Reporte sobre el consumo de agua

Control 2

0.1 Introducción

En el presente reporte se analizará el consumo de agua de cierta ciudad con respecto a su consumo de energía. Se iniciará el análisis con una visualización del conjunto de datos. Posteriormente se ajustará un modelo lineal a partir de los datos y por último se validará el modelo. Como se mostrará a continuación, será necesario considerar distintos modelos lineales transformando la variable de salida (el consumo de agua). Dicha transformación sera determinada a partir de los resultados del análisis de residuales.

0.2 Gráfica de los datos

La fig.1 es una gráfica de los datos en donde cada observación se representa como un punto en el plano, el eje horizontal corresponde al consumo de energía y el vertical al consumo de agua. Se puede observar una correlación positiva entre las variables. Además, parece ser que, conforme se avanza en la dirección positiva del eje horizontal, los datos se muestran más dispersos verticalmente. Lo anterior puede ser un indicio de que no se cumple el supuesto de homocedasticidad.

in Figure 2 we see examples of plotting in R.

0.3 Primer ajuste del modelo lineal

Se procederá ajustando un modelo lineal. En la fig.2 se muestran los datos
y la recta ajustada por medio de mínimos cuadrados.

En la fig.4 se despliega la gráfica de residuales contra los valores ajustados. Aproximadamente en el intervalo [8, 16] del eje horizontal, los puntos se muestran más dispersos verticalmente. En el caso de homocedasticidad los puntos no deberían mostrar ningún patron discernible. En este caso parece ser que el supuesto no se cumple

0.4 Transformación Box-Cox

```r
lim_inf = -3
lim_sup = 3
step= 0.005
alpha = 0.1
y = c.agua
lambdas = seq(from=lim_inf, to = lim_sup, by = step)
geo_mean = reduce(y, function(y1,y2) y1*y2)
geo_mean = geo_mean ^ ( 1/length(y))
y_arr = as.array(y)
rss <- array(1:length(lambdas))

box_cox_transf <- function(y, lambda){
  if(lambda != 0)
    y^lambda
  else
    log(y)
}

for (i in 1:length(lambdas)) {
  lambda <- lambdas[i]
```

