UTM

UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING

SEMESTER II 2020/2021

SECI2143 PROBABILITY & STATISTICAL

DATA ANALYSIS

SECTION: 07

PROJECT 2

LECTURER: DR. NIES HUI WEN

Group 6: Sisthreehood

Project Title: FIFA 19 Players

Group Members:

1. Eunice Lim Xian Ni (A20EC0034)

2. Shabrina Salsabila Sakroni (A20EC0334)

3. Teh Jing Ling (A20EC0228)

TABLE OF CONTENTS

1.0 INTRODUCTION AND BACKGROUND

Since we have always been curious about the factors and relationship that may affect the athletes' performance during their competitions, thus we have decided to choose FIFA 19 players dataset to conduct inference statistical analysis for our project 2. The title for this project is FIFA 19 Players as we have collected the dataset of FIFA 19 players from an open-source website.

Link: https://www.kaggle.com/karangadiya/fifa19

The purpose of this study is to identify the variables and characteristics that affect the FIFA 19 players' athletics performance. By doing so, we need to conduct several hypothesis tests to identify if the datasets for FIFA 19 players are accurate and can be generally applied to the public. We are also interested in testing with several selected variables such as potential, stamina, strength, balance, ball control, and preferred foot of FIFA 19 players from different nationalities. We will mainly focus on inferential statistics because it can provide suggestions or explanations for our hypothesis tests and determine the variables' relationship.

2.0 DATASET

The FIFA 19 players dataset consists of 89 variables. In our study, we did not use all the variables and columns given. The variables we have chosen are potential, stamina, strength, balance, ball control, and preferred foot, while for nationality we have chosen Guinea, Egypt, England, Belgium, and France. Below is the description of the selection of variables for each test from the FIFA 19 complete players dataset:

1. **Hypothesis testing with 1 sample**
   The potential of FIFA 19 players from Egypt are selected to conduct hypothesis testing for 1 sample. Since our sample size is 31, thus we used z-test for our test. By looking at their potential, we can identify whether the mean value of potential of players is greater than the null hypothesis so that we can determine to reject or fail to reject the null hypothesis.

2. **Correlation test**
   The stamina and strength of FIFA 19 players have been chosen for the correlation test. This is because Stamina is the strength that enables us to maintain the physical or mental effort for long period. Stamina is the independent variable while strength is the dependent variable. The strength of a player increases when the stamina of a player increases. The sample size for this test is 31.

3. **Regression test**
   In the regression test, the balance and ball control of the FIFA 19 players is selected. Balance is important because good balance can improve a player's ball control and further improve the player's kicking performance. Hence, balance is chosen as an independent variable, and ball control selected as the dependent variable. The sample size for this test is 31.

4. **Chi-square test of independence**
   The dataset chosen for this test is preferred foot among FIFA 19 players from 3 different nationality which are Egypt, Guinea, and England. This is because we want to find out whether 3 different nationality of FIFA 19 players and are independent with their preferred foot. The sample size for this test is 92 players.

5. **ANOVA test**
   For ANOVA test, we have chosen stamina of FIFA 19 players from 5 different nationality which are Guinea, Egypt, England, Belgium and France. The reason why we chose to test the mean of stamina is because we are curious if the FIFA 19 players from 5 nationality have the same mean of stamina when playing football as we all know stamina is one of the crucial physical fitness attributes for football players. The sample size for this test is 51 players.

3.0 DATA ANALYSIS AND RESULT

3.1 Hypothesis testing with 1 sample

We will test whether the mean for the potential of FIFA 19 players from Egypt is equal to the population mean, μ_0 for the potential of FIFA 19 players with variance unknown. With a sample size of 31, the test statistic formula is $z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$.

