

[Home](#) [Read](#) [Sign in](#)

NATURAL RESOURCES BIOMETRICS

CONTENTS

Chapter 2: Sampling Distributions and Confidence Intervals

Sampling Distribution of the Sample Mean

Inferential testing uses the sample mean (\bar{x}) to estimate the population mean (μ). Typically, we use the data from a single sample, but there are many possible samples of the same size that could be drawn from that population. As we saw in the previous chapter, the sample mean (\bar{x}) is a random variable with its own distribution.

- The distribution of the sample mean will have a mean equal to μ .
- It will have a standard deviation (standard error) equal to σ/\sqrt{n}

[Previous: Chapter 1: Descriptive Statistics and the Normal Distribution](#)

[Next: Chapter 3: Hypothesis Testing](#)

is not normally distributed or we don't know anything about the distribution of our population?

The Central Limit Theorem states that the sampling distribution of the sample means will approach a normal distribution as the sample size increases.

- So if we do not have a normal distribution, or know nothing about our distribution, the CLT tells us that the distribution of the sample means (\bar{x}) will become normal distributed as n (sample size) increases.
- How large does n have to be?
- A general rule of thumb tells us that $n \geq 30$.

The Central Limit Theorem tells us that regardless of the shape of our population, the sampling distribution of the sample mean will be normal as the sample size increases.

Sampling Distribution of the Sample Proportion

The population proportion (p) is a parameter that is as commonly estimated as the mean. It is just as important to understand the distribution of the sample proportion, as the mean. With proportions, the element either has the characteristic you are interested in or the element does not have the characteristic. The sample proportion (\hat{p}) is calculated by

$$\hat{p} = \frac{x}{n}$$

where x is the number of elements in your population with the characteristic and n is the sample size.

Example 1

$$\hat{p} = \frac{238}{950} = 0.25$$

The distribution of the sample proportion has a mean of $\mu_{\hat{p}} = p$

and has a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

The sample proportion is normally distributed if n is very large and \hat{p} isn't close to 0 or 1. We can also use the following relationship to assess normality when the parameter being estimated is p , the population proportion:

$$n\hat{p}(1 - \hat{p}) \geq 10$$

Confidence Intervals

In the preceding chapter we learned that populations are characterized by descriptive measures called parameters. Inferences about parameters are based on sample statistics. We now want to estimate population parameters and assess the reliability of our estimates based on our knowledge of the sampling distributions of these statistics.

Point Estimates

We start with a point estimate. This is a single value computed from the sample data that is used to estimate the population parameter of interest.

- The sample mean (\bar{x}) is a point estimate of the population mean (μ).
- The sample proportion (\hat{p}) is the point estimate of the population proportion (p).

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

- A confidence interval is an interval of values instead of a single point estimate.
- The level of confidence corresponds to the expected proportion of intervals that will contain the parameter if many confidence intervals are constructed of the same sample size from the same population.
- Our uncertainty is about whether our particular confidence interval is one of those that truly contains the true value of the parameter.

Example 2

We are 95% confident that our interval contains the population mean bear weight.

If we created 100 confidence intervals of the same size from the same population, we would expect 95 of them to contain the true parameter (the population mean weight). We also expect five of the intervals would not contain the parameter.

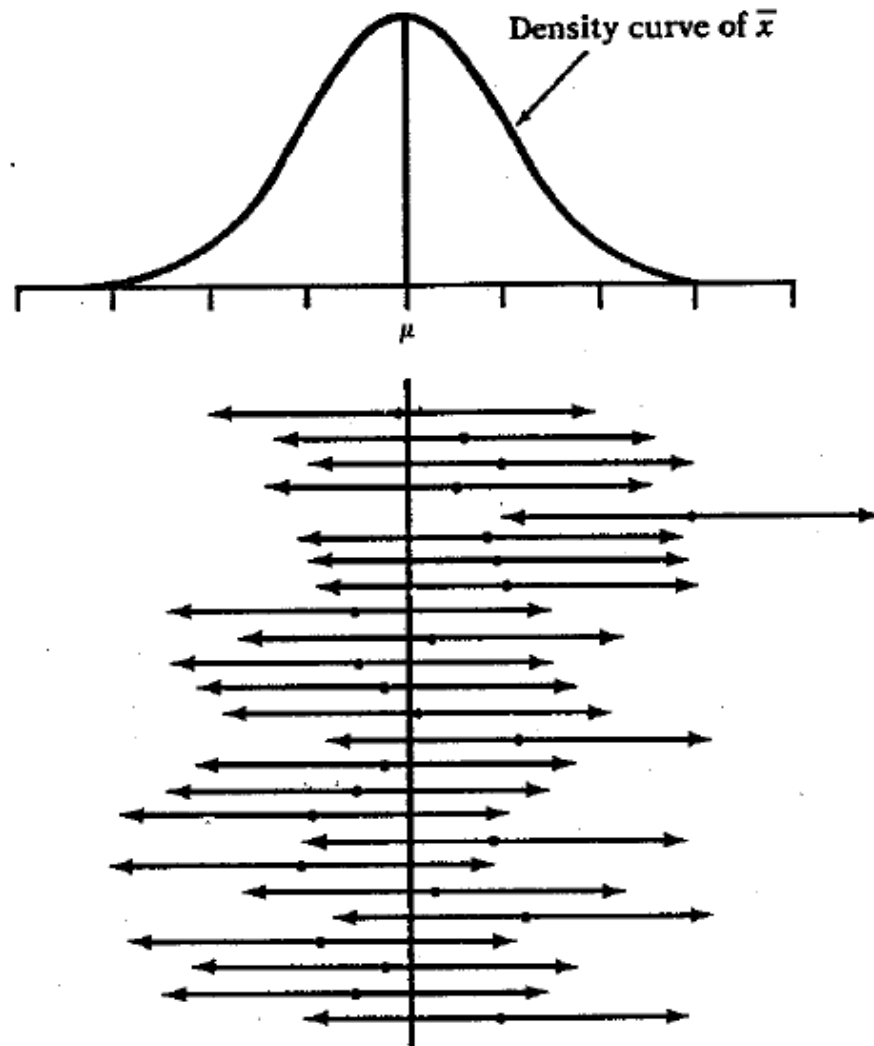


Figure 1. Confidence intervals from twenty-five different samples.

