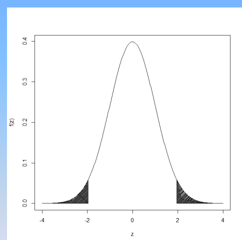
STATS 101
Introductory Statistics

Lecture 6

Confidence Interval Estimation

1-sample problems



In this lecture, following will be discussed:

- Population of interest
- Independent random sample from the population
- Use SAMPLE data to draw conclusions about the POPULATION of interest

STATISTICAL POPULATION

Look at a few examples:

1. A realtor needs to find the average selling price of homes in Las Vegas Metro.

POPULATION = {all single family homes in Las Vegas Metro}

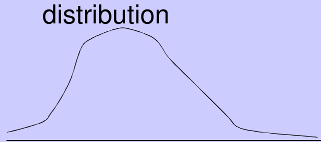
VARIABLE of interest = price of home,

2. A consumer group is interested in the number of failures of a new model of inline skates.

POPULATIUON = {all inline skates of this model}

VARIABLE of interest = # of failed inline skates in a given sample

Population probability distribution



Sample = $\{x_1, x_2, \ldots, x_n\}$

STATISTICAL INFERENCE : use sample information to draw conclusions about the population distribution.

TWO APPROACHES

1. Confidence Interval Estimation
2. Testing Statistical Hypotheses

Confidence Interval Estimation of Mean

Example 1: CI Estimation of the mean selling price of homes in a large city

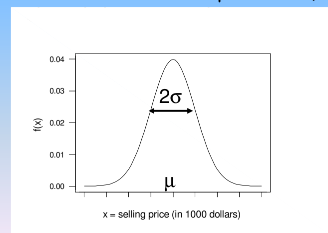
Suppose the realtor collects a representative random sample of n = 10 from the population:

$$\{x_1, x_2, \dots, x_n\}$$

Assumptions

[CI Estimation of mean– continued]

1. Assume that the random variable of interest X (selling price) is approximately normally distributed with mean μ and sd σ, or the sample



μ= unknown

parameter of interest

Ex 1:[CI Estimation of mean– continued]

μ= unknown population mean

is estimated by the SAMPLE MEAN:

$$\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ex 1:[CI Estimation of mean– continued]

A 95% confidence interval (CI) of the unknown mean μ is a random interval

(L, U) such that

$$P(L < \mu < U) = 0.95$$

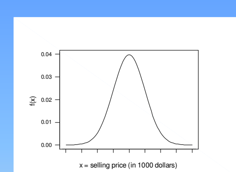
Ex 1:[CI Estimation of mean– continued]

We will use the CENTRAL LIMIT THEOREM (CLT) : pdf of sample mean is approximately normal with mean μ and sd = σ/√n for large n. For many situations, n ≥ 25 is sufficient.

$$\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \sim N(\mu, sd = \frac{\sigma}{\sqrt{n}})$$

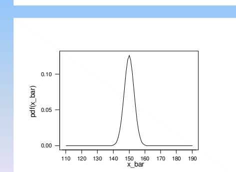
CLT

pdf of selling price x



pdf of mean selling price

$$\bar{x}$$



Ex 1:[CI Estimation of mean– continued]

1 $\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \sim N(\mu, sd = \frac{\sigma}{\sqrt{n}})$

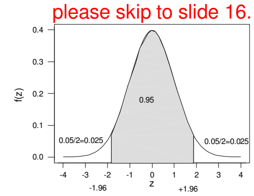
2 $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

In words: (1) estimate of μ is the sample mean, which has the normal distribution with mean μ and standard deviation σ/√n,
(2) (Sample mean - μ), standardized by its sd, has the standard normal distribution.

www.stats24x7.com 13

Ex 1:[CI Estimation of mean– continued]

If you don't care for algebra, please skip to slide 16.



To calculate 95% CI for mean μ, find z_0 from standard normal table so that $P(- z_0 \le Z \le z_0) = .95$.
$-z_0 = -1.96$, $z_0 = + 1.96$

www.stats24x7.com 14

Ex 1:[CI Estimation of mean– continued]

$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$P(-1.96 < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < +1.96) = 0.95$

Rewrite above as

$P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$

www.stats24x7.com 15

Ex 1:[CI Estimation of mean– continued]

95% CI for μ is

$(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$

Example calculation of CI:

Suppose the sample mean of 10 homes turned out to be $147,000, and suppose the population sd σ = $10,000

www.stats24x7.com 16

Ex 1:[CI Estimation of mean– continued]

95% CI for μ is

$(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$

$= (147 - \frac{1.96 \times 10}{\sqrt{10}}, 147 + \frac{1.96 \times 10}{\sqrt{10}})$

$= (147 - 6.20, 147 + 6.20)$

$= (140.8, 153.2)$

= ($140800,153200)

www.stats24x7.com 17

Ex 1:[CI Estimation of mean– continued]

The above formula assumed that σ is known – which is rarely the case.

When σ is unknown – it is estimated by the sample sd *s*. *Suppose for example 1 that sample sd of 10 observations was $9,000.*

www.stats24x7.com 18

Ex 1:[CI Estimation of mean– continued]

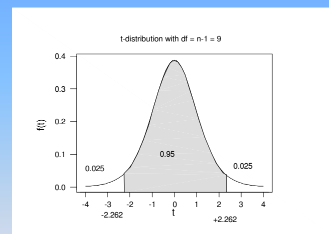
$$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \text{ becomes}$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t-distribution \text{ with degrees of freedom n-1}$$

In plain English, the distribution of (xbar - µ) standardized by sample sd is a t-distribution with DEGREES OF FREEDOM (n-1).

