Dpto. de Economía Cuantitativa
Universidad Complutense de Madrid

Examen de Introducción a la Econometría

16 de septiembre de 2004, 15:30h.
Duración: 2 HORAS y 30 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. (Consta de 5 apartados)
Sean los siguientes datos:

Empresa	y_i	x_i	$x_i y_i$	x_i^2
A	1	1	1	1
B	3	2	6	4
C	4	4	16	16
D	6	4	24	16
E	8	5	40	25
F	9	7	63	49
G	11	8	88	64
H	14	9	126	81
sumas	56	40	364	256

Cuadro 1:

donde y son beneficios, y x son gastos en formación de personal de una empresa.

Además se sabe que las varianzas y covarianzas muestrales son tales que:

$$T \cdot s_{\boldsymbol{y}}^2 = \sum (y_i - \overline{\boldsymbol{y}})^2 = 132,$$

$$T \cdot s_{\boldsymbol{x}}^2 = \sum (x_i - \overline{\boldsymbol{x}})^2 = 56,$$

$$T \cdot s_{\boldsymbol{x}\,\boldsymbol{y}} = \sum (x_i - \overline{\boldsymbol{x}})(y_i - \overline{\boldsymbol{y}}) = 84,$$

donde T es el tamaño muestral.

Suponga que se plantea el siguiente modelo

$$Y_i = a + bx_i + U_i,$$

donde U_i son otros factores que afectan a los beneficios distintos de sus gastos en formación (el término de error). Se sabe que la distribución conjunta de dichos factores es:

$$\boldsymbol{U} \sim N(\boldsymbol{0},\ \sigma^2\,\mathbf{I}),$$

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

donde $\mathbf{I}$ es una matriz identidad de orden 8, y σ^2 es la varianza de U_i, cuyo valor es desconocido.

(a) Estime por MCO los parámetros a y b del modelo.

Solución:

1. Por una parte:
$$\widehat{b} = \frac{s_{\boldsymbol{x}\boldsymbol{y}}}{s^2_{\boldsymbol{x}}} = \frac{84}{56} = 1.5$$
por otra, las medias muestrales son
$$\overline{\boldsymbol{x}} = \frac{\sum x_i}{8} = \frac{40}{8} = 5; \qquad \overline{\boldsymbol{y}} = \frac{\sum y_i}{8} = \frac{56}{8} = 7;$$
por lo que
$$\widehat{a} = \overline{\boldsymbol{y}} - \widehat{b} \cdot \overline{\boldsymbol{x}} = 7 - 1.5 \cdot 5 = -0.5.$$

Fin **ejercicio** 1

(b) ¿Cuál es el beneficio esperado para una empresa que incurriera en unos gastos de formación de personal de 3?

Solución: Según el modelo estimado, una empresa que incurra en unos gastos de 3 debería tener unos beneficios de
$$\widehat{y} = \widehat{a} + \widehat{b}x = -0.5 + 1.5 \cdot 3 = 4$$

Fin **ejercicio** 1

(c) Calcule los residuos de la empresa E y F. ¿Que indica en este caso el signo de los residuos? La comparación de los residuos para estas empresas ¿contradice el hecho de que F tiene mayores beneficios que E? Justifique su respuesta.

Solución: Los residuos de la empresa E serán:
$$y_E - \widehat{y_E} = y_E - \big(\widehat{a} + \widehat{b}x_E\big) = 8 - \big(-0.5 + 1.5 \cdot 5\big) = 8 - 7 = 1$$
y los de la empresa F:
$$y_F - \widehat{y_F} = y_F - \big(\widehat{a} + \widehat{b}x_F\big) = 9 - \big(-0.5 + 1.5 \cdot 7\big) = 9 - 10 = -1.$$

Puesto que
$$\widehat{y} = \mathrm{E}_{U|\mathbf{x}_{F\triangleright}}\left(\boldsymbol{Y} \mid \boldsymbol{x}_{F\triangleright}\right),$$
un signo positivo para el residuo de cierta empresa significa que ésta ha logrado unos beneficios mayores que los esperados (dado su nivel de gasto en formación de personal, x). Por el contrario, un residuo negativo significa que la empresa ha obtenido unos beneficios menores de los esperados por el modelo (dado su gasto en formación).

