

Punto 11 a 14

Deivid Zhang Figueroa

24/12/2021

```
library(car)
```

```
## Loading required package: carData
```

```
library(perturb)  
library(leaps)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':  
##  
##      recode
```

```
## The following objects are masked from 'package:stats':  
##  
##      filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

```
library(olsrr)
```

```
##  
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':  
##  
##      rivers
```

```
#LEER DATOS EN archivo asignado a su grupo, así  
library(readr)  
datos <- read_csv("archivos/winequality-red.csv")
```

```
## Rows: 1599 Columns: 12
```

```
## -- Column specification -----
## Delimiter: ",",
## dbl (12): fixed acidity, volatile acidity, citric acid, residual sugar, chlo...

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

#datos=read.table(file.choose(),header=T,sep=",",dec=".")
datos1<-datos[1:100,c(1:5,12)]
```

11. Ajuste el modelo de regresión sin las observaciones 17 y 34, suponga que se establece que hay un error de digitación con estas dos observaciones, presente sólo la tabla de parámetros ajustados resultante ¿Cambian notoriamente las estimaciones de los parámetros, sus errores estándar y/o la significancia? ¿Qué concluye al respecto? Evalúe el gráfico de normalidad para los residuales estudentizados para este ajuste ¿mejoró la normalidad? Concluya sobre los efectos de este par de observaciones.

RTA:

```
# Se elimina las observaciones 17 y 34
datos2 = datos1[-c(17,34),]
attach(datos2)

datos2 <- datos2 %>%
  rename(Fija = `fixed acidity`) %>%
  rename(Volatil = `volatile acidity`) %>%
  rename(Citrico = `citric acid`) %>%
  rename(Azucar = `residual sugar`) %>%
  rename(Clорuros = chlorides) %>%
  rename(Calidad = quality)

