Reporting Results

Stat 365: Statistical Communication

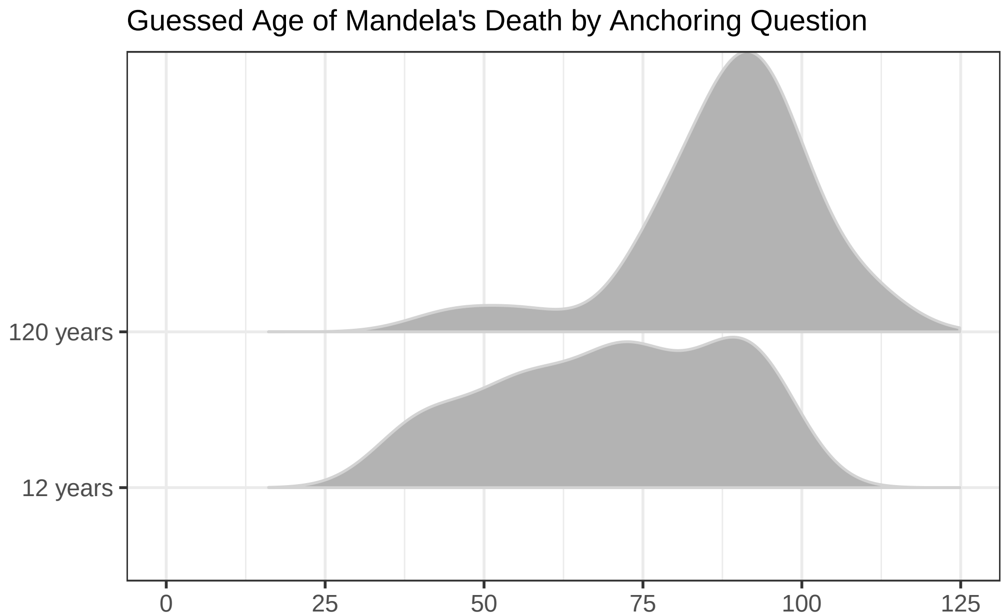
Statistical Tendencies

Dr. Keith Stanovich, a well-known psychologist and author of How to Think Straight about Psychology, describes probabilistic reasoning as the "Achilles heel of human cognition." His prime example is the statement: Men are taller than women. He goes on to describe this as a statistical tendency, not a statement that holds for all cases.

1. For each of the following questions, indicate whether or not it is a legitimate question about a statistical tendency. For the ones that are legitimate, answer the question about the tendency.

- Do you tend to go to bed before midnight?
- Do you tend to use fewer than 25 characters when you send a text message?
- Do you tend to have green eyes?
- Do you tend to have been born in California?
- Do Cal Poly students tend to have been born in California?
- Do Cal Poly students tend to be undergraduates?
- Do Cal Poly students tend to have applied to Cal Poly?
- Have U.S. presidents tended to be at least 35 years old?
- Have U.S. presidents tended to be at least 50 years old?
- Do you tend to be at least 20 years old?
- Do you tend to attend class?

2. Recall our Mandela study from the first week of class. Consider the following graph produced from the results:



- Does the data suggest that you tend to guess a higher age of death when anchored to a larger number (120 years)? Explain how you can tell.

- What inference method you would use to test this hypothesis?

Pesky p-values

One of the most important concepts and methods in statistics is also one of the hardest to understand and communicate clearly. We will examine what a p-value is and, at the same time, what a p-value is not.

3. Suppose we want to conduct a formal statistical analysis on our Mandela data. Use the output below to help you answer the following questions.

Table 1: Table X. Means and Standard Deviations of the Guesses for Mandela's Age of Death for each Numerical Anchoring of Participants to a Younger or Older Age

Numerical Anchor	Mean (Years)	Standard Deviation	N
12 Years	70.4	18.5	22
120 Years	88.0	14.2	25

Note:

Data was collected from the Spring 2023 Stat 365 class at Cal Poly.

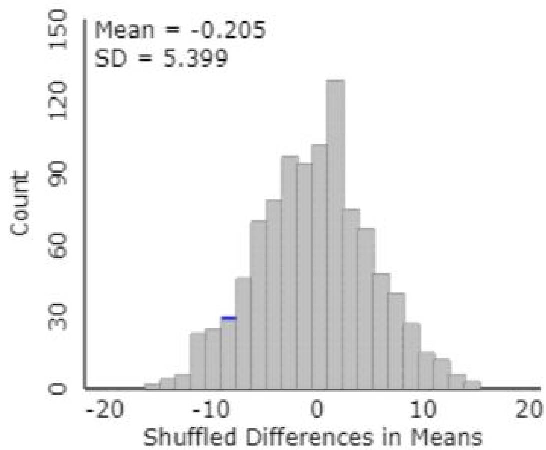
- Describe the null hypothesis to be tested.

- Did students who were anchored to an older age (120 years) guess a higher average age of death, as compared to students who were anchored to a younger age (12 years)? Justify your answer.

- Suppose for the moment that anchoring students to a specific age has no effect on the age of death guessed. Would it be *possible* to have obtained a result as extreme as we found?

- Describe how we could use a common device to simulate a repetition of this study, under the assumption that anchoring students to a specific age has no effect on the age of death guessed.

Below is a graph of simulated values (120 years - 12 years) from the Rossman Chance two-sample mean applet.



- Describe how to use the simulation analysis to approximate the p-value for this study. Then use the graph of simulation results to approximate the p-value.

Below are results from a two-sample t-test assuming equal variances.

```
    Two Sample t-test

data:  guess_yrs by anchor_q
t = 3.6733, df = 45, p-value = 0.000317
alternative hypothesis: true difference in means between group 120 years and group 12 years i
95 percent confidence interval:
 9.551312      Inf
sample estimates:
mean in group 120 years  mean in group 12 years
               87.96000                70.36364
```

- How does the p-value from this test compare to the p-value you obtained from the simulation above?

- Summarize your conclusion. Support your statement with the test statistics (and any specifications, e.g., degrees of freedom) and associated p-value.

Guidelines for p-value reporting

When reporting on statistical analyses, it's usually not appropriate to include all of this information about what a p-value means. But it is imperative not to overstate or otherwise mis-interpret what a p-value means or what conclusions the p-value allows. Some take-aways:

- A p-value is the probability, if the null hypothesis were true, of obtaining a test statistic at least as extreme as the one observed.
  - Null hypothesis is typically a statement of "no difference" or "no effect"
  - Null hypothesis statement is sometimes replaced with "by chance alone"
    - "Chance" can refer to random sampling or random assignment in data collection process, or hypothetical randomness with observational studies
    - "Alone" refers to the null hypothesis of "no difference" or "no effect"
- A p-value is one measure of the strength of evidence that the data provide against the null hypothesis.
  - There's nothing special about commonly used cut-offs such as .05 or .10 or .01.
  - A large (or otherwise non-small) p-value that is not small does not provide evidence in favor of the null hypothesis.
  - A p-value says nothing about the magnitude or importance of an effect.
- A p-value is NOT:
  - Probability that the null hypothesis is true, given the data
    - This is known as the prosecutor fallacy.
  - Probability that the data occurred by chance alone

When reporting on statistical analyses, it's usually not appropriate to include full information about what a p-value means. But it is imperative not to overstate, or otherwise mis-represent, what a p-value means or what conclusions the p-value allows.

Statistical Significance vs Practical Importance

4. What can we say about the difference in ages of death guessed between those anchored to an older age compared to a younger age in our Mandela study?

- Interpret the 95% confidence interval given above for the population average of all Statistics Majors at Cal Poly.

- Are the test decision from the previous section and the confidence interval consistent with each other? Explain.

- Under what circumstance is it especially important to keep in mind the distinction between statistical significance and practical importance?

Effect size for two-sample comparison of group means

For a two-sample comparison of group means, (Cohen's) effect size is defined as:

$$d = \frac{\text{sample mean 1} - \text{sample mean 2}}{\text{pooled standard deviation}}$$

where the pooled standard deviation

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

and measures the number of standard deviations (not standard errors) the two sample means are apart. Rough guidelines:

- $d = 0.2$: small difference
- $d = 0.5$: medium difference
- $d = 0.8$: large difference
- $d = 1.2$: very large difference
- $d = 2.0$: huge difference

- Calculate the effect size for comparing age of death guessed between the older and younger anchoring treatments. Then use the rough guidelines to select an adjective for describing this effect size.

- Rewrite the expression for the (pooled) t-test statistic ($t = \frac{(\bar{x}_1 - \bar{x}_2) - \text{null value}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$) as a product of two terms, one of which is the effect size d. According to this expression, what are two different ways to obtain a large t-test statistic, and therefore a small p-value?