Module E: Advanced Inference

Slide Deck E2:

Handling One and Two Proportions

The section in which we see how to perform hypothesis tests concerning population proportions using the R Statistical Environment. This deck will focus not only the code used to perform the analysis, it will also emphasize the analysis process.

Today's Objectives

By the end of this slidedeck, you should

1. understand the theory behind, and test hypotheses about:
   - a single population proportion
   - the difference between two population proportions
2. better understand the p-value and how to test hypotheses
3. clearly specify how confidence intervals and p-values both give important information about the population parameter

Note that we are moving beyond the general theory of confidence intervals and hypothesis testing. We are looking at how to specifically perform the procedures. Make sure you pay attention to the statistical process we follow.

One-Parameter Procedures: p

Parametric Procedure: Binomial procedure

- Graphic: Binomial plot
  `binom.plot(x, n)`
- Requires: Data generated from Binomial distribution
- `R` function: `binom.test(x, n)`

Note: This is *not* the procedure Hawkes covers. They use something called the Wald test, `wald.test` because it is easier to perform these calculations by hand.

Start of Lecture Material | The Test's Theory
One-Pop Props | Example 1: Coins
Two-Pop Props | Example 2: Juniors
End of Section Material | Example 3: Representativeness

The Test's Theory

Since we are trying to draw conclusions about a single population proportion, we should use a test statistic based on the sample proportion (Wald test). . . or upon the sample count (Binomial test).

If we know

$$X \sim \mathcal{B}in(n, p)$$

then we know that the number of observations is a particularly fine test statistic (we know its exact distribution). We only know the distribution of proportions approximately.

The following are three examples showing how to perform these calculations.

Example 1: Coins

Example

I have a coin that I think is fair. To test this, I flip it 10 times and count the number of heads in those 10 flips. A total of 3 heads actually came up. Is this sufficient evidence that the coin is not fair?

Here, the claim is $p = 0.500$ (fair). Since it contains the '=' sign, it is the null hypothesis. That means the two hypotheses are

$$H_0 : p = 0.500$$
$$H_a : p \neq 0.500$$

Example 1: Coins

We are trying to make a conclusion about p where the data are generated from a Binomial distribution. Thus, under the null hypothesis:

$$X \sim \mathcal{B}in(n = 10, p = 0.500)$$

We observed $X = 3$.

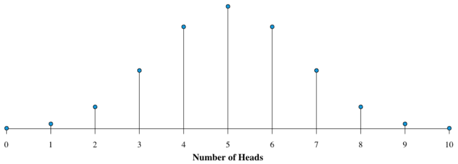
The p-value is defined as the probability of observing data this extreme — or more so — given that the null hypothesis is true. In this example, that means:

$$\text{p-value} = \mathbb{P}[\ X \leq 3\ ] + \mathbb{P}[\ X \geq 7\ ]$$

So, it is clear where the 3 came from, but where did the 7 come from???

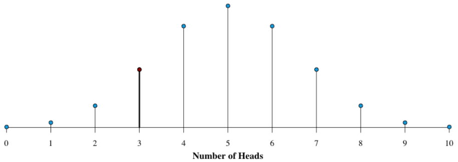
Example 1: Coins

$$X \sim \mathcal{B}in(n = 10, p = 0.500)$$



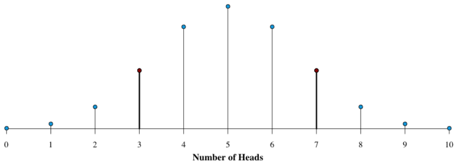
Example 1: Coins

$$\mathbb{P}[\ X = 3\ ]$$



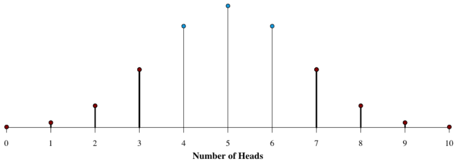
Example 1: Coins

$$\mathbb{P}[\ X = 3 \cup X = 7\ ]$$



Example 1: Coins

$$\mathbb{P}[\ X \leq 3 \cup X \geq 7\ ]$$



Example 1: Coins

Thus, the p-value is

$$\begin{aligned}
\text{p-value} &= \mathbb{P}[\ X \leq 3 \cup X \geq 7\ ] \\
&= \mathbb{P}[\ X \leq 3\ ] + \mathbb{P}[\ X \geq 7\ ] \\
&= \mathbb{P}[\ X \leq 3\ ] + \left(1 - \mathbb{P}[\ X < 7\ ]\right) \\
&= \mathbb{P}[\ X \leq 3\ ] + \left(1 - \mathbb{P}[\ X \leq 6\ ]\right) \\
&= \texttt{pbinom(3, size=10, prob=0.50)} \\
&\qquad \texttt{+ 1 - pbinom(6, size=10, prob=0.5)} \\
&= 0.34375
\end{aligned}$$

