Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid

Examen de Introducción a la Econometría

10 de septiembre de 2005, 9:15h.
Duración: 2 HORAS y 15 MINUTOS

Apellidos:	Nombre:	DNI:
Profesor/a:		Grupo:

No desgrape estas hojas

Problemas largos (2 puntos por problema)

Los siguientes problemas deben ser realizados en un **ÚNICO** cuadernillo a parte

(*escriba sus datos personales, grupo y profesor en dicho cuadernillo*[1])

Elija **SÓLO** dos (2) de los tres (3) problemas propuestos.

Ejercicio 1. Hemos encuestado a 500 empresarios de Madrid y a 450 de Barcelona sobre cuáles son sus expectativas de comercio para los próximos 12 meses con respecto a su situación actual. Las 5 posibles respuestas son: mucho peor (-2), peor (-1), similar (0), mejor (1) y mucho mejor (2). El resultado de la encuesta se resume en la siguiente tabla:

	mucho peor (-2)	peor (-1)	similar (0)	mejor (1)	mucho mejor (2)
Madrid	100	150	125	100	25
Barcelona	68	157	112	68	45

Queremos contrastar si la visión que tienen los empresarios madrileños acerca de su actividad económica futura es diferente de la de los catalanes. Proponga y resuelva dos contrastes apropiados (Kolmogorov y Chi-cuadrado) para abordar esta pregunta. Para cada contraste, use niveles de significación del 1 % y del 10 %. Comente los resultados.

Solución:

Kolmogorov-Smirnov Calculando la suma acumulada para el caso de Madrid, es decir:

Madrid	-2	-1	0	1	2
suma acumulada	100	250	375	475	500

y dividiendo por el numero de observaciones (500 para Madrid), obtenemos la función de distribución empírica para Madrid.

Procediendo de manera similar en el caso de Barcelona, obtenemos las siguientes func. de distribución empíricas y las diferencias entre ellas:

Func. dist	-2	-1	0	1	2
$F_{Madr}(x)$	0.20	0.50	0.75	0.95	1
$F_{Barc}(x)$	0.15	0.50	0.75	0.90	1
diferencias	0.05	0.00	0.00	0.05	0

Así pues el estadístico Kolmogorov-Smirnov es la máxima diferencia entre ambas funciones de distribución $D_{n,m} = \sup |F_{Madr}(x) - F_{Barc}(x)| = 0.05$.

Los valores críticos para niveles de significación del 1 % y 10 % son respectivamente

- $\alpha_{0.01,\ 500,\ 450} = 1.52 \cdot \sqrt{\frac{500+450}{500 \cdot 450}} = 0.098$
- $\alpha_{0.1,\ 500,\ 450} = 1.07 \cdot \sqrt{\frac{500+450}{500 \cdot 450}} = 0.069$

[1] Al finalizar entregue estas paginas dentro de dicho cuadernillo

Así pues no rechazamos H_0 ni al 1 % ni al 10 % de nivel de significación.

Chi cuadrado Las frecuencias teóricas son las siguientes:

	-2	-1	0	1	2
Madrid	88.4	161.6	124.7	88.4	36.8
Barcelona	79.6	145.4	112.3	79.6	33.2

La realización del estadístico Chi-cuadrado para este contraste es 16.163, que tenemos que comparar con los valores tabulados de una chi-cuadrado de 4 grados de libertad al 1 % y 10 % de significación. Estos valores son 7,78 y 13,28, respectivamente. En ambos casos, por tanto, rechazamos la hipótesis nula de homogeneidad de muestras. Nótese que la respuesta es la contraria que en el caso del Kolmogorov. En el caso del Chi-cuadrado, si la confianza fuese del 99,9 %, no rechazaríamos la nula de homogeneidad, por ejemplo.

Fin **ejercicio** 1

Ejercicio 2. (Consta de 5 apartados)

Sea X el precio de un determinado producto que su empresa va a comenzar a comercializar. Suponemos que dicho precio se distribuye $X \sim N(\mu\,, 100)$. Usted afirma que el precio medio más adecuado es de 30 euros, y para contrastar esta hipótesis frente a la de su jefe, que afirma que es de 40 euros, selecciona al azar una muestra de precios de compra de 25 clientes, y adopta la siguiente regla de decisión:

- Si la media muestral es inferior a 35, se considera adecuado fijar un precio de 30 euros.
- Si por el contrario la media muestral es igual o superior a 35, se considera adecuado fijar el precio *alternativo* de 40 euros.