Ex 1:[CI Estimation of mean– continued]



Ex 1:[CI Estimation of mean– continued]

95% CI for μ becomes

$$(\bar{x} - 2.262\frac{s}{\sqrt{n}}, \bar{x} + 2.262\frac{s}{\sqrt{n}})$$

which is sometimes expressed as

$$\bar{x} \pm 2.262\frac{s}{\sqrt{n}}$$

Estimate ±

Reliability coefficient from t-table with n-1 degrees of freedom

x (sd of estimate)

Ex 1:[CI Estimation of mean– continued]

Sample mean = $147,000, sd = $9,000

95% CI for μ is

$$(\bar{x} - 2.262\frac{\sigma}{\sqrt{n}}, \bar{x} + 2.262\frac{\sigma}{\sqrt{n}})$$

$$= (147 - \frac{2.262 \times 9}{\sqrt{10}}, 147 + \frac{2.262 \times 9}{\sqrt{10}})$$

$$= (147 - 6.44, 147 + 6.44)$$

$$= (140.6, 153.4)$$

= ($140600,153400)

CI Estimation of mean– continued]

95% confidence intervals in Ex.1 calculated by:

$$(\bar{x} - z_{0.975}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.975}\frac{\sigma}{\sqrt{n}}),$$

$z_{0.975} = 1.96$ from z-table

σ known

$$(\bar{x} - t_{df,0.975}\frac{s}{\sqrt{n}}, \bar{x} + t_{df,0.975}\frac{s}{\sqrt{n}})$$

$t_{df,0.975} = 2.262$ from t-table with df = n-1

σ unknown

CI Estimation of mean– continued]

90% confidence intervals in Ex.1 will be calculated by:

$$(\bar{x} - z_{0.95}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.95}\frac{\sigma}{\sqrt{n}}),$$

$z_{0.975} = 1.645$ from z-table

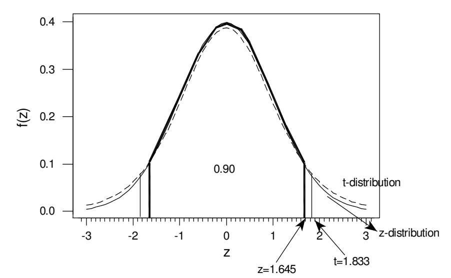
σ known

$$(\bar{x} - t_{df,0.95}\frac{s}{\sqrt{n}}, \bar{x} + t_{df,0.95}\frac{s}{\sqrt{n}})$$

$t_{df,0.975} = 1.833$ from t-table with df = n-1=9

σ unknown

z- and t- table values for 90% confidence



REMARK:

For n larger than 200, the t-table and the z-table will give identical values:

$$t_{df,1-\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}} \quad \text{for n} > 200$$

$$100(1\text{-}\alpha) = \text{confidence coefficient}$$

$$\alpha=0.05 \;\rightarrow\; 100(1\text{-}0.05) = 95\% \text{ confidence}$$

$$\alpha=0.10 \;\rightarrow\; 100(1\text{-}0.10) = 90\% \text{ confidence}$$

Example [CI Estimation of Binomial Proportion]

Suppose a random sample of n=50 inline skates are tested, and x=3 skates are found to be defective.

X = #(defective skates) in sample of 50 has the BINOMIAL distribution

Example [CI Estimation of Binomial Proportion – contd.]

$$X \sim BIN(n, p)$$

$$n = 50 = \#\text{ of samples}$$

$$x = \#\text{ of defectives in the sample}$$

$$p = \text{ proportion of defective skates in the population}$$

Example [CI Estimation of Binomial Proportion – contd.]

The unknown proportion of defectives *p* is estimated by the sample proportion of defectives:

$$\hat{p} = \frac{\#(\text{defectives}) \text{ in sample}}{\#\text{ of samples tested}} = \frac{x}{n}$$

Example [CI Estimation of Binomial Proportion – contd.]

Following result follows from CLT:

$$\hat{p} = \;\sim N(p, sd = \sqrt{\frac{p(1-p)}{n}})$$

$$\text{for n large so that } n\hat{p} \geq 5,\; n(1\text{-}\hat{p}) \geq 5$$

Example [CI Estimation of Binomial Proportion – contd.]

$sd(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$ is estimated by

$$sd(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example [CI Estimation of Binomial Proportion – contd.]

Using the formula for CI (see slide 21)

estimate ± reliability coefficient × sd(estimate)

we get the 95% CI for p as:

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The reliability coefficient of 1.96 is from the z-table or the standard normal table.

Example [CI Estimation of Binomial Proportion – contd.]

In the example, $x = 6$, n = 50, so

$$\hat{p} = \frac{x}{n} = \frac{3}{50} = 0.12$$

$$sd(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \sqrt{\frac{0.12(1-0.12)}{50}} = \sqrt{.002112} = .045957$$

Example [CI Estimation of Binomial Proportion – contd.]

95% CI:

$$\hat{p} \pm 1.96 \times sd(\hat{p})$$

$$= 0.12 \pm 1.96 \times .045957$$

$$= 0.12 \pm 0.09$$

$$= (0.03, 0.21)$$

In words: we are 95% confident that p lies between 3% and 21%.