

Instructions: Show all work to receive full credit. You should note any formulas used or calculator functions used, their inputs and outputs. I cannot grade work if I don't know where an answer came from. Be sure complete all parts of each questions, including requests for interpretation and explanations. Be as thorough as possible.

1. A researcher is interested in finding out about the buying habits of the American public. She decides to conduct a survey of 50 shoppers at a local mall, and asks them how much money they are planning to spend on gifts for an upcoming holiday. (3 points each)
- a. What is the population of interest in this study?

American shoppers

- b. What is the sample used in this study?

50 shoppers at local mall

- c. What is the parameter being studied?

their buying habits / amount of money American Shoppers will spend each

- d. What statistic is the researcher likely to use to estimate the parameter?

average spent by those in the sample

2. For each of the following variables, determine i) is the variable qualitative or quantitative? ii) the level of measurement: nominal, ordinal, interval, or ratio? iii) if the variable is quantitative, is it discrete or continuous? (4 points each)

- a. The time of day (i.e. 7:30 p.m.)

quantitative, continuous, interval

- b. Car model

qualitative, nominal

c. Phone number

qualitative, nominal

d. Number of students in a class

quantitative, discrete, ratio

3. A researcher notes a relationship in a study between getting married and weight gain. What is an example of a possible confounding variable the researcher would want to control for? Why? (4 points)

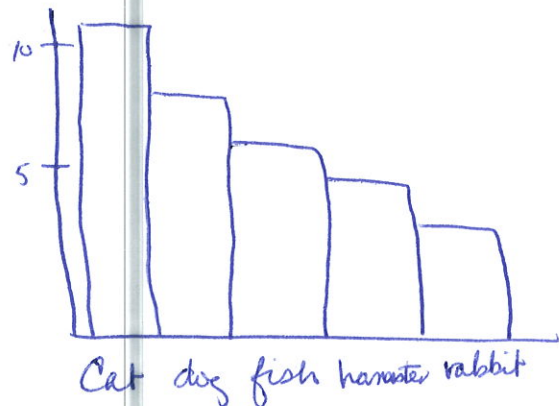
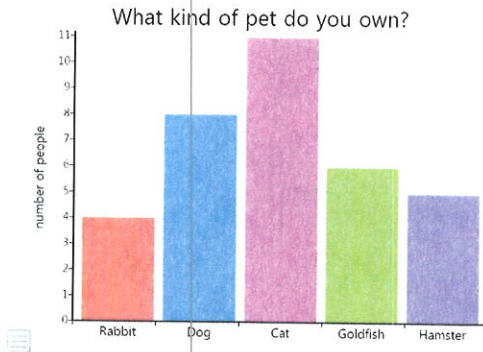
age (older people are both more likely to be married and to gain weight)

answers may vary

4. Why does a voluntary response sample tend to introduce bias into the results of a study? (5 points)

only people of very strong opinions are likely to respond

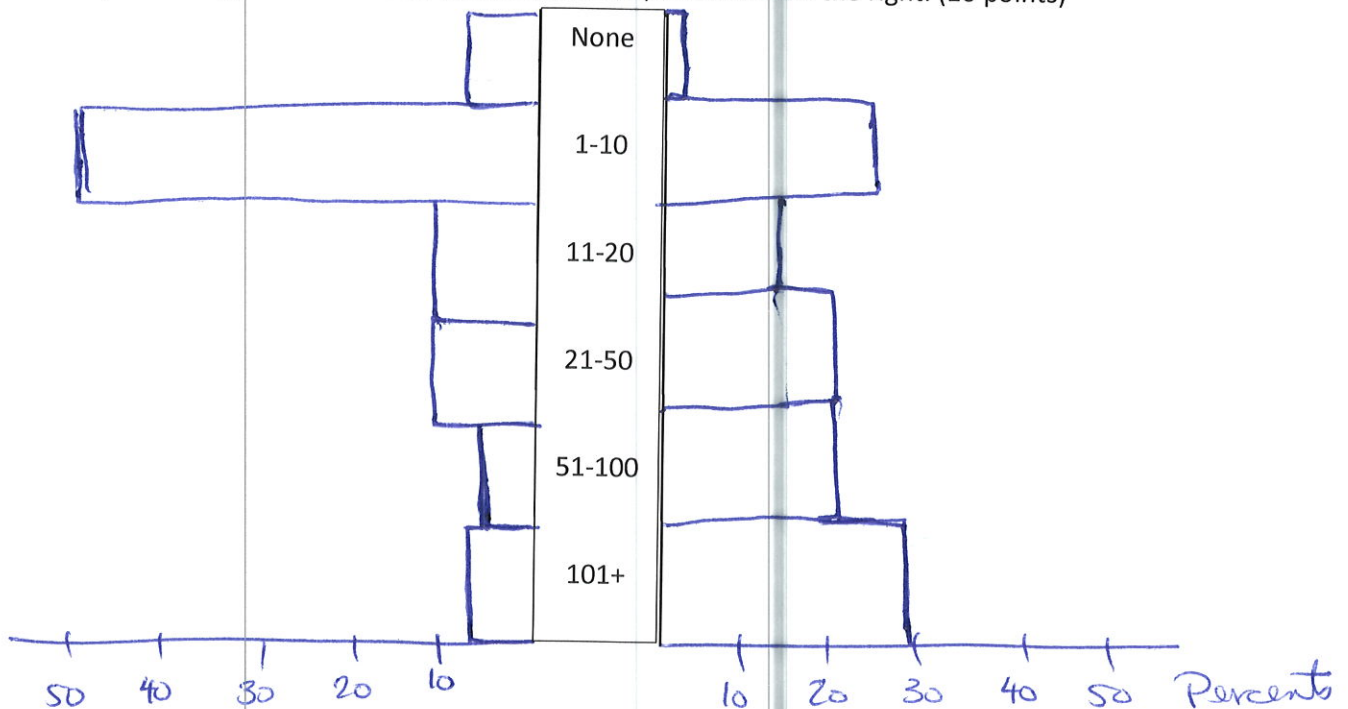
5. Below is a bar chart of the type of pets a sample of people own. In the space next to the graph, convert this graph to a Pareto chart. (8 points)