# Se ajusta el modelo para datos 2 con todas las variables
mod2 = lm(Calidad ~., data = datos2)
# Tabla de parámetros ajustados
summary(mod2)
```

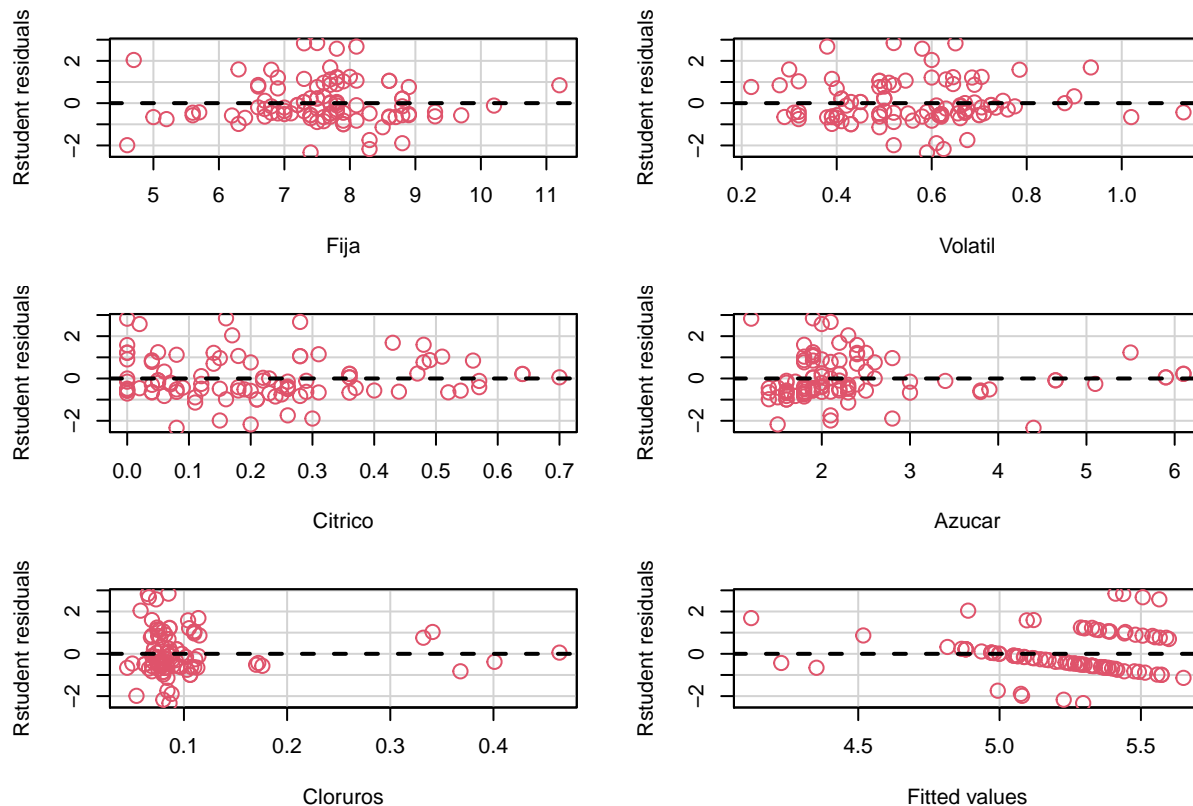
```
##
## Call:
## lm(formula = Calidad ~ ., data = datos2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2959 -0.3336 -0.1336  0.4274  1.5897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.94188    0.55924  10.625 < 2e-16 ***
## Fija           0.10754    0.06366   1.689 0.094535 .
## Volatil       -1.99180    0.47394  -4.203 6.1e-05 ***
## Citrico       -1.83987    0.49129  -3.745 0.000314 ***
## Azucar        -0.04494    0.05512  -0.815 0.416997
## Clорuros       0.91020    0.90062   1.011 0.314839
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5853 on 92 degrees of freedom
## Multiple R-squared:  0.1932, Adjusted R-squared:  0.1493
## F-statistic: 4.406 on 5 and 92 DF,  p-value: 0.001216
```

- En la anterior tabla se puede observar que, la estimación de los parámetros β_j no presentan cambios notorios, excepto para β_4 , pasa de 0.001826 (signo positivo) considerando las observaciones 17 y 34, a -0.04494 (signo negativo) sin considerar las observaciones 17 y 34. Ahora, para los errores estándar se puede observar que no cambian notoriamente sin considerar las observaciones 17 y 34, sin embargo, para la significancia, si se toma un nivel de significancia de 10%, $\alpha = 0.1$, se tiene que β_1 es significativo, pues $0.094535 < 0.1$, mientras que considerando las observaciones 17 y 34 esto no pasa, pues $0.14921 > 0.1$.

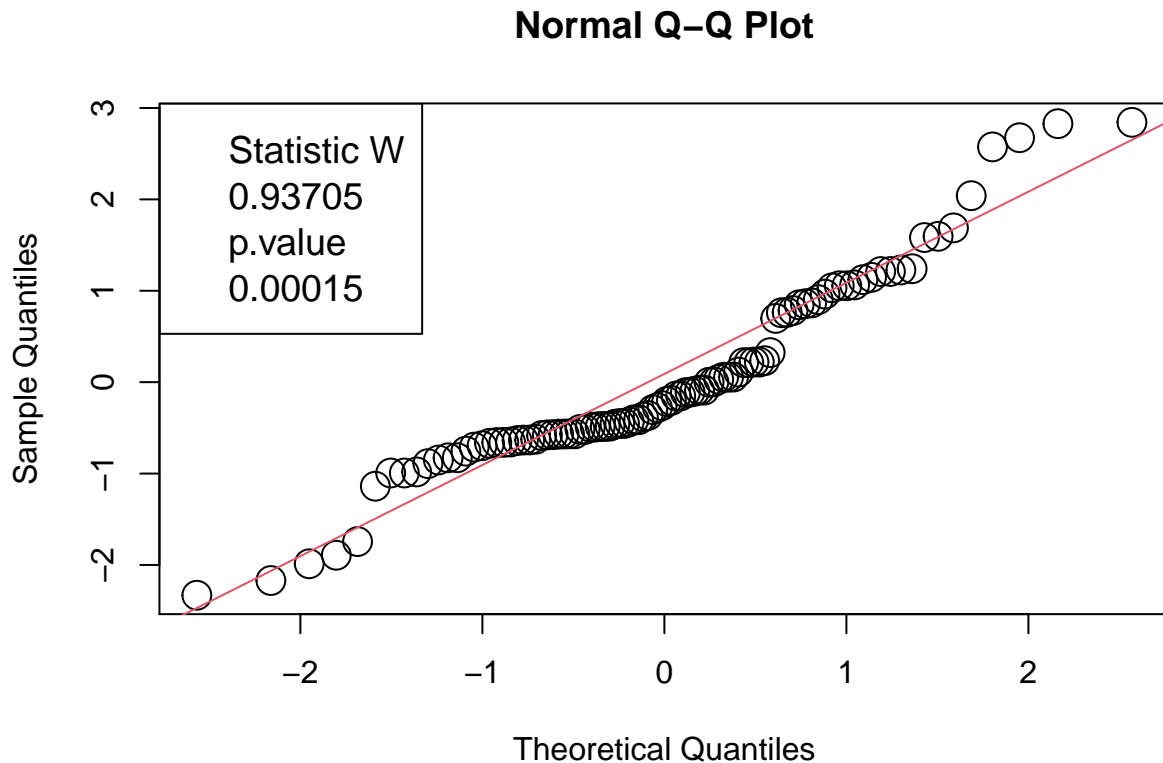
Gráfico de normalidad para los residuales estudentizados

