Lab 7: Comparing two means & Violations of assumptions

Lab Objectives:

♦ Compare the means of two groups
♦ Understand paired vs. two-sample designs
♦ Explore data transformation
♦ Detect violations of assumptions and use SPSS for nonparametric alternatives to one-sample, paired and two-sample t-tests

Exercise 1: Paired t-test

When comparing two means, if each data point in one group is uniquely paired with a data point in the other group, then the data are said to be paired. A **paired *t*-test** examines the mean difference between the members of each pair. Paired *t*-tests assume that the differences are normally distributed and that the pairs are randomly sampled.

A paired sampling design is usually more powerful than an unpaired design, as it reduces extraneous variation between sampling units. However, this sampling method is not always possible. When data are not paired, we compare the means of two groups using a **two-sample *t*-test**. The two-sample *t*-test assumes that there is an independent random sample from each group, and that the variable is normally distributed in each population with the same variance.

Welch's *t*-test is similar to a two-sample *t*-test, but does not require the assumption of equal variances. As you will see, the data output from a two-sample t-test in SPSS automatically includes the results of Levene's test for equal variances, the two-sample t-test, and Welch's approximate t-test.

Biological studies often examine the same patients before and after a particular treatment. These types of studies can be analyzed using a paired t-test. Let's begin with an example of weight gain in anorexia patients as a function of family therapy. We want to see if the difference in weight before and after a family therapy intervention is significantly different from zero.

Anorexia therapy

▪ Open the file *anorexia family therapy.sav*

▪ Select analyze, compare means, paired samples t-test.

▪ Drag BEFORE to Pair 1, Variable 1. Drag AFTER to Pair 1, Variable 2, as shown below. If we had additional measurements that we wanted to compare before and after treatment we could add these to pair 2 (e.g. depression before and after treatment).
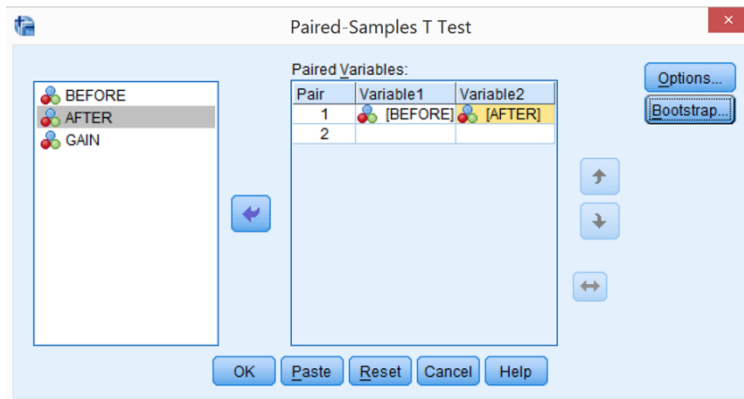
- Examine the output. The first table contains summary statistics, including mean, n, SD, and SEM for weight before and after therapy.

- The second table displays the correlation statistic. We haven't covered correlation in class yet; however it is fairly easy to interpret. The correlation coefficient is a measure of the strength and direction of a linear relationship between the two variables. The correlation coefficient can range from -1 to +1, with -1 indicating a perfect negative correlation, and +1 indicating a perfect positive correlation, and 0 indicating no correlation at all. This data shows a correlation of 0.538, indicating that the two variables are positively correlated, but not perfectly. 0.538 is a numerical description of how tightly around an imaginary linear line the points lie. Even though it is not a perfect correlation, you can see that the correlation is significant, with a P-value of 0.026.

- The third table contains the results of the paired samples t-test, which is what we are interested in. The mean difference is negative because it was calculated by subtracting the weight after therapy from the weight before. The 95% confidence interval of the difference is indicated, and shows that we are confident that there is a weight gain with therapy, as both the lower and upper limits are negative values. The last section of the table includes the t-statistic, the df, and the significance (P-value) for a two-tailed test.

- Remember from class that a paired t-test is identical to a one-sample t-test on the differences. Let's check this using the data in the gain column

- Select analyze, compare means, one sample t-test.

- Select Gain as your test variable, and use a test value of 0, as you are testing the null hypothesis that weight does not change with family therapy.

- Click OK and look at the t-statistic and P-value.

➢ Answer question 1 on Sakai.

