Project Report- Hypothesis Testing

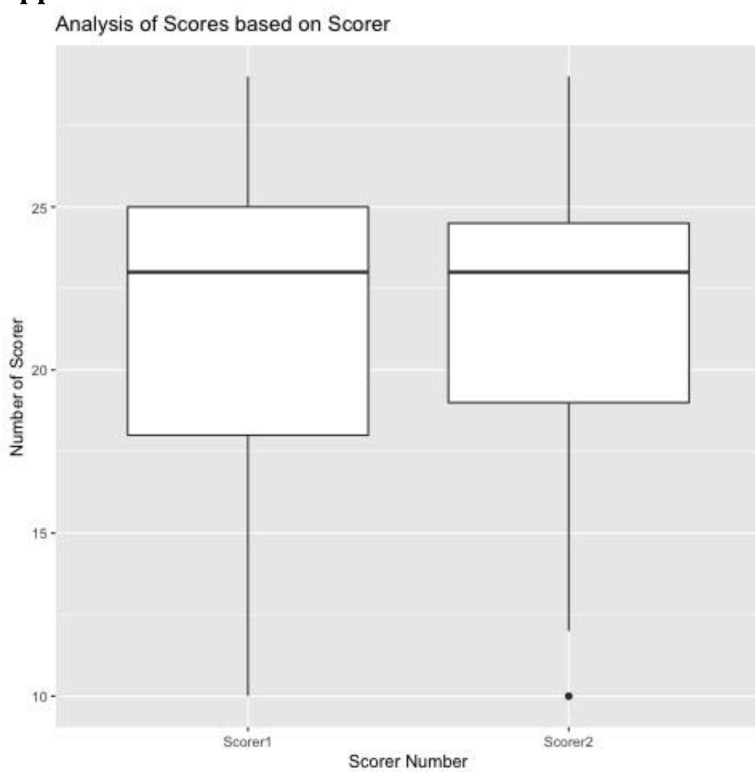
Analysis of Scores:

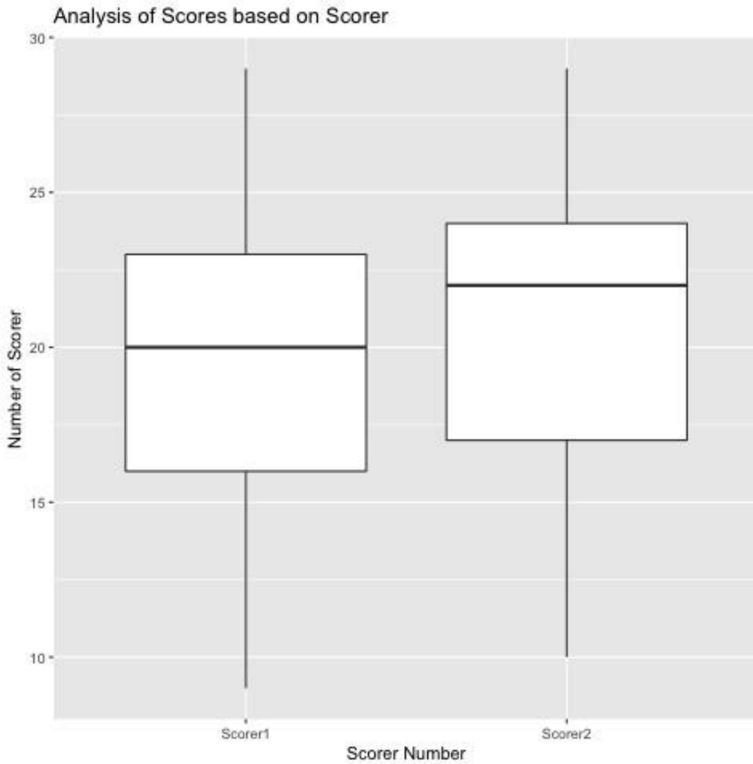
Hypothesis:

NullHypothesis : H_0 = The mean of scores is same for both the Scorers

AlternateHypothesis : H_1 = The mean of scores is different for both the Scorers

Approach:Wilcox Test:





Cutting : When performed Wilcox test, p-value is greater than 0.05, which applies the means have not changed
Suturing: When performed Wilcox test, p-value is less than 0.05, which states that the means of the scorers is different.