```
residualPlots(mod2, tests=FALSE, type="rstudent", quadratic=FALSE, col=2, cex=1.5)
```



```
test=shapiro.test(rstudent(mod2)) #Test de normalidad sobre residuales estudentizados
qqnorm(rstudent(mod2), cex=2)

qqline(rstudent(mod2), col=2)
legend("topleft", legend=rbind(c("Statistic W", "p.value"),
round(c(test$statistic, test$p.value), digits=5)), cex=1.2)
```



- En la anterior gráfica se puede observar que, sin considerar las observaciones 17 y 34, esto no mejora la normalidad para los residuales estudentizados, pues el valor p es muy pequeño < 0.05 , por lo que se rechaza la hipótesis nula H_0 : los residuales estudentizados se distribuyen como una normal.
- ¿Cuál sería el efecto de estas 2 observaciones?

12. Para el modelo con todas las variables y sin las observaciones 17 y 34, realice diagnósticos de multicolinealidad mediante

```
# borrar esto cuando se unan las partes
miscoeficientes=function(modeloreg,datosreg) {
  coefi=coef(modeloreg)
  datos2=as.data.frame(scale(datosreg))
  coef.std=c(0,coef(lm(update(formula(modeloreg),~.+0),datos2)))
  limites=confint(modeloreg,level=0.95)
  vifs=c(0,vif(modeloreg))
  resul=data.frame("Estimación"=coefi,"Límites"=limites,Vif=vifs,Coef.Std=coef.std)
  cat("Coeficientes estimados, sus I.C, Vifs y Coeficientes estimados estandarizados","\n")
  resul
}
```

a) Matriz de correlación de las variables predictoras

RTA:

```
cor(datos2)
```

```
##           Fija      Volatil      Citrico      Azucar      Cloruros
## Fija      1.00000000 -0.30395236  0.480084940  0.095188579  0.095194155
## Volatil  -0.30395236  1.00000000 -0.626393071  0.022231186 -0.028960549
## Citrico   0.48008494 -0.62639307  1.000000000  0.006279926  0.257846311
## Azucar    0.09518858  0.02223119  0.006279926  1.000000000 -0.059467342
## Cloruros  0.09519416 -0.02896055  0.257846311 -0.059467342  1.000000000
## Calidad   0.09399282 -0.25265651 -0.077737119 -0.080289809  0.004570133
##           Calidad
## Fija      0.093992822
## Volatil  -0.252656514
## Citrico  -0.077737119
## Azucar    -0.080289809
## Cloruros  0.004570133
## Calidad   1.000000000
```

- Matriz de correlaciones: Se detecta una asociación lineal alta entre las variables cítrico y volátil, con un valor de -0.626393071.

b) VIF's

RTA:

```
miscoeficientes(mod2,datos2)
```

```
## Coeficientes estimados, sus I.C, Vifs y Coeficientes estimados estandarizados
```

```
##           Estimación Límites.2.5.. Límites.97.5..      Vif      Coef.Std
## (Intercept)  5.94188241      4.8311878      7.0525770  0.000000  0.00000000
## Fija         0.10753755     -0.0188877      0.2339628  1.315018  0.18141877
## Volatil     -1.99179719     -2.9330746     -1.0505198  1.701883 -0.51343135
## Citrico     -1.83986936     -2.8156132     -0.8641255  2.128785 -0.51169020
## Azucar      -0.04493921     -0.1544098      0.0645314  1.016979 -0.07699702
## Cloruros     0.91020023     -0.8785036      2.6989040  1.111713  0.09978949
```

- Con los valores VIFs: no se observa valores superando la cota de 10. Por este método no se detecta multicolinealidad-

c) Proporciones de varianza

