5 Models lineals.

Exercicis de pràctiques

5.1 Regressió lineal simple i múltiple.

1) En un estudi per estudiar el nivell de colesterol en sang, en persones en edat de creixement (9-20 anys), s'ha plantejat el model:

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

on C, P, H i E són: el nivell de colesterol, el pes, l'alçada i l'edat, respectivament.
Les dades experimentals són independents i estan en el fitxer "`col.csv`".
Escrivint aquest model de la forma matricial $y = X\beta + \varepsilon$, ($dim\,(X) = (n, p)$ i té rang màxim), contesteu els apartats següents donant les expressions matricials que heu utilitzat i els resultats numèrics d'aquest model. A més comproveu que amb el programa estadístic `R`, obteniu els mateixos resultats. Quan es necessiti $\alpha = 0.05$.

a) Trobeu la funció de regressió $C = \hat{\beta}_0 + \hat{\beta}_1 P + \hat{\beta}_2 H + \hat{\beta}_3 E$.

b) Trobeu els *"valors predits"*, $\hat{y} = \widehat{E}\,[y|\beta]$, i també els "*valors residuals*", $\hat{r} = y - \hat{y}$. Amb ells feu la gràfica de *"residuals"* en front de *"predits"* i dieu si us sembla que el model resumeix bé el comportament del colesterol.

c) Contrasteu el test *"omnibus"* $\left.\begin{array}{l} H_0 : (\beta_1, \beta_2, \beta_3) = (0,0,0) \\ H_1 : (\beta_1, \beta_2, \beta_3) \neq (0,0,0) \end{array}\right\}$ per saber si el model que hem plantejat millora el model nul $C_i = \beta_0 + \varepsilon_i$ en el que el colesterol no depèn del pes ni de l'alçada ni de l'edat.
Nota: Haureu de calcular GLE, GLM, SQE, SQM, MQE, MQM i F_{test}.

- Per a cadascuna de les tres variables independents, dieu si, en variar-la però mantenint constant les altres dues, afecta al colesterol i si ho fa de forma positiva o negativa. És a dir, heu de contrastar els tests $\left.\begin{array}{l} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{array}\right\}$, $\left.\begin{array}{l} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{array}\right\}$ i $\left.\begin{array}{l} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 \neq 0 \end{array}\right\}$.
Comproveu que els test coincideixen amb el que dona el `summary(model)`, que coincideix amb el que dona la comanda `Anova(model)`, però no coincideixen amb els de la comanda `anova(model)`, que a més canvia si canviem l'ordre de les variables en

la fórmula del model

Notes:

Necessitareu la distribució de $\hat{\sigma}^2 = MQE$ que ja teniu calculada, la distribució de $\hat{\beta}$.

La comanda `Anova(model)` és de la llibreria "`car`" i calcula l'anova amb les sumes de quadrats tipus II, també se li podria demanar les tipus III. La comanda `anova(model)` calcula l'anova amb les sumes de quadrats tipus I (seqüencial) que depenen de l'ordre de les variables en la fórmula.

d) Torneu a fer l'apartat anterior de forma alternativa, calculant els intervals de confiança $IC_{1-\alpha}(\beta_1)$, $IC_{1-\alpha}(\beta_2)$ i $IC_{1-\alpha}(\beta_3)$.

e) Per una persona amb $P = 65$, $E = 15$ i $H = 150$ calculeu-ne:

- Regió de "*predicció del 95%*" del colesterol.
- L'interval de confiança $IC_{1-\alpha}(E[C|P = 65, E = 15, H = 150])$.
- Contrasteu el test $H_0 : E[C|P = 65, E = 15, H = 150] = 190$ vs $H_1 : E[C|P = 65, E = 15, H = 150] \neq 190$.

Nota: Necessitareu la distribució de $\widehat{E}[C|P = 65, E = 15, H = 150] = (1, 65, 15, 150)\,\hat{\beta}$.

