

# Homework 3: Intro to R for Epidemiologists

## Instructions

- Due at 11:59 PM on Wednesday, April 8, 2015
- Late assignments will not be accepted
- Covers topics in weeks 7-10
- Submit this assignment as a .R file by e-mail to [jenna.krall@emory.edu](mailto:jenna.krall@emory.edu)
- Name your file as “LASTNAME\_560R\_hw3.R”
- **Before submitting the assignment, comment out any `rm(list = ls())`, `setwd()`, or specific file paths to your computer (e.g. in `read.csv()`)**

You may work with your classmates, but you must type and turn in your own assignment. Do not copy and paste someone else’s work. Your code comments should be your own.

## Grading

The assignment is worth 15% of your final grade and will be graded out of 30 total points. **If you submit lines of code that produce errors in R, you will be docked one letter grade (3 points).**

Additionally, you will not receive credit for solutions such as:

```
mean_diab <- 5.2
```

You must, whenever possible, use functions on the right hand side:

```
mean_diab <- mean(diabetes)
```

## 1 Data

SUPPORT (Study to Understand Prognoses Preferences Outcomes and Risks of Treatment)

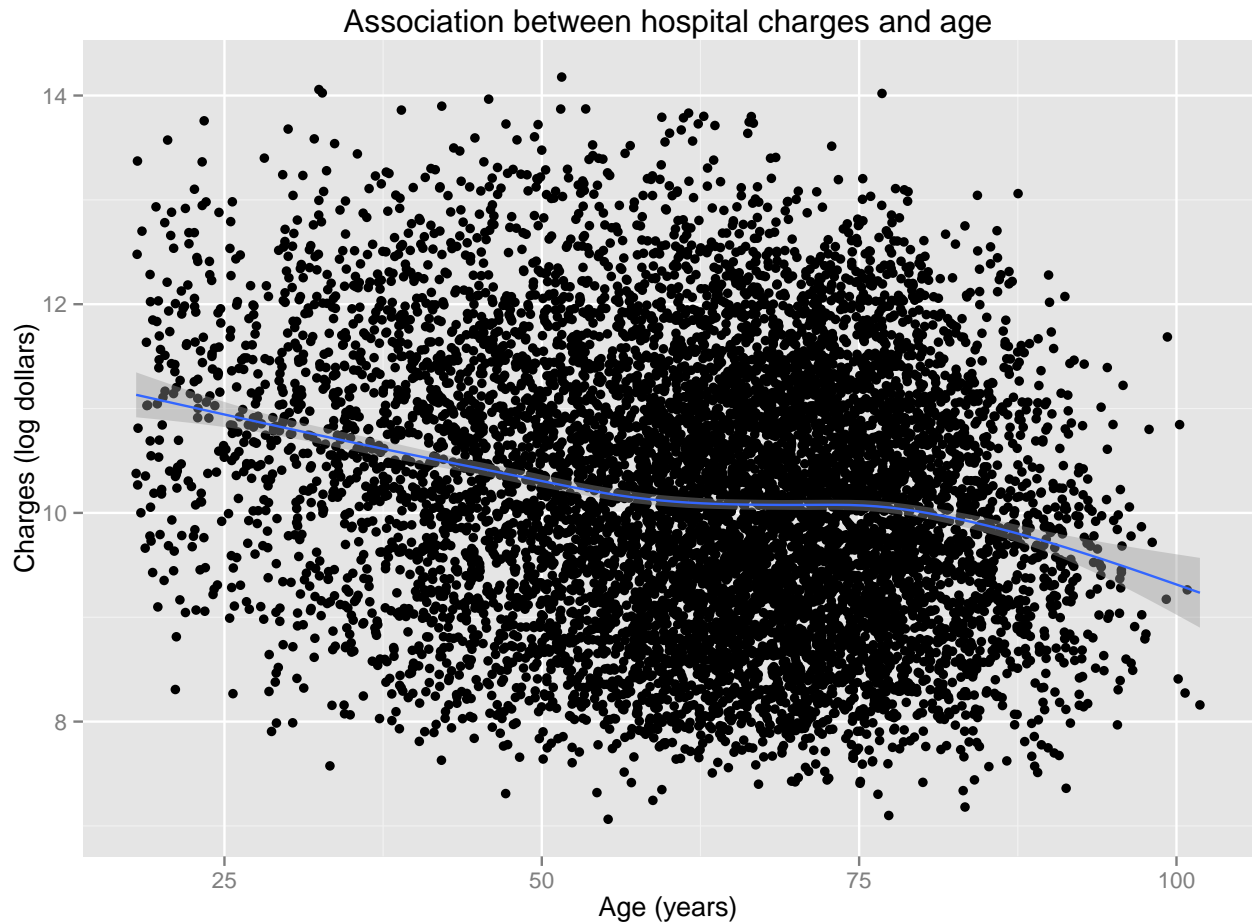
References

- Information about variables: <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/Csupport2.html>
- Data accessed February 6, 2015, <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>.

## 2 Part 0. Code formatting and commenting (2 points)

## 3 Part 1. Exploratory data analysis (3 points)

Recreate the following plot in `ggplot2` using the support dataset.



