POLS 5377 Scope & Method of Political Science

Week 11 Hypothesis Testing

The One-Sample Test

Healey. (2016) *Statistics: A Tool for Social Research,* Chapter 8

2

Key Questions:

- What is the logic behind hypothesis testing?
- What are the steps used to test hypotheses?
- How to interpret and report the results of hypothesis testing?
- What are the different formulas used for hypothesis testing?

3

Outline

- Logic of Hypothesis Testing
- Five-Step Model of Testing Hypothesis
- One-Tailed and Two-Tailed Tests
- Hypothesis Testing – Proportion
- Interpreting Test Results

4

Logic of Hypothesis Testing

Example:

- A researcher is studying the effectiveness of an influenza awareness campaign in Walker County that is designed to increase flu vaccination rate in the community.
- Since it is not feasible to survey each citizen, the research takes a random sample of 250 citizens who received the campaign messages.
- The researcher notes that citizens in the sample appear to have higher rate of receiving the flu vaccination.

Population **All citizens in Texas**	**Sample** **250 citizens drawn from those received the campaign messages**
$\mu = 43\%$ flu vaccination rate $\sigma = 2.2$	$\bar{X} = 48\%$ flu vaccination rate N = 250

5

Logic of Hypothesis Testing

- Question: are citizens who receive the influenza awareness campaign messages more likely to receive flu vaccination than are the citizens of Texas as a whole?
- There are two ways to interpret the difference between the 43% and the 48%.
  - First, there is **no difference** between citizens who received the campaign messages and the citizens in the state in general in terms of flu vaccination rate. The difference seen is trivial and due to the effects of random chance.
  - Second, the flue vaccination rate difference we see in our sample is **real**. These differences are statistically significant. The difference is very unlikely to have occurred by random chance. The awareness campaign works.
- How do we decide which interpretation is true? We set up a decision making process that enables us to choose the interpretation that is less likely to be incorrect.

6

Logic of Hypothesis Testing

- A **Test of Significance**, or **Hypothesis Testing** is the process we use for make this decision.
- To test the two possible interpretations, we always assume the first one is correct – that there is **NO difference** (statistically significant difference) between the sample and the population.
- This is known as the ***null hypothesis***.
- When we begin to test our hypothesis, we actually test the null hypothesis.
- The null hypothesis always specifies that any relationship or difference found is due solely to chance.
- In other words, that there is no difference between the two groups.

7

Logic of Hypothesis Testing

Two hypotheses:

- Null Hypothesis (H_0)
  - "The difference is caused by random chance" ; H_0 always states there is "no significant difference"
  - H_0: There is no difference between citizens in our sample and all citizens in the State as a whole in terms of flu vaccination rate.
- Alternative hypothesis (H_1)
  - "The difference is real"; H_1 always contradicts H_0
  - H_1 : There is a difference between citizens in our sample and all citizens in the State as a whole in terms of flu vaccination rate.
- One (and only one) of these explanations *must* be true, but which one?

8

Logic of Hypothesis Testing

- Assume the H_0 is true
  - What is the probability of getting the sample mean (48%) if the H_0 is true and all citizens received the campaign messages really have a mean of 43% vaccination rate?
  - If the probability is less than 0.05, reject the null hypothesis – which means reject the idea of there is no difference between the two groups
- Use the 0.05 value as a guideline to identify if the difference is rare or not.
- Use the normal curve table to determine the probability of getting the observed difference.
- If the observed value falls into the 0.05 area ($z > 1.96$; $z< -1.96$), the difference is large enough, that only 0.05 chance the H_0 is true.
- Reject H_0

9

Logic of Hypothesis Testing

Example:

A researcher is studying the effectiveness of a rehabilitation center that treats alcoholics. The researcher sampled 127 people who were treated in the center.

