Take Home Midterm 2

Sumeet Mishra

3/26/2019

```
handwashing=read.table("handwashing.txt",header=T)
#View(handwashing)
```

1.

(a) What are the experimental units in this study? What measurements are taken on each unit? How many independent samples are there?

The experimental units are students from Michigan State. The measurements are the 6 moral judgement questions that each student is asked to give their responses to that is Dog, Trolley, Wallet, Plane, Resume and Kitten. There are 2 independent samples one is the handwashing sample group and the other is the control group(no handwashing group).

(b) Perform a one-tailed significance test to study whether handwashing would lower the average answer to the trolley question. Carefully define hypotheses, calculate a P-value, and write a substantive conclusion.

$H_0 : \mu - \mu_1 >= 0$
$H_1 : \mu - \mu_1 < 0$

where
μ= average answer of students from handwashing group
μ_1= average answer of students from control group

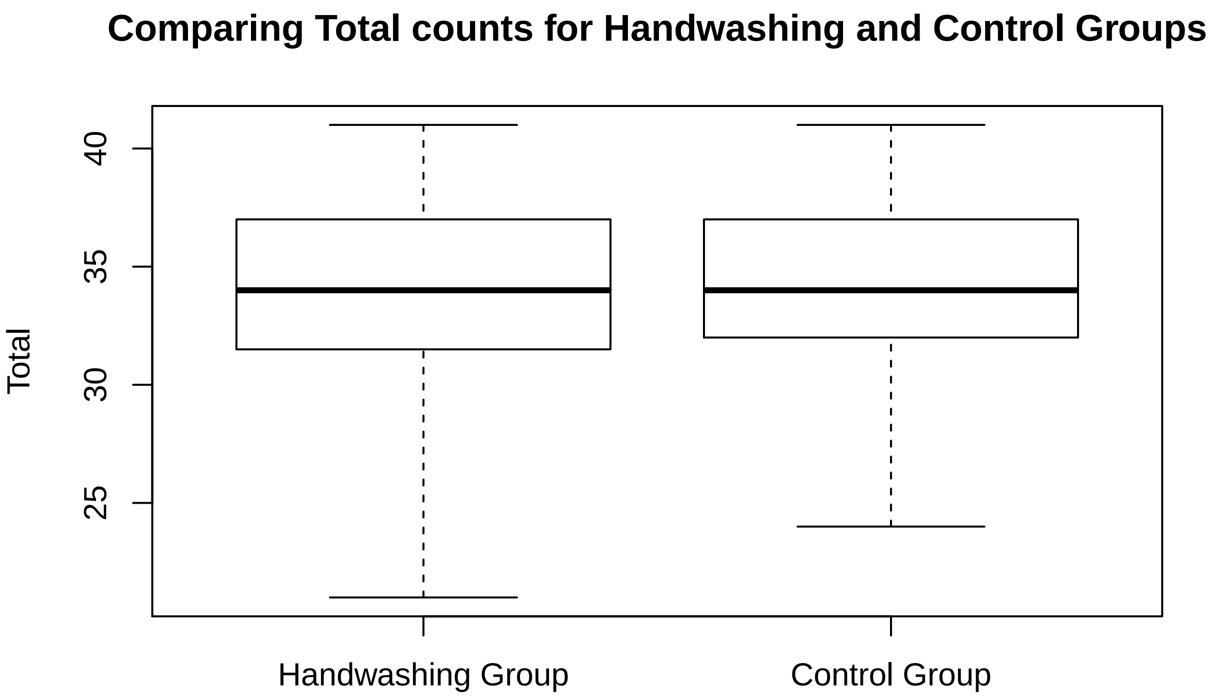
```
hw=subset(handwashing,Condition==1)
#View(hw)
nhw=subset(handwashing,Condition==0)
#View(nhw)
x_hw=mean(hw$Trolley)
x_nhw=mean(nhw$Trolley)
s_hw=sd(hw$Trolley)
s_nhw=sd(nhw$Trolley)
n_hw=length(hw$Trolley)
n_nhw=length(nhw$Trolley)

test_stat=(x_hw-x_nhw)/sqrt((s_hw^2/n_hw)+(s_nhw^2/n_nhw))
df=((s_hw^2/n_hw)+(s_nhw^2/n_nhw))^2/(((s_hw^2/n_hw)^2/(n_hw-1))+((s_nhw^2/n_nhw)^2/(n_nhw-1)))
p_value=pt(test_stat,df)
p_value
```

```
## [1] 0.7181691
```

The p_value is calculated as 0.7181691 which is greater than α (0.05). Therefore we failed to reject null hypothesis and conclude that handwashing would not lower the average answer to the trolley question.

