Section 6.1: Inference for a Single Proportion

STAT 1201
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

Chapter 6

Inference for categorical data

6.1 Inference for a single proportion

6.2 Difference of two proportions

6.3 Testing for goodness of fit using chi-square

6.4 Testing for independence in two-way tables

In this chapter, we apply the methods and ideas from Chapter 5 in several contexts for categorical data. We'll start by revisiting what we learned for a single proportion, where the normal distribution can be used to model the uncertainty in the sample proportion. Next, we apply these same ideas to analyze the difference of two proportions using the normal model. Later in the chapter, we apply inference techniques to contingency tables; while we will use a different distribution in this context, the core ideas of hypothesis testing remain the same.

For videos, slides, and other resources, please visit
www.openintro.org/os

6.1 Inference for a single proportion

We encountered inference methods for a single proportion in Chapter 5, exploring point estimates, confidence intervals, and hypothesis tests. In this section, we'll do a review of these topics and also how to choose an appropriate sample size when collecting data for single proportion contexts.

6.1.1 Identifying when the sample proportion is nearly normal

A sample proportion $\hat{p}$ can be modeled using a normal distribution when the sample observations are independent and the sample size is sufficiently large.

SAMPLING DISTRIBUTION OF $\hat{p}$

The sampling distribution for $\hat{p}$ based on a sample of size n from a population with a true proportion p is nearly normal when:

1. The sample's observations are independent, e.g. are from a simple random sample.
2. We expected to see at least 10 successes and 10 failures in the sample, i.e. $np \geq 10$ and $n(1-p) \geq 10$. This is called the **success-failure condition**.

When these conditions are met, then the sampling distribution of $\hat{p}$ is nearly normal with mean p and standard error $SE = \sqrt{\frac{p(1-p)}{n}}$.

Typically we don't know the true proportion p, so we substitute some value to check conditions and estimate the standard error. For confidence intervals, the sample proportion $\hat{p}$ is used to check the success-failure condition and compute the standard error. For hypothesis tests, typically the null value – that is, the proportion claimed in the null hypothesis – is used in place of p.

6.1.2 Confidence intervals for a proportion

A confidence interval provides a range of plausible values for the parameter p, and when $\hat{p}$ can be modeled using a normal distribution, the confidence interval for p takes the form

$$\hat{p} \pm z^{\star} \times SE$$

EXAMPLE 6.1

A simple random sample of 826 payday loan borrowers was surveyed to better understand their interests around regulation and costs. 70% of the responses supported new regulations on payday lenders. Is it reasonable to model $\hat{p} = 0.70$ using a normal distribution?

E

The data are a random sample, so the observations are independent and representative of the population of interest.

We also must check the success-failure condition, which we do using $\hat{p}$ in place of p when computing a confidence interval:

$$\text{Support: } np \approx 826 \times 0.70 = 578 \qquad \text{Not: } n(1-p) \approx 826 \times (1-0.70) = 248$$

Since both values are at least 10, we can use the normal distribution to model $\hat{p}$.

G

GUIDED PRACTICE 6.2

Estimate the standard error of $\hat{p} = 0.70$. Because p is unknown and the standard error is for a confidence interval, use $\hat{p}$ in place of p in the formula.[1]

E

EXAMPLE 6.3

Construct a 95% confidence interval for p, the proportion of payday borrowers who support increased regulation for payday lenders.

Using the point estimate 0.70, $z^\star = 1.96$ for a 95% confidence interval, and the standard error $SE = 0.016$ from Guided Practice 6.2, the confidence interval is

$$\text{point estimate} \ \pm \ z^\star \times SE \quad \rightarrow \quad 0.70 \ \pm \ 1.96 \times 0.016 \quad \rightarrow \quad (0.669, 0.731)$$

We are 95% confident that the true proportion of payday borrowers who supported regulation at the time of the poll was between 0.669 and 0.731.

CONFIDENCE INTERVAL FOR A SINGLE PROPORTION

Once you've determined a one-proportion confidence interval would be helpful for an application, there are four steps to constructing the interval:

Prepare. Identify $\hat{p}$ and n, and determine what confidence level you wish to use.

Check. Verify the conditions to ensure $\hat{p}$ is nearly normal. For one-proportion confidence intervals, use $\hat{p}$ in place of p to check the success-failure condition.

Calculate. If the conditions hold, compute SE using $\hat{p}$, find $z^\star$, and construct the interval.

Conclude. Interpret the confidence interval in the context of the problem.

For additional one-proportion confidence interval examples, see Section 5.2.

6.1.3 Hypothesis testing for a proportion

One possible regulation for payday lenders is that they would be required to do a credit check and evaluate debt payments against the borrower's finances. We would like to know: would borrowers support this form of regulation?

