Chi Square and ANOVA

Name: Shivani Vellanki

Date: 2023/03/11

INTRODUCTION:

The principles of chi - square and ANOVA testing are used to tackle numerous difficulties in this project. The project comprises hypothesis testing, crucial values, calculated test values, and decision making.

ANALYSIS:

Section 11-1 :

6) Blood types:

1. Hypothesis:
   H0 - The distribution of blood types is same as the general population
   H1 - The distribution of blood types is not same as the general population

2. Critical Value:
   $\alpha = 0.1$
   ```
   > alpha
   [1] 0.1
   ```

3. Compute the test value:

```
> observed <- as.vector(c(12, 8, 24, 6)) # Random Sample of 50 Patients
> p <- c(0.2, 0.28, 0.36, 0.16)
>
> result =chisq.test(x=observed, p=p)
> result

	Chi-squared test for given probabilities

data:  observed
X-squared = 5.4714, df = 3, p-value = 0.1404
```

4. Decision:

```
> ifelse(result$p.value>alpha, "Fail to reject null hypothesis",
+        "Reject the null hypothesis")
[1] "Fail to reject null hypothesis"
```

5. Summarise Decision:
   Since p-value>α, there is inadequate evidence to support the idea that blood type distribution differs from that of the general population.

8) On-time Performance by Airlines:

1. Hypothesis:
   H0 - Airlines on-time performance is same as the government's statistics
   H1 - Airlines on-time performance is not same as the government's statistics

2. Critical Value:
   $\alpha = 0.05$
   ```
   > alpha
   [1] 0.05
   ```

3. Compute the test value:
   ```
   > observed =c(125, 10,25, 40)
   > p = c(0.708, 0.082, 0.09, 0.12)
   >
   > result= chisq.test(x=observed, p=p)
   > result

           Chi-squared test for given probabilities

   data:  observed
   X-squared = 17.832, df = 3, p-value = 0.0004763
   ```

4. Decision:
   ```
   > ifelse(result$p.value>alpha, "Fail to reject the null hypothesis", "Reject the null hypothesis")
   [1] "Reject the null hypothesis"
   ```

5. Summarise Decision:
   Since the p-value<α, there is enough information to conclude that airline on-time performance differs from official data.

Section 11-2:

8) Ethnicity and Movie Admissions:

1. Hypothesis:
   H0 - The movie attendance is independent of ethnicity.
   H1 - The movie attendance is not independent of ethnicity.

2. Critical Value:

   α = 0.05

```
> alpha
[1] 0.05
```

3. Compute the test values:

```
> #Creating vectors for rows of matrix
> r1 =c(724, 370)
> r2= c(335,292)
> r3 = c(174, 152)
> r4 =c(107,140)
>
> numberOfRows=4
>
> #matrix from the rows
> mtrx = matrix(c(r1,r2,r3,r4), nrow = numberOfRows, byrow = TRUE)
>
> rownames(mtrx)=c("Caucasian","Hispanic","African American", "Other")
>
> colnames(mtrx)=c("2013","2014")
> mtrx
                 2013 2014
Caucasian         724  370
Hispanic          335  292
African American  174  152
Other             107  140
>
> result <- chisq.test(mtrx)
> result

	Pearson's Chi-squared test

data:  mtrx
X-squared = 60.144, df = 3, p-value = 5.478e-13
```

4. Decision:

```
> ifelse(result$p.value>alpha, "Fail to reject the null hypothesis", "Reject the null hypothesis")
[1] "Reject the null hypothesis"
```

5. Summarise Decision:

   Since p-value<α, there is adequate data to conclude that ethnicity has no effect on movie attendance.

