Lecture 7

Confidence Intervals for Single Population Parameters

Section A: Confidence Intervals for Population Means

2

Learning Objectives

- Upon completion of this lecture section, you will be able to:
  - Explain how the CLT sets the groundwork for computing a confidence interval for an unknown population parameter (mean, proportion, incidence rate) using the results from a single sample
  - Estimate a 95% confidence interval for a population mean, based on results of a single sample from the population
  - Estimate other level confidence intervals (99%, 90%..) for a population mean, based on results of a single sample from the population

3

Central Limit Theorem Revisited

- Recall, the CLT states that if all possible random samples of the same size, n, were taken from the same population, and a summary statistic were computed (mean, proportion, incidence rate) for each sample, then the distribution of the summary statistic values across these samples….

4

Central Limit Theorem Revisited

- Most (95%) of the summary statistic values fall within 2 standard errors of the truth they are estimating (even more (99%) fall within 3 standard errors)

5

Central Limit Theorem Revisited

- In research, only one sample will be taken from each population under study. So how will the CLT help research?

6

Example 1 : Blood Pressure Data, 113 Men

- Example 2: Systolic blood pressure (SBP) measurements from a random sample of 113 adult men taken from a clinical population

(Estimate of μ): $\bar{x} = 123.6$ mmHg

(Estimate of σ): $s = 12.9$ mmHg

7

Example 1 : Blood Pressure Data, 113 Men

- Example 2: Systolic blood pressure (SBP) measurements from a random sample of 113 adult men taken from a clinical population

(Estimate of μ): $\bar{x} = 123.6$ mmHg

(Estimate of σ): $s = 12.9$ mmHg

8

Example 2: Length of Stay Data

- Example 2: Length of stay claims at Heritage Health with an inpatient stay of at least one day in 2011 (12,928 claims)

(Estimate of μ): $\bar{x} = 4.3$ days

(Estimate of σ): $s = 4.9$ days

9

Example 2: Length of Stay Data

- Example 2: Length of stay claims at Heritage Health with an inpatient stay of at least one day in 2011 (12,928 claims)

(Estimate of μ): $\bar{x} = 4.3$ days

(Estimate of σ): $s = 4.9$ days

10

Example 3: Weight Change and Diet Type

- "A Low Carbohydrate as Compared with a Low Fat Diet in Severe Obesity"[1]
  - 132 severely obese subjects randomized to one of two diet groups
  - Subjects followed for six month period

[1] Samaha, F., et al. A Low-Carbohydrate as Compared with a Low-Fat Diet in Severe Obesity, *New England Journal of Medicine*, 348: 21

11

Example 3: Weight Change and Diet Type

- **Scientific Question**—Is Weight Change Associated with Diet Type?

	Diet Group	
	Low-Carb	Low-Fat
Number of subjects (n)	64	68
Mean weight change (kg) Post-diet less pre-diet	-5.7	-1.8
Standard deviation of weight changes (kg)	8.6	3.9

12

Example 3: Weight Change and Diet Type

- 95% CIs For weight change by diet group

Low Carb: $-5.7 \pm 2 \times \frac{8.6}{\sqrt{64}} \rightarrow -5.7 \pm 2 \times 1.08 \approx (-7.8\ kg,\ -3.5\ kg)$

Low Fat: $-1.8 \pm 2 \times \frac{3.9}{\sqrt{68}} \rightarrow -1.8 \pm 2 \times .47 \approx (-2.7\ kg, -0.9\ kg)$

13

Example 4

- Effect of Lower Targets for Blood Pressure and LDL Cholesterol on Atherosclerosis in Diabetes: The SANDS Randomized Trial[2]

- **"Objective** To compare progression of subclinical atherosclerosis in adults with type 2 diabetes treated to reach aggressive targets of low-density lipoprotein cholesterol (LDL-C) of 70 mg/dL or lower and systolic blood pressure (SBP) of 115 mm Hg or lower vs standard targets of LDL-C of 100 mg/dL or lower and SBP of 130 mm Hg or lower."

[2] Howard B et al., Effect of Lower Targets for Blood Pressure and LDL Cholesterol on Atherosclerosis in Diabetes: The SANDS Randomized Trial, *Journal of the American Medical Association (2008)*. 299(14).

