29

The Five-Step Model: Proportions

- When analyzing variables that are not measured at the interval-ratio level (and therefore a mean is inappropriate), we can test a hypothesis on a one sample proportion
  - For example, the flu vaccination rate mentioned at the beginning of this lecture
- The five step model remains primarily the same, with the following changes:
  - The assumptions are: random sampling, nominal level of measurement, and normal sampling distribution
  - The formula for Z(obtained) is:

$$Z(obtained) = \frac{P_s - P_u}{\sqrt{P_u(1 - P_u)/N}}$$

30

The Five-Step Model: Proportions

- A random sample of 122 households in a low-income neighborhood revealed that 53 (P_s = 0.43 = 53/122) of the households were headed by women
- In the city as a whole, the proportion of women-headed households is 0.39 (P_u)
- Are households in lower-income neighborhoods significantly different from the city as a whole?
- Conduct a 90% hypothesis test (alpha = 0.10)

31

The Five-Step Model: Proportions

Step 1: Make Assumptions and Meet Test Requirements

- Random sampling
  - Hypothesis testing assumes samples were selected according to EPSEM
  - The sample of 122 was randomly selected from all lower-income neighborhoods
- Level of measurement is nominal
  - Either woman-head or not women-headed
- Sampling distribution is normal in shape
  - This is a "large" sample ($N \geq 100$)

32

The Five-Step Model: Proportions

Step 2: State the Null and Alternative Hypothesis

- H_0: $P_u = 0.39$; There is no difference between the lower-income neighborhoods and the city as a whole in terms of the proportion of women-headed households
  - The sample of 122 comes from a population where 39% of households are headed by women; The difference between 0.43 and 0.39 is trivial and caused by random chance
- H_1: $P_u \neq 0.39$; There is a difference between the lower-income neighborhoods and the city as a whole in terms of the proportion of women-headed households (Two-tailed test)
  - The sample of 122 comes from a population where the percent of women-headed households is not 39; The difference between 0.43 and 0.39 reflects an actual difference between lower-income neighborhoods and all neighborhoods

33

The Five-Step Model: Proportions

Step 3: Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution= Z distribution (N is large)
- Alpha (α) = 0.10 (two-tailed)
- Critical region begins at ±1.65
  - This is the critical Z score associated with a two-tailed test and alpha equal to 0.10
  - If the obtained Z score falls in the critical region, reject H_0

34

The Five-Step Model: Proportions

Step 4: Compute the Test Statistic

$$Z(obtained) = \frac{P_s - P_u}{\sqrt{P_u(1-P_u)/N}} = \frac{0.43 - 0.39}{\sqrt{(0.39)(1-0.39)/122}}$$
$$= +0.91$$

Z(obtained) = +0.91

35

The Five-Step Model: Proportions

Step 5: Make a Decision and Interpret Results

- The obtained Z score did not fall in the critical region so we **fail to reject** the H_0
  - If the H_0 were true, a sample outcome of 0.43 would be likely
  - Therefore, the H_0 is not false and cannot be rejected
- Report result:
  - With α equals 0.10, the obtained Z score of 0.91, we **fail to reject** the null hypothesis. The data suggest that the population of women-headed households in lower-income neighborhoods is not significantly different from the city as a whole.

36

Reminders

- When report your findings, you always want to include the **level of significance** (α value)**,** the **obtained test statistic** ($Z_{obtained}$) and the **conclusion** in the statement.
- Never use the word "prove" in your conclusion. Instead, state that the "findings indicate …" or the "data suggest …"
- We do not claim H_0 is correct or incorrect, we only reject or fail to reject H_0. When dealing with samples, we have to be prepared for the possibility for error.

37

Type I and Type II Errors

- Type I, or alpha error (α):
  - Rejecting a true null hypothesis
- Type II, or beta error (β) :
  - Failing to reject a false null hypothesis
- Relationships between decision making and errors

	H_0 actually is True	H_0 actually is False
Reject H_0	**Type I error (α)**	**Correct**
Fail to Reject H_0	**Correct**	**Type II error (β)**

38

Type I and Type II Errors

- As you increase the level of significance, say from .10 to .01, the smaller the critical area and the harder it becomes to reject the null hypothesis.
- The harder it becomes to reject the null, the less likely the probability of Type I error.
- However, as the critical region becomes smaller, the chances of committing a Type II error increase.
- So, the selection of a level of significance must be conceived as a balance between the types of error.
- Normally, in the social sciences, we want to minimize Type I error, and the conventional significance level is set as 0.05.
- However, the researchers can always change the significance level (such as 0.04 or 0.03), if they think it is reasonable for the specific context.

39

After this lecture:

You should learn the following key concepts:

- The logic of hypothesis testing
  - How to state the null hypothesis and alternative hypothesis
  - How to test hypothesis using the five step process
- Distinguish different formulas that is suitable for different conditions: such as when the σ is known or unknown; when calculating sample mean or sample proportion
- Distinguish the critical values for one-tailed and two-tailed tests.
- How to determine if we reject or fail to reject the null hypothesis; how to report findings formally and appropriately
- The types of error we may commit when testing hypothesis