Trabajo 3

Ike Mercado y Fabián Ramírez

```
library('MASS')
library('car')
library('alr3')
library('faraway')
```

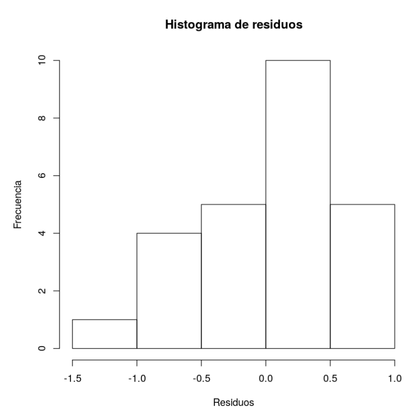
Ejercicio A, página 198.

(1)

```
y = c(10.98,11.13,12.51,8.40,9.27,8.73,6.36,8.50,
       7.82,9.14,8.24,12.19,11.88,9.57,10.94,9.58,
       10.09,8.11,6.83,8.88,7.68,8.47,8.86,10.36,
       11.08)
ypred = c(10.86,10.47,11.79,8.72,9.13,8.20,6.18,8.40,
          8.04,9.22,9.64,11.54,11.60,9.18,10.61,9.49,
          9.49,8.29,6.51,8.56,7.74,9.38,9.77,11.00,
          11.76)
res = c(0.12,0.66,0.72,-0.32,0.14,0.53,0.18,0.10,-0.22,-0.08,-1.40,0.65,0.28,0.39,0.33,0.
 ↪09,0.60,-0.18,0.32,0.32,-0.06,-0.91,-0.91,-0.64,-0.68)
```

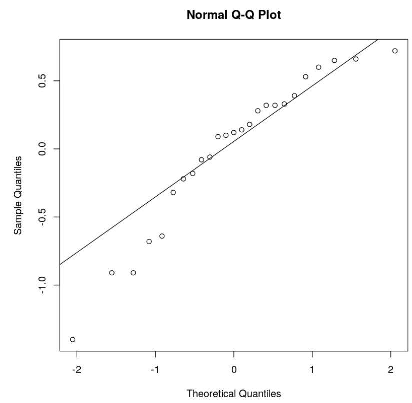
Histograma

```
hist(x=res,main="Histograma de residuos",xlab="Residuos",ylab="Frecuencia")
```



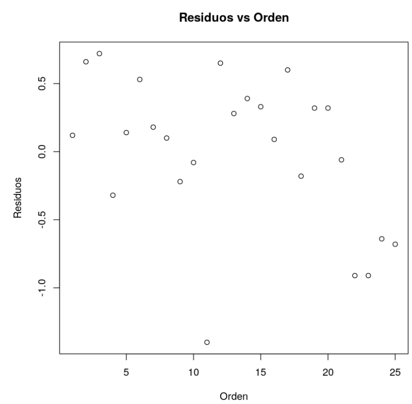
Nscore

```
qqnorm(res)
qqline(res)
```



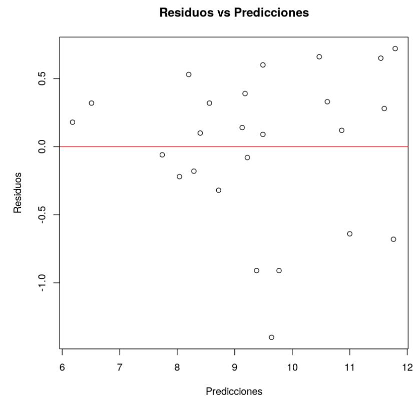
Residuos vs Orden

```
plot(x=res,main = "Residuos vs Orden",xlab="Orden",ylab="Residuos")
```



Residuos vs Prediccion

[9]:
```
plot(x=ypred,y=res,main="Residuos vs Predicciones",xlab="Predicciones",ylab="Residuos")
abline(h=0,col=2)
```



(2)

- **Histograma** Hay mas valores negativos que positivos.
- **Nscore** No se aprecia una simetría de los datos en función del 0, parece claramente estar mas cargado hacia un lado.
- **Residuos v/s Orden** No se ve un patrón o tendencia, probablemente podríamos decir que es un ruido pero hay un valor atípico muy bajo en comparación a los demás.
- **Residuos v/s Predicción** Aparece nuevamente el valor particularmente bajo, y se ve que los valores están mas presentes entre el 9 y 12.

