

STAT 450 Tutorial

Nam Hee Gordon Kim

January 11, 2018

About NHGK

My name is Nam. My name does not rhyme with Sam but with Islam and lip balm. I am originally from South Korea. However, I don't listen to K-Pop.

Favorite Equation

My favorite equation is the Bayes' Rule equation.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Rule is widely used in statistics and probability. The Bayesian approach to artificial intelligence and machine learning is rapidly gaining popularity.

Favorite Dataset

My favorite dataset is the `trees` dataset in R. Thirty-one (31) trees were measured. The features of interests are the height, girth, and volume of each tree.

```
summary(trees)
```

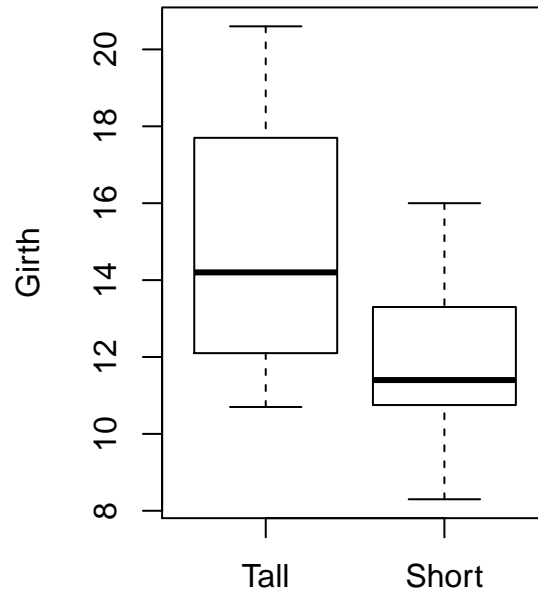
```
##      Girth      Height      Volume
##  Min.   : 8.30   Min.   :63   Min.   :10.20
##  1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40
##  Median :12.90   Median :76   Median :24.20
##  Mean   :13.25   Mean   :76   Mean   :30.17
##  3rd Qu.:15.25   3rd Qu.:80   3rd Qu.:37.30
##  Max.   :20.60   Max.   :87   Max.   :77.00
```

```
head(trees, 5)
```

```
##      Girth Height Volume
## 1      8.3     70    10.3
## 2      8.6     65    10.3
## 3      8.8     63    10.2
## 4     10.5     72    16.4
## 5     10.7     81    18.8
```

For this dataset, I show that tall trees tend to have larger girth. The definition of “tall” trees is above average height.

Girth of Tall vs. Short Trees



The median and mean girth was higher in the tall trees than those of the shorter trees. My initial intuition is that taller trees must support its height by growing in girth. I performed two-class student T test to test the hypothesis:

$$H_0 : \text{tall and short trees have the same population mean of girth}$$

which is formalized as

$$H_0 : \mu_{\text{tall}} = \mu_{\text{short}}$$

```
t.test(Girth ~ tall, data=trees)
```

```
##
##  Welch Two Sample t-test
##
## data:  Girth by tall
## t = -3.2911, df = 24.146, p-value = 0.003061
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.290160 -1.213174
## sample estimates:
## mean in group FALSE mean in group TRUE
##      11.67500      14.92667
```

Test statistic t returned -3.29. The p -value was 0.003. With significance level 95%, we reject H_0 in favor of the difference in population means.

Acknowledgments

This project was funded by My Parents Grant as well as Internship Money.