STAT 511A HW 2

Kathleen Wendt

Load packages

```
library(readr)
library(magrittr)
library(tibble)
library(ggplot2)
```

Part 1

Assume Z has a standard normal distribution.

Part 1A

P(Z <= 0.64)

```
pnorm(0.64)
```

```
## [1] 0.7389137
```

Part 1B

P(Z <= -0.37)

```
pnorm(-0.37)
```

```
## [1] 0.3556912
```

Part 1C

P(Z > 1.24)

```
pnorm(1.24, lower.tail = FALSE)
```

```
## [1] 0.1074877
```

Part 1D

P(-0.37 <= Z <= 1.15)

```
pnorm(1.15) - pnorm(-0.37)
```

```
## [1] 0.5192368
```

Part 1E

Find z such that P(Z <= z) = 0.3300

```
qnorm(0.3300)
```

```
## [1] -0.4399132
```

Part 1F

Find z such that P(Z > z) = 0.1841

```
qnorm(0.1841, lower.tail = FALSE)
```

```
## [1] 0.8998502
```

Part 2

Assume that Y has a normal distribution with mean of 5.4 and standard deviation of 0.2.

Part 2A

P(Y <= 5.7)

```
pnorm(5.7, mean = 5.4, sd = 0.2)
```

```
## [1] 0.9331928
```

Part 2B

P(Y > 5.3)

```
pnorm(5.3, mean = 5.4, sd = 0.2, lower.tail = FALSE)
```

```
## [1] 0.6914625
```

Part 2C

P(5.2 <= Y <= 5.5)

```
pnorm(5.5, mean = 5.4, sd = 0.2) - pnorm(5.2, mean = 5.4, sd = 0.2)
```

```
## [1] 0.5328072
```

Part 2D

Find the value y such that P(Y <= y) = 0.85.

```
qnorm(0.85, mean = 5.4, sd = 0.2)
```

```
## [1] 5.607287
```

Part 3

Let Y have a *skewed* distribution ($\mu = 80$, $\sigma = 5$). Suppose a random sample of $n = 100$ is drawn from the population.

Part 3A

Based on the Chebyshev's Rule (selected based on skewness), at least 75% of the sample data will lie between 70 and 90.

Part 3B

The sampling distribution would have mean $= \mu = 80$ and its standard deviation would be $\sigma/\sqrt(n) = 5/\sqrt(100) = 0.5$. Based on the Central Limit Theorem, for any variable Y, with a finite mean of μ and standard deviation of σ, $\bar{x}$ converges to a *normal distribution* with mean of μ and standard deviation of $\sigma/\sqrt(n)$, as n increases.

Part 4

Seed Data

Part 4A

Read Data

```
seeds <- read_csv("Seeds.csv")
```

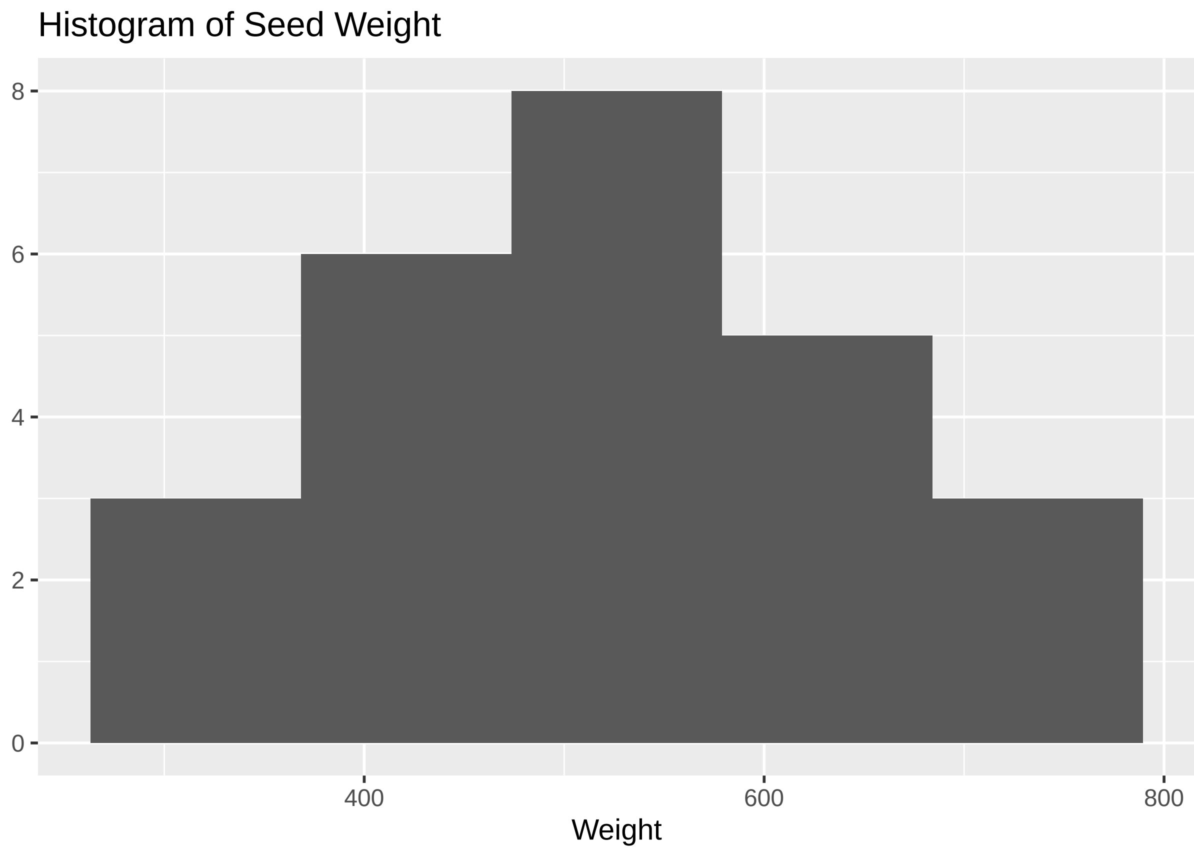
Check Data

```
glimpse(seeds)
```

```
## Observations: 25
## Variables: 1
## $ Weight <dbl> 343, 659, 348, 433, 755, 441, 469, 583, 431, 562, 545, ...
```

