

Lecture 26/Chapter 22

Hypothesis Tests for Proportions

- Null and Alternative Hypotheses
- Standardizing Sample Proportion
- P-value, Conclusions
- Examples

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).

Example: Revisiting the Wording of Questions

- **Background:** A Pew poll asked if people supported *civil unions* for gays; some were asked **before** a question about whether they supported *marriage* for gays; others **after**. Of 735 people asked **before** the marriage question, **55%** opposed civil unions. Of 780 asked **after** the marriage question, **47%** opposed.
- **Question:** What explains the difference?
- **Response:**

Example: Testing a Hypothesis about a Majority

- **Background:** In a Pew poll of 735 people, 0.55 opposed civil unions for gays.
- **Question:** Are we convinced that a majority (more than 0.5) of the population oppose civil unions for gays?
- **Response:** It depends; if the population proportion opposed were only ____, how improbable would it be for at least ____ in a random sample of 735 people to be opposed?

Example: Testing a Hypothesis about a Minority

- **Background:** In a slightly different Pew poll of 780 people, 0.47 opposed civil unions for gays.
 - **Question:** Are we convinced that a minority (less than 0.5) of the population oppose civil unions for gays?
 - **Response:** It depends; if the population proportion opposed were as high as ____, how improbable would it be for no more than ____ in a random sample of 780 people to be opposed?
- Note:** In both examples, we test a **hypothesis** about the larger population, and our conclusion hinges on the *probability* of observed behavior occurring in a random sample. This probability is called the **P-value**.

Testing Hypotheses About Pop. Value

1. Formulate hypotheses.
2. Summarize/standardize data.
3. Determine the P-value.
4. Make a decision about the unknown population value (proportion or mean).

Null and Alternative Hypotheses

- For a test about a single proportion,
- **Null hypothesis:** claim that the population proportion equals a proposed value.
 - **Alternative hypothesis:** claim that the population proportion is greater, less, or not equal to a proposed value.
- An alternative formulated with \neq is **two-sided**; with $>$ or $<$ is **one-sided**.

Testing Hypotheses About Pop. Value

1. Formulate hypotheses.
2. Summarize/standardize data.
3. Determine the P-value.
4. Make a decision about the unknown population value (proportion or mean).

Standardizing Normal Values (Review)

Put a value of a normal distribution into perspective by **standardizing** to its z-score:

$$z = \frac{\text{observed value} - \text{mean}}{\text{standard deviation}}$$

The observed value that we need to standardize in this context is the **sample proportion**. We've established Rules for its **mean** and **standard deviation**, and for when the **shape** is approximately normal, so that a probability (the **P-value**) can be assessed with the normal table.

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**.

Standardized Sample Proportion

□ To test a hypothesis about an unknown population proportion, find sample proportion and standardize to

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

□ z is called the **test statistic**.

Note that “sample proportion” is what we’ve observed, “population proportion” is the value proposed in the null hypothesis.

Conditions for Rule of Sample Proportions

1. Randomness [affects **center**]
 - Can't be biased for or against certain values
 2. Independence [affects **spread**]
 - If sampling without replacement, sample should be less than 1/10 population size
 3. Large enough sample size [affects **shape**]
 - Should sample enough to expect at least 5 each in and out of the category of interest.
- If 1st two conditions don't hold, the mean and sd in z are wrong; if 3rd doesn't hold, P -value is wrong.

Testing Hypotheses About Pop. Value

1. Formulate hypotheses.
2. Summarize/standardize data.
3. Determine the P -value.
4. Make a decision about the unknown population value (proportion or mean).

P -value in Hypothesis Test about Proportion

The P -value is the probability, assuming the null hypothesis is true, of a sample proportion at least as low/high/different as the one we observed. In particular, it depends on whether the alternative hypothesis is formulated with a less than, greater than, or not-equal sign.

Testing Hypotheses About Pop. Value

1. Formulate hypotheses.
2. Summarize/standardize data.
3. Determine the P -value.
4. Make a decision about the unknown population value (proportion or mean).

Making a Decision Based on a P -value

If the P -value in our hypothesis test is **small**, our sample proportion is **probably low/high/different**, assuming the null hypothesis to be true. We conclude it is **not** true: we reject the null hypothesis and believe the alternative.

If the P -value is **not small**, our sample proportion is **believable**, assuming the null hypothesis to be true. We are willing to believe the null hypothesis.

P -value **small** \longrightarrow **reject** null hypothesis

P -value **not small** \longrightarrow **don't reject** null hypothesis

Hypothesis Test for Proportions: Details

1. null hypothesis: pop proportion = proposed value
alt hyp: pop proportion $<$ or $>$ or \neq proposed value
2. Find sample proportion and standardize to z .
3. Find the P -value = probability of sample proportion as low/high/different as the one observed; same as probability of z this far below/above/away from 0.
4. If the P -value is small, conclude alternative is true. In this case, we say the data are **statistically significant** (too extreme to attribute to chance). Otherwise, continue to believe the null hypothesis.

Example: Testing a Hypothesis about a Majority

- **Background:** In a Pew poll of 735 people, 0.55 opposed civil unions for gays.
- **Question:** Are we convinced that a majority (more than 0.5) of the population oppose civil unions for gays?
- **Response:**
 1. Null: pop proportion _____ Alt: pop proportion _____
 2. Sample proportion = _____, $z =$ _____
 3. P -value = prob of z this far above 0: _____
 4. Because the P -value is small, we reject null hypothesis. Conclude _____

Example: Testing a Hypothesis about a Minority

- **Background:** In a Pew poll of 780 people, 0.47 opposed civil unions for gays.
- **Question:** Are we convinced that a minority (less than 0.5) of the population oppose civil unions for gays?
- **Response:**
 1. Null: pop proportion _____ Alt: pop proportion _____
 2. Sample proportion = _____, $z =$ _____
 3. P -value = prob of z this far below 0: approximately _____
 4. Because the P -value is _____

Example: Testing a Hypothesis about a M&Ms

- **Background:** Population proportion of red M&Ms is unknown. In a random sample, $15/75=0.20$ are red.
- **Question:** Are we willing to believe that $1/6 = 0.17$ of all M&Ms are red?
- **Response:**
 1. Null: pop proportion _____ Alt: pop proportion _____
 2. Sample proportion = _____, $z =$ _____
 3. P -value=prob of z this far away from 0 (either direction) _____
 4. Because the P -value isn't too small, _____