TOPIC 2: INFERENCES FOR ONE AND TWO POPULATIONS

Table of Contents for Topic 2

2.1 Inferences for One Mean (The One-Sample Case) – Parametric Methods

- We can make inferences about one population, based on one sample.
- One-population inferences may be of two types:
  - Inferences about one mean.
  - Inferences about one proportion (not covered in this course).
- Recall the z-score formula for a population and for a sampling distribution.

$$z = \frac{y-\mu}{\sigma} \quad \text{and} \quad z = \frac{\bar{y}-\mu}{\sigma/\sqrt{n}}$$

- Occasionally, sigma (**σ**) is known and, thus, the standardized version of the sample mean ($\bar{y}$) can be converted to the formula for the one-sample *z*-test as follows:

$$z = \frac{\bar{y}-\mu_0}{\sigma\ /\sqrt{n}}$$

- However, usually sigma (**σ**) is not known, so we estimate the population standard deviation using the sample standard deviation (***s***), obtaining the Student ***t*** version of the sample mean ($\bar{y}$).
- The Student ***t*** version of the sample mean can be converted to the one-sample t-test (below).

Student *t* (studentized) version of the Sample Mean, $\bar{y}$

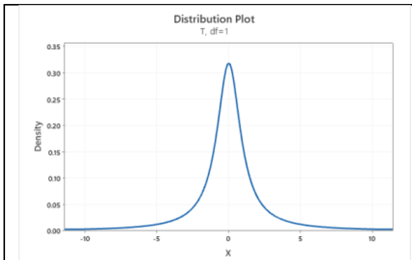
Suppose that a variable ***y*** of a population is normally distributed with mean $\bar{y}$, then for samples of size ***n***, the variable

$$t = \frac{\bar{y}-\mu}{s/\sqrt{n}}$$ follows the ***t***-distribution with ***n* – 1** degrees of freedom

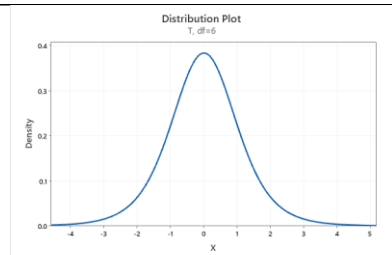
Introduction to the t-distribution

- The ***t***-distribution is very important in statistics and is applied, for example, in:
  - One-sample ***t***-test, two-sample ***t***-test, paired-sample ***t***-test, ***t***-test for the slope of a regression line
- The ***t***-distribution was developed by William Gosset in 1908. He published it under the name of 'student' so it became known as the student ***t*** distribution.
- The ***t***-curve is more spread out at the base than the normal distribution, particularly with small ***n***.

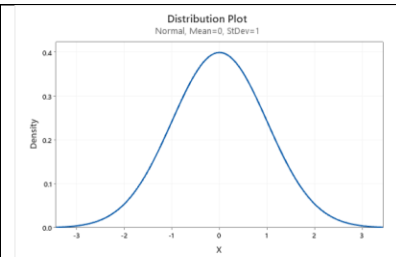
Properties of the *t*-curve (Properties 1 – 3 are the same as the standard normal distribution)
Property 1: The total area under a ***t***-curve = 1.
Property 2: A ***t***-curve is symmetrical about 0.
Property 3: A ***t***-curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis.
Property 4: There is a different ***t***-curve for each sample size, identified by its number of degrees of freedom (df), whereby ***df* = *n* – *1***.
Property 5: As ***df*** increases, the **t**-curve approaches the normal curve until, at ***df* = ∞**, the ***t***-curve coincides with the normal curve.



t-curve, df = 1



t-curve, df = 6



Normal curve

Using the *t*-Table

- A ***t***-table can be either one-tailed or two-tailed or both.
- However, the table used in this course is one-tailed (right-tailed).
- For cases where a required ***df*** is not shown, use the next lower ***df*** (to be conservative).

Examples:

1. For a one-tailed test, when α = 0.0005 and df = 12, $t_{0.0005} = 4.318$
2. For a one-tailed test, when α = 0.025 and df = 75, $t_{0.025} \approx 2.000$

2.1.1 Hypothesis Test for One Population Mean

The general formula for a test statistic is:

Test statistic: $$\frac{Estimate - H_0 value}{SE(Estimate)}$$

The One-Mean *t*-Test (also called the one-sample *t*-test)

Step 1: Check the purpose and assumptions (to see if this is the appropriate test for the research problem).

Purpose of the test: To test for the difference between a population mean (by taking a sample and calculating a sample mean) and some hypothesized (theoretical) mean or value.

Assumptions of the test:

1. Simple random sample.
2. The population under study is normally distributed or sample size is large.
3. σ is unknown.

Step 2: State the null and alternative hypotheses.

The null hypothesis is $H_0: \mu = \mu_0$ and the alternative hypothesis may be one of the following:

$H_a: \mu \neq \mu_0$ (two-tailed) or $H_a: \mu < \mu_0$ (left-tailed) or $H_a: \mu > \mu_0$ (right-tailed)

Step 3: Obtain the Calculated Value (or Observed Value) of the test statistic as follows:

$$t = \frac{\bar{y} - \mu_0}{s\ / \sqrt{n}} = \frac{\bar{y} - \mu_0}{SE(\bar{y})} \qquad df = n - 1$$

where μ_0 = some hypothesized (theoretical) mean or value.

Or the formula can be broken down into the following components:

Parameter	Estimate	SE(Estimate) of the mean	H_0 value	Reference Distribution
μ	$\bar{y}$	$\frac{s}{\sqrt{n}}$	μ_0	t_{n-1}

Step 4: Decide to reject H_0 or not reject H_0 and state the strength of the evidence against H_0.
Examine the *t*-table at $df = n - 1$
If the *P*-value ≤ α, we reject H_0 (otherwise do not reject H_0).

Step 5: Interpretation (Conclusion in Words).

Rounding Rules

1. Do not perform any rounding until all calculations are complete; otherwise, substantial rounding errors can occur.
2. When giving the final answer, keep at least one more decimal place than is given in the raw data.

Degrees of Freedom (df) = number of independent observations

- For standard deviation, ***df = n – 1*** (found in the denominator of the formula) because the sample mean is used as an estimate of the population mean in calculating it (using defining formula).
- You first calculate the mean and include that as part of the formula. When the sample mean is known, only **n – 1** observations are independent. One observation is <u>not</u> independent; it is fixed by knowing the other observations.
- For the one-sample t-test, also ***df = n – 1*** because standard deviation is included in the formula.

Example of a One-Sample t-test: Comparing Two-tailed and Left-tailed Hypothesis Tests (This example also shows the effect of sample size on the Power of the Test)

Acropora formosa (Staghorn Coral) forms large colonies. On mature reefs in Tanzania the average height of these colonies is about 75 cm, but they may reach 150 cm. (DSM, pp. 53, 55)

Research problem: A coral reef ecologist measures the height (in cm) of randomly selected *Acropora formosa* colonies along the reef crest of Mbudya Island (Dar es Salaam). Based on his experience in coral reef research, he knows that corals in shallow water, such as the reef crest, are generally shorter in height than average due to exposure to stronger wave action. Therefore, he formulates the **research question** as follows: At the 5% significance level, test whether the mean height of colonies of *Acropora formosa* colonies found on this reef crest of Mbudya Island is shorter than the mean height of 75 cm for such colonies throughout the country.

The reef ecologist first took a random sample of 9 colonies, obtaining the following data:

Heights (in cm) of random samples colonies of *A. formosa* at Mbudya (n = 9)										
Mbudya	81	48	74	69	79	56	59	61	84	

SPSS Output Based on Small Sample Size (n = 9)

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Mbudya_n9	9	67.8889	12.53440	4.17813

One-Sample Test							
	Test Value = 75						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Mbudya_n9	-1.702	8	.064	.127	-7.11111	-16.7459	2.5237

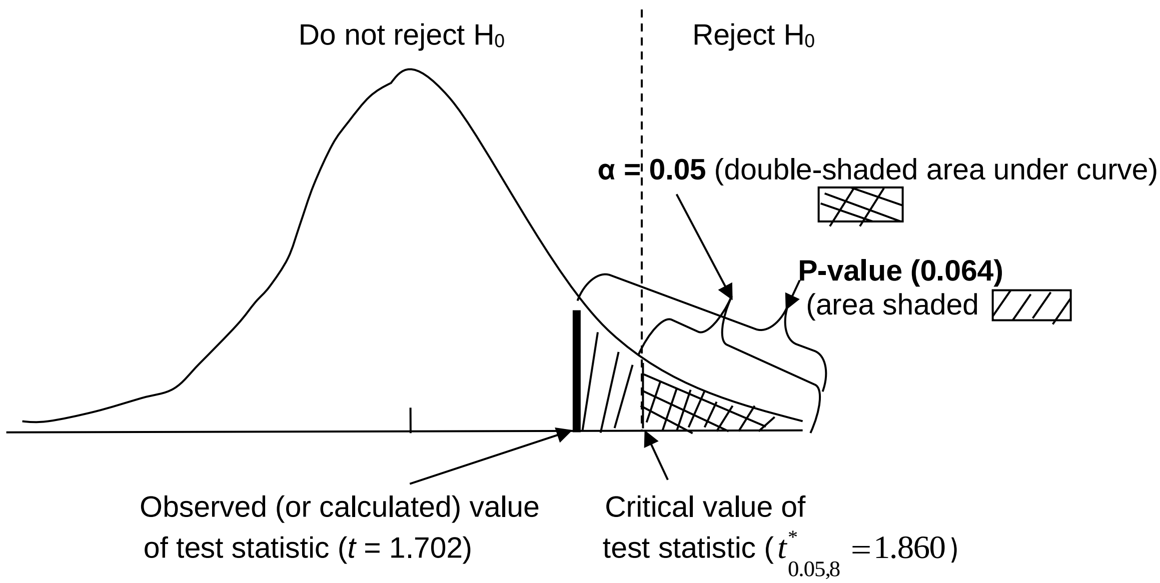
Based on these findings:

t = -1.702, df = 8, P = 0.064 (one-tailed)

Conclusion: At the 5% significance level, the data do not provide sufficient evidence to conclude that the mean height of colonies of *Acropora formosa* found on this reef crest of Mbudya Island is shorter than the mean height of 75 cm for such colonies throughout the country.

Illustration of the Case of Not Rejecting H_0

(Note: The test is left tailed, but the diagram is right tailed. This is essentially the same thing since the t-distribution is symmetric. The t-table provided only positive values.)



Note:

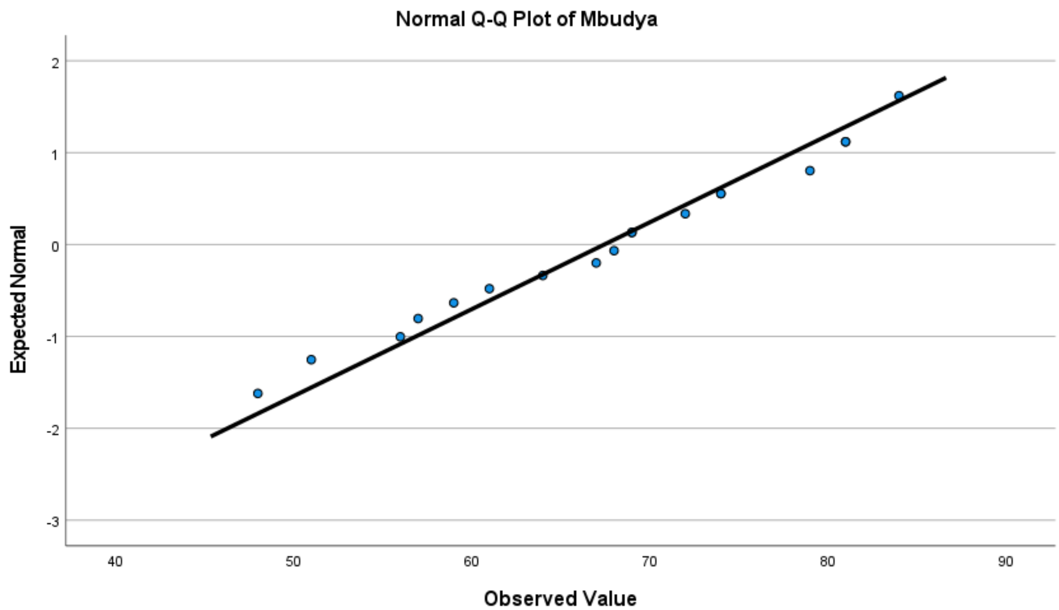
- Observed value of the test statistic < Critical value
  <u>AND</u>
- P-value > α

The Reef Ecologist was NOT convinced
To increase the power of the test, he took 9 more observations on the same reef to increased sample size to n = 18, obtaining the data shown below:

Heights (in cm) of random samples colonies of *A. formosa* at Mbudya (n = 18)																		
Mbudya	81	48	74	69	79	56	59	61	84	81	51	74	69	72	57	64	67	68

SPSS Output Based on Large Sample Size (n = 18) (For Normality tests & Hypothesis Test)

SPSS procedure for testing normality:
Select: Analyze >> Descriptive Statistics >> Explore >> Select Dependent Variable "Mbudya" >> Plots >> Tick Normal Probability Plots with Tests >> Continue >> OK. (This will give you the Normal Q-Q Plot and the Tests for Normality below).



Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Mbudya	.094	18	.200*	.969	18	.775

*. This is a lower bound of the true significance.

<u>**Shapiro Wilk Test for Normality**</u>**:**
H_0: The distribution is not different from a normal distribution.
Ha: The distribution is different from a normal distribution.

If the *P*-value for the Shapiro-Wilk Test is greater than 0.05, the distribution is <u>**NOT**</u> significantly different from a normal distribution.
If the *P*-value for the Shapiro-Wilk Test is less than or equal to 0.05, the distribution <u>**is**</u> significantly different from a normal distribution.
Note: For this test, always use $\alpha = 0.05$.

SPSS procedure to obtain the Descriptives (below) and Perform the Test (shown on page 14):
Select: Analyze >> Compare Means >> One-Sample T-test >> Select Test Variable “Mbudya“ >> Enter Test Value: 75 >> OK.

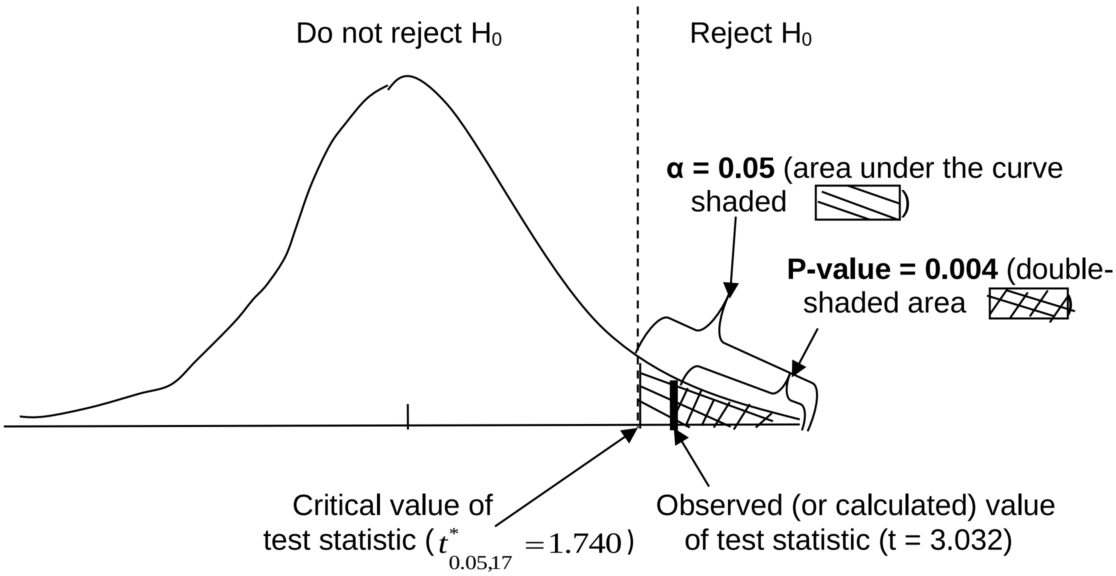
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Mbudya	18	67.4444	10.57312	2.49211

One-Mean t-test Showing All Steps (including checking assumptions)
At the 5% significance level, test whether the mean height of colonies of *Acropora formosa* colonies found on this reef crest of Mbudya Island is shorter than the mean height of 75 cm for such colonies throughout the country.
>>>>>>>>>>

>>>>>>>>>>

Illustration of Case of Rejecting H_0



Note:

- Observed value of the test statistic ≥ Critical value
  AND
- *P*-value ≤ α

IMPORTANT NOTE: P-value of One-tailed and Two-tailed tests
- A two-tailed can sometimes have a P-value > 0.50.
- However, for a one-tailed test, if you get a P-value > 0.5, you have tested the wrong tail.

Null Distribution

- If asked, "What is the distribution of the test-statistic under the null hypothesis?", you are just required to state the hypothesis test that you selected, the test statistic and degrees of freedom.
- For example: A one-sample t-test, "t(16)"; OR One-way ANOVA, "F(3, 28)".

Suppose you take a conservative approach and do a two-tailed test
At the 5% significance level, test whether the mean height of colonies of *Acropora formosa* colonies found on this reef crest at Mbudya Island is different from the mean height of 75 cm for such colonies throughout the country.

>>>>>>>>>>

>>>>>>>>>>
Note: In the above example, the evidence against H_0 is stronger when doing the one-tailed test as opposed to doing the two-tailed hypothesis. The lower P-value in the one-tailed test indicates stronger evidence. However, both cases are and considered as having very strong evidence ($0.001 < P \leq 0.01$).

Significance Level (α) versus P-Value

Distinction between Significance Level (α) and P-Value (both considered as Type I Error)

- When planning a hypothesis test, the significance level (α) = Probability of a Type I error. It is like a cut-off, the maximum error you can accept when rejecting H_0.
- P-Value is based on the calculated value of the test statistic = the observed Probability of the Type I error.
- That is why you compare the P-value and the significance level (α).
- If you reject H_0, the P-value = the probability of making a mistake by committing a Type I error.
- If you do not reject H_0, the P-value = the probability of the mistake you would have made if you had rejected H_0. That is why you did not reject H_0, because the chance of error would have been too large, i.e., greater than alpha (α).

2.1.2 Confidence Intervals for One Population Mean

- The general formula for a confidence interval is:

Confidence Interval: $Estimate \pm Critical\ Value \times SE(Estimate)$

One-Mean *t*-Interval Procedure (OR one-sample *t*-interval procedure)

Step 1: Find the Critical value: For a given confidence level (1 – α), use the t-table showing the critical values of the t-distribution to find $t_{\alpha/2}$(=t_{Crit}) (=t*) in the row for the appropriate df, where ***df = n – 1*** and ***n*** is the sample size.

Step 2: Calculate the two-sided confidence interval for ***μ*** to obtain the endpoints as follows:

$$\bar{y} - t_{\alpha/2} \times \frac{s}{\sqrt{n}} \quad \text{to} \quad \bar{y} + t_{\alpha/2} \times \frac{s}{\sqrt{n}}$$

OR $$\bar{y} \pm t_{\alpha/2} \times \frac{s}{\sqrt{n}} \quad \Rightarrow \quad \bar{y} \pm t_{\alpha/2,n-1} \times SE(\bar{y})$$

where **n** is the sample size; $\bar{y}$ and ***s*** are computed from the sample data.

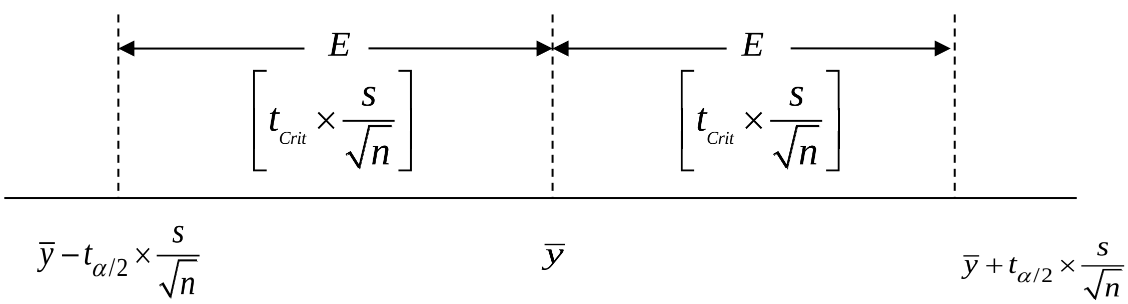
Or, the formula can be broken down as follows:

Parameter	Estimate	Critical value	SE(Estimate) of the mean
μ	$\bar{y}$	$t_{\alpha/2,n-1}$ (or t*)	$\frac{s}{\sqrt{n}}$

Step 3: Interpret the confidence interval in terms of the research problem being investigated.

Note: **Margin of Error (E)** = $t_{\alpha/2} \times \frac{s}{\sqrt{n}}$

Two-sided Confidence Interval and Margin of error can be **illustrated** as follows:



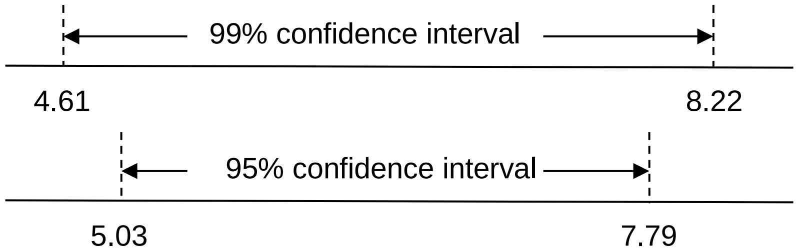
Margin of error = half the length of the confidence interval

- The margin of error determines the precision with which $\bar{y}$ estimates ***μ.***
- Increasing sample size:
  - Decreases the margin of error, and
  - Increases the precision of the estimate.

Confidence and Precision

- The length of the confidence interval is inversely proportional to precision.
  - For a given confidence level, a wide confidence interval indicates poor precision of the data.
- Thus, for a fixed sample size, decreasing the confidence level decreases the confidence interval, but increases the precision, and vice versa.

Compare the two confidence intervals and determine which has the greatest precision.



Conclusion about comparative precision: For the same sample size, since the 95% confidence interval is shorter, it has greater precision than the 99% confidence interval.

Examples of One-Mean Confidence Interval: Using the data given for the hypothesis test above **>>>>>>>>>>**

(a) Calculate a 95% confidence interval (using the same sample size n = 18) for the mean height of colonies of *Acropora formosa* found on the reef crest of Mbudya Island.

(b) Calculate a 99% confidence interval (using the same sample size n = 18).

(c) Suppose you increase the same size to $\boldsymbol{n}$ = 50 and you still get the same sample mean and sample standard deviation, calculate a 95% confidence interval.

>>>>>>>>>>

Meaning of a Confidence interval:

- For a certain percentage or level of confidence, that particular percentage of all samples of size $\boldsymbol{n}$, taken from the same population, will result in confidence intervals that will contain the population parameter.
- In the case of a one-mean confidence interval, that particular percentage of all samples of size $\boldsymbol{n}$ will result in confidence intervals (using the sample mean $\bar{y}$ as a point estimate) that will contain the population mean $\boldsymbol{\mu}$.
- **For a small number of samples:** E.g., if you take 20 samples of *Acropora formosa* ($\boldsymbol{n}$ = 18) from the same population, and calculate 95% confidence intervals for each, **approximately** 95% (or 19 out of 20) would contain the true population mean $\boldsymbol{\mu}$. However, just by chance, perhaps only 16 or perhaps all 20 of them would contain the population mean.
- **For a large number of samples:** E.g., if you take 1,000 or 10,000 samples ($\boldsymbol{n}$ = 18) from the same population, the result will get closer and closer to exactly 95% of the confidence intervals would contain the true population mean.

