Lista_8

Claudia Meneses

3/1/2020

a) Primero ajustamos el modelo lineal en los 4 regresores x's. Se puede observar que todos los regresores son no significantes, excepto x3. El p-value de estadístico F es significante, por lo cual tiene sentido rechazar la hipótesis nula. $X\beta = 0$.

```
mod1 <- lm(y~x1+x2+x3+x4, data=datos)
summary(mod1)
```

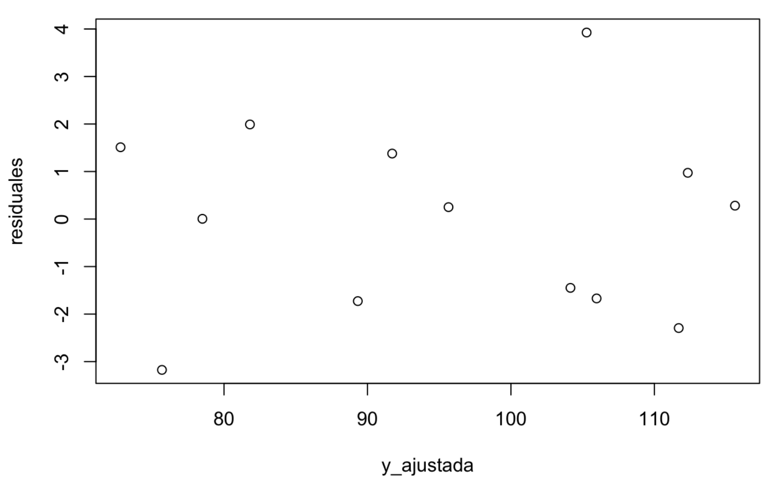
```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = datos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1750 -1.6709  0.2508  1.3783  3.9254
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  62.4054    70.0710   0.891   0.3991
## x1            1.5511     0.7448   2.083   0.0708 .
## x2            0.5102     0.7238   0.705   0.5009
## x3            0.1019     0.7547   0.135   0.8959
## x4           -0.1441     0.7091  -0.203   0.8441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.446 on 8 degrees of freedom
## Multiple R-squared:  0.9824, Adjusted R-squared:  0.9736
## F-statistic: 111.5 on 4 and 8 DF,  p-value: 4.756e-07
```

Calculamos los residuales, los residuales estandarizados, los studentizados interna y externamente.

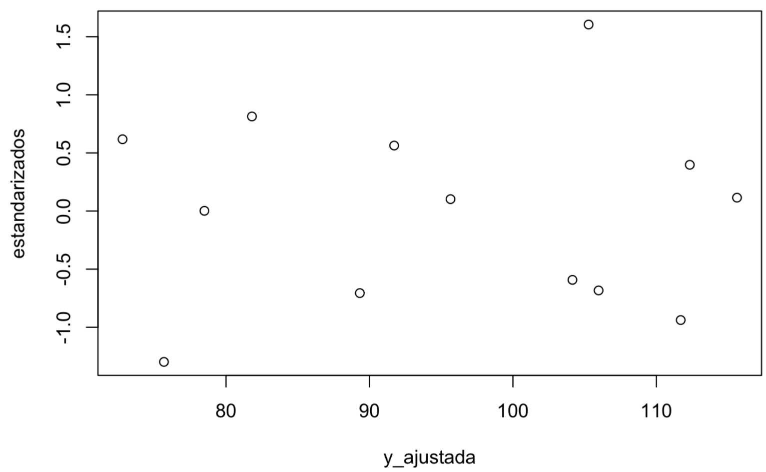
```
n=length(datos$y)
X=model.matrix(mod1)
H=X%*%solve(crossprod(X,X))%*%t(X)
hii <- diag(H)
residuales <- residuals(mod1)
estandarizados <- residuales/(2.446)
var_student <- 2.446*sqrt((1-hii))
studentizados_internamente <- residuales/var_student
s2_i <- ((n-5)*(2.446^2) - (residuales^2)/(1-hii))/(n-1-5)
studentizados_externamente <- residuales/(s2_i*sqrt(1-hii))
y_ajustada <- mod1$fitted.values
```

Graficamos los residuales para compararlos. Parece que los distintos residuales nos dan informacón parecida. Podemos notar que parecen cumplir con la aleatoriedad.

