STAT 511A HW 2

Kathleen Wendt

Load packages

```
library(readr)
library(magrittr)
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
```

```
1 - pnorm(1.24)
```

```
## [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

```r
qnorm(0.3300)
```

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

Part 1F

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

```r
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)

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

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

Part 2B

P(Y > 5.3)

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

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

Part 2C

P(5.2 <= Y <= 5.5)

```r
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.

```r
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 sample distribution would be similar to the population distribution, such that its μ would be close to 80 and its σ would be close to 5 with a similar level of skewness. Based on the Central Limit Theorem, if more and more samples of 100 are drawn randomly, the sampling distribution of the sample mean will approach normality.

Part 4

Seed Data

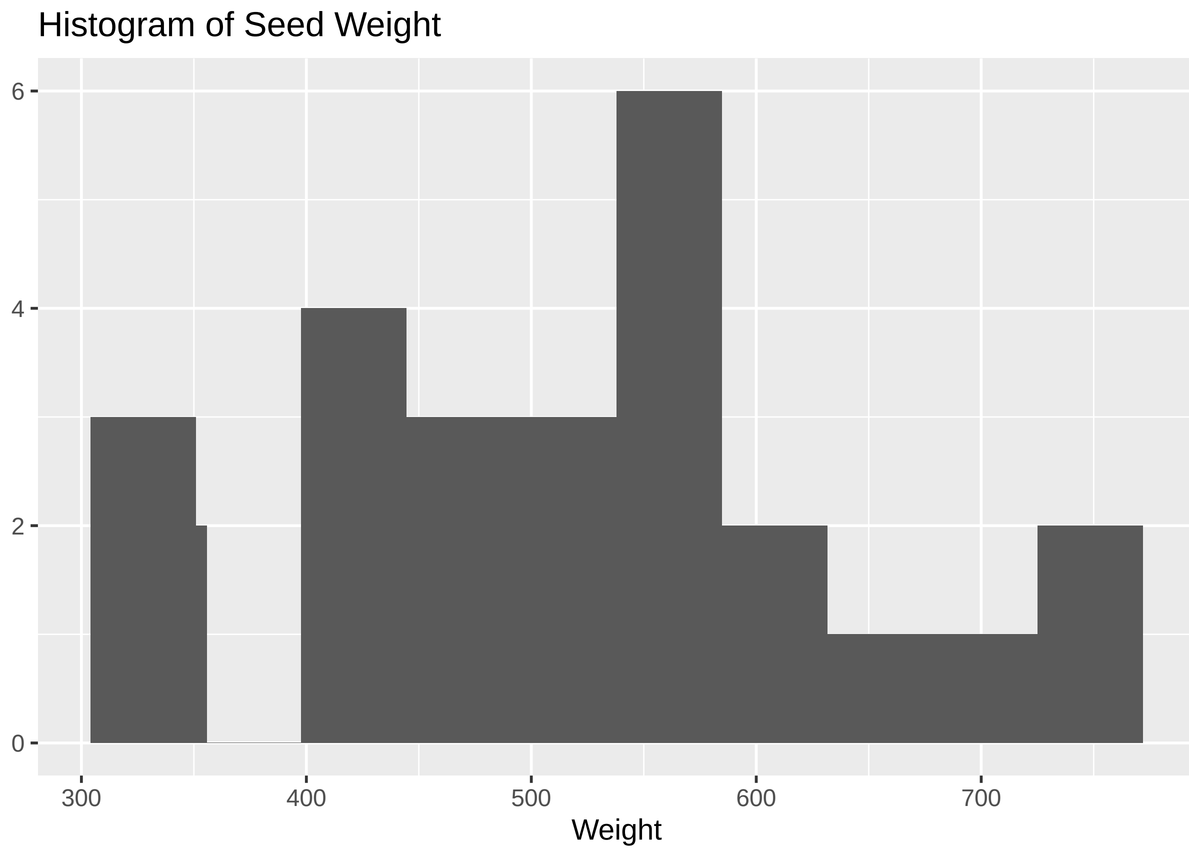
Part 4A

Read Data

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

Histogram

```
qplot(Weight, data = seeds) +
  stat_bin(bins = 10) +
  ggtitle("Histogram of Seed Weight")
```



Sample Mean

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

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

Sample Standard Deviation

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

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

Part 4B - I might need to redo some of this later this week

Construct a 95% confidence interval for seed μ.

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

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

The 95% confidence interval of seed weight is 479.12 and 573.06.

Part 4C

Interpret confidence interval.

A confidence interval is an interval of values computed from sample data that is almost sure to cover the true population parameter. In this case, if 100 samples of seeds' weights were collected, 95 of the samples would contain the true μ (population mean seed weight) (between 479.12 and 573.06).

Part 4D

The validity of the above confidence interval is medium. 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. It is unknown whether the seeds were somehow interdependent; however, it is unlikely.
3. Normally distributed data. The sample distribution is not normal but might approach normality with larger sample sizes.
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 not affect the width of a confidence interval; the width would *stay the same.*