Laboratorio 6
Estadistica Para Ingenieros
Francisco Sucre 10-10717
José Cipagauta 05-38040

```
> # Set De Datos: Salud
> library(rJava)
> library(xlsxjars)
> library(xlsx) # Este carga ya de por si las dependencias
> salud = read.xlsx("DatosI16.xlsx", sheetIndex = 3, rowIndex = c(3:31), colIndex = c(2:12),
header = T)
```

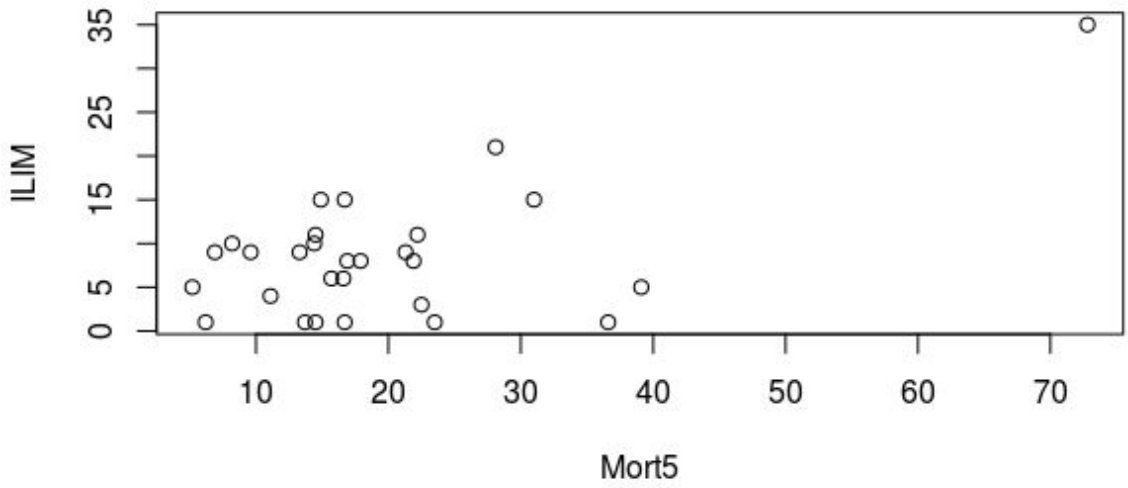
Los nombres de los paises no son relevantes para este estudio, y ademas no son datos numericos Por lo cual cor(salud) da errores

```
> attach(salud) # Attach para facil accesso, en realidad se usa es Mort
>
```

