STAT511 Hw2

Amy Fox

Question 1

Assume that Z has a standard normal distribution. Compute the following.

A. P(Z $\leq$ 0.64)

```r
pnorm(0.64)
```

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

B. P(Z $\leq$ -0.37)

```r
pnorm(-0.37)
```

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

C. P(Z $>$ 1.24)

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

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

D. P(-0.37 $\leq$ Z $\leq$ 1.15)

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

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

E. Find the value z such that P(Z $\leq$ z) $=$ 0.3300

```r
qnorm(0.3300)
```

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

F. Find the value z such that P(Z $>$ z) $=$ 0.1841

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

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

Question 2

Assume that Y has a normal distribution with mean 5.4 and standard deviation 0.2. Compute the following.

A. P(Y $\leq$ 5.7)

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

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

B. P(Y $>$ 5.3)

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

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

C. P(5.2 $\leq$ Y $\leq$ 5.5)

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

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

D. Find the value y such that P(Y $\leq$ y) = 0.85.

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

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

Question 3

Let Y have a skewed distribution with $\mu = 80$ and $\sigma = 5$. Suppose a random sample of size n=100 is drawn from the population.

A. Give an interval with the property that at least 75% of the data will be in that interval. What rule did you use to determine the interval?

Chebyshev's rule is used to determine the interval where for any distribution (non-normal), at least 75% of the data will lie within $\bar{y} \pm$ 2s. In this case, the data will lie within 80 $\pm$ 10, therefore the internval is (70, 90)

B. Describe the distribution of $\bar{y}$. Give the mean, standard deviation and shape of the distribution. (3 pts)

```r
# standard deviation
5/sqrt(100)
```

```
## [1] 0.5
```

Need to check the mean and if there is another calculation

The mean is 80. For non-normal data, the standard deviation is sigma/sqrt(n). The standard deviation is 0.5, and the shape of the distribution is approaching normal because according to the Central Limit Theorem, as n increases, the distribution becomes close to normal.

Checked above with Ariel

Question 4

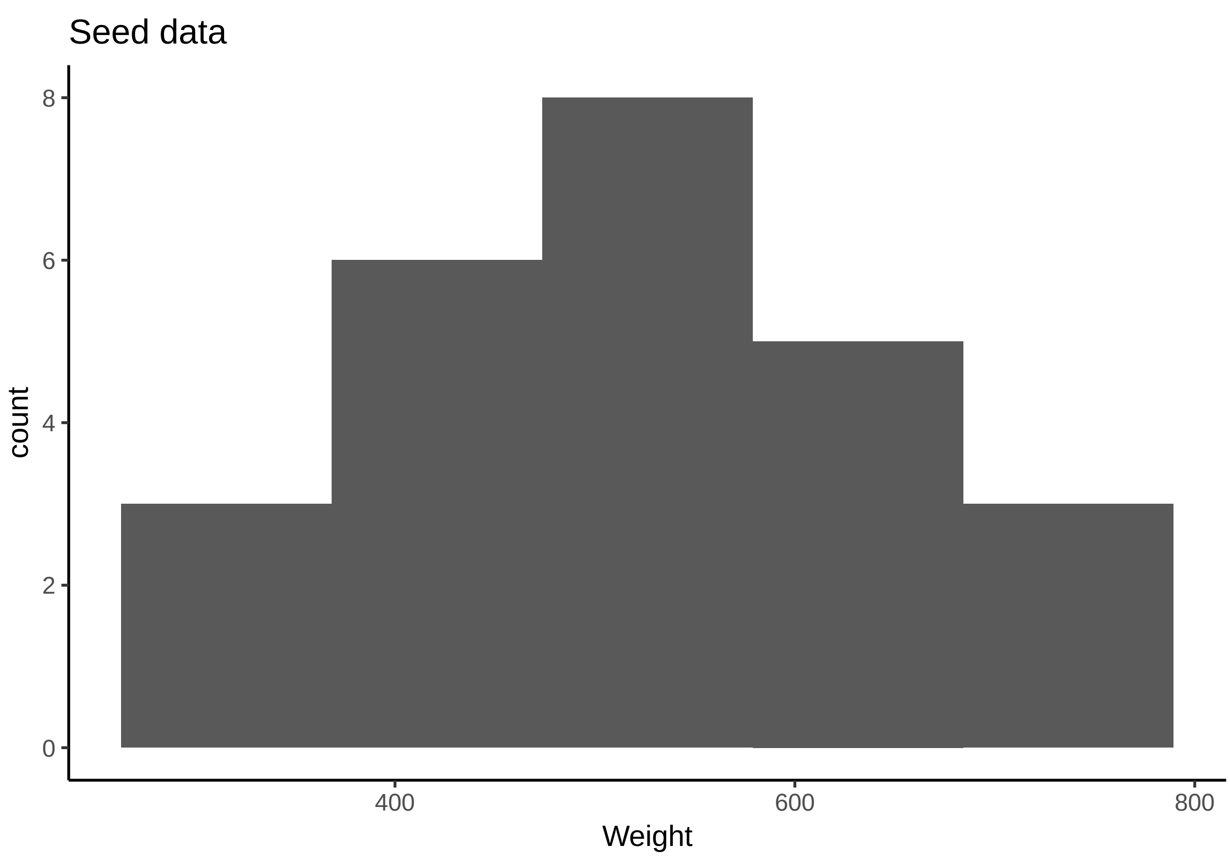
A random sample of n=25 seeds from a particular bean population is obtained. The weight (g) of each seed is recorded. The data is available from Canvas as "Seeds.csv".