La comparación entre empresas con distinta inversión en formación no es apropiada para valorar los datos sobre beneficios (sólo lo es entre empresas con mismo nivel de gasto en formación). La empresa F tiene mayores beneficios que los de E, pero, dado su nivel de gasto en formación (7), estos beneficios deberían haber sido aún mayores (el valor esperado es 10).

Fin **ejercicio** 1

(d) Estime por MCO un intervalo de confianza del 95 % para el parámetro b del modelo, sabiendo que la suma de los residuos al cuadrado es 6.

Solución: El estimador MCO se distribuye Normal con esperanza igual al verdadero valor de los parámetros estimados, y varianza desconocida.

- Buscamos los valores A y B tales que
$$P\left(A \leq \frac{\widehat{b}|\mathbf{x} - b}{\sqrt{\frac{\widehat{s^2_{\widehat{e}}}}{\sum(x_i - \overline{\boldsymbol{x}})^2}}} \leq B \right) = (1-\alpha)$$

Donde $\frac{\widehat{b}|\mathbf{x}-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}}$ se distribuye como una t de Student con $T-2$ grados de libertad; por tanto A y B son los valores que aparecen en las tablas, y que determinan un intervalo centrado en cero con una probabilidad asociada del 95 %; es decir, $A = -2.447$, y $B = 2.447$, y $\widehat{\mathfrak{s}^2_{\widehat{e}}} = 6/(T-2) = 1$. Así pues, la estimación del intervalo de confianza de parámetro desconocido b es

$$\mathrm{IC}^b_{0.95}(\boldsymbol{w}) = \left[1.5 \pm 2.447 \cdot \sqrt{1/56}\right]$$

Fin **ejercicio** 1

(e) Contraste la hipótesis de que "la pendiente del modelo es uno" frente a que "es menor que uno" con un nivel de significación del 10 %. ¿Cuál es el p-valor de la estimación de "dicha pendiente"?

Solución: Las hipótesis son:

$$H_0: \ b = 1$$
$$H_1: \ b < 1$$

La región critica de una sola cola es

$$RC = \left\{ \mathbf{x} \;\middle|\; \frac{\widehat{b}|\mathbf{x} - 1}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} < k \right\},$$

donde k es el valor de la tablas para una t de Student de seis grados de libertad, ya que el estadístico de la parte izquierda de la desigualdad tiene dicha distribución. Para $\alpha = 0.1$, tenemos que $k = t_{6,\ 0.1} = -1.44$. Sustituyendo tenemos que

$$\frac{1.5-1}{\sqrt{1/56}} = 3.74 > k = t_{6,0.1} = -1.44$$

por lo que no rechazamos H_0.

El p-valor es la probabilidad de

$$P\left(\widehat{b}|\mathbf{x} \le 1.5 \;\middle|\; H_0\right) = P\left(\frac{\widehat{b}|\mathbf{x} - b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} \le \frac{1.5-b}{\sqrt{\frac{\widehat{\mathfrak{s}^2_{\widehat{e}}}}{\sum(x_i-\overline{x})^2}}} \;\middle|\; H_0\right)$$
$$= P\left(W \le \frac{1.5-1}{\sqrt{1/56}} = 3.74\right) \simeq 0.999$$

donde W se distribuye como una t de Student con seis grados de libertad. Fin **ejercicio** 1

Ejercicio 2. (Consta de 5 apartados)