(3)

Notemos que:

```
D = sum((res[2:25]-res[1:24])^2)/sum(res^2)
print(D)
```

```
[1] 1.39289
```

Notemos que $1{,}39289 \in [1{,}654; 1{,}123]$ por tanto usando la tabla D-W con 5% de significancia el estadístico se tiene que:

La prueba no es concluyente.

Observación D-W es una estadística de prueba que se utiliza para detectar la presencia de auto-correlación (una relación entre los valores separados el uno del otro por un intervalo de tiempo dado) en los residuos (errores de predicción).

(4)

Notemos que:

```
posi = length(res[res>0])
neg = length(res[res<0])
n1 = min(posi,neg)
n2 = max(posi,neg)
r = 1
for(i in 1:24)
  if(res[i]*res[i+1]< 0){
    r = r +1
  }
u <- 2*n1*n2/(n1+n2) + 1
sigma <- 2*n1*n2*(2*n1*n2-n1-n2)/((n1+n2)^2*(n1+n2-1))
z <- (r-u+0.5)/sigma^0.5
pvalor <- 2*pnorm(z)
print(pvalor)
```

```
[1] 0.05500883
```

Para el test de rachas tenemos las hipotesis

$$H_0\text{: Los errores son independientes vs } H_1\text{:Los errores no son independientes}$$

En la prueba de Durbin-Watson anterior se obtuvo que es inconcluso, sin embargo a medida si tomamos una significancia del 1% se tiene que $1{,}39289 \in [0{,}90, 1{,}41]$ por lo que no se concluiría, sin embargo se puede ver que el estadístico está muy cercano a la zona de no rechazo de que $\rho = 0$(hipotesis nula del test Durbin Watson $\rho = 0$), es decir, se podría esperar que para significancias menores al 1% no rechazo la hipotesis de que los errores son independientes.
Con el test de rachas, se obtiene un p-valor de 0,05500883 que es el valor mínimo de significancia para rechazar H_0, por lo que para significancias menores al 5,5%, en particular, menores al 1% no se rechaza H_0, es decir ,no se rechaza la independencia de los errores, lo que es corrobora lo que se esperaba con el test de Durbin Watson.

1. Ejercicio H, Página 274.

```
x1 <- c(1.76,1.55,2.73,2.73,2.56,2.8,2.8,1.84,2.16,1.98,0.59,0.8,0.8,1.05,1.8,1.8,1.77,2.
  3,2.03,1.91,1.91,1.91,0.76,2.13,2.13,1.51,2.05)
x2 <- c(0.07,0.07,0.07,0.07,0.07,0.07,0.07,0.07,0.07,0.02,0.02,0.02,0.02,0.02,0.02,0.
  02,0.02,0.02,0.474,0.474,0.474,0.474,0.474,0.474,0.474,0.474,0.474)
x3 <- c(7.8,8.9,8.9,7.2,8.4,8.7,7.4,8.7,8.8,7.6,6.5,6.7,6.2,7,7.3,6.5,7.6,8.2,7.6,8.3,8.
  2,6.9,7.4,7.6,6.9,7.5,7.6)
y <- c(110.4,102.8,101,108.4,100.7,100.3,102,93.7,98.9,96.6,99.4,96.2,99,88.4,75.3,92,82.
  4,77.1,74,65.7,56.8,62.1,61,53.2,59.4,58.7,58)
```

