

# BIOL 458 Biometry

## Lab 4: HYPOTHESIS TESTS USING SPSS

---

### OBJECTIVE

The objective of this lab is to introduce you to simple hypothesis tests designed to test for differences in the location or central tendency between an experimental group and a theoretical value or between two experimental groups, and how to perform these tests using **SPSS**. Ordinarily, you would choose which test to apply to your data before you collect your data, or at least before you examine your data, and apply only that test. However, in this lab we will be using a number of different statistical test procedures to test the same null hypothesis on each data set. This will permit us to compare these different tests and to learn the assumptions involved in applying each test, when to apply them, and how to interpret the results.

### Hypothesis Tests

The **Central Limit Theorem** states that the sampling distribution of the sample mean,  $\bar{x}$ , for random samples of size  $n$  is approximately normally distributed with its mean equal to the underlying population mean,  $\mu$ , and its standard deviation is equal to  $\sigma/\sqrt{n}$ . Furthermore, as  $n$  increases this approximation improves. Hence, the sample mean,  $\bar{x}$ , provides a point estimate of the population mean ( $\mu$ ), and the sample standard deviation,  $s$ , provides a point estimate of the population standard deviation ( $\sigma$ ). Therefore, in the course of acquiring an independent random sample of observations from an underlying population and calculating the sample mean and standard deviation, one is estimating population parameters. If one is willing to assume that the underlying population is normally distributed, one also is estimating the parameters of a normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where,  $f(x)$  is the probability density of the random variable  $x$ .

Observations may be transformed from their raw units to standard deviation units using a  $z$ -transformation,  $z = (x - \bar{x})/s$ , so that one may use the standard normal distribution (a normal distribution with mean equal to zero and standard deviation equal to one) to determine if a particular observation falls above or below a specific percentile of the distribution. The sample mean,  $\bar{x}$ , is also a random variable and can be transformed in an analogous way  $z = (x - \mu)/(\sigma/\sqrt{n})$ , that is subtracting its mean and dividing by its standard deviation. If one knows the value of  $\sigma$ , the difference between the sample mean and an hypothesized value of the population mean,  $\mu$ , divided by  $\sigma$ , follows the

standard normal curve when the null hypothesis of no difference between the sample mean and hypothesized population mean is true. Of course this is only correct if we know the value of  $\sigma$ .

When we use the sample standard deviation,  $s$ , as an estimate of  $\sigma$ , the distribution of  $(\bar{x} - \mu) / (s / \sqrt{n})$  follows Student's  $t$ -distribution which is a distribution whose shape depends on its parameter, the degrees of freedom (df). For the test just described,  $df = n - 1$ , where  $n$  is the sample size. Degrees of freedom are the number of freely variable observations in the data. Given that the sample data are being used to estimate the sample standard deviation, one freely variable observation is lost from the data since given  $s$  and any  $n - 1$  observations, the  $n$ th observation is fixed. **Student's  $t$  distribution** is very similar to the  $z$ -distribution, or standard normal distribution. In particular, both distributions are symmetric, uni-modal, and have a mean of 0. As the sample size,  $n$ , increases, the shape of Student's  $t$  distribution converges on that of the standard normal distribution.

## BACKGROUND ON STATISTICAL TESTING

It is not possible to develop tests that are absolutely conclusive. All of the tests have a possibility of two kinds of error: rejecting a null hypothesis when it is not false or failing to reject a false null hypothesis. These errors are called **Type I and Type II errors**, respectively. The probability of a Type I error is denoted by  $\alpha$  and the probability of a Type II error by  $\beta$ . The **significance level** is defined as  $\alpha \times 100$  (in percent). In testing hypotheses, the probability of a Type I error can be specified; however, the probability of a Type II error can only be calculated when  $\alpha$ ,  $s$ ,  $n$ , and the "effect size" are specified in advance. In testing a hypothesis, we would ideally like the value of  $\alpha$  to be small and the power of the test ( $1 - \beta$ ) to be large.

When choosing which statistical test to apply to your data, you must consider several things. Selecting which test to use basically comes down to the purpose of the test and the consequences of making an error. It depends on the question being addressed, and the assumptions required of the statistical tests. Both the choices of the type of test to be used and the significance level depend on the problem at hand. The level of significance is usually chosen to be 0.10, 0.05, or 0.01.

