Introduction to Statistics Tutorial: Large-Sample Estimation

INCOGEN, Inc.

2008

INCOGEN

Outline

- Types of estimators
- Evaluating goodness of an estimator
- Calculating confidence intervals
- Choosing sample size
- Estimating the difference between two means
- Estimating a binomial proportion
- Estimating the difference between two binomial proportions

INCOGEN

Introduction

The objective of statistics is to make <u>inference</u> about a population based on collecting information from a sample.

We want to make decisions and predictions based on the observations or data in that sample.

Populations can be described numerically by <u>parameters</u>…and the value of a parameter can be <u>estimated</u> from the data from our sample.

Once a parameter is estimated, we want to have a <u>measure of goodness</u> that describes how correct the estimate is.

Introduction

Statistical Inference:

- the inference
- a measure of its goodness

Types of Estimates

1. Point Estimates – a single number describing a parameter calculated from the sample data

2. Interval estimates or confidence intervals- two numbers, defining an interval on a line within which the parameter is expected to lie

Point Estimators

Point Estimator is a statistic calculated from sample measurements

Margin of error = 1.96 x standard error of the estimator

Confidence Interval

A Large-sample **Confidence Interval** is a range of values used to estimate the true value of the population parameter.

For example,

$$\text{Lower bound} < \mu < \text{Upper bound}$$

where the lower and upper bounds are calculated using the data and μ is the unknown, but fixed, population parameter

INCOGEN

Degree of Confidence

Degree of Confidence, also called Level of Confidence or Confidence Coefficient - the probability 1 - α that is the relative frequency of times the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times

INCOGEN

Confidence Interval Interpretation

95% confidence interval:

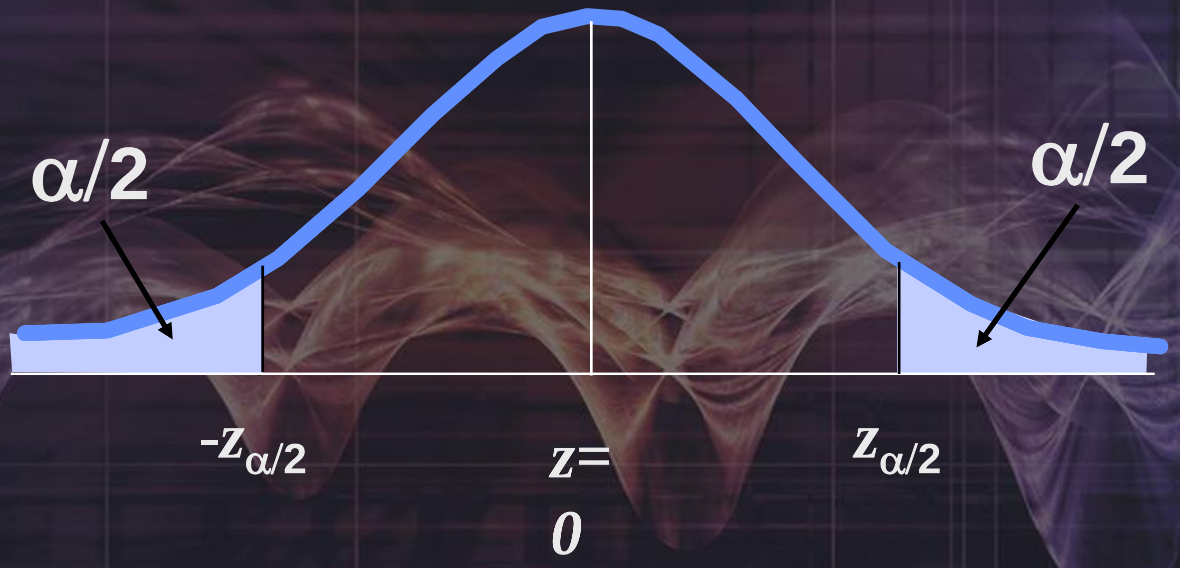
$$35 < \mu < 42$$

Interpret: We are 95% confident that the interval from 35 to 42 actually does contain the true value of μ. This means that if we were to select many different samples of same sample size and construct the confidence intervals, 95% of them would actually contain the value of the population mean μ.

Incorrect: The probability is 95% that the true value of μ will fall in the interval between 35 and 42.

