Normal Approximations for Hypothesis Testing

Ryan Miller

The Null Distribution of a Single Proportion

Consider the null hypothesis:

$$H_0 : p = 0.5$$

▶ Recall that the **null distribution** represents possible outcomes (sample proportions) that could have occurred *had the null hypothesis been true*
  ▶ For this hypothesis, the *center* of the null distribution is 0.5
  ▶ Notice how the *spread* of null distribution has a major impact on the *p*-value

Null Dist. when n = 10

Null Dist. when n = 40
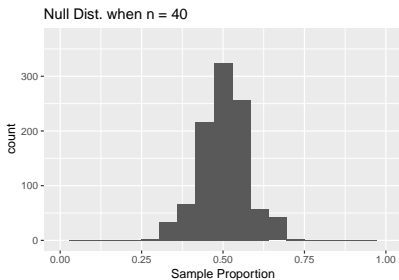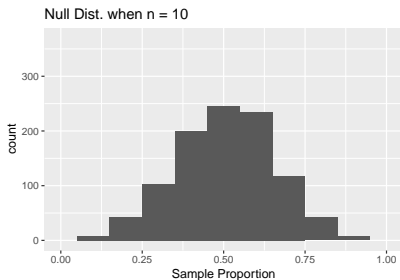
CLT for Null Distributions

The null distribution of a single proportion can be approximated by a normal curve. Assuming $H_0 : p = p_0$ is true, then:

$$\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

- ▶ We already used a similar approximation to calculate confidence intervals
    - ▶ This time, we are considering the sampling distribution of $\hat{p}$ in a hypothetical world where the null hypothesis is true
    - ▶ So the value specified in the null hypothesis (denoted p_0) should be used in the approximation

Practice - Finding *p*-values

For each of the following scenarios use the Theoretical Distribution Menu on StatKey

1. Suppose $H_0 : p = 0.5$ and the sample size is $n = 12$, use a normal approximation of the null distribution to find the two-sided *p*-value for a sample proportion of $\hat{p} = 0.75$
2. Suppose $H_0 : p = 0.1$ and the sample size is $n = 100$, use a normal approximation of the null distribution to find the two-sided *p*-value for a sample proportion of $\hat{p} = 0.02$

Practice - Solutions

1. Here $p_0 = 0.5$, so $SE = \sqrt{\frac{0.5*(1-0.5)}{12}} = 0.144$; leading us to approximate the null distribution with a $N(0.5, 0.144)$ curve, suggesting a two-sided p-value of 0.082

2. Here $p_0 = 0.1$, so $SE = \sqrt{\frac{0.1*(1-0.1)}{100}} = 0.03$; leading us to approximate the null distribution with a $N(0.1, 0.03)$ curve, suggesting a two-sided p-value of 0.0076

Standardization and Test Statistics

▶ Rather than considering a distinct null distribution for every different application, it is common to construct a standardized **test statistic** that can *always* be compared to the *standard normal distribution*

▶ Similar to *z*-scores, most test statistics take the form:

$$\text{Test Statistic} = \frac{\text{Observed Estimate} - \text{Null Value}}{\text{Standard Error}}$$

▶ To find the *p*-value, we locate this test statistic on the $N(0,1)$ curve

  ▶ The benefit of this approach is that the test statistic will *always* have the same null distribution

The z-test for a single proportion

▶ For a single proportion, this standardization gives rise to the *one-proportion* **z-test**, which uses the following test statistic:

$$z = \frac{\hat{p} - p_0}{SE}$$

▶ p_0 is specified in the null hypothesis, and $SE = \sqrt{\frac{p_0(1-p_0)}{n}}$

▶ The p-value of this test is found by comparing z to the standard normal distribution

Practice - Applying the *z*-test

Previously, we evaluated a statement on Wikipedia claiming that 70% of babies born at 25 weeks gestation (15 weeks early) go on to survive by using data from a study where researchers searched the Johns Hopkins' University Hospital medical records and found 31 of 39 such babies survived.