14

Example 4

- **"Design, Setting, and Participants** A randomized, open-label, blinded-to-end point, 3-year trial from April 2003-July 2007 at 4 clinical centers in Oklahoma, Arizona, and South Dakota. Participants were 499 American Indian men and women aged 40 years or older with type 2 diabetes and no prior CVD events.

- **Interventions** Participants were randomized to aggressive (n=252) vs standard (n=247) treatment groups with stepped treatment algorithms defined for both."

15

Example 4

- **Results Mean** target LDL-C and SBP levels for both groups were reached and maintained. **Mean** (95% **confidence interval**) levels for LDL-C in the last 12 months were 72 (69-75) and 104 (101-106) mg/dL and SBP levels were 117 (115-118) and 129 (128-130) mm Hg in the aggressive vs. standard groups, respectively.

16

Example 4

- **Lots of 95% CIS!**

Table 2. Differences in Mean Changes From Baseline to 36 Months, Aggressive vs Standard Groups[a]

	Mean (95% Confidence Interval)							
	Baseline		36 mo[b]		Change at 36 mo			P Value for Difference
	Aggressive	Standard	Aggressive	Standard	Aggressive	Standard	Difference	
Weight, kg	90 (88 to 93)	90 (88 to 92)	91 (89 to 94)	91 (88 to 93)	1.0 (−0.8 to 2.2)	1.0 (−0.3 to 2.3)	0.3 (−1.7 to 2.3)	.83
BMI[c]	34 (33 to 34)	33 (32 to 34)	34 (33 to 35)	34 (33 to 34.4)	0.3 (−0.3 to 0.9)	0.4 (−0.1 to 0.9)	0.1 (−0.6 to 0.9)	.77
Waist, cm	110 (108 to 112)	110 (108 to 112)	111 (109 to 113)	110 (108 to 112)	0.2 (−1.0 to 1.6)	0.6 (−0.7 to 2.0)	0.4 (−1.5 to 2.3)	.66
CRP mg/L[d]	2.7 (2.3 to 3.1)	2.8 (2.4 to 3.3)	2.2 (1.9 to 2.7)	3.3 (2.8 to 3.8)	−0.7 (11)[e]	0.9 (9)[e]	1.6 (−0.4 to 3.6)[e]	.12[e]
DBP, mm Hg	74 (73 to 76)	76 (75 to 78)	67 (66 to 68)	73 (72 to 74)	−7 (−8 to −6)	−3 (−4 to −1)	4.0 (2.5 to 5.5)[f]	<.001
SBP, mm Hg	128 (126 to 130)[g]	133 (131 to 135)[g]	117 (115 to 118)	129 (128 to 130)	−11 (−13 to −9)	−3 (−5 to −1)	8 (6 to 12)[f]	<.001
Glucose, mg/dL	159 (151 to 168)	156 (147 to 166)	169 (158 to 179)	169 (158 to 180)	11 (1 to 23)	14 (1 to 28)	4 (−14 to 22)	.68
HDL-C, mg/dL	46 (44 to 48)	46 (44 to 47)	48 (47 to 50)	48 (47 to 50)	3.0 (1.4 to 3.8)	3.0 (1.2 to 3.9)	0.1 (−1.9 to 1.8)	.94
LDL-C, mg/dL	104 (100 to 108)	104 (100 to 108)	72 (69 to 75)	104 (101 to 106)	−31 (−35 to −26)	1 (−3 to 6)	32 (26 to 38)[f]	<.001
Non-HDL-C, mg/dL	138 (134 to 142)	140 (136 to 144)	102 (98 to 106)	138 (135 to 141)	−35 (−40 to −30)	0.2 (−4.4 to 4.9)	35 (28 to −42)[f]	<.001
TC, mg/dL	184 (180 to 188)	185 (181 to 190)	150 (146 to 154)	187 (183 to 190)	−32 (−37 to −27)	3 (−2 to 8)	35 (27 to 42)[f]	<.001
TC/HDL-C, mg/dL	4.2 (4.1 to 4.4)	4.2 (4.1 to 4.4)	3.3 (3.1 to 3.4)	4.0 (3.9 to 4.2)	−1.0 (−1.1 to −0.8)	−0.1 (−0.3 to 0.0)	0.8 (0.6 to 1.0)[f]	<.001
Triglycerides, mg/dL[d]	158 (149 to 167)	168 (159 to 177)	137 (130 to 144)	160 (153 to 168)	−26 (78)[e]	−12 (84)[e]	14 (−3 to 29)[f,d]	.06[e]
Hemoglobin A_{1c}	8.2 (7.9 to 8.4)	7.9 (7.6 to 8.1)	8.3 (8.0 to 8.6)	8.2 (7.8 to 8.5)	0.1 (−0.2 to 0.4)	0.3 (−0.1 to 0.6)	0.2 (−0.3 to 0.6)	.45