```
colldiag(mod2)
```

```
## Condition
## Index      Variance Decomposition Proportions
##           intercept Fija  Volatil Citrico Azucar Cloruros
## 1    1.000  0.000    0.001  0.002   0.005   0.005   0.008
## 2    3.503  0.001    0.000  0.022   0.288   0.027   0.040
## 3    4.409  0.000    0.001  0.001   0.087   0.076   0.789
## 4    6.431  0.007    0.008  0.051   0.016   0.868   0.146
## 5   13.796  0.046    0.184  0.767   0.599   0.023   0.016
## 6   27.141  0.946    0.807  0.158   0.005   0.001   0.001
```

- Con las proporciones de descomposición de varianza: se puede observar que, en la quinta fila, π_{52} y π_{53} superan 0.5, y no existe otra fila i donde 2 π_{ij} superen esta cota, luego, con estos índices se detecta que volátil y cítrico están involucradas en una relación de multicolinealidad.

13. Sin las observaciones 17 y 34, construya modelos de regresión utilizando los métodos de selección (muestre de cada método sólo la tabla de resumen de este y la tabla ANOVA y la de parámetros estimados del modelo finalmente resultante):

```
#Todas las regresiones posibles; da información del Cp, R2, R2adj
k=ols_step_all_possible(mod2)
k
```

##	Index	N	Predictors	R-Square	Adj. R-Square
## 2	1	1	Volatil	6.383531e-02	0.054083599
## 1	2	1	Fija	8.834651e-03	-0.001489988
## 4	3	1	Azucar	6.446453e-03	-0.003903063
## 3	4	1	Citrico	6.043060e-03	-0.004310658
## 5	5	1	Cloruros	2.088611e-05	-0.010395563
## 10	6	2	Volatil Citrico	1.554956e-01	0.137716600
## 11	7	2	Volatil Azucar	6.941412e-02	0.049822840
## 6	8	2	Fija Volatil	6.416116e-02	0.044459295
## 12	9	2	Volatil Cloruros	6.384287e-02	0.044134295
## 7	10	2	Fija Citrico	2.845080e-02	0.007997130
## 8	11	2	Fija Azucar	1.687068e-02	-0.003826780
## 13	12	2	Citrico Azucar	1.241161e-02	-0.008379724
## 9	13	2	Fija Cloruros	8.853988e-03	-0.012012244
## 14	14	2	Citrico Cloruros	6.692076e-03	-0.014219670
## 15	15	2	Azucar Cloruros	6.446495e-03	-0.014470421
## 16	16	3	Fija Volatil Citrico	1.773604e-01	0.151105960
## 23	17	3	Volatil Citrico Cloruros	1.645202e-01	0.137855980
## 22	18	3	Volatil Citrico Azucar	1.599665e-01	0.133156898
## 17	19	3	Fija Volatil Azucar	7.010037e-02	0.040422727
## 24	20	3	Volatil Azucar Cloruros	6.946534e-02	0.039767430
## 18	21	3	Fija Volatil Cloruros	6.418056e-02	0.034313981
## 19	22	3	Fija Citrico Azucar	3.767912e-02	0.006966747
## 20	23	3	Fija Citrico Cloruros	2.938369e-02	-0.001593424
## 21	24	3	Fija Azucar Cloruros	1.698351e-02	-0.014389356
## 25	25	3	Citrico Azucar Cloruros	1.283068e-02	-0.018674726
## 27	26	4	Fija Volatil Citrico Cloruros	1.873597e-01	0.152407457
## 26	27	4	Fija Volatil Citrico Azucar	1.842320e-01	0.149145192
## 30	28	4	Volatil Citrico Azucar Cloruros	1.681609e-01	0.132382885
## 28	29	4	Fija Volatil Azucar Cloruros	7.019572e-02	0.030204139
## 29	30	4	Fija Citrico Azucar Cloruros	3.829514e-02	-0.003068515
## 31	31	5	Fija Volatil Citrico Azucar Cloruros	1.931893e-01	0.149340879
##	Mallow's Cp				
## 2	12.750133				
## 1	19.021816				
## 4	19.294141				
## 3	19.340139				
## 5	20.026843				
## 10	4.298177				
## 11	14.113986				
## 6	14.712977				
## 12	14.749272				

```
## 7    18.785002
## 8    20.105474
## 13   20.613938
## 9    21.019611
## 14   21.266132
## 15   21.294136
## 16    3.804954
## 23    5.269111
## 22    5.788371
## 17   16.035733
## 24   16.108145
## 18   16.710765
## 19   19.732704
## 20   20.678625
## 21   22.092608
## 25   22.566152
## 27    4.664740
## 26    5.021394
## 30    6.853966
## 28   18.024861
## 29   21.662460
## 31    6.000000
```