Hence Scorer has an effect for Suturing , not Cutting

Hypothesis testing – Linear Modelling

Linear modelling gives the relationship between the dependent and independent variables

In our data set we are finding the hypothesis between each attribute such as Age, sex, year and mean perspiration with the scores of scorer.

1.hypothesis : the score obtained does not depend on the demographics of the subject , session , age , year , sex and perspiration.

For this we did linear modelling of scores given by two scorers with all the other attributes

1.Linear modelling of Scores with all other attributes

```
> scorer1lm= lm(formula = df$Scores~df$Mean_Perspiration+df$Age+df$Sex+df$Session+df$Task+df$Scorer)
> print(scorer1lm)

Call:
lm(formula = df$Scores ~ df$Mean_Perspiration + df$Age + df$Sex +
    df$Session + df$Task + df$Scorer)

Coefficients:
         (Intercept)  df$Mean_Perspiration                df$Age                df$Sex            df$Session
             2.12823             -34.67233               0.46067               2.12787               1.91609
      df$TaskSuturing       df$ScorerScorer2
            -1.01505               0.02941
```

Intercept is 2.128 and coefficient for mean perspiration is -34.67 , coefficient for age is 0.4606 . coefficient for sex is 2.127 , coefficient for session is 1.916 . so the complete regression equation is

Score1= 2.128+ -34.67*mean perspiration+ 0.4607*Age + 2.127*Sex+1.916*Session+-1.015*task+0.029*scorer

This equation informs us that scores will increase by -34.67 for every one percent increase in mean Perspiration value , and score is directly proportional to age which states that the if older age people are hired the score would have increased.

2.linear modelling of scores Vs age

Hypothesis – the score obtained does not depend on the age of the subject

```
> score1_age=lm(formula = df$Scores~df$Age )
> print(score1_age)

Call:
lm(formula = df$Scores ~ df$Age)

Coefficients:
(Intercept)       df$Age
      5.865        0.641
```

3.linear modelling of scores Vs year

Hypothesis – the score obtained does not depend on the year of the subject

```
> score1_year=lm(formula = df$Scores~df$Year)
> print(score1_year)

Call:
lm(formula = df$Scores ~ df$Year)

Coefficients:
(Intercept)      df$Year
     18.816        1.341
```

4.linear modelling of scores with sex

Hypothesis – the score obtained does not depend on the sex of the subject

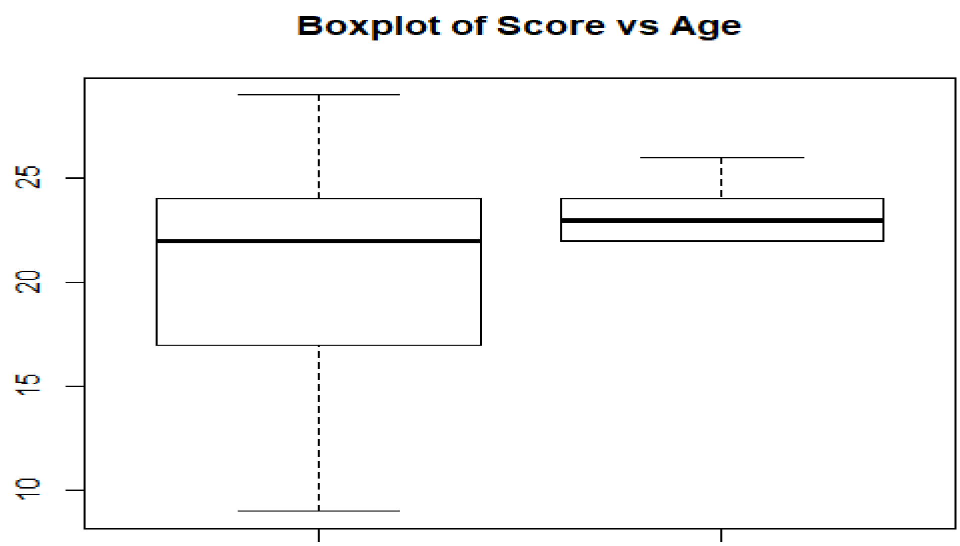
```
> score1_sex=lm(formula = df$Scores~df$Sex)
> print(score1_sex)

Call:
lm(formula = df$Scores ~ df$Sex)

Coefficients:
(Intercept)       df$Sex
      17.24         2.59
```

