Homework 3

Kai Aragaki

Problem 1

A

```
qt(0.95, 11)
```

```
## [1] 1.795885
```

The confidence intervals would then be calculated by $\bar{x} \pm 1.80 \times \frac{S}{\sqrt{12}}$

B

```
qt(0.975, 29)
```

```
## [1] 2.04523
```

The confidence intervals would then be calculated by $\bar{x} \pm 2.05 \times \frac{S}{\sqrt{30}}$

C

```
qt(0.9, 17)
```

```
## [1] 1.333379
```

The confidence intervals would then be calculated by $\bar{x} \pm 1.33 \times \frac{S}{\sqrt{17}}$

Problem 2

For a 95% confidence interval ($\alpha = 0.05$, so $\frac{\alpha}{2} = 0.025$)

```
n <- 400
x_bar <- 140
s <- 25
t <- qt(0.975, 399)
ll <- x_bar - (t * s / sqrt(n))
ul <- x_bar + (t * s / sqrt(n))
```

So the mean, with 95% confidence intervals, could be expressed as 140mm (137.54mm, 142.46mm)

Similarly, for a 99% confidence interval ($\alpha = 0.01$, $\frac{\alpha}{2} = 0.005$),

```
t <- qt(1 - 0.005, 399)
ll <- x_bar - (t * s / sqrt(n))
ul <- x_bar + (t * s / sqrt(n))
```

Giving a mean, with 99% confidence intervals, could be expressed as 140mm (136.76mm, 143.24mm)

Problem 3

A

```
# Our measurements
mm <- c(107, 101, 93, 94, 96, 114)
alpha <- 0.05
n <- length(mm)
df <- n - 1

x_bar <- mean(mm)
s <- sd(mm) # We should note: sd uses n-1 in the denominator as a default

t <- qt(1 - alpha / 2, df)
ll <- x_bar - (t * s / sqrt(n))
ul <- x_bar + (t * s / sqrt(n))
```

So a 95% confidence interval for the population mean ranges from 92.14 to 109.52

B

We know that a confidence interval for the standard deviation can be represented as

$$(S\sqrt{\frac{n-1}{U}}, S\sqrt{\frac{n-1}{L}})$$

We can find our upper (L) and lower (U) values using the χ^2_{n-1} distribution

```
l <- qchisq(alpha / 2, df)
u <- qchisq(1 - alpha / 2, df)

ll <- s * sqrt(df / u)
ul <- s * sqrt(df / l)
```

So a 95% confidence interval for parameter σ is 5.17 to 20.31

C

The confidence interval for the variance is simply the square of each end of the interval:

$$(\frac{(n-1)S^2}{U}, \frac{(n-1)S^2}{L})$$

So this can be calculated as

```
ll <- (s^2 * df) / u
ul <- (s^2 * df) / l
```

So a 95% confidence interval for parameter σ^2 is 26.72 to 412.45

Problem 4

First, let's start with an analytic derivation:

We calculate some standard values first:

```
a <- c(132, 72, 102, 115, 59, 103, 86, 159, 60, 94, 80, 97)
b <- c(101, 96, 93, 106, 81, 77, 106, 97, 74)
n_a <- length(a)
n_b <- length(b)
df_a <- n_a - 1
df_b <- n_b - 1
v_a <- var(a)
v_b <- var(b)
s_a <- sd(a)
s_b <- sd(b)
x_a <- mean(a)
x_b <- mean(b)
```

Then, since we have assumed that the population standard deviation with each strain is the same, we can pool the variances to get $\hat{\sigma}_{pooled}$:

```
s_pool <- sqrt((v_a * df_a + v_b * df_b) / (df_a + df_b))
```

We can then calculate an estimated standard error for the difference of the two sample means using

$\hat{SD}(\bar{X} - \bar{Y}) = \hat{\sigma}_{pooled}\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$

```
est_sd_diff <- s_pool * sqrt(1/n_a + 1/n_b)
```

The interval then comes to be

```
diff_mean <- x_a - x_b
diff_int_ll <- (x_a - x_b) - qt(0.975, df_a + df_b) * est_sd_diff
diff_int_ul <- (x_a - x_b) + qt(0.975, df_a + df_b) * est_sd_diff
```

So the 95% confidence interval for $\bar{X}_a - \bar{X}_b$ is -17.44 to 25.94

Now, as fun as this was, let's see if we can get the same results much more quickly using R's built in functions:

```
tt <- t.test(a, b, var.equal = TRUE)
tt$conf.int
```

```
## [1] -17.43919  25.93919
## attr(,"conf.level")
## [1] 0.95
```

A perfect match!