

Prueba de Nivel de R

Alumno: Diego Antonini

Padrón: 100013

Librerías utilizadas

```
In [46]: library("readxl")
library("ggplot2")
library("corplot")
library("car")
```

Lectura dataset

```
In [89]: tema<-read_excel("tema0502.xlsx")
```

Visualicacion de las primeras filas del dataset

```
In [89]: head(tema)

  vta  traf  preA  preB  preC  pub  promo  exhib
161 4.308  2.46  2.22  3.59  0.509  0  0
218 5.287  2.83  2.46  4.55  0.479  1  1
289 4.548  2.36  2.13  4.57  0.427  1  1
104 4.568  2.57  2.18  4.25  0.577  0  1
276 5.103  2.17  1.98  4.27  0.307  0  0
295 4.885  2.68  2.50  4.74  0.393  1  1
```

A) Aplicar un modelo lineal para predecir las ventas y hacer una evaluación general de la calidad de ajuste del modelo

Modelo

```
In [90]: modelo<-lm(formula = log(vta) ~ traf + preA + preB + promo + exhib, data = tema)
summary(model1)

Call:
lm(formula = log(vta) ~ traf + preA + preB + promo + exhib, data = tema)

Residuals:
    Min       1Q   Median       3Q      Max
-0.44751 -0.08662 -0.00931  0.09667  0.32774

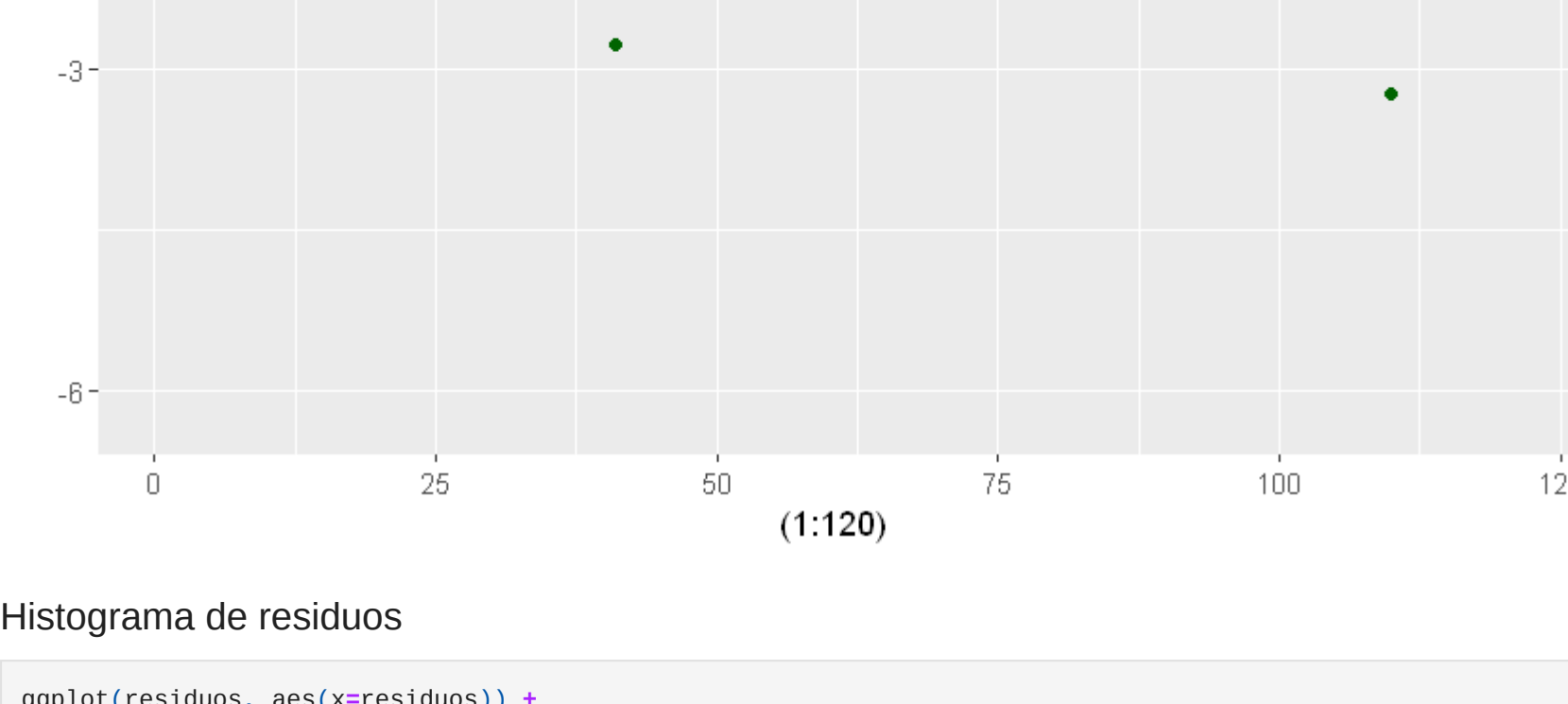
Coefficients:
(Intercept)  7.06397    0.22689   31.134   < 2e-16 ***
traf         0.19013    0.02549    7.469  1.84e-11 ***
preA        -2.73892    0.10198   -26.077   < 2e-16 ***
preB         1.74884    0.08989   19.475   < 2e-16 ***
promo        0.38621    0.03277    9.834   < 2e-16 ***
exhib        0.18941    0.03423    4.959  2.59e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1426 on 114 degrees of freedom
Multiple R-squared:  0.9138,    Adjusted R-squared:  0.91
F-statistic: 241.7 on 5 and 114 DF,  p-value: < 2.2e-16
```

