Inference for Two Proportions

Stat 120

May 08 2023

One Proportion or Two?

You want to...

... compare the proportion of students who use a Windows-based PC to the proportion who use a Mac.

a) Inference for one proportion

b) Inference for two proportions

▶ Click for answer

One Proportion or Two?

You want to...

... compare the proportion of students who study abroad between those attending public universities and those at private universities.

a) Inference for one proportion

b) Inference for two proportions

▶ Click for answer

One Proportion or Two?

You want to...

... compare the proportion of in-state students at a university to the proportion from outside the state.

a) Inference for one proportion

b) Inference for two proportions

▶ Click for answer

One Proportion or Two?

You want to...

... compare the proportion of in-state students who get financial aid to the proportion of out-of-state students who get financial aid

This is...

a) Inference for one proportion

b) Inference for two proportions

▶ Click for answer

Tagging Penguins

Are metal tags detrimental to penguins? ***A study looked at the 10 year survival rate of penguins tagged either with a metal tag or an electronic tag. 20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins.***

Is there a statistically significant difference in survival rates?

$$H_0 : p_M = p_E \quad H_A : p_M \neq p_E$$

$p =$ true survival rate

Source: Saraux, et. al. (2011). “Reliability of flipper-banded penguins as

Tagging Penguins

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins.

	Survived	Died	Total
Metal Tag	33	134	167
Electronic Tag	68	121	189
Total	101	255	356

Are the conditions met for using the normal distribution for inference?

a). Yes

b). No

▶ Click for answer

Pooled Proportion

We don't know p_M or p_E, so how do we compute the SE for our hypothesis test?

- *Assume the two proportions are equal and use one proportion for both groups.*
- *Our best guess of this one proportion comes from combining data from both groups and computing the overall proportion, called the pooled proportion p.*
- *Hint: the pooled proportion will always be somewhere in between the two sample proportions.*

Inference Using $\mathrm{N}(0,1)$

If the distribution of the sample statistic is normal: A confidence interval can be calculated by

$$\text{sample statistic } \pm z^* \times SE$$

where z^* ***is a*** $N(0,1)$ ***percentile depending on the level of confidence. A p-value is the area in the tail(s) of a*** $N(0,1)$ ***beyond***

$$z = \frac{\text{sample statistic } - \text{ null value}}{\text{SE}}$$

Test for a Difference in Proportions

$$H_0 : p_1 = p_2$$
$$H_a : p_1 \neq p_2$$

$$\hat{p}_{pooled} = \frac{33+68}{167+189} = 0.2837$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_1} + \frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_2}}}$$

If observed counts in the two-way table are at least 10, then the p-value can be computed as the area in the tail(s) of a standard normal beyond z. Always use pooled proportion for the SE.

Tagging Penguins

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins.

33 survived with metal tags and 68 with electronic The pooled proportion is:

$$\hat{p}_{pooled} = \frac{33+68}{167+189} = 0.2837$$

SE for our test:

$$SE = \sqrt{\frac{0.284(1-.284)}{167} + \frac{0.284(1-.284)}{189}} = 0.048$$

Tagging Penguins

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins. The pooled SE is 0.048.

$$\text{Standardized test stat:} \quad z = \frac{(0.2 - 0.36) - 0}{0.048} = -3.34$$

```r
# Standardized test statistic
((0.2 - 0.36) - 0)/(0.048)
[1] -3.333333
```

```r
# P-value
2*pnorm(-3.34,0,1)
[1] 0.0008377839
```

Reject the null

Tagging Penguins

A difference in survival rates as extreme, or more extreme, than 16% would occur by chance only about 0.08% of the time. There is a statistically significant difference
$(z = -3.34, p = 0.0008)$

How much do the rates differ?

- Compute a 95% CI for the difference...

How do we compute the SE?

- We **can't** use the pooled version since we've concluded the proportions differ!

Confidence Interval for $p_1 - p_2$

```
For large enough n₁ and n₂ :    statistic ± z* × SE
```

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Metal Tags and Penguins

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins. Give a 90% confidence interval for the difference in proportions (metal - electronic).

What is z^* for the confidence interval?

a) 1.280

b) 1.645

c) 1.960

d) 2.575

e) 0.90

Tagging Penguins: 90% C.I

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins. Give a 90% confidence interval for the difference in proportions (metal - electronic).

$90\% CI$ for $\mathrm{p}_M - \mathrm{p}_E$:

$$
\begin{aligned}
(0.2 - 0.36) \pm 1.645 \cdot \sqrt{\frac{0.2(1-0.2)}{167} + \frac{0.36(1-0.36)}{189}} \\
= -0.16 \pm 1.645 \times 0.047 \\
= (-0.237, -0.09)
\end{aligned}
$$

We are 90% confident that the survival rate is between 9 and 23.7 lower for metal tagged penguins, as opposed to electronically tagged.

Metal Tags and Penguins

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins. Give a 95% confidence interval for the difference in proportions (metal - electronic).

What is z^* for the confidence interval?

a) 1.280

b) 1.645

c) 1.960

d) 2.575

e) 0.90

Tagging Penguins: 95% C.I

20% of the 167 metal tagged penguins survived, compared to 36% of the 189 electronic tagged penguins. Give a 95% confidence interval for the difference in proportions (metal - electronic).

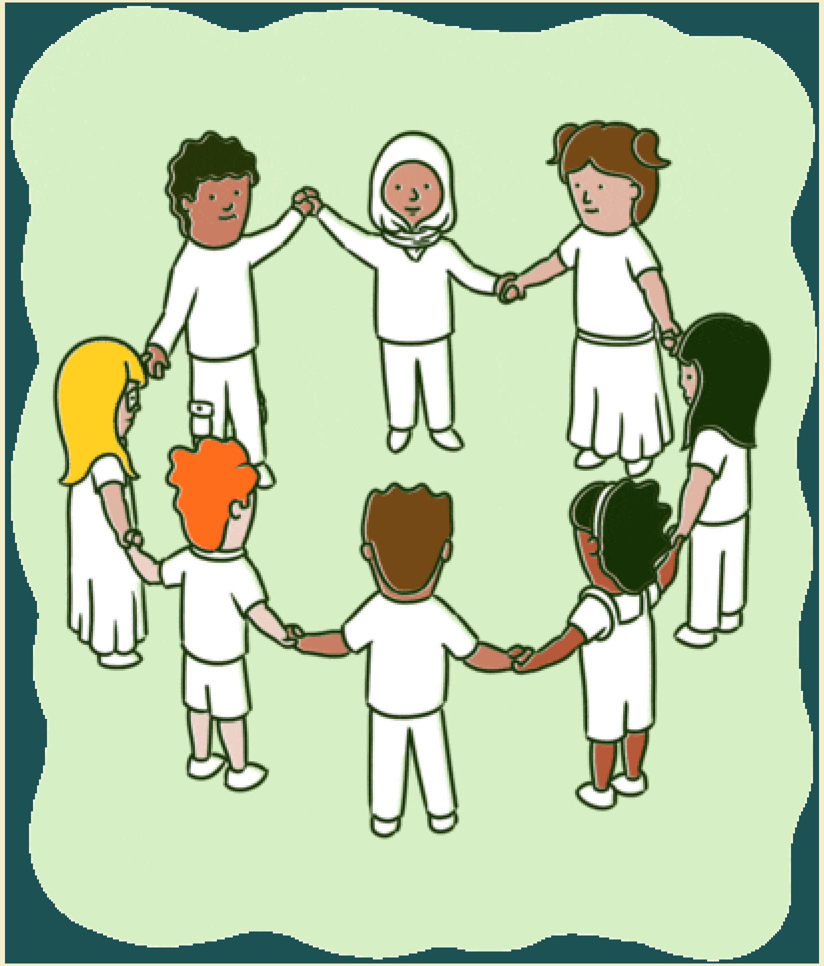
$95\% CI$ for $\mathrm{p}_M - \mathrm{p}_E$:

$$
\begin{aligned}
(0.20-0.36) &\pm 1.96\sqrt{\frac{0.20\times 0.80}{167}+\frac{0.36\times 0.64}{189}} \\
&= -0.16 \pm 1.96 \times 0.047 \\
&= (-0.251, -0.069)
\end{aligned}
$$

We are 95% confident that between 6.9% to 25.1% fewer penguins survive when metal tags are used compared to electronic tags.

YOUR TURN 1

10:00



Let's go over to the class activity .Rmd file and complete the tasks for today.