```
plot(k)
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

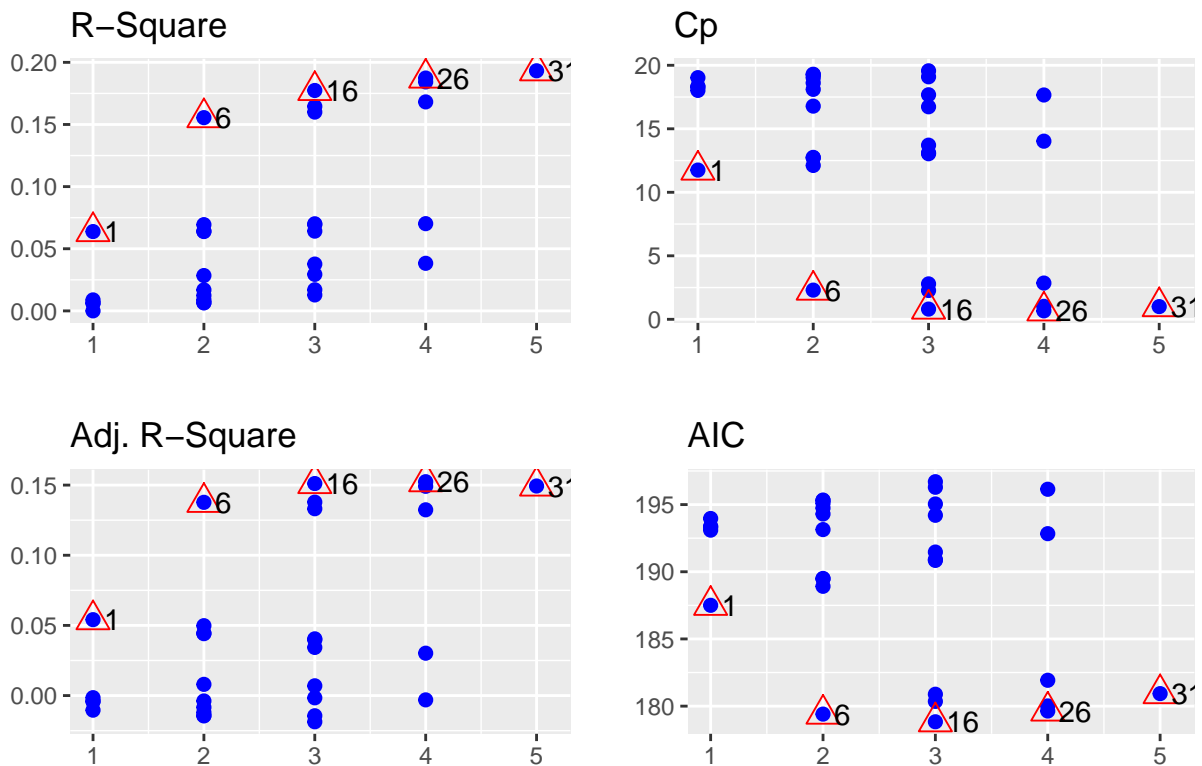
```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

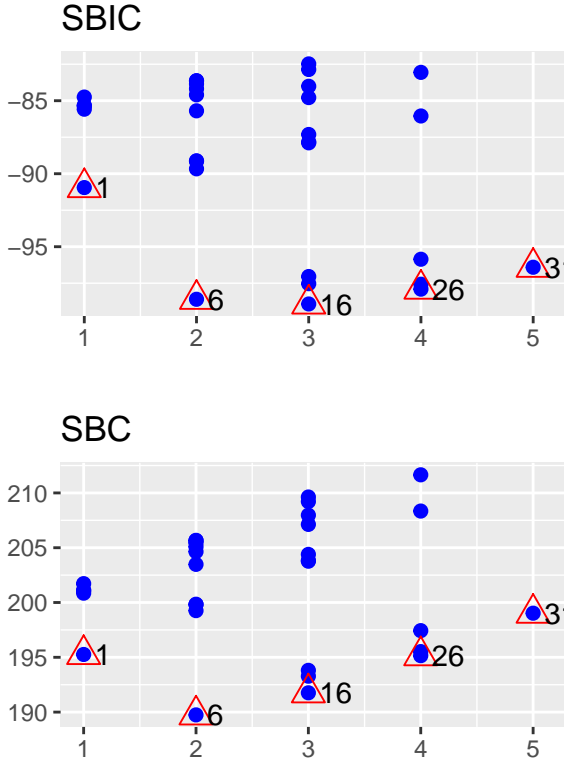
```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

page 1 of 2





a) Selección según el R_{adj}^2

RTA: Según el R_{adj}^2 , los mejores modelos son el 6, 16, 26 y 31, y como estos 3 últimos no muestran un incremento significativo en este estadístico, con respecto al modelo 6, entonces aplicando el principio de parsimonia, se escogería el modelo 6:

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + E_i$$

, $E_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

b) Selección según el estadístico C_p

RTA: teniendo en cuenta que con este estadístico se busca que el modelo con el menor valor $|C_p - p|$, los mejores candidatos son el modelo 6: $|C_p - p| = |4.298177 - 3| = 1.298177$, el modelo 16: $|C_p - p| = |3.804954 - 4| = 0.195046$, el modelo 26: $|C_p - p| = |4.664740 - 5| = 0.33526$ y el modelo 31: $|C_p - p| = |6 - 6| = 0$, pero de acuerdo con la ecuación

$$C_p = \frac{SSE_p}{MSE(X_1, X_2, \dots, X_k)} - (n - 2p)$$

, esto siempre ocurre con el modelo con todas las variables, por lo tanto, teniendo en cuenta que el modelo 16 tiene el valor más pequeño, entonces por este criterio se selecciona el modelo 16:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + E_i$$

, $E_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

c) Stepwise

```
# selección stepwie
ols_step_both_p(mod2, pent = 0.05, prem = 0.05, details=T)
```

```
## Stepwise Selection Method
```

```
## -----
```

```
##
```

```
## Candidate Terms:
```

```
##
```

```
## 1. Fija
```

```
## 2. Volatil
```

```
## 3. Citrico
```

```
## 4. Azucar
```

```
## 5. Cloruros
```

```
##
```

```
## We are selecting variables based on p value...
```

```
##
```

```
##
```

```
## Stepwise Selection: Step 1
```

```
##
```

```
## - Volatil added
```

```
##
```

```
## Model Summary
```

```
## -----
```

```
## R 0.253 RMSE 0.617
```

```
## R-Squared 0.064 Coef. Var 11.813
```

```
## Adj. R-Squared 0.054 MSE 0.381
```

```
## Pred R-Squared 0.028 MAE 0.460
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
```

```
## Sum of
## Squares DF Mean Square F Sig.
```

```
## -----
```

```
## Regression 2.493 1 2.493 6.546 0.0121
```

```
## Residual 36.568 96 0.381
```

```
## Total 39.061 97
```

```
## -----
```

```
##
```

```
## Parameter Estimates
```

```
## -----
```

```
## model Beta Std. Error Std. Beta t Sig. lower upper
```

```
## -----
```

```
## (Intercept) 5.776 0.224 25.732 0.000 5.331 6.222
```

```
## Volatil -0.980 0.383 -0.253 -2.559 0.012 -1.741 -0.220
```

```
## -----
```

```
##
```

```
##
```

```
##
```

```
## Stepwise Selection: Step 2
```

```

##
## - Citrico added
##
##                               Model Summary
## -----
## R                               0.394          RMSE              0.589
## R-Squared                       0.155          Coef. Var        11.279
## Adj. R-Squared                   0.138          MSE              0.347
## Pred R-Squared                   0.105          MAE              0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074          2          3.037      8.746      3e-04
## Residual        32.987         95          0.347
## Total           39.061         97
## -----
##
##                               Parameter Estimates
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)  6.608          0.336          19.654      0.000      5.941      7.276
## Volatil      -1.924          0.469          -0.496      -4.100      0.000      -2.855      -0.992
## Citrico      -1.397          0.435          -0.388      -3.211      0.002      -2.260      -0.533
## -----
##
##
##                               Model Summary
## -----
## R                               0.394          RMSE              0.589
## R-Squared                       0.155          Coef. Var        11.279
## Adj. R-Squared                   0.138          MSE              0.347
## Pred R-Squared                   0.105          MAE              0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074          2          3.037      8.746      3e-04
## Residual        32.987         95          0.347
## Total           39.061         97

