Problem set 10

Hrishikesh M Bharadwaj

2023-11-16

```
data <- read.csv("epl2223.csv")
head(data)
```

```
##   Div       Date  Time       HomeTeam      AwayTeam FTHG FTAG FTR HTHG HTAG HTR
## 1  E0 05/08/2022 20:00 Crystal Palace       Arsenal    0    2   A    0    1   A
## 2  E0 06/08/2022 12:30         Fulham     Liverpool    2    2   D    1    0   H
## 3  E0 06/08/2022 15:00    Bournemouth   Aston Villa    2    0   H    1    0   H
## 4  E0 06/08/2022 15:00          Leeds        Wolves    2    1   H    1    1   D
## 5  E0 06/08/2022 15:00      Newcastle Nott'm Forest    2    0   H    0    0   D
## 6  E0 06/08/2022 15:00      Tottenham   Southampton    4    1   H    2    1   H
##      Referee HS AS HST AST HF AF HC AC HY AY HR AR B365H B365D B365A   BWH  BWD
## 1   A Taylor 10 10   2   2 16 11  3  5  1  2  0  0  4.20   3.6  1.85  4.33 3.50
## 2   A Madley  9 11   3   4  7  9  4  4  2  0  0  0 11.00   6.0  1.25 10.00 5.75
## 3   P Bankes  7 15   3   2 18 16  5  5  3  3  0  0  3.75   3.5  2.00  3.75 3.40
## 4    R Jones 12 15   4   6 13  9  6  4  2  0  0  0  2.25   3.4  3.20  2.30 3.30
## 5   S Hooper 23  5  10   0  9 14 11  1  0  3  0  0  1.66   3.8  5.25  1.65 3.80
## 6 A Marriner 18 10   8   2 11  6 10  2  3  0  0  0  1.33   5.5  8.50  1.35 5.25
##    BWA   IWH  IWD  IWA   PSH  PSD  PSA   WHH WHD  WHA   VCH VCD  VCA  MaxH MaxD
## 1 1.87  4.30 3.55 1.85  4.50 3.65 1.89  4.40 3.5 1.83  4.60 3.5 1.87  4.60 3.78
## 2 1.28 12.00 5.75 1.27 11.20 6.22 1.28 12.00 5.5 1.27 13.00 6.0 1.25 13.00 6.40
## 3 2.00  3.65 3.45 2.05  3.93 3.58 2.04  3.75 3.3 2.05  3.75 3.3 2.00  4.00 3.66
## 4 2.95  2.30 3.30 3.15  2.39 3.33 3.30  2.25 3.3 3.20  2.30 3.2 3.10  2.42 3.54
## 5 5.50  1.65 3.80 5.50  1.71 3.74 5.83  1.67 3.7 5.25  1.62 3.7 5.50  1.72 3.96
## 6 8.25  1.37 5.25 7.75  1.37 5.39 9.11  1.35 5.0 8.50  1.33 5.0 9.00  1.40 5.50
##   MaxA  AvgH AvgD AvgA B365.2.5 B365.2.5.1 P.2.5 P.2.5.1 Max.2.5 Max.2.5.1
## 1 1.95  4.39 3.59 1.88     2.10       1.72  2.14    1.78    2.19      1.91
## 2 1.31 10.99 6.05 1.28     1.50       2.62  1.50    2.70    1.54      2.76
## 3 2.10  3.80 3.50 2.04     2.00       1.80  2.10    1.81    2.10      1.87
## 4 3.30  2.34 3.34 3.18     2.05       1.85  2.09    1.83    2.11      1.87
## 5 6.00  1.67 3.80 5.57     2.05       1.85  2.10    1.81    2.10      1.86
## 6 9.20  1.36 5.27 8.64     1.61       2.30  1.65    2.37    1.65      2.48
##   Avg.2.5 Avg.2.5.1   AHh B365AHH B365AHA PAHH PAHA MaxAHH MaxAHA AvgAHH AvgAHA
## 1    2.09      1.76  0.50    2.04    1.89 2.03 1.89   2.06   1.91   2.01   1.87
## 2    1.48      2.63  1.75    1.90    2.03 1.91 2.00   1.92   2.04   1.89   1.99
## 3    2.03      1.80  0.50    1.87    2.06 1.88 2.04   1.88   2.07   1.85   2.04
## 4    2.03      1.81 -0.25    2.05    1.88 2.04 1.89   2.06   1.90   2.01   1.87
## 5    2.03      1.81 -0.75    1.87    2.06 1.92 2.01   1.92   2.08   1.86   2.02
## 6    1.61      2.34 -1.50    2.04    1.89 2.08 1.85   2.08   1.91   2.03   1.85
##   B365CH B365CD B365CA BWCH BWCD BWCA  IWCH IWCD IWCA  PSCH PSCD PSCA  WHCH
## 1   4.50   3.60   1.80 4.50 3.50 1.83  4.40 3.55 1.85  4.58 3.63 1.88  4.80
## 2  11.00   5.75   1.28 9.25 6.00 1.29 11.00 5.50 1.30 10.50 6.50 1.29 11.00
## 3   4.00   3.50   1.95 3.90 3.40 1.95  3.85 3.45 2.00  4.09 3.59 2.00  4.00
## 4   2.37   3.30   3.00 2.40 3.30 2.75  2.45 3.30 2.95  2.45 3.44 3.09  2.40
## 5   1.53   4.00   6.00 1.58 3.90 6.00  1.63 3.80 6.00  1.57 4.22 6.60  1.53
## 6   1.36   5.00   8.50 1.36 5.25 8.25  1.37 5.25 8.00  1.39 5.34 8.55  1.33
##   WHCD WHCA  VCCH VCCD  VCCA MaxCH MaxCD MaxCA AvgCH AvgCD AvgCA B365C.2.5
## 1  3.4 1.78  4.75 3.50  1.85  5.01  3.70  1.91  4.56  3.57  1.85      2.10
## 2  5.5 1.27 11.50 6.00  1.29 11.95  6.93  1.30 10.33  6.20  1.28      1.50
## 3  3.4 1.95  4.10 3.40  2.00  4.25  3.63  2.06  3.99  3.49  2.00      2.10
## 4  3.3 2.90  2.40 3.40  3.00  2.50  3.55  3.18  2.43  3.36  3.02      1.95
## 5  3.9 6.50  1.57 3.90  7.00  1.67  4.30  7.00  1.59  4.07  6.15      1.94
## 6  4.8 9.50  1.33 5.25 10.00  1.40  5.50 10.00  1.37  5.24  8.59      1.61
##   B365C.2.5.1 PC.2.5 PC.2.5.1 MaxC.2.5 MaxC.2.5.1 AvgC.2.5 AvgC.2.5.1  AHCh
## 1        1.72   2.14     1.78     2.19       1.91     2.08       1.76  0.50
## 2        2.62   1.49     2.77     1.51       3.00     1.47       2.73  1.75
```

```
## 3        1.72   2.13     1.79     2.24       1.81     2.10       1.76  0.50
## 4        1.95   1.96     1.94     2.09       1.96     1.96       1.87 -0.25
## 5        1.96   1.97     1.93     2.06       1.97     1.94       1.89 -1.00
## 6        2.30   1.65     2.36     1.67       2.40     1.63       2.31 -1.50
##   B365CAHH B365CAHA PCAHH PCAHA MaxCAHH MaxCAHA AvgCAHH AvgCAHA
## 1     2.09     1.84  2.04  1.88    2.09    1.88    2.03    1.85
## 2     1.90     2.03  1.91  2.02    2.01    2.06    1.89    1.99
## 3     1.93     2.00  1.93  2.00    1.94    2.04    1.88    2.00
## 4     2.08     1.85  2.10  1.84    2.14    1.87    2.08    1.81
## 5     1.97     1.96  1.99  1.93    2.19    1.97    2.03    1.86
## 6     2.07     1.86  2.04  1.88    2.08    1.88    2.03    1.85
```