2) En el fitxer "`reg8.csv`" hi teniu les dades de X i Y, en 8 situacions diferents indicades per la variable REG, per a calcular les rectes de regressió de Y respecte X.

Nota: Teniu les mateixes dades en el fitxer `reg8h.csv`, però les 8 situacions estan en columnes diferents.

a) Calculeu les 8 rectes de regressió i observeu que totes donen les mateixes funcions de regressió i el mateix resultat del test.

b) Per a cada un dels casos, feu les gràfiques:

- "residuals" front "predits.
- valors observats de Y i X, amb la recta de regressió.

Observeu que de les 8 situacions només n'hi ha una en que el model sigui del tot correcte. Trobeu els motius pels que en els demés casos el model no és correcte, o té algun problema.

3) En un laboratori de química és molt corrent utilitzar potenciòmetres com a pH-metres, però si canviem els elèctrodes també es poden utilitzar per a determinar altres ions o molècules.

Així es pot connectar l'anomenat elèctrode selectiu d'amoníac que té una membrana especial que, com el seu nom indica, és sensible a l'amoníac gas. En aquest cas les lectures que dona el potenciòmetre es relacionen amb la concentració de nitrogen amoniacal mitjançant uns patrons (solucions de concentració de N. amoniacal coneguda) i la funció:

$$lectura = \beta_0 + \beta_1 \cdot \log(concentració)$$

En una experiència realitzada, els patrons, de concentracions conegudes han donat :

Concentració de NNH3 en ppm	*2*	*5*	*10*	*20*	*50*	*100*
lectura de l'elèctrode en mV	*85*	*107*	*124*	*140*	*161*	*179*

i un extracte aquós d'un compost ha donat una lectura de : $167mV$.

L'objectiu és estimar la concentració de N. amoniacal que té l'extracte, pel que necessitarem fer la regressió de les lectures (mV) dels patrons, en funció de les seves concentracions (ppm), segons la funció indicada. Quan es necessiti utilitzeu $\alpha = 0.05$.

a) Digueu quin és el model lineal d'aquesta regressió, les condicions que s'ha de suposar que es compleixen i plantegeu les hipòtesis del test ANOVA.

b) Calculeu la funció de regressió i contrasteu el test ANOVA.

c) Estudieu els punts estranys i els punts influents. Digueu si el model que hem escollit és adequat o no, i si cal revisar algun dels resultats obtinguts.

d) Digueu quina és la concentració que segons la regressió dona la mateixa lectura que l'extracte (167 mV).

e) És possible que la concentració de l'extracte sigui 60 ppm?

f) Digueu quina concentració pot tenir l'extracte, es a dir, trobeu totes les concentracions que poden donar la lectura de l'extracte (167 mV) amb probabilitat 0,95.

4) Per estudiar com evoluciona la quantitat de llet diària produïda per una vaca en funció dels dies que han passat des del part, s'han obtingut les dades del fitxer "`pllet.csv`", en el que la primera columna són els dies i la segona la producció en l/dia. Quan es necessiti utilitzeu $\alpha = 0.05$.

a) Amb el model de la recta $E[prod] = \beta_0 + \beta_1 dies$:

(1) Trobeu la funció de regressió i

- Plantegeu i contrasteu el test ANOVA.
- Estimeu la variància de l'Error.
- Calculeu el coeficient de determinació ajustat.

(2) Dibuixeu:

- La funció de regressió amb els punts.
- Els residuals front els valors predits.
- La banda de predicció$_{(95\%)}$ amb la funció de regressió i els punts observats.

(3) Justifiqueu si us sembla que el model és adequat o no.

b) Amb el model de la paràbola $E[prod] = \beta_0 + \beta_1 dies + \beta_2 dies^2$:

(1) Trobeu la funció de regressió i

- Plantegeu i contrasteu el test ANOVA.