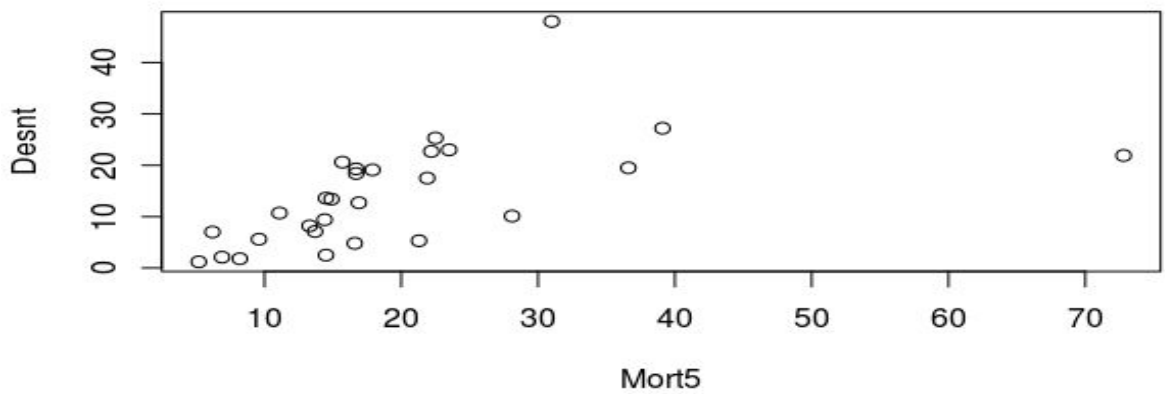
1. Realice un primer modelo con las variables cuyó |ρ| > 0,5 con respecto a la variable respuesta. Para esto calcule la matriz de correlación y estudie las gráficas de las variables.

```
> cor(salud)
          ILIDTP       ILIM      Mort5       Desnt      MortM      MortH      Tuber
ILIDTP  1.00000000  0.74255336  0.3885160  0.05478401  0.1921462  0.04079690
0.2544024
ILIM    0.74255336  1.00000000  0.5971364  0.18917959  0.3149995  0.07197191
0.4145543
Mort5   0.38851599  0.59713641  1.0000000  0.54749005  0.8333538  0.64251745
0.8851362
Desnt   0.05478401  0.18917959  0.5474900  1.00000000  0.4409646  0.48494594
0.3922875
MortM   0.19214618  0.31499945  0.8333538  0.44096458  1.0000000  0.88669515
0.7987874
MortH   0.04079690  0.07197191  0.6425174  0.48494594  0.8866951  1.00000000
0.5840950
Tuber   0.25440236  0.41455428  0.8851362  0.39228749  0.7987874  0.58409500
1.0000000
VIH     0.21745965  0.23271804  0.4944222 -0.02567071  0.4702123  0.42154205
0.3505904
EV60   -0.20576196 -0.25606229 -0.6925585 -0.21628634 -0.7680551 -0.71247546
-0.6754800
Med    -0.15522249 -0.21643050 -0.4885948 -0.31815734 -0.4282248 -0.40874813
-0.3883888
PIB    -0.21390107 -0.07284822 -0.1759778 -0.32777708 -0.1894637 -0.26140176
-0.1093355
             VIH       EV60        Med        PIB
ILIDTP  0.21745965 -0.2057620 -0.1552225 -0.21390107
ILIM    0.23271804 -0.2560623 -0.2164305 -0.07284822
```

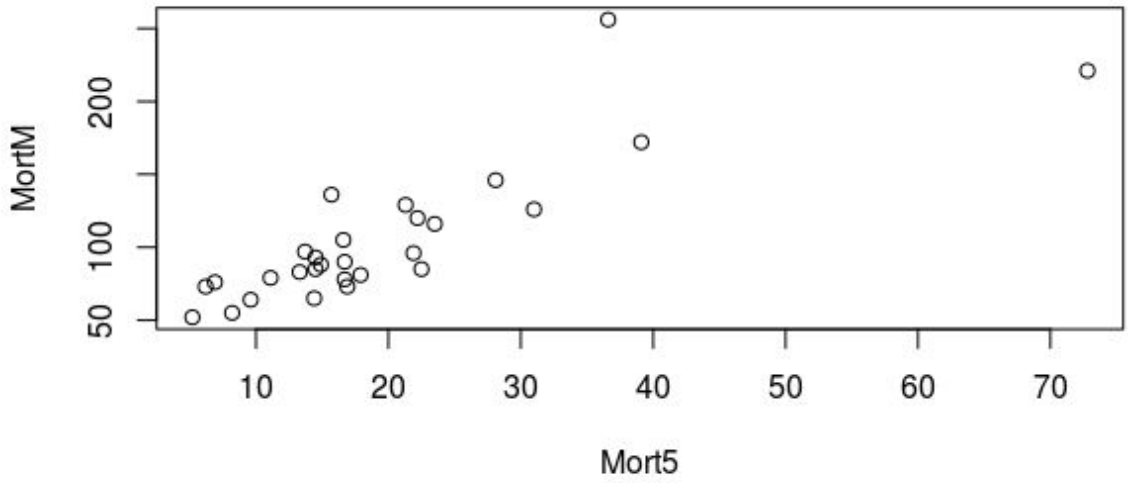
```
Mort5   0.49442222 -0.6925585 -0.4885948 -0.17597782
Desnt  -0.02567071 -0.2162863 -0.3181573 -0.32777708
MortM   0.47021235 -0.7680551 -0.4282248 -0.18946367
MortH   0.42154205 -0.7124755 -0.4087481 -0.26140176
Tuber   0.35059041 -0.6754800 -0.3883888 -0.10933554
VIH     1.00000000 -0.6586757 -0.3864572 -0.26767455
EV60   -0.65867571  1.0000000  0.3624455  0.30793204
Med    -0.38645718  0.3624455  1.0000000  0.20734415
PIB    -0.26767455  0.3079320  0.2073442  1.00000000
> # Se eligen solo los que tengan |ρ| > 0.5
> plot(Mort5, ILIM)
```



