STA13: Elementary Statistics

Lecture 13
Book Sections 5.1-5.2

Dmitriy Izyumin

February 12 2018

Estimators

An estimator is a rule or formula that tells how to calculate an estimate of a parameter based on the sample.

- The sample mean $\bar{x}$ is an estimator of the pop. mean μ
- The sample median is another estimator of μ
- The sample variance s^2 is an estimator of σ^2

Some Vocabulary

The term estimator will refer to the type of statistic, rule, orformula that is used to estimate a parameter.

- "The estimator used to estimate the average age of CA voters is the sample mean based on a sample of 500 voters."

The term estimate will refer to the specific value of the estimator based on the sample.

- "The sample estimate of the average age of CA voters is 42 years."

Point Estimators

The value of an estimator...

- depends on a random sample,
- varies from sample to sample,
- varies according to a sampling distribution.

Point Estimators

A point estimator is a single number calculated from the sample to estimate the parameter.

Basic Idea:
Based on the sample, give the most likely value of μ.

The sample mean $\bar{x}$ is a point estimator of μ.

Mean of an Estimator

The mean is the "average" or the "expected" value of an estimator.

This affects the accuracy of the estimation.

- mean $>$ parameter $\Rightarrow$ systematic overestimation
- mean $<$ parameter $\Rightarrow$ systematic underestimation

Unbiasedness

We say an estimator is unbiased if its expected value (or mean) is equal to the parameter being estimated.

Otherwise we say an estimator is biased.

An unbiased estimator does not systematically overestimate or underestimate the parameter.

Unbiased Estimators

$\bar{x}$ is an unbiased estimator of the population mean μ.

- $\bar{x}$ estimates μ
- the expected value (or mean) of $\bar{x}$ is μ

Standard Error of an Estimator

The SE of an estimator is its average distance from its mean.

For an unbiased estimator, the SE tells
"how far from the truth" it is on average.

Estimators with smaller SE are preferred (smaller spread).

FACT: $\bar{x}$ has the smallest SE of all unbiased estimates of μ.

Standard Error of an Estimator

When the parameters are unknown (most of the time), we use the sample to compute the standard error.

Sample mean $\bar{x}$:

- True SE: $\frac{\sigma}{\sqrt{n}}$ (use this if σ is known)
- Approximate SE: $\frac{s}{\sqrt{n}}$

Interval Estimators

An interval estimator is an interval calculated from the sample, meant to contain the parameter with a high (pre-specified) probability.

Basic Idea:
Based on the sample, give a plausible range of values for μ.

Confidence Intervals

A confidence interval is a type of interval estimators.

A confidence interval takes the form

$$\text{Point Estimator} \pm \text{Margin of Error}$$

Margin of Error

A confidence interval takes the form

$$\text{Point Estimator} \pm \text{Margin of Error}$$

The margin of error is chosen in such a way that it guarantees a certain degree of confidence that the interval contains the true value of the parameter.

Margin of Error

The ME depends on...

- the estimator's standard error
  (ME increases if SE increases)
- the sample size
  (ME decreases if n increases)
- the desired confidence
  (ME increases if we want larger confidence)

Constructing the Margin of Error

CLT implies that the sample mean is approximately $N(\mu, \sigma/\sqrt{n})$ when the sample is large.

- $P(-1.96 < Z < 1.96) = 0.95 \qquad Z \sim N(0,1)$
- $P(-1.96 < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$
- $P\left(\bar{x} - 1.96(\sigma/\sqrt{n}) < \mu < \bar{x} + 1.96(\sigma/\sqrt{n})\right) = 0.95$
- We are 95% confident that this interval contains μ:

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

General Case

We may want a different confidence level (other than 95%)

For a $(1-\alpha)100\%$ confidence interval

- Let $z_{\alpha/2}$ be the point such that $P(Z > z_{\alpha/2}) = \alpha/2$.
- We are $100(1-\alpha)\%$ confident that this interval contains μ:

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

What if σ is Unknown?

If $n \geq 30$, then...

- s is a fairly good estimate of σ,
- we can use s instead of σ if σ is unknown.

Some Useful Z-values

α	$(1-\alpha)100\%$	$z_{\alpha/2}$
0.1	90%	1.645
0.05	95%	1.96
0.02	98%	2.33
0.01	99%	2.58

CI for the Population Mean - Large Sample

Population: Mean μ, standard deviation σ

Sample: Size $n \geq 30$, sample mean $\bar{x}$, sample s.d. s.

- point estimator of μ is $\bar{x}$
- $(1-\alpha)100\%$ confidence interval for μ is $\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$
  (If σ is known, use it instead of s)
- margin of error is $z_{\alpha/2}\frac{s}{\sqrt{n}}$

Interpreting a Confidence Interval

We say:

We are ___1___% confident that the true value of the population mean is between ___2___ and ___3___ ___4___.

The blanks are:

1. Confidence level $100(1-\alpha)\%$
2. Lower bound of interval
3. Upper bound of interval
4. Units in which the population mean is measured