

Part I: 60 points

1. Compute the mean, variance, and standard deviation of the following sample values from a population: 16, 12, 14, -2. Show the formula for the mean, variance, and standard deviation.
2. The Bookstall, Inc., is a specialty bookstore concentrating on used books sold via the internet. Paperbacks are \$1.00 each, and hardcover books are \$3.50. Of the 50 books sold Tuesday morning, 40 were paperback and the rest were hardcover. What was the weighted mean price of a book?
3. According to the *2015 American Community Survey*, in 2012 the United States had 30.1 million households with one person, 37.1 million with two persons, 17.8 million with three persons, 15.0 million with four persons, and 10.4 million with five or more persons. (4 points)
 - a. Construct a relative frequency distribution.
 - b. Report and Interpret the (i) median, (ii) mode of household size.

Please complete the following questions that come from our textbook. To avoid confusion, I have labeled each numerically then the number as it appears in the text follows.

4. 1.2
5. 1.3
6. 1.10
7. 1.11
8. 1.29
9. 2.2
10. 2.12
11. 2.13
12. 2.17
13. 2.21

Part II - R Programming, Descriptive Statistics, and Data Visualization (40 points)

The following problems require you to use R to perform the indicated tasks. The data set is available in the "Datasets" folder in the Canvas /Files/datasets section and on the assign. You must submit your R Markdown code and output with each answer for the exercise.

Import the "General Social Survey 2008" data set file in to R using the `fread` function from the `data.table` package. Note that the file on Canvas is a Comma Separated Values file or .csv. Take a look at the dataset. You will notice that it contains dozens of variables and thousands of observations. You will only work with a few variables for this assignment, but we will look at this data again in the future. Below are relevant variable

definitions, but you can find all variable descriptions in the General Social Survey Codebook.

Whenever possible use the tidyverse package to accomplish a task. Don't stress too much about this. You will be graded on writing reproducible code that produces the right answer. Not on what package you used, how you used it, or what code you use to get there.

1. Derive basic descriptive statistics including mean, median, mode, range, variance, and standard deviation for the following variables: **AGE**, **HRS1**, **TVHOURS**, and **EDUC**. (10 points)
2. Generate histograms for the following variables: **AGE**, **HAPMAR**, and **LIFE**. (10 points)
3. What does the histogram for **AGE** tell us? Is there anything distinctive about its shape? (4 points).
4. What do you notice about the histograms for **HAPMAR** and **LIFE**? (4 points)
5. Do people in the survey generally find their lives exciting, routine, or dull? (4 points)
6. How do people describe their marriages? (4 points)
7. Approximately what percent of individuals in the sample are older than 75 years? (4 points)

Variables Definitions:

MARITAL: Current marital status

1=married, 2=widowed, 3=divorced, 4=separated, 5=never married, 9=no answer

EDUC: Years of education

0=no formal education, 1-20=years of school (up to 8 years in university study), 98=don't know, 99=no answer

HAPMAR: Taking things all together, how would you describe your marriage?

1=very happy, 2=pretty happy, 3=not too happy, 8=don't know, 9=no answer, 0=not married

LIFE: Do you find your life exciting, pretty routine, or dull?

1=exciting, 2=routine, 3=dull, 8, 9, and 0= no answer or N/A

AGE: In years

HRS1: Number of hours worked the previous week (1-89 hours):

98=don't know, 99=no answer, -1=not employed

TVHOURS: Total number of hours watched on a daily basis (0-24):

98=don't know, 99=no answer, -1=not applicable