In this example, twenty-five samples from the same population gave these 95% confidence intervals. In the long term, 95% of all samples give an interval that contains μ , the true (but unknown) population mean.

Level of confidence is expressed as a percent.

- The complement to the level of confidence is α (alpha), the level of significance.
- The level of confidence is described as $(1 - \alpha) * 100\%$.

- We use a point estimate (e.g., sample mean) to estimate the population mean.
- We attach a level of confidence to this interval to describe how certain we are that this interval actually contains the unknown population parameter.
- We want to estimate the population parameter, such as the mean (μ) or proportion (p).

$$\bar{x} - E < \mu < \bar{x} + E \text{ or } \hat{p} - E < p < \hat{p} + E$$

where E is the margin of error.

The confidence is based on area under a normal curve. So the assumption of normality must be met (see [Chapter 1](#)).

Confidence Intervals about the Mean (μ) when the Population Standard Deviation (σ) is Known

A confidence interval takes the form of: **point estimate \pm margin of error.**

The point estimate

- The point estimate comes from the sample data.
- To estimate the population mean (μ), use the sample mean (\bar{x}) as the point estimate.

The margin of error

- Depends on the level of confidence, the sample size and the population standard deviation.
- It is computed as $E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ where $Z_{\alpha/2}$ is the critical value from the standard normal table associated with α (the level of significance).

- This is a Z-score that bounds the level of confidence.
- Confidence intervals are ALWAYS two-sided and the Z-scores are the limits of the area associated with the level of confidence.

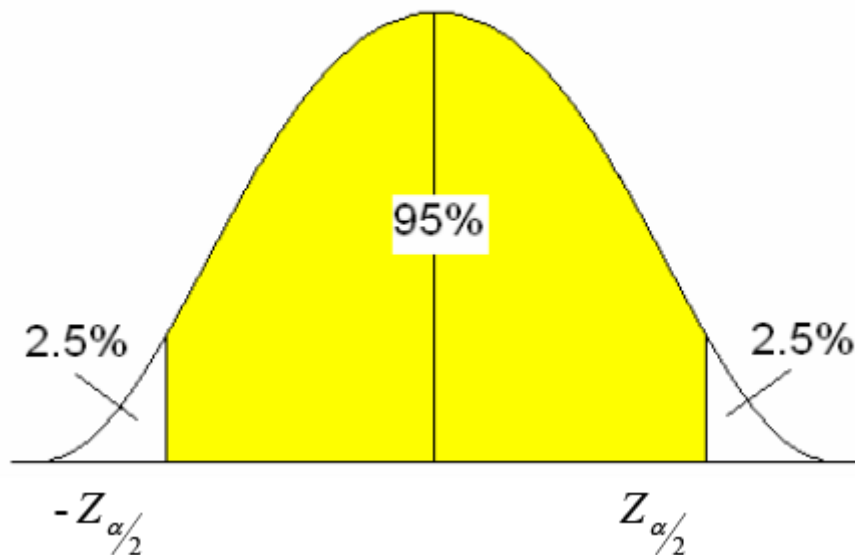


Figure 2. The middle 95% area under a standard normal curve.

- The level of significance (α) is divided into halves because we are looking at the middle 95% of the area under the curve.
- Go to your standard normal table and find the area of 0.025 in the body of values.
- What is the Z-score for that area?
- The Z-scores of ± 1.96 are the critical Z-scores for a 95% confidence interval.

| Confidence Level | α (level of significance) | $Z_{\alpha/2}$ |
|------------------|----------------------------------|----------------|
| 99% | 1% | 2.575 |
| 95% | 5% | 1.96 |
| 90% | 10% | 1.645 |

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

Construction of a confidence interval about μ when σ is known:

1. $Z_{\alpha/2}$ (critical value)
2. $E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$ (margin of error)
3. $\bar{x} \pm E$ (point estimate \pm margin of error)

Example 3

Construct a confidence interval about the population mean.

Researchers have been studying p-loading in Jones Lake for many years. It is known that mean water clarity (using a Secchi disk) is normally distributed with a population standard deviation of $\sigma = 15.4$ in. A random sample of 22 measurements was taken at various points on the lake with a sample mean of $\bar{x} = 57.8$ in. The researchers want you to construct a 95% confidence interval for μ , the mean water clarity.

$$1) Z_{\alpha/2} = 1.96$$

$$2) E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} = 1.96 * \frac{15.4}{\sqrt{22}} = 6.435$$

$$3) \bar{x} \pm E = 57.8 \pm 6.435$$

95% confidence interval for the mean water clarity is (51.36, 64.24).

We can be 95% confident that this interval contains the population mean water clarity for Jones Lake.

Now construct a 99% confidence interval for μ , the mean water clarity, and interpret.

$$1) Z_{\alpha/2} = 2.575$$

$\sigma = 15.4$

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

99% confidence interval for the mean water clarity is (49.35, 66.25).

