Lecture 24/Chapter 20
Estimating Proportions with Confidence

- Example: Importance of Margin of Error
- From Probability to Confidence
- Constructing a Confidence Interval
- Examples

Example: *What Can We Infer About Population?*

- **Background**: Gallup poll: "If you could only have one child, would you have a preference for the sex? If so, would you prefer a boy or a girl?"
  Of 321 men with a preference, 70% wanted a boy.
  Of 341 women with a preference, 47% wanted a boy.
- **Questions:** Can we conclude that a majority of *all* men with a preference wanted a boy? And that a minority of *all* women with a preference wanted a boy?
- **Response:**

Probability then Inference, Proportions then Means

Probability theory dictates behavior of sample proportions (categorical variable of interest) and sample means (quantitative variable) in random samples from a population with known values. Now

perform **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

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).

Rule for Sample Proportions *(Review)*

- **Center:** The mean of sample proportions equals the true population proportion.
- **Spread:** The standard deviation of sample proportions is standard error =

$$\sqrt{\frac{\text{population proportion} \times (1 - \text{population proportion})}{\text{sample size}}}$$

- **Shape:** (Central Limit Theorem) The frequency curve of proportions from the various samples is approximately normal.

Empirical Rule *(Review)*

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

Example: *Applied Rule to M&Ms (Review)*

- **Background**: Population proportion of blue M&M's is 1/6=0.17. Students repeatedly take random samples of size 1 Tablespoon (about 75) and record the proportion that are blue.
- **Question:** What does the Empirical Rule tell us?
- **Response:** The probability is
  - 68% that a sample proportion falls within 1×0.043 of 0.17: in [0.127, 0.213]
  - 95% that a sample proportion falls within 2×0.043 of 0.17: in [0.084, 0.256]
  - 99.7% that a sample proportion falls within 3×0.043 of 0.017: in [0.041, 0.299]

Note: This was a **probability** statement: population proportion was known to be 0.17; we stated what sample proportions do.

Example: *An Inference Question about M&Ms*

- **Background**: Population proportion of red M&Ms is unknown. In a random sample, 18/75=0.24 are red.
- **Question:** What can we say about the proportion of all M&Ms that are red?
- **Response:**

Note: We're 95% sure that it falls within 2 *standard errors* of 0.24. Unfortunately, the exact standard error is unknown.

Approximating Standard Error

The standard error of sample proportion is

$$\sqrt{\frac{\text{population proportion} \times (1-\text{population proportion})}{\text{sample size}}}$$

which we approximate with

$$\sqrt{\frac{\text{sample proportion} \times (1-\text{sample proportion})}{\text{sample size}}}$$

because the population proportion is unknown.

Example: *The Inference Question about M&Ms*

- **Background**: Population proportion of red M&Ms is unknown. In a random sample, 18/75=0.24 are red.
- **Question:** What can we say about the proportion of all M&Ms that are red?
- **Response:** The approximate standard error is

  We're 95% confident that the unknown population proportion of reds falls within 2 standard errors of 0.24, in the interval

Note: The "95%" part of our claim goes hand-in-hand with the number 2: for a normal distribution, 95% of the time, the values are within 2 standard deviations of their mean.

95% Confidence Interval for Population Proportion

An approximate 95% confidence interval for population proportion is

$$\text{sample proportion} \pm 2\sqrt{\frac{\text{sample proportion} \times (1 - \text{sample proportion})}{\text{sample size}}}$$

Example: *What Can We Infer About Population?*

- **Background**: Gallup poll: "If you could only have one child, would you have a preference for the sex? If so, would you prefer a boy or a girl?"
  Of 321 men with a preference, 70% wanted a boy.
- **Question:** What is a 95% confidence interval for the population proportion? Can we conclude that a majority of all men with a preference wanted a boy?
- **Response:** The 95% confidence interval is:

The interval suggests a majority of *all* men with a preference want a boy, because

Example: *More Inference for Proportions*

- **Background**: Gallup poll: "If you could only have one child, would you have a preference for the sex? If so, would you prefer a boy or a girl?"
  Of 341 women with a preference, 47% wanted a boy.
- **Question:** What is a 95% confidence interval for the population proportion? Can we conclude that a minority of all women with a preference wanted a boy?
- **Response:** The 95% confidence interval is:

  The interval contains so the population proportion could be in a majority or a minority.

