ISBN: 978-612-48342-4-0

INTRODUCCIÓN A LA BIBLIOMETRÍA PRÁCTICA

INTRODUCCIÓN A LA BIBLIOMETRÍA PRÁCTICA

FAUSTO FRANCISCO MATOS URIBE
FORTUNATO CONTRERAS CONTRERAS
JULIO CESAR OLAYA GUERRERO

INTRODUCCIÓN A LA BIBLIOMETRÍA PRÁCTICA

FAUSTO FRANCISCO MATOS URIBE
FORTUNATO CONTRERAS CONTRERAS
JULIO CESAR OLAYA GUERRERO

ASOCIACION DE BIBLIOTECÓLOGOS DEL PERU

Introducción a la Bibliometría Práctica

Autores:
©Fausto Francisco Matos Uribe
©Julio César Olaya Guerrero
©Fortunato Contreras Contreras

Editado por
©Asociación de Bibliotecólogos del Perú
Calle Hipólito Bernardette Nº 106 – Barranco

Telef. 997684606
contacto@abp.org.pe
Lima – Perú
Primera edición, marzo 2023
Tiraje: 100 ejemplares

Hecho en depósito Legal en la Biblioteca Nacional del Perú Nº.°2023-01866

ISBN: 978-612-48342-4-0

Se terminó de imprimir en marzo del 2023 en:
Oscar Ricardo Retamozo Ramos
Avenida Ignacio Merino Nº 1738 - Lince

PROLOGO

El presente libro, es producto de las experiencias en la enseñanza del curso de Bibliometria, que tiene como prerrequisito los cursos de matemática básica, estadística descriptiva y estadística inferencial, esperando que este material sirva de ayuda al estudiante y profesional de la carrera de bibliotecología y ciencia de la información, el cual tiene por objetivo realizar un análisis bibliométrico de la literatura científica.

El libro se divide en 4 capítulos, acompañado de teoría y práctica para su mejor compresión.

LOS AUTORES

Índice

Pág.

Capítulo 1

1.1. Bibliometría. 7

1.2. Cienciometría. 8

1.3. Informetría. 8

1.4. Análisis bibliométrico 9

Capítulo 2

2.1. Conceptos matemáticos. 10

2.1.1. Ecuaciones exponenciales, ejercicios resueltos y propuestos. 10

2.1.2. Logaritmos, ejercicios resueltos y propuestos. 11

2.1.3. Progresión geométrica, ejercicios resueltos y propuestos. 15

2.1.4. Graficas de funciones, ejercicios resueltos y propuestos. 17

2.2. Conceptos estadísticos 18

2.2.1. Media geométrica, ejercicios resueltos y propuestos 18

2.2.2. Media armónica, ejercicios resueltos y propuestos. 19

2.2.3. Hipótesis estadística: Prueba de hipótesis de la media y proporción poblacional. 20

2.2.4. Medias de asociación. 30

2.2.5. Regresión lineal múltiple y simple, ejercicios resueltos y propuestos. 37

2.2.6. Coeficiente de determinación, ejercicios resueltos y propuestos. 41

2.2.7. Regresión lineal con datos codificados, ejercicios resueltos y propuestos. 44

2.2.8. Regresión no lineal, ejercicios resueltos y propuestos. 48

Capítulo 3

3.1. Leyes bibliométricas. 54

3.2. Modelo de Price, ejercicios resueltos y propuestos. 54

3.3. Modelo de Lotka, ejercicios resueltos y propuestos. 60

3.4. Modelo de Bradford, ejercicios resueltos y propuestos. 66

3.5. Ley de Zipf, ejercicios resueltos y propuestos 71

Capítulo 4

4.1. Indicadores bibliométricos 74

4.2. El número de citas 74

4.3. Índice H de un investigador. 75

4.4. Índice i10 de un investigador. 77
4.5. Factor de impacto de una revista científica (FI) 77
4.6. Índice de inmediatez de una revista (Ii) 78
4.7. Vida media de un trabajo de investigación (Vm) 79
4.8. Cuartiles 80
4.9. Índice de colaboración (IC) 81
4.10. Grado de colaboración (GC) 81
4.11. Coeficiente de colaboración (CC) 81

Bibliográfica 83

Anexos
Anexo 01: Demostraciones 84
Anexo 02: Tabla estadísticas 86

Capítulo I

Los términos bibliometría y cienciometría fueron introducidos casi simultáneamente por Pritchard y por Nalimov y Mulchenko en 1969. Mientras que Pritchard explicó el término bibliometría como “la aplicación de métodos matemáticos y estadísticos a libros y otros medios de comunicación”, Nalimov y Mulchenko definieron la cienciometría como “la aplicación de aquellos métodos cuantitativos que se ocupan del análisis de la ciencia vista como un proceso de información”. De acuerdo con estas interpretaciones, la especialidad de la cienciometría se restringe a la medición de la comunicación científica, mientras que la bibliometría está diseñada para tratar con procesos de información más generales. Los límites de alguna manera difusos entre las dos especialidades casi se desvanecieron durante las últimas tres décadas, y hoy en día ambos términos se usan casi como sinónimos. En cambio, la informetría de campo tomó el lugar de la bibliometría especializada originalmente más amplia. El término informetría fue adoptado por VINITI (Gorkova, 1988) y representa un subcampo más general de las ciencias de la información que se ocupa del análisis estadístico matemático de los procesos de comunicación en la ciencia. En contraste con la definición original de bibliometría, la informetría también se ocupa de los medios electrónicos y, por lo tanto, incluye temas como el análisis estadístico del texto (científico) y los sistemas de hipertexto, las circulaciones de las bibliotecas, las medidas de información en las bibliotecas electrónicas, los modelos para los procesos de producción de información y los análisis cuantitativos. aspectos de la recuperación de información también (Glanzel, 2003: 6)

1.1. Bibliometria

La palabra “bibliometría” deriva de los vocablos griegos *biblos:* ‘ libro’, y *metron:* ’medir’. Es la aplicación de las matemáticas y el método estadístico a la publicación de los resultados de la investigación científica; dicho de otra manera, es la aplicación de tratamientos cuantitativos a la comunicación escrita, producto tangible de la investigación. La bibliometría parte de la necesidad de cuantificar ciertos aspectos de la ciencia para poder comparar, medir y objetivar la actividad científica (Dávila, et.al, 2009: 320).

El análisis estadístico de la literatura científica comenzó casi 50 años antes de que se acuñara el término “bibliometría”. En 1926, *Alfred J. Lotka* publicó su estudio pionero sobre la distribución de frecuencias de la productividad científica determinada a partir de un índice decenal (1907-1916) de *Chemical Abstracts.* Lotka concluyó que

“el número (de autores) que hacen n contribuciones es aproximadamente 1/n² de los que hacen una; y la proporción de todos los contribuyentes, que hace una sola contribución, es de alrededor del 60 por ciento.”

Este resultado puede considerarse como una regla empírica incluso hoy, 75 años después de su publicación.

Casi al mismo tiempo, en 1927, *Gross* y *Gross* publicaron su estudio basado en citas para ayudar a decidir qué publicaciones periódicas de química deberían comprar mejor las bibliotecas universitarias pequeñas. En particular, examinaron 3633 citas del volumen de 1926 del *Journal of the American Chemical Society.* Este estudio se considera el primer análisis de citas, aunque no es un análisis de citas en el sentido de la bibliometría actual.

Ocho años después de que apareciera el artículo de Lotka, *Bradford* (1934) publicó su estudio sobre la distribución de frecuencia de artículos sobre revistas. Encontró que

"si las revistas científicas se ordenan en orden de productividad decreciente sobre un tema dado, pueden dividirse en un núcleo de revistas más particularmente dedicadas al tema y varios grupos o zonas que contienen el mismo número de artículos que el núcleo cuando el número de publicaciones periódicas en el núcleo y las zonas sucesivas será como 1: b : b^2 …"

Zipf (1949) formuló una ley interesante en bibliometría y lingüística cuantitativa que derivó del estudio de la frecuencia de palabras en un texto. Según Zipf rf = C, donde r es el rango de una palabra, f es la frecuencia de aparición de la palabra y C es una constante que depende del texto analizado. Puede considerarse una generalización de las leyes de Lotka y Bradford. Formuló el siguiente principio subyacente de su ley, aunque nunca ha demostrado cómo se aplica este principio a su ecuación. 1

"El Principio del Mínimo Esfuerzo significa... que una persona... se esforzará por resolver sus problemas de tal manera que minimice el trabajo total que debe realizar para resolver tanto sus problemas inmediatos como sus probables problemas futuros...". (Zipf, 1949). (Glanzel, 2003: 6-7).

La Bibliometría estudia el comportamiento de la producción de la literatura científica (libros, revistas científicas, artículos científicos, autores, etc.) y para tal objetivo utiliza a las matemáticas y la estadística. La Bibliometria proporciona información sobre cómo se desenvuelve la producción de la literatura científica ya sea: local, nacional, nacional e internacional, con el objetivo de tomar decisiones.

.

1.2. Cienciometría

Es el estudio de la Ciencia como problema multidimensional que reúne a investigadores + textos + conocimiento. Es decir, es la Ciencia de la Ciencia. (Amezquita et. al, 2011: 12).

Mide el movimiento de la actividad social y tienen con objetos: el análisis de la producción, circulación y consumo, de trabajos científicos.

1.3. Informetría

La infometría es ampliamente utilizada para la medición de variables que intervienen en el análisis del contenido del quehacer investigativo, a través de la implementación de modelos teóricos y medidas de información. Por ello, la las matemáticas y la modulación son herramientas fundamentales en el proceso de gestión de la información. (Amézquita et. al, 2011: 19).

Toma en cuenta los dos conceptos anteriores y desarrolla métodos y procedimientos para medir y analizar los procesos de investigación científicas.

1.4. Análisis bibliométrico

Consiste en realizar el seguimiento de publicaciones científicas y/o autor, tomando como referencia la materia y el periodo de análisis con el objetivo de dar solución a un trabajo científico. Utilizando las bases de datos:

- Scopus,
- Web of Science (WOS),
- Google Scholar,
- PubMed
- MEDLINE
- Journal Citation Reports(JCR)
- Science Citation Index Expandel (SCI)
- Current Contents Connect (CC).

Diferencias entre Bibliometría, Cienciometría e Informetría

Tipología	Bibliometría	Cienciometría	Informetría
Objeto de estudio	Libros, documentos, revistas, artículos, autores y usuarios.	Disciplinas, materias, campos y esferas.	Palabras, documentos y bases de datos.
Variables	Números en circulación, citas, frecuencia de aparición las palabras, longitud de las oraciones. etc	Aspectos que diferencian a las disciplinas y a las subdisciplinas. Revistas autores, trabajos, forma en que se comunican los científicos.	Difieren de la cienciometría en los propósitos de las variables, por ejemplo, medir la recuperación, la relevancia, el recordatorio, etc.
Métodos	Clasificación, frecuencia, distribución.	Análisis de documento y correspondencia.	Modelo rector-espacio, modelos boleanos de recuperación, modelos probabilísticos, lenguaje del procesamiento, enfoques basados en el conocimiento, tesauros.
Objetivos	Asignar recursos, dinero, tiempo, etc.	Identificar esferas de interés; dónde se encuentran las materias; comprender cómo y con qué frecuencia se comunican los científicos.	Aumentar la eficiencia de la recuperación

Fuente: McGrath W. (1989). What bibliometricians, scientometricians and informetricians study; a typology for definition and classification; topics for discusión (citado por Amezquitam, et al, 2011: 22)

Capítulo 2

2.1. Conceptos matemáticos

2.1.1 Ecuaciones exponenciales

Se denomina así aquellas ecuaciones algebraicas, en la cual la incógnita aparece como exponente.

Ejemplos

$$(\frac{1}{8})^{2x+1} = 16$$

$$4^{2x-1} = 8^{3-x}$$

$$x^{x^3} = 3$$

Para resolver una ecuación exponencial nos valemos de algunas propiedades:

$$a^x = a^m \quad \rightarrow x = m$$

$$x^x = b^b \quad \rightarrow x = b$$

$$x^a = a^x \quad \rightarrow x = a$$

Ejercicios resueltos

Calcular el valor de "x"

$$7^{2x-3} = 49^{x^2+x-6}$$

$$(7^2)^{\frac{1}{2}(2x-3)} = 49^{x^2+x-6}$$

$$49^{(x-\frac{3}{2})} = 49^{x^2+x-6}$$

$$x - \frac{3}{2} = x^2 + x - 6 \rightarrow x^2 = \frac{9}{2} \rightarrow x = \pm\frac{3}{2}\sqrt{2}$$

Calcular el valor de "x"

$$3^{x+3} = (\frac{1}{27})^{2x-12}$$

$$3^{x+3} = (3^{-3})^{2x-12} \rightarrow 3^{x+3} = (3)^{-3(2x-12)} \rightarrow x+3 = -3(2x-12)$$

$$x + 3 = -6x + 36 \rightarrow 7x = 33 \rightarrow x = \frac{33}{7}$$

Ejercicios propuestos

En cada uno de los casos calcular el valor de "x"

1. $9^{x^2-7x+12}=1$

2. $9^{x+3}=(\frac{1}{81})^{2x-12}$

3. $(\frac{1}{8})^{2x+1}=16$

4. $x^{x^3}=3$

2.1.2. Logaritmos

El logaritmo de un número real positivo $N>0$, en una base dada b>0 diferente de la unidad, es igual a un exponente x, exponente que debe elevarse a la base para reproducir el número dado.

$$Log_b N = x \rightarrow b^x = N, \quad b>0, \quad b\neq 1 \quad N>0$$

Ejemplo-1

Hallar el logaritmo de 64 en base 2.

$Log_2 64 = x \rightarrow 2^x = 64 \rightarrow 2^x = 2^6 \rightarrow x = 6$

Es decir 6 es el logaritmo de 64 en base 2.

Ejemplo-2

Hallar el logaritmo de 0.001 en base 10

$Log_{10} 0.001 = x \rightarrow 10^x = 0.001 \rightarrow 10^x = 10^{-3} \rightarrow x = -3$

Propiedades de los logaritmos

a. El logaritmo de la unidad es igual a cero, $Log_b 1 = 0$

b. Logaritmo de la base es igual a la unidad, $Log_b b = 1$

c. El logaritmo del producto de varios factores en una base dada, es igual a la suma de los logaritmos de sus factores en la base dada.

d. $Log_b N.P.Q = Log_b N + Log_b P + Log_b Q$

e. El logaritmo del cociente de dos números en una base dada, es igual al logaritmo de la diferencia del numerador y denominador en la base dada.

f. $Log_b \frac{M}{N} = Log_b M - Log_b N$

g. El logaritmo de una potencia N^r en una base dada, es igual al exponente multiplicado por el logaritmo del número en la base dada.

h. $Log_b N^r = r.Log_b N$

i. $Log_{(b/a)}(a/b) = -1$

j. El cambio de base

El logaritmo de un número $"a"$ en una base dada $"b"$, es igual a un cociente, del logaritmo del número $"a"$ y el logaritmo de la base $"b"$ ambas en una base de interés $"c"$.

$$Log_b a = \frac{Log_c a}{Log_c b}, \quad c = una\ base\ cualquiera$$

Regla de la cadena (del cambio de base)

$$Log_b a.\log_c b = Log_c a$$

Corolario de la regla de la cadena

$$si \quad a = c \rightarrow Log_b c.\log_c b = Log_c c = 1$$

Nota

$Log_{10} N = Log\ N$, cuando la base es 10 se omite la base

$Log_e N = Ln\ N$, cuando la base es el número neperiano e=2.718….

Teorema

A ambos lados de una igualdad se puede aplicar logaritmo.

$$y = a^x \quad entonces \quad \log y = \log a^x$$

Demostración

$$y = a^x \quad entonces \quad \log_y a^x = 1 \rightarrow \frac{\log a^x}{\log y} = 1 \rightarrow \log y = \log a^x$$

Ejercicios resueltos

Ejemplo-1

Hallar x en: $2^{2x+1}=500$

Solución

$$2^{2x+1}=500\rightarrow \log_2 500=2x+1\rightarrow \frac{\log 500}{\log 2}=2x+1\rightarrow 8.96578=2x+1\rightarrow x=3.98289$$

Ejemplo-2

Despejar x en:

$$y=\frac{10^x-10^{-x}}{2}$$

Solución

$$2y=10^x-10^{-x}$$

$$2y=10^x-\frac{1}{10^x}\rightarrow 2y.10^x=10^{2x}-1\rightarrow (10^x)^2-2y(10^x)-1=0$$

Hacemos que $m=10^x$

$$m^2-2ym-1=0$$

$$m=\frac{2y\pm\sqrt{4y^2-4(1)(-1)}}{2}=\frac{2y\pm\sqrt{4(y^2+1)}}{2}=y\pm\sqrt{y^2+1}$$

$$m=y\pm\sqrt{y^2+1}\rightarrow 10^x=y\pm\sqrt{y^2+1}\rightarrow x=\log(y\pm\sqrt{y^2+1})$$

Ejemplo-3

Linealizar el modelo no lineal

$$y=ab^x \quad ,$$

Para linealizar el modelo no lineal, se aplica logaritmos a ambos lados

$$Log\ y=Log\ ab^x$$
$$Log\ y=Log\ a+Log\ b^x$$
$$Log\ y=Log\ a+x(Log\ b)$$
$$Z=A+Bx\leftarrow \text{(modelo linealizado)}$$

Siendo:

$$Z=Log\ y \quad A=Log\ a \quad B=Log\ b$$

Ejercicios propuestos

1. Hallar $Log_{\sqrt[3]{2}} 64$

2. Hallar el valor de x en:

$$12^{2x+3} = 6^{x-1}$$

3. Linealizar el siguiente modelo no lineal

$$y = \frac{a}{x^b}$$

4. Calcular: $\log_2 2\sqrt[3]{16}$

5. Calcular el valor de x en cada caso:

$$3^{x+1} = 100$$

$$7^{3x-5} = 20^{x+1}$$

6. Despejar el x en cada caso:

$$a^{x+1} = b^{2x-1}$$

$$y = \frac{e^x - e^{-x}}{2}$$

7. Si $Log\,64 = a \quad Log\,81 = b$, hallar el logaritmo de 2 en base 3

8. Hallar los valores de $x \quad e \quad y$

$$\begin{cases} 3^{x+y} = 5 \\ 3^{x-y} = 7 \end{cases}$$

9. Calcular x en la ecuación: $a^{x+3} = b^{2x-12}$

10. Hallar los valores de $x \quad e \quad y$

$$\begin{cases} \log_2(x+y) = 2 \\ \log_3(x-y) = 3 \end{cases}$$

11. Resolver

$$\log_2(2x+1) + \log_3(2x+1) = 1$$

12. Resolver

$$\log_2 x + \log_4 x + \log_8 x = 11$$

13. Resolver

$$\log_{x+1} 2 + \log_{x+1} 8 = 2$$

14. Resolver

$$\log_2 x + \log_4 x + \log_8 x = 11$$

15. Resolver

$$3^{\log_9 (x-8)} = 2$$

16. Resolver

$$1 + \log(x-1) = \log x$$

17. Resolver

$$2\log_3 (x+1) + \log_9 (x-1) = 2$$

18. Resolver

$$7^{5x+3} = 3^{x+1} \qquad x = -0.54908492$$

2.1.3. Progresión geométrica (PG)

Es una sucesión de números, en la cual un término cualquiera es igual al anterior multiplicado por una constante, a esta constante se le llama razón geométrica (cociente entre dos cantidades). Una PG de n términos tiene la siguiente forma:

$$a, \quad ar, \quad ar^2, \quad ar^3, ..., ar^{n-1}$$

Siendo

$a =$ Primer término de la progresión geométrica

$r =$ Razón de la progresión geométrica

$ar^{n-1} =$ Último término de la progresión geométrica

Nota

Termino del lugar k de una PG

$T_k = ar^{k-1}$

Suma de los términos

$$S = \frac{a(1-r^n)}{1-r}$$

Productos de los términos de una PG

$$P = a^n.r^{\frac{n(n-1)}{2}}$$

Convergencia de una PG cuando la razón $\{-1 < r < 1\} - \{0\}$. Cuando la suma de sus términos tiende al infinito.

$$\lim_{n\to\infty} S_n = \frac{a}{1-r}$$

Ejemplo-1

Sea la progresión geométrica: $3, \quad 3/4, \quad 3/16, \quad 3/48$

Calcular: la razón, suma, producto y la suma límite de sus términos.

