

# Notas de clase de Probabilidad y Estadística

## Volumen 5: Cap. 6 (Regresión lineal simple)

Versión 2 (Mayo 21, 2020)

### **Dr. rer. nat. Humberto LLinás Solano**

Doctor en Estadística (Mainz-Alemania)

Profesor Titular/Investigador Asociado

hllinas@uninorte.edu.co

Departamento de Matemáticas y Estadística

**Universidad del Norte**

([www.uninorte.edu.co](http://www.uninorte.edu.co)).



# ÍNDICE GENERAL

## **PREFACIO** **PÁGINA 3**

Introducción	3
El autor	3

## **6**

## **EL MODELO DE REGRESIÓN LINEAL SIMPLE** **PÁGINA 5**

6.1	Preliminares	5
6.2	El modelo de regresión lineal simple	7
6.3	Supuestos básicos para el modelo de regresión lineal	7
6.3.1	Supuestos	7
6.3.2	Comentarios	8
6.4	Estimación de los parámetros del modelo	10
6.5	Recursos online	11
6.6	Resumen de notaciones	11
6.7	Intervalos de confianza	13
6.8	Pruebas de hipótesis para la pendiente	14
6.8.1	Contraste para la pendiente de la regresión poblacional usando la prueba $t$	14
6.8.2	Contraste para la pendiente de la regresión poblacional usando el procedimiento del análisis de varianza	14
6.9	Pruebas de hipótesis para el intersección	15
6.10	Inferencias para la correlación poblacional	15
6.10.1	Contrastes para la correlación poblacional nula	15
6.10.2	Contrastes para la correlación poblacional (más general)	15
6.10.3	Intervalo de confianza para la correlación poblacional	16
6.11	Ejercicios	18

## **A**

## **APÉNDICE DE TABLAS** **PÁGINA 23**

1.	Distribución binomial	23
2.	Distribución de Poisson	26
3.	Distribución normal estándar	27
4.	Distribución $t$ de Student	29
5.	Distribución chi-cuadrada	30
6.	Distribución $F$ de Fisher	32
7.	Algunas distribuciones discretas	36

8. Algunas distribuciones continuas	36
9. Resumen de distribuciones muestrales e intervalos de confianza	37

**B****GUÍA RÁPIDA PARA TRABAJAR CON STATGRAPHICS** **PÁGINA 41**

B.1	Análisis de un solo conjunto de datos	41
B.2	Análisis simultáneo de dos o más conjuntos de datos	41
B.3	Gráficos de dispersión	42
B.4	Diagramas de presentación	42
B.5	Variables numéricas multidimensionales	43
B.6	Distribuciones de probabilidad	43
B.7	Inferencias basadas en una sola muestra	43
B.8	Inferencias basadas en dos muestras	44
B.9	Bondad de ajuste	44

**C****GUÍA RÁPIDA PARA TRABAJAR CON SPSS** **PÁGINA 45**

C.1	Definición de las variables	45
C.1.1	Transformación de una variable	46
C.1.2	Recodificación de una Variable	47
C.1.3	Filtrado de datos	47
C.2	Análisis exploratorio de datos	48
C.3	Inferencia sobre una o más poblaciones	49

**D****USO DE LA CALCULADORA EN LA ESTADÍSTICA** **PÁGINA 51****BIBLIOGRAFÍA & REFERENCIAS** **PÁGINA 53**



# Prefacio

## Introducción

---

Estas notas de clase hacen parte de un compendio de varios volúmenes y están dirigido a todo tipo de público que requiere de algún conocimiento básico en Estadística.

## El autor

---

Humberto Jesús Llinás Solano es Licenciado en Ciencias de la Educación, con énfasis en Matemáticas, Física y Estadística de la Universidad del Atlántico (Colombia). Magister en Matemáticas, convenio Universidad del Valle-Universidad del Norte (Colombia). Doctor en Estadística (Dr. rer. nat.) de la Universidad Johannes Gutenberg de Mainz (Alemania). Desde 1998 se desempeña como profesor de tiempo completo de la Universidad del Norte y forma parte de los grupos de investigación Matemáticas y Enfermedades tropicales de dicha institución. Autor de los productos<sup>1</sup>:

- *Estadística descriptiva y distribuciones de probabilidad* (2005, [6])
- *Estadística inferencial* (2006, [8])
- *Una visión histórica del concepto moderno de integral* (como editor, 2006, [4])
- *Medida e integración* (2007, [9])
- *Applets de estadística* (2007, [11])
- *Introducción a la estadística con aplicaciones en Ciencias Sociales* (2012, [12])
- *Procesos estocásticos con aplicaciones* (como coautor, 2013, [2])
- *Introducción a la estadística matemática* (2014, [13])
- *Introducción a la teoría de la probabilidad* (2014, [14])

**Para más referencias, pueden consultarse:**

- [CVLAC](#)
- [ORCID](#)
- [Google Scholar](#)

---

<sup>1</sup>Se cita el título del texto o applet, el año de publicación y la referencia bibliográfica respectiva. Cuando sea necesario, un comentario adicional.



# 6

## El modelo de regresión lineal simple

### 6.1 Preliminares

---

1. La relación más sencilla entre dos variables  $x$  y  $y$  es una relación lineal  $Y = \delta + \beta x$ .
2. Pendiente  $\beta$ . item  $\delta$  punto de corte con el eje  $Y$ .
3. Calculadora de ecuaciones de la recta.
4.  $x$  se llama VARIABLE INDEPENDIENTE (PREDICTORA O EXPLICATIVA).
5. Para  $x$  fija,  $y$  se llama VARIABLE DEPENDIENTE O DE RESPUESTA.
6. Por lo general, las observaciones se efectuarán para diversos valores  $x_1, \dots, x_n$  de la variable independiente  $x$ . Representaremos con  $Y_i$  y  $y_i$  la variable aleatoria y el valor asociado con  $x_i$ . Entonces, la información disponible está formada por las  $n$  parejas  $(x_i, y_i)$ ,  $i = 1, \dots, n$ .
7. Un primer paso en el análisis de regresión simple es trazar un diagrama de dispersión de los datos. En ese diagrama, cada pareja  $(x_i, y_i)$  es un punto ubicado en un sistema coordenado bidimensional.



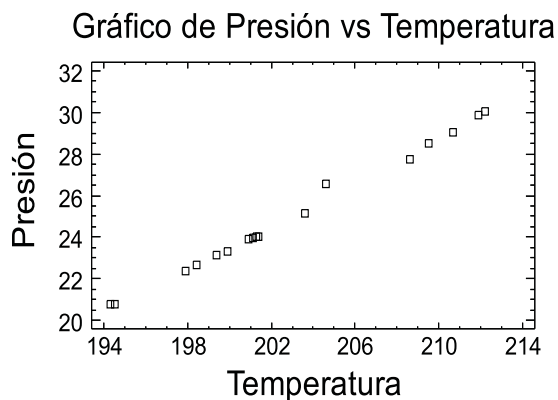
**Ejemplo 6.1**

Durante las décadas de los cuarenta y cincuenta del siglo XIX, el físico escocés J. D. Forbes realizó estudios con los que pretendía determinar la altitud sobre el nivel del mar de cumbres montañosas. Para realizar este estudio, Forbes tomó medidas de la presión y la temperatura de ebullición del agua en diferentes ubicaciones de los Alpes y Escocia como se muestra en la tabla de abajo (las medidas de presión fueron registradas en pulgadas de mercurio y las de temperatura las registró en grados Fahrenheit).

Observación $i$	Temperatura $x_i$	Presión $y_i$
1	194,5	20,79
2	194,3	20,79
3	197,9	22,4
4	198,4	22,67
5	199,4	23,15
6	199,9	23,35
7	200,9	23,89
8	201,1	23,99
9	201,4	24,02
10	201,3	24,01
11	203,6	25,14
12	204,6	26,57
13	209,5	28,49
14	208,6	27,76
15	210,7	29,04
16	211,9	29,88
17	212,2	30,06

Así,  $(x_1, y_1) = (194,5; 20,79)$ ,  $(x_2, y_2) = (194,3; 20,79)$ , etc. En la figura 6.1 mostramos el diagrama de dispersión para estos datos. Observemos lo siguiente, en los datos y en el diagrama:

- Algunas observaciones tienen valores  $x$  idénticos, pero valores diferentes de  $y$  (compárese con las dos primeras observaciones).
- Hay una fuerte tendencia de  $y$  a aumentar cuando aumenta  $x$ . Esto es, los valores mayores de presión tienden a estar asociados con valores de mayor temperatura.
- Al parecer se podría pronosticar el valor de  $y$  a partir del valor de  $x$  si determináramos una línea recta que pasara de manera razonable cerca de los puntos en la gráfica. En otras palabras, hay evidencia de que una apreciable (aunque no perfecta) relación lineal entre las dos variables. ◀



**Figura 6.1:** Diagrama de dispersión para los datos del ejemplo 6.1

## 6.2 El modelo de regresión lineal simple

### Definición 6.1

Supongamos que estamos interesados en conocer la relación entre una variable dependiente  $Y$  y una independiente  $X$ . Si la variable aleatoria toma los valores  $x_i$ , entonces, la ecuación de la RECTA VERDADERA (o POBLACIONAL) DE REGRESIÓN expresa los correspondientes valores  $Y_i$  como

$$Y_i = \delta + \beta x_i + \epsilon_i,$$

donde  $\delta$  y  $\beta$  son constantes y  $\epsilon_i$ , llamado TÉRMINO DE ERROR, es una variable aleatoria con media 0.

## 6.3 Supuestos básicos para el modelo de regresión lineal

### 6.3.1. Supuestos

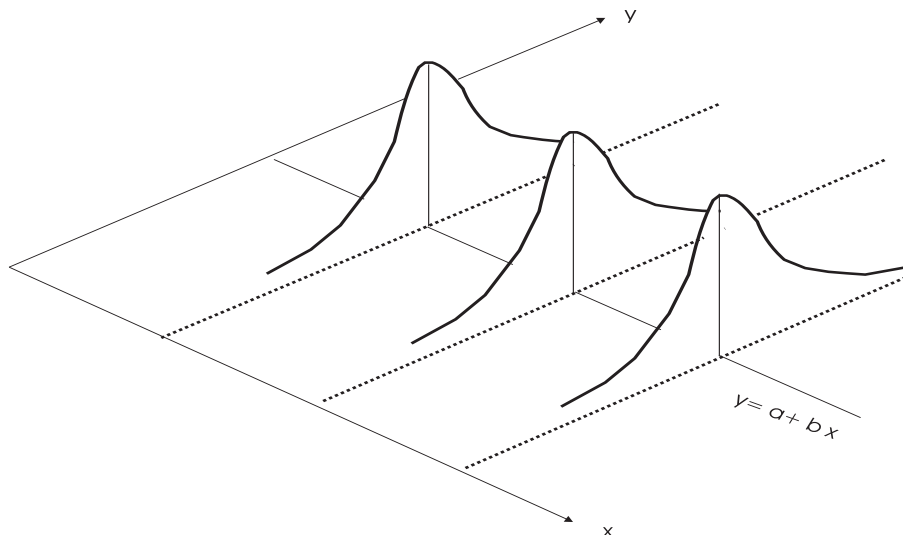
Denotemos la recta verdadera de regresión por

$$Y_i = \delta + \beta x_i + \epsilon_i$$

y asumamos que se dispone de  $n$  pares de observaciones. Suelen realizarse los siguientes supuestos:

1. Cada  $x_i$  es un número fijo (asignado, por ejemplo, por un investigador), o es la realización de una variable aleatoria  $X_i$  independiente del término error  $\epsilon_i$ . En el último caso, la inferencia se realiza condicionado al valor observado  $x_i$ .
2. Los términos de error  $\epsilon_i$  son variables aleatorias con media 0, es decir,

$$E(\epsilon_i) = 0, \quad \text{para todo } i = 1, \dots, n.$$



**Figura 6.2:** Ilustración del modelo de regresión: se representan funciones de densidad de la variable dependiente para valores  $x$  dados de la variable independiente

3. Las variables aleatorias  $\epsilon_i$  tienen todas las mismas varianzas  $\sigma^2$ , es decir,

$$V(\epsilon_i) = \sigma^2, \quad \text{para todo } i = 1, \dots, n.$$

4. Las variables aleatorias  $\epsilon_i$  no se hallan correlacionadas, luego,

$$E(\epsilon_i \epsilon_j) = 0, \quad \text{para todo } i, j = 1, \dots, n \text{ con } i \neq j.$$

En el resto de este capítulo, asumiremos que se verifican estos supuestos.

### 6.3.2. Comentarios

Ahora, los supuestos anteriores del modelo de regresión implican que, para un valor fijo  $x_i$  de  $X$ :

1. La esperanza condicional de  $Y_i$  dado que  $X = x_i$  viene dada por:

$$E(Y_i / X = x_i) = E(\delta + \beta x_i + \epsilon_i) = \delta + \beta x_i + E(\epsilon_i) = \delta + \beta x_i$$

2. La varianza condicional de  $Y_i$  dado que  $X = x_i$  viene dada por:

$$V(Y_i / X = x_i) = V(\delta + \beta x_i + \epsilon_i) = V(\delta + \beta x_i) + V(\epsilon_i) = 0 + \sigma^2 = \sigma^2,$$

#### Teorema 6.1

Denotemos la recta verdadera de regresión por  $Y_i = \delta + \beta x_i + \epsilon_i$ . Si se cumplen los supuestos en el modelo de regresión, entonces, para un valor fijo  $x_i$  de  $X$ , la esperanza y varianza condicionales de  $Y_i$  dado que  $X = x_i$  vienen dadas, respectivamente, por

$$E(Y_i / X = x_i) = \delta + \beta x_i \quad \text{y} \quad V(Y_i / X = x_i) = \sigma^2.$$

**Ejemplo 6.2**

Suponga que la relación entre la profundidad del océano  $X$  y la temperatura del agua  $Y$  está descrito por el modelo de regresión lineal simple con la verdadera recta de regresión  $Y = 65 - 1,2x + \epsilon$  y supongamos que se cumplen los supuestos del modelo de regresión y que  $\epsilon$  está normalmente distribuida con media 0 y desviación  $\sigma = 8$ .