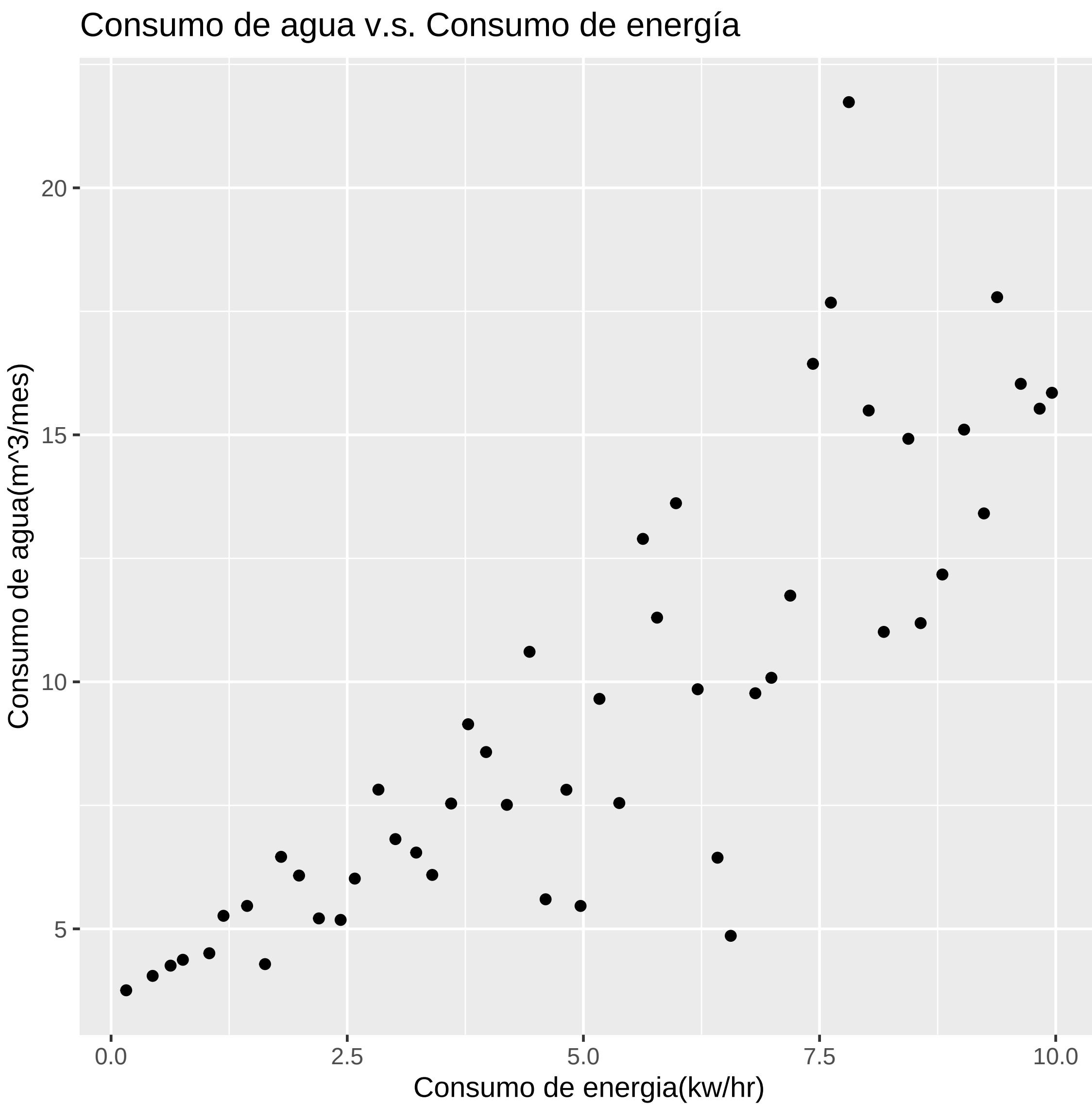


Figure 1: Consumo de agua v.s. Consumo de energía