10) Women in Military:

1. Hypothesis:
   H0 - There is no relationship exists between rank and branch of armed forces
   H1 - There is relationship between rank and branch of armed forces

2. Critical Value:
   α = 0.05
   ```
   > alpha
   [1] 0.05
   ```

3. Compute the test value:
   ```
   > #Create vectors of rows of matrix
   > r1 = c(10791, 62491)
   > r2 = c(7816, 42750)
   > r3 = c(932, 9525)
   > r4 = c(11819, 54344)
   >
   > numberOfRows=4
   >
   > #Create matrix of rows
   > mtrx =matrix(c(r1,r2,r3,r4), nrow = numberOfRows, byrow = TRUE) > mtrx
   >
   > #Name the rows and columns
   > rownames(mtrx)=c("Army","Navy","Marine Corps", "Air Force")
   >
   > colnames(mtrx)=c("Officers", "Enlisted")
   > result <- chisq.test(mtrx)
   Warning message:
   In chisq.test(mtrx) : Chi-squared approximation may be incorrect
   > result

           Pearson's Chi-squared test

   data:  mtrx
   X-squared = 0, df = 3, p-value = 1
   ```

4. Decision:
   ```
   > ifelse(result$p.value>alpha, "Fail to reject the null hypothesis", "Reject the null hypothesis")
   [1] "Fail to reject the null hypothesis"
   ```

5. Summarise Decision:

   Since p-value>α, there is insufficient evidence to establish a relationship between military rank and branch.

Section 12-1 :

8) Sodium Contents of Foods:

1. Hypothesis:
   H0 - There is no difference in the mean amounts of sodium content among snacks, cereals and desserts.
   H1 - There is a difference in the mean amounts of sodium content among snacks, cereals and desserts

2. Critical Value:
   α = 0.05

```
> alpha
[1] 0.05
```

3. Compute test values:

```
> #Creating dataframes for each food type
> condimentsDF <- data.frame("sodium"=c(270,130,230,180,80,70,200), "foodType"=rep("condiments",7), stringsAsFa
ctors=FALSE)
>
> cerealsDF <- data.frame("sodium"=c(260,220,290,290,200,320,140),"foodType"=rep("cereals",7), stringsAsFactors
 = FALSE)
>
> dessertsDF <- data.frame("sodium"=c(100,180,250,250,300,360,300,160),"foodType"=rep("desserts",8), stringsAsF
actors=FALSE)
>
> #Combining all the above data.frames into one
> sodiumDF <- rbind(condimentsDF, cerealsDF, dessertsDF)
> str(sodiumDF)
'data.frame':	22 obs. of  2 variables:
 $ sodium  : num  270 130 230 180 80 70 200 260 220 290 ...
 $ foodType: chr  "condiments" "condiments" "condiments" "condiments" ...
> sodiumDF$food <- as.factor(sodiumDF$food) # changing variable from char to factor
>
> #Running the ANOVA test
> sodiumAnova <- aov(sodium~foodType, data = sodiumDF)
> summary(sodiumAnova)
            Df Sum Sq Mean Sq F value Pr(>F)
foodType     2  27544   13772   2.399  0.118
Residuals   19 109093    5742
> #save summary to an object
> a.summary = summary(sodiumAnova)
>
> #Degrees of freedom
> # k-1: between group variance - numerator
> df.numerator = a.summary
> df.numerator
            Df Sum Sq Mean Sq F value Pr(>F)
foodType     2  27544   13772   2.399  0.118
Residuals   19 109093    5742
>
> #n-k: within group variance -denominator
>
> df.denominator <- a.summary
> df.denominator
            Df Sum Sq Mean Sq F value Pr(>F)
foodType     2  27544   13772   2.399  0.118
Residuals   19 109093    5742
```

```
> #Extract the F-test value from the summary
> F.value <- a.summary[[1]][1, "F value"]
> F.value
[1] 2.398538
>
> #Extract p-value from the summary
> p.value <- a.summary[[1]][1, "Pr(>F)"]
> p.value
[1] 0.1178108
```

4. Decision:

```
> ifelse(p.value>alpha, "Fail to reject null hypothesis", "Reject null hypothesis")
[1] "Fail to reject null hypothesis"
```

5. Summarise Result:

```
> TukeyHSD(sodiumAnova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sodium ~ foodType, data = sodiumDF)

$foodType
                          diff        lwr       upr     p adj
condiments-cereals  -80.000000 -182.89588  22.89588 0.1456674
desserts-cereals     -8.214286 -107.84279  91.41422 0.9761344
desserts-condiments  71.785714  -27.84279 171.41422 0.1866850
```

By using Tukey package in R, it shows as below that all the p-value of each pair of food is greater than alpha = 0.05, as a result, none of their differences are statistically significant. As a result, there is insufficient information to conclude that there is a statistically significant difference in salt level across condiments, cereals, and sweets.