**Mean—one population (n >30)

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

```
> setwd("D:/Software/2020-2021/Sems 2/Probability & Statistical Data Analysis/Project2")
> library(readxl)
> data <- read_excel("data.xlsx")
> View(data)
> library(readxl)
> PDATA <- read_xlsx("data.xlsx", sheet = "data", range = "I1:I18208")
> HDATA <- read_xlsx("data.xlsx", sheet = "H", range = "A1:A32")
> #population data
> p = PDATA$Potential
> x = HDATA$Potential
> DATA = data.frame(potential=c(x))
> DATA2= data.frame(potential=c(p))
> n=nrow(DATA)
> sd=sd(x)
> sd
[1] 5.363757
> # population mean
> mu=round(mean(p), digits = 4)
> # sample mean
> meanX=round(mean(x), digits = 4)
> #Level of significance
> alpha=0.05/2
> # Z statistics (n>30)
> z = (meanX-mu)/(sd/sqrt(n))
> # critical value
> z.alpha = round(qt(1-alpha, df=n-1),digits = 4)
> # Display the c.v. on both sides
> c(-z.alpha,z.alpha)
[1] -2.0423  2.0423
> # p-value
> pval=round((2*pt(z,df=n-1,lower.tail = FALSE)),digits = 4)
> # One-sample t-test
> res =t.test(x, mu=71.3073, alternative = "two.sided")
> res$statistic
       t
1.388746
> res$parameter
df
30
> res$p.value
[1] 0.175135
> res$conf.int
[1] 70.67772 74.61260
attr(,"conf.level")
[1] 0.95
> res$estimate
mean of x
 72.64516
```

Figure 3.1.1: The results of hypothesis testing and test statistics using R Studio.

The significance level for this testing is 0.05. This is a two-tailed test, the critical value is ±2.0423. The mean of FIFA 19 Egyptian players is 72.64516 and $\mu_o = 71.3073$. The standard deviation value is 5.363757. Since P-value > α (0.1751 > 0.05), thus we fail to reject H_0. Hence, we can conclude that the mean for potential of FIFA 19 Egyptian players is not equal to population mean for the potential of FIFA19 players.

3.2 CORRELATION

In correlation, we measure the strength of the relationship between the stamina and the strength of FIFA19 players in Egypt. The sample size of this dataset is 31. Pearson's product-moment correlation coefficient is used in this test because the two variables is ratio level.

```
> Data
   Nationality  x  y   xy   x2   y2
1        Egypt 84 70 5880 7056 4900
2        Egypt 78 67 5226 6084 4489
3        Egypt 92 68 6256 8464 4624
4        Egypt 73 85 6205 5329 7225
5        Egypt 54 87 4698 2916 7569
6        Egypt 74 69 5106 5476 4761
7        Egypt 66 67 4422 4356 4489
8        Egypt 86 72 6192 7396 5184
9        Egypt 61 42 2562 3721 1764
10       Egypt 76 86 6536 5776 7396
11       Egypt 62 80 4960 3844 6400
12       Egypt 71 71 5041 5041 5041
13       Egypt 68 56 3808 4624 3136
14       Egypt 70 69 4830 4900 4761
15       Egypt 60 70 4200 3600 4900
16       Egypt 73 77 5621 5329 5929
17       Egypt 59 64 3776 3481 4096
18       Egypt 23 64 1472  529 4096
19       Egypt 66 78 5148 4356 6084
20       Egypt 73 74 5402 5329 5476
21       Egypt 68 69 4692 4624 4761
22       Egypt 74 87 6438 5476 7569
23       Egypt 60 65 3900 3600 4225
24       Egypt 36 64 2304 1296 4096
25       Egypt 69 56 3864 4761 3136
26       Egypt 37 47 1739 1369 2209
27       Egypt 68 61 4148 4624 3721
28       Egypt 69 67 4623 4761 4489
29       Egypt 60 75 4500 3600 5625
30       Egypt 85 84 7140 7225 7056
31       Egypt 45 31 1395 2025  961
> Total_Data
  TotalX TotalY TotalXY TotalX2 Totaly2
1   2040   2122  142084  140968  150168
> n
[1] 31
```

Figure 3.2.1: Calculation of data from R Studio

Let x = stamina of FIFA 19 players and y = the strength of FIFA 19 players used to calculate the Pearson Correlation. Based on the Figure 3.2.1, the total of x is 2040 while total of y is 2122. The sum of x multiple with y is 142084. The power two of x and y are 140968 and 150168. The sample size, n is 31 Egypt players. All this data will be used in calculating Pearson Correlation Coefficient.