Residuos

```
In [94]: residuos<-data.frame(residuos=srstandard(model1))
ggplot(residuos,aes(y=residuos,x=(1:120)))+geom_point(color="darkgreen")+labs(title="Residuos Estandarizados")+ylim(-6,6)
```

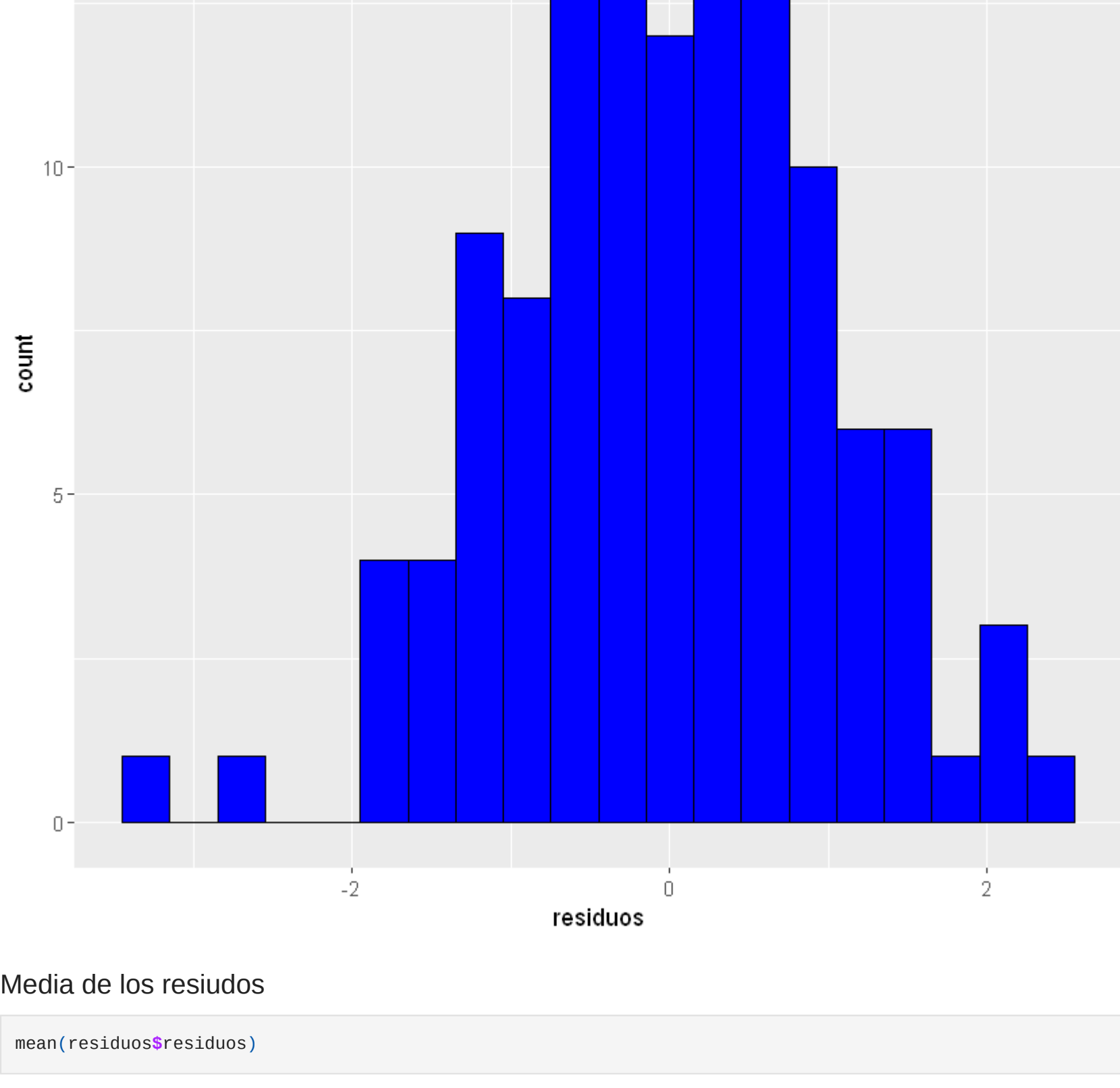
Residuos Estandarizados



Histograma de residuos

```
In [95]: ggplot(residuos, aes(x=residuos)) +
  geom_histogram(color="black", fill="blue", binwidth=0.3)+
  labs(title="Histograma de residuos")
```

Histograma de residuos



Media de los residuos

```
In [96]: mean(residuos$residuos)

[1] 0.00159992292884961
```

Validez del modelo

R2= 0.9138

En el "summary" del modelo se puede observar que los coeficientes son significativos

Se cumple la homocedasticidad y normalidad de los residuos

B) Evaluar la colinealidad y eventualmente buscar una solución

Matriz de correlaciones

```
In [97]: correlation_matrix<-cor(tema[, -1])
corplot(correlation_matrix, method = "number")
```



DET

```
In [98]: det(correlation_matrix)

[1] 0.263486089343122
```

VIF

```
In [99]: vif(model1)

traf      1.07183263389092
preA     2.01462545417718
preB     2.04187651506056
promo    1.66905098314009
exhib    1.71218529195072
```

El DET da 0.26. No es severo pero tampoco es ideal, hay cierto grado de correlación entre el precio de B y A

Los VIF son menores a 5. No se observa multicolinealidad severa

C) Evaluar la curvatura agregando términos cuadráticos y cúbicos para las variables explicativas