Aplicamos una transformación al modelo:

$$Y = \alpha X_1^{\beta} X_2^{\gamma} X_3^{\delta} \epsilon \Longrightarrow \log(Y) = \alpha + \beta \log(X_1) + \gamma \log(X_2) + \delta \log(X_3) + \log(\epsilon)$$

El cual es lineal para α, β, γ y δ luego ajustamos un modelo de regresión lineal:

```
logy <- log(y)
logx1 <-log(x1)
logx2 <- log(x2)
logx3 <- log(x3)
regr <- lm(logy~logx1+logx2+logx3)
summary(regr)
```

```
Call:
lm(formula = logy ~ logx1 + logx2 + logx3)
Residuals:
     Min       1Q   Median       3Q      Max
-0.32252 -0.07841  0.01167  0.10143  0.25879
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.00845    0.72881   4.128 0.000409 ***
logx1        0.03499    0.09001   0.389 0.701030
logx2       -0.14425    0.02405  -5.999 4.07e-06 ***
logx3        0.50649    0.36763   1.378 0.181555
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.1575 on 23 degrees of freedom
Multiple R-squared:  0.6148,Adjusted R-squared:  0.5646
F-statistic: 12.24 on 3 and 23 DF,  p-value: 5.453e-05
```

Para mantener el orden llamaremos modelo 1 a esta regresión.

```
anderson_model <- lm(y~x1+x2+x3+x1*x2)
summary(anderson_model)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x1 * x2)
Residuals:
    Min      1Q  Median      3Q     Max
-22.155  -3.770   1.458   5.503  16.983
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   82.173     20.338   4.040 0.000547 ***
x1             2.463      4.722   0.522 0.607190
x2           -75.378     39.144  -1.926 0.067168 .
x3             1.584      3.122   0.507 0.616997
x1:x2         -1.374     21.265  -0.065 0.949058
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 10.1 on 22 degrees of freedom
Multiple R-squared:  0.7549,Adjusted R-squared:  0.7103
F-statistic: 16.94 on 4 and 22 DF,  p-value: 1.784e-06
```

Para mantener el orden llamaremos modelo 2 a esta regresión. Con la finalidad de comparar los modelos notemos que:

1. El R^2 y R^2-ajustado es mayor en el modelo 2 que en el modelo 1.

Notemos que esto al parecer es lo único que podemos utilizar para argumentar correctamente que un modelo es mejor que el otro. Aún así existen similitudes negativas entre los modelos como por ejemplo que los valores p de las variables explicativas del modelo 2 son muy altas, en particular X_1X_2 no aporta casi nada a la explicación de la variable de respuesta. Recomendaría aplicar la rutina backward para estos modelos.

```
nuevo_modelo_1 = stepAIC(regr, trace=FALSE, direction="backward")
summary(nuevo_modelo_1)
```

```
Call:
lm(formula = logy ~ logx2 + logx3)
Residuals:
     Min       1Q   Median       3Q      Max
-0.31641 -0.08578  0.02058  0.09414  0.27945
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.87068    0.62548   4.590 0.000118 ***
```

```
logx2       -0.14288    0.02336  -6.115 2.57e-06 ***
logx3        0.58545    0.30096   1.945 0.063549 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1547 on 24 degrees of freedom
Multiple R-squared:  0.6123,Adjusted R-squared:   0.58
F-statistic: 18.95 on 2 and 24 DF,  p-value: 1.154e-05
```

```
nuevo_modelo_2 = stepAIC(anderson_model, trace=FALSE, direction="backward")
summary(nuevo_modelo_2)
```

```
Call:
lm(formula = y ~ x2)
Residuals:
     Min       1Q   Median       3Q      Max
-21.9819  -4.5396   0.1604   7.1104  17.0104
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   98.839      2.563  38.557  < 2e-16 ***
x2           -77.846      9.259  -8.408 9.38e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 9.779 on 25 degrees of freedom
Multiple R-squared:  0.7388,Adjusted R-squared:  0.7283
F-statistic: 70.69 on 1 and 25 DF,  p-value: 9.375e-09
```