Formula of Pearson Correlation Coefficient:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[(\sum x^2) - \frac{(\sum x)^2}{n}\right]\left[(\sum y^2) - \frac{(\sum y)^2}{n}\right]}}$$

where:

r = Sample correlation coefficient

n = Sample size

x = Value of the independent variable

y = Value of the dependent variable

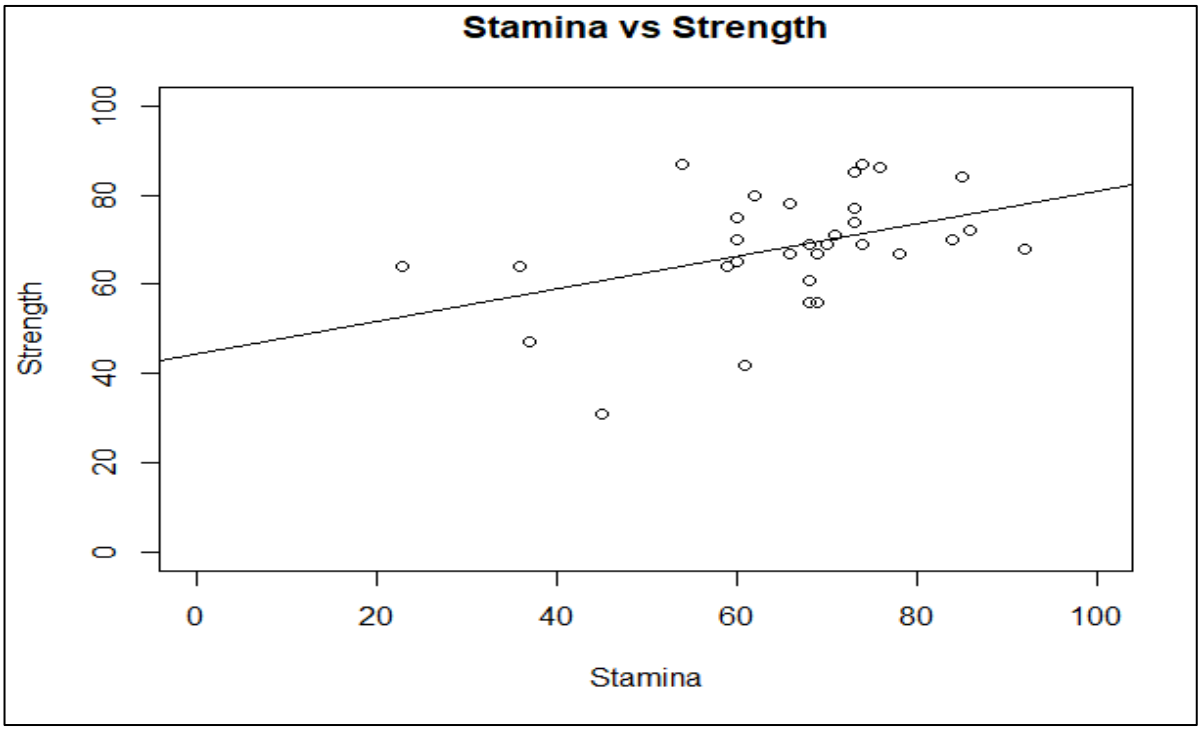

Figure 3.2.2: Scatter plot relationship between the stamina of FIFA19 players and the strength of FIFA19 players from Egypt

```
> r
[1] 0.4250032
> t
[1] 2.528428
> df
[1] 29
> #p-value
> pv
[1] 0.0171574
> #critical value
> cv
[1] 2.04523
> cor.test(x,y)

        Pearson's product-moment correlation

data:  x and y
t = 2.5284, df = 29, p-value = 0.01716
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.08319132 0.67733898
sample estimates:
      cor
0.4250032
```

Figure 3.2.3: Final result for Pearson's product-moment correlation calculated by R Studio

Based on the Figure 3.2.3, r is Pearson's product-moment correlation which calculated by the formula given in the previous pages. The correlation coefficient, r is 0.4250032. The scatter plot in the Figure 3.2.2 and correlation coefficient, r is related. The line in the scatter plot indicates weak positive correlation between the stamina of FIFA 19 complete player and strength of FIFA 19 complete player in Egypt. R falls between 0.08319132 and 0.67733898 which in the weak positive range. This explains that the strength increases when stamina increases.

Significance Test for correlation

Significance Test for correlation is used to test whether the evidence of a linear relationship between the stamina and strength of FIFA 19 complete player in Egypt.

