

1. S'han calculat els valors d'una variable resposta com funció lineal d'uns paràmetres. Les observacions són

$$\begin{aligned} 0.9 &= \mu + \alpha \\ 2.6 &= \mu + \beta - \gamma \\ 3.1 &= \mu + \beta \\ 0.5 &= \mu + \alpha - \gamma \end{aligned}$$

- (a) Quines diferències hi ha entre considerar aquesta informació com un sistema d'equacions lineals o com un model lineal?  
 (b) Podem calcular el valor dels paràmetres? Quina condició ha de verificar una funció paramètrica per a ser estimable en el model lineal?  
 (c) Calculeu l'estimació de la variància de l'estimador lineal òptim de  $\alpha - \beta$ .  
 (d) Feu els següents contrastos:

$$\text{i) } H_0 : \alpha = \beta \quad \text{ii) } H_0 : 2\alpha = \beta$$

2. S'ha estimat l'equació de regressió lineal entre el temps de duració d'un procés i una variable de control  $x_1$  amb una mostra de 21 dades i s'ha obtingut l'equació  $y = 2 + x_1$ ;  $\hat{\sigma}^2 = 1$ . (a) Si se sap que  $\hat{s}_y^2 = 5$ , calculeu el coeficient de correlació i el de determinació de la recta anterior; (b) Construïu un interval de confiança per al paràmetre que mesura l'efecte de  $x_1$  (99%); (c) Preveieu el valor de  $y$  per a  $x_1 = 10$  amb un interval de confiança de 99%. (A la mostra  $\bar{x}_1 = 0$ ); (d) Per tal de millorar l'equació anterior s'afegeixen dues variables  $x_2, x_3$  a les mateixes 21 dades, de manera que s'obté l'equació de la regressió múltiple

$$y = 2 + \underset{(0.1)}{1.05x_1} + \underset{(0.54)}{0.3x_2} + \underset{(0.2)}{0.1x_3} \quad \hat{\sigma}^2 = 0.5$$

Calculeu el coeficient de correlació i el de determinació i justifiqueu si aquest model és millor o no que l'anterior (els números entre parèntesis són les desviacions estàndard estimades). Quina conclusió podem extreure respecte de la relació entre  $x_2$  i  $x_3$ ?

3. En un estudi s'ha mesurat el calor emès en un experiment ( $y=\text{cal/gr}$ ) en funció del percentatge de 4 components ( $x_1, x_2, x_3, x_4$ ). Les dades són:

$x_1$	$x_2$	$x_3$	$x_4$	$y$
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

La suma de quadrats de l'error RSS per a diferents models és:

model	RSS	model	RSS	model	RSS
----	2715.8	12--	57.9	123-	48.11
1---	1265.7	1-3-	1227.1	12-4	47.97
-2--	906.3	1--4	74.8	1-34	50.84
--3-	1939.4	-23-	415.4	-234	73.81
---4	883.9	-2-4	868.9		
		--34	175.7	1234	47.86

a on ---- representa el model nul (sense cap variable), 1--- representa el model amb només  $x_1$ , etc. Es demana:

- Escolliu el millor subconjunt de variables explicatives amb l'algorisme *forward*.
  - Escolliu el millor subconjunt de variables explicatives amb l'algorisme *backward*.
  - Apliqueu el criteri *stepwise*.
  - Compareu els tres subconjunts trobats als apartats anteriors.
4. Sis individus participen en un experiment de discriminació perceptiva. Els experimentadors mesuren la taxa d'error en un procés de discriminació. Es proposen 3 condicions experimentals: en la primera els individus realitzen una determinada tasca en silenci (*no noise*), a la segona la realitzen amb un soroll intermitent (*intermittent*) i a la tercera el soroll és continu (*continuous*). Les dades es troben a la taula següent:

Subjects	No noise	Intermittent	Continuous
s1	117	119	127
s2	130	126	131
s3	122	118	129
s4	123	117	134
s5	126	120	137
s6	116	120	128

Qüestions:

- Pel tractament estadístic de les dades es consulta un estadístic inexpert que proposa dos models:  
 Model I:  $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ , on  $\alpha_i$  = efecte del soroll i  $\epsilon_{ij}$  = error experimental  
 Model II:  $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ , on  $\alpha_i$  = efecte del soroll,  $\beta_j$  = efecte de l'individu i  $\epsilon_{ij}$  = error experimental  
 Quin considereu que és el millor model? No seria millor un model amb interacció? Justifiqueu la resposta.
- Les sumes de quadrats dels dos models són:  
 Model I:  $SS_s = 403.11$ ,  $SS_{\text{error}} = 165.33$   
 Model II:  $SS_s = 403.11$ ,  $SS_{\text{ind}} = 164.44$ ,  $SS_{\text{error}} = 100.89$   
 Com serien les del model amb interacció? Esculliu els resultats adients per analitzar les dades.
- En un gràfic de mitjanes es veu que *intermittent* s'assembla bastant a *no noise*. Indiqueu si són o no diferents.