

Facultat de Informàtica, UPC

# Probabilitat i Estadística 2, GCED

Exercicis de Pràctiques

Marta Pérez-Casany i Víctor Peña

13 de setembre de 2024

# Índex

<b>1</b>	<b>Repàs Estimació: Moments i màxima versemblança</b>	<b>3</b>
<b>2</b>	<b>Distribucions: <math>\chi^2</math>, <math>t</math>-d'Student, <math>F</math>-Fisher-Snédecor</b>	<b>4</b>
<b>3</b>	<b>Intervals de confiança</b>	<b>5</b>
<b>4</b>	<b>Tests d'hipòtesis</b>	<b>7</b>
<b>5</b>	<b>Model lineal</b>	<b>10</b>
5.1	Regressió . . . . .	10
5.2	Anova . . . . .	14
5.3	Ancova . . . . .	17
<b>6</b>	<b>Model lineal Generalitzat</b>	<b>19</b>

# 1 Repàs Estimació: Moments i màxima versemblança

- 1) En un punt d'una carretera amb poc tràfic comptem el nombre de vehicles que passen durant 10 minuts, considerem que el recompte segueix una distribució *Poisson* ( $\lambda$ ) d'esperança  $\lambda$  desconeguda.

S'han fet 5 recomptes i s'ha obtingut la mostra  $\mathbf{y} = \{11, 6, 8, 5, 8\}$ .

- Trobeu l'estimador màxim versemblant de  $\lambda$  i comproveu que és únic. Coincideix amb el del mètode dels moments?
- Doneu la distribució asimptòtica de  $\hat{\lambda}$ .

- 2) Suposem que  $T$ , el temps de vida d'una bombeta, té distribució exponencial  $f_T(t) = \alpha e^{-\alpha t}$  amb  $\alpha$  desconeguda. Hem mesurat el temps que han durat 10 bombetes, els resultats obtinguts són  $\mathbf{t} = \{70, 822, 33, 344, 265, 1374, 481, 2494, 901, 909\}$ .

- Trobeu l'estimador màxim versemblant de  $\alpha$  i comproveu que és únic. Coincideix amb el del mètode dels moments?
- Doneu la distribució asimptòtica de  $\hat{\alpha}$ .

- 3) No sabem quina és la probabilitat ( $p$ ) de germinació d'unes llavors, però n'hem comptat el nombre de germinades de 3 safates de 144 alvèols/safata i hem obtingut  $\mathbf{y} = (106, 112, 124)$

- Trobeu l'estimador màxim versemblant de  $p$  i comproveu que és únic. Coincidexia amb el del mètode dels moments?
- Doneu la distribució asimptòtica de  $\hat{p}$ .

- 4) Sigui  $H$  l'alçada de les estudiants (entre 18 i 26 anys), considerem  $H \sim N(\mu, \sigma^2)$  però  $\mu$  i  $\sigma$  són desconegudes. N'hem mesurat 12 i hem obtingut  $\mathbf{H} = \{160, 168, 173, 164, 168, 162, 170, 155, 171, 156, 169, 175\}$ .

- Trobeu l'estimador màxim versemblant de  $\mu$  i  $\sigma$ . Comproveu que  $(\hat{\mu}, \hat{\sigma})$  és únic.
- Doneu la distribució asimptòtica de  $(\hat{\mu}, \hat{\sigma})$ .
- Demostreu que l'estimador màxim versemblant té biaix.

## 2 Distribucions: $\chi^2$ , $t$ -d'Student, $F$ -Fisher-Snédecor

1) Demostreu analíticament que la distribució Fisher-Snédecor compleix que:

$$F_{\alpha, n_1, n_2} = (F_{1-\alpha, n_2, n_1})^{-1}$$

2) Distribució  $\chi^2$

- Genereu 100 dades d'una  $\chi^2_{15}$ . Calculeu-ne la mitjana aritmètica i la variància i comproveu que estan properes als valors teòrics. Feu-ne l'histograma.
- Feu el mateix que abans amb una  $\chi^2_{50}$ . Dibuixeu l'histograma i el plot de densitat i compareu-lo amb el d'una Normal amb els paràmetres corresponents. Què deduiu?
- Calculeu la probabilitat que una distribució  $\chi^2_{15}$  acumula per sota del punt 7.3.
- Calculeu el punt que deixa una probabilitat per sobre igual a 0.2 en una distribució  $\chi^2_{15}$ .

3) Distribució  $t$  –  $d'$ Student

- Genereu 100 dades d'una  $t$ -d'Student<sub>13</sub>, i dibuixeu-ne l'histograma i el plot de densitat.
- Calculeu la probabilitat que una  $t$ -d'Student<sub>8</sub> acumula per sota del punt 3.5.
- Calculeu el punt que acumula una probabilitat per sota igual a 0.8 en una distribució  $t$ -d'Student<sub>8</sub>.

4) Distribució  $F$  – Fisher – Snédecor

- Genereu 100 dades d'una  $F$ -Fisher<sub>3,6</sub>, i dibuixeu-ne l'histograma i el plot de densitat.
- Calculeu la probabilitat que una  $F$ -Fisher<sub>3,6</sub> acumula per sota del punt 3.5.
- Calculeu el punt que acumula una probabilitat per sota igual a 0.8 en una distribució  $F$ -Fisher<sub>3,6</sub>.
- Comproveu que  $F_{0.05, 5, 10}$  és igual al recíproc de  $F_{0.95, 10, 5}$ .
- Quin és el punt que acumula un 50% de probabilitat per sota en una distribució  $F_{10, 10}$ ? i en una  $F_{25, 25}$ ?

