Hypothesis Testing

- Identify the research question (Do female ducks prefer green to plain bread?)
- Identify a quantity related to the research question whose value we don't know ('**parameter**').
- Writing the statistical hypotheses in terms of that parameter of interest. H_0: p=0.5 and H_a: p>0.5.
- Collect data and calculate a statistic (Number of ducks that preferred the green bread)
- Find the distribution of the statistic under the null hypothesis (Binomial)
- Find the p-value (probability that the result we got or a more extreme one happens just by chance given that the null hypothesis is true).
- Decide if the p-value is small or large
- Reject if p-value is lower than the significance threshold a.

Today: Inference for proportions

Practice

The government wants to introduce a new law, and wants to know if it is popular among voters.

A simple random sample of 670 voters was surveyed to understand their opinions.

Parameter and point estimate

We would like to estimate the proportion of all voters who support the law.

What are the parameter of interest and the point estimate?

Parameter of interest: proportion of *all* voters who support the law

p a population proportion

Point estimate: proportion of *sampled* voters who support the law

$\hat{p}$ a sample proportion

Inference on a proportion

What percent of all voters support the new law?

We can answer this research question using a confidence interval, which we know is always of the form

point estimate ± ME

And we also know that *ME = critical value times the SE* of the point estimate.

$$SE_{\hat{p}} = ?$$

Standard error of a sample proportion

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

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

- But of course this is true only under certain conditions

Sample proportions are also nearly normally distributed

Central limit theorem for proportions

Sample proportions will be nearly normally distributed with mean equal to the population mean, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

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

- But of course this is true only under certain conditions… any guesses?

independent observations: <10% of the total population for sampling without replacement

sample is very large or distribution is not very skewed: at least 10 successes and 10 failures ($n\hat{p}$, $n(1-\hat{p})>10$)

Note: Since *p* is unknown, we use p̂ in the calculation of the standard error.

Back to experimental design...

The survey found that 571 out of 670 (85%) of voters answered that they support the law. Estimate (using a 95% confidence interval) the proportion of all voters who support the law

Given: $n = 670$, $\hat{p} = 0.85$. First check conditions.

1. *Independence*: The sample is random, and 670 < 10% of all Americans, therefore we can assume that one respondent's response is independent of another.
2. *Success-failure*: 571 people answered correctly (successes) and 99 answered incorrectly (failures), both are greater than 10.

Practice

We are given that n = 670, $\hat{p}$ = 0.85, we also just learned that the standard error of the sample proportion is

$$SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$$

Which of the below is the correct calculation of the 95% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Practice

We are given that n = 670, $\hat{p}$ = 0.85, we also just learned that the standard error of the sample proportion is

$$SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$$

Which of the below is the correct calculation of the 95\% confidence interval?

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} \rightarrow (0.82, 0.88)$

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

Practice

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey is ±3%. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween." At 95% confidence level, is this news piece's statement justified?

(a) Yes
(b) No
(c) Can't tell

Practice

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey is ±3%. A news piece on this study's findings states: "More than 10% of all Americans have objections on religious grounds to celebrating Halloween." At 95% confidence level, is this news piece's statement justified?

(a) Yes
(b) No
(c) Can’t tell

Choosing a sample size

How many people should you sample in order to cut the margin of error of a 95% confidence interval down to 1%.

$$ME = z^{\star} \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}}$$ → *Use estimate for $\hat{p}$ from previous study*

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

$$n \geq 4898.04$$ → *n should be at least 4,899*

What if there isn't a previous study?

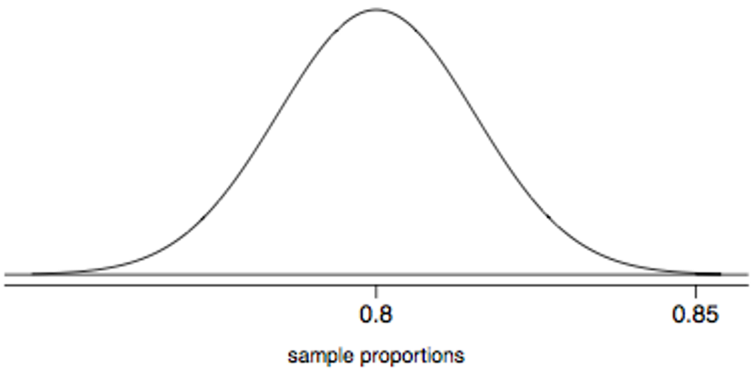
... use $\hat{p} = 0.5$

why?

- if you don't know any better, 50-50 is a good guess
- $\hat{p} = 0.5$ gives the most conservative estimate -- highest possible sample size

Hypothesis Testing for proportions

The survey found that 571 out of 670 (85%) of Americans support the law. Do these data provide convincing evidence that more than 80% of Americans support the law?



Hypothesis Testing

- Identify the research question (more than 80% of Americans support the law)
- Identify a quantity related to the research question whose value we don't know ('**parameter**').
- Writing the statistical hypotheses in terms of that parameter of interest. H_0: $p = p_0$ and H_a: $p > p_0$.
- Collect data and calculate a statistic Z-score: $\frac{\hat{p}-p}{SE_p}$
- Find the distribution of the statistic under the null hypothesis

  Under the null, $p = p_o$ so Z-score: $\frac{\hat{p}-.8}{SE_{p=0.8}}$
- Find the p-value (probability that the result we got or a more extreme one happens just by chance given that the null hypothesis is true).
- Decide if the p-value is small or large
- Reject if p-value is lower than the significance threshold a.

CI vs. HT for proportions

Success-failure condition:

- CI: At least 10 *observed* successes and failures
- HT: At least 10 *expected* successes and failures, calculated using the null value

Standard error:

- CI: calculate using observed sample proportion:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- HT: calculate using the null value:

$$SE = \sqrt{\frac{p_o(1-p_o)}{n}}$$

Practice

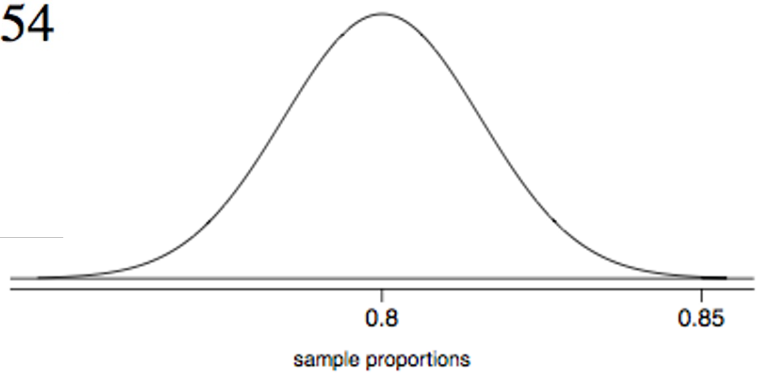
The survey found that 571 out of 670 (85%) of Americans support the law. Do these data provide convincing evidence that more than 80% of Americans support the law?

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$SE = \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154$$

$$Z = \frac{0.85 - 0.80}{0.0154} = 3.25$$

$$p - value = 1 - 0.9994 = 0.0006$$



Since the p-value is low, we reject H_0. The data provide convincing evidence that more than 80% of Americans support the law

Recap - inference for one proportion

Population parameter: p, point estimate: $\hat{p}$

Recap - inference for one proportion

Population parameter: p, point estimate: $\hat{p}$

Conditions

- independence
  - random sample and 10% condition
- at least 10 successes and failures

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$

- for CI: use $\hat{p}$
- for HT: use p_0

Difference of Two Proportions

Melting ice cap

Scientists predict that global warming may have big effects on the polar regions within the next 100 years. One of the possible effects is that the northern ice cap may completely melt. Would this bother you a great deal, some, a little, or not at all if it actually happened?

(a) A great deal
(b) Some
(c) A little
(d) Not at all

Results from the General Social Survey

The GSS asks the same question, below are the distributions of responses from the 2010 GSS as well as from a group of introductory statistics students at Duke University:

	GSS	Duke
A great deal	454	69
Some	124	30
A little	52	4
Not at all	50	2
Total	680	105

Difference between the proportions of *all* Duke students and *all* Americans who would be bothered a great deal by the northern ice cap completely melting.

Parameter and point estimate

- *Parameter of interest*: Difference between the proportions of *all* Duke students and *all* Americans who would be bothered a great deal by the northern ice cap completely melting.

$$p_{Duke} - p_{US}$$

- *Point estimate*: Difference between the proportions of *sampled* Duke students and *sampled* Americans who would be bothered a great deal by the northern ice cap completely melting.

$$\hat{p}_{Duke} - \hat{p}_{US}$$

Inference for comparing proportions

- The details are the same as before…
- CI: *point estimate ± margin of error*
- HT: Use *Z = (point estimate - null value) / SE* to find appropriate p-value.
- We just need the appropriate standard error of the point estimate $SE_{\hat{p}_{duke} - \hat{p}_{US}}$, which is the only new concept.

Standard error of the difference between two sample proportions

$$SE_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Conditions for CI for difference of proportions

1. *Independence within groups:*
   - The US group is sampled randomly and we're assuming that the Duke group represents a random sample as well.
   - $n_{Duke} < 10\%$ of all Duke students and $680 < 10\%$ of all Americans.

   We can assume that the attitudes of Duke students in the sample are independent of each other, and attitudes of US residents in the sample are independent of each other as well.

2. *Independence between groups:*
   The sampled Duke students and the US residents are independent of each other.
3. *Success-failure:*
   At least 10 observed successes and 10 observed failures in the two groups.

Sample proportions are also nearly normally distributed

Construct a 95% confidence interval for the difference between the proportions of Duke students and Americans who would be bothered a great deal by the melting of the northern ice cap $(p_{Duke} - p_{US})$.

Data	Duke	US
A great deal	69	454
Not a great deal	36	226
Total	105	680
$\hat{p}$	0.657	0.668

$$(\hat{p}_{\text{Duke}} - \hat{p}_{US}) \pm z^{\star} \times \sqrt{\frac{\hat{p}_{\text{Duke}}(1-\hat{p}_{\text{Duke}})}{n_{\text{Duke}}} + \frac{\hat{p}_{US}(1-\hat{p}_{US})}{n_{US}}} =$$

$$= \quad (0.657 - 0.668) \pm 1.96 \times \sqrt{\frac{0.657 \times 0.343}{105} + \frac{0.668 \times 0.332}{680}}$$

$$= \quad -0.011 \pm 1.96 \times 0.0497$$

$$= \quad -0.011 \pm 0.097 = \quad (-0.108, 0.086)$$

Practice

Which of the following is the correct set of hypotheses for testing if the proportion of all Duke students who would be bothered a great deal by the melting of the northern ice cap differs from the proportion of all Americans who do?

(a) $H_0 : p_{Duke} = p_{US}$
$H_A : p_{Duke} \neq p_{US}$

(b) $H_0 : \hat{p}_{Duke} = \hat{p}_{US}$
$H_A : \hat{p}_{Duke} \neq \hat{p}_{US}$

(c) $H_0 : p_{Duke} - p_{US} = 0$
$H_A : p_{Duke} - p_{US} \neq 0$

(d) $H_0 : p_{Duke} = p_{US}$
$H_A : p_{Duke} < p_{US}$

Practice

Which of the following is the correct set of hypotheses for testing if the proportion of all Duke students who would be bothered a great deal by the melting of the northern ice cap differs from the proportion of all Americans who do?

(a) $H_0 : p_{Duke} = p_{US}$
$H_A : p_{Duke} \neq p_{US}$

(b) $H_0 : \hat{p}_{Duke} = \hat{p}_{US}$
$H_A : \hat{p}_{Duke} \neq \hat{p}_{US}$

(c) $H_0 : p_{Duke} - p_{US} = 0$
$H_A : p_{Duke} - p_{US} \neq 0$

(d) $H_0 : p_{Duke} = p_{US}$
$H_A : p_{Duke} < p_{US}$

Both (a) and (c) are correct.

Flashback to working with one proportion

- When constructing a confidence interval for a population proportion, we check if the *observed* number of successes and failures are at least 10.

$$n\hat{p} \geq 10 \qquad\qquad n * (1 - \hat{p}) \geq 10$$

- When conducting a hypothesis test for a population proportion, we check if the *expected* number of successes and failures are at least 10.

$$np_0 \geq 10 \qquad n * (1 - p_0) \geq 10$$

Pooled estimate of a proportion

- In the case of comparing two proportions where H_0: $p_1 = p_2$, there isn't a given null value we can use to calculate the *expected* number of successes and failures in each sample.
- Therefore, we need to first find a common (*pooled*) proportion for the two groups, and use that in our analysis.
- This simply means finding the proportion of total successes among the total number of observations.

Pooled estimate of a proportion

$$\hat{p} = \frac{\#\ of\ successes_1 + \#\ of\ successes_2}{n_1 + n_2}$$

Practice

Calculate the estimated pooled proportion of Duke students and Americans who would be bothered a great deal by the melting of the northern ice cap. Which sample proportion ($\hat{p}_{Duke}$ *or* $\hat{p}_{US}$) the pooled estimate is closer to? Why?

