

Examen de Introducción a la Econometría (LECO).

Departamento de Economía Cuantitativa. Universidad Complutense de Madrid.

11 de junio de 2007. Duración: **2 horas y 30 minutos**.

Apellidos:

Nombre:

DNI:

Nombre del profesor:

Grupo:

No desgrape las hojas de este cuadernillo. El examen está compuesto por tres problemas y 10 cuestiones cortas. Elija **SÓLO DOS DE LOS TRES** problemas propuestos y respóndalos en hojas aparte. Las cuestiones cortas deben ser respondidas en el espacio debajo del enunciado en este mismo cuadernillo. Junto a cada pregunta y apartado se indica la puntuación.

Problemas

Problema 1. (2.5 pt.) Sea la función de densidad: $f(x, y) = k$, $0 < x < 6$, $\frac{x}{2} < y < 3$.

- (0.25 pt) Dibuje el soporte de $f(x, y)$. El soporte es el triángulo rectángulo de vértices $(0, 0)$, $(6, 3)$, $(0, 3)$.
- (0.25 pt) ¿Cuál es el valor de k que hace de $f(x, y)$ una función de densidad?

$$1 = \int_0^6 \int_{x/2}^3 k dy dx = \int_0^6 k \left(3 - \frac{x}{2}\right) dx = 9k; \quad k = \frac{1}{9}.$$

- (0.25 pt) Compruebe si X e Y son independientes.

$$f_1(x) = \int_{x/2}^3 \frac{1}{9} dy = \frac{6-x}{18}, 0 < x < 6; \quad f_2(y) = \int_0^{2y} \frac{1}{9} dx = \frac{2y}{9}, 0 < y < 3.$$

Es evidente que el producto $f_1(x) \cdot f_2(y)$ no coincide con la función de densidad conjunta por lo que **no son independientes**.

- (0.25 pt) Calcule $E(Y/X = x)$ y represéntela sobre el soporte.

$$E(Y/X = x) = \int_{x/2}^3 y \frac{\frac{1}{9}}{\frac{6-x}{18}} dy = \int_{x/2}^3 y \frac{2}{6-x} dy = \frac{9 - \frac{x^2}{4}}{6-x} = \frac{36 - x^2}{4(6-x)} = \frac{6+x}{4}, 0 < x < 6.$$

La representación gráfica es el segmento $(0, 3/2)$, $(6, 3)$.

- (0.5 pt) Calcule $P[X + Y < 3]$.

Una vez añadida la recta $y = 3 - x$ al soporte y solucionando para x , se aprecia que los límites del suceso $X + Y < 3$ son $0 < x < 2$ y $3 - x < y < x/2$. Así la probabilidad pedida es:

$$P[X + Y < 3] = \int_0^2 \left[\int_{x/2}^{3-x} \frac{1}{9} dy \right] dx = \int_0^2 \frac{1}{9} \left(3 - \frac{3}{2}x\right) dx = \frac{1}{3}.$$

6. (0.5 pt) Calcule $P[Y \leq 2/X \leq 2]$.

$$P[Y \leq 2/X \leq 2] = \frac{P[Y \leq 2, X \leq 2]}{P[X \leq 2]} = \frac{1/3}{5/9} = \frac{3}{5},$$

donde:

$$P[Y \leq 2, X \leq 2] = \int_0^2 \left[\int_{x/2}^2 \frac{1}{9} dy \right] dx = \frac{1}{3},$$

$$P[X \leq 2] = \int_0^2 \frac{6-x}{18} dx = \frac{5}{9}.$$

7. (0.5 pt) Calcule $P[Y \leq 2/X = 2]$.