Community (all alcoholics)	Sample drawn from all treated alcoholics
$\mu = 7.2$ days of absent per year $\sigma = 1.43$	$\bar{X} = 6.8$ days of absent per year N = 127

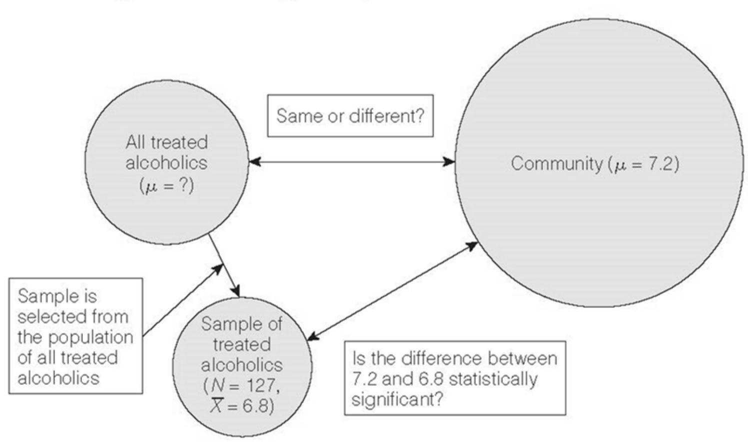
Question: Does the population of all treated alcoholics have different number of absent day than the community as a whole?

- What is the cause of the difference between 6.8 and 7.2?
  - Real difference or Random chance?

10

Logic of Hypothesis Testing

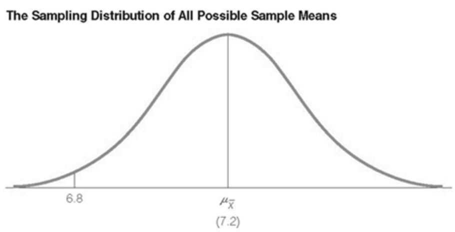
A Test of Hypothesis for Single-Sample Means

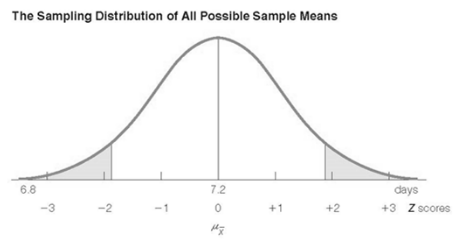


11

Logic of Hypothesis Testing

- H_0: There is no difference between treated alcoholics in our sample and the whole community in terms of number of absent day.
- H_1 : There is a difference between treated alcoholics in our sample and the whole community in terms of number of absent day.
- If H_0 is true, then the mean of the sampling distribution should be same as the population mean, $\mu = \mu_{\bar{X}} = 7.2$
- What is the probability of getting a sample that its mean ($\bar{X}$) equals 6.8?





12

Logic of Hypothesis Testing

- Calculate the z score with the formula:

Sample mean

Population mean

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

Population Standard Deviation

Standard deviation of sampling distribution

$$Z = \frac{6.8 - 7.2}{1.4/\sqrt{127}} = \frac{-0.4}{0.127} = -3.15$$

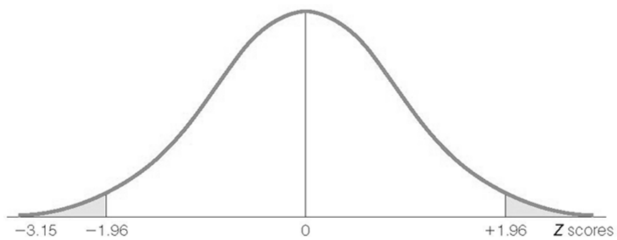
$$Reminder:\ Z = \frac{X_i - \bar{X}}{\sigma}$$

13

Logic of Hypothesis Testing

- With the sampling error 0.05 ($Z=\pm1.96$), the result Z = -3.15 shows the sample mean falls in the shaded area.
- This means, if H_0 is true, the probability of the sample mean to be 6.8 is less than 0.05, which is very low (H_0 is unlikely to be true).
- Therefore, we **reject** H_0 .