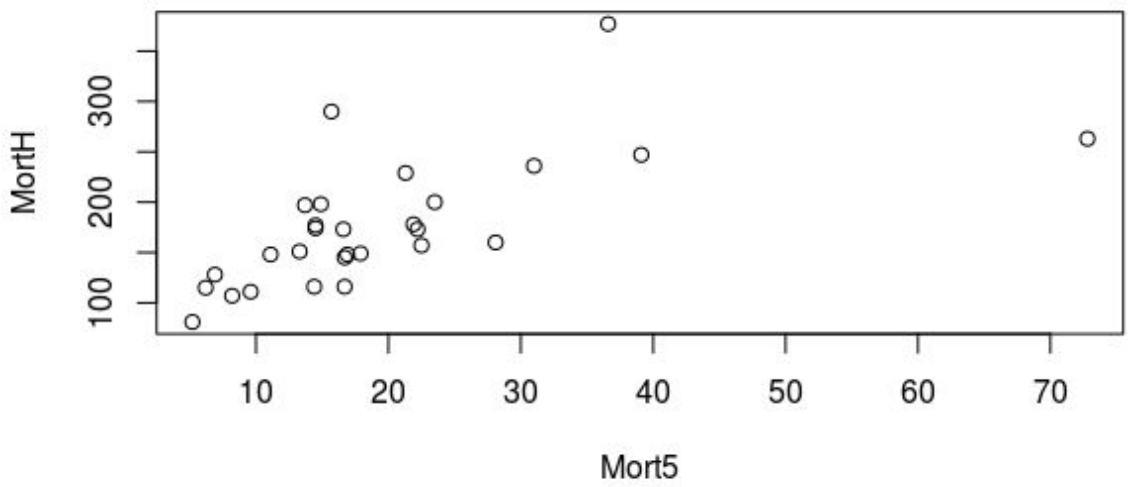

```
> plot(Mort5, Desnt)
```



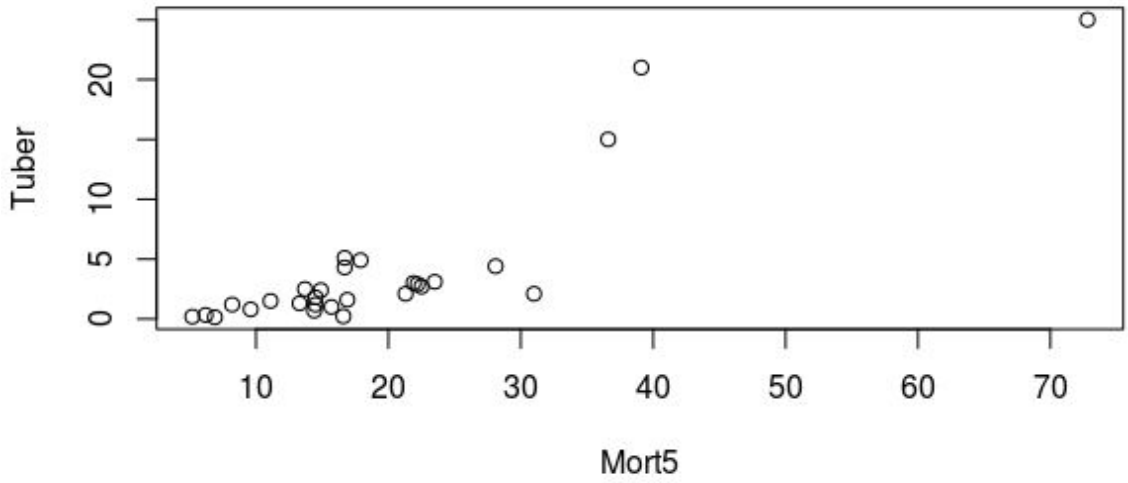
> plot(Mort5, MortM)



