PSY 201: Statistics in Psychology

Lecture 28

Hypothesis testing for independent proportions

What is a "margin of error"?

Greg Francis

Purdue University

Fall 2023

PROPORTIONS

- we want to test hypotheses about proportions of populations

$$H_0 : P_1 = P_2$$

$$H_a : P_1 \neq P_2$$

- or, the same thing is

$$H_0 : P_1 - P_2 = 0$$

$$H_a : P_1 - P_2 \neq 0$$

- (or we could use directional hypotheses)

SAMPLING DISTRIBUTION

- the statistic is the difference between two **independent** sample proportions

$$p_1 - p_2$$

- need to know sampling distribution and standard error
- turns out that for large sample sizes (and some constraints on the proportions), the sampling distribution is approximately normal with a mean equal to the difference of population proportions

$$P_1 - P_2$$

STANDARD ERROR

- the standard error of the difference between independent proportions is

$$s_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

  - where p is the average proportion across the groups

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

  - or an equivalent formula is

$$p = \frac{f_1 + f_2}{n_1 + n_2}$$

  - with f_1 and f_2 being the **frequencies** of occurrences in each sample, respectively.
  - Also

$$q = 1 - p$$

INDEPENDENT SAMPLES

- we now have everything we need to carry out hypothesis testing of proportions for the two-sample case when the samples are independent
- That is we can calculate the test statistic

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{s_{p_1 - p_2}}$$

- when $H_0 : P_1 - P_2 = 0$ this is simply

$$z = \frac{(p_1 - p_2)}{s_{p_1 - p_2}}$$

- and look up a p-value with the normal distribution calculator

CONFIDENCE INTERVALS

- we can create confidence intervals too
- the general formula is

$$CI = \text{statistic} \pm (\text{critical value }) \times (\text{standard error})$$

- for the difference of proportions it becomes

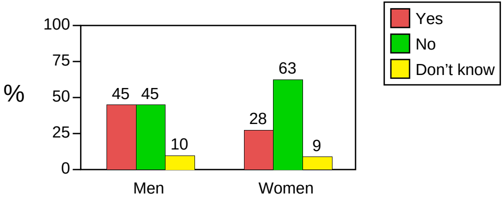
$$CI = (p_1 - p_2) \pm (z_{cv})(s_{p_1 - p_2})$$

AN EXAMPLE

- A Gallup poll sampled 1005 adults and asked, "Are you a fan of college football, or not?"

Men much more likely than women to be college football fans.

± 3 % Margin of Error
October 21-24, 1999
Sample Size=1,005



FOOTBALL

- Is the sex difference a real difference between populations?
- (1) State the hypothesis.

$$H_0 : P_1 - P_2 = 0$$

$$H_a : P_1 - P_2 \neq 0$$

- Set the criterion
  - we'll use $\alpha = 0.01$

FOOTBALL

- (2) Find the test statistic
- We are interested in the proportion of people who answer "yes" to the question.
- We really need to know n_1 and n_2, but we do not have that information. We'll assume $n_1 = 502$ and $n_2 = 503$, for males and females, respectively.
- We need p, the average proportion across the groups,

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(502)(0.45) + (503)(0.28)}{1005} = 0.364$$

- and q

$$q = 1 - p = 1 - 0.364 = 0.636$$

FOOTBALL

- we use p and q to get standard error

$$s_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$= \sqrt{(0.364)(0.636)\left(\frac{1}{502}+\frac{1}{503}\right)} = 0.03035$$

- and can now compute the test statistic

$$z = \frac{(p_1-p_2)-(P_1-P_2)}{s_{p_1-p_2}}$$

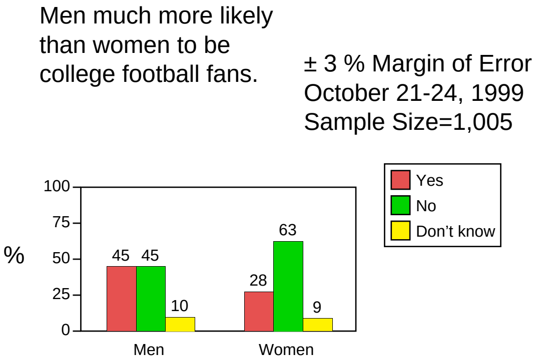
$$z = \frac{(0.45-0.28)-0}{0.03035} = 5.6013$$

- (3) Find the p-value from the normal distribution calculator

$$p \approx 0$$

FOOTBALL

- (4) Make a decision
- Reject H_0. The samples suggest that men and women have different proportions of being fans of college football



ONLINE CALCULATOR

- Small rounding differences

Enter data:

Sample size for group 1: n_1 = 502

Number of scores with trait for group 1: f_1 = 226

Sample size for group 2: n_2 = 503

Number of scores with trait for group 2: f_2 = 141

Specify hypotheses:

$H_0 : P_1 - P_2 =$ 0

H_a : Two-tails

$\alpha =$ 0.01

Run Test

Test summary	
Null hypothesis	$H_0 : P_1 - P_2 = 0$
Alternative hypothesis	$H_a : P_1 - P_2 \neq 0$
Type I error rate	$\alpha = 0.01$
Sample size for group 1	$n_1 = 502$
Sample size for group 2	$n_2 = 503$
Sample proportion for group 1	$p_1 = 0.4502$
Sample proportion for group 2	$p_2 = 0.2803$
Pooled proportion	$p = 0.3652$
Standard error	$s_{p_1 - p_2} = 0.030376$
Test statistic	$z = 5.592688$
p value	$p = 0.000000$
Decision	Reject the null hypothesis
Confidence interval critical value	$z_{cv} = 2.576236$
Confidence interval	$CI_{99} = (0.091626, 0.248136)$

FOOTBALL

- note, we take a sample of 1005 people, and we draw conclusions about *everyone* in the US
- that is remarkable, and it works because we know the properties of the sampling distribution
- of course, our conclusion could be wrong. There is a chance, $\alpha < 0.01$, that even if H_0 were true that we would get a difference of sample proportions like this.

FOOTBALL

- what's that margin of error about?
- You see in lots of polls that there is a "margin of error of $\pm 3\%$" (or $\pm 5\%$,...)
- It's the range of a confidence interval

$$CI = (p_1 - p_2) \pm (z_{cv})(s_{p_1 - p_2})$$

$$CI_{99} = (0.45 - 0.28) \pm (2.576)(0.03035)$$

$$CI_{99} = (0.45 - 0.28) \pm 0.07818$$

- going 0.08 (or 8%) above and below the difference of the sample proportions
- where does 3% come from?

FOOTBALL

- Try building a CI_{95}

$$CI = (p_1 - p_2) \pm (z_{cv})(s_{p_1-p_2})$$

$$CI_{95} = (0.45 - 0.28) \pm (1.96)(0.03035)$$

$$CI_{90} = (0.45 - 0.28) \pm 0.059486$$

- going 0.06 (or 6%) above and below the difference of the sample proportions
- where does 3% come from?

MARGIN OF ERROR

- In this particular case, the Gallup organization seems to be reporting the $\pm$ range of a 95% confidence interval for the proportion of the entire set of data, under the worst case (when s_p is as big as it possibly could be)

$$CI = p \pm (z_{cv})(s_p)$$

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

- is biggest for $p = 0.5$
- with $n = 1005$

$$s_p = \sqrt{\frac{0.5(0.5)}{1005}} = 0.01577$$

- for a 95% CI,

$$z_{cv} = 1.96$$

- so

$$CI_{95} = p \pm (1.96)(0.01577) = p \pm 0.03091$$

MARGIN OF ERROR

- Journalists often refer to a margin of error, but they often do not really explain how it is calculated
- It *does* give an indication of how much variability is in the data
- If described properly, it could give us some information about the confidence of the estimate. But the confidence is almost never given
- I can make a CI (and the margin of error) big or small by varying my desired confidence.
- Moreover, remember, margin of error is not absolute, but only with confidence! (Sometimes the CI is wrong, and the "margin of error" is much bigger than the CI indicates.)

POWER

- This study is rather old (1999). You might decide to check whether there are similar results today. To design a new experiment, you can use the previous data as estimates of the population proportions.

Specify the population characteristics:

$H_0 : P_1 - P_2 =$ 0

$H_a : P_{a1} =$ 0.45 $\quad P_{a2} =$ 0.28

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.01

Power= 0.9 — Calculate minimum sample size

Sample size for group 1, $n_1 =$ 239

Sample size for group 2, $n_2 =$ 239 — Calculate power

POWER

- You have to think carefully about what you are going to compare. For example, rather than repeat the same experiment for people in 2018, you might be interested in checking whether the proportion of women who like college football has changed since 1999:

$$H_0 : P_1 - P_2 = 0$$

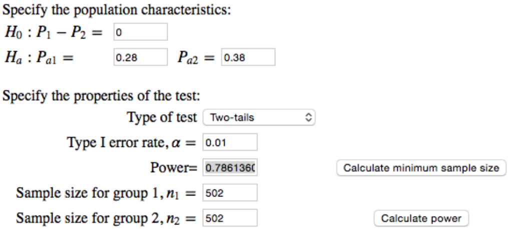
- That would be a different experiment, and thus it requires a different power analysis.

POWER

- For example, maybe an advertiser is willing to reconsider placing ads targeted to women during college football games, provided the proportion has increased by at least 0.1 since 1999.
- Then, our specific values for the alternative hypothesis are:

$$H_a : P_{a1} = 0.28, P_{a2} = 0.38$$

- If we use the same sample sizes as the 1999 study ($n_1 = 502$, $n_2 = 502$)



POWER

- To have 90% power for this study, we can find the minimum sample size

Specify the population characteristics:

$H_0 : P_1 - P_2 =$ 0

$H_a : P_{a1} =$ 0.28 $P_{a2} =$ 0.38

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.01

Power= 0.9 Calculate minimum sample size

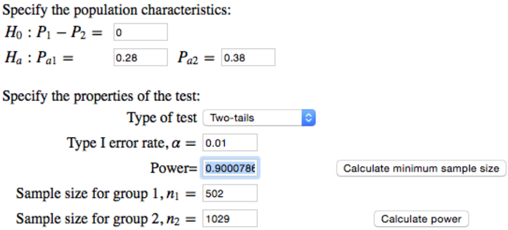
Sample size for group 1, $n_1 =$ 659

Sample size for group 2, $n_2 =$ 659 Calculate power

- but this does not help us much, as the sample from 1999 is fixed at $n_1 = 502$.

POWER

- We can fix n_1=502 and increase the n_2 value until we get a power of 0.9



- It takes a quite large sample to detect a 0.1 difference in proportions!

CONCLUSIONS

- two-sample case
- independent proportions
- confidence interval
- power

NEXT TIME

- two-sample case
- dependent proportions
- confidence interval
- power (tricky)

What do people think about death?