Stat/Math 415 Homework 7

Due Friday Nov 8; Joseph Sepich (jps6444)

1 Problem 8.1-1

Problem constraints:

- X ~ N(μ, 100)
- H_0: $\mu = 110$
- H_1: $\mu > 110$
- n = 16
- $\overline{x} = 113.5$

1.1 Part a

Given a significance level of $\alpha = 0.05$, we want to determine if we reject the null hypothesis. Let us use a critical region approach where we will only reject the null hypothesis if observed $Z > Z_\alpha$. The $Z_\alpha = Z_{0.05}$ is 1.645. Let's transform our sample mean into a test statistic using the known variance of 100.

$$Z = \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{113.5 - 110}{\sqrt{100/16}} = 1.40$$

Since 1.40 < 1.645, our test statistic is **not** within the critical region, therefore we fail to reject the null hypothesis of $\mu = 110$.

1.2 Part b

Given a significance level of $\alpha = 0.10$, we want to determine if we reject the null hypothesis. Let us use a critical region approach where we will only reject the null hypothesis if observed $Z > Z_\alpha$. The $Z_\alpha = Z_{0.10}$ is 1.28. Let's transform our sample mean into a test statistic using the known variance of 100.

$$Z = \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{113.5 - 110}{\sqrt{100/16}} = 1.40$$

Since 1.40 > 1.28, our test statistic is within the critical region, therefore we will reject the null hypothesis of $\mu = 110$.

1.3 Part c

The p-value for this test corresponds to the probability of having our test statistic or higher given our distribution: $P(Z > 1.40)$. Looking at the Z score table the p-value would be 1-0.9192 = **0.0808**.

2 Problem 8.1-3

Problem constraints:

- X ~ N(μ, 100)
- H_0: $\mu = 170$
- H_1: $\mu > 170$
- n = 25

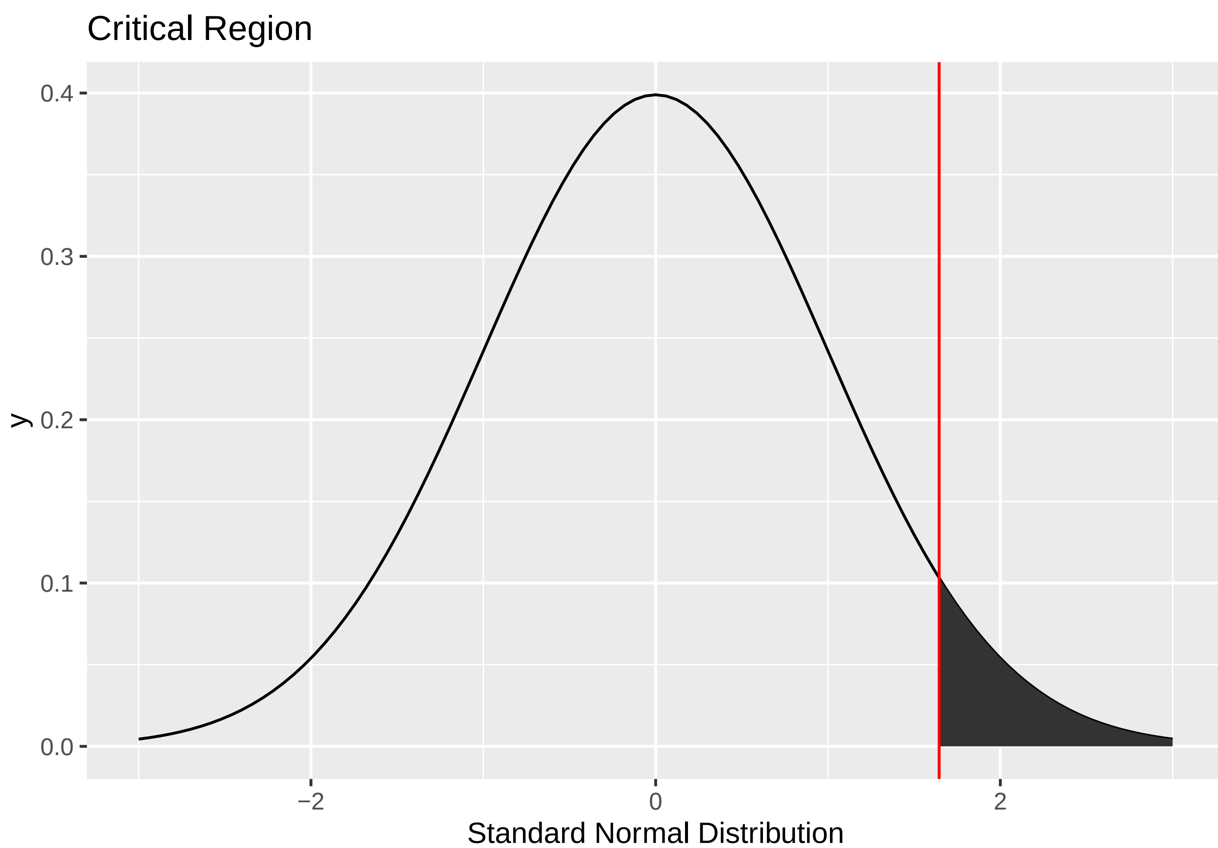
2.1 Part a

The test statistic for our distribution (varaince of 100) is as follows:

$$Z = \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\overline{x} - 170}{\sqrt{100/25}} = \frac{\overline{x} - 170}{2}$$

Given a significance level of $\alpha = 0.05$, the criticial region is $Z > Z_\alpha = 1.645$

```
ggplot(data.frame(x = c(-3, 3)), aes(x)) +
  stat_function(fun = dnorm) +
  stat_function(fun = dnorm,
                xlim = c(1.645,3),
                geom = "area") +
  geom_vline(xintercept=1.645, color="red") +
  ggtitle('Critical Region') +
  xlab('Standard Normal Distribution')
```



2.2 Part b

To calculate the value of the test statisic we must first calculate the value of the sample mean: $\overline{x} = \frac{1}{n}\Sigma_{i=1}^{25} x_i$.

```
x_data <- c(170, 167, 174, 179, 179, 156, 163, 156, 187, 156, 183, 179, 174, 179,
            170, 156, 187, 179, 183, 174, 187, 167, 159, 170, 179)
n <- 25

x_mean <- sum(x_data) / n
paste0('The vale of the sample mean is: ', x_mean)
```

```
## [1] "The vale of the sample mean is: 172.52"
```

$$Z = \frac{\overline{x} - 170}{2} = \frac{172.52 - 170}{2} = 1.26$$

Given the value of our test statistic and our critical region, we **cannot** reject the null hypothesis $\mu = 170$, because the test statistic 1.26 is not in the critical region $Z > 1.645$.

2.3 Part c

The p-value for this test corresponds to the probability of having our test statistic or higher given our distribution: $P(Z > 1.26)$. Looking at the Z score table the p-value would be 1-0.8962 = **0.1038**.

3 Problem 8.1-6

Problem constraints:

- X $\approx N(\mu = 3.4, \sigma^2)$
- n = 9

3.1 Part a

The null hypothesis is H_0: $\mu = 3.4$.

3.2 Part b

The coach's hypothesis or the null hypothesis is H_1:$\mu > 3.4$.