Modelo con todas las variables continuas sin transformar

```
In [101.]: modelo<-lm(formula = vta ~ traf + preA + preB + preC + pub , data = tema)
summary(modelC)

Call:
lm(formula = vta ~ traf + preA + preB + preC + pub, data = tema)

Residuals:
    Min       1Q   Median       3Q      Max
-87.58 -43.18 -20.14  25.21 297.59

Coefficients:
(Intercept)  581.558    130.966    3.839  0.000210 ***
traf        41.144     11.392    3.621  0.000439 ***
preA       -522.096    45.795   -11.401   < 2e-16 ***
preB       346.159    49.113    6.979  9.3e-14 ***
preC        4.013     18.012    0.223  0.824081
pub         64.471     58.522    1.102  0.272931
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

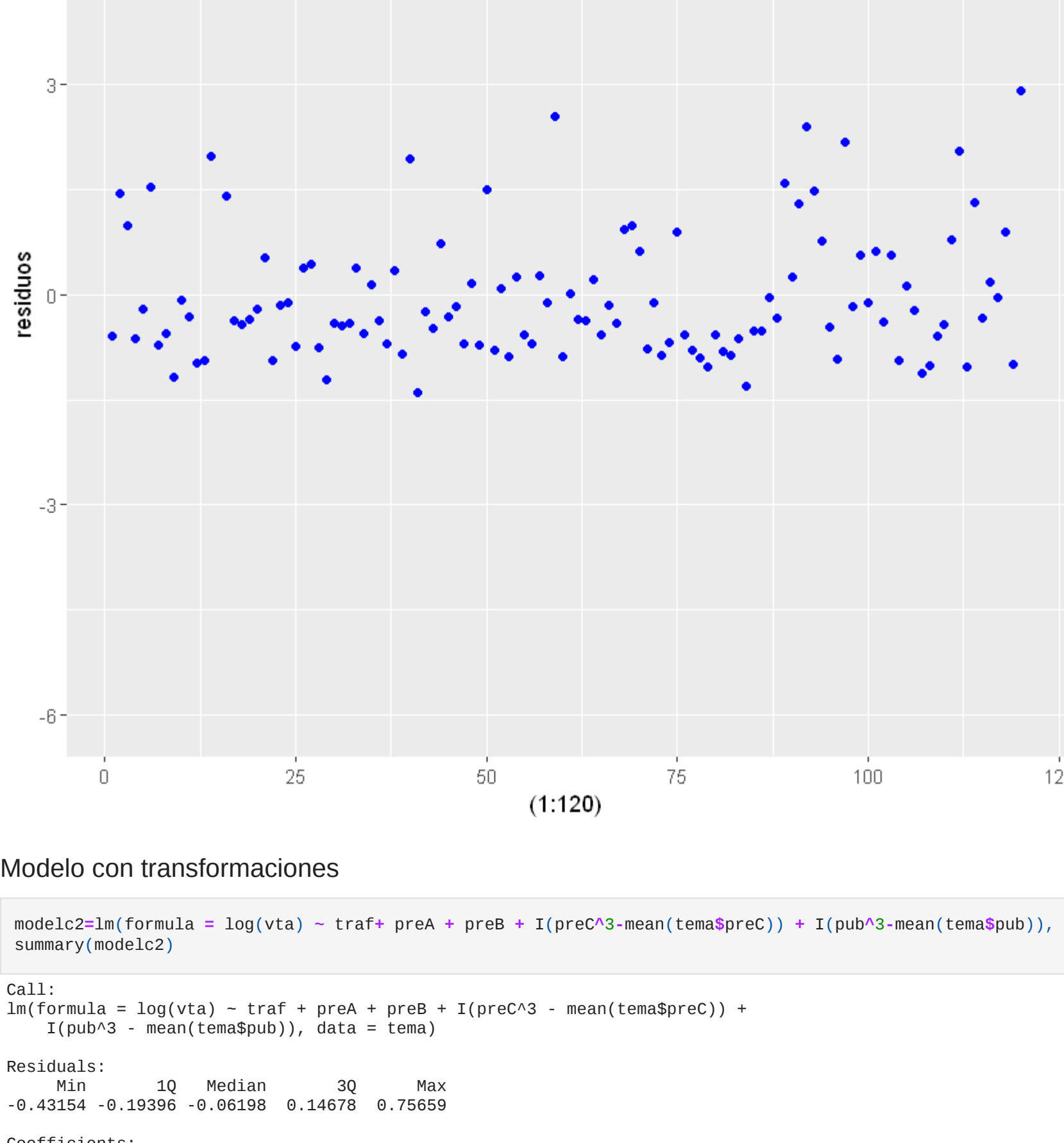
Residual standard error: 64.02 on 114 degrees of freedom
Multiple R-squared:  0.5886,    Adjusted R-squared:  0.5795
F-statistic: 32.62 on 5 and 114 DF,  p-value: < 2.2e-16
```

Residuos

```
In [69]: residuos_d<-data.frame(residuos=srstandard(modelC))
ggplot(residuos_d,aes(y=residuos,x=(1:120)))+geom_point(col = ifelse(residuos > 3,'red','blue'))+labs(title="Residuos Estandarizados" , subtitle="En rojo se p
```