Dado el valor $X = 2$, el soporte para $Y/X = 2$ es el intervalo $(1, 2)$. La probabilidad pedida es:

$$P[Y \leq 2/X = 2] = \int_1^2 f(y/X = 2) dy = \int_1^2 \frac{f(2, y)}{f_1(2)} dy = \int_1^2 \frac{1}{2} dy = \frac{1}{2}.$$

Problema 2. (2.5 pt.) Sea X el tiempo en minutos que transcurre entre dos llamadas que llegan a una centralita telefónica. Se supone que $X \sim f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$, $x \geq 0$; esto es, sigue una distribución exponencial con parámetro θ . Se pide:

1. (0.5 pt) Calcule $E(X)$ y $V(X)$.

$$E(X) = \int_0^{\infty} x \frac{1}{\theta} e^{-\frac{1}{\theta}x} dx = \int_0^{\infty} z e^{-z} \theta dz = \theta \Gamma(2) = \theta,$$

donde se ha hecho el cambio de variable $z = \frac{1}{\theta}x$. Para obtener la varianza es conveniente calcular primero el momento de segundo orden:

$$E(X^2) = \int_0^{\infty} x^2 \frac{1}{\theta} e^{-\frac{1}{\theta}x} dx = \int_0^{\infty} \theta z^2 e^{-z} \theta dz = \theta^2 \Gamma(3) = 2\theta^2.$$

Así, la varianza resulta: $V(X) = E(X^2) - E(X)^2 = 2\theta^2 - \theta^2 = \theta^2$.

2. (0.5 pt) Formule la función de verosimilitud para una muestra tamaño n .

Suponiendo muestreo aleatorio simple, la función de verosimilitud es el producto de las funciones de densidad de los elementos de la muestra:

$$\mathcal{L}(\theta/x_1, \dots, x_n) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{1}{\theta}x_i} = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum x_i}.$$

3. (0.5 pt) Encuentre la región crítica más potente para contrastar $H_0 : \theta = 1$ frente a la alternativa $H_1 : \theta = 3$.

Para encontrar la mejor región crítica recurrimos el Teorema de Neyman-Pearson, que prueba que dicha región crítica se obtiene del cociente de la función de verosimilitud en la nula entre la verosimilitud en la alternativa. Así:

$$\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)} = \frac{e^{-\frac{1}{\theta_0} \sum x_i} / \theta_0^n}{e^{-\frac{1}{\theta_1} \sum x_i} / \theta_1^n} = \left(\frac{\theta_1}{\theta_0} \right)^n \exp \left[- \left(\frac{1}{\theta_0} - \frac{1}{\theta_1} \right) \sum_i x_i \right] \leq k \Rightarrow RH_0.$$

Ahora es necesario transformar la región crítica en un suceso equivalente en función de un estadístico cuya distribución sea conocida bajo la nula:

$$\begin{aligned} \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left[-\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \sum_{i=1}^n x_i\right] &\leq k, \\ \exp\left[-\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \sum_{i=1}^n x_i\right] &\leq k', \\ -\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \sum_{i=1}^n x_i &\leq k'', \\ \sum_{i=1}^n x_i &\geq k'''. \end{aligned}$$

donde en el último paso hay que tener en cuenta que, dadas las hipótesis nula y alternativa, $-\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) < 0$.

Es posible que todavía haya que transformar el estadístico para conocer su distribución suponiendo que la hipótesis nula es cierta, como se ilustra en el siguiente apartado.

4. (0.5 pt) Sabiendo que el estadístico $\frac{2}{\theta} \sum x_i \sim \chi_{2n}^2$, donde x_i son los elementos de una muestra aleatoria simple tamaño n , encuentre el valor crítico para un contraste al 5% de significación con $n = 5$.

El nivel de significación es la probabilidad de rechazar la hipótesis nula suponiendo que es cierta. Bajo la hipótesis nula, la región crítica es $2 \sum x_i \geq c^*$ y la distribución del estadístico queda $2 \sum x_i \sim \chi_{2n}^2$. El valor crítico se obtiene de:

$$\alpha = 0,05 = P\left[2 \sum x_i \geq c^* / 2 \sum x_i \sim \chi_{10}^2\right],$$

y buscando en las tablas resulta $c^* = 18,31$.

5. (0.5 pt) Si con la muestra anterior se obtiene $\bar{x} = 1,5$, ¿cuál es el p-valor del contraste? ¿Qué decisión adoptaría sobre H_0 ?