Example: *Confidence Interval for Smaller Sample*

- **Background**: Gallup poll: "If you could only have one child, would you have a preference for the sex? If so, would you prefer a boy or a girl?" Suppose 70% of only 10 men with a preference wanted a boy.
- **Question:** What is a 95% confidence interval for the population proportion? Can we conclude that a majority of all men with a preference wanted a boy?
- **Response:** The 95% confidence interval is:

Now it's

Example: *Confidence Interval for Larger Sample*

- **Background**: Gallup poll: "If you could only have one child, would you have a preference for the sex? If so, would you prefer a boy or a girl?" Now suppose 47% of 2500 women with a preference wanted a boy.
- **Question:** What is a 95% confidence interval for the population proportion? Can we conclude that a minority of all women with a preference wanted a boy?
- **Response:** The 95% confidence interval is:

Now it looks like

Sample Size, Width of 95% Confidence Interval

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

$$\text{sample proportion} \pm 2\sqrt{\frac{\text{sample proportion} \times (1-\text{sample proportion})}{\text{sample size}}}$$

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 proportion is

$$\text{sample proportion} \pm 1.645\sqrt{\frac{\text{sample proportion}\times(1-\text{sample proportion})}{\text{sample size}}}$$

An approximate 99% confidence interval for population proportion is

$$\text{sample proportion} \pm 2.576\sqrt{\frac{\text{sample proportion}\times(1-\text{sample proportion})}{\text{sample size}}}$$

Example: *A 99% Confidence Interval*

- **Background**: According to "Helping Stroke Victims", German researchers who took steps to reduce the temps of 25 people who had suffered severe strokes found 14 survived instead of the expected 5.
- **Question:** Based on the treatment survival rate 14/25=0.56, what is a 99% confidence interval for the proportion of all such patients who would survive with this treatment? Does the interval contain 5/25=0.20?
- **Response:**

Example: *A 90% Confidence Interval?*

- **Background**: 100 people in Lafayette, Colorado volunteered to eat a good-sized bowl of oatmeal for 30 days to see if simple lifestyle changes---like eating oatmeal---could help reduce cholesterol. After 30 days, 98 lowered their cholesterol.
- **Question:** What is a 90% confidence interval for the proportion of all people whose cholesterol would be lowered in 30 days by eating oatmeal?
- **Response:**

Conditions for Rule of Sample Proportions

- Randomness [affects center]
  - Can't be biased for or against certain values
- Independence [affects spread]
  - If sampling without replacement, sample should be less than 1/10 population size
- Large enough sample size [affects shape]
  - Should sample enough to expect at least 5 each in and out of the category of interest.

Example: *Preview of a Hypothesis Test Question*

- **Background**: Population proportion of red M&Ms is unknown. In a random sample, 18/75=0.24 are red. The approximate standard error is $\sqrt{\frac{0.24(1-0.24)}{75}} = 0.05$ so we're 95% sure that the unknown population proportion of reds falls within 2 standard errors of 0.24, in the interval 0.24±2(0.05)=(0.14, 0.34).
- **Question:** Can we believe that the population proportion of reds is 0.30?
- **Response:**

Example: *Approximate Margin of Error*

- **Background**: Margins of error discussed:
  - 0.10 for 75 M&Ms, sample proportion 0.24
  - 0.05 for 341 men, sample proportion 0.70 wanted a boy
  - 0.05 for 371 women, sample proportion 0.47 wanted a boy
- **Question:** What are the approximate error margins, using 1 divided by square root of sample size; how accurate are they?
- **Response:**
  - Close to 0.10?
  - Close to 0.05?
  - Close to 0.05?________________

EXTRA CREDIT (Max. 5 pts.) Assuming the class to be a random sample of Pitt undergrads, set up a proportion confidence interval based on survey data of interest to you. Survey data is available at www.pitt.edu/~nancyp/stat-0800/index.html