

Estadística: Modelos Lineales

Final Enero 2.004, Tipo: A

Sección 1. Instrucciones

1. 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.
2. Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay personas que echan 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.

3. Al final, hay una Sección con unas pocas preguntas breves, que requieren cálculos no superiores a dos líneas: respóndelas directamente en el espacio libre.

Sección 2. Cuestiones de elección múltiple

4. El estimador *ridge*:
 - (a) Es un estimador no lineal: por eso puede lograr un mejor ECM (error cuadrático medio) que el estimador MCO.
 - (b) Es un estimador sesgado: por eso puede lograr un mejor ECM (error cuadrático medio) que el estimador MCO.
 - (c) No puede mejorar al MCO en términos de varianza, como quedó suficientemente demostrado en el teorema de Gauss-Markov.
 - (d) Es particularmente interesante en situaciones de fuerte multicolinealidad, aunque puede mejorar el ECM en todos los casos.
5. Un valor propio cercano a cero de la matriz $X'X$ sería indicio de:
 - (a) Multicolinealidad aproximada.
 - (b) No linealidad del diseño.
 - (c) R^2 muy pequeña.
 - (d) Varianza de la perturbación muy grande.

Apellidos y Nombre: _____

DNI: _____

Grupo: _____

Profesor : _____

6. Una observación con residuo MCO muy grande:
 - (a) Contribuye de modo importante a engrosar SSE.
 - (b) Siempre tendrá también un residuo borrado grande.
 - (c) Tendrá gran influencia sobre al menos uno de los β 's estimados.
 - (d) Corresponde a una observación que siempre tendremos interés en desechar.
7. Los factores de incremento de varianza (o VIF's, *variance inflation factors*):
 - (a) Son estadísticos que permiten diagnosticar multicolinealidad.
 - (b) Están relacionados con los residuos internamente studentizados.
 - (c) Están relacionados con los residuos externamente studentizados.
 - (d) Están relacionados con la R^2 de la regresión.
 - (e) Están relacionados con la R^2 de la regresión resultante de ajustar linealmente cada regresor regresándolo sobre todos los demás.
 - (f) Toman valores entre 0 y 1.
8. ¿Cuales de las siguientes propiedades posee *necesariamente* la matriz de covarianzas de los residuos de un modelo de regresión lineal ordinaria?
 - (a) Diagonal.
 - (b) Idempotente.
 - (c) Simétrica.
 - (d) De rango completo.

9. Para facilitar al analista la selección de modelos de regresión hay una porción de estrategias o algoritmos de búsqueda. ¿Cuál de ellos, si es que alguno, garantiza hallar *el mejor* de los modelos, de acuerdo con el criterio que hayamos especificado?

- (a) La estrategia *all subsets* o búsqueda entre todos los subconjuntos de regresores.
- (b) La regresión escalonada hacia delante.
- (c) La regresión esclonada hacia atrás.
- (d) La regresión escalonada mixta.
- (e) Todo falso.

10. A la vista de lo anterior,

- (a) La variable `Educación` parece tener un claro efecto, en el sentido de que un mayor nivel educativo conlleva un mayor prestigio social.
- (b) Una vez considerado el efecto de la educación y de la renta, el tipo de ocupación no parece tener efecto sobre el prestigio.
- (c) A igualdad de educación y renta, un profesional tendría en promedio un índice de prestigio en torno a 31 unidades superior al de un “white collar”.
- (d) Claramente, la variable “Tipo” de ocupación explica una mucho mayor porción de suma de cuadrados de Prestigio que cualquiera de las demás.

COMIENZO DE UN BLOQUE DE PREGUNTAS

Las preguntas hasta el siguiente trazo horizontal hacen referencia a los datos que siguen, relacionando un índice de prestigio de diferentes ocupaciones con sendos índices de ingresos, educación y una variable cualitativa (“Tipo”) tomando valores `prof` (= “Profesional”), `bc` (= “Blue collar”) y `wc` (= “White collar”, o empleado de “cuello blanco”).

