21 de Enero de 2000

Responde a las siguientes preguntas sobre papel ordinario, de forma breve y concisa. Al entregar tu exámen, has de entregar también la Tarea 10, que no fue posible finalizar en periodo ordinario de clases.

Estadística III (P33)

Exam, Tipo: A

Name: ______________________

Student Number: ______________________

TA: ______________________

Date: ______________________

Responde a las cuestiones breves haciendo un círculo sobre las respuestas correctas, directamente sobre este formulario de examen (que has de entregar). Puede haber más de una respuesta correcta, en cuyo caso has de señalarlas todas para obtener puntuación. Respuestas incorrectas tienen una penalización del 20 % de la puntuación que hubieran obtenido de ser acertadas.

1. Un modelo ANOVA con dos tratamientos cruzados, con tres y tres niveles respectivamente, *con interacción*, posee:

   (a) Catorce parámetros libres.

   (b) Todo falso.

   (c) Ocho parámetros libres.

   (d) Seis parámetros libres.

   (e) Diez parámetros libres.

2. La desigualdad de Bonferroni de primer orden estudiada en el curso proporciona:

   (a) Una cota superior del nivel de significación conjunto cuando contrastamos varias hipótesis, cada una de las cuáles con un estadístico diferente.

   (b) Una cota de la probabilidad de que la hipótesis alternativa sea falsa.

   (c) Una cota del p-value asociado a la varianza.

   (d) El p-value de $\hat{\beta}_0$.

3. ¿Cuál de las siguientes cosas NO tiene nada que ver con inferencia simultánea?

   (a) S de Scheffé.

   (b) Recorrido studentizado.

   (c) Validación cruzada.

   (d) Desigualdad de Bonferroni.

   (e) Máximo de k variables t de Student.

4. Al hacer regresión en componentes principales introducimos un sesgo pero reducimos la varianza de los estimadores.

   (a) Cierto

   (b) Falso

5. Al hacer regresión en componentes principales obtenemos no todo $\hat{\beta}$, sino sólo estimaciones para algunos β's seleccionados: es así como sorteamos el problema de la multicolinealidad.

   (a) Cierto.

   (b) Falso

6. La distribución del estadístico Q_h empleado en el contraste de hipótesis lineales es, bajo ciertos supuestos, una F de Snedecor cuando la hipótesis nula es cierta. Estos supuestos incluyen:

   (a) Regresores ortogonales.

   (b) Presencia de una columna de "unos".

   (c) Homoscedasticidad de las perturbaciones.

   (d) Normalidad de las perturbaciones.

7. Si reescalas los regresores, ¿cuáles de los siguientes estadísticos permanecen invariantes?

   (a) SSE

   (b) SST

   (c) SSR

   (d) R^2

   (e) Los $\hat{\beta}$'s.

8. Un residuo borrado d_i "grande" es síntoma inequívoco de que la observación i-ésima es muy influyente.

   (a) Cierto. Un residuo borrado grande indica que la estimación de los parámetros se altera sustancialmente en ausencia de esa observación.

   (b) No necesariamente: podría tratarse sólamente de una observación rara.

9. En ocasiones, las unidades experimentales en un análisis de varianza son de diferentes tipos y tienen un efecto previsible y conocido sobre la variable respuesta. En este caso, lo idóneo sería:
   - (a) Aleatorizar el o los tratamientos de interés, asignándolos al azar a los diferentes tipos.
   - (b) Considerar los diferentes tipos como niveles de un tratamiento adicional: aunque no nos interese su efecto por sí mismo, actuar así reduce SSE y mejora la precisión de las estimaciones.
   - (c) Asignar los niveles de los tratamientos de interés de modo que cada nivel se asigne a un solo tipo de unidad experimental.
   - (d) Todo falso.

10. El estimador *ridge*,
    - (a) Logra, para algún k, una reducción del ECM en relación al estimador MCO sólo en los casos en que hay acusada multicolinealidad.
    - (b) Todo falso.
    - (c) Logra, para algún k, una reducción del ECM en relación al estimador MCO en todos los casos.
    - (d) Logra, para todo $k > 0$, una reducción de las varianzas de los β's respecto a la obtenible mediante MCO.
    - (e) Da lugar a estimaciones sesgadas de los β's.

11. La inclusión de un regresor irrelevante en un modelo:
    - (a) Disminuye los grados de libertad, y hace, en general, más imprecisa la estimación.
    - (b) Sesga el estimador de la varianza de la perturbación.
    - (c) Sesga los estimadores de los parámetros correspondientes a regresores relevantes.
    - (d) Hace la matriz de diseño de rango deficiente.