Abbreviations: BMI, body mass index; CRP, C-reactive protein; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol.
SI conversions: for CRP to nmol/L, multiply by 9.524; for glucose to mmol/L, multiply by 0.0555; for HDL-C, LDL-C, and TC to mmol/L multiply by 0.0259; for hemoglobin A_{1c} to proportion of total hemoglobin, multiply by 0.01; and for triglycerides to mmol/L, multiply by 0.0113.
[a] Twenty-three baseline variables are compared and presented in Tables 1 and 2.
[b] N for the 36-mo lipids variables was 458 and the mean values were based on the average of 24-, 30- and 36-month observations.
[c] BMI is calculated as weight in kilograms divided by height in meters squared.
[d] Geometric mean (95% confidence interval).
[e] P value is based on arithmetic mean.
[f] Significant mean difference at 36 mo: DBP, LDL-C, non-HDL-C, SBP, TC, TC/LDL-C, and triglycerides (P < .001).
[g] Significant differences at baseline for SBP, P = .003.

17

Example 4

- **Lots of 95% CIS!**

Table 2. Differences in Mean Changes From Baseline to 36 Months, Aggressive vs Standard Groups[a]

	Mean (95% Confidence Interval)							
	Baseline		36 mo[b]		Change at 36 mo			P Value for Difference
	Aggressive	Standard	Aggressive	Standard	Aggressive	Standard	Difference	
Weight, kg	90 (88 to 93)	90 (88 to 92)	91 (89 to 94)	91 (88 to 93)	1.0 (−0.8 to 2.2)	1.0 (−0.3 to 2.3)	0.3 (−1.7 to 2.3)	.83
BMI[c]	34 (33 to 34)	33 (32 to 34)	34 (33 to 35)	34 (33 to 34.4)	0.3 (−0.3 to 0.9)	0.4 (−0.1 to 0.9)	0.1 (−0.6 to 0.9)	.77
Waist, cm	110 (108 to 112)	110 (108 to 112)	111 (109 to 113)	110 (108 to 112)	0.2 (−1.0 to 1.6)	0.6 (−0.7 to 2.0)	0.4 (−1.5 to 2.3)	.66
CRP mg/L[d]	2.7 (2.3 to 3.1)	2.8 (2.4 to 3.3)	2.2 (1.9 to 2.7)	3.3 (2.8 to 3.8)	−0.7 (11)[e]	0.9 (9)[e]	1.6 (−0.4 to 3.6)[e]	.12[e]
DBP, mm Hg	74 (73 to 76)	76 (75 to 78)	67 (66 to 68)	73 (72 to 74)	−7 (−8 to −6)	−3 (−4 to −1)	4.0 (2.5 to 5.5)[f]	<.001
SBP, mm Hg	128 (126 to 130)[g]	133 (131 to 135)[g]	117 (115 to 118)	129 (128 to 130)	−11 (−13 to −9)	−3 (−5 to −1)	8 (6 to 12)[f]	<.001
Glucose, mg/dL	159 (151 to 168)	156 (147 to 166)	169 (158 to 179)	169 (158 to 180)	11 (1 to 23)	14 (1 to 28)	4 (−14 to 22)	.68
HDL-C, mg/dL	46 (44 to 48)	46 (44 to 47)	48 (47 to 50)	48 (47 to 50)	3.0 (1.4 to 3.8)	3.0 (1.2 to 3.9)	0.1 (−1.9 to 1.8)	.94
LDL-C, mg/dL	104 (100 to 108)	104 (100 to 108)	72 (69 to 75)	104 (101 to 106)	−31 (−35 to −26)	1 (−3 to 6)	32 (26 to 38)[f]	<.001
Non-HDL-C, mg/dL	138 (134 to 142)	140 (136 to 144)	102 (98 to 106)	138 (135 to 141)	−35 (−40 to −30)	0.2 (−4.4 to 4.9)	35 (28 to −42)[f]	<.001
TC, mg/dL	184 (180 to 188)	185 (181 to 190)	150 (146 to 154)	187 (183 to 190)	−32 (−37 to −27)	3 (−2 to 8)	35 (27 to 42)[f]	<.001
TC/HDL-C, mg/dL	4.2 (4.1 to 4.4)	4.2 (4.1 to 4.4)	3.3 (3.1 to 3.4)	4.0 (3.9 to 4.2)	−1.0 (−1.1 to −0.8)	−0.1 (−0.3 to 0.0)	0.8 (0.6 to 1.0)[f]	<.001
Triglycerides, mg/dL[d]	158 (149 to 167)	168 (159 to 177)	137 (130 to 144)	160 (153 to 168)	−26 (78)[e]	−12 (84)[e]	14 (−3 to 29)[f,d]	.06[e]
Hemoglobin A_{1c}	8.2 (7.9 to 8.4)	7.9 (7.6 to 8.1)	8.3 (8.0 to 8.6)	8.2 (7.8 to 8.5)	0.1 (−0.2 to 0.4)	0.3 (−0.1 to 0.6)	0.2 (−0.3 to 0.6)	.45

