Introductory Statistics Lectures

Testing a claim about two proportions

Two sample hypothesis test of the proportion

Anthony Tanbakuchi
Department of Mathematics
Pima Community College

(Compile date: Tue May 19 14:50:52 2009)

Contents

1 Testing a claim about two proportions

1.1 Introduction

Example 1. A recent study published in the New England Journal of Medicine (Vol. 318, No. 4) of 22,000 male physicians studied the ability of aspirin to prevent heart attacks. The study ran for six years at a cost of 4.4 million. The study group was broken into two equal halves and one group received regular doses of aspirin and the other placebos. Of those receiving aspirin, 104 suffered heart attacks, while the placebo group had 189 heart attacks. Does the data indicate that aspirin reduces the rate of heart attacks?

Question 1. What is the given data in this problem?

Question 2. What do we want to test?

REVIEW

What is a hypothesis test?
A hypothesis test calculates the probability of observing our sample data **assuming the null hypotheses H_0 is true**. If the probability is low enough (p-value $\leq \alpha$), we reject H_0 and have evidence to support our alternative hypothesis H_a.

Eight simple steps

0. Write down what is **known**.
1. Determine which type of hypothesis **test** to use.
2. Check the test's **requirements**.
3. Formulate the **hypothesis**: H_0, H_a
4. Determine the **significance level** α.
5. Find the p**-value**.
6. Make the **decision**.
7. State the final **conclusion**.

K-T-R-H-S-P-D-C: "Know The Right Hypothesis So People Don't Complain"

Question 3. What is a test statistic?

Question 4. What is the test statistic's distribution based on?

Common form of a test statistic

$$\text{test statistic} = \frac{(\text{sample statistic}) - (\text{null hypothesis of parameter})}{(\text{standard deviation of sample statistic})} \quad (1)$$

Question 5. What is the p-value?

Question 6. If we reject H_0 what is the probability of a Type I error?

Question 7. When does a Type II error occur?

Question 8. What variable represents the probability of a Type II error?

1.2 Testing claims about 2 population proportions

USE

Often used to help answer:

1. Is the proportion of x the same in the two populations?
2. Is the proportion of x the same in the two populations?
3. Is process 1 equivalent to process 2? (Produces same proportion.)
4. Is the new process better than the current process? (Has a higher yield.)
5. Is the new process better than the current process by at least some predetermined threshold amount?

COMPUTATION

Notation

Since we have **two samples** (sample 1 & sample 2) we can define the number of successes, the sample size, and estimates of p in terms of each sample:

$$\text{successes: } x_1, x_2 \quad (2)$$

$$\text{sample sizes: } n_1, n_2 \tag{3}$$

$$\text{point estimate: } \hat{p_1} = \frac{x_1}{n_1}, \qquad \hat{p_2} = \frac{x_2}{n_2}, \qquad \Delta\hat{p} = \hat{p}_1 - \hat{p}_2 \tag{4}$$

And make hypothesis about the two populations:

$$\text{population parameters: } p_1, p_2 \qquad \Delta p = p_1 - p_2 \tag{5}$$

Dependent vs. independent samples

Definition 1.1 Independent samples.
The samples from one population are not related to or paired with the samples values in from the other population.

Definition 1.2 Dependent samples (matched pairs / paired samples).
When the samples from the first population have some relationship or pairing to the second population that is sampled.

Examples:

independent samples Randomly breaking the class into two groups, feeding group 1 only lettuce for a week and group 2 only steak and measuring their weights at the end of the week.

dependent samples Feeding the whole class lettuce for 1 week and then weighting everyone at the end of the week (sample 1). Then for week 2 feeding everyone steak and measuring their weight at the end of week 2 (sample 2).

Definition 1.3 Two sample proportion hypothesis test.

requirements (1) simple **independent** random samples, (2) normal approx. to binomial applies to both samples (np and $nq \geq 5$ for both groups).

null hypothesis $p_1 = p_2$ or $\Delta p = 0$ (**specify which is** p_1 **&** p_2**.**)

alternative hypothesis (1) $p_1 \neq p_2$, (2) $p_1 < p_2$, or (3) $p_1 > p_2$

Note that we will only deal with the common case $H_0 : \Delta = 0$. However, it is possible to have a null hypothesis $H_0 : \Delta p = \Delta p_0$ where Δp_0 is non-zero.

Test statistic

$$z = \frac{\Delta\hat{p} - \Delta p_0}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} \tag{6}$$

where the pooled estimate of p_1 and p_2 is:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}, \qquad \bar{q} = 1 - \bar{p} \tag{7}$$

Question 9. What distribution do you use to calculate the p-value?

Two sample proportion hypothesis test:
`prop.test(x, n, alternative="two.sided")`

x vector defined as: `x=c(`x_1`, `x_2`)`
n vector defined as: `n=c(`n_1`, `n_2`)`
alternative H_a $\neq$:`"two.sided"`, $<$:`"less"`, $>$:`"greater"`

R Command

Confidence interval for two proportions.

Definition 1.4

$$\text{CI for } \Delta p: \qquad \Delta\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p_1}\hat{q_1}}{n_1} + \frac{\hat{p_2}\hat{q_2}}{n_2}} \tag{8}$$

Back to our original example

Example 2. A recent study published in the New England Journal of Medicine (Vol. 318, No. 4) of 22,000 male physicians studied the ability of aspiring to prevent heart attacks. The study ran for six years at a cost of 4.4 million. The study group was broken into two equal halves and one group received regular doses of aspirin and the other placebos. Of those receiving aspirin, 104 suffered heart attacks, while the placebo group had 189 heart attacks. Does the data indicate that aspirin reduces the rate of heart attacks?