Sean dos variables aleatorias, (X,Y), discretas que reflejan la siguiente información: $X = 1$ si se da un shock positivo de oferta en la economía (por ejemplo, el precio del petroleo baja por debajo de los 35-30\$/barril), $X = 0$ si la economía no sufre ningún shock (por ejemplo, el precio del petroleo se mantiene en torno a los 35\$/barril) y $X = -1$ si el shock es negativo (por ejemplo, el precio del petroleo se coloca por encima de los 35-40\$/barril); $Y = 1$ si el nivel de empleo aumenta, $Y = 0$ si se mantiene e $Y = -1$ si disminuye. La siguiente tabla refleja su ley de probabilidades conjuntas:

$X \backslash Y$	-1	0	1
-1	5/24	3/24	0
0	2/24	6/24	2/24
1	1/24	2/24	3/24

(a) Sin tener información acerca del shock económico, ¿qué es más probable, que aumente o que disminuya el nivel de empleo?

Solución: Puesto que nos preguntan por probabilidades acerca de la evolución del empleo al margen de los shocks de oferta, necesitamos conocer las función de cuantía marginal de Y.

Calculamos las funciones de cuantía marginales:

$X \backslash Y$	-1	0	1	$P_X(x)$
-1	$5/24$	$3/24$	0	$8/24$
0	$2/24$	$6/24$	$2/24$	$10/24$
1	$1/24$	$2/24$	$3/24$	$6/24$
$P_Y(y)$	$8/24$	$11/24$	$5/24$	1

Por tanto, $P(Y = 1) = 5/24$ y $P(Y = -1) = 8/24$; así pues, "a priori" es más probable una disminución del nivel de empleo. Fin **ejercicio** 2

(b) Suponga, que apreciamos que el shock ha sido negativo, ¿Cambiaría la respuesta del apartado anterior? ¿Cuál sería ahora la probabilidad de que el empleo caiga, que se mantenga o que suba?

Solución: Ahora necesitamos la función de cuantía de Y condicionada al hecho de que ha habido un shock negativo. Por tanto nos interesa fijarnos en la primera fila de la función de cuantía conjunta. Dicha fila no es una función de cuantía por si sola, ya que no suma uno. Para obtener la función de cuantía de Y condicionada, debemos dividir dicha fila por la probabilidad $P(X = 1)$; por tanto

$$P_{Y|X}(y \mid -1) = \begin{pmatrix} \frac{5/24}{8/24} & \frac{3/24}{8/24} & 0 \end{pmatrix} = \begin{pmatrix} \frac{5}{8} & \frac{3}{8} & 0 \end{pmatrix}$$

que suma uno y por tanto es función de cuantía.

Es decir, $P_Y(-1) = \frac{5}{8}$, $P_Y(0) = \frac{3}{8}$, $P_Y(1) = 0$ que son probabilidades distintas de las descritas por la función de cuantía marginal de Y. Por tanto:

1. Sigue siendo más probable una caída del nivel de empleo (de hecho, un incremento tiene probabilidad nula).
2. Las probabilidades son $\frac{5}{8}$, $\frac{3}{8}$, y 0 respectivamente.

Fin **ejercicio** 2

(c) Sin tener información acerca del shock de la economía, ¿Cuál es el valor esperado de la variable que describe la evolución del empleo?

Solución: Es la esperanza incondicional de Y, por tanto,

$$\mathrm{E}(Y) = -1 \cdot \frac{8}{24} + 0 \cdot \frac{11}{24} + 1 \cdot \frac{5}{24} = -\frac{3}{24}$$

Fin **ejercicio** 2

(d) ¿Cuál sería el valor esperado en el movimiento del empleo si el shock ha sido negativo? ¿y su varianza?

