0 | 60 | 10 | 50 | 0.0624 | 0.56 | 0.885 | 0.250 | 2.50 | 0.190 | 3.0 | 1.5 | 1.5 |
| 5 N | Circular | 0 | 60 | 10 | 50 | 0.0620 | 0.42 | 1.683 | 0.248 | 1.86 | 0.238 | 3.0 | 1.75 | 1.5 |
| 6 N | Circular | 0 | 60 | 10 | 50 | 0.0611 | 0.54 | 0.332 | 0.245 | 3.06 | 0.138 | 2.0 | 1.5 | 0.75 |
| 1 P | Circular | 0 | 90 | 10 | 50 | 0.1037 | 0.52 | 3.507 | 0.415 | 2.32 | 0.304 | 12.0 | 9.0 | 7.0 |
| 2 P | Circular | 0 | 90 | 10 | 50 | 0.1044 | 0.56 | 3.849 | 0.418 | 2.50 | 0.314 | 10.5 | 7.5 | 6.5 |
| 3 P | Circular | 0 | 90 | 10 | 50 | 0.0565 | 0.70 | 1.028 | 0.227 | 3.12 | 0.202 | 5.5 | 4.5 | 4.0 |
| 4 P | Circular | 0 | 90 | 10 | 25 | 0.0458 | 0.54 | 1.526 | 0.732 | 2.01 | 0.608 | 10.5 | 8.0 | 8.0 |
| 4 V | Circular | 30 | 90 | 10 | 25 | 0.0637 | 0.70 | 3.844 | 1.010 | 4.37 | 0.628 | 10.0 | 10.0 | 9.0 |
| 5 V | Circular | 30 | 90 | 10 | 25 | 0.0768 | 0.72 | 4.575 | 1.230 | 4.52 | 0.662 | 11.0 | 11.0 | 11.0 |
| 6 V | Circular | 30 | 90 | 10 | 25 | 0.0557 | 1.80 | 1.510 | 0.893 | 5.01 | 0.500 | 6.0 | 6.0 | 5.5 |

- α Ángulo de incidencia del cuerpo del derrumbe
- θ Talud de la margen del derrumbe
- Z Altura de la base del derrumbe sobre el fondo del embalse.
- d Profundidad del embalse
- V Volumen de agua desplazado por el derrumbe
- t_c Tiempo de duración del derrumbe
- E Energía del movimiento ondulatorio
- g Aceleración de la gravedad
- γ Peso específico del agua
- h_m Elevación máxima de la cresta de la onda
- $h_{2.5}$ Altura de onda relativa a una distancia del origen de 2.5 veces el tirante
- h_5 Altura de onda relativa a una distancia del origen de 5 veces el tirante.

En la gráfica 3 se relacionan valores de energía con alturas de onda relativa a una distancia del origen de 5 veces el tirante. De esta gráfica se obtendrán los siguientes índices:

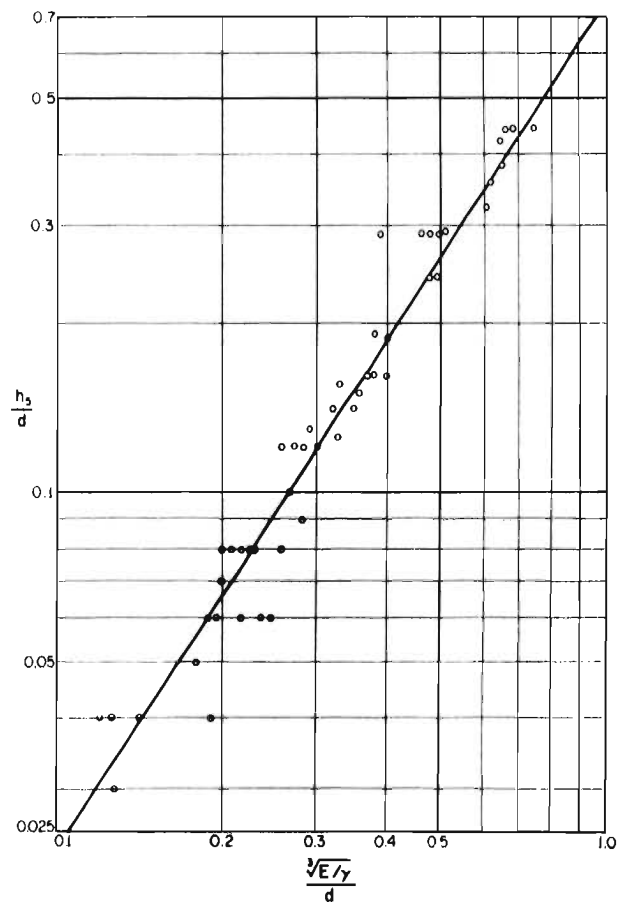
- 1) Correlación Real
- 2) Correlación Espuria
- 3) Grado espurio
- 4) Correlación exagerada
- 5) Correlación parcial implícita.