Razón $\rightarrow r = \dfrac{3/4}{3} = \dfrac{3/16}{3/4} = \dfrac{3/48}{3/16} = \dfrac{1}{4}$

Suma $\rightarrow S = \dfrac{3(1-(1/4)^4)}{1-\dfrac{1}{4}} = 3.984375$

Producto $\rightarrow P = (3)^4(\dfrac{1}{4})^{\frac{4(4-1)}{2}} = 0.01977539$

Suma límite de los infinitos términos $\rightarrow S = \dfrac{3}{1-\dfrac{1}{4}} = 4$

Ejemplo-2

Sea la sucesión de términos, supongamos que es una PG: $5, \quad 11, \quad 20, \quad 41$

Calcular la razón geométrica estimada

$$\rightarrow r = \frac{\frac{11}{5} + \frac{20}{11} + \frac{41}{20}}{3} = 2.023$$

2.1.4. Gráfica de funciones

Sea la función matemática de la línea recta

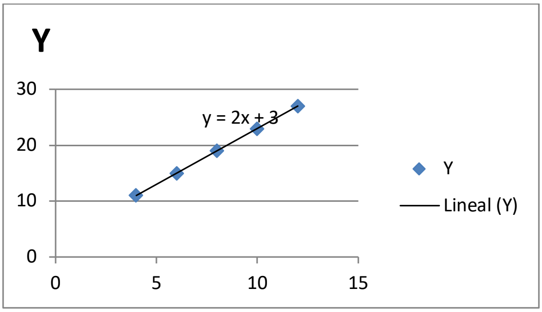
$y = 3 + 2x$

X=variable independiente e y=variable dependiente, el valor que toma” y” depende del valor que se asigna a “x”

La función proporciona pares ordenados (X, Y), para obtener estos pares ordenados , damos valores arbitrario a la variable X y calculamos los correspondiente valores para Y.

x	y
4	11
6	15
8	19
10	23
12	27

Los pares ordenados (X, Y), lo presentamos en un plano cartesiano (nube de puntos) y tiene por gráfica



Ejercicios propuestos

Graficar las siguientes funciones:

a. $y = \dfrac{400}{x^{2.5}}$

b. $y = 1.56\log(\dfrac{2.5}{x})$

c. $y = 1.25(2^x)$

d. $y = \dfrac{20}{2+3x}$

e. $y = x^3 + x + 1$

2.2. Conceptos estadísticos

2.2.1. Media geométrica $(\bar{x}_g)$

Se utiliza cuando los datos por su naturaleza, presentan un comportamiento de una serie en progresión geométrica o los datos provienen de una variación porcentual.

La fórmula:

Sea un conjunto de n datos positivos, tal que $x_i > 0$

$$x_1, x_2, x_3, \ldots, x_n$$

Su valor corresponde a la raíz $n-$enésima del producto de los n datos

$$\bar{x}_g = \sqrt[n]{x_1.x_2.x_3....x_n}$$

También se puede expresar por:

$$Log(\bar{x}_g) = Log(\sqrt[n]{x_1.x_2.x_3....x_n}) = \frac{1}{n}\sum_{i=1}^{n} Log(x_i) \rightarrow$$

$$\bar{x}_g = Antig(\frac{1}{n}\sum_{i=1}^{n} Log(x_i)$$

Ejemplo-1.

Calcular la media geométrica para los siguientes datos:

4, 9, 15, 32

$$\bar{x}_g = \sqrt[4]{(4)(9)(15)(32)} = \sqrt[4]{17280} = 11.465$$

La media geométrica para los datos es de 11.5

También, por el otro método

$$Log(\bar{x}_g) = \frac{\log(4) + \log(9) + \log(15) + \log(25)}{4} = 1.059385935$$

$$\bar{x}_g = 10^{1.059385935} = 11.465$$

Ejemplo-2.

La información de la tabla corresponde a la producción de artículos científicos en medicina entre los años 1990 al 2020 de cierto país. Calcular la producción promedio anual y el incremento anual promedio.

Años	Producción de revistas científicas en medicina	Variación=se divide la producción de un año cualquiera por el año anterior
1990	120	
1995	150	1.25
2000	187	1.246667
2005	234	1.251337
2010	293	1.252137
2015	366	1.249147
2020	458	1.251366

Calcular la producción promedio de revistas científicas

$$\overline{x}_g = \sqrt[7]{(120)(150)...(458)} = 234.23 \quad \textit{articulos}$$

Calcular la variación promedio de la producción de revistas científicas

$$\overline{x}_g = \sqrt[6]{(1.25)(1.246667)...(1.251366)} = 1.25$$

La variación promedio es de 1.25 (125%) y su incremento promedio cada 5 años es de 1.25-1=0.25 en porcentaje=25%.

2.2.2. Media armónica $(\overline{x}_a)$

Se utiliza cuando con cierta cantidad fija, se generan otras cantidades con dos unidades: km/hora, monto/articulo, horas/artículos, etc. y se pide calcular el promedio es esta última.

La fórmula es:

Sea un conjunto de n datos positivos (no cero)

$$x_1, x_2, x_3, ..., x_n$$

$$\overline{x}_a = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ... + \frac{1}{x_n}}$$

Ejemplo:

Se tiene 80 nuevos soles y se desea comprar cierto número de revistas científicas de la misma materia y en tres editoriales, tal como se presenta en la tabla. Calcular el precio promedio de las revistas.

Revistas científicas	Cantidad	Precio unitario por revistas=monto/cantidad
A	5	S/.16=x1
B	8	S/.10=x2
C	10	S/.8=x3

$$\bar{x}_a = \frac{3}{\frac{1}{16}+\frac{1}{10}+\frac{1}{8}} = S/.10.43$$

S/.10.43, es el precio promedio de las revistas.

2.2.3. Hipótesis estadística

Es un supuesto que se hace en la poblacion en estudio, con respecto a la distribución de una variable aleatoria o a sus parámetros, la validez de este supuesto, será confrontado con una muestra elegida al azar de la poblacion, un estadístico de prueba y un nivel de significación, que elige el investigador para rechazar o no la hipótesis que se supone cierta en la poblacion (hipótesis nula). Se supone que este supuesto lo hace el investigador con la experiencia en el campo de su competencia. Una hipótesis se representa con la letra mayúscula H.

Ejemplos

1. La distribución del número de artículos por revistas científica en matemática sigue una distribución normal.

H :La variable número de artículos por revista científica sigue una distribución normal.

2. El promedio del número de artículos por revista científica, en matemática es 16 artículos.

$$H : \mu = 16 \quad \text{Artículos}$$

3. La variabilidad del número de artículos por revistas con respecto a la media en la poblacion es de 23.5 artículos^2.

$$H : \sigma^2 = 23.5 \quad articulos^2$$

La Hipótesis estadística puede ser nula (llamada también planteada) simbolizada por $H_p \quad o \quad H_0$ y la hipótesis alternante simbolizada por $H_a \quad o \quad H_1$.

<u>Hipótesis Nula o Planteada</u> (H_0)

Es la hipótesis estadística, sobre la cual el investigador está dispuesto a creer a priori como verdadera, y cuya validez será sometida a comprobación a través de una muestra elegida al azar de la poblacion en estudio. La hipótesis nula se formula con el objetivo de ser rechazada, pero con una probabilidad bien pequeña (0.1, 0.05, 0.01), en caso de obtener una muestra desafortunada (es decir con valores muy altos o valores muy bajos).

Hipótesis Alternante (H_a .)

Es la contraparte de la hipótesis nula y representa la creencia o sospecha que el investigador quiere probar.

Ejemplo

Un bibliotecólogo afirma que el número promedio de artículos por revistas científicas es 14 artículos, esta afirmación corresponde a la hipótesis nula; la hipótesis alternante podría ser diferente a 14 artículos, mayor a 14 artículos o menor a 14 artículos:

$$H_0 : \mu = 14 \quad articulos$$
$$H_a : \mu \neq 14, \quad H_a : \mu > 14, \quad H_a : \mu < 14$$

TIPOS DE ERRORES

La decisión de rechazar o no rechazar la hipótesis nula, con base a la información de la muestra aleatoria extraída de la población en estudio, pude conllevar a decisiones erradas debido a las fluctuaciones del muestreo, y cometer los siguientes errores.

ERROR TIPO I

Consiste en rechazar la hipótesis nula H_0 a consecuencia de los datos de la muestra cuando esta es cierta en la población.

ERROR TIPO II

Consiste en no rechazar la hipótesis nula H_0 a consecuencia de los datos de la muestra cuando esta es falsa en la población.

Para estos dos tipos de errores se presenta el siguiente cuadro

CUADRO

DECISION TOMADA DEBIDO A LA MUESTRA	POBLACION EN ESTUDIO	
	H_0 ES CIERTA	H_0 ES FALSA
H_0 SE ACEPTA	DECISION CORRECTA	ERROR TIPO II
H_0 SE RECHAZA	ERROR TIPO I	DECISION CORRECTA

Nivel de significación de la prueba α

Es probabilidad de cometer el error tipio I, se denota con el símbolo α, es una medida de la confiabilidad de la decisión de rechazar la hipótesis nula. Las probabilidades utilizadas para rechazar la hipótesis nula H_0 para $\alpha = 1\%, \quad 5\%, \quad 10\%$

$$\alpha = P(cometer\ el\ error\ tipo\ I) = P(rechazar\ H_0 / H_0\ es\ cierta)$$

Nivel de significación β

Es la probabilidad de cometer el error tipio II, se le denota con el símbolo β

$$\beta = P(cometer\ el\ error\ tipo\ II) = P(aceptar\ H_0 / H_0\ es\ falsa)$$

Potencia de la prueba $(1-\beta)$

Es la probabilidad de aceptar la hipótesis alternante dado que esta es verdad

$$1-\beta = P(acepta\ H_0 / H_0\ es\ cierta)$$

Supuestos para una prueba estadística

Para realizar una prueba de hipótesis para determinados parámetros de una poblacion en estudio se requiere el cumplimiento de un conjunto de supuestos, para dar validez a la decisión que se toma:

a. Los datos de la muestra son elegidos al azar de la poblacion en estudio.
b. La variable de donde proviene el parámetro tiene una distribución normal.
c. Existe pruebas de ciertos parámetros, que requiere que el tamaño de la muestra debe ser grande.

Prueba de hipótesis

Para evaluar una prueba de hipótesis sobre determinados parámetros, se utiliza la distribución muestral de la variable aleatoria de donde proviene el parámetro, datos de una muestra elegida al azar de tamaño n y el nivel de significación α para rechazar la hipótesis nula H_0.

Del ejemplo anterior, supongamos la siguiente hipótesis nula y alternante:

$$H_p : \mu = 14 \quad articulos$$
$$H_p : \mu \neq 14 \quad articulos$$

Suponiendo que el parámetro a evaluar proviene de una distribución normal con variancia igual a 12.25 articulos^2 y el nivel de la significación $\alpha = 0.05$, bajo la hipótesis nula cierta. Planteamos:

$$P(rechazar \quad H_0 / H_0 verdadera \quad en \quad la \quad poblacion) = \alpha$$

$$P(\bar{x} < a \quad \cup \quad \bar{x} > b) = 0.05$$

$$P\left(\frac{\bar{x}-14}{\sqrt{12.25}/\sqrt{n}} < \frac{a-14}{\sqrt{12.25}/\sqrt{n}}\right) \quad \cup \quad \frac{\bar{x}-14}{\sqrt{12.25}/\sqrt{n}} > \frac{b-14}{\sqrt{12.25}/\sqrt{n}}) = 0.05$$

$$P(Z < \frac{a-14}{\sqrt{12.25}/\sqrt{n}}) \quad \cup \quad Z > \frac{b-14}{\sqrt{12.25}/\sqrt{n}}) = 0.05$$

$$\frac{a-14}{\sqrt{12.25}/\sqrt{n}} = -1.96 \qquad \frac{a-14}{\sqrt{12.25}/\sqrt{n}} = 1.96$$

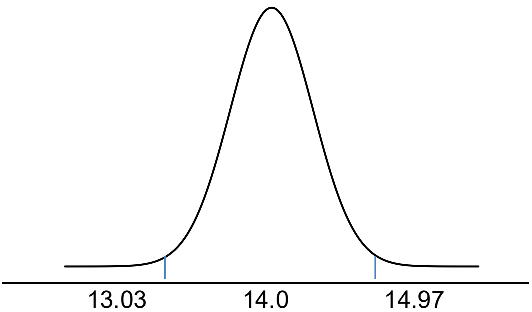
$$a = 14 - 1.96\frac{\sqrt{12.25}}{\sqrt{n}} \qquad b = 14 + 1.96\frac{\sqrt{12.25}}{\sqrt{n}}$$

Para un tamaño de muestra n=50. Se tiene:

$$a = 14 - 1.96\frac{\sqrt{12.25}}{\sqrt{50}} \qquad b = 14 + 1.96\frac{\sqrt{12.25}}{\sqrt{50}}$$

$$a = 13.03 \qquad b = 14.97$$

A los valores 13.03 y 14.97 se le conoce como valores críticos, si la muestra elegida al azar de tamaño 50 arroja que la media muéstral es de 12 o 16, entonces se rechaza la hipótesis nula con un nivel de significación del 5%, y se concluye que hay suficiente evidencia muestral para rechazar la hipótesis nula con una significación del 5%. Si el valor de la media muéstral se encuentra entre los dos valores críticos, entonces no se rechaza la hipótesis nula con un nivel de significación del 5%. Gráficamente se tiene:



Las áreas a la izquierda y derecha de los puntos 13.03 y 14.97, se le llama zona de rechazo de la hipótesis nula y la zona entre los puntos 13.03 y 14.97 se le llama zona de no rechazo de la hipótesis nula.

<u>Región critica o de rechazo de</u> H_0

Es la región que contienen los valores muestrales para lo cual se rechaza la hipótesis nula $(\bar{x} \leq 13.03 \quad o \quad \bar{x} \geq 14.97)$

Región aceptación de H_p

Es la región que contienen los valores muestrales para lo cual no se rechaza la hipótesis nula $(13.03 < \bar{x} < 14.97)$

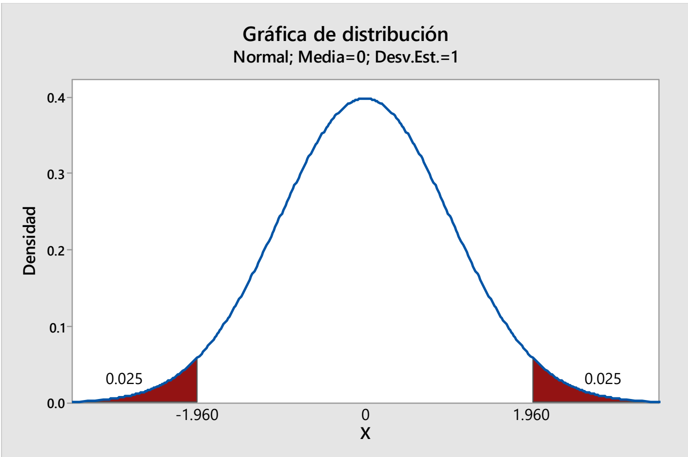
En lo sucesivo el procedimiento anterior se puede simplifican estandarizando el valor obtenido en la muestra y comparando con el valor tabular obtenido del nivel de significación.

Resumiendo: La prueba es la siguiente

$$H_p : \mu = 14 \quad articulos$$
$$H_p : \mu \neq 14 \quad articulos$$

Supongamos que para un tamaño de muestra $n = 50$, el valor de la media muestral es $\bar{x}$=15.2 y la variancia de la población se conoce σ^2=12.25 y el nivel de significación para la prueba es α=0.05. Entonces se estandariza el valor obtenido para la muestra que es un valor Z, y se compara con la región crítica determinada por el nivel de significación de la distribución Z. si el valor estandarizado cae en la zona de rechazo, entonces se rechaza la hipótesis nula de lo contrario no se rechaza la hipótesis nula. A la fórmula que me permite estandarizar el valor muestral y tomar la decisión de rechazar o no rechazar la hipótesis nula, se le llama estadístico de prueba.

$$Z_c = \frac{15.2 - 14}{\frac{\sqrt{12.25}}{\sqrt{50}}} = 2.424366$$



El valor Zc = 2.424366, cae en la zona de rechazo la hipotesis nula, luego se

rechaza la hipotesis nula.

Nota:
Otra forma de rechazar o no rechazar la hipótesis nula H_0 , es comparar el nivel de significación α establecido para la prueba, con la probabilidad de rechazar la hipótesis nula proporcionado por el estadístico de prueba calculado, valor de probabilidad que se conoce con el nombre del p valor. Del ejemplo anterior.

$$\alpha \geq p \rightarrow se \quad rechaza \quad la \quad H_0$$
$$\alpha < p \rightarrow se \quad no \quad se \quad rechaza \quad la \quad H_0$$

La hipótesis nula H_0 se rechaza si:

$$p(z_c < -\frac{15.2-14}{\frac{\sqrt{12.25}}{\sqrt{50}}} \cup z_c > \frac{15.2-14}{\frac{\sqrt{12.25}}{\sqrt{50}}}) = P(z_c < -\frac{15.2-14}{\frac{\sqrt{12.25}}{\sqrt{50}}}) + P(z_c > \frac{15.2-14}{\frac{\sqrt{12.25}}{\sqrt{50}}})$$
$$P(z_c < -2.424366) + P(z_c > 2.424366) = 0.007668 + 0.007668 = 0.015336 = 0.015$$
$$como \quad \alpha = 0.05 > p = 0.015 \rightarrow se \quad rechaza \quad la \quad H_0$$

Pasos a seguir para probar una hipótesis estadística

a. Plantear la hipótesis nula y alternante en términos del enunciado del problema
b. Elegir el nivel de significación para la prueba.
c. Región critica (Región de rechazo de la hipótesis planteada Ho)
d. Cálculos del estadístico de prueba, usando los datos de la muestra y bajo el supuesto de la hipótesis nula cierta.
e. Decisión y conclusión.

Pruebas de hipótesis para la media de una población

Sea una población en estudio, en la cual se define una variable aleatoria de interés X, supongamos que la variable tiene una distribución normal con media μ y variancia σ^2, de dicha población se toma una muestra aleatoria $x_1, x_2, \ldots x_n$, obteniéndose la media muestral $\overline{x}$. Se demuestra que los estadísticos:

$$z_c = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim Z(0\,;1) \quad .$$

Se usará cuando la variancia de la población se conoce.

$$t_c = \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim t_{(n-1)gl}$$

Se usará cuando la variancia de la población no se conoce y el tamaño de la muestra es menor a 30, para un tamaño de muestra mayor a 30, se puede usar la distribución Z.

Formulación de las hipótesis:

Pruebas bilaterales	*Pruebas unilaterales*	
$H_p : \mu = \mu_o$ $H_p : \mu \neq \mu_o$	$H_p : \mu \geq \mu_o$ $H_a : \mu < \mu_o$	$H_a : \mu \geq \mu_o$ $H_a : \mu < \mu_o$

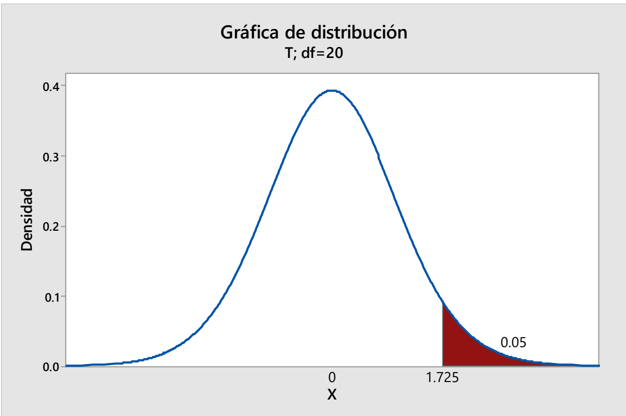
.
Ejemplo-1

Probar con un nivel de significación del 5% que la media del número de artículos de cierta revista científica es superior a 15 artículos. Una muestra de 21 revistas elegidas al azar de la poblacion, arroja que la media muestral es de 17 artículos con una varianza de 46.24 artículos^2.

$$H_0 : \mu \leq 15$$
$$H_a : \mu > 15$$

Nivel de significación para la prueba $\alpha = 0.05 = P(rechazar \quad H_0 / H_0, cierta)$

Región critica para la prueba, como la prueba de hipótesis es unilateral a la derecha (por la hipótesis alternante) la región critica o de rechazo de la hipótesis planteada es:



Estadístico de prueba: como la variancia de la población no se conoce y el tamaño de la muestra es menor a 30, el estadístico de prueba para este caso es:

$$T_c = \frac{\bar{x} - \mu_o}{s / \sqrt{n}} \sim t(n-1)gl / H_p \quad Cierta$$

$$T_c = \frac{17 - 15}{6.8 / \sqrt{21}} = 1.348$$

$T_c = 1.348 < T_{(0.05,20)gl} = 1.725$, entonces no se rechaza la hipótesis nula, por lo tanto no existe la suficiente evidencia estadística para rechazar la hipótesis nula de que la media poblacional del número de artículos es menor a 15 artículos.

