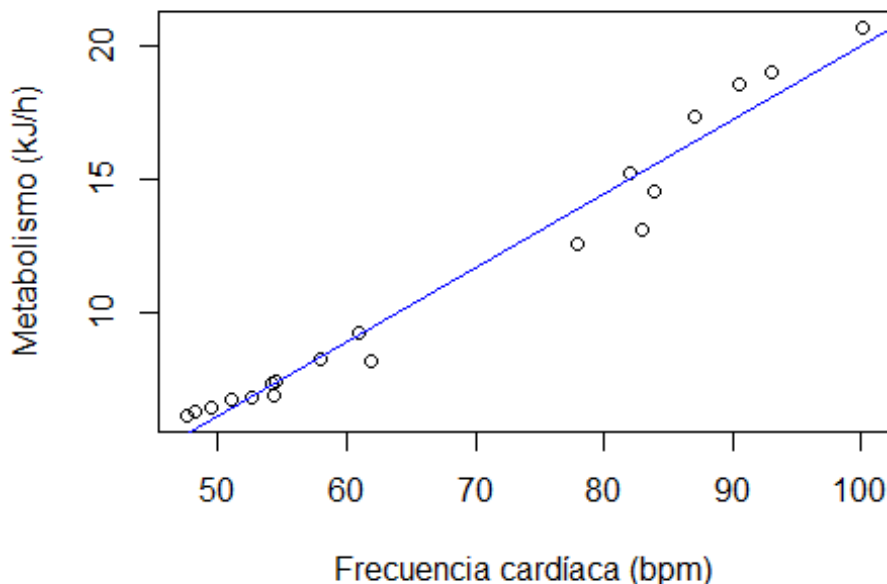


Intro ML Diego VZ

2023-03-01

```
heartbpm <- c(47.53, 48.27, 49.51, 51.09, 52.57, 54.30,  
54.25, 54.45, 57.95, 60.92, 61.91, 77.92,  
82.07, 82.95, 83.94, 86.96, 90.42, 92.93, 100.05)  
metabol <- c(6.15, 6.31, 6.43, 6.78, 6.86, 6.90, 7.37, 7.41,  
8.24, 9.22, 8.16, 12.61, 15.26, 13.09, 14.59,  
17.35, 18.57, 19.00, 20.70)  
vulture <- data.frame(heartbpm, metabol)  
rm(heartbpm, metabol)  
attach(vulture)  
# (a) Nube de puntos  
plot(heartbpm, metabol, main = "Relación entre la frecuencia cardíaca y  
el metabolismo",  
xlab = "Frecuencia cardíaca (bpm)", ylab = "Metabolismo (kJ/h)")  
# (b) Recta de regresión lineal  
modelo <- lm(metabol ~ heartbpm)  
abline(modelo, col = "blue")
```

Relación entre la frecuencia cardíaca y el metabolismo



```
# (c) Gráfico de residuos vs. valores ajustados  
plot(modelo, which = 1)  
# (d) Curva de regresión parabólica  
curve(modelo$coefficients[1] + modelo$coefficients[2]*x +
```

```

modelo$coefficients[3]*x^2,
add = TRUE, col = "red")

datos_ts <- c(114, 134, 124, 128, 116, 120, 138, 130, 139, 125, 132, 130,
140, 144, 110, 148, 124, 136, 150, 120, 144, 153, 134, 152, 158, 124,
128, 138, 142, 160, 135, 138, 142, 145, 149, 156, 159, 130, 157, 142,
144, 160, 174, 156, 158, 174, 150, 154, 165, 164, 168, 140, 170, 185,
154, 169, 172, 144, 162, 158, 162, 176, 176, 158, 170, 172, 184, 175,
180)
datos_edad <- c(17, 18, 19, 19, 20, 21, 21, 22, 23, 25, 26, 29, 33, 33,
34, 35, 36, 36, 38, 39, 39, 40, 41, 41, 41, 42, 42, 42, 44, 44, 45, 45,
46, 47, 47, 47, 47, 48, 48, 50, 50, 51, 51, 52, 53, 55, 56, 56, 56, 57,
57, 59, 59, 60, 61, 61, 62, 63, 64, 65, 65, 65, 66, 67, 67, 68, 68, 69,
70)
datos_tas <- data.frame(tas = datos_ts, Edad = datos_edad)
modelo_1 <- lm(datos_ts~datos_edad)
modelo_1

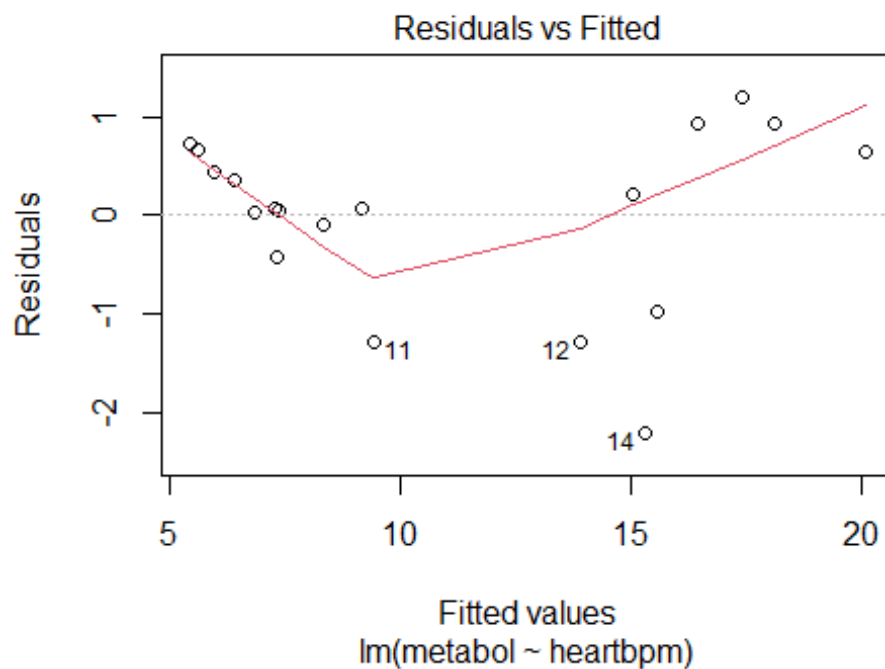
##
## Call:
## lm(formula = datos_ts ~ datos_edad)
##
## Coefficients:
## (Intercept)      datos_edad
##      103.3527         0.9836

summary(modelo_1)

##
## Call:
## lm(formula = datos_ts ~ datos_edad)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.794  -7.024   1.960   8.190  22.634
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  103.3527     4.3261   23.89  <2e-16 ***
## datos_edad    0.9836     0.0892   11.03  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.09 on 67 degrees of freedom
## Multiple R-squared:  0.6447, Adjusted R-squared:  0.6394
## F-statistic: 121.6 on 1 and 67 DF, p-value: < 2.2e-16

library(faraway)

