Estadística: ModelosLineales

Final Enero 2.002, Tipo: $\boxed{\text{A}}$

Apellidos y Nombre: ______________________

DNI: ______________________

Grupo: ______________________

Profesor : ______________________

Sección 1. Instrucciones

Salvo que se indique lo contrario, las preguntas bien contestadas valen un punto. Puede haber más de una respuesta correcta, y para obtener puntuación has de señalarlas todas. Preguntas que no estén bien contestadas puntúan -0.5 veces su valor.

Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay más de una persona que echa a perder una nota potencialmente buena por responder temeraria o atolondradamente.

¡Por favor, fíjate bien en todos los detalles!

Te ayudará proceder por exclusión de absurdos. Si una pregunta te parece ambigua, anota brevemente la razón al margen y no la contestes.

1. Para contrastar en un modelo de regresión la hipótesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, adoptamos el siguiente criterio: formar los cuatro estadísticos t (cocientes $\hat{\beta}_i/\hat{\sigma}_{\hat{\beta}_i}$) y rechazar H_0 cuando cualquiera de ellos excede en valor absoluto de $t_{N-p;\alpha}$ (cuantil dejando una cola de tamaño α en la distribución t_{N-p}). Tal modo de operar nos dará un nivel de significación:

   (a) Exactamente igual a α.

   (b) Mayor que α.

   (c) Menor que α.

   (d) Depende de los valores de los demás parámetros en el modelo.

2. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de grados de libertad de la suma de cuadrados de los residuos, SSE, es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

3. En un problema de estimación condicionada,

   (a) SSE_h (la suma de cuadrados en el modelo estimado con la condición) nunca será menor que SSE del modelo sin la condición.

   (b) Las estimaciones de los parámetros no afectados por la condición no siempre serán insesgadas.

   (c) Los estimadores tendrán menor varianza que los estimadores MCO si, y solo si, la condición impuesta es "correcta" (verificada por los parámetros del modelo).

   (d) Todo falso.

4. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de parámetros libres es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

5. ¿En cuál de las siguientes circunstancias estaría justificado emplear un modelo de regresión lineal sin columna de "unos" (y sin estimar β_0 por tanto)?

   (a) Cuando las unidades de medida de *todos* los regresores son idénticas.

   (b) Cuanto tanto regresores como regresando han sido centrados.

   (c) Cuando los regresores se expresan todos en las mismas unidades y el origen tomado es generalmente aceptado —como el origen de la era cristiana para medir tiempos o el punto de congelación del agua para medir temperaturas—.

   (d) Siempre, en toda circunstancia, hay que utilizar una columna de "unos" como regresor.

6. Al estimar un modelo de regresión logística, consideramos un modelo M_1 con cinco parámetros y otro M_2 con los cinco de M_1 y otros dos, que denotaremos β_5, β_6. Sean $L(M_1)$ y $L(M_2)$ los valores máximos de la verosimilitud al ajustar ambos modelos. Bajo la hipótesis de que $\beta_5 = \beta_6 = 0$,

$$-2 \ln \left(\frac{L(M_1)}{L(M_2)} \right)$$

se distribuye como:

   (a) χ^2_2
   (b) χ^2_7
   (c) $\mathcal{F}_{2,7}$
   (d) $\mathcal{F}_{2,5}$

7. ¿Cuál o cuáles de las siguientes son medidas de influencia?

   (a) Distancia de Cook, D_i.
   (b) DFFITS$_i$.
   (c) C_p de Mallows.
   (d) Ninguna de ellas.

8. En un diseño no completo como el de cuadrado latino se pierde:

   (a) La condición de ortogonalidad entre las sumas de cuadrados atribuibles a cada efecto.
   (b) La homoscedasticidad de las perturbaciones, que pasan a ser heteroscedásticas por construcción.
   (c) La posibilidad de estimar todas o algunas de las interacciones.
   (d) La linealidad de la especificación.

9. ¿Cuál o cuáles de los siguientes supuestos son necesarios para garantizar que la R^2 está entre 0 y 1?

   (a) Normalidad en las perturbaciones.
   (b) Modelo “correcto”.
   (c) Modelo incluye una columna de “unos” (u otros regresores que combinados linealmente generan $\vec{1}$).
   (d) Modelo con efectos simples (no interacciones de ningún género).

10. En un modelo de regresión lineal, la omisión de regresores que deberían formar parte de la especificación:

   (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.
   (b) Sesga, en general, los estimadores de los parámetros incluidos.
   (c) Puede no afectar al sesgo de los estimadores de los parámetros incluidos, si acontece que los regresores omitidos son ortogonales a los incluidos.
   (d) Disminuye, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo completo.

11. ¿En qué condiciones puede $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ (proyección de $\vec{\boldsymbol{y}}$ sobre el subespacio que generan las columnas de $\boldsymbol{X}$) no ser única?

   (a) Cuando haya acusada multicolinealidad entre las columnas de $\boldsymbol{X}$.
   (b) Siempre es única.
   (c) No tiene por qué ser única en ningún caso.
   (d) Todo falso.

12. Las ecuaciones normales cuando la matriz $\boldsymbol{X}$ es de rango completo:

   (a) Pueden tener más de una solución.
   (b) Tienen siempre solución única.
   (c) Dan lugar a múltiples estimadores de $\hat{\boldsymbol{\beta}}$.
   (d) Forman un sistema de ecuaciones no lineal, que puede resolverse mediante técnicas iterativas.

13. Al estimar un modelo lineal mediante regresión *ridge*, es decir, $\hat{\boldsymbol{\beta}}_{(\boldsymbol{k})} = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\vec{\boldsymbol{y}} \ldots$

   (a) Sólamente es posible mejorar el ECM (error cuadrático medio) en los casos en que la multicolinealidad es perfecta.
   (b) Un incremento de k acarrea un incremento del sesgo del estimador.
   (c) Un incremento de k acarrea un incremento de la varianza del estimador.
   (d) Un incremento de k acarrea una disminución de la varianza del estimador.

14. ¿Cuál de los siguientes criterios penaliza más la inclusión de regresores adicionales en un modelo?

    (a) Distancia de Cook, D_i.

    (b) DFFITS$_i$.

    (c) C_p de Mallows.

    (d) $\overline{R}^2$.

15. ¿Cuál o cuáles de los siguientes procedimientos para seleccionar la especificación de un modelo garantizan encontrar un óptimo global (es decir, que ningún otro modelo en la clase considerada mejora al al seleccionado de acuerdo con el criterio adoptado)?

    (a) *Stepwise* hacia adelante.

    (b) *All-subsets* o regresión sobre todos los subconjuntos.

    (c) *Stepwise* hacia atrás.