$H_o: \rho = 0$ (No linear correlation between stamina and strength)

$H_1: \rho \neq 0$ (Linear correlation exists between stamina and strength)

Formula for test statistic,

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Based on the Figure 3.2.3, test statistic, t is 2.5284. Degree of freedom, df is n-2, thus df is 29. The significance level, $\alpha = 0.05$. Since the hypothesis is two-tailed, α has to divided by two and significance level become $\alpha = 0.025$. P-value is the significant level of the test statistic, t which is 0.01716. 95% of the confidence interval for the correlation coefficient is (0.08319132, 0.67733898). The critical value, $t_{29,0.0025}$ is 2.04523.

Conclusion: Since P-value < 0.05 and critical value > 2.04523, we reject H_o. There is sufficient evidence of a linear relationship between the stamina and strength of FIFA 19 complete player in Egypt at the 5% level of significance.

3.3 REGRESSION

Regression analysis is used to analyse the value of ball control based on the value of balance for the FIFA19 players from Egypt. In this analysis, the sample size is 31. The independent variable, x is balance while the dependent variable, y is ball control. A simple linear regression usually to describe the relationship between the independent variable, x and dependent variable, y with a straight line by using only one independent variable, x.

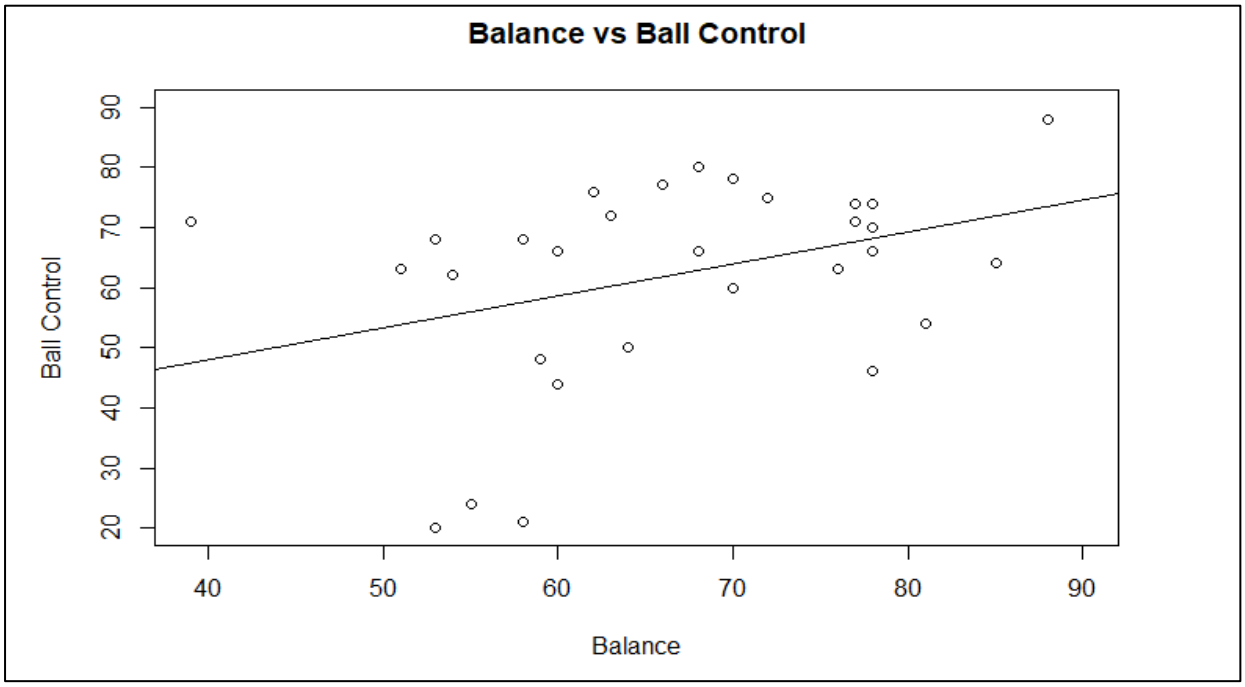

Figure 3.3.1: Scatter plot relationship between the balance and ball control of FIFA 19 complete player in Egypt

```
> Y

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    26.6134       0.5323

> summary(Y)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-36.485  -8.937   3.402  11.184  23.628

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.6134    17.1878   1.548   0.1324
x             0.5323     0.2531   2.103   0.0442 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.99 on 29 degrees of freedom
Multiple R-squared:  0.1324,    Adjusted R-squared:  0.1024
F-statistic: 4.424 on 1 and 29 DF,  p-value: 0.04422
```