(a) A partir de su afirmación y la de su jefe escriba la hipótesis nula y la alternativa; y a partir de la regla decisión escriba la región crítica del contraste.

Solución: H_0: $\mu = 30$; H_1: $\mu = 40$; $RC = \{\mathbf{x} \,|\, \overline{x} \geq 35\}$ Fin **ejercicio** 2

(b) Obtenga el nivel de significación del contraste.

Solución: Por definición $\alpha = P(\overline{x}() \geq 35)$ bajo H_0 $(\mu = 30)$; y puesto que $X \sim N(\mu\,, 100)$, su media muestral también se distribuye $N(\,,\,)$ con esperanza μ y varianza $100/25$; así pues, bajo H_0

$$\alpha = P(\overline{x}() \geq 35) = P\left(\frac{\overline{x}() - 30}{\sqrt{100/25}} \geq \frac{35 - 30}{\sqrt{100/25}}\right) = P(Z \geq 2.5) = \boxed{0.0062}$$

Fin **ejercicio** 2

(c) ¿Cúal es la potencia del contraste?

Solución: La potencia es $W(40) = P(\overline{x}() \geq 35)$ bajo H_1 $(\mu = 40)$;

$$W(40) = P(\overline{x}() \geq 35) = P\left(\frac{\overline{x}() - 40}{\sqrt{100/25}} \geq \frac{35 - 40}{\sqrt{100/25}}\right) = P(Z \geq -2.5) = \boxed{0.9938}$$

Fin **ejercicio** 2

(d) Obtenga la región crítica para un nivel de significación del 1 %.

Solución: Ahora buscamos un k tal que $P(\overline{x}() \geq k) = 0.01$ bajo H_0 $(\mu = 30)$; por tanto

$$P(\overline{x}() \geq k) = P\left(\frac{\overline{x}() - 30}{\sqrt{100/25}} \geq \frac{k - 30}{\sqrt{100/25}}\right) = P\left(Z \geq -\frac{k - 30}{\sqrt{100/25}}\right)$$

pero dicha probabilidad debe ser igual a 0.01, así que $\frac{k-30}{\sqrt{100/25}} = 2.32$, es decir $k = 34.64$, y la región crítica buscada es

$$RC = \{\mathbf{x} \text{ tales que } \overline{x} \geq 34.64\}$$

Fin **ejercicio** 2

(e) Para el nivel de significación anterior, resuelva el contraste suponiendo que la media muestral obtenida ha sido 36.

Solución: Puesto que la media es mayor que 34.64, la muestra pertenece a la región crítica RC y, por tanto, se rechaza H_0 con un nivel de significación del 1 %. Fin **ejercicio** 2

Ejercicio 3. (Consta de 5 apartados)

Sea (X, Y) una variable aleatoria bidimensional discreta, donde X e Y pueden tomar los valores 0, 1 ó 2. La variable aleatoria X hace referencia al volumen de venta de bebidas alcohólicas en un chiringuito de playa en un día cualquiera, mientras que Y hace mención al volumen de ventas de refrescos el mismo día en el mismo establecimiento. Definimos las variables de la siguiente manera: $X = 0$ si se vende poco, $X = 1$ si el volumen de ventas es medio y $X = 2$ si se vende mucho. De manera análoga definimos Y. La ley de probabilidad conjunta viene dada por la siguiente función de cuantía:

X \ Y	0	1	2
0	0.12	0.16	0.12
1	0.08	0.12	0.16
2	0.04	0.08	0.12

(a) Calcule las leyes de probabilidad marginales de X e Y. Según esta información, ¿que situación es más probable respecto a las ventas de alcohol?

Solución:

X \ Y	0	1	2	$P_X(x)$
0	0.12	0.16	0.12	0.40
1	0.08	0.12	0.16	0.36
2	0.04	0.08	0.12	0.24
$P_Y(y)$	0.24	0.36	0.40	

Por lo tanto,

- $P\,(\text{poco}) = P_X(0) = 0.40$
- $P\,(\text{medio}) = P_X(1) = 0.36$
- $P\,(\text{mucho}) = P_X(2) = 0.24$

es decir, lo más probable es vender pocas bebidas alcohólicas. Fin **ejercicio** 3

(b) Halle las leyes de probabilidad condicionadas de X con respecto a los diferentes valores que pueda tomar Y. ¿Varían entre ellas y con respecto a la ley de probabilidad marginal de X? ¿Que significa la respuesta a la pregunta anterior? Si $Y = 0$, ¿cambiaría la respuesta cualitativa del apartado anterior? ¿Y si $Y = 2$?

