Final Report

December 2022

Research Project: **Determining the Optimal Distance a Football Can Travel**

Prepared by: Dina Dinh

Research Goal

The experiment evaluates two major research questions:

1. What material of the football and method of causing the football to travel will lead to the football traveling the farthest distance?

2. Are the method of causing the football to travel and the material of the football significantly different?

Dataset Information

There are 12 data points in the data set. The input variable material of football and method of traveling are 2 levels and nominal. The output variable is continuous because the football can travel any distance within any range. The larger output values, or response values, are better because we want the football material and method that makes the football travel the farthest.

Data Exploration

A table was used to represent the dataset, and from Table 1, a test statistic is used to reject or not reject the null hypothesis. The null and alternative hypotheses for the experiment are the following:

Part 1: T-Test

$H_0$1: The average distance a person can cause a football to travel is 150 ft.

H_a1: The average distance a person can cause a football to travel is not 150 ft.

Part 2: ANOVA

$H_0$2: The average maximum distance the football can travel is not affected by the method of making the football travel or the material of the football.

H_a2: The average maximum distance the football can travel is affected by the method of making the football travel or the material of the football.

When using the normal probability plot, it can be seen, in Figure 1, that the output variable comes from a normal distribution.

Model and Method

Data for this experiment was analyzed using RStudio. For Part 1, a t-test shown in Code 1 of the appendix was used to reject the null hypothesis. Since the p-value is 0, the null hypothesis was rejected and the true average a person can make a football travel is most likely not 150 ft. For Part 2, a two-way ANOVA model shown in Code 2 and Code 3 was used to compare the two input variables and analyze the data.

Main Results

By running the ANOVA analysis, the results are shown in Code 2 and 3. The results of the Tukey method are shown in Code 4. Using the adjusted p-values in the Tukey method and comparing it to the alpha value 0.05, it can be concluded that the effects of the control variable are not different since both p-values are greater than 0.05. This analysis helps with Part 2 and shows the null hypothesis is not rejected.

Model Diagnostic

The diagnostic procedures can be seen in Code 5.

Model Comparison/Selection

A regression model was used on the dataset via function lm() in RStudio. From lm() shown in Code 6, the p-values for ControlVariable1 is 0.0636 and for ControlVariable2 is 0.2473, which means ControlVariable1 has a 6.36% chance of being not meaningful and 24.73% chance for ControlVariable2. As a result, these variables are most likely not useful for the model. Also, the R^2 values are low so the model can only explain little of the data variability.

Summary

To conclude, the null hypothesis was rejected in Part 1 due to the t-test, but the null hypothesis in Part 2 was not rejected based on the Tukey method. However, using a regression model proved that the variables used in the experiment were not meaningful to the model. Therefore, further adjustments to the experiment are needed for improvement.

EXST 7004 Final Report: Appendix

Dina Dinh

2022-12-10

Table 1:

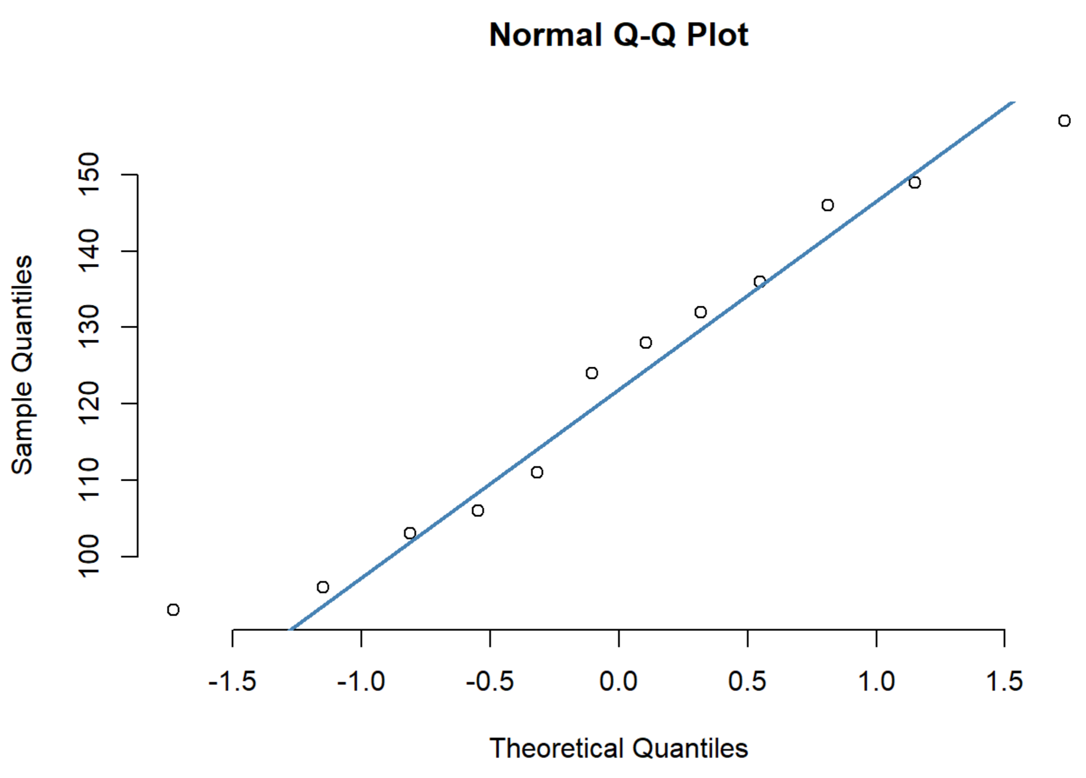
```
ControlVariable1=c("1","1","2","2","1","1","2","2","1","1","2","2")
ControlVariable2=c("1","2","1","2","1","2","1","2","1","2","1","2")
Outcome=c(124,136,157,96,93,103,146,128,111,106,149,132)

df<- data.frame(ControlVariable1, ControlVariable2, Outcome)

write.table(df, file='C:\\Users\\ddinh\\OneDrive\\Documents\\EXST 7004\\data.txt')
df
```

```
##    ControlVariable1 ControlVariable2 Outcome
## 1                 1                1     124
## 2                 1                2     136
## 3                 2                1     157
## 4                 2                2      96
## 5                 1                1      93
## 6                 1                2     103
## 7                 2                1     146
## 8                 2                2     128
## 9                 1                1     111
## 10                1                2     106
## 11                2                1     149
## 12                2                2     132
```

Figure 1:

Showing output variable is from normal distribution

```
qqnorm(Outcome, pch = 1, frame = FALSE)
qqline(Outcome, col = "steelblue", lwd = 2)
```



Code 1:

```
t_obs <-(mean(Outcome) - 1)/(sqrt((sd(Outcome)^2)/length(Outcome)))
p_value<-2*pt(t_obs,df=(length(Outcome)-1),lower.tail=FALSE)
p_value
```

```
## [1] 6.283519e-10
```

Code 2:

```
model.2.1 <- aov(Outcome ~ControlVariable1 + ControlVariable2)
summary(model.2.1,type=T)
```

```
##                  Df Sum Sq Mean Sq F value Pr(>F)  
## ControlVariable1  1 1518.7  1518.7   4.470 0.0636 .
## ControlVariable2  1  520.1   520.1   1.531 0.2473  
## Residuals         9 3058.1   339.8                 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Code 3:

```
model.2.2 <- aov(Outcome~ControlVariable1 + ControlVariable2 + ControlVariable1*ControlVariable2)
summary(model.2.2)
```

```
##                                  Df Sum Sq Mean Sq F value Pr(>F)  
## ControlVariable1                  1 1518.7  1518.7   6.093 0.0388 *
## ControlVariable2                  1  520.1   520.1   2.087 0.1866  
## ControlVariable1:ControlVariable2 1 1064.1  1064.1   4.269 0.0727 .
## Residuals                         8 1994.0   249.3                 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