Ejemplo-2

De la pregunta anterior, para que valores del estimador $\bar{x}$ se rechaza la hipótesis nula.

Se rechaza la hipótesis planteada siempre y cuando:

$$T_c > 1.725 \rightarrow \frac{\bar{x} - 15}{s / \sqrt{n}} > 1.725 \rightarrow \bar{x} > 1.725 . \frac{s}{\sqrt{n}} + 15 = 17.56$$

Se rechaza la hipótesis nula para valores de la media muestral superior a 17.56.

<u>Prueba de hipótesis para una proporción poblacional</u>

Sea la proporción muestral $\hat{p} = \frac{x}{n}$, siendo "x" el número de elementos con la cualitativa de interés en la muestra y "n" el tamaño de la muestra, si $n \rightarrow \infty \quad (n > 30), \quad entonces \ Z_C = \frac{\hat{p} - P}{\sigma_{\hat{p}}} \sim N(0\,;1)$, corresponde al estadístico de prueba.

Siendo:

$$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$$ Si el muestreo es con reemplazo

$$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$$ Si el muestreo es sin reemplazo

Ejemplo-1:

Los datos que a continuación se presenta corresponden a una muestra aleatoria de la cantidad de artículos publicados por cierta revistas científica en ciencias matemáticas.

13	15	13	12	20	23	10	12
19	24	23	12	14	15	17	13
18	12	13	14	12	17	18	12
14	19	16	25	25	11	13	14
17	21	18	23	15	16	15	15

¿Probar con un nivel de significación del 5% que la proporción de revistas con más de 15 artículos publicados es superior al 42%?

Hipótesis a probarse

$H_p : P \leq 0.42$

$H_A : P > 0.42$

1. Nivel de significación de la prueba

$\alpha = 0.05 = P(rechazar \quad la \quad H_p / H_p, \quad cierta, \quad en \quad la \quad población)$

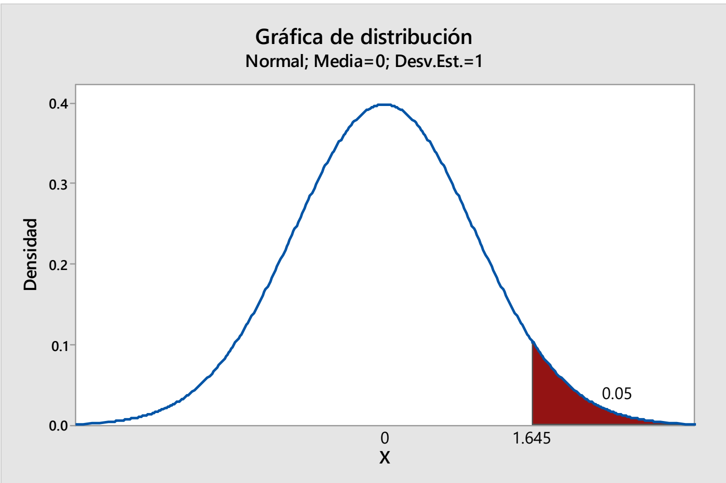
2. Estadístico de prueba

$$Z = \frac{\hat{p} - P}{\sigma_{\hat{p}_A}} \sim N(0\,,1)$$

3. Supuestos:

   a. Las observaciones son elegidas aleatoriamente
   b. El tamaño de muestra es grande

4. Criterio de decisión



Se acepta H_p , si $Z_c \leq 1.645$, en caso contrario se rechaza.

Cálculos, de los datos se tiene:

Tamaño de la muestra	Numero de revistas con más de 15 artículos
40	18

Con $n_a = 40 \quad p_A = \frac{18}{40} = 0.45$, se tiene que la

$$\sigma_{p_A} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.42(1-0.42)}{40}} = 0.078038$$

$$Z_c = \frac{0.45-0.42}{0.078038} = 0.3844$$

Conclusión

Como $Z_c = 0.3844 < 1.645$; luego no se rechaza la hipótesis planteada y se concluye, para $\alpha = 0.05$ evidencias muéstrales indican que la proporción de revistas con más de 15 artículos no es superior a 0.42.

Como para todas estas pruebas se utiliza es estadístico Z, se presenta la siguiente tabla para valores clásicos del nivel de significación:

Hipótesis alternante	Nivel de significación de la prueba		
	0.01 (1%)	0.05 (5%)	0.1 (10%)
2 colas	-2.576 y 2.576	-1.96 y 1.96	-1.645 y 1.645
1 cola a la derecha	2.326	1.645	1.282
1 cola a la izquierda	-2.326	-1.645	-1.282

Ejercicios propuestos:

a. Probar con un nivel de significación del 5% que la edad promedio de los usuarios que concurren a un centro de información es diferente a los 27 años. Una muestra de 50 usuarios elegidos al azar, proporciona la siguiente información:

b. $$\sum_{i=1}^{50} x_i = 1624 \qquad \sum_{i=1}^{50} x_i^2 = 55230$$

c. Probar con un nivel de significación del 5% que el número promedio de artículos por revistas en computación es superior a 8 artículos. Una muestra de 15 revistas elegidas al azar de una población con distribución normal en cuanto al número de artículos proporciona los siguientes datos:

10 8 9 15 17 7 12 14 16 9
8 10 15 12 7

d. Probar con un nivel de significación del 5% que el número promedio de artículos por revistas en computación es diferente a 18 artículos. Una muestra de 50 revistas elegidas al azar proporciona la siguiente información.

$$\bar{x} = 21 \qquad s^2 = 10$$

e. Probar con un nivel de significación del 5% que la proporción de revista científicas en idioma inglés es diferente de 0.24. Una muestra de 100 revistas elegidas al azar proporciona la siguiente información de que 28 están en idioma inglés.

f. Probar con un nivel de significación del 5% que el número promedio de personas que ingresan a un centro de información durante el día de atención es diferente a 630, se eligen al azar una muestra de 50 días y se obtiene por media aritmética 590 personas y desviación estándar 236 personas.

g. Probar con un nivel de significación del 5% que la proporción de revista científicas en idioma inglés es superior a 0.35. Una muestra de 200 revistas elegidas al azar proporciona la siguiente información de que 80 están en idioma inglés.

h. En una biblioteca desconocemos la proporción de libros escritos en español. De una muestra aleatoria de 125 libros, 80 de ellos están escritos en español, y el resto en otros idiomas. Según estos datos, ¿se puede afirmar que la proporción de libros escritos en español en dicha biblioteca es mayor a 0.6?. Use un nivel de significación de 0.05.

i. Probar con un nivel de significación del 5% que la proporción de revistas científicas en idioma inglés en la poblacion es superior a 0.38. De una muestra elegida al azar de 150 revistas, se encuentra que 60 están en idioma inglés. Calcular el valor calculado y la decisión tomada.

j. Probar con un nivel de significación del 5% que la proporción de revistas científica en física escrita en idioma inglés es superior a 0.45. Una muestra de tamaño 300 elegida al azar de la poblacion de revistas científica en física, arroja que 165 son revistas en idioma inglés.

2.2.4. Medidas de asociación.

Coeficiente de correlación lineal simple de Pearson

El coeficiente de correlación lineal simple mide la fuerza o grado de asociación lineal entre dos variables cuantitativas $X \; e \; Y$, que poseen una distribución normal bivariada. Su valor varía entre $[-1, \;\; 1]$, un valor próximo a 1 o -1 indica una alta asociación lineal entre la variable, el valor positivo indica una asociación lineal directa y un valor negativo indica una asociación lineal inversa, un valor cercano a cero indica que las variables no están asociadas linealmente.
Supuestos

- Las variables $X \; e \; Y$ son aleatorias
- Las variables $X \; e \; Y$ tienen distribuciones normales bivariada
- Existe una relación lineal entre las variables $X \; e \; Y$

Para calcular este coeficiente, se toma una muestra aleatoria de tamaño "n" de datos bivariados (x_i, y_i), correspondiente a las variables $X \; e \; Y$; luego un estimador para el parámetro ρ llamado, el coeficiente de correlación lineal poblacional es dado por:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}} \qquad (1)$$

El valor que toma este coeficiente varía en el intervalo $-1 \leq r \leq +1$

1. $r = +1$ nos indica que existe una asociación o correlación lineal perfecta positiva; es decir si "X" aumenta "Y" también aumenta; en otras palabras, todos los puntos están sobre una línea recta de pendiente positiva.
2. $r = -1$, nos indica que existe una asociación o correlación lineal perfecta negativa; es decir si "X" aumenta "Y" disminuye y viceversa; en otras palabras, todos los puntos están sobre una línea recta de pendiente negativa.
3. $r = 0$, nos indica que no existe relación lineal entre las variables $X \; e \; Y$; esto no implica que las variables $X \; e \; Y$ sean independientes.

Ejemplo

La información de la tabla corresponde a la producción de artículos científicos sobre enfermedades del corazón, durante los años 1990 al 2020

X=Años	Y=Producción de artículos científicos
1990	22
1995	30
2000	42
2005	55
2010	62
2015	70
2020	82

¿La producción de artículos científico está asociado linealmente al tiempo?

Para el cálculo del coeficiente de correlación lineal, se tiene las siguientes sumatorias que aparecen en negrita en la siguiente tabla.

X	Y	XY	X^2	Y^2
1990	22	43780	3960100	484
1995	30	59850	3980025	900
2000	42	84000	4000000	1764
2005	55	110275	4020025	3025
2010	62	124620	4040100	3844
2015	70	141050	4060225	4900
2020	82	165640	4080400	6724
14035	**363**	**729215**	**28140875**	**21641**

Aplicando la formula (1)

$$r=\frac{729215-\frac{(14035)(363)}{7}}{\sqrt{28140875-\frac{(14035)^2}{7}}\sqrt{21641-\frac{(363)^2}{7}}}=\frac{1400}{\sqrt{700}\sqrt{2816.857143}}=0.99700332$$

Interpretación: 0.997, este valor indica que los años y la producción de artículos científicos sobre la enfermedad del corazón están altamente asociados en forma directa.

Prueba de hipótesis para el coeficiente de correlación lineal

El valor del coeficiente de correlación lineal, por sí solo no es suficiente para evaluar el grado de asociación existente entre dichas variables. Este valor es resultado de una muestra elegida al azar de una poblacion, para dar validez a este coeficiente es necesario establecer una hipótesis con respecto a su parámetro y poder juzgar la asociación poblacional de las variables X e Y con la evidencia muestral y un nivel de significación α.

1. Cuando el valor hipotético del coeficiente de correlación población (parámetro) es cero $\rho=0$

   Cuando se desea verificar si el coeficiente de correlación estimado es o no nulo, lo cual sugiere la existencia o no de una asociación significativa entre las variables en estudio X e Y, se tienen las siguientes hipótesis:

   $$H_p:\rho=0\rightarrow(No\ hay\ correlacion\ lineal\ entre\ X\ e\ Y)$$
   $$H_a:\rho\neq 0\rightarrow(\ hay\ correlacion\ lineal\ entre\ X\ e\ Y)$$

   Nivel de significación $\alpha=P(rechazar\ H_p/H_p, es\ cierta)$

   El estadístico de prueba es:

   $$t_c=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\approx T_{(n-2)gl}$$

   Se rechaza la hipótesis nula, para una prueba de dos colas.
   Si $t_c<T_{(\frac{\alpha}{2},n-2)}$ o $t_c>T_{(\frac{\alpha}{2},n-2)}$

NOTA

Si no se rechaza Ho entonces nos indica que las evidencias muéstrales expresan que no existe una correlación lineal significativa entre las variables $X \; e \; Y$. Pero debe tenerse presente que el no rechazo de Ho no implica que las variables X e Y sean independientes.

También al rechazar la Ho, indicará que las evidencias muéstrales expresan que existe una correlación lineal significativa entre las variables $X \; e \; Y$. Debe tenerse presente que el rechazo de Ho no implica que la relación funcional entre las variables en estudio sea necesariamente una función lineal. En estos casos es recomendable obtener una grafico de la dispersión de los datos, para estudiar el tipo de relación funcional entre las variables.

Ejemplo

Del ejemplo anterior, probar con un nivel de significación del 5% que el coeficiente de correlación lineal es diferente de cero

Probar si existe una correlación lineal significativa entre las variables $X \; e \; Y$.
Use $\alpha = 0.05$

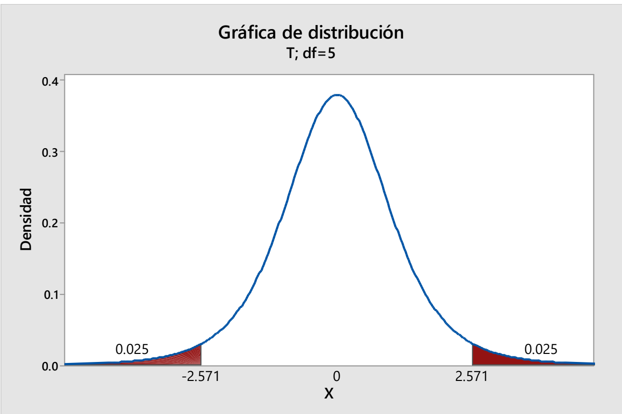
$$H_0 : \rho = 0 \rightarrow (No\ hay\ correlacion\ lineal\ entre\ X\ e\ Y)$$
$$H_a : \rho \neq 0 \rightarrow (\ hay\ correlacion\ lineal\ entre\ X\ e\ Y)$$
$$\alpha = 0.05$$

Usando la formula dada en (1) se tiene que $r = 0.997$

Calculo del estadístico de prueba

$$t_c = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.997 - 0}{\sqrt{\frac{1 - 0.997^2}{7 - 2}}} = 28.80$$

Región critica para la prueba: como la prueba es bilateral se tiene, se rechaza la hipótesis planteada si:



Como $t_c = 28.80 > 2.776$, entonces se rechaza la Ho. Las variables $X \; e \; Y$ están significativamente asociadas al nivel 5%.

Ejercicio propuesto

La siguiente tabla tiene información sobre el número de artículos publicados en matemáticas y el número de citas recibidas durante los años 1994 a 2000.

Años / Descripción	1994	1995	1996	1997	1998	1999	2000
N° artículos publicados	20	25	30	40	47	56	70
N° citas recibidas	4	9	20	14	21	23	28

1. Hallar el coeficiente de correlación.
2. Probar si existe una correlación significativa entre las variables X e Y. Use $\alpha = 0.05$.

Coeficiente de correlación de rangos de Spearman

Su valor mide la asociación entre dos variables $X \quad e \quad Y$, se usa cuando por lo menos unas de las variables son ordinal, y no requiere que se cumpla el supuesto de normalidad bivariada, tiene por formula:

$$r_s = 1 - \frac{6.\sum_{i=1}^{n}(Rx_i - Ry_i)^2}{n(n^2-1)} \qquad (1)$$

$(Rx_i - Ry_i) =$ Diferencias de los rangos de las variables $X \quad e \quad Y$

Para una muestra de tamaño n, se tiene (x_i, y_i), los valores que toman las variables $X \quad e \quad Y$, se consideran los órdenes se sus valores en forman ascendente o descendente, a estos ordenes se le llama rangos, en caso de que existan empates entre sus órdenes, se toma en cuenta su valor promedio.

.

La interpretación que del valor de r_s, es la misma que la del coeficiente de correlación de Pearson.

La fórmula anterior se usa cuando no hay empates en los rangos (ordenes), de lo contrario se usa la fórmula más general.

$$r_s = \frac{\sum_{i=1}^{n} R_{x_i}^2 + \sum_{i=1}^{n} R_{y_i}^2 - \sum_{i=1}^{n}(R_{x_i} - R_{y_i})^2}{2.\sqrt{R_{x_i}^2}.\sqrt{R_{y_i}^2}} \qquad (2)$$

Donde

$$\sum_{i=1}^{n} R_{x_i}^2 = \frac{n(n^2-1)}{12} - M_x \qquad M_x = \frac{\sum_{i=1}^{g} E_i(E_i^2-1)}{12}$$

$$\sum_{i=1}^{n} R_{y_i}^2 = \frac{n(n^2-1)}{12} - M_y \qquad M_y = \frac{\sum_{i=1}^{g} E_i(E_i^2-1)}{12}$$

$M_x =$ Es el término de corrección por empates en los rangos de x
$M_y =$ Es el termino de corrección por empates en los rangos de y

Supongamos que no hay empates en ambas variables la formula (2) se convierte en (1)

<u>Ejemplo</u>

Dos jurados evalúan a 12 candidatos para un puesto de trabajo, la calificación es de 0 a 20. Los datos se presentan en la tabla.

Postulante	X=Jurado A	Y=Jurado B	Rx_i	Ry_i	$(Rx_i - Ry_i)$	$(Rx_i - Ry_i)^2$
1	12(1)	14(1)	1.5	2.5	-1	1
2	14(4)	14(2)	4	2.5	1.5	2.25
3	15(6)	14(3)	6	2.5	3.5	12.25
4	13(3)	15(5)	3	6	-3	9
5	17(9)	16(8)	9.5	8.5	1	1
6	17(10)	18(12)	9.5	12	-2.5	6.25
7	16(8)	15(6)	8	6	2	4
8	15(5)	17(10)	6	10.5	-4.5	20.25
9	19(12)	17(11)	12	10.5	1.5	2.25
10	18(11)	15(7)	11	6	5	25
11	15(7)	16(9)	6	8.5	-2.5	6.25
12	12(2)	14(4)	1.5	2.5	-1	1
					total	90.5

Aplicando la formula (2), calculamos los términos de corrección para x e y
Para el rango de x: existe un grupo con 2 empates (g=1); otro grupo con 2 empates (g=2) y un tercer grupo con 3 empates (g=3).
Para el rango de y: existe un grupo con 4 empates (g=1); otro grupo con 3 empates (g=2), un tercer grupo con 2 empates (g=3) y un cuarto grupo con 2 empates (g=4).

$$M_x = \frac{\sum_{I=1}^{3} E_i(E_i^2-1)}{12} = \frac{2(2^2-1)+2(2^2-1)+3(3^2-1)}{12} = \frac{36}{12} = 3$$

$$M_x = \frac{\sum_{I=1}^{4} E_i(E_i^2-1)}{12} = \frac{4(4^2-1)+3(3^2-1)+2(2^2-1)+2(2^2-1)}{12} = \frac{96}{12} = 8$$

$$\sum_{i=1}^{12} R_{x_i}^2 = \frac{12(12^2-1)}{12} - 3 = 140$$
$$\sum_{i=1}^{n} R_{y_i}^2 = \frac{12(12^2-1)}{12} - 8 = 135$$

Existe una baja asociación entre ambas calificaciones de los jurados

$$r_s = \frac{140+135-90.5}{2.\sqrt{140}.\sqrt{135}} = 0.671020$$

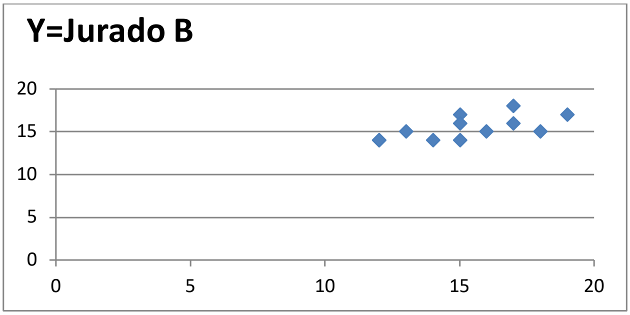
Entre las variables X e Y existe una baja correlación o asociación.

Usando la formula (1)

$$r_s = 1 - \frac{6.\sum_{i=1}^{n}(Rx_i - Ry_i)^2}{n(n^2-1)} = 1 - \frac{6(90.5)}{12(12^2-1)} = 0.683566 \qquad (1)$$

Entre los valores que toma la variable X e Y existe una baja correlación o asociación entre las opiniones de los jurados..
La correlación de Pearson entre los rangos es igual a 0.671020.

Graficas de las variables X e Y

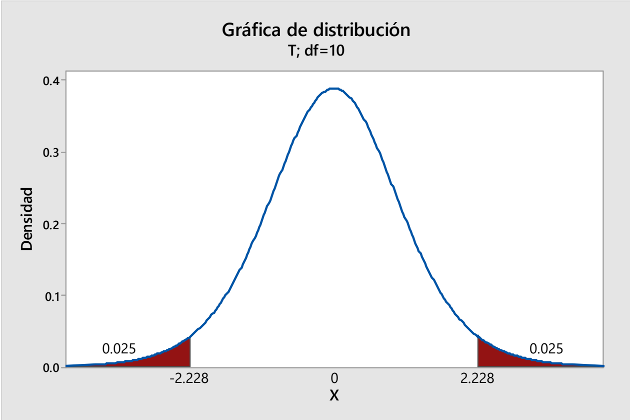


<u>Prueba de hipótesis</u>

$$H_0 : \rho = 0 \rightarrow (No\ hay\ correlacion\ entre\ X\ e\ Y)$$
$$H_a : \rho \neq 0 \rightarrow (\ hay\ correlacion\ entre\ X\ e\ Y)$$

Nivel de significación $\alpha = P(rechazar\ H_0 / H_0, es\ cierta)$

Se rechaza la hipótesis nula si el valor tabular cumple cualquiera de las siguientes condiciones:



El estadístico de prueba es:

$$t_c = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \approx T_{(n-2)gl}$$

$$t_c = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0.671020\sqrt{12-2}}{\sqrt{1-0.671020^2}} = 2.862$$

Conclusión se rechaza la hipótesis nula, existe una correlación significativa directa entre las variables.

