Examen de Introducción a la Econometría (LECO).

Departamento de Economía Cuantitativa. Universidad Complutense de Madrid.
8 de septiembre 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 cuardernillo. Junto a cada pregunta y apartado se indica la puntuación.

Problemas

Problema 1. (2.5 pt.) Sean X e Y las rentabilidades diarias de las acciones de dos empresas que *se cree* que siguen una distribución normal con esperanzas y varianzas desconocidas. La primera empresa (con rentabilidad X) es de pequeño tamaño y se dedica a negocios de alto riesgo. La segunda, cuya rentabilidad es Y, es de mayor tamaño y sus actividades comportan un bajo riesgo.

Se dispone de datos de 20 días de cotización para la primera empresa y 25 para la segunda, con los que se han calculado los siguientes estadísticos:

$\bar{x} = 0{,}1$	$\sum x_i^2 = 1{,}76$
$\bar{y} = 0{,}01$	$\sum y_i^2 = 1{,}28$

Se quiere contrastar si las acciones de ambas empresas tienen igual rentabilidad esperada con un nivel de confianza del 99 %.

1. (0.5 pt) Si cree que debe realizar algún contraste previo, explique porqué y plantee la hipótesis nula, la alternativa y la región crítica de dicho contraste preliminar.

   Aunque en el enunciado se dice que *se cree* que las variables aleatorias siguen una distribución normal, es conveniente comprobarlo antes de utilizar un contraste que exige el cumplimiento de tal supuesto.

   Para constrastar normalidad se pueden utilizar varios estadísticos vistos en el curso: chi-cuadrado de ajuste a una distribución teórica, Kolmogorov-Smirnov de una muestra y, para el caso específico de una normal, el contraste Jarque-Bera. Este **último resulta el más conveniente** al no requerir ningún supuesto o estimación de los parámetros de la normal teórica.

   El estadístico de contraste, denotado por JB, aparece en el formulario al final del examen. La hipótesis nula es que la muestra procede de una variable aleatoria normal frente a la alternativa de que no se ha extraído la muestra de dicha distribución. La región crítica es $JB \geq c^*$, donde el valor crítico c^* deja a su derecha el nivel de significación preseleccionado en una distribución χ^2_2.

2. (0.5 pt) Plantee la hipótesis nula, la alternativa y la región crítica del contraste óptimo de igualdad de rentabilidades esperadas.

   Bajo el supuesto de normalidad e independencia de las variables aleatorias, el contraste óptimo (más potente) es la t para igualdad de esperanzas con varianzas desconocidas. La hipótesis nula es $H_0 : \mu_X = \mu_Y$ frente a la alternativa $H_1 : \mu_X \neq \mu_Y$. En este problema sería razonable también una alternativa de una cola, puesto que cabe esperar que la rentabilidad de las acciones de la empresa con más riesgo sea superior a la rentabilidad para la empresa más conservadora ($H_1' = \mu_X > \mu_Y$).

   Si se denota por t el valor del estadístico, la región crítica es $\{t \leq -c^* \cup t \geq c^*\}$, donde los valores críticos cumplen $P[-c^* \leq t_{43} \leq c^*] = 1 - \alpha$, siendo α el nivel de signficación elegido. Si se utiliza la alternativa de una cola, la región crítica es $\{t \geq c^*\}$, y el valor crítico es tal que $P[t_{43} \leq c^*] = 1 - \alpha$.

3. (0.75 pt) Resuelva el contraste con los datos disponibles con un nivel de confianza del 99 %. ¿Qué consecuencias hubiera tenido realizar el contraste al 90 %?

   Para aplicar el contraste que aparece en el formulario se necesitan las cuasi-varianzas. Para obtenerlas hay que tener en cuenta que $\sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$, y resulta que $19 \cdot s_X^2 = 1{,}76 - 20 \cdot 0{,}1^2 = 1{,}56$ y $24 \cdot s_Y^2 = 1{,}28 - 25 \cdot 0{,}01^2 = 1{,}2775$.