There are two major types of statistical tests: parametric and non-parametric. Parametric tests make the assumption that the distribution of the underlying population is normal and this must be true for the results of the tests to be exact. However, even when the underlying population is non-normal, the sampling distribution of statistics based on the sample mean and standard deviation are approximately normal, so these tests still perform reliably. On the other hand non-parametric tests make no assumption about the shape of the underlying population. If the underlying population is normal then parametric tests are more powerful than the analogous non-parametric test.

Most **parametric tests** have 3 basic assumptions (although you should be certain about the exact nature of the assumptions for each specific test):

1. Independence - observations are collected independently and at random
2. Normality - underlying population are approximately normally distributed
3. Homogeneity of variances - equality of variances (homoscedasticity)

Alternatively, most non-parametric tests assume only that:

1. Independence - observations are collected independently and at random
2. Continuity - observations come from the same underlying continuous population.

As you can see it is always absolutely essential that observations are collected independently and at random for any test. The remaining assumptions for parametric tests are more stringent than for non-parametric tests. In general if the parametric assumptions are valid for your population(s) then these tests should be performed, since they are generally more powerful than the non-parametric tests. If you are uncertain or have reason to believe that the parametric assumptions do not hold for the population(s) under study, then the non-parametric tests may be more powerful and more reliable. The choice between these two basic types of tests is often made after a preliminary examination of the data or on the basis of previously acquired knowledge about the populations (*i.e.*, from the literature or other preliminary data).

## SOME STATISTICAL TESTS

### 1. Testing for differences between the sample mean and a theoretical value.

For situations where we know the value of  $\sigma$ , a test for differences between the sample mean and a theoretical value is based on the standard normal distribution,  $z$ . For a sample from which we use  $s$  to estimate  $\sigma$ , a test for differences between the sample mean and a theoretical value is based on Student's -  $t$  distribution. This is the test described and introduced above. Confidence intervals on the population mean,  $\mu$ , can also be derived from these statistics.

When  $\sigma$  is known:

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

so, the probability that the mean of the underlying population differs from a specific theoretical value by chance alone can be determined from the percentiles of the standard normal curve,  $z$ .

When  $s$  is used to estimate  $\sigma$ :

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

is the appropriate test statistic and it follows Student's -  $t$  distribution with degrees of freedom,  $df = n - 1$ .

## 2. Testing the differences between two sample means.

There are many times that you will want to test the differences between means. For example, you may wonder if two water samples came from the same stream, or if there is a difference between the amounts of time it takes men versus women to finish identical tasks.

### t-TEST

Given the assumption that both populations sampled have normal distributions, any hypothesis about a difference can be tested using the  $t$  distribution, regardless of sample size. However, one additional assumption becomes necessary: in order to use the  $t$  distribution for tests based on two (or more) samples, one must assume that the variances of both populations are equal. A pooled variance estimate ( $s_p$ ), an estimate that represents the weighted average of the two sample estimates of the underlying population variance, is used if the variances in the two normal populations under study are equal:

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

The  $t$  - statistic would then be:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with  $df = (n_1 + n_2) - 2$ .

If one is unwilling to assume that the variances of the two sampled groups are equal then separate variance estimates are used, however, this test only follows the  $t$  - distribution after correcting the degrees of freedom of the test to account for inequality of variances. The  $t$  - like statistic,  $t^*$ , when one assumes unequal variances would be:

$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

and follows the Behrens-Fisher distribution with  $df = (n_1 + n_2) - 2$ , or Student's -  $t$  distribution with

$$df = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left( \frac{s_1^2}{n_1} \right)^2}{n_1 + 1} + \frac{\left( \frac{s_2^2}{n_2} \right)^2}{n_2 + 1}} - 2$$

Another concept we will be examining today involves the difference between samples that are unrelated versus those that are paired (or repeated). Repeated measurements are useful when the variability between subjects is large. For example, if we took heart rates from a sample of people, and then administered a stimulant to another sample of people and took their heart rates, we would have two completely unrelated sets of observations from which to test the effect of this drug. However, we might suspect that individuals with normally high heart rates would also have higher heart rates when drugged. In the first design we have no control over this individual variability. If, however, we took an individual's heart rate before and after administration of the drug, we would have a set of paired observations from which to assess the effectiveness of this drug. In this design, the main assumption of parametric and non-parametric tests is still valid because subjects are chosen randomly and independently. Pairing observations is useful in many cases, when there is a positive correlation between subjects and their response to treatments, but it comes at a loss of degrees of freedom. In this case, you would use a paired  $t$  - test.