1. State the null and alternative hypothesis you'd use to determine if this study provides evidence against Wikipedia's claim
2. Calculate the test statistic for the *z*-test, and compare it to standard normal distribution to determine the *p*-value for this test
3. Use the *p*-value to make a conclusion regarding Wikipedia's claim

Practice - Solution

1. $H_0 : p = .7$ and $H_A : p \neq .7$
2. Here $\hat{p} = 31/39 = 0.795$ and $SE = \sqrt{\frac{.7(1-.7)}{39}}$, so
   $z = \frac{.795-.7}{.0734} = 1.29$
3. Locating $z = 1.29$ on the standard normal curve suggests a two-sided p-value of 0.198

The Exact Binomial Test

▶ For homework and exams in this class, we'll use the *z*-test when working with a single proportion
▶ However, Minitab allows you to use the **exact binomial test** via the "1 Proportion . . ." menu
  ▶ The exact binomial test is more accurate (it doesn't rely on a normal approximation) and should be preferred over the *z*-test in real-world applications
  ▶ In your final project (and future endeavors) you should pretty much always use the exact binomial test, especially when *n* is relatively small

Understanding the Exact Binomial Test

▶ Suppose $H_0 : p = .5$, $H_A : p > .5$, and we observe
   $\hat{p} = 3/4 = 0.75$

   ▶ Since $n = 4$, there are only five possible sample proportions that
     we could have observed
   ▶ Each has a probability that can be *calculated exactly* (under H_0)

Proportion	Probability	Calculation
0.00	0.0625	1 * 0.5^4 * 0.5^0
0.25	0.25	4 * 0.5^3 * 0.5^1
0.50	0.375	6 * 0.5^2 * 0.5^2
0.75	0.25	4 * 0.5^1 * 0.5^3
1.00	0.0625	1 * 0.5^0 * 0.5^4

▶ In this scenario, the one-sided exact binomial *p*-value of 0.3125
   is found by summing 0.25 and 0.0625 (the possible results that
   are at least as extreme as $\hat{p}$)

Practice - Comparing the Exact Binomial and *z*-test

▶ In the Johns Hopkins' example, researchers found that 31 of 39 babies born 15 weeks prematurely went on to survive.
▶ We used these data to test Wikipedia's claim that 70% of babies born 15 weeks early will survive, seeing a *p*-value of 0.198 when using the *z*-test. Compare this result with that from the exact binomial test:

1. Open Minitab and navigate: "Stat -> Basic Stats -> 1 Proportion"
2. Select "Summarized data" from the dropdown menu, and input the necessary information to conduct the exact binomial test
3. Compare your *p*-value with that of the *z*-test, why do you think they might be different?

Conditions for the *z*-test

▶ As we saw when constructing confidence intervals, normal approximations are only appropriately when certain conditions are met

▶ For testing a single proportion these conditions are:
   1. $n * p > 10$
   2. $n * (1 - p) > 10$

▶ These conditions are typically evaluated using p_0, since its value determines the null distribution

▶ When they aren't satisfied, the results of the exact binomial test might differ substantially from those of the *z*-test, and the exact binomial test should be preferred.

Null Distribution for a Difference in Proportions

Recall that the CLT suggests the following normal approximation for a difference in proportions:

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

▶ How might we use this approximation to determine the null distribution for $H_0 : p_1 - p_2 = 0$?

▶ Are there any difficulties in using the same approach we did for a single proportion to derive a *two-proportion z-test*?