```



teengamb

| ## | sex | status | income | verbal | gamble |
|-------|-----|--------|--------|--------|--------|
| ## 1 | 1 | 51 | 2.00 | 8 | 0.00 |
| ## 2 | 1 | 28 | 2.50 | 8 | 0.00 |
| ## 3 | 1 | 37 | 2.00 | 6 | 0.00 |
| ## 4 | 1 | 28 | 7.00 | 4 | 7.30 |
| ## 5 | 1 | 65 | 2.00 | 8 | 19.60 |
| ## 6 | 1 | 61 | 3.47 | 6 | 0.10 |
| ## 7 | 1 | 28 | 5.50 | 7 | 1.45 |
| ## 8 | 1 | 27 | 6.42 | 5 | 6.60 |
| ## 9 | 1 | 43 | 2.00 | 6 | 1.70 |
| ## 10 | 1 | 18 | 6.00 | 7 | 0.10 |
| ## 11 | 1 | 18 | 3.00 | 6 | 0.10 |
| ## 12 | 1 | 43 | 4.75 | 6 | 5.40 |
| ## 13 | 1 | 30 | 2.20 | 4 | 1.20 |
| ## 14 | 1 | 28 | 2.00 | 6 | 3.60 |
| ## 15 | 1 | 38 | 3.00 | 6 | 2.40 |
| ## 16 | 1 | 38 | 1.50 | 8 | 3.40 |
| ## 17 | 1 | 28 | 9.50 | 8 | 0.10 |
| ## 18 | 1 | 18 | 10.00 | 5 | 8.40 |
| ## 19 | 1 | 43 | 4.00 | 8 | 12.00 |
| ## 20 | 0 | 51 | 3.50 | 9 | 0.00 |
| ## 21 | 0 | 62 | 3.00 | 8 | 1.00 |
| ## 22 | 0 | 47 | 2.50 | 9 | 1.20 |
| ## 23 | 0 | 43 | 3.50 | 5 | 0.10 |
| ## 24 | 0 | 27 | 10.00 | 4 | 156.00 |

```
## 25 0 71 6.50 7 38.50
## 26 0 38 1.50 7 2.10
## 27 0 51 5.44 4 14.50
## 28 0 38 1.00 6 3.00
## 29 0 51 0.60 7 0.60
## 30 0 62 5.50 8 9.60
## 31 0 18 12.00 2 88.00
## 32 0 30 7.00 7 53.20
## 33 0 38 15.00 7 90.00
## 34 0 71 2.00 10 3.00
## 35 0 28 1.50 1 14.10
## 36 0 61 4.50 8 70.00
## 37 0 71 2.50 7 38.50
## 38 0 28 8.00 6 57.20
## 39 0 51 10.00 6 6.00
## 40 0 65 1.60 6 25.00
## 41 0 48 2.00 9 6.90
## 42 0 61 15.00 9 69.70
## 43 0 75 3.00 8 13.30
## 44 0 66 3.25 9 0.60
## 45 0 62 4.94 6 38.00
## 46 0 71 1.50 7 14.40
## 47 0 71 2.50 9 19.20
```

```
summary(teengamb)
```

```
##      sex      status      income      verbal
## Min.   :0.0000   Min.   :18.00   Min.    : 0.600   Min.     : 1.00
## 1st Qu.:0.0000   1st Qu.:28.00   1st Qu.: 2.000   1st Qu.: 6.00
## Median :0.0000   Median :43.00   Median : 3.250   Median : 7.00
## Mean   :0.4043   Mean   :45.23   Mean    : 4.642   Mean     : 6.66
## 3rd Qu.:1.0000   3rd Qu.:61.50   3rd Qu.: 6.210   3rd Qu.: 8.00
## Max.   :1.0000   Max.    :75.00   Max.    :15.000   Max.     :10.00
##      gamble
## Min.    : 0.0
## 1st Qu.: 1.1
## Median  : 6.0
## Mean    :19.3
## 3rd Qu.:19.4
## Max.    :156.0
```

```
cor(teengamb)
```

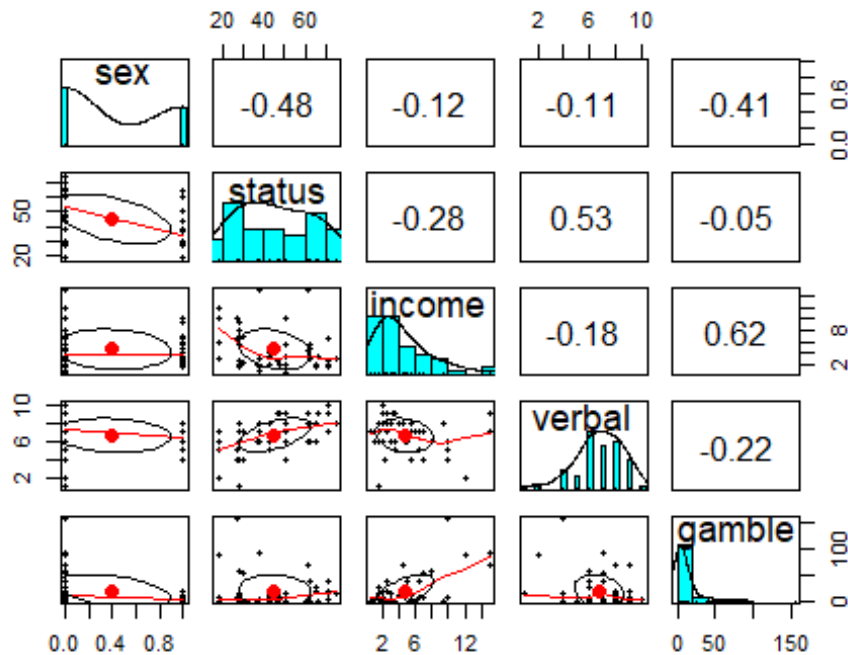
```
##      sex      status      income      verbal      gamble
## sex      1.0000000 -0.48093526 -0.1154586 -0.1069762 -0.40780533
## status -0.4809353  1.00000000 -0.2750340  0.5316102 -0.05042081
## income -0.1154586 -0.27503402  1.0000000 -0.1755707  0.62207690
## verbal -0.1069762  0.53161022 -0.1755707  1.0000000 -0.22005619
## gamble -0.4078053 -0.05042081  0.6220769 -0.2200562  1.00000000
```

```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following object is masked from 'package:faraway':
##
##      logit

pairs.panels(teengamb)
```



#Aquí están las relaciones entre las variables que se pueden inferir a partir de los coeficientes de correlación:

#Hay una correlación negativa moderada entre el género (sex) y el estado socioeconómico (status), lo que sugiere que las mujeres tienden a tener un estado socioeconómico más bajo que los hombres en la muestra.