### **MANN-WHITNEY U TEST or WILCOXON RANK SUM TEST**

These are mathematically equivalent non-parametric tests, used if the assumptions of the  $t$  - test cannot be met. They are both based on ranking all the cases in order of increasing size. The test statistic U for the Mann-Whitney U test is the number of times a score from Group 1 precedes a score from Group 2 in the joint ranking. If the samples are from the same population, the distribution of scores in the ranked list should be random; an extreme value of U indicates a nonrandom pattern. Within **SPSS**, for samples of less than 30 cases, the exact significance level for U is computed. For larger samples, U is transformed into a z- statistic.

### **WILCOXON SIGNED RANK TEST**

This is a non-parametric test, used if the assumptions of the paired  $t$  - test cannot be met. It computes the differences between the pair of variables, ranks the absolute differences, sums the positive and negative ranks, and computes the test statistic T from the positive and negative signed ranks. Under the null hypothesis, T is approximately normally distributed with mean 0 and variance 1 for large sample sizes.

---

## Lab 4 Assignment

---

1.) For each of the following rejection regions, determine the value of  $\alpha$ , the probability of a Type I error:

- a)  $z < -1.96$
- b)  $z > 1.645$
- c)  $-2.58 > z > 2.58$

2.) TESTING FOR DIFFERENCES BETWEEN THE SAMPLE MEAN AND A THEORETICAL VALUE.

Suppose you want to determine if the lead concentrations in drinking water do not exceed the public health standard of 5 ppb (parts per billion). Based on 10 water samples you estimate the sample mean lead concentration to be,  $\bar{x} = 1.17$ , and the sample standard deviation to be,  $s = 0.36$ . Test the null hypothesis that the average lead level in the water samples does not exceed the public health standard. Start by clearly specifying your null and alternative hypotheses explicitly in terms of the underlying population mean of lead concentration in drinking water. Should this be a one or two tailed test?

3). TESTING THE DIFFERENCES BETWEEN MEANS.

Two energy saving concepts in home building are solar powered homes and earth sheltered homes. Suppose you are drawing up plans for a new home and want to compare the expected annual heating costs for the two types of innovation. Independent random samples of solar-powered homes yielded the summary data on annual heating costs shown below. Assume random and independent sampling, and that the homes were comparable with respect to size, climatic conditions, etc.

Is there evidence (at the  $\alpha = 0.05$  significance level) that the mean annual cost of heating an earth sheltered home is significantly less than the corresponding cost of heating a solar powered home?

<b>SOLAR POWERED</b>	<b>EARTH-SHELTERED</b>
$n = 12$	$n = 6$
$\bar{x} = \$285$	$\bar{x} = \$234$
$s = \$55$	$s = \$26$

Hint: Would a one-tailed or two-tailed test be of interest to the potential homebuilder?

**CALCULATE THE TEST STATISTIC, DETERMINE THE REJECTION REGION, ACCEPT OR REJECT THE NULL HYPOTHESIS.**

4.) The following data was collected on nitrate concentrations in two streams:

<b>Stream 1</b>	15.42	13.57	9.49	13.95	16.77	21.17	6.80	7.34	15.93
<b>Stream 2</b>	13.68	13.68	13.18	14.00	11.84	11.49	15.51	13.29	14.22

Use **SPSS** to calculate: a) a *t*-test, b) a Mann-Whitney U test, c) Wilcoxon Rank Sum test. For each test, turn in the edited **SPSS** output. Additionally, interpret the results of each test. In your write-up describing and interpreting the results of each test, specify the value of each test statistic, the null-hypothesis tested by each statistic, whether or not you rejected the null-hypothesis, and what assumptions of the tests might not have been met.

5) The following data on nitrate concentrations was collected at specific points along the same stream before and after a potential pollution source was established on the banks of the stream.