Un extracto de los datos junto a la salida correspondiente al modelo que se ha ajustado se presenta a continuación:

	Tipo	Renta	Educacion	Prestigio
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90
minister	prof	21	84	87
professor	prof	64	93	93
dentist	prof	80	100	90
reporter	wc	67	87	52

```
Call:
lm(formula = Prestigio ~ Renta +
    Educacion + Tipo)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-14.890  -5.740  -1.754   5.442  28.972
```

```
Coefficients:
(Intercept)  -0.18503    3.71377   -0.050   0.96051
Renta         0.59755    0.08936    6.687  5.12e-08
Educacion     0.34532    0.11361    3.040   0.00416
Tipoprof     16.65751    6.99301    2.382   0.02206
Tipowc      -14.66113    6.10877   -2.400   0.02114
```

```
Residual standard error: 9.744 on 40 DF
Multiple R-Squared: 0.9131, Adj. R-squared: 0.9044
F-statistic: 105 on 4 and 40 DF, p-value: < 2.2e-16
```

11. Otros detalles que podemos deducir del estadillo anterior son:

- (a) Se ha empleado una muestra con 40 observaciones.
- (b) Se ha empleado una muestra con 45 observaciones.
- (c) La hipótesis de que los regresores considerados, en su conjunto, carecen de relación lineal con el regresando, puede ser contundentemente rechazada.

12. En el estadillo resumen de resultados correspondiente al modelo `mod1` tienes suficiente información para calcular el SSE correspondiente a dicho modelo. Resulta ser

- (a) 3797.82
- (b) 3576.34
- (c) 4012.34
- (d) 3696.34
- (e) Todo falso.

13. Con ayuda del SSE calculado en la pregunta anterior y el valor de R^2 es posible deducir que SST en el modelo `mod1` vale:

- (a) 42345.23
- (b) 43703.34
- (c) 38945.56
- (d) No es posible deducir lo que se pide.

14. Mirando la última columna, rotulada $Pr(>|t|)$, si los supuestos necesarios se verifican (el modelo ajustado es el correcto, las perturbaciones son normales, etc.) podemos afirmar que:

- (a) La estimación del parámetro correspondiente a la variable `Educacion` es claramente no significativa: es distinta de cero con probabilidad tan escasa como 0.00416.
- (b) Si realmente el β_i correspondiente a la variable `Educacion` fuera cero, un valor estimado como el obtenido de 0.34532 (o más grande en valor absoluto). se presentaría con probabilidad de tan solo 0.00416.
- (c) `Renta` es mucho más significativa que `Educacion`; por tanto, explica una mayor porción de suma de cuadrados del regresando.

15. Considerando la posibilidad de que `Prestigio` no crezca linealmente con `Renta`, decidimos ajustar el modelo cuyo resumen se incluye a continuación. (Nota el output ligeramente editado para que quepa en el espacio disponible: `Educacion` y `Educ` son la misma variable.)

```
mod2 <- lm(formula = Prestigio ~ Renta + Educacion
           + I(Educacion^2) + I(Educacion^3) + Tipo)

Residuals:
    Min       1Q   Median       3Q      Max
-15.042  -5.100  -1.985   5.870  28.103

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.605e+00  1.181e+01  0.474  0.6379
Renta        6.384e-01  9.421e-02  6.776 4.95e-08
Educacion    -3.842e-01  8.587e-01 -0.447  0.6571
I(Educ^2)    2.103e-02  1.773e-02  1.186  0.2429
I(Educ^3)   -1.480e-04  1.099e-04 -1.347  0.1859
Tipoprof     1.152e+01  7.667e+00  1.502  0.1413
Tipowc      -2.099e+01  7.347e+00 -2.857  0.0069
---
Residual standard error: 9.681 on 38 DF
Multiple R-Squared: 0.9185, Adj. R-squared: 0.9056
F-statistic: 71.36 on 6 and 38 DF, p-value: < 2.2e-16
```

Nos encontramos ahora con que ¡ninguna de las variables de educación es significativa, mientras que en el modelo anterior, `Educacion` lo era! ¿Qué conjeturarías al respecto?

- (a) La teoría ya no es aplicable, porque el nuevo modelo considerado no es lineal.
- (b) Cuando se emplean tres grados de libertad en ajustar el efecto de una sola variable, como en este caso, las estimaciones nunca pueden ser significativas.
- (c) Probablemente las variables `Educacion^2` y `Educacion^3` introducen una acusada multicolinealidad.
- (d) Todo falso.