#Hay una correlación negativa débil entre el estado socioeconómico (status) y el ingreso (income), lo que sugiere que los adolescentes con un mayor estatus socioeconómico tienden a tener mayores ingresos.

#Hay una correlación negativa débil entre el ingreso (income) y el puntaje verbal (verbal), lo que sugiere que los adolescentes con mayores ingresos tienden a tener puntajes verbales más bajos.

#Hay una correlación negativa débil entre el género (sex) y la frecuencia de juego (gamble), lo que sugiere que las mujeres tienden a jugar menos que los hombres en la muestra.

#Hay una correlación positiva moderada entre la frecuencia de juego (gamble) y el ingreso (income), lo que sugiere que los adolescentes con mayores ingresos tienden a jugar más.

```
str(uswages)
```

```
## 'data.frame':    2000 obs. of  10 variables:
## $ wage : num  772 617 958 617 902 ...
## $ educ : int  18 15 16 12 14 12 16 16 12 12 ...
## $ exper: int  18 20 9 24 12 33 42 0 36 37 ...
## $ race : int  0 0 0 0 0 0 0 0 0 0 ...
## $ smsa : int  1 1 1 1 1 1 1 1 1 0 ...
## $ ne   : int  1 0 0 1 0 0 0 0 0 0 ...
## $ mw   : int  0 0 0 0 1 0 0 1 0 1 ...
## $ so   : int  0 0 1 0 0 0 1 0 0 0 ...
## $ we   : int  0 1 0 0 0 1 0 0 1 0 ...
## $ pt   : int  0 0 0 0 0 0 1 1 1 0 ...
```

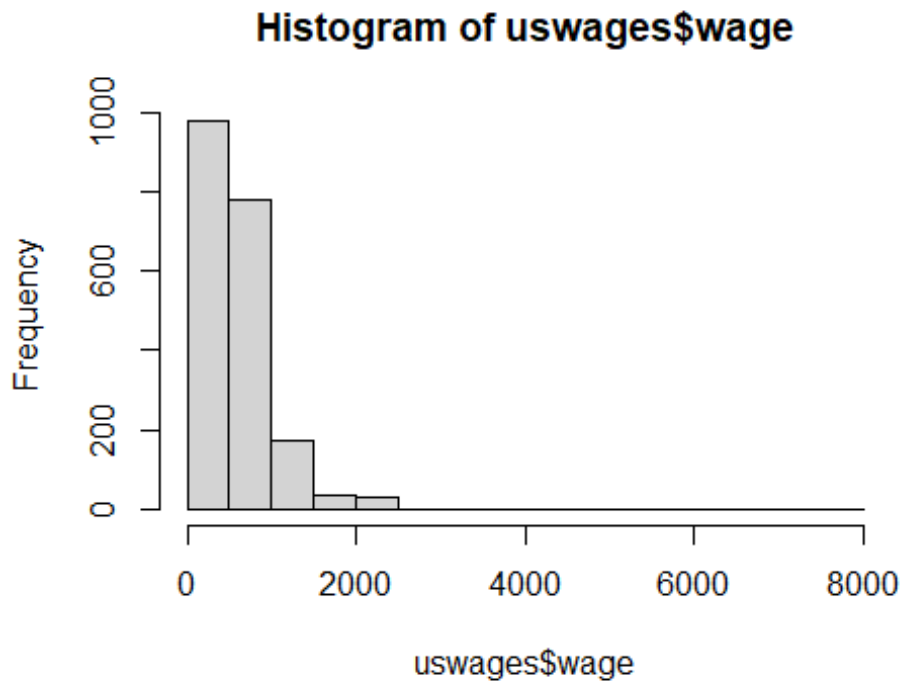
```
head(uswages)
```

```
##           wage educ exper race smsa ne mw so we pt
## 6085    771.60   18   18    0    1  1  0  0  0  0
## 23701   617.28   15   20    0    1  0  0  0  1  0
## 16208   957.83   16    9    0    1  0  0  1  0  0
## 2720    617.28   12   24    0    1  1  0  0  0  0
## 9723    902.18   14   12    0    1  0  1  0  0  0
## 22239   299.15   12   33    0    1  0  0  0  1  0
```

```
summary(uswages)
```

```
##           wage           educ           exper           race
## Min.      : 50.39   Min.      : 0.00   Min.      : -2.00   Min.      : 0.000
## 1st Qu.: 308.64   1st Qu.:12.00   1st Qu.:  8.00   1st Qu.: 0.000
## Median : 522.32   Median :12.00   Median :15.00   Median : 0.000
## Mean     : 608.12   Mean     :13.11   Mean     :18.41   Mean     : 0.078
## 3rd Qu.: 783.48   3rd Qu.:16.00   3rd Qu.:27.00   3rd Qu.: 0.000
## Max.     :7716.05   Max.     :18.00   Max.     :59.00   Max.     : 1.000
##           smsa           ne           mw           so
## Min.      :0.000   Min.      :0.000   Min.      :0.0000   Min.      :0.0000
## 1st Qu.: 1.000   1st Qu.: 0.000   1st Qu.: 0.0000   1st Qu.: 0.0000
## Median : 1.000   Median : 0.000   Median : 0.0000   Median : 0.0000
## Mean     : 0.756   Mean     : 0.229   Mean     : 0.2485   Mean     : 0.3125
## 3rd Qu.: 1.000   3rd Qu.: 0.000   3rd Qu.: 0.0000   3rd Qu.: 1.0000
## Max.     : 1.000   Max.     : 1.000   Max.     : 1.0000   Max.     : 1.0000
##           we           pt
## Min.      : 0.00   Min.      : 0.0000
## 1st Qu.: 0.00   1st Qu.: 0.0000
## Median : 0.00   Median : 0.0000
## Mean     : 0.21   Mean     : 0.0925
## 3rd Qu.: 0.00   3rd Qu.: 0.0000
## Max.     : 1.00   Max.     : 1.0000
```