### 3 Intervals de confiança

- 1) Després d'un tractament contra la obesitat, els pesos en Kg de vuit dones eren 58, 50, 60, 65, 64, 62, 56, 57. Suposem normalitat.
  - Trobeu un interval de confiança al 95% pel valor esperat de pes quan  $\sigma = 3$ .
  - Trobeu un interval de confiança al 90% pel valor esperat del pes si  $\sigma$  és desconeguda.
  - Trobeu un interval de confiança per la variància del pes.
  - Repetiu els dos apartats anteriors amb R
- 2) Es calcula que la quantitat de nicotina que tenen els cigarrets d'una determinada marca segueix una distribució  $N(30, \sigma^2)$ . Agafem deu cigarrets d'aquesta marca a l'atzar, i després d'analitzar-ne la nicotina obtenim que  $\frac{1}{10} \sum_{i=1}^{10} (X_i - 30)^2 = 12.4$ . Trobeu un IC per  $\sigma$  amb un nivell de confiança del 90%.
- 3) Es fan 5 mesures de la quantitat d'argent d'un mineral i s'obté 5.2, 4.8, 5.3, 5.7 i 5.0 mg d'argent. Determineu intervals de confiança a nivell 0.95 del valor esperat (quantitat d'argent) i de la variància (precisió de l'experiment), suposant normalitat.
- 4) Considerem una mostra de mida  $n$  d'una Normal amb valor esperat  $\mu$  i variància  $\sigma^2 = 16$ . Trobeu el menor valor de  $n$  per tal que  $(\bar{x} - 1, \bar{x} + 1)$  sigui un IC per a  $\mu$  a nivell 95%.
- 5) En una enquesta feta als Estats Units l'any 2021, es va preguntar a 235 persones tatuades si se'n penedien de fer-se els seus tatuatges. El resultat va ser que 28 persones van dir que se'n penedien, mentre que la resta va dir que no. Trobeu un interval de confiança a nivell 95% sobre la proporció poblacional de tatuats que se'n penedeixen de tatuar-se.
- 6) Una empresa vol fer un estudi per estimar una proporció poblacional. Volen que l'interval de confiança tingui una longitud de, com a molt, 0.2. Quina ha de ser la mida mostral mínima per a assegurar-nos que la condició es compleixi segur? Justifica la resposta.
- 7) A partir d'una mostra de 10 observacions d'una població Normal, es calcula un IC al 95% per la variància de la població i dona (0.5, 3.57). Si la mitjana aritmètica de la mostra és 4.92, dedueu l'extrem superior de l'IC pel valor esperat de la població al 95%.
- 8) Un material que prové de dos proveïdors que anomenem  $A$  i  $B$ , s'empaqueta en caixes. Els pesos de les caixes de  $A$  presenten una distribució  $N(\mu_1, 0.07^2)$  i els de la  $B$  una distribució  $N(\mu_2, 0.04^2)$ . Una mostra de 100 caixes de  $A$  dona una mitjana aritmètica de 0.99 kg, i una mostra de 300 caixes de  $B$  dona una mitjana aritmètica de 1.01 kg. Determineu un IC per a  $\mu_1 - \mu_2$  al 95%.
- 9) En la mateixa situació que al problema anterior, imagineu-vos que les variàncies dels pesos de les caixes de  $A$  i de  $B$  son desconegudes. Si es sap que per les dades de  $A$ ,  $\sum_{i=1}^{100} (x_i - \bar{x})^2 = 225.0$  i per les dades de  $B$   $\sum_{i=1}^{300} (y_i - \bar{y})^2 = 180.3$ ,
  - Estimeu  $\sigma_A^2$  i  $\sigma_B^2$  amb l'estimador màxim versemblant.

- Trobeu un IC al 99% per a  $\sigma_A^2$  i per a  $\sigma_B^2$ .
  - Si es suposa que les variàncies son iguals, trobeu una estimació raonable per aquesta variància comuna a partir de totes les dades.
  - Trobeu un IC pel quocient de les dues variàncies ( $\sigma_A^2/\sigma_B^2$ ).
- 10) Considera que l'alçada dels habitants d'un país pot modelitzar-se segons una v.a.  $X \sim N(\mu = 175, \sigma^2 = 10^2)$ . Utilitzant *R* porteu a terme els següents apartats:
- Genereu aleatòriament una mostra de mida 20 de *X*
  - Feu el "summary" de les dades i calculeu-ne la mitjana aritmètica, la variància i la desviació estàndard.
  - Calculeu a partir de la mostra i fent servir l'*R* com a calculadora, l'IC per  $\mu$  al 95% suposant  $\sigma^2$  coneguda.
  - Calculeu ara l'interval al 99%. Com expliqueu que ara aquest interval sigui més ampla?
  - Calculeu l'IC al 95% per  $\mu$  suposant que  $\sigma^2$  és desconeguda. Feu-ho utilitzant *R* com a calculadora i amb la comanda "t.test" de l'*R*.
  - Fent servir les dades del fitxer `carprice.csv`, a) feu el boxplot del preu del cotxe segons l'estat civil de la persona. Això ens permetrà intuir si els solters compren cotxes de preu més elevat que la resta, b) calculeu un IC per la diferència de preus d'un cotxe en els dos grups fent servir l'*R* com a calculadora, i fent servir la comanda "t.test".
- 11) La longitud d'un cert producte es pot modelar segons una llei normal. Imaginem-nos que els paràmetres son  $\mu = 3m$  i  $\sigma = 1.9m$ . Genereu 10 valors d'aquesta variable, i porteu a terme els següents apartats:
- Calcula un IC al 95% per  $\mu$ . Inclou aquest interval el veritable valor?
  - Compteu quants IC dels calculats a la classe contenen el veritable valor. S'aproxima al 95%?
  - Repetiu els apartats anteriors però ara amb una confiança del 50%.
  - L'interval de l'apartat anterior és més ampla o més estret que el del 95%? Perquè?
  - Calculeu l'IC al 95% per la variància, a) conté el veritable valor?, b) quina diferència principal hi ha entre aquest itnerval i els calculats per a l'esperança?
- 12) Una organització de consumidors va obtenir dades de 412 compradors de televisors de les marques A i B. En un període de dos anys, van recollir el nombre d'aparells que van necessitar reparacions. Dels aparells de la marca A, 111 van ser reparats i 162 no. En quant als aparells de la marca B, 85 van ser reparats i 54 no. Trobeu un interval de confiança al nivell 95% per la diferència de proporcions d'aparells que van ser reparats per les marques A i B. L'interval conté 0 o no? Comenteu els resultats.

## 4 Tests d'hipòtesis

Feu els següents exercicis a ma i amb l' $R$ .

- 1) S'ha fet un seguiment de iogurts en la fase de comercialització, per tal de comparar l'efecte de la temperatura de la fermentació en diverses variables.

Unitats experimentals: Unitats de iogurt analitzades alguns dies després de la fermentació.

Variables explicatives:

- *Dies*, que conté els dies transcorreguts entre la fermentació i l'anàlisi del iogurt.
- *Grups*, el grup queda determinat per la temperatura de fermentació. S'han considerat dues temperatures:  $42^\circ$  o  $43.5^\circ$ .

