

Midterm Details Statistics 506

The midterm will be Oct 26th, in-class. It will be with pen or pencil, closed notes.

Topics

The midterm will cover lecture notes 1-9, 11, 14-15. Notably this excludes “Other Statistical Software” and “R Visualization”. (15 is the second set of SAS notes, forthcoming.)

Example problems

Here’s some examples of the types of problems you can expect to see.

1. What value does `q` have after the following code is executed?

```
a <- 1:100 %% 2
b <- c(1, 2)
q <- sum(x * y)
```

2. Write an R function that takes in a vector of numbers and returns a labeled vector of the mean and median. Do not use the existing R functions `mean` or `median`.
3. For each of the following calls to `lm()`, provide the equivalent call to Stata’s `regress`.

- $y \sim x + z*q$
- $y \sim x:z + z$
- $y \sim I(x^2) - 0$

4. We want a regular expression which will match the following strings:

- `cat`.
- `a73`.
- `?=+`.

but not this string:

- abcd

For each of the follow regular expression, determine whether it will match the appropriate strings. If not, make a *minor* change that will fix it.

- a. `...\.`
- b. `(.){3}.`
- c. `[^.]{3}`
- d. `^[^1]`

5. The “orange” data, which contains 35 rows and 3 columns, records the growth of orange trees. The dataset has three columns:

- Tree: an ordered factor, identifying individual trees,
- age: a numeric vector giving the number of days since the tree was planted,
- circumference: a numeric vector recording the circumference of the trunk in mm.

- a. Write a tidyverse pipe to determine the number of observations per tree.
- b. Write a tidyverse pipe to change the units of age to “years” and circumference to “cm”.
- c. Write a tidyverse pipe to add a column assigning a z-score to each tree, centered around the mean for all trees at a given age.