

MODEL 1

1. Cas pràctic 1

Amb les dades associades als passatgers del Titànic es pretén estimar un model que expliqui la probabilitat de supervivència (*Survived*, 1 en cas positiu i 0 en cas negatiu) i es disposa de les següents variables explicatives: edat (*Age*, en anys), classe en la qual viatjaven (*Pclass*, 1 per a la 1a classe, 2 per a 2a classe i 3 per a la 3a classe), nombre de familiars amb els quals viatjava (*sibSp*) i finalment el sexe (*sex*, 1 si és home, 0 si és dona). En primer lloc, es va estimar el següent model per Mínims Quadrats Ordinaris:

```
lm(formula = Survived ~ Pclass + sex + Age + sibSp)

Residuals:
    Min       1Q   Median       3Q      Max
-0.98802 -0.21148 -0.07451  0.22785  1.00002

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.2076079   0.0519132   23.262 < 2e-16 ***
Pclass       -0.1761377   0.0167104  -10.541 < 2e-16 ***
sex          -0.5188206   0.0274746  -18.884 < 2e-16 ***
Age          -0.0026832   0.0007967   -3.368 0.00079 ***
sibSp        -0.0380853   0.0149149   -2.554 0.01083 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3843 on 886 degrees of freedom
Multiple R-squared:  0.3792, Adjusted R-squared:  0.3764
F-statistic: 135.3 on 4 and 886 DF, p-value: < 2.2e-16
```

a) Quina seria la probabilitat de sobreviure segons el model anterior per a una persona de 33 anys, que viatja a 3a classe, home i que viatja amb tres familiars més? Té sentit el resultat obtingut? Raona la teva resposta. Quins són els principals problemes d'utilitzar el Model Lineal de Probabilitat com al cas anterior?

Com a alternativa al model anterior es proposa l'estimació de l'especificació amb un Logit, obtenint-se els resultats següents:

```
glm(formula = Survived ~ Pclass + sex + Age + sibSp, family =
binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3821  -0.6199  -0.4183   0.6493   2.4531

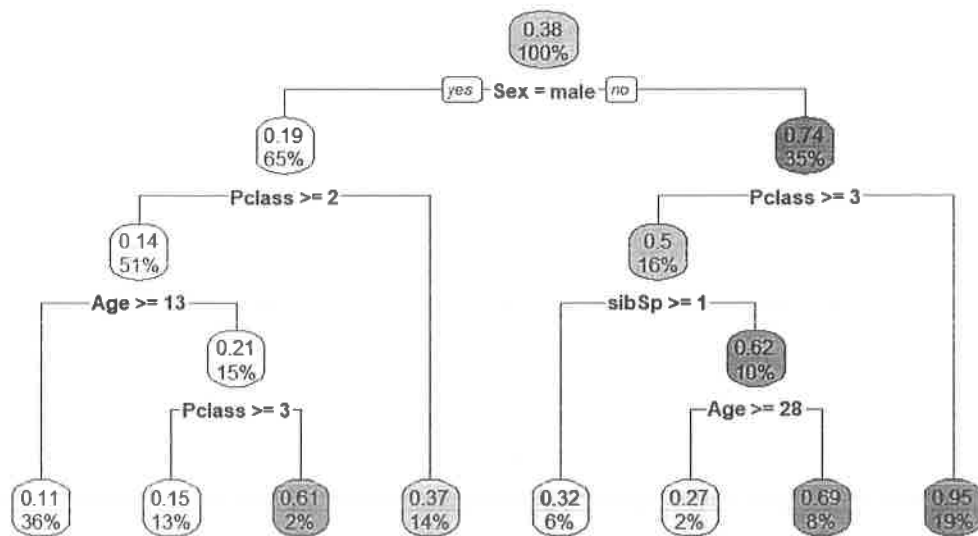
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.175718   0.392975  10.626 < 2e-16 ***
Pclass       -1.094619   0.117605  -9.308 < 2e-16 ***
sex          -2.719923   0.190705 -14.262 < 2e-16 ***
Age          -0.017314   0.005382  -3.217 0.00129 **
sibSp        -0.269533   0.110574  -2.438 0.01479 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.66  on 890  degrees of freedom
Residual deviance:  811.17  on 886  degrees of freedom
AIC: 921.17

Number of Fisher Scoring iterations: 4
```

- b) Calcula la probabilitat de l'apartat anterior amb aquest model. Tenen una interpretació directa els paràmetres estimats en aquest cas?
- c) Per a un individu amb característiques iguals a les dels apartats a) i b), calcular quin efecte tindria en la probabilitat de sobreviure el fet de passar de tenir 33 a 34 anys. Explica com s'interpretaria $\exp(-0.017314)$ és a dir, l'exponencial del paràmetre associat a l'edat.
- d) Explica què creus que s'està fent malament a l'hora d'introduir la variable *Pclass* en les estimacions anteriors i com podria solucionar-se.
- Finalment, amb les dades anteriors s'ha estimat un arbre (*Decision Tree* o *Regression Tree*) i s'han obtingut els resultats següents:



- e) Interpreta els resultats presentats al quadre anterior.