Estos 2 nuevos modelos son mas simples, tienen un poco menos de R^2 pero tienen un poco mas de $R^2 - ajustado$. Aún así se prefiere el modelo 2 al tener mejor R^2 y R^2-ajustado puesto que a pesar de aplicar la rutina backward, modelo 2 sigue siendo mejor y con valores p pequeños tanto para el test global como para las variables explicativas.

Conclusión: El modelo 2 es mejor que el modelo 1. Es decir es mejor el modelo Anderson.

Ejercicio C, página 250

Notemos que tenemos n=46, y el número de parámetros es 6, entones $v_1 = 5$ y $v_2 = 40$, y necesitamos F tal qué, como mínimo, cumpla $0{,}9 = \frac{5F}{5F+40}$, que es $F = 72$. Por lo tanto el valor de F tal que se tenga un R^2 de como mínimo 90 % debe ser 72.

```
72/qf(1-0.05,5,40)
```

29.3941567127951

Luego por lo anterior indica que al valor $F(5,20)_{5\%}$ que es el cuantil 0,95 de la distribución F debe agrandarse aproximadamente 29 veces para alcanzar R^2 superior al 0.9.

Ejercicio D, página 233

```
t = c(12, 23, 7, 8, 17, 22, 1, 11, 19, 20, 5, 2, 21, 15, 18, 3, 6, 10, 4, 9, 13, 14, 16)
Y = c(2.3, 1.8, 2.8, 1.5, 2.2, 3.8, 1.8, 3.7, 1.7, 2.8, 2.8, 2.2, 3.2, 1.9, 1.8, 3.5, 2.
↪8, 2.1, 3.4, 3.2, 3.0, 3.0, 5.9)
X = c(1.3, 1.3, 2.0, 2.0, 2.7, 3.3, 3.3, 3.7, 3.7, 4.0, 4.0, 4.0, 4.7, 4.7, 5.0, 5.3, 5.
↪3, 5.3, 5.7, 6.0, 6.0, 6.3, 6.7)
data <- data.frame(Y, X)
rownames(data) <- t
```

Realizamos la regresión considerando que se cumple la hipótesis de homoseasticidad (igual varianza de los errores)

```
summary(lm(Y~X,data=data))
```

```
Call:
lm(formula = Y ~ X, data = data)
Residuals:
     Min       1Q   Median       3Q      Max
-1.20457 -0.52300 -0.07827  0.34543  2.35859
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4256     0.5127   2.780   0.0112 *
X             0.3158     0.1149   2.749   0.0120 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.853 on 21 degrees of freedom
Multiple R-squared:  0.2647,Adjusted R-squared:  0.2297
F-statistic: 7.559 on 1 and 21 DF,  p-value: 0.01202
```

Ahora realicemos la regresión con pesos (mínimos cuadrados ponderados, página 33 del apunte.)

```
pesos = c(rep(1,length(X)-1), 0.25)
```

```
summary(lm(Y~X, data=data, weights=pesos))
```

```
Call:
lm(formula = Y ~ X, data = data, weights = pesos)
Weighted Residuals:
     Min       1Q   Median       3Q      Max
-1.04297 -0.54012 -0.06731  0.41216  1.33840
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

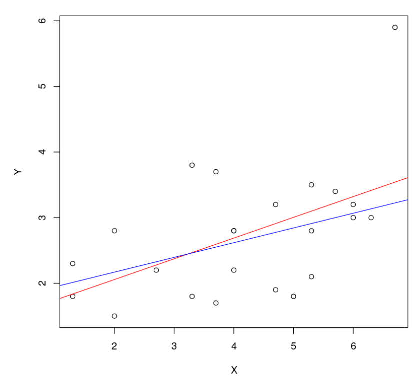
```
(Intercept)  1.72129    0.43673   3.941 0.000748 ***
X            0.22434    0.09997   2.244 0.035728 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.7086 on 21 degrees of freedom
Multiple R-squared:  0.1934,Adjusted R-squared:  0.155
F-statistic: 5.036 on 1 and 21 DF,  p-value: 0.03573
```

Obteniendo la ecuación:

$$\hat{Y} = 1{,}72129 + 0{,}22434X$$

Notemos que los valores de los parametros cambian, al igual que sus R^2. El R^2 bajo se puede deber a la poca cantidad de datos.