- (a) Halle la probabilidad de que  $Y > 50$  cuando  $X = 20$ .
- (b) Halle la probabilidad de que  $Y > 50$  cuando  $X = 25$ .
- (c) Suponga que  $Y_{25}$  representa una observación sobre la temperatura cuando  $X = 25$  y que  $Y_{24}$  simboliza una observación independiente con  $X = 24$ . Hallar la probabilidad de que  $Y_{25}$  exceda  $Y_{24}$ .

SOLUCION:

- (a) El supuesto de normalidad de los errores implica que la variable  $Y$  también es normal. Además, por el teorema 6.3.2, para  $X = 20$ ,  $Y$  tiene media y varianza

$$E(Y/X = 20) = 65 - (1,2)(20) = 41 \quad \text{y} \quad V(Y/X = 20) = 8^2 = 64,$$

respectivamente. Por consiguiente, teniendo en cuenta la tabla normal del apéndice,

$$P(Y > 50/X = 20) = P\left(Z > \frac{50 - 41}{8}\right) = P(Z > 1,125) = 0,1292.$$

- (b) Análogo a la parte (a), para  $X = 25$ ,  $Y$  tiene media y varianza

$$E(Y/X = 25) = 65 - (1,2)(25) = 35 \quad \text{y} \quad V(Y/X = 25) = 8^2 = 64,$$

respectivamente. Por lo tanto, otra vez teniendo en cuenta la tabla normal del apéndice, obtenemos

$$P(Y > 50/X = 25) = P\left(Z > \frac{50 - 35}{8}\right) = P(Z > 1,875) = 0,0301.$$

- (c) El supuesto de normalidad de los errores implica nuevamente que la variable  $Y_{25} - Y_{24}$  sea también normal. Además, por el teorema 6.3.2, tenemos que

$$E(Y_{25} - Y_{24}) = E(Y_{25}) - E(Y_{24}) = 35 - 36,2 = -1,2$$

y que

$$V(Y_{25} - Y_{24}) = V(Y_{25}) + V(Y_{24}) = 64 + 64 = 128,$$

respectivamente y en donde se ha tenido en cuenta la independencia de las variables  $Y_{25}$  y  $Y_{24}$ . Ahora, teniendo en cuenta la tabla normal del apéndice,

$$P(Y_{25} > Y_{24}) = P(Y_{25} - Y_{24} > 0) = P\left(Z > \frac{0 - (-1,2)}{\sqrt{128}}\right) = P(Z > 0,11) = 0,4562.$$

Con base en el resultado anterior, podemos concluir lo siguiente: aun cuando esperemos que  $Y$  disminuya al aumentar  $X$  en una unidad, no es probable que la  $Y$  observada en  $X = x + 1$  sea más grande que la  $Y$  observada en  $X = x$ .

## 6.4 Estimación de los parámetros del modelo

### Teorema 6.2

La estimación de mínimos cuadrados del coeficiente  $\beta$  de la recta de regresión real es

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}},$$

en donde

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}, \quad S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}.$$

La estimación de mínimos cuadrados de la ordenada al origen  $\delta$  de la recta de regresión verdadera es

$$\hat{\delta} = \bar{y} - \hat{\beta}\bar{x}.$$

### Ejemplo 6.3

Para obtener una idea de cómo reciben ciertas plantas abonos o minerales de la tierra, se hizo el experimento siguiente: se escogieron 9 tierras con plantas de ciertos cereales (tanto en tierras como plantas en condiciones lo más idénticas posibles). Se dio a cada tierra  $i$  una cantidad fija  $x_i$  de fósforo y se midieron después de 38 días las cantidades  $y_i$  de fósforo en plantas. Los datos del experimento fueron:

$x_i$	1	4	5	9	11	13	23	23	28
$y_i$	64	71	54	81	76	93	77	95	109

Supóngase que los  $y_i$  dependen linealmente de los  $x_i$  a través de un modelo de regresión lineal simple son parámetros  $\delta$  y  $\beta$ .

- Calcule  $\hat{\delta}$  y  $\hat{\beta}$ .
- Halle la ecuación verdadera de regresión.
- Calcule una estimación puntual de  $E(Y/X = 13)$ .
- Haga un diagrama de dispersión para los datos  $(x_i, y_i)$  y dibuje en él la línea de regresión.

SOLUCION:

Tenemos que  $n = 9$ ,  $\bar{x} = 13$  y  $\bar{y} = 80$ .

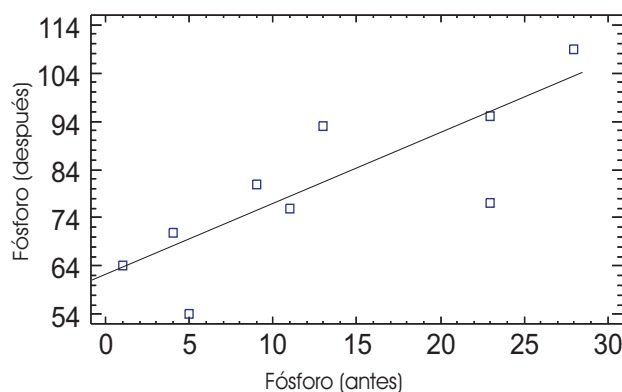
- Aplicando el teorema 6.4, encontramos que  $\hat{\beta} = 1,42$  y  $\hat{\delta} = 61,54$ .
- La ecuación verdadera de regresión viene dada por

$$E(Y/X = x) = 61,54 + 1,42x.$$

- Una estimación puntual de  $E(Y/X = 13)$  es

$$E(Y/X = 13) = 61,54 + (1,42)(13) = 80.$$

- En la figura 6.3 se encuentra el diagrama de dispersión y la línea de regresión para los datos dados.



**Figura 6.3:** Diagrama de dispersión y recta de regresión para los datos del ejemplo 6.4



## 6.5 Recursos online

1. Hacer click en el link: [Calculadora de ecuaciones de la recta](#).
2. Hacer click en los videos: [Video 1](#) o [Video 2](#) (solo el primer minuto) para crear una tabla de datos en Geogebra.
3. Hacer click en el video: [Regresión con Geogebra](#).

## 6.6 Resumen de notaciones

Se resumen las notaciones utilizadas hasta ahora y de las que se necesitan más adelante.

1. La ecuación de la recta de regresión muestral (estimada o de mínimo cuadrados):

$$E(Y/X = x) = \delta + \beta x$$

2.  $x$  se llama VARIABLE INDEPENDIENTE (PREDICTORA O EXPLICATIVA).
3. Para  $x$  fija,  $y$  se llama VARIABLE DEPENDIENTE O DE RESPUESTA.
4. Los parámetros del modelo son la pendiente  $\beta$  y el intercepto  $\delta$  (punto de corte con el eje  $Y$ ), entre otros.
5. El método para estimar los parámetros del modelo es el de MÍNIMOS CUADRADOS.
6. Suma de  $xy$  (Geogebra utiliza la notación  $nCov$ ):

$$nCov = S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

7. Suma de  $xx$  (Geogebra utiliza la notación  $nVarX$ ):

$$nVarX = S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

8. Estimación de de mínimo cuadrados de la pendiente:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

9. Intercepto estimado:

$$\hat{\delta} = \bar{y} - \hat{\beta}\bar{x}.$$

10. Suma de cuadrados total, suma de  $yy$  (Geogebra utiliza la notación  $nVarY$ ):

$$nVarY = S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

11. Suma de cuadrados de la regresión:

$$SSR = \hat{\beta}S_{xy}$$

12. Suma de cuadrados residual o del error (Geogebra utiliza la notación  $SECs$ ):

$$SECs = SSE = S_{yy} - SSR$$

13. Varianza estimada del error  $\epsilon$ :

$$S_{\epsilon}^2 = \frac{SSE}{n-2}$$

14. Varianza estimada de  $\hat{\beta}$ :

$$S_{\hat{\beta}}^2 = \frac{S_{\epsilon}^2}{S_{xx}}$$

15. Varianza estimada de  $\hat{\delta}$ :

$$S_{\hat{\delta}}^2 = \frac{S_{\epsilon}^2 \sum x_i^2}{n S_{xx}}$$

16. Coeficiente de correlación muestral  $r$  (o  $R$ ):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

17. Una propiedad de  $r$  es que siempre se cumple que:

$$-1 \leq r \leq 1$$

18. El coeficiente de determinación muestral se define como (siempre escribir en porcentajes):

$$R^2 \quad \text{o} \quad r^2 \quad (\text{multiplicado por } 100\%)$$

19. Interpretación de  $r^2$ : Cualquiera de las que aparecen abajo (escribir siempre lo que significa  $y$  y  $x$ ):

- El tanto por ciento de la variación observada en  $y$  (o se puede decir también: ... de la variabilidad de  $y$ ) se explica por su dependencia lineal de  $x$ .
- El tanto por ciento de la variación observada en  $y$  (o de la variabilidad de  $y$ ) se puede atribuir a la regresión lineal simple entre  $x$  y  $y$ .

## 6.7 Intervalos de confianza para la pendiente y el intersepto

La forma de los intervalos de confianza para  $\delta$  y  $\beta$  se muestra en el siguiente teorema:

### Teorema 6.3

Denotemos por  $\hat{\delta}$  y  $\hat{\beta}$  la estimación de mínimos cuadrados del intersepto  $\delta$  y de la pendiente  $\beta$  de la recta de regresión poblacional, respectivamente. Supongamos que se verifican los supuestos mencionados en seccionados anteriores y puede asumirse que los errores tienen distribución normal.

- (a) Un intervalo de confianza del  $(1 - \alpha)100\%$  para  $\delta$  se obtiene mediante:

$$\hat{\delta} - t_{\alpha/2} S_{\hat{\delta}} < \delta < \hat{\delta} + t_{\alpha/2} S_{\hat{\delta}}$$

- (b) Un intervalo de confianza del  $(1 - \alpha)100\%$  para  $\beta$  se obtiene mediante:

$$\hat{\beta} - t_{\alpha/2} S_{\hat{\beta}} < \beta < \hat{\beta} + t_{\alpha/2} S_{\hat{\beta}}$$

Aquí,  $t_{\alpha/2}$  es el valor de una variable aleatoria que deja un área de  $\alpha/2$  a la derecha de la distribución  $t$  de Student con  $n - 2$  grados de libertad.



## 6.8 Pruebas de hipótesis para la pendiente

### 6.8.1. Contraste para la pendiente de la regresión poblacional usando la prueba $t$

1. El estadístico de prueba tiene la forma

$$t = \frac{\hat{\beta} - \beta_0}{S_{\hat{\beta}}}$$

2. La distribución a considerar es la  $t$  de Student con  $n - 2$  grados de libertad.
3. La región crítica dependerá de tres casos, como se ilustra en la tabla 6.1:

**Cuadro 6.1:** Reglas de decisión para contrastes sobre  $\beta$  (usando  $t$  con  $n - 2$  grados de libertad)

Tipo de hipótesis	Regla de decisión
$H_0 : \beta \geq \beta_0$ $H_1 : \beta < \beta_0$ Cola a la izquierda	Si $t \leq -t_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \beta \leq \beta_0$ $H_1 : \beta > \beta_0$ Cola a la derecha	Si $t \geq t_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \beta = \beta_0$ $H_1 : \beta \neq \beta_0$ Dos colas	Si $t \leq -t_{\alpha/2}$ o $t \geq t_{\alpha/2}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .

### 6.8.2. Contraste para la pendiente de la regresión poblacional usando el procedimiento del análisis de varianza

1. Para probar la hipótesis  $H_0 : \beta = 0$ , calculamos:

$$F = \frac{SSR}{S_e^2}$$

2. En la tabla 6.2 resumimos la regla de decisión:

**Cuadro 6.2:** Regla de decisión para probar  $\beta = 0$  (usando la prueba  $F$ )

Tipo de hipótesis	Regla de decisión
$H_0 : \beta = 0$ $H_1 : \beta \neq 0$ Dos colas	Si $F > F_\alpha(1, n - 2)$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .

## 6.9 Pruebas de hipótesis para el intercepto

1. El estadístico de prueba tiene la forma

$$t = \frac{\hat{\delta} - \delta_0}{S_{\hat{\delta}}}$$

2. La distribución a considerar es la  $t$  de Student con  $n - 2$  grados de libertad.
3. La región crítica dependerá de tres casos, como se ilustra en la tabla 6.3:

**Cuadro 6.3:** Reglas de decisión para contrastes sobre  $\delta$  (usando  $t$  con  $n - 2$  grados de libertad)

Tipo de hipótesis	Regla de decisión
$H_0 : \delta \geq \delta_0$ $H_1 : \delta < \delta_0$ Cola a la izquierda	Si $t \leq -t_{\alpha}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \delta \leq \delta_0$ $H_1 : \delta > \delta_0$ Cola a la derecha	Si $t \geq t_{\alpha}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \delta = \delta_0$ $H_1 : \delta \neq \delta_0$ Dos colas	Si $t \leq -t_{\alpha/2}$ o $t \geq t_{\alpha/2}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .

## 6.10 Inferencias para la correlación poblacional

### 6.10.1. Contrastes para la correlación poblacional nula

La variable aleatoria correspondiente a

$$t = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

sigue una distribución  $t$  de Student con  $n - 2$  grados de libertad. En general y como es de suponerse, en esta situación, la región crítica dependerá de cada uno de los tres casos (como se puede ver en la tabla 6.4):

$$H_0 : \rho = 0, \quad H_0 : \rho \geq 0, \quad H_0 : \rho \leq 0.$$

### 6.10.2. Contrastes para la correlación poblacional (más general)

El estadístico de prueba es:

$$Z = \frac{\sqrt{n-3}}{2} \ln \left[ \frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)} \right],$$

la distribución a utilizar es la normal estándar. La región crítica dependerá, entonces, de cada uno de los tres casos señalados en (??), como se ilustra en la tabla 6.5:

**Cuadro 6.4:** Contrastes para  $\rho$  (usando la  $t$  de Student con  $n - 1$  grados de libertad)

Tipo de hipótesis	Regla de decisión
$H_0 : \rho \geq 0$ $H_1 : \rho < 0$ Cola a la izquierda	Si $t \leq -t_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \rho \leq 0$ $H_1 : \rho > 0$ Cola a la derecha	Si $t \geq t_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \rho = 0$ $H_1 : \rho \neq 0$ Dos colas	Si $t \leq -t_{\alpha/2}$ o $t \geq t_{\alpha/2}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .

**Cuadro 6.5:** Contrastes para  $\rho$  (usando la distribución normal)

Tipo de hipótesis	Regla de decisión
$H_0 : \rho \geq \rho_0$ $H_1 : \rho < \rho_0$ Cola a la izquierda	Si $Z \leq -Z_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \rho \leq \rho_0$ $H_1 : \rho > \rho_0$ Cola a la derecha	Si $Z \geq Z_\alpha$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .
$H_0 : \rho = \rho_0$ $H_1 : \rho \neq \rho_0$ Dos colas	Si $Z \leq -Z_{\alpha/2}$ o $Z \geq Z_{\alpha/2}$ , entonces, se rechaza $H_0$ ; de lo contrario, se acepta $H_0$ .

### 6.10.3. Intervalo de confianza para la correlación poblacional

También es posible construir un intervalo de confianza para  $\rho$ , utilizando la transformación de la ecuación (??), como se muestra en el teorema 6.10.3:

#### Teorema 6.4

Sea  $r$  el coeficiente de correlación muestral, calculado a partir de una muestra aleatoria de  $n$  pares de observaciones de una distribución conjunta normal. Si se define

$$\tanh(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}},$$

entonces, un intervalo de confianza del  $(1 - \alpha)100\%$  para  $\rho$  se obtiene mediante:

$$\tanh\left(\frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) - \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right) < \rho < \tanh\left(\frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) + \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right).$$

Aquí  $Z_{\alpha/2}$  es el valor de una variable aleatoria que deja un área de  $\alpha/2$  a la derecha de la distribución normal estándar.

**Ejemplo 6.4**

En los últimos 25 años, una cadena de tiendas de descuento en ropa para dama ha aumentado su porcentaje de mercado con el incremento en el número de tiendas. Nunca empleó su enfoque sistemático en la selección del lugar. Esta se basó, en esencia, en los que se consideró un gran sitio o una buena renta. Este año, con un plan estratégico para abrir varias tiendas nuevas, se pidió al director de proyectos especiales y planeación que desarrollara un enfoque para predecir las ventas anuales en todas las tiendas nuevas. Suponga que se decidió examinar la relación entre el tamaño (en metros cuadrados) de una tienda y sus ventas anuales con la selección de una muestra de 6 tiendas. Los resultados para estas 6 tiendas se resumen en la tabla de abajo.

$x$ (tamaño en metros cuadrados)	100	200	300	400	500	600
$y$ (ventas anuales en millones de pesos)	40	50	50	70	65	80

Suponiendo que se verifican los supuestos básicos para el modelo de regresión lineal (y el de normalidad) y que la desviación del error aleatorio del modelo es 8. Realizar los siguientes incisos.

*Cálculos básicos:*

1. Verifique que:  $n = 6$ ,  $\sum x_i = 2.100$ ,  $\sum y_i = 355$ ,  $\sum x_i y_i = 137.500$ ,  
 $\sum x_i^2 = 910.000$ ,  $nCov = S_{xy} = 13.250$ ,  $nVarX = S_{xx} = 175.000$ ,  $\bar{x} = 350$ ,  $\bar{y} = 59,16$ .
2. Verifique que:  $SSR = 1.003,21$ ,  $\sum y_i^2 = 22.125$ ,  $nVarY = S_{yy} = 1.120,833$ ,  $SECs = SSE = 117,6228$ ,  
 $S_e^2 = 29,4123$ ,  $S_e = 5,4233$ ,  $S_{\hat{\beta}}^2 = 0,000168$ ,  $S_{\hat{\beta}} = 0,0129641$ ,  $S_{\hat{\delta}}^2 = 25,49$ ,  $S_{\hat{\delta}} = 5,048$ .

*Estimación de los parámetros y gráfica:*

3. Halle el intercepto  $\hat{\delta}$  y la pendiente  $\hat{\beta}$ . Interprete estos valores.

$$R/. \quad \hat{\beta} = 0,0757143; \quad \hat{\delta} = 32,67$$

4. Halle la ecuación de regresión estimada.

$$y = 32,67 + 0,0757143x$$

5. Haga un bosquejo de la línea de regresión.

*Cálculo de probabilidades:*

6. Calcule la probabilidad de que la venta anual es mayor que 55 millones de pesos si el tamaño de la tienda fue de 400 metros cuadrados. Es decir, calcular  $P(Y > 55|X = 400)$ .

$$R/. \quad \sigma = 8; \quad \mu = E(Y|X = 400) = 62,947; \quad P(\dots) = P(Z > -0,99) = 0,8389$$

*Pruebas de hipótesis para  $\beta$ :*

7. Use la prueba  $F$  y  $\alpha = 0,05$  para determinar si los datos satisfacen el modelo de regresión lineal (es decir, si  $\beta \neq 0$ ).

$$R/. \quad F = 34,1085; \quad F_{0,05}(1,4) = 7,71; \quad \beta \neq 0$$

8. Use la prueba  $t$  y  $\alpha = 0,05$  para determinar si el modelo lineal es apropiado para los datos (es decir, si  $\beta \neq 0$ ).

$$R/. \quad t = 5,841; \quad t_{0,025}(4) = 2,776; \quad \beta \neq 0$$

*Prueba de hipótesis para  $\delta$ :*

9. Use  $\alpha = 0,05$  para determinar si la línea de regresión pasa por el origen (es decir, si  $\delta = 0$ ).

$$R/. \quad t = 6,474; \quad t_{0,025}(4) = 2,776; \quad \delta \neq 0$$

*Intervalo de confianza para  $\beta$ :*

10. Construya un intervalo de confianza del 95 % para  $\beta$  y, con ayuda del intervalo encontrado, verifique si puede concluir si los datos satisfacen el modelo de regresión lineal (es decir, si  $\beta \neq 0$ ).

$$R/. \quad 0,0398 < \beta < 0,1116; \quad \beta \neq 0$$

*Intervalo de confianza para  $\delta$ :*

11. Construya un intervalo de confianza del 95 % para  $\delta$  y, con ayuda del intervalo encontrado, verifique si puede concluir que la línea de regresión pasa por el origen (es decir, si  $\delta = 0$ ).

$$R/. \quad 18,66 < \delta < 46,671; \quad \delta \neq 0$$

*Coefficientes de correlación y de determinación (muestrales):*

12. Determine el coeficiente de correlación muestral e interprete su resultado.

$$R/. \quad r = 0,946$$

13. Determine el coeficiente de determinación muestral e interprete su resultado.

$$R/. \quad r^2 = 89,5\%$$

*Prueba de hipótesis para la correlación poblacional  $\rho$ :*

14. Use  $\alpha = 0,05$  para determinar si el coeficiente de correlación poblacional es distinto de cero. ¿Están dispersos los datos poblacionales?

$$R/. \quad t = 5,801; \quad t_{0,025}(4) = 2,776; \quad \rho \neq 0$$

## 6.11 Ejercicios

1. Los datos de 13 personas que aparecen en la siguiente tabla representan el incremento porcentual en sus inversiones en dos tipos de acciones, A y B, durante un mismo periodo de tiempo:

A:	2,8	2,2	-1,6	-1,3	5,6	-1,4	1,4	1,5	-4,7	1,1	1,5	0,2	-0,1
B:	20,3	-3,7	27,7	22,6	2,3	11,9	27,0	-4,3	20,3	4,2	14,9	-9,2	19,6

- (a) Estime la recta de regresión de  $y$  sobre  $x$ .  
 (b) Interprete el valor del intersección y de la pendiente de la recta de regresión muestral.

2. La siguiente tabla presenta datos sobre la resistencia a la comprensión ( $x$ ) y la permeabilidad ( $y$ ) de varias mezclas y tratamientos de concreto.

$y$	185	184	184	186	188	170	172	173	174	174	175	176
$x$	3,00	1,49	2,52	1,87	3,08	0,84	1,31	1,42	1,03	1,07	1,08	1,04
$y$	182	182	181	182	182	177	180	180	180	180	180	181
$x$	1,94	2,68	1,43	,90	1,81	180	1,45	1,60	1,61	2,13	2,15	0,84

- (a) Trace un diagrama de tallo y hojas de la permeabilidad y la resistencia a la comprensión y comente sobre las propiedades interesantes.
- (b) ¿Está determinado el valor de la resistencia a la comprensión en forma total y única por la permeabilidad de las mezclas y los tratamientos de concretos? Justifique su respuesta.
- (c) Trace un diagrama de dispersión de los datos. ¿Al parecer la permeabilidad se puede predecir bien por el valor de la resistencia? Justifique su respuesta.
3. Los datos siguientes representan el incremento porcentual del salario ( $y$ ) y el número de meses extras trabajados en la ejecución de un proyecto ( $x$ ):

$x$	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	1,0	1,1	1,2
$y$	9,8	9,5	8,9	8,6	10,2	9,3	9,2	10,5	8,1	7,8	8,5

- (a) Calcule la ecuación de la recta de regresión estimada.
- (b) Estime el incremento porcentual medio del salario cuando se han trabajado 1,75 meses en la ejecución de un proyecto.
4. Los siguientes datos representan los puntos acumulados por 20 equipos de fútbol al final de un torneo jugado en 2.004 ( $x$ ) y de otro, jugado en 2.005 ( $y$ ):

$x$	40	55	65	35	60	90	35	50	35	35
$y$	56	68	36	11	70	79	59	53	41	61

$x$	60	60	40	55	50	65	50	90	80	60
$y$	71	71	47	53	68	57	79	54	91	48

- (a) Grafique un diagrama de dispersión.
- (b) Encuentre la ecuación de la línea de regresión para predecir los puntos acumulados en el 2.005 a partir de los puntos acumulados en el 2.004.
- (c) Grafique la línea en el diagrama de dispersión.
5. Los siguientes datos representan el número de charlas  $x$  ofrecidas por una empresa de seguros durante seis meses y el número de afiliaciones a dicha empresa en ese periodo de tiempo:

$x$	20	30	50	40	20	50	40	25	50	40	20	25
$y$	365	475	440	490	420	560	525	480	510	385	400	395

- (a) Haga un diagrama de dispersión.
- (b) Encuentre la ecuación de la línea de regresión para predecir el número de afiliaciones a partir del número de charlas.

(c) Estime el número de afiliaciones cuando se realizaron 35 charlas.

6. La tabla adjunta contiene datos recopilados por un gerente de ventas en los reportes anuales (en millones de pesos) y en años de experiencia:

Años de experiencia ( $x$ )	1,2	1,4	1,2	1,0	1,5
Ventas anuales ( $y$ )	27	16	25	40	10

- (a) Estime la regresión lineal de las ventas sobre los años de experiencia.
- (b) Halle e interprete un intervalo de confianza del 90% para la pendiente de la recta de regresión poblacional.
- (c) ¿Sería útil la regresión muestral obtenida en el apartado (a) para predecir las ventas cuando se tenga 2,5 años de experiencia? Explique la respuesta.
7. Los siguientes datos representan el tiempo semanal  $y$  (en horas) de viaje para un mensajero y la cantidad de entregas  $x$  de cartas:

$x$	41	44	47	48	55	64	15	19	31	39
$y$	32	60	78	59	61	60	23	52	65	55

- (a) Calcule un intervalo del 95% de confianza para el cambio esperado en el tiempo total semanal de viaje, asociado con 1% de aumento en la cantidad de entregas de cartas. ¿Qué sugiere el intervalo acerca de la utilidad del modelo?
- (b) Realice una prueba de utilidad del modelo basado en el  $P$ -valor. ¿Utilizaría la recta de mínimos cuadrados para pronosticar el tiempo a partir de la cantidad de entregas  $x$  de cartas? Explique.
8. Los siguientes datos corresponden a la presión  $x$  de gas extraído (en micrones) y al tiempo de extracción  $y$  (en minutos):

$y$	3,3	3,7	4,1	4,3	4,8	5,0	5,4	2,5	3,0	3,1
$x$	160	260	275	325	370	420	480	40	130	155

- (a) Estime  $\sigma$  y la desviación estándar de  $\hat{\beta}$ .
- (b) Suponga que, antes del experimento, los investigadores creían que podría haber un incremento de 0,0060 minutos en el tiempo promedio de extracción, asociado con un incremento de 1 micrón en la presión. Use el  $P$ -valor y compárelo con un nivel de significancia de 0,10. ¿Contradice la información este supuesto?
9. Los siguientes datos representan el número  $x$  de proyectos presentados el año pasado por 12 universidades privadas y la ayuda recibida  $y$  (en millones de pesos) para la ejecución de estos proyectos:

$x$	15,7	17,2	13,8	24,2	15,0	12,7	13,8	18,7	10,8	11,8	25,4	17,2
$y$	4	3	6	5	3	12	5	1	12	11	2	4

- (a) Determine el coeficiente de correlación  $r$  de Pearson.
- (b) Haga una prueba, con  $\alpha = 0,05$ , para determinar si el número de proyectos está relacionado con las ayudas recibidas.
- (c) Pruebe la hipótesis nula  $H_0 : \rho \leq 0$  contra la hipótesis alternativa  $H_1 : \rho > 0$ , usando  $\alpha = 0,05$ .
- (d) Calcule el coeficiente de determinación e interprete el resultado.
10. Los siguientes datos representan los gastos anuales en publicidad  $x$  (en millones de pesos) y las ventas  $y$  (en millones de pesos):

$x$	$y$	$x$	$y$	$x$	$y$
4,17	96,97	3,63	179,18	6,09	196,67
10,04	154,70	4,65	171,81	3,08	289,59
6,02	151,61	2,97	200,23	1,76	105,71
1,52	163,92	1,57	125,19	3,09	275,97
4,81	147,82	0,98	120,49	4,18	95,83
7,70	141,77	1,57	98,61		

- (a) Halle el coeficiente de correlación muestral
- (b) Contraste, frente a una alternativa bilateral, la hipótesis nula de que la correlación poblacional es 0.

11. Los siguientes datos representan los gastos (en miles de pesos) en los servicios de agua ( $x$ ) y luz ( $y$ ), durante un determinado mes, de seis familias seleccionadas al azar:

$x$	80	74	65	83	70	92
$y$	63	87	78	90	74	84

- (a) Calcule e interprete el coeficiente de correlación de  $x$  y  $y$ .
- (b) Pruebe la hipótesis de que  $\rho = 0$  contra la alternativa de que  $\rho \neq 0$ . Utilice un nivel de significancia de 0,05.
12. Los siguientes datos se obtienen en un estudio de la relación entre el peso  $x$  (en kilogramos) y el volumen  $y$  (en centímetros cúbicos) de un tipo de recipiente:

$x$	5,52	3,21	4,32	2,31	4,30	3,71	2,75	2,15	4,41
$y$	36,5	27,2	27,7	28,3	30,3	29,7	29,5	26,3	32,2

- (a) Calcule el coeficiente de correlación muestral  $r$ .
- (b) Pruebe la hipótesis nula  $\rho = 0$  contra la alternativa  $\rho > 0$ , con un nivel de significancia de 0,01.
- (c) ¿Qué porcentaje de la variación en los volúmenes de los recipientes se explica por la diferencia en peso?
13. Los siguientes datos representan los ingresos mensuales por matrícula y el número de estudiantes matriculados en 19 escuelas privadas de cierto país. Utilice estos datos para ajustar una regresión lineal simple. Además, pruebe la hipótesis  $H_0 : \beta = 0$  contra  $H_1 : \beta \neq 0$ . Interprete sus resultados.

$x$	4.050	2.465	3.120	5.700	2.595	3.640	2.050	4.235	2.935	4.975
$y$	11,2	12,4	10,5	13,2	9,8	11,0	10,8	10,44	12,2	11,2

$x$	3.690	2.800	2.775	2.170	2.370	2.055	2.025	2.645	2.675
$y$	10,8	14,2	12,2	10,0	12,3	12,5	11,8	16,0	13,8

14. Se ha seleccionado una muestra aleatoria de 12 estudiantes de bachillerato que han repetido el séptimo grado este año y, para cada uno, se ha anotado el promedio global de las calificaciones de todas las asignaturas, tanto el de este año ( $y$ ) como el del año pasado ( $x$ ), así:

$x$	65	55	70	65	70	55	70	50	55	65	50	55
$y$	90	85	87	94	98	81	91	76	74	85	74	76

- (a) Calcule e interprete el coeficiente de correlación muestral.
- (b) Establezca las suposiciones necesarias sobre las variables aleatorias.



(c) Pruebe la hipótesis de que  $\rho = 0,5$  contra la alternativa de que  $\rho > 0,5$ . Utilice un  $P$ -valor en las conclusiones.

15. Los siguientes datos representan el volumen de lluvia  $x$  (en  $\text{m}^3$ ) y el volumen de escurrimiento  $y$  (en  $\text{m}^3$ ) para determinado lugar:

$x$	23	30	40	47	5	12	14	17	55	67	72	81	96	112	127
$y$	15	25	27	46	4	10	13	15	38	46	53	70	82	99	100

- ¿Respalda el diagrama de dispersión de los datos el uso del modelo de regresión lineal simple?
  - Calcule los estimados puntuales de la pendiente y la ordenada al origen de la recta de regresión poblacional.
  - Calcule un estimado puntual del volumen real de escurrimiento cuando el volumen de lluvia es 50.
  - Calcule un estimado puntual de la desviación estándar  $\sigma$ .
  - ¿Qué porción de la variación observada de volumen de escurrimiento se puede atribuir a la relación de regresión simple entre escurrimiento y lluvia?
16. Los siguientes datos representan el tiempo de vida  $x$  de un bombillo de marca A (en horas), así como el tiempo de vida  $y$  de un bombillo de marca B (en minutos):

$x$	4.200	3.600	3.750	3.675	4.050	2.770	4.870	4.500	3.450	2.700	3.750	3.300
$y$	370	340	375	310	350	200	400	375	285	225	345	285

- Calcule e interprete el valor del coeficiente de correlación muestral  $r$ .
  - ¿Cómo cambiaría el valor de  $r$  si  $x$  es el tiempo del bombillo de marca B y  $y$ , del bombillo de marca A?
  - ¿Cómo cambiaría el valor de  $r$  si el tiempo  $y$  se expresara en horas?
  - Trace gráficas de probabilidad normal y coméntelas.
  - Realice una prueba de hipótesis para decidir si el tiempo  $x$  y el tiempo  $y$  están correlacionados linealmente.
17. Un artículo mostró los siguientes datos sobre el tiempo de falla  $y$  (en segundos) de un componente electrónico y la temperatura  $x$  donde se emplea el componente (en grados centígrados):

$x$	57	60	72	81	85	94	46	48	109
$y$	2,28	2,34	2,53	2,28	2,62	2,63	2,18	2,10	2,50

$x$	121	132	137	148	149	184	185	187	55
$y$	2,66	2,79	2,80	3,01	2,98	3,34	3,49	3,26	2,13

- Calcule el valor del coeficiente de correlación muestral. Con base en este valor, ¿cómo describiría la naturaleza de la relación entre las dos variables?
- Si un primer componente tiene mayor valor de temperatura que un segundo componente, ¿qué se puede decir acerca del tiempo de falla real para los dos componentes?
- Si el tiempo de falla se expresa en minutos, ¿qué sucede con el valor  $r$ ? ¿Por qué?
- Si el modelo de regresión lineal simple se ajustara a estos datos, ¿qué proporción de la variación observada de tiempo de falla se podría explicar con la relación del modelo?
- Realice una prueba, con nivel de significancia de 0,01, para decidir si hay una relación lineal positiva entre las dos variables.

# A

## Apéndice de tablas

### 1. Distribución binomial

Las tablas (a)-(e) muestran la probabilidad  $P(X \leq k) = B(k; n, p)$  de que ocurran máximo  $k$  éxitos en  $n$  ensayos independientes, cada uno con probabilidad de éxito  $p$ .

Estas probabilidades se calculan para  $n = 5, 10, 15, 20$  y  $25$ , respectivamente.

#### (a) Tabla binomial para $n = 5$

	$p$												
k	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95
0	0,774	0,590	0,328	0,237	0,168	0,078	0,031	0,010	0,002	0,001	0,000	0,000	0,000
1	0,977	0,919	0,737	0,633	0,528	0,337	0,188	0,087	0,031	0,016	0,007	0,000	0,000
2	0,999	0,991	0,942	0,896	0,837	0,683	0,500	0,317	0,163	0,104	0,058	0,009	0,001
3	1,000	1,000	0,993	0,984	0,969	0,913	0,812	0,663	0,472	0,367	0,263	0,081	0,023
4	1,000	1,000	0,999	0,999	0,998	0,990	0,969	0,922	0,832	0,763	0,672	0,410	0,226
5	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

#### (b) Probabilidades binomiales acumuladas para $n = 10$

	$p$												
k	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95
0	0,599	0,349	0,107	0,056	0,028	0,006	0,001	0,000	0,000	0,000	0,000	0,000	0,000
1	0,914	0,736	0,376	0,244	0,149	0,046	0,011	0,002	0,000	0,000	0,000	0,000	0,000
2	0,988	0,930	0,678	0,526	0,383	0,167	0,055	0,012	0,002	0,000	0,000	0,000	0,000
3	0,999	0,987	0,879	0,776	0,650	0,382	0,172	0,055	0,011	0,004	0,001	0,000	0,000
4	1,000	0,998	0,967	0,922	0,850	0,633	0,377	0,166	0,047	0,020	0,006	0,000	0,000
5	1,000	1,000	0,994	0,980	0,953	0,834	0,623	0,367	0,150	0,078	0,033	0,002	0,000
6	1,000	1,000	0,999	0,996	0,989	0,945	0,828	0,618	0,350	0,224	0,121	0,013	0,001
7	1,000	1,000	1,000	1,000	0,998	0,988	0,945	0,833	0,617	0,474	0,322	0,070	0,012
8	1,000	1,000	1,000	1,000	1,000	0,998	0,989	0,954	0,851	0,756	0,624	0,264	0,086
9	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,994	0,972	0,944	0,893	0,651	0,401

**(c) Probabilidades binomiales acumuladas para  $n = 15$** 

k	$p$												
	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95
0	0,463	0,206	0,305	0,013	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1	0,829	0,549	0,167	0,080	0,035	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2	0,964	0,816	0,398	0,236	0,127	0,027	0,004	0,000	0,000	0,000	0,000	0,000	0,000
3	0,995	0,944	0,648	0,461	0,297	0,091	0,018	0,002	0,000	0,000	0,000	0,000	0,000
4	0,999	0,987	0,836	0,686	0,515	0,217	0,059	0,009	0,001	0,000	0,000	0,000	0,000
5	1,000	0,998	0,939	0,852	0,722	0,403	0,151	0,034	0,004	0,001	0,000	0,000	0,000
6	1,000	1,000	0,982	0,943	0,869	0,610	0,304	0,095	0,015	0,004	0,001	0,000	0,000
7	1,000	1,000	0,996	0,983	0,950	0,787	0,500	0,213	0,050	0,017	0,004	0,000	0,000
8	1,000	1,000	0,999	0,996	0,985	0,905	0,696	0,390	0,131	0,057	0,018	0,000	0,000
9	1,000	1,000	1,000	0,999	0,996	0,966	0,849	0,597	0,278	0,148	0,061	0,002	0,000
10	1,000	1,000	1,000	1,000	0,999	0,991	0,941	0,783	0,485	0,314	0,164	0,013	0,000
11	1,000	1,000	1,000	1,000	1,000	0,998	0,982	0,909	0,703	0,539	0,352	0,056	0,005
12	1,000	1,000	1,000	1,000	1,000	1,000	0,996	0,973	0,873	0,764	0,602	0,184	0,036
13	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,995	0,965	0,920	0,833	0,451	0,171
14	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,995	0,987	0,965	0,794	0,537

**(d) Probabilidades binomiales acumuladas para  $n = 20$** 

k	$p$												
	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95
0	0,358	0,122	0,012	0,003	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1	0,736	0,392	0,069	0,024	0,008	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2	0,925	0,677	0,206	0,091	0,035	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3	0,984	0,867	0,411	0,225	0,107	0,016	0,001	0,000	0,000	0,000	0,000	0,000	0,000
4	0,997	0,957	0,630	0,415	0,238	0,051	0,006	0,000	0,000	0,000	0,000	0,000	0,000
5	1,000	0,989	0,804	0,617	0,416	0,126	0,021	0,002	0,000	0,000	0,000	0,000	0,000
6	1,000	0,998	0,913	0,786	0,608	0,250	0,058	0,006	0,000	0,000	0,000	0,000	0,000
7	1,000	1,000	0,968	0,898	0,772	0,416	0,132	0,021	0,001	0,000	0,000	0,000	0,000
8	1,000	1,000	0,990	0,959	0,887	0,596	0,252	0,057	0,005	0,001	0,000	0,000	0,000
9	1,000	1,000	0,997	0,986	0,952	0,755	0,412	0,128	0,017	0,004	0,001	0,000	0,000
10	1,000	1,000	0,999	0,996	0,983	0,872	0,588	0,245	0,048	0,014	0,003	0,000	0,000
11	1,000	1,000	1,000	0,999	0,995	0,943	0,748	0,404	0,113	0,041	0,010	0,000	0,000
12	1,000	1,000	1,000	1,000	0,999	0,979	0,868	0,584	0,228	0,102	0,032	0,000	0,000
13	1,000	1,000	1,000	1,000	1,000	0,994	0,942	0,750	0,392	0,214	0,087	0,002	0,000
14	1,000	1,000	1,000	1,000	1,000	0,998	0,979	0,874	0,584	0,383	0,196	0,011	0,000
15	1,000	1,000	1,000	1,000	1,000	1,000	0,994	0,949	0,762	0,585	0,370	0,043	0,003
16	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,984	0,893	0,775	0,589	0,133	0,016
17	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,996	0,965	0,909	0,794	0,323	0,075
18	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,992	0,976	0,931	0,608	0,264
19	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,997	0,988	0,878	0,642

**(e) Probabilidades binomiales acumuladas para  $n = 25$**

k	p												
	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95
0	0,277	0,072	0,004	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1	0,642	0,271	0,027	0,007	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2	0,873	0,537	0,098	0,032	0,009	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
3	0,966	0,764	0,234	0,096	0,033	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,000
4	0,993	0,902	0,421	0,214	0,090	0,009	0,000	0,000	0,000	0,000	0,000	0,000	0,000
5	0,999	0,967	0,617	0,378	0,193	0,029	0,002	0,000	0,000	0,000	0,000	0,000	0,000
6	1,000	0,991	0,780	0,561	0,341	0,074	0,007	0,000	0,000	0,000	0,000	0,000	0,000
7	1,000	0,998	0,891	0,727	0,512	0,154	0,022	0,001	0,000	0,000	0,000	0,000	0,000
8	1,000	1,000	0,953	0,851	0,677	0,274	0,054	0,004	0,000	0,000	0,000	0,000	0,000
9	1,000	1,000	0,983	0,929	0,811	0,425	0,115	0,013	0,000	0,000	0,000	0,000	0,000
10	1,000	1,000	0,994	0,970	0,902	0,586	0,212	0,034	0,002	0,000	0,000	0,000	0,000
11	1,000	1,000	0,998	0,980	0,956	0,732	0,345	0,078	0,006	0,001	0,000	0,000	0,000
12	1,000	1,000	1,000	0,997	0,983	0,846	0,500	0,154	0,017	0,003	0,000	0,000	0,000
13	1,000	1,000	1,000	0,999	0,994	0,922	0,655	0,268	0,044	0,020	0,002	0,000	0,000
14	1,000	1,000	1,000	1,000	0,998	0,966	0,788	0,414	0,098	0,030	0,006	0,000	0,000
15	1,000	1,000	1,000	1,000	1,000	0,987	0,885	0,575	0,189	0,071	0,017	0,000	0,000
16	1,000	1,000	1,000	1,000	1,000	0,996	0,946	0,726	0,323	0,149	0,047	0,000	0,000
17	1,000	1,000	1,000	1,000	1,000	0,999	0,978	0,846	0,488	0,273	0,109	0,002	0,000
18	1,000	1,000	1,000	1,000	1,000	1,000	0,993	0,926	0,659	0,439	0,220	0,009	0,000
19	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,971	0,807	0,622	0,383	0,033	0,001
20	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,991	0,910	0,786	0,579	0,098	0,007
21	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,967	0,904	0,766	0,236	0,034
22	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,991	0,968	0,902	0,463	0,127
23	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,993	0,973	0,729	0,358
24	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,996	0,928	0,723

## 2. Distribución de Poisson

La tabla muestra la probabilidad  $P(X \leq k; \lambda)$  para algunos valores  $\lambda$ .