    (d) Validación cruzada.

16. Los "residuos borrados" *(deleted residuals)* d_i pueden obtenerse a partir de los mínimo cuadráticos $\hat{\epsilon}_i$, los términos diagonales p_{ii} de la matriz de proyección y estimadores adecuados de la varianza de la perturbación haciendo uso de la fórmula siguiente:

    (a) $d_i = \hat{\epsilon}_i / \sqrt{\hat{\sigma}^2(1 - p_{ii})}$

    (b) $d_i = \hat{\epsilon}_i / \sqrt{\hat{\sigma}^2_{(i)}(1 - p_{ii})}$

    (c) $d_i = \hat{\epsilon}_i / (1 - p_{ii})$

    (d) $d_i = \hat{\epsilon}_i / (\hat{\sigma}^2(1 - p_{ii}))$

17. Al ajustar una regresión, las observaciones i-ésima y j-ésima proporcionan los siguientes residuos externamente *studentizados* y borrados: $t_i = 3{,}3$, $t_j = 0{,}2$, $d_i = 3{,}4$ y $d_j = 3{,}9$. ¿Cuál de las dos observaciones señalarías como más sospechosa de ser influyente?

    (a) La i-ésima.

    (b) La j-ésima.

    (c) Ninguna.

    (d) Las dos son manifiestamente influyentes.

18. ¿Cuando es inadecuada la inclusión de una columna de "unos" entre los regresores?

    (a) Cuando el número de grados de libertad es relativamente elevado.

    (b) Cuando las escalas de los regresores sean muy similares.

    (c) Cuando tanto regresando como regresores se toman en desviaciones respecto a su media. En tal caso, β_0 sería siempre 0.

    (d) Cuando las escalas de los regresores sean muy diferentes.

19. Cuando contrastamos la hipótesis de nulidad de todos los parámetros de una regresión (salvo β_0), el estadístico de contraste Q_h toma una forma particularmente simple. Puede escribirse en función de N, p y. . .

    (a) La C_p de Mallows.

    (b) R^2

    (c) El criterio AIC.

    (d) SSE únicamente.

20. Su $\hat{\boldsymbol{\beta}}_*$ es un estimador lineal del vector de parámetros en un modelo de regresión, necesariamente:

    (a) Su matriz de covarianzas tiene en la diagonal principal elementos no menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$; lo garantiza el teorema de Gauss-Markov.

    (b) Puede tener en la diagonal principal de su matriz de covarianzas términos menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

    (c) Necesariamente ha de coincidir con el estimador MCO.

    (d) Necesariamente ha de ser insesgado.

21. Al ajustar un modelo con 4 regresores (incluida la columna de "unos") a $N = 120$ observaciones, se obtiene $SSE = 345{,}3$. Se añaden otros tres regresores al modelo, con parámetros $\beta_4, \beta_5, \beta_6$. La nueva regresión proporciona ahora $SSE^* = 312{,}3$. El estadístico Q_h para contrastar la hipótesis $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ toma el valor:

    (a) 2,35

    (b) 3,98

    (c) 2,853

    (d) 1,26

22. h es un subespacio del espacio vectorial M. ¿Cuáles de las siguientes igualdades son correctas?

    (a) $P_M = P_h$

    (b) $P_M P_h = P_h P_M$

    (c) $P_{M\cap h} = P_h$

    (d) $P_{M\cap h} = P_M - P_h$

23. Al objeto de diagnosticar la multicolinealidad aproximada y los regresores que pueden causarla, computamos los factores de incremento de varianza (VIF). Adicionalmente, practicamos las regresiones de cada uno de los regresores sobre todos los restantes. Para el regresor i-ésimo, obtenemos que la $R^2(i)$ de dicha regresión es 0.90. Ello quiere decir que su VIF será:

    (a) 0.10

    (b) 10

    (c) 1 / 0.90

    (d) 0.90 /(1 - 0.90), o sea 9

24. En un modelo de regresión, el método de estimación en componentes principales utilizando tantas componentes principales como regresores hay, daría:

    (a) El mismo resultado que el método de estimación *ridge* con $k \to \infty$.

    (b) El mismo resultado que el método de estimación *ridge* con $k = 0$

    (c) El mismo resultado que el método de estimación MCO.

    (d) Todo falso.

25. Una matriz de proyección es *siempre*:

    (a) Simétrica

    (b) Idempotente

    (c) De rango completo

    (d) Ortogonal

26. En un modelo de regresión lineal, la inclusión de regresores irrelevantes:

    (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

    (b) Sesga, en general, los estimadores de los parámetros incluidos.

    (c) Afecta a las varianzas de los estimadores de los parámetros "correctos", pero sólo si acontece que los regresores irrelevantes son ortogonales a los "correctos".

    (d) Aumenta, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo correcto.

27. Decimos que una combinación lineal de los parámetros, $\vec{a}'\hat{\boldsymbol{\beta}}$, es estimable cuando:

    (a) El vector $\vec{a}$ está unívocamente definido.

    (b) Las ecuaciones normales tiene más de una solución.

    (c) Hay una combinación lineal de las $\vec{Y}$ cuyo valor medio coincide con $\vec{a}'\hat{\boldsymbol{\beta}}$.

    (d) Las ecuaciones normales tienen una única solución.

28. Si las unidades experimentales con que contamos (por ej., parcelas) no son homogéneas (por ejemplo, hay unas que sabemos más fértiles que otras), al especificar un modelo de Análisis de Varianza:

    (a) Prescindiremos por completo de este hecho.

    (b) Introduciremos las diferencias de fertilidad como factor explicativo, recogiendo su efecto en un grupo de parámetros $\{\alpha_i^A\}$.

    (c) Nos abstendremos de hacer Análisis de Varianza: la inhomogeneidad de las unidades experimentales es una violación de los supuestos incompatible con el análisis.

    (d) Todo falso.

29. Un modelo ANOVA está jerárquicamente bien estructurado...

    (a) Cuando se ha consultado al catedrático de la asignatura sobre la especificación, y este ha dado su conformidad.

    (b) Cuando incluye columna de "unos".

    (c) Cuando verifica restricciones de marginalidad: la ausencia de un efecto implica la de todas las interacciones en las que dicho efecto tomaría parte.

    (d) Cuando el diseño es ortogonal y equilibrado.

30. ¿Cuales de los siguientes tipos de residuos tienen varianzas iguales?

    (a) Los internamente *studentizados*, r_i.

    (b) Los externamente *studentizados*, t_i.

    (c) Los borrados. d_i.

    (d) Los brutos, $\hat{\epsilon}_i$ o MCO.

