

Tarea 2: Modelos lineales generalizados y paramétricos

Angie Rodriguez Duque & Cesar Saavedra Vanegas

Octubre 22 de 2020

Actividad 1

Se dispone de los tiempos de vida (tiempos hasta que fallan, en horas) de 49 recipientes de presión sometidos a un nivel de carga del 70%

Distribución Weibull

Para estudiar este tipo de variable se acostumbra utilizar la distribución de Weibull, cuya función de densidad es:

$$f(y; \lambda, \theta) = \frac{\lambda y^{\lambda-1}}{\theta^\lambda} \exp \left[- \left(\frac{y}{\theta} \right)^\lambda \right]$$

Actividad 2

Base de datos

```
dim(Datos)
```

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

Este conjunto de datos de vino tinto consta de 1599 observaciones y 12 variables, 11 de las cuales son sustancias químicas. Las variables son:

1. **Acidez fija:** La mayoría de los ácidos implicados en el vino son fijos o no volátiles (no se evaporan fácilmente).
2. **Acidez volátil:** La cantidad de ácido acético en el vino, que en niveles demasiado altos puede provocar un sabor desagradable a vinagre.
3. **Ácido cítrico:** Encontrado en pequeñas cantidades, el ácido cítrico puede agregar “frescura” y sabor a los vinos.
4. **Azúcar residual:** Es la cantidad de azúcar que queda después de que se detiene la fermentación, es raro encontrar vinos con menos de 1 gramo / litro y los vinos con más de 45 gramos / litro se consideran dulces.
5. **Cloruros:** Es la cantidad de sal del vino.
6. **Dióxido de azufre libre:** La forma libre de SO_2 existe en equilibrio entre el SO_2 molecular (como gas disuelto) y el ion bisulfito; Previene el crecimiento microbiano y la oxidación del vino.

7. **Dióxido de azufre total:** Es la cantidad de formas libres y unidas de SO_2 ; en concentraciones bajas, el SO_2 es mayormente indetectable en el vino, pero en concentraciones de SO_2 libre superiores a 50 ppm, el SO_2 se hace evidente en la nariz y el sabor del vino.
8. **Densidad:** La densidad es cercana a la del agua dependiendo del porcentaje de alcohol y contenido de azúcar.
9. **pH:** Describe qué tan ácido o básico es un vino en una escala de 0 (muy ácido) a 14 (muy básico); la mayoría de los vinos están entre 3-4 en la escala de pH.
10. **Sulfatos:** Aditivo del vino que puede contribuir a los niveles de dióxido de azufre (SO_2), que actúa como antimicrobiano y antioxidante.
11. **Alcohol:** El porcentaje de contenido de alcohol del vino.
12. **Calidad:** Variable de respuesta (basada en datos sensoriales, puntuación entre 0 y 10).

Estadísticas descriptivas

```
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
```

Observaciones:

- Algunas de las variables tienen distribuciones normales (densidad, acidez fija, pH, acidez volátil).
- Algunas variables están un poco sesgadas hacia el extremo inferior de los valores (cloruros, ácido cítrico, azúcar residual, dióxido de azufre total).
- La variable calidad tiene solo 6 valores discretos.

```
corrplot(cor(Datos), method = "square")
```