Abbreviations: BMI, body mass index; CRP, C-reactive protein; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure; TC, total cholesterol.
SI conversions: for CRP to nmol/L, multiply by 9.524; for glucose to mmol/L, multiply by 0.0555; for HDL-C, LDL-C, and TC to mmol/L multiply by 0.0259; for hemoglobin A_{1c} to proportion of total hemoglobin, multiply by 0.01; and for triglycerides to mmol/L, multiply by 0.0113.
[a] Twenty-three baseline variables are compared and presented in Tables 1 and 2.
[b] N for the 36-mo lipids variables was 458 and the mean values were based on the average of 24-, 30- and 36-month observations.
[c] BMI is calculated as weight in kilograms divided by height in meters squared.
[d] Geometric mean (95% confidence interval).
[e] P value is based on arithmetic mean.
[f] Significant mean difference at 36 mo: DBP, LDL-C, non-HDL-C, SBP, TC, TC/LDL-C, and triglycerides (P < .001).
[g] Significant differences at baseline for SBP, P = .003.

18

A Note on the Level of Confidence

- 95% confidence intervals are the "industry standard" in research
- It is certainly possible, however, to estimate intervals with different levels of confidence. For example:
  - 90% confidence interval for a population mean

$$\bar{x} \pm 1.65 \times \frac{s}{\sqrt{n}}$$

  - 99% confidence interval for a population mean

$$\bar{x} \pm 2.58 \times \frac{s}{\sqrt{n}}$$

19

Summary

-

20

Section B: Confidence Intervals for Population Proportions and Incidence Rates

21

Learning Objectives

- Estimate a 95% (and other levels) confidence interval for a population proportion, based on results of a single sample from the population
- Estimate a 95% (and other levels) confidence interval for a population incidence rate, based on results of a single sample from the population

22

Example 1

- Response to therapy in random sample of 1,000 HIV+ positive patients from a citywide clinical population[1]

  So with 206 of the 1,000 responding, we have x_i=1 for 206 observations, and x_i=0 for 794 observations.

  So $\hat{p} = 0.206$

[1] http://inclass.kaggle.com/

23

Example 1

- Response to therapy in random sample of 1,000 HIV+ positive patients from a citywide clinical population

  So with 206 of the 1,000 responding, we have x_i=1 for 206 observations, and x_i=0 for 794 observations.

  So $\hat{p} = 0.206$

24

Example 2: Maternal/Infant HIV Transmission

- Results[2]

> ***Results.*** From April 1991 through December 20, 1993, the cutoff date for the first interim analysis of efficacy, 477 pregnant women were enrolled; during the study period, 409 gave birth to 415 live-born infants. HIV-infection status was known for 363 births (180 in the zido-

$$\hat{p} = \frac{53}{363} \approx 0.15 \ (15\%)$$

Of the 363 births whose HIV status was assessed (up to 18 months after birth) , 53 infants were HIV infected.