Section 12-2:

10) Sales of Leading Companies:

1. Hypothesis:
   H0: There is no significant difference in the means of the sales among three companies
   H1: There is a significant difference in the means of the sales among three companies

2. Significance level:
   $\alpha = 0.01$

3. Compute test values:

```
> #Create data.frame for the companies
> cereal = data.frame("Sales"=c(578,320,264,249,237), "Company"=rep("Cereal",5), stringsAsFactors = FALSE)
> Chocolate = data.frame("Sales"=c(311,106,109,125,173),"Company"=rep("Chocolate Candy", 5), stringsAsFactors =
FALSE)
> Coffee = data.frame("Sales"=c(261,185,302,689),"Company"=rep("Coffee",4), stringsAsFactors = FALSE)
> sales = rbind(cereal, Chocolate, Coffee)
> sales$Company = as.factor(sales$Company)
>
> anova = aov(Sales~Company, data=sales)
> #summary of the result
> summary(anova)
            Df Sum Sq Mean Sq F value Pr(>F)
Company      2 103770   51885   2.172   0.16
Residuals   11 262795   23890
> a.summary = summary(anova)
> df.numerator = a.summary
> df.numerator
            Df Sum Sq Mean Sq F value Pr(>F)
Company      2 103770   51885   2.172   0.16
Residuals   11 262795   23890
>
> #n-k: within group variance: denominator
> df.denominator <- a.summary
> df.denominator
            Df Sum Sq Mean Sq F value Pr(>F)
Company      2 103770   51885   2.172   0.16
Residuals   11 262795   23890
>
> #Extract the F-test value from the summary
> F.value <- a.summary[[1]][1, "F value"]
> F.value
[1] 2.171782
>
> #Extract p-value from the summary
> p.value <- a.summary[[1]][1, "Pr(>F)"]
> p.value
[1] 0.1603487
```

4. Decision:

```
> ifelse(p.value>alpha, "Fail to reject null hypothesis", "Reject null hypothesis")
[1] "Fail to reject null hypothesis"
```

5. Summarise Decision:

```
> TukeyHSD(anova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Sales ~ Company, data = sales)

$Company
                          diff        lwr       upr     p adj
Chocolate Candy-Cereal -164.80 -428.82409  99.22409 0.2535458
Coffee-Cereal            29.65 -250.38983 309.68983 0.9561014
Coffee-Chocolate Candy  194.45  -85.58983 474.48983 0.1916553
```

Since p-value for all companies is > α, as a result, there is a difference in sales between the three firms.

12) Per-Pupil Expenditures :

1. Hypothesis;
   H0 - There is no difference in the means of expenditures among three sections of the country.
   H1 - There is a difference in the means of expenditures among three sections of the country.

2. Significant level:
   α = 0.05

```
> Alpha
[1] 0.05
```

3. Compute Test Values:

```
> eastern = data.frame("Expenditures"=c(4946, 5953, 6202, 7243, 6113), "Section"=rep("Eastern third",5), string
sAsFactors = FALSE)
>
> middle = data.frame("Expenditures"=c(6149,7451,6000,6479), "Section"=rep("Middle third", 4), stringsAsFactors
 = FALSE)
> western = data.frame("Expenditures"=c(5282,8605,6528,6911), "Section"=rep("Western third", 4), stringsAsFacto
rs = FALSE)
> expenditure = rbind(eastern,middle, western)
> expenditure$Section = as.factor(expenditure$Section)
> anova = aov(Expenditures~Section, data=expenditure)
> a.summary = summary(anova)
>
> df.numerator = a.summary
> df.numerator
            Df  Sum Sq Mean Sq F value Pr(>F)
Section      2 1244588  622294   0.649  0.543
Residuals   10 9591145  959114
>
> #n-k: within group variance: denominator
> df.denominator <- a.summary
> df.denominator
            Df  Sum Sq Mean Sq F value Pr(>F)
Section      2 1244588  622294   0.649  0.543
Residuals   10 9591145  959114
>
> #Extract the F-test value from the summary
> F.value <- a.summary[[1]][1, "F value"]
> F.value
[1] 0.6488214
>
> #Extract p-value from the summary
> p.value <- a.summary[[1]][1, "Pr(>F)"]
> p.value
[1] 0.5433264
```

