

## Lecture 25/Chapter 21

### Estimating Means with Confidence

- Example: Meaning of Confidence Interval
- Reviewing Conditions and Rules
- Constructing a Confidence Interval for a Mean
- Matched Pairs & Two-Sample Studies

### Inference for Proportions then Means (Review)

**Probability** theory dictated behavior of sample proportions (categorical variable of interest) and sample means (quantitative variable) in random samples from a population with known values. Now we're performing **inference** with **confidence intervals**

- for proportions (Chapter 20)
  - for means (Chapter 21)
- or with **hypothesis testing**
- for proportions (Chapters 22&23)
  - for means (Chapters 22&23)

### Two Forms of Inference (Review)

**Confidence interval:** Set up a range of plausible values for the unknown population proportion (if variable of interest is categorical) or mean (if variable of interest is quantitative).

**Hypothesis test:** Decide if a particular proposed value is plausible for the unknown population proportion (if variable of interest is categorical) or mean (if variable of interest is quantitative).

### Example: The Meaning of a Confidence Interval

- **Background:** 625 households in a city were polled; their size (in persons) had mean 2.3, sd 1.75. A 95% confidence interval for pop. mean size is (2.16, 2.44).
- **Question:** Which of these is/are correct?
  - 95% of the households in the sample have 2.16 to 2.44 people.
  - 95% of the households in the city have 2.16 to 2.44 people.
  - The probability is 95% that mean household size in this city is between 2.16 and 2.44 people.
  - The probability is 95% that the interval we constructed by this method contains the unknown pop. mean household size.
  - We're 95% sure that pop. mean is btw. 2.16 and 2.14 people.

### Response:

To see why, we should follow steps in interval's construction...

## Conditions for Sample Means (Review)

- Randomness [affects center]
- Independence [affects spread]
  - If sampling without replacement, sample should be less than 1/10 population size
- Large enough sample size [affects shape]
  - If population shape is normal, any sample size is OK
  - If population is not normal, a larger sample is needed.

## Rule for Sample Means (if conditions hold)

- **Center:** The mean of sample means equals the true population mean.
- **Spread:** The standard deviation of sample means is standard error =  $\frac{\text{population standard deviation}}{\sqrt{\text{sample size}}}$
- **Shape:** (Central Limit Theorem) The frequency curve will be approximately normal, depending on how well 3rd condition is met.

## Empirical Rule; Probability to Inference

- For any normal curve, approximately
- 68% of values are within 1 sd of mean
  - 95% of values are within 2 sds of mean
  - 99.7% of values are within 3 sds of mean
- The **probability** is 95% that sample mean from a random sample falls within 2 sds of pop. mean.
- We are 95% **confident** that unknown population mean falls within 2 sds of the sample mean.
- In the long run, 95% of our 95% confidence intervals will contain the unknown pop. mean.

## Approximating Standard Error

- The sd (standard error) of sample mean is  $\frac{\text{population standard deviation}}{\sqrt{\text{sample size}}}$  which we approximate with  $\frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$  when the population standard deviation is unknown.

## 95% Confidence Interval for Population Mean

An approximate 95% confidence interval for population mean is

$$\text{sample mean} \pm 2 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

**Note:** the multiplier 2 comes from the 95% part of the 68-95-99.7 Rule, which only applies to normal curves. The interval will be incorrect if our sample is too small.

## Example: Confidence Interval for a Mean

- **Background:** 625 households in a city were polled; their size (in persons) had mean 2.3, standard deviation 1.75.
- **Question:** What is a 95% confidence interval for population mean household size?
- **Response:** sample mean  $\pm 2 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$   
= \_\_\_\_\_ . We're 95% confident that the unknown population mean household size falls in this interval; our method has a 95% success rate.

## 95% Confidence Interval for Population Mean

An approximate 95% confidence interval for population mean is

$$\text{sample mean} \pm 2 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

**Note:** the multiplier 2 comes from the 95% part of the 68-95-99.7 Rule, which only applies to normal curves. The interval will be incorrect if our sample is too small.