16. *Supongamos* que el modelo `mod2` es el más parametrizado de los razonables, y por ello adecuado para obtener una estimación $\hat{\sigma}^2$ no sesgada. Esta estimación sería:

- (a) 93.722
- (b) 90.830
- (c) 81.234
- (d) Todo falso.

17. El valor de la C_p de Mallows correspondiente a dicho modelo `mod2` haciendo uso del $\hat{\sigma}^2$ calculado en la pregunta anterior sería:
- 34.2
 - 52
 - 23.1
 - 6.4
18. Comparando la C_p calculada en la pregunta anterior con la que puedes calcular para el modelo `mod1` llegaríamos a la conclusión de que entre `mod1` y `mod2`, sobre la base del criterio C_p , es preferible:
- El modelo `mod1`.
 - El modelo `mod2`.
19. De acuerdo con el criterio de la \overline{R}^2 , el modelo preferido sería:
- El modelo `mod1`.
 - El modelo `mod2`.
20. Cuando se omite en un modelo de regresión lineal un regresor que hubiera debido aparecer:
- Siempre se sesgan los estimadores de los β 's correspondientes a los regresores incluidos.
 - En general, se sesgan los estimadores de los β 's correspondientes a los regresores incluidos; pero podría ocurrir que no fuera así si el regresor omitido es ortogonal a todos los presentes.
 - Siempre se sesga el estimador de σ^2 , varianza de la perturbación, a menos que el regresor omitido carezca en absoluto de capacidad predictiva sobre el regresando.
 - El estimador de σ^2 , varianza de la perturbación, continúa siendo insesgado si el regresor omitido es ortogonal a los incluidos.
21. La coexistencia de un R^2 muy elevado y unos t -ratios en su totalidad no significativos —salvo, quizá, el correspondiente a la columna de “unos”— es un síntoma indicativo de que:
- La especificación lineal es inadecuada, y se hace preciso probar una regresión no lineal.
 - El número de observaciones es claramente insuficiente.
 - Existen *outliers*.
 - Existe multicolinealidad.
22. ¿En cuál o cuales de las siguientes situaciones resultaría manifiestamente inadecuada la inclusión en el modelo de una columna de “unos”?
- Regresión de $Y =$ “Consumo en el periodo t ” sobre $X =$ “Renta en el periodo t ” para una muestra de N familias, con objeto de estimar la propensión media al consumo (supuesta constante).
 - Regresión de $Y =$ “Peso de un objeto” sobre $X =$ “Volumen del mismo objeto”, para una muestra de N objetos de la misma materia con el fin de estimar el peso específico de dicha materia.
 - Regresión de $Y =$ “Velocidad de sedimentación de una solución acuosa” sobre $X =$ “Temperatura en $^{\circ}\text{C}$ ” para una muestra de N soluciones de la misma materia y concentración.
23. Cuando se incluyen regresores irrelevantes en un modelo de regresión:
- Las estimaciones de cualquiera de los β 's pueden resultar sesgadas.
 - La estimación de σ^2 resultará sesgada por exceso.
 - La estimación de σ^2 resultará sesgada por defecto.
 - La estimación de σ^2 tendrá menos grados de libertad que los que hubiera tenido de ajustarse el modelo correcto.

