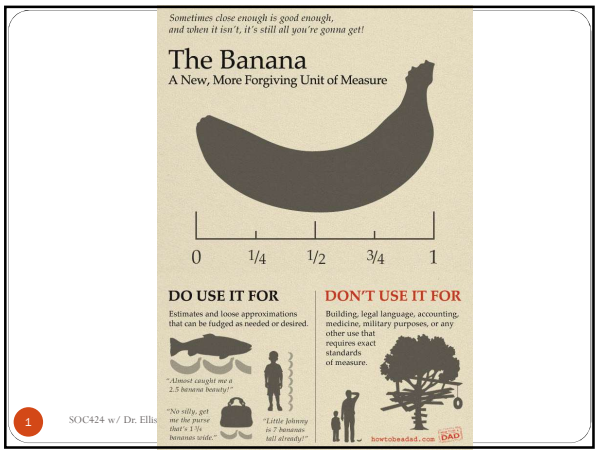


Admin Review Width Size Examples L

Outline for Today…

- **Preview *Next* Lecture: *Bivariate Beginnings***
- **Understanding a Confidence Interval**
  - Width of Interval
  - Impact of Sample Size
- **Choosing an Appropriate Sample Size**
  - To Estimate a Populatio Mean
  - To Estimate a Population Proportion
- **Data Example & SPSS Example**
- **Lab Exercise**

SOC424 – Statistics w/ Dr. Ellis Godard

Estimating n: Choosing a Sample Size

Admin Review Width Size Examples L

Next lecture: Bivariate starts

- **Introduction to Covariation**
  - Critical Juncture: Univariate to Bivariate
- **Will start w/ summary of class so far**
  - *Must* be comfortable with:
    - Level of measurement
    - sampling (*not* sample) distributions
    - standard errors
    - p values

Admin Review Width Size Examples L

Admin Updates

- LOTS of graded work coming
  - Prepare for another deluge :)
- PLEASE use office hours ☺ or… *whatever*
  - You have 34+ ways to contact me, 24/7 – use 'em!!

Admin Review Width Size Examples L

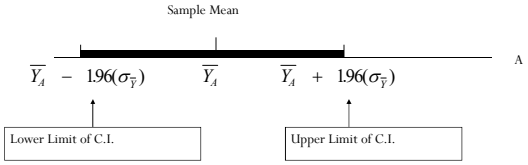
Rest of Semester: Significance Testing

- Inverts Estimation
- Four kinds of tests we'll discuss:
  - "Chi-square" (χ^2)
  - "t-test" (Student's "**t**")
  - ANOVA ("**F** test")
  - Regression ("**r**" correlation)
- All follow a common pattern:
  - Like a z score, but comparing the sample to the null
  - One # (usually an observed stat, such as a mean) minus another # (usually some other expectation), divided by some form of (standardized) error
  - Such as one *mean* minus another, or one *difference in means* minus another

Sampling Distrib's CT is unknown

We don't know where to place the sampling distribution, since we don't know the population mean – could go either direction, but CI around $\overline{Y}$ is same as CI around μ

A generic 95 percent confidence interval:



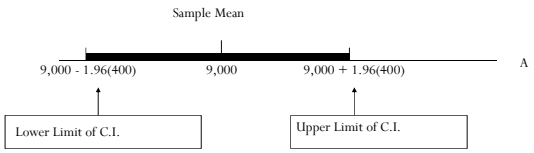
The *interval width* is the difference between the upper and lower confidence interval. We're 95% confident that the **population** mean is somewhere between these values.

Width of Confidence Interval…

- **Matters tremendously**
  - Want knowledge w/i some range
  - Want to know range of likely Mu's
  - But also want high level of confidence with as small a sample as possible
    - Time, money, manpower, etc. (497)

Our Example
(Mean Income of Migrant Workers in California)

A 95 percent confidence interval:

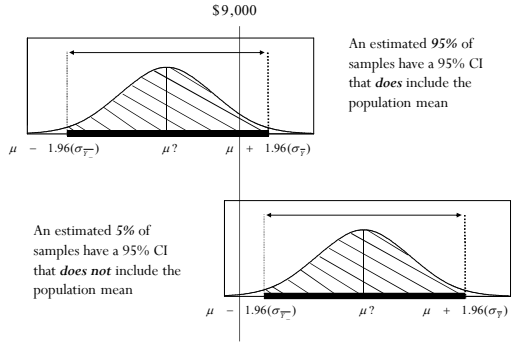


Conclusion: With 95 percent confidence, the (population) mean income of **all** migrant workers in California falls between $8,216 and $9,784.