4. Decision:

```
> ifelse(p.value>alpha, "Fail to reject null hypothesis", "Reject null hypothesis")
[1] "Fail to reject null hypothesis"
```

5. Summarise Decision:

```
> TukeyHSD(anova)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Expenditures ~ Section, data = expenditure)

$Section
                              diff       lwr      upr     p adj
Middle third-Eastern third  428.35 -1372.582 2229.282 0.7954670
Western third-Eastern third 740.10 -1060.832 2541.032 0.5203918
Western third-Middle third  311.75 -1586.599 2210.099 0.8954324
```

Because the p-values of the three portions of the nation are more than **α**, we may conclude that there is a variation in the means of spending among the three sections of the country.

Section 12-3 :

10) Increasing plant growth:

1. Hypothesis:

   H0 - There is no difference in the mean growth concerning light

   There is no difference in the mean growth concerning plant food

   There is no interaction between plant food and light

   H1 - There is a difference in the mean growth concerning light

   There is a difference in the mean growth concerning plant food

   There is an interaction between plant food and light

2. Significant Level:

```
> alpha
[1] 0.05
```

3. Compute Test values:

```
> data = data.frame(C1=c("A", "B"), C2=c("9.2,9.4,8.9","7.1,7.2,8.5"), C3=c("8.5,9.2,8.9","5.5,5.8,7.6"), strin
gsAsFactors = FALSE)
> names(data)=c("Plant_food", "Light1", "Light2")
```

```
> plant = plant%>% gather(Light, Inches,Light1:Light2 )
> anova_2 = aov(Inches~Plant_food+Light + Plant_food:Light, data=plant)
> a.anova2 = summary(anova_2)

> a.anova2
                  Df Sum Sq Mean Sq F value  Pr(>F)
Plant_food         1 12.813  12.813  24.562 0.00111 **
Light              1  1.920   1.920   3.681 0.09133 .
Plant_food:Light   1  0.750   0.750   1.438 0.26482
Residuals          8  4.173   0.522
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> p.value1 = a.anova2[[1]][1, "Pr(>F)"]
> p.value1
[1] 0.001112418
> p.value2 = a.anova2[[1]][2, "Pr(>F)"]
> p.value2
[1] 0.09133137
> p.value3 = a.anova2[[1]][3, "Pr(>F)"]
> p.value3
[1] 0.2648194
```

4. Decision:

```
> ifelse(p.value1>alpha,"There is no difference in the mean growth concerning plant food",
+        "There is a significant difference in the mean growth concerning plant food" )
[1] "There is a significant difference in the mean growth concerning plant food"
>
> ifelse(p.value2>alpha, "There is no difference in the mean growth concerning light",
+        "There is a significant difference in the mean growth concerning light")
[1] "There is no difference in the mean growth concerning light"
>
> ifelse(p.value3 >alpha,"There is no difference in the mean growth concerning plant food",
+        "There is a significant difference in the mean growth concerning plant food" )
[1] "There is no difference in the mean growth concerning plant food"
```

ON MY OWN;

Introduction:

The dataset is about baseball team victories from 1962 to 2012. The data collection includes information on all Major League Baseball clubs, including All National League and American League teams. This dataset is made up of a data frame with 1232 observations and 15 variables, three of which are categorical while the remaining twelve are numeric.