	$\lambda = 0,1$	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$k = 0$	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407	0,368
1	0,995	0,982	0,963	0,938	0,910	0,878	0,844	0,809	0,772	0,736
2	1,000	0,999	0,996	0,992	0,986	0,977	0,966	0,953	0,937	0,920
3	1,000	1,000	1,000	0,999	0,998	0,997	0,994	0,991	0,987	0,981
4	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,999	0,998	0,996
5	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999
6	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

	$\lambda = 2$	3	4	5	6	7	8	9	10	15	20
$k = 0$	0,135	0,050	0,018	0,007	0,002	0,001	0,000	0,000	0,000	0,000	0,000
1	0,406	0,199	0,092	0,040	0,017	0,007	0,003	0,001	0,000	0,000	0,000
2	0,677	0,423	0,238	0,125	0,062	0,030	0,014	0,006	0,003	0,000	0,000
3	0,857	0,647	0,433	0,265	0,151	0,082	0,042	0,021	0,010	0,000	0,000
4	0,947	0,815	0,629	0,440	0,285	0,173	0,100	0,055	0,029	0,001	0,000
5	0,983	0,916	0,785	0,616	0,446	0,301	0,191	0,116	0,067	0,003	0,000
6	0,995	0,966	0,889	0,762	0,606	0,450	0,313	0,207	0,130	0,008	0,000
7	0,999	0,988	0,949	0,867	0,744	0,599	0,453	0,324	0,220	0,018	0,001
8	1,000	0,996	0,979	0,932	0,847	0,729	0,593	0,456	0,333	0,037	0,002
9	1,000	0,999	0,992	0,968	0,916	0,830	0,717	0,587	0,458	0,070	0,005
10	1,000	1,000	0,997	0,986	0,957	0,901	0,816	0,706	0,583	0,118	0,011
11	1,000	1,000	0,999	0,995	0,980	0,947	0,888	0,803	0,697	0,185	0,021
12	1,000	1,000	1,000	0,998	0,991	0,973	0,936	0,876	0,792	0,268	0,039
13	1,000	1,000	1,000	0,999	0,996	0,987	0,966	0,926	0,864	0,363	0,066
14	1,000	1,000	1,000	1,000	0,999	0,994	0,983	0,959	0,917	0,466	0,105
15	1,000	1,000	1,000	1,000	0,999	0,998	0,992	0,978	0,951	0,568	0,157
16	1,000	1,000	1,000	1,000	1,000	0,999	0,996	0,989	0,973	0,664	0,221
17	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,995	0,986	0,749	0,297
18	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,998	0,993	0,819	0,381
19	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,997	0,875	0,470
20	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,917	0,559
21	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,947	0,644
22	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,967	0,721
23	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,981	0,787
24	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,989	0,843
25	1,000	1,000	1,000	0,994	0,970	0,902	0,586	0,212	0,034	0,994	0,888
26	1,000	1,000	1,000	0,998	0,980	0,956	0,732	0,345	0,078	0,997	0,922
27	1,000	1,000	1,000	1,000	0,997	0,983	0,846	0,500	0,154	0,998	0,948
28	1,000	1,000	1,000	1,000	0,999	0,994	0,922	0,655	0,268	0,999	0,966
29	1,000	1,000	1,000	1,000	1,000	0,998	0,966	0,788	0,414	1,000	0,978
30	1,000	1,000	1,000	1,000	1,000	1,000	0,987	0,885	0,575	1,000	0,987
31	1,000	1,000	1,000	1,000	1,000	1,000	0,996	0,946	0,726	1,000	0,992
32	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,978	0,846	1,000	0,995
33	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,993	0,926	1,000	0,997
34	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,971	1,000	0,999
35	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,991	1,000	0,999
36	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	1,000	1,000

### 3.Distribución normal estándar

La tabla muestra la probabilidad  $P(Z \leq z)$ .

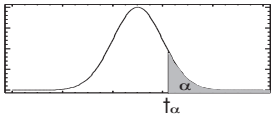
#### (a) Áreas para valores negativos de $Z$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0352	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0722	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641

**(b) Áreas para valores positivos de  $Z$**

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9278	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9948	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9961	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9971	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

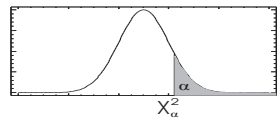
4. Distribución *t* de Student



	$\alpha$						
$\nu$	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	3,078	6,314	12,706	31,821	63,657	318,31	636,620
2	1,886	2,920	4,303	6,965	9,925	22,326	31,598
3	1,638	2,353	3,182	4,541	5,841	10,213	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,795
23	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	1,299	1,676	2,009	2,403	2,678	3,262	3,496
60	1,296	1,671	2,000	2,390	2,660	3,232	3,460
120	1,282	1,658	1,980	2,358	2,617	3,160	3,373
$\infty (= z)$	1,282	1,645	1,960	2,326	2,576	3,090	3,291



5. Distribución chi-cuadrada

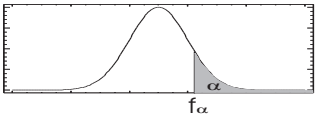


	$\alpha$									
$\nu$	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,000	0,000	0,000	0,001	0,00393	0,0158	0,0642	0,102	0,148	0,4550
2	0,010	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357
5	0,412	0,554	0,752	0,831	1,145	1,610	2,343	2,675	3,000	4,351
6	0,676	0,872	1,134	1,237	1,635	2,204	3,070	3,455	3,828	5,348
7	0,989	1,239	1,564	1,690	2,167	2,833	3,822	4,255	4,671	6,346
8	1,344	1,646	2,032	2,180	2,733	3,490	4,594	5,071	5,527	7,344
9	1,735	2,088	2,532	2,700	3,325	4,168	5,380	5,899	6,393	8,343
10	2,156	2,558	3,059	3,247	3,940	4,865	6,179	6,737	7,267	9,342
11	2,603	3,053	3,609	3,816	4,575	5,578	6,989	7,584	8,148	10,341
12	3,074	3,571	4,178	4,404	5,226	6,304	7,807	8,438	9,034	11,340
13	3,565	4,107	4,765	5,009	5,892	7,042	8,634	9,299	9,926	12,340
14	4,075	4,660	5,368	5,629	6,571	7,790	9,467	10,165	10,821	13,339
15	4,601	5,229	5,985	6,262	7,261	8,547	10,307	11,036	11,721	14,339
16	5,142	5,812	6,614	6,908	7,962	9,312	11,152	11,912	12,624	15,338
17	5,697	6,408	7,255	7,564	8,672	10,085	12,002	12,792	13,531	16,338
18	6,264	7,033	7,927	8,247	9,398	10,885	12,812	13,622	14,381	17,338
19	6,844	7,633	8,567	8,897	10,117	11,651	13,716	14,562	15,352	18,338
20	7,434	8,260	9,237	9,591	10,851	12,443	14,578	15,452	16,266	19,337
21	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337
24	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	10,520	11,524	12,692	13,120	14,611	16,473	18,940	19,939	20,867	24,337
26	11,160	12,198	13,409	13,844	15,379	17,292	19,820	20,843	21,792	25,336
27	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,336
28	12,461	13,565	14,847	15,308	16,928	18,939	21,588	22,657	23,647	27,336
29	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336
31	14,457	15,655	17,042	17,538	19,280	21,433	24,255	25,390	26,440	30,336
32	15,134	16,362	17,783	18,291	20,072	22,271	25,148	26,304	27,373	31,336
33	15,815	17,073	18,527	19,046	20,866	23,110	26,042	27,219	28,307	32,336
34	16,501	17,789	19,275	19,806	21,664	23,952	26,938	28,136	29,242	33,336
35	17,191	18,508	20,027	20,569	22,465	24,796	27,836	29,054	30,178	34,336
36	17,887	19,233	20,783	21,336	23,269	25,643	28,735	29,973	31,115	35,336
37	18,584	19,960	21,542	22,105	24,075	26,492	29,636	30,893	32,053	36,336
38	19,289	20,691	22,304	22,878	24,884	27,343	30,537	31,815	32,992	37,336
39	19,994	21,425	23,069	23,654	25,695	28,196	31,441	32,737	33,932	38,335
40	20,706	22,164	23,838	24,433	26,509	29,050	32,345	33,660	34,872	39,335

Valores críticos  $\chi^2_{\alpha}(\nu)$  (continuación)

$\nu$	$\alpha$									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11,345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18,465
5	6,064	6,626	7,289	9,236	11,070	12,832	13,388	15,086	16,750	20,517
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	10,656	11,389	12,242	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	11,781	12,549	13,442	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	12,899	13,701	14,631	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	14,011	14,845	15,812	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	15,119	15,984	16,985	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	16,222	17,117	18,151	21,064	23,685	26,119	26,873	29,141	31,319	36,123
15	17,322	18,245	19,311	22,307	24,996	27,488	28,259	30,578	32,801	37,697
16	18,418	19,369	20,465	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	19,511	20,489	21,615	24,769	27,587	30,191	30,995	33,409	35,718	40,790
18	20,601	21,605	22,760	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	21,689	22,718	23,900	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	22,775	23,828	25,038	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	23,858	24,935	26,171	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	24,939	26,039	27,301	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	26,018	27,141	28,429	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	27,096	28,241	29,553	33,196	36,415	39,364	40,270	42,980	45,558	51,179
25	28,172	29,339	30,675	34,382	37,652	40,646	41,566	44,314	46,928	52,620
26	29,246	30,434	31,795	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	30,319	31,528	32,912	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	31,391	32,620	34,027	37,916	41,337	44,461	45,419	48,278	50,993	56,893
29	32,461	33,711	35,139	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	33,530	34,800	36,250	40,256	43,773	46,979	47,962	50,892	53,672	59,703
31	34,598	35,887	37,359	41,422	44,985	48,231	49,226	52,190	55,003	61,098
32	35,665	36,973	38,466	42,585	46,194	49,480	50,487	53,486	56,328	62,487
33	36,731	38,058	39,572	43,745	47,400	50,724	51,743	54,774	57,646	63,870
34	37,795	39,141	40,676	44,903	48,602	51,966	52,995	56,061	58,964	65,247
35	38,859	40,223	41,778	46,059	49,802	53,203	54,244	57,340	60,272	66,619
36	39,922	41,304	42,879	47,212	50,998	54,437	55,489	58,619	61,581	67,985
37	40,984	42,383	43,978	48,363	52,192	55,667	56,731	59,891	62,880	69,346
38	42,045	43,462	45,076	49,513	53,384	56,896	57,969	61,162	64,181	70,703
39	43,105	44,540	46,173	50,660	54,572	58,119	59,204	62,426	65,473	72,055
40	44,165	45,616	47,269	51,805	55,758	59,342	60,436	63,691	66,766	73,402

6. Distribución  $F$  de Fisher



(a) Valores críticos  $F_\alpha(v_1, v_2)$  para  $\alpha = 0,05$

	$v_1$								
$v_2$	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
$\infty$	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

**(b) Valores críticos  $F_{\alpha}(v_1, v_2)$  para  $\alpha = 0,05$**

	$v_1$									
$v_2$	10	12	15	20	24	30	40	60	120	$\infty$
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
$\infty$	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

**(c) Valores críticos  $F_\alpha(v_1, v_2)$  para  $\alpha = 0,01$** 

	$v_1$								
$v_2$	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5981	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
$\infty$	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

**(d) Valores críticos  $F_\alpha(v_1, v_2)$  para  $\alpha = 0,01$**

$v_2$	$v_1$									
	10	12	15	20	24	30	40	60	120	$\infty$
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
$\infty$	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

## 7. Algunas distribuciones discretas

NOMBRE	FUNCIÓN	PARÁMETROS	$E(X)$	$V(X)$
Uniforme	$f(x_k) = \frac{1}{n},$ $k = 1, 2, \dots, n$	$x_i < x_{i+1}$ $n \in \mathbb{N}$	$\frac{1}{n} \sum_{k=1}^n x_k$	$\frac{1}{n} \sum_{k=1}^n x_k^2 - \left( \frac{1}{n} \sum_{k=1}^n x_k \right)^2$
De dos puntos	$f(x_1) = p,$ $f(x_2) = 1 - p$	$x_1 < x_2$ $0 < p < 1$	$x_1 p + x_2 (1 - p)$	$(x_1 - x_2)^2 p(1 - p)$
Bernoulli	$f(0) = p,$ $f(1) = 1 - p$	$p$	$p$	$p(1 - p)$
Binomial	$\binom{n}{k} p^k (1 - p)^{n-k}$ $k = 0, 1, 2, \dots, n$	$0 < p < 1$ $n \in \mathbb{N}$	$np$	$np(1 - p)$
Poisson	$f(k) = \frac{1}{k!} e^{-\lambda} \lambda^k$ $k = 0, 1, 2, 3, \dots$	$\lambda > 0$	$\lambda$	$\lambda$
Hiper-geométrica	$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ $k \in \mathbb{N}_0, k \leq n,$ $k \leq M$	$M \in \mathbb{N}_0,$ $n, N \in \mathbb{N}$ $n \leq M \leq N$	$n \cdot \frac{M}{N}$	$na \left( \frac{N-n}{N-1} \right)$ $p = \frac{M}{N}$ $a = p(1 - p)$
Binomial negativa	$\binom{k+r-1}{r-1} p^r (1 - p)^k$ $k = 0, 1, 2, \dots$	$r > 0,$ $0 < p < 1$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Geométrica	$f(k) = p(1 - p)^k$ $k = 0, 1, 2, \dots$	$0 < p < 1$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$