12. Para contrastar la hipótesis $H_0 : \beta_1 = \ldots = \beta_{p-1}$ en un modelo de regresión lineal con normalidad, conocido el tamaño muestral N y número de regresores p, basta saber:
    - (a) R^2.
    - (b) SSR.
    - (c) SST.
    - (d) Todos los $\hat{\beta}$'s.
    - (e) La varianza estimada de los residuos.

13. Al hacer regresión en componentes principales
    - (a) Todo falso.
    - (b) Obtenemos estimaciones de los β's con mayor varianza.
    - (c) Obtenemos estimaciones sesgadas de los β's.

14. Un modelo ANOVA con dos tratamientos cruzados, con tres y tres niveles respectivamente, sin interacción, posee:
    - (a) Cuatro parámetros libres.
    - (b) Siete parámetros libres.
    - (c) Cinco parámetros libres.
    - (d) Ocho parámetros libres.
    - (e) Nueve parámetros libres.

15. En un modelo de regresión lineal con columna de "unos", ¿cuáles de los siguientes estadísticos te bastarían para contrastar (bajo los supuestos habituales más normalidad) la hipótesis $H_0 : \beta_1 = \ldots = \beta_{p-1} = 0$?
    - (a) SSE y SST.
    - (b) R^2.
    - (c) $\hat{\beta}_0$.
    - (d) Nada de lo anterior bastaría.

16. ¿Cuál de las siguientes medidas de influencia es global (es decir, no desglosa la influencia sobre diferentes parámetros)?
    - (a) SIC
    - (b) Distancia de Cook.
    - (c) EIC
    - (d) Todo falso.

17. La imposición de restricciones lineales sobre los parámetros *siempre* sesga en alguna medida las estimaciones; pero puede compensar en ocasiones por la reducción de varianza que se logra.

    (a) Cierto.

    (b) Falso.

18. Una vez estimado un modelo lineal, podemos querer emplearlo en predicción. ¿De que depende *la amplitud* del intervalo de confianza para la predicción $\hat{y}_i$ asociada al vector de regresores $\vec{x}_i$?

    (a) De $\hat{\sigma}^2$.

    (b) De los valores de los $\hat{\beta}$'s.

    (c) De $\vec{x}_i$.

    (d) Nada de lo anterior.

19. Considera dos subespacios vectoriales anidados, $h \subseteq M$. Sean P_m y P_h las matrices asociadas a operadores de proyección sobre cada uno de ellos. Se verifica:

    (a) $P_M P_h = P_h P_M$.

    (b) $P_h = P_M P_h$.

    (c) $P_h = P_h P_M$.

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

20. Considera un modelo explicativo del nivel de colesterol en la sangre, en función de regresores como ingesta de diferentes alimentos, etc. Al ajustarlo a un colectivo de personas, mitad hombres y mitad mujeres, obtienes una suma de cuadrados de los residuos SSE. Al ajustarlo por separado a hombres y mujeres obtienes sendas sumas de cuadrados de los residuos SSE$_h$ y SSE$_m$. Acontece, además, que SSE$_h$ $\gg$ SSE$_m$ y SSE $\gg$ SSE$_h$+ SSE$_m$. ¿Cuáles de las siguientes conclusiones son congruentes (no necesariamente ciertas, pero no en contradicción) con estos resultados?

    (a) Es inadecuado ajustar un único modelo (mismos regresores y misma estimación de los parámetros) a hombres y mujeres.

    (b) Por algún motivo, la variabilidad del nivel de colesterol en los hombres (no ajustada por los regresores), parece mayor que en las mujeres.

    (c) Las mujeres tiene claramente un mayor nivel de colesterol que los hombres.

    (d) Es pertinente un ajuste diferenciado para hombres y mujeres: quizá el modelo es el mismo, y los parámetros toman valores diferentes, o quizá es sólo σ^2_ϵ lo que difiere de unos a otras.

    (e) Todo falso.

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

Responde a las cuestiones breves haciendo un círculo sobre las respuestas correctas, directamente sobre este formulario de examen (que has de entregar). Puede haber más de una respuesta correcta, en cuyo caso has de señalarlas todas para obtener puntuación. Respuestas incorrectas tienen una penalización del 20 % de la puntuación que hubieran obtenido de ser acertadas.