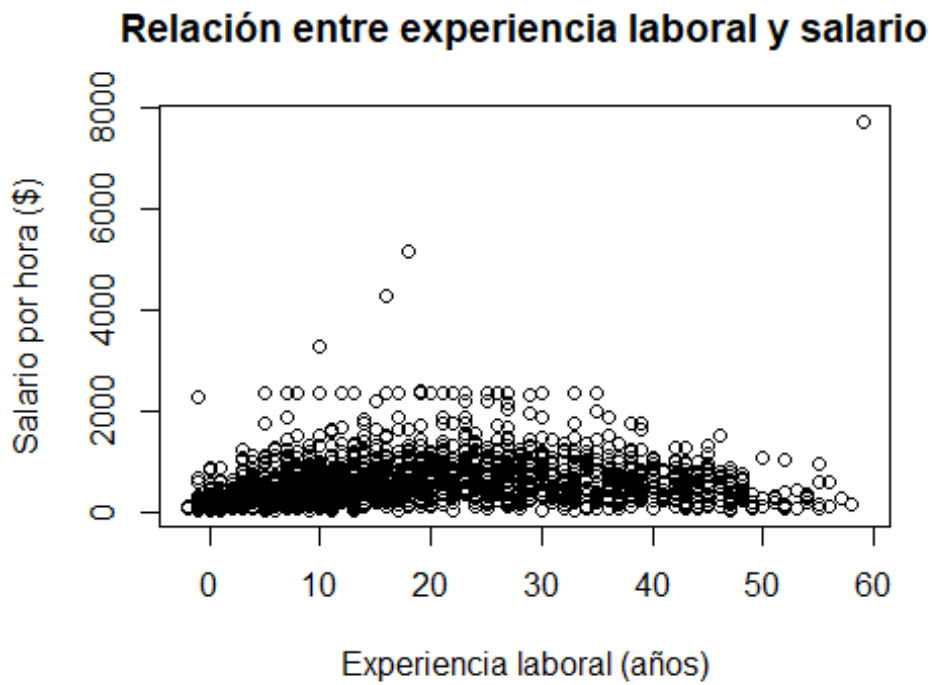
```
hist(uswages$wage)
```



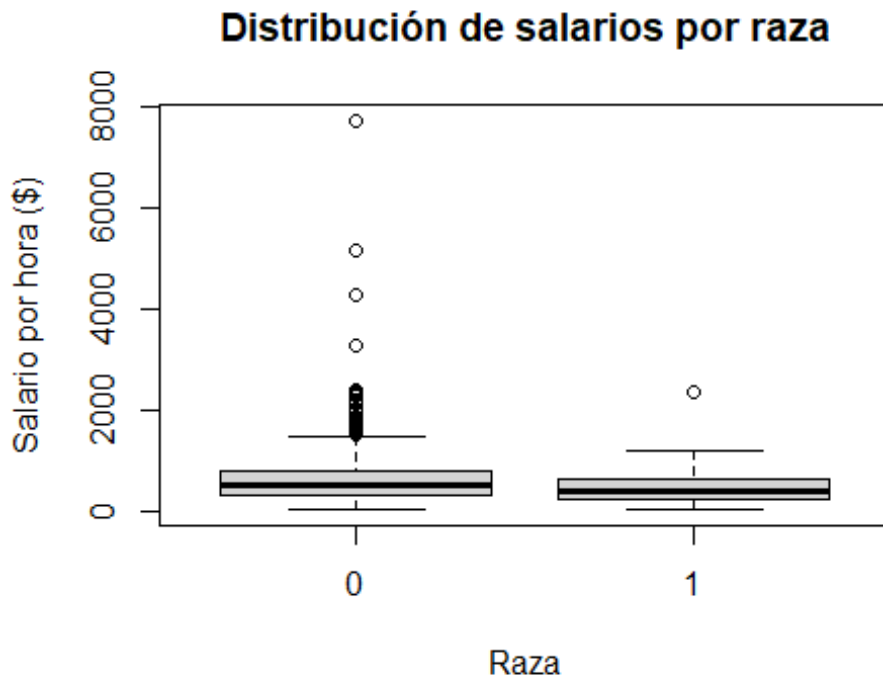
```
cor(uswages)
```

```
##          wage          educ          exper          race          smsa
## wage    1.00000000  0.248335778  0.183201154 -0.09622039  0.136325488
## educ    0.24833578  1.000000000 -0.302478788 -0.09703783  0.077974658
## exper   0.18320115 -0.302478788  1.000000000  0.03080806  0.008733907
## race   -0.09622039 -0.097037827  0.030808062  1.00000000  0.039347879
## smsa    0.13632549  0.077974658  0.008733907  0.03934788  1.000000000
## ne      0.02790797  0.031759525  0.034703174 -0.06089660  0.123988587
## mw     -0.02845714  0.031909948 -0.027688979 -0.05507879 -0.055850582
## so     -0.04001049 -0.063706161  0.024800191  0.18604055 -0.051488079
## we      0.04693321  0.005877411 -0.034645733 -0.09045259 -0.010060827
## pt     -0.22896844  0.024399860 -0.077940099  0.02297372 -0.019527933
##          ne          mw          so          we          pt
## wage    0.02790797 -0.02845714 -0.04001049  0.046933213 -0.228968444
## educ    0.03175953  0.03190995 -0.06370616  0.005877411  0.024399860
## exper   0.03470317 -0.02768898  0.02480019 -0.034645733 -0.077940099
## race   -0.06089660 -0.05507879  0.18604055 -0.090452586  0.022973717
## smsa    0.12398859 -0.05585058 -0.05148808 -0.010060827 -0.019527933
## ne      1.00000000 -0.31339297 -0.36743389 -0.280987276 -0.017927310
## mw     -0.31339297  1.00000000 -0.38769247 -0.296479593  0.036050860
## so     -0.36743389 -0.38769247  1.00000000 -0.347603997 -0.014194638
## we     -0.28098728 -0.29647959 -0.34760400  1.000000000 -0.003601402
## pt     -0.01792731  0.03605086 -0.01419464 -0.003601402  1.000000000
```

```
plot(uswages$exper, uswages$wage, xlab="Experiencia laboral (años)",  
ylab="Salario por hora ($)", main="Relación entre experiencia laboral y  
salario")
```



```
boxplot(wage ~ race, data=uswages, xlab="Raza", ylab="Salario por hora  
($)",  
main="Distribución de salarios por raza")
```

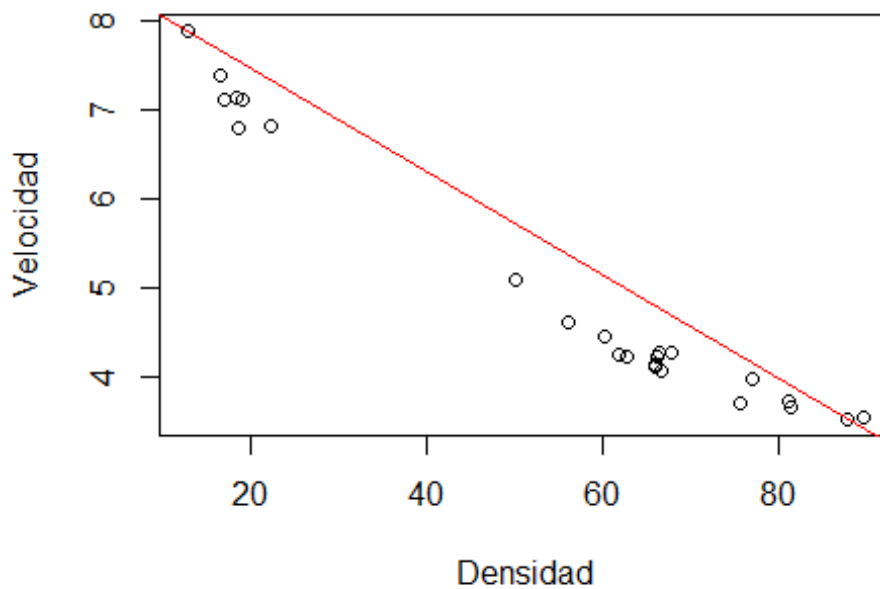



#Podemos ver que Las distribuciones de salarios por hora son diferentes para cada grupo de raza. Por ejemplo, La mediana del salario por hora es más alta para Los trabajadores blancos en comparación con Los trabajadores negros. También podemos ver que hay más valores atípicos para Los trabajadores blancos, Lo que sugiere que estos grupos tienen una mayor variabilidad en Los salarios por hora.

```
dens <- c(12.7,17.0,66.0,50.0,87.8,81.4,75.6,66.2,81.1,62.8,77.0,89.6,
18.3,19.1,16.5,22.2,18.6,66.0,60.3,56.0,66.3,61.7,66.6,67.8)
vel <- c(62.4,50.7,17.1,25.9,12.4,13.4,13.7,17.9,13.8,17.9,15.8,12.6,
51.2,50.8,54.7,46.5,46.3,16.9,19.8,21.2,18.3,18.0,16.6,18.3)
rvel <- sqrt(vel)

#(a) Dibujar la nube de puntos y la recta que pasa por los puntos (12.7,
√ 62.4) y (87.8, √ 12.4)
plot(dens, rvel, xlab="Densidad", ylab="Velocidad", main="Nube de
puntos")
puntos <- data.frame(dens=c(12.7, 87.8), rvel=c(sqrt(62.4), sqrt(12.4)))
abline(lm(rvel ~ dens, data=puntos), col="red")
```

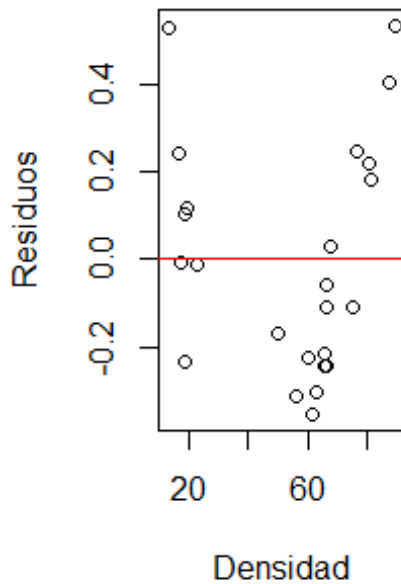
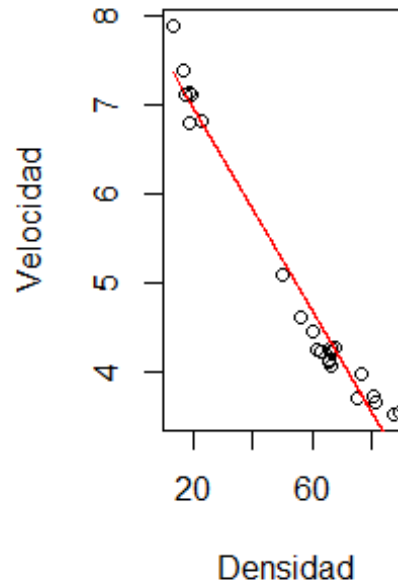
Nube de puntos



```
modelo <- lm(rvel ~ dens)
predicciones <- predict(modelo)
residuos <- rvel - predicciones

par(mfrow=c(1,2))
plot(dens, residuos, xlab="Densidad", ylab="Residuos", main="Gráfico de
residuos")
abline(h=0, col="red")

plot(dens, rvel, xlab="Densidad", ylab="Velocidad", main="Predicciones")
lines(dens, predicciones, col="red")
```

Gráfico de residuos**Predicciones**

```
sum(residuos^2)

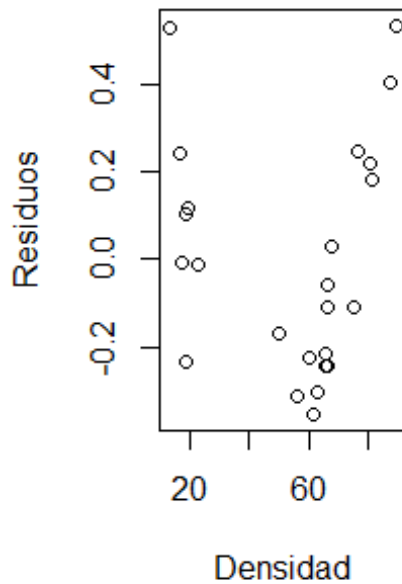
## [1] 1.591218

# (b) Hallar la recta de regresión simple
fit <- lm(rvel ~ dens)
summary(fit)

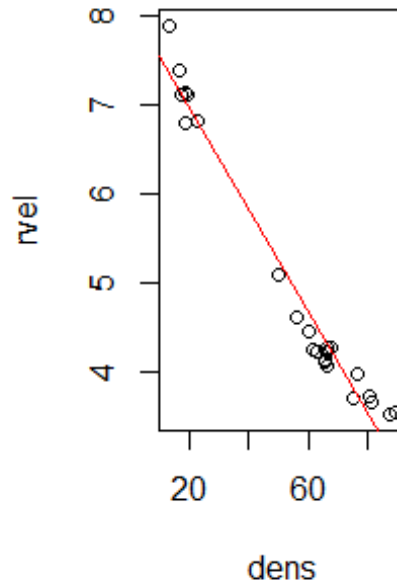
##
## Call:
## lm(formula = rvel ~ dens)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35337 -0.22722 -0.03566  0.18942  0.53349
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.089813   0.130629   61.93  <2e-16 ***
## dens        -0.056626   0.002177  -26.01  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2689 on 22 degrees of freedom
## Multiple R-squared:  0.9685, Adjusted R-squared:  0.9671
## F-statistic: 676.4 on 1 and 22 DF, p-value: < 2.2e-16
```

```
plot(dens, resid(fit), main="Residuos vs Densidad", xlab="Densidad",
ylab="Residuos")
abline(lm(rvel ~ dens), col="red")
plot(dens, rvel, main = "Regresión lineal simple")
abline(fit, col = "red")
```

