

Shannon Keenan
Exercises in R
June 2011

Part 1: Graphical Descriptive Measures

The data set used is taken from Databank.xlsx file. It includes these several variables:

Age – given in years

Education Level- (0-no high school degree, 1-high school degree, 2- college graduate, 3-graduate degree)

Smoking Status- (0-does not smoke, 1-smokes, less than 1 pack per day, 2- smokes one or more packs per day)

Exercise-(0-none, 1-light, 2-moderate, 3-heavy)

Weight- given in pounds

Serum Cholesterol- given in milligram percent

Systolic Pressure- given in millimeters of Mercury

IQ- given in standard IQ test score values

Sodium- given in miliequivalents per liter

Gender- male(M), female(F)

Marital Status -(M=married, W= widowed, S=Single, D=divorced)

The goal is to describe these variables and their relationships in R.

We must access the data. The data is an excel file. First, open the file in excel. Change the header Smoking Status in column D to Smoking-Status so there is no space in between words. Save the file as a text document and note the location of the file.

Now we need to open the data in R:

open R

type `databank= read.table(file="yourfilelocation ",header=TRUE)` in the command window.

note: in your file location you must put two backslashes between each folder.

For example your entry could read:

```
databank= read.table(file="C:\\Users\\Shannon\\Documents\\Databank.txt",header=TRUE)
```

type `databank` in the command window to see your table

type `attach(databank)` in the command window to be able to work with the data.

Now you can type the name of any header to view its data entries. (example, type `IQ` in the command window and the 100 data enties for IQ will appear.)

note: R is case-sensitive so each variable must be typed in all CAPS.

Exercise 1: Describe the categorical variable Marital Status.

1A: Frequency Table

Type `MARITAL.STATUS` in the command window to see the data.

Type `table(MARITAL.STATUS)` in the command window to see a frequency table.

Type `marriage= table(MARITAL.STATUS)` in the command window.

Type `marriage` to see that this is not assigned to the frequency table of Marital Status.

1B: Construct a pie chart for Marital Status

Type `pie(marriage)` in the command window to see a pie chart of Marital Status

Type `pie(marriage, main="Marital.Status")` to make a pie chart of Marital Status with a title.

Note: All graphs made in R can be put into other programs using copy and paste. Be sure to save your plot because R uses the same output window for all graphs.

1C: Construct a bar chart for Marital Status

Type `barplot(marriage)` to see a barplot of Marital Status

Type `barplot(marriage, main="MaritalStatus")` in the command window to see the barplot of Marital Status with a title.

Exercise 2: Describe the categorical variable Smoking

Repeat steps 1A through 1C for variable `SMOKING.STATUS`

Exercise 3: Describe the numerical variable Weight

3A: Construct a stemplot.

Type `stem(WEIGHT)` in the command window to produce a stemplot of weight

3B: Construct a frequency histogram

Type `hist(WEIGHT)` in the command window to make a histogram of weight.

Type `hist(WEIGHT, breaks=11)` to make a suggestion of 11 classes for the histogram of weight. To put exactly 11 classes in the histogram you should put a vector with the points you would like breaks.

Exercise 4: Describe the numerical variables Age, Systolic and IQ.

Repeat steps 3A-3B for the variables `AGE`, `SYSTOLIC` and `IQ`. Select an appropriate number of classes to make for each histogram.

Exercise 5: Describe the variables Weight and Sodium

5A: Construct a scatterplot

Type `plot(WEIGHT, SODIUM)` to make a scatterplot of weight vs sodium intake

Type `plot(WEIGHT, SODIUM, main="WeightvsSodium")` to add a title.

Exercise 6: Describe the variables Weight and IQ

Repeat step 5A for Weight and IQ.

PART 2: Numerical Descriptive Measures: Center, Variation, Skewness, Relative Standing

Exercise 1: Measures of Central Tendency- AGE

Step 1A: Compute the mean

Type `mean(AGE)` to find the mean age.

Step 1B: Compute the median

Type `median(AGE)` to find the median age.

Step 1C: Compute the mode

Type `which(table(AGE) == max(table(AGE)))` to find the mode of Age. The first number listed is the mode of the variable and the second is the place when ordered smallest to largest.

Exercise 2: Measures of Variability

Step 2A: Compute the Range

Type `range(AGE)` to find the highest and lowest values of the data set. Use R as a calculator to compute the range. For example just type `71-18` and R will give you 56.

Or—type `diff(range(AGE))` to get the range directly.

Step 2B: Compute the variance

Type `var(AGE)` to find the variance of age.

Step 2C: Compute the Standard Deviation

Type `sd(AGE)` to find the standard deviation of age.

Exercise 3: Measures of Relative Standing

Step 3A: Compute the lower quartile

Type `quantile(AGE, .25)` to get the lower quartile.

Step 3B: Compute median (middle quartile)

Type `quantile(AGE,.5)` to get the median.

Step 3C: Compute the upper quartile

Type `quantile(AGE,.75)` to get the upper quartile.

Alternatively, you can type `summary(AGE)` to get the five number summary and mean.

Repeat Exercises 1-3 of Part 2 for WEIGHT, SYSTOLIC and IQ.

PART 3: Correlation and Regression

Exercise 1: Describe the variables Weight and Sodium

Step 1A: Compute Correlation

Type `cor(WEIGHT,SODIUM)` to get the correlation coefficient for weight and sodium.

Step 1B: Compute y- intercept

Type `lm(formula= WEIGHT ~ SODIUM)` to compute the y- intercept. The number under intercept is the y-intercept and the number under sodium is the slope.

Step 1C: Place the equation on the graph

Type `res= lm(formula= WEIGHT ~ SODIUM)`

Type `plot(SODIUM,WEIGHT)`

Type `abline(res)` to add the trendline to the graph.

To add a title of to the chart type `plot(WEIGHT,SODIUM,main="yourintercept+slopes")`

