HW01 - Type of Variables and Histograms

Stat 20 & 131A, Spring 2017, Prof. Sanchez

Due Jan-26

1) Identify the type of variable (qualitative or quantitative) for the list of questions from a survey applied to college students in a statistics class. *1pt*

    a. Name of student.

    b. Birth date (e.g. 10/21/1995).

    c. Age (in years).

    d. Home Address (e.g. 1234 Shattuck Ave).

    e. Telephone number (e.g. 510-123-4567).

    f. Major field of study.

    g. College year-grade: freshman, sophomore, junior, senior.

    h. Score on midterm test (based on 100 possible points).

    i. Overall grade: A, B, C, D, F.

   j. Length of time—in minutes—to complete Stat 20 final test.

2) Consider a variable with numeric values describing electronic ways of expressing personal opinions: 1 = Twitter; 2 = email; 3 = text message; 4 = Facebook; 5 = blog. Is this a quantitative or a qualitative variable? Explain. *0.25pts*

3) The histogram below shows the distribution of final scores in a certain class. *0.6pts*
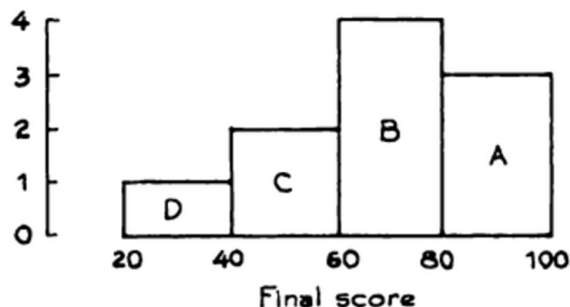

    a. Which block represents the people who scored between 60 and 80?

    b. Ten percent scored between 20 and 40. About what percentage scored between 40 and 60?

    c. About what percentage scored over 60?

4) A histogram of monthly wages for part-time employees is shown below (densities are marked in parentheses). Nobody earned more than $1,000 a month. The block over the class interval from $200 to $500 is missing. How tall must it be?. 0.2pts
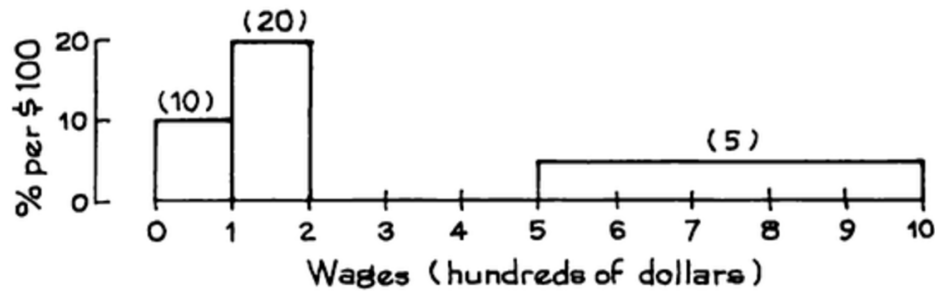

5) An investigator collects data on hourly wage rates for three groups of people. Those in group B earn about twice as those in group A. Those in group C earn about $10 an hour more than those in group A. Which histogram belongs to which group? (The histograms don't show wages above $50 an hour). 0.75pts
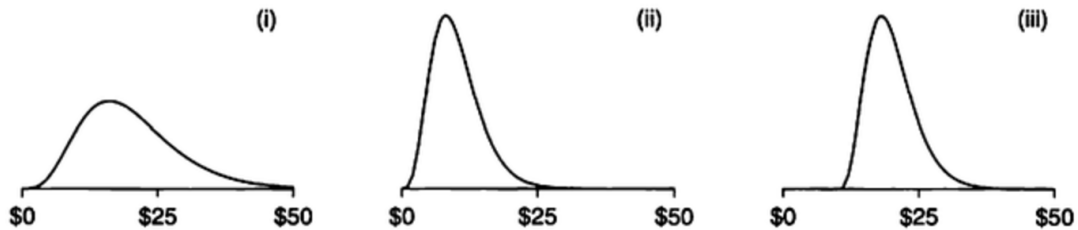
6) Three people plot histograms for the weights of subjects in a study, using the density scale. Only one is right. Which one, and why? 0.2pts
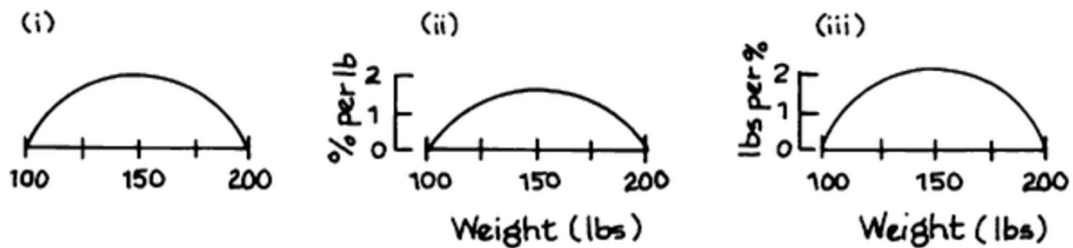
7) The figure below is a histogram showing the distribution of blood pressure for all 14,148 women.
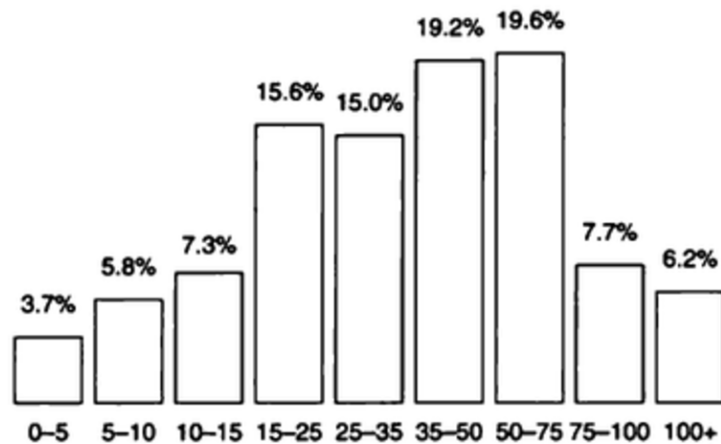3.5pts
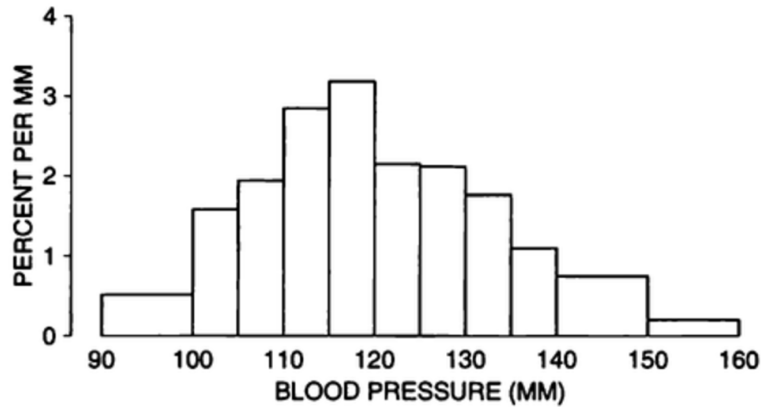
Use the histogram to answer the following questions:

  a. Is the percentage of women with blood pressures above 130 mm around 25%, 50%, or 75%?

  b. Is the percentage of women with blood pressures between 90 mm and 160 mm around 1%, 50%, or 99%?

  c. In which interval are there more women: 135-140 mm or 140-150 mm?

  d. Which interval is more crowded: 135-140 mm or 140-150 mm?

  e. On the interval 125-130 mm, the height of the histogram is about 2.1% per mm. What percentage of the women had blood pressures in this class interval?

  f. Which interval has more women: 97-98 mm or 102-103 mm?

  g. Which is the most crowded millimiter of all?

8) The figure below shows the distribution of American families by income. Ranges include the left endpoint but not the right. For example, 3.7% of the families had incomes in the range $0-$4,999, 5.8% had incomes in the range $5,000-$9,999, and so forth. True or False, and explain. $_{1.5pts}$

  a. Although American families are not spread evenly over the whole income range, the families that earn between $10,000 and $35,000 are spread fairly evenly over that range.

  b. The families that earn between $35,000 and $75,000 are spread fairly evenly over that range.

  c. The graph is a histogram.

The histogram bars show, from left to right:
- 0–5: 3.7%
- 5–10: 5.8%
- 10–15: 7.3%
- 15–25: 15.6%
- 25–35: 15.0%
- 35–50: 19.2%
- 50–75: 19.6%
- 75–100: 7.7%
- 100+: 6.2%

9) This problem involves working with R, and it assumes that you have read the script `03-histograms.pdf` available in the course's github repository. *2pts*

Consider the following vector `x` with 1000 random numbers.

```r
# run this code to generate the vector x
set.seed(9875)
size = 1000
x = runif(n = size, min = 0, max = 10)
```

Use R to make two density histograms. The first histogram should have the following interval classes (left endpoint included)
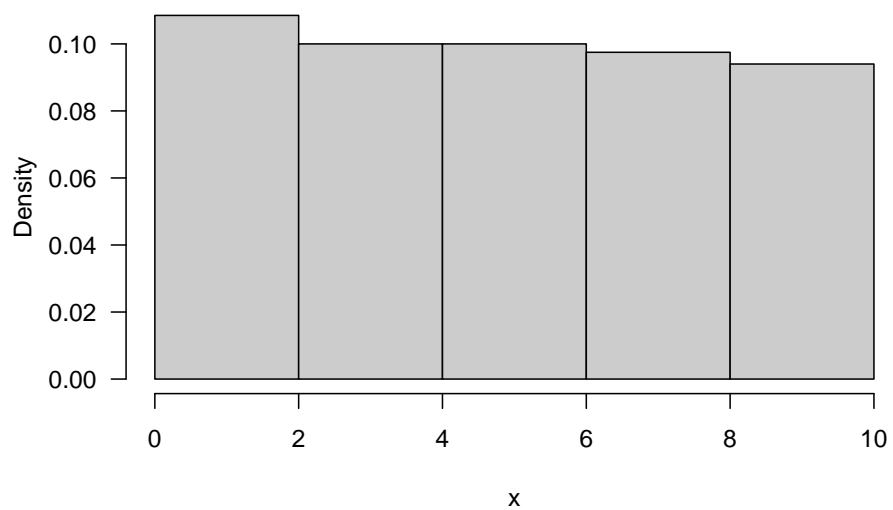
- `[0, 2)`
- `[2, 4)`
- `[4, 6)`
- `[6, 8)`
- `[8, 10)`

The second histogram should have the following interval classes (right endpoint included):

- `(0, 1]`
- `(1, 2]`
- `(2, 4]`
- `(4, 7]`
- `(7, 10]`

Your histograms should look like these:

4

Histogram of x

Histogram of x