Residuos vs Densidad

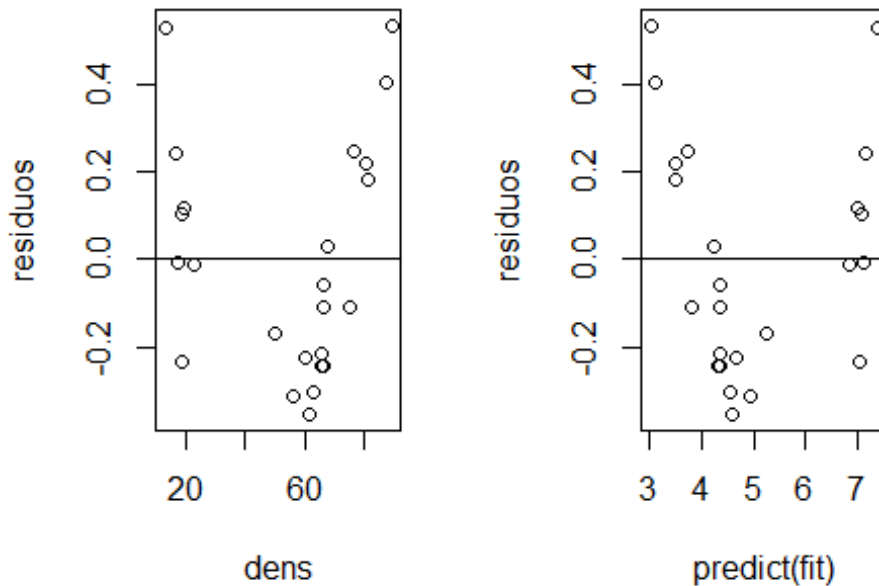


Regresión lineal simple



```
residuos <- resid(fit)
plot(dens, residuos, main = "Residuos vs. Densidad")
abline(h=0)
plot(predict(fit), residuos, main = "Residuos vs. Predicciones")
abline(h=0)
```

Residuos vs. Densidad Residuos vs. Predicción



```
# Suma de cuadrados de los residuos
sum(resid(fit)^2)

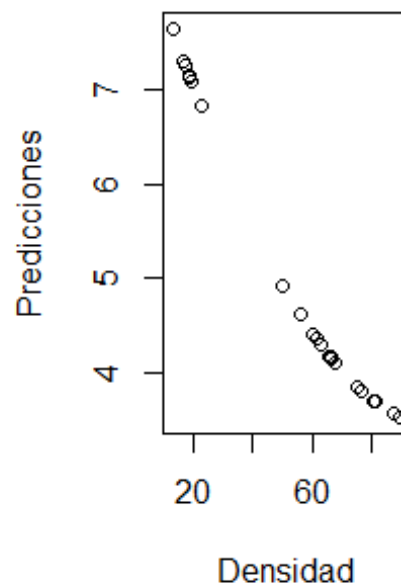
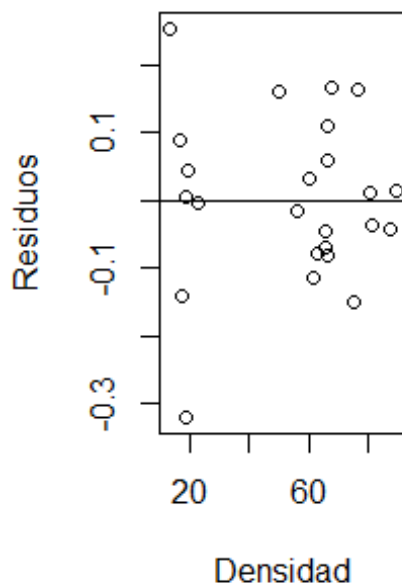
## [1] 1.591218

# (c) Mejorar el modelo anterior considerando una regresión parabólica
# Ajuste de un modelo de regresión parabólica
fit2 <- lm(rvel ~ dens + I(dens^2))

# Coeficientes del modelo de regresión parabólica
coefficients(fit2)

## (Intercept)          dens      I(dens^2)
## 8.8814208199 -0.1035152795  0.0004892585

# Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones.
# Gráfico de los residuos con la densidad
plot(dens, resid(fit2), xlab = "Densidad", ylab = "Residuos")
abline(h=0)
# Gráfico de las predicciones
plot(dens, fitted(fit2), xlab = "Densidad", ylab = "Predicciones")
abline(h=0)
```



Suma de cuadrados de Los residuos

```
sum(resid(fit2)^2)
```

```
## [1] 0.3534143
```

```
max_flujo <- 0
```

```
for (i in 1:length(dens)) {
```

```
  for (j in 1:length(vel)) {
```

```
    flujo <- dens[i] * vel[j]
```

```
    if (flujo > max_flujo) {
```

```
      max_flujo <- flujo
```

```
      dens_max_flujo <- dens[i]
```

```
      vel_max_flujo <- vel[j]
```

```
    }
```

```
  }
```

```
}
```

```
max_flujo
```

```
## [1] 5591.04
```

```
dens_max_flujo
```

```
## [1] 89.6
```

```
vel_max_flujo
```

```
## [1] 62.4
```

```

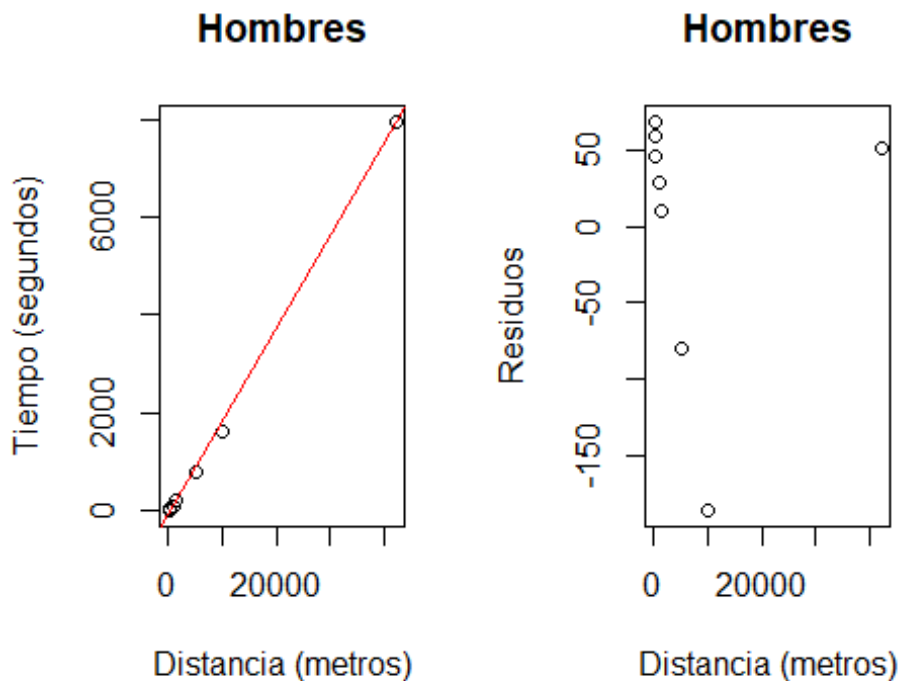
# (a) Datos de Los hombres
distancia <- c(100, 200, 400, 800, 1500, 5000, 10000, 42192)
tiempo <- c(9.84, 19.32, 43.19, 102.58, 215.78, 787.96, 1627.34, 7956.00)

# Recta de regresión simple
modelo <- lm(tiempo ~ distancia)

# Gráfico de La recta de regresión junto a Los datos
plot(distancia, tiempo, xlab = "Distancia (metros)", ylab = "Tiempo (segundos)", main = "Hombres")
abline(modelo, col = "red")

# Gráfico de Los residuos con La distancia
plot(distancia, resid(modelo), xlab = "Distancia (metros)", ylab = "Residuos", main = "Hombres")