Solución: Las leyes de probabilidad condicionadas están resumidas en las columnas de la siguiente tabla.

	$P_{X\|Y}(x\,\|\,0)$	$P_{X\|Y}(x\,\|\,1)$	$P_{X\|Y}(x\,\|\,2)$
X=0	0.500	$0.4\overset{\frown}{4}$	0.300
X=1	$0.33\overset{\frown}{3}$	$0.3\overset{\frown}{3}$	0.400
X=2	0.167	$0.2\overset{\frown}{2}$	0.300

Las probabilidades han cambiado. Esto nos indica que las variables son dependientes. Nueva información sobre X hace cambiar las leyes de probabilidad sobre las ventas de bebidas alcohólicas; no obstante, si la venta de refrescos ha sido baja (Y=0) sigue siendo más probable que la venta de alcohol sea baja. Sin embargo, si la venta de refrescos ha sido elevada (Y=2) lo más probable es que haya un volumen intermedio de ventas de bebidas alcohólicas.

Fin **ejercicio** 3

(c) Calcule la correlación entre la venta de refrescos y la de bebidas alcohólicas. Interprete el resultado.

Solución: Necesitamos calcular las desviaciones típicas y la covarianza de X e Y.

Por una parte, $\mathrm{E}(X) = 0.36 + 2 \cdot 0.24 = 0.84$ y por otra $\mathrm{E}\left(X^2\right) = 0.36 + 2^2 \cdot 0.24 = 1.32$

Por tanto

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \left[\mathrm{E}(X)\right]^2 = 1.32 - \left[0.84\right]^2 = 0.6144$$

En cuanto a Y, $\mathrm{E}(Y) = 0.36 + 2 \cdot 0.40 = 1.16$ y $\mathrm{E}\left(Y^2\right) = 0.36 + 2^2 \cdot 0.40 = 1.96$

Así pues,

$$\mathrm{Var}(Y) = \mathrm{E}\left(Y^2\right) - \left[\mathrm{E}(Y)\right]^2 = 1.96 - \left[1.16\right]^2 = 0.6144$$

La covarianza es

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) = \\ &= 0.12 + 2 \cdot 0.16 + 2 \cdot 0.08 + 4 \cdot 0.12 \;-\; 0.84 \cdot 1.16 = 0.1056\end{aligned}$$

Finalmente el coeficiente de correlación es

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{0.1056}{\sqrt{0.6144}\sqrt{0.6144}} = 0.17188$$

Existe una débil relación lineal de signo positivo entre las variables.

Fin **ejercicio** 3

(d) El dueño del chiringuito considera "mal día" a aquel en el que las ventas bajas de un tipo de bebida, no son compensadas con ventas elevadas del otro tipo de bebida (es decir si $X+Y<2$). ¿Cúal es la probabilidad de que se dé un "mal día"?

Solución: Los casos que generan un mal día se pueden resumir en aquellos que verifican $X+Y<2$. Así pues,

$$P(X+Y<2) = P_{XY}(0,0) + P_{XY}(0,1) + P_{XY}(1,0) = 0.36$$

Fin **ejercicio** 3

(e) ¿Cómo cambian las respuestas del apartado anterior si se sabe con anterioridad que $Y=0$, que $Y=1$, ó que $Y=2$ (use la información de las probabilidades condicionadas?)

Solución: Tenemos que reescribir estas probabilidades en términos de las leyes de probabilidad condicionadas:

- Si $Y=0$:

$$P(X+Y<2 \mid Y=0) = P_{X|Y}(0 \mid 0) + P_{X|Y}(1 \mid 0) = 0.50 + 0.333 = 0.833$$

- Si $Y=1$:

$$P(X+Y<2 \mid Y=1) = P_{X|Y}(0 \mid 1) = 0.44$$

- Si $Y=2$:

$$P(X+Y<2 \mid Y=2) = 0$$

Fin **ejercicio** 3

Preguntas cortas (0.5 puntos por pregunta)

Las siguientes cuestiones deben ser respondidas en el **recuadro** que queda a continuación de cada pregunta.

Ejercicio 4. Sea X los tipos de interés a un año. Tenemos una muestra aleatoria simple con 40 datos. El coeficiente de curtosis es 3,852 y el de asimetría de -0,305. Queremos contrastar si se distribuye normal usando el estadístico de Jarque-Bera. Halle el p-valor de este contraste y comente el valor obtenido en relación a los distintos posible niveles de significación para realizar el contraste.

Solución: Para la muestra disponible $JB = 40\left[\frac{(-.305)^2}{6} + \frac{((3.852)-3)^2}{24}\right] = 1.83$

Así pues, el p-valor de esta muestra es la probabilidad de observar un estadístico JB mayor o igual a 1.83 bajo la hipótesis nula de normalidad, es decir, la probabilidad de que una variable aleatoria con distribución $\chi^2_{(2)}$ sea mayor o igual a 1.83:

$$P\left(JB() \geq 1.83\right) = 0.40$$

Este es un p-valor elevado, esto quiere decir que para niveles de significación inferiores al 40 % (normalmente se emplean niveles del 10 % o menos) no se puede rechazar H_0.

Fin **ejercicio** 4

Ejercicio 5. La función generatriz de momentos de $Z \sim N\left(0\,,\,1\right)$ "evaluada en t" es: $M_Z(t) = e^{\frac{1}{2}t^2}$. A partir de esta información, demuestre que la función generatriz de momentos de $Y = \sigma Z + \mu$ es $e^{\mu t + \frac{1}{2}t^2\sigma^2}$.

Solución:

$$M_Y(t) = \mathrm{E}\left(e^{Yt}\right) = \mathrm{E}\left(e^{(\sigma Z+\mu)t}\right) = \mathrm{E}\left(e^{\sigma Zt}\cdot e^{\mu t}\right) = e^{\mu t}\mathrm{E}\left(e^{\sigma tZ}\right).$$

Pero $\mathrm{E}\left(e^{\sigma tZ}\right)$ es la función generatriz de momentos de una $N\left(0\,,\,1\right)$ "evaluada en σt", es decir, $\mathrm{E}\left(e^{\sigma tZ}\right) = M_Z(\sigma t)$; por tanto

$$M_Y(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t}e^{\frac{1}{2}(\sigma t)^2} = e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

Fin **ejercicio** 5

Ejercicio 6. Sean $X \sim N\left(\mu_X\,,\,\sigma^2_X\right)$ e $Y \sim N\left(\mu_Y\,,\,\sigma^2_Y\right)$ los rendimientos de dos activos financieros. Se dispone de dos muestras de tamaños n y m respectivamente y se quiere hacer un contraste para discutir si el primer rendimiento tiene más incertidumbre que el segundo. Proponga las hipótesis nula y alternativa más apropiadas para llevar a cabo este contraste. Proponga el estadístico de contraste y la región crítica para llevar a cabo este contraste con un nivel de significación α. (NO resuelva el contraste, limitese a contestar lo que se le pide)

Solución: Las hipótesis son $H_0 : \sigma^2_X = \sigma^2_Y$ y $H_1 : \sigma^2_X > \sigma^2_Y$

El estadístico de contraste es $\frac{\mathfrak{s}^2_{\mathbf{x}}/\sigma^2_X}{\mathfrak{s}^2_{\mathbf{y}}/\sigma^2_Y} \sim F_{n-1,m-1}$

La región crítica es $RC = \left\{\mathbf{x}\,\middle|\,\frac{\mathfrak{s}^2_{\mathbf{x}}}{\mathfrak{s}^2_{\mathbf{y}}} > k_\alpha\right\}$,

Fin **ejercicio** 6

Ejercicio 7. Queremos contrastar una hipótesis nula simple frente a una alternativa también simple. Para este propósito, disponemos de dos regiones críticas con idénticos niveles de significación. ¿Quiere esto decir que los dos contrastes son igualmente buenos? Razone la respuesta.

Solución: **NO.** Que ambos contrastes posean idénticos niveles de significación significa que en ambos existe la misma probabilidad de cometer el error tipo I (rechazar H_0 siendo cierta). Pero la probabilidad de cometer el error tipo II puede diferir. De entre ambos, preferiremos el contraste más potente (el que tenga mayor probabilidad de rechazar H_0 cuando es falsa).