<b>Sample Point</b>	1	2	3	4	5	6	7	8	9	10
<b>Time 1</b>	4.97	7.39	9.08	9.65	9.83	10.84	10.86	10.93	11.78	12.01
<b>Time 2</b>	2.92	5.98	6.75	7.72	11.98	14.07	14.65	14.94	17.80	19.52

Use **SPSS** to calculate:

a) a paired sample *t*-test , b) a Wilcoxon Matched-Pairs Signed-Rank test

For each test, turn in the **SPSS** edited output. Additionally, interpret the results of each test. In your write-up describing and interpreting the results of each test, specify the value of each test statistic, the null-hypothesis tested by each statistic, whether or not you rejected the null-hypothesis, and what assumptions of the tests might not have been met. Why couldn't you have used these tests in problem 4?

### Further Instructions about the Lab Exercise

In **Lab 4**, the first 3 questions cannot be answered using **SPSS**. For questions 4 and 5, you will need to use three different commands in the **Analyze** menu under the **Compare Means** submenu, and under the **Nonparametric Tests** submenu.

**Question 4** requires you to use the **Independent Samples *t*-Test** command in the **Compare Means** submenu. You will first need to enter the data. For independent samples the response variable or random variable you are attempting to compare between 2 groups must be entered in a single column. However, to identify which group

an observation comes from, you also need to enter an integer code (like 1 and 2) in a separate column. Any pair of integers will work since the integer values are essentially subscripts and their magnitudes are irrelevant to the calculations; just make sure that you use a single value to code for the observations in group 1 and a single value to code for the observations in group 2. Once your data are entered and the variable attributes are correctly set on the **Variable View** tab of the **Data Editor Window**, you can select the **Independent Samples t-Test** command from the **Compare Means** submenu of the **Analyze** menu. In the sub-window that opens, using the arrow buttons click over the response variable into the **Test Variable** box, and the integer code you created to label the observations from each group to the **Grouping Variable** box. You will notice that double “question marks” appear next to the grouping variable name in the **Grouping Variable** box and that the **Define Groups** button now becomes available. Click on the **Define Groups** button and enter the values of the 2 integers that you used to label observations from your 2 groups of subjects. After you enter the codes click on continue to return to the main **Independent Samples t-Test** sub-window. Click on “OK” to execute the procedure.

The results of the **Independent Samples t-Test** will appear in the **Output Viewer** as 2 tables. The first table contains the means, standard deviations, standard errors and sample sizes for each of your 2 groups. The second table includes the essentials of the test you performed along with other information **SPSS** thinks you might want. First, columns 4, 5, and 6, contain the value of the  $t$ -statistic, its degrees of freedom (df), and two tailed significance level, respectively. These values are provided for  $t$ -tests both under the assumption that the variances in each group are equal, and under the assumption that they are unequal. Column 7 and 8 provide estimates of the mean difference between treatment groups and its standard error. These values are the numerator and denominator of the  $t$ -ratio, respectively. Columns 9 and 10 contain the upper and lower limits of a 95% confidence interval placed of the underlying population difference between the means of the two groups.

**A note on significance levels; SPSS almost always gives the two-tailed significance levels for a test (except  $F$  tests). If you think it most appropriate to have a one-tailed alternative hypothesis for any particular test, then to obtain the 1-tailed significance level, divide the two-tailed significance value by 2 (in half).**

Columns 2 and 3 contain the results of Levene’s test for equality of variances. This test involves using an  $F$ -statistic (tests involving variances always use  $F$  or Chi-square statistics) which is reported in column 2, and its degrees of freedom in column 3. You could use the results of this test to help you decide which of the 2  $t$ -tests to use (the one assuming equal or the one assuming unequal variances). I usually assume unequal variances, since it is the safest assumption to make, and therefore ignore Levene’s test.

To perform the non-parametric test that is equivalent to the independent samples  $t$ -test, you need to use the **2 Independent Samples** command from the **Nonparametric Tests** submenu of the **Analyze** menu. The data file set-up is identical to that required for the independent samples  $t$ -test. Furthermore, the sub-window that appears when

you select the **2 Independent Samples** command in the **Nonparametric Tests** submenu is similar to that from the equivalent *t*-test. Use the arrow buttons to click over the response variable and the grouping variable to the appropriate boxes. Assign the codes to their respective groups using the **Define Groups** button. Leave the default setting with the **Mann-Whitney U test** selected. This is the most appropriate test for our purposes. Check the help information if you want to learn more about the other test options. Click “OK” to execute the procedure.