FIN DEL BLOQUE DE PREGUNTAS

24. El estimador MCO verifica:
- Es de mínima varianza entre todos los estimadores lineales.
 - Es de mínima varianza entre todos los estimadores insesgados.
 - Es de mínimo error cuadrático medio entre todos los estimadores lineales.
 - Es de mínimo error cuadrático medio entre todos los estimadores lineales e insesgados.
 - Es de mínima varianza entre todos los estimadores lineales e insesgados.
25. Cuando en un modelo que incluye la columna de “unos” tomamos una variable y la reescalamos (expresándola, por ejemplo, en unidades que son mil veces mayores que las primitivas, de modo que la variable correspondiente toma valores mil veces más pequeños), se verifica que:
- El ajuste es idéntico al primitivo; el coeficiente estimado de la variable reescalada será ahora mil veces más pequeño.
 - El ajuste es idéntico al primitivo; el coeficiente estimado de la variable reescalada será ahora mil veces más grande.
 - R^2 no varía.
 - SSE y SSR no varían, pero SST puede hacerlo.
26. Se conoce como distancia de Cook:
- El mayor número de millas náuticas que el capitán Cook recorrió durante un periodo de 24 horas consecutivas en sus exploraciones por los mares australes.
 - Una medida de influencia de la observación i -ésima sobre *uno* cualquiera de los β_i .
 - Una medida de influencia de la observación i -ésima sobre *el conjunto* de los β_j .
 - El denominador de los residuos borrados.
 - Todo falso.
27. Cuando se ajusta un modelo $y = X\beta + \epsilon$ y la columna de “unos” está entre los regresores, la R^2 ordinaria (no corregida) puede interpretarse como:
- El ángulo que forma y con el subespacio que generan las columnas de X .
 - El coseno del ángulo que forma y con el subespacio que generan las columnas de X .
 - El coseno al cuadrado del ángulo que forma y con el subespacio que generan las columnas de X .
 - El nivel de significación de los β estimados.
 - Una medida de la calidad del ajuste.
28. Si se hace regresión en componentes principales y se toman tantas componentes principales como regresores hay (excluida, en su caso, la columna de “unos”),
- Se obtiene una solución que reproduce la que obtendríamos mediante MCO.
 - Se soluciona radicalmente el problema de la multicolinealidad, porque al ser la componentes principales ortogonales por construcción, no puede existir multicolinealidad ni ninguno de sus perniciosos efectos.
 - Se obtiene un estimador no lineal y sesgado.
 - Se obtiene una solución idéntica a la que obtendríamos mediante el estimador *ridge* haciendo $k = p$.
29. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por intervalo 95 % para el índice de prestigio de una ocupación de tipo “cuello blanco” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)

30. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por punto para el índice de prestigio de una ocupación de tipo “cuello blanco” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)
31. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por punto para el índice de prestigio de una ocupación de tipo “blue collar” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)

Respuestas al examen de tipo A

Sección 1. Instrucciones

1. 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.
2. Intento medir conocimientos y no agudeza visual. Inevitablemente, en un examen de este tipo hay que prestar mucha atención. Cada curso hay personas que echan 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.

3. Al final, hay una Sección con unas pocas preguntas breves, que requieren cálculos no superiores a dos líneas: respóndelas directamente en el espacio libre.

Sección 2. Cuestiones de elección múltiple

4. El estimador *ridge*:
 - (a) Es un estimador no lineal: por eso puede lograr un mejor ECM (error cuadrático medio) que el estimador MCO.
 - (b) **Es un estimador sesgado: por eso puede lograr un mejor ECM (error cuadrático medio) que el estimador MCO.**
 - (c) No puede mejorar al MCO en términos de varianza, como quedó suficientemente demostrado en el teorema de Gauss-Markov.
 - (d) **Es particularmente interesante en situaciones de fuerte multicolinealidad, aunque puede mejorar el ECM en todos los casos.**
5. Un valor propio cercano a cero de la matriz $X'X$ sería indicio de:
 - (a) **Multicolinealidad aproximada.**
 - (b) No linealidad del diseño.
 - (c) R^2 muy pequeña.
 - (d) Varianza de la perturbación muy grande.
6. Una observación con residuo MCO muy grande:
 - (a) **Contribuye de modo importante a engrosar SSE.**
 - (b) Siempre tendrá también un residuo borrado grande.
 - (c) Tendrá gran influencia sobre al menos uno de los β 's estimados.
 - (d) Corresponde a una observación que siempre tendremos interés en desechar.
7. Los factores de incremento de varianza (o VIF's, *variance inflation factors*):
 - (a) **Son estadísticos que permiten diagnosticar multicolinealidad.**
 - (b) Están relacionados con los residuos internamente studentizados.
 - (c) Están relacionados con los residuos externamente studentizados.
 - (d) Están relacionados con la R^2 de la regresión.
 - (e) **Están relacionados con la R^2 de la regresión resultante de ajustar linealmente cada regresor regresándolo sobre todos los demás.**
 - (f) Toman valores entre 0 y 1.
8. ¿Cuales de las siguientes propiedades posee *necesariamente* la matriz de covarianzas de los residuos de un modelo de regresión lineal ordinaria?
 - (a) Diagonal.
 - (b) Idempotente.
 - (c) **Simétrica.**
 - (d) De rango completo.

9. Para facilitar al analista la selección de modelos de regresión hay una porción de estrategias o algoritmos de búsqueda. ¿Cuál de ellos, si es que alguno, garantiza hallar *el mejor* de los modelos, de acuerdo con el criterio que hayamos especificado?

