Author: Brenda Gunderson, Ph.D., 2015

Lab 8: Comparing Two Means

Objective:
In this lab, you will learn an important statistical technique that will allow you to compare two populations with respect to their means by looking at μ_1-μ_2. The aim is to help you understand the ideas behind confidence intervals, tests of significance, and the statistical language involved in the comparison of two population means.

Application:
Jim believes that the mean GPA for all English majors is different from the mean GPA for all Math majors. He takes a random sample of students in the Math and English buildings and he ends up with 32 English majors and 35 Math majors in his sample. He has more Math majors, which leaves Jim with no method to pair an English major with Math major and no way to perform the paired t test.

Overview:
The two independent samples *t* procedures (sometimes called Student's *t* procedures) are used when you want to compare the means of two populations that are not related or matched in any way. The idea is that we use the two sample means to estimate the corresponding population means.

We may want to construct a confidence interval to estimate the difference between the two population means. Or, we may wish to test if the difference in the two population means equals a value specified in a null hypothesis; generally, the hypothesized test value is 0.

Going back to the application, we will assume that we have a random sample from each of the two populations. This means Jim will assume that he has a random sample of English majors and a random sample of Math majors. To check this assumption, he will create a time plot with the GPAs for the English majors and assess the stability as well as creating a separate time plot with the GPAs for the Math majors.

The second assumption needed to perform the two independent sample t-test is that the variable being measured has a normal model for each population, although possibly with different means. To assess this assumption, Jim will create two separate QQ plots; one of the GPAs for English majors and one of the GPAs for Math majors.

In addition, the two random samples are assumed independent of each other. However, there are two versions of independent samples *t* procedures: pooled and unpooled (also known as general). In a pooled *t* procedure, we assume equal population variances for the two populations of responses.

Whether or not we can assume the two populations have equal variances will dictate which of the two procedures we use. Details on determining which version to use are provided next.

Checking the Equal Population Variance Assumption: There are several ways to check the equal population variances assumption required for a pooled *t*-test. Which one you will use will often depend on the information provided.

- **Side-by-side Boxplots:** Examine the IQRs of the sample data. If they are similar, then the assumption is valid. (NOTE: "Similar" in this sense does not mean that the boxes need to line up right next to each other. It means that the lengths or sizes of the boxes should be similar.) If one IQR is twice as large as the other, the assumption of equal population variances is not valid.

- **Sample Standard Deviations:** Since variance is standard deviation squared, if the sample standard deviations are similar, then the assumption is valid. If one standard deviation is twice as large as the other, the assumption of equal population variances doesn't hold.

- **Levene's Test for Homogeneity of Variances:** Levene's test is a hypothesis test, but it is a test about population variances rather than means. In Jim's case the null hypothesis is that the populations of English and Math majors do have equal variances (H_0: $\sigma_1^2 = \sigma_2^2$) and the alternative is that these populations have equal variances (H_a: $\sigma_1^2 \neq \sigma_2^2$). When the null hypothesis from the Levene's test is **NOT** rejected, it would be reasonable for Jim to perform the pooled *t*-test for the two population mean GPAs. If the null hypothesis from Levene's test **IS** rejected, then the assumption of equal population variances is unreasonable, and Jim should use the unpooled (general) version instead of the pooled. To make this determination, we will use an alpha of 10% for Levene's tests.

Common Population Standard Deviation: If Jim concludes that the populations of Math and English majors have the same variances, he is also stating that these populations have the same standard deviation. In order to run the pooled version of this test, he must estimate the common population standard deviation with the common sample standard deviation. This common standard deviation is written as s_p and Jim will calculate it by taking a weighted average of the standard deviation of English majors' GPAs and the standard deviation of Math majors' GPAs.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Formula Card

Two Population Means	
General	Pooled
Parameter $\mu_1 - \mu_2$	**Parameter** $\mu_1 - \mu_2$
Statistic $\bar{x}_1 - \bar{x}_2$	**Statistic** $\bar{x}_1 - \bar{x}_2$
Standard Error $\text{s.e.}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	**Standard Error** pooled $\text{s.e.}(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$
Confidence Interval $(\bar{x}_1 - \bar{x}_2) \pm t^*(\text{s.e.}(\bar{x}_1 - \bar{x}_2))$ df $= \min(n_1-1, n_2-1)$	**Confidence Interval** $(\bar{x}_1 - \bar{x}_2) \pm t^*(\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2))$ df $= n_1+n_2-2$
Two-Sample *t*-Test $t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\text{s.e.}(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ df $= \min(n_1-1, n_2-1)$	**Pooled Two-Sample *t*-Test** $t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ df $= n_1+n_2-2$

Warm-Up: Check Your Understanding

A student researcher, Jackie, is interested in learning about how much time students spend studying per week. She is curious to find out if study time is higher for female students (group 1) versus male students (group 2). Data from a survey of 47 female and 45 male undergraduate students is collected. One question asked students to report the number of hours that they study in a typical week. Assume that all the needed conditions are met to perform a pooled two independent samples t-Test.

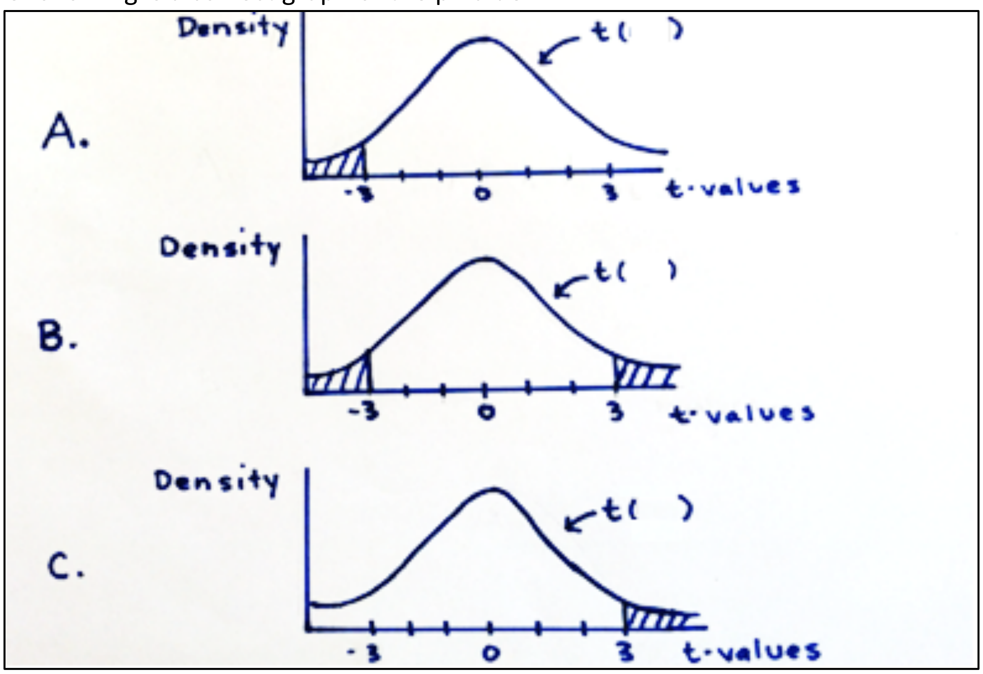
State the hypotheses to be tested and provide the test statistic value.

H_0: ______________________ versus H_a: ______________________

The test was run and the difference in the two sample means, $\bar{x}_1 - \bar{x}_2$ has been computed. This difference turned out to be 3 standard errors *above* the hypothesized difference for $\mu_1 - \mu_2$ of 0.

Test Statistic Value: ____________________

Which of the following is a correct graph of the p-value?



If the *p*-value for this test was 0.076,
are the results significant at a 5% level? **Yes** **No**

What are the appropriate degrees of freedom for the pooled t test? **91** **43** **90**

ILP: Do Men and Women Differ in Their SSHA Scores?

Background: A total of 38 college freshmen at a private college were administered the Survey of Study Habits and Attitudes (SSHA), a psychological test designed to measure motivation and attitude towards study habits in college students. The sampled students were a simple random sample consisting of 18 females and 20 males. Scores on the test range from a low of 0 to a high of 200 and it is known that they may explain collegiate success. School administrators are interested in whether or not there is a difference between the mean scores for males and females. The scores for females (Group 1) and males (Group 2) are listed in the **SSHA.Rdata** data set. (Source: Moore and McCabe (1999), pg 563).

