



Grau d'Estadística

S'ha d'entregar un document PDF o HTML amb les respostes i un arxiu R ben ordenat per apartats i amb tots els comentaris (amb el símbol #) que considereu oportuns. En el nom dels fitxers no poseu accents, només símbols ASCII del 0 al 127 (sense ñ,ç,...).

Problema 1

Amb el model lineal normal

$$\begin{aligned}10.98 &= 2\alpha + \beta + 3\gamma + \epsilon_1 \\11.03 &= 2\alpha + \beta + 3\gamma + \epsilon_2 \\0.95 &= \alpha - \beta + \epsilon_3 \\4.03 &= \alpha + \gamma + \epsilon_4 \\6.94 &= \alpha + \beta + 2\gamma + \epsilon_5 \\7.02 &= \alpha + \beta + 2\gamma + \epsilon_6\end{aligned}$$

contesteu les següents qüestions:

- (a) Quina condició ha de verificar una funció paramètrica per a que sigui estimable en aquest model?
- (b) Indiqueu si les funcions paramètriques següents són estimables i calculeu l'estimador MQ quan sigui possible:

(i) $\alpha + \gamma$ (ii) γ

- (c) Calculeu l'estimació de la covariància entre els estimadors lineals òptims de $\alpha + \gamma$ i $\beta + \gamma$ i la variància de l'estimador lineal òptim de $\alpha + \gamma$.
- (d) Calculeu un interval de confiança al 95% per a $\alpha + \gamma$.
- (e) Feu el contrast de la hipòtesi $H_0 : \alpha + \gamma = 4$.
Feu també el contrast de la hipòtesi $H_0 : \alpha + \gamma = 4, \beta + \gamma = 3$.

Problema 2

En el llibre de McDonald(2014) [?] s'utilitza una base de dades sobre un peix d'aigua dolça i clima suau que es troba en els rierols de Maryland: "longnose dace" (*Rhinichthys cataractae*).



La variable dependent és el nombre de peixos per secció de 75 metres del rierol. Les variables independents són: l'àrea (en acres) drenats del rierol, l'oxigen dissolt (en mg/litre), la profunditat màxima (en cm) del segment de 75 metres del rierol, la concentració de nitrat (mg/litre), la concentració de sulfat (mg/litre) i la temperatura de l'aigua a la data de mostreig (en graus C).

Un dels objectius biològics podria ser la de mesurar les característiques físiques i químiques d'un rierol i ser capaç de predir l'abundància de *rhinichthys cataractae*. Un altre objectiu podria ser el de generar hipòtesis sobre les causes de la variació en l'abundància de *rhinichthys cataractae*.

- (a) Feu un resum numèric de les variables implicades en la regressió. Feu també un gràfic que relacioni les variables dos a dos. Té sentit calcular la correlació entre variables? Quines són les variables regressores més correlacionades amb la variable resposta?

Encara que es poden contestar aquestes preguntes amb les funcions bàsiques de R, si feu servir la funció `chart.Correlation()` del paquet `PerformanceAnalytics` obtindreu un bonic gràfic i la resposta a la darrera pregunta.

- (b) Estudieu la normalitat de la variable resposta. Sabeu fer el test de normalitat de D'Agostino? Feu un histograma de les dades d'aquesta variable i pinte una corba de densitat estimada.

I si apliquem logaritmes¹?

- (c) Obteniu l'estimació dels paràmetres del model (també de σ^2) i calculeu el coeficient de determinació. Què mesura aquest coeficient?

És significativa la regressió? Què significa això últim?

- (d) Investigueu la validesa del model lineal.

Hi ha algun residu que podríem qualificar de atípic? Feu una definició del concepte *dada atípica* (outlier) en aquest context. Feu algun gràfic adient.

- (e) Doneu els intervals de confiança al 98% de β_0 , de β_{Temp} i de σ^2 . Com interpretem β_0 , β_{Temp} i σ^2 ? En quines unitats estan aquests coeficients?

- (f) Contrasteu la hipòtesi $H_0 : \beta_{Temp} = 3$. Contrasteu també la hipòtesi $H_0 : \beta_{DO2} = \beta_{NO3}$.

Utilitzeu un nivell de significació $\alpha = 0.05$.

- (g) Quina seria la predicció del número de peixos en cadascuna de les seccions estudiades si augmentem en 1 grau només la temperatura de l'aigua i mantenim els valors de les altres variables? Feu una predicció concreta amb IC al 90%.

- (h) Feu un test de permutacions de la hipòtesi de no significació de la regressió (tots els coeficients són zero, excepte el terme d'intercepció).

Referències

- [1] McDonald, J.H. 2014. Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland.

¹Encara que sembla millor aplicar logaritmes a la variable resposta i també a d'altres regressores, en aquest examen NO ho farem per no complicar l'estudi.