Reminders: (1) Use read.csv() to import the data. (2) Use str() to check the data after importing. (3) Use $ or with() to access the Weight column!

A. Construct a histogram of the data. Also give the sample mean and sample standard deviation. (3 pts)

```r
library(ggplot2)
seed_data <- read.csv("../Data/Seeds.csv")

# histogram of the data
ggplot(seed_data, aes(Weight)) +
  geom_histogram(bins = 5) +
  ggtitle("Seed data") +
  xlab("Weight") +
  theme_classic()
```




```r
# mean
mean(seed_data$Weight)
```

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

```r
# standard deviation
sd(seed_data$Weight)
```

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

B. Give a 95% confidence interval for μ (population mean seed weight).

For a 95% confidence interval, $\alpha = 0.05$ and $\alpha/2 = 0.025$. First t-$\alpha/2$ must be calculated. df = n-1 where n = 25

```r
# this is t-alpha/2
qt(1-0.025, df = 25-1)
```

```
## [1] 2.063899
```

```r
mean(seed_data$Weight) + qt(1-0.025, df = 25-1)*(sd(seed_data$Weight)/sqrt(25))
```

```
## [1] 573.0646
```

```r
mean(seed_data$Weight) - qt(1-0.025, df = 25-1)*(sd(seed_data$Weight)/sqrt(25))
```

```
## [1] 479.1754
```

```r
# or I can do this...
t.test(seed_data$Weight)
```

```
##
##  One Sample t-test
##
## data:  seed_data$Weight
## t = 23.131, df = 24, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  479.1754 573.0646
## sample estimates:
## mean of x
##    526.12
```

The CI is (479.1754457, 573.0645543).

C. Interpret your confidence interval from part B.

I am 95% confident that the true population mean of the seed weight is between 479g and 573g.

D. Do you think the CI is valid? In other words, discuss whether assumptions satisfied.

We must assume that this seed data is a random sample with independent observations. Further, the data must be normally distributed or have a large sample size.

We know that the data is a random sample as stated in the problem. The data also appears to be normally distributed as shown in the histogram.

Question 5

Describe how the following affect the width of the confidence interval (assuming everything else is held constant). Answer should be increase, decrease or stays the same.

A. Sample size increases.

Decrease

B. Confidence level increases.

Increase

C. Standard deviation increases.

Increase