2.2.5. Regresion lineal múltiple y simple

Regresion lineal múltiple

Estudia la relación funcional entre una variable Y, llamada dependiente o respuesta y un conjunto de k variables llamadas independientes (explicativas o predictoras), $X_1, X_{2,}, ..., X_k$ de tal forma que la primera depende de la segunda o dicho de otra forma $X_1, X_{2,}, ..., X_k$ influye en el comportamiento de Y; con el objetivo de:

- Determinar las variables o factores que influyen en la variable respuesta.
- Modelar la variable respuesta y las variables independientes
- Predecir o estimar la media de Y, conociendo un valor de X.

Modelo de regresion lineal múltiple con k variables independientes

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_{2,} +, ..., + \beta_k X_k + \varepsilon_i \qquad (1)$$

$Y =$ Variable aleatoria

$X_1, X_{2,}, ..., X_k$ =Variables no aleatorias que influyen en el comportamiento de la variable Y

$\beta_1, \beta_{2,}, ..., \beta_k$ =Coeficientes asociados a las variables independientes

ε_i =Error aleatorio, se espera de poca influencia en Y, con media cero y varianza homogénea.

De (1) por el supuesto del error aleatorio

$$E(Y / X_1, X_2, ...X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \qquad (2)$$

(2) en (1)

$$Y_i = E(Y_i / X_1, X_2, ..., X_k) + \varepsilon_i \qquad (3)$$

Una función $f(X_1, X_2, ..., X_k)$ estima a $E(Y_i / X_1, X_2, ..., X_k)$

Regresion lineal simple

Para una sola variable independiente

$$Y_i = \beta_0 + \beta_1 X + \varepsilon_i$$

La media de la variable respuesta tomando en cuenta el supuesto de los errores aleatorios

$$E(Y / X) = \beta_0 + \beta_1 X$$

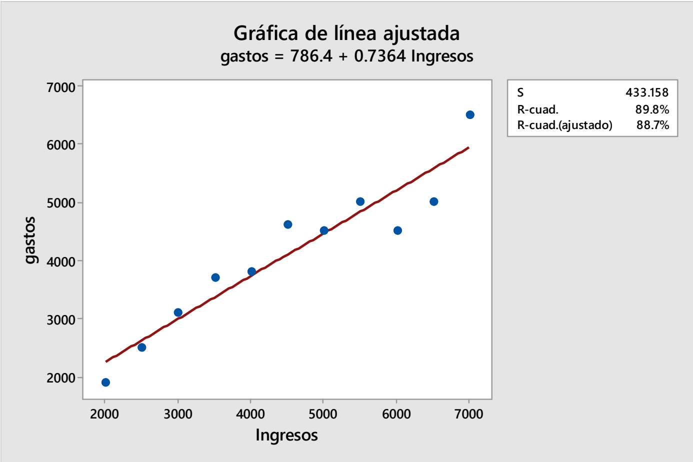
Un modelo estimado para $E(Y / X)$, es dado por $\hat{Y} = b_0 + b_1 X$

Ejemplo (1)

Y = gastos de las familias(S/.)   X = ingresos de los Jefes de familias(S/.)

Para estas dos variables, los gastos de las familias dependen de los ingresos, matemáticamente. Una muestra elegida al azar de 11 familias.

N°.	Ingresos	gastos
1	2000	1900
2	2500	2500
3	3000	3100
4	3500	3700
5	4000	3800
6	4500	4600
7	5000	4500
8	5500	5000
9	6000	4500
10	6500	5000
11	7000	6500



Una estimación para $E(Y/X)$, es dado por:: $\hat{Y} = f(X) = b_0 + b_1 X$

Supuestos para un modelo de regresión lineal simple

a. Los valores que toma X son fijados por el investigador.

b. Para cada valor de x_i fijo, se genera una población de valores para Y que tienen una distribución Normal con media y variancia, dada por:

$$Y \approx N(\quad \beta_0 + \beta_1 X_i \ ; \quad \sigma^2)$$

c. Para cada población, la diferencia entre un valor cualquiera y_i y su media $E(Y/x_i)$ se le llama error, es decir:

$$y_i - E(Y/x_i) = \varepsilon_i$$

$$y_i = E(Y/x_i) + \varepsilon_i$$

Si suponemos para la media $E(Y/x_i)$ es una recta lineal; es decir que una recta pase por todos los puntos de la media, luego:

$$f(x) = E(Y/x_i) = \beta_0 + \beta_1 x_i$$

Lo anterior es llamado la Ecuación de Regresión Lineal Poblacional de Y sobre X

Luego:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (1)$$

(1) Se le llama la ecuación de regresión lineal poblacional o modelo de regresión lineal poblacional.

Los errores ε_i se distribuyen normal e independientemente como media cero (0) y variancia (σ^2)

ECUACION DE REGRESION ESTIMADA

Para estimar el modelo (1) se extrae una muestra aleatoria de tamaño "n" y se registra los valores de los pares ordenados

$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$,

Con estos pares ordenados planteamos la función de regresión lineal muestral dada por (2), la que estima a la función de regresión poblacional; es decir:

$$y_i = b_0 + b_1 x_i + e_i \qquad (2)$$

La recta estimada es:

$$\hat{y} = b_0 + b_1 x \qquad (3)$$

(3) estima a la $E(Y/X)$, es decir:

$$\hat{y} \quad \rightarrow \quad E(Y/x_i)$$

Para estimar los coeficientes de la ecuación de regresión muestral dado por (2), se hará uso de la Técnica de los Mínimos Cuadros Ordinarios (MCO), técnica que consiste en calcular los valores de b_0 y b_1 tal que minimice la suma de los cuadrados de los errores; es decir de (2), se tiene:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = f(b_0, b_1)$$

Para obtener los valores que minimiza la función $f(b_0, b_1)$ derivamos parcialmente con respecto a b_0 y b_1 e igualamos a cero; es decir:

$$\frac{\partial f(b_0, b_1)}{\partial b_0} = 0 \quad \rightarrow \quad b_0 = \overline{y} - b_1 \overline{x}$$

$$\frac{\partial f(b_0, b_1)}{\partial b_1} = 0 \quad \rightarrow \quad b_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

2.2.6. Coeficiente de Determinación

Es un valor y mide la proporción o porcentaje de la variación de la variable Y explicada o atribuible al modelo estimado. Se calcula aplicando la siguiente formula.

$$R^2 = \frac{\sum_{I=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{I=1}^{n}(y_i - \overline{y})^2} * 100\% = \frac{b_1^2 \sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{I=1}^{n}(y_i - \overline{y})^2} * 100\%$$

Su valor varía de 0 al 100%, cuando esta próximo al 100%, indica que la ecuación de la recta estimada, se ajusta o se acomoda lo mejor posible a la nube de puntos.

Para una regresion lineal simple, el coeficiente de determinación se calcula a través del coeficiente de correlación de Pearson, elevando su valor al cuadrado, es decir,

. $R^2 = (r_{Pearson})^2 .*100\%$

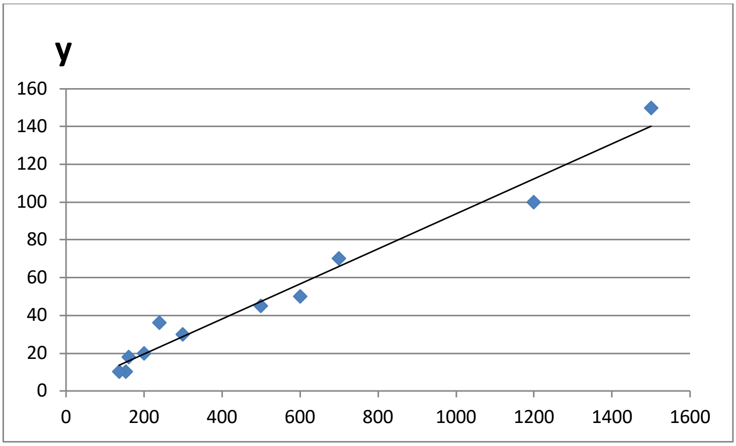
Ejemplo

Con la finalidad de estudiar la relación funcional entre la cantidad de docentes y la cantidad de artículos científicos publicados para 11 universidades elegidas al azar de cierto país, la tabla siguiente presenta la siguiente información.

X=N° de docentes	Y=N° de artículos publicados
200	20
300	30
700	70
1200	100
1500	150
600	50
154	10
136	10
160	18
240	36
500	45

a. Presentar los datos de ambas variables en un plano cartesiano (nube de puntos).
b. Estimar la ecuación de regresión lineal de y sobre x.
c. Calcular el coeficiente de determinación para la ecuación estimada.
d. Estimar la cantidad de artículos para una universidad que tienen 1000 docentes.

.Para la pregunta (a), se presenta la nube de puntos (X,Y)



De la nube de puntos, para las universidades en estudio existe una tendencia lineal positiva entre la Y=cantidad de docentes y la X=cantidad de artículos publicados.

Para la pregunta (b)

Para calcular los coeficientes de la recta estimada se debe calcular las siguiente sumatorias, resultados que aparecen en la última fila de la tabla

$$\sum_{i=1}^{7} x_i = 5690 \qquad \sum_{i=1}^{7} y_i = 539 \qquad \sum_{i=1}^{7} x_i y_i = 473920 \qquad \sum_{i=1}^{7} x_i^2 = 5045412$$

$$\sum_{i=1}^{7} y_i^2 = 45045$$

x	y	x*y	x*x	y*y
200	20	4000	40000	400
300	30	9000	90000	900
700	70	49000	490000	4900
1200	100	120000	1440000	10000
1500	150	225000	2250000	22500
600	50	30000	360000	2500
154	10	1540	23716	100
136	10	1360	18496	100
160	18	2880	25600	324
240	36	8640	57600	1296
500	45	22500	250000	2025
5690	**539**	**473920**	**5045412**	**45045**

Usando la formula se tiene:

$$b_1 = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{473920 - \frac{(5690)(539)}{11}}{5045412 - \frac{(5690)^2}{11}} = 0.09281537$$

$$b_0 = \overline{y} - b_1 \overline{x} = \frac{539}{11} - 0.09281537(\frac{5690}{11}) = 0.989140$$

Luego la recta de regresión estimada de y sobre x es:

$$\hat{y} = b_0 + b_1 x \quad = 0.98913812 + 0.09281537x$$

Interpretación de la pendiente

Por cada docente que se incrementa en las universidades, la cantidad artículos publicado se incrementa 0.093 en promedio, si el incremento es 20 docentes el incremento es de 1.86 artículos publicados en promedio.

Calculo del coeficiente de determinación

Aplicando la formula, el valor del coeficiente de determinación es de 97.18%, este valor indica que, el 97.18% de la variación de los valores que toma Y (cantidad de artículos publicados), se atribuye a la variación de los valores que toma la variable (cantidad de docentes).

Y la estimación de la cantidad promedio de artículos publicado para 1000 docentes es de 93.8, este valor se obtiene al reemplazar en la ecuación estimada el valor de X=1000.

Ejercicios propuestos.

1. La información de la tabla corresponde a la cantidad de artículos publicados en química, durante los años 1990 hasta 2020. Estimar el modelo lineal.

x=años	y=cantidad de artículos publicados
1990	13
1995	28
2000	42
2005	58
2010	77
2015	90
2020	110

2. Con la información de la pregunta 1. Calcular e interpretar el coeficiente de determinación.
3. Con la información de la pregunta 1. Estimar la cantidad de artículos publicados para el año 2022.

2.2.7. Regresión lineal simple con datos codificados para la variable X

Ocurre situaciones en la cual los valores que toma la variable independiente X están igualmente espaciados, cuando esto acurre los cálculos se simplifican para obtener el valor del intercepto y la pendiente de la recta de regresión estimada. Para logar estos objetivos los datos de la variable X deben estar codificados.

Metodología para codificar los valores que toma X

a. Se calcula el valor de $-(n-1)/2$

b. Al valor calculado en el punto (a) se le multiplica por el espaciado "r" obteniéndose: $-r(n-1)/2$, convirtiéndose en el primer valor codificado

c. Los restantes valores codificados se obtienen sumándole al valor obtenido en el punto (b) el valor del espaciado "r" .

d. Con los valores codificado se verifica que $\sum_{i=1}^{n} x_i^* = 0$, con este resultado se simplifican los cálculos del intercepto y pendiente.

Nota

Otra forma de obtener lo mismo es utilizar la siguiente relación:

$$X_i^* = x_i - \overline{x}$$

$$x_i^* = a + (i-1)r$$

$$\overline{x} = a + \frac{r(n-1)}{2}$$

$$x_i^* - \overline{x} = (i-1)r - \frac{r(n-1)}{2} = \frac{2r(i-1) - r(n-1)}{2} = \frac{r(2i-2-n+1)}{2} = r\left[\frac{-(n-2i+1)}{2}\right]$$

$$x_i^* - \overline{x} = r\left[\frac{-(n-2i+1)}{2}\right]$$

$$r = espacido$$

$$i = i - esima \quad observacion$$

$$n = tamaño \quad de \quad la \quad muestra$$

$$tambien \quad se \quad cumple \quad que:$$

$$\sum_{i=1}^{n}(x_i^* - \overline{x}) = -\frac{r}{2}\sum_{i=1}^{n}(n-2i+1) = -\frac{r}{2}\left[n^2 - 2\frac{n(n+1)}{2} + n\right] = -\frac{r}{2}\left[n^2 - n^2 - n + n\right] = 0$$

Ejemplo

La información que se presenta en la tabla, corresponde a la producción de artículos científicos en biológicas durante los años 1990 al 2015.

X=años	Y=producción
1990	15
1995	38
2000	54
2005	76
2010	92
2015	115

a. Estimar la ecuación de regresión lineal de Y sobre X

b. Calcular el coeficiente de correlación lineal de x e y

c. Calcular el coeficiente de determinación del modelo

d. Estimar la producción de artículos para el años 2017.

e. Obtener a partir de la ecuación de regresión codificada, la ecuación de regresión para datos sin codificar.

Regresión parar datos sin codificar

x	y	$x.y$	x^2	y^2
1990	15	29850	3960100	225
1995	38	75810	3980025	1444
2000	54	108000	4000000	2916
2005	76	152380	4020025	5776
2010	92	184920	4040100	8464
2015	115	231725	4060225	13225
12015	**390**	**782685**	**24060475**	**32050**

Cálculos:

$$b = 3.90857143$$

$$a = -7761.91429$$

$$r = 0.99877964$$

$$R^2 = 99.756$$

$$y = -7761.91429 + 3.90857143(2017) = 121.6742843 = 121.67$$

Codificando los datos

Solución (a):

Los valores de la variable x están igualmente espaciado, luego

$$-\frac{(6-1)}{2} = -2.5 \rightarrow 5(-2.5) = -12.5$$

$x^* = años$	$y =$ Producción
-12.5	15
-7.5	38
-2.5	54
2.5	76
7.5	92
12.5	115

Calculamos las sumas de las variables y su producto cruzado

N°	x^*	y	$x^*.y$	$(x^*)^2$	y^2
1	-12.5	15	-187.5	156.25	225
2	-7.5	38	-285	56.25	1444
3	-2.5	54	-135	6.25	2916
4	2.5	76	190	6.25	5776
5	7.5	92	690	56.25	8464
6	12.5	115	1437.5	156.25	13225
Total	**0**	**390**	**1710**	**437.5**	**32050**

Cálculos:

$$b = \frac{\sum_{i=1}^{6} x_c y}{\sum_{i=1}^{6} x_c^2} = \frac{1710}{437.5} = 3.90857$$

$$a = \frac{\sum_{i=1}^{6} y}{6} = \frac{390}{6} = 65$$

La ecuación de regresión lineal codificada es:

$$y = 65 + 3.90857\, x^*$$

<u>Solución (b):</u>

Calculo del coeficiente de correlación lineal de $x \quad e \quad y$

$$r = \frac{\sum x_c.y}{\sqrt{\sum x_c^2}.\sqrt{\sum y^2 - \frac{(\sum y)^2}{6}}} = \frac{1710}{\sqrt{437.5}.\sqrt{32050 - \frac{(390)^2}{6}}} = \frac{1710}{1712.089367} = 0.99878$$

<u>Solución (c):</u>

Coeficiente de determinación

$$R^2 = r^2(100\%) = (0.99878)^2(100\%) = 99.756\%$$

Solución (d):

Para la proyección se debe codificar año 2017 a través de la relación

$$x^* = x - \overline{x} \rightarrow x_c = 2017 - 2002.5 = 14.5$$

$$y = 65 + 3.90857\, x^* \rightarrow y = 65 + 3.90857\,(14.5) = 121.67$$

Solución (e):

Reemplazamos en la ecuación de regresión lineal codificada el valor de

$$x^* = x - \overline{x} \rightarrow x^* = x - 2002.5$$

Luego:

$$y = 65 + 3.90857(x - 2002.5) \rightarrow y = -7761.911425 + 3.90857\, x$$

Este último resultado, también se obtiene si utilizamos los datos originales.

Ejercicios propuestos

1. La información que a continuación se presenta corresponden a la producción de artículos científicos en química entre los años 1990 a 2020.

X=Años	Y=Producción de artículos científicos
1990	22
1995	30
2000	42
2005	55
2010	62
2015	70
2020	82

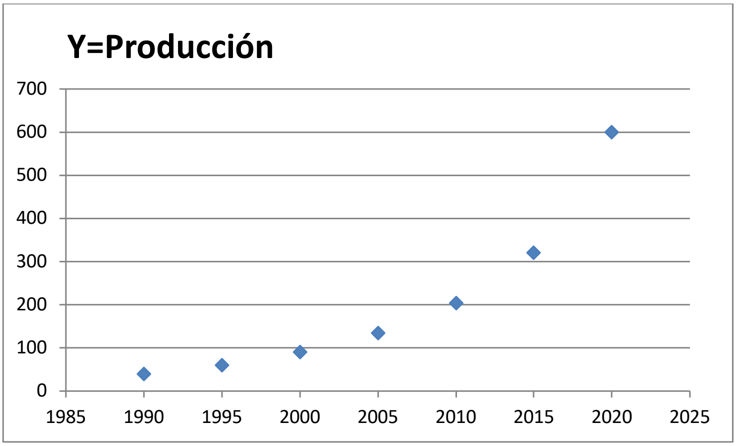
   a. Estimar la ecuación de regresión lineal de Y sobre X codificada.
   b. Calcular el coeficiente de correlación lineal
   c. Calcular el coeficiente de determinación
   d. Estimar la producción de revistas científicas para el año 2013 (utilizar 2 decimales).
   e. Estimar la producción de revistas para el año 2022 (utilizar 2 decimales).
   f. Estimar la producción de revistas para el año 2025(utilizar 2 decimales).

2.2.8. Regresion no lineal

Sea la información de la tabla

X=años	Y=producción
1990	40
1995	60
2000	90
2005	135
2010	203
2015	320
2020	600

Gráfica de la información de la tabla



La nube de puntos de la gráfica, nos indica una tendencia no lineal y para modelar este comportamiento, se busca un modelo matemático que mejor se ajuste a la nube de puntos.

Modelo no lineal

$$y = a.b^x$$

Aplicando logaritmo decimal a ambos lado de la igualdad con el objetivo de linealizar el modelo no lineal y transformarlo en un modelo lineal, modelo que permite calcular los valores de a y b.

$$y = ab^x \rightarrow \log y = \log a.b^x \rightarrow \log y = \log a + \log b^x$$
$$\log y = \log a + x(\log b) \rightarrow \log y = \log a + (\log b)x$$
$$Z = A + Bx \rightarrow \text{mod}elo \quad lineal$$

Dónde:

$$Z = \log y$$
$$A = \log a \rightarrow a = 10^{A}$$
$$B = \log b \rightarrow b = 10^{B}$$

Sus ecuaciones para calcular A y B, suponiendo la variable X codificada, son:

$$B = \frac{\sum_{i=1}^{n} x_i z_i}{\sum_{i=1}^{n} x_i^2} \qquad A = \frac{\sum_{i=1}^{n} z_i}{n}$$

Estima la ecuación de regresion no lineal de la forma $y = ab^{x^*}$

El modelo no lineal: $y = 141.5134851\,(1.091719056\,)^{x^*}$

Estimar la producción de revistas científicas para el año 2022.