[2]Connor E, et al. Reduction of Maternal-Infant Transmission of Human Immunodeficiency Virus Type 1 with Zidovudine Treatment. *New England Journal of Medicine* (1994). 331(18); 1173-1180

25

Example 2: Maternal/Infant HIV Transmission

- Results

Of the 363 births whose HIV status was assessed (up to 18 months after birth) , 53 infants were HIV infected.

$$\hat{p} = \frac{53}{363} \approx 0.15 \ (15\%)$$

26

Example 3: Colorectal Cancer Screening

- From Abstract[3]

> **Results:** Compared with those in the usual care group, participants in the intervention groups were more likely to be current for CRC screening for both years with significant increases by intensity (usual care, 26.3% [95% CI, 23.4% to 29.2%]; automated, 50.8% [CI, 47.3% to 54.4%]; assisted, 57.5% [CI, 54.5% to 60.6%]; and navigated, 64.7% [CI, 62.5% to 67.0%]; $P < 0.001$ for all pair-wise comparisons). Increases in screening were primarily due to increased uptake of FOBT being completed in both years (usual care, 3.9% [CI, 2.8% to 5.1%]; automated, 27.6% [CI, 24.9% to 30.0%]; assisted, 30.5% [CI, 27.9% to 33.2%]; and navigated, 35.8% [CI, 33.1% to 38.6%]).

[3]Green B, et al. An Automated Intervention With Stepped Increases in Support to Increase Uptake of Colorectal Cancer Screening: A Randomized Trial. *Annals of Internal Medicine* (2013). 158(5); 301-307

27

Example 3: Colorectal Cancer Screening

- From Abstract

> **Results:** Compared with those in the usual care group, participants in the intervention groups were more likely to be current for CRC screening for both years with significant increases by intensity (usual care, 26.3% [95% CI, 23.4% to 29.2%]; automated, 50.8% [CI, 47.3% to 54.4%]; assisted, 57.5% [CI, 54.5% to 60.6%]; and navigated, 64.7% [CI, 62.5% to 67.0%]; $P < 0.001$ for all pair-wise comparisons). Increases in screening were primarily due to increased uptake of FOBT being completed in both years (usual care, 3.9% [CI, 2.8% to 5.1%]; automated, 27.6% [CI, 24.9% to 30.0%]; assisted, 30.5% [CI, 27.9% to 33.2%]; and navigated, 35.8% [CI, 33.1% to 38.6%]).

$$\hat{p}_{automated} = 0.508 \ (50.8\%)$$
$$\hat{p}_{assisted} = 0.575 \ (57.5\%)$$
$$\hat{p}_{navigated} = 0.647 \ (64.7\%)$$

$$\hat{p}_{usual\_care} = 0.263 \ (26.3\%); \quad n = 1{,}166$$

28

Example 3: Colorectal Cancer Screening

- From Abstract

$$\hat{p}_{usual\_care} = 0.263 \ (26.3\%); \quad n = 1{,}166$$

29

Example 4: 95% CI for IR

- Mayo Clinic: Primary Biliary Cirrhosis (PBC treatment), randomized clinical trial[4]

Among the entire sample of 312 patients, there were 125 deaths for 1,715 years of follow-up time, for an estimated incidence rate:

$$\hat{IR} = \frac{125 \ \text{deaths}}{1{,}715 \ \text{years}} \approx 0.073 \ \text{deaths/year}$$

4 Dickson E, et al. Trial of Penicillamine in Advanced Primary Biliary Cirrhosis. *New England Journal of Medicine*. (1985) 312(16): 1011-1015

30

Example 4: 95% CI for IR

- The standard error for an incidence rate can be estimated as:.

$$\hat{SE}(\hat{IR}) = \frac{\sqrt{\text{number of events}}}{\text{follow-up time}}$$

- So for these PBC data,

$$\hat{SE}(\hat{IR}) = \frac{\sqrt{125}}{1{,}715 \text{ years}} = \frac{11.18}{1{,}715 \text{ years}} \approx 0.0065$$

31

Example 4: 95% CI for IR

- 95% confidence interval for IR

$$\hat{IR} \pm 2 \times \hat{SE}(\hat{IR})$$

- So for these PBC data,

$$0.073 \pm 2 \times 0.0065 \rightarrow 0.073 \pm 0.013$$

Which gives an interval from $(\frac{0.06 \text{ deaths}}{\text{year}}, \frac{0.086 \text{ deaths}}{\text{year}})$

32

Example 4: 95% CI for IR

- Just like the estimated rates can be rescaled, so can the confidence intervals

33

Example 5: 95% CI for IR

- Maternal Vitamin Supplementation and Infant Mortality[5]

Among the entire sample of 10,295 patients, there were 644 deaths for 1,627,725 days of follow-up time, for an estimated incidence rate:

$$\hat{IR} = \frac{644 \text{ deaths}}{1{,}627{,}725} \approx 0.0004 \text{ deaths/day}$$

5 Katz J, West K et al. Maternal low-dose vitamin A or ß-carotene supplementation has no effect on fetal loss and early infant mortality: a randomized cluster trial in Nepal. *American Journal of Clinical Nutrition* (2000) Vol. 71, No. 6, 1570-1576.