A.

```
data$TotalGoals <- data$FTHG + data$FTAG
```

B.

```
observed_counts <- table(data$TotalGoals)
observed_counts["7 or more"] <- sum(observed_counts[7:length(observed_counts)])
observed_counts <- observed_counts[1:7]
observed_counts
```

```
##  0  1  2  3  4  5  6
## 23 70 87 79 57 31 18
```

C.

```
m <- mean(data$TotalGoals)
m
```

```
## [1] 2.852632
```

D.

```
expected_probs <- dpois(0:6, m)
expected_probs[7] <- 1 - sum(expected_probs[1:6])
```

E.

```
# chi-squared test
chi_square_test <- chisq.test(observed_counts, p = expected_probs)

# chi-squared statistic and P-value
chi_square_statistic <- chi_square_test$statistic
p_value <- chi_square_test$p.value

cat("Chi-Squared Statistic:", chi_square_statistic, "\n")
```

```
## Chi-Squared Statistic: 4.269525
```

```
cat("P-Value:", p_value, "\n")
```

```
## P-Value: 0.6402535
```

```
# Conclusion
if (p_value < 0.05) {
  cat("Conclusion: Reject the null hypothesis. There is a significant association between the nu
mber of total goals and the observed counts.")
} else {
  cat("Conclusion: Fail to reject the null hypothesis. There is no significant association betwe
en the number of total goals and the observed counts.")
}
```

```
## Conclusion: Fail to reject the null hypothesis. There is no significant association between t
he number of total goals and the observed counts.
```