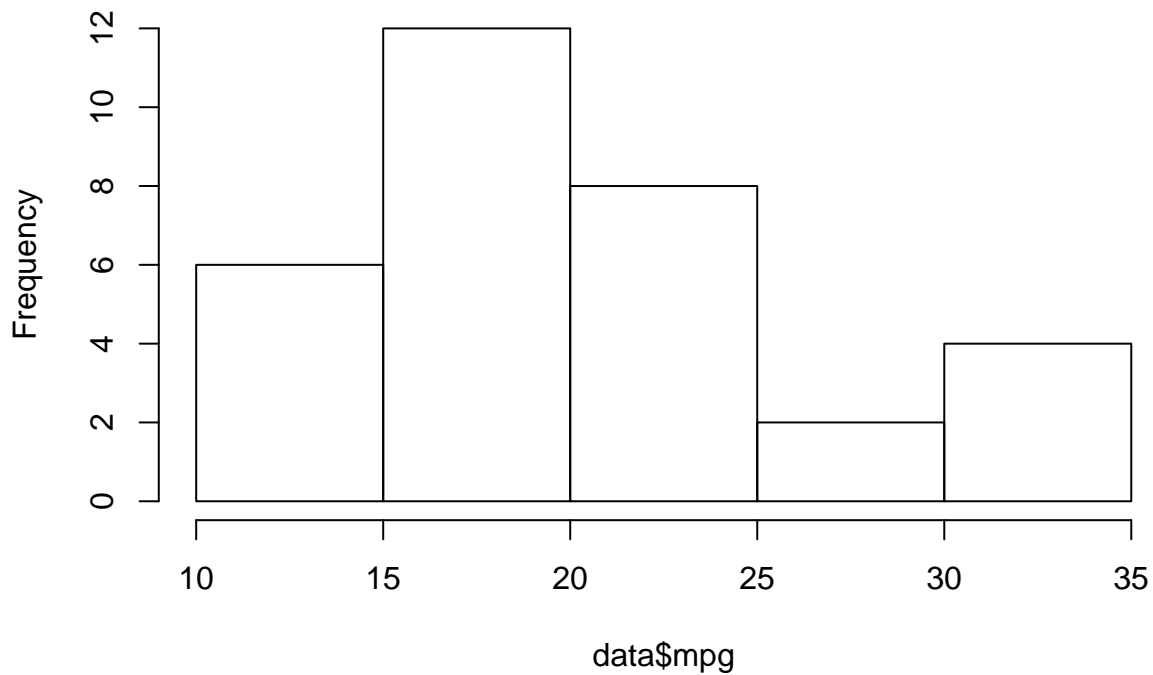# Central Limit Theorem

Patrick Kelly

2/12/2020

Central Limit Theorem

You're Not a Data Scientist If You Don't Know The Central Limit Theorem

Click Here"Central Limit Theorem"

```r
data <- mtcars
hist(data$mpg)
```

Histogram of data$mpg

```
# Test for normality
# H0 = distribution is normal
# HA = distribution is not normal
shapiro.test(data$mpg)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  data$mpg
## W = 0.94756, p-value = 0.1229
```

```
# If p-value>0.05, H0 is not rejected.
round(mean(data$mpg),2)
```
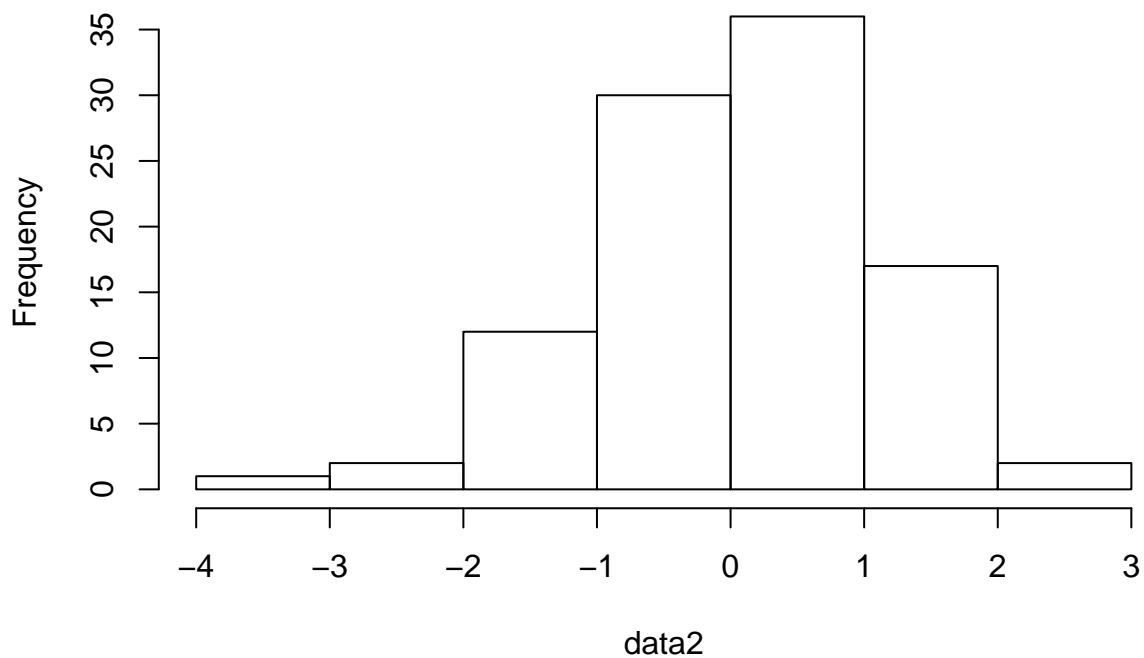
```
## [1] 20.09
```