- Estimeu la variància de l'Error.
- Calculeu el coeficient de determinació ajustat.

(2) Dibuixeu:

- La funció de regressió amb els punts.
- Els residuals front els valors predits.
- La banda de predicció$_{(95\%)}$ amb la funció de regressió i els punts observats.

(3) Justifiqueu si us sembla que el model és adequat o no.

c) Amb el model de la funció gamma transformant amb el logaritme
$E\left[\log\left(prod\right)\right] = \beta_0 + \beta_1 dies + \beta_2 \log\left(dies\right)$:

(1) Trobeu la funció de regressió i

- Plantegeu i contrasteu el test ANOVA.
- Estimeu la variància de l'Error.
- Calculeu el coeficient de determinació ajustat.
- Plantegeu i contrasteu el test per contestar les preguntes :
  · ajusta millor la funció gamma que la funció exponencial?
  · ajusta millor la funció gamma que la funció potencial?

(2) Dibuixeu:

- La funció de regressió amb els punts (amb transformació i sense).
- Els residuals front els valors predits.
- La banda de predicció amb la funció de regressió i amb els punts. (log i sense)
- Detecteu punts estranys i punts influents.

(3) Justifiqueu si us sembla que:

- el model és adequat o no.
- el model és més adequat, o no, que el de la funció exponencial.
- el model és més adequat, o no, que el de la funció potencial.

d) Entre tots els models d'aquest exercici, quin us sembla que és més adequat? per què?

5) En la mateixa situació que l'exercici **1)**.

És ben conegut que l'excés de pes es un dels factors que afecta negativament al nivell de colesterol de les persones. En un estudi amb nens i joves de 9 a 20 anys, s'ha obtingut el seu nivell de colesterol, C, pes, P, alçada, H, i edat, E. Els resultats són al fitxer "`col.csv`" i $\alpha = 0.05$.

a) Plantegeu i calculeu la recta de regressió del colesterol en funció del pes.

b) Dibuixeu la gràfica de la recta de regressió amb les bandes de confiança i de predicció.

c) Utilitzeu les gràfiques adequades pel diagnòstic de:

- Tendències i homogeneïtat de variàncies.
- Valores estranys.
- Valores influents.

d) Interpreteu els resultats, trobeu alguna contradicció? justifiqueu les conclusions. Indicació: us pot ajudar fer el diagrama de dispersió de C vs P, afegint les rectes de regressió agrupant per edats, la comanda és:

```
scatterplot(C~P|EDAT, reg.line=lm, smooth=F, spread=F, boxplots=F,
    span=0.5, data=dades)
```

6) Com en l'exercici **1)**.

En un estudi per estudiar el nivell de colesterol en sang, en persones en edat de creixement (9-20 anys), s'ha plantejat el model:

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

on C, P, H i E són: el nivell de colesterol, el pes, l'alçada i l'edat, respectivament.

Les dades experimentals són independents i estan en el fitxer ''`col.csv`''. Quan es necessiti utilitzeu $\alpha = 0.05$.

a) Comproveu, amb notació matricial, que si canviem les dades experimentals per combinacions lineals d'elles, $X_C = X \cdot C$ on C és una matriu (p, p) invertible, aleshores, en ajustar el model $Y = X_C\beta_C + \varepsilon$ s'obté que:

(1) Els valors predits, els residuals, $\hat{\sigma}^2$, R^2 i tots els elements del test "omnibus", donen exactament el mateix que sense transformar les dades.

(2) $\hat{\beta}_C = C^{-1}\hat{\beta}$, pel que també canvia la seva distribució, la col·linealitat, els tests dels paràmetres i els intervals de confiança dels paràmetres.

b) És ben conegut que l'excés de pes es un dels factors que afecta negativament al nivell de colesterol de les persones, però l'excés de pes no és exactament el pes. Suposant que $P_0 = -10 + 0.5H$ és un patró del pes respecta a l'alçada en les persones d'edat en el rang de la nostra experiència, definim l'excés de pes $EP = P - P_0 = P + 10 - 0.5H$. Ajusteu la regressió de C en funció de EP, H i E i comproveu que:

(1) Els valors predits, els residuals, $\hat{\sigma}^2$, R^2 i tots els elements del test "omnibus", donen exactament el mateix que sense transformar les dades.