Solución: Debemos calcular la esperanza de Y condicionada a $X = -1$, es decir,

$$\mathrm{E}_{Y|X}(Y \mid -1) = -1 \cdot \frac{5}{8} + 0 \cdot \frac{3}{8} + 1 \cdot 0 = -\frac{5}{8}$$

Para calcular la varianza condicionada necesitamos calcular primero el momento condicionado de segundo orden de Y, ya que como sabemos

$$\mathrm{Var}_{Y|X}(Y \mid -1) = \mathrm{E}_{Y^2|X}\left(Y^2 \mid -1\right) - \left(\mathrm{E}_{Y|X}(Y \mid -1)\right)^2.$$

Puesto que

$$\mathrm{E}_{Y^2|X}\left(Y^2 \mid -1\right) = (-1)^2 \cdot \frac{5}{8} + 0^2 \cdot \frac{3}{8} + 1^2 \cdot 0 = \frac{5}{8},$$

entonces

$$\operatorname{Var}_{Y|X}(Y|-1) = \operatorname{E}_{Y^2|X}\left(Y^2 \mid -1\right) - \left(\operatorname{E}_{Y|X}(Y|-1)\right)^2 = \frac{5}{8} - \left(-\frac{5}{8}\right)^2 = 0.234$$

Fin **ejercicio** 2

(e) Comparando las respuestas de los apartados anteriores, ¿podría decir algo acerca de la existencia o no de dependencia de las variables? Dé una medida del grado de relación lineal entre las variables. ¿Que signo tiene?

Solución: Puesto que la probabilidad condicional difiere de la incondicional, esto indica que existe "algún tipo" de dependencia entre las variables.

Para ver el grado de dependencia lineal debemos calcular el coeficiente de correlación lineal.

Por una parte, puesto que $\operatorname{Cov}(X,Y) = \operatorname{E}(XY) - \operatorname{E}(X)\operatorname{E}(Y)$ necesitamos conocer $\operatorname{E}(X)$ y $\operatorname{E}(XY)$:

$$\operatorname{E}(X) = -1 \cdot \frac{8}{24} + 0 \cdot \frac{10}{24} + 1 \cdot \frac{6}{24} = -\frac{2}{24}$$

$$\operatorname{E}(XY) = (-1)(-1) \cdot \frac{5}{24} + (1)(-1) \cdot \frac{1}{24} + (1)(1) \cdot \frac{3}{24} = \frac{7}{24}$$

(donde hemos omitido los sumandos nulos).

Por tanto

$$\operatorname{Cov}(X,Y) = \operatorname{E}(XY) - \operatorname{E}(X)\operatorname{E}(Y) = \frac{7}{24} - \frac{-2}{24} \cdot \frac{-3}{24}$$

El coeficiente de correlación se define como

$$\operatorname{Corr}(X,Y) = \rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Así pues, nos queda calcular $\operatorname{Var}(X)$ y $\operatorname{Var}(Y)$:

Puesto que

$$\begin{aligned}
\operatorname{E}\left(X^2\right) &= (-1)^2 \cdot \frac{8}{24} + 0 \cdot \frac{10}{24} + 1^2 \cdot \frac{6}{24} = \frac{14}{24} \\
\operatorname{E}\left(Y^2\right) &= (-1)^2 \cdot \frac{8}{24} + 0 \cdot \frac{11}{24} + 1^2 \cdot \frac{5}{24} = \frac{13}{24}
\end{aligned}$$

tenemos que $\operatorname{Var}(X) = \frac{14}{24} - \left(-\frac{2}{24}\right)^2$ y $\operatorname{Var}(Y) = \frac{13}{24} - \left(-\frac{3}{24}\right)^2$; por lo que finalmente

$$\operatorname{Corr}(X,Y) = \rho_{XY} = \frac{\frac{7}{24} - \frac{-2}{24} \cdot \frac{-3}{24}}{\sqrt{\left[\frac{14}{24} - \left(-\frac{2}{24}\right)^2\right]\left[\frac{13}{24} - \left(-\frac{3}{24}\right)^2\right]}} = 0.37$$

que indica cierta dependencia lineal *positiva*, pero no muy acusada, puesto que esta lejos de uno en valor absoluto.