   El estadístico de contraste, una vez particularizado en la hipótesis nula, es:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}} = \frac{0{,}1 - 0{,}01}{\sqrt{\left(\frac{1}{20} + \frac{1}{25}\right) \frac{20(1{,}56/19) + 25(1{,}2775/24)}{20+25-2}}} = \frac{0{,}09}{\sqrt{0{,}09 \frac{2{,}9728}{43}}} = 1{,}1410.$$

   El estadístico de contraste sigue una distribución t_{43}, pero considerando el elevado número de grados de libertad se puede aproximar bien por la $N(0,1)$. Para la alternativa de dos colas, los valores críticos son $-2{,}57$ y $2{,}57$. Si se opta por el contraste de una cola, el valor crítico para una significación del 1 % es 2,33. Por lo tanto, se acepta la hipótesis nula de igualdad de esperanzas con cualquiera de los enfoques.

   Si el nivel de confianza hubiese sido el 90 %, los valores críticos son $\pm 1{,}64$ para dos colas y 1,28 para una cola, por lo que también se acepta la hipótesis nula.

4. (0.75 pt) Se sabe que la rentabilidad de la primera empresa X, y el volumen negociado de sus acciones, V, tienen una distribución conjunta normal bivariante. También se dispone de la media del volumen negociado $\bar{V} = 4$, de la varianza muestral $S_V^2 = 100$ y de la covarianza muestral con X, $S_{XV} = 4$. Estime la esperanza de la rentabilidad si el volumen negociado de la empresa ha sido $V = 5$.

   Suponiendo normalidad, la esperanza condicionada es lineal: $E(X/V = v) = a + bv$. Utilizando la información muestral, las estimaciones de los parámetros son $\hat{b} = S_{XV}/S_V^2 = 4/100 = 0{,}04$ y $\hat{a} = \bar{x} - \hat{b}\bar{v} = 0{,}1 - 0{,}04 \cdot 4 = -0{,}06$. La esperanza condicionada estimada que se pide es $E(\widehat{X/V} = 5) = -0{,}06 + 0{,}04 \cdot 5 = 0{,}14$.

Problema 2. (2.5 pt.) Sea X el número de llamadas por minuto que llegan a una centralita telefónica. Se supone que $X \sim P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$, $x \geq 0, 1, 2, \ldots$; esto es, sigue una distribución de Poisson con parámetro λ. Se pide:

1. (0.5 pt) Utilice la función generatriz para calcular $E(X)$ y $V(X)$. **Nota**: La función generatriz de momentos de una v.a. $\mathcal{P}(\lambda)$ es $M(t) = e^{[\lambda(e^t-1)]}$.

$$E(X) = \left.\frac{\partial M(t)}{\partial t}\right|_{t=0} = e^{[\lambda(e^t-1)]}\lambda e^t\Big|_{t=0} = \lambda e^{\lambda(e^t-1)+t}\Big|_{t=0} = \lambda.$$

$$E(X^2) = \left.\frac{\partial^2 M(t)}{\partial t^2}\right|_{t=0} = \lambda e^{[\lambda(e^t-1)+t]}(\lambda e^t + 1)\Big|_{t=0} = \lambda^2 + \lambda.$$

La varianza resulta $V(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

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 marginal de los elementos de la muestra:

$$\pounds(\lambda/x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod x_i!}.$$

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

Para encontrar la mejor región crítica recurrimos el Teorma 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:

$$\frac{\pounds(\lambda_0)}{\pounds(\lambda_1)} = \frac{e^{-n\lambda_0}\lambda_0^{\sum x_i}/\prod x_i!}{e^{-n\lambda_1}\lambda_1^{\sum x_i}/\prod x_i!} = e^{-n(\lambda_0-\lambda_1)}\left(\frac{\lambda_0}{\lambda_1}\right)^{\sum x_i}.$$