- (a) **La estrategia *all subsets* o búsqueda entre todos los subconjuntos de regresores.**
- (b) La regresión escalonada hacia delante.
- (c) La regresión esclonada hacia atrás.
- (d) La regresión escalonada mixta.
- (e) Todo falso.

10. A la vista de lo anterior,

- (a) **La variable Educación parece tener un claro efecto, en el sentido de que un mayor nivel educativo conlleva un mayor prestigio social.**
- (b) Una vez considerado el efecto de la educación y de la renta, el tipo de ocupación no parece tener efecto sobre el prestigio.
- (c) **A igualdad de educación y renta, un profesional tendría en promedio un índice de prestigio en torno a 31 unidades superior al de un “white collar”.**
- (d) Claramente, la variable “Tipo” de ocupación explica una mucho mayor porción de suma de cuadrados de Prestigio que cualquiera de las demás.

COMIENZO DE UN BLOQUE DE PREGUNTAS

Las preguntas hasta el siguiente trazo horizontal hacen referencia a los datos que siguen, relacionando un índice de prestigio de diferentes ocupaciones con sendos índices de ingresos, educación y una variable cualitativa (“Tipo”) tomando valores `prof` (= “Profesional”), `bc` (= “Blue collar”) y `wc` (= “White collar”, o empleado de “cuello blanco”).

Un extracto de los datos junto a la salida correspondiente al modelo que se ha ajustado se presenta a continuación:

	Tipo	Renta	Educacion	Prestigio
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90
minister	prof	21	84	87
professor	prof	64	93	93
dentist	prof	80	100	90
reporter	wc	67	87	52

```
Call:
lm(formula = Prestigio ~ Renta +
    Educacion + Tipo)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-14.890  -5.740  -1.754   5.442  28.972
```

```
Coefficients:
(Intercept)  -0.18503    3.71377   -0.050   0.96051
Renta         0.59755    0.08936    6.687  5.12e-08
Educacion     0.34532    0.11361    3.040   0.00416
Tipoprof     16.65751    6.99301    2.382   0.02206
Tipowc      -14.66113    6.10877   -2.400   0.02114
```

```
Residual standard error: 9.744 on 40 DF
Multiple R-Squared: 0.9131, Adj. R-squared: 0.9044
F-statistic: 105 on 4 and 40 DF, p-value: < 2.2e-16
```

11. Otros detalles que podemos deducir del estadillo anterior son:

- (a) Se ha empleado una muestra con 40 observaciones.
- (b) **Se ha empleado una muestra con 45 observaciones.**
- (c) **La hipótesis de que los regresores considerados, en su conjunto, carecen de relación lineal con el regresando, puede ser contundentemente rechazada.**

12. En el estadillo resumen de resultados correspondiente al modelo `mod1` tienes suficiente información para calcular el SSE correspondiente a dicho modelo. Resulta ser

- (a) **3797.82**
- (b) 3576.34
- (c) 4012.34
- (d) 3696.34
- (e) Todo falso.

13. Con ayuda del SSE calculado en la pregunta anterior y el valor de R^2 es posible deducir que SST en el modelo `mod1` vale:

- (a) 42345.23
- (b) **43703.34**
- (c) 38945.56
- (d) No es posible deducir lo que se pide.

14. Mirando la última columna, rotulada $Pr(>|t|)$, si los supuestos necesarios se verifican (el modelo ajustado es el correcto, las perturbaciones son normales, etc.) podemos afirmar que:

- (a) La estimación del parámetro correspondiente a la variable `Educacion` es claramente no significativa: es distinta de cero con probabilidad tan escasa como 0.00416.
- (b) **Si realmente el β_i correspondiente a la variable `Educacion` fuera cero, un valor estimado como el obtenido de 0.34532 (o más grande en valor absoluto), se presentaría con probabilidad de tan solo 0.00416.**
- (c) `Renta` es mucho más significativa que `Educacion`; por tanto, explica una mayor porción de suma de cuadrados del regresando.