## 8. Algunas distribuciones continuas

NOMBRE	FUNCIÓN	PARÁMETROS	$E(X)$	$V(X)$
Uniforme	$f(x) = \frac{1}{b-a},$ $a < x < b$	$a < b$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $x \in \mathbb{R}$	$\mu \in \mathbb{R},$ $\sigma^2 > 0$	$\mu$	$\sigma^2$
Normal estándar	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$ $x \in \mathbb{R}$		0	1
Gamma	$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta},$ $x > 0$	$\alpha > 0,$ $\beta > 0$	$\alpha\beta$	$\alpha\beta^2$
Exponencial	$f(x) = \lambda e^{-\lambda x},$ $x > 0$	$\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$t$ de Student	$f(x) = a_n (1 + n x^2)^{-(n+1)/2},$ $a_n := \frac{\Gamma\left(\frac{n+1}{2}\right) \sqrt{n}}{\Gamma\left(\frac{n}{2}\right) \sqrt{\pi}}, x \in \mathbb{R}$	$n \in \mathbb{N}$	0, $n \geq 2$	$\frac{n}{n-2},$ $n \geq 3$
Chi-cuadrada	$\frac{1}{a_n} x^{\frac{n}{2}-1} e^{-x/2},$ $a_n := 2^{n/2} \Gamma\left(\frac{n}{2}\right), x > 0$	$n > 0$	$n$	$2n$
$F$ de Fisher	$f(x) = \frac{a_n x^{\frac{m}{2}-1}}{(n+mx)^{(m+n)/2}}$ $a_n := \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)}, x > 0$	$m, n \in \mathbb{N}$	$\frac{n}{n-2},$ $n \geq 3$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},$ $n \geq 5$
Erlang	$\frac{(\lambda x)^{k-1}}{(k-1)!} \cdot \lambda e^{-\lambda x}$	$k \in \mathbb{N}, \lambda > 0$	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}, x > 0$

9. Resumen de distribuciones muestrales e intervalos de confianza

Cuadro A.1: Distribución de la media muestral

	¿FORMA DE LA POBLACIÓN?	¿ES $\sigma^2$ CONOCIDA?	¿TAMAÑO DE LA MUESTRA?	¿DISTRIBUCIÓN MUESTRAL?	¿Z Ó t?
1.	Normal	Sí	No importa	Normal	$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
2.		No	Grande ( $n \geq 30$ )	Normal	$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$
3.			Pequeño ( $n < 30$ )	t de Student, $v = n - 1$ grados de libertad	$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$
4.	No normal o desconocida	Sí	Grande ( $n \geq 30$ )	Normal	$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
5.			Pequeño ( $n < 30$ )	Callejón sin salida	
6.		No	Grande ( $n \geq 30$ )	Normal	$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$
7.			Pequeño ( $n < 30$ )	Callejón sin salida	

Cuadro A.2: Distribución relacionadas con proporciones

	¿ESTADÍSTICO?	¿SUPUESTO?	¿DIST. MUESTRAL	¿Z?
1.	Proporción muestral	$n \geq 30$	Normal	$Z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$
2.		$np \geq 5,$ $n(1-p) \geq 5$	Normal	
3.	Diferencia de proporciones muestrales	$n_1 \geq 30,$ $n_2 \geq 30$	Normal	$Z = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$
4.		$n_1 p_1 \geq 5,$ $n_1 (1-p_1) \geq 5,$ $n_2 p_2 \geq 5,$ $n_2 (1-p_2) \geq 5$	Normal	



**Cuadro A.3:** Distribución de la diferencias de medias muestrales

$X$  representa la población y para las dos últimas posibilidades de la tabla:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad v' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

	¿X?	¿ $\sigma_1^2$ y $\sigma_2^2$ SE CONOCEN?	¿ $\sigma_1^2 = \sigma_2^2$ ?	¿ $n_1$ y $n_2$ ?	¿DISTRIBUCIÓN MUESTRAL?	¿Z Ó t? $d := \bar{x}_1 - \bar{x}_2$ , $\mu := \mu_1 - \mu_2$
1.	No normal	Sí	No im- porta	Grandes $n_1 \geq 30$ , $n_2 \geq 30$	Normal	$Z = \frac{d - \mu}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
2.		No	No im- porta	Grandes $n_1 \geq 30$ , $n_2 \geq 30$	Normal	$Z = \frac{d - \mu}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
3.	Normal	Sí		No importa	Normal	$Z = \frac{d - \mu}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
4.		No	Sí	Pequeño $n_1 < 30$ , $n_2 < 30$	t de Student con $v = n_1 + n_2 - 2$ grados de libertad	$t = \frac{d - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
5.			No	Pequeño $n_1 < 30$ , $n_2 < 30$	v' grados de libertad  (redondear al en- tero más cercano)	$t = \frac{d - \mu}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

**Cuadro A.4:** Distribución de la varianza muestral y de la razón de varianzas muestrales

	ESTADÍSTICO	¿POBLACIÓN?	¿DISTRIBUCIÓN MUESTRAL?	¿ $\chi^2$ Ó F?
1.	$s^2$	Normal	Chi-cuadrada con $v = n - 1$ grados de libertad	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$
2.	$s_1^2 / s_2^2$	Ambas normales	F de Fisher con $v_1 = n_1 - 1$ , $v_2 = n_2 - 1$ grados de libertad	$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$  Regla: $F_{1-\alpha}(a, b) = \frac{1}{F_{\alpha}(b, a)}$

**Cuadro A.5:** Intervalos de confianza para la media poblacional

	¿POBLACIÓN?	¿ $\sigma^2$ CONOCIDA?	¿TAMAÑO MUESTRAL?	¿DISTRIBUCIÓN MUESTRAL?	¿INTERVALO? $\bar{x} - b < \mu < \bar{x} + b$ , con:
1.	Normal	Sí	No importa	Normal	$b := Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
2.		No	Grande ( $n \geq 30$ )	Normal	$b := Z_{\alpha/2} \frac{s}{\sqrt{n}}$
3.			Pequeño ( $n < 30$ )	$t$ de Student, $v = n - 1$ grados de libertad	$b := t_{\alpha/2} \frac{s}{\sqrt{n}}$
4.	No normal o desco- nocida	Sí	Grande ( $n \geq 30$ )	Normal	$b := Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
5.			Pequeño ( $n < 30$ )	Callejón sin salida	
6.		No	Grande ( $n \geq 30$ )	Normal	$b := Z_{\alpha/2} \frac{s}{\sqrt{n}}$
7.			Pequeño ( $n < 30$ )	Callejón sin salida	

**Cuadro A.6:** Intervalos para la proporción y para la diferencia de proporciones

	¿ESTADÍSTICO?	¿SUPUESTOS?	¿DISTR. MUESTRAL?	¿INTERVALO DE CONFIANZA? $\bar{p} - b < p < \bar{p} + b$ , con:
1.	Proporción muestral	$n \geq 30$	Normal	$b := Z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$
2.		$np \geq 5$ , $n(1-p) \geq 5$	Normal	
3.	Diferencia de proporciones muestrales	$n_1 \geq 30$ , $n_2 \geq 30$	Normal	$\bar{p} := \bar{p}_1 - \bar{p}_2$
4.		$n_1 p_1 \geq 5$ , $n_1(1-p_1) \geq 5$ , $n_2 p_2 \geq 5$ , $n_2(1-p_2) \geq 5$	Normal	$b := Z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

**Cuadro A.7:** Intervalos para la varianza y para la razón de varianzas

		¿POBLACIÓN?	¿DISTRIBUCIÓN MUESTRAL?	¿INTERVALO DE CONFIANZA?
1.	$s^2$	Normal	Chi-cuadrada con $v = n - 1$ grados de libertad	$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}$
2.	$s_1^2 / s_2^2$	Ambas normales	$F$ de Fisher con $v_1 = n_1 - 1$ , $v_2 = n_2 - 1$ grados de libertad	$\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot F_{\frac{\alpha}{2}}(v_2, v_1)$  Regla: $F_{1-\alpha}(a, b) = \frac{1}{F_{\alpha}(b, a)}$

**Cuadro A.8:** Intervalos de confianza para la diferencias de medias poblacionales

$X$  representa la población y para las dos últimas posibilidades de la tabla:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad v' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

	¿X?	¿ $\sigma_1^2$ y $\sigma_2^2$ SE CONOCEN?	¿ $\sigma_1^2 = \sigma_2^2$ ?	¿ $n_1$ y $n_2$ ?	¿DISTRIBUCIÓN MUESTRAL?	¿INTERVALO? $d - b < \theta < d + b$ , donde $d := \bar{x}_1 - \bar{x}_2$ $\theta := \mu_1 - \mu_2$ y:
1.	No normal	Sí	No importa	Grandes ( $n_1 \geq 30$ , $n_2 \geq 30$ )	Normal	$b := Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
2.		No	No importa	Grandes ( $n_1 \geq 30$ , $n_2 \geq 30$ )	Normal	$b := Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
3.	Normal	Sí	No importa	No importa	Normal	$b := Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
4.		No	Sí	Pequeño ( $n_1 < 30$ , $n_2 < 30$ )	$t$ de Student con $v = n_1 + n_2 - 2$ grados de libertad	$b := t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
5.			No	Pequeño ( $n_1 < 30$ , $n_2 < 30$ )	$t$ de Student con $v'$ grados de libertad (redondear al en- tero más cercano)	$b := t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

# B

## Guía rápida para trabajar con Statgraphics

### B.1 Análisis de un solo conjunto de datos

---

1. Abrir el archivo de datos **calles.sf3**.
2. Seleccionamos *Describe ... Numeric Data ... One-Variable Analysis*.
3. Elegimos *Data = Longitud* y pulsamos la opción *OK*.
4. Sale la llamada *ventana del análisis*. Los íconos principales de esta ventana son:
  - *Input dialog* (ícono de diálogos): para seleccionar o cambiar variables dentro del archivo y análisis seleccionado.
  - *Tabular options* (ícono de opciones tabulares): medidas estadísticas, percentiles, tablas de frecuencia, inferencias, etc.
  - *Graphical options* (ícono de opciones gráficas): diagramas de dispersión, histogramas, etc.
  - *Save results* (ícono de salvar resultados): permite salvar los resultados del análisis.
5. Transformación de una variable:<sup>1</sup> *One Variable Analysis*, activar el botón *Transform* y, en *Operators*, elegir *logaritmo*.

### B.2 Análisis simultáneo de dos o más conjuntos de datos

---

1. *Compare ... Two Samples ... Two Sample Comparison ...*
2. Para obtener diagramas de cajas múltiples: *Compare ... Multiple Samples ... Multiple-Sample Comparison ... Multiple Data Columns ... Ok ... Samples=* (en esta última opción mencionar los datos que queremos comparar)
3. Para obtener diagramas de cajas múltiples: *Plot ... Exploratory Plots ... Multiple Box-and-Whisker Plot ... Data=distancia ... Level codes=year ...*

---

<sup>1</sup>Por ejemplo, si quisiéramos trabajar con el logaritmo de la variable escribimos LOG(**longitud**) en vez de **longitud**.

## B.3 Gráficos de dispersión

---

Con la opción *Plot...Scatterplots* se pueden realizar:

1. Gráficos univariantes (*Univariate Plot*). Por ejemplo, abrir archivo de datos **autos.sf3** y utilizar la variable *mpg*.
2. Gráficos bidimensionales *X-Y* simples (*X-Y plot*) y múltiples (*Multiple X-Y Plot*). Por ejemplo, abrir archivo de datos **autos.sf3** y hacer *Y=mpg* y *X=potencia*. Sobre la gráfica, pulsar botón derecho del ratón y elegir *Pane options*. Aparece una pantalla con varios campos. Elegir *Point Codes=model*.
3. Gráficos tridimensionales *X-Y-Z* simples (*X-Y-Z plot*) y múltiples (*Multiple X-Y-Z Plot*). Por ejemplo, abrir archivo de datos **autos.sf3** y hacer *X=accel*, *Y=cilindro*, *Z=price*. Sobre la zona gráfica: botón derecho, *Pane options*, *Point Codes=origin*.
4. Gráficos de matriz (*Matriz Plot*).
5. Gráficos en coordenadas polares (*Polar Coordinates Plot*).