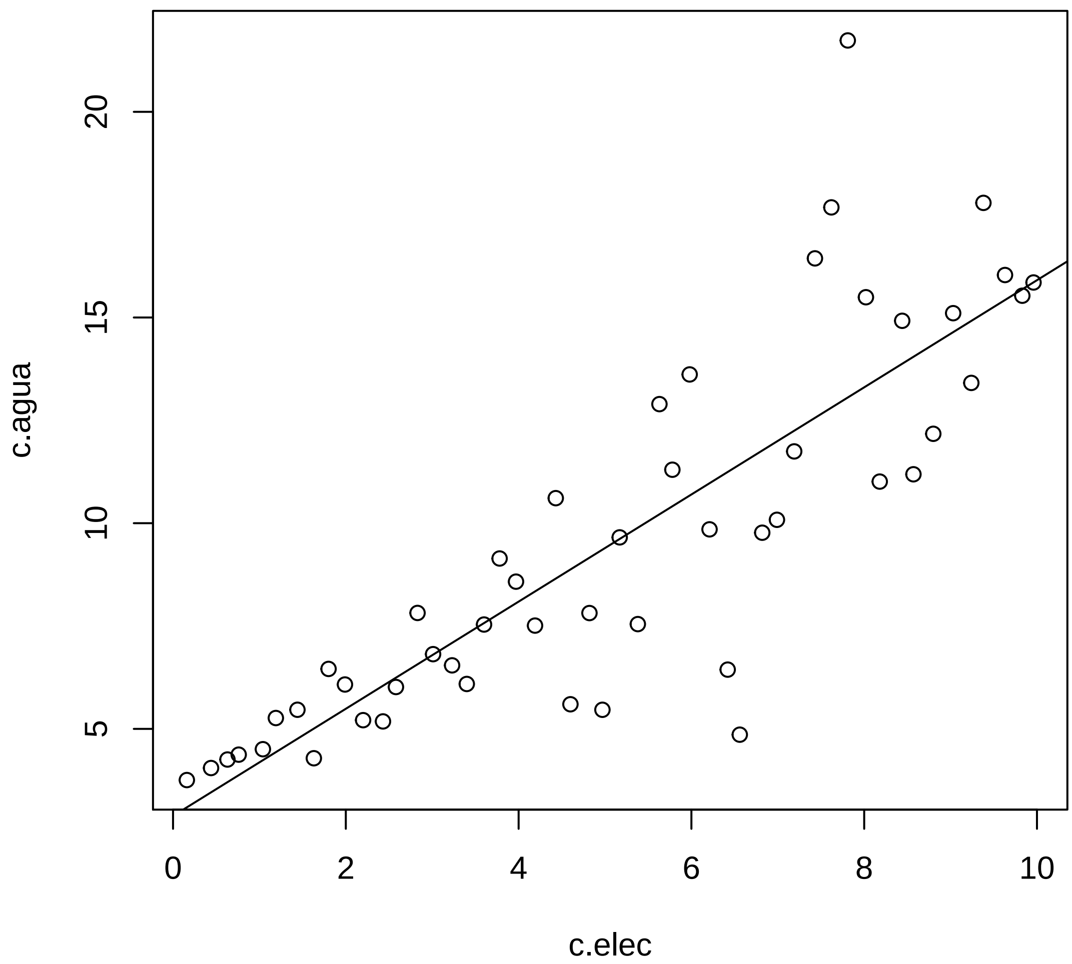
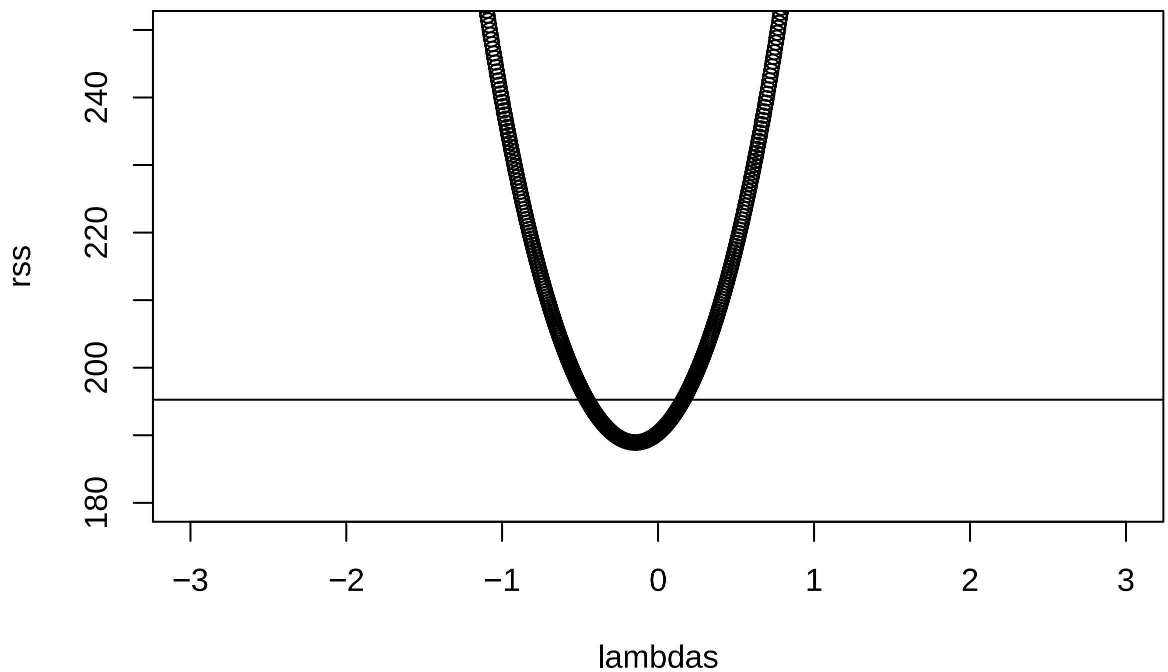


