Interval Estimation

SOC424 – Statistics w/ Dr. Ellis Godard

There is ***nothing*** *new in this lecture.*
We're just weaving together baby steps
☺

Where we are in the semester…

#	Date	Read (5th)	Due	Area	Lecture Topic	Lab #	Lab Assignment	T Lab	R Lab
1	Tue Aug 26			*Orientation*	Welcome & Orientation	-			
2	Thu Aug 28	1.1 to 1.4			Basic Terms	-		-	-
3	Tue Sep 2	2.1 & 2.5			Measurement Issues	1	Levels / Age 3x	1	-
4	Thu Sep 4	3.1			Data Reduction	2	Heaven & Hell		1, 2
5	Tue Sep 9	-	HW1	*Description*	Display & Analysis (Shapes & SPSS)	3	Total Miles	2, 3	
6	Thu Sep 11	3.2 & 3.5			Central Tendency	4	CT		3, 4
7	Tue Sep 16	3.3 & 3.4			Dispersion	5	Dispersion	4, 5	
8	Thu Sep 18	3.7			Indices & Data Cleaning	6	Music index		5, 6
9	Tue Sep 23	4.2		*Inference*	Probability & Z Scores	7	Standardizing Scores	6, 7	
10	Thu Sep 25	4.3	HW2		Zs & Ps	8	Table A		7, 8
11	Tue Sep 30	3.6 & 5.1			Parameters & Pt Estimation	9	Distances	8, 9	
12	Thu Oct 2	2.2 to 2.4 & 4.3			Sampling (Issues, Methods, Effects)	10	Sampling		9, 10
13	[illegible]	[illegible]			[illegible]	[illegible]	[illegible]	[illegible]	
14	Thu Oct 9	5.3		*Estimation*	**Confidence Intervals**	12	CI for Intervals		11*, 12
15	[illegible]		[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	
16	Thu Oct 16	6.1 & 6.4			**Hypotesting & Zs**	14	Writing Hypotheses		13, 14
17	Tue Oct 21				**Hypotesting for Large ns**	15	Two Tests	14, 15	
18	Thu Oct 23	6.3 & 6.8	HW4		**The "t" test, for small ns**	16	CI & Test Ages		15, 16
19	Tue Oct 28	5.4			**Sample Size Estimation**	17	Estimating n Needed	16, 17	
20	Thu Oct 30	7.1, 7.3, & 10.1		[illegible]	[illegible]	18	Comparing Means		17, 18
21	Tue Nov 4	7.2	HW5		Differences in Proportions	19	Comparing Proportions	18, 19	
22	Thu Nov 6	12.1			Analysis of Variance	20, 21-EC	ANOVA (+ MODELS EC)		19, 20, 21*
	Tue Nov 11				(Veteran's Day)				
23	Thu Nov 13	9.4 & 9.5			Scatterplots & Correlation	22	Grade Correlations		22
24	Tue Nov 18	9.1 to 9.3	HW6		Regression	23	Regression Lab	20,21*,22,23	
25	Thu Nov 20	10.2 & 11.1			Multiple Regression	24-EC	Multiple Reg (EC)		23, 24*
26	Tue Nov 25	8.1	HW7	*Association*	Crosstabulations	25	TBA (any)	24*, 25	
	Thu Nov 27				(Thanksgiving)				
27	Tue Dec 2	8.2 & p.233			Dependence	26, 27-EC	TBA (SCU) (& 27-EC)	26, 27*	
28	Thu Dec 4	pp.238 to 243	HW8		Association	28-EC	Measures of Assoc (EC)		25,26,27*,28
x	Tue Dec 9				(no lecture - work session only)			28*	
x	[illegible] Dec 11				(no lecture - work session only)				
x	[illegible] Dec 18	-	HW9		(no meetings - deadline only - 10am, firm!)				

Admin Intervals Pictures !!! Examples Lab

Outline for Today…

- **Key Steps** so far
- **Reminders** about Interval & Zs
- **Confidence** About Sample Means
  - Confidence Intervals
  - Confidence Coefficients
  - Z-scores and Confidence Intervals
- **SPSS Illustration**
- **Lab Exercise** (*two* parts… dos… II… 2… not just 1… *2*!)
  - And you need to DO BOTH PARTS to get credit

Admin Intervals Pictures !!! Examples Lab

Key Steps so far

- **Scientific** goal & procedures
- **Level of measurement** is critical
  - **Nominal, ordinal, interval**
- **Descriptions:** Shape, Center, *Spread*
- **Distances:** Sampling Error, Deviations, & *Z Scores*
  - **P values!**
- **Distributions:** Population, Sample, *Sampling*
  - Standard deviation of the sampling distribution is called the standard error (s / square root of n)
- **Estimates:** Point estimates, *Interval Estimates*