Si $\bar{x} = 1,5$, entonces $2n\bar{x} = 2 \sum x_i = 15$. El p-valor se calcula:

$$\alpha^* = P\left[2 \sum x_i \geq 15 / 2 \sum x_i \sim \chi_{10}^2\right] = 0,1321.$$

Con las tablas disponibles en el examen, sólo se puede decir que $\alpha^* > 0,10$. A los niveles de significación habituales (1%, 5% y 10%) no se puede rechazar la hipótesis nula: hay evidencia de que transcurre un minuto entre llamadas.

Problema 3. (2.5 pt.) Es frecuente oír comentarios sobre si el rendimiento de los estudiantes se evalúa de forma similar por diferentes profesores de la misma asignatura. Se dispone de las calificaciones de dos conjuntos de estudiantes **pertenecientes a dos grupos distintos de la misma asignatura** que aparecen en la siguiente tabla:

Grupo A	5.6	5.0	7.2	5.5	5.9	4.3	2.8	6.8	6.6	3.8
Grupo E	5.8	5.6	5.5	4.8	5.9	8.5	7.0			

Para los datos anteriores se han calculado los siguientes estadísticos:

	Media	Varianza	Asimetría	Curtosis	Mediana
Grupo A	5.35	1.76	-0.43	2.19	5.55
Grupo E	6.16	1.28	1.03	2.98	5.80

- (0.5 pt) Contraste si los datos del Grupo A proceden de una distribución normal. Calcule el p-valor.

Con la información disponible, el contraste de normalidad que se puede aplicar es el Jarque-Bera:

$$JB = n \left[\frac{AS^2}{6} + \frac{(K-3)^2}{24} \right] = 10 \left[\frac{(-0,43)^2}{6} + \frac{(2,19-3)^2}{24} \right] = 0,5763.$$

El estadístico sigue una distribución χ_2^2 , por lo que el valor crítico al 5% de significación es $c^* = 5,99$. El p-valor, con las tablas disponibles, es $\alpha^* > 0,7$. No se puede rechazar la normalidad de las calificaciones del Grupo A.

- (0.5 pt) Suponiendo que para las calificaciones del Grupo E se acepta el mismo resultado que ha obtenido en el apartado 1, ¿cómo contrastaría que el nivel de exigencia es igual en ambos grupos? Formule la hipótesis nula, la alternativa, escriba el estadístico que emplearía y justifique la respuesta.

Puesto que se supone que las calificaciones son normales (se ha contrastado para el Grupo A), el estadístico apropiado es el de igualdad de esperanzas bajo normalidad. En este caso, $H_0 : \mu_A = \mu_E$ frente a la alternativa $H_0 : \mu_A \neq \mu_E$. El estadístico de contraste en la hipótesis nula es:

$$z = \frac{\bar{x}_A - \bar{x}_E}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_E}\right) \frac{n_A s_A^2 + n_E s_E^2}{n_A + n_E - 2}}} \sim t_{n_A + n_E - 2}$$

Teniendo en cuenta la hipótesis alternativa, la región crítica es $\{z \leq -c^* \cup z \geq c^*\}$, donde c^* es el valor de la distribución $t_{n_A + n_E - 2}$ (t_{15}) que deja a su derecha $\alpha/2$ (la mitad de la significación).

- (0.5pt) Aplique el contraste propuesto en el apartado anterior. Indique si rechazaría la hipótesis nula y calcule el p-valor.

Para calcular el estadístico hay que tener en cuenta que se dan las varianzas en vez de las cuasi-varianzas muestrales. Las cuasi-varianzas resultan: $s_A^2 = \frac{10}{9} S_A^2 = 1,96$ y $s_E^2 = \frac{7}{6} S_E^2 = 1,49$. El estadístico de contraste es:

$$z = \frac{5,35 - 6,16}{\sqrt{\left(\frac{1}{10} + \frac{1}{7}\right) \frac{10 \cdot 1,96 + 7 \cdot 1,49}{10 + 7 - 2}}} = -1,16.$$

Utilizando las tablas de la distribución t_{15} , el valor crítico al 5% de significación (bilateral) es $c^* = 2,13$. Puesto que $z \in (-2,13; 2,13)$, no se puede rechazar la hipótesis nula. El p-valor de este contraste, con las tablas disponibles, es superior a 0,2.

4. (0.5 pt) ¿Cómo cambiaría su respuesta al **apartado 2** si **NO** pudiese suponer normalidad de los datos de calificaciones?

