

# *Introduction to R for Data Management and Analysis*

Marcel Ramos

Thursday, June 21, 2018

# *Announcements*

- Last in-person class on Tuesday
- Review Session - data analysis workflow
- Extra credit assignment

## Notes on Tuesday's lecture

- Questions on the material

### Pseudocode

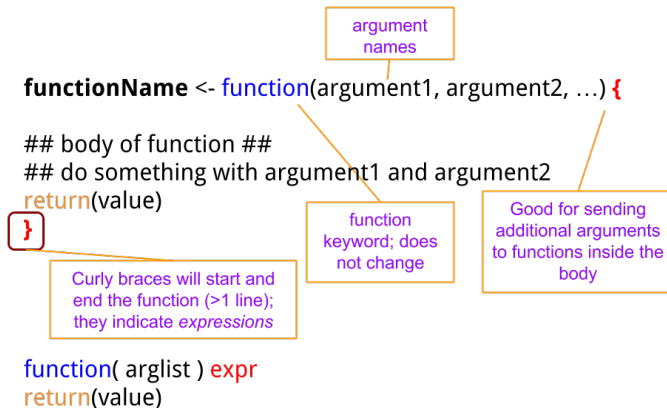
- for loop structure

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

- function structure

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

## More on functions



- ❖ The body of the function will include the operations to be performed with the input arguments

## *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

- 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*

- Use the broom package to clean up results from model functions
  - 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