(2) El *VIF* (col·linealitat) a canviat.

(3) Els resultats dels tests $\left.\begin{array}{l} H_0 : \beta_{C2} = 0 \\ H_1 : \beta_{C2} \neq 0 \end{array}\right\}$, $\left.\begin{array}{l} H_0 : \beta_{C3} = 0 \\ H_1 : \beta_{C3} \neq 0 \end{array}\right\}$ i $\left.\begin{array}{l} H_0 : \beta_{C4} = 0 \\ H_1 : \beta_{C4} \neq 0 \end{array}\right\}$ han canviat.

(4) Reinterpreteu com cadascuna de les variables independents afecta al colesterol, mantenint constants les demés variables independents.

(5) Tindria sentit eliminar l'alçada de la regressió i deixar només l'excés de pes i l'edat? quins avantatges i quins inconvenients tindria?

5.2 Anàlisi de la variància.

7) Para veure si afegir edulcorant millora l'engreix de garrins s'ha fet una experiència en la que s'ha mesurat el guany mig diari, *GMD*, de garrins en les mateixes condicions però afegint edulcorant. S'han provat les 5 dosis: *D00*, *D08*, *D15*, *D20* i *D30*, incloent la dosi D00 que de fet és la dieta sense edulcorant, i cada dieta s'ha provat en 5 garrins. Els resultats experimentals obtinguts són al fitxer "`gmd.csv`".
Tractant la dosis com els nivells d'un factor i $\alpha = 0.05$, calculeu:

a) Les matrius dels següents models lineals, les matrius del canvi dels demés models al primer, i per cada model, operant amb les matrius, el valor estimat dels paràmetres:

(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 $\alpha_1 = 0$

b) Utilitzant `R` ajusteu el model utilitzant algun dels models anteriors.
Notes: Per defecte el programa utilitza la 3a parametrització, per fer la primera s'ha d'indicar que es vol sense terme independent `lm(GMD~0+DOSI...)` i per la segona s'ha d'indicar que o faci amb la parametrització SUM, per fer-ho hi ha dues opcions, una és afegir a la comanda `lm(....,contrasts=list(DOSI="contr.sum"))` o bé a l'inici de la sessió executar la comanda `options(contrasts = c("contr.sum", "contr.treatment"))` així en aquesta sessió utilitzarà sempre aquesta parametrització.

c) Plantegeu, contrasteu i interpreteu, el test de l'anàlisi de la variància.
Comproveu que amb la 1a parametrització no contrasta el test per veure si els valors esperats de cada nivell són iguals o diferents, però si acceptem el terme independent, aleshores no importa amb quin dels models contrastem el test.

d) Utilitzant el mètode de Tukey:
Notes: No importa la parametrització escollida i es necessita el paquet del `R: emmeans`.

- Per cada nivell de dosi, estimeu els valors esperats del *GMD* i el seu interval de confiança.
  Nota: comanda `emmeans`.
- Feu les comparacions de totes les dosis entre elles.
  Dieu quines parelles de nivell de dosi, donen un GMD esperat diferent.
  Nota: comandes `emmeans i pairs.`
- Expresseu el resultat anterior de forma compacta.
  Nota: comandes `emmeans i cld.`

8) 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 són de 0 a 10 i l'ordenació és en ordre de preferència (de 1 a 5).

