STAT 113
Theoretical Approximations for Inferences About Differences

Colin Reimer Dawson

Oberlin College

November 14, 2019

Cases to Address

We will need standard errors to do CIs and tests for the following parameters:

1. Single Proportion (Last Time)
2. Single Mean (Just Finished)
3. Difference of Proportions (Today)
4. Difference of Means (Today)
5. Mean of Differences (Next Week)

Example: Penguins Again! But Different!

Penguin Breeding

The scientists who studied whether metal bands were harmful to penguin survival also examined whether they affected the penguins' breeding patterns. For the metal-band group, 39 of 122 penguin-seasons resulted in offspring (32%), whereas the controls had offpsring in 70 of 160 penguin-seasons (44%).

- If we are focused on the effect of metal bands on penguins' reproductive "success", what *population parameter* would we want to focus on?
- What is the corresponding sample statistic?

Outline

Variance and Standard Error of Differences

- The difference between two statistics is affected by both statistics.
- Each statistic differs from its corresponding parameter: it has some **sampling error**.
- If the two sources of sampling error always went in the same direction (i.e., either both statistics were overestimates or both were underestimates), some error would cancel out when we looked at the difference of proportions.
- If they always went in opposite directions, the errors would add.
- But sometimes they go the same way, sometimes opposite ways.
- Because of how standard errors work, if the two samples are unrelated, this is like moving in perpendicular directions.

"Pythagorean Theorem" of Standard Errors

When two sample statistics come from unrelated samples, A and B, the estimation error associated with their difference obeys a "Pythagorean Theorem":

$$SE^2_{\text{difference}} = SE^2_A + SE^2_B$$

In other words

$$SE_{\text{difference}} = \sqrt{SE^2_A + SE^2_B}$$

Standard Error of Difference of Proportions

The standard error for a proportion is:

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

By the "Pythagorean Theorem" of Standard Errors, the SE of a difference of two proportions (from two unrelated samples) is

$$SE_{\hat{p}_A - \hat{p}_B} = \sqrt{SE^2_{\hat{p}_A} + SE^2_{\hat{p}_B}}$$

Expanding this gives us

$$SE_{\hat{p}_A - \hat{p}_B} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

(You don't need to memorize this!)

Analytic Approximations of Sampling Distributions

Param.	Stat.	Theory SE	Distribution	df
p	$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$\mathcal{N}(0,1)$	–
μ	$\bar{x}$	$\sqrt{\frac{s^2}{n}}$	t	$n-1$
$p_A - p_B$	$\hat{p}_A - \hat{p}_B$	$\sqrt{SE^2_{\hat{p}_A} + SE^2_{\hat{p}_B}}$	$\mathcal{N}(0,1)$	–
$\mu_A - \mu_B$	$\bar{x}_A - \bar{x}_B$	$\sqrt{SE^2_{\bar{x}_A} + SE^2_{\bar{x}_B}}$	t	$\min(n_A, n_B) - 1$
μ_{diff}	$\bar{x}_{\text{diff}}$	$\sqrt{\frac{s^2_{\text{diff}}}{n_{\text{diff}}}}$	t	$n_{\text{diff}} - 1$
ρ	r	$\sqrt{\frac{1-r^2}{n-2}}$	t	$n-2$

$$CI: \quad \text{Observed Statistic} \pm \text{Standardized Quantile} \times \widehat{SE}$$

$$\text{Sandardized Test Statistic}: \quad \frac{\text{Observed Statistic} - \text{Null Param.}}{\widehat{SE}}$$

Distribution of $\hat{p}_A - \hat{p}_B$

- **Condition:** The Normal approximation for the distribution of $\hat{p}_A - \hat{p}_B$ holds reasonably well if there are at least 10 expected cases from all four combinations:

$$n_A p_A \geq 10 \qquad n_A(1 - p_A) \geq 10$$
$$n_B p_B \geq 10 \qquad n_B(1 - p_B) \geq 10$$

- Mean of sampling distribution: $p_A - p_B$
- Standard deviation (standard error):

$$SE_{\hat{p}} = \sqrt{\frac{p_A(1 - p_A)}{n_A} + \frac{p_B(1 - p_B)}{n_B}}$$

Outline

Standard Error of a Difference

CI and Test for Difference of Proportions

CI and Test for Difference of Means

CI Summary: Difference of Proportions

0. Check whether conditions for distribution approximation hold: All 4 of $n_A\hat{p}_A$, $n_A(1-\hat{p}_A)$, $n_B\hat{p}_B$ and $n_B(1-\hat{p}_B)$ are at least 10
1. If so, find the mean and SD of the theoretical distribution to replace the bootstrap distribution
   - Mean = Sample Statistic $= \hat{p}_A - \hat{p}_B$
   - SD = Estimated Standard Error $= \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$
2. Use the confidence level and the standardized distribution (Standard Normal) to get z-scores of the endpoints
3. Convert z scores to endpoints on the original scale using the mean and standard deviation found in step 1.

$$\text{Endpoint} = \text{Sample Statistic} + z_{endpoint} \cdot \widehat{SE}$$

Example: Penguins Again! But Different!

Penguin Breeding

The scientists who studied whether metal bands were harmful to penguin survival also examined whether they affected the penguins' breeding patterns. For the metal-band group, 39 of 122 penguin-seasons resulted in offspring (32%), whereas the controls had offpsring in 70 of 160 breeding seasons (44%).

Descriptive Statistics:

$$n_{metal} = \_\_\_\_\_\_ \qquad n_{control} = \_\_\_\_\_\_$$

$$\hat{p}_{breed|metal} = \_\_\_\_\_\_ \qquad \hat{p}_{breed|control} = \_\_\_\_\_\_$$

$$\hat{p}_{breed|metal} - \hat{p}_{breed|control} = \_\_\_\_\_\_$$

Penguins Breed Confidence (Interval)

Let's find a 90% CI for $p_A - p_B$.

$$CI : \text{Sample Stat} \pm z_{endpoint} \cdot \widehat{SE}$$

Conditions: Is the Normal approximation reasonable?

Sample Statistic and Standard Error:

$$\text{Sample Stat} = \underline{\hspace{4cm}}$$

$$\widehat{SE} = \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$$

$$= \underline{\hspace{4cm}}$$

z-scores of Endpoints:

$$z_{endpoint} = \underline{\hspace{4cm}}$$

Interval and Interpretation:

Hypothesis Test: Difference of Two Proportions

- For the hypothesis test, our distribution is going to replace a randomization distribution
- Centered at the null param., with a corresponding standard deviation

$$\widehat{SE}_{\hat{p}_A - \hat{p}_B} = \sqrt{\widehat{SE}^2_{\hat{p}_A} + \widehat{SE}^2_{\hat{p}_B}} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

- What should we plug in for p_A and p_B, according to H_0?

Null Standard Error

$$\widehat{SE}_{\hat{p}_A - \hat{p}_B} = \sqrt{\widehat{SE}^2_{\hat{p}_A} + \widehat{SE}^2_{\hat{p}_B}} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

- Problem: H_0 states what the *difference* is between p_A and p_B, but not what each is individually
- But, H_0 says that p_A and p_B are the *same thing* (whatever their value), so, according to H_0...

$$\widehat{SE}_{\hat{p}_A - \hat{p}_B} = \sqrt{\frac{p_{both}(1-p_{both})}{n_A} + \frac{p_{both}(1-p_{both})}{n_B}}$$

- We can estimate p_{both} by combining the groups and calculating the *combined* sample proportion
- Note: We held this proportion constant already when doing a randomization test.

Example: Penguins Again! But Different!

Penguin Breeding

The scientists who studied whether metal bands were harmful to penguin survival also examined whether they affected the penguins' breeding patterns. For the metal-band group, 39 of 122 penguin-seasons resulted in offspring (32%), whereas the controls had offpsring in 70 of 160 breeding seasons (44%).

What is $\hat{p}_{breed|combined}$?

$$\hat{p}_{breed|combined} = \frac{39+70}{122+160} = \frac{109}{282} = 0.39$$

P-values: Difference of Proportions

0. Check whether conditions for distribution approximation hold: All 4 of $n_A\hat{p}_{combined}$, $n_A(1-\hat{p}_{combined})$, $n_B\hat{p}_{combined}$ and $n_B(1-\hat{p}_{combined})$ are at least 10
1. If so, find the mean and SD of the distribution to replace the randomization distribution
   - Mean = Null Parameter Value $= p_A - p_B$ according to H_0
   - SD = Standard Error $= \sqrt{\hat{p}_{combined}(1-\hat{p}_{combined})(\frac{1}{n_A}+\frac{1}{n_B})}$
2. Convert the observed sample statistic to its z-score (our "test statistic") within this distribution

$$\text{Test Stat.} = \frac{\text{Observed Sample Statistic} - \text{Null Parameter}}{\text{Standard Error}}$$

3. The P-value is the area under the relevant theoretical curve (Standard Normal) past the test statistic

Example: Penguins Again! But Different!

Penguin Breeding

The scientists who studied whether metal bands were harmful to penguin survival also examined whether they affected the penguins' breeding patterns. For the metal-band group, 39 of 122 penguin-seasons resulted in offspring (32%), whereas the controls had offpsring in 70 of 160 breeding seasons (44%).

Descriptive Statistics:

$$n_{metal} = 122 \qquad n_{control} = 160$$
$$\hat{p}_{breed|metal} = 0.32 \qquad \hat{p}_{breed|control} = 0.44$$
$$\hat{p}_{breed|metal} - \hat{p}_{breed|control} = -0.12$$
$$\hat{p}_{breed|combined} = 0.39$$

Conditions: Are they satisfied?

Do metal bands reduce breeding chances?

Standard Error:

$$\widehat{SE} = \sqrt{\hat{p}_{combined}(1 - \hat{p}_{combined})(\frac{1}{n_A} + \frac{1}{n_B})}$$
$$= \underline{\hspace{6cm}}$$

Test Statistic:

$$\text{Test Stat.} = \frac{\text{observed difference} - \text{null difference}}{\widehat{SE}}$$
$$= \underline{\hspace{6cm}}$$

P-value and Conclusion in Context:

P-value = ________

Conclusion in Context? ________________

Outline

Analytic Approximations of Sampling Distributions

Param.	Stat.	Theory SE	Distribution	df
p	$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$\mathcal{N}(0,1)$	–
μ	$\bar{x}$	$\sqrt{\frac{s^2}{n}}$	t	$n-1$
$p_A - p_B$	$\hat{p}_A - \hat{p}_B$	$\sqrt{SE^2_{\hat{p}_A} + SE^2_{\hat{p}_B}}$	$\mathcal{N}(0,1)$	–
$\mu_A - \mu_B$	$\bar{x}_A - \bar{x}_B$	$\sqrt{SE^2_{\hat{x}_A} + SE^2_{\hat{x}_B}}$	t	$\min(n_A, n_B) - 1$
μ_{diff}	$\bar{x}_{\text{diff}}$	$\sqrt{\frac{s^2_{\text{diff}}}{n_{\text{diff}}}}$	t	$n_{\text{diff}} - 1$
ρ	r	$\sqrt{\frac{1-r^2}{n-2}}$	t	$n-2$

$$CI: \quad \text{Observed Statistic} \pm \text{Standardized Quantile} \times \widehat{SE}$$

$$\text{Sandardized Test Statistic}: \quad \frac{\text{Observed Statistic} - \text{Null Param.}}{\widehat{SE}}$$

Distribution of $\bar{x}$ (σs known)

- Two groups in the population with means μ_A and μ_B, and standard deviations σ_A and σ_B
- Conditions: Sampling distribution of $\bar{x}_A - \bar{x}_B$ is Normal if
  - Populations are Normal, or
  - Sample sizes in each group are large enough (can use $n_A, n_B \geq 27$)
- Mean of Sampling Dist.: $\mu_A - \mu_B$
- Standard deviation (standard error):

$$
\begin{aligned}
SE_{\bar{x}_A - \bar{x}_B} &= \sqrt{SE_{\bar{x}_A}^2 + SE_{\bar{x}_B}^2} \\
&= \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}
\end{aligned}
$$

Estimating σ_A and σ_B

- Since we don't know σ_A and σ_B, we use

$$\widehat{SE}_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

- Now, the standardized difference has a t-distribution

$$\frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \sim t_{df}$$

- A conservative rule is to set $df = \min(n_A, n_B) - 1$

CI Summary: Difference of Means

0. Check whether conditions for distribution approximation hold: Both groups have $n \geq 27$
1. If so, find the mean and SD of the theoretical distribution to replace the bootstrap distribution
   - Mean = Sample Statistic $= \bar{x}_A - \bar{x}_B$
   - SD = Estimated Standard Error $= \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$
2. Use the confidence level and the standardized distribution (t-distribution with $\min(n_A, n_B) - 1$ df) to get z-scores of the endpoints
3. Convert z scores to endpoints on the original scale using the mean and standard deviation found in step 1.

$$\text{Endpoint} = \text{Sample Statistic} + z_{endpoint} \cdot \widehat{SE}$$

P-values: Difference of Means

0. Check whether conditions for distribution approximation hold: Both groups have $n \geq 27$
1. If so, find the mean and SD of the distribution to replace the randomization distribution
   - Mean = Null Parameter Value = μ_0
   - SD = Estimated Standard Error = $\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$
2. Convert the observed sample statistic to its z-score (our "test statistic") within this distribution

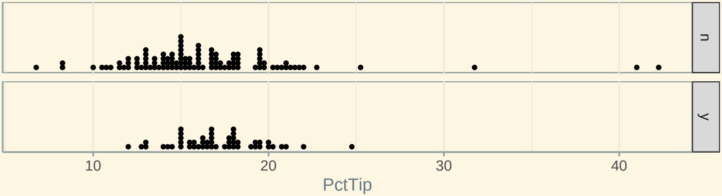
$$\text{Test Stat.} = \frac{\text{Observed Sample Statistic} - \text{Null Parameter}}{\text{Standard Error}}$$

3. The P-value is the area under the relevant theoretical curve (t distribution with $\min(n_A, n_B) - 1 df$) past the test statistic

Example: Credit Card Use and Tip Percentage

Credit Cards and Tip Percentage

We analyze the percent tip left on 157 bills from the First Crush bistro in Northern New York State. The mean percent tip left on the 106 bills paid in cash was 16.39 with a standard deviation of 5.05. The mean percent tip left on the 51 bills paid with a credit card was 17.10 with a standard deviation of 2.47.



Conditions: How are we doing here? Should be ok, since samples are not too small ($n \geq 27$ for both)

Credit Cards and Tips: Descriptive Statistics

Credit Cards and Tip Percentage

We analyze the percent tip left on 157 bills from the First Crush bistro in Northern New York State. The mean percent tip left on the 106 bills paid in cash was 16.39 with a standard deviation of 5.05. The mean percent tip left on the 51 bills paid with a credit card was 17.10 with a standard deviation of 2.47.

$n_{card} = \underline{\qquad}51 \qquad n_{cash} = \underline{\qquad}106$

$\bar{x}_{tip|card} = \underline{\qquad}17.10 \qquad \bar{x}_{tip|cash} = \underline{\qquad}16.39$

$\bar{x}_{tip|card} - \bar{x}_{tip|cash} = \underline{\qquad}0.71$

$s_{tip|card} = \underline{\qquad}2.47 \qquad s_{tip|cash} = \underline{\qquad}5.05$

Credit Cards and Tips: Standard Error

In this case the SE is the same for the CI and the test, so we'll compute it once.

$$SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$
$$= \sqrt{\frac{2.47^2}{51} + \frac{5.05^2}{106}}$$

```
estimatedStandardError <- sqrt(2.47^2/51 + 5.05^2/106)
estimatedStandardError

    [1] 0.6001792
```

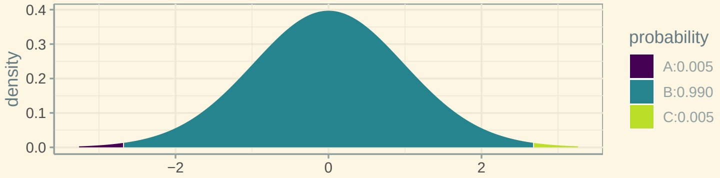
Credit Cards and Tips: Confidence Interval

Let's find a 99% CI for $\mu_{card} - \mu_{cash}$. So far we have

$$Sample\ Stat. = \bar{x}_{tip|card} - \bar{x}_{tip|cash} = 0.71$$
$$\widehat{SE} = 0.60$$

Next we need the standardized endpoints. What distribution?
t with $\min(n_{cash}, n_{card}) - 1$ df. $n_{cash} = 106$, $n_{card} = 51$

```
qdist("t", c(0.005, 0.995), df = 50)
```




```
[1] -2.677793  2.677793
```

Credit Cards and Tips: Confidence Interval

Let's find a 99% CI for $\mu_{card} - \mu_{cash}$. So far we have

$$\text{Sample Stat.} = \bar{x}_{tip|card} - \bar{x}_{tip|cash} = 0.71$$

$$\widehat{SE} = 0.60$$

$$z_{endpoint} = \pm 2.68$$

$$\text{CI: } 0.71 \pm 2.68 \cdot 0.60$$

```
0.71 - 2.68 * 0.60; 0.71 + 2.68 * 0.60

     [1] -0.898
     [1] 2.318
```

Interpretation: We are 99% confident that the mean tip percentage left by customers using a credit card at the First Crush Bistro is somewhere between 0.9 percentage points lower and 2.3 percentage points higher than the mean tip left by customers paying in cash.

Credit Cards and Tips: Hypothesis Test

Do people who pay differently tip differently, on average?

$$\text{Sample Stat.} = \bar{x}_{tip|card} - \bar{x}_{tip|cash} = 0.71$$
$$\widehat{SE} = 0.60$$

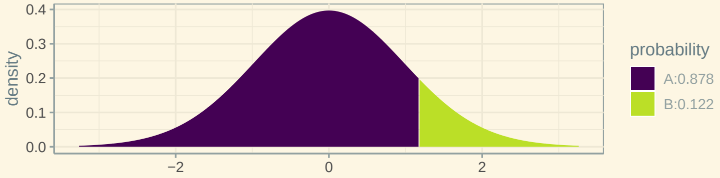
Test Statistic:

$$\text{Test Stat} = \frac{\text{Observed Difference} - \text{Null Difference}}{\widehat{SE}} = \frac{0.71 - 0}{0.60} = 1.18$$

Credit Cards and Tips: Hypothesis Test

P-value:

```
2 * pdist("t", 1.18, df = 50, lower.tail = FALSE)
```




```
[1] 0.2435839
```

Conclusion: We do not have statistically significant evidence ($t = 1.18$, $p = 0.24$, two-tailed) that people paying with a credit card for meals at the First Crush Bistro tip differently, on average, as a percentage of their bill, than people paying with cash.