Exam 1 study guide

Katie Schuler

2024-08-27

0.1 Additional study materials

- [Practice exam 1](#)
- [Provided reference sheets](#) (you will have this on the exam)
- Practice exam 1 solutions

The first exam will test the following learning objectives, divided into the following topic areas. For each topic area, you should be able to do the list that follows. You can think of this as a studying checklist!

1. **R Basics: general**

   - Assign an object to a valid variable name, list all variables in the environment and remove them
   - Use packages and differentiate between installing and loading
   - Get help with a function or package from R
   - Return information about an object, including its structure, data type, length, and attributes
   - Explain what functions and control flow are; differentiate between types of control flow

2. **R Basics: vectors, operations, and subsetting**

   - Distinguish between an atomic vector and a list
   - Create atomic vectors and determine their data types
   - Differentiate between implicit and explicit coercion and coerce an object to another type
   - Use arithmetic, comparison, and logical operators on vectors
   - Explain how more complex data structures are built from atomic vectors and create them
   - Distinguish between `NA` and `NULL`
   - Subset vectors and higher dimensional objects with the `[`, `[[` and `$` operators

3. **Data visualization: basics**

- Describe how to create a plot with `ggplot2` including the 3 basic requirements
- Distinguish between mapping and setting aesthetics
- Describe how `ggplot2` maps categorical variables to aesthetics and interpret the 3 common warnings people encounter in this process
- Interpret `ggplot()` calls with explicit or implicit arguments for data and mapping
- Recognize the geoms we discussed in class and select which to use for a given situation
- Differentiate between globally and locally defined mappings and recognize them in given plot (or code)

4. **Data visualization: layers**

- Use the `position` argument to modify the position of the geoms in `geom_bar()` or `geom_point()`
- Describe `stat="identity"` and describe the default transformations for `geom_bar()`, `geom_histogram()`, and `geom_smooth()`
- Set the smoothing method for `geom_smooth()` and the bins or bindwidth for `geom_histogram()`
- Facet a plot with `facet_wrap()` and `facet_grid()`
- Modify axis, legend, and plot labels with `labs()`
- Apply a given theme to a plot and adjust the base font size or family.
- Describe scales and recognize the outcome of adding a scale layer

5. **Data importing**

- Load the `tidyverse`, recognize the included packages, and critique code for redundant loading
- Construct a tidy dataset and critique whether a given dataset is tidy
- Use the map function from the `purr` package
- Create a tibble and distinguish between a tibble and a data frame
- Use `readr` to read delimited files and determine whether `readr` can read files of a given type
- Use `col_types` to add a column specifications and explain how readr guesses without it
- Solve the 3 most common importing problems we discussed in class

6. **Data wrangling**

- Describe the common structure of `dplyr` functions (aka verbs)
- Combine `dplyr` functions with the pipe operator to solve complex problems
- Manipulate rows with `filter()`, `arrange()`, and `distinct()`
- Maniuplate columns with `mutate()`, `select()`, and `rename()`
- Group and summarise data with `group_by()`, `summarise()`, and `ungroup()`
- Evaulate `dplyr` functions that include the common arguments we covered in class

7. **Sampling distribution**

- Explore a dataset with an appropriate figure (histogram, boxplot, scatterplot) and summary statistics appropriate for the distribution.
- Recognize uniform and Gaussian probability distributions in a plot or equation and use R's functions `d*()`, `p*()`, and `r*()` to work with these distributions
- Explain the difference between the parameter and the paramter estimate
- Construct the sampling distribution of a paramater estimate with `infer` and quantify the spread of the distribution with a confidence interval.
- Understand the difference between constructing a confidence interval the standard error method vs. the percentile method.

8. **Hypothesis testing**

- Given a set of data, implement the 3-step hypothesis testing framework **nonparametrically**: (1) Pose a null hypothesis, (2) quantify how likely a given pattern of results is under the null, and (3) determine whether to reject the null (conceptually and with the `infer` framework).
- Given a theoretical distriubiton (e.g. t), implement the 3-step hypothesis testing framework **parametrically**.
- Given an observed correlation, determine whether a correlation is positive, negative, or no correlation.