

SECTION 2: ONE OR TWO POPULATION MEANS

2.1 Inferences for One Mean (The One-Sample Case)

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

$$z = \frac{y - \mu}{\sigma} \quad \text{and} \quad z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \approx Z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} \rightarrow \text{replace } \sigma \text{ with } s \rightarrow t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}$$

- 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) and obtain the Student t (studentized) version of the sample mean (\bar{y})
- The Student t version of the sample mean can be converted to the one-sample t-test, which will be discussed in detail 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

$$t = \frac{\bar{y} - \mu}{s / \sqrt{n}}$$

has the t -distribution with $n - 1$ degrees of freedom (i.e., $df = n - 1$)

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 significance of 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 than the normal distribution, particularly at the base, and especially for small sample sizes.

Properties of the t -curve

(Properties 1 – 3 are in common with 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 $df = \infty$, the t -curve coincides with the normal curve.

Illustration of t -curve for $df = 1$, $df = 6$ and $df = 1000$ (Using MINITAB)

Using the t-Table

- A *t*-table can be either one-tailed or two-tailed or both.
- The table used in this course (NOT from the textbook) 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 $\alpha = 0.0005$ and $df = 12$, $t_{0.0005} = 4.318$
2. For a one-tailed test, when $\alpha = 0.025$ and $df = 75$, $t_{0.025} \approx 2.000$

2.1.1 Hypothesis Test for One Population

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:

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

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

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

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 $\leq \alpha$, we reject H_0 (otherwise do not reject H_0)

Step 5: Interpretation (Conclusion in Words)

The general formula for a test statistic is:

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

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).
- When the sample mean is known, only $n - 1$ observations are independent. One observation is not independent; it is fixed by knowing the other observations
- For the one-sample t-test, also $df = n - 1$ because standard deviation is in the denominator of the formula

Example of a One-Sample t-test: Comparing Two-tailed and Left-tailed Hypothesis Tests



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 researcher measures the height of randomly selected *Acropora formosa* colonies along the reef crest of Mbudya Island (Dar es Salaam), obtaining measurements (in cm) as shown in the table below. Summary statistics describing the data are also shown below. The population of heights of these colonies is known to be normally distributed, so even a small sample size is adequate. 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 different from the mean height of 75 cm for such colonies throughout the country.

Heights (in cm) of random samples colonies of <i>A. formosa</i> at Mbudya																		
Mbudya	81	48	74	69	79	56	59	64	51	61	72	74	84	81	67	57	69	68

Descriptive Statistics: *A.formosa*-reef crest

Variable	N	Mean	SE Mean	StDev	Min	Q1	Median	Q3	Max
Height (cm)	18	67.4444	2.4921	10.5731	48.00	58.50	68.50	75.25	84.00

Step 1: The one-sample t-test is selected

Purpose of the research problem: To test for the difference between one population mean (based on the sample mean) and the hypothesized mean of 75 cm.

Assumptions:

1. Random sample
2. Population normally distributed
3. σ is not known

>>>>>>

$$H_0: \bar{X}_{\text{cm}} = \mu \quad 0.05 \text{ p-value}$$

$$H_A: \bar{X}_{\text{cm}} \neq \mu \quad N = 18, df = 17$$

$$\bar{Y} = 67.444$$

$$SE(\bar{Y}) = 2.4921$$

$$t^* = -3.0318 = \frac{67.444 - 75.00}{2.4921}$$

$$(0.005 < p < 0.01) \times 2$$

p-value two-tailed

$$df(18) @ 0.025 = -2.110 : 2.110$$

$$p < \alpha = 0.05$$

$$|t^*| \leq df$$

therefore we reject H_0

>>>>>>

An Experienced Coral Reef Researcher's Approach

This researcher 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 re-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.

>>>>>>

$$H_0: \mu = 75\text{cm}$$

$$H_A: \mu < 75\text{cm}$$

$$N = 18$$

$$df = 17$$

$$P\text{-value } (0.005 > P > 0.0025)$$

$$P < \alpha = 0.05$$

So we reject H_0 .

>>>>>>

Note: In the above example, the evidence against H_0 is stronger than when performing the two-tailed hypothesis, but it is still within the same range of $0.001 < P \leq 0.01$ and considered as very strong evidence.

Exact probabilities for a Type I Error

- When we use statistical tables to find the Type I error, we can only say the p is greater than or less than or between certain values
 - 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
 - E.g. $P = 0.1332$ or $P = 0.00167$, etc.

Significance Level (α) versus P-Value

Distinction between Significance Level (α) and P-Value

- When planning a hypothesis test, the significance level (α) = Probability of a Type I error (the maximum that you can accept for a given question)
- P-Value read from a statistical table based on the calculated value of the test statistic = the observed Probability of the Type I error
- 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: $\text{Estimate} \pm \text{Critical Value} \times \text{SE(Estimate)}$

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

Step 1: Find the Critical value: For a given confidence level ($1 - \alpha$), 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: Two-sided confidence interval for μ is given by the endpoints:

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

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

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

This means a two-tailed test.

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)** = $t_{\alpha/2} \times \frac{s}{\sqrt{n}}$

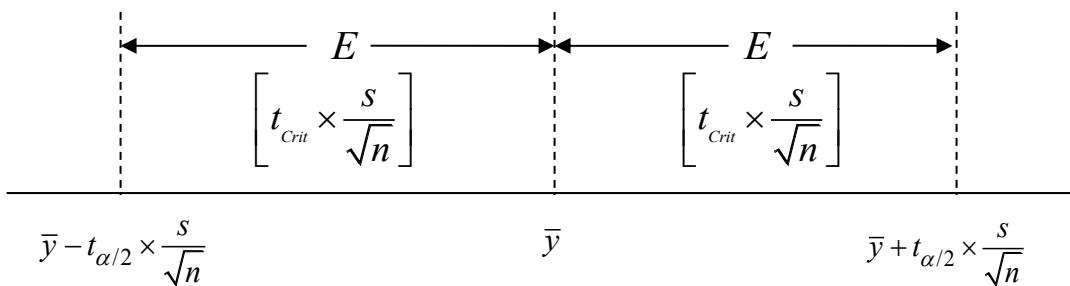
One-Sided t-intervals for One-Sample

Lower bound t-interval: $\bar{y} - t_{\alpha, n-1} \times \frac{s}{\sqrt{n}} \Rightarrow$ Consistent with a right-tailed test

Upper bound t-interval: $\bar{y} + t_{\alpha, n-1} \times \frac{s}{\sqrt{n}} \Rightarrow$ Consistent with a left-tailed test

Note: $t_{\text{Crit}} = t_{\alpha, n-1}$

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



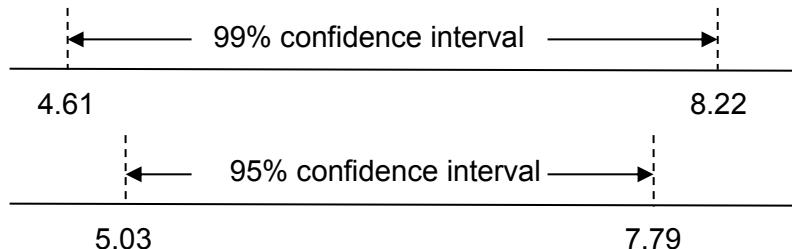
Margin of error = half the length of the confidence interval)

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

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 and improves 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.

Example: Using the data and information given for the hypothesis test above, calculate a 95% confidence interval for the mean height of colonies of *Acropora formosa* found on the reef crest of Mbudya Island (Dar es Salaam).

>>>>>>

$$95\% CI, \alpha = 0.05 @ df = n-1, 18-1 = 17$$

$$t_{\alpha/2, n-1} = t_{0.05/2, 18-1} = t_{0.025, 17} = 2.110$$

$$\text{Estimate} = \bar{y} = 67.444$$

$$SE \text{ of est.} = SEC\bar{y} = 2.4921$$

$$\bar{y} \pm t^* \times SEC\bar{y}$$

$$67.444 \pm 2.110 \times 2.4921$$

$$67.444 \pm 5.2583$$

$$(62.19, 72.70) \text{ cm}$$

>>>>>>

Relation between Hypothesis Tests and Confidence Intervals

SPSS is designed to only do two-tailed test and C.I.

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

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

- **The case of rejecting H_0 :** The null hypothesis will be rejected if and only if the $(1-\alpha)$ confidence interval for μ does not contain the theoretical mean or value (μ_0).
- **The case of not rejecting H_0 :** The null hypothesis will not be rejected if the $(1-\alpha)$ confidence interval for μ does contain the theoretical mean or value (μ_0).

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

For a right-tailed hypothesis, the null hypothesis will be rejected in favour of $H_a: \mu > \mu_0$ if and only if the $(1 - \alpha)$ -level lower confidence bound for the population mean μ is greater than μ_0

For a left-tailed hypothesis, the null hypothesis will be rejected in favour of $H_a: \mu < \mu_0$ if and only if the $(1 - \alpha)$ -level upper confidence bound for the population mean μ is less than μ_0

Note: Confidence level = $1 - \alpha = 1 - \text{significance level}$

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: 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.

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

H_0 was rejected at $0.005 > P > 0.0025$ (exact P-value = 0.004). So, again we concluded that there was a difference. Therefore, this is even greater confirmation that the two-sided confidence interval would not contain the hypothesized value of 75 cm.

SPSS Output for Hypothesis Test

One-Sample Test

	Test Value = 75					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Mbudya	-3.032	17	.008	-7.55556	-12.8134	-2.2977

Interpretation:

Note: "Sig." is actually the P-value (not significance level). The significance level is not input into SPSS. The P-value given is always the 2-tailed value and cannot be changed.

Two-tailed hypothesis test for the difference between the mean height of *Acropora formosa* on the reef crest at Mbudya Island and the hypothesized mean height of 75 cm for such colonies throughout the country

Result: $t = -3.032$, $df = 17$, $P = 0.008$

Note: The confidence interval shown when doing a hypothesis test is meaningless.

One-tailed (left-tailed) hypothesis test to determine whether the mean height of *Acropora formosa* on the reef crest at Mbudya Island is less than the hypothesized mean height of 75 cm for such colonies throughout the country

Result: $t = -3.032$, $df = 17$, $P = (0.008)/2 = 0.004$

you have to divide by 2

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

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Mbudya	27.063	17	.000	67.44444	62.1866	72.7023

- Note:** The test value must be changed from 75 to 0

*Set the test value to 0
to get P and the CI.*

Two-sided confidence interval

- The two-sided 95% confidence level is (62.19, 72.70) cm
- Though we don't input the test value of 75, since it is outside this interval, the mean height of *Acropora* at Mbudya is different from the hypothesized 75 cm
- Note:** the t-statistic (27.063) and P-value (0.000) for the hypothesis test are now meaningless.

SPSS Output for Confidence Interval: For one-sided (Upper bound 95% confidence interval)

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Mbudya	27.063	17	.000	67.44444	63.1092	71.7797

One-sided confidence interval (upper-bound)

- SPSS does not do one-sided confidence intervals or hypothesis tests, so to obtain this 95% upper bound, we had to "trick" SPSS by determining alpha as $0.05 \times 2 = 0.10$. Then, we input 90% (100% – 10%)
- So, the 95% upper bound is 71.78 cm, which can also be written as $(-\infty, 71.78)$ cm
- Since 75 is higher than the upper bound, that means that the mean height of Acropora on the reef crest at Mbudya is less than the hypothesized value of 75.
- Note:** the lower bound shown (63.11) is meaningless

>>>>>>

Illustration:

$$\text{For } 95\% \text{ C.I.} \quad \alpha = 1 - 0.95 = 0.05 \quad \text{df} = n - 1 = 17 \quad t_{\alpha} = t_{0.05} = 1.740$$

Hand calculations for a 95% upper bound confidence interval

$$U.B \bar{Y} + t_{0.05, 17} \times S.E(\bar{Y})$$

$$67.4444 + 1.740 \times 2.4921$$

$$64.4444 + 4.336$$

$$(-\infty, 71.78) \text{ cm}$$

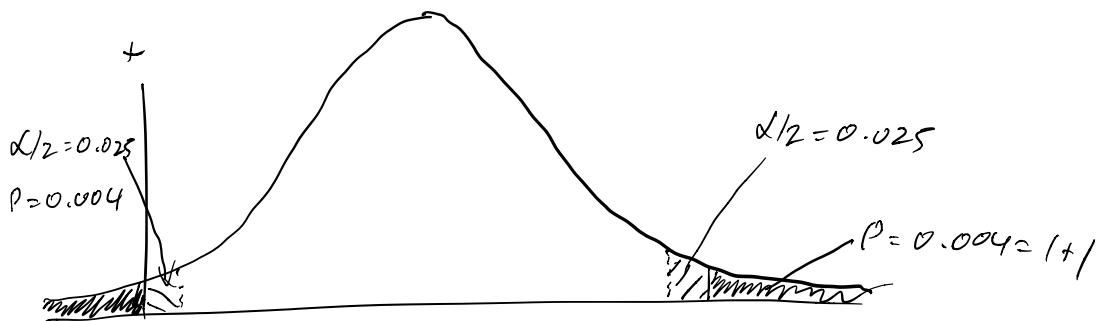
we can be 95% confident that ... mean is shorter than 71.78 cm

>>>>>>>

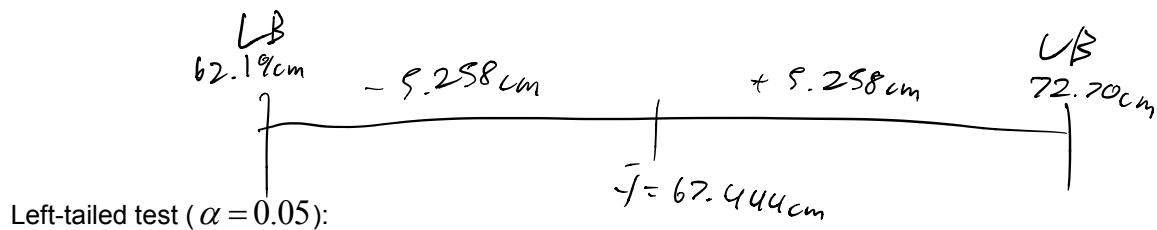
Illustrations

>>>>>>

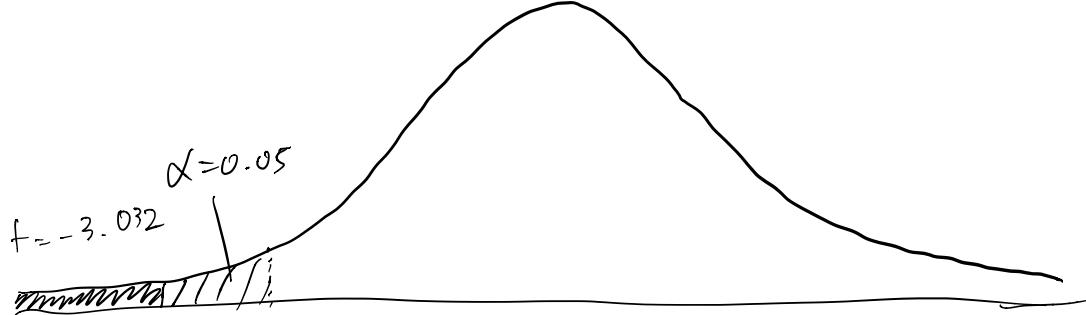
Two-tailed test ($\alpha = 0.05$)



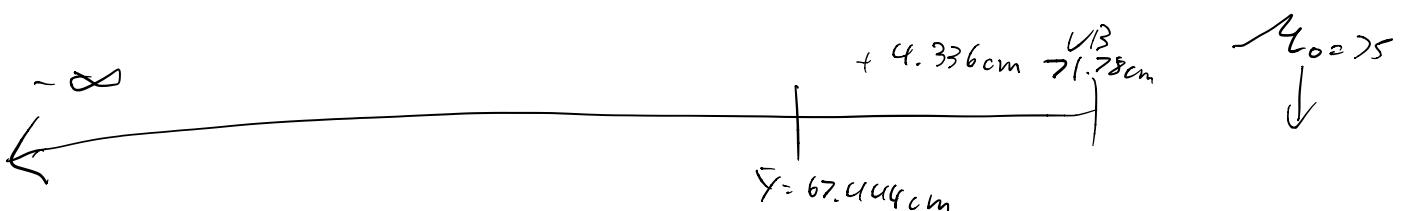
Two-sided confidence interval (95%):



Left-tailed test ($\alpha = 0.05$):



Upper bound confidence interval (95%):



A one-sided confidence interval will be shorter on one side than the two-sided confidence interval and it will go to infinity on the other side.

>>>>>>

2.2 Inferences for Two Means (The Two-Sample Case)

In this Section we are talking about the case where:

- We have **one variable**
- We are comparing **two populations** or groups by taking **two samples**
- Two sample inference:
 - Comparing **two population means**: **Independent samples**
 - Comparing **two population means**: **Paired samples**
 - Comparing **two proportions** (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

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

Suppose that x 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 equals the square root of the sum of the population variances, each divided by the corresponding sample size:

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}$$

- $\bar{y}_1 - \bar{y}_2$ 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 the general case of comparing two populations (independent samples) and can be applied to all situations, whether standard deviations are equal or not equal, and regardless of sample sizes

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

- Simple random samples (also implies also independent sampling within samples)
- Both populations are normally distributed or both samples are large
- Samples are independent

Test statistic:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}}$$

$$\text{Where } df = \frac{[(s_1^2 / n_1) + (s_2^2 / n_2)]^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}} \text{ (rounded down to nearest integer)}$$

- You will not need to perform this hypothesis test in this course, but you need to understand the difference between this test and the pooled t-test described below and to understand the assumptions of each.

Example of the Two-Mean t-test (Nonpooled t-test):

The weights of random samples of chickens (in kg) of two different breeds (A and B) are shown below. Both populations are normally distributed, so even small samples are adequate. At the 5% significance level, determine if there is a difference in the weights of Breed A and Breed B.

	Sample 1 (Breed A)	Sample 2 (Breed B)
	1.5	1.4
	1.0	1.1
	1.3	1.2
	1.9	1.3
	1.7	1.2
	1.5	1.1
		1.2
		1.4
		1.3
		1.2
		1.3
Mean	$\bar{x}_1 = 1.483333$	$\bar{x}_2 = 1.245455$
Standard deviation	$s_1 = 0.312517$	$s_2 = 0.103573$

Note: Standard deviations are very different and sample sizes are different, but the nonpooled t-test can still be applied.

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 sample 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)

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]
5. Sample sizes should be roughly equal

Null and Alternative Hypotheses:

The null hypothesis is $H_0: \mu_1 = \mu_2$ and the alternative hypothesis may be one of the following:

$$H_a: \mu_1 \neq \mu_2 \quad \text{or} \quad H_a: \mu_1 < \mu_2 \quad \text{or} \quad H_a: \mu_1 > \mu_2$$

(two-tailed) (left-tailed) (right-tailed)

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

Where s_p (pooled standard deviation) is given by:

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

$$df = n_1 + n_2 - 2$$

Or, the formula can be broken down as follows:

Parameter	Estimate	SE(Estimate) of the <u>difference</u> between the means	H_0 value	Reference Distribution
$\mu_1 - \mu_2$ <u>(Assume $\sigma_1 = \sigma_2$)</u>	$\bar{y}_1 - \bar{y}_2$	$s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ Where $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$	0 (or Δ_0)	$t_{n_1+n_2-2}$

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

Note: If you obtain a df that is not shown in the t -table, always go to the df below that.

If the P-value $\leq \alpha$, we reject H_0 (otherwise do not reject H_0)

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: For a given confidence level $(1 - \alpha)$ at $df = n_1 + n_2 - 2$

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

Two-sided interval:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \times s_p \sqrt{(1/n_1) + (1/n_2)} \quad \text{or} \quad (\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, n_1+n_2-2} \times SE(\bar{y}_1 - \bar{y}_2)$$

Lower bound t-interval: $(\bar{y}_1 - \bar{y}_2) - t_\alpha \times SE(\bar{y}_1 - \bar{y}_2) \Rightarrow$ Consistent with a right-tailed test

Upper bound t-interval: $(\bar{y}_1 - \bar{y}_2) + t_\alpha \times SE(\bar{y}_1 - \bar{y}_2) \Rightarrow$ Consistent with a left-tailed test

Where s_p = pooled standard deviation

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

Or, the formula can be broken down as follows:

Parameter	Estimate	Critical value	SE(Estimate) of the <u>difference</u> between the means
$\mu_1 - \mu_2$ (<u>Assume</u> $\sigma_1 = \sigma_2$)	$\bar{y}_1 - \bar{y}_2$	Two-sided $t_{\alpha/2, n_1+n_2-2}$ One-sided t_{α, 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.

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

Note: Confidence level = $1 - \alpha = 1 - \text{significance level}$

Example of the Pooled t-test (also a Two-Mean t-test)

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. Both populations have colony heights that are normally distributed, so even small sample sizes are adequate. 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.

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

Step 1: Pooled-sample t-test is selected because

Purpose of the study: To test for a difference between two population means

Assumptions:

1. Two independent random samples
2. Both populations are normally distributed
3. The two standard deviations are nearly equal ($10.57/10.38 < 2$) and sample sizes are nearly equal.

Step 2:

$H_0: \mu_1 = \mu_2$ (There is no difference in the mean heights of colonies of *Acropora formosa* colonies found on the reef crests of these two islands.)

$H_a: \mu_1 \neq \mu_2$ (There is a difference in the mean heights of colonies of *Acropora formosa* colonies found on the reef crests of these two islands.)

Parameter: $\mu_1 - \mu_2 = \mu_{Mbudya} - \mu_{Bongoyo}$

(mean height of *A. formosa* at Mbudya – mean height at Bongoyo)

>>>>>

Estimate of the diff between means = $\bar{Y}_1 - \bar{Y}_2 = 67.444 - 71.9375 = -4.4931\text{cm}$

Estimate of pooled std dev

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(18-1)(10.57312)^2 + (16-1)(10.38248)^2}{16+18-2}} = 10.48418$$

SE of the estimate of the diff between means

$$SE = 10.484189 \sqrt{\frac{1}{18} + \frac{1}{16}}$$

$$SE_{\bar{Y}_1 - \bar{Y}_2} = 3.60228$$

$$t = \frac{-4.49}{3.60228} = -1.247 = t$$

$$df = 18+16-2 = 32$$

these are 1-tailed test double the p-value for 2-tailed test

$$p\text{-value } (0.197 > p > 0.10) \times 2$$

$$0.30 > p > 0.20$$

there is weak evidence against H_0 $p > \alpha$ of 0.05

>>>>>

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).

>>>>>

95% C.I $\alpha = 0.05$

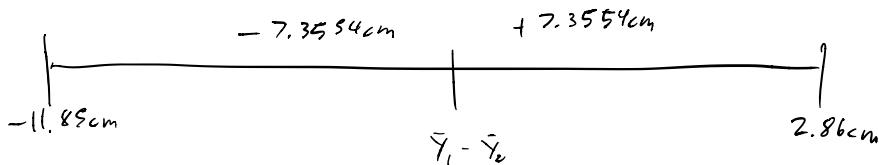
$$\textcircled{Q} df = n_1 + n_2 - 2 = 32 \quad t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.042$$

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{0.025} \times SE(\bar{Y}_1 - \bar{Y}_2)$$

$$-4.4931 \pm 2.042 \times 3.60228$$

$$-4.4931 \pm 7.35585576 \rightarrow ME$$

$$(-11.8489, 2.8628) \text{ cm}$$



>>>>>

Relation between Hypothesis Tests and Confidence Intervals

Relation between Hypothesis Tests and 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 : The null hypothesis will be rejected if and only if the $(1-\alpha)$ confidence interval for the difference between the population means ($\mu_1 - \mu_2$) does not contain 0, i.e. either both endpoints are negative OR both endpoints are positive.
- The case of not rejecting H_0 : The null hypothesis will not be rejected if the $(1-\alpha)$ confidence interval for the difference between the population means ($\mu_1 - \mu_2$) does contain 0, i.e. one endpoint is negative and the other is positive.

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

For a right-tailed hypothesis, the null hypothesis will be rejected in favour of $H_a: \mu_1 > \mu_2$ if and only if the $(1 - \alpha)$ -level lower confidence bound for the difference between $\mu_1 - \mu_2$ is positive (i.e., > 0)

For a left-tailed hypothesis, the null hypothesis will be rejected in favour of $H_a: \mu_1 < \mu_2$ if and only if the $(1 - \alpha)$ -level upper confidence bound for the difference between $\mu_1 - \mu_2$ is negative (i.e., < 0)

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

Return to the example of comparing the heights of *Acropora formosa* colonies at Mbudya and Bongoyo Islands.

Results from the hypothesis test showed that, at a significance level 5% ($\alpha = 0.05$), 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 $(1 - \alpha)\%$ 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: Hypothesis test and Confidence Interval for the Difference in Mean Height of *Acropora formosa* on the Reef Crests at Mbudya and Bongoyo Islands

is diff from the t-test

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
A.formosa	Equal variances assumed	.026	.873	-1.247	32	.221	-4.49306	3.60228	-11.83067	2.84456
	Equal variances not assumed			-1.249	31.662	.221	-4.49306	3.59831	-11.82565	2.83954

>>>>>

Levene's test

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

H_a : variance are diff for at least 2 pop.

Using the default α of 0.05

If p of sig. is > than 0.05, there is no sig diff.

>>>>>>

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
 - Measuring something like educational level of husbands and their wives
 - 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

The null hypothesis is $H_0: \mu_1 = \mu_2$ and the alternative hypothesis may be one of the following:

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

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

$$\text{Standard deviation of the mean difference} = s_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2 / n}{n-1}}$$

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

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

The formula can be broken down as follows:

Parameter	Estimate	SE(Estimate) of the mean difference	H ₀ value	Reference Distribution
μ_d (or $\mu_1 - \mu_2$)	\bar{d}	$SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$	$0 (or \Delta_0)$	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 $\leq \alpha$, we reject H_0 (otherwise do not reject H_0)

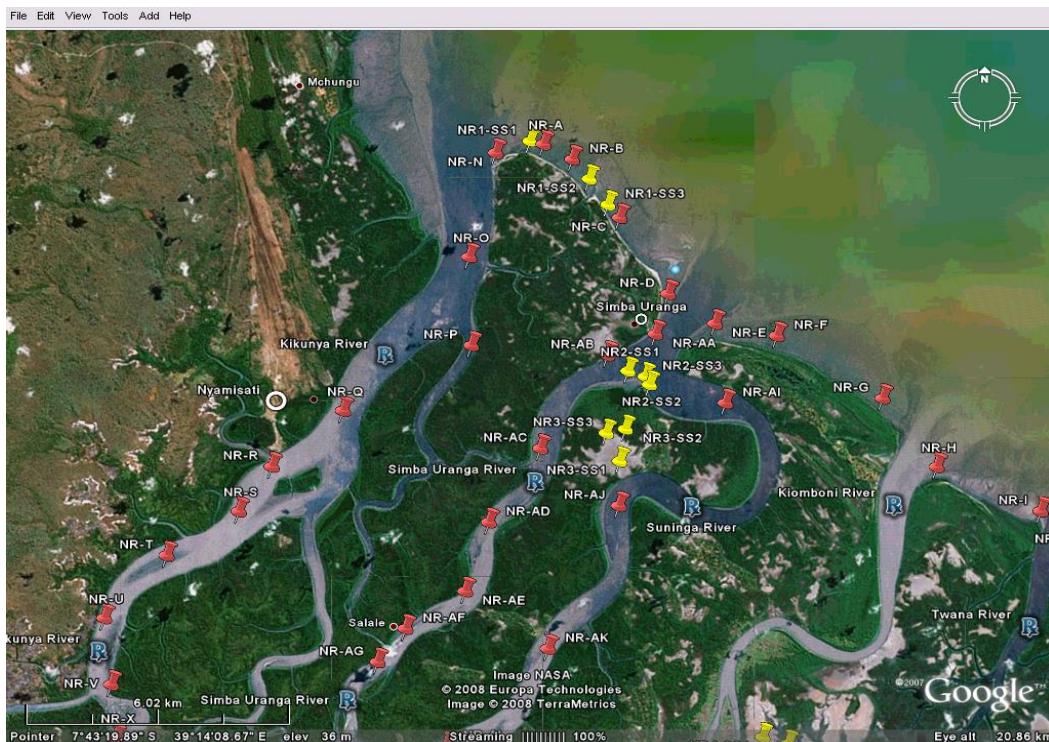
Step 5: Interpretation (conclusion) in words in terms of the research problem being investigated.

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:

- Permanent plots were established
- Girth at breast height of mangroves was measured to get mangrove basal area in permanent plots in 2007 and 2009.
- Thus, **the measurements are paired in space (same plots) for two time periods.**

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.



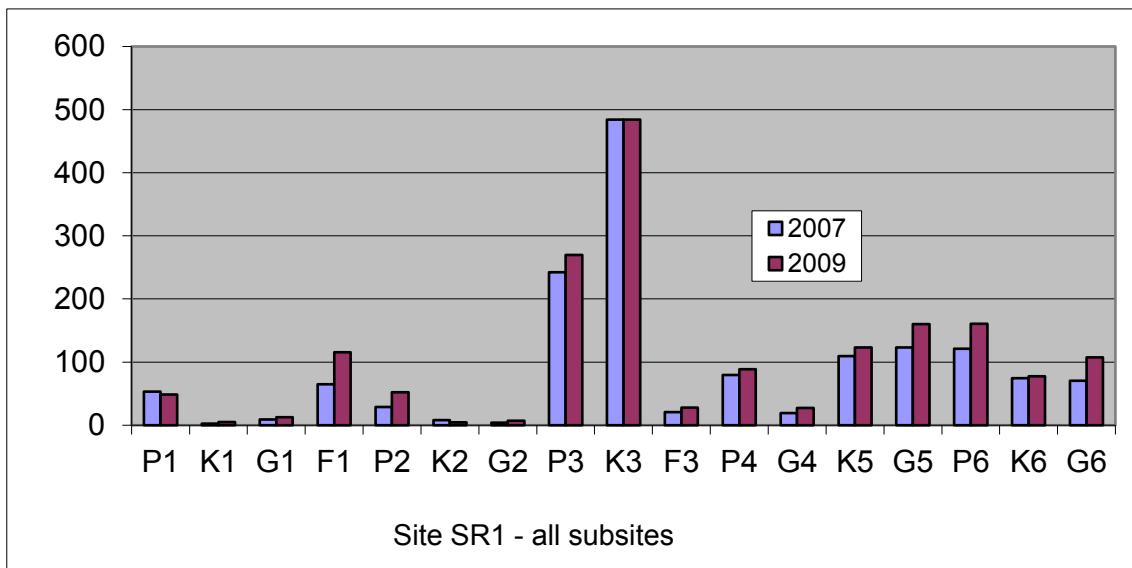
Avicennia mangroves invading into a saline flat area due to sea level rise, but some are also under stress due to decreased rainfall and humidity.

Research Hypothesis for landward edges: There was a difference/change in mangrove basal area (a measure of mangrove abundance) in Site SR1 (around the periphery of a saline flat area) between 2007 and 2009. It 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, since the direction of change was not predicted,

- this is a two-tailed test



Research Problem: At the 5% significance level, test whether mangrove basal area in Site SR1 changed between 2007 and 2009. (Reduced to 10 plots) Assume that the required assumptions are met.

Plot	Basal area (cm²/25-m² 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

Step 1: 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 paired samples
2. The differences between the paired observations are approximately normally distributed

Step 2: $H_0: \mu_1 = \mu_2$ (Mean mangrove basal area did not change between 2007 and 2009)

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

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

>>>>>>

Step 3:

	Basal area (cm ² /25-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			-82	1562

$$\bar{d} = \frac{-82}{10} = -8.2$$

SD of the d_i's

$$S_d = \sqrt{\frac{\sum d^2 - (\sum d)^2}{n-1}} = \sqrt{\frac{1562 - (-82)^2}{10}}$$

$$t = \frac{\bar{d}}{S_d / \sqrt{n}} = \frac{-8.2}{9.94205 / \sqrt{10}} = 3.143$$

df = 9

P-value (0.01 < P < 0.02) × 2 = 2.608
0.02 < P < 0.04

>>>>>>

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 all 17 plots 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 t-test that is applied for independent samples.

Note:

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

Paired t-Interval Procedure

Paired t-Interval Procedure

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

Assumptions: Same as for the Paired t-test

Step 1: For a given confidence level $(1 - \alpha)$, use the t-table to find $t_{\alpha/2}$ in the row for the appropriate df , where $df = n - 1$

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

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

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.

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

Example

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.

>>>>>>

For 95% CI, and $\alpha = 0.05$ @ $t_{\alpha/2} = t_{0.025} = 2.262$

$$\text{Parameters: } \mu_d = \mu_{2007} - \mu_{2009}$$

$$SE(\bar{d}) = 3.14395$$

$$\bar{d} \pm t_{\alpha/2} \times SE(\bar{d})$$

$$-8.2 \pm 2.262 \times 3.14395$$

$$(-15.31, -1.09)$$

It is estimated with a 95% C.I. that the diff. in ... between 2007 and 2009 was somewhere between -15.31 and -1.09. Note: ME = 7.11

Because 0 is not in the C.I., we reject H_0 .

>>>>>>

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	2007	89.9000	10	156.42282
	2009	98.1000	10	157.54396

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 2007 & 2009	10	.998	.000

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
2007 - 2009	-8.20000	9.94205	3.14395	-15.31212	-1.08788	-2.608	9	.028			

Research Problem on Exercise Program to Reduce Weight

It is claimed that a certain exercise program will reduce body weight by 20 kg or more 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

Step 1: Paired-sample t-test is selected because the purpose of the study is to test for a difference between two population means based on the difference in body weights that are paired on the same person (before and after).

>>>>>>

$$H_0: \mu_{\text{before}} - \mu_{\text{after}} = 20 \text{ kg}$$

$$H_A: \mu_{\text{before}} - \mu_{\text{after}} > 20 \text{ kg}$$

$$\text{Parameter: } \mu_d = \mu_{\text{before}} - \mu_{\text{after}}$$

$$\text{hypothesised diff} = \Delta_0 = 20 \text{ kg}$$

>>>>>>

Step 3:

Minitab Output

Paired T-Test and CI: Body wt-before, Body wt-after

	N	Mean	StDev	SE Mean
Body wt-before	15	123.07	22.68	5.86
Body wt-after	15	101.27	23.61	6.10
Difference	15	21.800	3.167	0.818

99% lower bound for mean difference: 19.654

T-Test of mean difference = 20 (vs > 20): T-Value = 2.20 P-Value = 0.022

By Hand Calculations

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

Step 4: $df = n - 1 = 15 - 1 = 14$

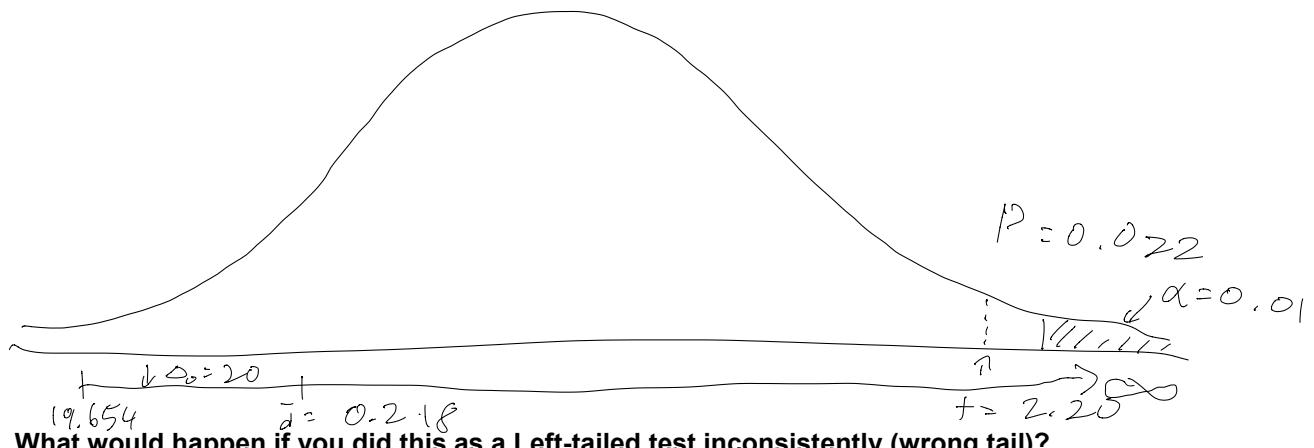
At $df = 14$, $P\text{-value} = 0.02 < P < 0.025$ (Minitab gives exact $P\text{-value} = 0.022$)

Since $P\text{-value} > \alpha (0.01)$, do not reject H_0 . There is strong evidence against H_0 but not very strong evidence as required by this research problem.

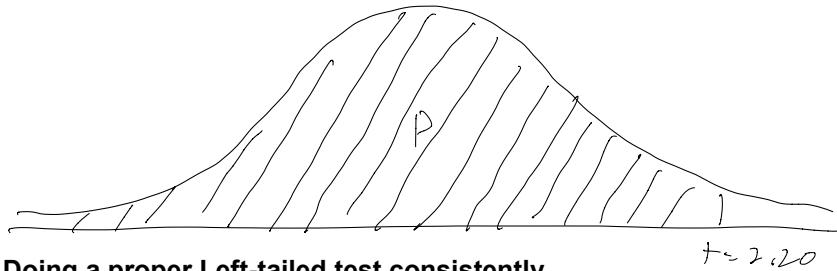
Step 5:

At the 1% significance level, the data do not provide sufficient evidence to prove the claim that the exercise program reduces body weight by 20 kg or more in seriously overweight people.

Sketch graph of Right-tailed test and Lower-bound t-interval



What would happen if you did this as a Left-tailed test inconsistently (wrong tail)?



Doing a proper Left-tailed test consistently

for a left tail test you need a negative diff

What would happen if you did this test without setting a hypothesized difference (in other words the H_0 value = 0)?

>>>>>

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

Research problem on corals: The heights of randomly selected *Acropora formosa* colonies at Mbudya Island and Bongoyo are shown below, along with summary statistics. However, this time the research design is planned such that it measures colonies at different depths (below mean sea level) along the slope of the reef. At the 5% significance level, test whether there is a difference in the mean heights of colonies of *Acropora formosa* colonies found on these two reefs.

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

Step 1: Paired-sample t-test is selected because the purpose of the study is to test for a difference between two population means, based on samples that are paired in space (paired at the same vertical depths on the reef).

Step 2: $H_0: \mu_{Mbudya} - \mu_{Bongoyo} = 0$

$H_a: \mu_{Mbudya} - \mu_{Bongoyo} \neq 0$

Parameter: $\mu_1 - \mu_2 = \mu_{Mbudya} - \mu_{Bongoyo}$ (mean heights of *A. formosa* colonies)

Step 3: (Minitab Output)

Paired T-Test and CI: Mbudya, Bongoyo

	N	Mean	StDev	SE Mean
Mbudya	13	78.31	7.49	2.08
Bongoyo	13	73.00	4.02	1.12
Difference	13	5.31	5.45	1.51

95% CI for mean difference: (2.01, 8.60)

T-Test of mean difference = 0 (vs not = 0): T-Value = 3.51 P-Value = 0.004

Step 4:

$$df = n - 1 = 13 - 1 = 12$$

At df = 12, P-value = $(0.001 < P < 0.0025) \times 2 = (0.002 < P < 0.005)$
(Minitab gives exact P-value = 0.004)

Since P-value < α (0.05), reject H_0 . There is very strong evidence against H_0 (< 0.01)

Note: the 95% confidence interval for the mean difference does not contain 0, which agrees with the conclusion of rejecting H_0 .

Step 5: Interpretation (Conclusion in Words)

At the 5% significance level, there is very strong evidence that there is a difference in the mean heights of *Acropora formosa* colonies along the reef slopes at Mbudya and Bongoyo Islands.

Coral reef researcher's interpretation: The greater height of *Acropora formosa* colonies along the reef slope at Mbudya compared to Bongoyo is likely explained by (or related to) two environmental factors:

1. Faster ocean currents at Mbudya due to its being more open to the ocean. Thus, coral polyps have access to a greater abundance of prey items and the symbiotic zooxanthellae (microscopic plants embedded in the coral tissue) can absorb nutrients more quickly.
2. Greater pollution of the seawater at Bongoyo due to a coastal sewage pipe. This pollution increases levels of phytoplankton, thus blocking light required by the zooxanthellae.

2.3 More on Assumptions of Statistical Inference (Including ANOVA)

2.3.1 Summary of the Assumptions for Various Hypothesis Tests/Confidence Intervals

- 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
- **Two-sample t-test for independent samples (Pooled or Nonpooled):** Both samples must come from populations that are normally distributed
- **Paired-sample t-test:** Differences between paired observations must be normally distributed (the two separate populations may not necessarily be normally distributed)
- **One-way ANOVA:** All samples being compared must come from normally distributed populations

Equal Standard Deviations (or Variances)

- **Two-sample t-test for Independent Samples (Pooled t-test only):** The standard deviations of the two samples must be equal (or approximately so)
 - Along with this the two sample sizes must be nearly equal
- **ANOVA:** All samples being compared must have equal variances (or approximately so)

2.3.2 Assessing/Testing for Violation of the Assumptions

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

- Required only for the **two-sample t-test for independent samples (pooled and nonpooled)** and **One-way ANOVA**
- Cannot be corrected by any transformation, but if it is realized later that there is some kind of pairing of the observations, you can switch to doing a paired test

Assessing/Testing for Normality

- **Histograms, Stem-and-leaf diagrams and Dotplots**
 - Perform a visual assessment by comparing the distribution of the population with a bell-shaped curve
 - Very subjective
- **Normal probability plot (also called the Q-Q Plot)**
 - Used to assess each data set separately
 - Perform a visual assessment by comparing the distribution of the population with a straight line
 - Normality assumption is not violated if the all data points fall approximately in a straight line
 - Normality assumption is is violated if there are **serious departures** from a straight-line pattern
 - Easier to evaluate than histograms, etc.
 - Still somewhat subjective, but much easier to determine a straight line than a bell-shaped curve

Assessing Normality Using a Normal Probability Plot (Similar to Q-Q Plots in interpretation)

- Normal probability plots can also be used to assess outliers

Guidelines for Assessing Normality Using a Normal Probability Plot

- If the plot is roughly linear, you can assume that the variable is approximately normally distributed.
- If the plot is not roughly linear, you can assume that the variable is not approximately normally distributed.

These guidelines should be applied:

- loosely for small samples, but
- strictly for large samples

• Hypothesis Tests

- Can make an objective decision
- Chi-square Goodness of Fit Test:** Expected frequencies are calculated using the proportions of the normal distribution and these are compared with the observed frequencies
- Anderson-Darling Test (AD Test)**
 - Very powerful for testing normality

>>>>> Hypotheses of AD test:

H_0 : The distribution is not diff from the normal distribution

H_1 : The distribution is diff from the normal distribution.

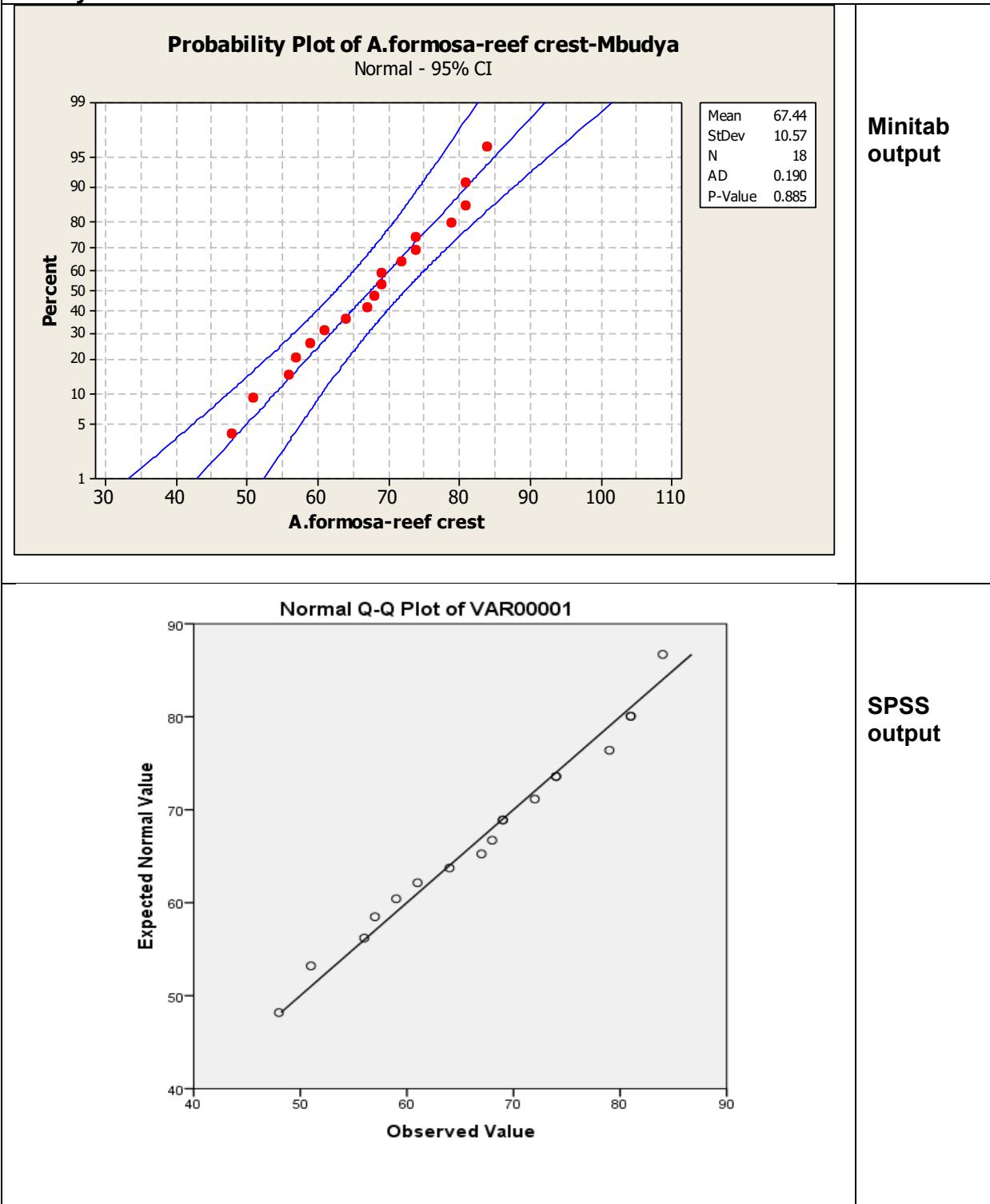
Default is $\alpha = 0.05$

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**
 - Calculate the ratio of the largest standard deviation divided by the smallest standard deviation. If this ratio is ≤ 2 , we may consider the standard deviations to be equal enough to perform the pooled t-test
 - Cannot be applied for ANOVA
- Levene's Test for the Equality of Variances**
 - Perform in SPSS
 - Very accurate way of assessing equality
 - Can be applied for ANOVA

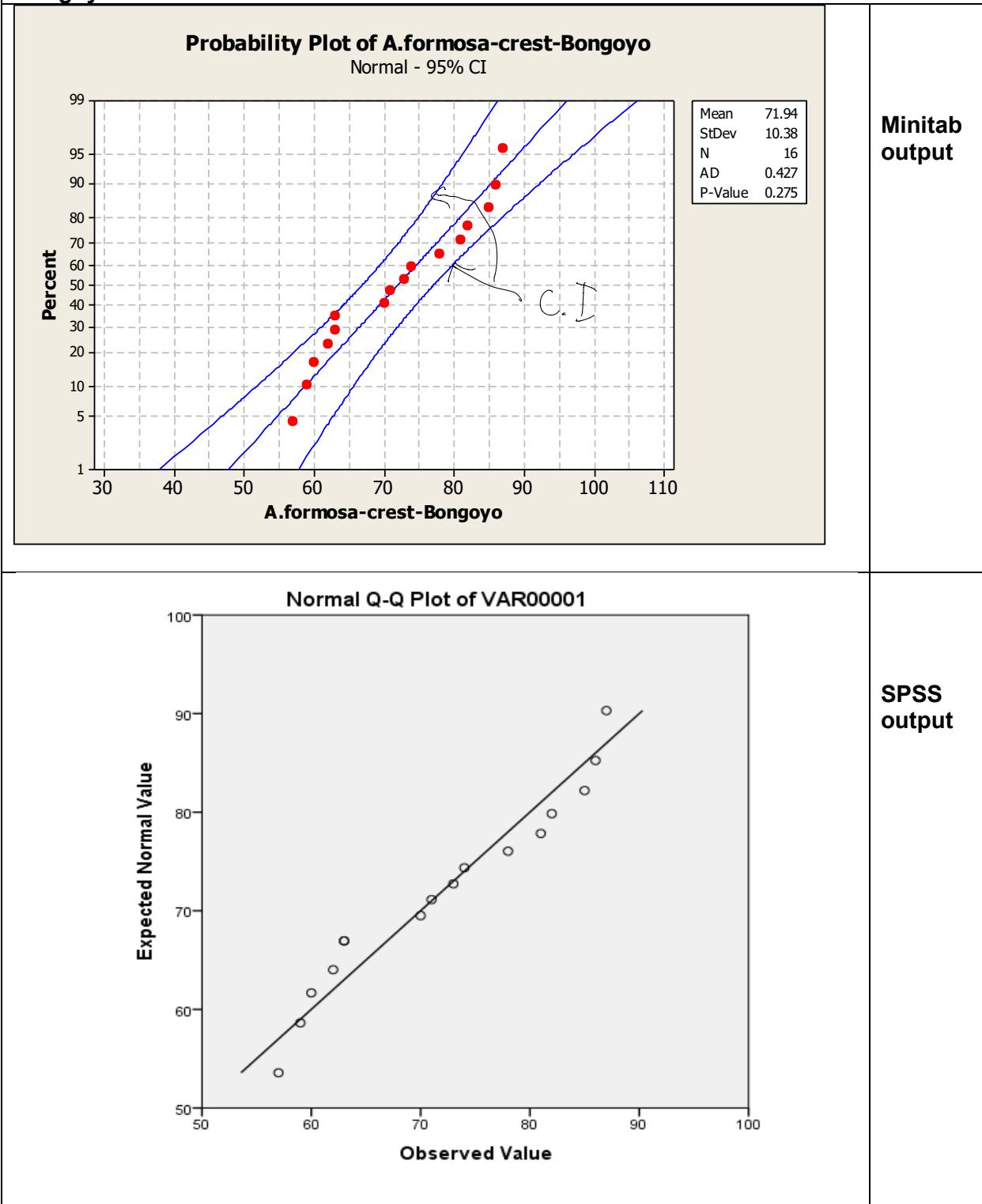
Testing for Normality in the Previous Examples

One-sample of data on the heights of Acropora formosa colonies on the reef crest at Mbudya Island



Testing for Normality in the Previous Examples

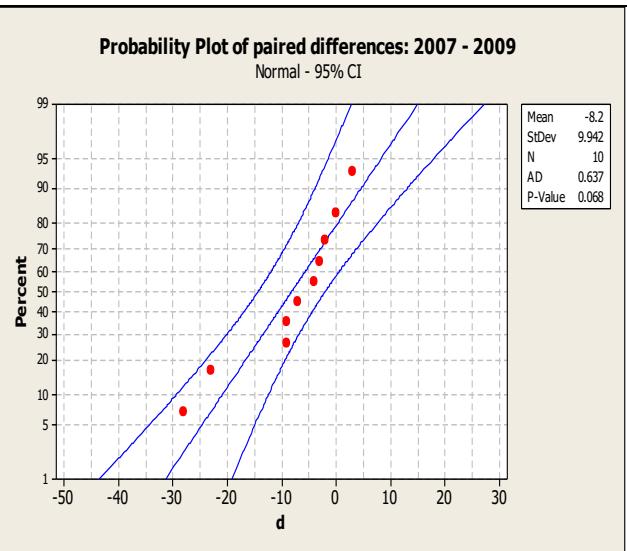
One-sample of data on the heights of *Acropora formosa* colonies on the reef crest at Bongoyo Island



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

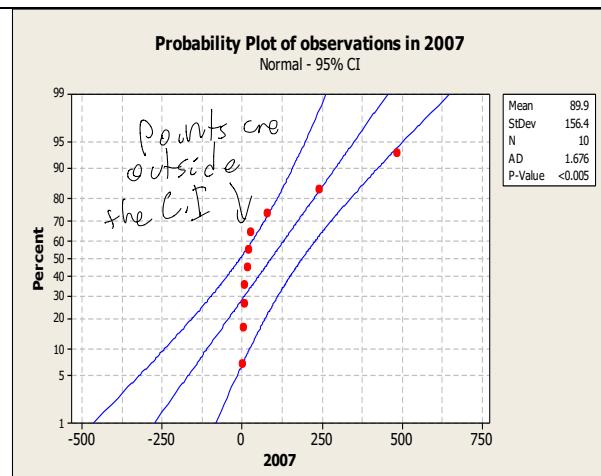
Paired design applied:

1. Much more powerful due to pairing in space
2. Separate samples for 2007 and 2009 seriously violate the assumption of normality



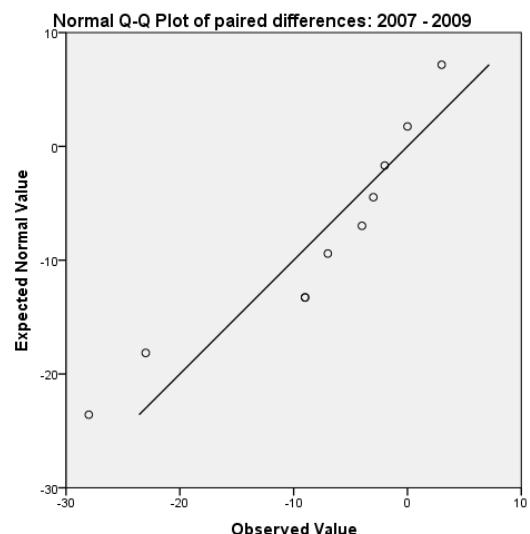
Paired differences between 2007 and 2009 are normally distributed because:

1. All data points fall within 95% CI lines
2. Anderson-Darling (AD) test statistic = 0.637; P-value = 0.068. Thus H_0 not rejected; no difference between this data set and a normal distribution



Mangrove abundance in 2007 are not normally distributed because:

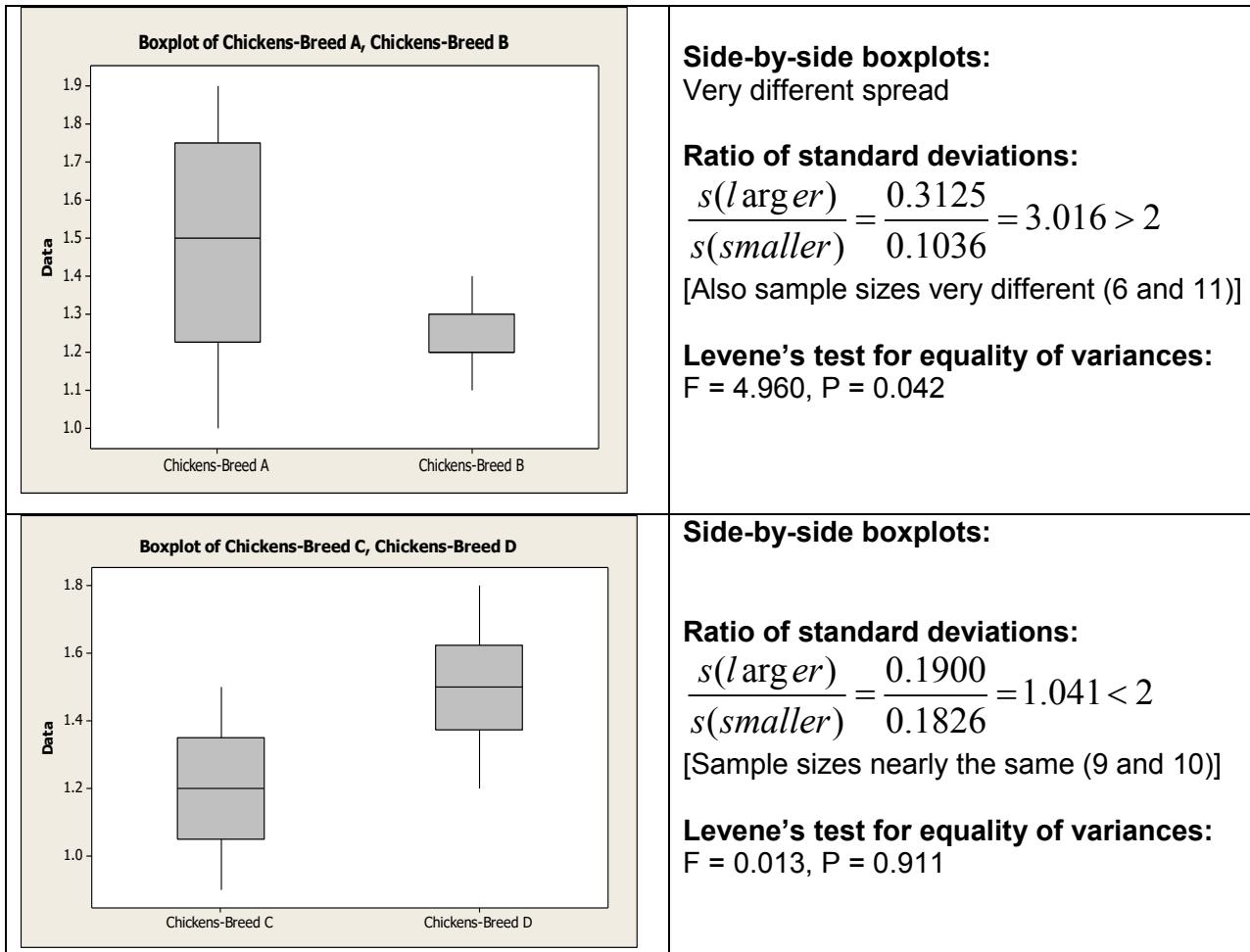
1. The plot is curved, and some data points fall outside the 95% CI lines
2. Anderson-Darling (AD) test statistic = 1.676; P-value < 0.005. Thus, H_0 is rejected; that is, there is a significant difference between this data set and a normal distribution



SPSS output:

Paired differences between 2007 and 2009

Assessing/Testing for Equal Standard Deviations



2.3.4 Robustness and Resistance of the t-Tools

- Virtually no data set will fit all the assumptions perfectly
- How much can a data set depart from the assumptions without the statistical procedures being significantly affected

Robustness = the ability of a statistical procedure to withstand departures or violations of the assumptions without being seriously or significantly affected

Departures from normality

- The t-tools are relatively robust to departures from normality
- Only if normality is seriously violated, will the t-test give unreliable results
- If normality is seriously violated, data transformation should be performed, or one can apply nonparametric tests

Departures from equal standard deviations

- This is the most 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 can often affect the t-tools more seriously than violations of some of the other assumptions
- Consider the effects of outliers on the sample mean:
Data set I: 10, 20, 30, 50, 70 Sample mean = 36
Data set II: 10, 20, 30, 50, 700 Sample mean = 162
- Sample mean is not resistant to outliers
- Standard deviation also 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, these tools are not resistant to serious outliers
- Thus, 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

2.4 Transformations

- 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

Types of Transformations

Logarithmic transformation

- The Log transformation is the most common type of transformation 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 lognormal

Square Root Transformation

- Take the square root of all observations
- Useful for data recorded in counts

Reciprocal transformation

- Useful for waiting times

Arcsine Square Root Transformation

- Useful for proportions

Analysis of Data After Transformations Have Been Applied

- After transformation, the t-tools are applied to the transformed data
- The analysis must be back-transformed in order to interpret the research problem and draw a conclusion
- Back-transformation is often the most difficult part
- However, with log transformed data, back transformation is relatively easy; therefore we will mainly be applying log transformations

Demonstration on Effects of Natural Log (base e) Transformations on Various Scenarios:

- 1. Data sets (A and B) that fit all assumptions**
 - The demonstration shows that, the two data sets were normally distributed, with no outliers, and had nearly equal standard deviations before transformation
 - The natural log transformation has little effect on the extent of normality and the equality of standard deviations
- 2. Data sets (C and D) with unequal standard deviations, but both were normal**
 - Before transformation, the ratio of larger SD to smaller SD = $29.31/1.912 = 15.33$ (much greater than 2, therefore unsuitable of the pooled t-test)
 - After transformation, the ratio is $0.4001/0.3626 = 1.10$ (suitable for the pooled t-test)
- 3. Data set that was very nonnormal (curved, right skewed) and had a very large SD**
 - Before transformation, the Anderson-Darling test statistic is 1.068 with P-value = 0.006, meaning that the distribution is very significantly different from normality
 - After transformation, AD is 0.202 with a P-value of 0.848, making it very normal (not significantly different from normality)
 - At the same time, the transformation changes SD from being huge (79.64) to being moderate 1.557, making it likely to be similar to other data sets with which you might want to compare it
 - Note: one natural log transformation corrected the data for 2 assumptions at the same time
- 4. Data set with two extreme outliers**
 - Before transformation, the AD test gave P-value < 0.005 and several data points are outside the 95% CI lines
 - After transformation, the AD test gave P-value = 0.028, which is just under 5%, making the transformed data not completely normal. However, the data points are almost within the CI lines and, as such the transformed data does not seriously violate normality. Apart from the two extreme outliers, the rest of the data are close to being normal.
 - This points to the fact that outliers are sometimes more difficult to correct than data that are nonnormal as indicated by a curved probability plot

2.5 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
5. State the conclusions on the original scale

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

$$\text{Mean } [\ln(Y)] = \text{Median } [\ln(Y)]$$

And since the log preserves ordering,

$$\text{Median } [\ln(Y)] = \ln[\text{Median}(Y)]$$

$\overline{\ln Y_1}$ and $\overline{\ln Y_2}$ represent the averages of the logged values of Sample 1 and Sample 2

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

And $\overline{\ln Y_1} - \overline{\ln Y_2}$ estimates $= \ln \left[\frac{\text{Median}(Y_1)}{\text{Median}(Y_2)} \right]$

And $e^{\overline{\ln Y_1} - \overline{\ln Y_2}}$ estimates $\left[\frac{\text{Median}(Y_1)}{\text{Median}(Y_2)} \right]$

Example: 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.

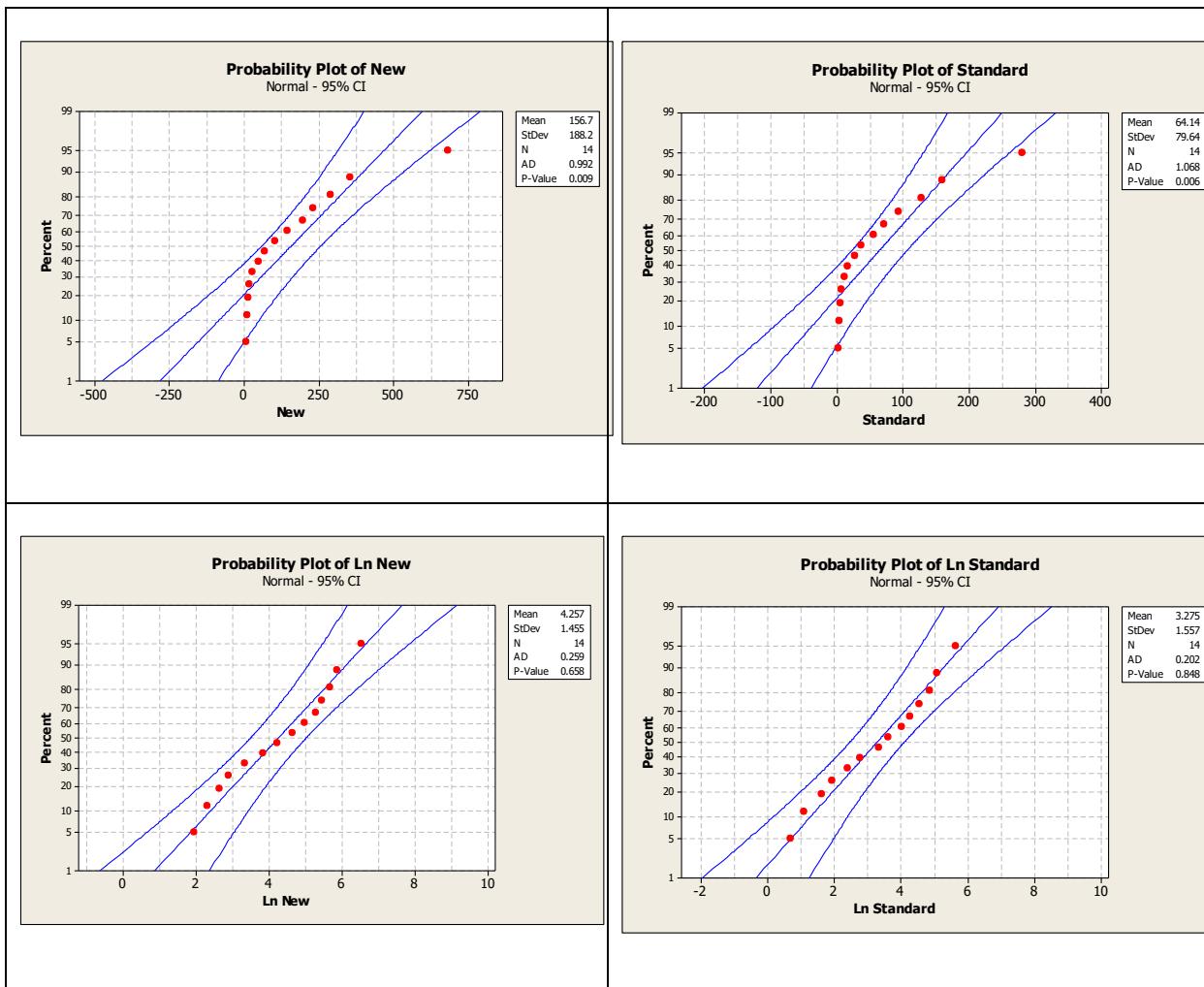
New	Standard	Ln(New)	Ln(Standard)
7	2	1.94591	0.693147
10	3	2.302585	1.098612
14	5	2.639057	1.609438
18	7	2.890372	1.94591
28	11	3.332205	2.397895
47	16	3.850148	2.772589
69	28	4.234107	3.332205
104	37	4.644391	3.610918
145	55	4.976734	4.007333
198	72	5.288267	4.276666
230	94	5.438079	4.543295
288	128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

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

Variable	Total								
	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)

>>>>>

$$\text{Estimate of } d\bar{Y} = e^{0.982} = 2.6698$$

$$L.B \text{ of C.I.} = e^{0.011} = 1.0111$$

$$U.B \text{ of C.I.} = e^{1.954} = 7.0569$$

Conclusions on the original scale (indicating the multiplicative effect of the treatment):

Conclusion of the hypothesis test:

At the 10% significance level: the median test score of those who took the new program is estimated to be 2.6698 times the median score of those in the test program.

At 90 % C.I for the ratio of the median in the original scale is

$$\left| \frac{\text{Med(new)}}{\text{Med(Old)}} \right| \text{ is } (e^{0.011}, e^{1.954}) = (1.0111, 7.0569)$$

Or it is estimated with 90% confidence that the median of the test program score of the new program is in between $\sqrt{ } \times$ times the median test score for the std prog.

>>>>>

OR, we can say that the median test score for the new program is between 1.1% $[(1.011 - 1) \times 100]$ and 605.7% $[(7.057 - 1) \times 100]$ 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)

[Please note that, 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 $[\ln(Y)]$ = $\ln[\text{Median}(Y)]$

For new program: The median of the logged values (4.439)

$\approx \ln \text{ of the median of the original data } (\ln 86.5 = 4.460)$

For standard program: The median of the logged values (3.472)

$\approx \ln \text{ of the median of the original data } (\ln 32.5 = 3.481)$

Furthermore:

$$e^{(\bar{\ln Y}_1 - \bar{\ln Y}_2)} = e^{(4.257 - 3.275)} = e^{0.982} = 2.670$$

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

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

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

Back Transformation in Reverse

[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:

>>>>> Estimate of the diff = $e^{-0.482} = 0.3746$

UB of the C.I. = $e^{-0.011} = 0.9891$

LB of the C.I. = $e^{-1.954} = 0.1417$

Conclusions on the original scale (indicating the multiplicative effect of the treatment)::

Conclusion based on the hypothesis test:

@ the 10% sig level the median test score of those who took the std program is estimated to be 0.3746 times the median test score of those in the new program .

Conclusion based on the confidence interval:

A ... ratio of the median is the original scale $\left| \frac{\text{med(std)}}{\text{med(new)}} \right|$
is $(e^{-1.954}, e^{-0.011}) = (0.1417, 0.9891)$

>>>>>

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 inverse of 0.3746

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

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

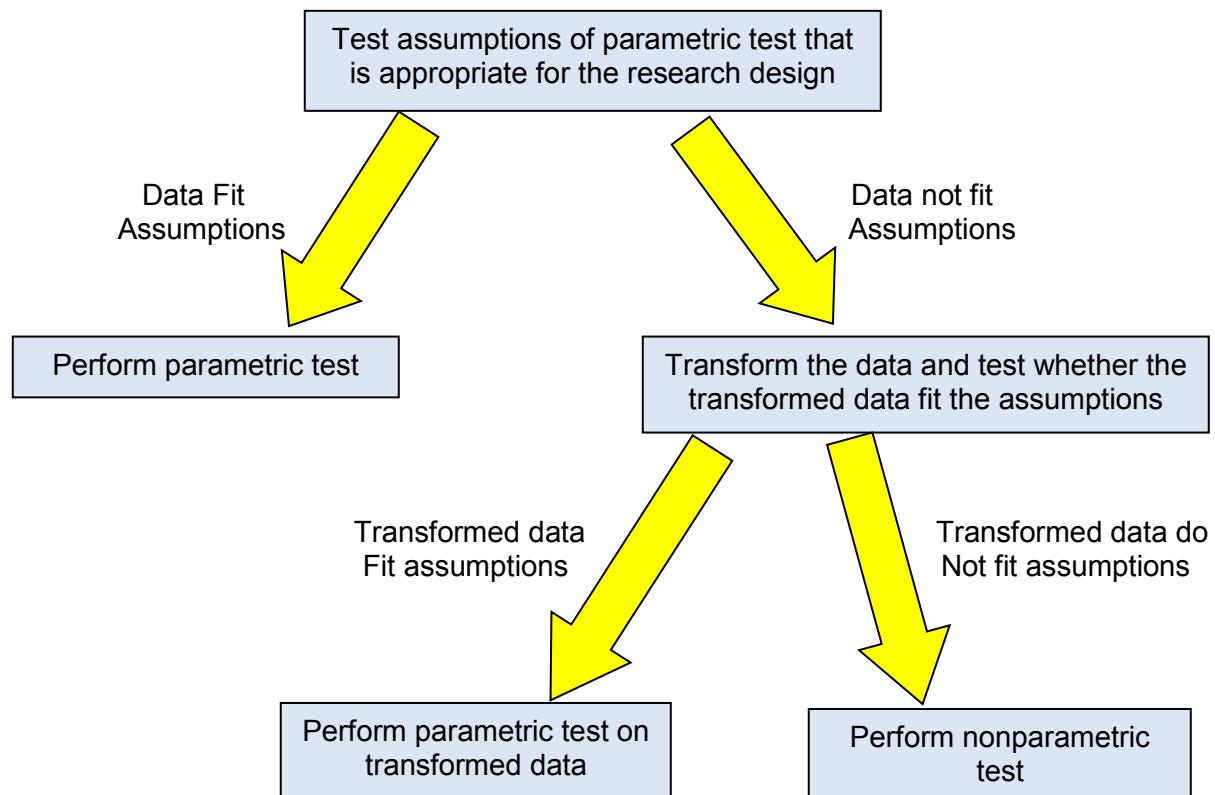
2.6 Nonparametric Methods

- If the data do not fit the assumptions of parametric tests, even after applying a transformation, then nonparametric methods should be performed
- Nonparametric methods do not use estimates of population parameters in their calculations
- Do not make any assumptions about the nature of the distribution of the populations being investigated
- Only assumption or requirement is that the samples must be selected randomly
- Therefore, they can be applied in many cases when the parametric methods are not valid
- Can be applied to categorical data, whereas parametric tests cannot.
- Slightly less powerful (approximately 95% as powerful)
- Most nonparametric tests convert the data to ranks and then calculations are performed on the ranks

Table: Parametric tests and their nonparametric equivalents

Parametric test	Nonparametric test
One-sample t-test	Wilcoxon signed-rank test
Two sample t-test (independent samples)	Mann-Whitney U test
Paired-sample t-test	Wilcoxon paired-sample test
One-factor ANOVA	Kruskal-Wallis test

2.6 Best Approach to Selecting Statistical Methods



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: One-way ANOVA, "F(3, 28)" OR a two-sample t-test, "t(16)"

Supplementary 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	Difference
Average	23.40	25.12	-1.72
Standard Deviation	5.42	5.87	1.43

Suppose that the numbers highlighted in yellow are not given.

- (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.

>>>>>

$$\text{ME} = \frac{UB - LB}{2}$$

$$\text{paired t-interval} = \left[-0.21906 - (-3.22094) \right]$$

$$= 1.50094$$

$$\bar{d} \pm t_{\alpha/2} \times \frac{s_d}{\sqrt{n}} = ME$$

$$t_{\alpha/2} = m \times \frac{\sqrt{n}}{s_d} = 1.50094 \times \frac{\sqrt{6}}{1.43}$$

$$t_{\alpha/2} = 2.571 = t_{0.025}$$

$$\alpha = (0.025) \times 2 = 0.05$$

>>>>>

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

>>>>>>

You get the idea. . .

>>>>>>

- (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.