Si no se puede suponer normalidad de los datos y dado que, además, las muestras son pequeñas, sería necesario recurrir a un contraste no paramétrico. En este caso se dispone de muestras independientes de variables aleatorias continuas, por lo que el contraste adecuado sería el de Mann-Whitney sobre igualdad de medianas.

Si m denota la mediana, la hipótesis nula es $H_0 : m_A = m_E$, frente a la alternativa de que las medianas de las calificaciones de los grupos son diferentes. El estadístico U (que aparece en el «formulario»), basado en las sumas de los rangos asignados a los elementos de las dos muestras, sería el apropiado.

5. (0.5 pt) Utilice el contraste de Kolmogorov-Smirnov para comprobar si los datos del Grupo E siguen la misma distribución que los del Grupo A. ¿Es equivalente el resultado del apartado 3 y el obtenido en éste?

Las calificaciones pertenecen al intervalo $[0, 10]$, por lo que una selección de puntos razonable sería $z = 2, 4, 6, 8$. Aplicándolo a los datos:

z	2.0	4.0	6.0	8.0
$F_{10}^*(z)$	0.00	0.20	0.70	1.00
$G_7^*(z)$	0.00	0.00	0.71	0.86
$ F_{10}^*(z) - G_7^*(z) $	0.0000	0.2000	0.0143	0.1429

El estadístico de contraste es $D_{10,7} = 0,2$. Aproximando el valor crítico al 5%, resulta $c_{0,05}^* = 1,22\sqrt{\frac{10+7}{10 \cdot 7}} = 0,6012$. Por lo tanto, no se puede rechazar la hipótesis nula de que las calificaciones de ambos grupos siguen la misma distribución.

El resultado de este contraste es más fuerte que el obtenido en el apartado 3, puesto que aquí se ha aceptado igualdad de distribución, bajo normalidad esto implica igualdad de esperanzas y varianzas.

Cuestiones cortas. Debe obtener más 2 puntos en estas preguntas para que se califiquen los problemas.

Cuestión 1. (0.5 pt) Una variable aleatoria bidimensional $(X, Y)'$, se distribuye normal bivalente con vector de esperanzas $(0, 0)'$ y matriz de varianzas covarianzas $\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$. ¿Cuál es la probabilidad de que $(\frac{1}{2}X - Y)$ sea mayor que 1,28?

Si se define $Z = (\frac{1}{2}X - Y)$, dadas las propiedades de la normal, $Z \sim N(0, \frac{1}{4}2 + \frac{1}{2})$; esto es, $Z \sim N(0, 1)$. La respuesta es: $P(Z > 1,28) = 0,10$.

Cuestión 2. (0.5 pt) Considere una economía con sólo dos bienes de consumo: A y B . Se tienen los siguientes datos:

	Bien A		Bien B	
	precio	cantidad	precio	cantidad
1998	5	30	2	20
1999	6	35	4	10

Tomando como base el año 1998, ¿cuál es la inflación anual en 1999?

El índice de precios de Laspeyres en el año 1 con base año 0 para un conjunto de bienes $i = 1, \dots, n$ viene dado por la expresión:

$$I_{1/0}^L = \frac{\sum_{i=1}^n p_i^1 q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} = \frac{6 \cdot 30 + 4 \cdot 20}{5 \cdot 30 + 2 \cdot 20} = 1,3684.$$

El índice de precios se suele expresar 136,84% y el índice en el año base es 100%. La inflación interanual, que es la tasa de variación, resulta $\pi_{99/98} = \frac{I_{1/0}^L - I_{0/0}^L}{I_{0/0}^L} = 36,8\%$.

Cuestión 3. (0.5 pt) Sea (X, Y) un par aleatorio que representa la demanda de dos bienes **sustitutivos**. Se sabe que $E(X) = 5$ y $E(Y) = 10$. Si se produce una demanda de tres unidades de X , ¿esperaría que la demanda de Y fuese mayor que 10 ó menor que 10? ¿Por qué?