Entonces, sea $y = \frac{h_5}{d}$ y $Z = \frac{\sqrt[3]{E/\gamma}}{d}$

Si se hace.

$$x_1 = h_5$$

$$x_2 = d$$



Gráfica 3.- Relación de la energía con la altura de onda en $X/d = 5.0$

$$x_3 = E$$

$$x_4 = \gamma$$

Queda una relación de $y = \frac{x_1}{x_2}$ contra $Z = \frac{\sqrt[3]{x_3/x_4}}{x_2}$

i) Correlación Real.-

Sea $y = f(x_1, x_2)$, la función es $f(m_1, m_2) = \frac{m_1}{m_2}$ y

$Z = f(x_2, x_3, x_4)$, la función es $F(m_2, m_3, m_4) = \frac{\sqrt[3]{m_3/m_4}}{m_2}$

De la ec. (1.1)

Para $a = 1$, $f_1 = \frac{1}{m_2}$

Para $a = 2$, $f_2 = -\frac{m_1}{m_2^2}$

De la ec. (1.2)

Para $b = 2$, $F_2 = -\frac{1}{m_2^2} \left(\frac{m_3}{m_4}\right)^{1/3}$

Para $b = 3$, $F_3 = \frac{1}{3m_2 \left(m_3^2/m_4\right)^{1/3}}$

Para $b = 4$, $F_4 = -\frac{1}{3m_2} \left(\frac{m_3}{m_4^3}\right)^{1/3}$

Ahora el numerador de la ec. (1) es.

$$\sum_{a=1}^2 \sum_{b=2}^4 f_a F_b r_{ab} \sigma_a \sigma_b =$$

$$= f_1 F_2 r_{12} \sigma_1 \sigma_2 + f_1 F_3 r_{13} \sigma_1 \sigma_3 + f_1 F_4 r_{14} \sigma_1 \sigma_4 +$$

$$+ f_2 F_2 r_{22} \sigma_2 \sigma_2 + f_2 F_3 r_{23} \sigma_2 \sigma_3 + f_2 F_4 r_{24} \sigma_2 \sigma_4$$

Sustituyendo los valores anteriores y tomando en cuenta que $r_{22} = 1$.

Luego, multiplicando todo por $\frac{3m_2^2}{m_1} \left(\frac{m_4}{m_3}\right)^{1/3}$ y sustituyendo $C_i = \frac{\sigma_i}{m_i}$

donde $1 \leq i \leq 4$, queda.

$$- 3 r_{12} C_1 C_2 + r_{13} C_1 C_3 - r_{14} C_1 C_4 + C_2^2 -$$

$$- r_{23} C_2 C_3 + r_{24} C_2 C_4 .$$

Ahora para el primer término del denominador de la ec. (1).

$$\left(\sum_{a=1}^2 \sum_{c=1}^2 f_a f_c r_{ac} \sigma_a \sigma_c \right)^{1/2} =$$

$$= \left(f_1 f_1 r_{11} \sigma_1 \sigma_1 + f_1 f_2 r_{12} \sigma_1 \sigma_2 + f_2 f_1 r_{21} \sigma_2 \sigma_1 + \right.$$

$$\left. + f_2 f_2 r_{22} \sigma_2 \sigma_2 \right)^{1/2}$$

Sustituyendo valores y multiplicando todo por $\left(\frac{m_2^2}{m_1^2}\right)^{1/2}$. Y luego

go sustituyendo las Ces, queda.