Figure 2: Consumo de agua v.s. Consumo de energía (con recta ajustada)

```
  if( lambda == 0){
      y_lambda = apply(y_arr, 1, function(y) geo_mean*log(y))
  }else{
    y_lambda = apply(y_arr, 1, function(y) (y^lambda)/(lambda*geo_mean^(lambda-1)) )
  }
  fm <- lm(y_lambda ~ c.elec)
  rss[i] = sum(resid(fm)^2)


}
plot(rss ~ lambdas, ylim =c(180, 250))
y_ic = min(rss)*(1 + qt(1-alpha/2, length(y) -2 )/length(y-2))
abline( h = y_ic)
```




```
ic <- match(rss[rss < y_ic], rss)
print("Lambdas en el ic")
```

```
## [1] "Lambdas en el ic"
```

```
print(lambdas[ic])
```

```
##   [1] -0.455 -0.450 -0.445 -0.440 -0.435 -0.430 -0.425 -0.420 -0.415 -0.410
##  [11] -0.405 -0.400 -0.395 -0.390 -0.385 -0.380 -0.375 -0.370 -0.365 -0.360
##  [21] -0.355 -0.350 -0.345 -0.340 -0.335 -0.330 -0.325 -0.320 -0.315 -0.310
##  [31] -0.305 -0.300 -0.295 -0.290 -0.285 -0.280 -0.275 -0.270 -0.265 -0.260
##  [41] -0.255 -0.250 -0.245 -0.240 -0.235 -0.230 -0.225 -0.220 -0.215 -0.210
##  [51] -0.205 -0.200 -0.195 -0.190 -0.185 -0.180 -0.175 -0.170 -0.165 -0.160
```

```
##  [61] -0.155 -0.150 -0.145 -0.140 -0.135 -0.130 -0.125 -0.120 -0.115 -0.110
##  [71] -0.105 -0.100 -0.095 -0.090 -0.085 -0.080 -0.075 -0.070 -0.065 -0.060
##  [81] -0.055 -0.050 -0.045 -0.040 -0.035 -0.030 -0.025 -0.020 -0.015 -0.010
##  [91] -0.005  0.000  0.005  0.010  0.015  0.020  0.025  0.030  0.035  0.040
## [101]  0.045  0.050  0.055  0.060  0.065  0.070  0.075  0.080  0.085  0.090
## [111]  0.095  0.100  0.105  0.110  0.115  0.120  0.125  0.130  0.135  0.140
## [121]  0.145  0.150  0.155  0.160
```