(c) Create a variable called Total that gives the total score on the six moral judgment questions for each individual. (Remember to give R code.) Draw labeled side-by-side boxplots (on the same plot) of Total for the handwashing and control groups.

```
handwashing$Total=(handwashing$Dog+handwashing$Trolley+
                     handwashing$Wallet+handwashing$Plane+handwashing$Resume+handwashing$Kitten)
total_hw=subset(handwashing,Condition==1)
#View(total_hw)
total_nhw=subset(handwashing,Condition==0)
#View(total_nhw)
boxplot(total_hw$Total,total_nhw$Total,range=0,
        main="Comparing Total counts for Handwashing and Control Groups",
        ylab="Total",names=c('Handwashing Group','Control Group'))
```



(d) Perform a one-tailed significance test to study whether handwashing would lower the average total score for the morality questions.Carefully define hypotheses,calculate a P-value, and write a substantive conclusion.

```
x_total_hw=mean(total_hw$Total)
x_total_nhw=mean(total_nhw$Total)
s_total_hw=sd(total_hw$Total)
s_total_nhw=sd(total_nhw$Total)
n_total_hw=length(total_hw$Total)
n_total_nhw=length(total_nhw$Total)

test_total_stat=(x_total_hw-x_total_nhw)/sqrt((s_total_hw^2/n_total_hw)+
                                                (s_total_nhw^2/n_total_nhw))
df_total=((s_total_hw^2/n_total_hw)+(s_total_nhw^2/n_total_nhw))^2/
```

```r
  (((s_total_hw^2/n_total_hw)^2/(n_total_hw-1))+((s_total_nhw^2/n_total_nhw)^2/(n_total_nhw-1)))
p_value=pt(test_total_stat,df_total)
p_value
```

```
## [1] 0.3999633
```

The p_value is calculated as 0.3999633 which is greater than α (0.05). Therefore we failed to reject null hypothesis and conclude that handwashing would not lower the average total score for the morality questions.

(e) Find 95% confidence intervals for (i) the population mean of Total for hand-washing, (ii) the population mean of Total for control, (iii) the difference in the population mean of Total between the two groups.

(i) 95% confidence intervals for the population mean of Total for hand-washing

```r
UL_hw=x_total_hw+qt(0.975,n_total_hw-1)*(s_total_hw/sqrt(n_total_hw))
LL_hw=x_total_hw-qt(0.975,n_total_hw-1)*(s_total_hw/sqrt(n_total_hw))
print(paste(" The 95% confidence interval for handwashing group for Total is between",
            LL_hw,"and",UL_hw))
```

```
## [1] " The 95% confidence interval for handwashing group for Total is between 32.6077890606924 and 34
```

(ii) 95% confidence intervals for the population mean of Total for control

```r
UL_nhw=x_total_nhw+qt(0.975,n_total_nhw-1)*(s_total_nhw/sqrt(n_total_nhw))
LL_nhw=x_total_nhw-qt(0.975,n_total_nhw-1)*(s_total_nhw/sqrt(n_total_nhw))
print(paste(" The 95% confidence interval for control group for Total is between",
            LL_nhw,"and",UL_nhw))
```

```
## [1] " The 95% confidence interval for control group for Total is between 32.9748010023142 and 34.706
```

(iii) 95% confidence intervals for the difference in the population mean of Total between the two groups.

```r
UL_d_nhw=(x_total_hw-x_total_nhw)+qt(0.975,df_total)*sqrt((s_total_hw^2/n_total_hw)+
                                                            (s_total_nhw^2/n_total_nhw))
LL_d_nhw=(x_total_hw-x_total_nhw)-qt(0.975,df_total)*sqrt((s_total_hw^2/n_total_hw)+
                                                            (s_total_nhw^2/n_total_nhw))
print(paste(" The 95% confidence interval for the difference in the population mean of Total between th
```

```
## [1] " The 95% confidence interval for the difference in the population mean of Total between the two
```

(f) Which test will be more reliable for determining whether handwashing affects moral judgments – the test for Trolley in part (b), or the test for Total in part (d)? Explain.

The test we have done for total in part (d) will be more reliable. Because the total score takes count of all the scores across measurements (6 judgemental questions) in which Trolley is also included. In doing so, total experiment reduces any bias that was there for the trolley experiment.