```
plot(X,Y, title ='')
abline(lm(Y~X, data=data), col='red')
abline(lm(Y~X, data=data, weights=wts), col='blue')
```



Note que ultimo dato al parecer es un dato atípico o un outlier, por tanto en vez de eliminar el dato se decidió ponderarlo.

Ejercicio 12.7, página 269

Leemos la data:

```
data(titanic)
attach(titanic)
```

12.7.1

```
log = glm(cbind(Surv,N-Surv)~Class+Age+Sex, data=titanic,family=binomial())
summary(log)
```

```
Call:
glm(formula = cbind(Surv, N - Surv) ~ Class + Age + Sex, family = binomial(),
    data = titanic)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.1356  -1.7126   0.7812   2.6800   4.3833
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1862     0.1586   7.480 7.40e-14 ***
ClassFirst    0.8577     0.1573   5.451 5.00e-08 ***
ClassSecond  -0.1604     0.1738  -0.923    0.356
ClassThird   -0.9201     0.1486  -6.192 5.93e-10 ***
AgeChild      1.0615     0.2440   4.350 1.36e-05 ***
SexMale      -2.4201     0.1404 -17.236  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 671.96  on 13  degrees of freedom
Residual deviance: 112.57  on  8  degrees of freedom
AIC: 171.19
Number of Fisher Scoring iterations: 5
```

Al parecer una gran cantidad de mujeres sobrevivieron, excepto en tercera clase, donde el ratio (dado por el parámetro) fue mucho mejor, ¿Podría esto dar una idea de correlación entre Class y Sex?,¿Puede que existan otras relaciones?

```
nlog = update(log,~(Class+Sex+Age)^2)
```

Usando la prueba chi cuadrado quitame las variables cruzadas que sean irrelevantes

```
drop1(nlog,test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
<none>	NA	2.513008e-10	70.62104	NA	NA
Class:Sex	3	6.501316e+01	129.63420	65.013158	4.983623e-14
Class:Age	2	3.726253e+01	103.88357	37.262529	8.101110e-09
Sex:Age	1	1.685397e+00	70.30644	1.685397	1.942088e-01

Según el análisis anterior la variable Age:Sex es irrelevante por lo que se elimina del modelo. Luego quítame la variable Age:Sex y luego quítame cualquier otra variable irrelevante.

```
nnlog = update(nlog,~.-Age:Sex)
drop1(nnlog,test="Chisq")
```

	Df	Deviance	AIC	LRT	Pr(>Chi)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
<none>	NA	1.685397	70.30644	NA	NA
Class:Sex	3	76.904040	139.52508	75.21864	3.252707e-16
Class:Age	2	45.899201	110.52024	44.21380	2.506655e-10