***Task*:** Perform a test to assess if there is a difference between the mean score for women on the SSHA and the mean score for men on the SSHA.

Procedure: Since the SSHA score is a quantitative response and we have two sets of scores to compare, the appropriate inference procedure for this scenario is the two independent samples t-test, and the specific parameter of interest is ___________. Think about why is this not a paired procedure.

Hypothesis Test

1. **State the Hypotheses:** H_0: ______________________ versus H_a: ________________________ ,

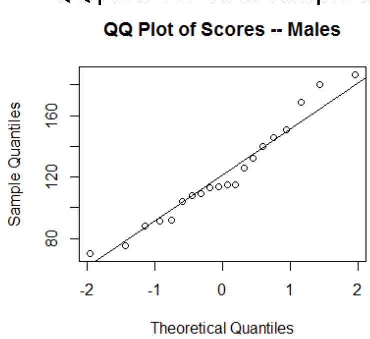
   where μ_1 represents____

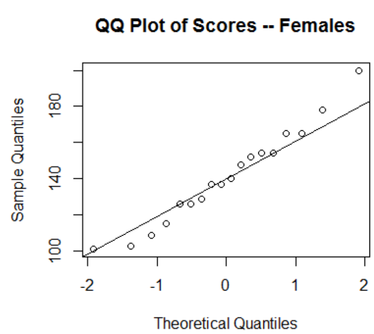
   and μ_2 represents____

   Note: In order to state the appropriate direction in the alternative hypothesis, it is important to know which population is being referred to as population 1 and which is population 2.
   Determine Alpha: We were told the significance level was 5%.

   > **Remember:** Your hypotheses and parameter definition should always be a statement about the **population(s)** under study.

2. **Check Assumptions (Exploratory Data Analysis)**
   a. Checking Normality
      i. Based on the information provided in the background of the problem, we will assume that we have independent, random samples of SSHA scores. One remaining assumption that needs to be checked is that each sample comes from a normally distributed population. The QQ plots for each sample are provided next.





      ii. Does it appear that the assumption that each sample comes from a normally distributed population is met? Why?

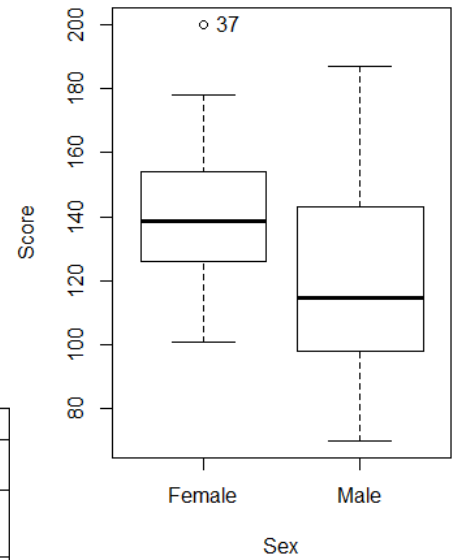
      iii. What remaining assumption needs to be checked before we can conduct the two independent samples t-test?

b. Now, we must determine if we can assume equal population variances or not. We will utilize all three methods discussed earlier to do this.

i. *Side-by-side Boxplots:*
The side-by-side boxplots show that the IQRs are:

similar **not similar**

ii. *Sample Standard Deviations:* Remember that we can generate summary statistics by using **Statistics > Summaries > Numerical Summaries,** and be sure to **summarize by groups** to get results for both sexes.



Summary Statistics			
Group	**Mean**	**Std. Dev**	**Sample Size**
Male			
Female			

These sample standard deviations are: **similar** **not simi**

iii. *Levene's Test:* Write the appropriate hypotheses for Levene's test using correct notation:

H_0: ____________________________ H_a: _________________________

We can perform Levine's test in RCommander using **Statistics > Variances > Levene's Test** (be sure to **use MEANS** and not MEDIANS).