Figure 3.3.2: Summary answer of Regression analysis calculated by R Studio

Based on the figure 3.3.1, the scatter plot shows a positive linear regression relationship. This means that the value of the balance increases, the value of the ball control also increases.

```
> k
[1] 1
> n
[1] 31
> MeanY
[1] 62.25806
> MeanX
[1] 66.96774
> beta1
[1] 0.5322664
> beta0
[1] 26.61339
> df
[1] 29
> SSE
[1] 7411.264
> RSE
[1] 15.98627
> SSR
[1] 1130.671
> SST
[1] 8541.935
> SB1
[1] 0.253051
> t
[1] 2.103396
> R2
[1] 0.1323671
> adjR2
[1] 0.1024487
> PE
[1] 23.8716
> f
[1] 4.424274
> pv
[1] 0.04422164
> cv
[1] 2.04523
```

Figure 3.3.3: Calculation of data from R Studio

Estimated Regression Model formula:

$$\hat{y}_i = b_0 + b_1 x$$

where

$\hat{y}$ = Estimated y value

b_0= Estimate of regression intercept

b_1= Estimate of the regression slope

x = Independent variable

The formula for b_1 and $b_0$are

$$b_1 = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \text{ and } b_0 = \bar{y} - b_1 \bar{x}$$

where

x = Independent variable

y = Dependent variable

n = Sample size

$\bar{x}$ = Mean of independent variable

$\bar{y}$ = Mean of dependent variable

From Figure 3.3.2, it shows that the estimated regression model is $\hat{y} = 26.6134 + 0.5323x$. This shows b_0= 26.6132 is estimated value when x is zero and the intercept of the regression slope. b_1= 0.5323 is the value of the slope and the estimated change of the estimated value with one-unit change in x. Hence, the ball control will not have the value of zero because b_0 = 26.6132. 26.6132 is the portion for the value of the ball control which not explained by the balance. b_1= 0.5323 shows that the average value of the ball control increases by 0.5323, on, average, for each additional one-unit.

Total variation formula:

SST = SSE + SSR

where

Total sum of squares, SST = $\Sigma(y - \bar{y})^2$

Error sum of squares, SSE = $\Sigma(y - \hat{y})^2$

Regression sum of squares, SSR = $\Sigma(\hat{y} - \bar{y})^2$

Based on the Figure 3.3.3, SST=8541.935 is the total variation of the regression slope which made up by SSE and SSR. SSE is an unexplained variation which is 7411.264 while SSR is an explained variation which is 1130.671. The sum of unexplained variation and explained variation is equal to total variation.

Model summary

Based on the Figure 3.3.2, the summary of linear regression model consist of 6 component.

1. **Call**

   This is a function call which use to calculate the regression model with "lm(y~x)" code.

2. **Residuals**

   Residuals provide a quick view with a mean which is zero. The median in residuals which is 3.402 that should not be far than zero. The mininum and maximum of residuals are -36.485 and 23.628 which are almost equal in absolute value.

3. **Coefficients**

   In the table, it shows the estimates of the beta coefficients which is the value in $\hat{y} = 26.6134 + 0.5323x$. 26.6134 is the intercept of the regression slope while 0.5323 is estimated change of the estimated value with one-unit change in x.

Standard Error of Estimate formula:

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}}$$

where
SSE = Sum of squares error
n = Sample size
k = Number of independent variables in the model

Standard error formula:

$$s_{b1} = \frac{s_\varepsilon}{\Sigma(x-\bar{x})^2}$$

Standard Error of the Estimate is used to calculate the standard error. Standard error is used to determine the accuracy of beta coefficients. 0.2351 of standard error means that the observed values is closer to the regression slope. T-statistic(or value) and p-value will be explained in significance test for regression.

4. **Residual standard error (RSE)**

   RSE can be calculated by using Standard Error of Estimate formula. RSE is used to define how the linear regression fit to the data with degrees of freedom.

   Percentage error formula:

$$Percentage\ error = \frac{RSE}{\bar{x}}$$

   15.99 is the approximate value for the actual distance deviate from the true regression line with 29 degrees of freedom. The percentage error of the linear regression is almost 23.87%.

5. **R-squared**($\boldsymbol{R^2}$)

   Coefficient of Determination , $R^2$and Adjusted R^2formula:

$R^2 = \frac{SSR}{SST}$ and $Adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{(n-k-1)}$

The range of R^2 is between 0 and 1 which means that the portion of variation in the dependent variable is explained by the variation in the independent variable. R^2in the data is 0.1324 which say that it is not a good indication. The adjusted R^2 is 0.1024 is closer to zero because the balance is not explain much on the ball control.

