Estadística III

Examen Final

11 de Febrero de 1998

INSTRUCCIONES Y CONVENCIONES

1. El examen consta de preguntas con respuesta de elección múltiple, que se señala en la hoja de codificación.
2. La totalidad de las cuestiones que siguen se refiere al modelo de regresión lineal y análisis de varianza.
3. Cuando hablemos de “los estimadores mínimo-cuadráticos” nos estamos refiriendo a los estimadores del vector de parámetros en el modelo,

   $$\vec{Y} = X\vec{\beta} + \vec{\epsilon}$$

   dados por

   $$\hat{\beta} = (X'X)^{-1}X'\vec{y}$$

   Siempre que no se indique lo contrario, supondremos las condiciones habituales, pero no normalidad.
4. El examen consta de 7 hojas numeradas; cerciórate de tenerlas todas.
5. Respuestas incorrectas puntúan negativo, el 20 % de lo que lo hubieran hecho positivamente de ser respondidas correctamente; no seas temerario(a).
6. No hay cuestiones deliberadamente ambiguas, pero si puede haber interpretaciones no previstas por el examinador; si encuentras alguna pregunta que no sea cierta ni falsa, responde “Ni cierto ni falso”, y explica brevemente al margen la razón.
7. Para contestar, debes señalar la(s) respuesta(s) correcta(s) (puede haber más de una, o ninguna) en la hoja de codificación.
8. Aprobar requiere 35 puntos. *Si los alcanzas* tu nota de curso ponderará tanto la nota del examen como las de las tareas periódicas que has entregado y te han sido corregidas. El tiempo previsto es de 1 hora 45’.

1. (1 punto) Los estimadores mínimo cuadráticos de los parámetros $\vec{\beta}$ son insesgados si, y sólo si, las perturbaciones $\vec{\epsilon}$ son normales.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

2. (1 punto) La estimación de la varianza de la perturbación $\hat{\sigma}^2 = \frac{SSE}{N-p}$ es insesgada si, y sólo si, las perturbaciones $\vec{\epsilon}$ son normales.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

3. (1 punto) La estimación de la varianza de la perturbación $\hat{\sigma}^2 = \frac{SSE}{N-p}$ deja de ser insesgada si hay multicolinealidad acusada entre los regresores. Hay entonces que recurrir a procedimientos de estimación distintos del MCO.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

4. (1 punto) La traza de la matriz de proyección es $N - p$.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

5. (1 punto) La matriz de proyección *siempre* es de rango completo.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

6. (1 punto) La matriz de proyección puede ser de rango deficiente sólo en el caso de que haya multicolinealidad exacta entre los regresores.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

7. (1 punto) *Siempre* se verifica en un problema de regresión que $SST = SSR + SSE$.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

8. (1 punto) ¿Cuál o cuáles de estas propiedades debe necesariamente tener una matriz de proyección?

   a) Simétrica
   b) Idempotente
   c) Iguales elementos a lo largo de la diagonal principal
   d) Diagonal
   e) Escalar (diagonal con iguales elementos en la diagonal).

9. (1 punto) En un modelo de regresión con matriz de diseño de rango deficiente es imposible (en ausencia de restricciones sobre los parámetros) estimar la varianza de la perturbación.

   (a) Cierto (b) Falso (c) Ni cierto ni falso

10. (1 punto) Cuando hacemos estimación condicionada de los parámetros, los estimadores *siempre* tienen menor varianza.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

11. (1 punto) Cuando hacemos estimación condicionada de los parámetros, los estimadores *siempre* son sesgados.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

12. (1 punto) Cuando hacemos estimación condicionada de los parámetros, la suma de cuadrados de los residuos *siempre* es igual o mayor que cuando hacemos estimación incondicionada (del mismo modelo, por supuesto).

    (a) Cierto (b) Falso (c) Ni cierto ni falso

13. (2 puntos) ¿Cuáles de los siguientes supuestos son cruciales para demostrar la insesgadez de $\hat{\beta}$ como estimador de $\vec{\beta}$?

    a) Todas las variables incluídas como regresores intervienen en la generación de $\vec{y}$.

    b) Todas las variables que intervienen en la generación de $\vec{y}$ han sido incluídas como regresores.

    c) Los residuos son homoscedásticos.

    d) Las perturbaciones son homoscedásticas.

14. (2 puntos) Se requiere igual varianza en las perturbaciones para demostrar el teorema de Gauss-Markov.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

15. (1 punto) Al estimar por MCO un modelo de regresión lineal con normalidad, son independientes:

    a) Los $\hat{\beta}_i$ entre sí.

    b) Los $\hat{\epsilon}_i$ entre sí.

    c) $\hat{\beta}$ y $\hat{\sigma}^2$.

    d) $\hat{\beta}$ y cada uno de los $\hat{\epsilon}_i$.