```



```

# Gráfico de Las predicciones
plot(predict(modelo), resid(modelo), xlab = "Predicciones", ylab = "Residuos", main = "Hombres")

# Suma de cuadrados de Los residuos
SSres <- sum(resid(modelo)^2)
SSres

## [1] 55189

```

```

# [1] 417362.4

# R2
SStot <- sum((tiempo - mean(tiempo))^2)
R2 <- 1 - SSres/SStot
R2

## [1] 0.9989417

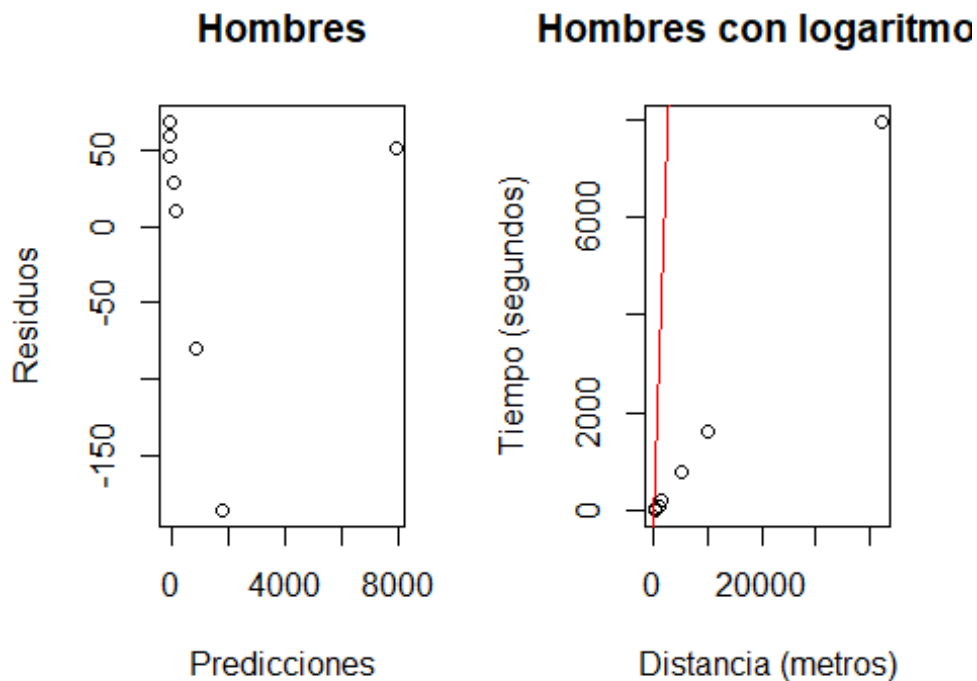
# [1] 0.9960131

# (b) Datos de Los hombres con Logaritmos
log_distancia <- log(distancia)
log_tiempo <- log(tiempo)

# Recta de regresión simple con Logaritmos
modelo_log <- lm(log_tiempo ~ log_distancia)

# Gráfico de La recta de regresión junto a Los datos
plot(distancia, tiempo, xlab = "Distancia (metros)", ylab = "Tiempo (segundos)", main = "Hombres con logaritmos")
abline(exp(coef(modelo_log)), col = "red")

```



```

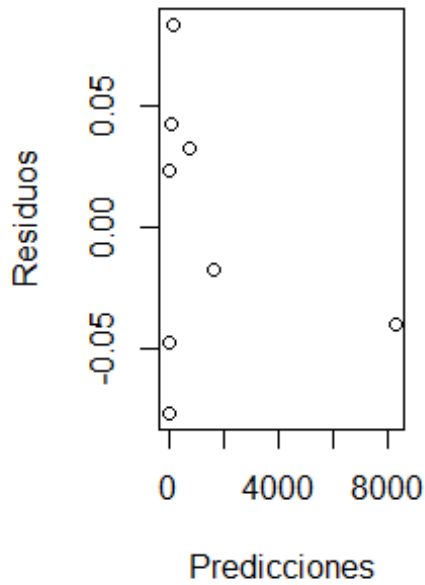
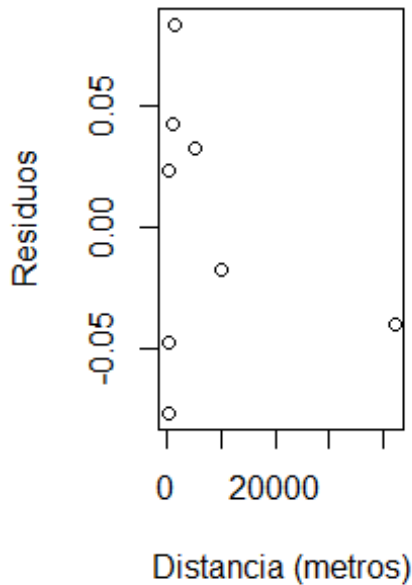
# Gráfico de Los residuos con La distancia
plot(distancia, resid(modelo_log), xlab = "Distancia (metros)", ylab = "Residuos", main = "Hombres con logaritmos")

```



```
# Gráfico de Las predicciones
plot(exp(predict(modelo_log)), resid(modelo_log), xlab = "Predicciones",
ylab = "Residuos", main = "Hombres con logaritmos")
```