Fin **ejercicio** 7

Ejercicio 8. Sea la siguiente función de densidad

$$f_{XY}(x,y) = \begin{cases} k & \text{si } 0 < x < 2y < 4 \\ 0 & \text{en el resto de casos} \end{cases}$$

Usando la función de densidad conjunta, ¿cómo calcularía la probabilidad $P\left(Y > 1\right)$? (NO lo calcule, pero indique correctamente los órdenes de integración)

Solución: Hay dos posibles respuestas

1. $$P\left(Y > 1\right) = \int_{1}^{2}\int_{0}^{2y} k\ dx\ dy$$

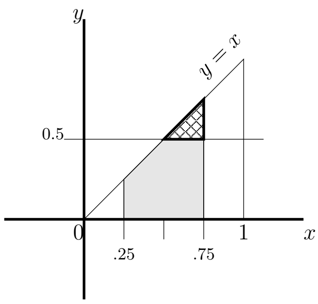
2. $$P\left(Y > 1\right) = \int_{0}^{2}\int_{1}^{2} k\ dy\ dx + \int_{2}^{4}\int_{x/2}^{2} k\ dy\ dx$$

Fin **ejercicio** 8

Ejercicio 9. Sea la siguiente función de densidad

$$f_{XY}\left(x,y\right) = \begin{cases} k(x+y) & \text{si } 0 < y < x < 1 \\ 0 & \text{en el resto de casos} \end{cases}.$$

¿Cómo calcularía la probabilidad condicionada $P\left(Y > 0.5 \mid X \in (0.25,\ 0.75)\right)$?
(NO lo calcule, pero indique correctamente los órdenes de integración)



Solución: Hay dos posibles respuestas

1. La probabilidad condicionada $P\left(Y > 0.5 \mid X \in (0.25,\ 0.75)\right)$ es

$$\frac{\int_{0.5}^{0.75}\int_{y}^{0.75} f_{XY}\left(x,y\right)\ dx\ dy}{\int_{0.25}^{0.75}\int_{0}^{x} f_{XY}\left(x,y\right)\ dy\ dx} = \frac{\int_{0.5}^{0.75}\int_{y}^{0.75} f_{XY}\left(x,y\right)\ dx\ dy}{\int_{0.25}^{0.75} f_{X}\left(x\right)\ dx}$$

donde la función de densidad marginal de X es $f_{X}\left(x\right) = \int_{0}^{x} f_{XY}\left(x,y\right)\ dy$.

2. La probabilidad condicionada: $P\left(Y > 0.5 \mid X \in (0.25,\ 0.75)\right)$ es

$$\frac{\int_{0.5}^{0.75}\int_{0.5}^{x} f_{XY}\left(x,y\right)\ dy\ dx}{\int_{0.25}^{0.75}\int_{0}^{x} f_{XY}\left(x,y\right)\ dy\ dx} = \frac{\int_{0.5}^{0.75}\int_{0.5}^{x} f_{XY}\left(x,y\right)\ dy\ dx}{\int_{0.25}^{0.75} f_{X}\left(x\right)\ dx}$$

donde la función de densidad marginal de X es $f_{X}\left(x\right) = \int_{0}^{x} f_{XY}\left(x,y\right)\ dy$.

Fin **ejercicio** 9

Ejercicio 10. Sea la siguiente función de densidad condicionada

$$f_{Y|X}\left(y \mid x\right) = \begin{cases} kx & \text{si } 0 < y < x,\ \text{para } x \in (0,1) \\ 0 & \text{en el resto de casos} \end{cases}.$$

¿Cómo calcularía el valor esperado de Y condicionado a que X tomase un valor arbitrario en el intervalo (0,1)?
(NO lo calcule, pero indique correctamente los órdenes de integración).

Solución:

$$\mathrm{E}_{Y|X}\left(Y \mid x\right) = \int_{0}^{x} y \cdot f_{Y|X}\left(y \mid x\right)\ dy = \int_{0}^{x} kxy\ dy$$

Fin **ejercicio** 10

Ejercicio 11. Los volúmenes de venta de dos bienes producidos por la misma empresa, X e Y, suponemos que tienen distribución normal bivariante con vector de medias $\begin{pmatrix}10 & 15\end{pmatrix}'$ y matriz de varianzas covarianzas $\left(\begin{smallmatrix}9 & 2\\ 2 & 9\end{smallmatrix}\right)$. ¿Cómo calcularía la probabilidad de que el volumen de ventas del primer producto supere el volumen de ventas del segundo?

