1

POLS 5377 Scope & Method of Political Science

Week 10 Inferential Statistics

Estimation Procedures

Healey. (2016) *Statistics: A Tool for Social Research,* Chapter 7

2

Key Questions:

- How to use the sample and sampling distribution to estimate the population?
- What is confidence level, how to compute and interpret it?
- How to compute the estimated confidence intervals?
- How to report the estimated confidence intervals?

3

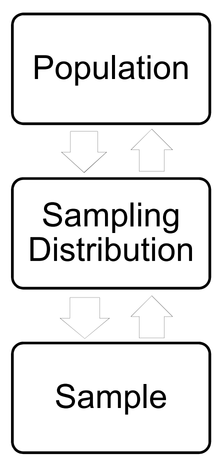
Outline

- Constructing Confidence Interval
- Confidence Interval Estimation
- Report Confidence Interval
- Width of Interval Estimates

4

Logic of Estimation

- Use the sample to estimate the population
- The sample should be unbiased (use EPSEM techniques)
- Every time we draw a random sample, we always have the possibility of sampling error.
- The sample is linked to the population via the sampling distribution
- According to the central limit theorem, if the sample size big enough, the sampling distribution will be
  - Normal in shape
  - $\mu_{\bar{X}} = \mu$
  - $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$

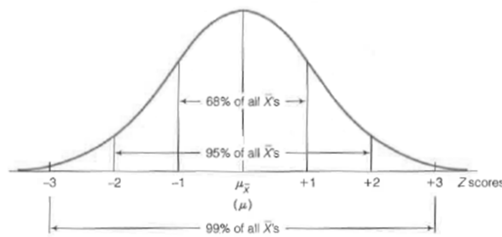


5

Sampling Distribution as a Normal Curve

- The sampling distribution as a normal curve
  - $\mu_{\bar{X}} = \mu$
  - 68% of all possible sample means ($\bar{X}$) is in the range of ±1 z score
  - 95% of all possible sample means ($\bar{X}$) is in the range of ±2 z score
  - 99% of all possible sample means ($\bar{X}$) is in the range of ±3 z score

- Example: Estimate the average income in a community
  - N = 500
  - $\bar{X}$ = $45,000
  - $\mu_{\bar{X}} = \mu = ?$



6

Constructing Confidence Interval

- Constructing confidence interval
  - Step 1: Decide the probability of error: α (alpha)
    - α=0.05 or 95% confidence level are commonly used
    - Sometimes, we may set the probability of error α=0.01 or 99% confidence level

7

Confidence Interval

- Step 2: Find the Z score associated with the α by using the normal curve table
  - If α is equal to 0.05, we would place half (0.025) of this probability in the lower tail and half in the upper tail of the distribution
  - Looking up this area in column c of the Table, we find a Z of 1.96

1.94	0.4738	0.0262
1.95	0.4744	0.0256
$\boxed{1.96}$	0.4750	$\boxed{0.0250}$
1.97	0.4756	0.0244

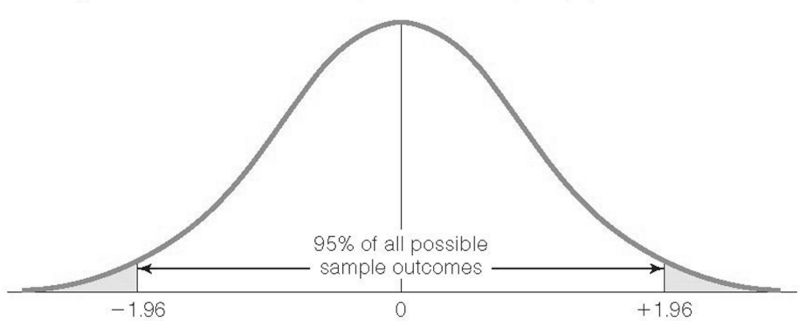
The Sampling Distribution with Alpha (α) Equal to 0.05

8

Constructing Confidence Interval

- Step 2: Find the Z score associated with the α by using the normal curve table

Finding the Z Score That Corresponds to an Alpha (α) of 0.05



9

Constructing Confidence Interval

Finding Z for the confidence level at 85%

- Step 1: decide the value of α
  - Given that the confidence level is 85%, we are willing to be wrong 15% of the time, and alpha equals 0.15
- Step 2: Find the Z score associated with the α
  - Dividing the total area of 0.15 across the two tails of the sampling distribution, we find an area in one tail of 0.0750 (0.15/2)
  - Looking in column c of normal curve table, we find a Z of 1.44 for an area of 0.749 and a Z of 1.43 for an area of 0.0764
  - Choosing between these two Z scores, we would pick 1.44, the larger one.
  - Z = ±1.44