G

GUIDED PRACTICE 6.4

Set up hypotheses to evaluate whether borrowers have a majority support or majority opposition for this type of regulation.[2]

To apply the normal distribution framework in the context of a hypothesis test for a proportion, the independence and success-failure conditions must be satisfied. In a hypothesis test, the success-failure condition is checked using the null proportion: we verify np_0 and $n(1 - p_0)$ are at least 10, where p_0 is the null value.

---

[1] $SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{0.70(1-0.70)}{826}} = 0.016$.

[2] H_0: $p = 0.50$. H_A: $p \neq 0.50$.

G

GUIDED PRACTICE 6.5

Do payday loan borrowers support a regulation that would require lenders to pull their credit report and evaluate their debt payments? From a random sample of 826 borrowers, 51% said they would support such a regulation. Is it reasonable to model $\hat{p} = 0.51$ using a normal distribution for a hypothesis test here?[3]

E

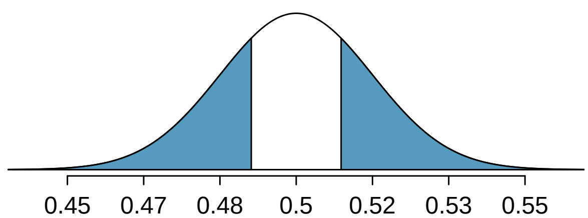
EXAMPLE 6.6

Using the hypotheses and data from Guided Practice 6.4 and 6.5, evaluate whether the poll provides convincing evidence that a majority of payday loan borrowers support a new regulation that would require lenders to pull credit reports and evaluate debt payments.

With hypotheses already set up and conditions checked, we can move onto calculations. The standard error in the context of a one-proportion hypothesis test is computed using the null value, p_0:

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.5(1 - 0.5)}{826}} = 0.017$$

A picture of the normal model is shown below with the p-value represented by the shaded region.



Based on the normal model, the test statistic can be computed as the Z-score of the point estimate:

$$Z = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{0.51 - 0.50}{0.017} = 0.59$$

The single tail area is 0.2776, and the p-value, represented by both tail areas together, is 0.5552. Because the p-value is larger than 0.05, we do not reject H_0. The poll does not provide convincing evidence that a majority of payday loan borrowers support or oppose regulations around credit checks and evaluation of debt payments.

HYPOTHESIS TESTING FOR A SINGLE PROPORTION

Once you've determined a one-proportion hypothesis test is the correct procedure, there are four steps to completing the test:

Prepare. Identify the parameter of interest, list hypotheses, identify the significance level, and identify $\hat{p}$ and n.

Check. Verify conditions to ensure $\hat{p}$ is nearly normal under H_0. For one-proportion hypothesis tests, use the null value to check the success-failure condition.

Calculate. If the conditions hold, compute the standard error, again using p_0, compute the Z-score, and identify the p-value.

Conclude. Evaluate the hypothesis test by comparing the p-value to α, and provide a conclusion in the context of the problem.

For additional one-proportion hypothesis test examples, see Section 5.3.

---

[3]Independence holds since the poll is based on a random sample. The success-failure condition also holds, which is checked using the null value ($p_0 = 0.5$) from H_0: $np_0 = 826 \times 0.5 = 413$, $n(1 - p_0) = 826 \times 0.5 = 413$.

6.1.4 When one or more conditions aren't met

We've spent a lot of time discussing conditions for when $\hat{p}$ can be reasonably modeled by a normal distribution. What happens when the success-failure condition fails? What about when the independence condition fails? In either case, the general ideas of confidence intervals and hypothesis tests remain the same, but the strategy or technique used to generate the interval or p-value change.

When the success-failure condition isn't met for a hypothesis test, we can simulate the null distribution of $\hat{p}$ using the null value, p_0. The simulation concept is similar to the ideas used in the malaria case study presented in Section 2.3, and an online section outlines this strategy:

www.openintro.org/r?go=stat_sim_prop_ht

For a confidence interval when the success-failure condition isn't met, we can use what's called the **Clopper-Pearson interval**. The details are beyond the scope of this book. However, there are many internet resources covering this topic.

The independence condition is a more nuanced requirement. When it isn't met, it is important to understand how and why it isn't met. For example, if we took a cluster sample (see Section 1.3), suitable statistical methods are available but would be beyond the scope of even most second or third courses in statistics. On the other hand, we'd be stretched to find any method that we could confidently apply to correct the inherent biases of data from a convenience sample.

While this book is scoped to well-constrained statistical problems, do remember that this is just the first book in what is a large library of statistical methods that are suitable for a very wide range of data and contexts.

