Assignment 3

1. 1- We can take the difference of the two means to find a point estimate for the mean of the two samples. This gives us our point estimate: $3.1 - 2.4 = 0.7$.

2.

3. 3-

   (a) First we must check whether the standard deviations of our two samples are significantly different. Since R can only handle raw data, we do this with:

   Now we find our CI using the R code:

   ```
   > tsum.test(18.5,8.03,265,16.5,6.96,298)
   ```

   We obtain the interval:

   $$(0.7484522, 3.2515478).$$

   (b) Since this interval does not contain 0, we can say with 95% confidence that there is a difference in the means of the posttest scores.

4.

5. 5-

   (a) Begin with

   $$\begin{aligned} X &\sim N(\mu_1, \sigma^2) \\ Y &\sim N(\mu_2, 3\sigma^2). \end{aligned}$$

   And let $W = 2\bar{X} + \bar{Y}$. We want to find $M_W(t)$.

   $$\begin{aligned} M_W(t) &= E\left[\exp(2t\bar{X} + t\bar{Y})\right] \\ &= E\left[\exp(2t\bar{X})\right] E[\exp(t\bar{Y})] && \bar{X} \perp \bar{Y} \\ &= M_{\bar{X}}(2t) M_{\bar{Y}}(t). \end{aligned}$$

   And we know that:

   $$\begin{aligned} M_{\bar{X}}(2t) &= \exp\left(2t\mu_1 + (\frac{2t^2\sigma^2}{n}\right) \\ M_{\bar{Y}}(t) &= \exp\left(t\mu_2 + \frac{3t^2\sigma^2}{m}\right). \end{aligned}$$

   Since each X_i, Y_i is normal.

   $$\begin{aligned} M_W(t) &= \exp\left(2t\mu_1 + (\frac{2t^2\sigma^2}{n}\right) \exp\left(t\mu_2 + \frac{3t^2\sigma^2}{m}\right) \\ &= \left(t(2\mu_1 + \mu_2) + t^2\left(\frac{2\sigma^2}{n} + \frac{3\sigma^2}{m}\right)\right). \end{aligned}$$

   And we recognise the MGF of a normal distribution, and say $W \sim N(2\mu_1 + \mu_2, \sigma^2\left(\frac{2}{n} + \frac{1}{m}\right)$

   (b) To find our confidence interval we first notice that since W is normally distributed:

   $$\frac{W - 2\mu_1 - \mu_2}{\sigma^2\left(\frac{4}{n} + \frac{3}{m}\right)} = Z.$$

And now we begin with the probability statement:

$$\begin{aligned}
1-\alpha &= P(a < 2\mu_1 + \mu_2 < b) \\
&= P(a - W < -W + 2\mu_1 + \mu_2 < b - W) \\
&= P(W - b < W - (2\mu_1 + \mu_2) < W - a) \\
&= P\left(\frac{W-b}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)} < \frac{W-(2\mu_1+\mu_2)}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)} < \frac{W-a}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)}\right) \\
&= P\left(\frac{W-b}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)} < Z < \frac{W-a}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)}\right).
\end{aligned}$$

Which gives us:

$$\begin{aligned}
Z_{\frac{\alpha}{2}} &= \frac{W-b}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)} \\
-Z_{\frac{\alpha}{2}} &= \frac{W-a}{\sigma^2\left(\frac{4}{n}+\frac{3}{m}\right)}.
\end{aligned}$$

And after isolating for a, b and substituting $W = 2\bar{X} + \bar{Y}$, we get $(1-\alpha)100\%$ confidence interval for $2\mu_1 + \mu_2$:

$$(2\bar{X} + \bar{Y}) \pm Z_{\frac{\alpha}{2}} \sigma^2 \left(\frac{4}{n} + \frac{3}{m}\right).$$

6.

7. 7-

8.

9. 8-

10.

11. 9-

12.

13. We begin by checking whether σ_1 and σ_2 are significantly different. Running `var.test(mydata$DelaysAA,mydata$DelaysUA)` in R gives the interval:

$$(0.4946789, 1.7401896).$$

And since our interval contains 1, we can assume the variances to be equal.

Now we can run `t.test(mydata$DelaysAA,mydata$DelaysUA,var.equal=TRUE)` , which gives us the interval:

$$(-14.8649214.52444).$$

Which contains 0, so we cannot say that the sample means are significantly different.

14.