6. **F-statistic**

   F-statistic is the duplicate value of test statistic which shows the overall significance of the simple linear regression model. F-statistic = 4.424 which produce p-value = 0.04422. This shows that the F-statistic is highly significant because p-value < 0.05.

Significance Test for regression

T-test is used to determine the linear relationship between balance and ball control for FIFA 19 players from Egypt.

H_o: $\beta_1 = 0$ (No linear relationship between balance and ball control)

H_1: $\beta_1 \neq 0$ (Linear relationship does exist between balance and ball control)

Test statistic formula for regression:

$$t = \frac{b_1 - \beta_1}{S_{b1}}$$

Based on the Figure 3.3.2, test statistic, t is 2.103. Degree of freedom, df is n-2, thus df is 29. The significance level, $\alpha = 0.05$. Due to the hypothesis is two-tailed, α has to be divided by two. Thus, significance level become $\alpha = 0.025$. P-value is the significant level of the test statistic, t whizh is 0.0442. The critical value, $t_{29,0.0025}$ is 2.04523.

Conclusion: Since P-value < 0.05 and critical value > 2.04523, we reject H_o. There is sufficient evidence of a linear relationship exist between the balance and ball control of FIFA 19 players in Egypt at the 5% level of significance.

3.4 CHI-SQUARE TEST OF INDEPENDENCE

The chi-square test of independence is used to determine whether 3 different nationality of FIFA 19 players (Egypt, England, Guinea) are independent with preferred foot when playing football. The sample size for this test is 92.

H_0 : FIFA 19 players from 3 different nationality are independent with preferred foot
H_1 : FIFA 19 players from 3 different nationality are not independent with preferred foot

```
> library(MASS)
> # get the contingency table
> tbl <- table(tb1$Nationality, tb1$`Preferred Foot`)
> tbl

         Left Right
  Egypt      7    24
  England    5    24
  Guinea     6    25
> # perform chi-square test on the data table
> chisq.test(tbl)

        Pearson's Chi-squared test

data:  tbl
X-squared = 0.27455, df = 2, p-value = 0.8717

> #critical value
> alpha <- 0.05
> x2.alpha <- qchisq(alpha, df=2, lower.tail=FALSE)
>
> output <- chisq.test(tbl)
> x2.alpha
[1] 5.991465
> output$statistic
X-squared
0.2745516
> output$parameter
df
 2
> output$observed

         Left Right
  Egypt      7    24
  England    5    24
  Guinea     6    25
> output$expected

             Left    Right
  Egypt   6.131868 24.86813
  England 5.736264 23.26374
  Guinea  6.131868 24.86813
```

Figure 3.4.1: Calculation of chi-square test of independence using R Studio

The formula of chi-square test statistic is:

$$x^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

Based on the Figure 3.4.1, we can observe that with degree of freedom, df = 2 and the significance level, $\alpha = 0.05$ (critical value = 5.991465), the chi-square test statistic obtained

is 0.2745516. Since test statistic < critical value (0.2745516 < 5.991465), thus fail to reject H_0. Therefore, we can conclude that there is sufficient evidence to claim that FIFA 19 players from 3 different nationality (Egypt, England, Guinea) is independent with preferred foot.

3.5 ANOVA

The objective of ANOVA test is to determine the significance differences between means of stamina among 5 different nationality of FIFA 19 players which are Guinea, Egypt, England, Belgium and France at significance level, $\alpha = 0.05$. The sample size for this test is 51.

H_0 : μ1 = μ2 = μ3 = μ4 = μ5 (All FIFA 19 players from 5 different nationality (Guinea, Egypt, England, Belgium, France) have the same mean of stamina)