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

$$\begin{aligned} e^{-n(\lambda_0-\lambda_1)}\left(\frac{\lambda_0}{\lambda_1}\right)^{\sum x_i} &\leq k, \\ -n(\lambda_0-\lambda_1) + \log\left(\frac{\lambda_0}{\lambda_1}\right)\sum x_i &\leq k', \\ \log\left(\frac{\lambda_0}{\lambda_1}\right)\sum x_i &\leq k'', \\ \sum_{i=1}^{n} x_i &\geq k'''. \end{aligned}$$

En el último paso se ha tenido en cuenta que $\log(\lambda_0/\lambda_1) < 0$ puesto que $0 < \lambda_0 < \lambda_1$. Como podemos encontrar la distribución del estadístico resultante, el último paso ya nos da directamente el estadístico y la región crítica.

4. (0.5 pt) Sabiendo que el estadístico $\sum x_i \sim \mathcal{P}(n\lambda)$, 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 $n = 5$.

Para encontrar el valor crítico recurrimos a la significación del contraste, que es la probabilidad de rechazar la hipótesis nula cuando es cierta. El valor crítico es el c^* que cumple:

$$P\left[\sum x_i \geq c^* \Big/ \sum x_i \sim \mathcal{P}(5)\right] = 0{,}05.$$

Buscando en las tablas de la $\mathcal{P}(5)$ se obtiene que $c^* = 9$.

5. (0.5 pt) Si con la muestra anterior se obtiene $\bar{x}$ =2.2, ¿cuál es el p-valor del contraste? ¿Qué decisión adoptaría sobre H_0 al 5 % de significación?

   Si la media de la muestra es 2,2, entonces $\sum x_i = 11$, que es mayor que el valor crítico y por tanto se rechaza la hipótesis nula. Buscando en las tablas, se obiene el p-valor:

$$\alpha^* = P\left[\sum x_i \geq 11 \Big/ \sum x_i \sim \mathcal{P}(5)\right] = 0{,}01.$$

Problema 3. (2.5 pt.) Es frecuente oir comentarios sobre si el rendimiento de los estudiantes se evalúa de forma similar en distintas asignaturas o por diferentes profesores de la misma asignatura.

1. (0.5 pt) Si se dispone de las calificaciones de **los mismos estudiantes en dos asignaturas *similares***, ¿cómo contrastaría que la posición central de las calificaciones es igual en las dos asignaturas? Formule la hipótesis nula, la alternativa y escriba el estadístico que emplearía.

   En principio se dispone de dos posibles contrastes el t de igualdad de esperanzas y el de Wilcoxon de igualdad de medianas. La elección de uno u otro depende del cumplimiento de los supuestos en que se basan. En el caso del contraste t se requiere que las muestras procedan de *distribuciones normales* e *independientes*. Por su parte el contraste de Wilcoxon no tiene requisitos previos sobre la distribución de la que se extraen las muestras.

   Aunque pudiera ser razonable suponer normalidad para las calificaciones, **no es razonable suponer que las calificaciones de los mismos estudiantes en asignaturas similares son independientes**. De hecho, el coeficiente de correlación lineal resulta 0.80 y la correlación por rangos de Spearman es 0.83, indicando claramente que las calificaciones tienen una elevada asociación y, por tanto, haciendo inadecuado el contraste t.

   El contraste a emplear es el de igualdad de mediadas, donde la hipótesis nula es $H_0 : m_1 = m_2$ frente a la alternativa $H_1 : m_1 \neq m_2$. El estadístico es $T = T^+ - T^-$, siendo T^+ la suma de los rangos para las observaciones $x_i - y_i > 0$ y análogamente T^- es la suma de los rangos asignados a las observaciones $x_i - y_i < 0$, donde los rangos se asignan al conjunto de elementos $|x_i - y_i| > 0$.