The Levene's test statistic is F = _____________ , with a *p*-value of _______________.

Therefore, we **can** **cannot** reject the hypothesis that the population variances are equal.

Based on these results, we can say:
the assumption of equal population variances **is** **is not** valid.

Thus, the procedure that we will use for this test is the **pooled** **unpooled** procedure.

The symbol for the estimate of the common population standard deviation is ____.

Calculate its value:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Next, use the estimate s_p to compute the pooled standard error:

$$\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

3. Compute the Test-Statistic and Calculate the *p*-value

Test Statistic and P-value

a. Generate the *t*-test output using **Statistics > Means > Independent-Samples T Test**. Make sure to choose the appropriate direction for this test. Recall, we want to assess **if there is a difference** between the mean score for women on the SSHA and the mean score for men on the SSHA. Go to the **Options** tab, and you will find an **"Allow equal variances?"** setting. We have already looked at the data in various ways to make our decision of a **Yes** (pooled) setting. Using this output, fill out the table below.

Two Sample T Results					
	t	***df***	***p-value***	***95% CI lower***	***95% CI upper***
Pooled					

b. What is the distribution *of the test statistic* if the null hypothesis is true?

Note: This is not the same as the distribution *of the population that the data were drawn from*, and will be the model used to find the *p*-value.

Visualize the *p*-value:

c. Draw a picture of the *p*-value,
with labels for the distribution and x-axis.

d. Provide an interpretation of the *p*-value.

e. Think about it … how would you report the *p*-value for this test if you were trying to determine if males' mean SSHA score is larger than the females SSHA score on average? (Hint: You would need to be careful on the order of the subtraction!)

4. **Evaluate the *p*-value and Conclusion**
Evaluate the *p*-value:
What is your decision at a 5% significance level? **Reject H_0** **Fail to Reject H_0**

Remember:	Reject H_0	⇔ Results statistically significant
	Fail to Reject H_0	⇔ Results not statistically significant

Conclusion:
What is your conclusion *in the context of the problem*?

Note: Conclusions should always include a reference to the population parameter of interest.
Conclusions should not be too strong; you can say that you have sufficient evidence,
but do NOT say that we have *proven* anything true or false.

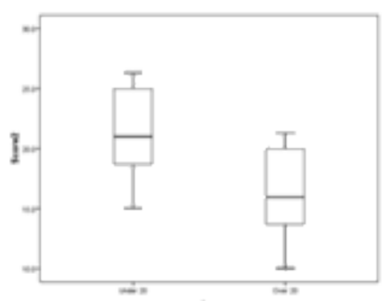
5. **Confidence Interval (CI):**
   a. Provide the corresponding 95% confidence interval from the output for the difference in the two population mean scores.

   a. Based on the confidence interval, would you reject the null hypothesis of no difference in population means at a 5% significance level?
   Circle one: **Yes** **No**
   Explain.

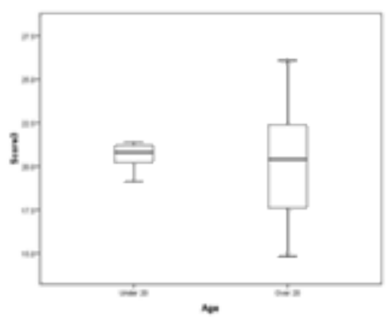
   Did your conclusion here match the one you made in the previous part 4? **Yes** **No**

Cool-Down: Check Your Understanding

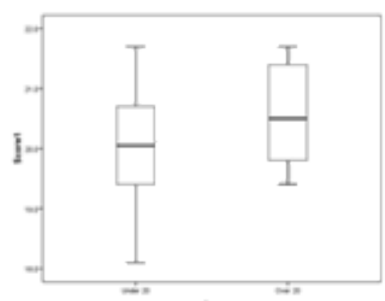
Consider the following sets of boxplots of scores between two age groups.



Set 1



Set 2



Set 3

a. Which set(s) that indicate a pooled test is appropriate: **Set 1** **Set 2** **Set 3**

b. For which set(s) are you most likely to reject the null hypothesis that the population mean scores are equal? **Set 1** **Set 2** **Set 3**