34

Example 5: 95% CI for IR

- The standard error for an incidence rate can be estimated as:.

$$\hat{SE}(\hat{IR}) = \frac{\sqrt{\text{number of events}}}{\text{follow-up time}}$$

- So for these infant mortality data,

$$\hat{SE}(\hat{IR}) = \frac{\sqrt{644}}{1{,}627{,}725 \text{ days}} = \frac{25.4}{1{,}627{,}725 \text{ days}} \approx 0.000016$$

35

Example 5: 95% CI for IR

- 95% confidence interval for IR

$$\hat{IR} \pm 2 \times \hat{SE}(\hat{IR})$$

- So for these PBC data,

$$0.0004 \pm 2 \times 0.000016 \rightarrow 0.0004 \pm 0.000032$$

Which gives an interval from $(\frac{0.000037 \text{ deaths}}{\text{day}}, \frac{0.000043 \text{ deaths}}{\text{day}})$

36

Summary

37

Section C: So What is a Confidence Interval, and How Should it Be Interpreted?

38

Learning Objectives

- Gain a conceptual and practical understanding of how to interpret a confidence interval for a single population parameter
- Think critically about when a confidence interval is necessary versus not
- Gain some insight as to why the 95% confidence interval became the standard for research

39

Confidence Intervals In General

- A confidence interval for a population level parameter (mean, proportion, incidence rate) is an interval that factors in the uncertainty in our estimate for the parameter, because we are using data from an imperfect sample
- A confidence interval can be interpreted as a "range of plausible values for the unknown truth"
- Confidence intervals can allow for difference levels of uncertainty: 90%, 95% 99% etc.. However, the standard is 95%, so this is what we will use in the class

40

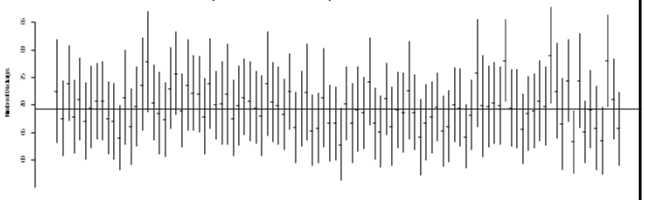
Interpretation

- When a confidence interval is created from a sample of data, the resulting interval either includes the value of the unknown, true parameter, or the interval does not include the truth
  - We will never know if a confidence interval contains the truth
- What about the 95% part?
  - The 95% (or other percent if other level intervals are used) refers to how often the approach to creating a 95% CI "works" in general: in other words, for 95% of samples randomly selected from a population, the 95% confidence interval created from the sample will contain the true value of the parameter (mean, proportion, incidence rate etc..) of interest

41

Interpretation: Example, Hospital Discharges 2011[1]

- True average discharge count: 69.2
  Below shows graphic of 100 95% confidence intervals, each based on a random sample of 250 hospitals

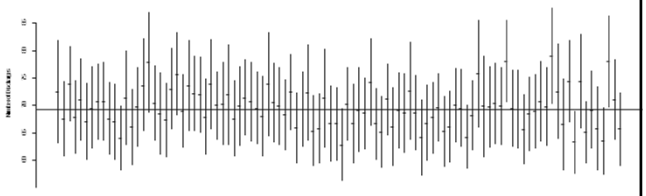


[1] https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html

42

Interpretation: Example, Hospital Discharges 2011

- Reminder: In this example, the population data were known, and simulation was used to repeatedly sample from the population



43

Real Life Research

- In research, however, the truth is not directly observed
- Researchers can only estimate the truth about a population from an (imperfect) data sample
- The confidence interval provides a method for combining the best sample based estimate of a population level quantity, with an estimate of the uncertainty in this quantity with regards to estimating the true, population level value

44

Example 1: Length of Stay Data

- Example 1: Length of stay claims at Heritage Health with an inpatient stay of at least one day in 2011 (12,928 claims)
  (Estimate of μ): $\bar{x} = 4.3$ days (Estimate of σ): $s = 4.9$ days

The 95% confidence interval for the true mean length of stay based on these data was 4.2 to 4.4 days.