16. (1 punto) Un modelo de regresión lineal con homoscedasticidad proporcionara siempre predicciones por intervalo de amplitud fija (supuesto que empleamos un nivel de confianza $(1 - \alpha)$ igual para todas ellas).

    (a) Cierto (b) Falso (c) Ni cierto ni falso

17. (1 punto) Un modelo estimado con regresores irrelevantes daría en general estimaciones sesgadas de los parámetros correspondientes a β's que sí deben ser incluidos en el modelo.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

18. (1 punto) Un modelo estimado con regresores irrelevantes daría en general estimaciones sesgadas de σ^2.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

19. (1 punto) Un modelo "escaso" (omitiendo regresores relevantes) daría en general estimaciones sesgadas de σ^2.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

20. (1 punto) Las consecuencia de estimar un modelo escaso pueden deducirse fácilmente si pensamos en lo que ocurre al hacer estimación condicionada con restricciones de nulidad sobre algunos parámetros.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

21. (2 puntos) El vector de residuos al ajustar un modelo de regresión lineal sólo sigue una distribución multivariante con matriz de covarianzas diagonal sólo cuando $\vec{\epsilon} \sim N(\vec{0}, \sigma^2 I)$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

22. (2 puntos) Dos residuos en particular, $\hat{\epsilon}_i$ y $\hat{\epsilon}_j$, pueden ser incorrelados si se verifican los supuestos habituales. Para comprobar su incorrelación se requiere, exclusivamente, conocer la matriz $X(X'X)^{-1}X'$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

23. (1 punto) Se demostró en clase que si P_M es la matriz de proyección sobre el subespacio M, puede siempre expresarse como el producto TT', siendo T una matriz cuyas columnas forman una base ortonormalizada de M. Consecuentemente, P_M es necesariamente una matriz unidad.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

24. (1 punto) Es posible, en el caso en que hay normalidad, construir intervalos de confianza $(1-\alpha)$ (para un α arbitrario) de cada residuo $\hat{\epsilon}_i$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

25. (2 puntos) Sea cual fuere la respuesta a la pregunta anterior, *supongamos* que fuera realmente posible construir tales intervalos de confianza. La probabilidad de que algún residuo se saliera de su respectivo intervalo sería:

    a) Precisamente α.
    b) Mayor que α, y tanto mayor cuantos más residuos hubiera.
    c) Menor que α, y tanto menor cuantos más residuos hubiera.

26. (1 punto) En un modelo de regresión, $\overline{R}^2$ nunca decrecerá al introducir un regresor adicional.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

27. (1 punto) El estimador *ridge* tiene la potencialidad de mejorar en todo caso el error cuadrático medio (ECM) de la estimación, aunque su utilización suele limitarse a casos en que hay acusada multicolinealidad.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

28. (1 punto) Sea $\hat{\beta}^{(k)} = ((X'X) + kI)^{-1}X'\vec{Y}$. La varianza de $\hat{\beta}^{(k)}$ (o traza de la matriz de covarianzas si $\hat{\beta}$ es un vector) para k adecuado, mejora la del estimador MCO.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

29. (1 punto) En el caso del modelo de Análisis de Varianza equilibrado con un único factor,
$$\vec{Y}_{i,k} = \mu + \sum_{i=1}^{I} \alpha_i^A + \epsilon_{i,k} \qquad (i = 1, \ldots, I)$$
la forma cuadrática $||\vec{Y}_{i.} - \vec{Y}_{..}||^2$ se distribuye como una χ^2_{I-1} cuando se verifican los supuestos habituales más normalidad, y la hipótesis $H_0\colon \alpha_i^A = 0$ es cierta.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

30. (1 punto) Lo anterior solo es cierto en el caso de que $\mu = 0$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

31. (2 puntos) Para que $\vec{c}\,'\hat{\beta}$ sea un estimador con reducida varianza, tiene que suceder que $\vec{c}$ sea aproximadamente colineal en R^p a un vector propio $\vec{v}_i$ de $X'X$ con $\lambda_i \approx 0$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

32. (1 punto) Para que $\vec{\beta}$ sea estimable en su totalidad, todos los valores propios de $X'X$ han de ser distintos de cero. Esto es lo mismo que exigir que $X'X$ sea de rango completo.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

33. (2 puntos) $X'X$ es de orden $p \times p$ y rango p siempre que X es de orden $N \times p$ ($N \geq p$) y las p columnas de X son linealmente independientes.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