Data	Duke	US
A great deal	69	454
Not a great deal	36	226
Total	105	680
$\hat{p}$	0.657	0.668

$$\begin{aligned}\hat{p} &= \frac{\#\ of\ successes_1 + \#\ of\ successes_2}{n_1 + n_2} \\ &= \frac{69 + 454}{105 + 680} = \frac{523}{785} = 0.666\end{aligned}$$

CI vs. HT for proportions

Do these data suggest that the proportion of all Duke students who would be bothered a great deal by the melting of the northern ice cap differs from the proportion of all Americans who do? Calculate the test statistic, the p-value, and interpret your conclusion in context of the data.

Data	Duke	US
A great deal	69	454
Not a great deal	36	226
Total	105	680
$\hat{p}$	0.657	0.668

$$Z = \frac{(\hat{p}_{Duke} - \hat{p}_{US})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{Duke}} + \frac{\hat{p}(1-\hat{p})}{n_{US}}}}$$

$$= \frac{(0.657 - 0.668)}{\sqrt{\frac{0.666 \times 0.334}{105} + \frac{0.666 \times 0.334}{680}}} = \frac{-0.011}{0.0495} = -0.22$$

$$p - value = 2 \times P(Z < -0.22) = 2 \times 0.41 = 0.82$$

Recap - comparing two proportions

- Population parameter: $(p_1 - p_2)$, point estimate: $(\hat{p}_1 - \hat{p}_2)$
- Conditions:

Recap - comparing two proportions

- Population parameter: $(p_1 - p_2)$, point estimate: $(\hat{p}_1 - \hat{p}_2)$

- Conditions:
  - independence within groups
    - random sample and 10% condition met for both groups
  - independence between groups
  - at least 10 successes and failures in each group

Recap - comparing two proportions

- Population parameter: $(p_1 - p_2)$, point estimate: $(\hat{p}_1 - \hat{p}_2)$

- Conditions:
  - independence within groups
    - random sample and 10% condition met for both groups
  - independence between groups
  - at least 10 successes and failures in each group

$$SE_{(\hat{p}_1-\hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

- for CI: use $\hat{p}_1$ and $\hat{p}_2$
- for HT:
  - when H_0: $p_1 = p_2$: us $\hat{p}_{pool} = \frac{\#\ suc_1 + \#suc_2}{n_1 + n_2}$
  - when H_0: $p_1 - p_2$ = (some value other than 0): use $\hat{p}_1$ and $\hat{p}_2$

Reference - standard error calculations

	one sample	two samples
mean	$SE = \frac{s}{\sqrt{n}}$	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
proportion	$SE = \sqrt{\frac{p(1-p)}{n}}$	$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

- When working with means, it's very rare that σ is known, so we usually use *s*.
- When working with proportions,
  - if doing a hypothesis test, *p* comes from the null hypothesis
  - if constructing a confidence interval, use $\hat{p}$ instead

Practice

A drone company is considering a new manufacturer for rotor blades. The new manufacturer would be more expensive, but they claim their higher-quality blades are more reliable, with 3% more blades passing inspection than their competitor. Set up appropriate hypotheses for the test.

Practice

- Identify the research question

- Identify a quantity related to the research question whose value we don't know ('**parameter**').
- Writing the statistical hypotheses in terms of that parameter of interest.

- Collect data and calculate a statistic

- Find the distribution of the statistic under the null hypothesis

- Find the p-value (probability that the result we got or a more extreme one happens just by chance given that the null hypothesis is true).
- Decide if the p-value is small or large
- Reject if p-value is lower than the significance threshold a.

Practice

- Identify the research question (C2 blades are better than C1 blades)
- Identify a quantity related to the research question whose value we don't know ('**parameter**'). $p_2 - p_1$
- Writing the statistical hypotheses in terms of that parameter of interest. H_0: $p_2 - p_1$=0.03 and H_a: $p_2 - p_1 > 0.03$.
- Collect data and calculate a statistic (Z-score:$\frac{\widehat{p_2}-\widehat{p_1}-(\mathrm{p}_2-\mathrm{p}_1)}{SE_{p_2-p_1}} = \frac{\widehat{p_2}-\widehat{p_1}-0.03}{SE_{p_2-p_1}}$)
- Find the distribution of the statistic under the null hypothesis $N(0,1)$
- Find the p-value (probability that the result we got or a more extreme one happens just by chance given that the null hypothesis is true).
- Decide if the p-value is small or large
- Reject if p-value is lower than the significance threshold a.

Practice

The quality control engineer collects a sample of blades, examining 1000 blades from each company, and she finds that 899 blades pass inspection from the current (C1) supplier and 958 pass inspection from the prospective (C2) supplier.

Find the p-value

Should we change suppliers?