

Module 7: lab

BIOSTAT 213: Introduction to Computing in the R Software Environment

Wednesday, February 27, 2019

Write your solutions in the provided code chunks. You may work in groups of 2–3.

1. Use the space below to import data and load libraries.

```
> library(knitr)
> #
```

2. Use *nhanes.csv*. Show the first five rows of low-density lipoprotein (LDL) cholesterol (**ldhdd**), age (**ridageyr**), gender (**gender**), history of smoking (**hs**), history of military service (**military**), and vigorous recreational activity (**vra**). Use **kable()** to output this data. Center-align all columns.

```
> #
```

3. Use *nhanes.csv*. Use **ggplot2** to create a scatter plot with LDL cholesterol (**ldhdd**) on the *y* axis and age (**ridageyr**) on the *x* axis. Include a LOESS smoother.

```
> #
```

4. Use *nhanes.csv*. Fit a linear model, with LDL cholesterol (**ldhdd**) as the outcome. Include adjustment for age (**age**), gender (**gender**), history of smoking (**hs**), history of military service (**military**), and vigorous recreational activity (**vra**).

```
> #
```

5. Use *nhanes.csv*. Use **kable()** to display the coefficient estimates, the standard errors, the test statistics, and the *p*-values. Omit the row representing the intercept. Center-align all columns. Round the first 3 columns to 1 digit and the last column to 2 digits.

```
> #
```

6. Use *nhanes.csv*. Use **cbind()** to create an object that contains the coefficient estimates and the lower and upper bounds of the 95% confidence intervals. Check the object type. Is this an object type we have discussed in this course? Use **as.data.frame()** to convert the object to a data frame. Use **names()** to give the columns new names: **Estimate**, **Lower**, and **Upper**. Omit the row representing the intercept. Use **kable()** to display the data frame. Center-align all columns and round all columns to 1 digit.

```
> #
```

7. Use *nhanes.csv*. Use the data frame defined immediately above to check whether each confidence interval includes 0.

```
> #
```

8. Use *nhanes.csv*. In Module 6's lab, you were asked to use **tapply()** and a custom function to find the percent of individuals in each **bpcat** group who smoke (**hs**). What is a more efficient way to find the percent of individuals in each **bpcat** group who smoke (**hs**)? Show this method.

```
> #
```

9. Use *nhanes.csv*. Write a function that takes as an argument a vector (you can think of the vector as a column in a data frame). The function should return the number of missing values that the variable contains. Check the function on the **ridageyr** and **bmwt** variables.

```
> #
```

10. Use *nhanes.csv*. Use one of the apply functions to apply the above function to every single column in the NHANES data.

```
> #
```