Hombres con logaritmo Hombres con logaritmo

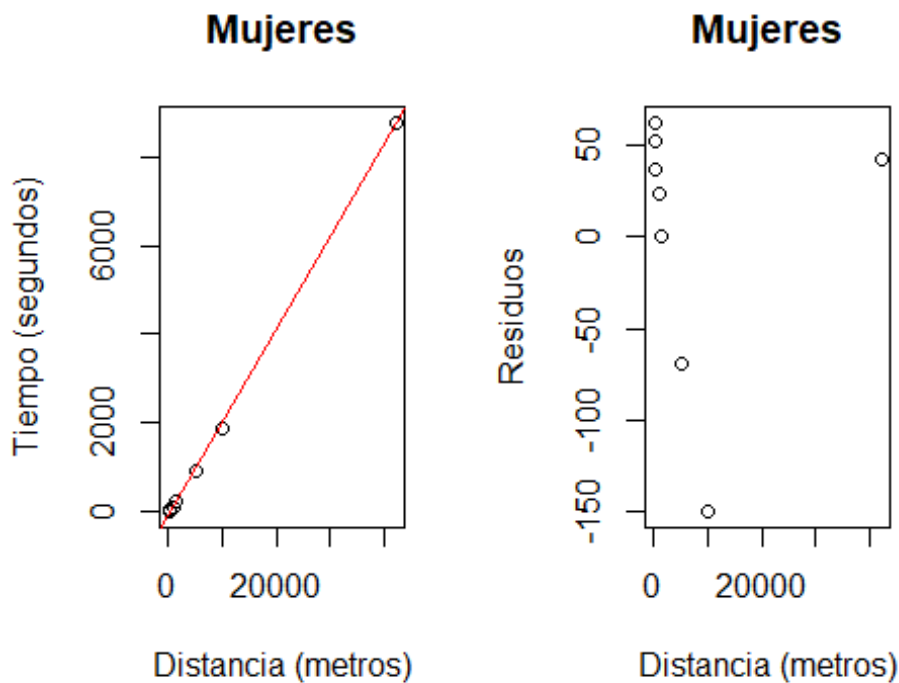


```
# (a) Datos de Las mujeres
distancia <- c(100, 200, 400, 800, 1500, 5000, 10000, 42195)
tiempo_m <- c(10.94, 22.12, 48.25, 117.73, 240.83, 899.88, 1861.63,
8765.00)

# Recta de regresión simple
modelo <- lm(tiempo_m ~ distancia)

# Gráfico de la recta de regresión junto a Los datos
plot(distancia, tiempo_m, xlab = "Distancia (metros)", ylab = "Tiempo
(segundos)", main = "Mujeres")
abline(modelo, col = "red")

# Gráfico de Los residuos con La distancia
plot(distancia, resid(modelo), xlab = "Distancia (metros)", ylab =
"Residuos", main = "Mujeres")
```



```
# Gráfico de Las predicciones
plot(predict(modelo), resid(modelo), xlab = "Predicciones", ylab =
"Residuos", main = "Mujeres")

# Suma de cuadrados de Los residuos
SSres <- sum(resid(modelo)^2)
SSres

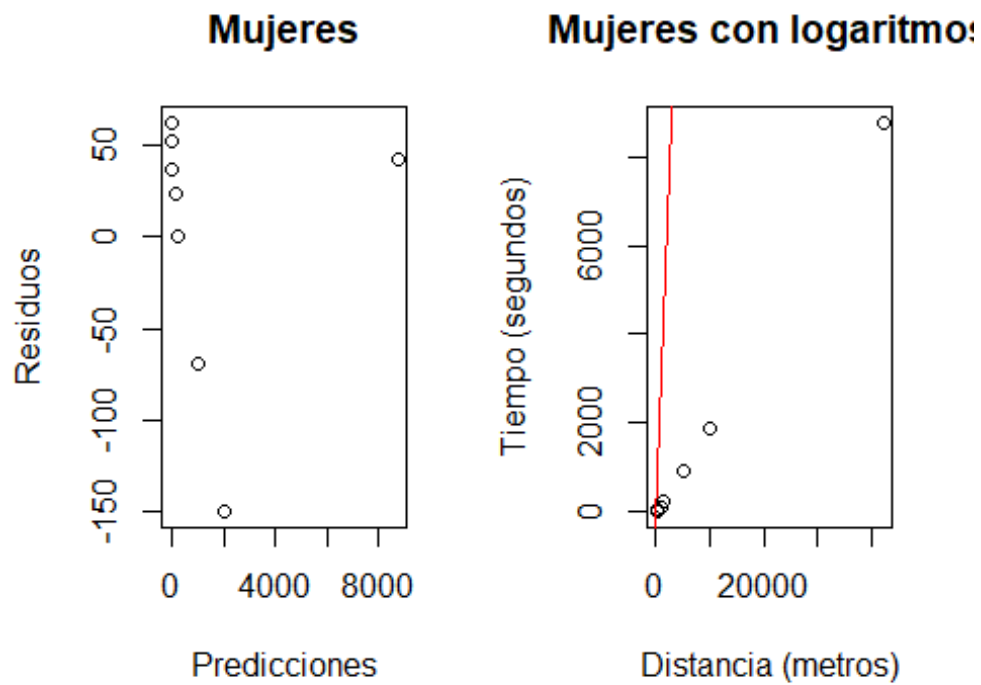
## [1] 37919.92

# [1] 441789.6

# R2
SStot <- sum((tiempo_m - mean(tiempo_m))^2)
R2 <- 1 - SSres/SStot
R2

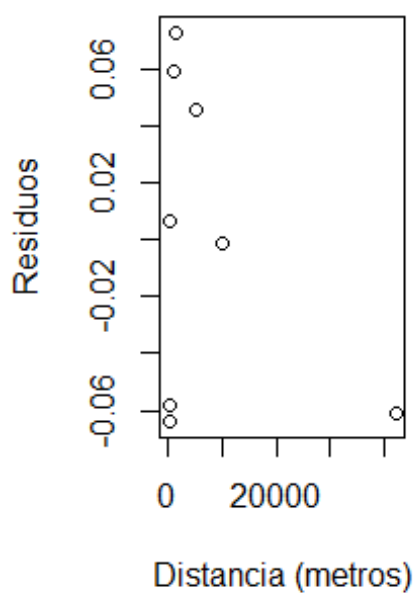
## [1] 0.9994007

# [1] 0.9941927
# Lo mismo pero con Los Logaritmos
log_distancia <- log(distancia)
log_tiempo_m <- log(tiempo_m)
modelo_log_m <- lm(log_tiempo_m ~ log_distancia)
plot(distancia, tiempo_m, xlab = "Distancia (metros)", ylab = "Tiempo
(segundos)", main = "Mujeres con logaritmos")
abline(exp(coef(modelo_log_m)), col = "red")
```

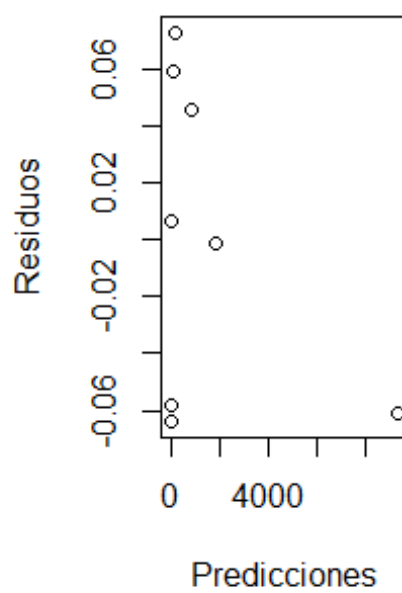


```
plot(distancia, resid(modelo_log_m), xlab = "Distancia (metros)", ylab =
"Residuos", main = "Mujeres con logaritmos")
plot(exp(predict(modelo_log_m)), resid(modelo_log_m), xlab =
"Predicciones", ylab = "Residuos", main = "Mujeres con logaritmos")
```

Mujeres con logaritmo:



Mujeres con logaritmo:



```
SSres <- sum(resid(modelo_log_m)^2)
SSres

## [1] 0.02211633

SStot <- sum((log_tiempo_m - mean(log_tiempo_m))^2)
R2 <- 1 - SSres/SStot
R2

## [1] 0.999404
```