

# Stats Demo 1

## Environment setup

```
knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 60), tidy = TRUE)

library(Sleuth3)
library(reshape2)
library(ggplot2)
library(dplyr)
library(plotly)
#library(ggplot2)
```

## 1. Creating data

Creating data for test. Prices for CS course books in OSU bookstore. Notice the difference between lists and Data frames.

```
Prices <- c(99.34, 51.53, 20.45, 97.22, 61.89, 58.17, 61.63, 44.63, 96.69, 48.88)
Prices
```

```
## [1] 99.34 51.53 20.45 97.22 61.89 58.17 61.63 44.63 96.69 48.88
```

```
typeof(Prices)
```

```
## [1] "double"
```

```
#check? data_frame(Prices)
is.data.frame(Prices)
```

```
## [1] FALSE
```

```
#as trans the type
PricesDF <- as.data.frame(Prices)
colnames(PricesDF) <- ("Price")
PricesDF$Price
```

```
## [1] 99.34 51.53 20.45 97.22 61.89 58.17 61.63 44.63 96.69 48.88
```

```
PricesDF
```

```
##      Price
## 1  99.34
## 2  51.53
## 3  20.45
## 4  97.22
## 5  61.89
## 6  58.17
## 7  61.63
## 8  44.63
## 9  96.69
## 10 48.88
```

```
Genres <- factor(c("Science", "Engineering", "Art", "Science",
"Math", "Engineering", "Art", "Math", "Engineering", "Art"))
BookStore <- data.frame(Price = Prices, Genre = Genres)
BookStore
```

```
##      Price      Genre
## 1  99.34    Science
## 2  51.53 Engineering
## 3  20.45        Art
## 4  97.22    Science
## 5  61.89        Math
## 6  58.17 Engineering
## 7  61.63        Art
## 8  44.63        Math
## 9  96.69 Engineering
## 10 48.88        Art
```

## 2. Working with data

Manipulating data frames is easier as each column becomes an object.

## 3. Plots

ggplot provides a platform to plot various plots by changing the graph types. Refer to <https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf> for common commands. The type of plot depends on the nature of the variable. Here we see how bar plots are better suited for discrete data:

## 4. Distribution of Data

Sampling from any distribution, we can examine a histogram of the data:

## 5. Central Limit Theorem

When repeated samples are drawn with a sample size  $n \approx 30$ , the sample mean distribution approximately represents a normal distribution:

## 6. One sample t-test

Test whether the mean price of your sample of CS books is significantly different from the mean price of all books in the OSU bookstore (population):

## 7. Two sample t-test

What about when we compare against the mean price of all books on Amazon? We can compare using a similarly sized sample from Amazon. We then test if the mean prices are significantly different between the two samples:

## 8. One and two-tail analysis

alternative: the alternative hypothesis. Allowed value is one of “two.sided” (default), “greater” or “less”.  
var.equal: a logical variable indicating whether to treat the two variances as being equal. If TRUE then the pooled variance is used to estimate the variance otherwise the Welch test is used.

## 9. Paired t-test

When there exists a natural pairing in the observations. For example, when comparing the prices of the same books in Amazon and OSU BookStore.

## 10. Reporting