Per la puntuació del color:

a) Plantegeu el model lineal additiu amb els factors tastador i paté, i escriviu-ne la matriu del model.

b) Ajusteu el model utilitzant el R, i mireu quina matriu ha utilitzat.
Nota: per poder veure quina matriu del model a utilitzat heu d'afegir a la comanda `lm(...,x=T,...)` així amb ``"nom del model"$x`` ens donarà la matriu.

c) Suposant que es compleixen les condicions per fer el test anova, contrasteu el test adequat per veure si els patés es poden distingir, o no, pel color, feu-ne les comparacions múltiples i interpreteu els resultats.

d) Compareu els resultats de l'apartat anterior amb els que s'obtindrien si ho féssim amb el model de només factor paté, sense tenir en compte els tastadors.

e) Creus que es compleixen les condicions del test anova? Justifiqueu-ho teòricament i amb gràfics de diagnòstic.

Per a cadascuna de les demés valoracions, torneu a fer els apartats anteriors.
Nota: El cas de l'ordenació és molt diferent de les altres valoracions, i queda una versió asimptòtica del test de Friedmann.

9) 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, per separat en llet de cabra, d'ovella i de vaca; com canvia el rendiment amb un tractament tèrmic (crua/pasteuritzada) i amb l'addició de CaCl2 (si/no). Les dades són al fitxer "`formatges.csv`".
Per cadascun dels tres tipus de llet per separat, suposant que es compleixen les hipòtesis dels models lineals i amb $\alpha = 0.05$:

a) Plantegeu el model factorial i trobeu-ne la matriu del model.
Ajusteu el model amb `R` i mireu quina és la matriu del model que ha utilitzat.
Tenim els subespais vectorials V_1, V_2, V_3 i V_4, on V_1 és el generat per les columnes de la matriu del model corresponents al terme independent, V_2 per les de l'efecte principal del tractament tèrmic, V_3 per les de l'efecte principal del clorur càlcic i V_4 per les de la interacció. Aquests 4 subespais són ortogonals entre ells? això depèn de la parametrització escollida?

b) Tenen algun efecte els tractaments (combinacions dels nivells dels dos factors) utilitzats?
En cas afirmatiu responeu justificadament les preguntes següents:
   - Quin és el millor o els millors tractaments?
   - El CaCl2 té algun efecte sobre el rendiment?

- Podem dir que afegir CaCl2 fa augmentar el rendiment?
- El tractament tèrmic té algun efecte sobre el rendiment?
- Podem dir que pasteuritzar fa augmentar el rendiment?
- 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?

10) Tenim els resultats de mesurar l'àrea que cobreixen plantes entapissants a talussos de carreteres. S'han estudiat dues especies *E1* i *E2* (factor *ESPECIE*) i haver posat, no, compost al talús (factor *COMPOST*). Les dades experimentals són al fitxer "`area.csv`". Utilitzant $\alpha = 0.05$:

a) Plantegeu el model lineal additiu de la variable $AREA$ respecte els factors *ESPECIE* i *COMPOST*.

- Ajusteu aquest model.
- Contrasteu el test anova. Què ens diu el test?
- Contrasteu el test adequat per comprovar que en el model no era necessària la interacció dels dos factors.
- Feu les comparacions múltiples. Què ens diuen?
- Mitjançant gràfics de diagnòstic, comproveu que no es compleixen les hipòtesis dels models lineals. On és el problema?

b) Plantegeu el model lineal additiu de la variable $\log(AREA)$ en lloc de $AREA$.

- Ajusteu aquest model.
- Contrasteu el test anova. Què ens diu el test?
- Contrasteu el test adequat per comprovar que en el model és necessària la interacció dels dos factors.
- Feu les comparacions múltiples del model adequat. Què ens diuen?
- Mitjançant gràfics de diagnòstic, comproveu que ara ja no hi ha el problema que hi havia en el model de la variable $AREA$.
- En estudiar el logaritme de la variable, canvia el sentit dels efectes del model?