Relation between Hypothesis Tests and Corresponding Confidence Intervals

Relation between Hypothesis Tests and Corresponding Confidence Intervals: Inferences for One Population Mean

For a two-tailed hypothesis test for comparing one population mean (μ) with some hypothesized mean or value (μ_0) at the significance level α:

- **The case of rejecting H_0:** If the null hypothesis is rejected, the (1-α) confidence interval for μ will <u>not</u> contain the hypothesized mean or value (μ_0).
- **The case of <u>not</u> rejecting H_0:** If the null hypothesis is not rejected, the (1-α) confidence interval for μ <u>will</u> contain the hypothesized mean or value (μ_0).

[If μ_0 is within the interval, there is no significant difference between μ and μ_0]

When we refer to <u>Corresponding</u> Confidence intervals
<u>Two conditions</u> that must be met to ensure that the conclusions will be the same for a hypothesis test and a confidence interval performed on the same data:

1. The confidence level must be a compliment of the significance level (α) applied in the hypothesis test.
2. They must be the same “sided”, that is, both two-sided or both one-sided (not covered).
   - If these two conditions are not met, the hypothesis test and confidence level may still give the same conclusion, but there is no guarantee.

>>>>>>>>>>

Comparison of the Two-tailed Test & Two-sided 95% Confidence Interval (*n* = 18)

>>>>>>>>>>

Compare the conclusions of the hypothesis tests and confidence interval for the research problem on *Acropora formosa* on the reef crest at Mbudya Island.

Two-tailed hypothesis test for $\mu \neq 75$ cm (at $\alpha = 0.05$):

H_0 was rejected at $0.01 > P > 0.005$ (exact P-value = 0.008). So, we concluded that there was a difference.

Confidence Interval (at 95% level):

The interval (62.19, 72.70) cm does not contain 75 cm and thus confirms that the mean height of the population of colonies of Acropora formosa on the reef crest is different from the hypothesized value or mean height of this species countrywide.

SPSS procedure to perform a One-Mean t-test:
Select: Analyze >> Compare Means >> One-Sample T-test >> Select Test Variable "Mbudya" >> Enter Test Value: 75 >> Select Options and set 95% Confidence level (default) >> OK.

SPSS Output for Hypothesis Test

One-Sample Test

	Test Value = 75						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Mbudya	-3.032	17	.004	.008	-7.55556	-12.8134	-2.2977

Note: "Sig." is actually the P-value (not significance level). The significance level is not input into SPSS. Both one-tailed and two-tailed P-values are provided.

Compare the results with the hand calculations on pp. 6-8:
Same t-statistic, and df and exact P-values lie within those ranges.
Note: The confidence interval shown when doing a hypothesis test is NOT correct.

- You can manually calculate: [-12.8314 + 75, -2.2977 + 75] = [62.1866, 72.7023].
- OR, per obtain SPSS output as shown below.

SPSS procedure to obtain a One-Mean Confidence Interval:
Select: Analyze >> Compare Means >> One-Sample T-test >> Select Test Variable "Mbudya" >> Enter Test Value: Leave as 0 >> Select Options and set 95% Confidence level (default) >> OK.
Note: The Test Value must be changed from 75 to 0.

SPSS Output for Confidence Interval: For two-sided 95% confidence interval

One-Sample Test

	Test Value = 0						
			Significance			95% Confidence Interval of the Difference	
	t	df	One-Sided p	Two-Sided p	Mean Difference	Lower	Upper
Mbudya	27.063	17	<.001	<.001	67.44444	62.1866	72.7023

Two-sided confidence interval

- The two-sided 95% confidence level is (62.19, 72.70) cm.

Note: The t-statistic (27.063) and P-values (< 0.001) shown here are **meaningless**.

Obtaining P-values from Statistical Tables and Computer Output

- When we use statistical tables, we will only find the P-value within certain ranges.
  - E.g., P < 0.001 or P > 0.10 or 0.05 < P < 0.01, etc.
- However, when we use a computer program it will tell us the exact P-values.
  - E.g., P = 0.1332 or P = 0.00167, etc.
- The exact P-value from the output will always be within the P-value range shown in the table.

2.2 Inferences for Two Means (Two-Sample Cases)

- This involves **one response variable**, which is **quantitative** and usually continuous.
- The **explanatory variable is categorical**, which has **two levels or categories**, which can be considered as two populations.
- We are testing for the effect of the explanatory variable on the response variable.
- We are comparing **two populations** or groups by taking **two samples.**
- Two sample inferences:
    - Comparing **two population means: Independent samples**
    - Comparing **two population means: Paired samples**
    - Comparing **two proportions** (not means, not covered in this course).

2.2.1 Inferences for Two Population Means: Based on Two Independent Samples

2.2.1.1 The Sampling Distribution of the Difference between Two Sample Means for Independent Samples

Table: Notation for parameters and statistics when comparing two populations.

Parameter/Statistic	Population 1	Population 2
Population mean	μ_1	μ_2
Population standard deviation	σ_1	σ_2
Sample mean	$\bar{y}_1$	$\bar{y}_2$
Sample standard deviation	s_1	s_2
Sample size	n_1	n_2

Sampling Distribution of the Difference Between Two Independent Sample Means, $\bar{y}_1 - \bar{y}_2$:

Suppose that ***y*** is a normally distributed variable on each of two populations, then, for independent samples of size n_1 and n_2 from the two populations,

- The mean of all possible differences between the two sample means equals the difference between the two population means: $\mu_{\bar{y}_1-\bar{y}_2} = \mu_1 - \mu_2$.
- The standard deviation of all possible differences between the two sample means is:
$$\sigma_{\bar{y}_1-\bar{y}_2} = \sqrt{(\sigma_1^2/n_1)+(\sigma_2^2/n_2)}$$
- $\bar{y}_1 - \bar{y}_2$ is assumed to be normally distributed.

2.2.1.2 Inferences for Two Population Means Using Independent Samples, Standard Deviations Not Assumed Equal

- This is a **general case** of comparing two population means and can be applied whether standard deviations are equal or not equal.
- You are NOT responsible for performing this test; you ONLY need to know the assumptions.

Nonpooled Two-Mean t-test (or Nonpooled Two-Sample t-test)

Assumptions:

1. Simple random samples (also implies also independent sampling within samples).
2. Both populations are normally distributed or both samples are large.
3. Samples are independent.

2.2.1.3 Inferences for Two Population Means Using Independent Samples, Standard Deviations Assumed Equal

- This is the special case of comparing two populations (independent samples), which can be applied when the standard deviations of the two populations are similar and samples sizes are nearly equal
- Where the assumptions of the pooled t-test are met, it is slightly more powerful than nonpooled t-test.

Pooled t-test (Two-sample t test assuming equal variances; also called two-mean t-test assuming equal variances)

Step 1: Check the purpose and assumptions (to see if this is the appropriate test):

Purpose: To test for the difference between two population means (μ_1 and μ_2) based on two sample means ($\bar{y}_1$ and $\bar{y}_2$).

Assumptions:

1. Simple random samples.
2. Both populations are normally distributed or both samples are large.
3. Samples are independent.
4. Equal population standard deviations (A rule of thumb: if the ratio of the larger to the smaller sample standard deviation < 2, we can say the assumption has been met) [OR apply Levene's Test for Equality of Variances – more reliable].
5. Sample sizes should be roughly equal.

Step 2: State Null and Alternative Hypotheses:

Null hypothesis is H_0: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$

Alternative hypothesis may be one of the following:

H_a: $\mu_1 \neq \mu_2$ or	H_a: $\mu_1 < \mu_2$ or	H_a: $\mu_1 > \mu_2$
Or Ha: $\mu_1 - \mu_2 \neq 0$	Ha: $\mu_1 - \mu_2 < 0$	Ha: $\mu_1 - \mu_2 > 0$
(two-tailed)	(left-tailed)	(right-tailed)

Step 3: Calculate the test statistic:

Test statistic:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - \Delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2) - \Delta_0}{SE(\bar{y}_1 - \bar{y}_2)}$$

Where s_p (pooled standard deviation) is:

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

Degrees of freedom:

$$df = n_1 + n_2 - 2$$

The formula can be broken down into these components:

Parameter	Estimate	SE(Estimate) of the difference between the means	H_0 value	Reference Distribution
$\mu_1 - \mu_2$ (Assume $\sigma_1 = \sigma_2$)	$\bar{y}_1 - \bar{y}_2$	$s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$0 (or \Delta_0)$	$t_{n_1 + n_2 - 2}$

Note: Δ_0 = some hypothesized difference, which is almost always 0

Step 4: Decide to reject H_0 or not reject H_0 and state the strength of the evidence against H_0

If you obtain a *df* that is not shown in the *t*-table, go to the next lower *df*.

If the P-value ≤ α, we reject H_0 (otherwise do not reject H_0)

Step 5: Interpretation (conclusion) in words in terms of the research problem being investigated.

Confidence Intervals for the Difference Between the Means of Two Populations, Using Independent Samples, Standard deviations Assumed Equal

Two-Mean t-Interval Procedure (=Pooled t-Interval)

Purpose: To find a confidence interval for the difference between two population means (μ_1 and μ_2) based on two sample means ($\bar{y}_1$ and $\bar{y}_2$).

Assumptions: Same as for the Pooled t-test

Step 1: Obtain the Critical Value $t_{\alpha/2}$ (=t*) (=t_{Crit}) for a given confidence level (1 – α) at

$$df = n_1 + n_2 - 2.$$

Step 2: Calculate the endpoints of the two-sided confidence interval of $\mu_1 - \mu_2$:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

or $$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, n_1+n_2-2} \times SE(\bar{y}_1 - \bar{y}_2)$$

Where s_p = pooled standard deviation

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

The formula can be broken down into these components:

Parameter	Estimate	Critical value	SE(Estimate) of the difference between the means
$\mu_1 - \mu_2$ (Assume $\sigma_1 = \sigma_2$)	$\bar{y}_1 - \bar{y}_2$	Two-sided $t_{\alpha/2, n_1+n_2-2}$	$s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ Where $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

Step 3: Interpret the confidence interval in terms of the research problem.

Margin of Error (E) (two-sided) = $t_{\alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

Note: Confidence level = 1 – α = 1 – significance level

Example of the Pooled t-test (a Two-Mean t-test for Independent Samples)

Research problem: The heights of randomly selected *Acropora formosa* colonies along the reef crests of Mbudya Island and Bongoyo (Dar es Salaam) are shown below, along with summary statistics. At the 5% significance level, test whether there is a difference in the mean heights of colonies of *Acropora formosa* colonies found on the reef crests of these two islands. Assume that **all assumptions are met** for the required analysis.

Heights (in cm) of random samples colonies of A. formosa at Mbudya and Bongoyo																		
Mbudya	81	48	74	69	79	56	59	61	84	81	51	74	69	72	57	64	67	68
Bongoyo	86	87	70	62	73	71	85	57	82	74	81	60	63	59	63	78		

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Mbudya	18	48.00	84.00	67.4444	2.49211	10.57312
Bongoyo	16	57.00	87.00	71.9375	2.59562	10.38248
Valid N (listwise)	16					

>>>>>>>>>>

>>>>>>>>>>

Example of Calculating a Pooled t-interval

Using the data and information given for the pooled t-test above, calculate a 95% confidence interval for the difference between mean heights of colonies of *Acropora formosa* found on the reef crests at Mbudya Island and Bongoyo Island (Dar es Salaam).

>>>>>>>>>>

>>>>>>>>>>

Relation between Hypothesis Tests and Corresponding Confidence Intervals

Relation between Hypothesis Tests and Corresponding Confidence Intervals: Inferences for Two Population Means

For a two-tailed hypothesis test for comparing two population means at the significance level α:

- The case of rejecting H_0**:** If the null hypothesis is rejected, the (1-α) confidence interval for the difference between the population means ($\mu_1 - \mu_2$) will not contain 0, i.e. either both endpoints will be negative OR both endpoints will be positive.

- The case of not rejecting H_0**:** If the null hypothesis is not be rejected, the (1-α) confidence interval for the difference between the population means ($\mu_1 - \mu_2$) will contain 0, i.e. one endpoint will be negative and the other will be positive.

  [If 0 is within the interval, there is 0 difference OR no significant difference.]

When we refer to Corresponding Confidence intervals

Two conditions that must be met to ensure that the conclusions will be the same for a hypothesis test and a confidence interval performed on the same data:

1. The confidence level must be a compliment of the significance level (α) applied in the hypothesis test.
2. They must be the same "sided", that is, both two-sided or both one-sided.

- If these two conditions are not met, the hypothesis test and confidence level may still give the same conclusion, but there is no guarantee.

Example of Relating Hypothesis Test and Confidence Interval for the Difference Between Two Means

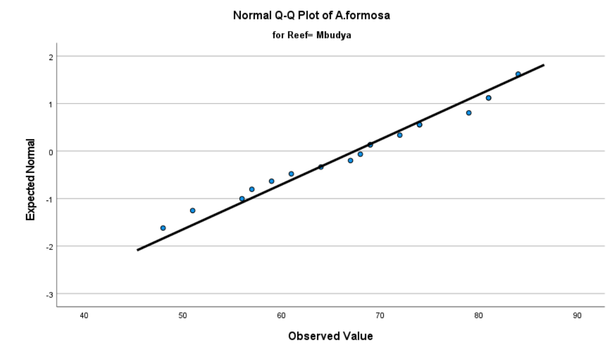
Results from the hypothesis test showed that, at a significance level 5% (α = 0.05), there was no difference in the mean heights of *Acropora formosa* colonies at Mbudya and Bongoyo Islands.

Results from calculating the confidence interval showed that the (1 – α)% or 95% confidence interval for the difference between the two means (-11.85, 2.86 cm) includes 0. Therefore, we can be 95% confident that the difference between the two means is 0 (not significantly different from 0).

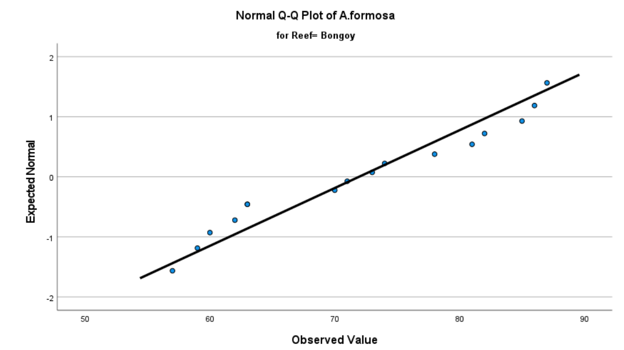
Therefore, the two types of inferential statistics give the same conclusion.

SPSS Output & Checking Assumptions for Above Example on Testing for the Difference in Mean Height of *Acropora formosa* on the Reef Crests at Mbudya and Bongoyo Islands

SPSS procedure for testing normality:
Select: Analyze >> Descriptive Statistics >> Explore >> Select Dependent Variable “Acropora Formosa” >> Select Factor List “Reef” >> Plots >> Tick Normal Probability Plots with Tests >> Continue >> OK. (This will give you the Normal Q-Q Plot and the Tests for Normality below).



Mbudya



Bongoyo

Tests of Normality[a]

		Kolmogorov-Smirnov[b]			Shapiro-Wilk		
	Reef	Statistic	df	Sig.	Statistic	df	Sig.
A.formosa	Bongoy	.180	16	.173	.922	16	.185
	Mbudya	.094	18	.200*	.969	18	.775

*. This is a lower bound of the true significance.

a. There are no valid cases for A. formosa when Reef = .000. Statistics cannot be computed for this level.

b. Lilliefors Significance Correction

SPSS procedure for Two-Mean Hypothesis Test and Confidence Interval (shown on next page):
Select: Analyze >> Compare Means >> Independent Samples T Test >> Select Test Variable “Acropora formosa“ >> Select Grouping Variable “Reef” >> Click Define Groups >> Group 1, type “Mbudya”; Group 2, type Bongoyo”, Continue >> Click Options, select Confidence level >> OK.

SPSS Output: Hypothesis test and Confidence Interval

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
A.formosa	Equal variances assumed	.026	.873	-1.247	32	.111	.221	-4.49306	3.60228	-11.83067	2.84456
	Equal variances not assumed			-1.249	31.662	.110	.221	-4.49306	3.59831	-11.82565	2.83954

Levene's Test for Equality of Variances:

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$ (based on $s_1^2 = s_2^2 = \ldots = s_k^2$)

Ha: Variances are different between at least 2 populations.

If the P-value for Levene's Test is greater than 0.05, there is **NOT** sufficient evidence of a difference between variances.

If the P-value for Levene's Test is less than or equal to 0.05, there **is** sufficient evidence of a difference between variances.

Note: For this test, always use α = 0.05.

Checking Assumptions in this Example:

(Although the question states that all the assumptions are met, we will now verify this.)

1. Two independent random samples.
   The question states the samples were taken randomly and there does not appear to be any link or pairing between the two samples.

2. Since the P-value for the Shapiro-Wilk Test is greater than 0.05 for both Bongoyo (P = 0.185) and Mbudya (P = 0.775), neither distribution is significantly different from a normal distribution.

3. Equal standard deviations. Levene's Test is most reliable.

   In this example, P = 0.873 (> 0.05), so there are no significant differences between variances.

   An approximate method: The ratio of the standard deviations (10.57/10.38) is < 2.

4. Sample sizes roughly equal (n_1 = 18, n_2 = 16).

2.2.2 Inferences for Two Population Means: Using Two Paired Samples

- Applies when two populations or measurements are **paired** in space or in time or by some relationship or paired on the same subject (study unit)
- Examples of pairing:
  - **Ecological monitoring** (coral reefs, mangroves, etc.) – same plots observed over time.
  - Taking measurements at the same time in different sites (pairing in time).
  - Taking measurements on the same patient, such as blood pressure, before and after treatment.
  - Heights of fathers and their oldest sons.

Paired-Sample t-test (or Paired t-test) [Sometimes called **Matched Pairs t-test**]

Step 1: Check the purpose and assumptions (to see if this is the appropriate test)

Purpose: To test for the difference between two populations means, μ_1 and μ_2 (based on the differences between two paired samples).

Assumptions:

1. Simple random sample.
2. Samples are paired (random paired sample) (not independent).
3. Differences between paired observations are normally distributed or sample size is large.

Step 2: State the null and alternative hypotheses

Null hypothesis is $H_0: \mu_1 - \mu_2$ or $H_0: \mu_d = 0$, where $\mu_d = \mu_1 - \mu_2$

Alternative hypothesis may be one of the following:

$H_a: \mu_1 \neq \mu_2$	or	$H_a: \mu_1 < \mu_2$	or	$H_a: \mu_1 > \mu_2$
$H_a: \mu_d \neq 0$		$H_a: \mu_d < 0$		$H_a: \mu_d > 0$
(two-tailed)		(left-tailed)		(right-tailed)

Step 3: Obtain the Calculated Value (or Observed Value) of the test statistic as follows:

Mean difference = $\bar{d} = \dfrac{\sum d_i}{n}$ [Note: $d = y_1 - y_2$ for each pair]

Standard deviation of the differences = $s_d = \sqrt{\dfrac{\sum d_i^2 - (\sum d_i)^2 / n}{n-1}}$

$$t = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}} = \frac{\bar{d} - \Delta_0}{SE(\bar{d})}$$

n = number of paired observations and $df = n-1$

The formula can be broken down into these components:

Parameter	Estimate	SE(Estimate) of the mean difference	H_0 value	Reference Distribution
μ_d $(or\, \mu_1 - \mu_2)$	$\bar{d}$	$SE(\bar{d}) = \dfrac{s_d}{\sqrt{n}}$	$0 (or\, \Delta_0)$	t_{n-1}

Note: Δ_0 = some hypothesized difference, which is almost always 0

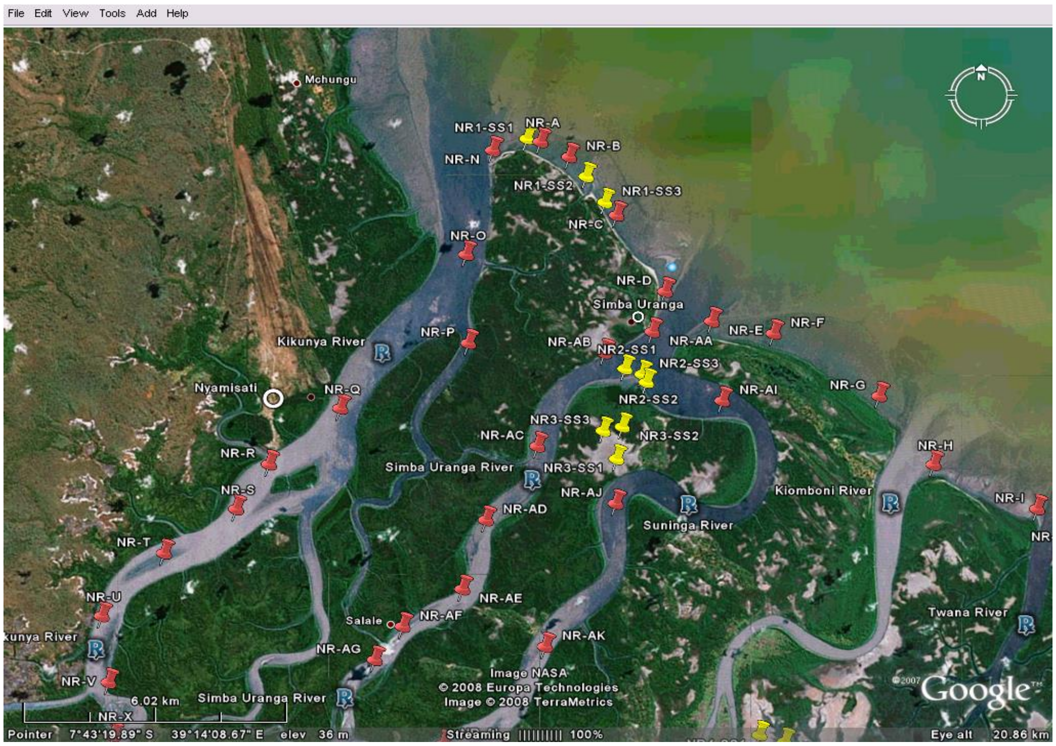
Step 4: Decide to reject H_0 or not reject H_0 and state the strength of the evidence against H_0
If you obtain a *df* that is not shown in the *t*-table, go to the next lower *df*.
If the P-value ≤ α, we reject H_0 (otherwise do not reject H_0)

Step 5: Interpretation (conclusion) in words in terms of the research problem being investigated.

Note on Power of the Paired-sample t-test:

- Where there is strong pairing (significant correlation) between paired observations, this paired-sample t-test is more powerful than the pooled or nonpooled t-test for independent samples.
- Where correlation between pairs is not significant, the t-tests for independent samples are more powerful.

Example of Paired Design:
Research on Impacts of Climate Change on Mangroves in Tanzania



Research design:

- This was a climate change impact assessment.
- Quantitative ecological plot assessment in 60 sites and subsites (a total of 480 plots).
- Permanent plots (each 5 m x 5 m) were established.
- Repeated in 2007 and 2009, measuring girth at breast height of the mangroves in each permanent plot to get mangrove basal area.
- Thus, **the measurements are paired in space (same plots) for two time periods.**
- So, a **paired-sample t-test** or **Wilcoxon paired-sample test** was used for all sites.

Two Trends:

1. **On seaward edges** of mangroves, there is drastic erosion in many areas due to sea level rise combined with increase in storms and wave activity.
2. **On landward edges** there is landward migration of mangroves in some areas due to sea level rise into the open area (sea level is rising approximately 4 mm per year vertically, which can mean as much as 1 m per year influx of water horizontally in a very flat area).

Long stretch of coastline at Subsite NR-SS3 being eroded away by increased wave action and sea level rise due to climate change.

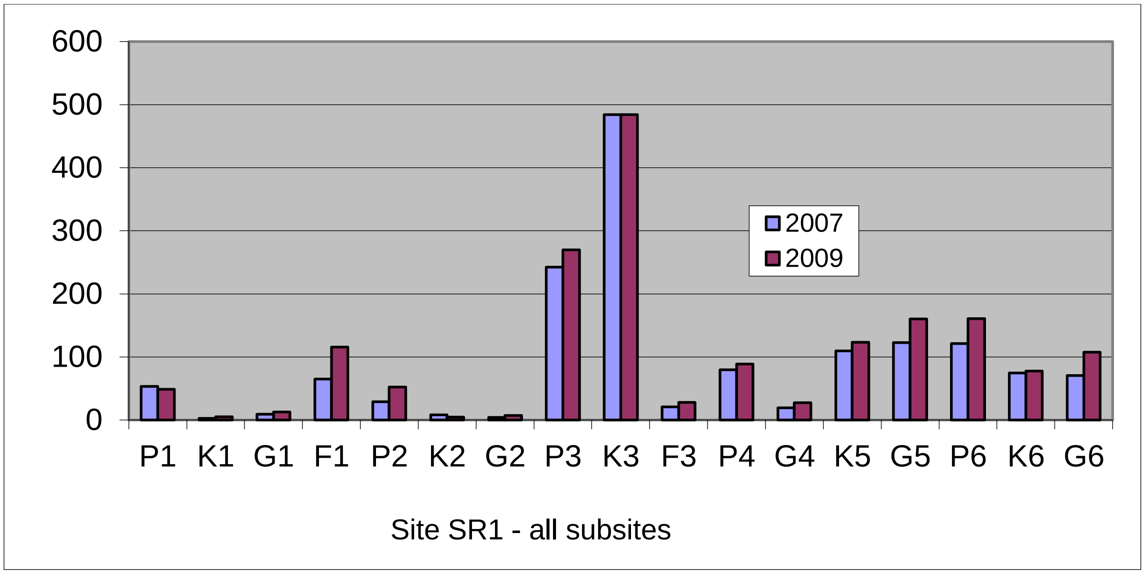
Avicennia mangroves invading into a saline flat area due to sea level rise, but some are also under stress due to decreased rainfall and humidity.

Research Question: Was there a change between 2007 and 2009 in mangrove basal area (a measure of mangrove abundance) in a Site SR1 on the landward edge? Random plots were set up in 2007 and girth at breast height of the mangroves was measured to obtain basal area. The same measurements were repeated in the same plots in 2009.

Mangroves may have:

- **Increased** due to sea level rise into the open area, or
- **Decreased** due to decreased rainfall and humidity and increased temperatures, which cause desiccation and stress on the mangroves.

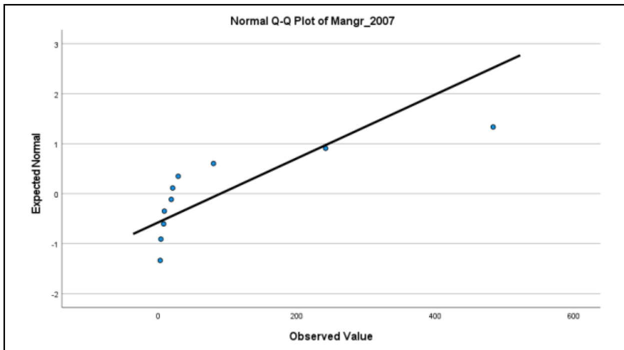
Therefore, a **two-tailed test** should be done.



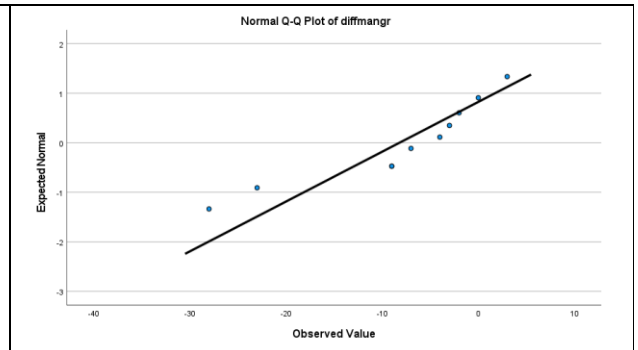
Research Problem: At the 5% significance level, test whether mangrove basal area in Site SR1 changed between 2007 and 2009. (Reduced to 10 plots here for analysis.)

	Basal area (cm^2/25-m^2 plot)	
Plot	**2007**	**2009**
1	3	5
2	9	13
3	29	52
4	8	5
5	4	7
6	242	270
7	484	484
8	21	28
9	80	89
10	19	28

SPSS procedure for testing normality:
Select: Analyze >> Descriptive Statistics >> Explore >> Select Dependent Variables: “Mangroves 2007”, “Mangroves 2009”, and “Diffmangroves” >> Plots >> Tick Normal Probability Plots with Tests >> Continue >> OK. (This will give you the Normal Q-Q Plot and the Tests for Normality below).



Mangrove 2007 (& 2009 is similar)



Paired differences

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Mangr_2007	.351	10	<.001	.629	10	0.000123
Mangr_2009	.323	10	.004	.661	10	0.000300
diffmangr	.268	10	.040	.863	10	0.083

a. Lilliefors Significance Correction

Solution for Paired t-test Showing All Steps (including checking assumptions)

Selection of Test: Paired-sample t-test is selected because:
Purpose of the study: To test for a difference between two population means, based on two samples that were paired in space (in the same permanent plots) in 2007 and 2009.
Assumptions:

1. Random sample of plots.
2. Samples (observations) are paired on the plots.
3. The differences between the paired observations are approximately normally distributed as indicated by the Shapiro-Wilk Test ($P = 0.083 > \alpha = 0.05$).
   [**Note**: Data recorded for each year are NOT normal: 2007 ($P = 0.000123$), 2009 ($P = 0.000300$).]

H_0: $\mu_d = 0$ or $\mu_1 = \mu_2$ (Mean mangrove basal area did not change between 2007 and 2009)

H_a: $\mu_d \neq 0$ or $\mu_1 \neq \mu_2$ (Mean mangrove basal area changed between 2007 and 2009)

Parameter: $\mu_d = \mu_{2007} - \mu_{2009}$

	Basal area ($cm^2/25\text{-}m^2$ plot)			
Plot	**2007**	**2009**	**d**	**d^2**
1	3	5	-2	4
2	9	13	-4	16
3	29	52	-23	529
4	8	5	3	9
5	4	7	-3	9
6	242	270	-28	784
7	484	484	0	0
8	21	28	-7	49
9	80	89	-9	81
10	19	28	-9	81
Sums			$\sum d = -82$	$\sum d^2 = 1562$

>>>>>>>>>>

>>>>>>>>>>

Note: For questions on the paired t-test, you may be provided with summary statistics, e.g., the mean difference and either the standard deviation of the differences or SE (the standard error of the mean difference).

Comparison of the Power of the Paired-sample t-test and Pooled t-test in this case where there is very strong pairing (correlation)

[Based on <u>all 17 plots</u> in this site]

- Correlation coefficient: r = 0.9898 (this is very strong correlation)

	Paired-sample t-test	**Pooled t-test**
Calculated t	t = - 3.637	t = - 0.369
Exact P-value	P = 0.002271	P = 0.7147
Decision	Reject H_0	Do not Reject H_0
Evidence to reject H_0	Very strong	Very weak
Power	Very powerful	Not powerful

Therefore, where pairing is very strong, such as in this example, the paired-sample t-test is much more powerful than a two-mean t-test for independent samples.

Note:

1. Paired design made the study <u>very sensitive</u> to small changes over time.
2. If you had not paired these in fixed plots, the <u>variation among plots</u> within the same time period would have <u>masked</u> (been much greater than) the <u>difference over time</u> (2007 to 2009)

Paired t-Interval Procedure

Paired t-Interval Procedure

Purpose: To find a confidence interval for the <u>difference</u> between two population means, μ_1 and μ_2 based on paired observations.

Assumptions: Same as for the Paired t-test.

Step 1: Obtain the Critical Value $t_{\alpha/2}$ (=t*) (=t_{Crit}) for a given confidence level (1 – α) at **df** = n – 1.

Step 2: The endpoints of the confidence interval of $\mu_1 - \mu_2$ are defined by:

$$\bar{d} \pm t_{\alpha/2} \times \frac{s_d}{\sqrt{n}} \quad \textbf{or} \quad \bar{d} \pm t_{\alpha/2,n-1} \times SE(\bar{d})$$

Or, the formula can be broken down as follows:

Estimate	Critical value	SE(Estimate) of the mean difference
$\bar{d}$	$t_{\alpha/2,n-1}$ (or t*)	$\frac{s_d}{\sqrt{n}}$

Step 3: Interpret the confidence interval in terms of the research problem being investigated.

Margin of Error (E) = $t_{\alpha/2,n-1} \times \frac{s_d}{\sqrt{n}}$

Example of a Paired *t*-interval
Based on the mangrove data shown above, calculate a 95% confidence interval for the difference in mangrove basal area in Site SR1 between 2007 and 2009.
>>>>>>>>>>

>>>>>>>>>>
SPSS procedure for Paired-sample Hypothesis Test and Confidence Interval:
Select: Analyze >> Compare Means >> Paired-samples T Test >> Select Mangroves2007 as Variable 1 and Mangroves2009 as Variable 2 >> Click Options, select Confidence level >> OK.

SPSS Output: Paired t-test and Confidence Interval for the Difference in Mangrove Basal Area in Site SR1 between 2007 and 2009

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Mangr_2007	89.9000	10	156.42282	49.46524
	Mangr_2009	98.1000	10	157.54396	49.81978

Paired Samples Correlations

		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	Mangr_2007 & Mangr_2009	10	.998	3.354×10^{-11}	6.708×10^{-11}

Paired Samples Test

	Paired Differences					t	df	Significance	
				95% Confidence Interval of the Difference					
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			One-Sided p	Two-Sided p
2007 - 2009	-8.20000	9.94205	3.14395	-15.31212	-1.08788	-2.608	9	.014	.028

Additional Example

A fuel manufacturer wanted to test the effectiveness of a new gasoline additive. A random sample of 6 cars were driven one week without the additive and one week with the additive, obtaining summary statistics as shown in the table below (in miles per gallon). Note: You might not need all of the statistics shown.

Summary statistics	Without additive	With additive	Pooled	Difference
Mean	23.40	25.12		-1.72
Standard deviation	5.42	5.87	5.65	1.43

Note: Perform the analysis **WITHOUT** using the numbers highlighted in yellow.

(a) The confidence interval for the difference in mileage without and with the additive is $(-3.22094, -0.21906)$. Determine the confidence level at which this interval was calculated.

>>>>>>>>>>

>>>>>>>>>>

(b) Calculate a 99% confidence interval for the difference in mileage without and with the additive.
>>>>>>>>>>

>>>>>>>>>>

(c) Compare the confidence interval given in part (a) and the confidence interval you calculated in part (b). Based on each of these confidence intervals, is there a difference in mileage without and with the additive. Explain why you either got the same conclusion or different conclusions from the two intervals.

The confidence interval given in part (a), $(-3.22094, -0.21906)$ is shorter and more precise than the confidence interval calculated in part (b), $(-4.074, 0.634)$.

Based on the 95% confidence interval given in part (a), we would conclude that there is a difference in mileage without and with the additive, it does not contain 0.
However, based on the 99% confidence interval calculated in part (b), we would conclude that there is no difference, because it does contain 0.

Research Problem on Exercise Program to Reduce Weight

It is claimed that a certain exercise program will reduce body weight by more than 20 kg within 6 months in seriously overweight people. The table below shows the body weights of a random sample of 15 people before and after undertaking this program. At the 1% significance level, test whether this claim is true.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before	88	94	121	160	138	115	105	112	109	99	123	135	150	155	142
After	65	68	94	136	118	95	88	92	89	73	100	114	134	133	120
Diff.	23	26	27	24	20	20	17	20	20	26	23	21	16	22	22

Summary statistics	Before	After	Difference
Mean	123.07	101.27	21.80
Standard deviation	22.68	23.61	3.167

>>>>>>>>>>

>>>>>>>>>>

If this was done as a left-tailed test (consistently), you would get the following:

Parameter: $\mu_d = \mu_{After} - \mu_{Before}$

H0: $\mu_{After} - \mu_{Before} = -20kg$

Ha: $\mu_{After} - \mu_{Before} < -20kg$

$$t = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}} = \frac{-21.8 - (-20)}{3.167 / \sqrt{15}} = \frac{-1.8}{0.818} = -2.200$$

What would happen if you did this test without setting a hypothesized difference (in other words the Ho value = 0)?

$$t = \frac{\bar{d} - \Delta_0}{s_d / \sqrt{n}} = \frac{21.8 - 0}{3.167 / \sqrt{15}} = \frac{21.8}{0.818} = 26.650$$

P = 1.059×10^{-13}

Another examples of Pairing: The heights of randomly selected *Acropora formosa* colonies at Mbudya Island and Bongoyo are shown below. However, unlike the example on page 18 where all measurement were recorded at the reef crest, here the colonies were measured at various depths (below mean sea level) along the slope of the reef.

Depth (m)	1	1	2	2	3	3	4	5	6	7	8	9	10
Mbudya (cm)	68	67	70	73	88	83	82	74	83	77	84	79	90
Bongoyo (cm)	70	71	69	71	79	74	77	73	69	72	72	70	82
Diff. In height	-2	-4	1	2	9	9	5	1	14	5	12	9	8

Note: Measurements are paired at the same depths on the two reefs, therefore, the paired t-test should be performed.

2.3 Parametric Methods, Transformations, and Nonparametric Methods (Including application in Comparing Several Populations, Topic 3)

2.3.1 Selection of the Most Appropriate Statistical Procedures

Criteria to be considered when selecting of the most appropriate analysis are as follows:

- **Type of response variables being recorded, or data scale used**
  - categorical, ordinal, quantitative
- **Research/Sampling design**
  - Types of explanatory variables/factors (categorical, quantitative, etc.)
  - Number of explanatory variables/factors and response variables
  - Number of populations/levels/treatments/groups of each factor being compared
  - Independent design versus paired or blocked design
- **Type of hypothesis being tested:**
  - Difference between groups/populations/treatments
  - Relationships between variables
- **Nature of the data obtained (check the assumptions)**
  - Do they fit the assumptions of a parametric test?
  - After transformation, do they fit the assumptions of a parametric test?
  - If not, do they fit the assumptions of a nonparametric test?

Parametric Statistical methods:

- Involve the estimation of population parameters, (e.g., population mean) based on sample statistics (e.g., sample mean), thus the name "parametric".
- Have certain underlying assumptions about the distributions of the populations being tested.
- When all assumptions are met, these are slightly more powerful than nonparametric tests.

Transformations

- Nonnormality, outliers and unequal standard deviations can often be corrected by transformations.
- Often one transformation can correct several violations of the assumptions at the same time.
- If the data fit the assumptions of a parametric test after transformation, perform the parametric test on the transformed data, and back-transform to the original scale when drawing conclusions.

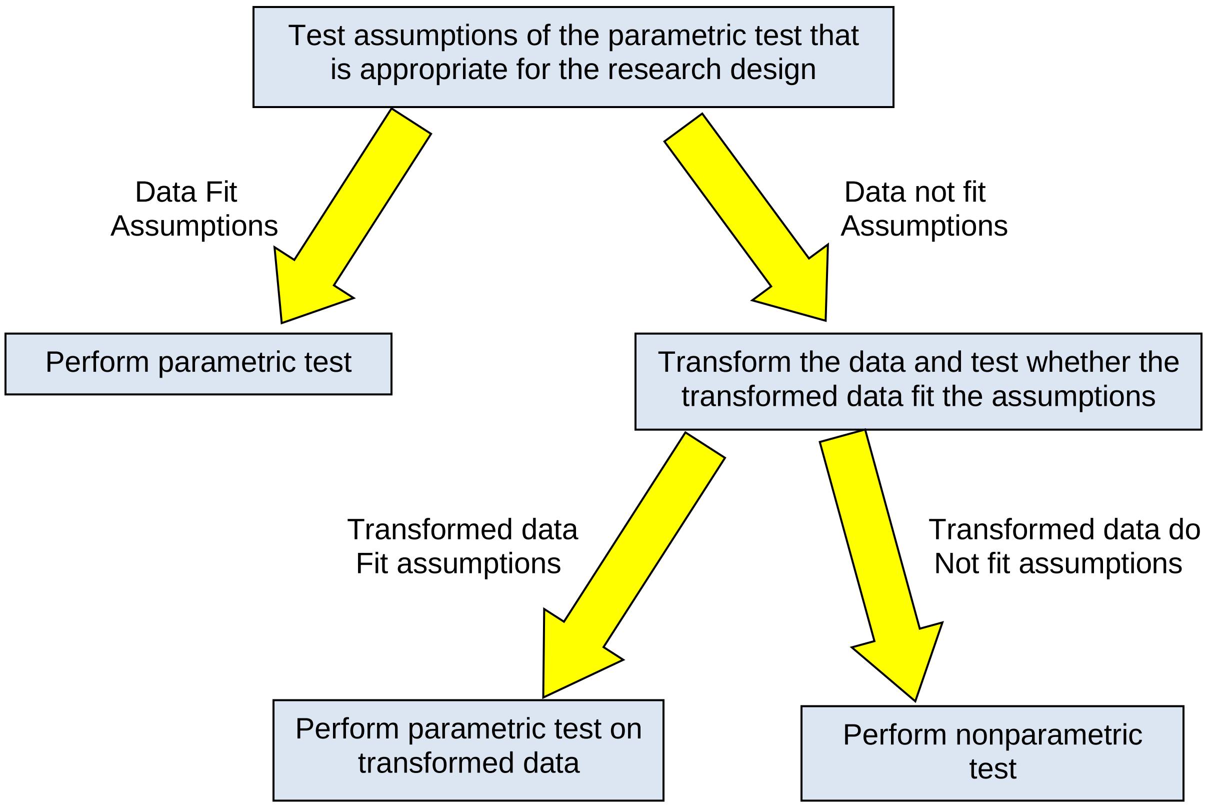
Nonparametric Methods

- Do NOT use estimates of population parameters in their calculations.
- Have fewer assumptions about the nature of the distributions of the populations being studied.
- Therefore, they can often be applied in cases when the parametric methods are not valid.
- If the assumptions of a parametric test are met, the corresponding nonparametric test can also be performed, but it is slightly less powerful (approximately 95% as powerful).
- Where data do NOT fit the assumptions of parametric tests, the corresponding nonparametric test is sometimes more powerful.
- Most nonparametric tests convert the data to ranks to perform the calculations; therefore, they are <u>not affected by outliers</u>.

Table: Parametric tests and their nonparametric equivalents.

Parametric test	Nonparametric test
One-sample t-test	Wilcoxon signed-rank test (for one sample)
Two sample t-test (independent samples)	Mann-Whitney U test (also called Wilcoxon rank sum test)
Paired-sample t-test	Wilcoxon paired-sample test
One-factor ANOVA	Kruskal-Wallis test
Pearson Correlation Test	Spearman rank correlation test
Randomized block ANOVA	Friedman's Test

Overall Approach to Choosing Parametric Tests, Data Transformations, or Nonparametric Tests



2.3.2 Summary of the Assumptions for Various Parametric Methods (Including ANOVA)

- Generally, the assumptions are the same for any given hypothesis test and its corresponding confidence interval (e.g., Pooled t-test and Pooled t-confidence interval).

Simple Random Sampling

- Required by all statistical inference.

Independence of Sampling

- Required only for the **two-sample t-test for independent samples (pooled and nonpooled).**
- And **One-way ANOVA.**

Normally Distributed Data

- **One-sample t-test:** The one sample must come from a population that is normally distributed, or the sample size should be large ($n \geq 30$, CLT).
- **Two-sample t-test for independent samples (Pooled or Nonpooled):** Both samples must come from normally distributed populations, or both sample sizes must be large ($n \geq 30$, CLT).
- **Paired-sample t-test:** Differences between paired observations must be normally distributed (the two separate populations may not necessarily be normally distributed), or the sample size should be large ($n \geq 30$, CLT).
- **One-way ANOVA:** All samples being compared must come from normally distributed populations

Equal Standard Deviations (or Variances)

- Required for the **Pooled Two-sample t-test for Independent Samples**
- **ANOVA:** All samples being compared must have equal variances (or approximately so).

No Serious Outliers

- Required by all parametric tests.

2.3.3 Assessing/Testing for Violations of the Assumptions & Robustness of the t-Tools and ANOVA

Planning for and Assessing Simple Random Sampling
- Must be planned as a basic part of the research design.
- No transformation can correct the sampling design once the research has been conducted.

Planning for and Assessing Independence of Sampling
- Should be planned as a basic part of the research design**.**
- If it is realized after the study that there is some kind of pairing of the observations, you can switch to doing a paired test.
- Cannot be corrected by any transformation.

Assessing/Testing for Normality
- **Histograms, Stem-and-leaf diagrams and Dotplots**
  - Perform a visual assessment by to see if the distribution forms a bell-shaped curve.
  - Very subjective.
- **Normal probability plot (also called the Q-Q Plot)**
  - Also used to <u>detect outliers</u>.
  - Used to assess each data set separately.
  - Normality assumption is <u>not</u> violated if all data points fall approximately in a straight line.
  - Normality assumption <u>is</u> violated if there are **serious departures** from a straight-line .
  - This guideline should be applied <u>loosely for small samples</u>, but <u>strictly for large samples</u>
  - Easier to evaluate than histograms, etc.
  - Somewhat subjective, but much easier to determine a straight line than a bell-shaped curve.
- **Hypothesis Tests**
  - Can be used to <u>make a very objective decision</u> about normality.
  - **Shapiro-Wilk Test** (Explained on page 5)
    - <u>Very powerful</u> and <u>accurate </u>for testing normality.
  - **Anderson-Darling Test (AD Test)**
    - Very similar to the Shapiro-Wilk Test

Assessing/Testing for Equal Standard Deviations
- **Boxplots**
  - Construct side-by-side boxplots on the same scale and compare their spread
  - Subjective method.
- **Ratio of the standard deviations**
  - If the ratio of the largest standard deviation divided by the smallest standard deviation is ≤ 2, we consider the standard deviations to be equal enough to perform the pooled t-test.
  - Not a very accurate method.
  - <u>Cannot</u> be applied for ANOVA because there are more than two populations.
- **Levene's Test for the Equality of Variances**
  - Perform in SPSS or other computer programs.
  - Very accurate way of assessing equality.
  - <u>Can</u> be applied for ANOVA.

Robustness and Resistance of the t-Tools
- Virtually no data set will fit all the assumptions perfectly.
- Generally, the t-tools and ANOVA are <u>**robust**</u> enough to withstand moderate violations of the assumptions without being seriously or significantly affected.

Departures from normality
- The t-tools are relatively robust to departures from normality.
- If normality is seriously violated, data transformation should be performed, or one can apply nonparametric tests.

Departures from equal standard deviations

- This is a crucial assumption.
- If standard deviations are significantly unequal, then the pooled estimate of the population standard deviation does not result in an accurate SE(Estimate) and the results of the pooled t-test are then invalid.

Departures from independence

- The SE(Estimate) in a t-test is calculated based on the independence assumption, so if the samples were not obtained independently, the t-test analysis may give very misleading results.
- The experiment may have to be repeated or, if significant pairing of the data occurs, the paired-sample t-test may be performed in place of the t-test for independent samples.

Resistance of t-tools to Outliers

- Drastic outliers often affect the t-tools more seriously than the violations of other assumptions.
- Consider the effects of outliers on the sample mean:

Data set I:	10, 20, 30, 50, 70	Sample mean = 36; Sample SD = 24.08
Data set II:	10, 20, 30, 50, 700	Sample mean = 162; Sample SD = 301.11

- Sample mean and sample standard deviation are NOT resistant to outliers.
- However, the median is resistant to outliers: For both samples, median = 30.
- Since t-tools are based on calculation of the sample mean and sample standard deviation, one or two drastic outliers can greatly affect the confidence interval as well as the t-statistic, changing the P-value and sometimes completely changing the conclusion.
- Transformations or nonparametric methods can be used when there are serious outliers.

2.3.4Transformations & Inference after a (Natural) Log Transformation

- Nonrandomness and Nonindependence - Cannot be corrected by transformations.
- Nonnormality, outliers and unequal standard deviations can often be corrected by transformations
- Often one transformation can correct several violations of the assumptions at the same time.

Logarithmic transformation

- The Log transformation is the most common and will be emphasized in this course.
- Both natural logarithms and log base 10 (common) can be used, but in this course, we will primarily use natural logarithms to the base ***e*** (where ***e*** = 2.7183).
- Data that becomes normally distributed after this transformation is referred to as <u>lognormal</u>.

Other Types of Transformations

- Other types of transformations include the square root transformation, reciprocal transformation, and arcsine square root transformation.

Inference after a (Natural) Log Transformation

Steps in Transforming Data and Making Inferences

1. Transform the data.
2. Check whether the transformed data fit the assumptions of the required test.
3. Perform the hypothesis test/ the confidence interval calculations on the transformed data.
4. Back-transform the estimate and the confidence interval to the original scale.
5. State the interpretations/conclusions on the original scale.

If the transformation is successful, the log-transformed data will be approximately symmetric such that:

Mean [ln(Y)] = Median [ln(Y)]

And since the log preserves ordering,

Median [ln(Y)] = ln[Median(Y)]

$\overline{LnY}_1$ and $\overline{LnY}_2$ represent the averages of the logged values of Sample 1 and Sample 2

Thus, $\overline{LnY}_1 - \overline{LnY}_2$ estimates $\ln[Median(Y_1)] - \ln[Median(Y_2)]$

And $\overline{LnY}_1 - \overline{LnY}_2$ estimates $= \ln\left[\dfrac{Median(Y_1)}{Median(Y_2)}\right]$

And $e^{\overline{LnY}_1 - \overline{LnY}_2}$ estimates $\left[\dfrac{Median(Y_1)}{Median(Y_2)}\right]$

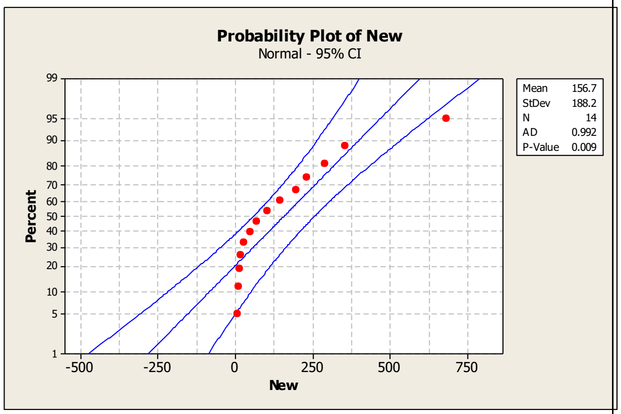
Example on Transformation & Inferences After Transformations:
Students were randomly allocated to two groups, one group to a new program and the other group to a standard program. At the end of the experiment, their test scores were recorded (scale = 0 – 700) as shown below. Since the test scores are not normally distributed and the standard deviations are very different, the log (natural) transformed data are also shown as well as summary statistics. At the 10% significance level, test whether there is a difference in the test scores of students undertaking the two programs and calculate 90% confidence limits.

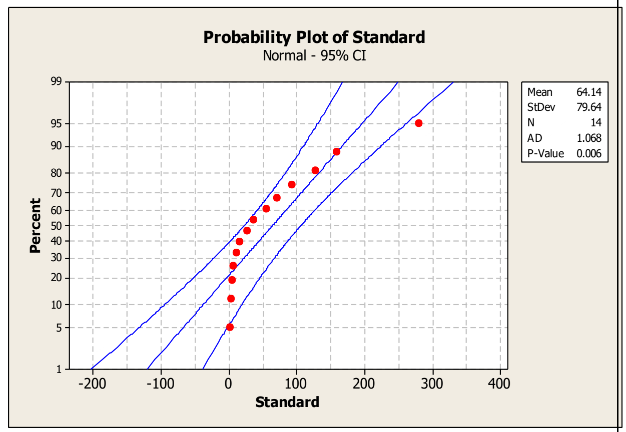
New	Standard	Ln(New)	Ln(Standard)
7	2	1.94591	0.693147
10	3	2.302585	1.098612
14	5	2.639057	1.609438
18	7	2.890372	1.94591
28	11	3.332205	2.397895
47	16	3.850148	2.772589
69	28	4.234107	3.332205
104	37	4.644391	3.610918
145	55	4.976734	4.007333
198	72	5.288267	4.276666
230	94	5.438079	4.543295
288	128	5.66296	4.85203
356	160	5.874931	5.075174
680	280	6.522093	5.63479

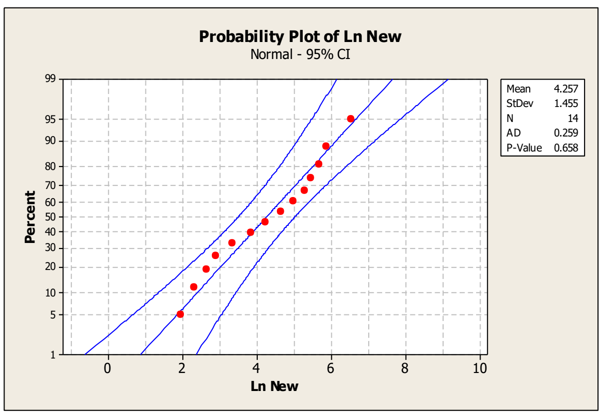
Descriptive Statistics: New, Standard, Ln New, Ln Standard

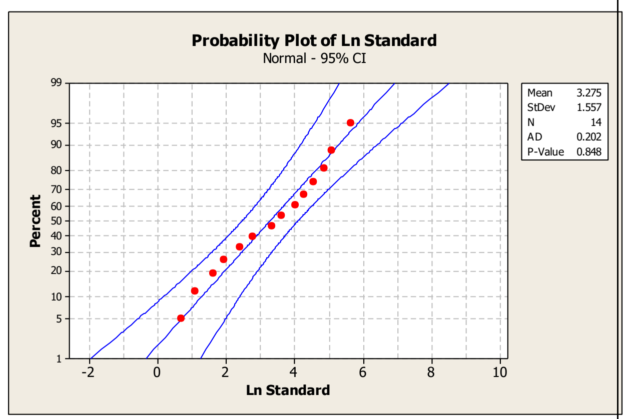
```
                Total
Variable        Count   Mean  SE Mean  StDev  CoefVar  Minimum     Q1  Median
New                14  156.7     50.3  188.2   120.11      7.0   17.0    86.5
Standard           14   64.1     21.3   79.6   124.15      2.0    6.5    32.5
Ln New             14  4.257    0.389  1.455    34.18    1.946  2.828   4.439
Ln Standard        14  3.275    0.416  1.557    47.55    0.693  1.862   3.472


Variable          Q3  Maximum  Skewness
New            244.5    680.0      1.85
Standard       102.5    280.0      1.78
Ln New         5.494    6.522     -0.17
Ln Standard    4.620    5.635     -0.21
```

Checking normality and equal standard deviations before and after transformations









Two-Sample T-Test and CI: Ln New, Ln Standard (Performed on log transformed data)

```
               N  Mean  StDev  SE Mean
Ln New        14  4.26   1.45     0.39
Ln Standard   14  3.28   1.56     0.42


Difference = mu (Ln New) - mu (Ln Standard)
Estimate for difference:  0.982
90% CI for difference:  (0.011, 1.954)
T-Test of difference = 0 (vs not =): T-Value = 1.72  P-Value = 0.096  DF = 26
Both use Pooled StDev = 1.5070
```

Conclusions based on the log transformed data:

Hypothesis test:
At the 10% significance level, there is moderate evidence that there is a difference in the means of the logged test scores between the new program and the standard program (Two-sample pooled t-test: t = 1.72, df = 26, P = 0.096).

Confidence interval:
The estimate of the difference between the means of the logged test scores of the new program and the standard program is 0.982 and the 90% confidence interval for the additive effect of the new program on the test scores is between 0.011 and 1.954. [Also, we can be 90% confident that there is a difference between the means of new and standard programs because 0 is not inside this confidence interval.]

Back Transformation of the Estimate and Confidence Int. and Interpretation on the original scale (Done by taking antilogs)
>>>>>>>>>>

>>>>>>>>>>
OR, we can say that the median test score for the new program is between 1.1% [(1.011 – 1) x 100)] and 605.7% [(7.057 – 1) x 100)] higher than the median test score for the standard program.

[**NOTE:** Also, this means that we can be 90% confident that there is a difference between the medians of new and standard programs because 1 (NOT 0) is not inside this confidence interval (1.011, 7.057).
This is because ln 1 = 0 and the antilog of 0 = 1(that is, $e^0 = 1$)

Note that:
The log-transformed data for both distributions are approximately symmetric
Mean [ln(Y)] = Median [ln(Y)]
For new program: 4.257 (mean) $\approx$ 4.439 (median)
For standard program: 3.275 (mean) $\approx$ 3.472 (median)
[**Note**: for symmetric distributions, the mean will always be approximately equal to the median; however, if a distribution has mean = median, this does not guarantee that it is symmetric.]
And since the log transformation preserves ordering,
Median [ln(Y)] = ln[Median(Y)]
For new program: The median of the logged values (4.439)
$\approx$ ln of the median of the original data (ln 86.5 = 4.460)
For standard program: The median of the logged values (3.472)
$\approx$ ln of the median of the original data (ln 32.5 = 3.481)

Furthermore:

$$e^{(\overline{LnY}_1-\overline{LnY}_2)}=e^{(4.257-3.275)}=e^{0.982}=2.670$$

estimates $\left[\dfrac{Median(Y_1)}{Median(Y_2)}\right]$ (population parameters)

estimated by $\dfrac{Median(sample1)}{Median(sample2)}=\dfrac{86.5}{32.5}=2.662$

This points to the <u>multiplicative</u> interpretation of the <u>ratio</u> of the population medians. Recall that the median is a better measure of the center of a skewed distribution than the mean.

<u>Back Transformation in Reverse</u>

[Reverse means: Subtracting Standard minus New, instead of New minus Standard (as above)]

Back Transformation of the estimate and confidence interval to the original data:

>>>>>>>>>>

>>>>>>>>>>

Comparing the two results

Estimate of the difference for New vs. Standard = 2.670

Estimate of the difference for Standard vs. New = 0.3746

2.670 is the <u>inverse</u> of 0.3746

Ratio of the endpoints of the confidence interval for New vs. Standard $=\dfrac{1.011}{7.057}=0.1433$

Ratio of the endpoints of the confidence interval for Standard vs. New $=\dfrac{0.1417}{0.9891}=0.1433$

2.3.5 Nonparametric Methods

2.3.5.1 Mann-Whitney U test

- Nonparametric equivalent of the two-sample t-test (Nonpooled or pooled)

Mann-Whitney U Test (= Wilcoxon Rank Sum Test) for Two Independent Samples

Step 1: Check the purpose and assumptions.

Purpose: To test for the difference between two populations.

Assumptions:
1. Simple random samples.
2. Independent samples.
3. Same-shape populations.

Step 2: State the null and alternative hypotheses,
Hypotheses are best stated in words, without referring to means.

Ho: The population distributions are not different. (Or the medians of the populations are not different.)
Ha: The population distributions are different, or one is greater or less than the other.
(Or the medians of the populations are different, or one is greater or less than the other.)

Step 3: Obtain the Calculated Value (or Observed Value) of the test statistic as follows:

First, rank all the data from both samples combined, from lowest to highest.
Assign average ranks where there are tied observations
M = sum of the ranks for sample data from Population 1

Step 4: Decide to reject H_0 or not reject H_0 and state the strength of the evidence against H_0

If $n_1 \le 10$ and $n_2 \le 10$, you must use technology or use the M-distribution table (though that will not be provided in this course).

If $n_1 > 10$ or $n_2 > 10$, use the standardized version of M (and use the P-value approach):
The sampling distribution of M is approximately normal with a mean of $\frac{n_1(1+n_1+n_2)}{2}$

and a standard deviation of $\sqrt{\frac{n_1 n_2(1+n_1+n_2)}{12}}$.

Standardized version of M is: $Z = \dfrac{M - \frac{n_1(1+n_1+n_2)}{2}}{\sqrt{\frac{n_1 n_2(1+n_1+n_2)}{12}}}$ ➔ **use standard normal table**

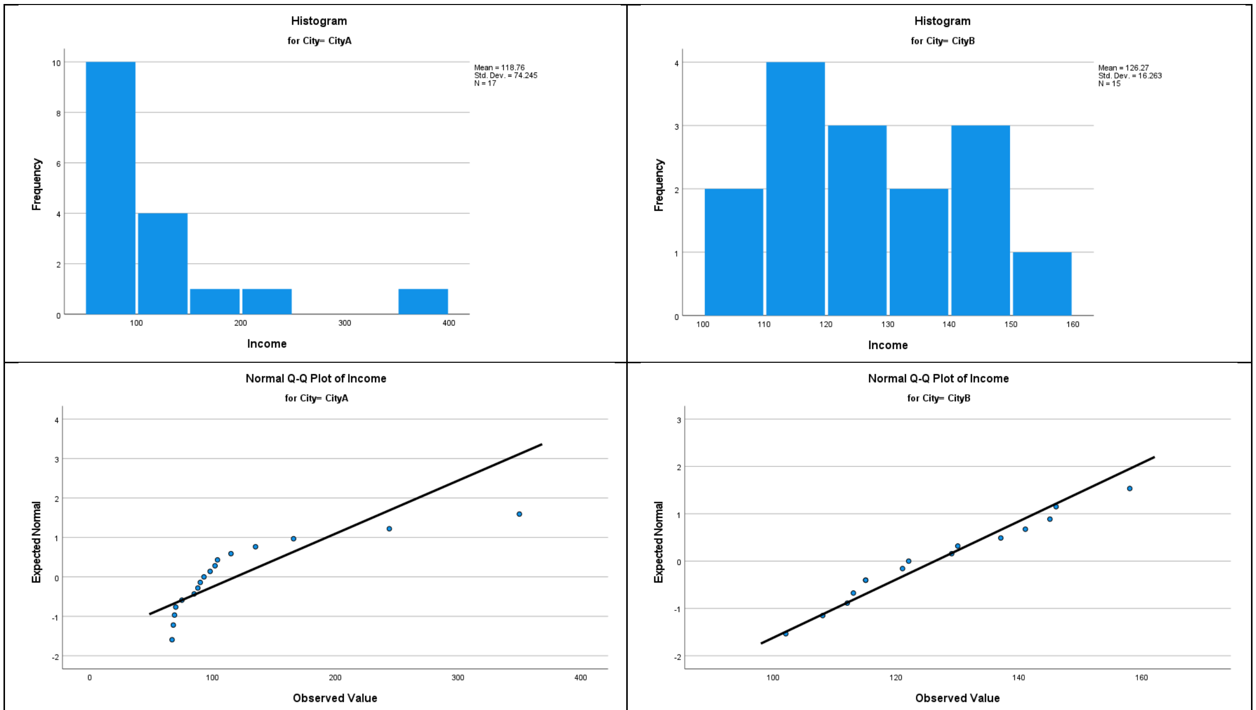
For a two-tailed test, **double** the P-value.

Step 5: Interpretation (conclusion) in words in terms of the research problem being investigated.

Example on the Mann-Whitney Test (both sample sizes > 10)
A study was conducted to compare the annual household incomes in two cities (City A and City B). A random sample of 17 households in City A and a random sample of 15 households in City B were selected. The table below shows the annual incomes of these households (in thousands of dollars) At the 5% significance level, perform the most appropriate test to determine whether there is a difference in annual household incomes between City A and City B?

	Annual household income (in thousands of dollars)																
CityA	102	135	98	67	350	75	90	68	244	70	166	85	115	69	93	88	104
CityB	145	122	115	102	130	112	115	129	137	146	108	158	141	121	113		

SPSS Output for Checking Assumptions on Original Data



City A	**City B**

Tests of Normality							
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	City	Statistic	df	Sig.	Statistic	df	Sig.
Income	CityA	.285	17	<.001	.682	17	<.001
	CityB	.156	15	.200*	.957	15	.647

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
Income	Based on Mean	6.429	1	30	.017
	Based on Median	2.603	1	30	.117
	Based on Median and with adjusted df	2.603	1	16.593	.126
	Based on trimmed mean	4.128	1	30	.051

Assumptions NOT Met for the parametric test (pooled t-test)

1. Shapiro-Wilk results show that, for City A, P < 0.001 (which is < 0.05); therefore, the distribution is different from a normal distribution. For City B, P = 0.647, so it is normally distributed.
2. Levene's test (based on means) gives P = 0.017 (which is < 0.05); therefore, there is sufficient evidence of a difference in variances.

Try Logged Data & Check the Assumptions Again

Tests of Normality							
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	City	Statistic	df	Sig.	Statistic	df	Sig.
LnIncome	CityA	.216	17	.034	.843	17	.008
	CityB	.150	15	.200*	.967	15	.805
*. This is a lower bound of the true significance.							
a. Lilliefors Significance Correction							

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
LnIncome	Based on Mean	7.855	1	30	.009
	Based on Median	5.008	1	30	.033
	Based on Median and with adjusted df	5.008	1	17.137	.039
	Based on trimmed mean	6.351	1	30	.017

STILL Assumptions NOT Met for the parametric test

1. Shapiro-Wilk results show that, for City A, P =0.008 (which is < 0.05); therefore, the distribution is different from a normal distribution. For City B, P = 0.805, so it is normally distributed.
2. Levene's test (based on means) gives P = 0.009 (which is < 0.05); therefore, there is sufficient evidence of a difference in variances.

Perform the Mann-Whitney U Test

Check purpose and assumptions
The purpose is to test for the difference between two means for populations that are independent. Random independent samples were taken. Examining the histograms, both distributions are right-skewed. Although one is more skewed than the other, that is OK.

>>>>>>>>>>

Calculations

City A		City B	
Income	**Rank**	**Income**	**Rank**
67		102	
68		108	
69		112	
70		113	
75		115	
85		115	
88		121	
90		122	
93		129	
98		130	
102		137	
104		141	
115		145	
135		146	
166		158	
244			
350			
Sum			

What would happen if we interchanged the two columns in the table of data and found the sum of the ranks for City B?

>>>>>>>>>>

SPSS procedure for the Mann-Whitney U Test:
Select: Analyze >> Nonparametric Tests >> Legacy Dialogs >> 2 Independent samples >> Select Test Variable "Income" >> Select Grouping Variable "City" >> Click Define Groups [Note: for this test, groups must be defined by numbers, NOT names] >> For Group 1, Type "1" for City A; Group 2, type "2" for City B >> Continue >> OK,

Ranks

	Group	N	Mean Rank	Sum of Ranks
Income	1	17	12.62	214.50
	2	15	20.90	313.50
	Total	32		

Test Statistics[a]

	Income
Mann-Whitney U	61.500
Wilcoxon W	214.500
Z	-2.493
Asymp. Sig. (2-tailed)	.013
Exact Sig. [2*(1-tailed Sig.)]	.011[b]

a. Grouping Variable: Group

b. Not corrected for ties.

2.3.5.2 Wilcoxon Paired-sample test (Paired Wilcoxon Signed Rank Test)

- Nonparametric equivalent of the paired-sample t-test,

Paired Wilcoxon Signed-Rank Test (Also called Wilcoxon paired-sample test)

Purpose: To test for the difference between two populations based on two paired samples,

Assumptions:

1. Simple random paired samples.
2. The paired differences should form a symmetric distribution.

Hypotheses

H_0: The population distributions (or the medians) are identical.
Ha: The population distributions (or medians) are different, or one is greater or less than the other

Obtaining the Observed Value of the test statistic and deciding to reject/not reject H_0.

If performing hand calculations (not in this course):

Find the absolute value of the paired differences, discard paired differences that equal 0, rank the absolute value of the differences, determine the signed ranks, calculate W as the sum of the positive ranks, and use the table of critical values of the *W*-distribution.

Performing the test using SPSS:

When you obtain the SPSS output (the procedure for doing that is shown below), it will provide you with the mean ranks, a ***z***-statistic and the ***P***-value based on the standard normal distribution table.

Conclusion:

Write an interpretation (conclusion) in words in terms of the research problem being investigated.

Example on the Wilcoxon Paired-Sample Test

Which movie is more interesting? A random sample of 16 people were taken from a certain population. They were provided with two movies to watch (Movie A and Movie B) in a random order, and then they were asked a number of questions. The responses of each participant were compiled to obtain a rating for each movie on a scale of 1 to 50, where 1 is the lowest rating and 50 was the highest. At the 5% significance level, perform the most appropriate test to determine whether there is a difference in rating between these two movies.

Table: Ratings given by each person, the paired differences, and absolute value of the paired differences.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
MovieA	22	24	25	27	27	29	30	32	34	35	36	36	37	39	40	41
MovieB	25	29	32	30	30	32	37	39	41	42	42	42	44	42	44	45
Diff	-3	-5	-7	-3	-3	-3	-7	-7	-7	-7	-6	-6	-7	-3	-4	-4
|Diff|	3	5	7	3	3	3	7	7	7	7	6	6	7	3	4	4

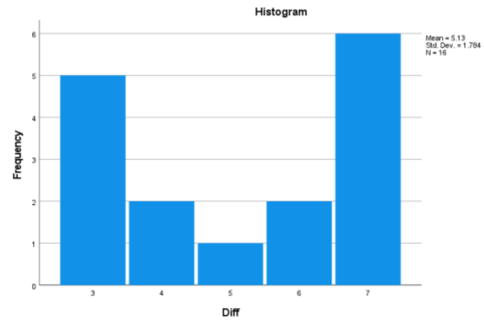
Test for Normality of Differences Based on Original Data

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MovieA	.123	16	.200*	.953	16	.540
MovieB	.218	16	.042	.881	16	.040
Diff	.228	16	.025	.790	16	.002

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Histogram of Absolute Value of Paired Differences (Original Data)



Test for Normality of Differences Based on Logged Data

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
LnMovieA	.148	16	.200*	.945	16	.409
LnMovieB	.226	16	.029	.870	16	.027
LnDiff	.221	16	.035	.785	16	.002

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Assumptions for the Parametric Paired-sample t-test NOT met

- Based on the Shapiro-Wilk Test, the paired differences are not normally distributed for either the original data (P = 0.002) or the logged data (P = 0.002).
- Therefore, perform the nonparametric equivalent, the Wilcoxon paired-sample test.

SPSS procedure for the Wilcoxon Paired-Sample Test:
Select: Analyze >> Nonparametric Tests >> Legacy Dialogs >> 2 Related Samples >> Select Test Pairs: “Movie A” for Variable 1; “Movie B” for Variable 2 >> Tick Test Type “Wilcoxon” >> Continue >> OK.

SPSS Output for the Wilcoxon Paired-Sample Test

Ranks				
		N	Mean Rank	Sum of Ranks
MovieA - MovieB	Negative Ranks	16[a]	8.50	136.00
	Positive Ranks	0[b]	.00	.00
	Ties	0[c]		
	Total	16		

a. MovieA < MovieB

b. MovieA > MovieB

c. MovieA = MovieB

Test Statistics[a]	
	MovieA - MovieB
Z	-3.550[b]
Asymp. Sig. (2-tailed)	0.000385
a. Wilcoxon Signed Ranks Test	
b. Based on positive ranks.	

Note: Perform the analysis **WITHOUT** using the numbers highlighted in yellow.

Perform the Wilcoxon Paired-Sample Test

Checking Assumptions

- The study was based on random paired samples/observations.
- The paired differences form a symmetric distribution (see Histogram above).

>>>>>>>>>>

>>>>>>>>>>