1.41	0.4207	0.0793
1.42	0.4222	0.0778
1.43	0.4236	0.0764
1.44	0.4251	0.0749
1.45	0.4265	0.0735

10

Constructing Confidence Interval

- Z score of common confidence levels

Z Scores for Various Levels of Alpha (α)

Confidence Level	Alpha (α)	$\alpha/2$	Z Score
90%	0.10	0.05	±1.65
95%	0.05	0.025	±1.96
99%	0.01	0.005	±2.58
99.9%	0.001	0.0005	±3.32
99.99%	0.0001	0.00005	±3.90

11

Confidence Interval Estimation

- Confidence intervals for sample means
- Large (N≥100) Samples, σ known

$$c.i. = \bar{X} \pm Z(\frac{\sigma}{\sqrt{N}})$$

- c.i. = confidence interval
- $\bar{X}$ = sample mean
- Z = the Z score determined by α level
- $\frac{\sigma}{\sqrt{N}}$ = the standard deviation of the sample distribution (standard error of the mean)
- $Z = \frac{X_i - \bar{X}}{\sigma} \quad \rightarrow \quad X_i = \bar{X} + Z * \sigma$

12

Confidence Interval Estimation

- Confidence intervals for sample means
- Large (N≥100) Samples, σ Unknown

$$c.i. = \bar{X} \pm Z(\frac{S}{\sqrt{N-1}})$$

- c.i. = confidence interval
- $\bar{X}$ = sample mean
- Z = the Z score determined by α level
- S = the standard deviation of the sample
- *N-1* = number of sample case - 1

13

Confidence Interval Estimation

- Example 1: Large Sample, σ known
  - From a sample of 200 residents, the sample mean IQ is 105 and the population standard deviation is 15
  - Calculate a 95% (alpha = 0.05) confidence interval

$$c.i. = \bar{X} \pm Z\left(\frac{\sigma}{\sqrt{N}}\right)$$
$$c.i. = 105 \pm 1.96\left(\frac{15}{\sqrt{200}}\right)$$
$$c.i. = 105 \pm 1.96\left(\frac{15}{14.14}\right)$$
$$c.i. = 105 \pm (1.96)(1.06)$$
$$c.i. = 105 \pm 2.08$$

14

Confidence Interval Estimation

- Example 2: Large Sample, σ unknown
  - The mean income for a random sample of 500 community residents is $45,000 with a standard deviation of $200
  - Calculate a 95% (alpha = 0.05) confidence interval

$$c.i. = \bar{X} \pm Z(\frac{S}{\sqrt{N-1}})$$
$$c.i. = 45000 \pm 1.96\left(\frac{200}{\sqrt{499}}\right)$$
$$c.i. = 45000 \pm (1.96)(8.95)$$
$$c.i. = 45000 \pm 17.54$$

15

Confidence Interval Estimation

- How to report the confidence interval estimation?
- Example 2:
  - Based on a sample of 500 residents, we can estimate that the average income of residents in this community is $\$45{,}000 \pm 17.54$, at 95% confidence level.
  - Another way to state the interval:
    - $44{,}982.46 \le \mu \le 45{,}017.54$
    - Based on a sample of 500 residents, we estimate that the population mean is greater than or equal to \$44,982.46 and less than or equal to \$45,017.54, at 95% confidence level

16

Confidence Interval Estimation

- Reminder:
  - For 95% confidence level, we provide the estimated values with ± 1.96 standard deviation from the mean of the sampling distribution. It indicates that 95% chance the population mean (μ) is actually in this interval.
  - However, there is 5% chance that the population mean is not included in this interval.

17

Confidence Interval Estimation

- Confidence intervals for sample proportions (large sample)
  - Step 1: decide the α level
  - Step 2: find the z score associated with the α level
  - Step 3: calculate the value with the following formula

$$c.i. = P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}}$$

- P_s = sample proportion
- P_u = population proportion

$$c.i. = \bar{X} \pm Z\left(\frac{\sigma}{\sqrt{N}}\right)$$

18

Confidence Interval Estimation

- Example 1
  - We random sample 200 residents of Houston and find 40% in support of a new public transportation system
  - Calculate a 95% (alpha = 0.05) confidence interval
  - $c.i. = P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}}$

$$c.i. = 0.4 \pm 1.96\left(\sqrt{\frac{0.5(1-0.5)}{200}}\right)$$

$$c.i. = 0.4 \pm 1.96\left(\sqrt{\frac{(0.5)(0.5)}{200}}\right)$$

$$c.i. = 0.4 \pm 1.96\left(\sqrt{\frac{0.25}{200}}\right)$$

$$c.i. = 0.4 \pm (1.96)(0.04)$$

$$c.i. = 0.4 \pm 0.08$$

c.i. = 40% $\pm$ 8%

Based on a sample of 200 residents, we estimate that Houston residents that in support the new public transportation system is between 32% and 48%, with a confidence level of 95%.