## B.4 Diagramas de presentación

---

Con la opción *Plot...Business Charts* se pueden realizar (abrir siempre el archivo **autos.sf3**):

1. Gráficos de barras simples (*Barchart*). Por ejemplo, realizar un gráfico de barras para la variable *origin* del archivo **autos.sf3**, que contiene el país de origen de los autos. Los valores de la variable *origin* son 1 para los autos norteamericanos, 2 para autos europeos y 3 para autos japoneses. En esta opción sale, entre otros, el campo *Counts* (Frecuencias) que permite introducir la variable que contiene las frecuencias absolutas de los valores de la variable a graficar. Como las frecuencias absolutas de los valores de la variable *origin* son: 85 para autos norteamericanos, 26 para autos europeos y 44 para autos japoneses, entonces, por esta razón, debemos escribir en este campo *join3(85;26;44)*. Además, el campo *Labels* (Etiquetas) permite introducir el nombre de la variable que contiene las etiquetas a utilizar para cada barra del gráfico. Como las etiquetas de los valores de la variable *origin* están contenidas *carmakers*, que son *America*, *Europe* y *Japan*, hacemos *Labels=carmakers*.
2. Gráficos de barras múltiples (*Multiple Barchart*). Por ejemplo, realizaremos un gráfico de barras dobles para las variables *origin* y *year* del archivo **autos.sf3**, que contienen el país de origen de los autos y el año de construcción, respectivamente. Los valores de la variable *year* son los intervalos 1978, [1979,1980] y [1981,1982]. Aparecen, entre otros, los siguientes campos:
  - *Columns* (Columnas): En este campo se introducen las variables que contienen las frecuencias absolutas de los valores de las variables a graficar, o una expresión de Statgraphics que contiene operadores y que genera sus valores. Como las frecuencias absolutas de los valores de la variable *origin* son: 85 para autos norteamericanos, 26 para autos europeos y 44 para autos japoneses, y como las frecuencias absolutas de los valores de la variable *year* son: 36 para 1978, 58 para [1979,1980] y 61 para [1981,1982], entonces, por esta razón, debemos escribir en este campo *join3(85;26;44)* y *join3(36;58;61)*.
  - *Labels* (Etiquetas): Hacemos *Labels=carmakers*.
3. Gráficos de sectores (*Piechart*). Por ejemplo, realizaremos un gráfico de sectores para la variable *origin* del archivo **autos.sf3**, que contienen el país de origen de los autos y el año de construcción, respectivamente. Los valores de la

variable *origin* son 1 para los autos norteamericanos, 2 para autos europeos y 3 para autos japoneses. Aparecen, entre otros, los siguientes campos:

- **Counts** (Frecuencias): En este campo se introducen las variables que contienen las frecuencias absolutas de los valores de las variables a graficar, o una expresión de Statgraphics que contiene operadores y que genera sus valores. Como las frecuencias absolutas de los valores de la variable *origin* son: 85 para autos norteamericanos, 26 para autos europeos y 44 para autos japoneses, entonces, por esta razón, debemos escribir en este campo *join3(85;26;44)*.
- **Labels** (Etiquetas): En este campo se debe introducir el nombre de la variable que contiene las etiquetas a utilizar para cada grupo de barras del gráfico. Como las etiquetas de los valores de la variable *origin* están contenidas *carmakers*, que son *America*, *Europe* y *Japan*, hacemos *Labels=carmakers*.

4. Gráficos de componentes de líneas (*Component Line Chart*)

5. Gráficos de escogencias alta y baja (*High-Low-Chose Chart*).

## B.5 Variables numéricas multidimensionales

---

Seleccione la siguiente secuencia de opciones: *Describe...Numeric Data...Multiple-Variable Analysis* y aparecen todas las variables del archivo. Aparece una ventana de diálogo en cuyo campo *Data* introducimos la variables *origin*, *price* y *year*. Luego, pulsamos el botón OK.

## B.6 Distribuciones de probabilidad

---

*Plot ... Probability Distributions*. Escogemos la distribución deseada. Los valores de los parámetros que definen la distribución (están fijados por defecto por el programa) los podemos modificar si pulsamos el botón derecho del ratón y escogemos la opción *Analysis Options*.

## B.7 Inferencias basadas en una sola muestra

---

1. Se escoge *Describe ... Numeric Data ... One Variable Analysis*. Elegimos la variable que va a ser objeto del análisis y pulsar OK. Al pulsar el ícono *Tabular options* aparecen, entre otros:

- **Confidence Intervals**.  
Calcula intervalos para la media (*Confidence Interval for Mean*) y la desviación típica (*Confidence Interval for Standard Deviation*) de la distribución. Pulsando el botón derecho del ratón y escogiendo *Pane Options* se puede modificar el nivel de confianza (*Confidence Level*) y el tipo de intervalo (*Interval Type*).
- **Hypothesis Testing**  
Se realizan los contrastes de la media y de la desviación típica. Pulsando el botón derecho del ratón y escogiendo *Pane options* se pueden modificar el valor del parámetro para la hipótesis nula (por ejemplo  $Mean = \mu_0$ ), del nivel de significancia  $\alpha$  (*Alpha*) y de la hipótesis alternativa:

## 2. Cálculo de la curva de potencia.

Describe ... Hypothesis Test ... Normal Mean y en Null Hypothesis se elige el valor de la media bajo la hipótesis nula. En la casilla Sample Sigma se escoge el valor de la desviación típica de la población. El tamaño de muestra se fija a través de Sample Size. Seleccionando el ícono de gráficos se selecciona la única gráfica posible (curva de potencia - Power Curve) y se pulsa OK.

## B.8 Inferencias basadas en dos muestras

---

1. Elegir Compare ... Two Samples, en donde aparecen cuatro (4) opciones: Two Sample Comparison, Paired-Sample Comparison, Hypothesis Tests, Sample-Size Determination.
2. Cuando seleccionamos Two Sample Comparison<sup>2</sup> el programa pide al usuario que especifique las dos columnas de datos a comparar (Sample 1 y Sample 2). Seleccionando Tabular options aparece, entre otros:
  - Comparison of Means: Intervalo de confianza para la diferencia de medias y contraste de igualdad de medias.
  - Comparison of Standard Deviations: Intervalo de confianza para el cociente de varianzas y contraste de igualdad de varianzas.
  - Kolmogorov-Smirnov Test: Prueba de hipótesis para saber si las distribuciones de ambas muestras son idénticas.

## B.9 Bondad de ajuste

---

1. Se selecciona Describe... Distribution Fitting...Uncensored Data. Al pulsar OK se obtiene, entre otras, la salida de los contrastes de bondad de ajuste.
2. Si, estando situados sobre esta salida, pulsamos el botón derecho del ratón y elegimos la opción Analysis Options del menú emergente resultante, obtenemos la caja de diálogo Probability Distributions Options, que presenta todas las posibles distribuciones a considerar para el ajuste (observamos que por defecto el ajuste se realiza a una distribución normal).
3. También aparecen los siguientes campos:
  - Number of Trials (número de ensayos).  
Se rellena con el número de tiradas cuando la distribución elegida para el ajuste es binomial;
  - Number of Successes (número de eventos).  
Se rellena con el número de éxitos cuando la distribución elegida es una binomial negativa.
  - Population Size (tamaño de la población).  
Se rellena con el tamaño de la población cuando la distribución elegida es una hipergeométrica.
4. La opción tabular Tests for Normality: realiza los contrastes de normalidad.
5. Opción tabular Goodness-of-Fit Tests: realiza los contrastes de la bondad de ajuste de los datos a una distribución dada.

---

<sup>2</sup>El procedimiento es idéntico cuando seleccionamos la opción Paired-Sample Comparison

# C

## Guía rápida para trabajar con SPSS

### C.1 Definición de las variables

---

Para definir cada variable hay dos procedimientos:

- Hacer doble clic sobre el encabezamiento de la variable o
- Seleccionar, en la parte inferior, la pestaña *vista de variables*.

Cuando se hace esto, observamos que hay una fila para cada variable del conjunto de datos y que existen 10 columnas: *Nombre, Tipo, Anchura, Decimales, Etiqueta, Valores, Perdidos, Columnas, Alineación y Medida*. La definición de una variable se basa en las opciones que se ofrecen en esa ventana:

1. *Asignar un nombre a cada variable*, cumpliendo las siguientes reglas:

- Nombres con no más de 8 caracteres (el primero debe ser una letra o @).
- No utilizar símbolos como &, /, \$, etc.
- No utilizar nunca espacios en blanco.
- No utilizar expresiones como ALL, AND, BY, EQ, GE, GT, LE, NE, NOT, OR, TO, o WITH.

2. *Asignar un tipo a cada variable*, indicando el máximo número de dígitos que deseamos para anotar las observaciones de la variable y el tipo de la variable con la que vamos a trabajar (*alfanumérica, fecha, moneda o numérica*) indicando en este caso el número de cifras decimales con que queremos que aparezca en el editor. SPSS permite trabajar con los siguientes tipos de variables:

- *Núméricas*: formato numérico estándar.
- *Coma*: comas de separación cada tres posiciones. Un punto para la parte decimal.
- *Punto*: al contrario que el anterior.
- *Notación Científica*: uso de la E para exponente.
- *Cadena*: variable alfanumérica (de más de 8 caracteres se considera larga).
- Además están los formatos de *fecha, dólar y moneda personalizada*.



Si no escogemos el tipo, el sistema lo asigna automáticamente, siendo el formato por defecto: *Númerica 8.2* que significa: Anchura: 8 y Decimales: 2; es decir, una amplitud de columna de 8 espacios, siendo los 2 últimos para los decimales.

3. *Asignar una Etiqueta a cada variable* de no más de 120 caracteres (entre 30 y 40 es el valor recomendado) que nos permita tener más información sobre esa variable.
4. *Asignar Valores*: se trata de asignar etiquetas a los valores de cada variable. No es obligatorio, pero sí muy útil en algunos casos.
5. *Definir Perdidos*: permite definir los valores de los datos especificados como perdidos por el usuario. Sitúese en el campo correspondiente a *Perdidos* de cualquier variable y pulse sobre el recuadro coloreado, aparece: Los códigos asignados a los valores ausentes deben de ser coherentes con el tipo de variables declarado: numéricos para las numéricas y alfanuméricos para las alfanuméricas (máximo 9 caracteres). Se pueden introducir hasta 3 valores perdidos (individuales) de tipo discreto, un rango de valores perdidos o un rango más un valor de tipo discreto. Sólo pueden especificarse rangos para las variables numéricas. Estos valores ausentes son denominados por SPSS “valores ausentes definidos por el usuario” (*user-defined missing values*), a diferencia de los definidos por el sistema (*system-missing values* o *sysmis*). Estos últimos corresponden a los que establece el sistema para los espacios en blanco y caracteres ilegales que puedan haber en el archivo de datos. Aparecen en los listados representados por comas.
6. *Definir Columnas*: consiste en especificar la amplitud de la columna. Podemos hacerlo también desde el propio archivo de datos.
7. *Definir Alineación*: seleccionar la justificación de las entradas de la columna: *Izquierda*, *Derecha* y *Centrado*.
8. *Especificar medida*. Se puede seleccionar uno de los tres niveles de medida:
  - *Escala*: los valores de datos son numéricos en una escala de intervalo. Las variables de escala deben ser numéricas.
  - *Ordinal*: los valores de datos representan categorías con un cierto orden intrínseco (bajo, medio, alto; totalmente de acuerdo, de acuerdo, en desacuerdo). Las variables ordinales pueden ser de cadena o valores numéricos. Notar que para variables de cadena ordinales, se asume que el orden alfabético de los valores de cadena indica el orden correcto de las categorías; en el caso de bajo, medio y alto el orden sería alto, bajo y medio (orden que no es correcto), por lo que es más fiable utilizar códigos numéricos para representar datos ordinales que usar etiquetas de estos códigos.
  - *Nominal*: los valores de datos representan categorías sin un cierto orden intrínseco. Las variables nominales pueden ser de cadena o valores numéricos que representan categorías diferentes, por ejemplo *1 = Hombre* y *2 = Mujer*.

### C.1.1. Transformación de una variable

Elegimos *Transformar ... Calcular*, y realizamos los siguientes pasos:

- a) Asignar un nombre y un tipo (por defecto será numérica) a la nueva variable en el cuadro de texto de la *Variable de destino*.
- b) Definir la expresión numérica que va a permitir calcular los valores de la misma. Para ello utilizaremos los nombres de las variables del archivo (podemos escribirlos o seleccionarlos del listado que aparece), constantes, operadores y funciones.
- c) Pulsar *Aceptar*.

Para construir estas expresiones pueden usarse operadores aritméticos como +, -, \*, /, \*\* y funciones como SQRT, EXP, LG10, LN, ARTAN, COS, SIN, ABS, MOD10, TRUNC, RND, entre otras:

- MOD10 (Resto resultante de dividir entre 10).
- TRUNC (Parte entera de un número).
- RND (Redondeo al entero más cercano).

Pulsando el botón derecho sobre el nombre de la función, aparece su descripción. El argumento de las funciones debe ir entre paréntesis. Existen funciones particulares como UNIFORM y NORMAL, que se utilizan para la generación de variables aleatorias. Son de bastante utilidad en estudios de simulación.

Es importante tener cuidado con el orden de utilización de los operadores y no olvidar que los valores antiguos pierden su vigencia al recodificar una variable sobre el mismo nombre.

El botón *SI...* permite realizar modificaciones similares, pero sujetas a que se verifique una condición lógica. Se incluirán aquellos casos que verifiquen la condición. Los que no la cumplan pasarán a ser valores ausentes definidos por el sistema.

Una expresión lógica es una expresión que puede ser evaluada como verdadera o falsa en función de los valores de las variables en ella relacionadas. El nexos de las variables son los operadores de relación: =, >, <=, <, >, ~=. Es posible formar expresiones complejas, utilizando los operadores lógicos: AND (&), OR (|), NOT (~).

### C.1.2. Recodificación de una Variable

A partir de una variable podemos crear otra cuyos valores sean una recodificación de los de la primera. Esta recodificación podemos hacerla tanto en la misma variable como en variables diferentes. Para ello, seleccionaremos *Transformar ... Recodificar ... En distintas variables*. Se abre una ventana en la que deberemos asignar un nombre (y una etiqueta si queremos) a la nueva variable.<sup>1</sup>

### C.1.3. Filtrado de datos

El programa SPSS permite seleccionar determinados casos para un próximo proceso, bien temporalmente o de forma permanente, sobre la base de un criterio lógico o de una decisión aleatoria. Para ello seleccionaremos el menú *Datos ... Seleccionar casos*. La selección de individuos puede ser temporal (*filtrados*) o permanente (*eliminados*). En la selección permanente eliminamos del archivo activo los individuos deseados, mientras que en la temporal, la selección es recuperable (los casos son filtrados). En esta última situación, los individuos (casos) del archivo que no satisfacen la condición aparecerán marcados como excluidos mediante una línea que cruza en diagonal su número de fila. Aparece también una variable llamada *filter\$* que el sistema crea para controlar el filtrado de datos.