34. (3 puntos) Para que en el modelo de Análisis de Varianza
$$\vec{Y}_{i,k} = \mu + \sum_{i=1}^{I} \alpha_i^A + \epsilon_{i,k} \qquad (i = 1, \ldots, I)$$
la restricción $\sum_{i=1}^{I} \alpha_i^A = 0$ identifique los parámetros —es decir, los haga estimables—, es preciso que el modelo sea equilibrado.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

35. (3 puntos) Para que en el modelo de Análisis de Varianza (con normalidad)
$$Y_{i,k} = \mu + \sum_{i=1}^{I} \alpha_i^A + \epsilon_{i,k} \qquad (i = 1, \ldots, I)$$

sean independientes $||\vec{Y}_{i.} - \vec{Y}_{..}||^2$ y $||\vec{Y}_{ik} - \vec{Y}_{i.}||^2$, es preciso que el diseño sea equilibrado.

(a) Cierto (b) Falso (c) Ni cierto ni falso

36. (2 puntos) El método de regresión en raíces latentes convierte la multicolinealidad en una ventaja, al aprovechar las combinaciones lineales casi exactas entre los regresores para la predicción. En realidad, cuando hacemos regresión en raíces latentes, no solo no nos molesta, sino que *deseamos* que haya combinaciones lineales casi exactas entre los regresores.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

37. (2 puntos) El método de regresión *ridge* proporciona estimadores $\hat{\beta}^{(k)}$ sesgados cuando $k = 0$.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

38. (2 puntos) Para contrastar la nulidad de todos los parámetros en una regresión, salvo el β_0 que afecta a la columna de "unos", al nivel de significación α, compararemos los respectivos t-ratios con $t_{N-p-1}^{\alpha/2}$ (p = número total de parámetros, incluyendo β_0). El contraste anterior requiere normalidad en las perturbaciones.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

39. (1 punto) La regresión en componentes principales está indicada cuando la matriz de diseño que nos proporcionan esta formada por columnas ortogonales, sin que sea posible tomar filas adicionales.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

40. (1 punto) Si podemos ampliar la matriz de diseño con filas adicionales, tendremos siempre interés en tomarlas de modo que $||\vec{x}_{N+1}||^2$ sea lo más reducido posible.

    (a) Cierto (b) Falso (c) Ni cierto ni falso

41. (2 puntos) Si en un modelo de Análisis de Varianza de dos tratamientos A y B no replicáramos ninguno de los experimentos:

    a) No podrían estimarse los α_i^A.
    b) No podrían estimarse los α_j^B.
    c) No podrían estimarse los α_i^A ni los α_j^B..
    d) No podrían estimarse los α_{ij}^{AB}.
    e) No podría estimarse más que la media general α.

42. (4 puntos) Has de estudiar los posibles efectos de la temperatura (alta, media, o baja), la presión (alta, media, o baja) y duración de la síntesis (larga, media, o corta) sobre la concentración del producto en un determinado proceso industrial.

Tales efectos, de existir, sabes que no interaccionan unos con otros. Para ello, cuentas tan solo con las siguientes observaciones:

Temperatura	Presión	Duración	Concentración
Alta	Alta	Corta	22.4 mg/cc
Alta	Media	Media	21.5 mg/cc
Alta	Baja	Larga	20.1 mg/cc
Media	Alta	Media	19.2 mg/cc
Media	Media	Larga	18.9 mg/cc
Media	Baja	Corta	19.4 mg/cc
Baja	Alta	Larga	22.1 mg/cc
Baja	Media	Corta	20.1 mg/cc
Baja	Baja	Media	20.3 mg/cc

¿Cuántas columnas tendría tu matriz de diseño?

(a) 4 (b) 7 (c) 10 (d) 18

43. (1 punto) Si la observación $(\vec{x}_i, y_i)$ es muy influyente, necesariamente su residuo externamente estudentizado al cuadrado, t_i^2 será grande.

(a) Cierto (b) Falso (c) Ni cierto ni falso

44. (2 puntos) Si d_i^2 (residuo borrado al cuadrado correspondiente a la observación $(\vec{x}_i, y_i)$ es muy grande, *podríamos* hallarnos en presencia de una observación influyente,

(a) Cierto (b) Falso (c) Ni cierto ni falso

45. (2 puntos) El comparar las sumas de cuadrados de los residuos para dos modelos proprocionaria un criterio razonable para escoger entre ellos si ambos tuvieran el mismo número de parámetros.

(a) Cierto (b) Falso (c) Ni cierto ni falso