Variables resposta mesurades durant l'experiment:

- *pH*, conté el valor del ph del iogurt en el moment de fer l'anàlisi.
- *Strep.*, conté la concentració de *Streptococcus salivarius thermophilus* en el moment de fer l'anàlisi.
- *Lactob.*, conté la concentració de *Lactobacillus delbrueckii bulgaricus* en el moment de fer l'anàlisi.

El dia de la fermentació a temperatura  $42^\circ$  les mesures del *pH* han sigut  $\{4.44, 4.47, 4.45, 4.42, 4.44, 4.48\}$ . Suposant que  $pH \sim N(\mu, \sigma^2)$  amb  $\mu$  i  $\sigma$  desconeguts i amb  $\alpha = 0.05$ , plantegeu i porteu a terme els tests corresponents per tal de contestar les preguntes següents:

- a) El seu valor esperat és 4.4? o no?
- b) El seu valor esperat és inferior a 4.4? o no?
- c) La seva desviació tipus és 0.02? o no?

- 2) Es posen 12 llavors a germinar en unes determinades condicions experimentals. Transcorregut un temps, s'observa que 7 d'elles han germinat. Trobeu: a) l'estimació de la probabilitat  $p$  de germinació sota les esmentades condicions, 2) l'IC aproximat per aquest probabilitat, c) Podem dir que aquesta probabilitat és igual a  $1/2$ ?
- 3) La velocitat de dissolució (en mg/s) d'unes pastilles efervescents en aigua té una desviació tipus poblacional  $\sigma = 2\text{mg/s}$ . Es vol comprovar si una reformulació dels ingredients de les pastilles afecta la velocitat de dissolució. Es vol ser capaç de detectar diferències de  $1\text{mg/s}$ . Si volem tenir els errors  $\alpha, \beta$  controlats i iguals a 0.05 quina és la grandària mínima de mostra que necessitem?
- 4) Continuant amb l'exercici 1), suposeu que a més de les mesures del *pH* dels iogurts fermentats a  $42^\circ$  tenim la dels fermentats a  $43.5^\circ$  que són  $\{4.38, 4.37, 4.33, 4.34, 4.31, 4.39\}$ . Suposant que  $pH|T \sim N(\mu_T, \sigma_T^2)$  amb  $\mu_T$  i  $\sigma_T$  desconeguts i amb  $\alpha = 0.05$ , plantegeu i porteu a terme els tests corresponents per tal de contestar les preguntes següents:
- a) Les dues esperances són iguals o no? (suposem variàncies iguals)

- b) Podem assegurar que el  $pH$  és més alt al fermentar a  $42^\circ$ ? (suposem variàncies iguals)
- c) Les variàncies del  $pH$  a les dues temperatures són iguals o no?
- d) Si a l'apartat anterior us ha sortit que no ho son, repetiu la comparació de valors esperats i treieu-ne conclusions.
- 5) S'ha fet un tast per conèixer les preferències dels consumidors respecte a dos vins que denotem, respectivament per A i B. Cada tastador puntua els dos vins. Les puntuacions atorgades han estat les següents:

Consumidor	1	2	3	4	5	6	7	8	9	10
A	5.3	4.1	5	5.7	5.8	7.2	3.4	5	6.2	5.8
B	4.7	5	6.2	5.6	6.4	7.9	4.4	5.5	6.5	5.7

- a) Amb  $\alpha = 0.05$ , feu el test escaient per tal de saber si hi ha preferència o no per algú dels dos vins.
- b) Especifiqueu quines assumpcions heu fet per tal de poder fer el test anterior.
- 6) Continuant amb l'exercici dels Iogurts i suposant que la concentració d'*Streptococcus* segueix una distribució Normal per a cada temperatura de fermentació, porteu a terme els contrastos necessaris per a contestar les dues preguntes següents:
- a) La variabilitat de la concentració d'*Streptococcus* es pot assumir igual en les dues temperatures de fermentació?
- b) En base al resultat obtingut a l'apartat anterior, digueu si és possible afirmar que quan la temperatura és major, la concentració d'*Streptococcus* també serà major.
- 7) Es sotmet un grup de 23 hamsters al fum del tabac, mentre que un altre grup format per 32 hamsters és considerat control (no s'exposa a les condicions de l'anterior). Després d'un període d'un any s'observa que 21 hamsters del primer grup desenvolupen càncer de pulmó mentre que del grup control el desenvolupen un total de 19. Calculeu: a) L'estimació de la diferència de desenvolupar càncer de pulmó en els dos grups, b) l'estimació de la desviació estàndard de l'anterior diferència, c) l'IC per a la diferència de les dues probabilitats, d) A partir de l'IC podeu deduir que les probabilitats son estadísticament diferents? Perquè? d) Porteu a terme la comparació  $H_0 : p_1 = p_2$  vs  $H_1 : p_1 \neq p_2$ .
- 8) Bishop va estudiar la mortalitat dels nadons durant el primer més, en funció de l'atenció pre-natal que havia tingut la mare. Concretament es van dividir les mares en dos grups, en funció de si havien rebut una atenció pre-natal de menys o de mes d'un mes. La proporció de mortalitat dels nadons per les mares que havien rebut menys d'un mes d'atenció pre-natal va ser de 20/393 mentre que la proporció pel grup que havia rebut més d'un mes d'atenció pre-natal va ser de 6/322. Podem dir que la probabilitat de que el nadó mori si s'ha tingut menys d'un mes d'atenció pre-natal és superior a la de l'altre grup?.
- 9) En l'exercici 1) teníem:  
*En un punt d'una carretera amb poc tràfic comptem el nombre de vehicles que passen durant 10 minuts, considerem que el recompte segueix una distribució Poisson ( $\lambda$ ) d'esperança  $\lambda$  desconeguda. S'han fet 5 recomptes i s'ha obtingut la mostra  $y = \{11, 6, 8, 5, 8\}$ .*  
Aquests recomptes s'havien fet en dies sense pluja. S'han fet també recomptes en dies de pluja, obtenint-se  $z = \{8, 11, 7, 10\}$ .

- a) A través del test de raó de versemblança, digueu si és possible assumir que els dies que no plou, les dades segueixen una Poisson de paràmetre igual a 6.5.



- b) Estimeu per màxima versemblança el nombre esperat de vehicles els dies de pluja.
- c) Utilitzant el test de raó de versemblança, amb  $\alpha = 0.05$ , dieu si passen el mateix nombre de cotxes tant els dies de pluja com els dies que no plou.