Codificando el valor X=2022

X*=X-media de los valores X

X*=2022-2005=17años

$$y_{2022} = 141.5134851(1.091719056)^{17}$$

Sea el modelo de regresión no lineal

$$Y_i = \beta_0 . \beta_1^{x_i} \varepsilon_i$$

Haciendo la transformación con el objeto de llevarlo a un modelo lineal.

$$\log Y_i = \log \beta_0 . \beta_1^{x_i} \varepsilon_i = \log \beta_0 + x_i \log \beta_1 + \log \varepsilon_i$$
$$z_i = \beta_0^* + \beta_1^* x_i + \varepsilon_i^* \qquad como \quad E(\varepsilon_i^*) = 0$$

Una estimación para $E(Z/X) = \beta_0^* + \beta_1^* x_i$, es dado por

$$\hat{Z} = \hat{\beta}_0^* + \hat{\beta}_1^* X = A + BX$$
$$z = A + Bx \quad \text{mod}elo \quad linealizado \rightarrow z = \log y \quad A = \log \hat{\beta}_0^* \quad B = \log \hat{\beta}_1^*$$

Ejemplo

Sea la información de la tabla

X=años	Y=producción de artículos científicos en matemática
1990	5
1995	10
2000	20
2005	40
2010	80
2015	160

a. Estimar la ecuación de regresión no lineal $y = ab^x$
b. Calcular el coeficiente de correlación lineal de x e y
c. Calcular el coeficiente de determinación del modelo
d. Proyectar la producción de artículos científicos para el año 2017 y 2020
e. Obtener a partir de la ecuación de regresión codificada, la ecuación de regresión para datos sin codificar

Como los datos de la variable x están igualmente espaciados, entonces se trabaja con datos codificados:

x^*	y	$z = \log y$	$z.x^*$	$(x^*)^2$	z^2
-12.5	5	0.69897	-8.73712505	156.25	0.48855907
-7.5	10	1	-7.5	56.25	1
-2.5	20	1.30103	-3.25257499	6.25	1.69267905
2.5	40	1.60205999	4.00514998	6.25	2.56659622
7.5	80	1.90308999	14.2731749	56.25	3.6217515
12.5	160	2.20411998	27.5514998	156.25	4.8581449
0	315	8.70926996	26.3401246	437.5	14.2277307

Solución (a):

$$B = \frac{\sum x_c z}{\sum x_c^2} = \frac{26.3401246}{437.5} = 0.060206 \rightarrow 0.060206 = \log b \rightarrow b = 1.148698$$

$$A = \frac{\sum z}{6} = \frac{8.709270}{6} = 1.451545 \rightarrow 1.451545 = \log a \rightarrow a = 28.284272$$

El modelo codificado es:

$$y = 28.28424272\,(1.148698)^{x^*}$$

Solución (b):

Calculo del coeficiente de correlación entre x_c e z

$$r=\frac{\sum x_c.z}{\sqrt{\sum x_c^2}.\sqrt{\sum z^2-\frac{(\sum z)^2}{6}}}=\frac{26.3401246}{\sqrt{437.5}.\sqrt{14.2277307-\frac{(8.70926996)^2}{6}}}=\frac{26.3401246}{26.3401244}=1.0$$

Solución (c):

Coeficiente de determinación

$$R^2=r^2(100\%)=(1.0)^2(100\%)=100\%$$

Solución (d):

Para la proyección se debe codificar año 2017 a través de la relación

$$x^*=x-\bar{x}\rightarrow x_c=2017-2002.5=14.5$$

$$y=28.284272(1.148698)^{x^*}\rightarrow y=28.284272(1.148698)^{14.5}=211.12$$

Para la proyección se debe codificar año 2020 a través de la relación

$$x^*=x-\bar{x}\rightarrow x_c=2020-2002.5=17.5$$

$$y=28.284272(1.148698)^{x^*}\rightarrow y=28.284272(1.148698)^{17.5}=320$$

Solución (e):

Reemplazamos en la ecuación de regresión codificada el valor de

$$x^*=x-\bar{x}\rightarrow x^*=x-2002.5 \quad (3)$$

(1) en (1)

$$y=28.284272(1.148698)^{x_c}\rightarrow y=28.284272(1.148698)^{(x-2002.5)}$$

Ejemplos propuestos

1. Los datos que a continuación se presenta, corresponden a la producción de artículos científicos en matemáticas para los años 1980 al 2005.

X= años	Y=producción de artículos científicos en matemática
1980	20
1985	38
1990	60
1995	82
2000	100
2005	122

a. Estimar el modelo de regresión lineal de Y sobre X

b. Calcular el coeficiente de correlación lineal de X e Y, interpretar

c. Calcular el coeficiente de determinación, interpretar

d. Estimar la producción promedio del número de artículos científicos para los años 1992 y 2003

e. Proyectar la producción promedio para el año 2016

2. La información que a continuación se presenta, corresponden a la producción de artículos científicos científicas en matemáticas para los años 1980 al 2005.

X= años	Y=producción de artículos científicos en matemática (1980-2005)
1980	20
1985	38
1990	80
1995	164
2000	320
2005	642

a. Estimar el modelo de regresión no lineal de Y sobre X ($y = a.b^x$)

b. Calcular el coeficiente de correlación de X e Y, interpretar

c. Calcular el coeficiente de determinación, interpretar

d. Estimar la producción promedio para los años 1992 y 2003

e. Proyectar la producción promedio para los años 2015 y 2016

3. Sean los datos de la tabla, estimar el modelo de regresion no lineal $y = \frac{ab}{a+bx}$.

X	Y
2	0.75
4	0.43
6	0.30
8	0.23
10	0.18

4. Con los datos de la pregunta 3, estimar el valor de “y” para x=5

5. Estimar la ecuación de Lotka $y = \frac{a}{x^b}$, para los datos que se presenta a continuación.

X=N° Artículos publicados	Y=N° Autores que publican
1	950
2	174
3	64
4	32
5	18
6	12
7	8
8	6
9	4
10	3
11	3
12	2
13	2
14	1
15	1
16	1

a. Con la información de la pregunta 5. Calcular el coeficiente de determinación.

b. Con la información de la pregunta 5. Calcular la elite de autores más productivos e interpretar.

c. Calcular el índice de productividad de autores.

Capítulo 3

3.1. Leyes bibliometricas

Estudia el comportamiento estadísticamente regular de la producción y el consumo de la literatura científica, de una determinada materia y en un periodo de tiempo determinado.

A través de la ley bibliometricas, se busca un modelo matemático que mejor se ajuste a la nube de puntos, de tal forma que se pueda modelar y realizar predicciones.

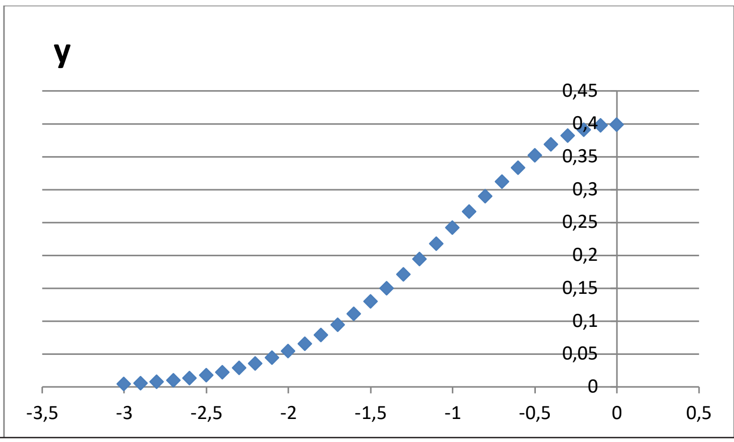
3.2 Ley de Price (ley del crecimiento exponencial de la literatura científica)

Derek J. de solla Price (1956), constato que el crecimiento de la literatura científica se produce a un ritmo muy superior respecto a otros fenómenos sociales, pero muy similar a otros fenómenos observables en contextos naturales, como los procesos biológicos. Dicho crecimiento de la literatura científica de una determinada materia, es tal que cada 10 a 15 años la literatura científica existente se duplica con un crecimiento exponencial, aunque esto depende en gran medida del área de conocimiento de la que se trate.

Cada materia de estudio tiene una evolución natural, pasando por diversas etapas:

- Precursores: corresponde a las primeras publicaciones de la materia en estudio.
- Crecimiento exponencial: la materia de estudio por su necesidad requiere ser investigado a profundidad.
- El crecimiento lineal: se hace lento la producción de las publicaciones.
- Saturación: ya no existe más producción, ´porque la materia ya es conocida.

Gráfica



Para el curso se analiza la parte que tienen un crecimiento exponencial, a través de un modelo no lineal de la forma:

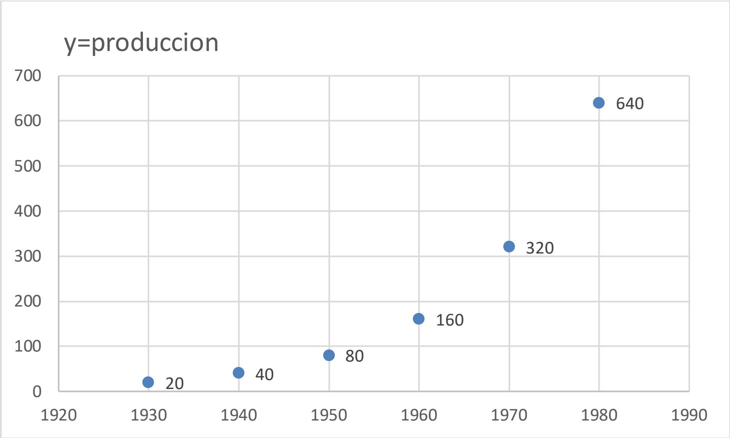
Modelo Matemático:

$$y = a.e^{bx}$$

Dónde:

Y= producción de la literatura científica (revistas, artículos, autores, etc.) de un tema específico y durante un período de tiempo.
X=años
e=base de logaritmos neperianos=2.718281828…

Gráfica:



Para estimar el modelo, se dispone de una serie cronológica X=años y Y =producción de literatura científica de una materia determinada y en un período de tiempo.

$$(x_1, y_2), (x_2, y_2), \ldots (x_n, y_n)$$

Para estimar el modelo exponencial de Price, se linealiza aplicando logaritmos neperianos ambos lados de la ecuación.

$$\begin{aligned}
&y = a.e^{bx} \rightarrow \text{mod}\,elo \quad no \quad lineal \\
&\ln y = \ln ae^{bx} \\
&\ln y = \ln a + \ln e^{bx} \\
&\ln y = \ln a + bx(\ln e) \\
&\ln y = \ln a + bx \\
&Z = A + Bx \rightarrow \text{mod}\,elo \quad lineal \\
&B = \frac{\sum xz}{\sum x^2} = b \\
&A = \frac{\sum z}{n} = \ln a \rightarrow a = e^A = \exp(A)
\end{aligned}$$

$$\ln y = \ln a.e^{bx} = \ln a + \ln .e^{bx} = \ln a + bx(\ln e)$$
$$\ln y = \ln a + bx$$
$$z = A + Bx \quad ,siendo:$$
$$z = \ln y, \quad A = \ln a, \quad B = b$$

Luego las ecuaciones para B y A son:

$$B = \frac{\sum xz - \frac{(\sum x)(\sum z)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$
$$A = \bar{z} - B\bar{x}$$

Si los valores x son codificados, entonces $\sum x_c = 0$ las ecuaciones se simplifican, quedando:

$$B = \frac{\sum x_c z}{\sum x_c^2} = b$$

$$A = \ln a \rightarrow a = e^A = \exp(A)$$

Luego se calcula el coeficiente de determinación, para conocer la bondad de ajuste del modelo estimado. Para el modelo codificado:

$$R^2 = \frac{B^2 \sum_{i=}^{n} (x_i^*)^2}{\sum_{i=1}^{n} z_i^2 - \frac{(\sum_{i=1}^{n} z_i)^2}{n}} (100\%)$$
$$o \quad R^2 = (r - Pearson \quad con \quad datos \quad codificados)^2$$

Ejemplo-1

La producción de artículos científicos de cierta materia, durante los años 1930 al 1980, se presenta en el siguiente cuadro

X =años	Y=producción de artículos científicos	$x_c =$ años codificado	$z = \ln y$	$x_c.z$	x_c^2
1930	20	-25	Ln20	-25.Ln20	625
1940	40	-15	Ln40	-15.Ln40	225
1950	80	-5	Ln80	-5.Ln80	25
1960	160	5	Ln160	5.Ln160	25
1970	320	15	Ln320	15.Ln320	225
1980	640	25	Ln640	25.Ln640	625
Total		0	28.37160	121.30075	1750

Estimar la parte exponencial del modelo de Price $y = ae^{bx}$

$$B = \frac{\sum x_c z}{\sum x_c^2} = \frac{121.30075}{1750} = 0.06931472 = b$$

$$A = \bar{z} = \frac{28.37160}{6} = 4.72860022 \rightarrow a = e^{4.72860022} = 113.137085$$

$$y = 113.137085.e^{0.06931472\,x^*}$$

Calculo del coeficiente de determinación:

$$r_{pearson} = 1 \rightarrow R^2 = (r_{pearson})^1 = 1*100\% = 100\%$$

El valor del coeficiente de determinación indica un ajuste perfecto del modelo hacia la nube de puntos.

Estimar la producción de artículos científicos para los años 1975 y 1990.

Para utilizar el modelo estimado codificado, se debe codificar los años 1975 y 1990.

$$X_i^* = X - \overline{X} \quad , \quad \overline{X} = 1955$$

Para el año 1975

$$X_{1975}^* = 1975 - 1955 = 20$$

$$y = 113.137085.e^{0.06931472(20)} = 452.55$$

La proyección para 1990, entonces $x_c = 35$

$$y = 113.137085.e^{0.06931472 x_c} = 113.137085.e^{0.06931472(35)} = 1280$$

De otra forma para el año 1990:

$$z = 4.7286 + 0.06931 x_c$$
$$z = 4.7286 + 0.06931(x - \bar{x})$$
$$z = 4.7286 + 0.06931(1990 - 1955)$$
$$z = 7.15445 \rightarrow z = \ln y = 7.15445 \rightarrow y = e^{7.15445} = 1279.788 = 1280$$

Calculo del tiempo de duplicación de la producción de literatura científica para el modelo exponencial

$$D = \frac{\ln 2}{b} = \frac{0.69314718}{0.06931} = 10 \text{ años}$$

El tiempo de duplicación de la producción de la literatura científica es cada 10 años.
Indica que cada 10 años en promedio la producción científica se duplica.

Calculo de la tasa de incremento anual (T)

$$T = 100\%(e^{b} - 1) = 100\%(e^{0.06931} - 1) = 7.176\%$$

Anualmente la producción de la literatura científica crece en promedio en un 7.176%.

Ejemplo-2

La producción de revista científica de cierta materia, durante los años 1930 al 1980, se presenta en la siguiente cuadro

X =años	Y=producción de artículos científicos.	$x^* =$ años codificado	$z = \ln y$	$x^*.z$	$(x^*)^2$
1960	20	-30	2.99573227	-89.8719682	900
1970	30	-20	3.40119738	-68.0239476	400
1980	83	-10	4.41884061	-44.1884061	100
1990	155	0	5.04342512	0	0
2000	325	10	5.78382518	57.83825182	100
2010	650	20	6.47697236	129.5394473	400
2020	1270	30	7.14677218	214.4031654	900
13930	2533	0	35.2667651	199.6965425	2800

Calculo del modelo de Price

$$y = 154.1782346 * e^{0.071320194\, x^*}$$

Calculo de coeficiente de determinación

$$R^2 = 99.55\%$$

Buen ajuste del modelo a la nube de puntos

Predecir la producción de la literatura científica para los años 2022 y 2025

Calculo de la media de los años, la media de los años es 1990

Años x	Años codificados $x-\bar{x}$	Producción estimada(reemplazando el respectivo año codificado al modelo estimado)
2022	32	1510.7
2025	35	1871.2

Calculo del tiempo de duplicación de la producción científica

$$D=\frac{\ln 2}{b}=\frac{\ln 2}{0.071320194}=9.7 \text{ años}$$

Calculo de la tasa de incremento anual

$$R=100\%(e^{b}-1)=100\%(e^{0.071320194}-1)=7.39\%$$

<u>Ley de Price: obsolescencia (envejecimiento) de la producción de la literatura científica</u>

Price, observa sobre la rapidez con la que la literatura científica de un tema dado pierde vigencia. Cuanto más tiempo tenga un recurso científico menos se cita (mención en los trabajos científicos).No es uniforme para todas las disciplina.

Tabla

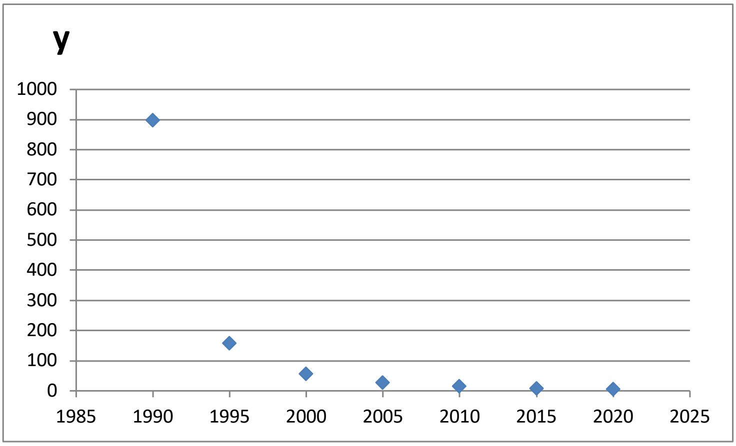
Años	Nro. De artículos publicados de un tema dado	Número de Referencias (B)	Cantidad del número de referencia de los 5 años anteriores(B)	Índice de Price= es el cociente entre (B/A)
2000	300	900	1996,1997,1998 ,1999,2000 igual 600	(600/900)100%=66.7% Un índice de Price superior a 30% indica que los artículos serán obsoletos rápidamente
2001	200	400	2001,2000,1999 ,1998,1997=50	(50/400)100%=12.5%

<u>Índice de Price</u>

$$I_{\text{año } t}=\frac{Cantidad \quad de \quad referencia \quad para \quad los \quad años \quad (t-5+1)}{total \quad de \quad referencia \quad hasta \quad el \quad año \quad t}(100\%)$$

<u>Grafica</u>

<u>Del comportamiento de la producción científica para una determinada materia</u>



3.3 Ley de Lotka (1926)

<u>Ley de la productividad de los autores</u>

<u>Enunciado</u>

Lotka observo que para una determinada materia de estudio y en un determinado periodo de tiempo, lo siguiente: Muchos autores publican pocos artículos científicos y pocos autores publican gran cantidad de artículos científicos. La ley de Lotka relaciona la cantidad de artículos científicos publicados (X) con la cantidad de autores (Y).

Ejemplo

Supongamos, que un período de 5 años, se ha publicado 78 artículos científicos sobre física para un total de 665 investigadores.

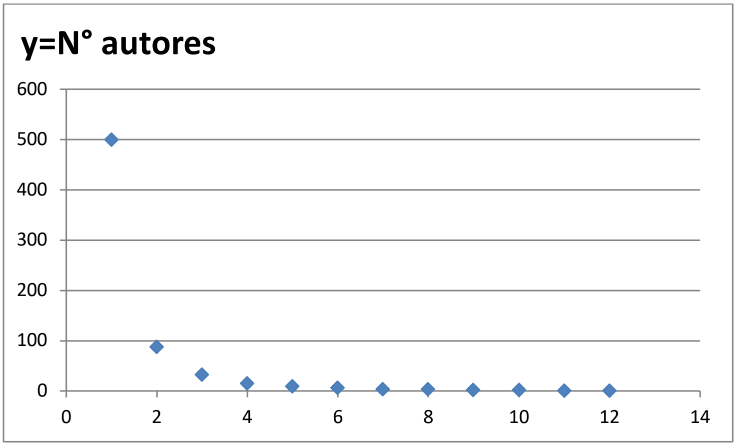
Tabla: Cantidad de artículos publicados y Cantidad de autores investigadores

X=cantidad de artículos científicos	Y= cantidad de autores investigadores
1	500
2	88
3	32
4	16
5	9
6	7
7	4
8	3
9	2
10	2
11	1
12	1
78	665

Interpretación de la tabla

Un grupo de 500 investigadores durante los 5 años, publicaron 1 solo trabajo científico, 88 publicaron 2 y un solo autor en ese mismo período de estudio publico 12 trabajos científico.