Thus, if the coin is fair, we would expect to observe a result this extreme or more so more than a third of the time.

Example 1: Coins

From the `allProcedures` document and the SCA examples posted, we know we could also just use

```
binom.test(x=3, n=10, p=0.50)
```

The results of this line tell us that the p-value is 0.3438.

Since this p-value is greater than $\alpha = 0.05$, we fail to reject the null hypothesis. There is no sufficient evidence that the coin is unfair. Furthermore, a 95% confidence interval for the probability of a flip landing head is from 0.067 to 0.652.

Example 2: Juniors

Example

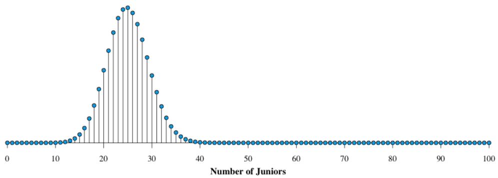
I contend that more than a quarter of the students at Knox are Juniors. To test this, I randomly sample from the student body asking class year. In my sample of 100 students, 30 stated they were Juniors.

Here, the claim is $p > 0.250$. Since it contains the '>' sign, it is not the null hypothesis. That means the two hypotheses are

$$H_0 : p \leq 0.250$$
$$H_a : p > 0.250$$

Example 2: Juniors

$$X \sim \mathcal{B}in(n = 100, p = 0.25)$$



Example 2: Juniors

Because

$$H_a : p > 0.250$$

The p-value is

$$\begin{aligned} \text{p-value} &= \mathbb{P}[\ X \geq 30\ ] \\ &= 1 - \mathbb{P}[\ X < 30\ ] \\ &= 1 - \mathbb{P}[\ X \leq 29\ ] \\ &= 0.1495 \end{aligned}$$

Because the p-value of 0.1495 is greater than our $\alpha = 0.05$, we cannot reject the hypothesis that the proportion of Juniors is greater than a quarter. In fact, we are 95% confident that the proportion of Juniors at Knox College is greater than 0.2249.

Example 2: Juniors

Using the power of R:

```
binom.test(x=30, n=100, p=0.25, alternative="greater")
```

The resulting output is

```
        Exact binomial test

data:  30 and 100
number of successes = 30, number of trials = 100, p-value = 0.1495
alternative hypothesis: true probability of success is greater than 0.25
95 percent confidence interval:
 0.2249232 1.0000000
sample estimates:
probability of success
                   0.3
```

Example 3: Representativeness

One use of the Binomial test is to check if your sample is representative. The data file someCollege was sent to me by the registrar of some college. I was supposed to model success (sufficiently high GPA), given some of the other variables.

Let us perform a quick check to see if the data are *reasonably* representative of the population.

Example 3: Representativeness

Example

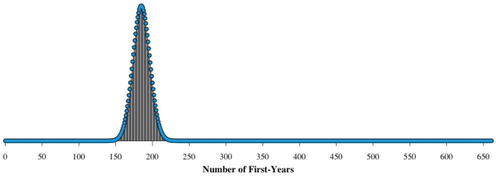
My (provided) sample consisted of 661 students, of which 22 were Freshmen. Given that the proportion of Freshmen at SCU is 28%, are the data representative in terms of Freshmen?

Here, the claim is $p = 0.280$. Since it contains the '=' sign, it is the null hypothesis. That means the two hypotheses are

$$H_0 : p = 0.280$$
$$H_a : p \neq 0.280$$

Example 3: Representativeness

$$X \sim \mathcal{B}in(n = 661, p = 0.28)$$



Example 3: Representativeness

The p-value is the probability of observing data this extreme — or more so — given that the null hypothesis is true.