#### 4 Part 2. Estimating associations with in-hospital death (10 points)

##### 1. Univariate logistic regression

- (a) Using simple logistic regression, estimate associations between in-hospital death (hospdead) and age, sex, coma score, number of comorbidities (treated as continuous), diabetes status, and years of education.
- (b) Extract out odds ratios and corresponding confidence intervals

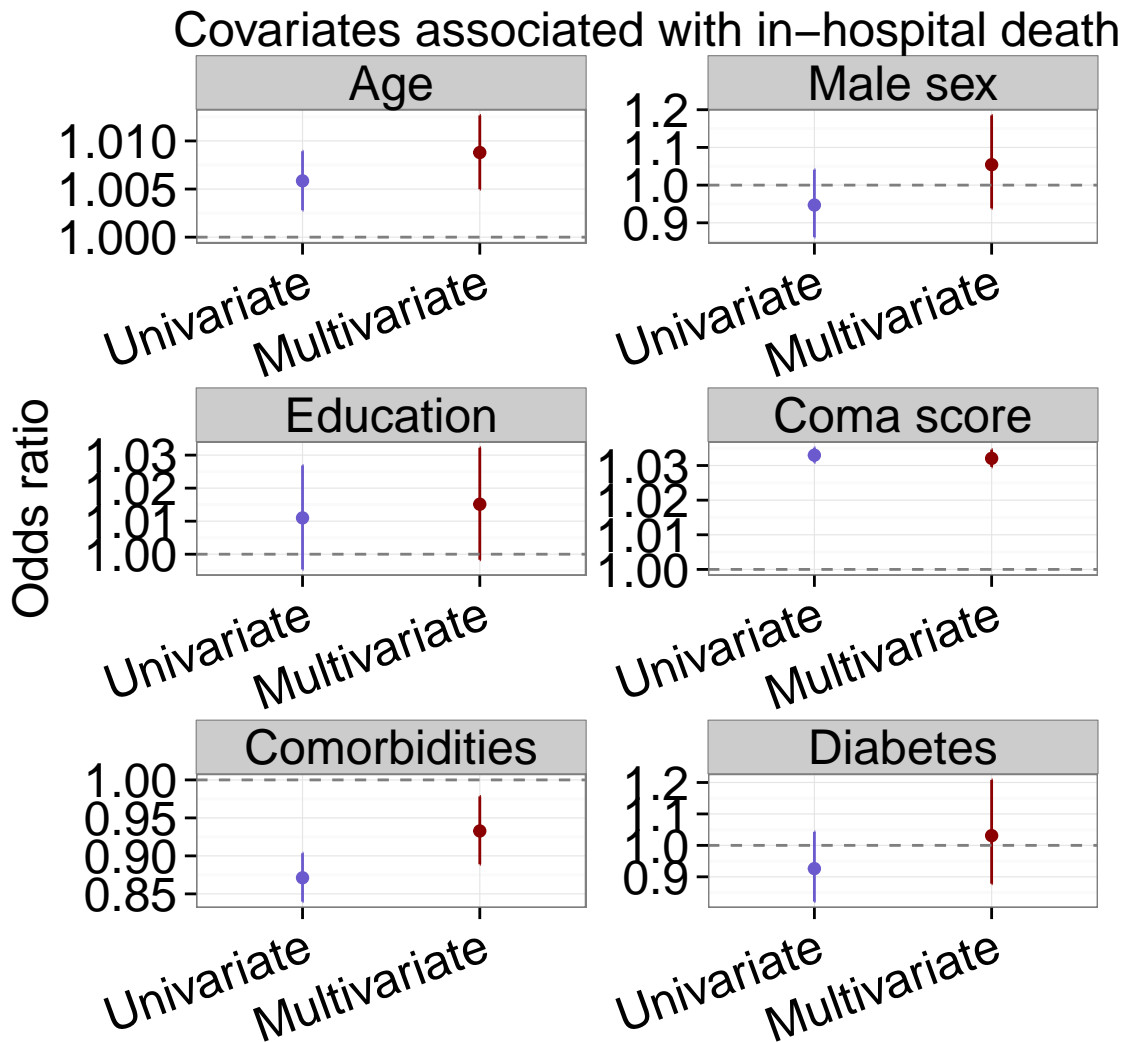
##### 2. Multiple logistic regression

- (a) Using multiple logistic regression, estimate associations between in-hospital death and age, sex, coma score, number of comorbidities (treated as continuous), diabetes status, and years of education.
- (b) Extract out odds ratios and corresponding confidence intervals

3. Combine results from 1-2 into one data frame called `death_OR` with 5 columns corresponding to the variable (e.g. age), odds ratio, lower 95% confidence bound, upper 95% confidence bound, and regression type (Univariate or Multivariate). Note that for each variable, you will have two rows: one corresponding to the results from part 1 and one corresponding to the results from part 2.

## 5 Part 3. Plotting results (15 points)

Plot your results from Part 2 using ggplot2 as below. You will need to use `cols <- c("slateblue", "darkred")`. Note: the size of all text (labels and legends) should be set to 18 pt.



## 6 Bonus (1 - 5 points)

Create a plot of your choosing (but different from the plots above) using ggplot2 and the support dataset.

- 1-2 points: any plot discussed in class
- 2-5 points: adding complexity beyond what was covered in class