## Example: Confidence Interval for Mean Weight

- **Background:** Weights (in lbs) for a sample of 52 college women had mean 129, sd 20.
- **Question:** What can we say about the mean weight of all college women?
- **Response:** We're 95% confident that the unknown population mean weight falls in the interval \_\_\_\_\_

## Example: Confidence Interval for Mean Male Weight

- **Background:** Weights (in lbs) for a sample of 28 college men had mean 168, sd 27.
- **Question:** What can we say about the mean weight of all college men?
- **Response:** We're 95% confident that the unknown population mean weight falls in the interval \_\_\_\_\_

### Example: Width of a Confidence Interval

- **Background:** 95% confidence intervals for pop. mean wts are  $129 \pm 2 \frac{20}{\sqrt{52}} = 129 \pm 5.6 = (123.4, 134.6)$  for women, and  $168 \pm 2 \frac{27}{\sqrt{28}} = 168 \pm 10.2 = (157.8, 178.2)$  for men.
- **Question:** Why is the interval wider for men?
- **Response:** First, \_\_\_\_\_  
Second, \_\_\_\_\_

### Example: What Can We Infer About Population?

- **Background:** 95% confidence intervals for pop. mean wts are  $129 \pm 2 \frac{20}{\sqrt{52}} = 129 \pm 5.6 = (123.4, 134.6)$  for women, and  $168 \pm 2 \frac{27}{\sqrt{28}} = 168 \pm 10.2 = (157.8, 178.2)$  for men.
- **Questions:** Is 160 lbs a plausible population mean weight for all women? For all men?
- **Responses:** For women: \_\_\_\_\_  
For men: \_\_\_\_\_

### Sample Size, Width of 95% Confidence Interval

Because sample size appears in the denominator of the confidence interval for population mean  $\text{sample mean} \pm 2 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$  smaller samples (less info) produce wider intervals; larger samples (more info) produce narrower intervals.

### Empirical Rule (Review)

- For any normal curve, approximately
- 68% of values are within 1 sd of mean
  - 90% of values are within 1.645 sd of mean
  - 95% of values are within 2 sds of mean
  - 99% of values are within 2.576 sds of mean
  - 99.7% of values are within 3 sds of mean

Fine-tune the information near 2 sds, where probability % is in the 90's.

## Intervals at Other Levels of Confidence

An approximate 90% confidence interval for population mean is

$$\text{sample mean} \pm 1.645 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

An approximate 99% confidence interval for population mean is

$$\text{sample mean} \pm 2.576 \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

**Example:** *A 90% Confidence Interval*

- **Background:** Suppose amount spent on textbooks in a semester by a random sample of 25 students had mean \$500, standard deviation \$100.
- **Question:** What is a 90% confidence interval for the mean amount spent by all students?
- **Response:**

**Example:** *A 99% Confidence Interval?*

- **Background:** The mean exam score for the 64 female Stat 800 students is 120, with standard deviation 19.
- **Question:** Is  $120 \pm 2.576(19)/8 = (114, 126)$  a 99% confidence interval for the mean score of the entire class of 100 students?
- **Response:** \_\_\_\_\_
  - 1st condition: \_\_\_\_\_
  - 2nd condition: \_\_\_\_\_

Paired Studies (or Matched Pairs)

To estimate the overall difference in pairs of measurements for a variable, focus on the single sample of differences.  
An approximate 95% confidence interval for the population mean of differences is  
sample mean diff  $\pm 2 \frac{\text{standard deviation of sample diffs}}{\sqrt{\text{sample size}}}$

### Example: Confidence Interval in a Paired Study

- **Background:** For a sample of 400 college students, we consider fathers' age minus mothers' age. The age differences have mean 2.4, sd 4.0.
- **Questions:** What is a 95% confidence interval for the mean of differences (in percentages) for all college students? How do we interpret the interval?
- **Response:** We are 95% confident that for all students, fathers are older by \_\_\_ to \_\_\_ years, on average.

### Two-Sample Studies

To estimate the difference between population means for two separate groups, we use the difference between sample means, the two sample standard deviations (1st and 2nd sd) and the two sample sizes.

An approximate 95% confidence interval for the difference between population means is

$$\text{diff btw. sample means} \pm 2 \sqrt{\frac{(\text{1st sd})^2}{\text{1st sample size}} + \frac{(\text{2nd sd})^2}{\text{2nd sample size}}}$$

### Example: CI for Difference btw Two Means

- **Background:** No. of cigarettes in a day by 8 female smokers: mean 11, sd 10; 4 males had mean 7, sd 5.
- **Question:** How many more cigarettes do female students smoke in general compared to males?
- **Response:** We're 95% confident that the unknown difference between population means falls in the interval \_\_\_\_\_ so on average \_\_\_\_\_ they might smoke anywhere from \_\_\_\_\_ to \_\_\_\_\_ there isn't necessarily a difference btw the 2 groups.