Fin **ejercicio** 2

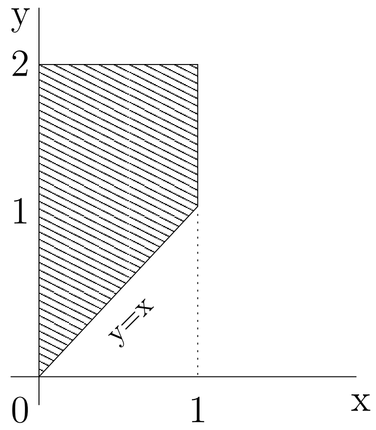
Ejercicio 3. (Consta de 7 apartados)

Sea (X,Y) una variable aleatoria bidimensional continua con la siguiente función de densidad:

$$f_{XY}(x,y) = \begin{cases} 2/3 & \text{si } x \in [0,1]; \qquad y \in [0,2]; \qquad y > x \\ 0 & \text{en los restantes casos} \end{cases}$$

(a) Dibuje el recinto en el que $f_{XY}(x, y)$ toma valores distintos de cero. Especifique claramente los rangos de definición (incondicionales y condicionales) de las variables aleatorias X e Y. (¿valores tomados por y restringen en cualquier caso los posibles valores que puede tomar x?)

Solución:



Nótese que x puede tomar valores entre 0 y 1, siempre y cuando además $x < y$. Cuando $y < 1$, la variable x estará en el intervalo $x \in (0,\ y)$; por lo que y restringe los posibles valores de x. Sin embargo, cuando $y \geq 1$, la variable x puede tomar cualquier valor entre cero y uno (y NO restringe, en este segundo caso, los posibles valores de x.) Fin **ejercicio** 3

(b) Verifique que $f_{XY}(x, y) = 2/3$ es efectivamente una función de densidad definida en el recinto especificado.

Solución: $\int_0^1 \int_x^2 2/3\ dydx = 1$ y además $f_{XY}(x, y)$ es estrictamente positiva en todo su dominio de definición. Fin **ejercicio** 3

(c) Calcule las funciones de densidad marginal de X e Y. Indique explícitamente sus dominios de definición.

Solución:

$$f_X(x) = \begin{cases} \int_x^2 2/3\ dy = \frac{4}{3} - \frac{2}{3}x, & \text{en el intervalo } x \in (0, 1) \\ 0 & \text{en el resto de casos.} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_0^y 2/3\ dx = \frac{2}{3}y, & \text{en el intervalo } y \in (0, 1) \\ \int_0^1 2/3\ dx = \frac{2}{3}, & \text{en el intervalo } y \in [1, 2) \\ 0 & \text{en el resto de casos.} \end{cases}$$

Fin **ejercicio** 3

(d) Calcule el valor esperado de X e Y.

Solución:

$$\mathrm{E}(Y) = \int_0^1 y \frac{2}{3} y\ dy + \int_1^2 y \frac{2}{3}\ dy = \frac{11}{9}.$$

$$\mathrm{E}(X) = \int_0^1 x \left(\frac{4}{3} - \frac{2}{3}x \right)\ dx = \frac{4}{9}.$$

Fin **ejercicio** 3

(e) Verifique que la probabilidad de que Y sea mayor que 1 es dos tercios, y la probabilidad de que sea menor o igual que 1 un tercio.

Solución:

$$\int_1^2 \frac{2}{3}\ dy = \frac{2}{3};$$

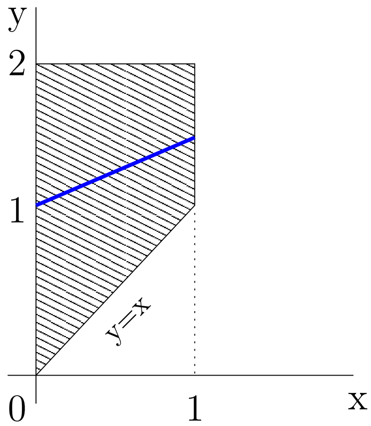
$$\int_0^1 \frac{2}{3} y dy = \frac{1}{3}; \quad \text{como cabe esperar por ser el suceso complementario al anterior.}$$