Admin Intervals Pictures !!! Examples Lab

Review: Interval Data

- **Level of Measurement**
  - Ordered values (more information than nominal)
  - Equal increments between (more precise than ordinal)
- **Describing a Distribution** (Sample or Pop.)
  - Central tendency: mean
    (median if skew; mode if very high %)
  - Dispersion: standard dev.
    (IQR if skew, range if small)
  - Errors: Sampling ("Mu – Ybar")
    & Standard (sigma/sqrt n)

Review: Parts of Distributions

- **Distance from Mean**
  - Z-score for Yi (vs. Z-score for Ybar)
- **Range of Scores**
  - In a <u>sample</u> distribution (a range of <u>Yi's</u>):
    Ybar +/- z * standard <u>deviation</u>
  - In a <u>sampling</u> distribution (a range of <u>Ybars</u>):
    Ybar +/- z * standard <u>error</u>
- **Associated Probability** ("p-value")
  - Likelihood of getting a value (Yi or Ybar) beyond that range (> z std devs from mean)

Description > Inference

- **Zs -> Ps**
  - For each number of *<u>standard deviations</u>* from the mean of a normal distribution (that is, for each "z")…
    …a known area under the curve ("p") constitutes an estimated percentage of the entire area
- **Example:**
  - In a class with mean of 22 and std dev of 1.5, an age of 25 is 2 stdevs from mean (25 is 2x1.5 from 22 --- 25-22/1.5 = 2)
  - Its z-score is thus 2, with 95.44% of the *values* under the curve occurring with that interval
- **Inference:**
  - Same Zs & Ps, but for sample means (sampling distribution) rather than individual case scores (sample distribution)
  - Since it's a sampling distribution, the denominator is the standard error (std dev of samples means) not the standard deviation (of cases)

Inferential Estimates

- Confidence interval is a range of scores above and below the mean in terms of standard errors
  - Ybar +/- (z x StdError)
    - SE = std dev of sampling distribution,
    - Estimate with stdev / sqrt n
  - e.g. <u>95.44%</u> CI for age: 22 +/- (<u>2</u> x 2.449)
  - e.g. <u>95%</u> CI for age: 22 +/- (<u>1.96</u> x 2.449)
- That's it. ☺
  - The rest of the lecture explains why, and how to interpret it

Confidence Elements

- ***Confidence interval:* interval within which the value of the parameter is believed to lie**
  - $\bar{Y}$ +/- some z # of stdevs, e.g. $\bar{Y}$ +/- 1.96 * SE
  - Width of confidence (z – can look @ w/o p's!)
  - Width of estimate (2* z * std dev)
- ***Confidence coefficient*: the likelihood that the interval contains the true value of the parameter**
  - 1 – probability beyond interval; "% in the middle"
  - e.g. 95% = 1 – 2*(0.0250) and $z_{p=0.0250}$=1.96

Zs + CLT -> C.I.s

- **Z is chosen so that the probability concentrated under a normal curve w/i z standard devs of the mean equals the confidence coefficient.**
  - For 90%, 95%, 98%, and 99%, use the z-values 1.64, 1.96, 2.33, 2.57, respectively.
- **For now, focus on 95% conf. intervals (z=1.96)**
  - Based on the Central Limit Theorem, there is a 95% probability that the population mean falls within 1.96 standard errors of the sample mean
  - $CI_{95\%}$ = Ybar +/- 1.96 * SE

Large-Sample CI for Mu

The confidence interval for the population mean is

$$\bar{Y} \pm z\sigma_{\bar{Y}}$$

or

$$\bar{Y} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$$ (by substitution)

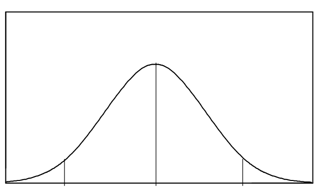
Since we don't know pop. mean, how could we know pop. stdev?

$$\bar{Y} \pm z\sigma_{\bar{Y}} \quad \text{or} \quad \bar{Y} \pm z\left(\frac{\sigma}{\sqrt{n}}\right) \quad \Longrightarrow \quad \bar{Y} \pm z\hat{\sigma}_{\bar{Y}} \quad \text{or} \quad \bar{Y} \pm z\left(\frac{s}{\sqrt{n}}\right)$$

In almost all applied situations, the sample standard deviation is used in lieu of the population standard deviation to compute the standard error.

Relation between Sampling Distribution of Sample Means & Confidence Interval

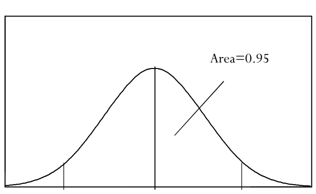
Sampling Distribution of Sample Means



$$\mu - z_{\alpha/2}\sigma_{\bar{Y}} \qquad \mu \qquad \mu + z_{\alpha/2}\sigma_{\bar{Y}}$$

What is the area in each tail?

How likely is it that a sample mean falls more than $z_{\alpha/2}$ standard errors away from the population mean?



$$\mu - 1.96\sigma_{\bar{Y}} \qquad \mu \qquad \mu + 1.96\sigma_{\bar{Y}}$$

If we drew one sample and calculated a 95% C.I., how likely is it that the population mean would fall within the confidence interval?

Confidence Intervals: Logic

- **Have been saying…**
  - **An estimated 95% of the sample means fall within 1.96 standard errors of the pop. mean**
- **Therefore, can imply that…**
  - **A 95% chance that the pop. mean will fall within 1.96 standard errors of the sample mean**

Isn't this counterintuitive?

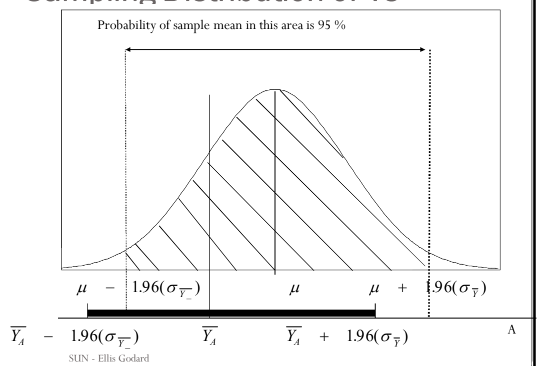
Previously, we began with the population mean of a variable and then asked how likely it was we would obtain a certain sample mean for a sample drawn from that population.

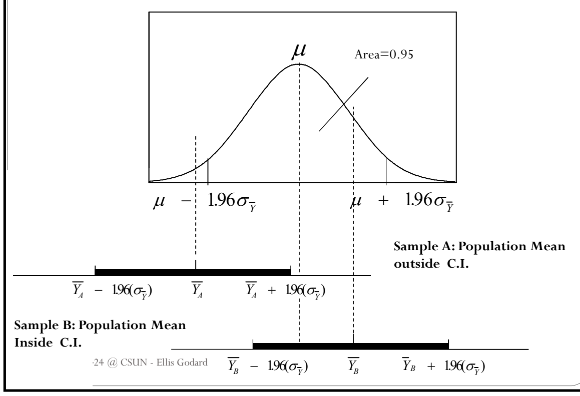
If we already know the population mean, why bother obtaining the sample mean?

No - this is the *reverse*…

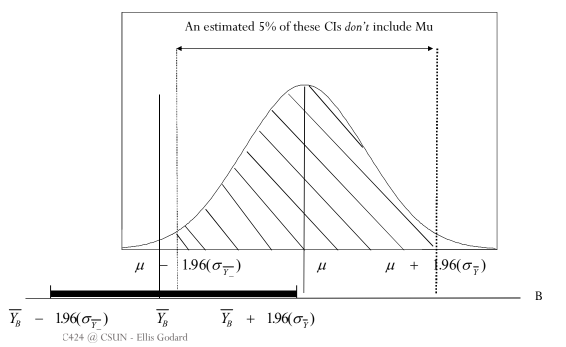
- Have been looking at Mu and asking how unusual a Yi or Ybar would be
- Now looking at Ybar and wanting to know what Mu is likely to be

Sampling Distribution of $\overline{Y}$s





Sampling Distribution of $\overline{Y}$s



That graph in prose…

- If we drew repeated samples of the same size from a population and calculated the 95 percent confidence intervals for each sample

…then…

- we would expect that the population mean would fall in the confidence interval 95 percent of the time.

Confidence Intervals: Leftover

- **Since 95% leaves some tails,**
  - **That is, since an estimated 5% of sample means fall more than 1.96 standard errors from the population mean…**
- **there is a chance our sample's weird.**
  - **…there is *a* 5% chance that we would be wrong in saying that the population mean is within 1.96 standard errors of the sample mean.**

Confidence Intervals: Calculation

- **Formula:**

$$C.I._{95\%} = \overline{Y} \pm (1.96 \times stderror)$$

- Identifies a range of actual scores, varying w/ confidence *and* standard error:
  - The larger the standard error, the wider the conf. Interval
  - If want a narrower range, must accept lower confidence

Example 1

- **For Family A** (from previous lecture)**:**
  - 10 +/- (1.96 x 2.449) =
    - 10 +/- 4.8 =
    - 5.2 to 14.8
  - If the six cases in that family represented a random sample from some larger population, then we could be 95% confident that that full population watches an average (mean) that falls between 5.2 and 14.8 hours of television per week

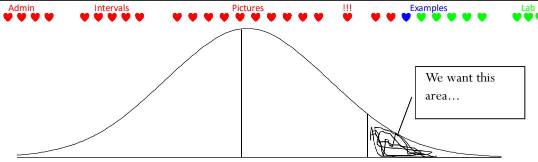
Example 2 (Prob 15 part b, last lecture)

(b) If we plan on taking a random sample of 64 migrant workers, what is the sampling distribution of the sample mean income? Find the probability that the sample mean exceeds $9,000.

