

Tarea 2: Modelos lineales generalizados y paramétricos

Angie Rodriguez Duque & Cesar Saavedra Vanegas

Octubre 22 de 2020

Actividad 1

Actividad 2

Base de datos

Estadísticas descriptivas

```
dim(Datos)
```

```
## [1] 1599 12
```

```
summary(Datos)
```

```
## fixed.acidity volatile.acidity citric.acid residual.sugar
## Min. : 4.60 Min. :0.1200 Min. :0.000 Min. : 0.900
## 1st Qu.: 7.10 1st Qu.:0.3900 1st Qu.:0.090 1st Qu.: 1.900
## Median : 7.90 Median :0.5200 Median :0.260 Median : 2.200
## Mean : 8.32 Mean :0.5278 Mean :0.271 Mean : 2.539
## 3rd Qu.: 9.20 3rd Qu.:0.6400 3rd Qu.:0.420 3rd Qu.: 2.600
## Max. :15.90 Max. :1.5800 Max. :1.000 Max. :15.500
## chlorides free.sulfur.dioxide total.sulfur.dioxide density
## Min. :0.01200 Min. : 1.00 Min. : 6.00 Min. :0.9901
## 1st Qu.:0.07000 1st Qu.: 7.00 1st Qu.: 22.00 1st Qu.:0.9956
## Median :0.07900 Median :14.00 Median : 38.00 Median :0.9968
## Mean :0.08747 Mean :15.87 Mean : 46.47 Mean :0.9967
## 3rd Qu.:0.09000 3rd Qu.:21.00 3rd Qu.: 62.00 3rd Qu.:0.9978
## Max. :0.61100 Max. :72.00 Max. :289.00 Max. :1.0037
## pH sulphates alcohol quality
## Min. :2.740 Min. :0.3300 Min. : 8.40 Min. :3.000
## 1st Qu.:3.210 1st Qu.:0.5500 1st Qu.: 9.50 1st Qu.:5.000
## Median :3.310 Median :0.6200 Median :10.20 Median :6.000
## Mean :3.311 Mean :0.6581 Mean :10.42 Mean :5.636
## 3rd Qu.:3.400 3rd Qu.:0.7300 3rd Qu.:11.10 3rd Qu.:6.000
## Max. :4.010 Max. :2.0000 Max. :14.90 Max. :8.000
```