Null Distribution for a Difference in Proportions

- For a difference in proportions, the null hypothesis doesn't explicitly specify the values of p_1 and p_2
  - This means we can't simply plug-in the "null value" in the CLT result and get the null distribution
- In fact, there are infinitely many values of p_1 and p_2 that will satisfy $H_0 : p_1 - p_2 = 0$
  - The proper choice will both satisfy the null hypothesis, and be consistent with our data
- Using a **pooled proportion**, $\hat{p}_{1+2}$, in place of *both* p_1 and p_2 accomplishes this
  - $\hat{p}_{1+2}$ is calculated by *ignoring the grouping variable* that defines the two proportions
  - For example, if $\hat{p}_1 = 12/20$ and $\hat{p}_2 = 7/15$, then $\hat{p}_{1+2} = \frac{7+12}{20+15}$

Practice - z-test for a Difference in Proportions

Previously, we introduced and analyzed data from a 2010 study which randomly assigned doctors to use either electronic prescription forms, or continue using written prescriptions. After 1 year, the error rate of each group was recorded:

	Error	Non-errors	Total
Electronic	254	3594	3848
Hand-written	1478	2370	3848

1. State the null and alternative hypotheses for a two-sided test evaluating the relationship between form type and errors
2. What is the *pooled proportion* in this application?
3. Use the pooled proportion and CLT normal approximation to conduct a z-test, report your p-value and conclusion

Practice - Solution

1. $H_0 : p_1 - p_2 = 0$ and $H_A : p_1 - p_2 \neq 0$
2. $\hat{p}_{1+2} = \frac{\text{Frequency in } n_1 + \text{Frequency in } n_2}{n_1 + n_2} = \frac{254 + 1479}{3848 + 3848} = 0.225$
3. $z_{\text{test}} = \frac{-0.318 - 0}{\sqrt{\frac{0.225(1-0.225)}{3848} + \frac{0.225(1-0.225)}{3848}}} = -33.4$; the p-value is near zero, providing overwhelming evidence that use of the electronic form leads to a lower error rate

Conditions for the z-test (differences in proportions)

As we've now discussed for many procedures based upon CLT results, certain sample size conditions should be checked to help us gauge whether the approximation is appropriate. For differences in proportions these conditions are:

1. $n_1 * p_1 \geq 10$
2. $n_1 * (1 - p_1) \geq 10$
3. $n_2 * p_2 \geq 10$
4. $n_2 * (1 - p_2) \geq 10$

These conditions are typically evaluated using the pooled proportion.

Fisher's Exact Test (differences in proportions)

- As was the case with a single proportion, we should prefer an exact over the z-test whenever we have access to software (ie: your final project and your future work outside of this class)
    - For a difference in proportions, the appropriate test is **Fisher's exact test**
- When the conditions on the prior slide are met, these two tests will produce similar results
- The mathematical details are beyond the scope of this course, but it is basically a generalization of the exact binomial test to two-way frequency tables

Practice - Fisher's Exact Test

In Minitab, use the "2 Proportions . . . " menu to analyze the
electronic vs. written prescription experiment by performing Fisher's
exact test.

	Error	Non-errors	Total
Electronic	254	3594	3848
Hand-written	1478	2370	3848

- ▶ How does the p-value of this test compare to the p-value of
  the z-test?
- ▶ Why do you think these p-values are similar/different?

Practice - Solution

1. The *p*-value of the exact test is approximately 0.000
2. The *p*-values of Fisher's exact test and *z*-test are similar in this application because the sample size is large

Summary

- For both the *one-proportion z-test* and the *two-proportions z-test* we "plug-in" values satisfying the null hypothesis into the CLT normal approximation

    - For a single proportion this is straightforward, p_0 is directly stated
    - For a difference in proportions, we must use the **pooled proportion**, $\hat{p}_{1+2}$

- In both cases, we compare the *z-test* statistic to the *standard normal curve*, $N(0,1)$, to determine the test's *p*-value
- The *z*-test is simple and convenient, but whenever possible we should prefer an exact test (implemented in Minitab)

    - Exact tests are especially important when the sample size is small

Conclusion

Right now you should. . .

1. Understand how a normal approximation can be used to describe the sampling and null distributions
2. Know how to perform hypothesis tests for one proportion, or a difference in proportions, using the CLT normal approximation
3. Be aware of the assumptions required for these normal approximations to be reasonable
4. Be aware of the exact binomial test and Fisher's exact test (and when to use them)

These notes cover Ch 5.1 and parts of Ch 6 of our textbook, I encourage you to read through those sections and their examples