$$\text{Como } r_{11} = r_{22} = 1.$$

$$= \left(C_1^2 + C_2^2 - 2 r_{12} \sigma_1 \sigma_2 \right)^{1/2}$$

y para el segundo término del denominador de la ec. (1).

$$\left(\sum_{a=2}^4 \sum_{b=2}^4 F_a F_b r_{ab} \sigma_a \sigma_b \right)^{1/2} =$$

$$= \left(F_2 F_2 r_{22} \sigma_2 \sigma_2 + F_2 F_3 r_{23} \sigma_2 \sigma_3 + F_2 F_4 r_{24} \sigma_2 \sigma_4 + \right.$$

$$+ F_3 F_2 r_{32} \sigma_3 \sigma_2 + F_3 F_3 r_{33} \sigma_3 \sigma_3 + F_3 F_4 r_{34} \sigma_3 \sigma_4 +$$

$$\left. + F_4 F_2 r_{42} \sigma_4 \sigma_2 + F_4 F_3 r_{43} \sigma_4 \sigma_3 + F_4 F_4 r_{44} \sigma_4 \sigma_4 \right)^{1/2}$$

Sustituyendo los valores correspondientes; multiplicando todo por -

$$\left[9m_2^2 \left(\frac{m_4^2}{m_3^2} \right)^{1/2} \right]^{1/2} \text{ y sustituyendo las Ces.}$$

queda.

$$\begin{aligned} \text{Ya que } r_{22} &= r_{33} = r_{44} = 1 \\ &= (9C_2^2 - 6r_{23}C_2C_3 + 3r_{24}C_2C_4 + C_3^2 \\ &\quad - 2r_{34}C_3C_4 + C_4^2)^{1/2} \end{aligned}$$

Ahora sustituyendo los valores anteriores en la ec. (1).

$$\begin{aligned} r_{yz} &= \frac{-3r_{12}C_1C_2 + r_{13}C_1C_3 - r_{14}C_1C_4 + C_2^2 - r_{23}C_2C_3}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2} (9C_2^2 - 6r_{23}C_2C_3 + \\ &\quad + r_{24}C_2C_4 \\ &\quad + 3r_{24}C_2C_4 + C_3^2 - 2r_{34}C_3C_4 + C_4^2)^{1/2}} \quad (12) \end{aligned}$$

Siendo la correlación real más general; ahora que si solo van a variar x_1 , x_2 y x_3 , dejando constante a x_4 , entonces la correlación real va a ser (teniéndose $C_4 = 0$).

$$r_{y,z} = \frac{-3r_{12}C_1C_2 + r_{13}C_1C_3 + C_2^2 - r_{23}C_2C_3}{(C_1^2 + C_2^2 - 2r_{12}C_1C_2)^{1/2} (9C_2^2 - 6r_{23}C_2C_3 + C_3^2)^{1/2}} \quad (13)$$

Ahora de la Tabla 2.

Para la variable $h_5 = x_1$, se obtiene.

$$m_1 = 4.857 \text{ y } \sigma_1 = 3.112$$

Como más del 63% de los valores de x_1 se encuentran dentro del intervalo.

$$m_1 - \sigma_1 \leq x_1 \leq m_1 + \sigma_1$$

$$1.745 \leq x_1 \leq 7.969$$

Se puede decir que x_1 es una variable distribuida normalmente. Por lo tanto.

$$C_1 = \frac{\sigma_1}{m_1} = \frac{3.112}{4.857} = 0.641$$

Para la variable $d = x_2$; como en todo el conjunto de valores de dicha variable la mitad de éstos tienen un valor de 25 y la otra mitad de 50, se hará la hipótesis de que existe una función de distribución binomial con $p = 1/2$ y $q = 1/2$. Luego tomando en cuenta el siguiente Teorema - (16):

"A medida que el número de pruebas, n , aumenta indefinidamente, la distribución de la variable binomial estandarizada se aproxima a la llamada distribución normal estándar". Se puede suponer que la variable x_2 sigue una distribución normal con $n = 63$, la media y la desviación estándar de la función de distribución binomial tienen los siguientes valores:

$$m_2 = np = 63 \left(\frac{1}{2}\right) = 31.5, \text{ y}$$

$$\sigma_2 = \sqrt{npq} = \sqrt{63 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)} = 3.975$$

Por lo tanto.