**Note:** Because the samples are small, we should have first checked that the histograms are roughly normal (they are).

**EXTRA CREDIT** (Max. 5 pts.) Assuming the class to be a random sample of Pitt undergrads, set up a confidence interval for the population mean based on survey data of interest to you. Alternatively, you can set up a confidence interval for the difference between two means. Do **not** feature the variables discussed in class (weights or cigarettes). Survey data is available at [www.pitt.edu/~nancyp/stat-0800/index.html](http://www.pitt.edu/~nancyp/stat-0800/index.html)

**CIVIL UNIONS VS. GAY MARRIAGE PROPOSALS** In a Gallup survey, conducted March 5-7 [2004], we found a surge in public support for gay civil unions, which would give gay and lesbian couples "some of the legal rights of married couples." Last July, Americans opposed gay civil unions by 57% to 40%, but the March poll showed that Americans favored the idea by 54% to 42% -- an increase in support of 14 percentage points. However, in the new survey, unlike the one last July, respondents were first asked if they favored or opposed *marriage* for gay couples, and then respondents were asked if they favored or opposed civil unions for gay couples. Given the sharp increase in support for gay civil unions, we wondered: Could the change in question format have contributed to that increase? Our theory was that in general, many people support gay civil unions as an alternative to gay marriages. If people are asked first about gay civil unions, many might indicate their opposition even though they really oppose gay marriage more generally, and just don't want the interviewer to think that their support for gay

**(continued)** question was asked before or after the civil unions question. The Pew experiment showed that the effect of asking the gay marriage question first, followed by the civil unions question, was to increase the measure of public support for civil unions by eight percentage points. Because the Pew questions and the Gallup questions are not identical, we cannot assume that Gallup would find the exact same effect. Nevertheless, the Pew findings suggest that somewhere about half of the 14-point increase found by Gallup in its current poll may be due simply to question context - adding the gay marriage question before the civil unions question. The other half of the increase is most likely due to a real change in public opinion. The Gallup and Pew polls raise another question: Which measure of public opinion is the "real" one? The answer to this question may be unsatisfactory to most poll watchers: There is no one "real" measure. Both measures provide insights into what the public is thinking. In Massachusetts, where the state legislature is trying to deal with a state supreme court ruling mandating the state to allow gay

**(continued)** civil unions means support for gay marriage. But if respondents are asked first about gay marriage, and are able to indicate their opposition to that idea, then they might be more likely to say they support civil unions -- implicitly as an alternative to gay marriage. As it turns out, the Pew Research Center had already tested this notion in a poll conducted last fall, Oct. 15-19, and it confirmed our suspicions. Half of the respondents in the Pew poll were asked about gay marriage first, followed by a question about civil unions. The other half of respondents were asked the question about civil unions and then about gay marriage. When the civil unions question was asked first, only 37% of respondents said they favored the idea, while 55% were opposed. But among respondents who had already been able to say they opposed gay marriage and were then asked about civil unions, support for civil unions was eight points higher (45%) and opposition (47%) eight points lower. Responses to the gay marriage question, according to the Pew report, were not affected by question order. Opposition remained at nearly 2-to-1, regardless of whether the gay marriage

**(continued)** marriages, support for gay civil unions among state legislators is quite high -- as an alternative to allowing gay marriages. In that trade-off context, the latest Gallup results suggest that a majority of Americans support gay civil unions.

**Views on Civil Unions Depend on the Context**

**Pew Poll**  
Oct. 15-19, 2003  
Civil unions for gays?

	Asked before gay marriage	Asked after gay marriage	%
Favor	37	45	45
Oppose	55	47	47
Don't know	8	8	8
<b>TOTAL</b>	<b>100</b>	<b>100</b>	<b>100</b>

ABOUT THE OCTOBER, 2003 NEWS INTEREST INDEX SURVEY

Results for the survey are based on telephone interviews conducted under the direction of Princeton Survey Research Associates among a nationwide sample of 1,515 adults, 18 years of age or older, during the period October 15-19, 2003. Based on the total sample, one can say with 95% confidence that the error attributable to sampling and other random effects is plus or minus 3 percentage points. For results based on either Form 1 (N=735) or Form 2 (N=780), the sampling error is plus or minus 4 percentage points.

In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls. <http://people-press.org/report/?pageid=762>