Exercise 2: Two-sample t-test

Often in Biology, you will want to compare means of two independent samples. In this case each measurement is not directly paired with another measurement. Instead we have several measurements in each group, and we want to compare the mean of one group to the other.

Spinocerebellar ataxia

The data in the file *ataxia.sav* are from an experiment on a mouse model of the neurodegenerative disease Spinocerebellar ataxia type 1. Researchers examined whether life span differed between mice with no exercise, compared with mice with a mild exercise regime.

▪ Open the file *ataxia.sav*

▪ Select analyze, compare means, independent samples t-test.

▪ In this case the test variable is lifespan and the grouping variable is treatment. Add both to the appropriate box. This ensures we will be comparing the mean lifespan (test variable) between the two treatment groups.

▪ After adding the grouping variable click on the button "define groups".

▪ In this file the treatment groups are assigned numeric values of 0 or 1, and the values are then labelled no exercise or exercise. We will therefore need to add 0 to the group 1 box, and 1 to the group 2 box, as shown below. Going forward, if you are unsure about what numeric values are assigned to your groups, consult the variable view.
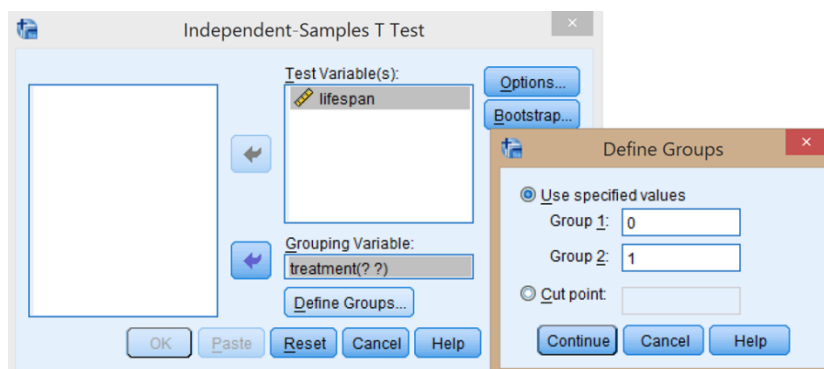
▪ Click OK to run the test.

▪ In the output file, the first table shows summary statistics for each group.

▪ The second table shows the results of the comparison of means. The first thing you should always examine is the result for **Levene's test** for equality of variances. Remember that an important assumption of a two-sample t-test is that the two groups have equal variances. Levene's test checks this. If the results of Levene's test are significant (P<0.05), this indicates that the two groups **DO NOT** have equal variances. In this case we need to use the t-test result for "equal variances not assumed". This row displays the result for **Welch's approximate t-test**, which does not assume equal variances.

▪ If the results of Levene's test are not significant (P>0.05), then we can assume that the variances are equal and can use the t-test data from that row. The P-value can be found under "Sig (2-tailed)".

		Independent Samples Test				
		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
Use if Levene's test is not significant (P >0.05)	lifespan Equal variances assumed	1.900	.198	-3.153	10	.010
Use if Levene's test is significant (P<0.05)	Equal variances not assumed			-3.153	7.407	.015

➢ Answer question 2 on Sakai.

Exercise 3: Choose the right T-test

Homophobia

In a paper published in 1996, Adams *et al* examined whether homophobia is associated with homosexual arousal. Men were assigned to either the homophobic or nonhomophobic group based on their scores on a Homophobia Index test. The men then watched sexually explicit erotic homosexual videotapes, and an arousal score was determined. These data are found in the file *homophobia.sav*. Conduct a statistical test to determine if there is any difference in arousal between the two groups.

➢ Answer question 3 on Sakai.

NZ Noggins

The head sizes of 18 New Zealand army recruits were measured with both cardboard and metal calipers. Researchers want to determine whether there is a significant difference between head size when measured with these two methods. The data is found in *NZnoggins.sav*. Open *NZnoggins.sav* and conduct an appropriate test to determine if there is a difference in head size when measured with cardboard and metal calipers.

➢ Answer question 4 on Sakai.

Beer & mosquitos

A study in West Africa on mosquitos that cause malaria examined whether drinking beer influenced attractiveness to mosquitos. Mosquitos were released next to 25 alcohol-free participants, and the proportion of mosquitos flying toward the participants was measured. The procedure was repeated 15 minutes after each of the same participants had consumed a liter of beer. Open *mosquitos beer.sav* and conduct an appropriate test to determine if there is a change in attractiveness to mosquitos following drinking beer.