Respuestas al examen de tipo $\boxed{\text{A}}$

Sección 1. Instrucciones

Salvo que se indique lo contrario, las preguntas bien contestadas valen un punto. Puede haber más de una respuesta correcta, y para obtener puntuación has de señalarlas todas. Preguntas que no estén bien contestadas puntúan -0.5 veces su valor.

Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay más de una persona que echa a perder una nota potencialmente buena por responder temeraria o atolondradamente.

¡Por favor, fíjate bien en todos los detalles!

Te ayudará proceder por exclusión de absurdos. Si una pregunta te parece ambigua, anota brevemente la razón al margen y no la contestes.

1. Para contrastar en un modelo de regresión la hipótesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, adoptamos el siguiente criterio: formar los cuatro estadísticos t (cocientes $\hat{\beta}_i/\hat{\sigma}_{\hat{\beta}_i}$) y rechazar H_0 cuando cualquiera de ellos excede en valor absoluto de $t_{N-p;\alpha}$ (cuantil dejando una cola de tamaño α en la distribución t_{N-p}). Tal modo de operar nos dará un nivel de significación:

   (a) Exactamente igual a α.

   (b) Mayor que α.

   (c) Menor que α.

   (d) Depende de los valores de los demás parámetros en el modelo.

2. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de grados de libertad de la suma de cuadrados de los residuos, SSE, es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

3. En un problema de estimación condicionada,

   (a) SSE_h (la suma de cuadrados en el modelo estimado con la condición) nunca será menor que SSE del modelo sin la condición.

   (b) Las estimaciones de los parámetros no afectados por la condición no siempre serán insesgadas.

   (c) Los estimadores tendrán menor varianza que los estimadores MCO si, y solo si, la condición impuesta es "correcta" (verificada por los parámetros del modelo).

   (d) Todo falso.

4. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de parámetros libres es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

5. ¿En cuál de las siguientes circunstancias estaría justificado emplear un modelo de regresión lineal sin columna de "unos" (y sin estimar β_0 por tanto)?

   (a) Cuando las unidades de medida de *todos* los regresores son idénticas.

   (b) Cuanto tanto regresores como regresando han sido centrados.

   (c) Cuando los regresores se expresan todos en las mismas unidades y el origen tomado es generalmente aceptado —como el origen de la era cristiana para medir tiempos o el punto de congelación del agua para medir temperaturas—.

   (d) Siempre, en toda circunstancia, hay que utilizar una columna de "unos" como regresor.

6. Al estimar un modelo de regresión logística, consideramos un modelo M_1 con cinco parámetros y otro M_2 con los cinco de M_1 y otros dos, que denotaremos β_5, β_6. Sean $L(M_1)$ y $L(M_2)$ los valores máximos de la verosimilitud al ajustar ambos modelos. Bajo la hipótesis de que $\beta_5 = \beta_6 = 0$,

$$-2 \ln \left(\frac{L(M_1)}{L(M_2)} \right)$$

se distribuye como:

(a) χ^2_2

(b) χ^2_7

(c) $\mathcal{F}_{2,7}$

(d) $\mathcal{F}_{2,5}$

7. ¿Cuál o cuáles de las siguientes son medidas de influencia?

(a) Distancia de Cook, D_i.

(b) DFFITS$_i$.

(c) C_p de Mallows.

(d) Ninguna de ellas.

8. En un diseño no completo como el de cuadrado latino se pierde:

(a) La condición de ortogonalidad entre las sumas de cuadrados atribuibles a cada efecto.

(b) La homoscedasticidad de las perturbaciones, que pasan a ser heteroscedásticas por construcción.

(c) La posibilidad de estimar todas o algunas de las interacciones.

(d) La linealidad de la especificación.

9. ¿Cuál o cuáles de los siguientes supuestos son necesarios para garantizar que la R^2 está entre 0 y 1?

(a) Normalidad en las perturbaciones.

(b) Modelo "correcto".

(c) Modelo incluye una columna de "unos" (u otros regresores que combinados linealmente generan $\vec{1}$).

(d) Modelo con efectos simples (no interacciones de ningún género).

10. En un modelo de regresión lineal, la omisión de regresores que deberían formar parte de la especificación:

(a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

(b) Sesga, en general, los estimadores de los parámetros incluidos.

(c) Puede no afectar al sesgo de los estimadores de los parámetros incluidos, si acontece que los regresores omitidos son ortogonales a los incluidos.

(d) Disminuye, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo completo.

11. ¿En qué condiciones puede $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ (proyección de $\vec{\boldsymbol{y}}$ sobre el subespacio que generan las columnas de $\boldsymbol{X}$) no ser única?

(a) Cuando haya acusada multicolinealidad entre las columnas de $\boldsymbol{X}$.

(b) Siempre es única.

(c) No tiene por qué ser única en ningún caso.

(d) Todo falso.

12. Las ecuaciones normales cuando la matriz $\boldsymbol{X}$ es de rango completo:

(a) Pueden tener más de una solución.

(b) Tienen siempre solución única.

(c) Dan lugar a múltiples estimadores de $\hat{\boldsymbol{\beta}}$.

(d) Forman un sistema de ecuaciones no lineal, que puede resolverse mediante técnicas iterativas.

13. Al estimar un modelo lineal mediante regresión *ridge*, es decir, $\hat{\boldsymbol{\beta}}_{(\boldsymbol{k})} = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\vec{\boldsymbol{y}} \ldots$

(a) Sólamente es posible mejorar el ECM (error cuadrático medio) en los casos en que la multicolinealidad es perfecta.

(b) Un incremento de k acarrea un incremento del sesgo del estimador.

(c) Un incremento de k acarrea un incremento de la varianza del estimador.

(d) Un incremento de k acarrea una disminución de la varianza del estimador.

14. ¿Cuál de los siguientes criterios penaliza más la inclusión de regresores adicionales en un modelo?

   (a) Distancia de Cook, D_i.

   (b) DFFITS$_i$.

   (c) C_p de Mallows.

   (d) $\overline{R}^2$.

15. ¿Cuál o cuáles de los siguientes procedimientos para seleccionar la especificación de un modelo garantizan encontrar un óptimo global (es decir, que ningún otro modelo en la clase considerada mejora al al seleccionado de acuerdo con el criterio adoptado)?

   (a) *Stepwise* hacia adelante.

   (b) *All-subsets* o regresión sobre todos los subconjuntos.

   (c) *Stepwise* hacia atrás.

   (d) Validación cruzada.