Fin **ejercicio** 3

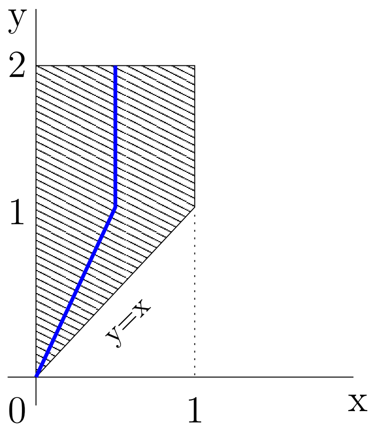
(f) Calcule la esperanza de Y condicionada a X y la esperanza de X condicionada a Y. Deje bien claro los dominios de definición de las dos esperanzas condicionadas. Dibuje ambas esperanzas condicionadas.

Solución:

$$\mathrm{E}_{Y|X}\left(Y|\, x\right) = \begin{cases} \int\limits_x^2 y \frac{2/3}{\frac{4}{3}-\frac{2}{3}x} dy = \frac{x^2}{2x-4} - \frac{2}{x-2} = \frac{1}{2}\left(x+2\right), & \text{definida para } x \in (0,1), \\ 0 & \text{en otro caso.} \end{cases}$$



$$\mathrm{E}_{X|Y}(X\,|\,y) = \begin{cases} \int_0^y x \frac{2/3}{2/3y} dx = \frac{1}{2}y & \text{definida para } y \in (0,1), \\ \int_0^1 x \frac{2/3}{2/3} dx = \frac{1}{2} & \text{definida para } y \in [1,2), \\ 0 & \text{en otro caso.} \end{cases}$$



Fin **ejercicio** 3

(g) ¿Son lineales las esperanzas condicionadas calculadas en el apartado anterior?

Solución: $\mathrm{E}_{Y|X}\left(Y|\, x\right)$ es lineal; sin embargo $\mathrm{E}_{X|Y}(X\,|\,y)$ no lo es. Fin **ejercicio** 3

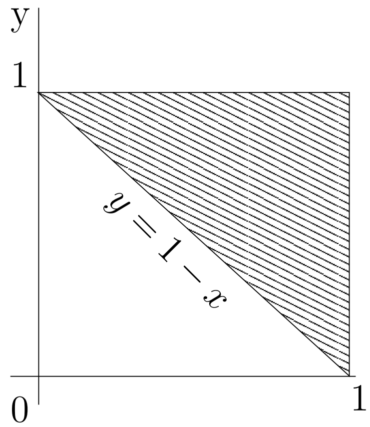
Preguntas cortas (2/3 de punto por pregunta)

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

Ejercicio 4. Sean X e Y con función de densidad $f_{XY}(x,y)$ en el triángulo $0 \leq 1-x \leq y \leq 1$. ¿Cuál (o cuales) de las siguientes expresiones es (o son) correcta (correctas)?

1. $\mathrm{E}(X) = \int_0^1 x \int_{1-x}^1 f_{XY}(x,y)\,dydx$
2. $\mathrm{E}(X) = \int_0^1 x \int_0^1 f_{XY}(x,y)\,dydx$
3. $\mathrm{E}(X) = \int_0^1 x \int_{1-x}^y f_{XY}(x,y)\,dydx$
4. $1 = \int_0^1 \int_0^1 f_{XY}(x,y)\,dydx$

Solución:



Sólo la primera. Fin **ejercicio** 4

Ejercicio 5. De 1000 encuestados, 680 son favorables a que bajen los impuestos sobre la gasolina y 320 no lo son. Un político manifiesta que el 75 % del total de la población es favorable a la propuesta y un 25 % no lo es. Contraste la veracidad de la afirmación del político según la encuesta realizada [Pista: lleve a cabo un contraste chi-cuadrado de ajuste a una distribución teórica].

