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?

(a) 95% of the households in the sample have 2.16 to 2.44 people.
(b) 95% of the households in the city have 2.16 to 2.44 people.
(c) The probability is 95% that mean household size in this city is between 2.16 and 2.44 people.
(d) The probability is 95% that the interval we constructed by this method contains the unknown pop. mean household size.
(e) We're 95% sure that pop. mean is bet. 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 if 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:**

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

= ____________________. We're 95% confident that the unknown population mean household size falls in this interval; our method has a 95% success rate.

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 Wt*

- **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±5.6=(123.4, 134.6) for women, and $168 \pm 2\frac{27}{\sqrt{28}}$ =168±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±5.6=(123.4, 134.6) for women, and $168 \pm 2\frac{27}{\sqrt{28}}$ =168±10.2=(157.8, 178.2) for men.
- **Question:** Is 160 lbs a plausible population mean weight for all women or all men?
- **Response:** For women: ____________________

  For men: ________________________

Sample Size, Width of 95% Confidence Interval

Because sample size appears in the denominator of the confidence interval for population meann

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

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 99% Confidence Interval?*

- **Background**: The mean exam score for the 64 female Stat 800 students is 120, with standard deviation 19.
- **Question:** Is 120±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: ______________________________

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 if not normal, a larger sample is needed.

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

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

Example: *Confidence Interval in a Paired Study*

- **Background**: For a sample of 9 public colleges, we compare the % of African American vs. % of Hispanic students. The % differences had mean 14, sd 12.
- **Question:** What is a 95% confidence interval for the mean of differences (in percentages) for all public colleges?
- **Response:** _______________

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