Vemos que nada ha cambiado, por lo que m3 es el modelo final,finalmente realizamos un summary y veamos que ocurre con los ratio de cada variable

```
summary(nnlog)
```

```
Call:
glm(formula = cbind(Surv, N - Surv) ~ Class + Sex + Age + Class:Sex +
    Class:Age, family = binomial(), data = titanic)
Deviance Residuals:
       1          2          3          4          5          6          7          8
 0.00000    0.00000    0.00000    0.00005    0.00000    0.00000    0.00001    0.00007
       9         10         11         12         13         14
 0.00000    0.00000   -0.87452    0.82651    0.38065   -0.30431
Coefficients: (1 not defined because of singularities)
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           1.897e+00  6.191e-01   3.064 0.002183 **
ClassFirst            1.658e+00  8.003e-01   2.072 0.038264 *
ClassSecond          -8.004e-02  6.876e-01  -0.116 0.907325
ClassThird           -2.115e+00  6.370e-01  -3.319 0.000902 ***
SexMale              -3.147e+00  6.245e-01  -5.039 4.68e-07 ***
AgeChild              3.379e-01  2.692e-01   1.255 0.209391
ClassFirst:SexMale   -1.136e+00  8.205e-01  -1.385 0.166162
ClassSecond:SexMale  -1.068e+00  7.466e-01  -1.431 0.152539
ClassThird:SexMale    1.762e+00  6.516e-01   2.704 0.006860 **
ClassFirst:AgeChild   2.242e+01  1.650e+04   0.001 0.998915
ClassSecond:AgeChild  2.442e+01  1.301e+04   0.002 0.998502
ClassThird:AgeChild          NA         NA      NA       NA
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 671.9622  on 13  degrees of freedom
Residual deviance:   1.6854  on  3  degrees of freedom
AIC: 70.306
Number of Fisher Scoring iterations: 21
```

A partir del summary vemos que el ratio de perdida de los de primera y segunda clase es mas grande que los de tercera clase, análogamente las mujeres de primera clase tienen mayor ratio que las mujeres de tercera clase por lo que tiene mayor sobrevivientes en general las mujeres tienen mayor ratio que los hombres ya que el parámetro que acompaña el sexo masculino es negativo. En el caso de los niños y las clases, se nota que en la tercera clase de los niños es NA, eso es consistente porque solo hay tres clases para los niños en lugar de cuatro, y también se deduce que niños de primera clase sobrevivieron mas que los niños de tercera clase porque su parámetro que lo acompaña es positivo.

2. Ejercicio C, página 318

(1)

```
velocidad <- c(150,275,200,150,175,200,150,175,200)
apariencia <- c(255,246,249,260,223,231,265,247,256)
```

Como tengo 3 variables cualitativas, utilizo dos variables dummy. A mano seria de la siguiente forma:

```
z1 <- c(1,1,1,0,0,0,0,0,0)
z2 <- c(0,0,0,1,1,1,0,0,0)
operador <- c(1,1,1,2,2,2,3,3,3)
dummy <- lm(apariencia~z1+z2)
summary(dummy)
```

```
Call:
lm(formula = apariencia ~ z1 + z2)
Residuals:
   Min     1Q Median     3Q    Max
   -15     -7     -1      5     22
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  256.000      7.311  35.018 3.61e-08 ***
z1            -6.000     10.339  -0.580    0.583
z2           -18.000     10.339  -1.741    0.132
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 12.66 on 6 degrees of freedom
Multiple R-squared:  0.3438,Adjusted R-squared:  0.1251
F-statistic: 1.572 on 2 and 6 DF,  p-value: 0.2826
```

Y mediante la función de R:

```
summary(lm(appearance~dummy(operator),data=data))
```

```
Call:
lm(formula = appearance ~ dummy(operator), data = data)
Residuals:
   Min     1Q Median     3Q    Max
   -15     -7     -1      5     22
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)       250.000      7.311  34.197 4.16e-08 ***
dummy(operator)2  -12.000     10.339  -1.161    0.290
dummy(operator)3    6.000     10.339   0.580    0.583
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 12.66 on 6 degrees of freedom
Multiple R-squared:  0.3438,Adjusted R-squared:  0.1251
F-statistic: 1.572 on 2 and 6 DF,  p-value: 0.2826
```

Nota que esta fija de forma distinta las variables, pero se obtienen los mismos resultados de $R^2, R^2 - ajustado$ y F^* pues a pesar de ser otra ecuación, explica la misma cantidad de información en conjunto.