Residuos Estandarizados

En rojo se puede observar el outlier



Modelo con transformaciones

```
In [79]: modelo2<-lm(formula = log(vta) ~ traf+ preA + preB + I(preC^3 - mean(tema$preC)) + I(pub^3 - mean(tema$pub)), data = tema)
summary(modelC2)

Call:
lm(formula = log(vta) ~ traf + preA + preB + I(preC^3 - mean(tema$preC)) + I(pub^3 - mean(tema$pub)), data = tema)

Residuals:
    Min       1Q   Median       3Q      Max
-0.43154 -0.19396 -0.06198  0.14678  0.75659

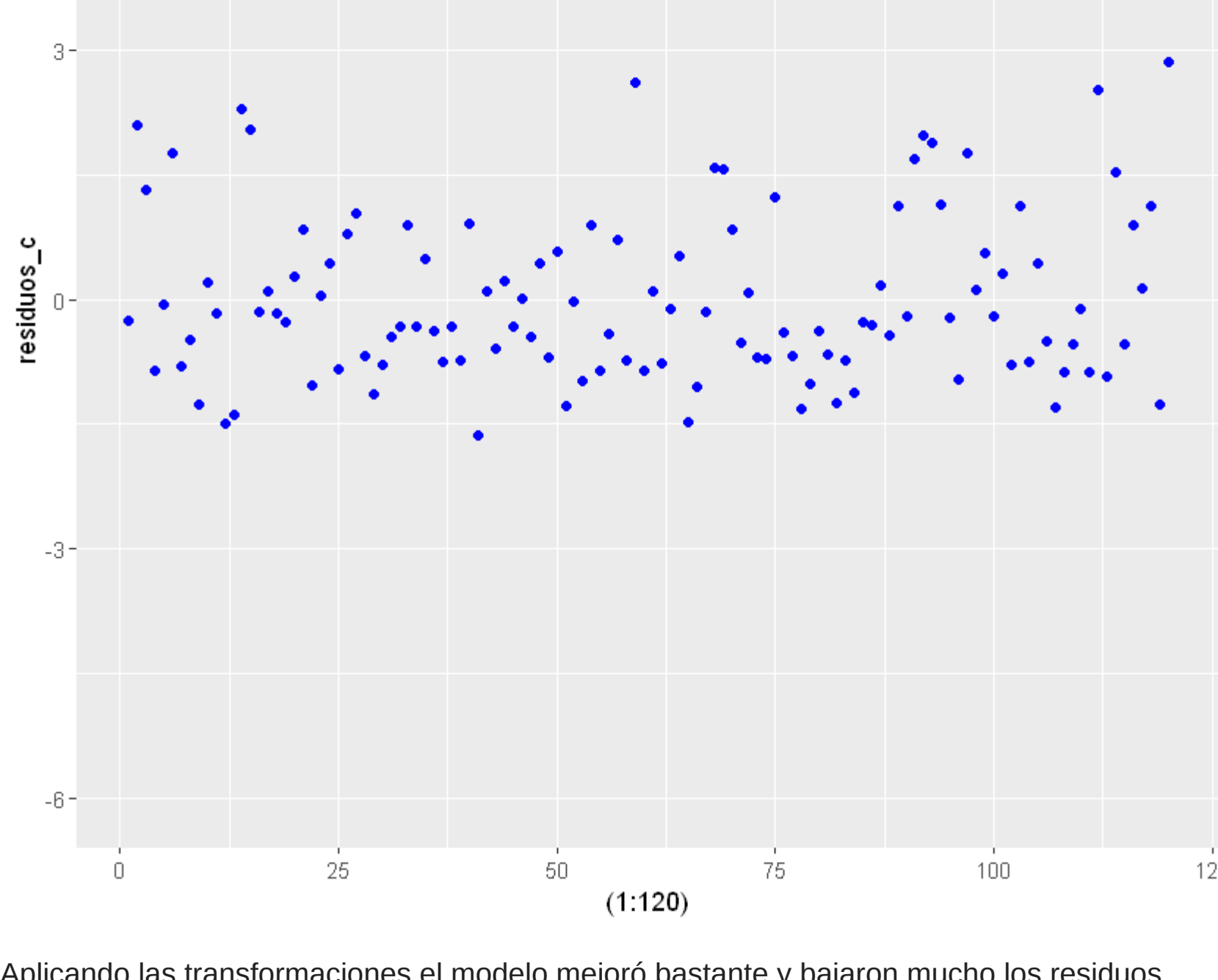
Coefficients:
(Intercept)  615.473    87.051    7.079  1.41e-18 ***
traf        37.465     7.644    4.901  3.24e-06 ***
preA       -493.295    27.003   -18.268   < 2e-16 ***
preB       292.844    23.515    12.453   < 2e-16 ***
I(preC^3 - mean(tema$preC))  0.0008351  0.0012569  0.664  0.508
I(pub^3 - mean(tema$pub))    0.3978102  0.3040755  1.308  0.193
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2684 on 114 degrees of freedom
Multiple R-squared:  0.6944,    Adjusted R-squared:  0.681
F-statistic: 51.82 on 5 and 114 DF,  p-value: < 2.2e-16
```

```
In [103.]: residuos_c<-data.frame(residuos=srstandard(modelC2))
ggplot(residuos_c,aes(y=residuos_c,x=(1:120)))+geom_point(col = ifelse(residuos_c > 3,'red','blue'))+labs(title="Residuos Estandarizados" , subtitle="En rojo se p
```