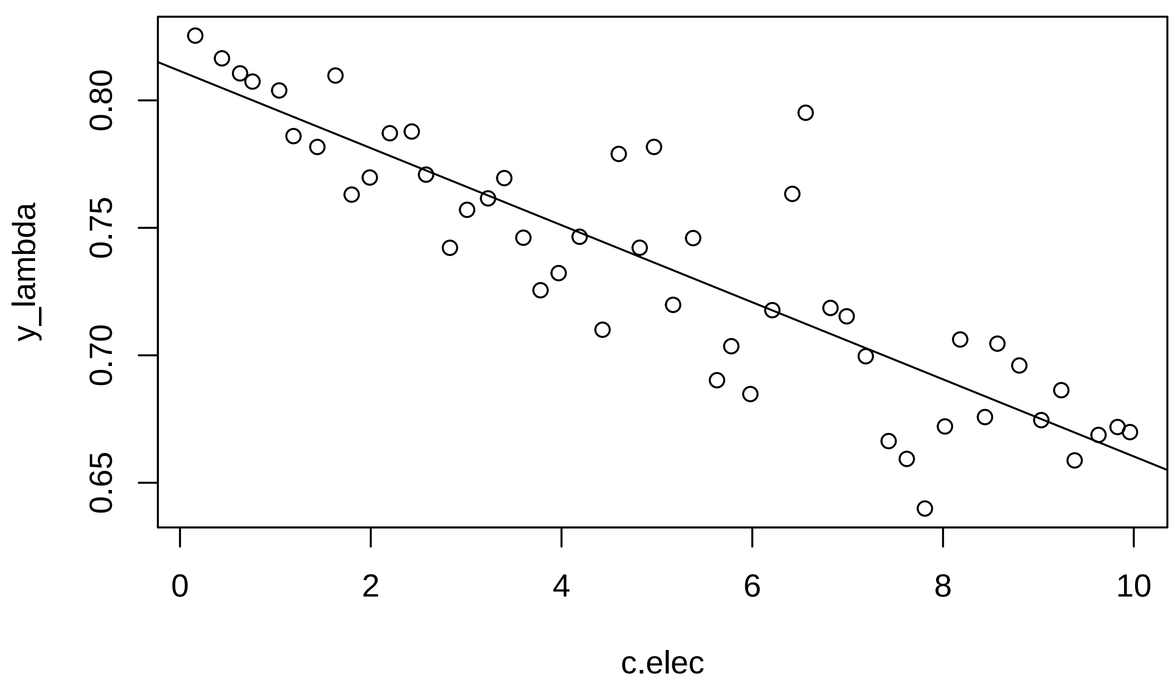
```r
min(rss)
```

```
## [1] 188.9265
```

```r
min_lambda <- lambdas[which.min(rss)]
paste("lambda", min_lambda)
```

```
## [1] "lambda -0.145"
```

```r
 y_lambda = box_cox_transf(y_arr, min_lambda)
 fm <- lm( y_lambda ~ c.elec )
 plot( y_lambda ~ c.elec )
 abline(fm)
```



Intervalos de confianza

Para calcular los intervalos de confianza utilizamos el primer modelo.

```r
alpha <- 0.1
x <- c.elec
n <- length(y)
fm <- lm(c.agua ~ c.elec)
```

```r
b0 <- coef(fm)[1]
b1 <- coef(fm)[2]
x0 <- 7.57
y_hat <- b0 + b1*x0
y_hat
```

```
## (Intercept)
##    12.74271
```

```r
quant <- qt(1-alpha/2, n -2 )
quant
```

```
## [1] 1.677224
```

```r
print("Intervalo de confianza al nivel")
```

```
## [1] "Intervalo de confianza al nivel"
```

```r
print(alpha)
```

```
## [1] 0.1
```

```r
ms_res = sum(resid(fm)^2)/(n-2)
se_y = sqrt(ms_res*(1/n + (x0 - mean(x))^2/sum((x-mean(x))^2) ) )
se_y
```

```
## [1] 0.4575344
```

```r
print(y_hat - quant*se_y)
```

```
## (Intercept)
##    11.97532
```

```r
print(y_hat + quant*se_y)
```

```
## (Intercept)
##    13.51009
```

```r
alpha <- 0.05
quant <- qt(1-alpha/2, n -2 )
print("Intervalo de confianza al nivel")
```

```
## [1] "Intervalo de confianza al nivel"
```

```r
print(alpha)
```

```
## [1] 0.05
```

```r
x0 <- 5.1
y_hat <- b0 + b1*x0
se_y = sqrt(ms_res*(1+ 1/n + (x0 - mean(x))^2/sum((x-mean(x))^2) ) )
print(y_hat - quant*se_y)
```

```
## (Intercept)
##     4.53151
```

```r
print(y_hat + quant*se_y)
```

```
## (Intercept)
##    14.52039
```