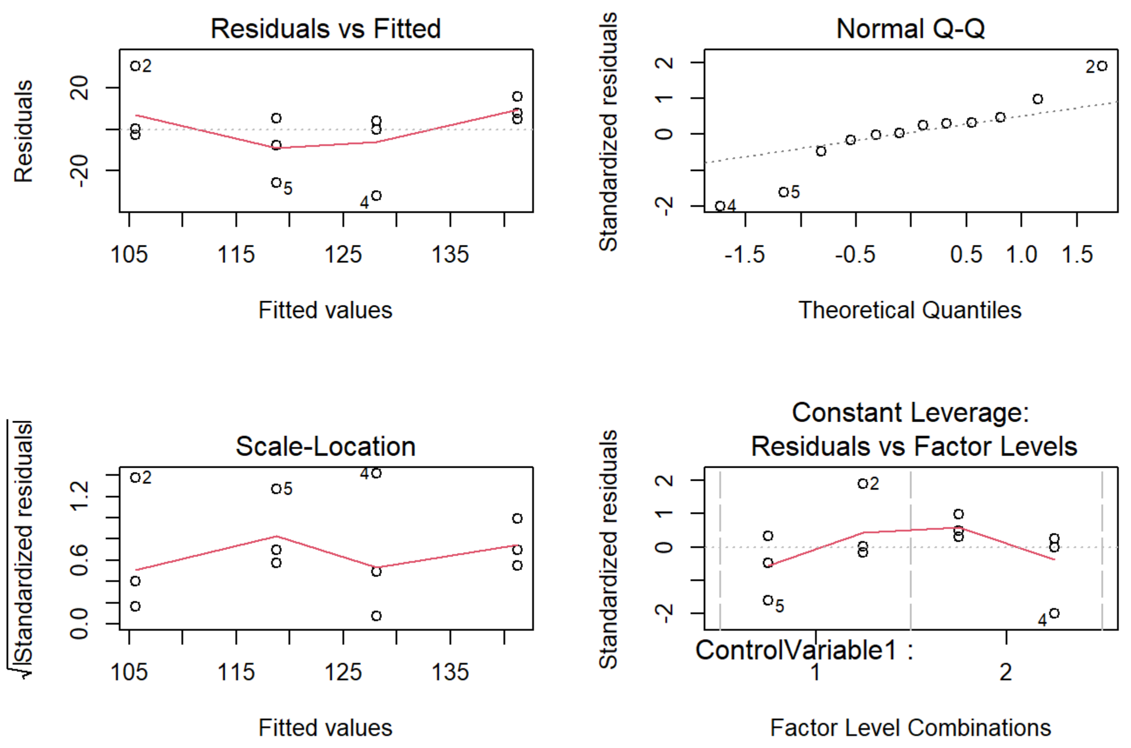
Code 4:

```
TukeyHSD(model.2.1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Outcome ~ ControlVariable1 + ControlVariable2)
##
## $ControlVariable1
##     diff       lwr      upr     p adj
## 2-1 22.5 -1.574959 46.57496 0.0636375
##
## $ControlVariable2
##          diff       lwr      upr     p adj
## 2-1 -13.16667 -37.24163 10.90829 0.2473266
```

Code 5:

```
par(mfrow=c(2,2))
plot(model.2.1)
```



Code 6:

```
lm.model<-lm (Outcome~ControlVariable1 + ControlVariable2,data = df)
summary (lm.model)
```

```
##
## Call:
## lm(formula = Outcome ~ ControlVariable1 + ControlVariable2, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.083  -3.875   2.167   5.875  30.417
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)    
## (Intercept)        118.750      9.217  12.884 4.19e-07 ***
## ControlVariable12   22.500     10.642   2.114   0.0636 .  
## ControlVariable22  -13.167     10.642  -1.237   0.2473    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.43 on 9 degrees of freedom
## Multiple R-squared:    0.4,  Adjusted R-squared:  0.2667
## F-statistic:     3 on 2 and 9 DF,  p-value: 0.1004
```