Solution:

1. Compute the standard error:

$$\sigma_{\bar{Y}} = \sigma / \sqrt{n} = 3{,}200 / \sqrt{64} = 400$$



$\mu = \$8{,}000$ $\bar{Y} = \$9{,}000$

3. Compute the z-score and use Table A to find area:

$$z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \frac{9{,}000 - 8{,}000}{400} = 2.50$$

$$P(z > 2.50) = .0062$$

4. State conclusion in full sentence(s): The probability that the mean income of a sample of 64 migrant worker is above $9,000 is equal to 0.006 or 0.6 percent.

Computing the 95 % C.I.

$$\bar{Y} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$9{,}000 \pm 1.96\left(\frac{3{,}200}{\sqrt{64}}\right) =$$

$$9{,}000 \pm 784 = \{8{,}216 - 9{,}784\}$$

The 95% confidence interval for the population mean is from $8,216 to $9,784.
That is, we can be 95% confident that this interval contains the population mean of yearly income for migrant workers in California.

Example 3

Using a sample of 384 farms in Brazil, the mean acreage was 450 and the sample standard deviation was 100.

Construct a 95 percent confidence interval for the mean acreage of farms.

Draw a Picture!

$\bar{Y} - 1.96(\sigma_{\bar{Y}})$ $\bar{Y}$ $\bar{Y} + 1.96(\sigma_{\bar{Y}})$

(Lower Limit)	(Upper Limit)
450 - 1.96(5.10)	450 + 1.96(5.10)
450 - 10	450 + 10
440	460

Conclude: "With a probability of 95 percent, the mean acreage of farms in Brazil falls between 440 and 460."

SPSS Illustration

- Calculate a 95% confidence interval for RSHOWS
  - Request the mean and standard deviation
    - Use ANALYZE – DESCRIPTIVES – FREQUENCIES
  - Calculate standard error
  - Check by having SPSS calculate standard error
    - Use the "S.E. mean" option in "Statistics…"
  - Calculate 95% Confidence Interval for the mean
  - Check by having SPSS calculate that interval
    - Use ANALYZE – DESCRIPTIVES – EXPLORE
  - Draw a picture to summarize

Lab, Part I (Qs 1- 3)

If the mean for a variable is 1000 and the standard *error* is 10, construct and interpret…

1. **a 95.44% confidence interval**
2. **a 95% confidence interval**
3. **a 90% confidence interval**

DON'T STOP – THERE ARE TWO PARTS
(AND DON'T JUST DO THE OTHER ONE) –
YOU MUST DO BOTH FOR CREDIT

That math from SPSS…

CI95 = 1.1923 +/- (1.96 * (1.23730/sqrt(52)))
= 1.1923 +/- (1.96 * (1.23730/7.21))
= 1.1923 +/- (1.96 * (.171))
= 1.1923 +/- (.33516)
= .857 to 1.52

Check w/ "S.E. Mean" under "Statistics"

We are 95% confident that, in the population, people watch between .857 and 1.52 of those reality shows.

Check w/ Analyze > Desc > Explore

Lab, Part II (did you skip Part I?) in SPSS

- Using MUSIC index from an earlier lab
  - Measuring the number of genres that respondents like
  - Use musicb.sav, or you'll have to recode & compute again
  - If you already have mean & std. dev, don't need SPSS again
  - Check w/ SPSS, but *do the math* & *show it!*

4. Calculate the standard error for MUSIC
5. Calculate $CI_{95\%}$ for MUSIC
6. Draw a picture!
   - Refer to lecture b4 saying "I don't know how"
   - And start paying attention to lecture… dude, we're drawing PICTURES!! How cool is THAT?!? ☺

Calculation Reminders…

- Must do parts in order
  - PPMDAS – Pretty Please My Dear Aunt Sally
    - Parentheses (then) Powers (then) Multiplication (then) Division (then) Addition (then) Subtraction
  - PEMDAS – Please Excuse (Eat?) My Dear Aunt Sally
    - Parentheses (then) Exponents (then) Multiplication (then) Division (then) Addition (then) Subtraction
- For formula Ybar +/- z * s.e.
  - Don't do the +/- z, and THEN multiply by s.e.
    - Must multiply z times s.e., and THEN +/- it to Ybar
  - Don't do it in *reverse* – don't +/- Ybar to the rest!