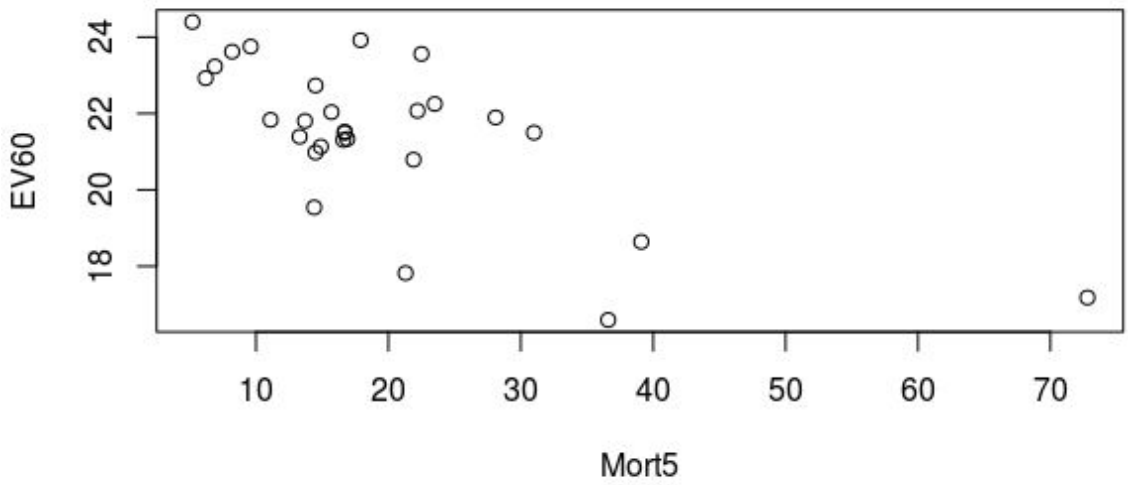
> plot(Mort5, MortH)



> plot(Mort5, Tuber)



> plot(Mort5, EV60)



Modelo 1, Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + EV60 con ρ1 = 0.5971, ρ2 = 0.5475, ρ3 = 0.8334, ρ4 = 0.6425, ρ5 = 0.8851, ρ6 = -0.6926 respectivamente

```
> m1 = lm(Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + EV60)
> m1

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + EV60)

Coefficients:
(Intercept)         ILIM        Desnt        MortM        MortH        Tuber         EV60
   19.54719      0.49314      0.27493      0.07758     -0.01173      0.92747     -0.82582
```

2. Realice un segundo modelo con regresión paso a paso. En cada paso elimine la variable menos significativa y continue hasta que todas las restantes tengan un nivel de signifancia menor a 0.05.

Modelo 2, Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB

```
> m2 = lm(Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB)
> summary(m2)

Call:
lm(formula = Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH +
    Tuber + VIH + EV60 + Med + PIB)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4157 -1.5417 -0.4587  2.0722  6.3063

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.15411   18.15272  -0.064  0.95005
ILIDTP       0.02964    0.30353   0.098  0.92335
ILIM         0.43078    0.17322   2.487  0.02357 *
Desnt        0.35450    0.09746   3.638  0.00204 **
MortM        0.06299    0.05348   1.178  0.25508
MortH       -0.01199    0.03307  -0.362  0.72150
Tuber        0.98215    0.23781   4.130  0.00070 ***
VIH          5.01090    1.92488   2.603  0.01856 *
EV60        -0.10694    0.73001  -0.146  0.88526
Med         -0.03573    0.06324  -0.565  0.57947
PIB          0.39702    0.33874   1.172  0.25735
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.672 on 17 degrees of freedom
Multiple R-squared:  0.9524,  Adjusted R-squared:  0.9244
```