Bad News: Researcher will NEVER know if true mean length of stay is within the 95% CI.

Good News: The method used to create the interval works most (95%) of the time.

45

Example 2

- Response to therapy in random sample of 1,000 HIV+ positive patients from a citywide clinical population

  Recall, $\hat{p} = 0.206$, with 95% CI (for p) of 0.18 to 0.23

  Bad News: Researcher will NEVER know if true proportion who would respond to therapy is in the given interval.

  Good News: The method used to create the interval works most (95%) of the time.

46

Does One Always Need to Consider Uncertainty?

- Are confidence intervals always necessary?

47

Why 95%?

- Why 95% as the "industry" Standard?

48

Summary

-

49

Section D: A Note on Confidence Intervals for Population Quantities Based on Small Samples

50

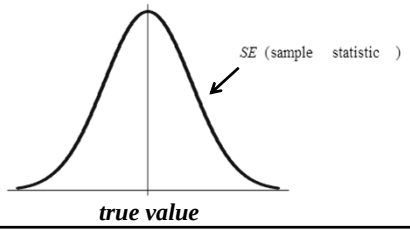
Learning Objectives

- Appreciate and note the role of corrections to the CLT methods when estimating a confidence interval for a mean from a small data sample
- Appreciate and note the role of exact, computer based computations as an alternative to the CLT methods when estimating a confidence interval for a proportion or incidence rate from a small data sample

51

Recap: CLT

- So the CLT tells us the following:
  When taking a random sample of size *n* from a population the theoretical sampling distribution of a sample statistic (mean, proportion, incidence rate) across all possible random samples of size *n* is:



52

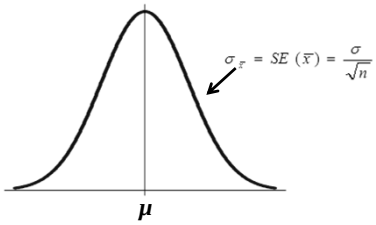
Recap: CLT

- The CLT requires "enough data" to kick in ; for example, with sample means, a sample size cutoff often used is n=60
- Why does CLT require a larger sample? Sample needs to be large enough so that influence of any single sample value is relatively small
- "Large enough" varies depending on data type. The good news is that computers can compute correct 95% (and other level) CIs regardless of the sample size, and the interpretation is the same.

53

Recap: CLT: for means

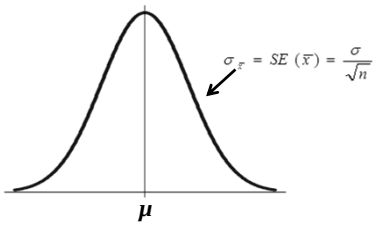
- So the CLT tells us the following:
  When taking a random sample of continuous measures of size *n* from a population with true mean μ and true sd σ the theoretical sampling distribution of sample means from all possible random samples of size *n* is:



54

Recap: CLT

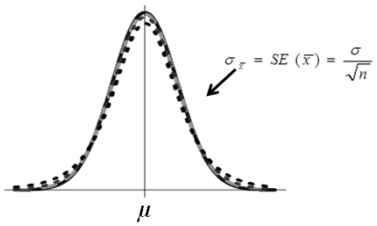
- Technically this is true for "large *n*" : for this course, we'll say n > 60; but when *n* is smaller, sampling distribution not quite normal, but follows a *t-distribution*



55

t-distributions

- The t-distribution is the "fatter, flatter cousin" of the normal: t-distribution uniquely defined by degrees of freedom



56

Why the t?

- Basic idea: remember, the true SE($\bar{x}$) is given by the formula
$$\sigma_{\bar{x}} = SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$
- But of course we don't know σ, and replace with s to estimate
$$S\hat{E}(\bar{x}) = \frac{s}{\sqrt{n}}$$
- In small samples, there is a lot of sampling variability in *s* as well: so this estimates is less precise
- To account for this additional uncertainty, we have to go slightly more than $\pm 2 \times S\hat{E}(\bar{x})$ to get 95% coverage under the sampling distribution

57

Underlying Assumptions

- How much bigger the 2 needs to be depends on the sample size
- You can look up the correct number in a "t-table" or "t-distribution" with *n-1* degrees of freedom, *but in reality, the computer will hand this detail*