This translates as:

$$\begin{aligned} \text{p-value} &= 2 \times \mathbb{P}[\ X \leq 22\ ] \\ &= 2 \times 0.0000 \\ &= 0.0000 \end{aligned}$$

Since the observed p-value is less than our $\alpha = 0.05$, we reject the null hypothesis in favor of the alternative. In terms of Freshmen, this sample is *not* representative of the population at SCU.

Example 3: Representativeness

Using the power of R:

```
binom.test(x=22, n=661, p=0.28)
```

This results in this output:

```
        Exact binomial test

data:  22 and 661
number of successes = 22, number of trials = 661, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.28
95 percent confidence interval:
 0.02097345 0.04995873
sample estimates:
probability of success
             0.0332829
```


Example 3: Representativeness

Brief conclusion:

Since the observed p-value is much less than our $\alpha = 0.05$, we reject the null hypothesis in favor of the alternative.

In terms of Freshmen, this sample is not representative of the population at SCU. In fact, the proportion of Freshmen at SCU would need to be between 2.1% and 5.0% for this sample to be possibly representative.

Two-Parameter Procedures: $p_1 - p_2$

Parametric Procedure: Proportions Procedure

- Graphic: Binomial plot
  `binom.plot(x=c(x1,x2), n=c(n1,n2))`
- Requires: Expected number of successes is at least 5 in each group
- R function: `prop.test(x=c(x1,x2), n=c(n1,n2))`

Note: This is *not* the procedure Hawkes covers. They use something close to this, but *this* procedure makes adjustments for the fact that the Binomial distribution is discrete and the Normal distribution is not. As such, you will need to use the Wald test, `wald.test`, to perform Hawkes homework estimating $p_1 - p_2$.

The Theory

This section creates the theory behind the proportions test.

Since we are trying to draw conclusions about the *difference* between two population proportions, it would make sense to use a test statistic that is based on the difference in the two *sample* proportions.

The Theory

Thus, to start our theory, let us know the distributions of the two random variables:

$$X \sim \mathcal{B}in(n_x, p_x)$$
$$Y \sim \mathcal{B}in(n_y, p_y)$$

Since we are estimating the difference in population proportions, let's start with the difference in sample proportions. This is our first step, create a test statistic based on the difference in sample proportions:

$$\frac{X}{n_x} - \frac{Y}{n_y}$$

The second step is to know the distribution of this test statistic. So... what is its distribution???

The Theory

No clue. =(

In fact, it can be proven that this sample statistic does not have a proper (knowable) distribution (see STAT 322: Mathematical Statistics, II).

As such, this is a dead end.

However, we are saved by the power of the **Central Limit Theorem**!

The Theory

By the CLT, we know the *approximate* distribution of X and Y

$$X \dot\sim \mathcal{N}\big(n_x p_x,\ n_x p_x(1-p_x)\big)$$
$$Y \dot\sim \mathcal{N}\big(n_y p_y,\ n_y p_y(1-p_y)\big)$$

… and the *approximate* distribution of the sample proportions

$$P_x = \frac{X}{n_x} \dot\sim \mathcal{N}\left(p_x,\ \frac{p_x(1-p_x)}{n_x}\right)$$
$$P_y = \frac{Y}{n_y} \dot\sim \mathcal{N}\left(p_y,\ \frac{p_y(1-p_y)}{n_y}\right)$$

⋱

The Theory

… and the approximate distribution of the difference in sample proportions:

$$P_x - P_y \dot\sim \mathcal{N}\left(p_x - p_y,\ \frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}\right)$$

Finally, standardizing the random variable on the left gives us an acceptable test statistic and its distribution:

$$Z = \frac{(P_x - P_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}} \dot\sim \mathcal{N}(0,\ 1)$$

The Theory

If $p_x = p_y$ is a part of our null hypothesis (as usual), then this test statistic simplifies to

$$\frac{P_x - P_y}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}} \dot{\sim} \mathcal{N}(0,\ 1)$$

This is the "Z-procedure version" of the proportions test.

Note that there is an equivalent test. If we square both sides, we have the "Chi-square version" of the proportions test:

$$\frac{(P_x - P_y)^2}{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}} \dot{\sim} \Big(\mathcal{N}(0,\ 1)\Big)^2 = \chi^2_1$$

The one you use depends on the software. There is absolutely no difference between the two of them. With that said, the Chi-square version will be expanded and used later.