2. A statistician who knows much more about cricket and rugby than basketball and football wishes to compare the size of NBA and NFL players. He randomly selects twelve players from each league and collects the data shown in the tables on the next page.

(a) Find a 95% confidence intervals for: i.The mean height of all NBA players; ii.The mean height of all NFL players.

(i) 95% confidence intervals for mean height of all NBA players

```
height_NBA=c(80,82,82,77,74,81,83,80,76,82,81,81)
x_nba=mean(height_NBA)
s_nba=sd(height_NBA)
n_nba=length(height_NBA)
UL_nba=x_nba+qt(0.975,n_nba-1)*(s_nba/sqrt(n_nba))
LL_nba=x_nba-qt(0.975,n_nba-1)*(s_nba/sqrt(n_nba))
print(paste(" The 95% confidence interval for mean height of all NBA players is between",
            LL_nba,"and",UL_nba))
```

```
## [1] " The 95% confidence interval for mean height of all NBA players is between 78.1513337558994 and
```

(ii) 95% confidence intervals for mean height of all NFL players

```
height_NFL=c(71,71,74,73,76,74,79,75,70,77,77,68)
x_nfl=mean(height_NFL)
s_nfl=sd(height_NFL)
n_nfl=length(height_NFL)
UL_nfl=x_nfl+qt(0.975,n_nfl-1)*(s_nfl/sqrt(n_nfl))
LL_nfl=x_nfl-qt(0.975,n_nfl-1)*(s_nfl/sqrt(n_nfl))
print(paste(" The 95% confidence interval for mean height of all NFL players is between",
            LL_nfl,"and",UL_nfl))
```

```
## [1] " The 95% confidence interval for mean height of all NFL players is between 71.6668011675622 and
```

(b) (3 points) Find approximate 95% confidence intervals for: i.The median height of all NBA players; ii.The median height of all NFL players. If you cannot achieve exactly 95% confidence, get as close to that level of confidence as you can.

(i) 95% confidence intervals for median height of all NBA players

```
sort_height_NBA=sort(height_NBA)
alpha=1-0.95
k=qbinom(alpha/2,n_nba,0.5)
k
```

```
## [1] 3
```

By trial and error k found to be 2 for 96% which is near to 95% -

```r
1-2*pbinom(k,n_nba,0.5)
```

```
## [1] 0.8540039
```

```r
1-2*pbinom(k-1,n_nba,0.5)
```

```
## [1] 0.9614258
```

```r
kn=k-1
kn
```

```
## [1] 2
```

```r
print(paste("The 95% confidence interval for mean height of all NBA players is between",sort_height_NBA
```

```
## [1] "The 95% confidence interval for mean height of all NBA players is between 77 and 82"
```

(ii) 95% confidence intervals for median height of all NFL players

```r
sort_height_NFL=sort(height_NFL)
alpha=1-0.95
k=qbinom(alpha/2,n_nfl,0.5)
k
```

```
## [1] 3
```

By trial and error k found to be 2 for 96% which is near to 95% -

```r
1-2*pbinom(k,n_nfl,0.5)
```

```
## [1] 0.8540039
```

```r
1-2*pbinom(k-1,n_nfl,0.5)
```

```
## [1] 0.9614258
```

```r
kn=k-1
kn
```

```
## [1] 2
```

```r
print(paste("The 95% confidence interval for mean height of all NFL players is between",sort_height_NFL
```

```
## [1] "The 95% confidence interval for mean height of all NFL players is between 71 and 77"
```

(c) Suppose we are willing to assume that both NBA heights and NFL heights have close to normal distributions. Choose an appropriate statistical test for the null hypothesis that NBA players and NFL players have the same average height. Calculate the test statistic and P-value, and explain in words what you may conclude from this analysis.

$H_0 : \mu_0 - \mu_1 = 0$
$H_1 : \mu - \mu_1 \neq 0$

where μ_0= average height of NBA players μ_1= average height of NFL players

```r
test_stat_nn=(x_nba-x_nfl)/sqrt((s_nba^2/n_nba)+(s_nfl^2/n_nfl))
df_nn=((s_nba^2/n_nba)+(s_nfl^2/n_nfl))^2/
  (((s_nba^2/n_nba)^2/(n_nba-1))+((s_nfl^2/n_nfl)^2/(n_nfl-1)))
p_value=2*(1-pt(test_stat_nn,df_nn))
p_value
```

```
## [1] 6.082016e-05
```

The p_value is calculated as 0.0000608201 which is less than α (0.05). Therefore we reject null hypothesis and conclude that the average height for NBA players and NFL players are not same.