Width of Confidence Interval…

- **Is controllable by researcher**
  - increases as the confidence coefficient increases (wider if higher confidence)
  - decreases as the sample size increases (wider if small sample size)

Central Tendency of Pop = Unknown



1st Way to Increase the Accuracy

- Smaller confidence coefficient (e.g. 95% instead of 99%) leaves more room in the tails, so involves
  - a smaller distance from the sample mean,
  - which means a lower z score,
  - which means less +/- in each direction,
  - which results in a smaller interval

Decrease Confidence -> Narrower Interval

Change confidence > new z's

If confidence coefficient is …

98 percent, then $z_{0.01} = 2.33$;

99 percent, then $z_{0.005} = 2.57$;

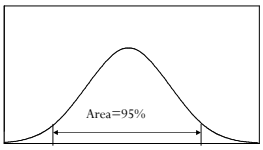
These are the z-scores used to calculate confidence intervals.

$0.005 = (1.00 - 0.99)/2$ ← There are two tails

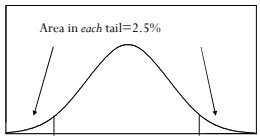
Total Area Under Normal Curve ← Confidence Coefficient

Where do these z's come from?

Table A, of course!



(If Area in middle is 95% (0.95), then the area in *both* tails combined is 0.05.)



(by Symmetry)

$z_{0.025} = 1.96$ ——(from Table A)

Width Increases as Coefficient Increases

95 Percent C.I. : $z_{0.95} = 1.96$

$\overline{Y}_A - 1.96(\sigma_{\overline{Y}})$ $\overline{Y}_A$ $\overline{Y}_A + 1.96(\sigma_{\overline{Y}})$

98 Percent C.I. : $z_{0.98} = 2.33$

$\overline{Y}_A - 2.33(\sigma_{\overline{Y}})$ $\overline{Y}_A$ $\overline{Y}_A + 2.33(\sigma_{\overline{Y}})$

99 Percent C.I.: $z_{0.99} = 2.57$

$\overline{Y}_A - 2.58(\sigma_{\overline{Y}})$ $\overline{Y}_A$ $\overline{Y}_A + 2.58(\sigma_{\overline{Y}})$

Trade-Off: Confidence vs. Accuracy

The width of the confidence interval increase as the confidence coefficient increases.

Extreme Cases:

What confidence coefficient would we select to be *almost certain* that the confidence interval includes the population mean?

What is the *smallest possible* confidence interval, and what is the confidence coefficient associated with it?

2nd Way to Increase the Accuracy

A larger sample reduces the size of the standard error *(because it's the denominator, the bottom part of the fraction)*, which in turn reduces the width of the confidence interval.

Increase n -> Narrower Interval

Examples

Calculate the 95% confidence interval, if the sample of migrant workers in California was …

$n = 64;\quad 9{,}000 \pm 1.96\left(\frac{3{,}200}{\sqrt{64}}\right) = 9{,}000 \pm 784$

$n = 100;\quad 9{,}000 \pm 1.96\left(\frac{3{,}200}{\sqrt{100}}\right) = 9{,}000 \pm 627$

$n = 1000;\quad 9{,}000 \pm 1.96\left(\frac{3{,}200}{\sqrt{1000}}\right) = 9{,}000 \pm 101$

Admin Review Width Size Examples

Why "Large Sample"?

If the sample isn't "large" (30+ cases), we cannot assume that the sampling distribution is approximately normal.

And if we cannot make this key assumption, we can't invoke the CLT, and the relation between the z-scores and confidence coefficients is not applicable.

Admin Review Width Size Examples

Alpha – the tails

- Alpha is the "ideal p"
  - How big can the "tails" be?
  - Alpha + Confidence Coefficient = 1 (100%)
- Confident Coefficient = 1 – alpha
  - e.g. for an alpha of .05, CC = 95%
- Alpha = 1 – Confidence Coefficient
  - e.g. for a CC of 95%, alpha is .05

Admin Review Width Size Examples

Sample Size Estimation

Key questions in research…

- How many cases do I have to observe?
- What's the minimum sample size needed to be "confident" in the results?
- Is our sample size sufficient to make conclusions about the central tendency, dispersion, and/or relationships observed?

Admin Review Width Size Examples

Margin of Error – how wrong?