19

Confidence Interval Estimation

- Example 2
  - Estimate **the proportion** of students at the university who missed at least one day of classes because of illness last semester
  - In a random sample of 200 students, 60 students reported missing one day of class due to illness
  - Calculate a 90% (alpha = 0.1) confidence interval

20

Confidence Interval Estimation

- Example 2

$$c.i. = P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}}$$

  - According to the statement, the sample proportion is 0.30 (60/200)

$$c.i. = 0.3 \pm 1.65\left(\sqrt{\frac{0.5(1-0.5)}{200}}\right)$$

$$c.i. = 0.3 \pm 1.65\left(\sqrt{\frac{(0.5)(0.5)}{200}}\right)$$

$$c.i. = 0.3 \pm 1.65\left(\sqrt{\frac{0.25}{200}}\right)$$

$$c.i. = 0.3 \pm (1.65)(0.04)$$

$$c.i. = 0.3 \pm 0.066$$

Based on a sample of 200 college students, we are 90% confident that the estimated proportion of students that had missed one day of class due to illness last year is between 0.234 and 0.366, or between 23.4% and 36.6%

21

Report Confidence Interval

- Interpret a confidence interval in a sentence or two that includes information on:
  - The sample statistic (mean or proportion)
  - The calculated confidence interval
  - The sample size (N)
  - The population to which you are estimating
  - The confidence level (often 95%)

22

Report Confidence Interval

Examples:

- c.i. = 45,000 ± 17.55 = $44,982.45 to $45,017.55
  - "Based on a sample of 200 community residents, we are 95% confident that the estimated average income for the entire community is between $44,982.45 and $45,017.55"
- c.i. = 0.30 ± 0.066 = 0.234 to 0.366
  - "Based on a sample of 200 college students, we are 90% confident that the estimated proportion of students that had missed one day of class due to illness last year is between 0.234 and 0.366, or between 23.4% and 36.6%"

23

Width of Interval Estimates

- The width of confidence intervals can be controlled by manipulating two terms in the equation:
  - **The confidence level (or alpha)**
  - When the confidence level increases (and the alpha decreases) the calculated interval is wider

Confidence Intervals Grow Wider as Confidence Levels Increase ($\bar{X}$ = \$45,000, s = \$200, N = 500 throughout)

Alpha (α)	Confidence Level	Interval	Interval Width
0.10	90%	\$45,000 ± \$14.77	\$29.54
0.05	95%	\$45,000 ± \$17.54	\$35.08
0.01	99%	\$45,000 ± \$23.09	\$46.18
0.001	99.9%	\$45,000 ± \$29.71	\$59.42

24

Width of Interval Estimates

- The width of confidence intervals can be controlled by manipulating two terms in the equation:
  - **The sample size**
  - When sample size increases the calculated interval is narrower

Confidence Intervals Grow Narrower as Sample Size Increases ($\bar{X}$ = \$45,000, s = \$200, α = 0.05 throughout)

N	Confidence Interval	Interval Width
100	c.i. = \$45,000 ± 1.96(200/$\sqrt{99}$) = \$45,000 ± \$39.40	\$78.80
500	c.i. = \$45,000 ± 1.96(200/$\sqrt{499}$) = \$45,000 ± \$17.55	\$35.10
1000	c.i. = \$45,000 ± 1.96(200/$\sqrt{999}$) = \$45,000 ± \$12.40	\$24.80
10,000	c.i. = \$45,000 ± 1.96(200/$\sqrt{9999}$) = \$45,000 ± \$3.92	\$7.84

25

Summary

- The formulas for calculating confidence intervals with large sample size

Choosing Formulas for Confidence Intervals

If the Sample Statistic Is a	And the Population Standard Deviation (σ) Is	Use Formula
Mean ($\overline{X}$)	Known	7.1 c.i. $= \overline{X} \pm Z(\sigma/\sqrt{N})$
Mean ($\overline{X}$)	Unknown	7.2 c.i. $= \overline{X} \pm Z(s/\sqrt{N-1})$
Proportion (P_s)		7.3 c.i. $= P_s \pm Z\sqrt{\frac{P_u(1-P_u)}{N}}$

26

After this lecture:

You should be able to:

- Interpret the concept of confidence levels
- Estimate confidence interval of sample means and sample proportions
- Report confidence intervals formally