The Sampling Distribution of Sample Means with the Sample Outcome ($\bar{X}$ = 6.8) Noted in *Z* Scores



14

Five-Step Model

- Make assumptions and meet test requirements
- State the null hypothesis (H_0) and alternative hypothesis (H_1)
- Select the sampling distribution and establish the critical region
- Compute the test statistic
- Make a decision and interpret the results of the test

15

Five-Step Model - Example

- Citizens in a city complain about the poor design of the transportation system which has caused the long commute-to-work time.
- The nationwide average commute-to-work time is 25.3 minutes (μ).
- A random sample of 182 (N) residents in the city has a mean commute-to-work time 28.0 minutes, with a standard deviation (s) of 9.4 minutes.
- Is there a difference between the city and the nationwide average in terms of commute time?

16

Five-Step Model - Example

Step 1: Make Assumptions and Meet Test Requirements

- Random sampling
  - Hypothesis testing assumes samples were selected according to EPSEM
  - The sample of 182 was randomly selected from all residents of the city
- Level of measurement is interval-ratio
  - Commute time measured in minutes is an interval-ratio level variable, so the mean is an appropriate statistic
- Sampling distribution is normal in shape
  - This is a “large” sample ($N \geq 100$)

17

Five-Step Model - Example

Step 2: State the Null and Alternative hypothesis

- H_0: There is no difference between the city and the nationwide average in terms of commute-to-work time.
  - The sample of 182 comes from a population (the City citizens) that has an average commute-to-work time of 25.3 minutes
  - The difference between 25.3 and 28 is trivial and caused by random chance
- H_1: There is a difference between the city and the nationwide average in terms of commute-to-work time.
  - The sample of 182 comes from a population that does not have a commute-to-work time of 25.3 minutes
  - The difference between 25.3 and 28 reflects an actual difference of commute time between the city and nationwide.

18

Five-Step Model - Example

Step 3: Select Sampling Distribution and Establish the Critical Region

- Sampling Distribution= Z
  - Alpha (α) = 0.05
  - Alpha is the indicator of “rare” events
  - Any difference with a probability less than α is rare and will cause us to reject the H_0
- Critical region begins at ± 1.96
  - This is the critical Z score associated with a two-tailed test and α equal to 0.05
  - If the Z score of the sample mean falls in the critical region, reject H_0

19

Five-Step Model - Example

Step 4: Compute the Test Statistic

- Sample size: 182 (>100, large sample)
- Population standard deviation (σ) : unknown
- Sample standard deviation (s): 9.4
- Formula:

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{N-1}}$$

$$Z(obtained) = \frac{28.0 - 25.3}{9.4/\sqrt{182-1}} = 3.86$$

Z (obtained): 3.86

20

Five-Step Model - Example

Step 5: Make a Decision and Interpret Results

- The obtained Z score fell in the critical region so we reject the H_0
  - If the H_0 were true, a sample outcome of 28.0 would be unlikely
  - Therefore, the H_0 is false and must be rejected
- Report result:
  - With α equals 0.05, the obtained Z score of 3.86, the null hypothesis is rejected. The data suggests that there is a difference between the city and the nationwide in terms of the commute-to-work time. The decision to reject the null hypothesis has a 0.05 probability of being wrong.

21

Hypothesis Testing - Summary

- It is important to know that, there are different types of test of significance, and to know when to use what:
  - When comparing samples to the population, use
    - one-sample z test (large sample; N≥100) or
    - one-sample *t*-test (small sample; N<100)
  - When comparing two samples, use a two-sample *t*-test.
- Two formulas of calculating z scores:
  - When σ is known: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$
  - When σ is unknown: $Z = \frac{\bar{X}-\mu}{S/\sqrt{N-1}}$

22

Hypothesis Testing - Summary

- Each test produces a score, either a z score or a t score depending upon the test. This is called the *test statistic*.
- What the test statistic tells us is whether the difference between the means we are comparing is statistically significant or not.
  - If the score is <u>significant</u>, we **reject** the <u>null hypothesis</u> – there is a difference between the two data sets.
  - If the score is <u>not significant</u>, we **do not reject** the <u>null hypothesis</u> – there is no significant difference between the two data sets.

