UNIVERSIDAD CARLOS III DE MADRID
ECONOMETRÍA I
Curso 2004/05
PRUEBA DE CLASE 1: SOLUCIONES

Distribución porcentual de respuestas por pregunta[1]				
PREGUNTA	(a)	(b)	(c)	(d)
1.	13	**28**	31	23
2.	6	12	**67**	12
3.	8	10	**67**	13
4.	18	**33**	8	37
5.	2	0	2	**95**
6.	**92**	0	5	1
7.	5	**68**	20	6
8.	16	**27**	28	23
9.	**50**	22	12	12
10.	3	**40**	42	13

[1] Estas estadísticas se han obtenido en base a los 82 alumnos que han respondido a la prueba en los grupos impartidos por César Alonso. En negrita figura la respuesta correcta.

1. Sean dos variables aleatorias X, Y. Indique cuál de las siguientes afirmaciones es FALSA.

   a) Si X es independiente en media de Y, entonces X e Y no están correlacionadas.
   Cierto: Recordemos que independencia en media es un concepto más fuerte que independencia lineal o ausencia de correlación.

   b) **Si X es independiente en media de Y, entonces Y es independiente en media de X.**
   Falso: A diferencia de la independencia estadística y de la ausencia de correlación, la independencia en media no es un concepto simétrico.

   c) Si X e Y son normales y $C(X,Y) = 0$, entonces X e Y son (estadísticamente) independientes.
   Cierto: la ausencia de correlación es una condición necesaria y suficiente de independencia en el caso de variables aleatorias normales, dado que toda relación entre variables normales es necesariamente lineal.
   (Este es el error más frecuente, resultado de olvidar la relación existente entre variables normales).

   d) Que X sea independiente en media de Y no implica necesariamente que X e Y son (estadísticamente) independientes.
   Cierto: la independencia en media es más débil que la independencia propiamente dicha o independencia estadística. En particular, Y no tiene por qué ser independiente en media de X; además, X puede ser independiente en media de Y pero no ser independiente en varianza de Y (es decir: qunque X sea independiente en media de X, la varianza condicional de X dado Y puede depender de los valores que tome Y).

2. Suponga que $Y = X + Z$, donde

$$\begin{array}{l} E(X) = 100, \quad E(Z) = 20, \quad V(X) = 1000, \\ V(Z) = 600, \quad C(X,Z) = 0. \end{array}$$

   Entonces:

   a) $PLO(Y|X) = 0$, porque $C(X,Z) = 0$.

   b) $PLO(Y|X) = 20$.

   c) $\mathbf{PLO(Y|X) = 20 + X}$
   (véase solución a continuación).

   d) $PLO(Y|X) = 120 + X$.

$$\begin{aligned} PLO\left(Y|\, X\right) &= E\left(Y\right)+\frac{C\left(Y,X\right)}{V\left(X\right)}\left(X-E\left(X\right)\right) \\ &= E\left(X\right)+E\left(Z\right)+\frac{V\left(X\right)+C\left(X,Z\right)}{V\left(X\right)}\left(X-E\left(X\right)\right) \\ &= E\left(Z\right)+X \\ &= 20+X \end{aligned}$$

3. Suponga que queremos encontrar el mejor predictor de Y dado X (considerando el criterio de minimizar el error cuadrático medio de la predicción).

   a) El mejor predictor es el $PLO(Y|X)$, porque es lineal.
   Falso: Recordemos que el $PLO(Y|X)$ es el mejor predictor lineal, pero el mejor predictor es $E(Y|X)$ que no tiene por qué ser lineal.

   b) El mejor predictor es $E(Y)$.
   Falso: $E(Y)$.es sólo el mejor predictor constante, y sólo coincidirá con el mejor predictor, $E(Y|X)$, cuando Y sea independiente en media de X.

   c) **El $PLO(Y|X)$ será el mejor predictor en el caso de que $E(Y|X)$ sea lineal.**
   Cierto: Si $E(Y|X)$ es lineal, entonces coincide con el $PLO(Y|X)$.

   d) Todas las respuestas anteriores son ciertas.
   Falso.

