MA206, Lesson 5 - Strength of Evidence

Review: What are the p-value guidelines for evaluating the **strength of evidence?**.

Weak Evidence against the null	0.1	$< p$
Moderate Evidence against the null	0.05	$< p \leq \quad 0.1$
Strong Evidence against the null	0.01	$< p \leq \quad 0.05$
Very Strong Evidence against the null		$< p \leq \quad 0.01$

Define and describe the **standardized statistic** for a proportion.

$$z = \frac{\hat{p} - mean(null)}{SD(null)}$$

the Standardized Statistic describes how far an observation is from the null hypothesis in terms of standard deviations

What are the standardized test statistic guidelines for evaluating the **strength of evidence**?

Weak Evidence against the null		$\lvert z \rvert$	≤ 1.5
Moderate Evidence against the null	$1.5 <$	$\lvert z \rvert$	≤ 2
Strong Evidence against the null	$2 <$	$\lvert z \rvert$	≤ 3
Very Strong Evidence against the null	$3 <$	$\lvert z \rvert$	

What factors impact the **strength of evidence**?

Distance from the Null
Sample Size
1 Tail vs 2 Tail test

1) Read "Inferences of Competence from Faces Predict Election Outcomes" by Todorov et al in Science (June 2008).

a) What is the author's research question? Why is it important?

Can snap-judgement inferences of competence based on facial appearance predict election results?
Answers may vary; it provides insight into our election process and questions how rational and deliberative they are

b) Identify the observational units in this study

Each individual congressional election

c) Identify the variable. Is the variable quantitative or categorical?

The variable is if the candidate who was deemed more competent was the winner of the election; it is categorical

d) Describe the parameter of interest in words.

The long run proportion of candidates who are deemed more competent that win their respective congressional election

e) List the Null and Alternate hypotheses for this study.

$H_0 : \pi = 0.5$
$H_a : \pi > 0.5$

f) For 2004 races only, list the observed statistic and sample size for both Senate and House races.

$\hat{p}_{Senate} = \frac{22}{32} = 0.688; \quad n_{Senate} = 32$
$\hat{p}_{House} = \frac{189}{279} = 0.677; \quad n_{House} = 279$

g) Which races have a higher proportion of candidates deemed more competent that won their respective elections, the Senate or House elections?

The Senate elections had a higher proportion $(0.688 > 0.677)$

h) Conduct a simulation analysis using the 3S Strategy. Comment on the center, variability and shape of the null distributions generated for Senate and House races.

Senate The Senate is centered at 0.5 with values from 0.28 to 0.75, it is mound shaped and symmetric.

House The House is centered at 0.5 with values from 0.39 to 0.60, it is mound shaped and symmetric

Senate The p-value should be around 0.0290, which is strong evidence against the null hypothesis that the long run proportion of candidates that look more competent win their elections is 50%

House: the p-value should be around (computationally) 0, which is very strong evidence against the null hypothesis that the long run proportion of candidates that look more competent win their elections is 50%

j) Report the mean and standard deviation of the simulated proportions.

Senate: the mean should be about 0.5 with a SD about 0.088

House: The mean should be about 0.5 with a SD about 0.0299

Senate: $z \approx 2.136364$, Strong Evidence against the Null - it is 2.14 standard deviations from the null

House: $z \approx 5.919732$, Very Strong Evidence against the Null - it is 5.9 standard deviations from the null

l) Is the evidence stronger for the Senate 2004 races or the House 2004 races? Why?

The evidence is stronger for the House races - this is because the sample size is significantly larger (279 compared to 32)

Our results **do not prove anything**, but provide strong evidence that the proportion of candidates that appear more competent who win their election is not 50% .

n) How broadly are you willing to generalize your conclusion?

This result is specific to naive voters at Princeton University and cannot be generalized outside of that population based solely on these results. Princeton Students may not be representative of the total voting population to determine elections.

2) An article published in *College Mathematics Journal* (Eyler, Shalla, Doumaux, and McDevitt, 2009) found that players tend to not prefer scissors when playing Rock-Paper-Scissors. You want to test if people really choose scissors less, and conduct a test. You played 120 games and your friend chose scissors 31 times.

a) List the null and alternate hypothesis in words and symbols.

$H_0 : \pi = 0.3333$. The true proportion of times that a person choose scissors in Rock-Paper-Scissors is $\frac{1}{3}$.
$H_a : \pi < 0.333$. The true proportion fo times that a person chooses scissors in Rock-Paper-Scissors is less than $\frac{1}{3}$.

b) Run a simulation using the WileyPlus applet and report the mean and standard deviation of your null distribution.

Mean ≈ 0.333, SD ≈ 0.043.

c) What is the standardized statistic (z) for your test? Comment on the strength of evidence.

The z-score should be close to -1.74, which is moderate evidence against the null hypothesis that scissors is chosen randomly at a rate of $\frac{1}{3}$.
With Mean and SD above, $z = \frac{\frac{31}{120} - 0.333}{0.043} = \frac{0.258333 - 0.33333}{0.043} = -1.74$

d) If you repeated the test another 240 times and your friend chose scissors the same proportion of times ($\hat{p} = 0.258333$), would you expect your strength of evidence to increase, decrease, or stay the same?

We would expect our strength of evidence to increase if the sample size is larger but the observed proportion is the same.

e) If we repeated the experiment with a different friend and our sample size stayed the same (120), but the number of times he chose scissors was 38, would the strength of evidence increase, decrease, or stay the same?

We would expect the strength of evidence to decrease if the observed statistic is closer to the null (less distance) and the sample size stayed the same.

f) What if we used our original experimental data ($\frac{31}{120}$ scissors), but instead we wanted to do a two-sided test instead of a one-sided test. Would our strength of evidence increase, decrease, or stay the same?

The strength of evidence would decrease if we go from a one-sided to a two-sided test.
Of note, our p-value will roughly double (higher p-value is less strength of evidence), but our z-statistic is still the same.

3) After Al Franken (D-MN) resigned from the U.S. Senate in 2018, one might wonder if Minnesota has had a larger number of congressional resignations than one would expect compared to other states. On Teams, download the resignations.csv file and upload it into R using the Tidyverse Tutorial or Course Guide as reference. This file[1] lists all of the congressional resignations from state senators back to 1905. As each state has exactly 2 senators, it is reasonable to assume that if a resigned senator is chosen at random, each state has an equal chance of being represented if state has nothing to do with resignation.

The following code follows the Halloween Candy example in the Course Guide (Significance: How Strong is the Evidence?) and can be used to find proportions of the dataset.

```
library(tidyverse)
library(janitor)

resign %>%
  count(State) %>%
  mutate(Proportion = n/sum(n)) %>%
  adorn_totals()
```

Alternatively,

```
resign % > %
mutate(MN = State == "MN") % > %
count(MN) % > %
mutate(Proportion = n/sum(n)) % > %
adorn_totals()
```

a) In your own words, explain what the research question is. What are our observational units and variable(s) of interest? Classify the variables as either categorical or quantitative.

The question is do senators from MN resign their office at an unusually high rate?
Our observational units are senators who have resigned since 1905. The variable is which state they are from, which is categorical.

b) List your null and alternate hypotheses, observed statistic, sample size, and the standard deviation of your null distribution.

$H_0 : \pi = \frac{1}{50} = 0.2$
$H_a : \pi \neq 0.2$
$\hat{p} = \frac{5}{120} = 0.04166667$
$n = 120$
$sd \approx 0.013$

c) List your simulated p-value. Comment on the strength of evidence.

Answers may vary slightly, but should be around 0.092. This is moderate evidence against the null that the proportion of resigned senators from MN is 0.02.

d) Calculate your standardized statistic. Comment on the strength of evidence.

$z = \frac{0.041666667 - 0.02}{0.013} = 1.666667$
This is moderate evidence against the null hypothesis that the proportion of resigned senators from MN is 0.02.

---

[1]From Kaggle, https://www.kaggle.com/yamqwe/congressional-resignationse
Can also be loaded directly into R from github https://raw.githubusercontent.com/mrocha316/MA206/main/Resignations.csv