58

The t-distribution

- So if we have a smaller sample size, we will have to go out more than 2 SEs to achieve 95% confidence
- How many standard errors we need to go depends on the degrees of freedom—this is linked to sample size
- The appropriate degrees of freedom are *n - 1*
- One option: You can look up the correct number in a "t-table" or "t-distribution" with *n-1* degrees of freedom

$$\bar{x} \pm t_{.95,n-1} \times S\hat{E}(\bar{x}) \Rightarrow \quad \bar{x} \pm t_{.95,n-1} \times \frac{s}{\sqrt{n}}$$

59

Notes on the t-distribution

- The particular t-table gives the number of SEs needed to cut off 95% under the sampling distribution

df	t	df	t
1	12.706	12	2.179
2	4.303	13	2.160
3	3.182	14	2.145
4	2.776	15	2.131
5	2.571	20	2.086
6	2.447	25	2.060
7	2.365	30	2.042
8	2. 360	40	2.021
9	2.262	60	2.000
10	2.228	120	1.980
11	2.201	∞	1.960

60

Notes on the t-Correction

- Can easily find a t-table for other cutoffs (90%, 99%) in any stats text or by searching the internet
- Also, using a computer
- The point is not to spent a lot of time looking up t-values: more important is a basic understanding of why slightly more needs to be added to the sample mean in smaller samples to get a valid 95% CI
- *The interpretation of the 95% CI (or any other level) is the same as discussed before*

61

Example 1

- Small study on response to treatment among 12 patients with hyperlipidemia (high LDL cholesterol) given a treatment
- Change in cholesterol post – pre treatment computed for each of the 12 patients
- Results: $\bar{x}_{change} = -1.4 \ mmol/L$

  $s_{change} = 0.55 \ mmol/L$

62

Example

- 95% confidence interval for true mean change

$$\bar{x} \pm t_{.95,11} \times \hat{SE}(\bar{x}) \Rightarrow$$

$$\bar{x} \pm 2.2 \times \hat{SE}(\bar{x}) \Rightarrow$$

$$-1.4 \pm 2.2 \times \frac{0.55}{\sqrt{12}} \Rightarrow$$

$$(-1.75 \ , mmol/L, \quad -1.05 \ mmol/L)$$

63

Small Samples for Binary and Time-To-Event Data

- For small samples of binary and time-to-event data, there is no adjustment analogous to the t-distribution for creating confidence intervals
- Exact methods need to be employed when creating small sample intervals for proportions and incidence rates; these are handled by the computer
  - Traditionally these were done only with small samples, and the CLT results used otherwise (even by computers)
  - Now, computers generally universally report the results of exact methods, but these are (nearly) identical for CLT based results in large samples

64

Small Samples for Binary and Time-To-Event Data

- Cutoff for small versus large with binary and time-to-event data is not so "cut and dry" as with continuous data

65

Example 2

- Small sample of binary data

Suppose a random sample of 20 MPH students was taken in February of 2013. Students were asked if they currently had a cold. Of the 20 students, 3 had cold symptoms. What is a 95% confidence interval for the true prevalence (proportion) of colds among MPH students at this time?

Here, $\hat{p} = \frac{3}{20} = 0.15 \ (15\%)$

66

Example 2

- Exact 95% confidence interval (from computer)

  0.03 to .38 (3% to 38%)

- CLT based 95% CI

  $0.15 \pm 2\sqrt{\frac{0.15 \times 0.85}{20}}$, which gives a 95% CI of -0.01 to 0.31.

67

Example 3

- Small sample of time to event data

  A pilot study was performed to evaluate the effect of a new therapy on allowing cigarette smokers to quit. 20 smokers were enrolled and followed for up to one month after the start of the therapy. Three in the sample quit smoking in the follow-up period, and there was a total of 313 days of follow-up.

Here, $\hat{IR} = \frac{3 \text{ events}}{313 \text{ days}} \approx 0.0096 \text{ events/day}$

68

Example 3

- Exact 95% confidence interval (from computer)

  0.002 to .028 (events/day)

- CLT based 95% CI

  $0.0096 \pm 2\left(\frac{\sqrt{3}}{313}\right)$, which gives a 95% CI of -0.0014 to 0.021 (events/day)

69

Summary

- With a small sample, adjustments need to be made to the CLT based approaches to estimating confidence intervals for means, proportions, and incidence rates
- Computers can handle these computations
- The interpretation of the resulting 95% (and other level) CIs is exactly the same, regardless of the how the confidence interval is constructed

70