**10)** En l'exercici 3) teníem:

*No sabem quina és la probabilitat ( $p$ ) de germinació d'unes llavors, però n'hem comptat el nombre de germinades de 3 safates de 144 alvèols/safata i hem obtingut  $\mathbf{y} = (106, 112, 124)$ .*

Tractàvem les tres safates com si les llavors fossin les mateixes, però sabem que abans de l'experiència s'havien conservat de formes diferents que anomenarem  $\mathbf{C} = (C1, C2, C3)$ .

- a) Per a cada tipus de conservació estimeu per màxima versemblança les probabilitats de germinació  $p = (p_1, p_2, p_3)$
- b) Utilitzant el test de raó de versemblança contesteu la pregunta: *El tipus de conservació té algun efecte sobre la probabilitat de germinació?*
- c) En cas que tingui efecte, dieu quines diferències hi ha entre les probabilitats segons el tipus de conservació.

## 5 Model lineal

### 5.1 Regressió

- 1) En un experiment de retenció de memòria es va demanar a 13 participants que memoritzessin una llista de paraules. A continuació, els participants havien d'intentar recordar totes les paraules de la llista en diferents moments. Les dades estan desades al fitxer `wordrecall.csv`. Al fitxer, hi trobareu la proporció de paraules recordades correctament (prop) en diversos intervals de temps des que es va memoritzar la llista (time, en minuts).
  - a) Grafiqueu la proporció a l'eix  $y$  i el temps a l'eix  $x$ . Creieu que té sentit ajustar un model lineal?
  - b) Transformeu la variable temps amb la funció logaritme. Repetiu el gràfic de l'apartat anterior amb el logaritme del temps. Què hi veieu? Té sentit ajustar un model lineal ara?
  - c) Ajusteu un model lineal (amb o sense transformació del temps, segons el vostre criteri). Interpreteu l'ordenada a l'origen ajustada i el pendent ajustat. Comenteu la  $R^2$  del model i la seva bondat d'ajust tot mirant els residus.
  - d) Feu una predicció puntual i un interval de predicció a nivell 95% de la proporció de paraules que un individu recorda passades 3h des que va veure la llista.
- 2) En un estudi per estudiar el nivell de colesterol en sang, en persones en edat de creixement (9-20 anys), s'ha plantejat un model per explicar el nivell de colesterol ( $C$ ) en funció del pes ( $P$ ), edat ( $E$ ) i alçada ( $H$ ). És conegut que el pes és un dels factors que influeixen en l'augment del nivell de Colesterol, i es vol aprofundir en aquest fet. Les dades experimentals són independents i estan en el fitxer "`col.csv`".
  - a) Per començar plategeu la recta de regressió del colesterol respecte el pes.
    - (1) Trobeu l'estimació dels coeficients de la recta de regressió i de la variància del error, i interpreteu-los.
    - (2) Dieu si el pes explica, en part, la variabilitat en el nivell de colesterol. Veieu alguna contradicció en els resultats?
    - (3) Dibuixeu, juntament amb les dades la banda de confiança del 99 juntament amb la de predicció del nivell de colesterol en funció del pes.
    - (4) Utilitzant l'anàlisi dels residus, dieu si el model es pot acceptar com a correcta o no.

Nota: per entendre millor el que passa, feu la gràfica (amb recta de regressió) del colesterol respecte el pes agrupats per edats, podeu fer-la amb la comanda:

```
scatterplot(C~P|E,smooth=F,col=1:20,data=...)
```

Quines conclusions treieu del gràfic anterior?

- 3) Els potenciòmetres (PH-metre) son sensors utilitzats en el mètode electroquímic per tal de medir el PH d'una dissolució. Ara bé, si es canvien els electrodos també es pot utilitzar per a mesurar la presència

d'altres ions o molècules. L'*electrode sensitiu d'amoniac* permet detectar a través de la seva membrana l'amoniac en forma de gas. En aquest cas, les lectures (L) que dona el potenciòmetre es relacionen amb la concentració de nitrogen amoniacal (C) seguint el model següent:

$$L = \beta_0 + \beta_1 \log(C).$$

S'ha fet un experiment de lectures a diferents concentracions de nitrogen amoniacal i els resultats han estat els següents:

C (en ppm)	2	5	10	20	50	100
L (en mV)	85	107	124	140	161	179

Ajusteu el model de regressió lineal simple associat al model anterior, i contesteu les següents preguntes per a un  $\alpha = 0.05$ :

- Quines son les estimacions dels paràmetres obtingudes? son significatius ambdós paràmetres?
- Interpreteu els dos paràmetres.
- Estimeu la variància del model.
- Creieu que hi ha *outliers* o punts influents?
- Feu l'anàlisi de residus i treieu-ne conclusions.
- Reformuleu el model si ho creieu oportú.
- En base al vostre model, digueu quina és la concentració que dona una lectura de 167 mV?
- Per a la concentració anterior, trobeu l'interval de predicció i l'interval de confiança per a la lectura estimada.

4) Amb les mateixes dades de l'exercici anterior del colesterol, plantegeu ara la regressió múltiple següent:

$$C_i = \beta_0 + \beta_1 P_i + \beta_2 H_i + \beta_3 E_i + \varepsilon_i \quad \text{amb } \varepsilon_i \sim N(0, \sigma^2)$$

- Trobeu l'estimació dels coeficients de l'hiperplà de regressió i de la variància del error, i interpreteu-los.
- Dieu si alguna de les variables ( $P, H, E$ ) explica, en part, la variabilitat existent entre els nivells de colesterol de les persones.
- Hi ha col·linealitat entre les variables explicatives?
- Utilitzant diagnòstics dieu si el model es pot acceptar com a vàlid.
- Imagineu-vos ara que es vol veure si enlloc del pes de la persona, és l'excés de pes el que afecta el nivell de colesterol. Es sap que pel rang d'edats en que ens movem, el pes d'una persona en funció de la seva alçada es modela de forma satisfactòria amb la següent recta de regressió:

$$P = -10 + 0.5H.$$

Calculeu el valor de la variable excés de pes ( $EP$ ) com la diferència entre el pes real de la persona i el que s'esperaria per la seva alçada, i plantegeu la regressió múltiple

$$C_i = \beta_0 + \beta_1 EP_i + \beta_2 H_i + \beta_3 E_i + \varepsilon_i \quad \text{amb } \varepsilon_i \sim N(0, \sigma^2)$$