16. Los "residuos borrados" *(deleted residuals)* d_i pueden obtenerse a partir de los mínimo cuadráticos $\hat{\epsilon}_i$, los términos diagonales p_{ii} de la matriz de proyección y estimadores adecuados de la varianza de la perturbación haciendo uso de la fórmula siguiente:

   (a) $d_i = \hat{\epsilon}_i / \sqrt{\hat{\sigma}^2(1 - p_{ii})}$

   (b) $d_i = \hat{\epsilon}_i / \sqrt{\hat{\sigma}^2_{(i)}(1 - p_{ii})}$

   (c) $d_i = \hat{\epsilon}_i / (1 - p_{ii})$

   (d) $d_i = \hat{\epsilon}_i / (\hat{\sigma}^2(1 - p_{ii}))$

17. Al ajustar una regresión, las observaciones i-ésima y j-ésima proporcionan los siguientes residuos externamente *studentizados* y borrados: $t_i = 3{,}3$, $t_j = 0{,}2$, $d_i = 3{,}4$ y $d_j = 3{,}9$. ¿Cuál de las dos observaciones señalarías como más sospechosa de ser influyente?

   (a) La i-ésima.

   (b) La j-ésima.

   (c) Ninguna.

   (d) Las dos son manifiestamente influyentes.

18. ¿Cuando es inadecuada la inclusión de una columna de "unos" entre los regresores?

   (a) Cuando el número de grados de libertad es relativamente elevado.

   (b) Cuando las escalas de los regresores sean muy similares.

   (c) Cuando tanto regresando como regresores se toman en desviaciones respecto a su media. En tal caso, β_0 sería siempre 0.

   (d) Cuando las escalas de los regresores sean muy diferentes.

19. Cuando contrastamos la hipótesis de nulidad de todos los parámetros de una regresión (salvo β_0), el estadístico de contraste Q_h toma una forma particularmente simple. Puede escribirse en función de N, p y...

   (a) La C_p de Mallows.

   (b) R^2

   (c) El criterio AIC.

   (d) SSE únicamente.

20. Su $\hat{\boldsymbol{\beta}}_*$ es un estimador lineal del vector de parámetros en un modelo de regresión, necesariamente:

   (a) Su matriz de covarianzas tiene en la diagonal principal elementos no menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$; lo garantiza el teorema de Gauss-Markov.

   (b) Puede tener en la diagonal principal de su matriz de covarianzas términos menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

   (c) Necesariamente ha de coincidir con el estimador MCO.

   (d) Necesariamente ha de ser insesgado.

21. Al ajustar un modelo con 4 regresores (incluida la columna de "unos") a $N = 120$ observaciones, se obtiene $SSE = 345{,}3$. Se añaden otros tres regresores al modelo, con parámetros $\beta_4, \beta_5, \beta_6$. La nueva regresión proporciona ahora $SSE^* = 312{,}3$. El estadístico Q_h para contrastar la hipótesis $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ toma el valor:

   (a) 2,35

   (b) 3,98

   (c) 2,853

   (d) 1,26

22. h es un subespacio del espacio vectorial M. ¿Cuáles de las siguientes igualdades son correctas?

    (a) $P_M = P_h$

    (b) $P_M P_h = P_h P_M$

    (c) $P_{M \cap h} = P_h$

    (d) $P_{M \cap h} = P_M - P_h$

23. Al objeto de diagnosticar la multicolinealidad aproximada y los regresores que pueden causarla, computamos los factores de incremento de varianza (VIF). Adicionalmente, practicamos las regresiones de cada uno de los regresores sobre todos los restantes. Para el regresor i-ésimo, obtenemos que la $R^2(i)$ de dicha regresión es 0.90. Ello quiere decir que su VIF será:

    (a) 0.10

    (b) **10**

    (c) 1 / 0.90

    (d) 0.90 /(1 - 0.90), o sea 9

24. En un modelo de regresión, el método de estimación en componentes principales utilizando tantas componentes principales como regresores hay, daría:

    (a) El mismo resultado que el método de estimación *ridge* con $k \to \infty$.

    (b) **El mismo resultado que el método de estimación *ridge* con** $k = 0$

    (c) **El mismo resultado que el método de estimación MCO.**

    (d) Todo falso.

25. Una matriz de proyección es *siempre*:

    (a) **Simétrica**

    (b) **Idempotente**

    (c) De rango completo

    (d) Ortogonal

26. En un modelo de regresión lineal, la inclusión de regresores irrelevantes:

    (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

    (b) Sesga, en general, los estimadores de los parámetros incluidos.

    (c) Afecta a las varianzas de los estimadores de los parámetros "correctos", pero sólo si acontece que los regresores irrelevantes son ortogonales a los "correctos".

    (d) **Aumenta, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo correcto.**

27. Decimos que una combinación lineal de los parámetros, $\vec{a}'\hat{\beta}$, es estimable cuando:

    (a) El vector $\vec{a}$ está unívocamente definido.

    (b) Las ecuaciones normales tiene más de una solución.

    (c) **Hay una combinación lineal de las $\vec{Y}$ cuyo valor medio coincide con $\vec{a}'\hat{\beta}$.**

    (d) Las ecuaciones normales tienen una única solución.

28. Si las unidades experimentales con que contamos (por ej., parcelas) no son homogéneas (por ejemplo, hay unas que sabemos más fértiles que otras), al especificar un modelo de Análisis de Varianza:

    (a) Prescindiremos por completo de este hecho.

    (b) **Introduciremos las diferencias de fertilidad como factor explicativo, recogiendo su efecto en un grupo de parámetros $\{\alpha_i^A\}$.**

    (c) Nos abstendremos de hacer Análisis de Varianza: la inhomogeneidad de las unidades experimentales es una violación de los supuestos incompatible con el análisis.

    (d) Todo falso.

29. Un modelo ANOVA está jerárquicamente bien estructurado. . .

    (a) Cuando se ha consultado al catedrático de la asignatura sobre la especificación, y este ha dado su conformidad.

    (b) Cuando incluye columna de "unos".

    (c) **Cuando verifica restricciones de marginalidad: la ausencia de un efecto implica la de todas las interacciones en las que dicho efecto tomaría parte.**

    (d) Cuando el diseño es ortogonal y equilibrado.

30. ¿Cuales de los siguientes tipos de residuos tienen varianzas iguales?

    (a) **Los internamente *studentizados*, r_i.**

    (b) **Los externamente *studentizados*, t_i.**

    (c) Los borrados. d_i.

    (d) Los brutos, $\hat{\epsilon}_i$ o MCO.

Estadística: ModelosLineales

Final Enero 2.002, Tipo: $\boxed{\text{B}}$

Apellidos y Nombre: ______________________

DNI: ______________________

Grupo: ______________________

Profesor : ______________________

Sección 1. Instrucciones

Salvo que se indique lo contrario, las preguntas bien contestadas valen un punto. Puede haber más de una respuesta correcta, y para obtener puntuación has de señalarlas todas. Preguntas que no estén bien contestadas puntúan -0.5 veces su valor.

Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay más de una persona que echa a perder una nota potencialmente buena por responder temeraria o atolondradamente.

¡Por favor, fíjate bien en todos los detalles!

Te ayudará proceder por exclusión de absurdos. Si una pregunta te parece ambigua, anota brevemente la razón al margen y no la contestes.

1. En un modelo de regresión lineal, la omisión de regresores que deberían formar parte de la especificación:

   (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

   (b) Sesga, en general, los estimadores de los parámetros incluidos.

   (c) Puede no afectar al sesgo de los estimadores de los parámetros incluidos, si acontece que los regresores omitidos son ortogonales a los incluidos.

   (d) Disminuye, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo completo.

2. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de parámetros libres es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

3. h es un subespacio del espacio vectorial M. ¿Cuáles de las siguientes igualdades son correctas?

   (a) $P_M = P_h$

   (b) $P_M P_h = P_h P_M$

   (c) $P_{M\cap h} = P_h$

   (d) $P_{M\cap h} = P_M - P_h$

4. Al ajustar una regresión, las observaciones i-ésima y j-ésima proporcionan los siguientes residuos externamente *studentizados* y borrados: $t_i = 3{,}3$, $t_j = 0{,}2$, $d_i = 3{,}4$ y $d_j = 3{,}9$. ¿Cuál de las dos observaciones señalarías como más sospechosa de ser influyente?

   (a) La i-ésima.

   (b) La j-ésima.

   (c) Ninguna.

   (d) Las dos son manifiestamente influyentes.

5. Una matriz de proyección es *siempre*:

   (a) Simétrica

   (b) Idempotente

   (c) De rango completo

   (d) Ortogonal

6. ¿Cuales de los siguientes tipos de residuos tienen varianzas iguales?

   (a) Los internamente *studentizados*, r_i.

   (b) Los externamente *studentizados*, t_i.

   (c) Los borrados. d_i.

   (d) Los brutos, $\hat{\epsilon}_i$ o MCO.

7. Al ajustar un modelo con 4 regresores (incluida la columna de "unos") a $N = 120$ observaciones, se obtiene $SSE = 345{,}3$. Se añaden otros tres regresores al modelo, con parámetros $\beta_4, \beta_5, \beta_6$. La nueva regresión proporciona ahora $SSE^* = 312{,}3$. El estadístico Q_h para contrastar la hipótesis $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ toma el valor:

   (a) 2,35

   (b) 3,98

   (c) 2,853

   (d) 1,26

8. ¿Cuál o cuáles de los siguientes supuestos son necesarios para garantizar que la R^2 está entre 0 y 1?

   (a) Normalidad en las perturbaciones.

   (b) Modelo "correcto".

   (c) Modelo incluye una columna de "unos" (u otros regresores que combinados linealmente generan $\vec{\mathbf{1}}$).

   (d) Modelo con efectos simples (no interacciones de ningún género).

9. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de grados de libertad de la suma de cuadrados de los residuos, SSE, es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

10. Cuando contrastamos la hipótesis de nulidad de todos los parámetros de una regresión (salvo β_0), el estadístico de contraste Q_h toma una forma particularmente simple. Puede escribirse en función de N, p y...

    (a) La C_p de Mallows.

    (b) R^2

    (c) El criterio AIC.

    (d) SSE únicamente.

11. ¿Cuando es inadecuada la inclusión de una columna de "unos" entre los regresores?

    (a) Cuando el número de grados de libertad es relativamente elevado.

    (b) Cuando las escalas de los regresores sean muy similares.

    (c) Cuando tanto regresando como regresores se toman en desviaciones respecto a su media. En tal caso, β_0 sería siempre 0.

    (d) Cuando las escalas de los regresores sean muy diferentes.

12. En un modelo de regresión lineal, la inclusión de regresores irrelevantes:

    (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

    (b) Sesga, en general, los estimadores de los parámetros incluidos.

    (c) Afecta a las varianzas de los estimadores de los parámetros "correctos", pero sólo si acontece que los regresores irrelevantes son ortogonales a los "correctos".

    (d) Aumenta, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo correcto.

13. Al estimar un modelo lineal mediante regresión *ridge*, es decir, $\hat{\boldsymbol{\beta}}_{(\boldsymbol{k})} = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\vec{\boldsymbol{y}}\ldots$

    (a) Sólamente es posible mejorar el ECM (error cuadrático medio) en los casos en que la multicolinealidad es perfecta.

    (b) Un incremento de k acarrea un incremento del sesgo del estimador.

    (c) Un incremento de k acarrea un incremento de la varianza del estimador.

    (d) Un incremento de k acarrea una disminución de la varianza del estimador.

14. Al estimar un modelo de regresión logística, consideramos un modelo M_1 con cinco parámetros y otro M_2 con los cinco de M_1 y otros dos, que denotaremos β_5, β_6. Sean $L(M_1)$ y $L(M_2)$ los valores máximos de la verosimilitud al ajustar ambos modelos. Bajo la hipótesis de que $\beta_5 = \beta_6 = 0$,

$$-2 \ln \left(\frac{L(M_1)}{L(M_2)} \right)$$

    se distribuye como:

    (a) χ_2^2

    (b) χ_7^2

    (c) $\mathcal{F}_{2,7}$

    (d) $\mathcal{F}_{2,5}$

15. Su $\hat{\boldsymbol{\beta}}_*$ es un estimador lineal del vector de parámetros en un modelo de regresión, necesariamente:
   (a) Su matriz de covarianzas tiene en la diagonal principal elementos no menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$; lo garantiza el teorema de Gauss-Markov.
   (b) Puede tener en la diagonal principal de su matriz de covarianzas términos menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
   (c) Necesariamente ha de coincidir con el estimador MCO.
   (d) Necesariamente ha de ser insesgado.

16. ¿Cuál de los siguientes criterios penaliza más la inclusión de regresores adicionales en un modelo?
   (a) Distancia de Cook, D_i.
   (b) DFFITS$_i$.
   (c) C_p de Mallows.
   (d) $\overline{R}^2$.

17. ¿En qué condiciones puede $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ (proyección de $\vec{\boldsymbol{y}}$ sobre el subespacio que generan las columnas de $\boldsymbol{X}$) no ser única?
   (a) Cuando haya acusada multicolinealidad entre las columnas de $\boldsymbol{X}$.
   (b) Siempre es única.
   (c) No tiene por qué ser única en ningún caso.
   (d) Todo falso.