1. **Importing the dataset into R**

```
> # On Your Own - baseball.csv
> baseballDF= read.csv("/Users/shivanivellanki/Downloads/baseball.csv",1)
```

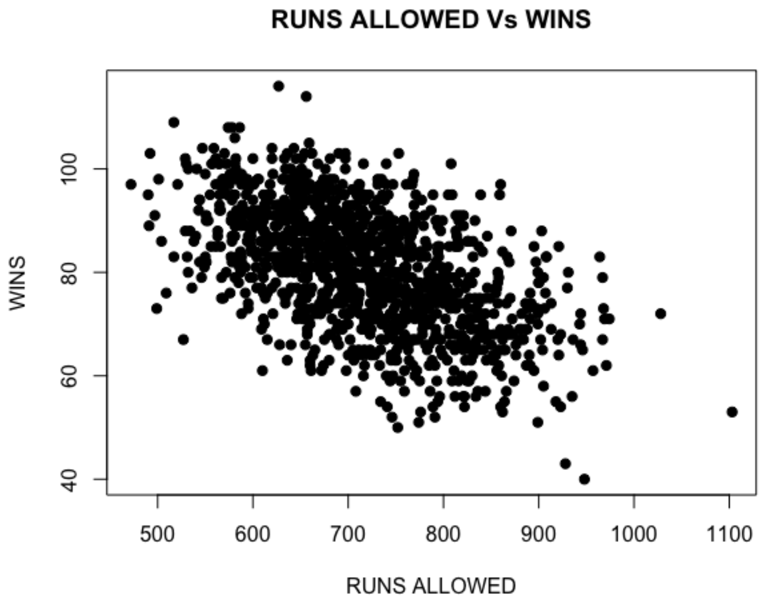
```
> baseballDF
   Team League Year  RS  RA  W   OBP   SLG    BA Playoffs RankSeason RankPlayoffs   G  OOBP  OSLG
1   ARI     NL 2012 734 688 81 0.328 0.418 0.259        0         NA           NA 162 0.317 0.415
2   ATL     NL 2012 700 600 94 0.320 0.389 0.247        1          4            5 162 0.306 0.378
3   BAL     AL 2012 712 705 93 0.311 0.417 0.247        1          5            4 162 0.315 0.403
4   BOS     AL 2012 734 806 69 0.315 0.415 0.260        0         NA           NA 162 0.331 0.428
5   CHC     NL 2012 613 759 61 0.302 0.378 0.240        0         NA           NA 162 0.335 0.424
6   CHW     AL 2012 748 676 85 0.318 0.422 0.255        0         NA           NA 162 0.319 0.405
7   CIN     NL 2012 669 588 97 0.315 0.411 0.251        1          2            4 162 0.305 0.390
8   CLE     AL 2012 667 845 68 0.324 0.381 0.251        0         NA           NA 162 0.336 0.430
9   COL     NL 2012 758 890 64 0.330 0.436 0.274        0         NA           NA 162 0.357 0.470
10  DET     AL 2012 726 670 88 0.335 0.422 0.268        1          6            2 162 0.314 0.402
11  HOU     NL 2012 583 794 55 0.302 0.371 0.236        0         NA           NA 162 0.337 0.427
12  KCR     AL 2012 676 746 72 0.317 0.400 0.265        0         NA           NA 162 0.339 0.423
13  LAA     AL 2012 767 699 89 0.332 0.433 0.274        0         NA           NA 162 0.310 0.403
14  LAD     NL 2012 637 597 86 0.317 0.374 0.252        0         NA           NA 162 0.310 0.364
15  MIA     NL 2012 609 724 69 0.308 0.382 0.244        0         NA           NA 162 0.327 0.399
16  MIL     NL 2012 776 733 83 0.325 0.437 0.259        0         NA           NA 162 0.326 0.414
17  MIN     AL 2012 701 832 66 0.325 0.390 0.260        0         NA           NA 162 0.333 0.442
18  NYM     NL 2012 650 709 74 0.316 0.386 0.249        0         NA           NA 162 0.315 0.401
19  NYY     AL 2012 804 668 95 0.337 0.453 0.265        1          3            3 162 0.311 0.419
20  OAK     AL 2012 713 614 94 0.310 0.404 0.238        1          4            4 162 0.306 0.378
21  PHI     NL 2012 684 680 81 0.317 0.400 0.255        0         NA           NA 162 0.306 0.407
22  PIT     NL 2012 651 674 79 0.304 0.395 0.243        0         NA           NA 162 0.314 0.390
23  SDP     NL 2012 651 710 76 0.319 0.380 0.247        0         NA           NA 162 0.319 0.398
24  SEA     AL 2012 619 651 75 0.296 0.369 0.234        0         NA           NA 162 0.308 0.394
25  SFG     NL 2012 718 649 94 0.327 0.397 0.269        1          4            1 162 0.313 0.393
26  STL     NL 2012 765 648 88 0.338 0.421 0.271        1          6            3 162 0.313 0.387
```