Especificaciones:

- *Todos los casos*: indica que quiere procesar todos los casos del archivo de datos de trabajo.
- *Si se satisface la condición*: indica que quiere procesar sólo los casos que satisfagan una condición lógica. Para especificar o cambiar la condición, pulse en *Sí*. Esta alternativa crea la variable *filter\$*, que el sistema crea para controlar el filtrado de datos.

---

<sup>1</sup> Cuidado!, si se selecciona ... borrarás la variable original.

- **Muestra aleatoria de casos:** indica que queremos seleccionar los casos de forma aleatoria para su procesamiento. Si ha tecleado las especificaciones de muestreo, éstas aparecerán junto al botón de comando Muestra. Si no, o si quiere cambiarlas, pulse en *Muestra* (véase más adelante). Esta alternativa también crea la variable *filter\$*.
- **Basándose en el rango del tiempo o de los casos:** permite seleccionar los casos deseados siempre que sean consecutivos.
- **Usar variable de filtro:** indica que quiere utilizar los valores de una variable numérica existente para controlar el filtrado de casos. Seleccione la variable de la lista de la izquierda. Los casos cuyo valor sea 0, o ausentes, en la variable de filtro se excluyen del análisis.

## C.2 Análisis exploratorio de datos

---

Primero abrir el archivo de datos.

- a) **Tablas de frecuencias:** *Analizar ... Estadísticos descriptivos ... Frecuencias*. SPSS también cuenta con el menú alternativo *Analizar ... Tablas personalizadas* que posibilita alterar el formato del resultado.
- b) **Estadísticos:** *Analizar ... Estadísticos descriptivos ... Descriptivos* donde hay que seleccionar la variable o variables de interés y después *Opciones* para escoger los estadísticos que interesan. Sin embargo con este menú no se pueden obtener los percentiles. Para obtenerlos hay que usar *Analizar ... Estadísticos descriptivos ... Frecuencias* y entrar en la opción *Estadísticos* en donde se seleccionan los percentiles deseados.
- c) **Gráficos de sectores:** *Gráficos ... Sectores* y seleccionaremos una o varias variables apareciendo un cuadro de diálogo, cuyas opciones pasamos a comentar:
  - 1) *Resúmenes para grupos de casos:* Genera un gráfico en el que cada sector corresponde a un valor de la variable seleccionada. El tamaño del sector se determina por la opción *Los sectores representan*, esta opción aparece en el cuadro de diálogo que surge después de pulsar el botón *Definir* del cuadro anterior. También es posible que los sectores representen otra cosa, como la media de los valores de otra variable, el valor máximo, etc.; esto se consigue con la opción *Otra función resumen*. Se puede también editar el gráfico haciendo doble clic sobre él, con posibilidad de cambiar colores, tramas, desgajar sectores, etc.
  - 2) *Resúmenes para distintas variables.* Permite que los sectores representen variables en lugar de grupos de casos. Cada sector representa una función de una determinada variable (por ejemplo, la suma de los valores de sus casos).
  - 3) *Valores individuales de los casos.* Se resume una única variable, los casos ya son valores agrupados de la variable. Cada sector representa el valor de un caso individual. Con *Gráficos ... Interactivos ... Sectores* podemos obtener representaciones con efectos más llamativos.
- d) **Diagramas de barras:** *Gráficos ... Barras* y *Gráficos ... Interactivos ... Barras*.
- e) **Histogramas:** *Gráficos ... Histograma* o *Gráficos ... Interactivos ... Histograma*.
- f) **Gráficos de tallo y hojas:** *Analizar ... Estadísticos descriptivos ... Explorar*.
- g) **Diagramas de caja:** *Gráficos ... Diagrama de cajas*.

- h) **Diagramas de dispersión:** *Gráficos ... dispersión ... simple* o *Gráficos ... Interactivos ... Diagrama de dispersión*, en donde aparece un cuadro de diálogo en el que se puede elegir qué variable ocupará el eje X y qué otra el eje Y.

## C.3 Inferencia sobre una o más poblaciones

---

Primero abrir el archivo de datos.

- a) **Análisis de una muestra:** *Analizar ... Comparar medias ... Prueba T para una muestra*. Aparece una pantalla en cuyo campo *Contrastar Variables* introducimos las variables que queremos contrastar. En esta ventana, seleccionamos *Opciones*, para introducir el grado de confianza deseado (por defecto es del 95%). Al final se pulsa *Aceptar*.
- b) **Análisis de dos muestras emparejadas o relacionadas (Prueba T para muestras relacionadas).** Para efectuar la prueba T para muestras relacionadas se necesita una columna en los datos para cada una de las variables a comparar. Seleccionamos *Analizar ... Comparar medias ... Prueba T para muestras relacionadas*. Aparece la ventana en donde seleccionamos las variables en cuya comparación estamos interesados. Al hacer la primera selección en la columna de variables, esta aparece en el recuadro selecciones actuales como *variable 1*, y al realizar la segunda selección aparecerá como *variable 2*. En ese momento, ya seleccionadas las dos, es cuando las podemos introducir en la columna variables relacionadas. Se pulsa *Aceptar*.
- c) **Análisis de dos muestras independientes (Prueba T para muestras independientes).** El programa necesita una columna en el editor de datos que contenga los valores de la variable cuyas medias en las dos poblaciones se desea comparar, y otra que indica la población o grupo a que pertenece cada individuo. A continuación, seleccionamos *Analizar ... Comparar medias ... Prueba T para muestras independientes*. Aparece una ventana en donde, en primer lugar seleccionamos una variable numérica y con el puntero la situamos en la ventana de *Contrastar variables*. A continuación, seleccionamos una única variable de agrupación y pulsamos *Definir grupos*. En esta ventana debemos especificar los dos grupos de la variable de contraste, eligiendo entre:
- *Usar valores especificados*. Escribimos un valor para el Grupo 1 y otro para el Grupo 2. Los casos con otros valores quedarán excluidos.
  - *Punto de corte*. Escribimos un número que divida los valores de la variable de agrupación en dos conjuntos.

Si la variable de agrupación es de cadena corta, por ejemplo, *SI* y *NO*, podemos escribir una cadena para el Grupo 1 y otra para el Grupo 2. Los casos con otras cadenas quedarán excluidos del análisis. Una vez completada la ventana y tras pulsar *Continuar*, volvemos a la ventana de *Prueba T para muestras independientes*. Pulsando el botón *Opciones* podemos introducir un valor entre 1 y 99 para el coeficiente de confianza de un intervalo, cuyo valor por defecto es del 95%. Tras pulsar el botón *Aceptar*.

- d) **Pruebas de normalidad.** *Analizar ... Estadísticos descriptivos ... Explorar*. Aparece la ventana *Explorar*. En el caso de una muestra situamos la variable en la ventana *Dependientes*, y dejamos *Factores* en blanco. Para dos muestras independientes, situamos la variable a contrastar en la ventana *Dependientes*, y la variable que forma los grupos en la de *Factores*. Para dos muestras emparejadas situamos una variable con la diferencia de las dos originales en la ventana *Dependientes*, y dejamos *Factores* en blanco. A continuación, debemos pulsar el botón *Gráficos* y en la nueva ventana escoger la opción de *Histograma* y activar la opción de *Gráficos con pruebas de normalidad*.



# D

## Uso de la calculadora en la estadística

Las explicaciones las basaremos en la utilización de las calculadoras Casio fx-82MS, fx-83MS, fx-85MS, fx-270MS, fx-300MS y fx-350MS.

### Cálculos estadísticos

Para realizar cálculos estadísticos en la calculadora, tenga en cuenta los siguientes comentarios:

- Utilice **MODE** **2** para ingresar el modo estadístico SD.
- Utilice **SHIFT** **CLR** **1** **=** para borrar la memoria.
- Ingrese los datos usando la secuencia de tecla siguiente: <Dato> **DT**.
- Tenga en cuenta la tabla siguiente para los cálculos que se necesiten:

Para llamar este tipo de valor:	Realice esta operación:
$\sum x^2$	<b>SHIFT</b> <b>S-SUM</b> <b>1</b>
$\sum x$	<b>SHIFT</b> <b>S-SUM</b> <b>2</b>
$n$	<b>SHIFT</b> <b>S-SUM</b> <b>3</b>
$\bar{x}$	<b>SHIFT</b> <b>S-VAR</b> <b>1</b>
$\sigma_n$	<b>SHIFT</b> <b>S-VAR</b> <b>2</b>
$\sigma_{n-1}$	<b>SHIFT</b> <b>S-VAR</b> <b>3</b>

### Ejemplo D.1

Calcule  $n$ ,  $\sum x$ ,  $\sum x^2$ ,  $\bar{x}$ ,  $\sigma_n$  y  $\sigma_{n-1}$  para los datos siguientes: 55, 54, 51, 55, 53, 53, 54 y 52.

SOLUCION:

- Primero, ingresamos al modo SD con las teclas **MODE** **2**.
- Luego, borramos la memoria con la secuencia de teclas **SHIFT** **CLR** **1** **=**.
- Posteriormente, ingresamos los datos: 55 **DT** 54 **DT** 51 **DT** 55 **DT** 53 **DT** 53 **DT** 54 **DT** 52 **DT**.
- Por último, calculamos las medidas estadísticas pedidas:

Suma de los cuadrados de los valores  $\sum x^2 = 22,805$

SHIFT S-SUM 1 =

Suma de valores  $\sum x = 427$

SHIFT S-SUM 2 =

Número de datos  $n = 8$

SHIFT S-SUM 3 =

Media aritmética  $\bar{x} = 53,375$

SHIFT S-VAR 1 =

Desviación estándar poblacional  $\sigma_n = 1,316956719$

SHIFT S-VAR 2 =

Desviación estándar muestral  $\sigma_{n-1} = 1,407885953$

SHIFT S-VAR 3 =

## Precauciones con el ingreso de datos

- **DT DT** ingresa el mismo dato dos veces.
- También puede ingresar múltiples entradas del mismo dato usando **SHIFT ;**. Por ejemplo, para ingresar el dato 110 diez veces presiones 110 **SHIFT ;** 10 **DT**.
- Mientras ingresa datos o después de completar el ingreso de datos, puede usar las teclas **Δ** y **▽** para ir visualizando a través de los datos que ha ingresado.
- Si ingresa múltiples ingresos del mismo dato usando **SHIFT ;** para especificar la frecuencia de datos (número de ítemes de datos) como se describe anteriormente, pasando a través de los datos muestra el ítem de dato y una pantalla separada para la frecuencia de datos (freq).
- Los datos visualizados pueden editarse, si así lo desea. Ingrese el valor nuevo y presione la tecla **=** para reemplazar el valor antiguo por el valor nuevo. Esto también significa que si desea realizar alguna otra operación (cálculo, llamada de resultados de cálculos estadísticos, etc.), siempre deberá presionar primero la tecla **AC** para salir de la presentación de datos.
- Presionando la tecla **DT** en lugar de **=** después de cambiar un valor sobre la presentación, registra el valor que ha ingresado como un elemento de dato nuevo, y deja el valor antiguo tal como está.
- Puede borrar el valor del dato visualizado usando **Δ** y **▽**, y luego presionando **SHIFT CL**. Borrando un valor de dato ocasiona que todos los valores siguientes se desplacen hacia arriba.
- Después de ingresar los datos en el modo SD, no podrá visualizar o editar más los datos ítemes de datos individuales, después de cambiar a otro modo.

# Bibliografía

- [1] AGRESTI, A., *Categorical data analysis*. John Wiley and Sons, Inc., New York, 1990.
- [2] BARBOSA, R.; LLINÁS, H., *Procesos estocásticos con aplicaciones*, Barranquilla: Editorial Universidad del Norte, 2013.
- [3] HOSMER, D. and LEMESHOW S., *Applied Logistic Regression*, Segunda edición, John Wiley and Sons, 2000.
- [4] KALB, K. y KONDER, P., *Una visión histórica del concepto moderno de integral*, Barranquilla: Editorial Universidad del Norte, 2006 (editor: Dr. rer. nat. Humberto Llinás).
- [5] KLEINBAUM, D. and KLEIN, M., *Logistic Regression: A self Learning Text*, Segunda edición, Ed. Springer, 2002.
- [6] LLINÁS, H.; ROJAS, C., *Estadística descriptiva y distribuciones de probabilidad*. Barranquilla: Ediciones Uninorte, 2005.
- [7] LLINÁS, H., *Precisiones en la teoría de los modelos logísticos*, Revista Colombiana de Estadística, Volumen 29, Número 2, pág. 239-265, 2006.
- [8] LLINÁS, H., *Estadística inferencial*, Barranquilla: Editorial Universidad del Norte, 2006.
- [9] LLINÁS, H., *Medida e integración*. Barranquilla: Editorial Universidad del Norte, 2007.
- [10] LLINÁS, H., *Applet: La ley de los grandes números*. Se puede encontrar en el siguiente link:  
<http://ylang-ylang.uninorte.edu.co/Objetos/Estadistica/LeyDeGrandesNumeros/index.html>
- [11] LLINÁS, H., *Applets de estadística*, 2007. Se puede encontrar en el siguiente link:  
<http://ylang-ylang.uninorte.edu.co:8080/drupal/?q=node/238>
- [12] LLINÁS, H.; ALONSO, J. FLÓREZ, K., *Introducción a la estadística con aplicaciones en Ciencias Sociales*, Barranquilla: Editorial Universidad del Norte, 2012.
- [13] LLINÁS, H., *Introducción a la estadística matemática*, Barranquilla: Editorial Universidad del Norte, 2014.
- [14] LLINÁS, H., *Introducción a la teoría de probabilidad*, Barranquilla: Editorial Universidad del Norte, 2014.
- [15] NELDER, J.A. and WEDDERBURN, R.W.M., *Generalized linear models*. The Journal of the Royal Statistical Society, serie A 135, pág.370-384, 1972.
- [16] PÉREZ, C., *Estadística práctica con Statgraphics*. España: Prentice Hall, 2002.
- [17] Página web de datos estadísticos del Institute for Digital Research and Education (IDRE) de la Universidad de California en Los Angeles (UCLA): <https://stats.idre.ucla.edu/>. En especial, consultar: <https://stats.idre.ucla.edu/other/examples/alr2/>