15. Considerando la posibilidad de que `Prestigio` no crezca linealmente con `Renta`, decidimos ajustar el modelo cuyo resumen se incluye a continuación. (Nota el output ligeramente editado para que quepa en el espacio disponible: `Educacion` y `Educ` son la misma variable.)

```
mod2 <- lm(formula = Prestigio ~ Renta + Educacion
           + I(Educacion^2) + I(Educacion^3) + Tipo)

Residuals:
    Min       1Q   Median       3Q      Max
-15.042  -5.100  -1.985   5.870  28.103

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.605e+00  1.181e+01  0.474  0.6379
Renta        6.384e-01  9.421e-02  6.776 4.95e-08
Educacion    -3.842e-01  8.587e-01  -0.447  0.6571
I(Educ^2)     2.103e-02  1.773e-02  1.186  0.2429
I(Educ^3)    -1.480e-04  1.099e-04  -1.347  0.1859
Tipoprof     1.152e+01  7.667e+00  1.502  0.1413
Tipowc      -2.099e+01  7.347e+00  -2.857  0.0069
---
Residual standard error: 9.681 on 38 DF
Multiple R-Squared: 0.9185, Adj. R-squared: 0.9056
F-statistic: 71.36 on 6 and 38 DF, p-value: < 2.2e-16
```

Nos encontramos ahora con que ¡ninguna de las variables de educación es significativa, mientras que en el modelo anterior, `Educacion` lo era! ¿Qué conjeturarías al respecto?

- (a) La teoría ya no es aplicable, porque el nuevo modelo considerado no es lineal.
- (b) Cuando se emplean tres grados de libertad en ajustar el efecto de una sola variable, como en este caso, las estimaciones nunca pueden ser significativas.
- (c) **Probablemente las variables `Educacion^2` y `Educacion^3` introducen una acusada multicolinealidad.**
- (d) Todo falso.

16. *Supongamos* que el modelo `mod2` es el más parametrizado de los razonables, y por ello adecuado para obtener una estimación $\hat{\sigma}^2$ no sesgada. Esta estimación sería:

- (a) **93.722**
- (b) 90.830
- (c) 81.234
- (d) Todo falso.

17. El valor de la C_p de Mallows correspondiente a dicho modelo `mod2` haciendo uso del $\hat{\sigma}^2$ calculado en la pregunta anterior sería:
- (a) 34.2
 - (b) **52**
 - (c) 23.1
 - (d) 6.4
18. Comparando la C_p calculada en la pregunta anterior con la que puedes calcular para el modelo `mod1` llegaríamos a la conclusión de que entre `mod1` y `mod2`, sobre la base del criterio C_p , es preferible:
- (a) **El modelo `mod1`.**
 - (b) El modelo `mod2`.
19. De acuerdo con el criterio de la \overline{R}^2 , el modelo preferido sería:
- (a) El modelo `mod1`.
 - (b) **El modelo `mod2`.**
- FIN DEL BLOQUE DE PREGUNTAS
-
20. Cuando se omite en un modelo de regresión lineal un regresor que hubiera debido aparecer:
- (a) Siempre se sesgan los estimadores de los β 's correspondientes a los regresores incluidos.
 - (b) **En general, se sesgan los estimadores de los β 's correspondientes a los regresores incluidos; pero podría ocurrir que no fuera así si el regresor omitido es ortogonal a todos los presentes.**
 - (c) **Siempre se sesga el estimador de σ^2 , varianza de la perturbación, a menos que el regresor omitido carezca en absoluto de capacidad predictiva sobre el regresando.**
 - (d) El estimador de σ^2 , varianza de la perturbación, continúa siendo insesgado si el regresor omitido es ortogonal a los incluidos.
21. La coexistencia de un R^2 muy elevado y unos t -ratios en su totalidad no significativos —salvo, quizá, el correspondiente a la columna de “unos”— es un síntoma indicativo de que:
- (a) La especificación lineal es inadecuada, y se hace preciso probar una regresión no lineal.
 - (b) El número de observaciones es claramente insuficiente.
 - (c) Existen *outliers*.
 - (d) **Existe multicolinealidad.**
22. ¿En cuál o cuales de las siguientes situaciones resultaría manifiestamente inadecuada la inclusión en el modelo de una columna de “unos”?
- (a) Regresión de $Y =$ “Consumo en el periodo t ” sobre $X =$ “Renta en el periodo t ” para una muestra de N familias, con objeto de estimar la propensión media al consumo (supuesta constante).
 - (b) **Regresión de $Y =$ “Peso de un objeto” sobre $X =$ “Volumen del mismo objeto”, para una muestra de N objetos de la misma materia con el fin de estimar el peso específico de dicha materia.**
 - (c) Regresión de $Y =$ “Velocidad de sedimentación de una solución acuosa” sobre $X =$ “Temperatura en $^{\circ}\text{C}$ ” para una muestra de N soluciones de la misma materia y concentración.
23. Cuando se incluyen regresores irrelevantes en un modelo de regresión:
- (a) Las estimaciones de cualquiera de los β 's pueden resultar sesgadas.
 - (b) La estimación de σ^2 resultará sesgada por exceso.
 - (c) La estimación de σ^2 resultará sesgada por defecto.
 - (d) **La estimación de σ^2 tendrá menos grados de libertad que los que hubiera tenido de ajustarse el modelo correcto.**

