**Problema 1**

1. Y = Xβ + ϵ

Y1

Y2

Y3

Y4

1 1 1

1 0 1

0 1 0

2 -3 2

β1

β2

β3

ϵ1

ϵ2

ϵ3

ϵ4

**=** +

Les hipòtesis que cal fer per considerar-ho un model lineal normal són les condicions de Gauss - Markov:

1. E(ϵi) = 0, i = 1,...,n

Es tracta d’una condició natural sobre un error.

D’aquesta manera ens assegurem que E(*y*i) = β0 + β1xi, el model lineal és correcte i la distribució dels errors serà aleatòria (quan els representem gràficament no tindran cap forma concreta).

1. var(ϵi) = σ2, i = 1,...,n

És la propietat d’homoscedasticitat.

En cas contrari, si tinguéssim heteroscedasticitat, var(ϵi) variaria en funció de xi.

També es pretén prevenir-se de possibles punts influents.

1. E(ϵi · ϵj) = 0, per a tot i ≠ j

Les observacions han de ser incorrelacionades. Amb dos punts tenim una recta de regressió. Amb 20 còpies d’aquests dos punts, tenim 40 punts i la mateixa recta, poc fiable.

Aquestes condicions poden expressar-se en forma matricial com

E(ϵ) = 0, var(ϵ) = σ2 In

on E(ϵ) és el vector d’esperances matemàtiques i var(ϵ) és la matriu de covariàncies de ϵ=(ϵ1,...,ϵn)’.

També suposarem que cada ϵi es N(0, σ) i que ϵ1, ..., ϵn son estocàsticament independents.

La adopció d’aquestes condicions evitarà teòricament aquestes situacions anòmales que hem descrit.

1. En primer lloc, introduïm les dades a l’R i calculem el rang de la matriu de disseny.

Veiem que el rang de la matriu de disseny és 2. Per tant, com que hi ha 3 paràmetres (β1, β2, β3) perquè una funció paramètrica sigui estimable haurà de complir 1 condició (nº condicions = nº paràmetres – rang(X)).

A continuació plantegem el sistema d’equacions per a trobar la condició que s’ha de complir:

a1 = b1 + b2 + 2b4

a2 = b1 + b3 + -3b4 a1 = a3 (és la condició que s’ha de complir)

a3 = b1 + b2 + 2b4

|  |  |  |
| --- | --- | --- |
| **Funció** | **Estimable** | **Justificació** |
| β1 | No | No compleix la condició per a ser fpe |
| β1 – 2β2 + β3 | Sí | Compleix la condició per a ser fpe |
| β1 – β2 + β3 | Sí | Compleix la condició per a ser fpe |

1. Les equacions normals d’aquest model són han de ser de la forma X’XB = X’Y (on B és l’estimador de β). Per tant, tenim que:

6 -5 6

-5 11 -5

6 -5 6

β1

β2

β3

6.95

1.14

6.95

6β1 - 5β2 + 6β3 = 6.95

= -5β1 + 11β2 - 5β3 = 1.14

6β1 - 5β2 + 6β3 = 6.95

Com que el rang de la matriu de disseny és 2 i tenim 3 paràmetres podem fixar-ne un d’ells, per exemple β3 = 0.

Per tant, considerarem la mateixa matriu de disseny, però eliminant la columna de β3 i buscarem els paràmetres.

β1 = 2.003659

β2 = 1.014390

|  |  |
| --- | --- |
| **Funció** | **MQ** |
| β1 – 2β2 | -0.025 |
| β1 – β2 | 0.989 |

Aquestes estimacions són invariants

perquè β1 i β2 són invariants.

Tenim que la covariància de β1 – 2β2 i β1 – β2 és 7.49e-05.

**PROBLEMA 2**

1. Estimació de la variància de l’error = (Residual standard error)2 = 514.8516

Coeficient de determinació ajustat = Adjusted R-squared = 0.4816

1. > pf(sg$fstatistic[1], 4, 42, lower.tail=F)

value

1.814703e-06

El *F* estadístic és clarament significatiu, el que fa significativa la regressió.

La hipòtesi nul·la d’aquest contrast és H0: sex = status = income = verbal = 0. Hipòtesi que es rebutja clarament i acceptem la hipòtesi alternativa, que és el model que hem plantejat.

1. Sembla que la variància dels errors és normal, tot i que s’observen tres punts anòmals, estranys (24, 36 i 39).
2. També es verifica la hipòtesi de normalitat tot i que tornem a trobar els mateixos punts estranys.
3. Observem tres punts amb influència potencial, el 5, el 24 i el 39.
4. Tenim 1 *outlier*, que és el 24.
5. Tenim 2 punts influents (el 35 i el 42).