11) En el fitxer prac2f.csv, hi teniu 8 columnes, las dues primeres ($F1, F2$) són dos factors de 3 i 4 nivells respectivament, i les 6 restants ($V1, V2, \ldots, V6$) són resultats experimentals (simulats) de 6 variables aleatòries.
Per cadascuna de les variables aleatòries poseu l'enunciat d'una experiència, de forma que les seves dades experimentals puguin ser les del fitxer. Per separat, amb $\alpha = 0.05$, feu els apartats següents:

a) Descriviu el comportament de les dades mitjançant:

- una taula de mitjanes de l'estil:

$F1$ \ $F2$	1	2	3	4	Total
1					
2					
3					
Total					

- el diagrama de mitjanes-errors estàndard, $\overline{X} - S_{\overline{X}}$.

b) Plantegeu i ajusteu el model lineal factorial.

(1) Contrasteu el test anova.

(2) Feu les comparacions múltiples dels tractaments i les dels factors.
Nota: Pot ajudar transcriure el resultat de les comparacions múltiples a la taula de mitjanes.

(3) Dibuixeu el diagrama de dispersió dels residus vs els valors predits per poder veure si hi ha tendències no explicades i/o heterogeneïtat de variàncies.

c) Amb els resultats de l'apartat anterior contesteu justificadament les preguntes següents:

- Hi ha algun efecte? Com es veu en la taula de mitjanes?
- Quina és la variància de l'error?
- El factor *F1* té efecte? Com es veu a la taula de mitjanes? i al diagrama $\overline{X} - S_{\overline{X}}$?
- El factor *F2* té efecte? Com es veu a la taula de mitjanes? i al diagrama $\overline{X} - S_{\overline{X}}$?
- Hi ha interacció? Com es veu a la taula de mitjanes? i al diagrama $\overline{X} - S_{\overline{X}}$?
- Quin és el millor o millors tractaments?
- Quin és el millor o millors nivells del factor *F1*?
- Quin és el millor o millors nivells del factor *F1*?
- El millor tractament és la combinació dels millors nivells de *F1* i *F2*? per què?
- Si haguessis d'utilitzar el nivell 1 del factor *F1*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 2 del factor *F1*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 3 del factor *F1*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 1 del factor *F2*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 2 del factor *F2*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 3 del factor *F2*, quin tractament escolliries?
- Si haguessis d'utilitzar el nivell 4 del factor *F2*, quin tractament escolliries?
- Es veu alguna tendència anòmala en els residus?
- Es veu alguna tendència en la variància dels residus?

d) Utilitzant el model de dos factors sense interacció feu l'equivalent de l'apartat b) i responeu les preguntes que tinguin sentit de l'apartat c). Dieu si aquest model és adequat.

e) Utilitzant el model de dos factors encaixats ($F1 + F1 : F2$) feu l'equivalent de l'apartat b) i responeu les preguntes que tinguin sentit de l'apartat c). Dieu si aquest model és adequat.

12) Es vol estudiar si la degradació, *W2* , de les farines de blat depèn de la varietat del blat i també si afecta la presència, o no, d'insectes a la farina. Les dades experimentals obtingudes són al fitxer "`blat.csv`", on *VAR* indica la varietat, *W2* el valor de la degradació i *PRES* la presència d'insectes. Una característica d'aquestes dades és que hi diferències grans entre el nombre de repeticions dels tractaments.

a) Suposant que es compleixen les condicions del test anova, plantegeu i ajusteu un model lineal factorial de la variable *W2* amb els factors *VAR* i *PRES*. Amb els resultats obtinguts contesteu les preguntes següents:

- La degradació es veu afectada per la varietat i/o la presència d'insectes?
- La degradació es veu afectada per la presència d'insectes?
- La degradació es veu afectada per la varietat?
- Compara els resultats utilitzant sumes de quadrats tipus I i les de tipus II. L'ordre dels factors en la fórmula del model, té algun efecte sobre els resultats dels tests?
- En aquest cas, quin paper juga la interacció? podríem prescindir d'ella?

b) Suposant que es compleixen les condicions del test anova, plantegeu i ajusteu un model lineal de la variable *W2* amb els factors *VAR* i *PRES* sense interacció. Amb els resultats obtinguts contesteu les preguntes de l'apartat a) que tinguin sentit.