3.3 Part c

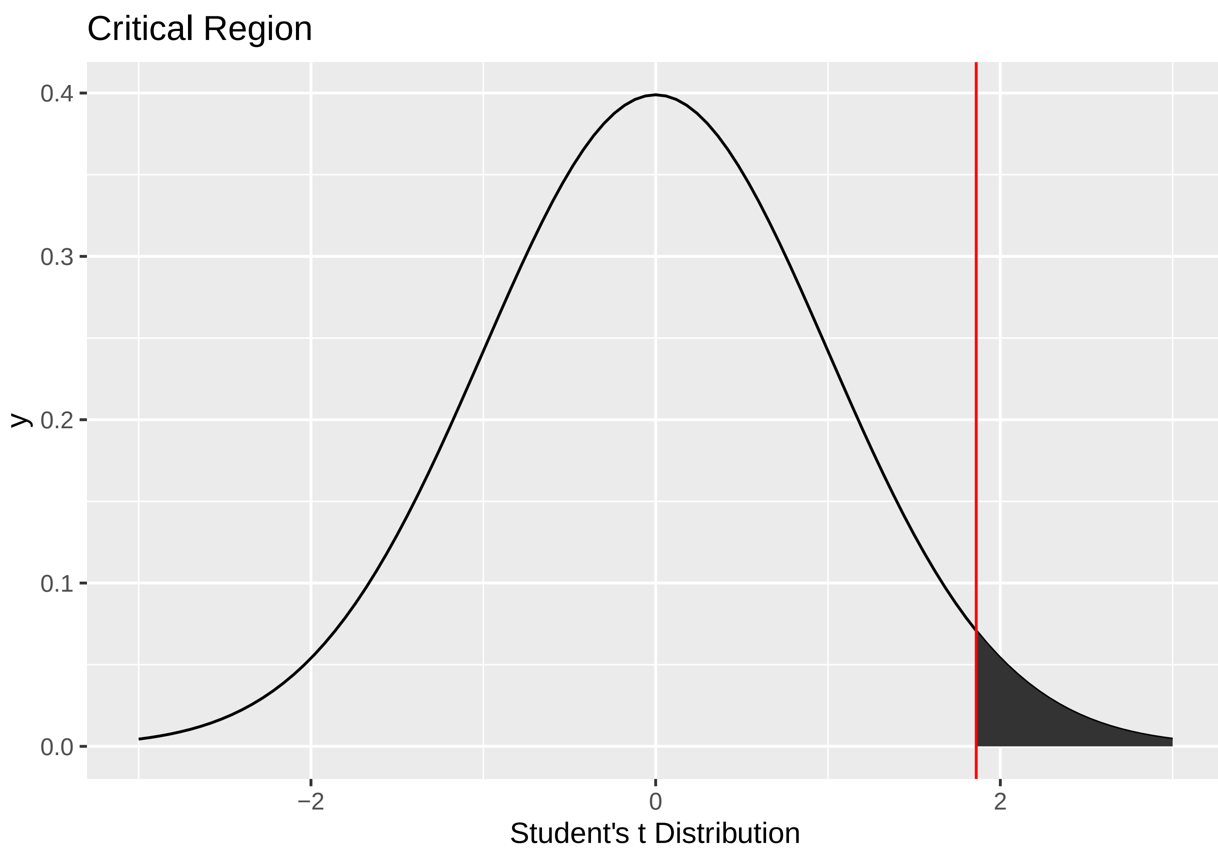
Test test statistic follows the t distribution, since we are using an approximation and have an unknown variance. Transforming to student's t would be:

$$t = \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\overline{x} - 3.4}{\sqrt{\frac{S^2}{9}}}$$

3.4 Part d

A critical region for a value of $\alpha = 0.05$ would be $t > t_\alpha = t_{0.05} = 1.860$. This would come from a t table with n-1 = 8 degrees of freedom.

```
ggplot(data.frame(x = c(-3, 3)), aes(x)) +
  stat_function(fun = dnorm) +
  stat_function(fun = dnorm,
                xlim = c(1.860,3),
                geom = "area") +
  geom_vline(xintercept=1.860, color="red") +
  ggtitle('Critical Region') +
  xlab('Student\'s t Distribution')
```