4. Sea el modelo cuadrático

$$Y=\beta_0+\beta_1 X+\beta_2 X^2+\varepsilon$$

   donde $E\left(\varepsilon|\, X\right)=0$, $\beta_1\neq 0$, $\beta_2\neq 0$.

   a) Se verifica que $E(Y\,|X\,)=PLO(Y\,|X\,)$.
   Falso: Es evidente que $E(Y\,|X\,)$ es cuadrática (y por tanto no lineal) en X, de manera que no puede coincidir con la proyección lineal de Y dado X.

   b) **Se verifica que** $E(Y\,|X\,)=PLO(Y\,|X,X^2\,)$.
   Verdadero: Considerando las variables $X, Z=X^2$, $E(Y\,|X\,)=E(Y\,|X,Z\,)$ es lineal en dichas variables, y por tanto coincide con la correspondiente proyección lineal.

   c) La función de esperanza condicional de Y dado X es lineal en X.
   Falso: Es evidente que $E(Y\,|X\,)$ es cuadrática (y por tanto no lineal) en X.

   d) Se verifica que $\beta_1=C(X,Y)/\,V(X)$.
   Falso: $C(X,Y)/\,V(X)$ es por definición la pendiente del $PLO(Y\,|X\,)$, y coincidirá con la pendiente de X en el $PLO(Y\,|X,X^2\,)$ si y solamente si $C(X,X^2)=0$, lo que en general no tiene por qué ocurrir (véase la regla de la variable omitida).

(Este es el error más frecuente, resultado de confundir el coeficiente de la regresión múltiple de una determinada variable con los del predictor lineal óptimo condicional a esa única variable).

5. En el modelo de regresión lineal simple

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

donde se cumplen los supuestos habituales:

a) El término de error verifica que $E(\varepsilon) = 0$, $C(X, \varepsilon) = 0$.

b) $\beta_0 = E(Y) - \beta_1 E(X)$.

c) $\beta_1 = C(X, Y)/V(X)$.

d) **Todas las afirmaciones anteriores son ciertas.**
Cierto: Las afirmaciones anteriores son consecuencia de los supuestos del modelo de regresión simple.

6. Considere el modelo

$$E\,(\text{Salario}\,|\text{Sexo, Educación}) = 9,5 - 4,1 \text{ Sexo} + 0,8 \text{ Educación},$$

donde el Salario (anual) está expresado en miles de euros, la Educación está medida en años y el Sexo toma el valor 0 para hombres y 1 para mujeres. Entonces, el salario esperado de un varón con 10 años de educación es:

a) **17500 euros.**
Cierto (véase la solución a continuación).

b) 9500 euros.
Falso (véase la solución a continuación): Se está considerando el salario esperado de un varón con 0 años de educación.

c) 13400 euros.
Falso (véase la solución a continuación): Se está considerando el salario esperado de una mujer con 10 años de educación.

d) 21600 euros.
Falso (véase la solución a continuación).

$E\,(\text{Salario}\,|\text{Sexo, Educación}) = 9{,}5 - 4{,}1 \times 0 + 0{,}8 \times 10 = 17{,}5$ miles de euros $=$ 17500 euros.

7. En el modelo de la pregunta 6, la diferencia esperada en el salario de un hombre y de una mujer con igual nivel de educación es:

a) 9500 euros.
Falso (véase la solución a continuación): Se está considerando por error el término constante.

b) **4100 euros.**
Cierto (véase la solución a continuación).

c) −4100 euros.
Falso (véase la solución a continuación): El signo es el opuesto al correcto.

d) 5400 euros.
Falso (véase la solución a continuación)

$E\,(\text{Salario}\,|\text{Sexo } = 0, \text{ Educación}) - E\,(\text{Salario}\,|\text{Sexo } = 1, \text{ Educación}) =$ 4,1 miles de euros = 4100 euros.

8. En el modelo de la pregunta 6, suponga además que

$$PLO\,(\text{Educación}\,|\text{Sexo}) = 8,0 - 3,0 \text{ Sexo}.$$

La pendiente del $PLO(\text{Salario}|\text{Sexo})$ sería:

a) $-4,1 + 0,8 \times 3 = -1,7.$

b) $\mathbf{-4,1 - 0,8 \times 3 = -6,5.}$

c) $-4,1.$
(Este es el error más frecuente, resultado de confundir regresión simple y regresión múltiple).

d) No se puede calcular con la información disponible.