Solución: El valor del estadístico del contraste es: $\frac{(750-680)^2}{750} + \frac{(250-320)^2}{250} = 26.13$, que excede con mucho el valor crítico de las tablas correspondientes a una chi-cuadrado de un grado de libertad, por lo que rechazamos la hipótesis nula. Fin **ejercicio** 5

Ejercicio 6. Queremos contrastar la independencia entre el nivel de ingresos de una familia y su ideología política. Para ello encuestamos a un total de 400 familias y el siguiente cuadro resume la información obtenida:

	derecha	izquierda
$<$ 18 mil euros	80	110
$\geq$18 mil euros	125	85

A partir de esta información muestral, ¿cuáles son las frecuencias teóricas absolutas para llevar a cabo un contraste Chi-cuadrado?

Solución:

	derecha	izquierda
$<$ 18 mil euros	$\frac{190}{400}\frac{205}{400}400 = 97.4$	$\frac{190}{400}\frac{195}{400}400 = 92.6$
$\geq$18 mil euros	$\frac{210}{400}\frac{205}{400}400 = 107.6$	$\frac{210}{400}\frac{195}{400}400 = 102.4$

Fin **ejercicio** 6

Ejercicio 7. Dado un nivel de significación α, proponga la región crítica que crea más conveniente para resolver el siguiente contraste: $H_0 : \sigma_X^2 = \sigma_Y^2$, $H_1 : \sigma_X^2 \neq \sigma_Y^2$. (suponga mismo tamaño muestral, n, para la X e Y) ¿Cambiaría la región crítica si $H_1 : \sigma_X^2 > \sigma_Y^2$? Razone la respuesta.

Solución: $RC_1 = \left\{\frac{s_X^2}{s_Y^2} < k_1 \text{ o } \frac{s_X^2}{s_Y^2} > k_2\right\}$, con $k_1 = F_{n,n,\alpha/2}$ y $k_2 = F_{n,n,1-\alpha/2}$. En el segundo caso, $RC_2 = \left\{\frac{s_X^2}{s_Y^2} > F_{n,n,1-\alpha}\right\}$, ya que si usáramos la región de dos colas el contraste perdería potencia. Fin **ejercicio** 7

Ejercicio 8. En el siguiente modelo de regresión simple, $Y = \alpha + \beta X + \varepsilon$, Y mide la tasa de crecimiento del PIB real y X la tasa de crecimiento de la masa monetaria. Tenemos una muestra de tamaño 35. Queremos contrastar la hipótesis de neutralidad monetaria en esta muestra (es decir, la masa monetaria no tiene efecto real sobre el nivel del PIB). Proponga la hipótesis nula y alternativa que, según la t$^{\text{a}}$ económica, crea más conveniente para realizar el contraste, así como la región crítica. ¿Cual es la distribución del estadístico de contraste?

Solución: $H_0: \ \beta = 0$, frente a $H_1: \ \beta > 0$. Nótese que, por t$^{\text{a}}$ económica, lo más razonable es proponer un contraste de cola superior. La región crítica más conveniente es por tanto: $RC = \left\{\hat{\beta} > k\right\} = \left\{\frac{\hat{\beta}-\beta}{dt(\hat{\beta})} > t_{1-\alpha}\right\}$, donde $\hat{\beta}$ es el estimador MCO y $t_{1-\alpha}$ es el valor tabulado de una t-Student con 33 grados de libertad que deja a la izquierda un área de $1-\alpha$. Fin **ejercicio** 8

Ejercicio 9. Sea una población X que se distribuye $N(\mu,\ 1)$. Se desea contrastar H_0: $\mu = 5$ frente a H_1: $\mu > 5$. Se tiene una muestra aleatoria simple de tamaño 16, con $\overline{\mathbf{x}} = 5.5$, calcule el p-valor del contraste [para su resolución, proponga la región crítica más apropiada].

