Baseball_1

Thomas Young

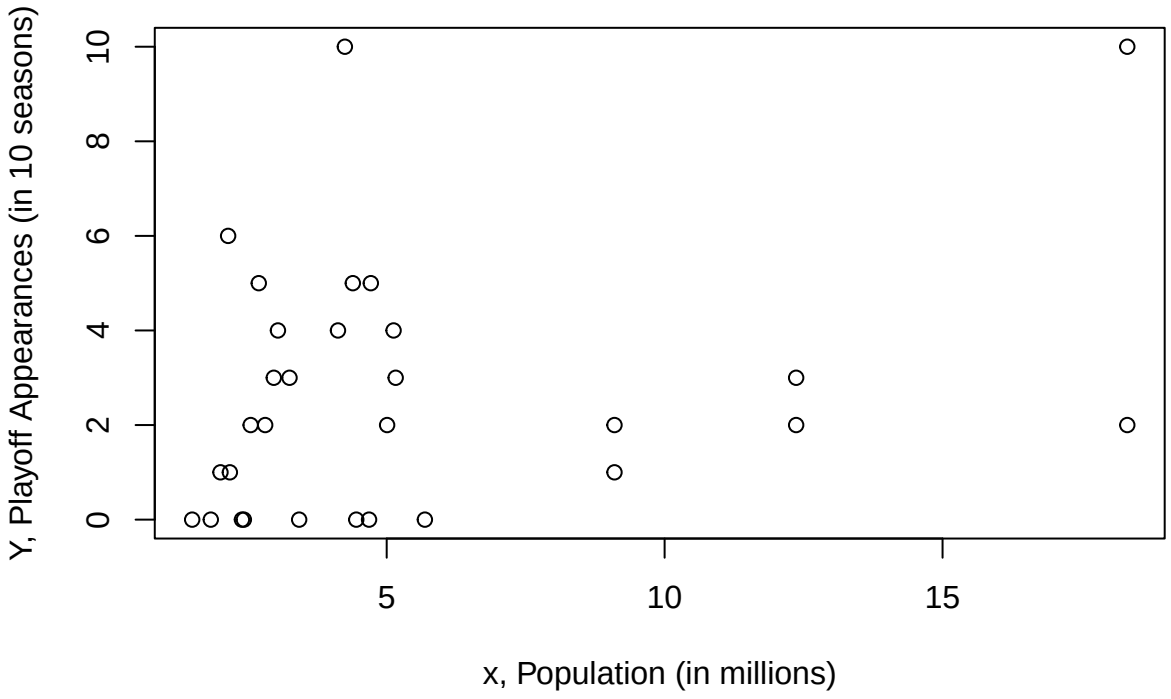
2025-03-07

Question about baseball from stats book using minimal packages

Chapter 6 of Bradbury (2007), a book on baseball, uses regression analysis to compare the success of the 30 Major League Baseball teams. For example, the author considers the relationship between x i , market size (i.e., the population in millions of the city associated with each team) and Y i , the number of times team i made the post-season playoffs in the m i =10 seasons between 1995 and 2004. Chapter 6 of Bradbury (2007), a book on baseball, uses regression analysis to compare the success of the 30 Major League Baseball teams. For example, the author considers the relationship between x i , market size (i.e., the population in millions of the city associated with each team) and Y i , the number of times team i made the post-season playoffs in the m i =10 seasons between 1995 and 2004.

```
# Read in data set and look at plot
playoffs<- read.table("https://gattonweb.uky.edu//sheather//book//docs//datasets//playoffs.txt",
                      header=TRUE)

plot(PlayoffAppearances~Population,data=playoffs,xlab="x, Population (in millions)"
     , ylab="Y, Playoff Appearances (in 10 seasons)")
```



We can see a major concern that there looks to be a moderate level of linear correlation between Population and Playoff. Another concern is the distribution does not look normal which violates our assumptions. One way we can check this to to make a model

```
# Make a model and check for evidence of a relationship
m1 = glm(PlayoffAppearances ~ Population, data = playoffs)
summary(m1)
```

```
##
## Call:
## glm(formula = PlayoffAppearances ~ Population, data = playoffs)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.7547     0.7566   2.319   0.0279 *
## Population    0.1684     0.1083   1.555   0.1311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6.859802)
##
##     Null deviance: 208.67  on 29  degrees of freedom
## Residual deviance: 192.07  on 28  degrees of freedom
## AIC: 146.84
##
## Number of Fisher Scoring iterations: 2
```

Our test statistic (0.1311) fails to reject the null hypothesis. There is not enough evidence to suggest there is a relationship between Y and x, meaning playoff appearance and population of the baseball city. The residual deviance is also very large so the model is not the best fit.