5.3 Anàlisi de la covariància.

13) Es vol comparar les valoracions, v, de dos mètodes pedagògics, m, en funció del coeficient d'intel·ligència, c, amb les dades del fitxer "`comrect.csv`" .
Plantegeu i ajusteu el model lineal factorial de la variable v en funció del factor m i de la covariable c. Suposant que es compleixen les condicions dels models lineal, amb $\alpha = 0.05$, a més d'estimar els paràmetres i contrastar els test de cada paràmetre (paràmetre=0, o no), contrasteu també els test adequats per contestar les preguntes següents:

a) Les dues rectes són iguals? o no?

b) Les dues rectes són paral·leles? o no?

c) Les dues rectes tenen el mateix terme independent? o no?

d) Per cadascun dels següents valores de la variable c: 90, 105 i 120, quines diferències hi ha en la valoració segons el mètode?

Repetiu l'exercici, però amb les valoracions vv.

14) En un estudi per veure com afecta, a primera hora del matí, l'estrès hídric sobre la fisiologia d'unes plantes, s'ha plantejat una experiència per mesurar la fotosíntesi, *FS*, i la transpiració, *TR*, de les plantes sotmeses a un factor estrès hídric, *SHIDR*, amb 4 nivells *S1*, *S2*, *S3* i *S4*, corresponents al nombre de dies sense ser regades. En total s'han mesurat 20 plantes, 5 per cada nivell d'estrès. Un problema que es va presentar en fer l'experiència és que per mesurar la *FS* i la *TR* d'una planta es triga aproximadament 3 minuts, o sigui en total una hora. Però a primera hora del matí les condicions atmosfèriques canvien ràpidament, per poder compensar-ho es va mesurar també el temps, *T*, des de l'inici de prendre mesures i la radiació solar, *RAD*, en el moment de fer la mesura. Totes les dades són al fitxer "`shidr.csv`" on hi ha també la variable *REP*, que és el número de repetició de la planta dins del nivell de SHIDR.
Suposant que es compleixen les condicions dels models lineals, amb $\alpha = 0.05$, per a cascuna de les parelles (FS, RAD), (FS, T), (TR, RAD) i (TR, T), que anomenarem (Y, X):

a) Dibuixeu la gràfica de dispersió (Y, X), junt amb les rectes de regressió de $Y \sim X$ dins de cada nivell del factor *SHIDR*.

b) Plantegeu, i ajusteu, el model factorial de la variable Y en funció del factor $SHIDR$ i de la covariable X.

c) Contrasteu els test adequats per a contestar les preguntes:

- El factor $SHIDR$ i la covariable X afecten a la variable Y?
- Les 4 rectes són paral·leles? en el cas que no ho siguin dieu quines diferències hi ha en els pendents.

- En el valor mitja de la covariable X, hi ha diferències en els valors estimats de la Y? en cas afirmatiu, quines diferències hi ha?
- Per un valor petit de X, $RAD = 350$ o bé $T = 0$, hi ha diferències en els valors estimats de la Y? en cas afirmatiu, quines diferències hi ha?
- Per un valor alt de X, $RAD = 650$ o bé $T = 60$, hi ha diferències en els valors estimats de la Y? en cas afirmatiu, quines diferències hi ha?
- Es podria plantejar un altre model de dimensió més petita? per què?

d) Contesteu també:

- Acompanyant el factor $SHIDR$, quina covariable, RAD o T, afecta més a la variable FS?
- i a la variable TR?

