STAT 100 Lecture 14 : Drawing Inferences from Large Samples

Nate Strawn

October 27

Last Time...

1. The Sampling Distribution of $\overline{X}$ for a Normal Population
2. The Central Limit Theorem

Today's Agenda

1. Statistical Inference (Section 8.1)
2. Point estimation of a population mean (Section 8.2)
3. Examples

Statistical Inference

Definition

Statistical inference *deals with drawing conclusions about population parameters from an analysis of the sample data.*

The two most important types of inferences:

- Estimation of parameters
- Testing of statistical hypothesis

Statistical Inference

Definition

Statistical inference *deals with drawing conclusions about population parameters from an analysis of the sample data.*

The two most important types of inferences:

- Estimation of parameters
- Testing of statistical hypothesis

Statistical Inference

Definition

Statistical inference *deals with drawing conclusions about population parameters from an analysis of the sample data.*

The two most important types of inferences:

- Estimation of parameters
- Testing of statistical hypothesis

Example 1.Types of Inference: Point Estimation, Interval Estimation, and Testing Hypotheses

To study the growth of pine trees at an early stage, a nursery worker records 40 measurements of the heights of one-year-old red pine seedlings. This set of measurements is

Heights of One-Year-Old Red Pine Seedlings Measured in Centimeters

2.6	1.9	1.8	1.6	1.4	2.2	1.2	1.6
1.6	1.5	1.4	1.6	2.3	1.5	1.1	1.6
2.0	1.5	1.7	1.5	1.6	2.1	2.8	1.0
1.2	1.2	1.8	1.7	0.8	1.5	2.0	2.2
1.5	1.6	2.2	2.1	3.1	1.7	1.7	1.2

(Courtesy of Professor Alan Ek.)

Example 1.Types of Inference: Point Estimation, Interval Estimation, and Testing Hypotheses

More specifically, depending on the purpose of the study, we may wish to do one, two, or all three of the following:

- Estimate a single value for the unknown μ (point estimation).
- Determine an interval of plausible values for μ (interval estimation).
- Decide whether or not the mean height μ is 1.9 centimeters, which was previously found to be the mean height of a different stock of pine seedlings (testing statistical hypotheses).

Example 1.Types of Inference: Point Estimation, Interval Estimation, and Testing Hypotheses

More specifically, depending on the purpose of the study, we may wish to do one, two, or all three of the following:

- Estimate a single value for the unknown μ (point estimation).
- Determine an interval of plausible values for μ (interval estimation).
- Decide whether or not the mean height μ is 1.9 centimeters, which was previously found to be the mean height of a different stock of pine seedlings (testing statistical hypotheses).

Example 1.Types of Inference: Point Estimation, Interval Estimation, and Testing Hypotheses

More specifically, depending on the purpose of the study, we may wish to do one, two, or all three of the following:

- Estimate a single value for the unknown μ (point estimation).
- Determine an interval of plausible values for μ (interval estimation).
- Decide whether or not the mean height μ is 1.9 centimeters, which was previously found to be the mean height of a different stock of pine seedlings (testing statistical hypotheses).

Example 1.Types of Inference: Point Estimation, Interval Estimation, and Testing Hypotheses

More specifically, depending on the purpose of the study, we may wish to do one, two, or all three of the following:

- Estimate a single value for the unknown μ (point estimation).
- Determine an interval of plausible values for μ (interval estimation).
- Decide whether or not the mean height μ is 1.9 centimeters, which was previously found to be the mean height of a different stock of pine seedlings (testing statistical hypotheses).

Example 2. Inferences about an Unknown Proportion

Example

In California in November 2006, the gubernatorial race pitted the Republican incumbent Arnold Schwarzenegger against the Democratic candidate, Phil Angelides. The exit poll on which TV networks relied for their projections found, after sampling 2705 voters, that 56.5 % said they voted for Schwarzenegger (www.cnn.com/ELECTION/2006). At the time of the exit poll, the percentage of entire voting population (nearly 7 million people) that voted for Schwarzenegger was unknown.

Example 2. Inferences about an Unknown Proportion

Example

1. *How close can we expect a sample percentage to be to the population percentage? For instance, if 56.5% of 2705 sampled voters supported Schwarzenegger, how close to 56.5% is the percentage of entire population of 7 million voters who voted for him?*
2. *The sample proportion $\hat{p}$=56.5 % sheds some light on p, but it is subject to some error since it draws only on a part of the population. We would like to evaluate its margin of error and provide an interval of plausible values of p.*

Point Estimation of a Population Mean

Definition

A statistic intended for estimating a parameter is called a **point estimator**, *or simply an* **estimator.** *The standard deviation of an estimator is called its* **standard error**: *S.E.*

A population mean μ is estimated by sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

The properties of the sample mean $\overline{X}$.

Fact

1. $E(\overline{X}) = \mu$
2. $sd(\overline{X}) = \sigma/\sqrt{n}$ *so* $S.E(\overline{X}) = \sigma/\sqrt{n}$
3. *With large n,* $\overline{X}$ *is nearly normally distributed with mean* μ *and standard deviation* $\sigma/\sqrt{n}$.

