Test

Equipo 34

Acorde a una muestra aleatoria de 72 hospitales de los cuales se tienen los datos: Y: Riesgo de infección, X1: Duración de la estadía, X2: Rutina de cultivos, X3: Número de camas, X4: Censo promedio diario, X5: Número de enfermeras. (Mejorar introducción!!!)

Deseamos analizar primeramente la matriz de correlaciones entre las diferentes variables aleatorias

```
source("funciones.R")
base <- read.table("Equipo34.txt", header = T)
pairs(base, lower.panel = myPanel.cor, upper.panel = panel.smooth, diag.panel = myPanel.box, labels = na
```

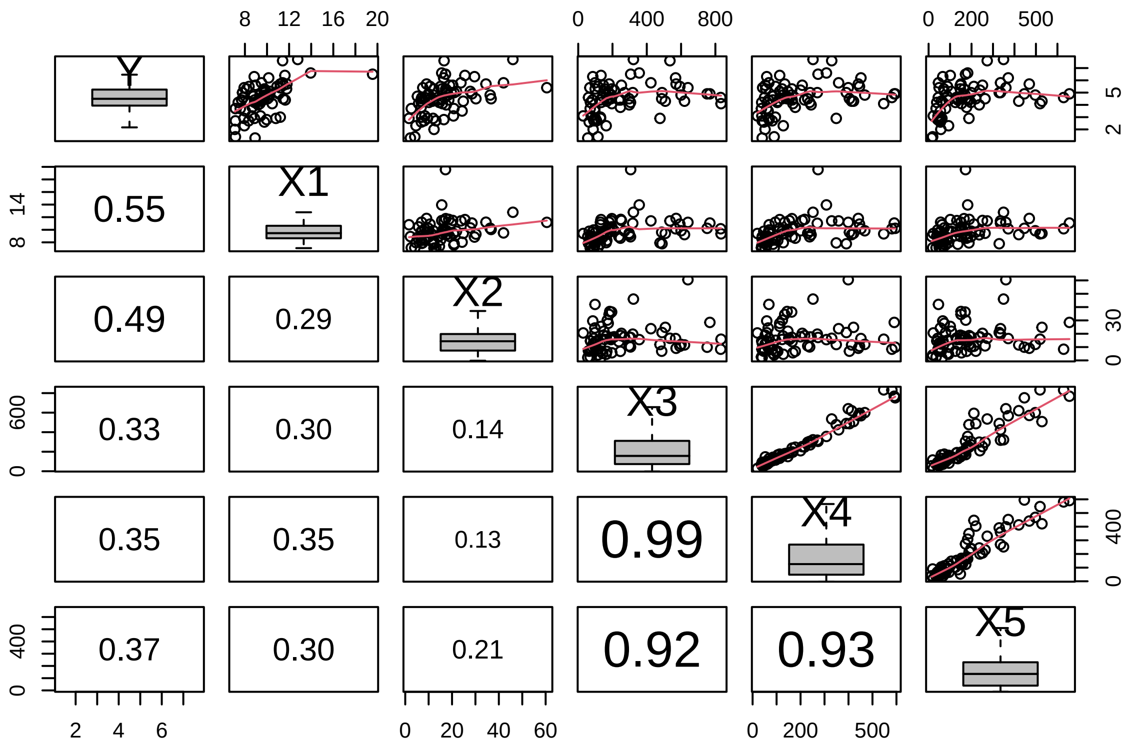

Figure 1: Matriz de correlaciones

Se puede notar una correlación lineal suficiente entre la respuesta y las covariables X1 y X2, además de una fuerte dependencia lineal entre las variables X3 y X4, X3 y X5, y X4 y X5.

Es de nuestro interés ajustar el siguiente modelo de regresión lineal múltiple:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i,\ \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2);\ 1 \leq i \leq 72$$

```r
modelo <- lm(Y ~ ., base)
myAnova(modelo)
```

```
##       Sum_of_Squares DF Mean_Square F_Value     P_value
## Model        52.3209  5   10.464176 10.7672 1.35475e-07
## Error        64.1423 66    0.971853
```

```r
modelo$coefficients
```

```
##   (Intercept)            X1            X2            X3            X4
##  0.9585583976  0.2681612919  0.0408523617 -0.0007105611  0.0009211901
##            X5
##  0.0015270982
```

Una vez ajustado el modelo se obtuvo lo siguiente:

$$\hat{y_i} = 0.9586 + 0.2682 x_{1i} + 0.0409 x_{2i} + -0.0007 x_{3i} + 0.00094 x_{4i} + 0.0015 x_{5i}\ ;\ 1 \leq i \leq 72$$

Del cual nos interesa conocer la significancia marginal de cada uno de los parametros para lo cual realizaremos la siguiente prueba para $j = 0, \cdots, 5$:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$$

con $T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{72-6}$ bajo H_0

```r
tabla.coef <- summary(modelo)$coefficients
tabla.coef
```

```
##                  Estimate  Std. Error    t value    Pr(>|t|)
## (Intercept)  0.9585583976 0.659854581  1.4526813 0.151050230
## X1           0.2681612919 0.074210153  3.6135391 0.000584111
## X2           0.0408523617 0.011983221  3.4091304 0.001115451
## X3          -0.0007105611 0.003490746 -0.2035557 0.839327047
## X4           0.0009211901 0.004978748  0.1850244 0.853777272
## X5           0.0015270982 0.002148440  0.7107939 0.479716370
```

Notar que X1 y X2 son las únicas variables con efecto significativo en la respuesta.

Además nos interesa hacer un análisis de varianza para verificar la significancia de la regresión, esto con ayuda de la tabla ANOVA que nos ayudará a constrastar las siguientes hipotesis:

$$\begin{cases} H_0 : \beta_1 = \cdots = \beta_5 = 0 \\ H_1 : \text{Al menos un } \beta_j \neq 0 \end{cases}$$

Obteniendo así lo siguiente de la tabla ANOVA:

```r
tabla.anova <- myAnova(modelo)
tabla.anova
```

```
##       Sum_of_Squares DF Mean_Square F_Value     P_value
## Model        52.3209  5   10.464176 10.7672 1.35475e-07
## Error        64.1423 66    0.971853
```

!!!Falta F

De la cual podemos concluir que al menos uno de los coeficientes es significativo, esto con un nivel de significancia del 5%