➢ Answer question 5 on Sakai.

More beer & mosquitos

The West African study was expanded by including a second control group of 18 participants who were given water instead of beer. The change in activation of mosquitos was measured in both the

group of 25 participants who drank beer, and the 18 participants who drank water. Open *mosquitos beer water.sav* and conduct a statistical test to examine the difference in the change in mosquito activation between the beer group and the water group.

➢ Answer question 6 on Sakai.

Exercise 4: Violations of assumptions and nonparametric tests

Many parametric tests, including *t*-tests, are robust to their assumptions. This means that the test works well even when the assumptions are violated. For example, with large sample sizes, the two-sample *t*-test works acceptably even if the two groups' standard deviations differ by threefold. However, using a parametric test when the assumptions are not met increases the type I error rate (probability of rejecting a true null hypothesis). If the assumptions are violated too severely to ignore, data can sometimes be transformed to make new variables that better match the assumptions of the methods. For example, with the log-transformation, the natural logarithm (Ln) of the value of a variable is computed, and is used in parametric tests instead of the variable itself.

If the assumptions are violated and data transformation doesn't produce data that better meet the assumptions, then nonparametric tests can be used. These include the **Mann-Whitney U test**, which examines the distributions of two independent groups, and the **Sign test**, which can be used in place of both a one-sample t-test and a paired samples t-test. Nonparametric tests are not as powerful as parametric tests. That is, they are less likely to reject a false null hypothesis (higher type II error). It is therefore better to use parametric tests if the assumptions can be met, and turn to nonparametric methods only after data transformation has failed to meet the assumptions of the parametric methods.

A **Shapiro-Wilk** test is a formal method for testing for departures from normality. A P-value of less than 0.05 indicates that the data are not a good fit for the normal distribution. Formal normality tests should be used with caution, as small samples may not have enough power to reject normality, and large samples may reject normality even when the departure is not severe to give up on methods that assume normality.

Biology & sports

A group of French researchers was interested in whether people who play a lot of sports have more sexual partners than those who do not. They asked two groups of students how many sexual partners they had in the previous year (Faurie et al 2004). One group was composed of 103 physical education majors who regularly participated in sports, and the other group was composed of 100 biology majors who did not regularly participate in sports. The data is found in the file *Biology sports.sav*

▪ Open the file *Biology sports.sav*

▪ Select analyze, descriptive statistics, **explore**.

▪ Add major to the Factor List and NumberSexPartners to the dependent list.

- Click on Plots. Uncheck Stem-and-leaf. Check the boxes for "histogram", and "normality plots with tests". Click continue.

- Click OK.

- You should see a table of descriptive statistics for the two groups, indicating the mean, 95% confidence interval, etc.

- If you scroll down, the next table displays the results for two tests of normality. We are interested in the results of the **Shapiro-Wilk** test. If the Sig. value of this test is less than 0.05, then the data are not a good fit for the normal distribution.

- Scroll down to visualize the two histograms of the data. You should be able to see that the data are not normally distributed. The important consideration for statistical tests is whether the variable is normally distributed in the *population*. However, random samples that are very skewed are often reflective of a non-normal population distribution.

- Finally, look at the two Normal Q-Q plots. These plots are another method for detecting deviations from normality. If the data are normally distributed, the data points should fall roughly along a straight line. The plots include a straight line as a guide.

- Now that we know the data is not normal, let's try a data transformation. Click Transform, **compute variable**.

- In the Target variable box, write "LogTransform"

- Under function group, select "Arithmetic"

- In the lower box, select Ln, then click the up arrow to move this function into the "Numeric expression" box.

- Select "NumberSexPartners" and click the right arrow to move it into the Ln function