2. (0.75 pt) Aplique su respuesta anterior al siguiente conjunto de datos sobre las calificaciones de 10 estudiantes sin hacer ningún supuesto sobre la distribución de las mismas:

Estudiante	1	2	3	4	5	6	7	8	9	10
Asignatura 1	5.6	5.0	7.2	5.5	5.9	4.3	2.8	6.8	6.6	3.8
Asignatura 2	5.9	4.5	6.8	4.6	5.5	4.3	4.6	5.6	5.9	3.2
$x_i - y_i$	−0.3	0.5	0.4	0.9	0.4	0	−1.8	1.2	0.7	0.6
$rang(\|x_i - y_i\|)$	1.0	4.0	2.5	7.0	2.5	–	9.0	8.0	6.0	5.0

   Se han añadido dos filas a la tabla de datos. La primera contiene las diferencias en las calificaciones y la segunda los rangos asignados a dichas diferencias en valor absoluto. Como se puede observar, el estudiante 6 ha obtenido la misma calificación en ambas asignaturas, por lo que se excluye de la asignación de rangos.

   Los estadísticos intermedios son $T^+ = 4 + 2{,}5 + 7 + 2{,}5 + 8 + 6 + 5 = 35$ y $T^- = 1 + 9 = 10$. El estadístico de Wilcoxon es $T = T^+ - T^- = 35 - 10 = 25$. Para encontrar el valor crítico se debería recurrir a las tablas específicas de este contraste, que no están disponibles, por lo que se usa la aproximación $\frac{T}{\sqrt{V(T)}} \to N(0,1)$. La expresión de la varianza se encuentra en el formulario ($V(T) = 285$), y el valor del estadístico es $w = \frac{25}{\sqrt{285}} = 1{,}48$.

Usando la aproximación normal y con un 5 % de significación, los valores críticos son $\pm1{,}96$, por lo que se acepta la hipótesis nula de que las medianas de las calificaciones son iguales en ambas asignaturas.

3. (0.5 pt) ¿Cómo cambiaría su respuesta al **apartado 1** si pudiese suponer normalidad de los datos de calificaciones?

   No cambiaría la respuesta. Para poder aplicar el contraste t no basta con la normalidad, sino que además debe darse la independencia entre las variables aleatorias de las que proceden las muestras, cosa que en este caso no ocurre como se ha ilustrado en el apartado 1.

4. (0.75 pt) Utilice el contraste de Kolmogorov-Smirnov para comprobar si los datos de la asignatura 1 siguen la misma distribución que los de la asignatura 2. ¿Es equivalente el resultado del apartado 2 y el obtenido en éste?

   Las funciones de distribución empíricas, evaluadas en los puntos $z = 2, 4, 6, 8$ que representan bien el rango de variación de los datos, son:

z	2	4	6	8
$F^*_{10}(z)$	0/10	2/10	7/10	1
$G^*_{10}(z)$	0/10	1/10	9/10	1
$\lvert F^*_{10}(z) - G^*_{10}(z)\rvert$	0	1/10	2/10	0

   El estadístico resulta $D_{10,10} = sup|F^*_{10}(z) - G^*_{10}(z)| = 0{,}2$ y el valor crítico aproximado para una significación del 5 % es $c^* = 1{,}22\sqrt{\frac{20}{100}} = 0{,}5456$. Como la región crítica es $\{D_{10,10} \geq c^*\}$, no se rechaza la hipótesis nula de igualdad de distribuciones.

   No es equivalente el resultado. En el apartado 2 sólo se comparaban medianas, mientras que en éste se compara toda la distribución. Cabe esperar que si se acepta igualdad de distribución también se acepte igualdad en la posición central, pero lo contrario no tendría porqué ocurrir.

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

El siguiente enunciado sirve para las siguientes dos cuestiones. La siguiente tabla recoge la evolución de los precios de tres bienes (A, B y C) desde 2003 hasta 2006.

	A	B	C
2003	800	9	55
2004	850	10	50
2005	1200	18	65
2006	1400	24	75

Cuestión 1. (0.5 pt) Construya la serie de **índices simples** para el bien A con base en 2003 y otra con base en 2006.

Cuestión 2. (0.5 pt) Calcule las tasas de variación anuales para las dos series de **índices simples**. ¿Dependen las tasas de variación de los índices simples del año base que hayamos considerado? Razone la respuesta

En la siguiente tabla aparecen los índices simples y las tasas de variación interanuales pedidas:

t	$I_{t/2003}$	$I_{t/2006}$	$T^1_{t/2003}$	$T^1_{t/2006}$
2003	100,00 %	57,14 %	–	–
2004	106,25 %	60,71 %	6,25 %	6,25 %
2005	150,00 %	85,71 %	41,18 %	41,18 %
2006	175,00 %	100,00 %	16,67 %	16,67 %

Llamando P_t al precio del bien A en cada año, los índices simples con base 2003 son $I_{t/2003} = \frac{P_t}{P_{2003}}$ y con base 2006 son $I_{t/2006} = \frac{P_t}{P_{2006}}$, con $t = 2003, \ldots, 2006$. Las tasas de variación interanual usando los índices base 2003 son:

$$T^1_{t/2003} = \frac{I_{t/2003} - I_{t-1/2003}}{I_{t-1/2003}} = \frac{\frac{P_t}{P_{2003}} - \frac{P_{t-1}}{P_{2003}}}{\frac{P_{t-1}}{P_{2003}}} = \frac{P_t - P_{t-1}}{P_{t-1}}, \qquad t = 2004, 2005, 2006.$$

De forma análoga se puede proceder para las tasas con base en 2006, con lo que se demuestra que son iguales con independencia del año base utilizado para los índices.

El siguiente enunciado sirve para las siguientes tres cuestiones. Para cierto par de variables aleatorias, la función de densidad de Y condicionada a X es $f(y/X = x) = \frac{1}{3-(x/2)}$, $0 < x < 6$, $\frac{x}{2} < y < 3$.

Cuestión 3. (0.5 pt) Halle la esperanza de Y condicionada a X. ¿Cuánto es el valor esperado de Y cuando X toma valores 2, 4 y 6? **Nota:** recuerde que $a^2 - b^2 = (a + b)(a - b)$.

$$E(Y/X = x) = \int_{x/2}^{3} y \frac{1}{3-(x/2)} dy = \frac{1}{3-(x/2)} \left[\frac{x^2}{2}\right]_{x/2}^{3} = \frac{1}{2} \frac{3^2 - (x/2)^2}{3-(x/2)} = \frac{1}{2}\left(3 + \frac{x}{2}\right), 0 < x < 6.$$

La esperanza evaluada en los puntos pedidos es $E(Y/X = 2) = 2$, $E(Y/X = 4) = 2{,}5$ y $E(Y/X = 6) = 3$.

Cuestión 4. (0.5 pt) Represente gráficamente el soporte de la función de densidad y la función esperanza condicionada. A la vista del resultado anterior ¿son independientes las variables X e Y?

El soporte de la función de densidad es el triángulo de vértices $(0,0)$, $(6,3)$ y $(0,3)$. La esperanza condicionada es el segmento entre los puntos $(1{,}5,0)$ y $(6,3)$. Además no son independientes, puesto que la esperanza condicionada depende de los valores de x.

Cuestión 5. (0.5 pt) Sabiendo que $E(Y^2/X=x)=\frac{1}{12}x^2+\frac{1}{2}x+3$, $0<x<6$, calcule la varianza de Y condicionada a X. Si se sabe que $X=6$ ¿cuánta dispersión habrá en los valores de Y? Razone el resultado.

$$\begin{aligned} V(Y/X=x) &= E(Y^2/X=x)-E(Y/X=x)^2=\frac{1}{12}x^2+\frac{1}{2}x+3-\left[\frac{1}{2}\left(3+\frac{x}{2}\right)\right]^2 \\ &= \frac{x^2+6x+36}{12}-\frac{x^2+12x+36}{16} \\ &= \frac{x^2-12x+36}{48}, \qquad 0<x<6. \end{aligned}$$

No hay incertidumbre sobre el valor de Y cuando $X=6$, seguro que es 3, puesto que $V(Y/X=6)=0$.

Cuestión 6. (0.5 pt) Una variable aleatoria bidimensional $(X,Y)'$, se distribuye normal bivariante con vector de esperanzas $(0,0)'$ y matriz de varianzas covarianzas $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. ¿Cuál es la probabilidad de que $(X-Y)$ sea mayor que 1.28 condicionado a que $X=1$?

Llamando $Z=X-Y$, es necesario encontrar la distribución de $Z/X=1$. Bajo normalidad sabemos que $Z/X=1\sim N\left[E(Z/X=1),V(Z/X=1)\right]$, por lo que en primer lugar necesitamos caracterizar la distribución conjunta de X y Z.

$$E(Z)=E(X-Y)=0, \qquad V(Z)=V(X-Y)=V(X)+V(Y)-2cov(X,Y)=2,$$
$$cov(X,Z)=E(XZ)-E(X)E(Z)=E(XZ)=E(X^2)-E(XY)=1.$$

Teniendo en cuenta que la distribución conjunta de X y Z es normal, los momentos condicionados son:

$$E(Z/X=x)=E(Z)-\frac{cov(X,Z)}{V(X)}E(X)+\frac{cov(X,Z)}{V(X)}x=x.$$

$$V(Z/X=x)=V(Z)(1-\rho_{XZ}^2)=V(Z)\left[1-\left(\frac{cov(X,Z)}{\sqrt{V(X)V(Z)}}\right)^2\right]=2\left(1-\frac{1}{2}\right)=1.$$

Así, $(Z/X=1)\sim N(1,1)$ y $P(Z>1{,}28/X=1)=0{,}39$.

Cuestión 7. (0.5 pt) La función generatriz de momentos de una v.a. Poisson con parámetro λ es $M(t)=\exp\left[\lambda(e^t-1)\right]$. Demuestre que si se tienen n variables aleatorias independientes $X_i\sim\mathcal{P}(\lambda_i)$, $i=1,\ldots,n$, entonces $\sum_{i=1}^n X_i\sim\mathcal{P}(\sum\lambda_i)$.

La función generatriz de $Y=\sum X_i$, cuando las X_i son v. a. independientes es $M_Y(t)=\prod M_{X_i}(t)$. Así:

$$M_Y(t)=\prod_{i=1}^n M_{X_i}(t)=\prod_{i=1}^n e^{\lambda_i(e^t-1)}=e^{(e^t-1)\sum_{i=1}^n\lambda_i},$$

que es la función generatriz de una v. a. $\mathcal{P}(\sum\lambda_i)$.

El siguiente enunciado sirve para las tres últimas cuestiones. Sea la posible función de cuantía conjunta para X e Y, $P(X=x,Y=y)=\frac{x\cdot y}{k}$, en el soporte $x=0,1$, $y=0,1,2$.

Cuestión 8. (0.5 pt) Encuentre k para que la función anterior sea una función de cuantía.

La función de cuantía debe cumplir que la suma de las probabilidades para todos los pares de valores de x e y sea unitaria. Así:

$$\sum_{x=0}^{1}\sum_{y=0}^{2}\frac{x\cdot y}{k}=\frac{0\cdot 0}{k}+\frac{0\cdot 1}{k}+\frac{x\cdot 2}{k}+\frac{1\cdot 0}{k}+\frac{1\cdot 1}{k}+\frac{1\cdot 2}{k}=\frac{3}{k}=1 \qquad \Rightarrow \qquad k=3.$$

La función de cuantía resulta $P(X=x,Y=y)=\frac{x\cdot y}{3}$, $x=0,1$, $y=0,1,2$.