The properties of the sample mean $\overline{X}$.

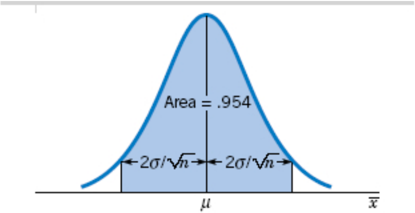

Figure 1
Approximate normal distribution of $\overline{X}$.

When we are approximating μ by $\overline{X}$, the 95.4% **error margin** is $2\sigma/\sqrt{n}$

The $100(1-\alpha)\%$ error margin.

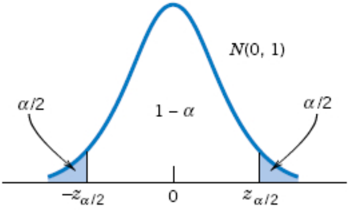

Figure 2 The notation $z_{\alpha/2}$.

Definition

$z_{\alpha/2}$ = *Upper* $\alpha/2$ *point of standard normal distribution.*
That is, the area to the right of $z_{\alpha/2}$ *is* $\alpha/2$*, and the area between* $-z_{\alpha/2}$ *and* $z_{\alpha/2}$ *is* $1-\alpha$.

Values of $z_{\alpha/2}$.

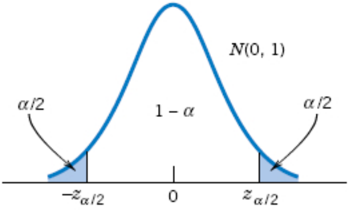

Figure 2 The notation $z_{\alpha/2}$.

$1-\alpha$	.80	.85	.90	.95	.99
$z_{\alpha/2}$	1.28	1.44	1.645	1.96	2.58

Point Estimation of the Mean.

Definition

Parameter: *Population mean* μ.
Data: $X_1, X_2, \ldots, X_n$ *(a random sample of size n)*
Estimator: $\overline{X}$ *(sample mean)*

$$S.E.(\overline{X}) = \sigma/\sqrt{n}$$

Estimated $S.E.(\overline{X}) = S/\sqrt{n}$
For large n, the $100(1-\alpha)\%$ *error margin is* $z_{\alpha/2}\sigma/\sqrt{n}$. *(If* σ *is unknown, use S in place of* σ*.)*

Point Estimation of the Mean Height of Seedings.

Example

From the data of Height of Seedings, consisting of 40 measurements of the heights of one-year-old red pine seedlings, give a point estimate of the population mean height and state a 95 % error margin.

Solution

$$\overline{x} = \sum x_i/40 = 1.715$$

$$s = \sqrt{\frac{\sum x_i - \overline{x}}{40 - 1}} = \sqrt{.2254} = .475$$

To calculate the 95% error margin, we set $1 - \alpha = .95$ *so that* $\alpha/2 = .025$ *and* $z_{\alpha/2} = 1.96$*. Therefore, the 95% error margin is* $1.96s/\sqrt{n} = 1.96 \times .475/\sqrt{40} = .15$ *cm*

Caution.

Note

1. *Standard error should not be interpreted as the "typical" error in a problem of estimation as the word "standard" may suggest. For instance, when* $S.E.(\overline{X}) = 0.3$ *, we should not think that the error* $(\overline{X} - \mu)$ *is likely to be .3, but rather, prior to observing the data, the probability is approximately .954 that the error will be within* $\pm 2(S.E.) = \pm .6$.
2. *An estimate and its variability are often reported in either of the forms: estimate* $\pm S.E.$ *or estimate* $\pm 2(S.E.)$. *In reporting a numerical result such as 53.4* $\pm$ *4.6, we must specify whether 4.6 represents* $S.E.$*,*$2(S.E.)$, *or some other multiple of the standard error.*

Determining the sample size

Definition

To be $100(1-\alpha)\%$ *sure that the error of estimation* $|\overline{X}-\mu|$ *does not exceed* d, *the* **required sample size** *is*

$$n = \frac{(z_{\alpha/2}\sigma)^2}{d^2}$$

Example. How many graham crackers need to be sampled?

Example

Assume that the standard deviation of calories in a graham cracker is 3.5. How many graham crackers need to be sampled if we want to be 90% sure that the population mean calories is estimated within .5 calories?

Solution

We have $d = 0.5$, $\sigma = 3.5$ *and* $z_{\alpha/2} = z_{0.05} = 1.645$. *Hence,*

$$\frac{(1.645 \times 3.5)^2}{0.5^2} = 132.6$$

So, the required sample size is $n = 133$.

For Next Time

- Read Section 8.3 from Johnson and Bhattacharyya
- Group Problems:

Group	1	2	3	4	5
Problem	8.1	8.3	8.5	8.7	8.13
Group	6	7	8	9	10
Problem	8.1	8.3	8.5	8.7	8.13