3.5 Part e

Before calculating the test statistic we must first calculate the sample mean and sample variance:

```
x_data <- c(3.4, 3.6, 3.8, 3.3, 3.4, 3.5, 3.7, 3.6, 3.7)
n <- 9

x_mean <- sum(x_data) / n
x_var <- sum((x_data - x_mean)^2) / (n-1)

paste0('The sample mean: ', x_mean)
```

```
## [1] "The sample mean: 3.55555555555556"
```

```
paste0('The sample variance: ', x_var)
```

```
## [1] "The sample variance: 0.0277777777777778"
```

$$t = \frac{3.556 - 3.4}{\sqrt{\frac{0.0278}{9}}} \approx 2.80$$

3.6 Part f

Given that our test statistic of $t_{0.05} = 2.80$ is within our critical region $t > 1.860$, we will reject the null hypothesis that $\mu = 3.4$.

3.7 Part g

The p-value for this test corresponds to the probability of having our test statistic or higher given our distribution: $P(t > 2.80)$ at n-1 = 8 degrees of freedom. Looking at the t table the p-value would be 1-0.9884 = **0.0116**.

4 Problem 8.2-3

Problem constraints:

- X and Y are normally distributed
- n = 9, m = 13
- H_0: $\mu_x = \mu_y$

Let's assume a significance level of $\alpha = 0.05$. Let's begin by defining our test statistic and relevant values.

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}}$$

$$\overline{x} = \frac{1}{n}\Sigma_{i=1}^n x_i$$

$$S^2 = \frac{1}{n-1}\Sigma_{i=1}^n (x_i - \overline{x})^2$$

$$S_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

```
x_data <- c(21.7, 21.0, 21.2, 20.7, 20.4, 21.9, 20.2, 21.6, 20.6)
y_data <- c(21.5, 20.5, 20.3, 21.6, 21.7, 21.3, 23.0, 21.3, 18.9, 20.0, 20.4, 20.8,
            20.3)
n <- 9
m <- 13

x_mean <- sum(x_data) / n
x_var <- sum((x_data - x_mean)^2) / (n-1)
y_mean <- sum(y_data) / m
y_var <- sum((y_data - y_mean)^2) / (m-1)
pool_var <- ((n - 1) * x_var + (m - 1) * y_var) / (n + m - 2)

paste0('The x sample mean: ', x_mean)
```

```
## [1] "The x sample mean: 21.0333333333333"
```

```
paste0('The x sample variance: ', x_var)
```

```
## [1] "The x sample variance: 0.3675"
```

```r
paste0('The y sample mean: ', y_mean)
```

```
## [1] "The y sample mean: 20.8923076923077"
```

```r
paste0('The y sample variance: ', y_var)
```

```
## [1] "The y sample variance: 1.01410256410256"
```

```r
paste0('The pooled sample variance: ', pool_var)
```

```
## [1] "The pooled sample variance: 0.755461538461539"
```

```r
t <- (x_mean - y_mean) / (sqrt(pool_var * (1/n + 1/m)))

paste0('Our test statistic is: ', t)
```

```
## [1] "Our test statistic is: 0.374173902560697"
```

Given our signifance level of $\alpha = 0.05$ we can check the t table with n + m - 2 = 20 degrees of freedom to get a critical region of $|t| > 2.086$. Since our sample test statistic of $t = 0.374$ is not within the critical region we **fail** to reject the null hypothesis.

5 Problem 8.2-5

Problem constraints:

- X and Y are normally distributed with the same variance
- H_0: $\mu_x - \mu_y = 0$
- H_1: $\mu_x - \mu_y < 0$
- $\alpha = 0.05$

5.1 Part a

The test statisitc for comparing these two sample means would be:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}}$$

Given a significance level of $\alpha = 0.05$ and degrees of freedom of n + m -2 = 26. The crtical region would be $t < -t_\alpha = -t_{0.05} = -1.706$.

5.2 Part b

```r
x_data <- c(49, 108, 110, 82, 93, 114, 134, 114, 96, 52, 101, 114, 120, 116)
y_data <- c(133, 108, 93, 119, 119, 98, 106, 131, 87, 153, 116, 129, 97, 110)
n <- 14
m <- 14

x_mean <- sum(x_data) / n
x_var <- sum((x_data - x_mean)^2) / (n-1)
y_mean <- sum(y_data) / m
y_var <- sum((y_data - y_mean)^2) / (m-1)
pool_var <- ((n - 1) * x_var + (m - 1) * y_var) / (n + m - 2)

paste0('The x sample mean: ', x_mean)
```

```
## [1] "The x sample mean: 100.214285714286"
```

```r
paste0('The x sample variance: ', x_var)
```

```
## [1] "The x sample variance: 604.489010989011"
```

```r
paste0('The y sample mean: ', y_mean)
```

```
## [1] "The y sample mean: 114.214285714286"
```

```r
paste0('The y sample variance: ', y_var)
```

```
## [1] "The y sample variance: 329.258241758242"
```

```r
paste0('The pooled sample variance: ', pool_var)
```

```
## [1] "The pooled sample variance: 466.873626373626"
```

```r
t <- (x_mean - y_mean) / (sqrt(pool_var * (1/n + 1/m)))

paste0('Our test statistic is: ', t)
```

```
## [1] "Our test statistic is: -1.71426273710483"
```

Given the value of our test statistic of $t = -1.714$ we can **reject** the null hypothesis $\mu_x - \mu_y = 0$, since $-1.714 < -1.706$.

5.3 Part c

The p-value for this test corresponds to the probability of having our test statistic or higher given our distribution: $P(t < -1.714)$ at n+m-2 = 26 degrees of freedom. Looking at the t table the p-value would be **0.049**.

6 Problem 8.2-9

Problem constraints:

- X and Y are normally distributed with same variance
- n = m = 13
- H_0: $\mu_x = \mu_y$
- H_1: $\mu_x < \mu_y$
- $\alpha = 0.05$

First let us calculate the test statistic t:

```
x_data <- c(2.9, 14.9, 1.0, 12.6, 9.4, 7.6, 3.6, 3.1, 2.7, 4.8, 3.4, 7.1, 7.2)
y_data <- c(7.8, 4.2, 2.4, 12.9, 17.3, 10.4, 5.9, 4.9, 5.1, 8.4, 10.8, 23.4, 9.7)
n <- 13
m <- 13

x_mean <- sum(x_data) / n
x_var <- sum((x_data - x_mean)^2) / (n-1)
y_mean <- sum(y_data) / m
y_var <- sum((y_data - y_mean)^2) / (m-1)
pool_var <- ((n - 1) * x_var + (m - 1) * y_var) / (n + m - 2)

paste0('The x sample mean: ', x_mean)
```

```
## [1] "The x sample mean: 6.17692307692308"
```

```
paste0('The x sample variance: ', x_var)
```

```
## [1] "The x sample variance: 17.2502564102564"
```

```
paste0('The y sample mean: ', y_mean)
```

```
## [1] "The y sample mean: 9.47692307692308"
```

```
paste0('The y sample variance: ', y_var)
```

```
## [1] "The y sample variance: 33.6852564102564"
```

```
paste0('The pooled sample variance: ', pool_var)
```

```
## [1] "The pooled sample variance: 25.4677564102564"
```

```
t <- (x_mean - y_mean) / (sqrt(pool_var * (1/n + 1/m)))

paste0('Our test statistic is: ', t)
```

```
## [1] "Our test statistic is: -1.66715229469244"
```

Given this one sided test at $\alpha = 0.05$, $P(t < -1.6672)$ at n + m - 2 = 24 degress of freedom is roughly greater than 0.05 (0.056), since $t_{0.05} = -1.711$. Under this significance level we **fail** to reject the null hypothesis, since our p-value is greater than $\alpha = 0.05$.