H_1 : At least one FIFA 19 player among 5 different nationality has different mean of stamina

$$F = \frac{variance\ between\ samples}{variance\ within\ samples}$$

```
>
> #read in xlsx tables
> my_data<-read_excel("data.xlsx",sheet="ANOVA",range = "A1:B51")
>
> library(readxl)
> #compute the analysis of variance
> res.aov<-aov(Stamina~Nationality, data = my_data)
>
> #summarise the analysis of variance model
> summary(res.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
Nationality   4   2569   642.3   2.693 0.0428 *
Residuals    45  10734   238.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure 3.5.1: Calculation of ANOVA test using R Studio

Based on Figure 3.5.1, the numerator, k-1 =5-1 =4 and denominator, k(n-1) =5(10-1) =45. The test statistic, F obtained is 2.693. From the table we can also observe that p-value is 0.0428, which is less than the significance level, $\alpha = 0.05$ (0.0428 < 0.05). Thus, we reject H_0. There is insufficient evidence to conclude that all FIFA 19 players from 5 different nationality (Guinea, Egypt, England, Belgium, France) have the same mean of stamina.

4.0 DISCUSSION AND CONCLUSION

Egypt and Guinea are the samples that we have chosen from many countries that involved in FIFA 19, because both nationalities have a sample size of 31 and if seen spontaneously (not from a statistical point of view) Egypt and Guinea have very little difference, and we ended up researching more so that we can find an accurate count for the two data. In addition, we also selected England, Belgium, and France, for accurate comparisons. Because these three nationalities are known to excel in the field of football.

We have learned that several factors must be considered when selecting a data set for each test such as deciding the purpose of our study and sample size. For example, if we want to do a z-test to test the hypothesis of 1 sample, our sample size should be more than 30, thus we can practice directly what is the difference between t-test and z-test. In addition, we have understood the theory which says that the variables used in the chi-squared test of independence must be two nominal variables.

For the analysis process, we have understood and put into practice how to use R Studio to calculate each test, which is a new experience for us as it we have to meet the project requirements. Before writing code with R-studio, we need to study the code explanation and learn how to analyse the output. It is indeed difficult to understand and learn a new program, but after several experiments, conducting discussions, and re-reading the instructions given by the lecturer, we were finally able to understand the output produced by R-studio.

Our curiosity is getting stronger to analyse the data and find the results. So, we tried R-studio on 1 sample Hypothesis testing, Correlation, Regression, Chi-Square Independence Test, ANOVA. And for the regression we get some summary of the model. We keep on testing the code and find the results in R-studio. Our group worked well together, apart from dividing the work, we also held discussions and share knowledge.

One of the very interesting finding is from the ANOVA test. From the test, we get to realise that at least one of the FIFA 19 players among 5 different nationalities (Egypt, Guinea, England, Belgium, and France) have different stamina. This is very interesting because majority of the football players from these country is excellent and we always thought they have the same mean of stamina. In our opinion, we think that the stamina of FIFA 19 players varies and is not based on their nationality and culture.

Summary of conclusions with a significance level of 5%:

i. The potential of FIFA 19 Egyptian players is not equal to population mean for the potential of FIFA19 players.
ii. Correlational, the strength of the Egyptian players increases with the increase in stamina.
iii. There is a linear relationship between stamina and strength of complete FIFA 19 players in Egypt.
iv. There is a linear relationship between balance and ball control of FIFA 19 players in Egypt.
v. There is sufficient evidence of a linear relationship exists between the balance and ball control of FIFA 19 players in Egypt.

vi. There is sufficient evidence to claim that FIFA 19 players from 3 different nationality (Egypt, England, Guinea) is independent with preferred foot.

vii. There is insufficient evidence to conclude that all FIFA 19 players from 5 different nationalities (Guinea, Egypt, England, Belgium, France) have the same mean of stamina.

5.0 REFERENCES

1. Felipe Rego. Quick guide: Interpreting simple linear model output in R. Retrieved 23 October 2015, from https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R
2. Kassambara. Articles – Regression Analysis. Retrieved from 10 March 2018, from http://www.sthda.com/english/articles/40-regression-analysis/167-simple-linear-regression-in-r/

6.0 APPENDIX

1. Karan Gadiya. FIFA 19 complete player dataset. https://www.kaggle.com/karangadiya/fifa19

7.0 Video presentation link & E-portfolio Link

Video presentation link

https://youtu.be/Den1CKvL_bk

E-portfolio Link

1. Eunice Lim Xian Ni: https://eportfolio.utm.my/user/eunice-lim-xian-ni/seci2143-probability-statistical-data-analysis
2. Shabrina Salsabila Sakroni: https://eportfolio.utm.my/user/shabrina-salsabila-sakroni/seci2143-probability-statistical-data-analysis
3. Teh Jing Ling: https://eportfolio.utm.my/user/jacklyn-jlt/fifa-19-data-analysis