Midterm Exam

Xijia Luo

11/2/2020

Instruction

This is your midterm exam that you are expected to work on it alone. You may NOT discuss any of the content of your exam with anyone except your instructor. This includes text, chat, email and other online forums. We expect you to respect and follow the GRS Academic and Professional Conduct Code.

Although you may NOT ask anyone directly, you are allowed to use external resources such as R codes on the Internet. If you do use someone's code, please make sure you clearly cite the origin of the code.

When you finish, please compile and submit the PDF file and the link to the GitHub repository that contains the entire analysis.

Introduction

In this exam, you will act as both the client and the consultant for the data that you collected in the data collection exercise (20pts). Please note that you are not allowed to change the data. The goal of this exam is to demonstrate your ability to perform the statistical analysis that you learned in this class so far. It is important to note that significance of the analysis is not the main goal of this exam but the focus is on the appropriateness of your approaches.

Project and Data Description (10pts)

Please explain what your data is about and what the comparison of interest is. In the process, please make sure to demonstrate that you can load your data properly into R.

PROJECT DESCRIPTION

Recently, a group of middle school students are interested in making a better paper helicopter. The way to judge a paper helicopter is good or bad is the length it can stay in the air. Our task is to help them design the best paper helicopter.

Data Description

The data was collected by ourselves. We have made 16 paper helicopters with different parameters including paper type, rotor length, leg length, leg width, and paper clip on leg. Detailed description of these parameters can be found in 1.3 - VARIABLES parts and 7.1 Assembly Instructions parts. We chose a wide and closed interior to test fly these 16 paper helicopters one by one in order to ensure the final experimental results will not be disturbed by external factors.

EDA (10pts)

Please create one (maybe two) figure(s) that highlights the contrast of interest. Make sure you think ahead and match your figure with the analysis. For example, if your model requires you to take a log, make sure you take log in the figure as well.

Our EDA includes a table summarizing the response variables under the explanatory variables we are interested in, and various plots explaining the relationship between our explanatory variables and our response variables.

```
#Paper Type
data%>%group_by(Paper.Type)%>%summarize(mean=mean(Flight.Time),max=max(Flight.Time),min=min(Flight.Time
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 5
##   Paper.Type  mean   max   min    sd
##   <fct>      <dbl> <dbl> <dbl> <dbl>
## 1 Heavy       1.57  1.78  1.18 0.149
## 2 Light       1.89  2.32  1.52 0.241
```

```
#Rotor Length
data%>%group_by(Rotor.Length)%>%summarize(mean=mean(Flight.Time),max=max(Flight.Time),min=min(Flight.Ti
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 5
##   Rotor.Length  mean   max   min    sd
##          <int> <dbl> <dbl> <dbl> <dbl>
## 1            6  1.65  2.13  1.18 0.264
## 2            8  1.82  2.32  1.56 0.220
```

```
#Leg Length
data%>%group_by(Leg.Length)%>%summarize(mean=mean(Flight.Time),max=max(Flight.Time),min=min(Flight.Time
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 5
##   Leg.Length  mean   max   min    sd
##        <int> <dbl> <dbl> <dbl> <dbl>
## 1          8  1.86  2.32  1.56 0.251
## 2         12  1.61  1.92  1.18 0.196
```

```
#Leg Width
data%>%group_by(Leg.Width)%>%summarize(mean=mean(Flight.Time),max=max(Flight.Time),min=min(Flight.Time)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 5
##   Leg.Width  mean   max   min    sd
##       <int> <dbl> <dbl> <dbl> <dbl>
## 1         3  1.71  2.32  1.18 0.306
## 2         5  1.76  2.13  1.52 0.195
```

```
#Paper Clip on Leg
data%>%group_by(Paper.Clip.On)%>%summarize(mean=mean(Flight.Time),max=max(Flight.Time),min=min(Flight.T
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 5
##   Paper.Clip.On  mean   max   min    sd
##   <fct>         <dbl> <dbl> <dbl> <dbl>
## 1 NO             1.79  2.32  1.22 0.273
## 2 Yes            1.68  1.99  1.18 0.228
```

A preliminary review of the data reveals interesting dynamics in the relationships between the different groups. It appears, at least from the dataset, that paper type, rotor length, leg length, and paper clip on are significant influence factors to the flight time.

Power Analysis (10pts)

Please perform power analysis on the project. Use 80% power, the sample size you used and infer the level of effect size you will be able to detect. Discuss whether your sample size was enough for the problem at hand. Please note that method of power analysis should match the analysis. Also, please clearly state why you should NOT use the effect size from the fitted model.

```r
#Paper Type
d1=(abs(1.57-1.89)/sqrt(0.148^2+0.241^2))
pwr.t.test(power=0.8,d=d1,sig.level=0.05,type='two.sample')
```

```
##
##      Two-sample t test power calculation
##
##               n = 13.29394
##               d = 1.131477
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#Rotor Length
d2=(abs(1.65-1.82)/sqrt(0.263^2+0.22^2))
pwr.t.test(power=0.8,d=d2,sig.level=0.05,type='two.sample')
```

```
##
##      Two-sample t test power calculation
##
##               n = 64.83488
##               d = 0.4957954
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#Leg Length
d3=(abs(1.86-1.61)/sqrt(0.251^2+0.196^2))
pwr.t.test(power=0.8,d=d3,sig.level=0.05,type='two.sample')
```

```
##
##      Two-sample t test power calculation
##
##               n = 26.4678
##               d = 0.7850271
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```r
#Leg Width
d4=(abs(1.708-1.757)/sqrt(0.306^2+0.195^2))
pwr.t.test(power=0.8,d=d4,sig.level=0.05,type='two.sample')
```

```
##
##      Two-sample t test power calculation
##
##               n = 861.7601
##               d = 0.1350416
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