Solución: La región crítica propuesta es: $RC = \{\overline{\mathbf{x}} > k\}$, por lo que el p-valor es igual a $P\left(\overline{\mathbf{x}} > 5.5 \mid H_0\right) = P\left(\frac{\overline{\mathbf{x}}-5}{\sqrt{1/16}} > \frac{5.5-5}{\sqrt{1/16}}\right) = P\left(Z > 2\right) = 1 - P\left(Z < 2\right) = 1 - 0.9772 = 0.023$, donde $Z \sim N\left(0\,,\,1\right)$.

Fin **ejercicio** 9

Ejercicio 10. Discuta la veracidad o falsedad de la siguiente afirmación:

Si la hipótesis nula no es rechazada, entonces es cierta.

Solución: **Falso:** El test de hipótesis sólo puede indicar que con los datos observados no rechazamos la hipótesis nula H_0 con una probabilidad β, es decir, asumiendo el riesgo de que, con probabilidad β, no rechacemos H_0 aún siendo falsa (error tipo II). Fin **ejercicio** 10

Ejercicio 11. Sea X una variable aleatoria con distribución simétrica alrededor de cero; es decir, $P_X\left(x\right) = P_X\left(-x\right)$ en caso discreto, ó $f_X\left(x\right) = f_X\left(-x\right)$ en caso continuo. Sea $Y = X^2$. Demuestre que Cov $\left(X, Y\right)$ =0. ¿Son X y Y independientes? [Pista: simetría implica $\mathrm{E}\left(X^3\right) = 0$]

Solución: Por ser X simétrica respecto a cero necesariamente $\mathrm{E}\left(X\right) = 0$, pero también su tercer momento respecto al origen debe ser cero (simetría implica $\mathrm{E}\left(X^3\right) = 0$). Así pues,

$$\mathrm{Cov}\left(X, Y\right) = \mathrm{E}\left(X^3\right) - \mathrm{E}\left(X\right)\mathrm{E}\left(X^2\right) = 0 - 0 \cdot \mathrm{E}\left(X^2\right) = 0$$

Obviamente son dependientes puesto que $Y = X^2$. Fin **ejercicio** 11

Ejercicio 12. Sean X e Y variables aleatorias con distribución conjunta normal con vector de medias $\boldsymbol{\mu} = (1, 0)$ y matriz de varianzas-covarianzas $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 2 \end{smallmatrix}\right)$. ¿Cuál es la probabilidad de que $Y > 1$? ¿Cómo debería ser la matriz $\boldsymbol{\Sigma}$ para garantizar que $P\left(Y > 1\right) = P\left(Y > 1 \mid X = 1\right)$?

Solución: $P\left(Y > 1\right) = P\left(\frac{Y-0}{\sqrt{2}} > \frac{1-0}{\sqrt{2}}\right) = P\left(\frac{Y-0}{\sqrt{2}} > \frac{1-0}{\sqrt{2}}\right) = 1 - 0.761 = 0.239$, donde hemos considerado que Y es normal (0,2).

Cuando la matriz es diagonal (es decir, cuando las covarianzas son cero) necesariamente ambas probabilidades son iguales (puesto que bajo normalidad conjunta covarianza nula implica independencia). Fin **ejercicio** 12

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 $h^{-1}(h(x)) = x$.

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

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

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 $(\mathbf{U}|\,\mathbf{X}) \sim N\left(\mathbf{0}\,,\,\sigma^2\,\mathbf{I}\right)$; 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 $\left(\frac{\sum(x-\bar{x})^2}{n-1}\right)$.

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: $D_n = \sup |F_n(x) - F(x)|$, 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: $D_n = \sup |F_1(x) - F_2(x)|$. Cuando n_1 y n_2 son grandes; y donde $F_1(x)$ y $F_2(x)$ son funciones de distribución empíricas (o muestrales)

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

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

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

Puntos porcentuales de la **fun. distribución** t de STUDENT 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	0.325	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