1. Un modelo ANOVA con dos tratamientos cruzados, con tres y tres niveles respectivamente, *con interacción*, posee:

   (a) Catorce parámetros libres.
   $\boxed{\text{(b)}}$ Todo falso.
   (c) Ocho parámetros libres.
   (d) Seis parámetros libres.
   (e) Diez parámetros libres.

2. La desigualdad de Bonferroni de primer orden estudiada en el curso proporciona:

   $\boxed{\text{(a)}}$ Una cota superior del nivel de significación conjunto cuando contrastamos varias hipótesis, cada una de las cuáles con un estadístico diferente.
   (b) Una cota de la probabilidad de que la hipótesis alternativa sea falsa.
   (c) Una cota del p-value asociado a la varianza.
   (d) El p-value de $\hat{\beta}_0$.

3. ¿Cuál de las siguientes cosas NO tiene nada que ver con inferencia simultánea?

   (a) S de Scheffé.
   (b) Recorrido studentizado.
   $\boxed{\text{(c)}}$ Validación cruzada.
   (d) Desigualdad de Bonferroni.
   (e) Máximo de k variables t de Student.

4. Al hacer regresión en componentes principales introducimos un sesgo pero reducimos la varianza de los estimadores.

   $\boxed{\text{(a)}}$ Cierto
   (b) Falso

5. Al hacer regresión en componentes principales obtenemos no todo $\hat{\beta}$, sino sólo estimaciones para algunos β's seleccionados: es así como sorteamos el problema de la multicolinealidad.

   (a) Cierto.
   $\boxed{\text{(b)}}$ Falso

6. La distribución del estadístico Q_h empleado en el contraste de hipótesis lineales es, bajo ciertos supuestos, una F de Snedecor cuando la hipótesis nula es cierta. Estos supuestos incluyen:

   (a) Regresores ortogonales.
   (b) Presencia de una columna de "unos".
   $\boxed{\text{(c)}}$ Homoscedasticidad de las perturbaciones.
   $\boxed{\text{(d)}}$ Normalidad de las perturbaciones.

7. Si reescalas los regresores, ¿cuáles de los siguientes estadísticos permanecen invariantes?

   $\boxed{\text{(a)}}$ SSE
   $\boxed{\text{(b)}}$ SST
   $\boxed{\text{(c)}}$ SSR
   $\boxed{\text{(d)}}$ R^2
   (e) Los $\hat{\beta}$'s.

8. Un residuo borrado d_i "grande" es síntoma inequívoco de que la observación i-ésima es muy influyente.

   (a) Cierto. Un residuo borrado grande indica que la estimación de los parámetros se altera sustancialmente en ausencia de esa observación.
   $\boxed{\text{(b)}}$ No necesariamente: podría tratarse sólamente de una observación rara.

9. En ocasiones, las unidades experimentales en un análisis de varianza son de diferentes tipos y tienen un efecto previsible y conocido sobre la variable respuesta. En este caso, lo idóneo sería:

   (a) Aleatorizar el o los tratamientos de interés, asignándolos al azar a los diferentes tipos.
   $\boxed{\text{(b)}}$ Considerar los diferentes tipos como niveles de un tratamiento adicional: aunque no nos interese su efecto por sí mismo, actuar así reduce SSE y mejora la precisión de las estimaciones.
   (c) Asignar los niveles de los tratamientos de interés de modo que cada nivel se asigne a un solo tipo de unidad experimental.
   (d) Todo falso.

10. El estimador *ridge*,

    (a) Logra, para algún k, una reducción del ECM en relación al estimador MCO sólo en los casos en que hay acusada multicolinealidad.

    (b) Todo falso.

    $\boxed{\text{(c)}}$ Logra, para algún k, una reducción del ECM en relación al estimador MCO en todos los casos.

    $\boxed{\text{(d)}}$ Logra, para todo $k > 0$, una reducción de las varianzas de los β's respecto a la obtenible mediante MCO.

    $\boxed{\text{(e)}}$ Da lugar a estimaciones sesgadas de los β's.

11. La inclusión de un regresor irrelevante en un modelo:

    $\boxed{\text{(a)}}$ Disminuye los grados de libertad, y hace, en general, más imprecisa la estimación.

    (b) Sesga el estimador de la varianza de la perturbación.

    (c) Sesga los estimadores de los parámetros correspondientes a regresores relevantes.

    (d) Hace la matriz de diseño de rango deficiente.