Dado que son sustitutivos, existe una relación inversa; esto es, mayores valores de una de las variables implica mayores probabilidades de valores pequeños de la otra. Así, dado un valor de X inferior a su esperanza marginal, cabe esperar que $E(Y/X = 3) > E(Y) = 10$.

Cuestión 4. (0.5 pt) La función generatriz de momentos de una v.a. chi-cuadrado de r grados de libertad es $M(t) = (1 - 2t)^{-r/2}$. Demuestre que si se tienen n variables aleatorias independientes $X_i \sim \chi_{r_i}^2$, $i = 1, \dots, n$, entonces $\sum_{i=1}^n X_i \sim \chi_{\sum r_i}^2$.

Si $Y = \sum a_i X_i$, con X_i independientes y las a_i constantes conocidas, $M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t)$. Así:

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (1 - 2t)^{-r_i/2} = (1 - 2t)^{-\frac{1}{2} \sum r_i},$$

que es la generatriz de una v. a. $\chi_{\sum r_i}^2$.

Cuestión 5. (0.5 pt) Si una variable aleatoria sigue una distribución exponencial con parámetro θ , $X \sim f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$, $x \geq 0$, obtenga su función generatriz de momentos.

La función generatriz de una v.a. X se define $M_X(t) = E(e^{tX})$, pero además debe cumplirse que $M(0) = 1$ y $\exists M_X(t) \forall t \in (-h, h)$, esto es, debe existir en un intervalo alrededor de cero. En este caso:

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \frac{1}{\theta} e^{-\frac{1}{\theta}x} dx = \frac{1}{\theta} \int_0^{\infty} e^{-(\frac{1}{\theta}-t)x} dx,$$

si se hace el cambio de variable $(\frac{1}{\theta} - t)x = z \Rightarrow x = \frac{\theta}{1-\theta t}z$, resulta:

$$M_X(t) = \frac{1}{\theta} \int_0^{\infty} e^{-(\frac{1}{\theta}-t)x} dx = \frac{1}{\theta} \int_0^{\infty} e^{-z} \frac{\theta}{1-\theta t} dz = \frac{1}{\theta} \frac{\theta}{1-\theta t} \Gamma(1) = (1-\theta t)^{-1}.$$

Se puede comprobar que $M_X(0) = 1$ y que $\exists M_X(t) \forall t \in (-\frac{1}{\theta}, \frac{1}{\theta})$, por lo que es una función generatriz.

Cuestión 6. (0.5 pt) **Utilizando los resultados de las dos cuestiones previas**, para una muestra aleatoria simple de tamaño n obtenida de una distribución exponencial con parámetro θ , demuestre que $\frac{2n}{\theta} \bar{x} \sim \chi_{2n}^2$.

En primer lugar, es necesario notar que $Y = \frac{2n}{\theta} \bar{x} = \frac{2}{\theta} \sum X_i = \sum \frac{2}{\theta} X_i$. También, como se indicó antes, si $Y = \sum a_i X_i$, con X_i independientes y las a_i constantes conocidas, $M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t)$. Entonces:

$$M_Y(t) = \prod_{i=1}^n M_X\left(\frac{2}{\theta}t\right) = \left[M_X\left(\frac{2}{\theta}t\right)\right]^n = \left[\left(1 - \theta \left\{\frac{2}{\theta}t\right\}\right)^{-1}\right]^n = (1 - 2t)^{-n},$$

que es la función generatriz de momentos de una χ_{2n}^2 .

Enunciado para las dos cuestiones siguientes. Un profesor de la Facultad, saliendo a las 8:00 de su casa (y evitando los túneles de la M-30), ha empleado 25 minutos de media en llegar al Campus de Somosaguas las últimas seis veces que ha venido. La desviación típica de dichos tiempos de trayecto ha sido de ocho minutos. Suponga normalidad de los tiempos de trayecto.

Cuestión 7. (0.5 pt) Proporcione un intervalo del 95 % de confianza para el tiempo que empleará el próximo día que venga a la Facultad.