Solución: Puesto que $P\left(X > Y\right) = P\left(X - Y > 0\right)$, basta con calcular la probabilidad de que W=X-Y sea mayor que cero. Ahora bien, puesto que W es combinación lineal de normales, $W \sim N\left(\mu\,,\,\sigma^2\right)$, donde $\mu = 10 - 15 = -5$; y $\sigma^2 = 9 + 9 - 2 \cdot 2 = 14$. Por tanto

$$P\left(W > 0\right) = P\left(\frac{W+5}{\sqrt{14}} > \frac{0+5}{\sqrt{14}}\right) = P\left(Z > 1.3363\right) =$$
$$= 1 - P\left(Z \leq 1.3363\right) = 1 - 0.9082 = 0.0918$$

Fin **ejercicio** 11

Ejercicio 12. Usando los mismos datos del problema anterior, calcule el valor esperado de la Y condicionado a que la X=12. ¿Difiere este valor del valor esperado de la Y? ¿Porqué? (En este problema **SÍ** debe calcular el valor esperado condicionado)

Solución: Puesto que X e Y tienen distribución conjunta normal, la esperanza condicionada en la siguiente función lineal

$$\mathrm{E}_{Y|X}\left(Y|\,x\right) = \mathrm{E}\left(Y\right) - \mathrm{E}\left(X\right)\frac{\mathrm{Cov}\left(Y,X\right)}{\mathrm{Var}\left(X\right)} + \frac{\mathrm{Cov}\left(X,Y\right)}{\mathrm{Var}\left(X\right)}\cdot x = 15 - 10\frac{2}{9} + \frac{2}{9}\cdot x$$

que evaluada para X=12 es $\mathrm{E}_{Y|X}\left(Y|\,12\right) = 15 - 10\frac{2}{9} + \frac{2}{9}\cdot 12 = 15.44$; que difiere de $\mathrm{E}\left(Y\right)$ como era de esperar, ya que covarianza distinta de cero implica *asociación* lineal entre ambas variables. Además, de que en este caso $x = 12 \neq \mathrm{E}\left(X\right)$. Fin **ejercicio** 12

Ejercicio 13. La variable aleatoria Z es combinación lineal de Y (es decir, $Z = a + bY$, donde a,b son constantes no nulas). Considere además, que hay otra variable X. Demuestre que si X e Y son independientes, entonces $\mathrm{E}_{Z|X}\left(Z\,|\,x\right) = \mathrm{E}\left(Z\right)$.

Solución:

$$\begin{aligned}
\mathrm{E}_{Z|X}\left(Z\,|\,x\right) =& \mathrm{E}_{Y|X}\left(a + bY|\,X\right) && \text{sustituyendo } Z \text{ por la combinación lineal}\\
=& a + b\mathrm{E}_{Y|X}\left(Y|\,x\right) && \text{por ser la esperanza un operador lineal}\\
=& a + b\mathrm{E}\left(Y\right) && \text{por ser } X \text{ e } Y \text{independientes}\\
=& \mathrm{E}\left(a + bY\right) && \text{por ser la esperanza un operador lineal}\\
=& \mathrm{E}\left(Z\right) && \text{sustituyendo la combinación lineal por } Z
\end{aligned}$$

Fin **ejercicio** 13

Ejercicio 14. El siguiente modelo de regresión lineal simple intenta captar la relación entre el PIB per capita de los países de la OCDE, Y, y el grado de escolarización como medida del nivel de educación de la economía, $\mathbf{X}$:

$$\ln(Y_i) = \alpha + \beta\ln(X_i) + U_i;$$

donde U_i son otros factores desconocidos o perturbaciones del modelo, que suponemos se distribuyen normal $(0, \sigma^2_U)$ condicionados a $\mathbf{X}$. Las estimaciones MCO de los parámetros del modelo son $\widehat{\alpha} = 0.89$ y $\widehat{\beta} = 0.28$, y sus desviaciones típicas son 0.025 y 0.08, respectivamente. Para un país j con un grado de escolarización $X_j = 10$ ¿Cuál sería el valor esperado de su PIB per capita? (Nótese que se pide el valor esperado de Y_j, y no de $\ln Y_j$)

Solución: Puesto que $\widehat{\ln Y_i} = 0.89 + 0.28\ln(X_i)$ entonces

$$Y_i = e^{0.89+0.28\ln(X_i)}$$

En el caso del país j-ésimo: $Y_j = e^{0.89+0.28\cdot\ln(10)} = 4.64$ Fin **ejercicio** 14

Ejercicio 15. Usando la misma información del apartado anterior, proponga la hipótesis nula y alternativa más apropiada para discutir si el nivel de educación de un país es significativo para explicar su PIB per capita. Indique también la región crítica más apropiada para el contraste.