12. Para contrastar la hipótesis $H_0 : \beta_1 = \ldots = \beta_{p-1}$ en un modelo de regresión lineal con normalidad, conocido el tamaño muestral N y número de regresores p, basta saber:

    $\boxed{\text{(a)}}$ R^2.

    (b) SSR.

    (c) SST.

    (d) Todos los $\hat{\beta}$'s.

    (e) La varianza estimada de los residuos.

13. Al hacer regresión en componentes principales

    (a) Todo falso.

    (b) Obtenemos estimaciones de los β's con mayor varianza.

    $\boxed{\text{(c)}}$ Obtenemos estimaciones sesgadas de los β's.

14. Un modelo ANOVA con dos tratamientos cruzados, con tres y tres niveles respectivamente, sin interacción, posee:

    (a) Cuatro parámetros libres.

    (b) Siete parámetros libres.

    $\boxed{\text{(c)}}$ Cinco parámetros libres.

    (d) Ocho parámetros libres.

    (e) Nueve parámetros libres.

15. En un modelo de regresión lineal con columna de "unos", ¿cuáles de los siguientes estadísticos te bastarían para contrastar (bajo los supuestos habituales más normalidad) la hipótesis $H_0 : \beta_1 = \ldots = \beta_{p-1} = 0$?

    $\boxed{\text{(a)}}$ SSE y SST.

    $\boxed{\text{(b)}}$ R^2.

    (c) $\hat{\beta}_0$.

    (d) Nada de lo anterior bastaría.

16. ¿Cuál de las siguientes medidas de influencia es global (es decir, no desglosa la influencia sobre diferentes parámetros)?

    (a) SIC

    $\boxed{\text{(b)}}$ Distancia de Cook.

    (c) EIC

    (d) Todo falso.

17. La imposición de restricciones lineales sobre los parámetros *siempre* sesga en alguna medida las estimaciones; pero puede compensar en ocasiones por la reducción de varianza que se logra.

    (a) Cierto.

    $\boxed{\text{(b)}}$ Falso.

18. Una vez estimado un modelo lineal, podemos querer emplearlo en predicción. ¿De que depende *la amplitud* del intervalo de confianza para la predicción $\hat{y}_i$ asociada al vector de regresores $\vec{x}_i$?

    $\boxed{\text{(a)}}$ De $\hat{\sigma}^2$.

    (b) De los valores de los $\hat{\beta}$'s.

    $\boxed{\text{(c)}}$ De $\vec{x}_i$.

    (d) Nada de lo anterior.

19. Considera dos subespacios vectoriales anidados, $h \subseteq M$. Sean P_m y P_h las matrices asociadas a operadores de proyección sobre cada uno de ellos. Se verifica:

    $\boxed{\text{(a)}}$ $P_M P_h = P_h P_M$.

    $\boxed{\text{(b)}}$ $P_h = P_M P_h$.

    $\boxed{\text{(c)}}$ $P_h = P_h P_M$.

    $\boxed{\text{(d)}}$ $P_{M \cap h^\perp} = P_M - P_h$.

20. Considera un modelo explicativo del nivel de colesterol en la sangre, en función de regresores como ingesta de diferentes alimentos, etc. Al ajustarlo a un colectivo de personas, mitad hombres y mitad mujeres, obtienes una suma de cuadrados de los residuos SSE. Al ajustarlo por separado a hombres y mujeres obtienes sendas sumas de cuadrados de los residuos SSE_h y SSE_m. Acontece, además, que $\text{SSE}_h \gg \text{SSE}_m$ y $\text{SSE} \gg \text{SSE}_h + \text{SSE}_m$. ¿Cuáles de las siguientes conclusiones son congruentes (no necesariamente ciertas, pero no en contradicción) con estos resultados?

    $\boxed{\text{(a)}}$ Es inadecuado ajustar un único modelo (mismos regresores y misma estimación de los parámetros) a hombres y mujeres.

    $\boxed{\text{(b)}}$ Por algún motivo, la variabilidad del nivel de colesterol en los hombres (no ajustada por los regresores), parece mayor que en las mujeres.

    (c) Las mujeres tiene claramente un mayor nivel de colesterol que los hombres.

    $\boxed{\text{(d)}}$ Es pertinente un ajuste diferenciado para hombres y mujeres: quizá el modelo es el mismo, y los parámetros toman valores diferentes, o quizá es sólo σ^2_ϵ lo que difiere de unos a otras.

    (e) Todo falso.