2. Descriptive Statistics:

```
> # Descriptive Statistics
> str(baseballDF)
'data.frame':   1232 obs. of  15 variables:
 $ Team        : chr  "ARI" "ATL" "BAL" "BOS" ...
 $ League      : chr  "NL" "NL" "AL" "AL" ...
 $ Year        : int  2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
 $ RS          : int  734 700 712 734 613 748 669 667 758 726 ...
 $ RA          : int  688 600 705 806 759 676 588 845 890 670 ...
 $ W           : int  81 94 93 69 61 85 97 68 64 88 ...
 $ OBP         : num  0.328 0.32 0.311 0.315 0.302 0.318 0.315 0.324 0.33 0.335 ...
 $ SLG         : num  0.418 0.389 0.417 0.415 0.378 0.422 0.411 0.381 0.436 0.422 ...
 $ BA          : num  0.259 0.247 0.247 0.26 0.24 0.255 0.251 0.251 0.274 0.268 ...
 $ Playoffs    : int  0 1 1 0 0 0 1 0 0 1 ...
 $ RankSeason  : int  NA 4 5 NA NA NA 2 NA NA 6 ...
 $ RankPlayoffs: int  NA 5 4 NA NA NA 4 NA NA 2 ...
 $ G           : int  162 162 162 162 162 162 162 162 162 162 ...
 $ OOBP        : num  0.317 0.306 0.315 0.331 0.335 0.319 0.305 0.336 0.357 0.314 ...
 $ OSLG        : num  0.415 0.378 0.403 0.428 0.424 0.405 0.39 0.43 0.47 0.402 ...
```