Solving the problem: step-by-step
Step 0 known information: (I will make the placebo group the **first** group)

```
R: n1 = 11000
R: n2 = 11000
R: x1 = 189
R: x2 = 104
```

Step 1 Test: two sample proportion.
Step 2 Requirements: (1) simple independent random samples, (2) normal approximation to binomial
Step 3 Hypothesis: $H_0 : p_1 = p_2$, $H_a : p_1 > p_2$. I suspect more heart attacks with the placebo. (Since $p_1 > p_2$ then $p_1 - p_2 < 0$ or $\Delta p = 0$.)
Step 4 Significance: $\alpha = 0.005$ (Effect of Type I error may hurt people, therefore make probability low.)

Step 5 p-value: using R

```
R: x = c(x1, x2)
R: n = c(n1, n2)
R: alpha = 0.005
R: res = prop.test(x, n, alternative = "greater")
R: res
        2-sample test for equality of proportions with
        continuity correction

data:  x out of n
X-squared = 24.407, df = 1, p-value = 3.899e-07
alternative hypothesis: greater
95 percent confidence interval:
 0.0050953 1.0000000
sample estimates:
```

```
    prop 1     prop 2
0.0171818 0.0094545
```

Step 6 Decision: since p-val $\leq \alpha$ **reject** H_0. ($3.9e - 07 \leq 0.005$)

Step 7 Conclusion: "The sample data supports the claim that aspiring does reduce the rate of heart attacks."

The probability that we made the wrong decision (a Type I error) is 3.9e-07.

Question 10. What is the problem with this conclusion?

Question 11. How is our use of `prop.test` different than the 1-sample case?

Question 12. Find the test statistic z and the p-value manually. (Check: $\bar{p}$=0.0133, z= 5, p-value= 2.88e-07)

Question 13. Construct a 95% CI for Δp. (Check $\Delta\hat{p} = 0.00773$ $z = 1.96$, $E = 0.00303$, $CI : (0.0047, 0.0108)$)

Question 14. Does the confidence interval support our hypothesis test conclusion?

1.3 Finding required sample size

2-PROP TEST SAMPLE SIZE (AD-HOC):
`power.prop.test(p1=`p_1`, p2=`p_2`, power=`$1-\beta$`,`
`sig.level=`α`, alternative="two.sided")`

R COMMAND

To determine the required sample size at the desired power:

p1 estimated true proportion of interest in sample 1. (see note below)

p2 estimated true proportion of interest in sample 2. (see note below)

power desired power, typically 0.8 or 0.9).

sig.level significance level, typically 0.05.

alternative set to `"two.sided"` or `"one.sided"` .

Note: `p1` and `p2` describe the minimum effect size of interest. Make $p_1 - p_2$ the smallest value you are interested in detecting, and make p_1 and p_2 centered about 0.5 for the worst case scenario.

Example 3. Determine the required sample size for a study that wishes to test the hypothesis that proportion of democrats and republicans who support proposition 187 is not the same. It is believed that the proportion of democrats and republicans who support the proposition is close to 50% for both, however it would be meaningful if the proportion of support differed by more than 10%.

```
R: res = power.prop.test(p1 = 0.45, p2 = 0.55, power = 0.9,
+      sig.level = 0.05, alternative = "two.sided")
R: res
     Two-sample comparison of proportions power calculation

              n = 523.29
             p1 = 0.45
             p2 = 0.55
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

 NOTE: n is number in *each* group
```

Thus, we need both the democrat and the republican group sample sizes to be 524

1.4 Summary

Two sample proportion hypothesis test

requirements (1) simple **independent** random samples, (2) normal approx. to binomial applies to both samples (np and $nq \geq 5$ for both groups).

null hypothesis $p_1 = p_2$ (**specify which is** p_1 **&** p_2**.**)

alternative hypothesis (1) $p_1 \neq p_2$, (2) $p_1 < p_2$, or (3) $p_1 > p_2$

test in R : `prop.test(x, n, alternative="two.sided")`

Note that `x, n` are **ordered vectors**.

1.5 Additional examples

Independent random samples of $n_1 = 140$ and $n_2 = 140$ observations were randomly selected from binomial populations 1 and 2, respectively. Sample 1 had 74 successes, and sample 2 had 81 successes. Suppose that you have no preconceived theory concerning which parameter p_1 or p_2, is the larger and you wish to detect only a difference between the two parameters if one exits.

Question 15. What is your null and alternative hypothesis?

Question 16. Find the p-value? (Check: 0.471)

Question 17. What is your formal decision?

Question 18. What is your conclusion?

Question 19. Karl Pearson, who developed many important concepts in statistics, collected crime data in 1909. Of those convicted of arson, 50 were drinkers and 43 abstained. Of those convicted of fraud, 63 were drinkers and 144 abstained. Use a 0.01 significance level to test the claim that the proportion of drinkers among convicted arsonists is greater than the proportion of drinkers among those convicted of fraud. Does is seem reasonable that drinking might have an effect on the type of crime? Why? (Check: p-value=9.65e-05)