$PLO(\text{Salario}|\text{Sexo}) = \gamma_0 + \gamma_1 Sexo$*, donde*

$$\begin{aligned} \gamma_0 &= E(\text{Salario}) - \gamma_1 E(\text{Sexo}) \\ \gamma_1 &= \beta_1 + \beta_2 \delta_1 \text{ (regla de la variable omitida)} \end{aligned}$$

donde β_1, β_2 *son los coeficientes de Sexo y Educación, respectivamente, en* $E\,(\text{Salario}\,|\text{Sexo, Educación})$ *y* δ_1 *es la pendiente del* $PLO\,(\text{Educación}\,|\text{Sexo})$*, es decir,*

$$\delta_1 = \frac{C(\text{Sexo, Educación})}{V(\text{Sexo})}.$$

Por tanto,

$$\gamma_1 = -4{,}1 + 0{,}8 \times (-3{,}0) = -6{,}5$$

9. Teniendo en cuenta la información de las preguntas 6 y 8, considere el coeficiente del Sexo en $E\,(\text{Salario}\,|\text{Sexo, Educación})$ y la pendiente del $PLO(\text{Salario}|\text{Sexo})$. Indique cuál de las siguientes afirmaciones es FALSA.

a) **Ambas medidas son equivalentes.**
Falso: A diferencia del coeficiente del Sexo en $E\,(\text{Salario}\,|\text{Sexo, Educación})$*, la pendiente del* $PLO(\text{Salario}|\text{Sexo})$ *no tiene en cuenta que la distribución de los niveles de educación difiere por sexo (como se ve en la pregunta anterior).*

b) La pendiente del $PLO(\text{Salario}|\text{Sexo})$ sobrevalora la discriminación salarial entre hombres y mujeres, dado que las mujeres tienen en promedio menor nivel de educación.
Cierto: Como se ve en el enunciado de la pregunta anterior, el Sexo está negativamente correlacionado con la Educación.

c) El coeficiente del sexo en $E(\text{Salario}\,|\text{Sexo, Educación})$, mide la diferencia en el salario medio entre hombres y mujeres manteniendo constante la educación, y por tanto es una medida más apropiada de discriminación.
Cierto: dicha medida tiene en cuenta que la distribución de niveles de educación difiere por sexos.

d) La pendiente del $PLO(\text{Salario}|\text{Sexo})$ no proporciona el efecto puro de la discriminación porque la distribución de niveles de educación difiere por sexos.
Cierto.

10. Considere el modelo de regresión múltiple

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

con los supuestos habituales. Indique cuál de las siguientes afirmaciones es FALSA.

a) Para que la función de esperanza condicional de Y dados X_1, X_2 sea lineal (y coincida por tanto con el predictor lineal óptimo) es necesario que $E(\varepsilon|\, X_1, X_2) = 0$.
Cierto: véanse los supuestos del modelo de regresión.

b) **Para que la función de esperanza condicional de Y dados X_1, X_2 sea lineal (y coincida por tanto con el predictor lineal óptimo) es necesario que $\mathbf{E}(\varepsilon|\, X_1, X_2) = \mathbf{0}$ y que $\mathbf{V}(\varepsilon|\, X_1, X_2) = \boldsymbol{\sigma}^2$.**
Falso: el supuesto de homocedasticidad condicional no afecta a la forma funcional de la función de esperanza condicional.

c) El incumplimiento del supuesto $V(\varepsilon|\, X_1, X_2) = \sigma^2$.no afecta a la linealidad de la función de esperanza condicional de Y dados X_1, X_2.
Cierto.

d) El hecho de que la función de esperanza condicional de Y dados X_1, X_2 sea lineal no implica que la función de esperanza condicional de Y dado X_1 sea también lineal
Cierto: la forma funcional de la función de esperanza condicional de Y dados X_1, X_2 no dice nada sobre la forma funcional de la función de esperanza condicional de Y dado X_1