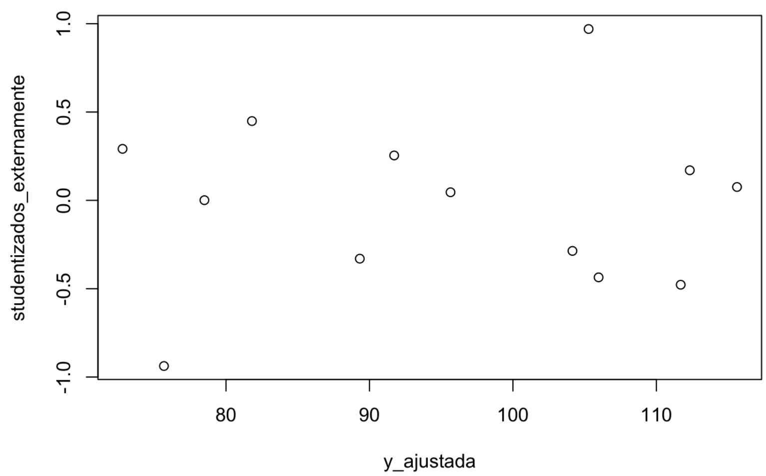
```
plot(y_ajustada, residuales)
```



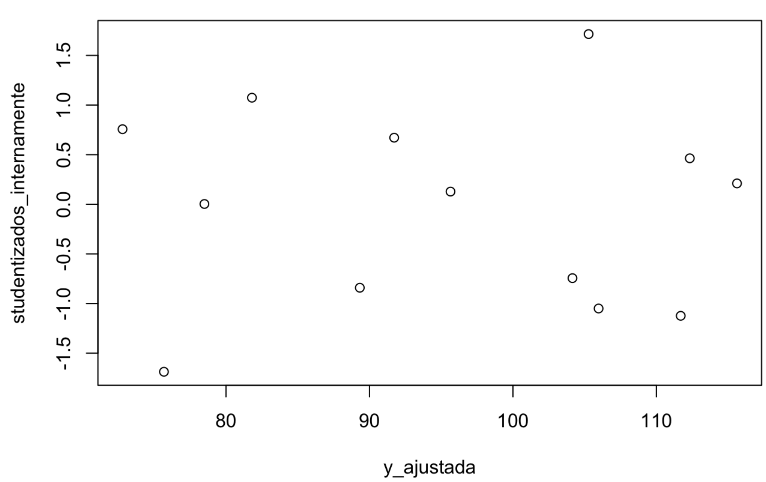

```
plot(y_ajustada, estandarizados)
```




```
plot(y_ajustada, studentizados_externamente)
```

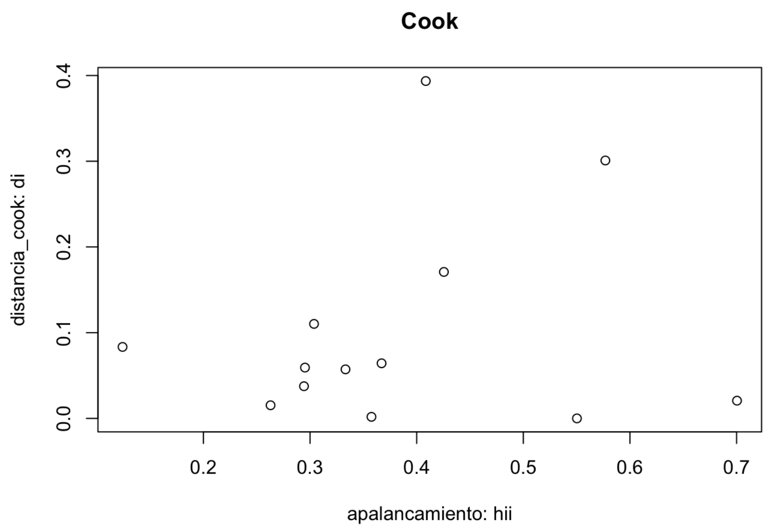



```
plot(y_ajustada,studentizados_internamente)
```



Calculamos los estadísticos de Cook. Se puede observar que dos observaciones destacan.

```
di <- (studentizados_internamente)^2*(hii/(1-hii))*1/5
plot(hii,di, main="Cook",ylab="distancia_cook: di", xlab="apalancamiento: hii")
```



El residual estandarizado y studentizado de la octava observación parece ser más alto de lo normal.

```
order(abs(estandarizados))
```

```
##  [1]  1  5 10 12  9  7  2  3  4 11 13  8  6
```

```
abs(estandarizados[8])
```

```
##        8
## 1.298033
```

```
order(abs(studentizados_externamente))
```

```
##  [1]  1  5 10 12  9  7  2  4  3 11 13  8  6
```

```
abs(studentizados_externamente[8])
```

```
##         8
## 0.9376636
```

Además la distancia de cook más grande es la de la octava observación.

```
order(di)
```

```
##  [1]  1  5 12 10  9  2  4  7  6 13 11  3  8
```

b) Eliminamos la octava observación y repetimos lo anterior. Al elmininar la octava observación las estimaciones de la Beta estimada cambiaron. En especial β_0 se redujo casi a la mitad. La desviación estándar se redujo un 20% aproximadamente.

```
mod2 <- lm(y~x1+x2+x3+x4, data=datos[-8,])
summary(mod2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4, data = datos[-8, ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1165 -1.1423 -0.1202  0.8024  3.9589
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2454    61.4554   0.606   0.5636
## x1            1.8797     0.6604   2.846   0.0248 *
## x2            0.7338     0.6312   1.162   0.2831
## x3            0.5187     0.6812   0.761   0.4713
## x4            0.1042     0.6212   0.168   0.8715
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.098 on 7 degrees of freedom
## Multiple R-squared:  0.9856, Adjusted R-squared:  0.9774
## F-statistic: 120.1 on 4 and 7 DF,  p-value: 1.578e-06
```

Apartir de los BFBETAS y DFFIT podemos notar que el dffit de la octava observación es la más alta.

```
print(influence.measures(mod1))
```

```
## Influence measures of
##   lm(formula = y ~ x1 + x2 + x3 + x4, data = datos) :
##
##      dfb.1_    dfb.x1   dfb.x2    dfb.x3   dfb.x4   dffit cov.r   cook.d
## 1  -0.00106  0.000773  0.00112  0.000698  0.00122  0.0030 4.335 2.06e-06
## 2   0.19947 -0.245131 -0.19378 -0.213899 -0.17826  0.5193 2.017 5.72e-02
## 3  -1.09529  1.033072  1.08281  1.083708  1.09704 -1.2356 2.195 3.01e-01
## 4  -0.23674  0.178058  0.24819  0.221515  0.22551 -0.5333 1.741 5.93e-02
## 5  -0.01884  0.002793  0.02378  0.002484  0.02264  0.0894 3.004 1.82e-03
## 6   0.19047 -0.123878 -0.18544 -0.177619 -0.19800  0.7594 0.225 8.34e-02
## 7   0.05380 -0.004812 -0.07168 -0.054930 -0.04362 -0.5497 2.151 6.43e-02
## 8   0.41856 -0.514380 -0.36015 -0.643741 -0.40814 -1.6352 0.365 3.94e-01
## 9   0.24840 -0.262227 -0.24268 -0.224818 -0.25144  0.4171 2.068 3.75e-02
## 10 -0.05297  0.116875  0.04138  0.078926  0.04637  0.3016 6.330 2.07e-02
## 11 -0.32727  0.378689  0.29816  0.463023  0.30992  0.9345 1.558 1.71e-01
## 12 -0.13589  0.135302  0.14134  0.128027  0.13177  0.2625 2.309 1.53e-02
## 13  0.35692 -0.300024 -0.38654 -0.289655 -0.35523 -0.7568 1.185 1.10e-01
##      hat inf
## 1  0.550   *
## 2  0.333
## 3  0.577   *
## 4  0.295
## 5  0.358   *
## 6  0.124
## 7  0.367
## 8  0.409
## 9  0.294
## 10 0.700   *
## 11 0.426
## 12 0.263
## 13 0.304
```