

Postwork - Sesión 1

Orlando Antonio Aguilar Puerto
Mario Alberto Encinas Cardona
Emanuel Flores Martínez
Andrés Benjamín Sánchez Alvarado

1/9/2021

1.Importación de los datos de soccer de la temporada 2019/2020 de la primera división de la liga española recopilados en <https://www.football-data.co.uk/spainm.php>. Los datos se almacenan en el dataframe `df.primera_division`.

```
df.primera_division <- read.csv('https://www.football-data.co.uk/mmz4281/1920/SP1.csv')
names(df.primera_division)
```

```
##      [1] "Div"          "Date"          "Time"          "HomeTeam"      "AwayTeam"
##      [6] "FTHG"         "FTAG"          "FTR"           "HTHG"          "HTAG"
##     [11] "HTR"          "HS"            "AS"            "HST"           "AST"
##     [16] "HF"           "AF"            "HC"            "AC"            "HY"
##     [21] "AY"           "HR"            "AR"            "B365H"         "B365D"
##     [26] "B365A"        "BWH"           "BWD"           "BWA"           "IWH"
##     [31] "IWD"          "IWA"           "PSH"           "PSD"           "PSA"
##     [36] "WHH"          "WHD"           "WHA"           "VCH"           "VCD"
##     [41] "VCA"          "MaxH"          "MaxD"          "MaxA"          "AvgH"
##     [46] "AvgD"         "AvgA"          "B365.2.5"      "B365.2.5.1"    "P.2.5"
##     [51] "P.2.5.1"      "Max.2.5"       "Max.2.5.1"    "Avg.2.5"       "Avg.2.5.1"
##     [56] "AHh"          "B365AHH"       "B365AHA"       "PAHH"          "PAHA"
##     [61] "MaxAHH"       "MaxAHA"        "AvgAHH"        "AvgAHA"        "B365CH"
##     [66] "B365CD"       "B365CA"        "BWCH"          "BWCD"          "BWCA"
##     [71] "IWCH"         "IWCD"          "IWCA"          "PSCH"          "PSCD"
##     [76] "PSCA"         "WHCH"          "WHCD"          "WHCA"          "VCCCH"
##     [81] "VCCD"         "VCCA"          "MaxCH"         "MaxCD"         "MaxCA"
##     [86] "AvgCH"        "AvgCD"         "AvgCA"         "B365C.2.5"     "B365C.2.5.1"
##     [91] "PC.2.5"       "PC.2.5.1"      "MaxC.2.5"     "MaxC.2.5.1"   "AvgC.2.5"
##     [96] "AvgC.2.5.1"   "AHCh"          "B365CAHH"     "B365CAHA"     "PCAHH"
##    [101] "PCAHA"        "MaxCAHH"       "MaxCAHA"      "AvgCAHH"       "AvgCAHA"
```

2.Del dataframe completo se extraen las variables correspondientes a los goles anotados por el equipo de casa (FTHG) y los anotados por el equipo visitante (FTAG) en cada partido, y se almacenan ambas columnas en el dataframe `df.goles`.

```
df.goles <- df.primera_division[, c("FTHG", "FTAG")]
```

3.Una vez extraídos los datos de los goles aislados, se generan las tablas de frecuencia para posteriormente calcular los valores de probabilidad del número de goles anotados. Con la función `table`, se agrupan el número de renglones en las que cada valor del número de goles sucedió. Esto es análogo a las funciones `GROUP BY` y `COUNT` de MySQL. Primero, se genera la tabla de frecuencias de los goles anotados por el equipo de casa:

```
table(df.goles$FTHG)
```

```
##
##    0    1    2    3    4    5    6
## 88 132  99  38  14   8   1
```

Utilizando el enfoque de probabilidad clásica, se calculan las siguientes probabilidades:

- La probabilidad (marginal) de que el equipo que juega en casa anote x goles ($x = 0, 1, 2, \dots$)
Se divide el vector de frecuencias entre el número total de partidos:

```
# table(df.goles$FTHG) / nrow(df.primera_division)
prop.table(table(df.goles$FTHG))
```

```
##
##           0           1           2           3           4           5
## 0.231578947 0.347368421 0.260526316 0.100000000 0.036842105 0.021052632
##           6
## 0.002631579
```

La función *prop.table* realiza esta operación de división automáticamente.

- La probabilidad (marginal) de que el equipo que juega como visitante anote y goles ($y = 0, 1, 2, \dots$)
La misma operación se realiza para los goles anotados por el equipo visitante:

```
prop.table(table(df.goles$FTAG))
```

```
##
##           0           1           2           3           4           5
## 0.357894737 0.352631579 0.213157895 0.047368421 0.023684211 0.005263158
```

- La probabilidad (conjunta) de que el equipo que juega en casa anote x goles y el equipo que juega como visitante anote y goles ($x = 0, 1, 2, \dots, y = 0, 1, 2, \dots$)

Por último, se aplica la misma función para calcular la probabilidad bivariada de que ambos equipos anoten x y y número de goles:

```
prop.table(table(df.goles))
```

```
##      FTAG
## FTHG    0    1    2    3    4    5
##  0 0.086842105 0.073684211 0.039473684 0.021052632 0.005263158 0.005263158
##  1 0.113157895 0.128947368 0.084210526 0.013157895 0.007894737 0.000000000
##  2 0.102631579 0.092105263 0.052631579 0.007894737 0.005263158 0.000000000
##  3 0.036842105 0.036842105 0.018421053 0.005263158 0.002631579 0.000000000
##  4 0.010526316 0.013157895 0.010526316 0.000000000 0.002631579 0.000000000
##  5 0.005263158 0.007894737 0.007894737 0.000000000 0.000000000 0.000000000
##  6 0.002631579 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
```