F-statistic: 34.03 on 10 and 17 DF, p-value: 2.877e-09

Modelo 3, Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB - 1
Eliminando intercept con p-valor de 0.9501

```
> m3 = lm(Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB - 1)
> summary(m3)

Call:
lm(formula = Mort5 ~ ILIDTP + ILIM + Desnt + MortM + MortH +
    Tuber + VIH + EV60 + Med + PIB - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4802 -1.5446 -0.4459  2.0926  6.2975

Coefficients:
       Estimate Std. Error t value Pr(>|t|)
ILIDTP  0.02868    0.29465   0.097 0.923533
ILIM    0.43035    0.16823   2.558 0.019762 *
Desnt   0.35531    0.09392   3.783 0.001361 **
MortM   0.06296    0.05198   1.211 0.241420
MortH  -0.01268    0.03035  -0.418 0.680971
Tuber   0.97762    0.22050   4.434 0.000321 ***
VIH     4.94867    1.61095   3.072 0.006571 **
EV60   -0.15180    0.18202  -0.834 0.415245
Med    -0.03658    0.06007  -0.609 0.550108
PIB     0.39790    0.32896   1.210 0.242099
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.569 on 18 degrees of freedom
Multiple R-squared:  0.9854,  Adjusted R-squared:  0.9773
F-statistic: 121.5 on 10 and 18 DF,  p-value: 2.048e-14
```

Modelo 4, Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB - 1
Eliminando ILIDTP con p-valor de 0.9235

```
> m4 = lm(Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + VIH + EV60 + Med + PIB - 1)
> summary(m4)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + VIH +
    EV60 + Med + PIB - 1)

Residuals:
```

```
    Min      1Q  Median      3Q     Max
-5.6051 -1.5351 -0.5034  2.0598  6.2944

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
ILIM   0.44211    0.11401   3.878 0.001012 **
Desnt  0.35334    0.08929   3.957 0.000845 ***
MortM  0.06257    0.05045   1.240 0.229994
MortH -0.01233    0.02933  -0.420 0.679014
Tuber  0.97845    0.21452   4.561 0.000213 ***
VIH    4.95006    1.56834   3.156 0.005199 **
EV60  -0.14682    0.17007  -0.863 0.398764
Med   -0.03654    0.05848  -0.625 0.539505
PIB    0.38892    0.30742   1.265 0.221136
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.475 on 19 degrees of freedom
Multiple R-squared:  0.9854,  Adjusted R-squared:  0.9785
F-statistic: 142.4 on 9 and 19 DF,  p-value: 1.654e-15
```

Modelo 5, Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 + Med + PIB - 1
Eliminando MortH con p-valor de 0.6790

```
> m5 = lm(Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 + Med + PIB - 1)
> summary(m5)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 +
    Med + PIB - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2815 -1.7702 -0.3496  2.0590  6.4636

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
ILIM   0.46309    0.10036   4.614 0.000168 ***
Desnt  0.34193    0.08330   4.105 0.000550 ***
MortM  0.04424    0.02483   1.782 0.089963 .
Tuber  1.01379    0.19324   5.246 3.91e-05 ***
VIH    4.84442    1.51586   3.196 0.004539 **
EV60  -0.16577    0.16057  -1.032 0.314210
Med   -0.03562    0.05722  -0.622 0.540670
PIB    0.39657    0.30050   1.320 0.201841
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 3.402 on 20 degrees of freedom
Multiple R-squared: 0.9853, Adjusted R-squared: 0.9794
F-statistic: 167 on 8 and 20 DF, p-value: < 2.2e-16

Modelo 6, Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 + PIB - 1
Eliminando Med con p-valor de 0.5407

```
> m6 = lm(Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 + PIB - 1)
> summary(m6)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 +
    PIB - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1368 -2.0558 -0.2577  2.1881  6.9425

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
ILIM   0.46506    0.09884   4.705 0.000121 ***
Desnt  0.35278    0.08025   4.396 0.000252 ***
MortM  0.04402    0.02446   1.799 0.086328 .
Tuber  1.02172    0.18999   5.378 2.47e-05 ***
VIH    5.05208    1.45697   3.468 0.002302 **
EV60  -0.20438    0.14593  -1.401 0.175952
PIB    0.39116    0.29596   1.322 0.200505
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.352 on 21 degrees of freedom
Multiple R-squared:  0.985,    Adjusted R-squared:   0.98
F-statistic: 196.6 on 7 and 21 DF,  p-value: < 2.2e-16
```

Modelo 7, Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 - 1
Eliminando PIB con p-valor de 0.2005

```
> m7 = lm(Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 - 1)
> summary(m7)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH + EV60 -
    1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0214 -1.7409 -0.1543  1.6828  7.2200