Cuestión 9. (0.5 pt) Obtenga las funciones de cuantía marginal de X e Y. ¿Son independientes X e Y?

$$P(X=x)=\sum_{y=0}^{2}\frac{x\cdot y}{3}=\frac{x\cdot 0}{3}+\frac{x\cdot 1}{3}+\frac{x\cdot 2}{3}=x, \qquad x=0,1.$$

$$P(Y=y)=\sum_{x=0}^{1}\frac{x\cdot y}{3}=\frac{0\cdot y}{3}+\frac{1\cdot y}{3}=\frac{y}{3}, \qquad y=0,1,2.$$

Son independientes, ya que se cumple que $\frac{x\cdot y}{3}=x\cdot\frac{y}{3}$ para todos los valores de x e y en el soporte.

Cuestión 10. (0.5 pt) Halle la función de cuantía condicionada $P(Y=y/X=1)$. Calcule $E(Y/X=1)$.

$$P(Y=y/X=1)=\frac{x\cdot y/3}{x}=\frac{y}{3}=P(Y=y),$$

puesto que son independientes. La esperanza condicionada, en este caso, coincide con la marginal: $E(Y/X=1)=0\frac{0}{3}+1\frac{1}{3}+2\frac{2}{3}=\frac{5}{3}$.

Fórmulas de posible utilidad

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

Aproximación lineal a la esperanza condicional.

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

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

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

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

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:

$$\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1); \qquad \frac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}; \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}; \qquad \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_s^2} \sim F_{n_1-1,n_2-1}$$

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1); \qquad \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}}} \sim t_{n_1+n_2-2}$$

donde s^2 denota la *cuasivarianza* muestral.

Proporciones. $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \to N(0,1)$. Con dos poblaciones y muestras de tamaños n_1 y n_2:

$$\frac{(\hat{p_1}-\hat{p_2})-(p_1-p_2)}{\sqrt{\left(\frac{n_1+n_2}{n_1\cdot n_2}\right)\hat{p_T}(1-\hat{p_T})}} \to N(0,1),$$

donde $\hat{p_T} = \frac{n_1\hat{p_1}+n_2\hat{p_2}}{n_1+n_2}$.

Contraste de Jarque-Bera. $JB = n\left[\frac{AS^2}{6}+\frac{(K-3)^2}{24}\right] \to \chi^2_2$.

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. $U = \text{mín}(U_1, U_2)$, donde $U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$ y $U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2$. Bajo H_0 se cumple $E(U) = \frac{n_1n_2}{2}$ y $V(U) = \frac{n_1n_2(n_1+n_2+1)}{12}$.

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

	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	$\boxed{0.6103}$	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	$\boxed{0.9495}$	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	$\boxed{0.9901}$	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	$\boxed{0.9949}$	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 1: Función de distribución de la $N(0,1)$

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
3	0.2767	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	10.2145
5	0.2672	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	5.8934
7	0.2632	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	4.7853
9	0.2610	0.5435	0.8834	1.3830	1.8331	2.2622	2.8214	4.2968
11	0.2596	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	4.0247
13	0.2586	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.8520
15	0.2579	0.5357	0.8662	1.3406	1.7531	2.1314	2.6025	3.7328
17	0.2573	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	3.6458
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 2: Función de distribución de la t_r.

λ	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.999
1	1	1	1	1	2	2	2	2	3	3	3	4	5
2	2	2	2	3	3	3	3	4	4	5	5	6	8
3	3	3	4	4	4	4	5	5	6	6	7	8	10
4	4	4	5	5	5	6	6	7	7	8	8	9	11
5	5	5	6	6	6	7	7	8	8	$\boxed{9}$	10	$\boxed{11}$	13
7	7	8	8	8	9	9	10	10	11	12	13	14	16
9	9	10	10	10	11	11	12	13	13	14	15	17	20
11	11	12	12	13	13	14	14	15	16	17	18	19	23
13	13	14	14	15	15	16	17	18	18	19	21	22	25
15	15	16	16	17	18	18	19	20	21	22	23	25	28
20	20	21	22	22	23	24	25	26	27	28	29	31	35

Cuadro 3: Función de distribución de la $\mathcal{P}(\lambda)$.