Normal Distribution (Gaussian/Bell Curve)

```
# Create dataset with mean=0 and standard deviation = 1
n <- 100
Mean <- 0
Standard_Deviation <- 1
data2 <- rnorm(n,Mean,Standard_Deviation)
hist(data2)
```


Histogram of data2

```
round(mean(data2))
```

```
## [1] 0
```

```
round(sd(data2))
```

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

```
shapiro.test(data2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data2
## W = 0.98746, p-value = 0.4694
```

```
# If p-value>0.05, H0 is not rejected.
```

```
n <- 100
mpg_samples <- c()
for (i in 1:n) {
    mpg_samples[i] = mean(sample(data$mpg, 10, replace = TRUE))
}
class(mpg_samples)
```

```
## [1] "numeric"
```

```
hist(mpg_samples, col = 'lightgrey', xlab = "MPG")
m <- round(mean(mpg_samples),2)
std <- round(sd(mpg_samples),2)
st_error <- std/sqrt(n)
ci_ll <- m-1.96*st_error
ci_ul <- m+1.96*st_error
abline(v=round(m,2),col="blue")
abline(v=round(ci_ll,2),col="red")
abline(v=round(ci_ul,2),col="red")
```

Histogram of mpg_samples
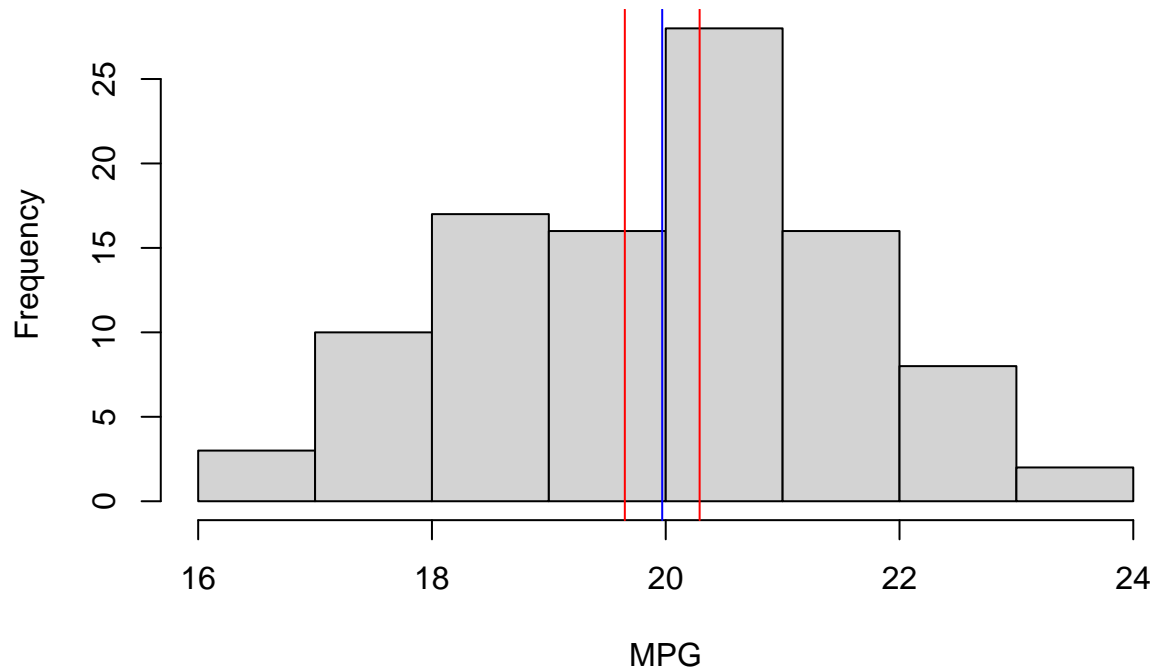

```r
shapiro.test(mpg_samples)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mpg_samples
## W = 0.98913, p-value = 0.5949
```

```r
# If p-value>0.05, H0 is not rejected.
cat("Mean of the 100 samples =",m)
```

```
## Mean of the 100 samples = 19.97
```

```r
cat("Lower limit of the 95% confidence interval =",round(ci_ll,2))
```

```
## Lower limit of the 95% confidence interval = 19.65
```

```r
cat("Upper limit of the 95% confidence interval =",round(ci_ul,2))
```

```
## Upper limit of the 95% confidence interval = 20.29
```

Blue line = mean of the 100 samples

Red lines = Lower and upper limits of the 95% confidence interval around the mean.