Solución: Puesto que se pide un contraste de significación:

- H_0: $\beta = 0$

En cuanto a H_1, en principio no es verosímil que el efecto de la educación sobre la renta sea negativo (es decir, a mayor educación menor renta); por tanto para garantizar una mayor potencia del contraste, éste debe ser de una sóla cola (la derecha en este caso)

- H_1: $\beta > 0$

$$RC = \left\{ \text{el conjunto de muestras } \mathbf{x} \text{ tales que } \frac{\widehat{\beta} - 0}{\sqrt{\frac{\widehat{s^2}}{\sum(x_i - \overline{x})^2}}} > t_{T-2,\ 1-\alpha} \right\}$$

donde T es el tamaño de la muestra. Fin **ejercicio** 15

Formulas de posible utilidad

Transformación de variables Sea $f_X(x)$ en el soporte $\mathbb{R}_X = (a, b)$. Sea $Y = h(X)$, entonces $f_Y(y) = \left|\frac{dh^{-1}(y)}{dy}\right| \cdot f_X\left(h^{-1}(y)\right)$; en el soporte $\mathbb{R}_Y = (h(a),\ h(b))$; donde $h^{-1}(\cdot)$ es la *función inversa* de $h(\cdot)$, es decir $h^{-1}\left(h(x)\right) = x$.

Función generatriz de momentos conjunta del vector aleatorio $\boldsymbol{X} = (X_1, \ldots, X_n)$ evaluada en el vector $\boldsymbol{t} = (t_1, \ldots, t_n)$: $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \mathrm{E}\left(e^{t_1 X_1 + \cdots + t_n X_n}\right)$

Aproximación lineal a la esperanza condicional

$$\mathrm{E}(Y|X = x) \approx g^*(x) = \mathrm{E}\left(Y\right) - \mathrm{E}\left(X\right)\frac{\mathrm{Cov}\left(X, Y\right)}{\mathrm{Var}\left(X\right)} + \frac{\mathrm{Cov}\left(X, Y\right)}{\mathrm{Var}\left(X\right)} \cdot x; \qquad \forall x \in \mathbb{R}_X.$$

Varianza condicional de la normal bivariante: $\mathrm{Var}_{Y|X}(Y|\, x) = \sigma_Y^2 - \frac{(\sigma_{XY})^2}{\sigma_X^2} = \sigma_Y^2(1 - \rho_{XY}^2)$,

Modelo de regresión lineal Sea el modelo $Y_t = a + bX_t + U_t$, donde $\left(\mathbf{U}|\,\mathbf{X}\right) \sim N\left(\mathbf{0}\,,\, \sigma^2\,\mathbf{I}\right)$; y $t = 1,\ 2, \ldots, T$. Sean $\widehat{a}$ y $\widehat{b}$ los estimadores MCO de a y b. Entonces

$$\widehat{a} \sim N\left(a\,,\, \frac{\sigma^2 \sum x_t^2}{T\sum(x_t - \overline{x})^2}\right); \qquad \widehat{b} \sim N\left(b\,,\, \frac{\sigma^2}{\sum(x_t - \overline{x})^2}\right).$$

Distribuciones de funciones de variables aleatorias si $X \sim N\left(\mu\,,\,\sigma^2\right)$, $X_1 \sim N\left(\mu_1\,,\,\sigma_1^2\right)$ y $X_2 \sim N\left(\mu_2\,,\,\sigma_2^2\right)$ con muestras de tamaños n, n_1 y n_2 respectivamente:

$$\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0\,,\,1)\,; \qquad \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0\,,\,1)$$

$$\frac{\bar{x}-\mu}{\sqrt{\mathfrak{s}^2/n}} \sim t_{n-1}; \qquad \frac{(n-1)\mathfrak{s}^2}{\sigma^2} \sim \chi^2_{n-1}; \qquad \frac{\mathfrak{s}_1^2/\sigma_1^2}{\mathfrak{s}_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\frac{(n_1-1)\mathfrak{s}_1^2+(n_2-1)\mathfrak{s}_2^2}{n_1+n_2-2}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}, \text{ si } \sigma_1^2=\sigma_2^2$$

donde $\mathfrak{s}^2$ denota la *cuasivarianza* muestral ($\frac{\sum(x-\bar{x})^2}{n-1}$).