Para calcular el intervalo de confianza se parte del estadístico pivote:

$$P\left[\lambda_1 \leq \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \leq \lambda_2\right] = 0,95,$$

donde los valores $\lambda_1 = -\lambda_2$ proceden de la distribución t_5 . Despejando en el intervalo definido dentro de la probabilidad se obtiene:

$$\left[\bar{x} - \lambda_2 \sqrt{s^2/n}, \bar{x} - \lambda_1 \sqrt{s^2/n}\right],$$

y sustituyendo por los datos teniendo en cuenta que $s^2/n = S^2/(n-1)$:

$$\left[25 - 2,5706 \sqrt{\frac{8^2}{5}}, 25 + 2,5706 \sqrt{\frac{8^2}{5}}\right] = [15,80; 34,20]$$

Cuestión 8. (0.5 pt) ¿A qué hora debe salir de casa este profesor si quiere tener una confianza del 95 % de llegar a tiempo a su clase de las 8:30?

Debería salir con la anticipación necesaria para que la hora de salida más la peor duración del trayecto con un nivel de confianza del 95 % fuese anterior a la hora de comienzo de la clase. En el apartado anterior se ha estimado que con un nivel de confianza del 95 % puede tardar 34,2 minutos (34 minutos y 12 segundos) en el peor de los casos, por lo que debería salir 4,2 minutos antes de las 8 para llegar a clase a su hora.

Cuestión 9. (0.5 pt) Sea la siguiente función de cuantía conjunta para X e Y :

$P(X = x, Y = y)$	$y = 0$	$y = 1$	$y = 2$
$x = 0$	0,33	0,12	0,05
$x = 1$	0	0,25	0,25

¿Son independientes? Calcule $E(Y/X = 1)$.

A la vista de la tabla, $P(X = 1, Y = 0) = 0$, pero $P(X = 1) = 0,50$ y $P(Y = 1) = 0,12$, por lo que el producto de marginales no coincide con la conjunta: **no son independientes**.

Las función de cuantía condicionada es: $P(Y = y/X = 1) = \frac{P(Y=y, X=1)}{P(X=1)}$, $y = 1, 2, 3$. La esperanza condicionada pedida resulta:

$$E(Y/X = 1) = \sum_{y=1}^3 y \frac{P(Y = y, X = 1)}{P(X = 1)} = 0 \frac{0}{0,5} + 1 \frac{0,25}{0,5} + 2 \frac{0,25}{0,5} = 1,5.$$

Cuestión 10. (0.5 pt) A partir de la tabla de la **cuestión 9**, obtenga la función generatriz de momentos marginal de Y .

$$M_Y(t) = E(e^{tY}) = \sum_{y=1}^3 e^{ty} P(Y = y) = e^{t0}0,33 + e^{t1}0,37 + e^{t2}0,30 = 0,30e^{2t} + 0,37e^t + 0,33.$$

Es fácil comprobar que $M_Y(0) = 1$ y que existe para cualquier valor $t \in \mathbb{R}$.

Fórmulas de posible utilidad

Transformación de variables. Sea $X \sim f_X(x)$ y se define $Y = h(X)$. Entoces $f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right|$ donde $h^{-1}(\cdot)$ es la *función inversa* de $h(\cdot)$.

Aproximación lineal a la esperanza condicional.

$$E^*(Y/X = x) = E(Y) - \frac{\text{Cov}(X, Y)}{V(X)} E(X) + \frac{\text{Cov}(X, Y)}{V(X)} \cdot x.$$

Varianza condicional de la normal bivalente. $V(Y/X = x) = V(Y)(1 - \rho_{XY}^2)$.

Modelo de regresión lineal. Sea $E(Y_i/X_i = x_i) = a + bx_i$ (o $Y_i = a + bx_i + U_i$, $U_i \stackrel{iid}{\sim} N(0, \sigma^2)$). Si \hat{a} y \hat{b} son los estimadores por el método de los momentos de a y b , entonces:

$$\frac{\hat{a} - a}{\sqrt{\frac{\hat{\sigma}^2 \sum x_i^2}{T \sum (x_i - \bar{x})^2}}} \sim t_{n-2}; \quad \frac{\hat{b} - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}} \sim t_{n-2}.$$

Distribuciones de funciones de variables aleatorias. Si $X \sim N(\mu, \sigma^2)$, $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$ e independientes y se dispone de muestras de tamaños n , n_1 y n_2 respectivamente:

$$\frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1), \quad \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \sim t_{n-1}, \quad \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1},$$

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1), \quad \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1 + n_2 - 2},$$

donde en todas las expresiones anteriores, s^2 denota la *cuasivarianza* muestral.