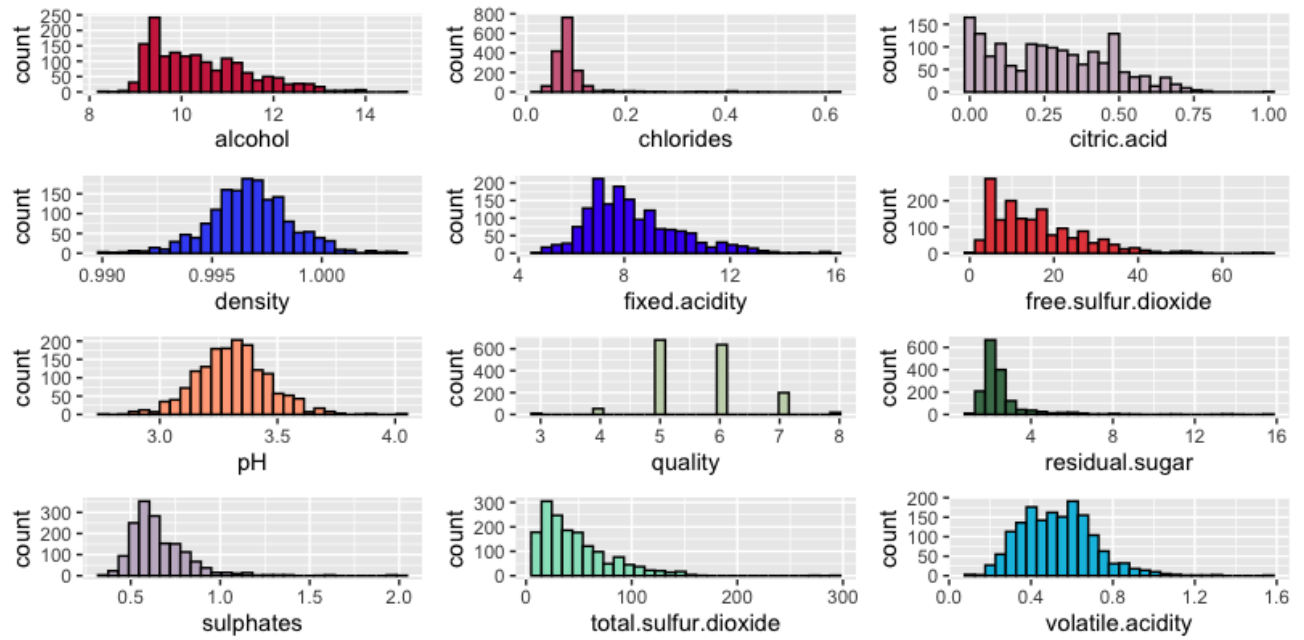
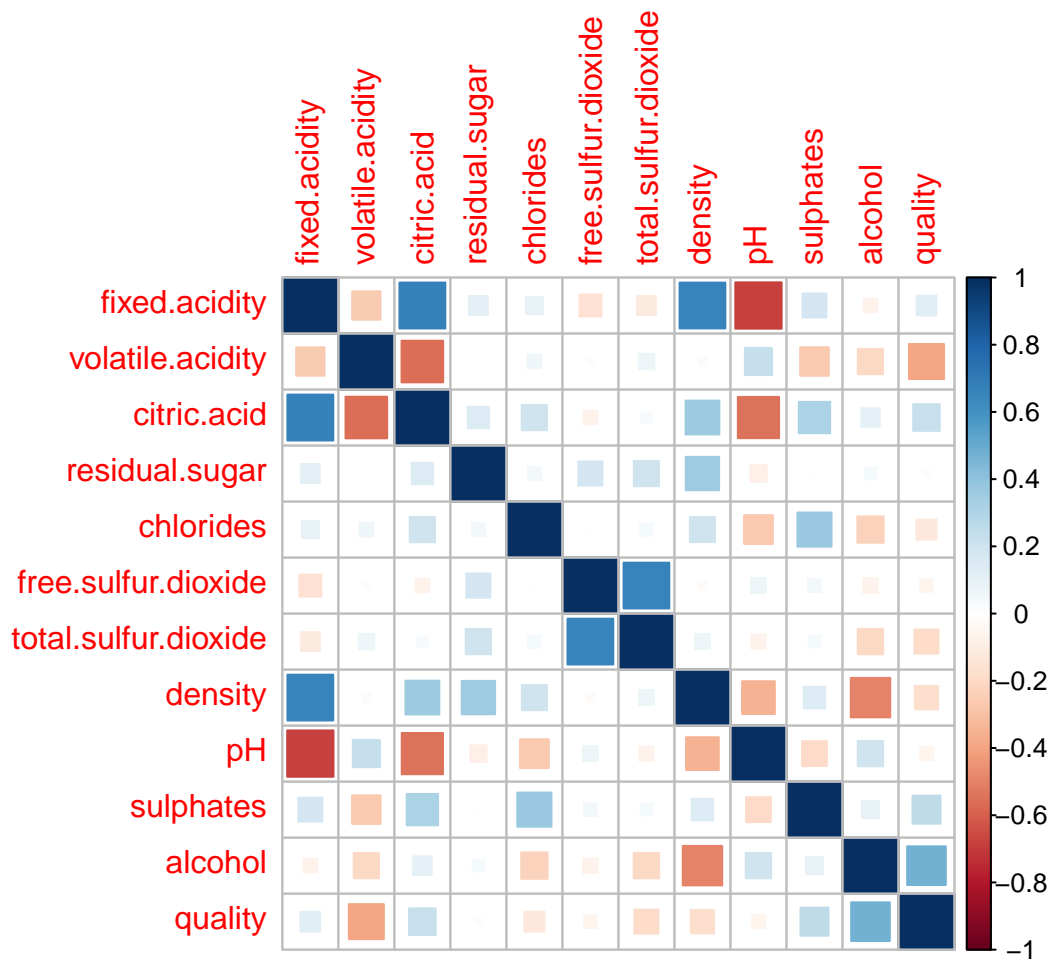


