Introduction to R for Data Management and Analysis

Marcel Ramos

June 19, 2016

Announcements

- Exercises mostly graded
- Extra credit due Thursday
- Final class today

Notes on Tuesday's lecture

- Questions on the material

Pseudocode

- `for` loop structure

```
for (variable in vector) {
    # < enter code here >
}
```

- `function` structure

```
functionname <- function(argument1 = "default",
                         argument2 = "default") {
    myresult <- anotherfunction(argument1, argument2)
        return(myresult)
}
```

Notes on Tuesday's lecture

- Functions are powerful tools
- Minimize errors
- Create a set of operations to achieve a goal
- Easy to write
- Loops are useful but are not easily extendable

Data Analysis

- R is particularly good at statistics
- Packages with new methods get published faster
- Extensibility is an advantage compared to other software

The basics

- Frequency tables
- Calculating odds ratios
- `relevel`

Frequencies

- `gmodels` package
- `CrossTable` function
- `table` function

Tests

- `chisq.test` function
- `fisher.test` function
- `t.test` function

Useful functions to apply on model objects

- Functions that work on `lm` class objects
  - `summary`
  - `fitted`
  - `resid`
  - `predict`

Tidy model results with broom

- `tidy` - model coefficients
- `augment` - fitted/residual values and more
- `glance` - model level statistics

Linear Regression

- `lm` function
- UCLA example

Logistic Regression and Odds Ratios

- `glm` function
- `Odds Ratio` calculation
- UCLA example