- (1) Trobeu l'estimació els coeficients de la regressió i de la variància del error i interpreteu-los.
  - (2) Quines diferències hi ha respecte la regressió múltiple anterior? Les prediccions són iguals?
  - (3) Dieu si alguna de les variables ( $EP, H, E$ ) explica de forma significativa la variabilitat existent en el nivell de colesterol.
  - (4) Hi ha col·linealitat entre les variables explicatives d'aquest model?
  - (5) Utilitzant l'anàlisi de residus dieu si el model és pot considerar correcte.
  - (6) Si es pot, simplifiqueu el model anterior i analitzeu-lo.
  - (7) Amb el model més simplificat
    - A. Pels que no tenen excés de pes, dibuixeu la banda de predicció (PI) del colesterol 95% respecte l'edat.
    - B. Supposem tres persones amb  $P = \{60, 65, 70\}$ ,  $E = 15$  i  $H = 150$ , i denotem per  $\mu_1$  el valor esperat del nivell de colesterol d'una persona que compleixi:  $P = 60, E = 15$ , i  $H = 150$ ,  $\mu_2$  el d'una persona que compleixi  $P = 65, E = 15$ , i  $H = 150$ ,  $\mu_2$ , i  $\mu_3$  el d'una persona que compleixi:  $P = 70, E = 15$ , i  $H = 150$ . Utilitzant el model simplificat, calculeu:
      - Els tres PI del colesterol al 95%.
      - Els 3 intervals de confiança,  $IC_{95\%}$ , pel nivell de colesterol esperat per a cadascun dels perfils.
      - Amb un nivell de significació del 5%, contrasteu els 3 tests:
        - $H_0 : \mu_1 = 200$  vs  $H_1 : \mu_1 \neq 200$ .
        - $H_0 : \mu_2 = 200$  vs  $H_1 : \mu_2 \neq 200$ .
        - $H_0 : \mu_3 = 200$  vs  $H_1 : \mu_3 \neq 200$ .
- 5) Una empresa que gestiona l'arbrat de diverses localitats de l'Àrea metropolitana de Barcelona, rutinàriament controla el diàmetre de la capçada ( $DCap$ ) dels lledoners.
- Donat que és molt més costos mesurar el  $DCap$  que d'altres mesures més simples, es proposen predir el  $DCap$  en funció de les mesures que son més fàcils de recollir. En aquest sentit es tenen en compte les següents variables, mesurades en metres les tres primeres i en anys la darrera,:  $PT$  perímetre del tronc a 1,5m d'alçària,  $PB$  perímetre de la base,  $HT$  alçada del tronc,  $E$  edat de l'arbre. S'ha obtingut una mostra que conté les mesures reals de les cinc variables en 311 lledoners. Les dades estan al fitxer `dcap.csv`.
- Primer de tot comproveu que els perímetres  $PT$  i  $PB$  estan molt correlacionats. Per tal d'evitar problemes de col·linealitat s'ha definit la variable *relació dels perímetres*  $RP = PB/PT$  que de fet és un indicador de la forma del tronc.
- S'han planejat tres possibles models per explicar l'evolució de la variable  $DCap$  en funció de les altres variables:
- modA*: Regressió lineal de la variable aleatòria  $DCap$  respecte les variables explicatives  $PT$ ,  $RP$ ,  $HT$  i  $E$ .
- modB*: Regressió lineal de la variable aleatòria  $\log(DCap)$  respecte les variables explicatives  $\log(PT)$ ,  $\log(RP)$ ,  $\log(HT)$  i  $\log(E)$ .
- modC*: Regressió lineal de la variable aleatòria  $\log(DCap)$  respecte les variables explicatives  $PT$ ,  $RP$ ,  $HT$  i  $E$ .

- a) Ajusteu el *modA*, suposant que es compleixen les condicions dels models lineals, i contesteu les preguntes següents:
- (1) Els coeficients de les 4 variables explicatives, amb  $\alpha = 0.05$ , són significativament diferents de 0?
  - (2) Hi ha col·linealitat?
  - (3) Quina és l'estimació de la variància residual?
  - (4) Les tendències que s'observen a la gràfica dels residuals front dels predits són les adequades? I les de la gràfica qqnorm?
  - (5) Mirant el residuals studentitzats, quantes observacions surten fora de l'interval  $(-2, 2)$ ? Quin percentatge d'elles representa? És raonable aquest percentatge?
- b) Ajusteu el *modB*, suposant que es compleixen les condicions dels models lineals, i contesteu les preguntes següents:
- (1) Els coeficients de les 4 variables explicatives, amb  $\alpha = 0.05$ , són significativament diferents de 0?
  - (2) Hi ha col·linealitat?
  - (3) Quina és l'estimació de la variància residual?
  - (4) Les tendències que s'observen a la gràfica dels residuals front dels predits són les adequades? I les de la gràfica qqnorm?
  - (5) Mirant el residuals studentitzats, quantes observacions surten fora de l'interval  $(-2, 2)$ ? Quin percentatge d'elles representa? És raonable aquest percentatge?
- c) Ajusteu el *modC*, suposant que es compleixen les condicions dels models lineals, i contesteu les preguntes següents:
- (1) Els coeficients de les 4 variables explicatives, amb  $\alpha = 0.05$ , són significativament diferents de 0?
  - (2) Hi ha col·linealitat?
  - (3) Quina és l'estimació de la variància residual?
  - (4) Les tendències que s'observen a la gràfica dels residuals front dels predits són les adequades? I les de la gràfica qqnorm?
  - (5) Mirant el residuals studentitzats, quantes observacions surten fora de l'interval  $(-2, 2)$ ? Quin percentatge d'elles representa? És raonable aquest percentatge?
- d) Entre els models *modA*, *modB* i *modC*, quin creieu que és millor? Per què?
- e) Amb el model escollit, calculeu la predicció, i l'interval de predicció del 95%, del Dcap dels lledoners de 10 anys amb  $PT = 0.4$ ,  $PB = 0.6$  i  $HT = 2.3$ .
- 6) Es vol predir la variable *Temps en recórrer 1600 metre* dels jugadors professionals de rugby. A tal afecta s'han mesurat el *Pes*, l'*Alçada*, el *Pes que és capaç d'aixecar* i el *Temps en recórrer 400 m* d'un total de 30 jugadors. Les dades es troben en el fitxer RUGBY2.csv
- a) Feu l'anàlisi exploratòria de dades i treieu-ne conclusions.
  - b) Ajusteu el model lineal que conté totes les explicatives i aneu-lo retocant fins a obtenir un model que ajusti de forma satisfactòria les dades. Interpreteu els paràmetres del model obtingut.
  - c) Trobeu l'IP al 90% pel temps en recórrer 1600m d'un jugador de rugby que medeixi 1,80, pesi 87kg, sigui capaç d'aixecar 130 kg, i hagi trigat 75 segons en recórrer 400 metres. Trobeu un IC al 95% pel valor esperat de la variable resposta per algú amb el mateix perfil que el jugador anterior.