Histogram

```
qplot(Weight, data = seeds) +
  stat_bin(bins = 5) +
  ggtitle("Histogram of Seed Weight")
```



Sample Mean

```r
mean(seeds$Weight)
```

```
## [1] 526.12
```

Sample Standard Deviation

```r
sd(seeds$Weight)
```

```
## [1] 113.7279
```

Part 4B

Construct a 95% confidence interval for seed μ.

```r
seed_ci <- t.test(seeds$Weight)
seed_ci$conf.int
```

```
## [1] 479.1754 573.0646
## attr(,"conf.level")
## [1] 0.95
```

Part 4C

Interpret confidence interval.

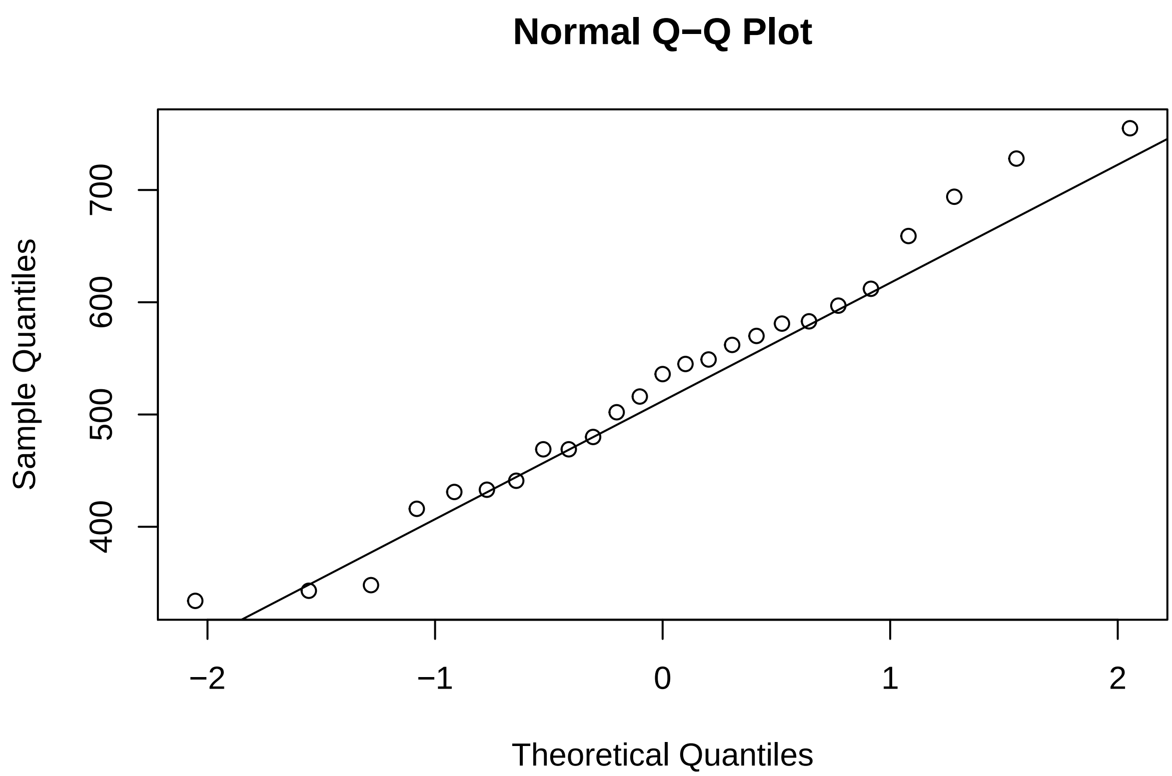
A 95% confidence interval is a random interval which, if the model is true, would include the true value of the population parameter μ with a probability of 95%. In the case of seed weight μ, the 95% confidence interval 479.1754457, 573.0645543 is an interval calculated by a method such that under repeated sampling, 95% of such intervals would include μ. Stated differently, 95% of all 95% confidence intervals for seed weight μ contain μ.

Part 4D

The validity of the above confidence interval is adequate. In order to construct a valid confidence interval of a population mean (μ), four assumptions need to be met:

1. Random sample. In this instance, the seeds were randomly sampled.
2. Independent observations. Because the seeds were randomly sampled, independence of observations can be assumed.
3. Normally distributed data. The distribution of the sample is approaching normal (with smaller bins) and will approach normality with larger sample sizes due to the Central Limit Theorem. The below Q-Q plot indicates limited deviation in the tails.

```
qqnorm(seeds$Weight)
qqline(seeds$Weight)
```



4. Large sample size. The current sample size (n = 25) does not meet this assumption.

Part 5

Describe how the following affect the width of a confidence interval.

Part 5A

Increasing sample size would *decrease* the width of a confidence interval.

Part 5B

Increasing the confidence level would *increase* the width of a confidence interval.

Part 5C

Increasing standard deviation would *increase* the width of a confidence interval.