

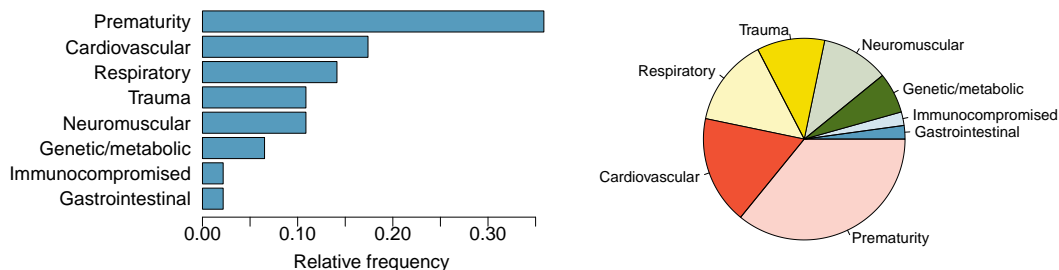
HOMEWORK №2

Math 107, Spring 2016

Due: April 5 by 4:00 pm

Problem 1

The bar chart and the pie chart below show the distribution of pre-existing medical conditions of children involved in a study on the optimal duration of antibiotic use in treatment of tracheitis, which is an upper respiratory infection.



- (a) What features are apparent in the bar chart but not in the pie chart?
- (b) What features are apparent in the pie chart but not in the bar chart?
- (c) Which graph would you prefer to use for displaying these categorical data?

Problem 2

For each of the following, describe whether you expect the distribution to be symmetric, right-skewed, or left-skewed. Also specify whether the mean or median would best represent a typical observation.

- (a) Housing prices in a country where 25% of the houses cost below \$350,000, 50% of the houses cost below \$450,000, 75% of the houses cost below \$1,000,000 and there are a meaningful number of houses that cost more than \$6,000,000.
- (b) Housing prices in a country where 25% of the houses cost below \$300,000, 50% of the houses cost below \$600,000, 75% of the houses cost below \$900,000 and very few houses that cost more than \$1,200,000.
- (c) Number of alcoholic drinks consumed by college students in a given week. Assume that most of these students don't drink since they are under 21 years old, and only a few drink excessively.
- (d) Annual salaries of the employees at a Fortune 500 company where only a few high level executives earn much higher salaries than all the other employees.

Problem 3

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey of 350,000 people in the United States. As its name implies, the BRFSS is designed to identify risk factors in the adult population and report emerging health trends. For example, respondents are asked about their diet and weekly physical activity, their HIV/AIDS status, possible tobacco use, and even their level of healthcare coverage. The BRFSS Web site (<http://www.cdc.gov/brfss>) contains a complete description of the survey, including the research questions that motivate the study and many interesting results derived from the data.

In this problem, we will focus on a random sample of 20,000 people from the BRFSS survey conducted in 2000. While there are over 200 variables in this data set, we will work with a small subset, which can be downloaded from the course website and contains the following variables:

Variable	Description
genhlth	Evaluation of general health (excellent, very good, good, fair or poor)
exerany	Have you exercised in the past month? (1 = yes, 0 = no)
hlthplan	Do you have some form of health coverage (1 = yes, 0 = no)
smoke100	Have you smoked at least 100 cigarettes in your lifetime (1 = yes, 0 = no)
height	Height (inches)
weight	Weight (pounds)
wt desire	Desired weight (pounds)
age	Age (years)
gender	Gender (m = male, f = female)

Use R to help you answer the following questions. To submit the plots, please export them from R (using the export tab in the plots pane of RStudio) and copy them into a Word document. This will allow you to submit multiple plots on one page. Please clearly label your plots.

- What is the case here?
- For each variable, indicate whether it is quantitative or categorical.
- Create a frequency table for the general health evaluations.
- Create a bar chart displaying the distribution of the general health evaluations.
- Create a histogram displaying the distribution of weights. Briefly describe this distribution, mentioning the number of modes, the shape, and the absence/presence of outliers.
- Calculate the mean and median of weight. Which is the most appropriate measure of center for these data?