Residuos Estandarizados

En rojo se puede observar el outlier



Aplicando las transformaciones el modelo mejoró bastante y bajaron mucho los residuos

D) Evaluar la presencia de outliers y eventualmente eliminarlos, y explicar los cambios que se producen en el modelo.

Modelo con todas las variables

```
In [20]: modeld<-lm(formula = vta ~ traf + preA + preB + preC + pub + promo + exhib, data = tema)
summary(modeld)

Call:
lm(formula = vta ~ traf + preA + preB + preC + pub + promo + exhib, data = tema)

Residuals:
    Min       1Q   Median       3Q      Max
-113.539 -23.398  -1.572   17.103  206.636

Coefficients:
(Intercept)  615.473    87.051    7.079  1.41e-18 ***
traf        36.265     6.715    5.400  3.83e-07 ***
preA       -493.295    27.003   -18.268   < 2e-16 ***
preB       292.844    23.515    12.453   < 2e-16 ***
preC        4.592     10.620    0.432  0.66628
pub         21.249     34.421    0.609  0.36591
promo       89.605    10.440    7.728  5.49e-12 ***
exhib       27.545     9.021    3.053  0.00283 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

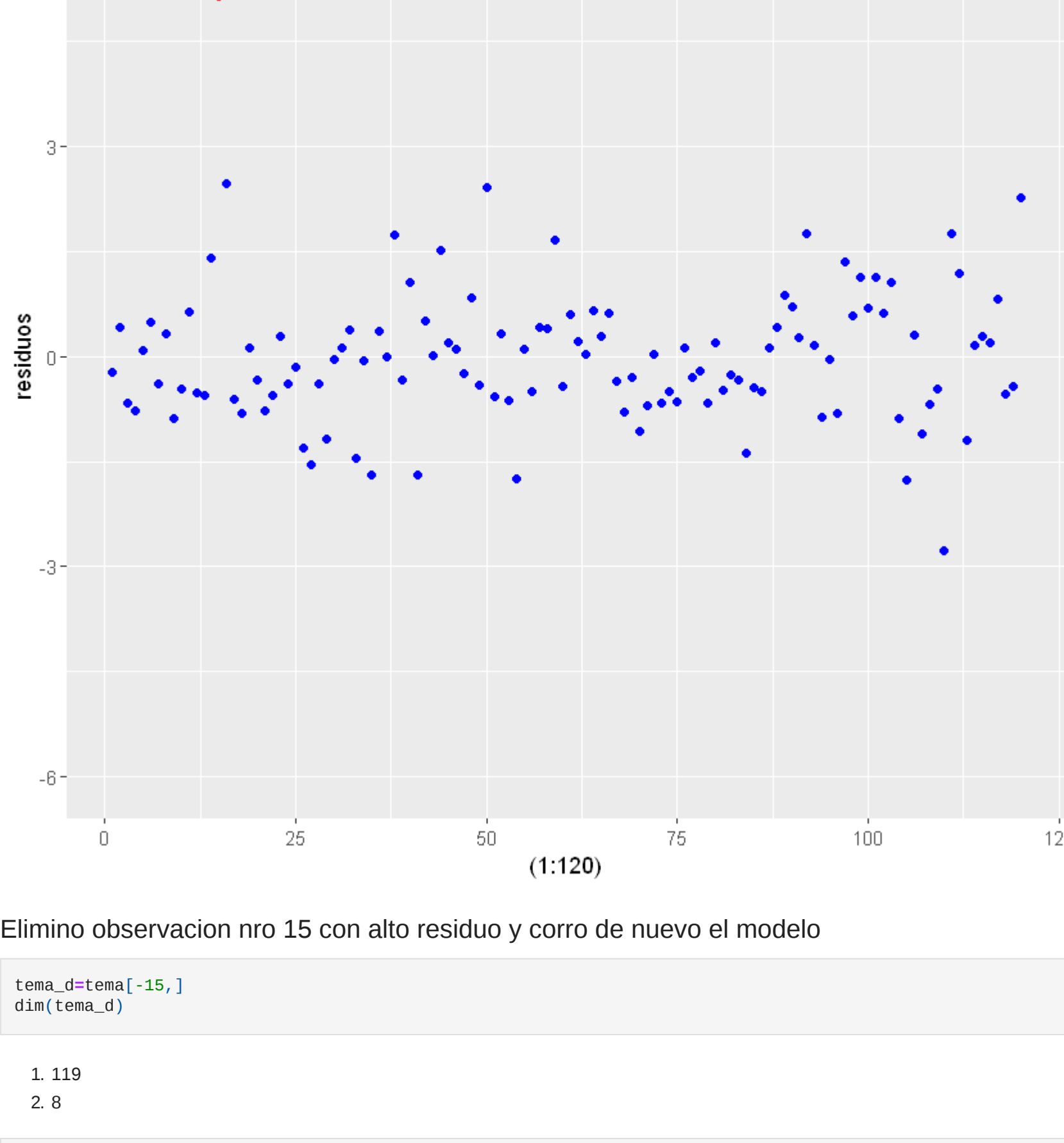
Residual standard error: 37.13 on 111 degrees of freedom
Multiple R-squared:  0.8366,    Adjusted R-squared:  0.8263
F-statistic: 81.19 on 7 and 111 DF,  p-value: < 2.2e-16
```