$$C_2 = \frac{\sigma_2}{m_2} = \frac{3.975}{31.5} = 0.126$$

Para la variable $E = x_3$, se obtiene

$$m_3 = 2.078 \text{ y } \sigma_3 = 2.852$$

Siguiendo aproximadamente una distribución normal. Por lo tanto.

$$C_3 = \frac{r_3}{m_3} = \frac{2.852}{2.078} = 1.372$$

Los coeficientes de correlación son:

$$r_{12} = -0.2367$$

$$r_{13} = 0.8332$$

$$r_{23} = 0.7149$$

Substituyendo
Despejando en la ec. (13) los valores anteriores, la correlación real queda de $r_{yz} = 0.88$, (88%). Por lo que se pueden inferir con cierta aproximación valores de las variables no repetidas de la gráfica 3, teniendo un porcentaje del 12% de error, el cual es muy bajo.

Debido al alto grado de correlación real en dicha gráfica, se puede suponer que la correlación parcial espuria existente tendrá un valor mínimo. Lo cual se comprobará en el siguiente inciso.

2).- CORRELACION ESPURIA.-

De la ec. (13), para $r_{12} = r_{13} = r_{23} = 0$, la correlación espuria es.

$$\rho_{yz} = \frac{C_2'}{(C_1^2 + C_2^2)^{1/2} (9C_2^2 + C_3^2)^{1/2}}$$

Despejando los valores correspondientes, queda $\rho_{yz} = 0.017$, (1.7%). La cual resulta una correlación espuria bastante baja, como se había supuesto anteriormente. Este valor de 1.7% también implica un valor muy pequeño de grado espurio, como se ve en el siguiente inciso.

3) GRADO ESPURIO.-

Como r_{yz} (real) = 0.880

y ρ_{yz} (esp.) = 0.017,

la gráfica 3 tendrá una correlación parcialmente espuria, con un grado de falsedad del 1.93%.

4) CORRELACION EXAGERADA.-

Si se obtiene la correlación de la gráfica 3 como la de dos variables únicas $X = \frac{\sqrt[3]{E/y}}{d}$ y $Y = \frac{h5}{d}$

Las variables X y Y siguen distribuciones normales, con.

$$m_x = 0.3265$$

$$\sigma_x = 0.1473$$

$$m_y = 0.1488$$

$$\sigma_y = 0.1108$$

$$\text{Luego, } C_x = \frac{m_x}{\sigma_x} = \frac{0.3265}{0.1473} = 2.2166$$

$$\text{y } C_y = \frac{m_y}{\sigma_y} = \frac{0.1488}{0.1108} = 1.3430$$

Teniendo una correlación entre X y Y de.

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.016}{(0.1473)(0.1108)}$$

$$r_{x,y} = 0.98 \quad (\text{correlación exagerada})$$

Cuando en realidad se tiene una correlación de.

$$r_{x,y} (\text{real}) = 0.88 \quad (\text{inciso 1})$$

Por lo que habrá una exageración del 10%, debido a no hacer uso de la fórmula de correlación real (ec. (13))

5) CORRELACION PARCIAL IMPLICITA.-

Suponiendo que solo se tienen los ensayos E de la tabla 2, y que se quiere saber la correlación entre las variables $h_5 = x_1$ y $\sqrt{E/y} = x_3$ pero que la correlación entre x_1 y x_3 fuera debido a la asociación entre cada una de las dos variables y una tercera variable $d = x_2$.

Como los datos disponibles no cumplen con la suposición de normalidad, además de que la muestra es pequeña, se utilizarán los coeficientes de correlación no paramétricos. Así el coeficiente de correlación parcial implícito de rango de Kendall $z_{13,2}$ (deducido de la ec. (9)) es.

$$z_{13,2} = \frac{z_{13} - z_{23} z_{12}}{\sqrt{1 - z_{23}^2} \sqrt{1 - z_{12}^2}} \quad (14)$$

Entonces, teniendo para z_{13} .

| | | | | |
|-------|-------|------|--------|------|
| x_1 | 7 | 4.5 | 9 | 4 |
| x_3 | 17.55 | 13.4 | 10.925 | 10.2 |

Ordenando para que los rangos de x_1 aparezcan en orden natural.