Gráfica



Modelo no lineal de Lotka

$$y = \frac{a}{x^b}$$

Dónde:

Y= cantidad de autores (investigadores)
X=cantidad de artículos publicados

Para estimar los valores de a y b del modelo de Lotka, se toma una muestra de tamaño n y se registran los valores de (X, Y). Como el modelo de Lotka es no lineal, para calcular los valores de a y b, se linealiza aplicando logaritmos decimales a ambos miembros de la igualdad.

$$y = \frac{a}{x^b} \rightarrow \quad \log y = \log(\frac{a}{x^b}) = \log a - \log x^b \rightarrow \log y = \log a - b\log x$$

$$\log y = \log a + (-b)\log x \rightarrow$$

$$Z = A + BW \rightarrow \text{mod}\,elo \quad lineal \qquad (2)$$

$$donde : Z = \log y \quad W = \log x \quad A = \log a \quad B = -b$$

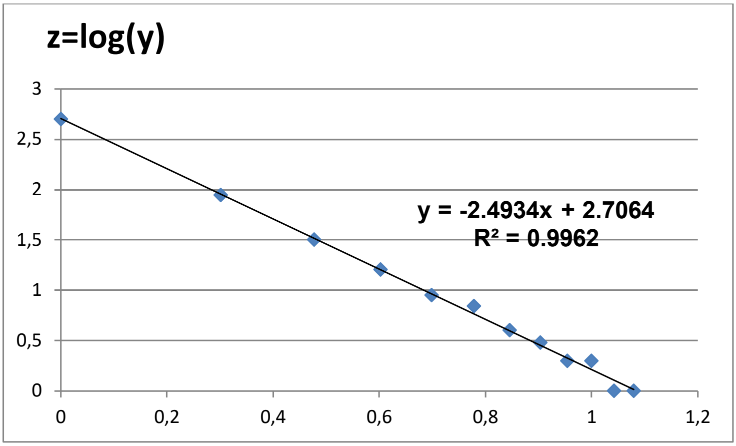
De (2) se calculan A y B:

$$B=\frac{\sum WZ-\frac{(\sum W)(\sum Z)}{n}}{\sum X^2-\frac{(\sum x)^2}{n}}=-b\rightarrow b=-B$$

$$A=\frac{\sum Z}{n}-B(\frac{\sum W}{n})=\log a\rightarrow a=10^A$$

Gráfica del modelo linealizado de Lotka

$$y=\frac{a}{x^b}\qquad linealizando\qquad z=A+Bw$$



Calculo del coeficiente de correlación de Pearson

Se calcula en el modelo transformado (2)

$$r=\frac{\sum_{i=1}^{n}w_iz_i-\frac{(\sum_{i=1}^{n}w_i)(\sum_{i=1}^{n}z_i)}{n}}{\sqrt{\sum_{i=1}^{n}w_i^2-\frac{(\sum_{i=1}^{n}w_i)^2}{n}}*\sqrt{\sum_{i=1}^{n}z_i^2-\frac{(\sum_{i=1}^{n}z_i)^2}{n}}}$$

Si r tiende a -1, los datos de w y z están asociados linealmente en forma inversa, esta asociación lineal con los datos transformados, también se transmite a las variables x e y.

Si r tiende a 0 los datos de z y w no están asociados linealmente, esta asociación lineal, también se transmite a las variables x e y.

Calculo del coeficiente de determinación

$$R^2=(coefciente\quad de\quad correlacion\quad de\quad pearson=r)^2$$

Elite de autores más productivos

Elite de autores es la cantidad de autores que publican más trabajos. Se calcula a través de la raíz cuadrada del total de autores.

$$E = \sqrt{total\ de\ autores\ en\ el\ \ periodo\ de\ estudio}$$

Para calcular la cantidad de trabajos publicados por los autores más productivos, se obtiene el acumulado de autores, del más productivo hasta el menos productivo.

Índice de productividad de autores (IP)

El índice de Lotka, permite identificar a los autores con baja, mediana y alta productividad, a través de la siguiente relación:

$$IP = \log(x) = w = \begin{cases} IP = 0 & autores \quad con \quad baja \quad productividad \\ 0 < IP < 1 & autores \quad con \quad mediana \quad productividad \\ IP \geq 1 & autores \quad con \quad alta \quad productividad \end{cases}$$

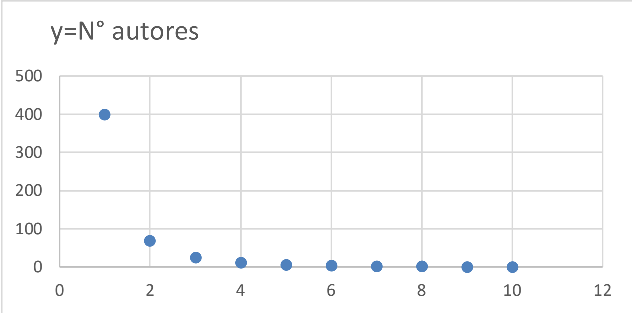
Ejemplo aplicativo-1

Los datos que a continuación se presentan corresponden a 526 autores que durante de 10 han publicado artículos, sobre una determinada materia. La primera fila de la tabla significa que se encontró a 400 autores que solo publicaron durante el periodo de estudio un solo artículo, y en la última fila a un solo autor que público 10 artículos durante el periodo de estudio.

x = N° de artículos
y = N° de autores

x=N° artículos	y=N° autores
1	400
2	70
3	26
4	12
5	7
6	4
7	3
8	2
9	1
10	1

Gráfica: artículos publicados versus autores



a. Estimar la ecuación de Lotka

b. Calcular el coeficiente de determinación

c. Calcular la elite de autores más productivo

Para calcular la ecuación de Lotka de debe realizar los siguientes cálculos:

x Artículos	y Autores	$w=\log x$	$z=\log y$	wz	w^2	z^2
1	400	0	2.60205999	0	0	6.7707162
2	70	0.30103	1.84509804	0.55542985	0.09061906	3.4043868
3	26	0.47712125	1.41497335	0.67511386	0.22764469	2.0021496
4	12	0.60205999	1.07918125	0.64973185	0.36247623	1.1646322
5	7	0.69897	0.84509804	0.59069818	0.48855907	0.7141907
6	4	0.77815125	0.60205999	0.46849374	0.60551937	0.3624762
7	3	0.84509804	0.47712125	0.40321424	0.7141907	0.2276447
8	2	0.90308999	0.30103	0.27185717	0.81557152	0.0906191
9	1	0.95424251	0	0	0.91057877	0
10	1	1	0	0	1	0
55	**526**	**6.55976303**	**9.16662191**	**3.61453889**	**5.21515941**	**14.736815**

Calculo de

$$B=\frac{\sum wz-\frac{(\sum w)(\sum z)}{10}}{\sum w^2-\frac{(\sum w)^2}{10}}=\frac{3.61453889-\frac{(6.55976303)(9.16662191)}{10}}{5.21515941-\frac{(6.55976303)^2}{10}}=\frac{-2.398547862}{0.912110309}=-2.629668624=-b$$

$$b=2.629668624$$

$$A=\frac{\sum z}{10}-b(\frac{\sum w}{10})=\frac{9.16662191}{10}-(-2.629668624)(\frac{6.55976303}{10}=2.641662493=\log a$$

$$\log a=2.641662493\rightarrow a=10^{2.641662493}=438.1900311$$

La ecuación de Lotka estimada es:

$$y = \frac{438.19}{x^{2.62967}}$$

Coeficiente de correlación
$r = -0.997887$

Coeficiente de determinación

$R^2 = (r)^2(100\%) = 99.58\%$

Elite de autores que más producen artículos científicos.

$Elite = \sqrt{total\ de\ autores} = \sqrt{526} = 22.93 = 23$ Autores publican 4 o más artículos

Tabla 01

x=N° artículos	y=N° autores	IP=log(x)	acumulados autores
1	400	0	526
2	70	0.30103	126
3	26	0.47712125	56
4	12	0.60205999	30
5	7	0.69897	18
6	4	0.77815125	11
7	3	0.84509804	7
8	2	0.90308999	4
9	1	0.95424251	2
10	1	1	1

Tabla 02

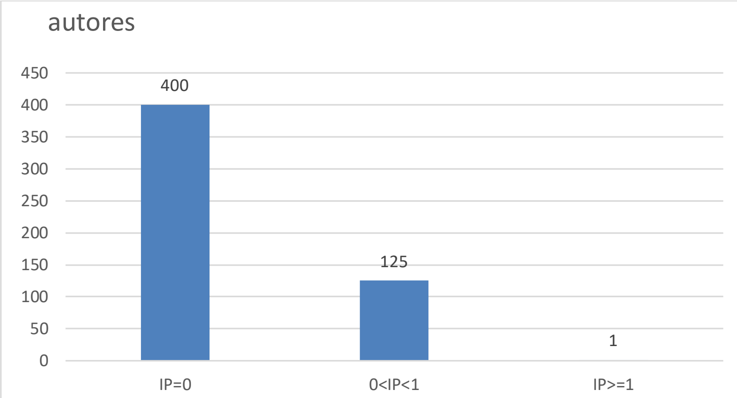
Índice de productividad	Autores	%
IP=0	400	76.05%
0<IP<1	125	23.76%
IP>=1	1	0.19%
Total	526	100.00%

La tabla 1, la última columna y la fila 5 se interpreta de la siguiente forma: 18 autores publican 5 o más artículos. Para 23 autores, publican 4 o más artículos

La tabla 02, indica la cantidad de autores más productivos y se obtienen a través del índice de productividad (IP), se calcula utilizando el logaritmo del número de artículos.

$$IP = \log(x) = \begin{cases} IP = 0 & baja \quad productividad \\ 0 < IP < 1 & media \quad productividad \\ IP \geq 1 & alta \quad productividad \end{cases}$$

Grafica



3.4 Ley de Bradford (1934)

La ley de Bradford relaciona revistas científicas con artículos publicados con el objetivo de determinar las revistas más productivas sobre un tema dado

Enunciado

Si un conjunto de revistas científicas se ordena asociadas a la productividad decreciente al número de artículos sobre un tema dado. Estas revistas se pueden agrupar en zonas que se forman dependiendo aproximadamente de la misma cantidad de artículos. Mientras que el conjunto de revistas utilizadas para cada zona aumenta en progresión geométrica $ak^0, ak^1, ak^2, ak^3, ...siendo \quad a = menor \quad N° \quad de \quad revistas \quad k = multiplicador = razon$

Ejemplo aplicativo

164 revistas científicas se ordenan en forman decreciente a la cantidad de artículos sobre un tema dado. (tomado del ejemplo de Bradford-Bibliografía de lubricación-1933)

N°	N° revistas	N° artículos	total, N° artículos	x=acum. N° revistas	y=acum total, N° artículos	w=ln(x)	w.y	w.w
1	1	22	22	1	22	0	0	0
2	1	18	18	2	40	0.69314718	27.7258872	0.48045301
3	1	15	15	3	55	1.09861229	60.4236759	1.20694896
4	2	13	26	5	81	1.60943791	130.364471	2.59029039
5	2	10	20	7	101	1.94591015	196.536925	3.78656631
6	1	9	9	8	110	2.07944154	228.73857	4.32407713
7	3	8	24	11	134	2.39789527	321.317967	5.74990174
8	3	7	21	14	155	2.63905733	409.053886	6.96462359
9	1	6	6	15	161	2.7080502	435.996082	7.33353589
10	7	5	35	22	196	3.09104245	605.844321	9.55454345
11	2	4	8	24	204	3.17805383	648.322981	10.1000261
12	13	3	39	37	243	3.61091791	877.453053	13.0387282
13	25	2	50	62	293	4.12713439	1209.25037	17.0332382
14	102	1	102	164	395	5.09986643	2014.44724	26.0086376
total	164		395	375	2190	34.2785669	7165.47543	108.171571

Pasos:

1. Calculamos para la 3ra columna: total de artículos por revistas, luego el total general, que resulta 395
2. Al total general de artículos (395), lo dividimos en 4 zonas: $\frac{395}{4} = 98.75 \approx 99 \quad articulos$
3. Los grupos de revistas con aproximadamente 99 artículos, los ubicamos utilizando la 4ta y 5ta columna (acumulado revistas y acumulado total de artículos) los grupos de artículos acumulados para cada zona serán: 101,196,293 y 395, cantidades que permite identificar las zonas:
   a. La primera zona tiene :101 artículos para 7 revistas
   b. La segunda zona tiene :196-101=95 artículos para 22-7=15 revistas
   c. La tercera zona tiene:293-196=97 artículos para 62-22=40 revistas
   d. La cuarta zona :395-293=102 artículos para 164-62=102 revistas
   Los resultados anteriores se presentan en el siguiente cuadro

Zonas	Total revistas	Total artículos
Zona-1 o núcleo: lugar donde se concentran las revistas más productivas en cuanto a la cantidad de artículos sobre un tema dado	7	101
Zona-2	15	95
Zona-3	40	97
Zona-4 :lugar donde se concentran las revistas menos productivas	102	102
Total	164	395

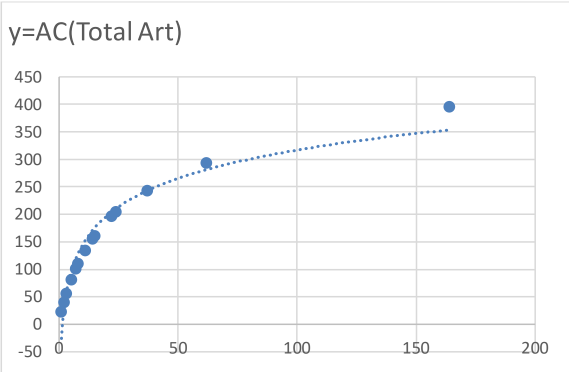
Multiplicador de Bradford=

$$k = \frac{\frac{15}{7} + \frac{40}{15} + \frac{102}{40}}{3} = 2.45$$

$7(2.45)^0, 7(2.45)^1, 7(2.45)^2, 7(2.45)^3$

$7, \quad 17, \quad 42, \quad 103$

Gráfica de acumuladas revistas (x) versus acumulados total de artículos (y)



X=AC(total revistas)

Ecuación de Brookes

$$y = b\ln(\frac{x}{a})$$

Linealizado la ecuación de Brookes, desarrollando.

$$y = b\ln(\frac{x}{a}) \rightarrow y = b(\ln x - \ln a) = b\ln x - b\ln a = (-b\ln a) + b\ln x$$

$y = c + bw$ Modelo lineal linealizado,

Dónde:

$$c = -b\ln a \rightarrow a = e^{-\frac{c}{b}} = \exp(-\frac{c}{b})$$

$$w = \ln x$$

Sus ecuaciones para calcular b y c son:

$$b = \frac{\sum wy - \frac{(\sum w)(\sum y)}{n}}{\sum w^2 - \frac{(\sum w)^2}{n}}$$

$$c = \frac{\sum y}{n} - b\frac{\sum w}{n}$$

Con los totales de la tabla (de la 4ta hasta la 8va columna)

$$b = 74.3899394$$

$$c = -25.7128939 \rightarrow a = e^{-(\frac{-25.7128939}{74.3899394})} = 1.412908227$$

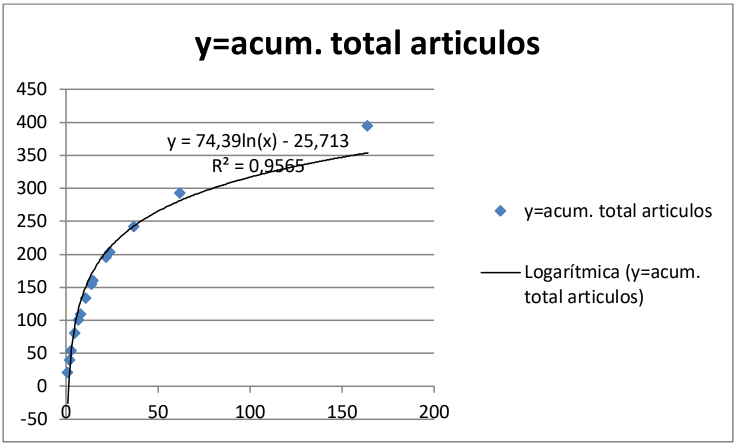
Luego la ecuación es:

$$y = 74.3899394\ln(\frac{x}{1.412908227})$$

Calculo del coeficiente de determinación

$$R^2 = (coeficientedePearson)^2 = (0.97801145)^2 = 0.9565 \rightarrow 95.65\%$$

Buen ajuste del modelo matemático de Brookes.



Ejercicios

Aplicación del modelo matemático de Bradford a la productividad de autores (datos tomados del informe profesional presentado por el Lic. Elías Reynaldo Barrenechea Gallardo)

Productividad de las revistas citadas en la
Revista Anales de la Facultad de Medicina

Revistas	Artículos
1	128
1	90
1	51
1	39
1	35
1	30
1	29
1	28
1	27
1	26
1	25
1	23
1	19
1	12
2	24
2	18
2	17
3	46
3	22
3	20
3	14
4	16
6	15
6	11
8	13
14	10
14	9
20	8
22	6
40	5
52	4
99	3
266	2
1116	1

Fuente: datos tomados del informe profesional presentado por el licenciado Elías Reynaldo Barrenechea Gallardo (Revista Anales de la Facultad de Medicina de la UNMSM)

a. Obtener las 4 zonas de Bradford

b. Calcular el multiplicador de Bradford

c. Estimar la ecuación de Brookes.

3.5 Ley de Zipf(1902-1950)

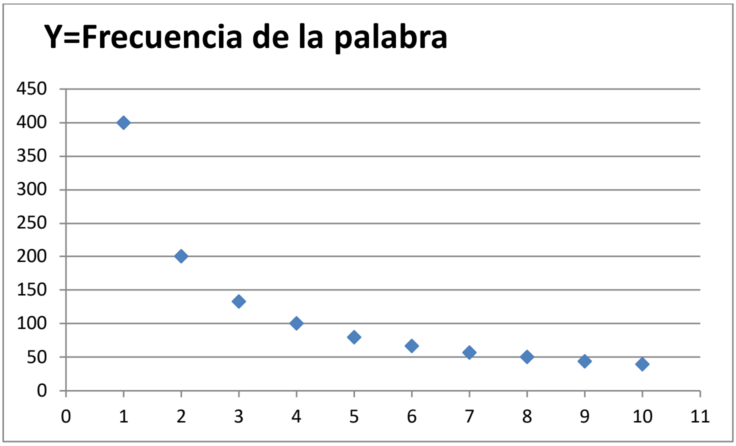
Enunciado

En todo documento que se formula en cualquier idioma, existen palabras cortas que ocurren con mayor frecuencia, en segundo lugar del ranking la palabra que se repite con mayor frecuencias es la mitad de la primera; para la tercera palabra que se repite es igual a la tercera parte de la primera, y así sucesivamente.

Esto se presenta en la siguiente tabla en forma teórica.

Palabras que se repiten en un documento o una conversación	X= Ranking de repetición de las palabras	Y=Frecuencias de las palabras
‘de	1	400
‘la	2	400(1/2)=200
‘en	3	400(1/3)=133
‘y	4	400(1/4)=100
‘para	5	400(1/5)=80
él	6	400(1/6)=67
´los	7	400(1/7)=57
‘las	8	400(1/8)=50
‘a	9	400(1/9)=44
‘por	10	400(1/10)=40

Gráfica



Modelo matemático de Zipf

$$X..Y = C \rightarrow Y = \frac{C}{X}$$

C= Valor constante
X=Ranking de las palabras que se repiten con la mayor frecuencia
Y=frecuencia de las palabras que se repiten.

Para un documento cualquiera se registran los datos (X ; Y), para estimar el modelo de Zipf, se linealiza aplicando logaritmo decimal..

$$Y = \frac{C}{X} \rightarrow \log y = \log(\frac{C}{x}) \rightarrow \log y = \log C + (-)(\log x)$$
$$Z = A + BW \rightarrow \mathrm{mod}\,elo \quad linealizado$$
$$Z = \log y$$
$$W = \log x$$
$$A = \log C \rightarrow C = 10^{A}$$
$$B = -1$$

Para que se cumpla la ley de Zipf, el coeficiente de determinación(la bondad de ajuste) debe ser próximo al 100%, y el valor de B debe aproximar por la derecha o por la izquierda al valor -1.

Ejemplo aplicativo.