6.1.5 Choosing a sample size when estimating a proportion

When collecting data, we choose a sample size suitable for the purpose of the study. Often times this means choosing a sample size large enough that the **margin of error** – which is the part we add and subtract from the point estimate in a confidence interval – is sufficiently small that the sample is useful. For example, our task might be to find a sample size n so that the sample proportion is within ± 0.04 of the actual proportion in a 95% confidence interval.

EXAMPLE 6.7

A university newspaper is conducting a survey to determine what fraction of students support a $200 per year increase in fees to pay for a new football stadium. How big of a sample is required to ensure the margin of error is smaller than 0.04 using a 95% confidence level?

E

The margin of error for a sample proportion is

$$z^\star \sqrt{\frac{p(1-p)}{n}}$$

Our goal is to find the smallest sample size n so that this margin of error is smaller than 0.04. For a 95% confidence level, the value $z^\star$ corresponds to 1.96:

$$1.96 \times \sqrt{\frac{p(1-p)}{n}} < 0.04$$

There are two unknowns in the equation: p and n. If we have an estimate of p, perhaps from a prior survey, we could enter in that value and solve for n. If we have no such estimate, we must use some other value for p. It turns out that the margin of error is largest when p is 0.5, so we typically use this *worst case value* if no estimate of the proportion is available:

$$\begin{aligned}
1.96 \times \sqrt{\frac{0.5(1-0.5)}{n}} &< 0.04 \\
1.96^2 \times \frac{0.5(1-0.5)}{n} &< 0.04^2 \\
1.96^2 \times \frac{0.5(1-0.5)}{0.04^2} &< n \\
600.25 &< n
\end{aligned}$$

We would need over 600.25 participants, which means we need 601 participants or more, to ensure the sample proportion is within 0.04 of the true proportion with 95% confidence.

When an estimate of the proportion is available, we use it in place of the worst case proportion value, 0.5.

GUIDED PRACTICE 6.8

G

A manager is about to oversee the mass production of a new tire model in her factory, and she would like to estimate what proportion of these tires will be rejected through quality control. The quality control team has monitored the last three tire models produced by the factory, failing 1.7% of tires in the first model, 6.2% of the second model, and 1.3% of the third model. The manager would like to examine enough tires to estimate the failure rate of the new tire model to within about 1% with a 90% confidence level. There are three different failure rates to choose from. Perform the sample size computation for each separately, and identify three sample sizes to consider.[4]

EXAMPLE 6.9

E

The sample sizes vary widely in Guided Practice 6.8. Which of the three would you suggest using? What would influence your choice?

We could examine which of the old models is most like the new model, then choose the corresponding sample size. Or if two of the previous estimates are based on small samples while the other is based on a larger sample, we might consider the value corresponding to the larger sample. There are also other reasonable approaches.

Also observe that the success-failure condition would need to be checked in the final sample. For instance, if we sampled $n = 1584$ tires and found a failure rate of 0.5%, the normal approximation would not be reasonable, and we would require more advanced statistical methods for creating the confidence interval.

GUIDED PRACTICE 6.10

G

Suppose we want to continually track the support of payday borrowers for regulation on lenders, where we would conduct a new poll every month. Running such frequent polls is expensive, so we decide a wider margin of error of 5% for each individual survey would be acceptable. Based on the original sample of borrowers where 70% supported some form of regulation, how big should our monthly sample be for a margin of error of 0.05 with 95% confidence?[5]

---

[4]For a 90% confidence interval, $z^\star = 1.65$, and since an estimate of the proportion 0.017 is available, we'll use it in the margin of error formula:

$$1.65 \times \sqrt{\frac{0.017(1-0.017)}{n}} < 0.01 \qquad \rightarrow \qquad \frac{0.017(1-0.017)}{n} < \left(\frac{0.01}{1.65}\right)^2 \qquad \rightarrow \qquad 454.96 < n$$

For sample size calculations, we always round up, so the first tire model suggests 455 tires would be sufficient.

A similar computation can be accomplished using 0.062 and 0.013 for p, and you should verify that using these proportions results in minimum sample sizes of 1584 and 350 tires, respectively.

[5]We complete the same computations as before, except now we use 0.70 instead of 0.5 for p:

$$1.96 \times \sqrt{\frac{p(1-p)}{n}} \approx 1.96 \times \sqrt{\frac{0.70(1-0.70)}{n}} \leq 0.05 \qquad \rightarrow \qquad n \geq 322.7$$

A sample size of 323 or more would be reasonable. (Reminder: always round up for sample size calculations!) Given that we plan to track this poll over time, we also may want to periodically repeat these calculations to ensure that we're being thoughtful in our sample size recommendations in case the baseline rate fluctuates.