```

```

## -----
##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)    6.608        0.336                19.654    0.000      5.941      7.276
##      Volatil   -1.924        0.469        -0.496    -4.100    0.000     -2.855     -0.992
##      Citrico   -1.397        0.435        -0.388    -3.211    0.002     -2.260     -0.533
## -----
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----
##
##                                     Model Summary
## -----
## R                0.394      RMSE                0.589
## R-Squared        0.155      Coef. Var          11.279
## Adj. R-Squared   0.138      MSE                0.347
## Pred R-Squared   0.105      MAE                0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                                     ANOVA
## -----
##      Sum of
##      Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074        2        3.037      8.746      3e-04
## Residual       32.987       95        0.347
## Total          39.061       97
## -----
##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)    6.608        0.336                19.654    0.000      5.941      7.276
##      Volatil   -1.924        0.469        -0.496    -4.100    0.000     -2.855     -0.992
##      Citrico   -1.397        0.435        -0.388    -3.211    0.002     -2.260     -0.533
## -----
##
##                                     Stepwise Selection Summary
## -----
##      Added/
##      Removed      R-Square      Adj.
##      Step      Variable      R-Square      C(p)      AIC      RMSE

```

```
## -----
##      1   Volatil      addition      0.064      0.054      12.7500      187.5034      0.6172
##      2   Citrico      addition      0.155      0.138      4.2980      179.4054      0.5893
## -----
```

- Según el método stepwise, el modelo a usar es el modelo 6:

$$Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + E_i$$

$$, E_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2).$$

d) Selección hacia adelante o forward

```
# selección forward
ols_step_forward_p(mod2, penter=0.05, details = T)
```

```
## Forward Selection Method
## -----
##
## Candidate Terms:
##
## 1. Fija
## 2. Volatil
## 3. Citrico
## 4. Azucar
## 5. Cloruros
##
## We are selecting variables based on p value...
##
## Forward Selection: Step 1
##
## - Volatil
##
##                               Model Summary
## -----
## R                0.253          RMSE                0.617
## R-Squared         0.064          Coef. Var           11.813
## Adj. R-Squared    0.054          MSE                0.381
## Pred R-Squared    0.028          MAE                0.460
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression         2.493          1          2.493      6.546      0.0121
## Residual           36.568         96          0.381
## Total              39.061         97
```

```

## -----
##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)    5.776      0.224              25.732    0.000      5.331      6.222
##      Volatil   -0.980      0.383      -0.253    -2.559    0.012     -1.741     -0.220
## -----
##
##
##
## Forward Selection: Step 2
##
## - Citrico
##
##                                     Model Summary
## -----
## R              0.394      RMSE              0.589
## R-Squared      0.155      Coef. Var      11.279
## Adj. R-Squared 0.138      MSE              0.347
## Pred R-Squared 0.105      MAE              0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                                     ANOVA
## -----
##      Sum of
##      Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074      2      3.037      8.746      3e-04
## Residual        32.987      95      0.347
## Total           39.061      97
## -----
##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)    6.608      0.336              19.654    0.000      5.941      7.276
##      Volatil   -1.924      0.469      -0.496    -4.100    0.000     -2.855     -0.992
##      Citrico   -1.397      0.435      -0.388    -3.211    0.002     -2.260     -0.533
## -----
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + Volatil
## + Citrico

```

```
##
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                0.394          RMSE                0.589
## R-Squared        0.155          Coef. Var            11.279
## Adj. R-Squared   0.138          MSE                 0.347
## Pred R-Squared   0.105          MAE                 0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                Sum of      DF      Mean Square      F      Sig.
##                Squares
## -----
## Regression      6.074        2          3.037      8.746      3e-04
## Residual       32.987       95          0.347
## Total          39.061       97
## -----
##
##                               Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)    6.608        0.336                19.654    0.000      5.941      7.276
## Volatil        -1.924        0.469         -0.496    -4.100    0.000     -2.855     -0.992
## Citrico        -1.397        0.435         -0.388    -3.211    0.002     -2.260     -0.533
## -----
##
##                               Selection Summary
## -----
##      Variable      Adj.
## Step  Entered      R-Square      R-Square      C(p)      AIC      RMSE
## -----
## 1  Volatil        0.0638      0.0541     12.7501     187.5034     0.6172
## 2  Citrico        0.1555      0.1377      4.2982     179.4054     0.5893
## -----
```

- Según el método forward, nuevamente, el modelo seleccionado es el modelo 6.

e) Selección hacia atrás o backward

```
# selección backward
ols_step_backward_p(mod2,prem=0.05,details = T)
```