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
```

```
ILIM   0.46462    0.10050   4.623 0.000132 ***
Desnt  0.31152    0.07518   4.144 0.000425 ***
MortM  0.04770    0.02471   1.930 0.066557 .
Tuber  1.05575    0.19140   5.516 1.53e-05 ***
VIH    4.63134    1.44569   3.204 0.004098 **
EV60  -0.04864    0.08752  -0.556 0.584021
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.409 on 22 degrees of freedom
Multiple R-squared:  0.9837,  Adjusted R-squared:  0.9793
F-statistic: 221.5 on 6 and 22 DF,  p-value: < 2.2e-16
```

Modelo 8, Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH - 1
Eliminando EV60 con p-valor de 0.5840

```
> m8 = lm(Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH - 1)
> summary(m8)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-6.1881 -1.9202 -0.4016  1.9247  7.1877

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
ILIM   0.44695    0.09390   4.760 8.46e-05 ***
Desnt  0.30371    0.07273   4.176 0.000363 ***
MortM  0.03846    0.01800   2.136 0.043508 *
Tuber  1.12517    0.14282   7.878 5.57e-08 ***
VIH    4.65212    1.42332   3.268 0.003377 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.357 on 23 degrees of freedom
Multiple R-squared:  0.9835,  Adjusted R-squared:  0.9799
F-statistic:   274 on 5 and 23 DF,  p-value: < 2.2e-16
```

3. Compare los modelos.

```
> summary(m1)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + MortH + Tuber + EV60)

Residuals:
   Min     1Q Median     3Q    Max
```

-7.882 -1.660 -0.528 2.137 6.910

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.54719 17.19541 1.137 0.26844
*ILIM 0.49314 0.13344 3.696 0.00134 **
Desnt 0.27493 0.09663 2.845 0.00969 **
MortM 0.07758 0.05901 1.315 0.20276
MortH -0.01173 0.03652 -0.321 0.75113
Tuber 0.92747 0.25797 3.595 0.00170 **
EV60 -0.82582 0.68279 -1.209 0.23992*
---
*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 4.1 on 21 degrees of freedom
Multiple R-squared: 0.9267, Adjusted R-squared: 0.9058
F-statistic: 44.27 on 6 and 21 DF, p-value: 7.554e-11

R cuadrado adjustado de 0,9058, aunque no todas las variables tienen significancia menor a 0.5
Solo 3 lo tienen, y las restantes 4 con significancia entre 0.20 a 0.75

> summary(m8)

Call:
lm(formula = Mort5 ~ ILIM + Desnt + MortM + Tuber + VIH - 1)

Residuals:
Min 1Q Median 3Q Max
-6.1881 -1.9202 -0.4016 1.9247 7.1877

Coefficients:
Estimate Std. Error t value Pr(>|t|)
ILIM 0.44695 0.09390 4.760 8.46e-05 ***
Desnt 0.30371 0.07273 4.176 0.000363 ***
MortM 0.03846 0.01800 2.136 0.043508 *
Tuber 1.12517 0.14282 7.878 5.57e-08 ***
VIH 4.65212 1.42332 3.268 0.003377 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.357 on 23 degrees of freedom
Multiple R-squared: 0.9835, Adjusted R-squared: 0.9799
F-statistic: 274 on 5 and 23 DF, p-value: < 2.2e-16

R cuadrado adjustado de 0.9799, todas las variables con significancia menor a 0.5

Este es un modelo mas util a usar, que se acerca mas a los datos, seria mejor

a la hora de hacer predicciones

4. Use los datos que aparecen en la misma hoja de datos, subrayados en azul, para hacer una predicción con ambos modelos.

Creando vector de datos para la prediccion y realizando la prediccion para el modelo 1

```
> datam1 = data.frame(ILIM = 3, Desnt = 10, MortM = 102, MortH = 150, Tuber = 1, EV60 = 20)
> predict(m1, newdata = datam1, interval = c("prediction"), level = .95)
      fit      lwr      upr
1 14.3405 4.612187 24.06882
```

Creando vector de datos para la prediccion y realizando la prediccion para el modelo 8

```
> datam8 = data.frame(ILIM = 3, Desnt = 10, MortM = 102, Tuber = 1, VIH = 0.4)
> predict(m8, newdata = datam8, interval = c("prediction"), level = .95)
       fit      lwr      upr
1 11.28705 4.044231 18.52988
```