Confidence Intervals

Gene Quinn

Confidence Intervals

One of the most common problems in statistics is estimation of a population mean

Confidence Intervals

One of the most common problems in statistics is estimation of a population mean

The process of obtaining such an estimate from a sample is known as *statistical inference*.

Confidence Intervals

One of the most common problems in statistics is estimation of a population mean

The process of obtaining such an estimate from a sample is known as *statistical inference*.

We want to use the information from a sample to generalize to the population under study.

Confidence Intervals

One of the most common problems in statistics is estimation of a population mean

The process of obtaining such an estimate from a sample is known as *statistical inference*.

We want to use the information from a sample to generalize to the population under study.

In particular, we will use the

sample mean $\overline{x}$

as a *point estimate* of the population mean μ.

Confidence Intervals

Definition: A **point estimate** is the value of a statistic that provides an estimate of the value of a parameter.

Confidence Intervals

Definition: A **point estimate** is the value of a statistic that provides an estimate of the value of a parameter.

The usual point estimate for the population mean μ is the sample mean, $\overline{x}$.

Confidence Intervals

Definition: A **point estimate** is the value of a statistic that provides an estimate of the value of a parameter.

The usual point estimate for the population mean μ is the sample mean, $\overline{x}$.

A shortcoming of a point estimate is that it provides no information about the precision of the estimate.

Confidence Intervals

Definition: A **point estimate** is the value of a statistic that provides an estimate of the value of a parameter.

The usual point estimate for the population mean μ is the sample mean, $\overline{x}$.

A shortcoming of a point estimate is that it provides no information about the precision of the estimate.

An alternative is to construct an *interval estimate*.

Confidence Intervals

Definition: A **point estimate** is the value of a statistic that provides an estimate of the value of a parameter.

The usual point estimate for the population mean μ is the sample mean, $\overline{x}$.

A shortcoming of a point estimate is that it provides no information about the precision of the estimate.

An alternative is to construct an *interval estimate*.

An *interval estimate*, consists of a *confidence interval* and an associated *level of confidence*.

Confidence Intervals

Actually, the confidence interval is determined by four quantities:

- The sample mean $\overline{x}$

Confidence Intervals

Actually, the confidence interval is determined by four quantities:

- The sample mean $\overline{x}$
- The population standard deviation σ

Confidence Intervals

Actually, the confidence interval is determined by four quantities:

- The sample mean $\overline{x}$
- The population standard deviation σ
- The sample size n

Confidence Intervals

Actually, the confidence interval is determined by four quantities:

- The sample mean $\overline{x}$
- The population standard deviation σ
- The sample size n
- The level of confidence $1 - \alpha$

Confidence Intervals

The lower bound of the confidence interval is:

$$\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

The lower bound of the confidence interval is:

$$\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The upper bound of the confidence interval is:

$$\overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

The lower bound of the confidence interval is:

$$\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The upper bound of the confidence interval is:

$$\overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

The constant $z_{\alpha/2}$ is determined by the level of confidence $1 - \alpha$

Confidence Intervals

The level of confidence is written as $1 - \alpha$, so

- α=.05 if we want a confidence level of 95%

Confidence Intervals

The level of confidence is written as $1 - \alpha$, so

- α=.05 if we want a confidence level of 95%
- α=.01 if we want a confidence level of 99%

Confidence Intervals

The level of confidence is written as $1-\alpha$, so

- α=.05 if we want a confidence level of 95%
- α=.01 if we want a confidence level of 99%
- α=.10 if we want a confidence level of 90%

Confidence Intervals

Once we have determined the value of α, all that remains is to determine the constant $z_{\alpha/2}$

Confidence Intervals

Once we have determined the value of α, all that remains is to determine the constant $z_{\alpha/2}$

The constant $z_{\alpha/2}$ is defined as the z-value that has an area of $\alpha/2$ to the right.

Confidence Intervals

Once we have determined the value of α, all that remains is to determine the constant $z_{\alpha/2}$

The constant $z_{\alpha/2}$ is defined as the z-value that has an area of $\alpha/2$ to the right.

This can be computed using either of the spreadsheet formulas:

=NORMSINV$(1 - \alpha/2)$

or

=-NORMSINV$(\alpha/2)$

Confidence Intervals

In summary, the confidence interval with confidence level $1-\alpha$ has lower limit

$$\overline{x} - NORMSINV(1-\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

In summary, the confidence interval with confidence level $1 - \alpha$ has lower limit

$$\overline{x} - NORMSINV(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}}$$

The upper limit is

$$\overline{x} + NORMSINV(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

Note that we have assumed that we **know** the population standard deviation σ

Confidence Intervals

Note that we have assumed that we **know** the population standard deviation σ

Often this is not realistic, and we will develop a similar technique for the case where we **do not know** the population standard deviation σ

Confidence Intervals

Note that we have assumed that we **know** the population standard deviation σ

Often this is not realistic, and we will develop a similar technique for the case where we **do not know** the population standard deviation σ

As we will see, in this case we have to estimate both the population mean μ and the population standard deviation σ from the sample.

Margin of Error

The **margin of error** associated with a $1 - \alpha$ level confidence interval with known σ is

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = NORMSINV(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}}$$

Margin of Error

The **margin of error** associated with a $1 - \alpha$ level confidence interval with known σ is

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = NORMSINV(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}}$$

In words, the margin of error $1/2$ the width of the confidence interval for μ

Sample Size

The margin of error E depends on:

- The confidence level $1 - \alpha$

Sample Size

The margin of error E depends on:

- The confidence level $1 - \alpha$
- The population standard deviation σ

Sample Size

The margin of error E depends on:

- The confidence level $1 - \alpha$
- The population standard deviation σ
- The sample size n

Sample Size

The margin of error E depends on:

- The confidence level $1 - \alpha$
- The population standard deviation σ
- The sample size n

We cannot control σ, but we are free to choose α and n within the constraints of available resources.

Sample Size

The margin of error E depends on:

- The confidence level $1 - \alpha$
- The population standard deviation σ
- The sample size n

We cannot control σ, but we are free to choose α and n within the constraints of available resources.

An important practical consideration is: How do we choose the sample size to produce a confidence interval with a specified margin of error (given α and σ)?

Sample Size

The required sample size is determined by solving the margin of error equation for n:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Sample Size

The required sample size is determined by solving the margin of error equation for n:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

In terms of a spreadsheet formula, the sample size is:

$$n = \left(\frac{NORMSINV(1 - \alpha/2) \cdot \sigma}{E}\right)^2$$

Sample Size

Example: We want to estimate the mean SAT score of freshmen in 2008 with a margin of error of 5 points.

What size sample is required if the population standard deviation is 100 and the desired level of confidence is 95%?

Sample Size

Example: We want to estimate the mean SAT score of freshmen in 2008 with a margin of error of 5 points.

What size sample is required if the population standard deviation is 100 and the desired level of confidence is 95%?

$1-\alpha$ in this case is $.95$, so $\alpha = .05$ and

$$n = \left(\frac{NORMSINV(1-\alpha/2) \cdot \sigma}{E} \right)^2$$

Sample Size

Example: We want to estimate the mean SAT score of freshmen in 2008 with a margin of error of 5 points.

What size sample is required if the population standard deviation is 100 and the desired level of confidence is 95%?

$1 - \alpha$ in this case is $.95$, so $\alpha = .05$ and

$$n = \left(\frac{NORMSINV(1 - \alpha/2) \cdot \sigma}{E} \right)^2$$

With $\alpha = .05$, $\sigma = 100$, and $E = 5$, the sample size is:

$$n = \left(\frac{NORMSINV(0.975) \cdot 100}{5} \right)^2 = 1537$$