5.linear modelling of scores with perspiration

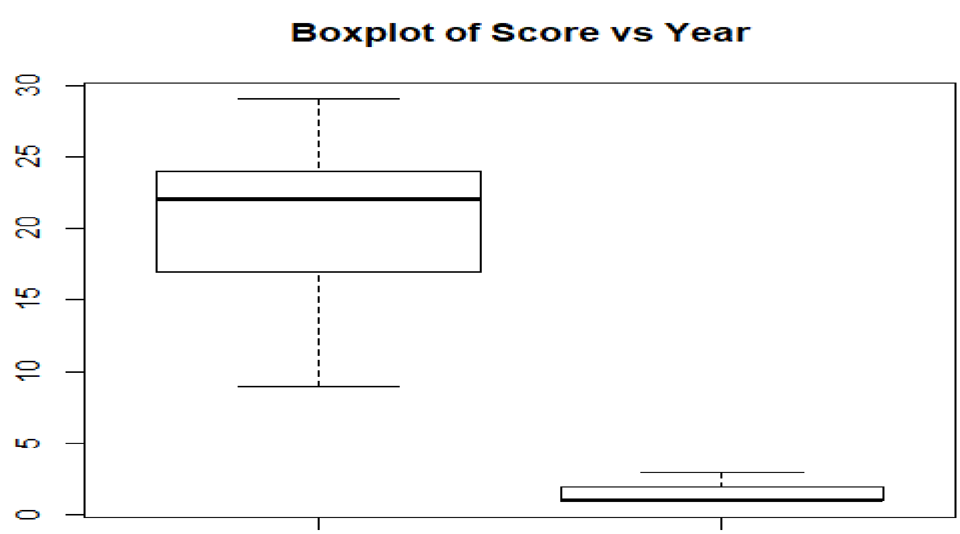
Hypothesis – the score obtained does not depend on the perspiration value of the subject

```
> score1_meanp=lm(formula = df$Scores~df$Mean_Perspiration)
> print(score1_meanp)

Call:
lm(formula = df$Scores ~ df$Mean_Perspiration)

Coefficients:
         (Intercept)  df$Mean_Perspiration
               20.82                368.27
```

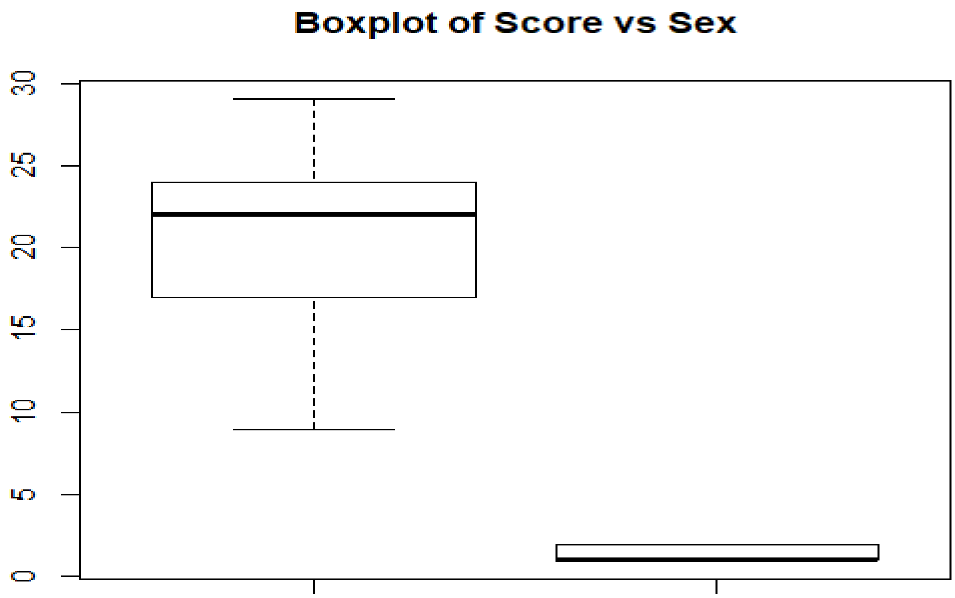
Box plot of Score Vs Age



Box plot of score vs year



Box plot of score vs Sex



Box plot of score vs Mean Perspiration