Example 4: A Hat Trick

Example

I would like to determine if the proportion of males who wear hats is the same as the proportion of females who do. To test this, I sample 100 males and 100 females. Ten males and 16 females were wearing hats.

The hypotheses are

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

Example 4: A Hat Trick

Step 1: Assemble our Information

Note that we are given the following information from the problem:

- $p_x = 10/100 = 0.10$
- $p_y = 16/100 = 0.16$
- $n_x = 100$
- $n_y = 100$
- $\alpha = 0.05\ldots\ Z_{\alpha/2} = \pm 1.96$

Since we are testing a hypothesis comparing two proportions, this is our test statistic formula:

$$\text{TS} = \frac{(p_x - p_y)^2}{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}$$

Example 4: A Hat Trick

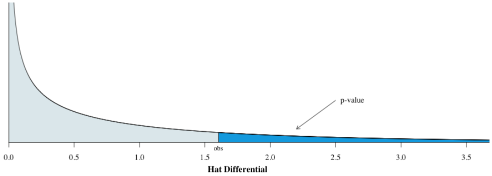
Step 2: Calculate the Test Statistic

$$\begin{aligned}\text{TS} &= \frac{(p_x - p_y)}{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}} \\ &= \frac{(0.10 - 0.16)^2}{\frac{0.10(1-0.10)}{100} + \frac{0.16(1-0.16)}{100}} \\ &= \frac{0.0036}{0.0009 + 0.001344} \\ &= 1.604278\end{aligned}$$

Example 4: A Hat Trick

Step 3: Calculate the p-value

The p-value for this test is the probability of observing data (test statistic) this extreme — or more so — given the null hypothesis is true:



Example 4: A Hat Trick

Step 3: Calculate the p-value

The p-value for this two-tailed test is the probability of observing data (test statistic) this extreme — or more so — given the null hypothesis is true:

$$\begin{aligned} \text{p-value} &= \mathbb{P}[\ TS \geq 1.604278\ ] \\ &= 0.2053 \end{aligned}$$

Example 4: A Hat Trick

Step 4: Interpret the Results

Brief conclusion.
Because the p-value of 0.2053 is greater than our $\alpha = 0.05$, we cannot reject the null hypothesis. There is not enough evidence to claim that the females wear hats at a different rate than males.

Final Notes and Comments

Note: This procedure is the one used by Hawkes, `wald.test`. It is very straight-forward and easy to understand. It also does *not* make some rather important adjustments to take into consideration that the Normal only approximates the Binomial.

Example 4: A Hat Trick

Final Notes and Comments

Doing this in R

```
prop.test( x=c(10,16), n=c(100,100) )
```

The resulting output

```
  2-sample test for equality of proportions
  without continuity correction

data:  c(10, 16) out of c(100, 100)
X-squared = 1.5915, df = 1, p-value = 0.2071
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.15284521  0.03284521
sample estimates:
prop 1 prop 2
  0.10   0.16
```

Today's Objectives

Now that we have concluded this lecture, you should be able to

1. understand the theory behind, and test hypotheses about:
   - a single population proportion
   - the difference between two population proportions
2. better understand the p-value and how to test hypotheses
3. clearly specify how confidence intervals and p-values both give important information about the population parameter

Today's R Functions

Here is what we used the following R functions:

- `binom.test(x, n)`
  performs the Binomial test for one proportion
- `prop.test(x=c(x1,x2), n=c(n1,n2))`
  performs the proportions test for comparing two proportions

Supplemental Activities

The following activities are currently available from the STAT 200 website to give you some practice in performing hypothesis tests concerning population proportions.

- SCA-12
- SCA-22

Source: `https://www.kvasaheim.com/courses/stat200/sca/`

In addition to the SCAs, there are **Laboratory Activity E** (confidence intervals) and **Laboratory Activity F** (hypothesis testing).

Source: `https://www.kvasaheim.com/courses/stat200/labs/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of understanding today's slidedeck:

- Hawkes Learning: Chapters 10 and 11
- Intro to Modern Statistics: Chapters 16 and 17
- R for Starters: Chapters 5 and 6
- Wikipedia: Confidence Intervals
  Hypothesis Testing

Please do not forget to use the `allProcedures` document that lists all of the statistical procedures we will use in R.