```
#Paper Clip on Leg
d5=(abs(1.78-1.68)/sqrt(0.272^2+0.228^2))
pwr.t.test(power=0.8,d=d5,sig.level=0.05,type='two.sample')
```

```
##
##      Two-sample t test power calculation
##
##               n = 198.706
##               d = 0.2817539
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Since we got 80 sample in total, for paper type and leg length we got enough samples. But for the others, I am afraid that we dont have enough samples to satisfied the 80% power. I should NOT use the effect size from the fitted model mainly becasue the requested sample is too large(861*2=1522), everyone wants a smaller sample so that could lower the experiment's opportunity cost.

Modeling (10pts)

Please pick a regression model that best fits your data and fit your model. Please make sure you describe why you decide to choose the model. Also, if you are using GLM, make sure you explain your choice of link function as well.

I will choose the category regression

```
data$Paper.Type<-as.character(data$Paper.Type)
data$Rotor.Length<-as.character(data$Rotor.Length)
data$Leg.Length<-as.character(data$Leg.Length)
data$Leg.Width<-as.character(data$Leg.Width)
data$Paper.Clip.On<-as.character(data$Paper.Clip.On)
model <- lm(Flight.Time ~ Paper.Type+Rotor.Length+Leg.Length+Leg.Width+Paper.Clip.On, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = Flight.Time ~ Paper.Type + Rotor.Length + Leg.Length +
##     Leg.Width + Paper.Clip.On, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.178666 -0.092024  0.005941  0.107976  0.177952
##
## Coefficients:
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.39572    0.03056  45.672  < 2e-16 ***
## Paper.TypeLight   0.32294    0.02546  12.686  < 2e-16 ***
## Rotor.Length8     0.17244    0.02546   6.774 2.59e-09 ***
## Leg.Length8       0.25094    0.02546   9.858 4.07e-15 ***
## Leg.Width5        0.05244    0.02546   2.060   0.0429 *
## Paper.Clip.OnYes -0.12766    0.02549  -5.008 3.62e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1138 on 74 degrees of freedom
## Multiple R-squared:  0.8149, Adjusted R-squared:  0.8024
## F-statistic: 65.14 on 5 and 74 DF,  p-value: < 2.2e-16
```

Validation (10pts)

Please perform a necessary validation and argue why your choice of the model is appropriate.

It is the most appropriate model because all of my independent variables are categorical since these helicopters are created by myself, it is impossible to make these variable continuesly. What's more, as the p-value = 2.2E-16 which is much less than 0.05, we reject the null hypothesis. Hence there is a significant relationship between these variables in the linear regression model of the data set faithful.

Inference (10pts)

Based on the result so far please perform statistical inference to compare the comparison of interest.

Compared to Paper Type that were used in paper helicopter, we would expect the lighter paper will cause the fight time increase 0.322 second compared to the heavy one, on average, at the same level of other factors.

Compared to Rotor Length that were used in paper helicopter, we would expect the 8cm rotor will cause the fight time increase 0.172 second compared to the 6cm one, on average, at the same level of other factors.

Compared to Leg Length that were used in paper helicopter, we would expect the 8cm leg will cause the fight time increase 0.251 second compared to 12cm leg, on average, at the same level of other factors.

Compared to Leg Width that were used in paper helicopter, we would expect the 5cm leg width paper will cause the fight time increase 0.052 second compared to 3cm leg width, on average, at the same level of other factors.

Compared to whether PaperClip were used in paper helicopter, we would expect that paperclip will cause the fight time derease 0.128 second compared to those without a paperclip, on average, at the same level of other factors.

Discussion (10pts)

Please clearly state your conclusion and the implication of the result.

Since we have proved the significance of the result and based on the inference part above, we can conclude that:

To get the best mean flight time, we need to build a paper helicopter with light paper type, 8 cm rotor length, 8 cm leg length, 5 cm leg width, and no paper clip on. This is not considering the interaction effect, which means that looking at these options individually, those are the better options in average.

To expand this conclusion, we can speculate that lighter paper, longer rotor length, shorter leg length, longer leg width and no paper clip on the bottom of the helicopter are better choice if we want this paper helicopter to stay in the air longer.

Limitations and future opportunity. (10pts)

Please list concerns about your analysis. Also, please state how you might go about fixing the problem in your future study.

There are several additional considerations to help to fully understand this study. Firstly, due to the paper size and type limit, we can't find a larger or more lighter paper, thus we can't be 100% sure the experiment is accurate. For example, maybe too light or too heavy is both not a good choice.

Secondly, the way we calculate time may be not accurate, our reaction time and timer and the tacit understanding between the experimenter and the timekeeper may lead to errors.

Finally, we can't not guarantee that each paper helicopter made by students are all the same. For example, they may have a better design for how to make a paper helicopter. All the parameters may be change and the final result will be much different.

In order to improve my project, I have a good idea. I will try to hold a paper helicopter competition to collect more datas. Since everyone is trying to win the reward, each of them will try different helicopters in lots of different parameters. Therefore, I can make my independent variable into continuous ones.(rotor length, leg length, leg width) What's more, I have an estimate of the best outcome based on the champion's helicopter.

For the model improvement, I think I need to add interactions between independent variables. That can make the model more accurate and convincing.

Comments or questions

If you have any comments or questions, please write them here.