15) En el fitxer ``pracovar.csv'', hi teniu les dades (simulades) d'una experiència amb:

- un factor, *FACTOR*.
- una covariable o variable independent, *X*.
- diverses variables dependents *Y1, Y2, ..., Y8*.

Per a cadascuna de les variables dependents, *Y1, Y2, ..., Y8*, per separat:

a) Poseu un enunciat d'una experiència, de forma que les seves dades experimentals puguin ser les del fitxer i feu la descriptiva següent:

- Gràfica dels punts amb les rectes de regressió de cada nivell del factor.

b) Plantegeu un model d'anàlisi de la covariància factorial, ajusteu-lo i contrasteu els tests necessaris per contestar les preguntes següents:

- Hi ha algun efecte?
- Hi ha algun efecte del factor? $\leftrightarrow$ Les rectes són iguals?
- Hi ha algun efecte de la covariable? $\leftrightarrow$ Les rectes són horitzontals?
- El factor i la covariable interaccionen? $\leftrightarrow$ Les rectes són paral·leles?
- El model és adequat? segons la gràfica de diagnòstic: residuals vs predits.

c) Plantegeu un model d'anàlisi de la covariància sense interacció, ajusteu-lo i contrasteu els tests necessaris per contestar les preguntes següents:

- Hi ha algun efecte?
- Hi ha algun efecte del factor? $\leftrightarrow$ Les rectes són iguals?
- Hi ha algun efecte de la covariable? $\leftrightarrow$ Les rectes són horitzontals?
- El model és adequat? segons la gràfica de diagnòstic: residuals vs predits.
- Quin model us sembla més adequat? el factorial o sense interacció.

- Hi ha algun altre model de dimensió més petita que sigui adequat? quin?

16) Es vol comparar com evoluciona, en passar el temps, el contingut de vitamina C d'un suc de taronja natural, depenent del tipus d'envàs i la temperatura de conservació, dels que hem escollit tres tractaments, ***"a"*, *"b"*** i ***'c"*** .

Durant 12 setmanes, començant al cap d'una setmana després de l'envasat dels sucs, s'-ha analitzat el contingut de vitamina C de dues unitats de cada tractament. Les dades experimentals (simulades) són al fitxer "`vitc.csv`".
En general sabem que el contingut de vitamina C evoluciona seguint una funció exponencial, $vitc = \alpha e^{-\beta \cdot set}$,on $\alpha > 0$, $\beta > 0$, i el valor d'aquests paràmetres depèn del tractament de conservació. Amb $\alpha = 0.05$, feu:

a) Sabent que a l'envasat tots els sucs tenien el mateix contingut de vitamina C, plantegeu, i ajusteu, un model lineal adequat per veure si en els tres tractaments la pèrdua de vitamina C es fa amb la mateixa rapidesa, o no. Amb aquest model:
   - Estimeu el contingut de vitamina C esperat, en el moment de l'envasat.
   - Per a cada tractament calculeu el valor estimat de la β.
   - Dieu si les tres $\beta's$ són totes iguals, o no, i quines diferències hi ha.
   - Quin tractament, o tractaments, conserven més la vitamina C?

b) Test blanc: Plantegeu, i ajusteu, un altre model de dimensió més gran, amb el que es pugui comprovar que a l'envasat, els sucs dels tres tractaments tenien realment el mateix nivell de vitamina C. Amb aquest model:
   - Estimeu, per cada tractament, el contingut esperat de vitamina C que tenien en ser envasats, es a dir quan $set = 0$.
   - Els continguts de vitamina C a $set = 0$, són iguals? o no?
   - Amb aquest model, quin tractament, o tractaments, conserven més la $vitc$? El resultat coincideix amb la mateixa pregunta de l'apartat a)? per què?

c) Justifiqueu si els models lineals utilitzats compleixen totes les hipòtesis dels *"models lineals"* i dieu si ajusten bé les dades.