Stat 500 **Lab 5 Solutions**

1.

a. Describe the independence assumption for the two-sample inference procedure. Explain why this assumption is valid for these data.

The independence assumption states that both samples must be independent and identically distributed and the two samples must be independent from each other. Both samples were taken randomly from the target populations, so each sample should be independent. The target populations are separate, and so that samples from each should be independent.

b. Describe the homogeneous variance assumption for the two-sample inference procedure. Check whether this assumption is appropriate or not for these data. Explain why violating this assumption has little impact on our results.

The homogeneous variance assumption states that the population variances are the same.

The adult age group has a larger standard deviation than the teenager group. The ratio of the two standard deviations is $23.3959/11.7440 = 1.992158$. *This is very close to 2.*
The Folded F-test and the BF test both have a small p-value (0.0145 for the Folded-F test and 0.0077 for the BF test), meaning we would reject the null hypothesis of equal population variances.
We also see differences in the boxplots between the two groups, with the adult group having a much larger range and IQR than the teenager group.

However, violating this assumption will have minor impact on our result since the sample sizes are the same and there is a large difference in the sample means between the two groups. If we use the Satterthwaite approximation, there is a small difference in the degrees of freedom from the pooled method, but the p-values are so small, it doesn't matter. Likewise, for the confidence intervals, the t values will be a little different between the two intervals, but the overall confidence intervals are very similar.*

c. Describe the normal distribution assumption for the two-sample inference procedure. Check whether this assumption is appropriate or not for these data.

The normal distribution assumption states that the distribution of the variance in both populations must be normally distributed. The Q-Q plots for both groups are fairly close to the lines. Both histograms look fairly symmetric and bell-shaped and are similar to the normal distribution and empirical distributions as well. The tests for normality all have relatively large p-values. So the normality assumption is okay.

Next we will explore the differences in mean times when participants used the control phone. Use the output from SAS to conduct a hypothesis test for the difference in the mean times to send the text message between the adults and teenagers.

Assumptions and Diagnostics

Independence – same as above.

Homogeneous Variance – same as above. The ratio of the sample standard deviations is around 2 and the folded F-test rejects the null hypothesis of equal variances. This seems *to be due to two very large observations in the adult group. Again, the impact of violating this assumption will be small, since the sample sizes are the same and there is a large difference in the sample mean values.*

Normal Distribution – The values for the teenager group seems to be normally distributed, but the values from the adult group do not. Again, this seems to be coming from the two large values in this group, which are denoted as outliers in the boxplot and histogram. The overall impact to the results above is minimal, but it might not be a bad idea to run the analysis again without these two observations to study the differences in the results.

2. Here is data from an experiment on the effect of sodium diet levels on systolic blood pressure. Below are the data:

Low Sodium Group	High Sodium Group
143	161
136	161
126	179
165	176
145	162
135	155
143	172
157	156
132	146
149	162

Perform a Wilcoxon rank-sum test to determine if the distributions of systolic blood pressures are the same for the two groups. Calculate the test statistic W and find a corresponding p-value using the low sodium group. The SAS code can be found in the file **bloodpressure1.sas**.

Here is the data table again with the ranks for the observations in the two groups.

Low Sodium Group		*High Sodium Group*	
Observed Value	*Rank*	*Observed Value*	*Rank*
143	*5.5*	*161*	*13.5*
136	*4*	*161*	*13.5*
126	*1*	*179*	*20*
165	*17*	*176*	*19*
145	*7*	*162*	*15.5*
135	*3*	*155*	*10*
143	*5.5*	*172*	*18*
157	*12*	*156*	*11*
132	*2*	*146*	*8*
149	*9*	*162*	*15.5*

Summing the ranks for the Low Sodium Group gives the test statistic $W = 66$. The mean and expected value of W under the null hypothesis of equal distributions are

$$E_0(W) = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{10(21)}{2} = 105$$

$$V_0(W) = \frac{n_1 n_2(n_1 + n_2 + 1)}{12} = \frac{100 * 21}{12} = 175$$

The large sample z-test statistic is

$$Z = \frac{W - E_0(W) - 0.5}{\sqrt{V_0(W)}} = \frac{|66 - 105| - 0.5}{\sqrt{175}} = 2.91$$

The p-value is $2 * P(Z > 2.91) = 0.0036$.