| | | | | |
|-------|------|------|-------|--------|
| x_1 | 4 | 4.5 | 7 | 9 |
| x_3 | 10.2 | 13.4 | 17.55 | 10.925 |

donde. $S = (3 - 0) + (1 -) + (0 - 1) = 2$

de la ec. (10), $z_{13} = \frac{2}{\frac{1}{2} 4 (4 - 1)} = 0.33$

Para z_{23} .

| | | | | |
|-------|-------|------|--------|------|
| X_3 | 17.55 | 13.4 | 10.925 | 10.2 |
| X_2 | 50 | 50 | 25 | 25 |

Ordenando para que los rangos de X_3 aparezcan en orden natural.

| | | | | |
|-------|------|--------|------|-------|
| X_3 | 10.2 | 10.925 | 13.4 | 17.55 |
| X_2 | 25 | 25 | 50 | 50 |

Debido a que en X_2 hay repeticiones en 25 y 30, el coeficiente de correlación se obtendrá de la ec. (11).

$$\text{Donde. } S = (1 - 0) + (1 - 0) + (0 - 0) = 2$$

$$T_{X_3} = 0$$

$$T_{X_2} = \frac{1}{2} [2(2 - 1) + 2(2 - 1)] = 2$$

Sustituyendo en la ec. (11)., queda.

$$z_{32} = z_{23} = 0.408$$

Para z_{12}

| | | | | |
|-------|----|-----|----|----|
| X_1 | 7 | 4.5 | 9 | 4 |
| X_2 | 50 | 50 | 25 | 25 |

Ordenando para que los rangos de X_1 aparezcan en orden natural.

| | | | | |
|-------|-----|-----|----|----|
| X_1 | 4 | 4.5 | 7 | 9 |
| X_2 | 2.5 | 50 | 50 | 55 |

donde, $S = (1 - 0) + (0 - 1) + (0 - 1) = -1$

$$T_{v_1} = 0$$

$$T_{v_2} = 2$$

De la ec. (11), $Z_{1,2} = -0.201$

Finalmente sustituyendo valores en la ec. (14), la correlación parcial implícita queda $Z_{1,2} = 0.46$. Lo que quiere decir que la correlación entre las variables x_1 y x_2 , dejando constante a la variable x_3 es de 0.46 ó 46 %.

De los resultados generales encontrados anteriormente:

$$r_{x_1} = 0.88 \text{ Correlación Real}$$

$$r'_{x_1} = 0.017 \text{ Correlación Espuria y}$$

$$\text{grado Espurio} = 1.93 \%$$

Se puede afirmar que la gráfica 3 está bien obtenida, ya que los valores de las correlaciones y el grado espurio son aceptables.

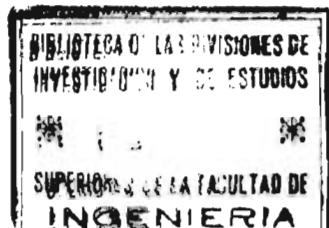
En lo que respecta a los resultados particulares:

$$r_{x_1} = 0.98 \text{ Correlación exagerada, con una exageración del 10\% sobre la real, y}$$

$$Z_{1,2} = 0.46 \text{ Correlación parcial implícita.}$$

La exageración del 10% resulta aceptable.

La correlación parcial implícita $Z_{1,2} = 0.46$ es mayor que la correlación $Z_{1,3} = 0.33$, por lo que la correlación entre las variables $x_1 = h_1$ y $x_2 = \sqrt[3]{E/\gamma}$, dejando como constante a la variable $x_3 = d$ resulta mayor que entre las mismas variables, dejándose ver la acción de la variable x_3 sobre las otras.



C A P I T U L O VIII

CONCLUSIONES

En este trabajo se analizó la llamada CORRELACION ESPURIA que se -- presenta cuando en una gráfica existen variables comunes a ambos ejes. Es ta es de gran importancia en la Ingeniería Experimental, ya que de dicha - correlación dependerá la inferencia de las variables no-repetidas. Se ca| cula la correlación espuria a partir de la correlación real de Pearson y - se analizan los modelos más comunes en la Hidráulica y la Mecánica de los- Fluidos. Llegando a las siguientes conclusiones:

1).- Si en un plano coordenado se están relacionando grupos de va-- riables aleatorias en donde existe alguna o algunas comunes a ambos ejes,- existirá una cierta correlación espuria, ya que la repetición de éstas va- riará la correlación nula debido a que se están correlacionando variables- iguales y por lo tanto se está alterando toda la interacción entre ellas.