- 7) (Starbucks) Treballarem amb un conjunt de dades que conté informació nutricional sobre el menjar que es ven a la cadena Starbucks. Les dades les trobareu al fitxer `starbucks.csv`. Les variables són el nom del producte, les calories, els grams de greix, fibra i proteïna, i el tipus de producte (*bakery* o *other*).
- Feu gràfics bivariants entre les calories i les altres variables numèriques de la base de dades. Comenteu els resultats.
  - Ajusteu un model que prediu les calories donades totes les variables a la base de dades excepte `item` i `type`. A partir d'aquest model, aneu reduint el nombre de variables una a una fins trobar un model en què tots els coeficients siguin significatius a nivell de significació  $\alpha = 0.05$ . Amb quines variables us quedeu?
  - Interpreteu el model que heu triat a l'apartat anterior. Concretament, interpreteu-ne els coeficients.
  - Identifiqueu l'observació amb el residu més gran (en valor absolut). Per què creieu que té el residu més gran?
  - Feu servir el model ajustat amb les dades de Starbucks per predir les calories dels següents productes de McDonalds: Big Mac (28g de greix, 45g d'hidrats de carboni, 3g de fibra, 24g de proteïna), Cheeseburger (11g de greix, 33g d'hidrats, 2g de fibra, 15g de proteïna), galeta de xocolata (7g de greix, 22g d'hidrats, 1g de fibra, 2g de proteïna). En realitat, les calories dels productes són 520 (Big Mac), 290 (Cheeseburger) i 160 (galeta). Creieu que el model està funcionant bé?
  - Una nutricionista diu que cada gram de greix té 9 calories i cada gram d'hidrats o proteïna en té 4. Creieu que el vostre model està d'acord amb la nutricionista? Justifiqueu la resposta.

## 5.2 Anova

- Per veure si afegir edulcorant millora l'engreix de garrins s'ha fet una experiència en la que s'ha mesurat el guany mig diari en pes, *GMD*, de garrins en les mateixes condicions però afegint edulcorant. S'han provat les 5 dosis: *D00*, *D08*, *D15*, *D20* i *D30*. La dosi *D00* és de fet la dieta sense edulcorant. Cada dieta s'ha provat en 5 garrins, i els resultats experimentals obtinguts els podeu trobar al fitxer "`gmd.csv`". Tractant la dosis com els nivells d'un factor i prenent  $\alpha = 0.05$ , calculeu:
  - Les matrius dels següents models lineals:
    - (1)  $y_{ij} = \mu_i + \varepsilon_{ij}$
    - (2)  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$  amb  $\sum_{i=1}^k \alpha_i = 0$
    - (3)  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$  amb  $\alpha_1 = 0$
  - Quina és la pregunta que té sentit contestar en aquesta situació? Escribiu-la i plantegeu-ne el test d'hipòtesis corresponent.
  - Porteu a terme la Taula Anova que us permet contestar la pregunta formulada i treieu-ne conclusions.
  - Utilitzant el mètode de Tukey i utilitzant el paquet de R `emmeans`,
    - Per cada nivell de dosi, estimeu els valors esperats del *GMD* i el seu interval de confiança (aquest no depen de la parametrització escollida).
    - Feu les comparacions dos a dos de totes les dosis entre elles. Dieu quines parelles de nivell de dosi, donen un GMD esperat diferent.
    - Expresseu el resultat anterior de forma compacta.

- 2) En un tast de patés, cada persona del panell ha tastat i puntuat 5 patés. Per a no influir sobre el sobre les persones del panell, els patés es varen presentar de forma aleatòria i identificats amb codis numèrics que indueixin a ser ordenats. En el fitxer “`pate.csv`” hi ha els resultats del tast, les seves columnes són:

- tastador
- codi del paté
- les valoracions de la qualitat de: color, aroma, textura, sabor i la ordenació.

Les puntuacions de les valoracions són de 0 a 10 i l'ordenació és en ordre de preferència (de 1 a 5).

- a) Calculeu la *valoració global* com la mitjana de color, aroma, textura i sabor. Per la valoració global, plantegeu el model lineal amb el factor paté com a variable explicativa. Suposant que es compleixen les condicions per fer el test anova, amb  $\alpha = 0.05$  i contrastant els tests adequats, contesteu:
- (1) El tipus de paté té algun efecte significatiu sobre la valoració?
  - (2) Si és el cas que sí, feu-ne les comparacions múltiples i interpreteu els resultats.
  - (3) Creus que es compleixen les condicions del test anova? Justifiqueu-ho teòricament i amb gràfics de diagnòstic.
- b) També per la valoració global, plantegeu el model lineal additiu amb els factors tastador i paté. Suposant que es compleixen les condicions per fer el test anova, amb  $\alpha = 0.05$  i contrastant els tests adequats, contesteu:
- (1) El tipus de paté i/o el tastador son variables que tenen algun efecte significatiu sobre la valoració?
  - (2) Si el tipus de paté té un efecte significatiu, feu-ne les comparacions múltiples i interpreteu els resultats.
  - (3) Compareu els resultats amb els obtinguts a amb el model de només factor paté sense tenir en compte els tastadors.
  - (4) Creus que es compleixen les condicions del test anova? Justifiqueu-ho teòricament i amb gràfics de diagnòstic.
  - (5) Escriviu a mà la matriu del model, i comproveu que és la matriu que utilitza l'R.
- c) (Optatiu) Repetiu les anàlisis anteriors per les valoracions de color, aroma, textura i sabor per separat.
- d) Per la variable ordenació dieu perquè no es compleixen les condicions dels models lineals.
- 3) En l'elaboració de formatges, el rendiment és la relació entre el pes del formatge obtingut i el de la llet utilitzada. Es vol veure si utilitzar llet crua o pasteuritzada (pasteuritzar la llet consisteix en aplicar un tractament tèrmic de la llet) canvia el rendiment i també si ho fa l'addició de  $\text{CaCl}_2$  (si/no) a la llet. Les dades són al fitxer “`formatges.csv`”. Observis que es tenen dades per a tres diferents tipus de llet: cabra, vaca i ovella. Per als formatges elaborats amb llet de cabra, suposant que es compleixen les hipòtesis del model lineal i amb  $\alpha = 0.05$ , plantegeu el model factorial (model amb interacció) i contesteu les preguntes següents:
- a) Tenen algun efecte en el rendiment els dos factors considerats?
  - b) La interacció és significativa? Interpreteu el resultat obtingut.
  - c) Si definim com *tractament* una combinació dels nivells dels dos factors considerats, contesteu les preguntes següents:

- 1) Quin és el millor o o son els millors tractaments?
- 2) El  $\text{CaCl}_2$  té algun efecte sobre el rendiment?
- 3) Podem dir que afegir  $\text{CaCl}_2$  fa augmentar el rendiment?
- 4) El tractament tèrmic té algun efecte sobre el rendiment?
- 5) Podem dir que pasteuritzar fa augmentar el rendiment?
- 6) Podríem utilitzar un altre model lineal de dimensió més petita? per què? quin? canviarien els resultats del test anova? i els de les comparacions múltiples?

Si ho creieu possible i convenient, simplifiqueu el model i contesteu les mateixes preguntes que a l'apartat anterior.

- 4) Repetiu les anàlisis de l'exercici anterior amb els altres tipus de llet.
- 5) Es volen cobrir determinades àrees dels talusos de la carretera amb plantes. Per tal de decidir com fer-ho, es volen estudiar dos tipus de plantes diferents, i també es vol analitzar si posant-hi compost aquestes cobreixen més àrea o no. Les dades experimentals de l'estudi es troben al fitxer "area.csv". Per a un  $\alpha = 0.05$ , porteu a terme les següents anàlisis:
  - a) Especifiqueu quina té sentit que sigui la variable resposta i quines les explicatives. Especifiqueu quines tres preguntes té sentit plantejar-se en aquesta situació.
  - b) Plantegeu els dibuixos que creieu convenient per tal d'intuir si hi ha o no interacció entre els dos factors del model.
  - c) Plantegeu el model additiu (sense interacció) associat a les variables anteriorment esmentades, i contesteu les següents preguntes:
    - Son significatives les dues variables explicatives o factors?
    - Què diu la taula anova? Interpreteu-la.
    - Porteu a terme les comparacions múltiples. Què dedueix?
    - Porteu a terme l'anàlisi de residus per tal de decidir si el model és adequat o no.
  - d) Plantegeu el model lineal additiu però ara considerant que la variable resposta és el logaritme de l'àrea.
    - Ajusteu el model i digueu si son significatius o no els dos factors.
    - Què concluiu del test anova?
    - Afegiu-hi el terme corresponent a la interacció i estudeu si és significativa o no.
    - Feu les comparacions múltiples corresponents.
    - Acompanyeu el que heu dit en l'apartat anterior amb un dibuix que ho recolzi.
    - Feu l'anàlisi de residus i justifiqueu si el model es pot donar com a vàlid o no.
    - En base a aquest model, com canvia l'àrea coberta esperada si enlloc de fer servir l'espècie 1 sense compost faig servir l'espècie dos amb compost?
- 6) Es vol estudiar si la degradació de les fariens de blat depèn de la varietat, i també si es veu afectada per la presència d'insectes a la farina. Les dades de l'estudi experimental es troben en el fitxer "blat.csv". En aquest fitxer, VAR indica la varietat de blat, W2 indica el valor de la degradació i PRES indica si hi ha presència o no d'insectes a la farina.



- a) Ajusteu un model factorial (amb interacció), i amb un  $\alpha = 0.05$ , contesteu les preguntes següents:
  - La degradació es veu afectada pel tipus de blat i per la presència o no d'insectes?
  - Creieu que els dos factors interaccionen? Com s'interpreta la interacció en aquest context?
  - Porteu a terme l'anàlisi de residus. Què concloeu?
- b) Ajusteu el model sense interacció i compareu-lo amb el model anterior.

## 5.3 Ancova

- 1) Es volen comparar les valoracions obtingudes seguint dos mètodes pedagògics ( $M$ ) diferents. També es vol saber la influència del coeficient d'intel·ligència ( $C$ ) en la valoració. Les dades reals obtingudes estan al fitxer "correct.csv". Observis que hi ha dues columnes corresponent a valoracions, la  $V$  i la  $VV$ .
  - a) Plantegeu i ajusteu el model lineal factorial (amb interacció) de la variable  $V$  en funció del factor  $M$  i de la covariable  $C$ . Suposant que es compleixen les condicions dels models lineal i amb  $\alpha = 0.05$ , contesteu les següents preguntes:
    - (1) Quines són les preguntes que volem contestar (especifiqueu-ne tres)?
    - (2) interaccionen el coeficient d'intel·ligència i el mètode pedagògic? Com ho veus? Interpreta el resultat.
    - (3) Hi ha diferències en les valoracions depenent del mètode pedagògic?
    - (4) Hi ha diferències entre les valoracions segons el coeficient intel·lectual? Què hauries de veure si no n'hi haguessin?
    - (5) Les dues rectes són la mateixa o no?
    - (6) Les dues rectes són paral·leles o no?
    - (7) Les dues rectes tenen el mateix terme independent o no?
    - (8) Per cadascun dels següents valors de la variable  $C$ : 90, 105 i 120, quines diferències hi ha en la valoració segons el mètode?
    - (9) És possible simplificar el model? Si és que sí feu-ho i torneu a respondre les mateixes preguntes.
  - b) Repetiu l'exercici, però amb les valoracions de la variable  $VV$ .
- 2) Carregueu la llibreria MASS del R i visualitzeu el dataset que porta per nom *anorexia*. Aquest dataset conté el pes abans (Prewt) i després (Postwt) del tractament d'un total de 72 noies diagnosticades amb trastorns alimentaris. Hi va haver tres tipus de tractaments: control (sense tractament), CBT (Cognitive Behavioural treatment) i FT ("Family treatment"). Ajusteu un model lineal per tal d'estudiar la relació entre el pes post i pre-tractament, i contesteu les següents preguntes:
  - a) Porteu a terme el diagrama de punts de les dues variables sense tenir en compte el tractament, juntament amb la corresponent recta de regressió. Què observeu?
  - b) Repetiu el plot del diagrama anterior però ara trobeu la recta de regressió per a cada grup de tractament. Què observeu?
  - c) Ajusteu un model lineal amb el mateix pendent pels tres grups i interpreteu-lo. Amb l'anàlisi de residus concluiu que és un bon model?