24. El estimador MCO verifica:
- Es de mínima varianza entre todos los estimadores lineales.
 - Es de mínima varianza entre todos los estimadores insesgados.
 - Es de mínimo error cuadrático medio entre todos los estimadores lineales.
 - Es de mínimo error cuadrático medio entre todos los estimadores lineales e insesgados.**
 - Es de mínima varianza entre todos los estimadores lineales e insesgados.**
25. Cuando en un modelo que incluye la columna de “unos” tomamos una variable y la reescalamos (expresándola, por ejemplo, en unidades que son mil veces mayores que las primitivas, de modo que la variable correspondiente toma valores mil veces más pequeños), se verifica que:
- El ajuste es idéntico al primitivo; el coeficiente estimado de la variable reescalada será ahora mil veces más pequeño.
 - El ajuste es idéntico al primitivo; el coeficiente estimado de la variable reescalada será ahora mil veces más grande.**
 - R^2 no varía.**
 - SSE y SSR no varían, pero SST puede hacerlo.
26. Se conoce como distancia de Cook:
- El mayor número de millas náuticas que el capitán Cook recorrió durante un periodo de 24 horas consecutivas en sus exploraciones por los mares australes.
 - Una medida de influencia de la observación i -ésima sobre *uno* cualquiera de los β_i .
 - Una medida de influencia de la observación i -ésima sobre el conjunto de los β_j .**
 - El denominador de los residuos borrados.
 - Todo falso.
27. Cuando se ajusta un modelo $y = X\beta + \epsilon$ y la columna de “unos” está entre los regresores, la R^2 ordinaria (no corregida) puede interpretarse como:
- El ángulo que forma y con el subespacio que generan las columnas de X .
 - El coseno del ángulo que forma y con el subespacio que generan las columnas de X .
 - El coseno al cuadrado del ángulo que forma y con el subespacio que generan las columnas de X .**
 - El nivel de significación de los β estimados.
 - Una medida de la calidad del ajuste.**
28. Si se hace regresión en componentes principales y se toman tantas componentes principales como regresores hay (excluida, en su caso, la columna de “unos”),
- Se obtiene una solución que reproduce la que obtendríamos mediante MCO.**
 - Se soluciona radicalmente el problema de la multicolinealidad, porque al ser la componentes principales ortogonales por construcción, no puede existir multicolinealidad ni ninguno de sus perniciosos efectos.
 - Se obtiene un estimador no lineal y sesgado.
 - Se obtiene una solución idéntica a la que obtendríamos mediante el estimador *ridge* haciendo $k = p$.
29. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por intervalo 95 % para el índice de prestigio de una ocupación de tipo “cuello blanco” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)

Sección 3. Preguntas breves

Respuesta: $-0,18503 + 0,59755 \times 33 + 0,34532 \times 74 - 14,6611$

31. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por punto para el índice de prestigio de una ocupación de tipo “blue collar” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)

Respuesta: Para esto se haría preciso disponer de la matrix $(X'X)^{-1}$, que no está disponible ni puede deducirse de los estadillos anteriores.

30. Haciendo uso del modelo `mod1` más arriba, y con los supuestos habituales más normalidad, construye una predicción por punto para el índice de prestigio de una ocupación de tipo “cuello blanco” cuyo índice de Renta fuera 33 y cuyo índice de Educación fuera 74. (Si no tuvieras información suficiente para hacerlo, indica qué es lo que te falta.)

Respuesta: $-0,18503 + 0,59755 \times 33 + 0,34532 \times 74$