Variable indicadora: pH_i

```
table(pHi)
```

```
## pHi
```

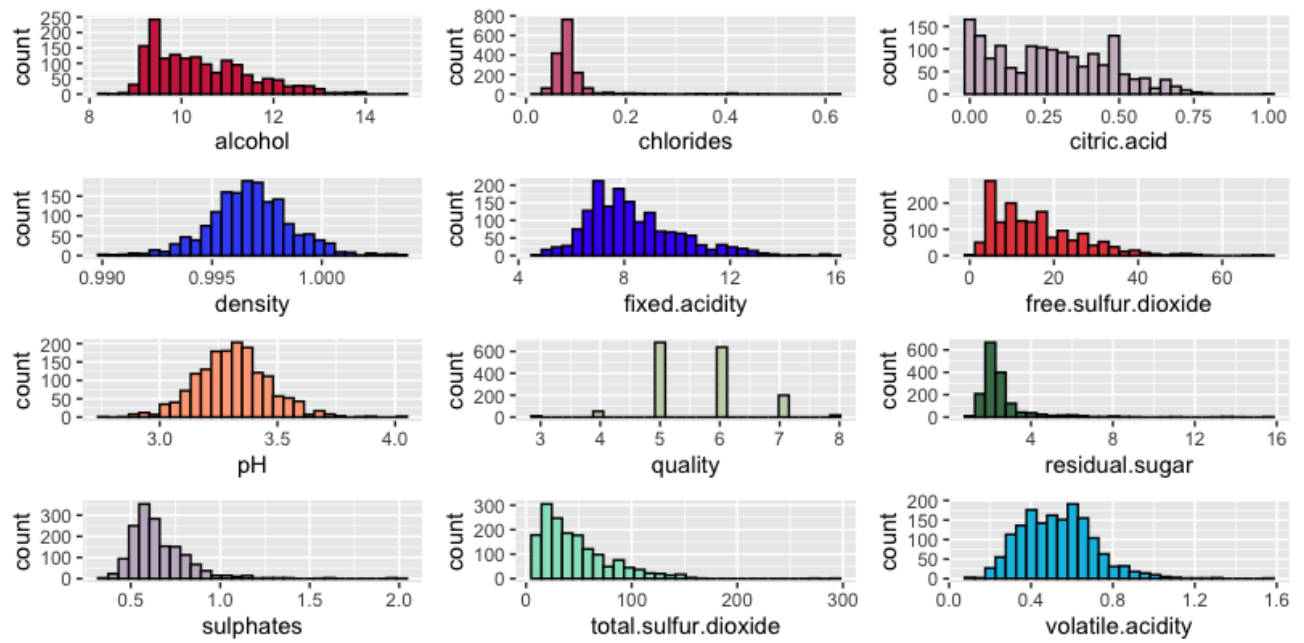


Figure 1: Distribución de las variables

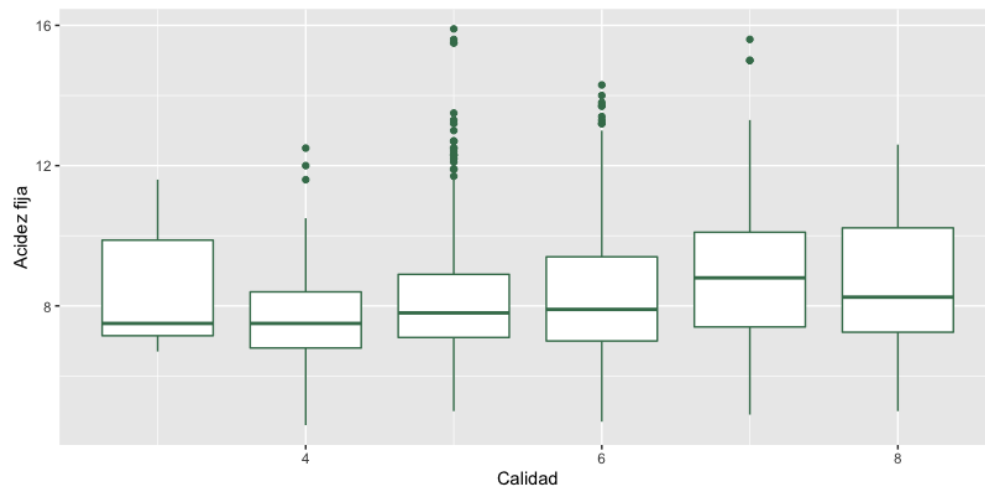


Figure 2: Relación acidez fija - calidad

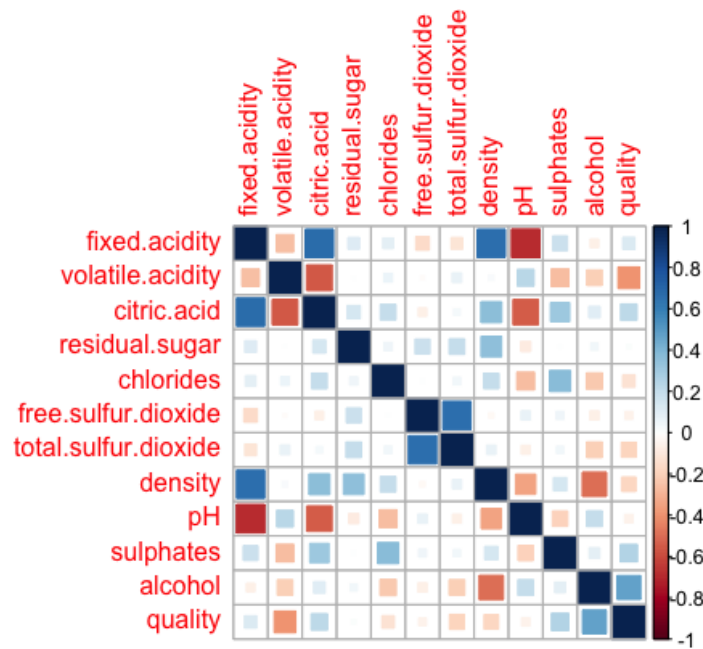


Figure 3: Correlación

```
## Bajo Medio Alto
## 134 1417 48
```

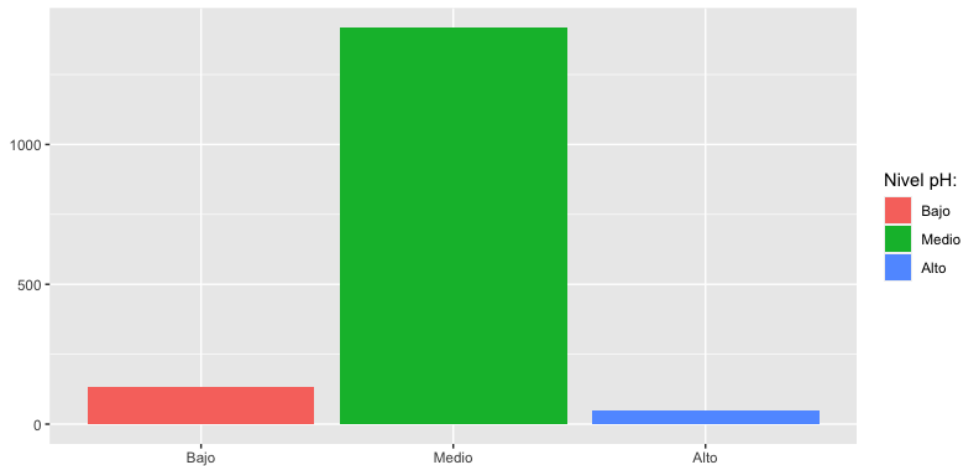


Figure 4: Variable indicadora

Modelo lineal generalizado (GLM)

```
Modelo <- glm(Datos$quality ~ Datos$fixed.acidity + pH, data=Datos)
summary(Modelo)
```

```
##
## Call:
## glm(formula = Datos$quality ~ Datos$fixed.acidity + pH, data = Datos)
```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8636  -0.6083   0.1899   0.4373   2.5442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.96912    0.15647  31.757 < 2e-16 ***
## Datos$fixed.acidity 0.06685    0.01308   5.113 3.56e-07 ***
## pHiMedio         0.11896    0.07979   1.491   0.136
## pHiAlto          0.17674    0.14878   1.188   0.235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.642382)
##
##      Null deviance: 1042.2  on 1598  degrees of freedom
## Residual deviance: 1024.6  on 1595  degrees of freedom
## AIC: 3836.1
##
## Number of Fisher Scoring iterations: 2
```

Conlusiones

Bibliografía