18. Las ecuaciones normales cuando la matriz $\boldsymbol{X}$ es de rango completo:
   (a) Pueden tener más de una solución.
   (b) Tienen siempre solución única.
   (c) Dan lugar a múltiples estimadores de $\hat{\boldsymbol{\beta}}$.
   (d) Forman un sistema de ecuaciones no lineal, que puede resolverse mediante técnicas iterativas.

19. ¿Cuál o cuáles de las siguientes son medidas de influencia?
   (a) Distancia de Cook, D_i.
   (b) DFFITS$_i$.
   (c) C_p de Mallows.
   (d) Ninguna de ellas.

20. Al objeto de diagnosticar la multicolinealidad aproximada y los regresores que pueden causarla, computamos los factores de incremento de varianza (VIF). Adicionalmente, practicamos las regresiones de cada uno de los regresores sobre todos los restantes. Para el regresor i-ésimo, obtenemos que la $R^2(i)$ de dicha regresión es 0.90. Ello quiere decir que su VIF será:
   (a) 0.10
   (b) 10
   (c) 1 / 0.90
   (d) 0.90 /(1 - 0.90), o sea 9

21. En un modelo de regresión, el método de estimación en componentes principales utilizando tantas componentes principales como regresores hay, daría:
   (a) El mismo resultado que el método de estimación *ridge* con $k \to \infty$.
   (b) El mismo resultado que el método de estimación *ridge* con $k = 0$
   (c) El mismo resultado que el método de estimación MCO.
   (d) Todo falso.

22. ¿Cuál o cuáles de los siguientes procedimientos para seleccionar la especificación de un modelo garantizan encontrar un óptimo global (es decir, que ningún otro modelo en la clase considerada mejora al al seleccionado de acuerdo con el criterio adoptado)?
   (a) *Stepwise* hacia adelante.
   (b) *All-subsets* o regresión sobre todos los subconjuntos.
   (c) *Stepwise* hacia atrás.
   (d) Validación cruzada.

23. En un diseño no completo como el de cuadrado latino se pierde:
   (a) La condición de ortogonalidad entre las sumas de cuadrados atribuibles a cada efecto.
   (b) La homoscedasticidad de las perturbaciones, que pasan a ser heteroscedásticas por construcción.
   (c) La posibilidad de estimar todas o algunas de las interacciones.
   (d) La linealidad de la especificación.

24. Un modelo ANOVA está jerárquicamente bien estructurado. . .

    (a) Cuando se ha consultado al catedrático de la asignatura sobre la especificación, y este ha dado su conformidad.

    (b) Cuando incluye columna de "unos".

    (c) Cuando verifica restricciones de marginalidad: la ausencia de un efecto implica la de todas las interacciones en las que dicho efecto tomaría parte.

    (d) Cuando el diseño es ortogonal y equilibrado.

25. Para contrastar en un modelo de regresión la hipótesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, adoptamos el siguiente criterio: formar los cuatro estadísticos t (cocientes $\hat{\beta}_i/\hat{\sigma}_{\hat{\beta}_i}$) y rechazar H_0 cuando cualquiera de ellos excede en valor absoluto de $t_{N-p;\alpha}$ (cuantil dejando una cola de tamaño α en la distribución t_{N-p}). Tal modo de operar nos dará un nivel de significación:

    (a) Exactamente igual a α.

    (b) Mayor que α.

    (c) Menor que α.

    (d) Depende de los valores de los demás parámetros en el modelo.

26. En un problema de estimación condicionada,

    (a) SSE_h (la suma de cuadrados en el modelo estimado con la condición) nunca será menor que SSE del modelo sin la condición.

    (b) Las estimaciones de los parámetros no afectados por la condición no siempre serán insesgadas.

    (c) Los estimadores tendrán menor varianza que los estimadores MCO si, y solo si, la condición impuesta es "correcta" (verificada por los parámetros del modelo).

    (d) Todo falso.

27. ¿En cuál de las siguientes circunstancias estaría justificado emplear un modelo de regresión lineal sin columna de "unos" (y sin estimar β_0 por tanto)?

    (a) Cuando las unidades de medida de *todos* los regresores son idénticas.

    (b) Cuanto tanto regresores como regresando han sido centrados.

    (c) Cuando los regresores se expresan todos en las mismas unidades y el origen tomado es generalmente aceptado —como el origen de la era cristiana para medir tiempos o el punto de congelación del agua para medir temperaturas—.

    (d) Siempre, en toda circunstancia, hay que utilizar una columna de "unos" como regresor.

28. Decimos que una combinación lineal de los parámetros, $\vec{\boldsymbol{a}}'\hat{\boldsymbol{\beta}}$, es estimable cuando:

    (a) El vector $\vec{\boldsymbol{a}}$ está unívocamente definido.

    (b) Las ecuaciones normales tiene más de una solución.

    (c) Hay una combinación lineal de las $\vec{\boldsymbol{Y}}$ cuyo valor medio coincide con $\vec{\boldsymbol{a}}'\hat{\boldsymbol{\beta}}$.

    (d) Las ecuaciones normales tienen una única solución.

29. Los "residuos borrados" *(deleted residuals)* d_i pueden obtenerse a partir de los mínimo cuadráticos $\hat{\epsilon}_i$, los términos diagonales p_{ii} de la matriz de proyección y estimadores adecuados de la varianza de la perturbación haciendo uso de la fórmula siguiente:

    (a) $d_i = \hat{\epsilon}_i/\sqrt{\hat{\sigma}^2(1-p_{ii})}$

    (b) $d_i = \hat{\epsilon}_i/\sqrt{\hat{\sigma}^2_{(i)}(1-p_{ii})}$

    (c) $d_i = \hat{\epsilon}_i/(1-p_{ii})$

    (d) $d_i = \hat{\epsilon}_i/(\hat{\sigma}^2(1-p_{ii}))$

30. Si las unidades experimentales con que contamos (por ej., parcelas) no son homogéneas (por ejemplo, hay unas que sabemos más fértiles que otras), al especificar un modelo de Análisis de Varianza:

    (a) Prescindiremos por completo de este hecho.

    (b) Introduciremos las diferencias de fertilidad como factor explicativo, recogiendo su efecto en un grupo de parámetros $\{\alpha_i^A\}$.

    (c) Nos abstendremos de hacer Análisis de Varianza: la inhomogeneidad de las unidades experimentales es una violación de los supuestos incompatible con el análisis.

    (d) Todo falso.

Respuestas al examen de tipo $\boxed{\text{B}}$

Sección 1. Instrucciones

Salvo que se indique lo contrario, las preguntas bien contestadas valen un punto. Puede haber más de una respuesta correcta, y para obtener puntuación has de señalarlas todas. Preguntas que no estén bien contestadas puntúan -0.5 veces su valor.

Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay más de una persona que echa a perder una nota potencialmente buena por responder temeraria o atolondradamente.

¡Por favor, fíjate bien en todos los detalles!

Te ayudará proceder por exclusión de absurdos. Si una pregunta te parece ambigua, anota brevemente la razón al margen y no la contestes.

1. En un modelo de regresión lineal, la omisión de regresores que deberían formar parte de la especificación:

   (a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

   (b) Sesga, en general, los estimadores de los parámetros incluidos.

   (c) Puede no afectar al sesgo de los estimadores de los parámetros incluidos, si acontece que los regresores omitidos son ortogonales a los incluidos.

   (d) Disminuye, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo completo.

2. Sea un modelo de Análisis de Varianza

   $$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \epsilon_{ijk}.$$

   Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de parámetros libres es:

   (a) 18

   (b) 12

   (c) 16

   (d) 24

3. h es un subespacio del espacio vectorial M. ¿Cuáles de las siguientes igualdades son correctas?

   (a) $P_M = P_h$

   (b) $P_M P_h = P_h P_M$

   (c) $P_{M\cap h} = P_h$

   (d) $P_{M\cap h} = P_M - P_h$

4. Al ajustar una regresión, las observaciones i-ésima y j-ésima proporcionan los siguientes residuos externamente *studentizados* y borrados: $t_i = 3{,}3$, $t_j = 0{,}2$, $d_i = 3{,}4$ y $d_j = 3{,}9$. ¿Cuál de las dos observaciones señalarías como más sospechosa de ser influyente?

   (a) La i-ésima.

   (b) La j-ésima.

   (c) Ninguna.

   (d) Las dos son manifiestamente influyentes.

5. Una matriz de proyección es *siempre*:

   (a) Simétrica

   (b) Idempotente

   (c) De rango completo

   (d) Ortogonal

6. ¿Cuales de los siguientes tipos de residuos tienen varianzas iguales?

   (a) Los internamente *studentizados*, r_i.

   (b) Los externamente *studentizados*, t_i.

   (c) Los borrados. d_i.

   (d) Los brutos, $\hat{\epsilon}_i$ o MCO.

7. Al ajustar un modelo con 4 regresores (incluida la columna de "unos") a $N = 120$ observaciones, se obtiene $SSE = 345{,}3$. Se añaden otros tres regresores al modelo, con parámetros $\beta_4, \beta_5, \beta_6$. La nueva regresión proporciona ahora $SSE^* = 312{,}3$. El estadístico Q_h para contrastar la hipótesis $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$ toma el valor:

   (a) 2,35

   (b) 3,98

   (c) 2,853

   (d) 1,26

8. ¿Cuál o cuáles de los siguientes supuestos son necesarios para garantizar que la R^2 está entre 0 y 1?

(a) Normalidad en las perturbaciones.

(b) Modelo "correcto".

(c) Modelo incluye una columna de "unos" (u otros regresores que combinados linealmente generan $\vec{1}$).

(d) Modelo con efectos simples (no interacciones de ningún género).

9. Sea un modelo de Análisis de Varianza

$$y_{ijk} = \alpha + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB} + \epsilon_{ijk}.$$

Supongamos que hay 3 niveles de A y 4 de B, y se replica cada combinación de niveles de A y B dos veces. El número de grados de libertad de la suma de cuadrados de los residuos, SSE, es:

(a) 18

(b) 12

(c) 16

(d) 24

10. Cuando contrastamos la hipótesis de nulidad de todos los parámetros de una regresión (salvo β_0), el estadístico de contraste Q_h toma una forma particularmente simple. Puede escribirse en función de N, p y...

(a) La C_p de Mallows.

(b) R^2

(c) El criterio AIC.

(d) SSE únicamente.

11. ¿Cuando es inadecuada la inclusión de una columna de "unos" entre los regresores?

(a) Cuando el número de grados de libertad es relativamente elevado.

(b) Cuando las escalas de los regresores sean muy similares.

(c) Cuando tanto regresando como regresores se toman en desviaciones respecto a su media. En tal caso, β_0 sería siempre 0.

(d) Cuando las escalas de los regresores sean muy diferentes.

12. En un modelo de regresión lineal, la inclusión de regresores irrelevantes:

(a) Afecta (sesgándola) a la estimación de β_0, pero no a la de los demás parámetros presentes en el modelo.

(b) Sesga, en general, los estimadores de los parámetros incluidos.

(c) Afecta a las varianzas de los estimadores de los parámetros "correctos", pero sólo si acontece que los regresores irrelevantes son ortogonales a los "correctos".

(d) Aumenta, en general, las varianzas de los estimadores, respecto a las que hubieran sido en el caso de ajustar el modelo correcto.

13. Al estimar un modelo lineal mediante regresión *ridge*, es decir, $\hat{\boldsymbol{\beta}}_{(\boldsymbol{k})} = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\vec{\boldsymbol{y}}\dots$

(a) Sólamente es posible mejorar el ECM (error cuadrático medio) en los casos en que la multicolinealidad es perfecta.

(b) Un incremento de k acarrea un incremento del sesgo del estimador.

(c) Un incremento de k acarrea un incremento de la varianza del estimador.

(d) Un incremento de k acarrea una disminución de la varianza del estimador.

14. Al estimar un modelo de regresión logística, consideramos un modelo M_1 con cinco parámetros y otro M_2 con los cinco de M_1 y otros dos, que denotaremos β_5, β_6. Sean $L(M_1)$ y $L(M_2)$ los valores máximos de la verosimilitud al ajustar ambos modelos. Bajo la hipótesis de que $\beta_5 = \beta_6 = 0$,

$$-2\ln\left(\frac{L(M_1)}{L(M_2)}\right)$$

se distribuye como:

(a) χ_2^2

(b) χ_7^2

(c) $\mathcal{F}_{2,7}$

(d) $\mathcal{F}_{2,5}$

15. Su $\hat{\boldsymbol{\beta}}_*$ es un estimador lineal del vector de parámetros en un modelo de regresión, necesariamente:

    (a) Su matriz de covarianzas tiene en la diagonal principal elementos no menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$; lo garantiza el teorema de Gauss-Markov.

    (b) Puede tener en la diagonal principal de su matriz de covarianzas términos menores que los de $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

    (c) Necesariamente ha de coincidir con el estimador MCO.

    (d) Necesariamente ha de ser insesgado.

16. ¿Cuál de los siguientes criterios penaliza más la inclusión de regresores adicionales en un modelo?

    (a) Distancia de Cook, D_i.

    (b) DFFITS$_i$.

    (c) C_p de Mallows.

    (d) $\overline{R}^2$.

