Lesson 11

Confidence Intervals and Hypothesis Tests of Proportions from Two groups

Lesson 11 overview

- Often researchers are interested in estimating the difference between two independent groups.
- When the variable of interest is continuous, appropriate inference methods are
  - Confidence Intervals of the difference between two means
  - Hypothesis tests (t-tests) of equal means
  - Nonparametric when sample size is small and normality assumption is not met (not covered in PH6414)
- When the variable of interest is categorical (binary), appropriate inference methods are
  - Confidence Intervals of the difference in proportions (Lesson 11)
  - Hypothesis tests (z-tests) of equal proportions (Lesson 11)
  - Chi-square tests of Independence (Lesson 12)

Lesson 11 Outline

- Two sample Z-test of proportions to test the null hypothesis of equal proportions
- Confidence Interval for the difference of two proportions
- Both inference methods require that the normal approximation to the binomial distribution are met for both groups
  - $n^*\pi > 10$ and $n^*(1-\pi) \geq 10$ (for CI)
  - $n^*\pi > 5$ and $n^*(1-\pi) \geq 5$ (for Hypothesis test)

Two sample z-test of proportions: Notation

- We have 2 distinct populations of some binary variable.

- The population proportions (π) are the true proportions in each population, the sample proportions (p) are estimates of the population proportions.

Population	Population proportion	Sample size	Number in sample with success	Sample proportion
1	π_1	n_1	x_1	$p_1 = \frac{x_1}{n_1}$
2	π_2	n_2	x_2	$p_2 = \frac{x_2}{n_2}$

Steps in Hypothesis Testing Comparing Two Proportions

1. State the null hypothesis H_0 and the alternative hypothesis H_a.
2. Calculate the value of the test statistic on which the test will be based.
3. Find the p-value for the observed data.
4. State a conclusion.

Hypothesis Testing Steps: Comparing two Proportions

- Step 1: State your hypotheses
  - H_0: $\pi_1-\pi_2= \pi_0$
  - Ha: $\pi_1-\pi_2\neq \pi_0$ or Ha: $\pi_1-\pi_2 < \pi_0$ or Ha: $\pi_1-\pi_2 > \pi_0$

  .

Hypothesis Testing Steps: Comparing two Proportions

Step 2 : Calculate your test statistic (z statistic is appropriate provided you have a SRS that meet this condition

$$n_1 p_1 \geq 5 \text{ and } n_1(1-p_1) \geq 5 \text{ and } n_2 p_2 \geq 5 \text{ and } n_2(1-p_2) \geq 5$$

We are testing under the null hypothesis that $\pi_1=\pi_2$, so we create

$$p = \frac{X_1 + X_2}{n_1 + n_2}$$

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Hypothesis Testing Steps: Comparing Two Proportions

- Step 3: Calculate the p-value

  One or two-sided.

- Step 4: Make your conclusion.

  If p-value is small, you have evidence against the null.

  If p-value is large, you do not have evidence against the null.

Two-sample z-test of difference of two proportions example

- A study was designed to investigate whether low birth weight (LBW: wt < 2500 gm) have the same rate of lower respiratory tract infection (LRTI) in the first 2 years as normal birth weight infants (NBW).

- A random sample of 1000 infants from several urban clinics were enrolled in the study.

- Infants were identified as LBW or NBW and followed for 2 years

- Test the hypothesis that there is ***no difference*** in the proportion of infants with any LRTI in the first 2 years between LBW and NBW infants.

Step 1: State the Hypotheses

1. State the Hypotheses
   - Let π_1 represent the population proportion of LRTI among LBW infants and let π_2 represent the population proportion of LRTI among NBW infants

     - H_0:
     - Ha:

Alpha and Critical Values

Use α = 0.05 as the significance level

The critical values are from the standard normal distribution:

=NORMSINV(0.025) and = NORMSINV(1-0.025)

Collect the data

1000 infants were enrolled in the study

- 64 of the infants were LBW, 936 were NBW
- After 2 years of follow-up -
  - 15 of the LBW infants had at least one LRTI
  - 112 of the NBW infants had at least one LRTI

Group	Total n	N with LRTI	Proportion with LRTI
LBW			
NBW			

Two-sample z-test of proportions

Group	Total n	N with LRTI	Proportion with LRTI
LBW	64	15	0.234
NBW	936	112	0.120

- Check that the normal approximation is valid for both groups:
- Calculate the overall proportion (p) of LRTI:

$$p = \frac{x_1 + x_2}{n_1 + n_2} =$$

Step 2: Calculate the z-statistic

Calculate the z-statistic: the difference between the sample proportions divided by the SE of the difference. Use the overall p to calculate the SE (diff).

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}} =$$

- The z-statistic for this test =

Step 3: P-value

By the p-value method:

- p-value =

Conclusion:

Two-sample z-test of proportions: Summary statement

In this study 23% of low birth weight infants had at least one LRTI in the first two years of life and 12% of normal birth weight infants had at least one LRTI in the first two years of life. A significantly different proportion of low birth weight infants had at least one LRTI by age 2 than normal birth weight infants ($p = 0.008$).

Two-sample z-test of proportions in Excel

- In Excel, work through the steps to calculate the z-statistic for the two-sample z-test of proportions
- There is no Excel function or test in the Data Analysis Tool for a z-test of proportions.