Residuos

```
In [32]: residuos_d<-data.frame(residuos=srstandard(modeld))
ggplot(residuos_d,aes(y=residuos,x=(1:120)))+geom_point(col = ifelse(residuos > 3,'red','blue'))+labs(title="Residuos Estandarizados" , subtitle="En rojo se p
```

Residuos Estandarizados

En rojo se puede observar el outlier



Elimino observacion nro 15 con alto residuo y corro de nuevo el modelo

```
In [42]: tema_d<-tema[-15,]
dim(tema_d)
```

1 119
2 8

```
In [44]: modeld2<-lm(formula = vta ~ traf + preA + preB + preC + pub + promo + exhib, data = tema_d)
summary(modeld2)
```

```
Call:
lm(formula = vta ~ traf + preA + preB + preC + pub + promo + exhib, data = tema_d)

Residuals:
    Min       1Q   Median       3Q      Max
-101.369 -22.334  -2.891   17.623   182.756

Coefficients:
(Intercept)  637.608    77.595    6.998  2.91e-18 ***
traf        36.265     6.715    5.400  3.83e-07 ***
preA       -493.295    27.003   -18.268   < 2e-16 ***
preB       292.844    23.515    12.453   < 2e-16 ***
preC        4.592     10.620    0.432  0.66628
pub         21.249     34.421    0.609  0.36591
promo       89.605    10.440    7.728  5.49e-12 ***
exhib       27.545     9.021    3.053  0.00283 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.13 on 111 degrees of freedom
Multiple R-squared:  0.8366,    Adjusted R-squared:  0.8263
F-statistic: 81.19 on 7 and 111 DF,  p-value: < 2.2e-16
```

El R2 aumenta y el RSE disminuye