Proporciones $\frac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{a}{\sim} N(0\,,\,1)\,.$ Con muestras de tamaños n y m:

$$\frac{(\widehat{p_1}-\widehat{p_2})-(p_1-p_2)}{\sqrt{\frac{(n+m)\widehat{p_T}(1-\widehat{p_T})}{n\cdot m}}} \underset{H_0}{\overset{a}{\sim}} N(0\,,\,1)\,; \quad \text{donde } \widehat{p_T}=\frac{n\widehat{p_1}+m\widehat{p_2}}{n+m};\ H_0:\ p_1=p_2$$

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

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

Kolmogorov-Smirnov para una sola muestra: $D_n = \sup\left|F_n(x)-F(x)\right|$, donde $F_n(x)$ es la función de distribución empírica (o muestral), y $F(x)$ es la función de distribución de H_0.

	nivel de significación				
Tamaño muestral	0.2	0.1	0.05	0.02	0.01
8	0.3583	0.4096	0.4543	0.5065	0.5418
9	0.3391	0.3875	0.4300	0.4796	0.5133
10	0.3226	0.3687	0.4092	0.4566	0.4889
> 40	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$

Kolmogorov-Smirnov para dos muestras: $D_n = \sup\left|F_1(x)-F_2(x)\right|$. Cuando n_1 y n_2 son grandes; y donde $F_1(x)$ y $F_2(x)$ son funciones de distribución empíricas (o muestrales)

$$\text{Nivel crítico: } D_{\alpha,n_1,n_2} \approx k\sqrt{\frac{n_1+n_2}{n_1n_2}},$$

donde $k = 1.07;\ 1.22;\ 1.52$; para α igual a $10\,\%$, $5\,\%$ y $1\,\%$ respectivamente.

z	x.x0	x.x1	x.x2	x.x3	x.x4	x.x5	x.x6	x.x7	x.x8	x.x9
	Probabilidad acumulada desde $-\infty$ hasta z para $Z \sim N(0\,,\,1)$									
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	0.9279	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	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

ν	60.0 %	66.7 %	75.0 %	80.0 %	87.5 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
	Puntos porcentuales de la **fun. distribución** t de STUDENT con ν grados de libertad										
1	0.325	0.577	1.000	1.376	2.414	3.078	6.314	12.706	31.821	63.657	318.31
2	0.289	0.500	0.816	1.061	1.604	1.886	2.920	4.303	6.965	9.925	22.327
3	0.277	0.476	0.765	0.978	1.423	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.464	0.741	0.941	1.344	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.457	0.727	0.920	1.301	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.453	0.718	0.906	1.273	1.440	1.943	2.447	3.143	3.707	5.208
7	0.263	0.449	0.711	0.896	1.254	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.447	0.706	0.889	1.240	1.397	1.860	2.306	2.896	3.355	4.501
9	0.261	0.445	0.703	0.883	1.230	1.383	1.833	2.262	2.821	3.250	4.297
10	0.260	0.444	0.700	0.879	1.221	1.372	1.812	2.228	2.764	3.169	4.144
11	0.260	0.443	0.697	0.876	1.214	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.442	0.695	0.873	1.209	1.356	1.782	2.179	2.681	3.055	3.930
13	0.259	0.441	0.694	0.870	1.204	1.350	1.771	2.160	2.650	3.012	3.852
14	0.258	0.440	0.692	0.868	1.200	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.439	0.691	0.866	1.197	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.439	0.690	0.865	1.194	1.337	1.746	2.120	2.583	2.921	3.686
17	0.257	0.438	0.689	0.863	1.191	1.333	1.740	2.110	2.567	2.898	3.646
18	0.257	0.438	0.688	0.862	1.189	1.330	1.734	2.101	2.552	2.878	3.610
19	0.257	0.438	0.688	0.861	1.187	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.437	0.687	0.860	1.185	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.437	0.686	0.859	1.183	1.323	1.721	2.080	2.518	2.831	3.527
22	0.256	0.437	0.686	0.858	1.182	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.436	0.685	0.858	1.180	1.319	1.714	2.069	2.500	2.807	3.485
24	0.256	0.436	0.685	0.857	1.179	1.318	1.711	2.064	2.492	2.797	3.467
∞	0.253	0.431	0.674	0.842	1.150	1.282	1.645	1.960	2.326	2.576	3.090

	Puntos porcentuales de la **fun. distribución** CHI-CUADRADO con ν grados de libertad										
ν	60.0 %	66.7 %	75.0 %	80.0 %	87.5 %	90.0 %	95.0 %	97.5 %	99.0 %	99.5 %	99.9 %
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179