

Dice Data

Two-Sample T-Test

Brady Fisher

Section 1: Choose The Model

The goal of this project is to see if there is a difference in the 50 dice rolls made by my TA and the 50 dice rolls I made. To accomplish this I will import data for my TA's rolls, and my dice rolls.

The results of me rolling a die 50 times was:

```
myList = c(6, 6, 1, 1, 1, 1, 2, 4, 4, 5, 6, 2, 4, 1, 1, 5, 6, 5, 6, 2, 5, 1, 2, 3, 3,
           5, 3, 3, 3, 5, 5, 6, 4, 1, 5, 2, 6, 1, 5, 6, 5, 1, 4, 1, 5, 1, 4, 3, 3, 6)
```

Next I will import my TAs dice rolls. My TAs 50 dice rolls were:

```
TAList = read.csv("C:/Users/Brady Fisher/Documents/R/Stat3022/TA_Dice_Rolls.csv")
TAList$x
```

```
## [1] 3 3 5 6 3 4 2 4 2 3 2 3 3 4 5 2 1 6 6 1 5 6 3 5 1 4 6 1 5 5 3 3 4 3 4
## [36] 6 1 5 5 3 6 6 1 4 2 1 4 2 4 2
```

Now lets look at a summary of both data sets.

```
summary(myList)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00    2.00    4.00    3.52    5.00    6.00
```

```
summary(TAList$x)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00    2.00    3.50    3.56    5.00    6.00
```

I will combine the TA's and my dice rolls into one data frame of 100 rows and 2 columns.

```
dat = data.frame(Name = rep(c("TA", "Student"), each = 50), Outcome = c(TAList$x, myList))
```

Check the first 6 and last 6 rows to see how the data looks.

```
head(dat)
```

```
##      Name Outcome
## 1     TA         3
## 2     TA         3
## 3     TA         5
## 4     TA         6
## 5     TA         3
## 6     TA         4
```

```
tail(dat)
```

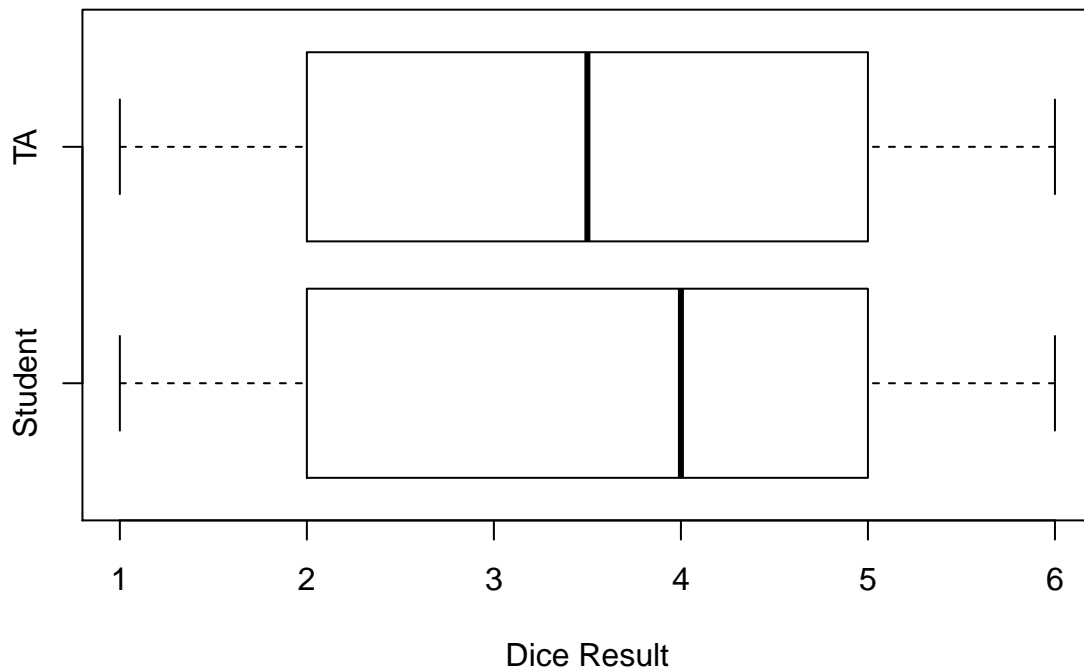
```
##      Name Outcome
## 95 Student      5
## 96 Student      1
## 97 Student      4
## 98 Student      3
## 99 Student      3
## 100 Student     6
```

Now for my data there are 2 variables. The first variable is called **Name**, which identifies who rolled the dice (Student or TA). This will be used as my explanatory variable and is binary.

The second variable is called **Outcome**, which specifies what number the die landed on. This will be my response variable and is quantitative.

Here is a quick look at the data through a side to side boxplot of the TA's and my data.

```
boxplot(Outcome~Name, data = dat, xlab="Dice Result", horizontal = TRUE)
```



Based on these graphs the hypotheses for my model will be:

Null Hypothesis $H_0 : \mu_{TA} = \mu_{Student}$

Alternative Hypothesis $H_a : \mu_{TA} \neq \mu_{Student}$

where μ_{TA} is the mean outcome of the TA's rolls, and $\mu_{Student}$ is the mean outcome of my rolls.

Section 2: Fit The Model

To fit the model I will use R's `t.test` with equal variance.

```
t.test(Outcome~Name, data=dat, conf.level = 0.95, alternative = "two.sided", var.equal = TRUE)

##
## Two Sample t-test
##
## data: Outcome by Name
## t = -0.11378, df = 98, p-value = 0.9096
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.737652 0.657652
## sample estimates:
## mean in group Student      mean in group TA
##                3.52                3.56
```

Section 3: Asses The Model

The assumptions of the two-sample t-test are that:

1. These groups each are from an independent sample.
2. Each group has a sample size greater than 30 or is approximately normally distributed.
3. The two groups have approximately the same variance.

The first assumption is meet since none of the TAs rolls or my rolls depended on any other roll of the dice.

The second assumption is meet since both groups had a sample size of 50.

The third assumption is meet since the boxplot I presented earlier consisting of the two groups with the Dice outcome shows that both groups had a similar range and interquartile range, and thus a similar variance.

Section 4: Use The Model

Since all the assumptions of the two-sample t-test are meet, the model seems appropriate. Therefore I can view the results of the test and make conclusions.

Since my p-value of the two-sample t-test is 0.9096, which is greater than 0.05, there is not enough evidence to reject the Null Hypothesis, H_0 . The 95% confidence interval is (-0.737652, 0.657652), which means that the difference in the mean outcome between the two groups is expected to be between -0.737652 and 0.657652. This indicates that the mean outcome of the dice rolls from the TA and me are not significantly different.