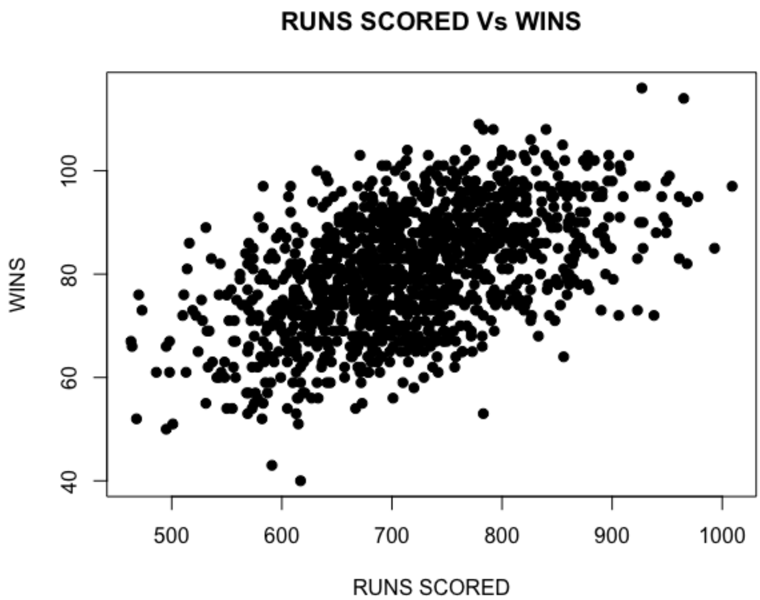
```
> summary(baseballDF)
     Team              League               Year            RS               RA               W        
 Length:1232        Length:1232        Min.   :1962   Min.   : 463.0   Min.   : 472.0   Min.   : 40.0  
 Class :character   Class :character   1st Qu.:1977   1st Qu.: 652.0   1st Qu.: 649.8   1st Qu.: 73.0  
 Mode  :character   Mode  :character   Median :1989   Median : 711.0   Median : 709.0   Median : 81.0  
                                       Mean   :1989   Mean   : 715.1   Mean   : 715.1   Mean   : 80.9  
                                       3rd Qu.:2002   3rd Qu.: 775.0   3rd Qu.: 774.2   3rd Qu.: 89.0  
                                       Max.   :2012   Max.   :1009.0   Max.   :1103.0   Max.   :116.0  

      OBP              SLG               BA             Playoffs        RankSeason     RankPlayoffs  
 Min.   :0.2770   Min.   :0.3010   Min.   :0.2140   Min.   :0.0000   Min.   :1.000   Min.   :1.000  
 1st Qu.:0.3170   1st Qu.:0.3750   1st Qu.:0.2510   1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:2.000  
 Median :0.3260   Median :0.3960   Median :0.2600   Median :0.0000   Median :3.000   Median :3.000  
 Mean   :0.3263   Mean   :0.3973   Mean   :0.2593   Mean   :0.1981   Mean   :3.123   Mean   :2.717  
 3rd Qu.:0.3370   3rd Qu.:0.4210   3rd Qu.:0.2680   3rd Qu.:0.0000   3rd Qu.:4.000   3rd Qu.:4.000  
 Max.   :0.3730   Max.   :0.4910   Max.   :0.2940   Max.   :1.0000   Max.   :8.000   Max.   :5.000  
                                                                     NA's   :988     NA's   :988    
       G              OOBP             OSLG       
 Min.   :158.0   Min.   :0.2940   Min.   :0.3460  
 1st Qu.:162.0   1st Qu.:0.3210   1st Qu.:0.4010  
 Median :162.0   Median :0.3310   Median :0.4190  
 Mean   :161.9   Mean   :0.3323   Mean   :0.4197  
 3rd Qu.:162.0   3rd Qu.:0.3430   3rd Qu.:0.4380  
 Max.   :165.0   Max.   :0.3840   Max.   :0.4990  
                 NA's   :812      NA's   :812     
```

Scatterplot of Runs Allowed VS Wins:



- We can observe from the above graph that the variable Runs Allowed has a negative correlation with Wins.

Scatterplot of Runs Scores VS Wins:



- We can observe from the above graph that the variable Runs Scored has a negative correlation with Wins.

3. Hypothesis:
   H0 - There is no difference in the wins by decade
   H1 - There is a difference in the wins by decade

   Critical Value:
   $\alpha = 0.05$

   Compute Test Value:

```
> r1 <- baseballDF$Team
> r2 <- baseballDF$RA
> r3 <- baseballDF$RS
> r4 <- baseballDF$W
> #matrix from the rows
> mtrx = matrix(c(r1,r2,r3,r4), nrow = rows, byrow = TRUE)
> #-naming rownames and colnames
> rownames(mtrx)=c("TEAM","RA","RS", "W")
> colnames(mtrx)= baseballDF$Year
> View(mtrx)
> result <- chisq.test(mtrx)
> result

        Pearson's Chi-squared test

data:  mtrx
X-squared = 19690, df = 3693, p-value < 2.2e-16
```

   Decision:

```
> ifelse(result$p.value>alpha, "Fail to reject the null hypothesis", "Reject the null hypothesis")
[1] "Reject the null hypothesis"
```

   Summarise Decision:

   Since p-value<α , there is no evidence to suggest that there is a difference in victories by decade.

References:

Team, D. (2021, August 25). *Chi-Square Test in R | Explore the Examples and Essential concepts!* DataFlair. https://data-flair.training/blogs/chi-square-test-in-r/

ANOVA in R. (n.d.). Stats and R. https://statsandr.com/blog/anova-in-r/