- d) Ajusteu un model lineal tenint en compte que el pendent de la recta que relaciona el pes post i pre pot canviar al canviar de tractament. Què concluïu? Es un millor model que l'anterior?
- e) Escribiu la forma matricial del model a  $ma$  i amb  $R$  i comproveu que és la mateixa.
- f) Si la Maria forma part del grup FT i pesa abans del tractament tres kg més que la Laura, quin pes s'espera que tingui la Maria post tractament en referència al pes de la Laura?
- g) Si la Joana ara pesa 75,8 lliures i li aconsellem seguir el tractament CBT, deduïu un IC pel seu pes esperat post-tractament al 95%.

## 6 Model lineal Generalitzat

- 1) Estudieu si les distribucions: exponencial i binomial negativa poden ser utilitzades com a distribucions de la variable resposta en MLG. En el cas que la resposta sigui afirmativa determineu per a cada família: el paràmetre canònic, l'enllaç canònic i la funció variància.
- 2) Com es defineix la *Deviance*?. Demostreu que en el cas que el vector resposta segueixi una distribució Normal amb mateixa variància per a totes les components, la deviance coincideix amb la suma de quadrats residual.
- 3) Com es defineix l'estadístic de Pearson generalitzat? Trobeu l'expressió d'aquest estadístic quan ajustem un MLG amb resposta normal (variàncies iguals) i funció enllaç la identitat.
- 4) S'ha portat a terme un estudi de la resistència a la compressió d'un element de fixació d'aliatge utilitzat en la construcció. Deu nivells de càrrega van ser considerats, que anaven de 2500 psi a 4300 psi amb increments de 200 psi. Per cadascun dels nivells de càrrega es van testar diferents elements de fixació. Els resultats obtinguts en l'estudi apareixen en la taula següent:

Càrrega	Mida de la mostra	Nombre dels que van fallar
2500	50	10
2700	70	17
2900	100	30
3100	60	21
3300	40	18
3500	85	43
3700	90	54
3900	50	33
4100	80	60
4300	65	51

- a) Dibuixa la proporció dels que van fallar com a funció de la càrrega.
  - b) Defineix el MLG que creguis que pot ser apropiat per ajustar aquestes dades i justifica el perquè.
  - c) Assumeix que ja coneixes el vector  $\hat{\beta}$ . Dedueix analíticament la fórmula per obtindre  $\hat{p}_i$ , és adir, les probabilitats de que el dispositiu falli amb una càrrega igual a  $i$ .
  - d) Ajusta les dades amb el *glm* de R i treieu conclusions de l'ajust obtingut.
- 5) Breslow i Day al 1980 van recollir les següents dades corresponents a la incidència del càncer d'esòfag en població masculina de França. Els homes van ser classificats depenent de si consumien alcohol o no, entenent que una persona era declarada consumidora quan bebia un litre o més de begudes alcohòliques al dia.
- a) Explica quin model lineal generalitzat consideres apropiat per ajustar aquestes dades.
  - b) Escribeu el model en forma matricial.

- c) Ajusta el model anterior amb R. Quin són els valors de les estimacions dels paràmetres?. Son significatius els paràmetres?
- d) Quina és la probabilitat predida per algú que no begui alcohol i tingui una edat entre 55 i 64 anys?

edat	consumissió d'alcohol	càncer	
		sí	no
25-34	alta	1	9
	baixa	0	106
35-44	alta	4	26
	baixa	5	164
45-54	alta	25	29
	baixa	21	138
55-64	alta	42	27
	baixa	34	139
65-74	alta	19	18
	baixa	36	88
+ 75	alta	5	0
	baixa	8	31

- 6) Uns ratolins van ser exposats a diferents dosis dund eternat tractament. Al cap d'un determinat període de temps es van comptar els ratolins que presentaven algun tipus de tumor. Els resultats corresponents a les diferents dosis apareixent a la taula següent:

període de temps		dose		
		0	0.45	0.75
16 mesos	ratolins amb tumors	1	3	7
	ratolins exposats	205	304	193
24 mesos	ratolins amb tumors	20	98	118
	ratolins exposats	762	888	587

- a) Explica quina és la variable resposta i quina/quines les explicatives i de quin tipus son.
- b) Què és el que es vol respondre en aquest experiment? Excribiu-ho com a un test d'hipòtesis (hi ha diverses coses a respondre).
- c) Explica quin creus que és el model apropiat per ajustar aquestes dades, i escriu-lo en forma matricial.
- d) Ajusta el model anterior amb R i contesta les preguntes que t'has fet en l'apartat b).
- 7) Assumeix un model amb funció enllaç *complementary log-log*, resposta binària i només una variable explicativa numèrica. Obté la fórmula matemàtica que correspon a les probabilitats estimades ( $\hat{p}_i$ ).
- 8) En un determinat estudi, l'objectiu és testar l'efectivitat de diferents dosis d'insecticida **DOSE**. Per això es van recollir el nombre total d'insectes que havien rebut una determinada dosi **T**, i el nombre dels que van morir a una dosi concreta **DIED**. Les dades estan al fitxer *insecticida.csv*.

Es sap que el que té sentit és assumir que la probabilitat de que un insecte mori a la dosi  $i$  — *ssima*,  $\pi_i$ , està relacionada amb la dosi de la forma següent:

$$\pi = \Phi(\alpha + \beta \cdot \log(DOSI_i))$$

, on  $\Phi$  és la funció de distribució (probabilitats acumulades) d'una distribució  $N(0, 1)$ .

- a) Ajusteu les dades amb un MLG. Estimeu els paràmetres, calculeu les probabilitats estimades i també el vector de mortalitat esperada per a cada dosi. Feu l'anàlisi de residus. Aquests apartats s'han de fer en les dues següents situacions:
- (1) Assumint resposta binomial amb funció enllaç **probit**.
  - (2) Assumint resposta binomial binomial amb funció enllaç **logit**.
- b) Porteu a terme una anàlisi comparativa dels dos models i treieu-ne conclusions.