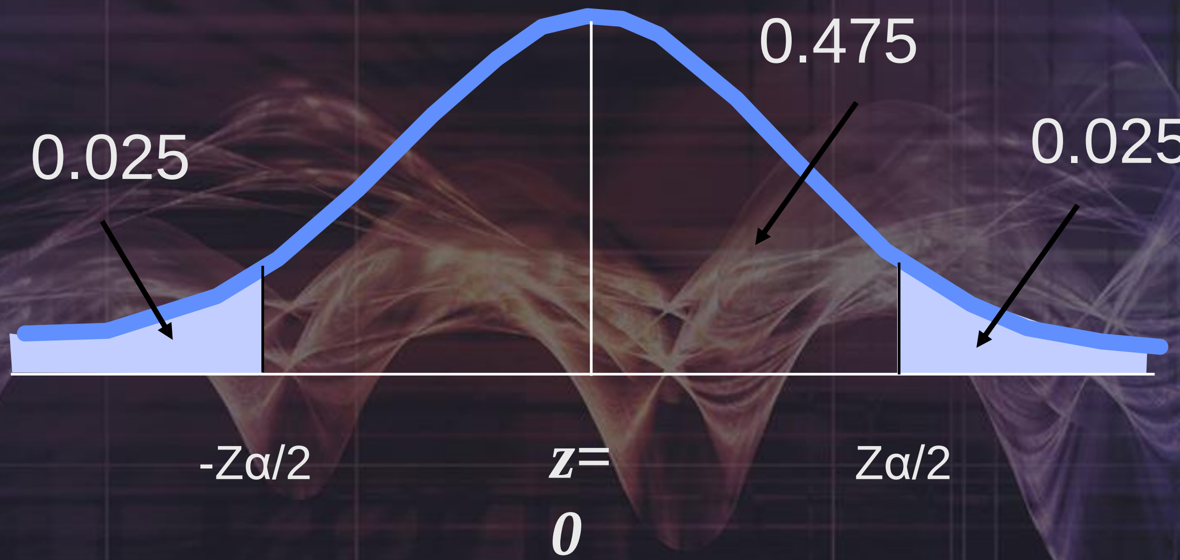
INCOGEN

The Critical Value: $z_{\alpha/2}$



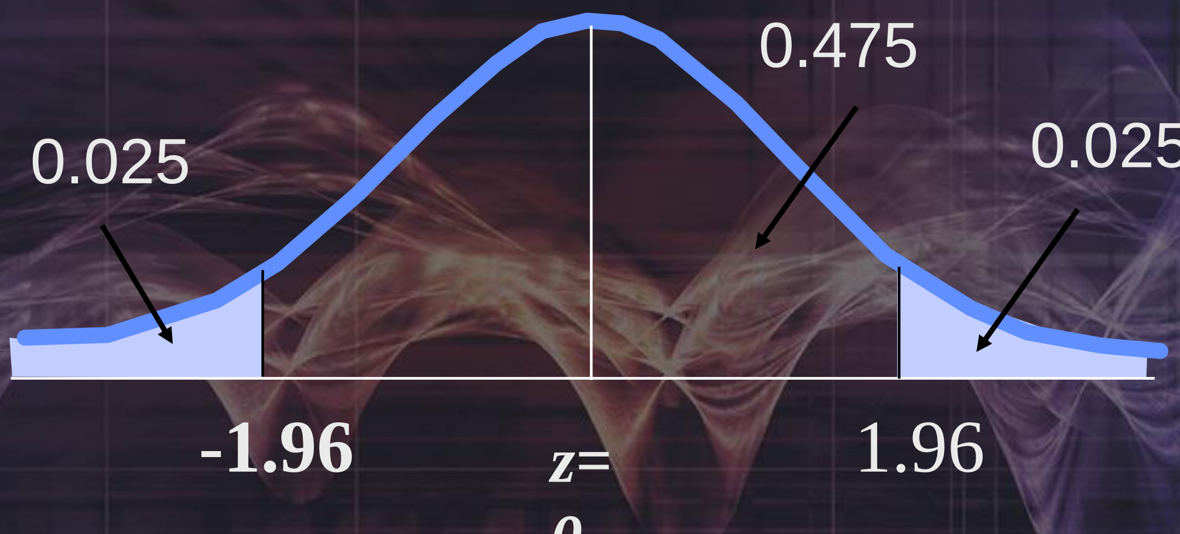
95% Confidence Interval

$\alpha = 0.05$ $\alpha/2 = 0.025$ $Z\alpha/2 = ?$



95% Confidence Interval

$\alpha = 0.05$ $\alpha/2 = 0.025$ $Z_{\alpha/2} = 1.96$



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.45
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.46
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.46
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.47
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.48
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.4
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.48