- Since the data include zeros, we will try log transforming the variable + 1. Simply add a +1 in the numeric expression as follows.
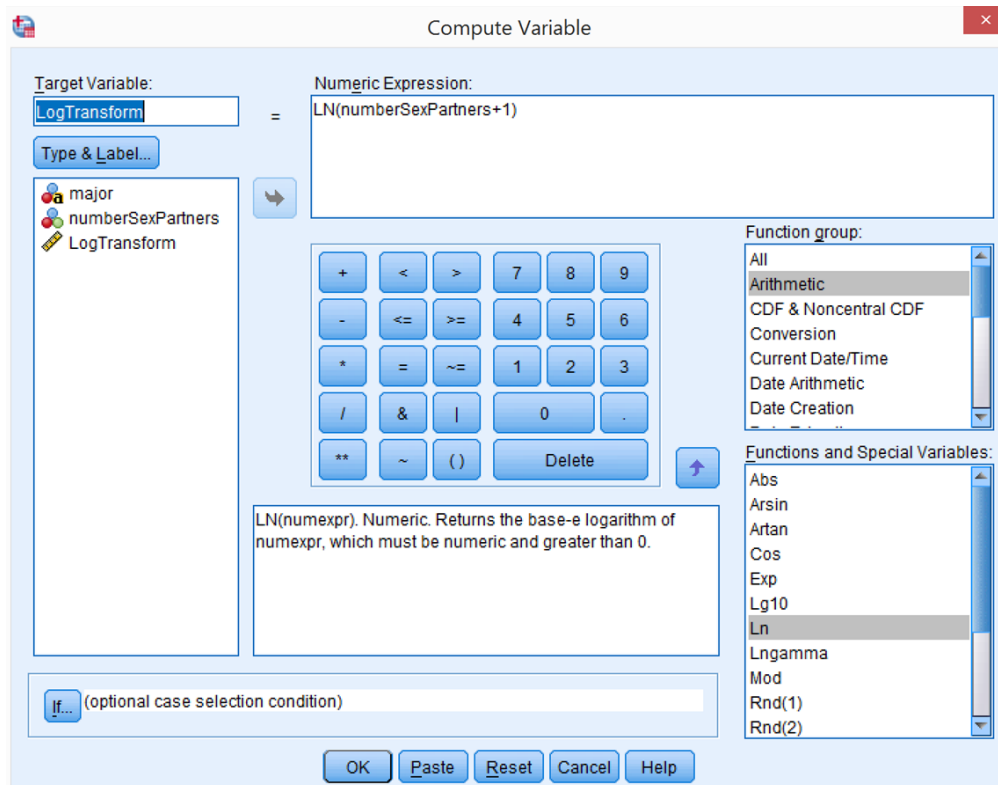
- Click OK to run the data transformation. The Log transformed data will be visible in a new column on your data view.

- Another common transformation for data that are skewed right is a reciprocal transformation. To compute this, click transform, compute variable again. In the target variable box write "Reciprocal". In the numeric expression box, adjust the formula so that it equals 1/(numberSexPartners+1). Click OK.

- Finally, let's try the square root transformation. Click transform, compute variable. In the target variable box write "SQRT". In the numeric expression box, adjust the formula so that it equals SQRT(numberSexPartners+0.5). You can find the Sqrt function under "Arithmetic", or you can simply type it in yourself.

- Now let's check to see if any of these transformations improved the normality of our data. Select Analyze, descriptive statistics, explore.

- Your settings from the last time you ran this test should still be in place. Add LogTransform, Reciprocal and SQRT to the Dependent list and click OK.

- In your output, you will see the summary statistics, followed by a table containing the results of the Shapiro-Wilks test for normality. This will be followed by the graphical outputs for each of the variables. Examine these results to determine if any of the transformations improved the fit of the data to a normal distribution.

- When a deviation from normality is too strong to ignore, and transformation does not meet the assumptions of a two-sample t-test (normal distribution and equal standard deviation), the next option is to try a non-parametric test.

- Select analyze, nonparametric tests, legacy dialogs, two-independent samples.

- Add Number of sex partners to the test variable list.

- Add major to the grouping variable. Define groups as 0 and 1.

- Ensure the box is checked for "**Mann-Whitney U**". Click OK.

- The output of the test statistics gives you the Mann-Whitney U value, a related statistic called Wilcoxon W, the equivalent Z value from the standard normal distribution, and the P-value (Asymp. Sig. 2-tailed). Note that SPSS uses the smaller U value (rather than the larger U value as we will do in class). When using the smaller U value, a different critical table is used. With smaller sample sizes, an additional P-value for exact significance will also be provided. If this value is provided, you should use it when considering whether or not to reject the null hypothesis.

- If the significance value is less than 0.05, then we can conclude that the distribution of the variable (number of sex partners) is not the same between the two majors. If both distributions have the same shape (in this case they do, both are right-skewed) then we can also conclude that the two groups do not have the same median.

➢ Answer question 7 on Sakai.