X=Ranking	Y=frecuencias	W=LOG(X)	Z=LOG(Y)	WZ	WW
1	240	0	2.38021124	0	0
2	120	0.30103	2.07918125	0.62589592	0.09061906
3	80	0.47712125	1.90308999	0.90800468	0.22764469
4	60	0.60205999	1.77815125	1.07055373	0.36247623
5	48	0.69897	1.68124124	1.17513719	0.48855907
6	40	0.77815125	1.60205999	1.24664499	0.60551937
7	34	0.84509804	1.53147892	1.29424983	0.7141907
8	30	0.90308999	1.47712125	1.33397341	0.81557152
9	27	0.95424251	1.43136376	1.36586815	0.91057877
10	24	1	1.38021124	1.38021124	1
11	22	1.04139269	1.34242268	1.39798916	1.08449872
12	20	1.07918125	1.30103	1.40404717	1.16463216
13	18	1.11394335	1.25527251	1.39830246	1.24086979
14	17	1.14612804	1.23044892	1.41025201	1.31360947
15	16	1.17609126	1.20411998	1.41615499	1.38319065
16	15	1.20411998	1.17609126	1.41615499	1.44990493
17	14	1.23044892	1.14612804	1.41025201	1.51400455
18	13	1.25527251	1.11394335	1.39830246	1.57570906
19	13	1.2787536	1.11394335	1.42445907	1.63521077
20	12	1.30103	1.07918125	1.40404717	1.69267905
		18.3861246	**29.2066915**	**24.4805006**	**19.2694686**

$$B = \frac{(24.4805006 - \dfrac{(18.3861246)(29.3861246)}{20}}{19.2694686 - \dfrac{(18.3861246)^2}{20}} = -1.00101528$$

$$A = \frac{29.3861246}{20} - 1.00101528(\frac{18.3861246}{20}) = 2.38057416 \rightarrow C = 10^{2.38057416} = 240.2000638$$
$$R^2 = 99.98\%$$

Por las estadísticas calculadas los datos de la tabla, cumple con la ley de Zipf.

4.1. Indicadores bibliométricos

Son medidas que proporcionan información sobre los resultados de la producción científica. Existe herramientas bibliometricas tales como: Web Science, Scopus, Google académico y Dimensions, que proporcional tales indicadores.

Los indicadores bibliométricos más importantes son:

4.2. El número de citas recibidas por los artículos publicados en una revista

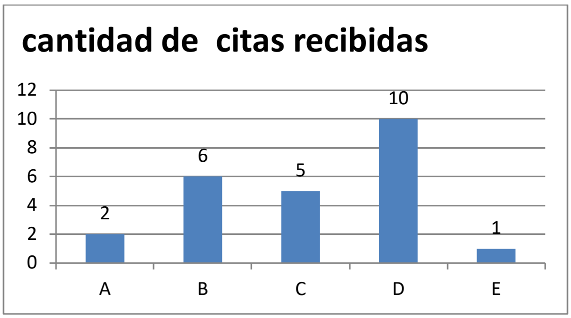
Su valor indica el número de veces que una revista o artículo publicado ha sido reconocido por otro autor. Su valor se obtiene haciendo un recuento de cuantas veces un artículo o revista fue citado en un trabajo de investigación y está publicado.

Ejemplo-1:

Un biólogo durante los años 2015 al 2022, ha publicado 5 artículos, el cuadro muestra la cantidad de citas que ha recibido por parte de otros profesionales.

Artículos publicados (trabajos de investigación)	cantidad de citas recibidas
A	2
B	6
C	5
D	10
E	1

Gráfica: cantidad de citas recibidas por sus trabajos publicados durante los años 2015 al 2022



El número de citas:

Mide cómo influye o cómo impacta un trabajo científico en la comunidad científica (consumidor).

El número de citas es un reconocimiento del trabajo realizado por un científico.

El número de citas se para determinar otros indicadores, tales como el factor de impacto (FI), índice h, índice i10, etc.

Aporte del número de citas

Indica la repercusión de una investigación en la comunidad científica.

Mide el interés y la influencia, aunque no siempre sea positiva.

Limitaciones

No debe utilizarse como un indicador de calidad

No debe usarse como una medida de reputación positiva para los investigadores individuales.

4.3. El índice H de un investigador.

El índice h o también llamado índice de Hirsch fue propuesto en el año 2005 por Jorge E Hirsch, físico de la universidad de California en San Diego.

Es un indicador numérico que evalúa la producción científica de un investigador, utilizando la cantidad de publicaciones con el número de citas recibidas.

Cálculo del índice h

a. En una tabla de frecuencias se colocan todas las publicaciones de un autor, con su correspondiente número de citas recibidas por sus trabajos.

b. Las publicaciones se ordenan tomando como referencia a las citas de mayor a menor (orden decreciente de las citas recibidas).

c. A las publicaciones del punto anterior, se enumera de menor a mayor, empezando por el número uno.

d. Se compara al número de citas con la numeración dada en el punto anterior, comparación que debe ser mayor o igual.

e. Cuando se encuentra una publicación cuyo número de citas sea inferior al número de orden; el índice h para el autor será el orden inmediatamente anterior.

Del ejemplo-1, calcular el índice h para el investigador e interpretarlo.

Artículos publicados a la fecha	Cantidad de citas recibidas a la fecha
A	2
B	6
C	5
D	10
E	1

Para su cálculo del índice h, de un autor, se ordenan sus artículos publicados por el autor, tomando en cuenta a la cantidad de citas recibida en forma decreciente. La cantidad de citas recibida debe ser mayor o igual al nuevo orden de los artículos, el primero de los artículos publicado que no cumple la condición, entonces el anterior orden de los artículos publicados es el índice h para el autor.

Artículos publicados ordenados tomando en cuenta a la cantidad de citas recibidas	cantidad de citas recibidas	criterio
D(1)	10	10>1er orden(Verdad)
B(2)	6	6> 2do orden(Verdad)
C(3)	5	5>3er orden(Verdad)
A(4)	2	2>=4to orden(Falso)
E(5)	1	1>=4to orden(Falso)

El índice h para el autor o investigador es igual a 3, significa que tres trabajos científicos publicados por el investigador (Artículos: D, B, C) han recibido 3 o más citas

Ventajas

Combina en un único indicador una medida de cantidad y otra de impacto de la producción científica.

Detecta los investigadores más destacados de un área de conocimiento.

Favorece las carreras científicas más dilatadas, investigadores genios.

Limitaciones

No es adecuado para comparar investigadores de diferentes áreas (biólogos publican más que los matemáticos)

Perjudica a los investigadores con actividad científica reciente.

Depende del número total de publicaciones. El índice h nunca podrá ser superior al número total de documentos.

Penaliza a los autores que publican menos, pero publican mejor

Notas:

Las bibliotecas universitarias, de hospitales y centro de investigación cuentan con profesionales que ayudan a:

Conocer las citas recibidas por su investigación.

Calcular el índice h.

Crear perfiles "profesionales" con indicador bibliométrico.
.
Asesorar sobre evaluación de la investigación, revistas en los que publican, artículos muy citados, comparativos, etc.

4.4. El índice i10 de un investigador.

Su valor que indica el número de trabajos científicos que un autor ha escrito y a la fecha recibió 10 a más citas. La base de datos que proporcionan este índice es Google Scholar.

Ejemplo

La tabla presenta 4 trabajos publicados por un investigador y las citas recibidas.

N° de trabajos científicos publicados a la fecha	N° de citas recibidas por los trabajos publicados a la fecha.
A	3
B	9
C	12
D	10

De la tabla el índice i10 para el autor es 2 (trabajos: C y D), existen 2 trabajos que recibieron 10 a más citas

4.5. Factor de impacto de una revista científica (FI)

El factor de impacto de una revista se utiliza para comparar la importancia de dicha revistas con otras, en el mismo campo científico. El factor de impacto de una revista en un año determinado, se calcula como el cociente del total de citas recibidas por las publicaciones en los dos años anteriores y el total de trabajo publicados en esos dos años. El resultado del cociente se interpreta como el promedio anual de citas de artículos por una revista:

.

$$FI_x = factor \quad de \quad impacto \quad por \quad la \quad revista \quad en \quad el \quad año \quad x$$

$$FI_x = \frac{Citas \quad recibidas \quad por \quad la \quad revistas \quad en \quad los \quad años(x-2, x-1)}{Total \quad de \quad ariculos \quad publicados \quad por \quad la \quad revista \quad en \quad los \quad años(x-2, x-1)}$$

Ejemplo

Calcular el factor de impacto de una determinada revista científica para el año 2020.

2 años anteriores al año 2020	Total de artículos publicados por la revista en los años 2018 y 2019	Total de citas recibidas en el 2020 por los artículos publicados en los años 2018 y 2019
2018	20	500
2019	50	800

$$FI_{2020} = \frac{Citas \quad recibidas \quad por \quad la \quad revistas \quad en \quad los \quad años(2018,2019)}{Total \quad de \quad ariculos \quad publicados \quad por \quad la \quad revista \quad en \quad los \quad años(2018,2019)}$$

$$FI_{2020} = \frac{500+800}{20+50} = \frac{1300}{70} = 18.57$$

Interpretación:

En promedio en el año 2020, la revista científica recibió 18.57 citas, por cada artículo publicado.

4.6. Índice de inmediatez de una revista (Ii)

Es un valor que nos muestra que tan rápido es citada una revista científica. Se calcula dividiendo el total de citas recibida por un artículo publicado en un año determinado y el total de artículos que público la revista en ese año.

$$Ii = \frac{Total \quad de \quad citas \quad que \quad reciben \quad en \quad un \quad año \quad los \quad articulos \quad publicados}{Total \quad de \quad articulos \quad publicados \quad por \quad esa \quad revista \quad en \quad ese \quad año}$$

Ejemplo

La revista científica ABC, tienen los siguiente datos con respecto al número de artículos que publico y al número de citas que ha recibido desde que fue publicada para el mismo año. Calcular el índice de inmediatez para cada año.

años	artículos publicados	citas recibidas	Índice de Inmediatez Ii
2000	150	80	0.53
2001	100	200	2.00
2002	120	120	1.00

$$Ii(2000)=\frac{80}{150}=0.53 \quad citas/articulos$$
$$Ii(2001)=\frac{200}{100}=2.0 \quad citas/articulos$$
$$Ii(2002)=\frac{120}{120}=1.0 \quad citas/articulos$$

Interpretación

Para el año 2000, el valor 0.53, indica que por cada 2 artículos publicados en las revistas ABC, se recibió 1 cita.

Para el año 2001, el valor 2, indica que por cada artículo publicado en la revista ABC, se recibió 2 cita.

Para el año 2002, por cada artículo publicado en la revista ABC, se recibió una cita.

4.7. Vida media de un trabajo de investigación (Vm)

La vida media de un documento publicado sobre un tema dado, se refiere al tiempo durante el cual fue publicada la mitad de la literatura activa o circulante. Para calcular la vida media toma en cuenta el tiempo de circulación de un documento y las citas recibidas. Se utiliza la fórmula:

$$Vm=k+\frac{0.5-a}{b-a}$$

k=número de años que toma acumular el 50% de las citas recibidas.
a=Proporción del número de citas acumuladas antes del 50% de citas recibidas.
b= Proporción del número de citas acumuladas después del 50% de citas recibidas.

Ejemplo

Calcular la vida media de un artículo publicado sobre matemática, entre los años 2000 al 2020.

Años	Citas recibidas	Acumuladas citas recibidas en forma decreciente
2000	0	302
2001	2	302
2002	8	300
2003	12	292
2004	26	280
2005	42	254
2006	30	212
2007	56	182
2008	70	126

Años	Citas recibidas	Acumuladas citas recibidas en forma decreciente
2009	15	56
2010	10	41
2011	7	31
2012	5	24
2013	4	19
2014	4	15
2015	3	11
2016	2	8
2017	2	6
2018	2	4
2019	1	2
2020	1	1
Total	302	

El artículo desde que se publicó 2000 al 2020 recibió un total de 302 citas, el 50% es 302/2=151, en la tabla este valor se busca en las frecuencias acumuladas de citas, encontrándose entre 126 menor número de citas y 182 mayor número de citas. El valor entero (k) es igual al número de años transcurrido desde el menor número de citas hasta el último año de estudio (2008 hasta 2020) el valor de k=13 años. El valor de a es igual al menor número de citas recibidas antes de 50%, dividido por el total de citas recibidas (302) y b es igual al mayor número de citas recibidas después del 50%, dividido por el total de citas (302), es decir,

$$a = \frac{126}{302} = 0.41721854 \quad y \quad b = \frac{182}{302} = 0.60264901$$

$$Vm = k + \frac{0.5 - a}{b - a} = 13 + \frac{(0.5 - 126/302)}{(182/302 - 126/302)} = 13.45 \quad años$$

Interpretación: 13.45 años, es la vida media del artículo publicado en el periodo de 2000 a 2020.

4.8. Cuartil

Es una medida de posición de una revista científica en comparación con todas las revistas las de sus áreas. Se obtiene al ordenar a las revistas de mayor a menor factor de impacto, al total de revistas ordenadas la dividimos en 4 partes iguales, cada una de estas partes será un cuartil. Las revistas con el factor de impacto más alto pertenecen al primer cuartil, los cuartiles medios serán el segundo y tercer cuartil y las revistas de bajo impacto pertenecen al cuarto cuartil.

4.9. Índice de colaboración (IC)

La colaboración científica se define como un proceso social de investigadores que se juntan con el propósito de realizar un nuevo trabajo de investigación, compartiendo sus recursos intelectuales y materiales. Se calcula, dividiendo el total de firmas (autores) por número de trabajos publicados en un período de estudio; para una tabla de frecuencias se tiene:

$$IC = \frac{\sum_{i=1}^{k} x_i . f_i}{n}$$

$x_i =$ Es el número de autores de un documento publicado.

$f_i =$ Es la cantidad de documentos publicados para el i-ésimo número de autores

$n =$ Es el total de documentos publicados en el período de estudio.

.

Se interpreta como el promedio de colaboradores en la formulación de documentos en el período en estudio.

La colaboración puede ser local, nacional e internacional

Para medir y mostrar las relaciones que se establecen entre los productores en la investigación puede aplicarse el análisis de las redes sociales científicas. Los indicadores de las redes nos permiten analizar el grado de conectividad de la red, número de interacciones de actores, así como el acercamiento entre éstos a través de sus interacciones. De estos indicadores se desprende que autores son más influyentes para establecer trabajos en colaboración.

4.10. El grado de colaboración (GC)

Su valor se interpreta como el aporte de cada grupo de autores, se calcula a través de la formula

$$GC_i = x_i \left(\frac{f_i}{n}\right)$$

4.11. Coeficiente de colaboración (CC)

$$CC = 1 - \frac{\sum_{i=1}^{k} \left(\frac{1}{x_i}\right) . f_i}{n}$$

Ejemplo aplicativo

Ejemplo durante los años 2010 al 2015, una Universidad publico 18 trabajos científicos, tal como se detalla en la siguiente tabla. Calcular el índice de colaboración, el grado de colaboración y coeficiente de colaboración de autores.

Tabla: Cantidad de autores y cantidad de trabajos publicados durante 2010-2015.

x_i = N° autores(colaboradores)	f_i =número de documentos	$x_i . f_i$	GC	$(\frac{1}{x_i}).f_1$
1	1	1	0.06	1.00
2	4	8	0.44	2.00
3	6	18	1.00	2.00
4	5	20	1.11	1.25
5	2	10	0.56	0.40
Total	18	57	3.17	6.65

Para el índice de colaboración IC:

$$IC = \frac{\sum_{i=1}^{k} x_i . f_i}{n} = \frac{57}{18} = 3.17 \quad autores$$

Durante los años 2010 al 2015 el promedio de colaboradores para los 18 documentos formulados es de 3.17 autores.

Calculo del grado de colaboración

$$GC = 1.(\frac{1_i}{18}) = 0.06 \rightarrow 6.0\%$$

6.0%, es el grado de participación para 1 autor

$$GC = 2.(\frac{4}{18}) = 0.44 \rightarrow 44.0\%$$

44.0%, es el grado de participación para 2 autores

$$GC = 3.(\frac{6}{18}) = 1.0 \rightarrow 100.0\%$$

100.0%, es el grado de participación para 3 autores

$$GC = 4.(\frac{5}{18}) = 1.111 \rightarrow 111.1\%$$

111.1%, es el grado de participación para 4 autores

$$GC = 5.(\frac{2}{18}) = 0.556 \rightarrow 55.6\%$$

55.6%, es el grado de participación para 5 autores

Calculo del coeficiente de colaboración

$$CC = 1 - \frac{\sum_{i=1}^{k} (\frac{1}{x_i}).f_i}{n} = 1 - \frac{6.65}{18} = 0.631$$

Bibliografía

ARDANUY, J. (2012). Breve Introducción a la Bibliometria. Barcelona: Universidad de Barcelona.

AMEZQUITA, J., ET AL (2011). Bibliometría, informetría y cienciometría. https://repositorio.unicartagena.edu.co/bitstream/handle/11227/245/modulo%20%20CTS%20No4-cienciometria.pdf?sequence=1

DAVILA, M. et al (2009). Bibliometría: conceptos y utilidades par el estudio médico y la formación profesional. https://www.redalyc.org/pdf/817/81712365011.pdf

EVANS, M., ROSENTHAL, J. (2014). Probabilidad y Estadística, La Ciencia de la Incertidumbre. Barcelona: Editorial Reverte.

GORBEA, P. (1996). El modelo Matemático de Bradford. México: Centro Universitario de Investigaciones bibliotecológicas.

GRANZEL, W. (2003). Bibliometrics as a research field. A course on theory and application of bibliometrics indicators. https://www.cin.ufpe.br/~ajhol/futuro/references/01%23_Bibliometrics_Module_KUL_BIBLIOMETRICS%20AS%20A%20RESEARCH%20FIELD.pdf

HINES, W. Y OTROS (2009). Probabilidad y Estadística para Ingeniería. México: Grupo Editorial Patria.

HOEL, P., BARBANCHO, A. (1976). Introducción a la estadística Matemática. Barcelona: Editorial Ariel.

LEVINE, D., KREHBIEL, T., BERENSON, M. (2014). Estadística para Administradores. México: Pearson Educación.

MARIN, J. (2008). Estadística Aplicada a las Ciencia de la Documentación. Madrid: Diego Marín.

MARTIN-PLIEGO, F., MONTERO, J., RUIZ-MAYA, L. (1998). Problemas de Probabilidad. Madrid: Paraninfo S.A.

MONTGOMERY, D., PECK, E., VINING, G. (2004). Introducción al Análisis de Regresión Lineal. México: Compañía Editorial Continental.

QUEZADA, N. (2017). Estadística con SPSS 24. Lima: Editorial MACRO.

RODRIGUEZ, L., TOMEO, V., JUAREZ, I., (2011). Métodos Estadísticos para Ingeniería. Madrid: Garceta grupo editorial.

URBIZAGASTEGUI, R. (2014). La bibliometría en el Perú. http://revista.letras.unmsm.edu.pe/index.php/le/article/view/269/265

URBIZAGASTEGUI, R. (2022). Bibliometría para bibliómetras. https://www.revistaotlet.com/bibliometria-para-bibliometras/

URBIZAGASTEGUI, R. (2016). El crecimiento de la literatura sobre la ley Bradford. http://rev-ib.unam.mx/ib/index.php/ib/article/view/54595/48539

URBIZAGASTEGUI, R. (2011). La ley Zipf y el punto de transición de Goffman en la indización automática.
http://rev-ib.unam.mx/ib/index.php/ib/article/view/27482/25470

URBIZAGASTEGUI, R. (2022). La productividad científica de los autores: un modelo de aplicación de la ley de lotka por el método del poder inverso generalizado. http://revistascientificas.filo.uba.ar/index.php/ICS/article/view/904/881

WALPOLE, R. Y OTROS (2012). Probabilidad y Estadística para ingeniería y ciencias. México: Pearson Educación.

