Inferences from Two Proportions

It's possible that we don't think that a proportion for a population is equal to a particular value, but it may be possible that we want to confirm that proportions from two populations are in fact the same. (For example, we may want to show that the incidence of male drivers and female drivers causing accidents are in fact the same, even if we don't happen to know what either proportion is.)

We can apply the same concepts of estimating the *difference* between two proportions with a confidence interval, and we can challenge the idea that two proportions are the same with a test of hypotheses. These applications simply need slightly different formulas for the new context.

CONFIDENCE INTERVALS FOR TWO PROPORTIONS

Recall that a confidence interval starts with an estimate of the parameter that we are investigating, and adds and subtracts a margin of error to establish an interval. This margin of error is the product of two numbers: the *critical value* which represents the *confidence level* that we've chosen for the published estimate, multiplied by the *standard error* for the type of estimate we're making. With proportions, we can calculate the standard deviation σ, so we may use a critical z score.

None of this changes with a comparison of two proportions. The main changes are the way we calculate standard error, and of course calculating the parameter itself.

Example 1: Are teens abandoning Facebook for other social media? The Pew Research Centre interviewed teenagers aged 13–17 in 2015 and 2018. In 2018, they interviewed 743 teens, and 51% said they used Facebook. In 2015, 71% of the 1060 teenagers they surveyed said they used Facebook. Express the amount of the decline in a 95% confidence interval.

Solution: The first step is one of those things that's so trivial to do that we often forget to do it, but it's vital. We need to decide which is the "first" statistic and which is the "second" so we can subtract them. Here, we're expressing the decline in usage over time, so we'll let p_1 be the proportion from 2015, and p_2 the proportion from 2018. We have been asked to estimate the population parameter $(p_1 - p_2)$.

The estimate of the parameter is just the difference between the observed proportions: $\hat{p}_1 = 0.71$ and $\hat{p}_2 = 0.51$, so $(\hat{p}_1 - \hat{p}_2) = 0.20$.

Next we need the critical z score. On Table A-2, it tells us the critical z score for a 95% confidence interval, $z_{0.025}$, is 1.960.

Last we need the standard error. This formula has changed:

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = \sqrt{\frac{0.71 \cdot 0.29}{1060} + \frac{0.51 \cdot 0.49}{743}} = 0.023034\ldots$$

From this information we can compile the confidence interval. Multiply the standard error by the critical z score to get margin of error: $0.023034 \times 1.960 = 0.045147\ldots$.

Add and subtract this number from the estimate to get the boundaries of the confidence interval, and always round to three decimal places:

$0.20 - 0.045147\ldots = 0.155$

$0.20 + 0.045147\ldots = 0.245$

$\therefore 0.155 < (p_1 - p_2) < 0.245$

or $(p_1 - p_2) = 0.200 \pm 0.045$

TESTS OF HYPOTHESES FOR TWO PROPORTIONS

Recall that we test a claim about parameters by writing null and alternative hypotheses, conducting a study (or two) that attempts to draw the same conclusion that the claim does, assuming the null hypothesis is true, and finding the probability that we get the results we got under that assumption. If the probability tells us the study is out of line with the null hypothesis, we consider the study to be evidence to reject it.

For the tests you do in this unit, the claim will always take the same form, i.e. that two population proportions are the same, or we hope they're different and assume they're the same. The "same" statement then is the null hypothesis, and the test is two-tailed. The assumption that the null hypothesis is true has an extra consequence — the best estimate for the overall sample proportion throws all the observations into the same pot. (If we claim that the proportion of redheads in Canada is equal to that of the U.S., then there's no reason to break down the study by country — take all respondents from either country as n and calculate p, the proportion of redheads from there.) This proportion is called the **pooled sample proportion**, with the symbol $\bar{p}$. We use $\bar{p}$ in calculating the z score for an observed difference in proportions.

Example 2: No one is as happy as their friends are on social media. This is the "Happiness Paradox", and some sociologists wonder if it's one reason behind feelings of loneliness and depression among social media users. A group of researchers analyzed data from Twitter consisting of the tweets of several thousand users in a single connected network of people who followed each other and interacted with each other on the site — friends, in other words. They analyzed the word choices in the tweets to measure the users' overall mood while using the site. Then each user had their own happiness score compared to the mean score of their friends in the network. If the friends' mean score was higher, then that user experienced the Happiness Paradox.

An interesting result of the study is that if personal happiness is graphed along the horizontal axis, and mean friend happiness on the vertical axis, 95% of the people in the network collected themselves into two distinct groups: happy people with happy friends ($n_1 = 29{,}033$) and unhappy people with unhappy friends ($n_2 = 8{,}018$). Both groups tended to experience the Happiness Paradox ($\hat{p}_1 = 0.577$, $\hat{p}_2 = 0.655$). Test the claim that these groups experience the Happiness Paradox with equal probability, $\alpha = 0.10$.

