Math 140
Introductory Statistics

Professor Bernardo Ábrego
Lecture 27
Sections 8.4

Example: Two AIDS treatments

- Consider a clinical trial experiment comparing two treatments for AIDS-related complex (ARC). The investigators want to find out if there is a significant difference on the survival rates of patients who had already developed AIDS. They have two treatments, patients who were given AZT and patients who were given AZT + ACV. Here's the data

Treated with

Survived?		AZT	AZT + ACV	Total
	No	28	13	41
	Yes	41	49	90
	Total	69	62	131

Notation:Proportions, Sample Proportions, and Sample Sizes.

$n_1 = 69$ sample size of patients treated with AZT

$n_2 = 62$ sample size of patients treated with AZT+ACV

$\hat{p}_1 = \frac{41}{69}$ proportion of patients in the sample treated with AZT that survived

$\hat{p}_2 = \frac{49}{62}$ proportion of patients in the sample treated with AZT+ACV that survived

p_1 True proportion of survival if all patients were treated with AZT (unknown)

p_2 True proportion of survival if all patients were treated with AZT+ACV (unknown)

Assumptions about the difference of two proportions.

- If we have obtain two **independent** sample proportions $\hat{p}_1$ and $\hat{p}_2$, then the distribution of the difference of the two proportions $\hat{p}_1 - \hat{p}_2$ is **approximately normal** as long as each proportion satisfies the following three conditions:
  - Each sample $\hat{p}_1$ and $\hat{p}_2$ is a simple random sample from a binomial population, and they are independent from each other.
    or, in case of experiments, subjects were randomly assigned to their treatments.
  - All the numbers $n_1\hat{p}_1, n_2\hat{p}_2, n_1(1-\hat{p}_1), n_2(1-\hat{p}_2)$ are at least 5.
  - Each of the two population sizes is at least 10 times the sample size.

Checking the conditions

- Check the conditions.
  - We assume that subjects were randomly assigned to treatments.
  - All of the following are greater than 5

$$n_1\hat{p}_1 = 69(41/69) = 41, n_2\hat{p}_2 = 62(49/62) = 49$$

$$n_1(1-\hat{p}_1) = 69(1-41/69) = 28$$

$$n_2(1-\hat{p}_2) = 62(1-49/62) = 13$$

  - The population size of AIDS patients that could potentially be treated with AZT is clearly greater than 10 times 69 = 690

    The population size of AIDS patients that could potentially be treated with AZT+ACV is clearly greater than 10 times 62 = 620

Writing the hypothesis

- The null hypothesis
  - H_0 : The new therapy (AZ+ACV) is as good as the old therapy (AZT). That is the survival rate if all patients were treated with AZT+ACV would be equal to the survival rate if all patients were treated with AZT. In symbols $p_1 = p_2$ or $p_1 - p_2 = 0$
- The alternate hypothesis
  - H_a : The new therapy (AZ+ACV) is better than the old therapy (AZT). That is the survival rate if all patients were treated with AZT+ACV would be greater than the survival rate if all patients were treated with AZT. In symbols $p_1 < p_2$ or $p_1 - p_2 < 0$

The Test Statistic

- The test statistic in general follows the form

$$\boxed{\text{Test Statistic} = \frac{\text{estimate} - \text{parameter}}{\text{standard deviation of estimate}}}$$

- The parameter is $p_1 - p_2$, and if the null hypothesis is true then $p_1 - p_2 = 0$.
- The estimate is what we obtain for the difference of the two proportions according to our samples.

$$\begin{aligned}\text{estimate} &= \hat{p}_1 - \hat{p}_2 \\ &= 0.594 - 0.790 \\ &= -0.196\end{aligned}$$

- Now, the standard deviation of the estimate under the assumption that $p_1 = p_2$ can be approximated by

$$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

- Where $\hat{p}$ is the **pooled estimate**

$$\hat{p} = \frac{\text{total survived}}{\text{grand total}} = \frac{90}{131} \approx 0.687$$

The Test Statistic

$$\boxed{\text{Test Statistic} = \frac{\text{estimate} - \text{parameter}}{\text{standard deviation of estimate}}}$$

- The parameter is $p_1 - p_2 = 0$,

$$\begin{aligned}\text{estimate} &= \hat{p}_1 - \hat{p}_2 \\ &= 0.594 - 0.790 \\ &= -0.196\end{aligned}$$

- The standard deviation of the estimate

$$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

- $\hat{p}$ is the **pooled estimate**

$$\hat{p} = \frac{\text{total survived}}{\text{grand total}} = \frac{90}{131} \approx 0.687$$

- so the standard deviation of the estimate is:

$$\sqrt{(0.687)(1-0.687)\left(\frac{1}{69}+\frac{1}{62}\right)}$$

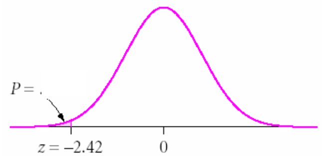
$$= 0.081145$$

- and the test statistic:

$$\frac{-0.196 - 0}{0.081145} = -2.415429$$

P-Value

- To get the P-Value we use the fact that the distribution of the difference is approximately normal.



- Since our test is one-sided we need to calculate the probability that the difference is *less than* our test-statistic.
- We can find P by doing

`normalcdf`(– 999999,– 2.4154)=.007858

Conclusion

- Since the P-Value equals 0.78% and this is definitely less than 5% we reject the null hypothesis.
- If both treatments were equally effective (null hypothesis is true) then there is only a 0.78% chance of getting a difference $\hat{p}_1 - \hat{p}_2$ as small or smaller than – 0.196. This probability is so small that we are confident that if all subjects in the experiment had been given AZT+ACV there would be a larger survival rate than if they had received only AZT.

Components of a Significance Test for the Difference of two Proportions

- See pages 531-532 in your book.

Example E65 (page 537)

- A poll of 256 boys and 257 girls age 12 to 17 asked, “Do you feel like you are personally making a positive difference in your community?” More girls (195) than boys (161) answered “yes.”
  - a. Using a one-sided test, is this a statistically significant difference? That is, if all teens were asked, are you confident that a larger proportion of girls than boys would say “yes”? Assume that the samples were selected randomly.
  - b. The report says, “Participants were selected through random digit dialing.” Do you have any concerns about whether such a procedure would give a random sample?
  - c. Find a 95% confidence interval for the proportion of all teens who would answer yes. What additional assumption do you need to make to do this?