

STP 226 ELEMENTARY STATISTICS

CHAPTER 8

CONFIDENCE INTERVAL FOR ONE POPULATION MEAN

8.1 Estimating a Population Mean

A **point estimate** of a parameter - is the value of a statistic used to estimate the parameter

A **confidence-interval estimate** of a parameter consists of an interval of numbers obtained from a point estimate of the parameter together with a percentage that specifies how confident we are that the parameter lies in the interval.

Confidence level - confidence percentage

8.2 Confidence Intervals for one Population Mean when σ is known

The One-Sample z-interval Procedure for a Population Mean

ASSUMPTIONS

1. Normal population or large sample
2. σ known

STEPS

1. For a confidence level of $1 - \alpha$, use Table II to find $z_{\alpha/2}$
2. The confidence level for μ is from

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{to} \quad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where $z_{\alpha/2}$ is found in step 1, n is the sample size, and \bar{x} is computed from the sample data.

The confidence interval is exact for normal populations and is approximately correct for large samples from non-normal populations.

When to use the z-interval procedure

1. For small samples, say, of size less than 15, the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.
2. For moderate-size samples, say, between 15 and 30, the z-interval procedure can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.
3. For large samples, say, of size 30 or more, the z-interval procedure can be used essentially without restriction. However, if outliers are present and their removal is not justified, the effect of the outliers on the confidence interval should be examined; that is, you should compare the confidence intervals obtained with and without the outliers. If the effect is substantial, then it is probably best to use a different procedure or take another sample.
4. If the outliers are present but their removal is justified and results in a data set for which the z-interval procedure is appropriate, then the procedure can be used.

A Fundamental Principle of Data Analysis

Before performing a statistical-inference procedure, look at the sample data. If any of the conditions required for using the procedure appear to be violated, do not apply the procedure. Instead use a different, more appropriate procedure, or, if you are unsure of one, consult a statistician.

Confidence and Precision

For a fixed sample size, decreasing the confidence level increases the precision, and vice-versa.

8.3 Margin of Error

Margin of Error for the Estimate of μ

The **margin of error** for the estimate of μ is $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

The margin of error is half the length of the confidence interval.

Margin of Error, Precision, and Sample Size

The length of a confidence interval for a population mean, μ , and hence the precision with which \bar{x} estimates μ , is determined by the margin of error, E . For a fixed confidence level, increasing the sample size increases the precision, and vice-versa.

Sample size for estimating μ

The sample size required for $(1 - \alpha)$ -level confidence interval for μ with a specified margin of error, E , is given by the formula

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2, \text{ rounded to the nearest whole number.}$$

8.4 Confidence Intervals for one Population Mean when σ is Unknown

If x is a normally distributed variable with mean μ and standard deviation σ , then, for samples of size n , the variable \bar{x} is also normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The standardized version of \bar{x} is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ has a standard normal distribution.

Since we have to use the sample standard deviation s , to estimate the population σ , the standardized version of \bar{x} is then called the studentized version of \bar{x} . $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

Student's t-distribution (t-distribution) - William Gosset (1908)

There is a different t-distribution for each sample size and each is identified by its degrees of freedom (df) $-(n - 1$ or one less than the sample size)

Each of these curves for the t-distributions is called a t-curve.

Studentized Version of the Sample Mean

Suppose a variable x of a population is normally distributed with mean μ . Then, for samples of size n , the studentized version of \bar{x} , $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ has the t-distribution with $n-1$ degrees of freedom.

Properties of t-curves

1. The total area under a t-curve is equal to 1.
2. A t-curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
3. A t-curve is symmetric about 0.
4. As the number of degrees of freedom becomes larger, t-curves look increasingly like the standard normal curve.

Using the t-tables

$t_{0.05}$ - t-value having area of 0.05 to its right

The one-sample t-interval procedure for a population mean

ASSUMPTIONS

1. Normal population or large sample
2. σ unknown

Step 1 For a confidence level of $1 - \alpha$, use Table IV to find $t_{\alpha/2}$ with $df = n - 1$, where n is the sample size.

Step 2 The confidence interval for μ is from

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

where $t_{\alpha/2}$ is found in step 1 and \bar{x} and s are computed from the sample data.

The confidence is exact for normal populations and is approximately correct for large samples from non-normal populations.

What if the assumptions are not satisfied?

Assumptions: population normally distributed or samples large

If population standard deviation is known, we can use the z-interval procedure.

If population standard deviation is unknown, we can use the t-interval procedure.

If samples are small the z-interval and t-interval procedures should not be used in the presence of outliers.

If sample size is small, or the population is far from being normally distributed, other methods need to be used.

One other method of estimating the population mean is the one sample **Wilcoxon confidence-interval procedure** (non-parametric procedure)

Non-parametric procedures (even with their own assumptions) are resistant to outliers and other extreme values and sample size.