2).- En el caso de que las variables involucradas no fueran aleato- rias sino que tuvieran una cierta intercorrelación, la repetición de las - variables también afectará la correlación general.

3).- La presencia de la correlación espuria puede causar una limita- ción de la dispersión real, debido a la presencia de los elementos comunes.

4).- La relación espuria puede estar basada parcial o totalmente en las características de los elementos comunes.

5).- Si existe espureidad habrá poca o ninguna confianza en una extensión de la relación más allá del rango de datos.

6).- En la tabla 1 se presentan los modelos de relaciones más comunes en la Hidráulica, y en la gráfica 1 se pueden obtener rápidamente los valores de correlación espuria de dichos modelos.

7).- La correlación exagerada es una correlación mal obtenida, debido a no hacer uso de la fórmula de correlación real (Pearson) para grupos de variables (columna 2, tabla 1).

8).- La correlación parcial implícita no es una forma de correlación espuria, sino más bien es un índice de asociación entre dos variables y -- una tercera que se mantiene constante. Esta correlación parcial depende de las variables en sí, en cambio la correlación espuria depende del acomodo de las variables que se están correlacionando.

9).- La correlación real de Pearson en comparación con las anteriores correlaciones depende de las variables en sí y del acomodo de las mismas.

10).- La correlación espuria se puede presentar no solamente bidimensionalmente sino en 3 o más dimensiones.

B I B L I O G R A F I A

1. Spiegel, M.R., "Statistics", Schaum's Outlines Series, 1961.
2. Pearson, K., "On a form of spurious correlation which may arise when indices are used in the measurement of organs", Proceedings, Royal Soc. of London, Vol. 60, 489 - 502.
3. Reed, L.J., "On the correlation between any two functions and its application to the general case of spurious correlation" -- Mathematis, Journal of the Washington Academy of Sciences, Vol. 11, No. 19, Noviembre 1921, 449 - 455.
4. Chayes, F., "On ratio correlation in petrography", Journal of - Geology, Vol. 57, No. 3, 1949, 239 - 254.
5. Benson, M.A., " Spurious correlation in hydraulics and hydrology", Journal of the Hydraulics Division, Julio 1965, 35 - 42.
6. Yalin, M.S. y Kamphuis, J.W., "Theory of dimensions and spurious correlation", Journal of Hydraulic Research (IARR), Vol. 9, 1971 No. 2, 248 - 265.
7. Yalin, M. Selim, "Geometrical properties of sand waves", Journal of the Hydraulics Division, ASCE, Vol. 90, No. HY5, Proc. Paper 4055, Septiembre, 1964.
8. Hely, A.G., y Olmsted, F.H., " Some relations between Streamflow characteristics and the environment in the Delaware river basin", Professional Paper 417-B, Geol Survey, U.S. Dept. of interior, - Washington, D.C., 1963.

9. Yevdjevich, F.M., "Fluctuations of wet and dry years, Part. I : Research data assembly and mathematical models", Hydrology --- Papers, Colorado State Univ., Fort Collins, Colo., 1963.
10. Mirajgaoker, A.G., y Charlu, K.L.N., "Natural roughness effects in rigid open channels", Journal of the Hydraulics Division, - ASCE, Vol. 89, No. HY5, Proc. Paper 3630, Septiembre, 1963.
11. Ezequiel, M. y Fox, K.A., "Methods of correlation and regression analysis", Wiley and Sons, 1959.
12. Siegel, S., "Estadística no Paramétrica", Trillas, 1972.
13. Daily, J.W. y Harleman, D.R.F., "Dinámica de los fluidos", Trillas, México, 1969.
14. Bowman, C.C. y Hansen, V.E., "Simplification of dimensional -- analysis", Journal of the Engineering Mechanics Division, ASCE, EM 1, Enero, 1959.
15. Cruickshank, C.V., "Derrumbes en embalses", Instituto de Ingeniería, No. 152, Enero, 1969.

16- Falta. (ver pag. 42)