The results of applying the **2 Independent Samples** command will appear in the **Output Viewer** appended to the bottom of the viewer. Two tables are produced. The first table gives the sample sizes, the average rank assigned to observations in each group, and the sum of ranks assigned to observations in each group. Remember that Non-parametric tests are based on jointly ranking the observations in the experimental groups from least to greatest. Table 2 contains on the 1<sup>st</sup> line the value of the **Mann-Whitney U** statistic, on the 2<sup>nd</sup> line the value of **W**, from the **Wilcoxon Rank Sum Test**, and on the 3<sup>rd</sup> line the value of **Z** (Normal approximation). The 4<sup>th</sup> and 5<sup>th</sup> lines contain the significance values for these tests. If one has no tied observations then one can report the **U** or **W** statistic and the exact significance from the 5<sup>th</sup> line. If however, there are tied observations, then one must report the **Z** value and the asymptotic significance from the 4<sup>th</sup> line.

For **Question 5**, you need to use the **Paired Samples t-test** from the **Compare Means** submenu of the **Analyze** menu, and the **2 Related Samples** command from the **Nonparametric Tests** submenu of the **Analyze** menu.

You must enter the data for these procedures differently than you did for the **Independent Samples t-Test** or the **2 Independent Samples** nonparametric test. Look at the table of data in the lab exercise for question 5 and imagine rotating this table 90° counterclockwise. You would then have 3 columns. The first would have an integer code identifying the sample point, and the 2 subsequent columns would contain the 2 observations on each sample point at each time (before and after a pollution source was established). This is how the data should be entered in the **Data Editor Window** (strictly speaking the integer code for sample point is not necessary). Note that this test requires that each “subject” be measured twice. Any subject with one of the pair of observations missing cannot be used in the analysis.

After you have entered the data in the above described format, select **Paired- Samples t-Test** from the **Compare Means** submenu of the **Analyze** menu. Click on either of the variable names associated with the two columns of data (not the integer code for sample point) and hold down the shift key and click on the second variable name. Use the arrow key to transfer this pair of variables to the **Paired Variables** box. Note that you must transfer two variables at the same time. Once you have the correct two variables in the **Paired Variables** box, click “OK” to execute the command.

The results will appear in the **Output Viewer** as 3 separate tables. The first table contains the sample sizes, and group means, standard deviations, and standard errors.

The 2<sup>nd</sup> table contains the correlation coefficient between the pairs of observations and its significance level. The 3<sup>rd</sup> table gives the details of the paired *t*-test. Columns 7, 8, and 9 contain the *t*-value, its degrees of freedom, and its significance level (2-tailed), respectively. Column 2 contains the mean difference between the paired observations. Column 3 contains the standard deviation of the mean difference, and column 4 contains the standard error of the mean differences. Column 2 divided by column 4 should equal the computed *t*-value. Finally, columns 5 and 6 contain the lower and upper limits on a 95% confidence interval on the underlying population mean difference between the paired observations.

To apply the equivalent non-parametric test to the data in question 5, choose the **2 Related Samples** command on the ***Nonparametric Tests*** submenu of the **Analyze** menu. Use the approach described above to click over the same pair of variables to the box labeled “**Test Pairs List.**” To obtain the **Wilcoxon Signed Rank** test, leave the default “**Wilcoxon**” box checked. To obtain the sign test as well, click on the box labeled “**Sign.**” Click “OK” to execute the command.

The results of applying the **2 Related Samples** command will appear in the **Output Viewer** appended to the bottom of the viewer. Two tables are produced. The first table gives the sample sizes, the average of the negative and positive ranks assigned to the pairs of observations, and the sum of negative and positive ranks assigned to the pairs of observations. Table 2 contains on the 1<sup>st</sup> line the **Z** value (normal approximation to **Wilcoxon’s T** statistic) and the 2<sup>nd</sup> line contains the significance value for this test. No exact test is provided.