17. ¿En qué condiciones puede $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ (proyección de $\vec{\boldsymbol{y}}$ sobre el subespacio que generan las columnas de $\boldsymbol{X}$) no ser única?

    (a) Cuando haya acusada multicolinealidad entre las columnas de $\boldsymbol{X}$.

    (b) Siempre es única.

    (c) No tiene por qué ser única en ningún caso.

    (d) Todo falso.

18. Las ecuaciones normales cuando la matriz $\boldsymbol{X}$ es de rango completo:

    (a) Pueden tener más de una solución.

    (b) Tienen siempre solución única.

    (c) Dan lugar a múltiples estimadores de $\hat{\boldsymbol{\beta}}$.

    (d) Forman un sistema de ecuaciones no lineal, que puede resolverse mediante técnicas iterativas.

19. ¿Cuál o cuáles de las siguientes son medidas de influencia?

    (a) Distancia de Cook, D_i.

    (b) DFFITS$_i$.

    (c) C_p de Mallows.

    (d) Ninguna de ellas.

20. Al objeto de diagnosticar la multicolinealidad aproximada y los regresores que pueden causarla, computamos los factores de incremento de varianza (VIF). Adicionalmente, practicamos las regresiones de cada uno de los regresores sobre todos los restantes. Para el regresor i-ésimo, obtenemos que la $R^2(i)$ de dicha regresión es 0.90. Ello quiere decir que su VIF será:

    (a) 0.10

    (b) 10

    (c) 1 / 0.90

    (d) 0.90 /(1 - 0.90), o sea 9

21. En un modelo de regresión, el método de estimación en componentes principales utilizando tantas componentes principales como regresores hay, daría:

    (a) El mismo resultado que el método de estimación *ridge* con $k \to \infty$.

    (b) El mismo resultado que el método de estimación *ridge* con $k = 0$

    (c) El mismo resultado que el método de estimación MCO.

    (d) Todo falso.

22. ¿Cuál o cuáles de los siguientes procedimientos para seleccionar la especificación de un modelo garantizan encontrar un óptimo global (es decir, que ningún otro modelo en la clase considerada mejora al al seleccionado de acuerdo con el criterio adoptado)?

    (a) *Stepwise* hacia adelante.

    (b) *All-subsets* o regresión sobre todos los subconjuntos.

    (c) *Stepwise* hacia atrás.

    (d) Validación cruzada.

23. En un diseño no completo como el de cuadrado latino se pierde:

    (a) La condición de ortogonalidad entre las sumas de cuadrados atribuibles a cada efecto.

    (b) La homoscedasticidad de las perturbaciones, que pasan a ser heteroscedásticas por construcción.

    (c) La posibilidad de estimar todas o algunas de las interacciones.

    (d) La linealidad de la especificación.

24. Un modelo ANOVA está jerárquicamente bien estructurado. . .

    (a) Cuando se ha consultado al catedrático de la asignatura sobre la especificación, y este ha dado su conformidad.

    (b) Cuando incluye columna de "unos".

    (c) Cuando verifica restricciones de marginalidad: la ausencia de un efecto implica la de todas las interacciones en las que dicho efecto tomaría parte.

    (d) Cuando el diseño es ortogonal y equilibrado.

25. Para contrastar en un modelo de regresión la hipótesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, adoptamos el siguiente criterio: formar los cuatro estadísticos t (cocientes $\hat{\beta}_i/\hat{\sigma}_{\hat{\beta}_i}$) y rechazar H_0 cuando cualquiera de ellos excede en valor absoluto de $t_{N-p;\alpha}$ (cuantil dejando una cola de tamaño α en la distribución t_{N-p}). Tal modo de operar nos dará un nivel de significación:

    (a) Exactamente igual a α.

    (b) Mayor que α.

    (c) Menor que α.

    (d) Depende de los valores de los demás parámetros en el modelo.

26. En un problema de estimación condicionada,

    (a) SSE_h (la suma de cuadrados en el modelo estimado con la condición) nunca será menor que SSE del modelo sin la condición.

    (b) Las estimaciones de los parámetros no afectados por la condición no siempre serán insesgadas.

    (c) Los estimadores tendrán menor varianza que los estimadores MCO si, y solo si, la condición impuesta es "correcta" (verificada por los parámetros del modelo).

    (d) Todo falso.

27. ¿En cuál de las siguientes circunstancias estaría justificado emplear un modelo de regresión lineal sin columna de "unos" (y sin estimar β_0 por tanto)?

    (a) Cuando las unidades de medida de *todos* los regresores son idénticas.

    (b) Cuanto tanto regresores como regresando han sido centrados.

    (c) Cuando los regresores se expresan todos en las mismas unidades y el origen tomado es generalmente aceptado —como el origen de la era cristiana para medir tiempos o el punto de congelación del agua para medir temperaturas—.

    (d) Siempre, en toda circunstancia, hay que utilizar una columna de "unos" como regresor.

28. Decimos que una combinación lineal de los parámetros, $\vec{a}'\hat{\beta}$, es estimable cuando:

    (a) El vector $\vec{a}$ está unívocamente definido.

    (b) Las ecuaciones normales tiene más de una solución.

    (c) Hay una combinación lineal de las $\vec{Y}$ cuyo valor medio coincide con $\vec{a}'\hat{\beta}$.

    (d) Las ecuaciones normales tienen una única solución.

29. Los "residuos borrados" *(deleted residuals)* d_i pueden obtenerse a partir de los mínimo cuadráticos $\hat{\epsilon}_i$, los términos diagonales p_{ii} de la matriz de proyección y estimadores adecuados de la varianza de la perturbación haciendo uso de la fórmula siguiente:

    (a) $d_i = \hat{\epsilon}_i/\sqrt{\hat{\sigma}^2(1-p_{ii})}$

    (b) $d_i = \hat{\epsilon}_i/\sqrt{\hat{\sigma}^2_{(i)}(1-p_{ii})}$

    (c) $d_i = \hat{\epsilon}_i/(1-p_{ii})$

    (d) $d_i = \hat{\epsilon}_i/(\hat{\sigma}^2(1-p_{ii}))$

30. Si las unidades experimentales con que contamos (por ej., parcelas) no son homogéneas (por ejemplo, hay unas que sabemos más fértiles que otras), al especificar un modelo de Análisis de Varianza:

    (a) Prescindiremos por completo de este hecho.

    (b) Introduciremos las diferencias de fertilidad como factor explicativo, recogiendo su efecto en un grupo de parámetros $\{\alpha_i^A\}$.

    (c) Nos abstendremos de hacer Análisis de Varianza: la inhomogeneidad de las unidades experimentales es una violación de los supuestos incompatible con el análisis.

    (d) Todo falso.