- ***Not* the same as….**
  - Standard error, standard deviation, sampling error
  - In a confidence interval, it's the standard error * z (or t)
- **Answers questions such as…**
  - How wrong can we be?
  - How "off" can our point estimate be?
  - How much can our estimate deviate from the population parameter?
  - How wide can the confidence interval be, *not* in z's or p's, but in *units of the actual variable*?
- **Examples:**
  - +/- 3% in a poll to predict an election outcome
  - +/- 2 years of age for a marketing study

Admin Review Width Size Examples

Factors to Consider in Choosing n

- Resources
  - *How much research money do you have*
- Amount and Type of Analysis
  - *More complicated studies require larger samples*
- Degree of Variability
  - *Use larger samples for more heterogeneous populations*
- Confidence
  - *Probability that interval will contain population mean*
  - **How confident do we want or need to be? (alpha, CC)**
- Precision
  - *the width of the confidence interval*
  - **How *wrong* can we be? (margin of error)**

Admin Review Width Size Examples

Sample Size Required for Estimating Population Mean

If we want the "margin of error" to be smaller than some value B, with a probability of at least $1-\alpha$, the appropriate sample size is obtained as follows:

$$n = \sigma^2 \left(\frac{z_{\alpha/2}}{B} \right)^2$$

You did something like that on a recent homework, solving for n – this just moves those pieces around, and gives the Margin of Error a name & symbol

Sample Size Required for Estimating Population Proportion

If we want the "margin of error" to be smaller than some value B, with a probability of at least 1 -α , the appropriate sample size is obtained as follows

$$n = .25\left(\frac{z_{\alpha/2}}{B}\right)^2$$

Why .25?

- Because the largest possible *(and therefore most conservative)* variance for the sampling distribution of a proportion is when the proportion is 50%
  - *see earlier lecture on interval estimates for proportions, slide 17*
  - *A 50/50 split would be 0.5 * 0.5 = 0.25*
- May sometimes have a better (smaller) estimate of the variance, *if* you have an estimate of the *actual* population proportion *(maybe it's 10%, not 50%)*
  - *A 10/90 split would be 0.1 * 0.9 = 0.09*

Estimating Min. Necessary Sample Size

The formula for n is the *variance*, multiplied by the square of the ratio of *the z score associated with the desired confidence level* and the *allowed margin of error*:

$$n = \hat{\sigma}^2 * \left(\frac{z_{a/2}}{B}\right)^2 = p*(1-p)*\left(\frac{z_{a/2}}{B}\right)^2$$

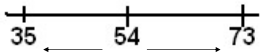
When the population proportions are unknown and we must use the conservative estimate of 50% (0.5), then to have a 95% level of confidence in a margin of error of not more than +/- 4% requires a sample size of 601 cases*:

$$0.5*(0.5)*\left(\frac{1.96}{.04}\right)^2 = .25(49)^2 \approx 601$$

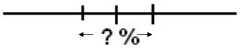
* Per subgroup or disaggregated unit

Class Survey Example (cont'd)

- What % of CSUN students have smoked cannabis?
- Calculated estimate ~35-73% based on n=26
- Allows our interval estimation to be wrong by 19% in either direction of the point estimate (54% +/- 19%)



- How many students must we survey if we want to be 95% confident within a margin of error of only 3% (not 19)?



Class Survey Example (cont'd, II)

- Conservative approach:

$$n = .25\left(\frac{z_{\alpha/2}}{B}\right)^2$$

$$= .25\left(\frac{1.96}{.03}\right)^2 = .25(65.3)^2$$

$$= .25(4268.4) = 1067.1$$

- Can't have 0.1 of a respondent, and want to be conservative, so ***round up*** (!) to 1068

Class Survey Example (cont'd, III)

- Using our point estimate, an *informed* approach can take advantage of a smaller estimated variance:

$$= p*(1-p) = .54*.46 = .2484$$

- So…

$$n = .2484\left(\frac{z_{\alpha/2}}{B}\right)^2$$

$$= .2484\left(\frac{1.96}{.03}\right)^2 = .2484\,(65.3)^2$$

$$= .2484\,(4268.4) = 1060.3 \approx 1061$$

- Note that a smaller sample size is needed *(HW hint!)*

SPSS Example (using musicb.sav)

Remember the index MUSIC?

The mean of *that sample* was 6.1, w/ a std dev of 2.6479

If we want to estimate the *population* mean,

and want to be 95% confident

that that estimate is accurate to within 1 genre,

how many people would we need to survey?

$$\sigma^2 * \left(\frac{z_{\alpha/2}}{B}\right)^2 = 7.0161 * \left(\frac{1.96}{1}\right)^2 = 26.953.. \approx 27$$

To have 95% confidence that the sample mean will be within one year of the population mean, we would need to survey 27 respondents.

For your next lab…

- **Use child.sav** (a study of child abductions by family members)
- **Filter out the (control group) cases that are not abductions**, using DATA – SELECT CASES and "If condition satisfied" to select only those for which "faprbs > -1". (You should soon see that there are 93 *cases*. If you don't, you're probably confusing cases and variables.)
- **Find a 95% C.I. for the mean age at which children are abducted by family members.**
- **Find a 95% C.I. for the proportion of family-abducted children likely to be female** (1 is male, 2 is female).
- **Estimate the minimum sample size necessary if we want to estimate the mean age at which children are family-abducted, and we want our point estimate to have a margin of error of not more than two years, with 95% confidence.**