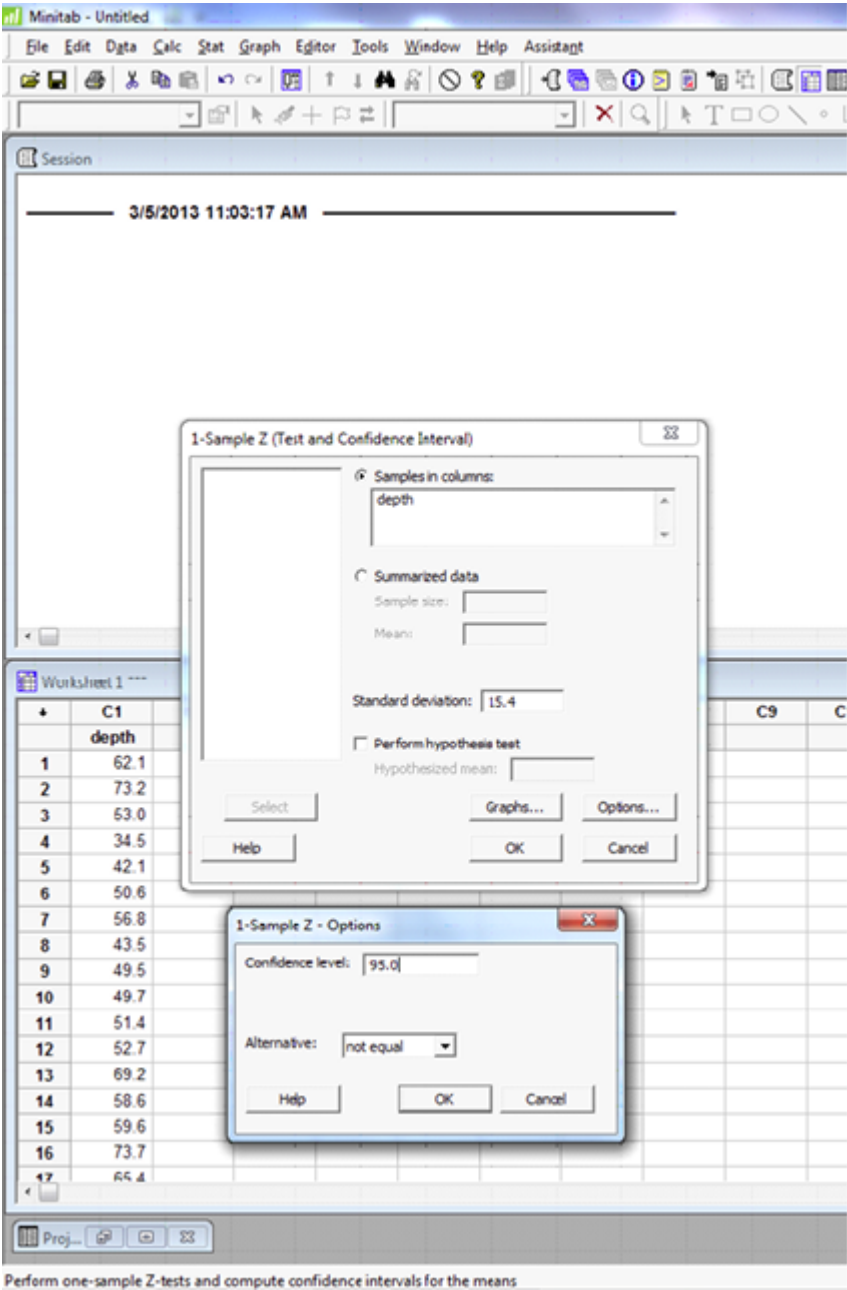
We can be 99% confident that this interval contains the population mean water clarity for Jones Lake.

As the level of confidence increased from 95% to 99%, the width of the interval increased. As the probability (area under the normal curve) increased, the critical value increased resulting in a wider interval.

Software Solutions

Minitab

You can use Minitab to construct this 95% confidence interval (Excel does not construct confidence intervals about the mean when the population standard deviation is known). Select Basic Statistics>1-sample Z. Enter the known population standard deviation and select the required level of confidence.



Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

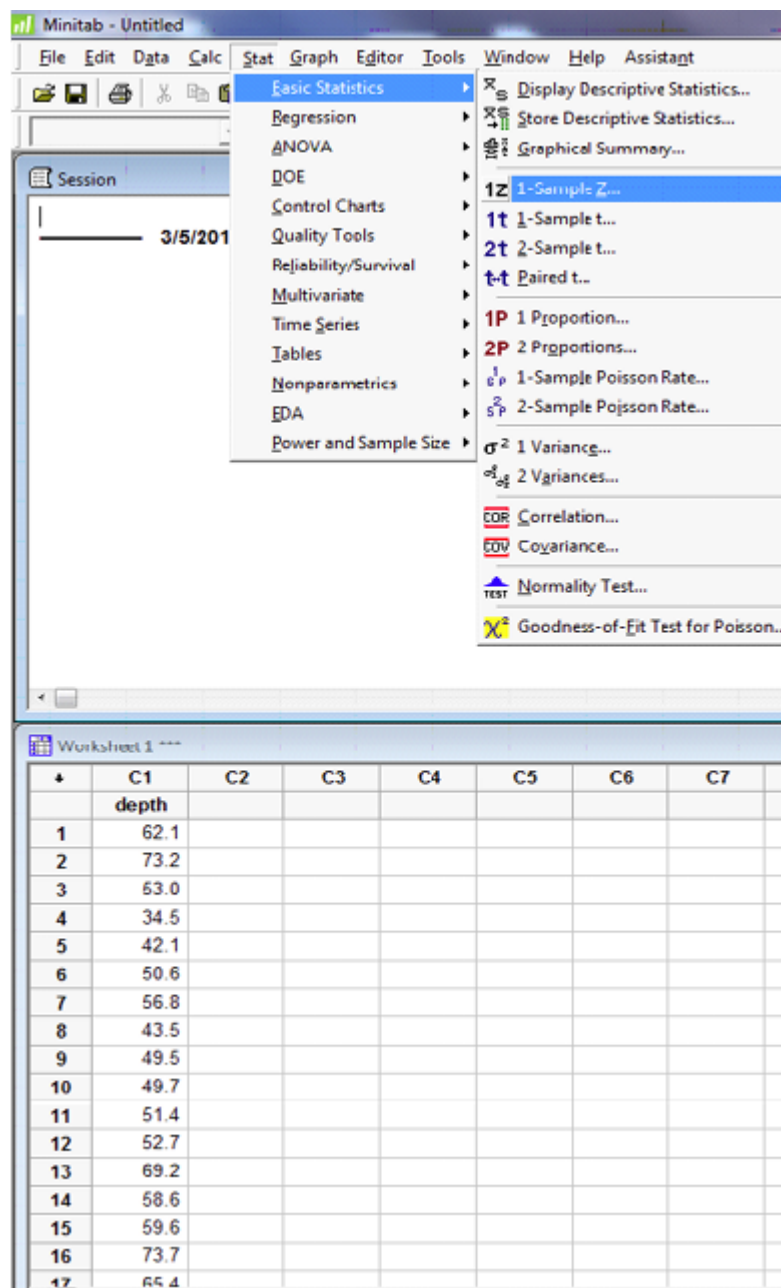


Figure 3. Minitab screen shots for constructing a confidence interval.

One-Sample Z: depth

The assumed standard deviation = 15.4

| Variable | N | Mean | StDev | SE Mean | 95% CI |
|----------|----|-------|-------|---------|----------------|
| depth | 22 | 57.80 | 11.60 | 3.28 | (51.36, 64.24) |

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

Confidence Intervals about the Mean (μ) when the Population Standard Deviation (σ) is Unknown

Typically, in real life we often don't know the population standard deviation (σ). We can use the sample standard deviation (s) in place of σ . However, because of this change, we can't use the standard normal distribution to find the critical values necessary for constructing a confidence interval.

The Student's t-distribution was created for situations when σ was unknown. Gosset worked as a quality control engineer for Guinness Brewery in Dublin. He found errors in his testing and he knew it was due to the use of s instead of σ . He created this distribution to deal with the problem of an unknown population standard deviation and small sample sizes. A portion of the t-table is shown below.

Area in Right Tail

| df | 0.10 | 0.05 | 0.025 | 0.02 | 0.01 | 0.005 |
|----|-------|-------|--------|--------|--------|--------|
| 1 | 3.078 | 6.314 | 12.706 | 15.894 | 31.821 | 63.657 |
| 2 | 1.886 | 2.920 | 4.303 | 4.849 | 6.965 | 9.925 |
| 3 | 1.638 | 2.353 | 3.182 | 3.482 | 4.541 | 5.841 |
| 4 | 1.533 | 2.132 | 2.776 | 2.999 | 3.747 | 4.604 |
| 5 | 1.476 | 2.015 | 2.571 | 2.757 | 3.365 | 4.032 |

Table 2. Portion of the student's t-table.

Example 4

Find the critical value $t_{\alpha/2}$ for a 95% confidence interval with a sample size of $n=13$.

- Degrees of freedom (down the left-hand column) is equal to $n-1 = 12$
- $\alpha = 0.05$ and $\alpha/2 = 0.025$
- Go down the 0.025 column to 12 df
- $t_{\alpha/2} = 2.179$

The critical values from the students' t-distribution approach the critical values from the standard normal distribution as the sample size (n) increases.

| n | Degrees of freedom | t .025 |
|----------|---------------------------|---------------|
| 11 | 10 | 2.228 |
| 51 | 50 | 2.009 |
| 101 | 100 | 1.984 |
| 1001 | 1000 | 1.962 |

Table 3. Critical values from the student's t-table.

Using the standard normal curve, the critical value for a 95% confidence interval is **1.96**. You can see how different samples sizes will change the critical value and thus the confidence interval, especially when the sample size is small.

Construction of a Confidence Interval about μ when σ is Unknown

1. $t_{\alpha/2}$ critical value with n-1 df
2. $E = t_{\alpha/2} * \frac{s}{\sqrt{n}}$
3. $\bar{x} \pm E$

Example 5

Researchers studying the effects of acid rain in the Adirondack Mountains collected water samples from 22 lakes. They measured the pH (acidity) of the water and want to construct a 99% confidence interval about the

ution of the pH of this population, and the sample is small ($n < 30$), so they look at a normal probability plot.

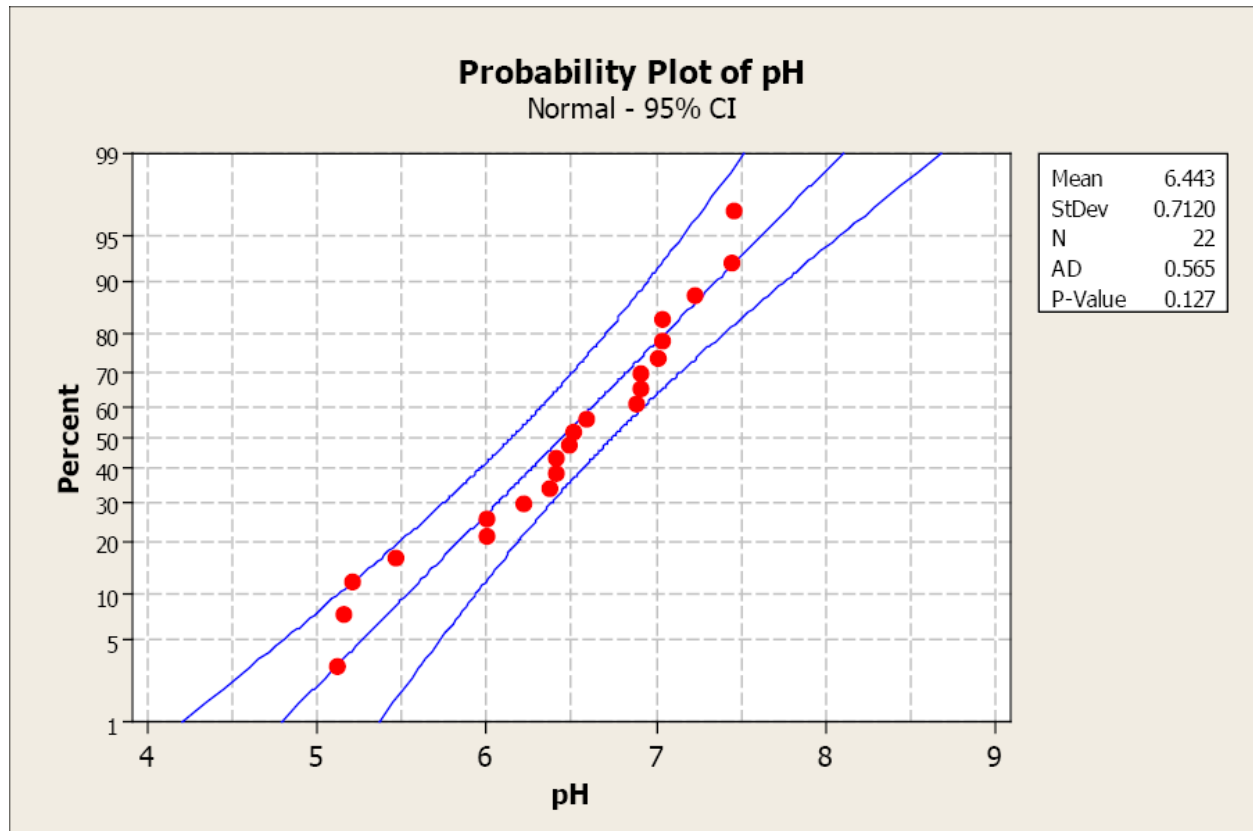


Figure 4. Normal probability plot.

The data is normally distributed. Now construct the 99% confidence interval about the mean pH.

$$1) t_{\alpha/2} = 2.831$$

$$2) E = t_{\alpha/2} * \frac{s}{\sqrt{n}} = 2.831 * \frac{0.7120}{\sqrt{22}} = 0.4297$$

$$3) \bar{x} \pm E = 6.443 \pm 0.4297$$

The 99% confidence interval about the mean pH is (6.013, 6.863).

We are 99% confident that this interval contains the mean lake pH for this lake population.

Now construct a 90% confidence interval about the mean pH for these

$$1) t_{\alpha/2} = 1.721$$

$$2) E = t_{\alpha/2} * \frac{s}{\sqrt{n}} = 1.721 * \frac{0.7120}{\sqrt{22}} = 0.2612$$

$$3) \bar{x} \pm E = 6.443 \pm 0.2612$$

The 90% confidence interval about the mean pH is (6.182, 6.704).

We are 90% confident that this interval contains the mean lake pH for this lake population.

Notice how the width of the interval decreased as the level of confidence decreased from 99 to 90%.

Construct a 90% confidence interval about the mean lake pH using Excel and Minitab.

Software Solutions

Minitab

For Minitab, enter the data in the spreadsheet and select Basic statistics and 1-sample t-test.

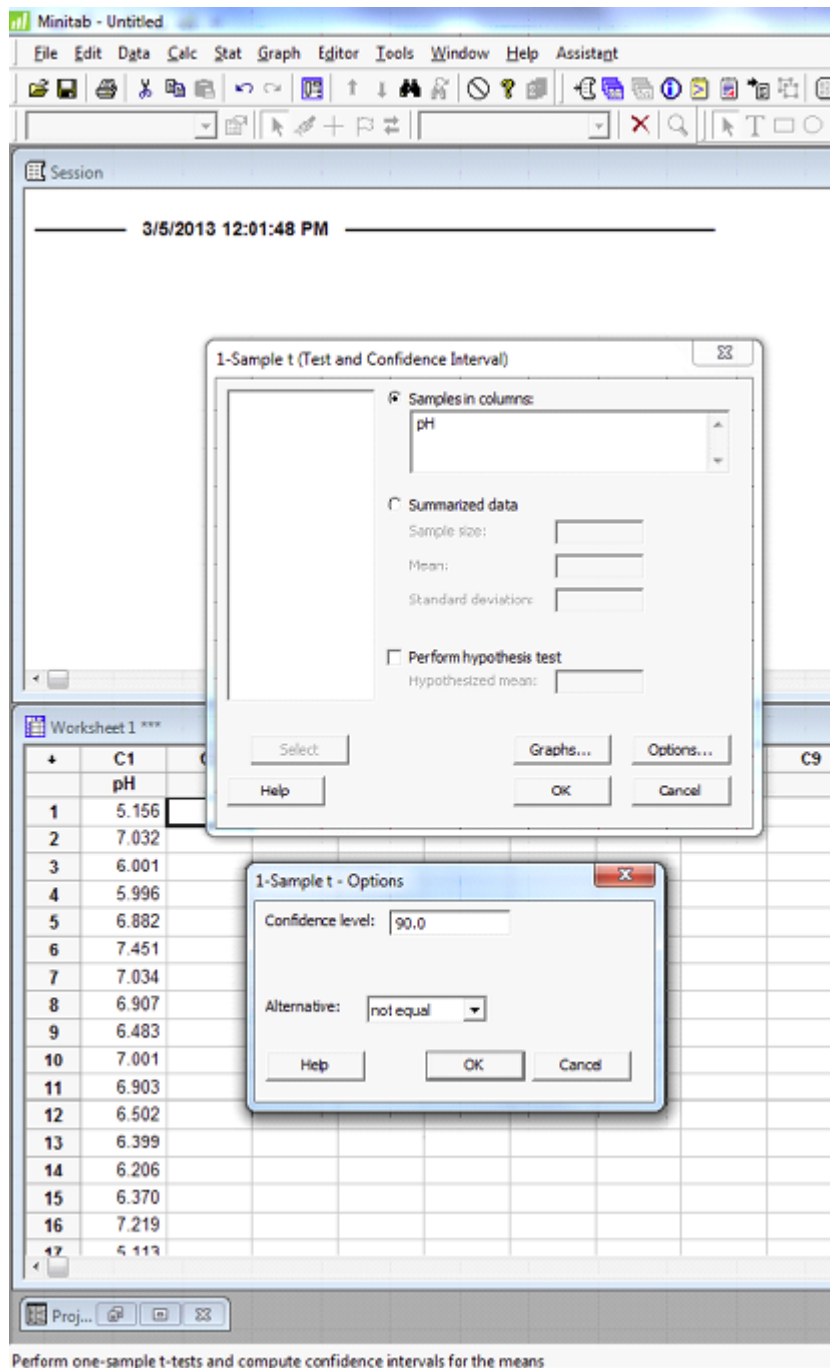
The screenshot shows the Minitab software interface. The 'Stat' menu is open, and 'Basic Statistics' is selected. The 'Basic Statistics' submenu is also open, showing various statistical tests. The '1t' option, representing a one-sample t-test, is highlighted. Below the menu, the 'Worksheet 1' is visible, containing a table of pH data. The table has columns C1 through C8, with C1 labeled 'pH'. The data is as follows:

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 |
|----|-------|----|----|----|----|----|----|----|
| | pH | | | | | | | |
| 1 | 5.156 | | | | | | | |
| 2 | 7.032 | | | | | | | |
| 3 | 6.001 | | | | | | | |
| 4 | 5.996 | | | | | | | |
| 5 | 6.882 | | | | | | | |
| 6 | 7.451 | | | | | | | |
| 7 | 7.034 | | | | | | | |
| 8 | 6.907 | | | | | | | |
| 9 | 6.483 | | | | | | | |
| 10 | 7.001 | | | | | | | |
| 11 | 6.903 | | | | | | | |
| 12 | 6.502 | | | | | | | |
| 13 | 6.399 | | | | | | | |
| 14 | 6.206 | | | | | | | |
| 15 | 6.370 | | | | | | | |
| 16 | 7.219 | | | | | | | |
| 17 | 5.113 | | | | | | | |

At the bottom of the Minitab window, a status bar reads: 'Perform one-sample t-tests and compute confidence intervals for the means'.

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing



One-Sample T: pH

| Variable | N | Mean | StDev | SE Mean | 90% CI |
|----------|----|-------|-------|---------|----------------|
| pH | 22 | 6.443 | 0.712 | 0.152 | (6.182, 6.704) |

Additional example:

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

Statistics - Constructing a Confidence Model



For Excel, enter the data in the spreadsheet and select descriptive statistics. Check Summary Statistics and select the level and confidence. Excel

Descriptive Statistics

Input
 Input Range:

Grouped By: ☒ Columns ☐ Rows

☐ Labels in first row

Output options
☒ Output Range:
☐ New Worksheet Ply:
☐ New Workbook

☒ Summary statistics
☒ Confidence Level for Mean: %
☐ Kth Largest:
☐ Kth Smallest:

Buttons: OK, Cancel, Help

| | A | B | C | D | E | F | G | H |
|----|-------|---|---|---|---|---|---|---|
| 1 | pH | | | | | | | |
| 2 | 5.156 | | | | | | | |
| 3 | 7.032 | | | | | | | |
| 4 | 6.001 | | | | | | | |
| 5 | 5.996 | | | | | | | |
| 6 | 6.882 | | | | | | | |
| 7 | 7.451 | | | | | | | |
| 8 | 7.034 | | | | | | | |
| 9 | 6.907 | | | | | | | |
| 10 | 6.483 | | | | | | | |
| 11 | 7.001 | | | | | | | |
| 12 | 6.903 | | | | | | | |
| 13 | 6.502 | | | | | | | |
| 14 | 6.399 | | | | | | | |
| 15 | 6.206 | | | | | | | |
| 16 | 6.37 | | | | | | | |
| 17 | 7.219 | | | | | | | |
| 18 | 5.113 | | | | | | | |
| 19 | 7.438 | | | | | | | |
| 20 | 6.582 | | | | | | | |
| 21 | 6.405 | | | | | | | |
| 22 | 5.459 | | | | | | | |
| 23 | 5.205 | | | | | | | |

Data Analysis

Analysis Tools

- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics**
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Moving Average
- Random Number Generation

Buttons: OK, Cancel, Help

| | A | B | C | D | E | F | G | H |
|----|-------|---|---|---|---|---|---|---|
| 1 | pH | | | | | | | |
| 2 | 5.156 | | | | | | | |
| 3 | 7.032 | | | | | | | |
| 4 | 6.001 | | | | | | | |
| 5 | 5.996 | | | | | | | |
| 6 | 6.882 | | | | | | | |
| 7 | 7.451 | | | | | | | |
| 8 | 7.034 | | | | | | | |
| 9 | 6.907 | | | | | | | |
| 10 | 6.483 | | | | | | | |
| 11 | 7.001 | | | | | | | |
| 12 | 6.903 | | | | | | | |
| 13 | 6.502 | | | | | | | |
| 14 | 6.399 | | | | | | | |
| 15 | 6.206 | | | | | | | |
| 16 | 6.37 | | | | | | | |
| 17 | 7.219 | | | | | | | |
| 18 | 5.113 | | | | | | | |
| 19 | 7.438 | | | | | | | |
| 20 | 6.582 | | | | | | | |
| 21 | 6.405 | | | | | | | |
| 22 | 5.459 | | | | | | | |
| 23 | 5.205 | | | | | | | |

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

| | |
|-------------------------|----------|
| Mean | 6.442909 |
| Standard Error | 0.151801 |
| Median | 6.4925 |
| Mode | #N/A |
| Standard Deviation | 0.712008 |
| Sample Variance | 0.506956 |
| Kurtosis | -0.5007 |
| Skewness | -0.60591 |
| Range | 2.338 |
| Minimum | 5.113 |
| Maximum | 7.451 |
| Sum | 141.744 |
| Count | 22 |
| Confidence Level(90.0%) | 0.26121 |

Excel gives you the sample mean in the first line (6.442909) and the margin of error in the last line (0.26121). You must complete the computation yourself to obtain the interval (6.442909 ± 0.26121).

Confidence Intervals about the Population Proportion (p)

Frequently, we are interested in estimating the population proportion (p), instead of the population mean (μ). For example, you may need to estimate the proportion of trees infected with beech bark disease, or the proportion of people who support “green” products. The parameter p can be estimated in the same ways as we estimated μ , the population mean.

- The sample proportion is the best point estimate for the true population proportion.
- Sample proportion $\hat{p} = \frac{x}{n}$ where x is the number of elements in the sample with the characteristic you are interested in, and n is the sample size.

The Assumption of Normality when Estimating Proportions

- The assumption of a normally distributed population is still important, even though the parameter has changed.
- Normality can be verified if:

- $$n * \hat{p} * (1 - \hat{p}) \geq 10$$

Constructing a Confidence Interval about the Population Proportion

Constructing a confidence interval about the proportion follows the same three steps we have used in previous examples.

1. $Z_{\alpha/2}$ (critical value from the standard normal table)
2. $E = Z_{\alpha/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (margin of error)
3. $\hat{p} \pm E$ (point estimate \pm margin of error)

Example 6

A botanist has produced a new variety of hybrid soybean that is better able to withstand drought. She wants to construct a 95% confidence interval about the germination rate (percent germination). She randomly se-

$$\hat{p} = \frac{x}{n} = \frac{421}{500} = 0.842$$

Check normality: $n * \hat{p} * (1 - \hat{p}) \geq 10 = 500 * 0.842 * (1 - 0.842) = 66.5$

You can assume a normal distribution.

Now construct the confidence interval:

$$1) Z_{\alpha/2} = 1.96$$

$$2) E = Z_{\alpha/2} * \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = 1.96 * \sqrt{\frac{.842(1 - .842)}{500}} = 0.032$$

$$3) \hat{p} \pm E = 0.842 \pm 0.032$$

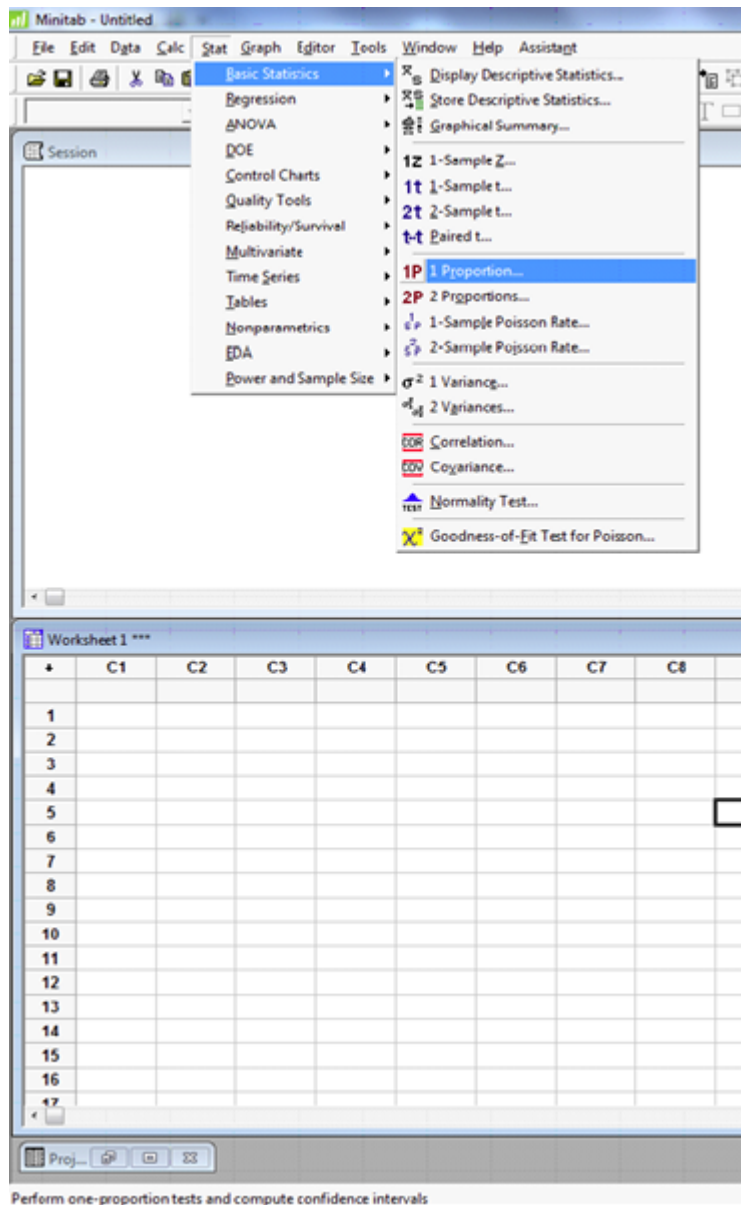
The 95% confidence interval for the germination rate is (81.0%, 87.4%).

We can be 95% confident that this interval contains the true germination rate for this population.

Software Solutions

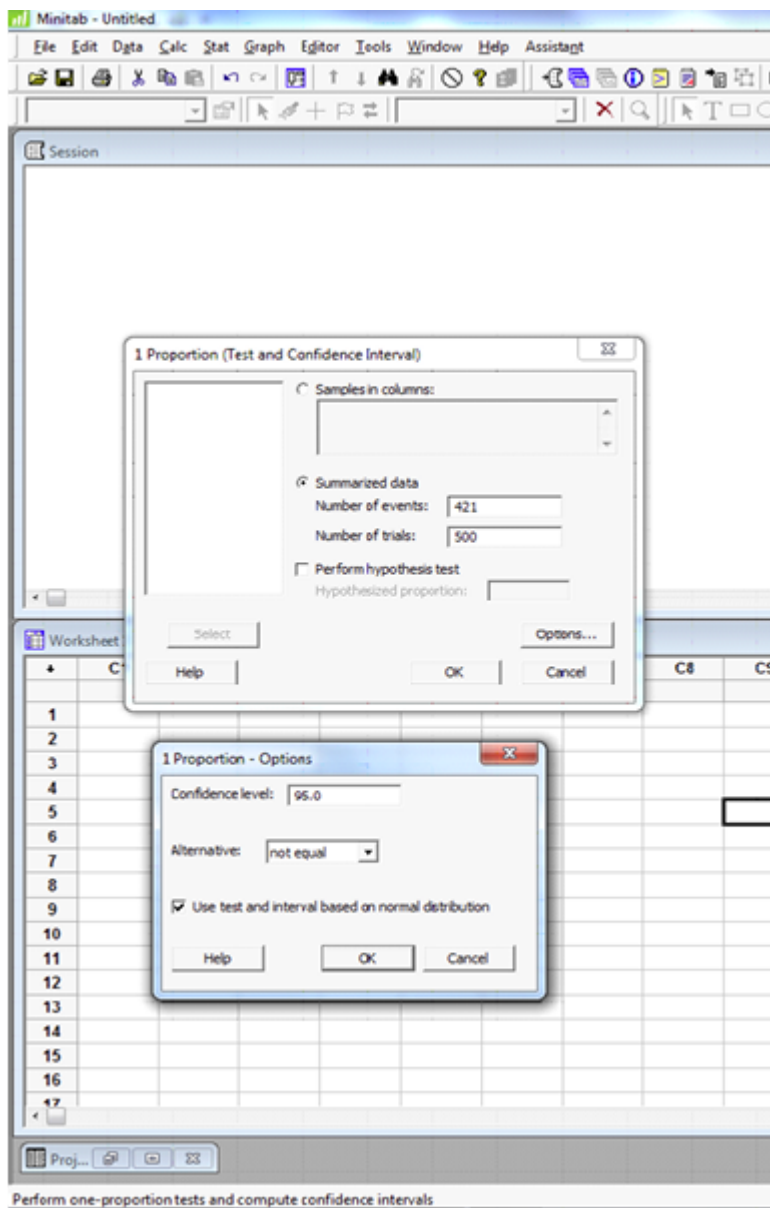
Minitab

You can use Minitab to compute the confidence interval. Select STAT>Basic stats>1-proportion. Select summarized data and enter the number of events (421) and the number of trials (500). Click Options and select the correct confidence level. Check “test and interval based on normal distribution” if the assumption of normality has been verified.



Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing



Test and CI for One Proportion

| Sample | X | N | Sample p | 95% CI |
|--------|-----|-----|----------|----------------------|
| 1 | 421 | 500 | 0.842000 | (0.810030, 0.873970) |

Using the normal approximation.

Excel

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

proportion.

Confidence Interval Summary

Which method do I use?

The first question to ask yourself is: **Which parameter are you trying to estimate?** If it is the mean (μ), then ask yourself: **Is the population standard deviation (σ) known?** If yes, then follow the next 3 steps:

Confidence Interval about the Population Mean (μ) when σ is Known

1. $Z_{\alpha/2}$ critical value (from the standard normal table)
2. $E = Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$
3. $\bar{x} \pm E$

If no, follow these 3 steps:

Confidence Interval about the Population Mean (μ) when σ is Unknown

LICENSE

1. $t_{\alpha/2}$ critical value with n-1 df from the student t-distribution

$$E = t_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

3. $\bar{x} \pm E$

Natural Resources Biometrics by Diane Kiernan is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#). If you want to consult a confidence interval about the population proportion, follow these 3 steps.

Confidence Interval about the Proportion

Previous: Chapter 1: Descriptive Statistics and the Normal Distribution

Next: Chapter 3: Hypothesis Testing

1. $Z_{\alpha/2}$ critical value from the standard normal table

2. $E = Z_{\alpha/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

3. $\hat{p} \pm E$

Powered by Pressbooks

[Guides and Tutorials](#) | [Contact](#)

Remember that the assumption of normality must be verified.