Solution: The claim is that $p_1 = p_2$, so the claim serves as our null hypothesis. The alternative hypothesis is that $p_1 \neq p_2$. In performing the test of hypotheses, we'll assume the null hypothesis is true: in words, that it doesn't matter if you're in the happy group or the unhappy group, your probability of experiencing the Paradox is the same. In that case, we need to calculate the over all value of p for these people. We need to know the actual numbers of people that experience the Paradox to do that:

$x_1 = 0.577 \cdot 29033 \approx 16{,}752$
$x_2 = 0.655 \cdot 8018 \approx 5{,}252$

overall $x = 16{,}752 + 5{,}252 = 22{,}004$
overall $n = 29{,}033 + 8{,}018 = 37{,}046$
$\therefore \bar{p} = 22{,}004 / 37{,}046 = 0.594$

We use this value of $\bar{p}$ (and its companion value, $\bar{q} = 0.406$) in calculating the standard error in finding the z score associated with getting the separation in proportions we observed between the two groups:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{(0.577 - 0.655) - 0}{\sqrt{\frac{0.594 \cdot 0.406}{29{,}033} + \frac{0.594 \cdot 0.406}{8{,}018}}} = -12.590\ldots$$

At this point we would ordinarily look that z score up in Table A-2, but that number is obviously off the scale; the table says we should be using 0.0001, and then doubling it to 0.0002 for the two-tailed test. (Technology tells us the P-value is actually so close to 0 that it can't give a reasonable decimal equivalent.) Our P-value is way under α, so we reject the null hypothesis. This study constitutes evidence to warrant rejection of the claim that the happy and unhappy Twitter users are equally likely to experience the Happiness Paradox.

Sources: https://www.pewresearch.org/wp-content/uploads/sites/9/2015/04/PI_TeensandTech_Update2015_0409151.pdf
https://www.pewinternet.org/wp-content/uploads/sites/9/2018/05/PI_2018.05.31_TeensTech_FINAL.pdf
https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-017-0100-1

EXERCISES

A. Related to the Happiness Paradox is the Friendship Paradox, which states that your friends on social media have, on average, more friends than you do, and that this is true of everyone, including your friends. Once again we can divide the survey of Twitter users into the happy group and the unhappy group. Among happy users, 95.8% of them suffer from the Friendship Paradox according to the study; among unhappy users, 88.8% do.

1) Express the disparity between the groups with a 99% confidence interval.

2) The paradox is based on comparing the number of friends each user has with the mean number of friends taken from that user's friends. Given this method, do you have a theory to explain the paradox?

B. A study into recidivism (continuing to commit crimes after being released from jail) in the UK found that out of 1,435 prisoners surveyed, 48% of them reported feeling worried or confused when they entered prison the first time. Among those who reported these feelings, 39% were convicted of another crime within a year, compared to 57% of prisoners who were convicted again within a year among those who did not report feeling worried or confused.
Source: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491119/re-offending-release-waves-1-3-spcr-findings.pdf

1) Approximately how many prisoners reported worry or confusion? Approximately how many didn't?

2) Give a 90% confidence interval for the rate of increase of recidivism for prisoners who did not report feelings of worry or confusion on entering prison the first time.

3) Use this study to test the claim that feelings of worry or confusion have no impact on recidivism with a 0.10 significance level. What P-value results from the study, and what conclusion would you write in terms of this claim?

C. Are migraines and obesity linked? A study in Denmark took a sample of 51 032 hospital patients who had migraines, and generated a second sample of 510 320 patients who did not report having migraines, and were demographically similar to the first sample. (Since 29.4% of the patients with migraines were men, 29.4% of the control group were men, and so on.) Among the patients who reported migraines, 1639 were obese. Among those who didn't, 12 821 patients were obese.

1) Test the claim that incidence of migraines and obesity are dependent at the 0.10 significance level. Write a sentence expressing the results of your test in the context of the previous paragraph.

2) State $\hat{p}_1$ and $\hat{p}_2$ and comment on the declaration of significance or non-significance in (1).

Source: https://www.bmj.com/content/360/bmj.k96

SOLUTIONS

A: (1) std. err. = 0.0037135…; $z_{0.005}$ = 2.575; CI = (0.060, 0.080)

(2) Means are vulnerable to outliers. Many people follow a few accounts with abnormally high numbers of friends (celebrities, influencers, companies) and these outliers skew the results. The paradoxes might go away if a more reasonable measure of centre, like a median, were used instead.

B: (1) Approximately 689 did report; approximately 746 did not.

(2) std. err. = 0.025958…; $z_{0.05}$ = 1.645; (0.137, 0.223)

(3) $\bar{p}$ = 0.4836; z = 6.75; P = 0.0002 (or 0 with tech);

This study constitutes evidence to warrant rejection of the claim that feelings of worry or confusion have no impact on recidivism.

C: (1) $\bar{p}$ = 0.025764…; z = 9.51; This study constitutes evidence to support the claim that migraines and obesity are dependent.

(2) $\hat{p}_1$ = 1639/51032 = 0.032117…; $\hat{p}_2$ = 12 821/510 320 = 0.025123…;

This study demonstrates statistical significance, but it's hard to see what the link would be good for — the percentages are so low that treating obesity doesn't seem like a reasonable way to treat migraines — and the two percentages are quite close (less than one percentage point apart). The connection has no practical significance.