Proporciones. $\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0, 1)$. Con dos poblaciones y muestras de tamaños n y m :

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\binom{n+m}{n,m} \hat{p}_T (1 - \hat{p}_T)}} \rightarrow N(0, 1),$$

donde $\hat{p}_T = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m}$.

Contraste de Jarque-Bera. $JB = n \left[\frac{AS^2}{6} + \frac{(K-3)^2}{24} \right] \rightarrow \chi_2^2$.

Contraste Chi cuadrado. $\sum_{i=1}^k \frac{(T_i - O_i)^2}{T_i} \sim \chi^2$ donde T_i y O_i son, respectivamente las i -ésimas frecuencias absolutas esperadas y observadas.

Contrastes de Kolmogorov-Smirnov. Para una muestra $D_n = \sup |F_n^*(x) - F(x)|$. Para dos muestras $D_{n,m} = \sup |F_n^*(x) - G_m^*(x)|$. $F_n^*(x)$ y $G_m^*(x)$ son funciones de distribución empíricas (o muestrales) y $F(x)$ es una función de distribución teórica.

Contraste de Wilcoxon. El estadístico $T = T^+ - T^-$, bajo H_0 cumple $E(T) = 0$ y $V(T) = \frac{n(n+1)(2n+1)}{6}$.

Contraste de Mann-Whitney. $U = \min(U_1, U_2)$, donde $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ y $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$. Bajo H_0 se cumple $E(U) = \frac{n_1 n_2}{2}$ y $V(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$.

Aproximación a los valores críticos en los contrastes de Kolmogorov-Smirnov. Para el contraste de una muestra, el valor crítico c^* con un nivel de significación α se aproxima mediante $c_\alpha^* = k_\alpha \sqrt{1/n}$, donde k_α es 1.07, 1.22, 1.36, 1.52 y 1.63 para niveles de significación del 20 %, 10 %, 5 %, 2 % y 1 %, respectivamente.

Para el contraste de dos muestras, el valor crítico aproximado se calcula:

$$c_\alpha^* = k_\alpha \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}},$$

donde k_α es 1.07, 1.22 y 1.52 para niveles de significación α del 10 %, 5 % y 1 %, respectivamente.

Tablas estadísticas

r	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.999
1	0.3249	0.7265	1.3764	3.0777	6.3138	12.7062	31.8205	318.3088
2	0.2887	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	22.3271
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	10.2145
4	0.2707	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	7.1732
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	5.8934
6	0.2648	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	5.2076
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	4.7853
8	0.2619	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	4.5008
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	4.2968
10	0.2602	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	4.1437
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	4.0247
12	0.2590	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.9296
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.8520
14	0.2582	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	3.7874
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	3.7328
16	0.2576	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	3.6862
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	3.6458
18	0.2571	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	3.6105
19	0.2569	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	3.5794
20	0.2567	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	3.5518

Cuadro 1: Función de distribución de la t_r .

r	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.999
2	0.45	0.71	1.02	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	13.82
4	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	11.14	13.28	18.47
6	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.64	12.59	14.45	16.81	22.46
8	4.59	5.53	6.42	7.34	8.35	9.52	11.03	13.36	15.51	17.53	20.09	26.12
10	6.18	7.27	8.30	9.34	10.47	11.78	13.44	15.99	18.31	20.48	23.21	29.59
12	7.81	9.03	10.18	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	32.91
14	9.47	10.82	12.08	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	36.12
16	11.15	12.62	13.98	15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	39.25
18	12.86	14.44	15.89	17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	42.31
20	14.58	16.27	17.81	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	45.31

Cuadro 2: Función de distribución de la χ_r^2 .

	x.xx	x.x1	x.x2	x.x3	x.x4	x.x5	x.x6	x.x7	x.x8	x.x9
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	0.8413	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	0.9279	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	0.9772	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	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	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	0.9987	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

Cuadro 3: Función de distribución de la $N(0, 1)$