INCOGEN

95% Confidence Interval for the Population Mean

$\bar{X}$ – Margin of Error < μ < $\bar{X}$ + Margin of Error

So we also need the Margin of Error:

$$1.96\ \sigma / \sqrt{n}$$

INCOGEN

95% Confidence Interval for the Population Mean

$\bar{X} - 1.96\sigma/\sqrt{n} < \mu < \bar{X} + 1.96\sigma/\sqrt{n}$

Example: Suppose $\bar{X} = 38.5$, $n = 100$ and $\sigma = 17.9$

95% Confidence Interval is

$38.5 - 1.96 * 17.9/\sqrt{100} < \mu < 38.5 + 1.96 * 17.9 \sqrt{100}$

$38.5 - 3.5 < \mu < 38.5 + 3.5$

$35 < \mu < 42$

INCOGEN

Confidence Intervals for Large n Step-by-Step

1. Find the critical value $z_{\alpha/2}$ that corresponds to the desired degree of confidence.

2. Evaluate the margin of error $ME = z_{\alpha/2} \cdot \sigma / \sqrt{n}$. If the population standard deviation is unknown, use the value of the sample standard deviation s provided that $n > 30$.

3. Find the values of $\bar{x}$ - ME and $\bar{x}$ + ME. Substitute those values in the general format of the confidence interval:

$$\bar{x} - ME < \mu < \bar{x} + ME$$

INCOGEN

Confidence Intervals Example

A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error ME and the 95% confidence interval. (From Triola, **Chapter 6,** Elementary Statistics, Eighth Ed.)

Confidence Intervals Example

A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error ME and the 95% confidence interval. (From Triola, **Chapter 6,** Elementary Statistics, Eighth Ed.)

n = 106
x = 98.20 degrees
s = 0.62 degree

α = 0.05
$\alpha/2$ = 0.025
$Z_{\alpha/2} = 1.96$

Confidence Intervals Example

A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error ME and the 95% confidence interval. (From Triola, **Chapter 6,** Elementary Statistics, Eighth Ed.)

n = 106
x = 98.20 degrees
s = 0.62 degree

α = 0.05
α/2 = 0.025
Z**α/2 = 1.96**

$$ME = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{0.62}{\sqrt{106}} = 0.12$$

INCOGEN

Confidence Intervals Example

A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error ME and the 95% confidence interval. (From Triola, **Chapter 6,** Elementary Statistics, Eighth Ed.)

n = 106
x = 98.20 degrees
s = 0.62 degree

$$\bar{x} - ME < \mu < \bar{x} + ME$$

α = 0.05
α/2 = 0.025
$Z_{\alpha/2}$ = 1.96

$$98.2^{\circ} - 0.12^{\circ} < \mu < 98.2^{\circ} + 0.12^{\circ}$$

$$98.08^{\circ} < \mu < 98.32^{\circ}$$

***ME* = 0.12**

Choosing Sample Size

1. Choose B, the bound on the margin of error

2. Choose $1-\alpha$, the confidence coefficient

3. Find n, using the formula

$Z_{\alpha/2}$ x (standard error of the estimator) = B

INCOGEN

Choosing Sample Size Example

Suppose you would like to estimate the population mean based on a random sample of *n* observations and prior experience suggests that $\sigma = 18.4$. If you want to estimate μ correct to within 2.1, with probability 0.95, how many observations should be in your sample?

$B = 2.1$

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05$

$Z_{\alpha/2} = 1.96$

$\sigma = 18.4$

$$1.96 \times (18.4/\sqrt{n}) = 2.1$$

Solve for *n*

Choosing Sample Size Example

$1.96\text{x}(18.4/\sqrt{n}) = 2.1$

Solve for n

$n = (1.96 \, x \, 18.4 / 2.1)^2$

$= 245$

So you need 245 observations to estimate the population mean to be correct within your desired 95% confidence interval.

INCOGEN

Conclusion

Other parameters of interest:

- Estimating the difference between two means
- Estimating a binomial proportion

  standard error: sqrt(p(1-p)/n)

- Estimating the difference between two binomial proportions

INCOGEN

References

This tutorial is comprised of materials from the following sources:

Introduction to Probability and Statistics by Mendenhall and Beaver. ITP/Duxbury.

The Cartoon Guide to Statistics by Gonick and Smith. HarperCollins.

Basic Statistics: an abbreviated overview by Ackerman, Bartz, and Deville. 2006 Accountability Conference

Elementary Statistics, Eighth Ed. by Triola. Addison-Wesley-Longman. 2001

INCOGEN