

Els alumnes amb el primer parcial aprovat han de fer els exercicis 2, 3 i 4.

La resta han de fer els exercicis 1, 2 i 3.

Cada exercici es pot fer en una hora. El temps també és un factor en l'avaluació.

Problema 1

Considereu un model lineal amb $n = 3$ observacions Y_1 , Y_2 i Y_3 i quatre paràmetres μ , α_1 , α_2 i α_3 . Suposem que

$$\begin{aligned}E(Y_1) &= \mu + \alpha_1 + \alpha_2 + \alpha_3 \\E(Y_2) &= \mu + \alpha_1 + \alpha_2 \\E(Y_3) &= \mu + \alpha_1\end{aligned}$$

Sigui $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$ el vector de paràmetres.

- (a) Escriviu aquest model en la forma $\mathbf{Y} = \mathbf{X}\beta + \epsilon$.
- (b) Caracteritzeu cadascuna de les següents funcions com estimable o no estimable: μ , α_1 , $\mu + \alpha_1$, $\mu + \alpha_1 + \alpha_2 + \alpha_3$. Justifiqueu la resposta.
- (c) Escriviu les equacions normals d'aquest model i trobeu dues solucions β_1 i β_2 si $Y_1 = 7.03$, $Y_2 = 5.98$ i $Y_3 = 4.23$.
- (d) Calculeu les estimacions mínimo-quadràtiques de les funcions estimables de l'apartat (b). Verifiqueu que aquestes estimacions són invariants per a qualsevol solució com β_1 o β_2 .
- (e) Feu el contrast de la hipòtesi $H_0: \mu + \alpha_1 = 4$.

Problema 2

La base de dades `sat` del paquet `faraway` conté les dades d'un treball titulat *Getting What You Pay For: The Devate Over Equity in Public School Expenditures* que preten estudiar les relacions entre la despesa en educació pública i els resultats del test d'avaluació de coneixements que s'anomena SAT.

- (a) Feu un resum numèric i gràfic de les dades i comenteu tot el que penseu que és rellevant.
- (b) Ajusteu un model de regressió amb la puntuació `total` com a variable resposta i les variables `expend`, `ratio` i `salary` com predictores. Contrasteu la hipòtesi $\beta_{salary} = 0$. Contrasteu la hipòtesi $\beta_{salary} = \beta_{ratio} = \beta_{expend} = 0$. Hi ha alguna predictora que tingui efecte sobre la resposta?
- (c) Afegiu la variable `takers` com a regressora. Contrasteu ara la hipòtesi $\beta_{salary} = 0$. Compareu aquest model amb el model anterior amb un test F . Aquest darrer test és equivalent a un test t , quin és aquest test?
- (d) Proporcioneu un interval de confiança al 95% per al coeficient β_{salary} en el darrer model.
- (e) Considereu el model que té la puntuació `total` SAT com a resposta i `expend`, `salary`, `ratio` i `takers` com predictores. Amb els gràfics o els valors adients, investigueu la dignosi d'aquest model en el següents punts:
 - (i) Variància constant dels errors.
 - (ii) Hipòtesi de normalitat.
 - (iii) Punts amb influència potencial (leverage).
 - (iv) Outliers.
 - (v) Punts influents.
 - (vi) Problemes de colinearitat.

Problema 3

La base de dades `stackloss` del paquet `faraway` presenta les observacions de 21 dies d'operació d'una planta per l'oxidació de l'amoníac (NH_3) en àcid nítric (HNO_3). Els òxids de nitrogen produïts s'absorbeixen en una torre d'absorció a contracorrent. (Brownlee, citat per Dodge, lleugerament reordenada per MM.)

`Air.Flow` és el flux d'aire i representa la taxa d'operació de la planta.

`Water.Temp` és la temperatura de l'aigua de refredament que circula per un serpentí a la torre d'absorció.

`Acid.Conc.` és la concentració dels metabòlits de l'àcid.

`stack.loss` (variable dependent) és 10 vegades el percentatge d'amoníac a la planta que s'escapa de la columna d'absorció sense ser absorbit, és a dir, una mesura (inversa) de l'eficiència de conjunt de la planta.

Ajusteu un model amb la variable `stack.loss` com a resposta i les altres tres variables com a predictores amb els següents mètodes:

- (a) Mínims quadrats *OLS*.
- (b) *Least absolute deviations*.
- (c) Mètode de Huber.
- (d) *Least trimmed squares*.

Compareu els resultats. Feu servir els mètodes de diagnosi en el model ordinari *OLS* per detectar outliers i punts influents. Elimineu aquests punts i recalculeu el model *OLS*. Compareu els resultats.

Problema 4

Prenem la base de dades `fat` del paquet `faraway` i feu servir el percentatge de grasa corporal com a resposta i les altres variables com a potencials regressores. Elimineu una de cada 10 observacions per utilitzar-les com a mostra de prova (*sample test*). Feu servir la resta d'observacions com a mostra d'ajust (*training sample*) per ajustar els següents models:

- (a) Regressió lineal amb totes les predictores.
- (b) Regressió lineal amb les variables seleccionades per AIC.
- (c) Regressió per components principals.
- (d) *Partial least squares*.
- (e) *Ridge regression*.

Podeu trobar més informació d'aquesta base de dades en l'adreça

<http://www.stat.unc.edu/faculty/cji/664/HW5-faraway.pdf>

Feu servir les instruccions finals per eliminar les variables que no són regressores i muntar les dues mostres `train` i `test`. La variable resposta és `siri`.