23

Choices in Hypothesis Testing

- Although you can follow the five-step model to test hypotheses, there are two choices that would affect your testing process:
  - One-tailed or two-tailed test
  - Alpha (α) level

24

One-Tailed or Two-Tailed Test

- Two-tailed: States that population mean is “not equal” to value stated in null hypothesis
  - H_1: $\mu \neq 7.2$ (a difference in the number of absent day)
  - H_1: $\mu \neq 25.3$ (a difference in the commute-to-work time)
- One-tailed: Differences in a specific direction
  - H_1: $\mu < 7.2$ (treated people have lower number of absent day)
  - H_1: $\mu > 25.3$ (the citizens in the city spend more time on commuting to work)
- The choice between one and two-tailed test is based on the researcher’s expectation about the two groups.
- If the direction is not clear, we use the two-tailed test.

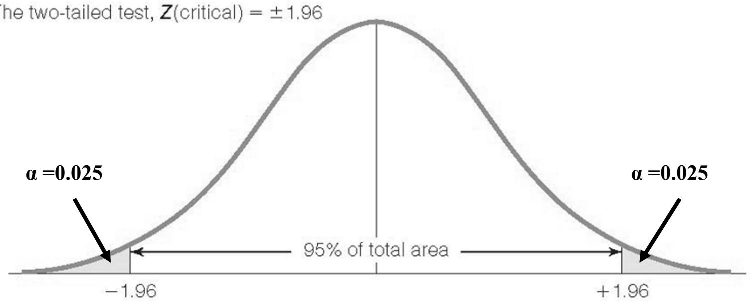
25

One-Tailed or Two-Tailed Test

- Your choices of one-tailed or two-tailed test would affect the Z score of the given α value

Establishing the Critical Region, One-Tailed Tests Versus Two-Tailed Tests (alpha = 0.05)

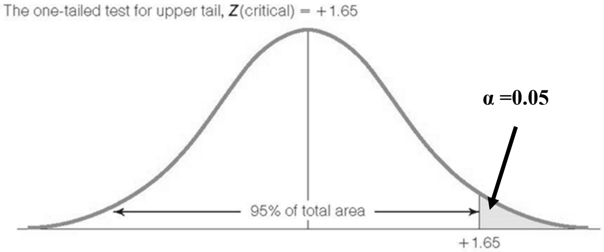
A. The two-tailed test, Z(critical) = ±1.96



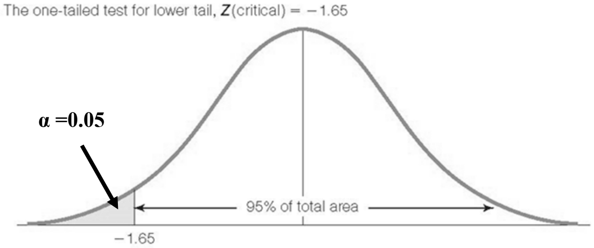
26

One-Tailed or Two-Tailed Test

B. The one-tailed test for upper tail, Z(critical) = +1.65



C. The one-tailed test for lower tail, Z(critical) = −1.65



27

One-Tailed or Two-Tailed Test

- Your choices of one-tailed or two-tailed test would affect the Z score of the given α value

One- vs. Two-Tailed Tests, $\alpha = 0.05$

If the Research Hypothesis (H_1) Uses	The Test Is	Concern Is on	Z(critical) Is
$\neq$	Two-tailed	Both tails	±1.96
$>$	One-tailed	Upper tail	+1.65
$<$	One-tailed	Lower tail	−1.65

Finding Critical Z Scores for One- and Two-Tailed Tests

		One-Tailed Value	
Alpha	Two-Tailed Value	*Upper Tail*	*Lower Tail*
0.10	±1.65	+1.29	−1.29
0.05	±1.96	+1.65	−1.65
0.01	±2.58	+2.33	−2.33
0.001	±3.32	+3.10	−3.10
0.0001	±3.90	+3.70	−3.70

28

Selecting an Alpha Level

- By assigning an alpha level, one defines an “unlikely” sample outcome
- Alpha level is the probability that the decision to reject the null hypothesis is incorrect

The Relationship Between Alpha and Z(Critical) for a Two-Tailed Test

If Alpha =	The Two-Tailed Critical Region Will Begin at Z(Critical) =
0.100	±1.65
0.050	±1.96
0.010	±2.58
0.001	±3.32

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