Backward Elimination Method

```

## -----
##
## Candidate Terms:
##
## 1 . Fija
## 2 . Volatil
## 3 . Citrico
## 4 . Azucar
## 5 . Cloruros
##
## We are eliminating variables based on p value...
##
## - Azucar
##
## Backward Elimination: Step 1
##
## Variable Azucar Removed
##
##
## Model Summary
## -----
## R                0.433      RMSE                0.584
## R-Squared        0.187      Coef. Var          11.182
## Adj. R-Squared   0.152      MSE                0.341
## Pred R-Squared   0.104      MAE                0.439
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
## ANOVA
## -----
## Sum of
## Squares      DF      Mean Square      F      Sig.
## -----
## Regression    7.319      4      1.830      5.36      6e-04
## Residual     31.743     93      0.341
## Total        39.061     97
## -----
##
## Parameter Estimates
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upper
## -----
## (Intercept)  5.881      0.553      10.631      0.000      4.782      6.979
## Fija         0.102      0.063      0.172      1.617      0.109     -0.023      0.228
## Volatil     -2.009      0.473     -0.518     -4.252      0.000     -2.948     -1.071
## Citrico     -1.841      0.490     -0.512     -3.755      0.000     -2.815     -0.867
## Cloruros      0.960      0.897      0.105      1.070      0.288     -0.822      2.741
## -----
##
##
## - Cloruros
##
## Backward Elimination: Step 2

```



```

##
## Variable Cloruros Removed
##
##                               Model Summary
## -----
## R                               0.421          RMSE              0.585
## R-Squared                       0.177          Coef. Var        11.191
## Adj. R-Squared                   0.151          MSE              0.342
## Pred R-Squared                   0.104          MAE              0.445
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.928          3          2.309      6.755      4e-04
## Residual        32.133         94          0.342
## Total           39.061         97
## -----
##
##                               Parameter Estimates
## -----
##                               model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)      5.911          0.553              10.691      0.000      4.813      7.009
## Fija              0.100          0.063              0.169      1.581      0.117      -0.026      0.225
## Volatil           -1.920          0.466             -0.495     -4.125      0.000      -2.845     -0.996
## Citrico           -1.685          0.469             -0.469     -3.597      0.001      -2.616     -0.755
## -----
##
##
## - Fija
##
## Backward Elimination: Step 3
##
## Variable Fija Removed
##
##                               Model Summary
## -----
## R                               0.394          RMSE              0.589
## R-Squared                       0.155          Coef. Var        11.279
## Adj. R-Squared                   0.138          MSE              0.347
## Pred R-Squared                   0.105          MAE              0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----

```

```

##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074        2          3.037      8.746      3e-04
## Residual       32.987       95          0.347
## Total          39.061       97
## -----
##
##              Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upper
## -----
## (Intercept)      6.608        0.336              19.654      0.000      5.941      7.276
## Volatil          -1.924        0.469       -0.496      -4.100      0.000     -2.855     -0.992
## Citrico          -1.397        0.435       -0.388      -3.211      0.002     -2.260     -0.533
## -----
##
##
## No more variables satisfy the condition of p value = 0.05
##
## Variables Removed:
##
## - Azucar
## - Cloruros
## - Fija
##
## Final Model Output
## -----
##
##              Model Summary
## -----
## R              0.394      RMSE              0.589
## R-Squared       0.155      Coef. Var      11.279
## Adj. R-Squared  0.138      MSE              0.347
## Pred R-Squared  0.105      MAE              0.446
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##              ANOVA
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      6.074        2          3.037      8.746      3e-04
## Residual       32.987       95          0.347
## Total          39.061       97
## -----
##
##              Parameter Estimates

```

```
## -----
```

##	model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
##	-----							
##	(Intercept)	6.608	0.336		19.654	0.000	5.941	7.276
##	Volatil	-1.924	0.469	-0.496	-4.100	0.000	-2.855	-0.992
##	Citrico	-1.397	0.435	-0.388	-3.211	0.002	-2.260	-0.533
##	-----							

```
##
```

```
##
```

```
## Elimination Summary
```

```
## -----
```

##		Variable		Adj.			
##	Step	Removed	R-Square	R-Square	C(p)	AIC	RMSE
##	-----						
##	1	Azucar	0.1874	0.1524	4.6647	179.6362	0.5842
##	2	Cloruros	0.1774	0.1511	3.8050	178.8347	0.5847
##	3	Fija	0.1555	0.1377	4.2982	179.4054	0.5893
##	-----						

- Según el método backward, nuevamente, el modelo seleccionado es el modelo 6.

14. Con base en los anteriores numerales, ¿Cuál modelo sugiere para la variable respuesta? ¿por qué?

RTA: