

Lab 5: Hypothesis testing

Not graded, just practice

Katie Schuler

2024-09-26

Practice your new stats skills with these practice exam questions! Best to open a fresh Google Colab notebook and test things out! Refer to the study guide to find answers as well.

Tip

More than one answer may be correct!

If you would like to practice with a set of data, you can import the following dataset with `read_csv`. Note that females are coded as NA in this dataset!

```
# brain volumes simulated from Ritchie et al  
"http://kathrynschuler.com/datasets/brain_volume.csv"
```

0.1 Visualize a categorical variable

(a) Which of the following is the best choice to visualize a categorical variable? Choose one.

- (A) `geom_rug()`
- (B) `geom_histogram()`
- (C) `geom_boxplot()`
- (D) `geom_point()`
- (E) `geom_smooth()`

(b) Which of the following figures shows a box plot?

Rows: 5216 Columns: 2

-- Column specification -----

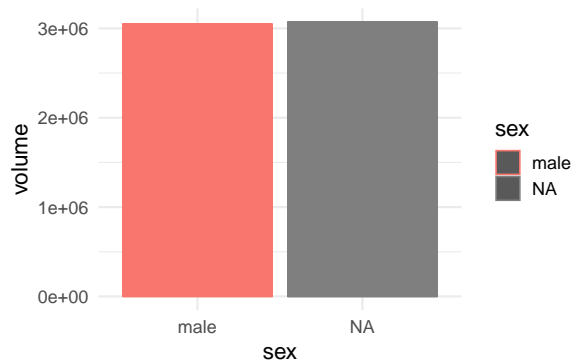
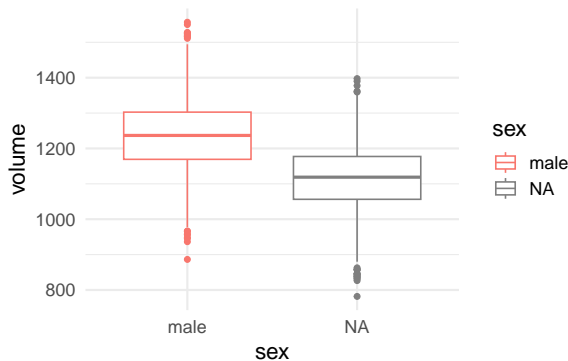
Delimiter: ","

chr (1): sex

dbl (1): volume

i Use ``spec()`` to retrieve the full column specification for this data.

i Specify the column types or set ``show_col_types = FALSE`` to quiet this message.



- (A) left
 - (B) right
 - (C) both the left and right
 - (D) neither the left and right
- (c) True or false, the difference in means between male and female participants in our sample is subject to sampling variability?
- (A) True
 - (B) False

0.2 Hypothesis testing framework

- (a) Fill in the blanks about the 3-step hypothesis testing framework: (1) First we pose a _____, (2) then we ask if true, how likely is our observed pattern of results? This likelihood is quantified with a (_____), and (3) finally if the likelihood is less than some threshold, we (_____) the null hypothesis.
- (b) What is the practical reason we pose a null hypothesis?

- (A) It is the hypothesis most likely to be true.
 - (B) It allows us to generate predictions based on prior beliefs.
 - (C) It is the hypothesis for which we can simulate data.
 - (D) It ensures that the alternative hypothesis is proven false.
- (c) True or false, Randomization simulates a world in which there is no relationship between brain volume and sex.
- (A) True
 - (B) False
- (d) Which of the following would compute a p-value?
- (A) Count the number of values in our null distribution that are more extreme than our observed value.
 - (B) Call `get_p_value()` on an `infer` object in which we have constructed a null distribution.
 - (C) p-values are not computed.
 - (D) Add up all of the differences in means
- (e) A large p-value means our observed value is very _____ under the null hypothesis. A small p-value means our observed value is very _____ under the null hypothesis.
- (f) True or false, a p-value of less than 0.05 indicates obtaining our observed value under the null is *impossible*
- (A) True
 - (B) False
- (g) A type I error is also known as a _____ (wrongly thinking that the effect is present); a type II error is also known as a _____ (wrongly thinking the effect is absent)

0.3 There is only one test

- (a) There are two ways we can construct a sampling distribution: (1) _____, via brute computational force; and (2) _____, by assuming the data were sampled from a known probability distribution.
- (b) True or false, we can use the `t.test()` function or `assume(distribution = "t")` with an `infer` object to calculate a p-value for a t-test.

0.4 Correlation

- (a) We can explore the relationship between two quantities visually with a scatter plot. Which geom is best suited for this?
- (A) `geom_histogram()`
 - (B) `geom_boxplot()`
 - (C) `geom_point()`
 - (D) `geom_dotplot()`
 - (E) `geom_density()`
- (b) If there is no relationship between variables, we say they are _____
- (c) One way to quantify _____ relationships is with correlation.
- (d) Correlation ranges from -1 to 1, where:
- (A) -1 means mostly negative and 1 means mostly positive
 - (B) -1 means perfectly negative and 1 means perfectly positive
 - (C) -1 means low correlation and 1 means high correlation
- (e) True or false, on an `infer` object, we can calculate the correlation with `calculate(stat = "correlation")`
- (A) True
 - (B) False
- (f) Suppose we quantify how likely our observed correlation is under the null hypothesis and our p-value is 0.68. Should we reject the null hypothesis?

- (A) Yes
- (B) No