2. Cas pràctic 2

Aquest exercici consisteix en emplenar correctament la fitxa de l'Inventari d'activitats estadístiques d'una nova operació estadística d'acord amb la normativa vigent, i amb especial atenció al que estableix el Pla d'estadística de les Illes Balears 2018-2021, aprovat pel Decret 40/2018, de 23 de novembre, d'acord amb el següent supòsit:

L'Institut d'Estadística de les Illes Balears (IBESTAT), en tant que unitat estadística del Sistema Estadístic de les Illes Balears (Sestib), ha decidit dur a terme el 2023 una nova operació estadística per conèixer els usos lingüístics entre la població de les illes Balears, la qual s'ha donat d'alta en el Programa anual i Pla estadístic corresponent.

Aquesta operació es durà a terme cada cinc anys, i consistirà en una enquesta telefònica que es farà a la població de les Illes Balears major de 16 anys. Les dades que es volen obtenir han

de poder ser desagregades per a la seva difusió per illa (Mallorca, Menorca, Eivissa i Formentera) en les seves principals variables; tot i que en el cas de Mallorca també s'obtindrà informació per poder fer la desagregació entre el municipi de Palma i la Part Forana a efectes d'anàlisis interns.

L'enquesta contemplarà el nivell/grau de capacitat lingüística respecte del català, el castellà i l'anglès (l'entén, el parla, el sap llegir, l'escriu); així com el seu ús en distints àmbits per part dels enquestats (família, amics, treball, xarxes socials, mitjans de comunicació, etc.). També mesurarà el coneixement d'altres llengües estrangeres per part de la població, encara que no en valorarà l'ús. Pel que fa a la població, s'atendrà a les variables de sexe, edat, lloc de naixement propi i dels pares, i —en cas de no ser natural de les Illes Balears— anys de residència a la comunitat autònoma, així com el nivell acadèmic i el grup/nivell professional al qual s'adscriu. La metodologia es detalla en un document tècnic que es difondrà al web de l'IBESTAT, juntament amb les dades obtingudes, les quals s'emmagatzemaran a la base de dades de l'IBESTAT.

Per dur a terme aquesta enquesta es contractarà una empresa per fer els treballs de camp, amb un pressupost de 35.000 euros. La resta de tasques les duran a terme el personal ordinari de l'IBESTAT.

EXERCICI

Emplenau els apartats que corresponguin de la fitxa d'inventari de l'activitat estadística

La fitxa que integra el inventari d'activitats estadístiques s'ha d'elaborar d'acord amb les definicions que s'estableixen en els annexos que s'adjunten del Pla estadístic 2018-2021.

Nota: Posar **NC** als camps que no correspon emplenar, ateses les característiques de l'operació d'acord amb la informació que es dona respecte d'aquest supòsit.

A. Part identificativa de l'activitat

1. Número de l'activitat o de l'operació:
2. Àrea i secció temàtica:
3. Denominació de l'activitat:
4. Òrgan responsable:
5. Departament:

B. Conceptes tècnics

Finalitat

6. Objectius:

7. Població o col·lectiu objecte d'estudi:

Procedència de les dades

8. Organismes o unitats estadístiques de la comunitat autònoma que hi participen:
9. Altres organismes que hi participen:
10. Convenis de col·laboració:
11. Entitats privades i particulars obligats a col·laborar:
12. Compensació:

Part metodològica de producció

13. Tipus d'activitat:
14. Fonts o tipus de l'activitat originària:
15. Forma de recollida de les dades:
16. Periodicitat de recollida de les dades:
17. Emmagatzemament de les dades:
18. Àmbit geogràfic de l'activitat:
19. Dades geolocalitzades:
20. Principals variables que conté:
21. Abast del secret estadístic:
22. Document metodològic:

Difusió

23. Suport en què es difon:
24. Periodicitat de la difusió:
25. Períodes disponibles:
26. Nivell màxim de desagregació territorial que es difon:
27. Desagregació de resultats per la variable sexe:

Elements de planificació

28. Inclòs a Pla:
29. Inclòs a programa:
30. Estat:
31. Cost estimatiu de l'activitat:

32. Observacions:

Únicament en cas d'existir relació amb operacions estadístiques estatals

- Número d'operació:
- Denominació:
- Unitat responsable: