Lab 1b: Histograms

Stat 131A, Fall 2018

Learning Objectives:

- Keep getting to know RStudio work environment
- Understand the help documentation in R
- Learn to install R packages
- Analyze the distribution of a quantitative variable using a histogram.
- Describe shape, give a general "guesstimate" of center and the overall range.

General Instructions

- This lab consists of two parts. The first part (first 25 mins approx) involves a first contact with R and RStudio. The second part (second 25 mins) involves answering questions about types of variables.
- Your GSI will continue to guide/help you with the *knitting* process to get the html file that you will submit for this lab.

Part 1: More intro to R and RStudio

Installing Packages

R comes with a large set of functions and packages. A package is a collection of functions that have been designed for a specific purpose. One of the great advantages of R is that many analysts, scientists, programmers, and users can create their own pacakages and make them available for everybody to use them.

You can install a package using the `install.packages()` function. Just give it the name of a package, surrounded by quotes, and R will look for it in CRAN, and if it finds it, R will download it to your computer. In the R console, type the following command:

```r
# installing
install.packages("knitr")
```

You can also install a bunch of packages at once:

```r
install.packages(c("readr", "ggplot2"))
```

Once you installed a package, you can start using its functions by *loading* the package with the function `library()`

```
library(knitr)
```

You only have to install a package once. In other words, there is no need to reinstall a package. In order to use the functions of a package, simply *load* it with `library()`. Also, when invoking `install.packages()`, do this from the console, NOT from an 'Rmd file; this will prevent knitting problems.

Your turn

- Install packages `"stringr"`, `"RColorBrewer"`, and `"XML"` (do this from the R console!)
- Calculate: $3x^2 + 4x + 8$ when $x = 2$
- Calculate: $3x^2 + 4x + 8$ but now with a numeric sequence for x using `x <- -3:3`
- Find out how to look for information about math binary operators like `+` or `^` (without using `?Arithmetic`).
- There are several tabs in the pane `Files, Plots, Packages, Help, Viewer`. Find what does the tab **Files** is good for?
- What about the tab **Help**?
- In the tab **Help**, what happens when you click the button with the icon of a house?
- Now go to the tab **History**. What is it good for? and what about the buttons of its associated menu bar?
- Likewise, what can you say about the tab **Environment**?

Part 2: Histograms

- Write your solutions in an `Rmd` (R markdown) file.
- Name this file as `lab01b-first-last.Rmd`, where `first` and `last` are your first and last names (e.g. `lab01b-gaston-sanchez.Rmd`).
- Knit your `Rmd` file as an html document (default option).
- Submit your `Rmd` and `html` files to bCourses, in the corresponding lab assignment.

Problem 1

The following figure shows four histograms (A, B, C, D).
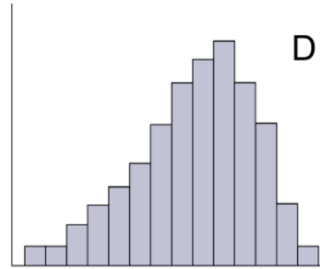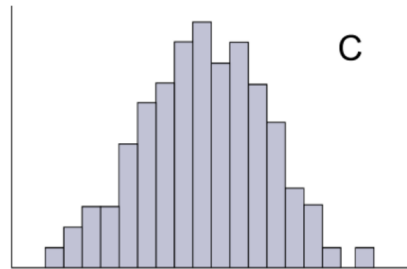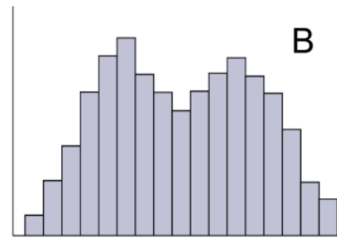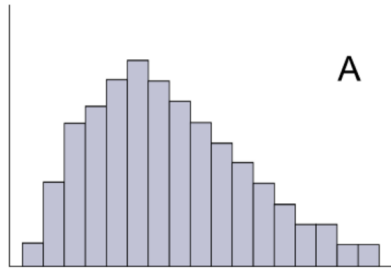
a. Which distribution seems to be skewed right?

b. Which distribution seems to be skewed left?

c. Which distribution seems to be symmetrical or "bell" shaped?

d. Which distribution seems to be bimodal?

e. Which distribution shows a gap?

Problem 2

Match the following descriptions to the histograms 1-3.

distribution 1   distribution 2   distribution 3
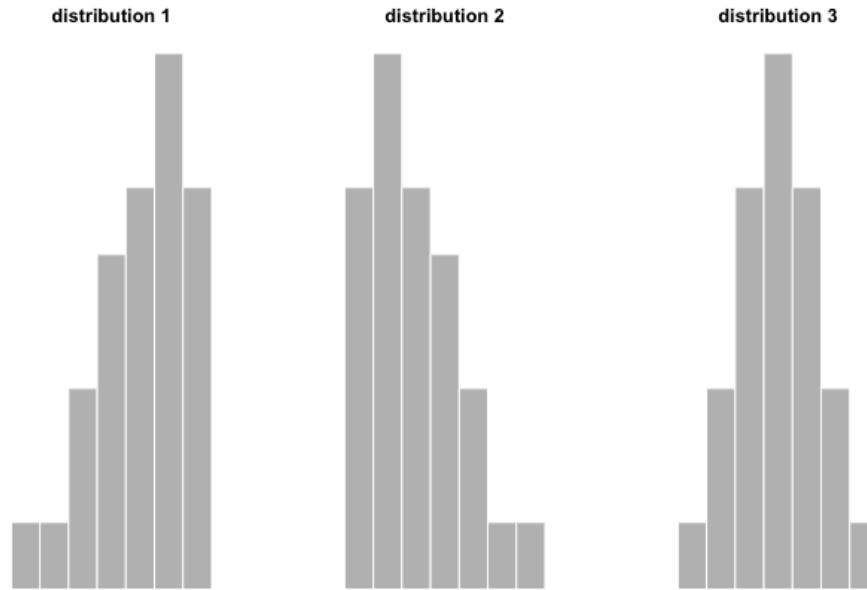
a. Scores on an easy exam for a class of students

b. Scores on a hard exam for a class of students

c. Number of siblings for a large sample of U.S. adults

d. Exact volume of soda in a one-liter bottle for a case of 24 bottles

e. Dates on the pennies I have in my car ashtray

f. Weights for a large sample of newborn babies

Problem 3

This histogram shows the distribution of exam scores for 15 students in a Biology class.

Histogram of grades
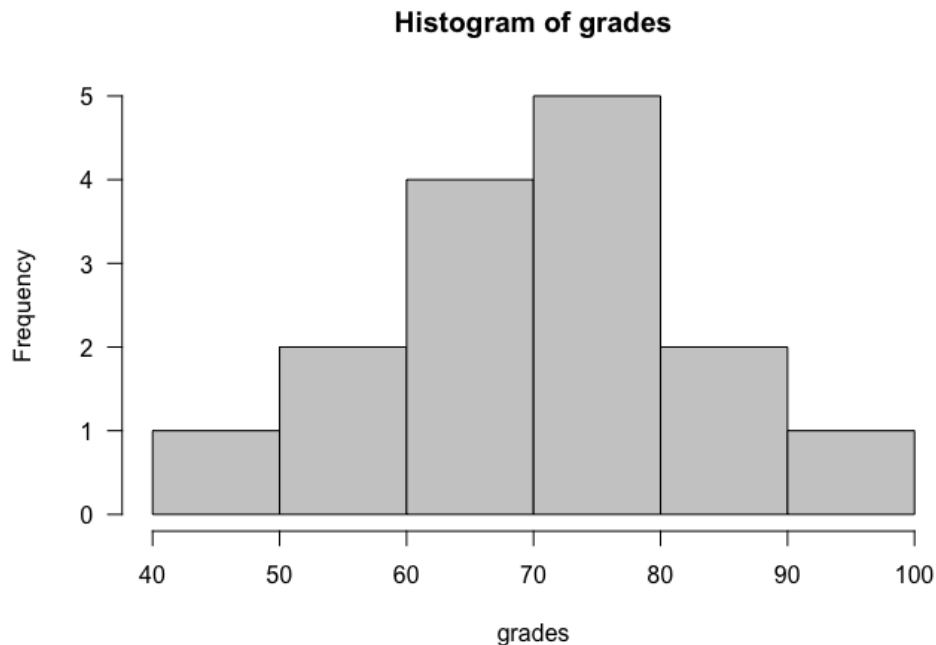
a) How would you describe the shape of this distribution of exam scores? (Use statistical vocabulary.)

b) Give an interval that describes typical grades on this exam.

c) Estimate the overall range of grades on this exam. (Range = Max – Min)

d) What percentage of the students made a D on the exam (a grade of 60-69)?

e) What percentage of the students passed the exam with a 70 or better?

f) What percentage of the students made an A ($>=90$) or a B (80-89)?

g) What percentage of the students who passed the exam made an A or a B?

h) What percentage of the students who failed the exam (grades lower than 70) made a D (a grade of 60-69%)?

Problem 4

True or False

a. A histogram allows you to visualize how values are distributed.

b. Bars in a histogram can be rearranged.

c. The area of the bars in a histogram represent the frequency of values in a class interval.

d. Class intervals must be of equal width in a histogram.

e. The height of the bars in a histogram must be equal to the frequency of the corresponding class intervals.

f. The total area of the rectangles in a histogram must be equal to 100%.

Problem 5

Below are sketches of histograms for test scores in three different classes. The scores range from 0 to 100; a passing score was 50.
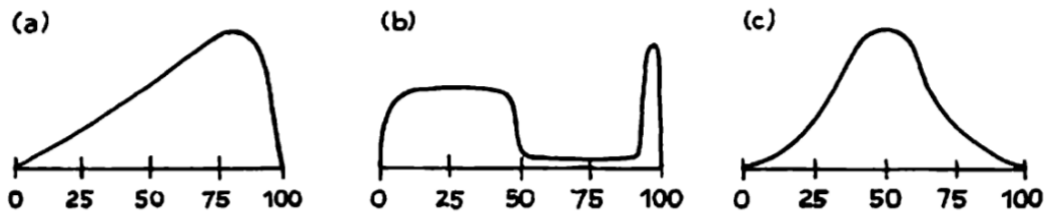
a. For each class, was the percentage who passed about 50%, well over 50%, or well under 50%?

b. One class had two quite distinct groups of students, with one group doing rather poorly on the test, and the other group doing very well. Which class was it?

c. In class (b), Were there more people with scores in the range 40-50 or 90-100?

Problem 6

In one hypothetical study, 100 people had their heights measured to the nearest eighth of an inch. A histogram for the results is shown below. Two of the following lists have this histogram. Which ones, and why?
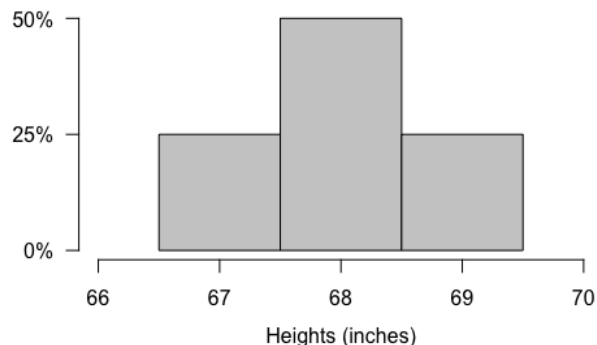
a. 30 people, 67 inches tall; 40 people, 68 inches tall; 30 people, 69 inches tall.

b. 10 people, $66\frac{3}{4}$ inches tall; 15 people, $67\frac{1}{4}$ inches tall; 50 people, 68 inches tall; 25 people, 69 inches tall.

6

c. 25 people, 67 inches tall; 50 people, 68 inches tall; 25 people, 69 inches tall.

Don't forget to submit your `Rmd` and knitted `html` file to bCourses.