CI for the Difference in Population Proportions

- Confidence interval for the difference in population proportions:

$$(p_1 - p_2) \pm Z * SE(p_1 - p_2)$$

where :

$$SE(p_1 - p_2) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$n_1 p_1 \geq 10 \text{ and } n_1(1-p_1) \geq 10 \text{ and } n_2 p_2 \geq 10 \text{ and } n_2(1-p_2) \geq 10$$

SE (diff) and coefficient for CI of difference between proportions

- Your book is *incorrect* in saying the SE for the confidence interval is:

$$SE(p_1 - p_2) = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

- To do this you must assume ($\pi_1 = \pi_2$).

CI for the Difference between 2 Population Proportions: MI Example

- Binge drinking on college campuses is a public health concern.

- A survey of 17, 096 students revealed the following results.

Summarize the data

Group	N	X	p
Men	**7180**	**1630**	
Women	**9916**	**1684**	

Check that the normal approximation to the binomial distribution is valid for both groups:

CI for the Difference

- Construct the 95% CI for the difference

$$(p_1 - p_2) \pm Z^* \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$(0.227 - 0.170) \pm 1.96 \sqrt{\frac{0.227\ \ (0.773)}{7180} + \frac{0.170(0.830)}{9916}}$$

$$0.057 \pm 0.012$$

$$(0.045, 0.069)$$

Interpretation of 95% CI for the difference of proportions

The data suggest that men are more likely to be frequent binge drinkers with an estimated risk difference of 5.7% and 95% confidence interval from 4.5% to 6.9%.

Confidence Interval and z-test

- Typically, if the 95% confidence interval does not contain 0, the two-tailed z-test will result in the decision to reject the null hypothesis of equal proportions.
- However; the standard error is different for the confidence interval and the hypothesis test, just as with the one sample proportion.

Determining which test to use

Death certificate accuracy

- **Across two different hospitals, the results of 575 autopsies were compared to the cause of death listed on the death certificates.**
- **Based on this comparison, the death certificates were classified as accurate or inaccurate.**
- **One hospital was a community based hospital, the other a university hospital.**
- **Is the University hospital more accurate with the cause of death listed on the death certificate than the community based hospital?**

Death certificate accuracy – The data

	Accurate Death Certificate	Inaccurate Death Certificate
Community Hospital	157	72
University Hospital	268	78

WHICH TEST BEST ANSWERS THE SCIENTIFIC QUESTION:

1. Hypothesis test for one proportion (Z test)
2. Hypothesis test for one mean (t test).
3. Hypothesis test for two independent proportions (Z test)
4. Hypothesis test for the difference of means (t test).
5. Hypothesis test for the difference of means (t test).

MI in the Navajo population

- **The population of interest is Navajos residing in the United States.**
- **Out of a random sample of 144 Navajo people 15 had experienced acute myocardial infarction (MI).**
- **Researchers wanted to know if the proportion of Navajos who experience MI is greater than the national average for all Americans (π_0).**

WHICH TEST IS MOST APPROPRIATE FOR THIS STUDY?

1. Hypothesis test for one proportion (Z test)
2. Hypothesis test for one mean (t test).
3. Hypothesis test for two independent proportions (Z test)
4. Hypothesis test for the difference of means (t test).
5. Hypothesis test for the difference of means (t test).

Weight Gain and Cognitive Behavior Therapy

Everitt (1994) compared several different therapies as treatments for anorexia. One condition was Cognitive Behavior Therapy, and he collected data on weights before and after therapy.

29 people with anorexia were randomly assigned to cognitive behavior therapy plus standard of care or to standard of care alone.

Everitt was interested in testing the experimental hypothesis that cognitive behavior therapy would lead to weight gain.

WHICH TEST BEST ANSWERS THE SCIENTIFIC QUESTION?

1. Hypothesis test for one proportion (Z test)
2. Hypothesis test for one mean (t test).
3. Hypothesis test for two independent proportions (Z test)
4. Hypothesis test for the difference of means (t test).
5. Hypothesis test for the difference of means (t test).

CHANGES IN SMOKING STATUS

- **Smoking status was examined over a two year period.**
- **In 2007, a sample of 212 adults, ages 18-22 years, was asked to identify themselves as smokers or nonsmokers.**
- **In 2009, the *same individuals* were again asked whether they were currently smokers or nonsmokers.**
- **Researchers want to know if an equal number of participants switched from being smokers to nonsmokers and from nonsmokers to smokers.**

Changes in smoking status– The data

	2009	
2007	Smoker	Not Smoker
Smoker	62	10
Not Smoker	8	132

WHICH TEST BEST ANSWERS THE SCIENTIFIC QUESTION: DID AN EQUAL NUMBER OF PARTICIPANTS SWITCHED FROM BEING SMOKERS TO NONSMOKERS AND FROM NONSMOKERS TO SMOKERS?

1. Hypothesis test for one proportion (Z test)
2. Hypothesis test for one mean (t test).
3. Hypothesis test for two independent proportions (Z test)
4. Hypothesis test for the difference of means (t test).
5. Hypothesis test for the difference of means (t test).
6. None of the above.

Readings and Assignments

- Reading
  - Chapter 6: pgs. 146-149
- Lesson 11 Practice Exercises
- Lesson 11 Excel Module