Figure 1: Distribución de las variables



- La densidad tiene una correlación muy fuerte con la acidez fija.
- Las variables más fuertemente correlacionadas con la calidad son la acidez volátil y el alcohol.
- El alcohol tiene una correlación negativa con la densidad. Esto es evidente por el hecho de que la densidad del agua es mayor que la densidad del alcohol.

Variable indicadora: pHi

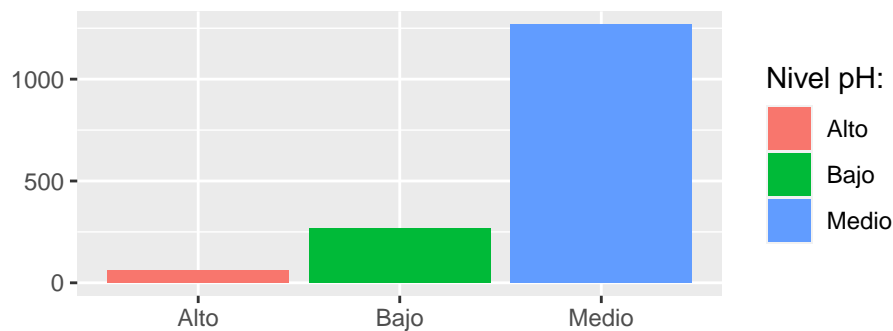
Se hace necesario crear una variable indicadora partiendo de los valores presentados por pH , esta variable indicadora cuenta con tres niveles los cuales son, bajo, medio, alto.

A partir de la siguiente figura es posible observar como el nivel de pH con mayor frecuencia es aquel que se denomina como “medio” con 1417 observaciones, mientras que los niveles “bajo” y “alto”, presentan frecuencias muy bajas, esto es, 134 y 48 respectivamente.

```
table(pHi)
```

```
## pHi
## Alto Bajo Medio
## 63 267 1269
```

G3



Modelo lineal generalizado (GLM)

En esta sección, se ajusta un modelo lineal generalizado usando como respuesta la variable “calidad” (quality) y como variables de predicción las variables “acidez fija” (fixed acidity) y “pHi”, tal como sigue:

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

```
##
## Call:
## glm(formula = Datos$quality ~ Datos$fixed.acidity + pHi, data = Datos)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8968  -0.6091   0.1887   0.4341   2.6015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.03884    0.13061  38.579  < 2e-16 ***
## Datos$fixed.acidity  0.07194    0.01365   5.272 1.54e-07 ***
## pHiBajo          -0.11957    0.12566  -0.952   0.341
## pHiMedio           0.02348    0.10672   0.220   0.826
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.6411399)
##
##      Null deviance: 1042.2  on 1598  degrees of freedom
## Residual deviance: 1022.6  on 1595  degrees of freedom
## AIC: 3833
##
## Number of Fisher Scoring iterations: 2
```

Conlusiones

Bibliografía