(2)

```
summary(aov(dummy))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
z1           1     18    18.0   0.112  0.749
z2           1    486   486.0   3.031  0.132
Residuals    6    962   160.3
```

(3)

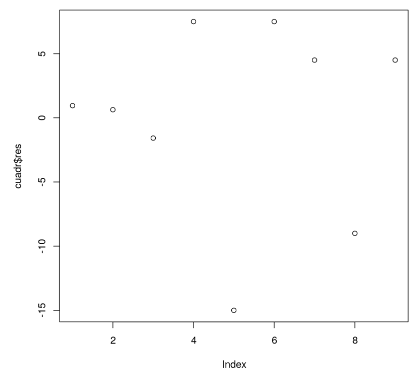
```
vel_model <- lm(apariencia~velocidad)
summary(vel_model)
```

```
Call:
lm(formula = apariencia ~ velocidad)
Residuals:
    Min      1Q  Median      3Q     Max
-26.143  -2.143   3.286   8.286  13.286
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 267.1429    23.2969  11.467 8.62e-06 ***
velocidad    -0.1029     0.1227  -0.838     0.43
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 13.8 on 7 degrees of freedom
Multiple R-squared:  0.09121,Adjusted R-squared:  -0.03862
F-statistic: 0.7026 on 1 and 7 DF,  p-value: 0.4296
```

(4)

Dado los residuos.

[79]:
```
plot(cuadr$res)
```



Proponemos un modelo cuadrático con la finalidad de incrementar el R^2

```
cuadr = lm(apariencia~(velocidad+z1+z2)^2)
summary(cuadr)
```

```
Call:
lm(formula = apariencia ~ (velocidad + z1 + z2)^2)
Residuals:
      1        2        3        4        5        6        7        8
 0.9474   0.6316  -1.5789   7.5000 -15.0000   7.5000   4.5000  -9.0000
      9
 4.5000
Coefficients: (1 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  287.5000    61.8940   4.645   0.0188 *
velocidad     -0.1800     0.3513  -0.512   0.6437
z1           -23.0263    68.7607  -0.335   0.7598
z2            52.0000    87.5313   0.594   0.5943
velocidad:z1   0.1105     0.3780   0.292   0.7890
velocidad:z2  -0.4000     0.4968  -0.805   0.4796
z1:z2              NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 12.42 on 3 degrees of freedom
Multiple R-squared:  0.6843,Adjusted R-squared:  0.1582
F-statistic: 1.301 on 5 and 3 DF,  p-value: 0.4412
```

De esta forma aumentamos casi al doble el R^2 y un poco el $R^2 - ajustado$. Pero el valor-p de la significancia global del modelo (test-F) es demasiado grande por tanto recomendaría aplicar una rutina backward para dejar las mejores variables explicativas.

```
final = stepAIC(cuadr, trace=FALSE, direction="backward")
```

```
summary(final)
```

```
Call:
lm(formula = apariencia ~ velocidad + z2 + velocidad:z2)
Residuals:
      1       2       3       4       5       6       7       8       9
 -2.135   1.269  -3.173   7.500 -15.000   7.500   7.865  -7.654   3.827
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  272.01923   18.64991  14.586 2.74e-05 ***
velocidad     -0.09923    0.09500  -1.045    0.344
z2            67.48077   52.68453   1.281    0.256
velocidad:z2  -0.48077    0.29536  -1.628    0.165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 9.888 on 5 degrees of freedom
Multiple R-squared:  0.6666,Adjusted R-squared:  0.4665
F-statistic: 3.332 on 3 and 5 DF,  p-value: 0.114
```

De esta forma tenemos un modelo con valor p para el test F mucho mejor y con R^2 ajustado mejor que el modelo anterior. Por tanto este modelo es el que proponemos para el problema.