Health expenditure

Health expenditure per person from a random sample of 20 countries is found in the file *health expenditure.sav*. You suspect that this data is not normally distributed. Use SPSS to conduct a natural log transformation of this data, and conduct a Shapiro-Wilk test on both the untransformed and transformed data to determine if this improves the distribution.

➢ Answer question 8 on Sakai.

Killer lions

When an intruding male lions take over a pride of females, they often kill most or all of the infants in the group. Following takeover, the stability of the pride is unpredictable, and scientists hypothesized that females may delay ovulation until the uncertainty has passed. To examine this, a study of lions in the Serengeti measured the time to reproduction of female lions after losing cubs to infanticide, compared to the time to reproduction after losing cubs to accidental death. The data are not normally distributed, even following transformation. Conduct an appropriate statistical test on the data in the file *Lions.sav* to determine if there is a significant difference in reproduction time.

➢ Answer question 9 on Sakai.

Parasitic ducks

Female goldeneye ducks lay eggs in other females' nests, in addition to the eggs they produce and raise in their own nests. Females lay just one egg per day, so the total time span for all eggs in one nest to hatch can be affected by whether the eggs were laid early or late. To determine whether a female tends to lay her earliest eggs in her own nest, Andersson and Ahlund (2012) measured eggs produced by 14 female goldeneyes and assigned each a parasitism index. The index is negative if she tends to lay her last eggs in another female's nest, and positive if she tends to lay her last eggs in her own nest. An index of zero indicates that she alternates laying eggs in her nest and the other female's nest. This data is not normally distributed, and transformation does not make it normal.

To determine whether the parasitism index is significantly different from zero, we could conduct a **one-sample sign test**. Unfortunately, SPSS does not have a procedure option specifically for a one-sample sign test. However, there is a fairly easy workaround.

- Open the file *parasitic ducks.sav*

- We will add a new variable for the null hypothesis. Name the new variable "Null", and enter a value of "0" for each female, as we are testing our values against the null hypothesis that the parasitism index is zero.

- Under variable view, change null to a scale variable.

- Select analyze, nonparametric tests, legacy dialogs, two-related samples.

- Add Null to variable 1, and Parasitism Index to variable 2.

- Uncheck Wilcoxon. Check the box for "Sign" instead. Click OK.

- The first output box will tell you the number of ducks with a parasitism index less than the null value (zero), the number with a parasitism index greater than zero, and the number of ties.

- In this case, 12 ducks had an index greater than zero. Is this enough to reject the null hypothesis? Examine the test-statistic box, which provides the P-value. If this is less than 0.05, we reject the null hypothesis and conclude that the parasitism index is significantly different from zero.

➤ Answer question 10 on Sakai.

Incompatible mates

There are two genetically distinct types of Gouldian finches, one having a red face and the other having a black face. Previous experiments have shown that females have a strong preference for mating with males with the same colour face as themselves, and that the offspring are more likely to survive when both parents are the same type. To examine stress caused by having an incompatible mate, researchers measured corticosterone concentration in 43 females following breeding with a compatible male (same colour face), and in the same 43 females following breeding with an incompatible male (different colour face). The data are not normally distributed. We will therefore conduct a **paired-samples sign test**.

- Open the file *incompatible mates.sav*

- Select analyze, nonparametric tests, legacy dialogs, two-related samples.

- Add the two variables to the first test pair.

- Uncheck Wilcoxon and check the box for "Sign" instead.

- Consult the output to determine if there is a significant difference in the amount of corticosterone produced in females with incompatible vs compatible mates.

➢ Answer question 11 on Sakai.

Exercise 5: Additional Practice

Contraception prevalence

The data in the file contraceptive prevalence provides the percentage of married women ages 15 – 49 who are practicing, or whose sexual partners are practicing, any form of contraception. Data have been compiled from World Bank for 57 countries for the year 2002 and the year 2012. You are interested in analyzing the data to determine if there is a difference in contraception between these two years.

First, analyze the data to determine whether it violates the assumptions of the t-test that would apply to this data.

Then, analyze the data using the appropriate statistical test to determine whether there is a change in contraception prevalence between 2002 and 2012.

➢ Answer questions 12 & 13 on Sakai.

Have you imported your research project data into SPSS yet?

If not, today would be a great time to get started. Stick around and give it a try while you have a TA who can offer some help ☺

Projects are due in **2 weeks!**