Anexos-1

1. Demostrar que: $\log_b M.N = \log_b M + \log_b N$

   Sean dos números:

   $$M > 0 \quad , \quad N > 0 \quad \wedge \quad b \neq 1 > 0 \rightarrow$$
   $$\log_b M = x \rightarrow b^x = M$$
   $$\log_b N = y \rightarrow b^y = N$$
   $$b^{x+y} = M.N \rightarrow \log_b M.N = x + y \rightarrow \log_b M.N = \log_b M + \log_b N \quad demostrado$$

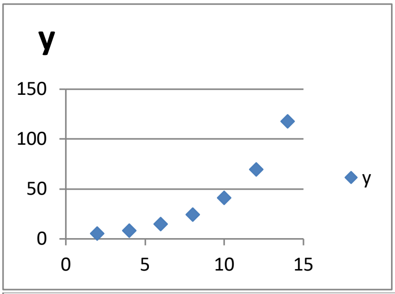
2. Demostrar que: $\log_b \log_a N = x \rightarrow N = a^{b^x}$

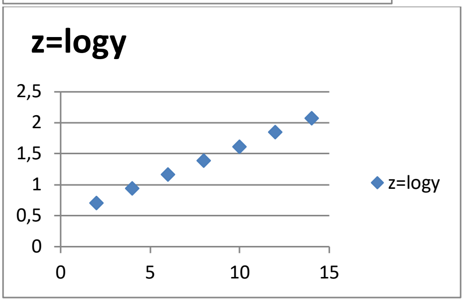
   $$\log_b \log_a N = x \rightarrow b^x = \log_a N \rightarrow N = a^{b^x} \quad demostrado$$

3. Gráfica de un modelo no lineal linealizado

   $$y = ab^x \leftarrow \text{mod}\,elo \quad no \quad lineal$$
   $$\log y = \log a + (\log b)x \quad z = A + Bx \leftarrow \text{mod}\,elo \quad no \quad lineal \quad linealizado$$

Modelo No lineal Modelo linealizado





4. Estimación de los parámetros $\beta_o \quad y \quad \beta_1$, por mínimo cuadrados ordinarios (MCO)

   Sea el modelo de regresion lineal

$$y_i = \beta_o + \beta_1 x_i + \varepsilon_i$$

Para estimar los parámetros $\beta_o \quad y \quad \beta_1$, se extrae una muestra aleatoria de tamaño n y se toman las medidas $(x_i; y_i) \quad i = 1,2,3,...,n$

Sea el modelo estimado $y_i = b_o + b_1 x_i + e_i$. Siendo $b_o \quad estima \quad \beta_0; \quad b_1 \quad estima \quad \beta_1 \quad y \quad e_i \quad estima \quad \varepsilon_i$, el método por mínimo cuadrados ordinarios (MCO) consiste en minimizar la suma de cuadrados de los residuos:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = f(b_0, b_1)$$

$$\frac{\partial f(b_0, b_1)}{\partial b_0} = 0 \rightarrow b_0 = \bar{y} - b_1 \bar{x}$$

$$\frac{\partial f(b_0, b_1)}{\partial b_1} = 0 \rightarrow b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

5. Coeficiente de determinación R^2

$$y_i = b_0 + b_1 x_i + e_i \rightarrow y_i = \bar{y} - b_1 \bar{x} + b_1 x_i + e_i \rightarrow y_i - \bar{y} = b_1 (x_i - \bar{x}) + e_i$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} e_i^2 \quad dividiendo \quad por \quad \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$1 = \frac{b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} + \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$Sea \quad R^2 = \frac{b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \rightarrow R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$Si \quad \sum_{i=1}^{n} e_i^2 \approx 0 \rightarrow R^2 \approx 1 \rightarrow un \quad buen \quad ajuste \quad del \quad modelo \quad estimado$$

6. Demostrar que $R^2 = [r]^2$

$$Como \quad R^2 = \frac{b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right)^2 . \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}\right) =$$

$$R^2 = \frac{\left(\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right)^2}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^2} . \left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}\right) = \frac{\left(\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} . \left(\frac{1}{\sum_{i=1}^{n} (y_i - \bar{y})^2}\right)$$

$$R^2 = \frac{\left(\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} . \left(\frac{1}{\sum_{i=1}^{n} (y_i - \bar{y})^2}\right) = \left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}\right]^2 = [r]^2 \quad demostrado$$

Anexo-2

Las siguientes tablas estadísticas han sido obtenidas por el autor, mediante el paquete estadístico MINITAB Y MICROSOFT EXCEL

1. NORMAL ESTANDAR

2. T-STUDENTS

PROBABILIDADES ACUMULADAS DE LA DISTRIBUCION NORMAL ESTANDAR
(PARA VALORES POSITIVOS DE Z)

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8079	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,00	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,10	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,20	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,30	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,40	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,50	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,60	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,70	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,80	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	1,0000
3,90	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,00	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Profesor: Fausto Matos Uribe de la Universidad Nacional Mayor de San Marcos

PROBABILIDADES ACUMULADAS DE LA DISTRIBUCION NORMAL ESTANDAR
(PARA VALORES NEGATIVOS DE Z)

-0,0900	-0,0800	-0,0700	-0,0600	-0,0500	-0,0400	-0,0300	-0,0200	-0,0100	0,0000	Z
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	-4,0
0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	-3,9
0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	-3,8
0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	-3,7
0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	-3,6
0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	-3,5
0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	-3,4
0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005	0,0005	-3,3
0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007	0,0007	-3,2
0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009	0,0010	-3,1
0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013	0,0014	-3,0
0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018	0,0019	-2,9
0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025	0,0026	-2,8
0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034	0,0035	-2,7
0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045	0,0047	-2,6
0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060	0,0062	-2,5
0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080	0,0082	-2,4
0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104	0,0107	-2,3
0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136	0,0139	-2,2
0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174	0,0179	-2,1
0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222	0,0228	-2,0
0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281	0,0287	-1,9
0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351	0,0359	-1,8
0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436	0,0446	-1,7
0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537	0,0548	-1,6
0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655	0,0668	-1,5
0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793	0,0808	-1,4
0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934	0,0951	0,0968	-1,3
0,0985	0,1003	0,1020	0,1038	0,1057	0,1075	0,1093	0,1112	0,1131	0,1151	-1,2
0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314	0,1335	0,1357	-1,1
0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539	0,1562	0,1587	-1,0
0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788	0,1814	0,1841	-0,9
0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061	0,2090	0,2119	-0,8
0,2148	0,2177	0,2207	0,2236	0,2266	0,2297	0,2327	0,2358	0,2389	0,2420	-0,7
0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676	0,2709	0,2743	-0,6
0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015	0,3050	0,3085	-0,5
0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372	0,3409	0,3446	-0,4
0,3483	0,3520	0,3557	0,3594	0,3632	0,3669	0,3707	0,3745	0,3783	0,3821	-0,3
0,3859	0,3897	0,3936	0,3974	0,4013	0,4052	0,4090	0,4129	0,4168	0,4207	-0,2
0,4247	0,4286	0,4325	0,4364	0,4404	0,4443	0,4483	0,4522	0,4562	0,4602	-0,1
0,4641	0,4681	0,4721	0,4761	0,4801	0,4840	0,4880	0,4920	0,4960	0,5000	0,0

Profesor: Fausto Francisco Matos Uribe
Universidad Nacional Mayor de San Marcos.

	PROBABILIDADES ACUMULADAS DE LA DISTRIBUCION T-STUDENTS													
GL.	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95	0.975	0.99	0.995
1	-63.6567	-31.8205	-12.7062	-6.3138	-3.0777	-1.3764	-0.7265	0.7265	1.3764	3.0777	6.3138	12.7062	31.8205	63.6567
2	-9.92484	-6.96456	-4.30265	-2.91999	-1.88562	-1.06066	-0.61721	0.61721	1.06066	1.88562	2.91999	4.30265	6.96456	9.92484
3	-5.84091	-4.5407	-3.18245	-2.35336	-1.63774	-0.97847	-0.58439	0.58439	0.97847	1.63774	2.35336	3.18245	4.5407	5.84091
4	-4.6041	-3.74695	-2.77645	-2.13187	-1.53321	-0.94096	-0.56865	0.56865	0.94096	1.53321	2.13187	2.77645	3.74695	4.6041
5	-4.03215	-3.36493	-2.57058	-2.01505	-1.47592	-0.91954	-0.55943	0.55943	0.91954	1.47592	2.01505	2.57058	3.36493	4.03215
6	-3.70745	-3.14267	-2.44691	-1.94318	-1.43976	-0.9057	-0.55338	0.55338	0.9057	1.43976	1.94318	2.44691	3.14267	3.70745
7	-3.49948	-2.99795	-2.36463	-1.89459	-1.41492	-0.89603	-0.54911	0.54911	0.89603	1.41492	1.89459	2.36463	2.99795	3.49948
8	-3.35539	-2.89646	-2.306	-1.85955	-1.39682	-0.88889	-0.54593	0.54593	0.88889	1.39682	1.85955	2.306	2.89646	3.35539
9	-3.24985	-2.82144	-2.26216	-1.83311	-1.38304	-0.8834	-0.54348	0.54348	0.8834	1.38304	1.83311	2.26216	2.82144	3.24985
10	-3.16928	-2.76378	-2.22814	-1.81246	-1.3722	-0.87906	-0.54153	0.54153	0.87906	1.3722	1.81246	2.22814	2.76378	3.16928
11	-3.10582	-2.71808	-2.20099	-1.79589	-1.36343	-0.87555	-0.53994	0.53994	0.87555	1.36343	1.79589	2.20099	2.71808	3.10582
12	-3.05456	-2.681	-2.17882	-1.78229	-1.35622	-0.87261	-0.53862	0.53862	0.87261	1.35622	1.78229	2.17882	2.681	3.05456
13	-3.01228	-2.65031	-2.16037	-1.77093	-1.35017	-0.87015	-0.5375	0.5375	0.87015	1.35017	1.77093	2.16037	2.65031	3.01228
14	-2.97686	-2.6245	-2.1448	-1.76131	-1.34503	-0.86805	-0.53655	0.53655	0.86805	1.34503	1.76131	2.1448	2.6245	2.97686
15	-2.94672	-2.60248	-2.13145	-1.75305	-1.34061	-0.86625	-0.53573	0.53573	0.86625	1.34061	1.75305	2.13145	2.60248	2.94672
16	-2.92078	-2.58349	-2.11991	-1.74589	-1.33679	-0.86467	-0.53501	0.53501	0.86467	1.33679	1.74589	2.11991	2.58349	2.92078
17	-2.89824	-2.56694	-2.10982	-1.73961	-1.3334	-0.86328	-0.53438	0.53438	0.86328	1.3334	1.73961	2.10982	2.56694	2.89824
18	-2.87844	-2.55238	-2.10093	-1.73407	-1.3304	-0.86205	-0.53382	0.53382	0.86205	1.3304	1.73407	2.10093	2.55238	2.87844
19	-2.86095	-2.53948	-2.09303	-1.72914	-1.32773	-0.86095	-0.53331	0.53331	0.86095	1.32773	1.72914	2.09303	2.53948	2.86095
20	-2.84534	-2.52798	-2.08598	-1.72473	-1.32534	-0.85997	-0.53286	0.53286	0.85997	1.32534	1.72473	2.08598	2.52798	2.84534
21	-2.83137	-2.51765	-2.07963	-1.72075	-1.3232	-0.85908	-0.53246	0.53246	0.85908	1.3232	1.72075	2.07963	2.51765	2.83137
22	-2.81878	-2.50832	-2.0739	-1.71715	-1.32124	-0.85828	-0.53209	0.53209	0.85828	1.32124	1.71715	2.0739	2.50832	2.81878
23	-2.80734	-2.49987	-2.06866	-1.71388	-1.31946	-0.85755	-0.53175	0.53175	0.85755	1.31946	1.71388	2.06866	2.49987	2.80734
24	-2.79694	-2.49217	-2.0639	-1.71089	-1.31784	-0.85686	-0.53144	0.53144	0.85686	1.31784	1.71089	2.0639	2.49217	2.79694
25	-2.78744	-2.48511	-2.05954	-1.70814	-1.31635	-0.85624	-0.53115	0.53115	0.85624	1.31635	1.70814	2.05954	2.48511	2.78744
26	-2.77871	-2.47863	-2.05553	-1.70562	-1.31499	-0.85567	-0.53089	0.53089	0.85567	1.31499	1.70562	2.05553	2.47863	2.77871
27	-2.77068	-2.47266	-2.05183	-1.70331	-1.3137	-0.85514	-0.53065	0.53065	0.85514	1.3137	1.70331	2.05183	2.47266	2.77068
28	-2.76326	-2.46714	-2.04841	-1.70113	-1.31253	-0.85465	-0.53043	0.53043	0.85465	1.31253	1.70113	2.04841	2.46714	2.76326
29	-2.75639	-2.46202	-2.04523	-1.69914	-1.31143	-0.85419	-0.53023	0.53023	0.85419	1.31143	1.69914	2.04523	2.46202	2.75639
30	-2.75000	-2.45726	-2.04227	-1.69726	-1.31042	-0.85377	-0.53002	0.53002	0.85377	1.31042	1.69726	2.04227	2.45726	2.75
31	-2.74406	-2.45282	-2.03951	-1.69552	-1.30946	-0.85337	-0.52984	0.52984	0.85337	1.30946	1.69552	2.03951	2.45282	2.74406
32	-2.73848	-2.44869	-2.03693	-1.6939	-1.30857	-0.853	-0.52967	0.52967	0.853	1.30857	1.6939	2.03693	2.44869	2.73848
33	-2.73328	-2.4448	-2.03453	-1.69236	-1.30774	-0.85265	-0.5295	0.5295	0.85265	1.30774	1.69236	2.03453	2.4448	2.73328

	PROBABILIDADES ACUMULADAS DE LA DISTRIBUCION T-STUDENTS													
GL.	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95	0.975	0.99	0.995
34	-2.72839	-2.44116	-2.03225	-1.69092	-1.30695	-0.85232	-0.52935	0.52935	0.85232	1.30695	1.69092	2.03225	2.44116	2.72839
35	-2.72381	-2.43774	-2.03011	-1.68957	-1.30621	-0.85201	-0.52921	0.52921	0.85201	1.30621	1.68957	2.03011	2.43774	2.72381
36	-2.71948	-2.4345	-2.02809	-1.6883	-1.30552	-0.85172	-0.52908	0.52908	0.85172	1.30552	1.6883	2.02809	2.4345	2.71948
37	-2.71541	-2.43145	-2.0262	-1.68709	-1.30486	-0.85144	-0.52895	0.52895	0.85144	1.30486	1.68709	2.0262	2.43145	2.71541
38	-2.71156	-2.42857	-2.02439	-1.68596	-1.30424	-0.85118	-0.52883	0.52883	0.85118	1.30424	1.68596	2.02439	2.42857	2.71156
39	-2.70791	-2.42585	-2.02269	-1.68488	-1.30364	-0.85094	-0.52871	0.52871	0.85094	1.30364	1.68488	2.02269	2.42585	2.70791
40	-2.70446	-2.42326	-2.02108	-1.68385	-1.30308	-0.8507	-0.52861	0.52861	0.8507	1.30308	1.68385	2.02108	2.42326	2.70446
41	-2.70118	-2.4208	-2.01954	-1.68288	-1.30254	-0.85048	-0.5285	0.5285	0.85048	1.30254	1.68288	2.01954	2.4208	2.70118
42	-2.69807	-2.41847	-2.01808	-1.68195	-1.30204	-0.85026	-0.5284	0.5284	0.85026	1.30204	1.68195	2.01808	2.41847	2.69807
43	-2.6951	-2.41625	-2.01669	-1.68107	-1.30155	-0.85006	-0.52831	0.52831	0.85006	1.30155	1.68107	2.01669	2.41625	2.6951
44	-2.69228	-2.41413	-2.01537	-1.68023	-1.30109	-0.84987	-0.52822	0.52822	0.84987	1.30109	1.68023	2.01537	2.41413	2.69228
45	-2.68959	-2.41212	-2.0141	-1.67943	-1.30065	-0.84968	-0.52814	0.52814	0.84968	1.30065	1.67943	2.0141	2.41212	2.68959
46	-2.68701	-2.41019	-2.0129	-1.67866	-1.30023	-0.84951	-0.52805	0.52805	0.84951	1.30023	1.67866	2.0129	2.41019	2.68701
47	-2.68456	-2.40835	-2.01174	-1.67793	-1.29982	-0.84934	-0.52798	0.52798	0.84934	1.29982	1.67793	2.01174	2.40835	2.68456
48	-2.6822	-2.40658	-2.01063	-1.67722	-1.29944	-0.84917	-0.5279	0.5279	0.84917	1.29944	1.67722	2.01063	2.40658	2.6822
49	-2.67995	-2.40489	-2.00958	-1.67655	-1.29907	-0.84902	-0.52783	0.52783	0.84902	1.29907	1.67655	2.00958	2.40489	2.67995
50	-2.67779	-2.40327	-2.00856	-1.67591	-1.29871	-0.84887	-0.52776	0.52776	0.84887	1.29871	1.67591	2.00856	2.40327	2.67779
51	-2.67572	-2.40172	-2.00758	-1.67528	-1.29837	-0.84873	-0.52769	0.52769	0.84873	1.29837	1.67528	2.00758	2.40172	2.67572
52	-2.67373	-2.40022	-2.00665	-1.67469	-1.29805	-0.84859	-0.52763	0.52763	0.84859	1.29805	1.67469	2.00665	2.40022	2.67373
53	-2.67182	-2.39879	-2.00575	-1.67412	-1.29773	-0.84846	-0.52757	0.52757	0.84846	1.29773	1.67412	2.00575	2.39879	2.67182
54	-2.66998	-2.39741	-2.00488	-1.67356	-1.29743	-0.84833	-0.52751	0.52751	0.84833	1.29743	1.67356	2.00488	2.39741	2.66998
55	-2.66822	-2.39608	-2.00404	-1.67303	-1.29713	-0.84821	-0.52745	0.52745	0.84821	1.29713	1.67303	2.00404	2.39608	2.66822
60	-2.66028	-2.39012	-2.0003	-1.67065	-1.29582	-0.84765	-0.5272	0.5272	0.84765	1.29582	1.67065	2.0003	2.39012	2.66028
65	-2.6536	-2.3851	-1.99714	-1.66864	-1.29471	-0.84719	-0.52698	0.52698	0.84719	1.29471	1.66864	1.99714	2.3851	2.6536
70	-2.6479	-2.38081	-1.99444	-1.66691	-1.29376	-0.84679	-0.5268	0.5268	0.84679	1.29376	1.66691	1.99444	2.38081	2.6479
75	-2.64298	-2.3771	-1.9921	-1.66543	-1.29294	-0.84644	-0.52664	0.52664	0.84644	1.29294	1.66543	1.9921	2.3771	2.64298
80	-2.63869	-2.37387	-1.99006	-1.66412	-1.29222	-0.84614	-0.5265	0.5265	0.84614	1.29222	1.66412	1.99006	2.37387	2.63869
85	-2.63491	-2.37102	-1.98827	-1.66298	-1.29159	-0.84587	-0.52637	0.52637	0.84587	1.29159	1.66298	1.98827	2.37102	2.63491
90	-2.63157	-2.3685	-1.98667	-1.66196	-1.29103	-0.84563	-0.52626	0.52626	0.84563	1.29103	1.66196	1.98667	2.3685	2.63157
95	-2.62858	-2.36624	-1.98525	-1.66105	-1.29053	-0.84542	-0.52616	0.52616	0.84542	1.29053	1.66105	1.98525	2.36624	2.62858
100	-2.62589	-2.36422	-1.98397	-1.66023	-1.29007	-0.84523	-0.52608	0.52608	0.84523	1.29007	1.66023	1.98397	2.36422	2.62589
150	-2.609	-2.35146	-1.97591	-1.65508	-1.28722	-0.84402	-0.52552	0.52552	0.84402	1.28722	1.65508	1.97591	2.35146	2.609
200	-2.60063	-2.34514	-1.9719	-1.65251	-1.2858	-0.84342	-0.52524	0.52524	0.84342	1.2858	1.65251	1.9719	2.34514	2.60063
250	-2.59564	-2.34136	-1.9695	-1.65097	-1.28495	-0.84306	-0.52507	0.52507	0.84306	1.28495	1.65097	1.9695	2.34136	2.59564

	PROBABILIDADES ACUMULADAS DE LA DISTRIBUCION T-STUDENTS													
GL.	0.005	0.01	0.025	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95	0.975	0.99	0.995
300	-2.59232	-2.33884	-1.9679	-1.64995	-1.28438	-0.84282	-0.52496	0.52496	0.84282	1.28438	1.64995	1.9679	2.33884	2.59232
400	-2.58818	-2.33571	-1.96591	-1.64867	-1.28367	-0.84252	-0.52482	0.52482	0.84252	1.28367	1.64867	1.96591	2.33571	2.58818
500	-2.5857	-2.33383	-1.96472	-1.64791	-1.28325	-0.84234	-0.52473	0.52473	0.84234	1.28325	1.64791	1.96472	2.33383	2.5857
600	-2.58405	-2.33258	-1.96393	-1.6474	-1.28296	-0.84222	-0.52468	0.52468	0.84222	1.28296	1.6474	1.96393	2.33258	2.58405
700	-2.58287	-2.33169	-1.96336	-1.64703	-1.28276	-0.84214	-0.52464	0.52464	0.84214	1.28276	1.64703	1.96336	2.33169	2.58287
800	-2.58199	-2.33102	-1.96293	-1.64676	-1.28261	-0.84207	-0.52461	0.52461	0.84207	1.28261	1.64676	1.96293	2.33102	2.58199
900	-2.5813	-2.3305	-1.9626	-1.64655	-1.28249	-0.84202	-0.52459	0.52459	0.84202	1.28249	1.64655	1.9626	2.3305	2.5813
1000	-2.58075	-2.33008	-1.96234	-1.64638	-1.2824	-0.84198	-0.52457	0.52457	0.84198	1.2824	1.64638	1.96234	2.33008	2.58075
2000	-2.57829	-2.32821	-1.96115	-1.64562	-1.28198	-0.8418	-0.52448	0.52448	0.8418	1.28198	1.64562	1.96115	2.32821	2.57829