6. Use the data in the table below to construct a comparative bar chart/histogram following the steps indicated.
- a. Compute the relative frequencies for adults and teens and record them in the table. Round the relative frequencies/percentages to one decimal place. (12 points)

| Number of Texts | Adult Frequency <i>Total - 1936</i> | Percent of Adults | Teen Frequency <i>Total - 627</i> | Percent of Teens |
|-----------------|--|-------------------|--------------------------------------|------------------|
| None | 173 | 8.9% | 13 | 2.1% |
| 1-10 | 978 | 50.5% | 138 | 22.0% |
| 11-20 | 249 | 12.9% | 69 | 11% |
| 21-50 | 249 | 12.9% | 113 | 18% |
| 51-100 | 134 | 6.9% | 113 | 18% |
| 101+ | 153 | 7.9% | 181 | 28.9% |

- b. Using the relative frequencies from the table above, construct the comparative bar chart/histogram with the number of texts as the vertical axis down the middle of the graph (as shown), and the bars for adults on the left, and teens on the right. (10 points)



- c. Use this information to compare the texting habits of adults and teens. (4 points)
- teens are much more likely to text more often*
- d. Explain why breaking up the numerical categories in this fashion could be considered misleading. (4 points)

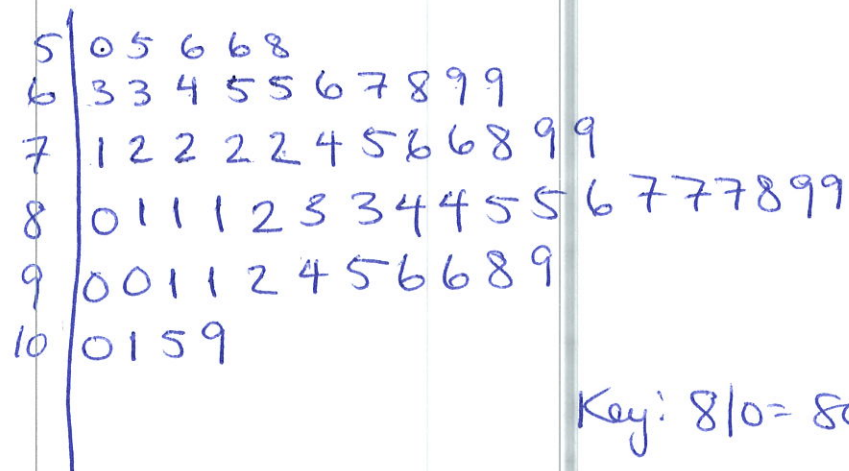
*the category widths are not equal
this distorts the shape of the distribution*

7. A sample of retirees were asked their age. The results are shown below in a table.

Ages of Retirees

| | | | | |
|-----|-----|----|-----|-----|
| 79 | 80 | 86 | 72 | 76 |
| 50 | 81 | 71 | 96 | 81 |
| 88 | 78 | 72 | 65 | 95 |
| 74 | 105 | 55 | 83 | 63 |
| 56 | 68 | 65 | 90 | 69 |
| 58 | 98 | 64 | 84 | 91 |
| 72 | 75 | 94 | 91 | 79 |
| 56 | 82 | 90 | 85 | 87 |
| 99 | 72 | 96 | 76 | 109 |
| 83 | 81 | 89 | 67 | 69 |
| 100 | 89 | 84 | 101 | 92 |
| 63 | 87 | 85 | 87 | 66 |

Use the data in the table to construct a stemplot of the data. [Hint: You may want to use your calculator to sort the data first.] Be sure to include a key in your graph. (12 points)



8. For the same data as in Problem #7, find the following statistics:

- a. The mean, median and mode (6 points)

$$\bar{x} = 79.98 \approx 80.0$$

$$\tilde{x} = 81$$

$$\text{mode} = 72$$

- b. The standard deviation and range (4 points)

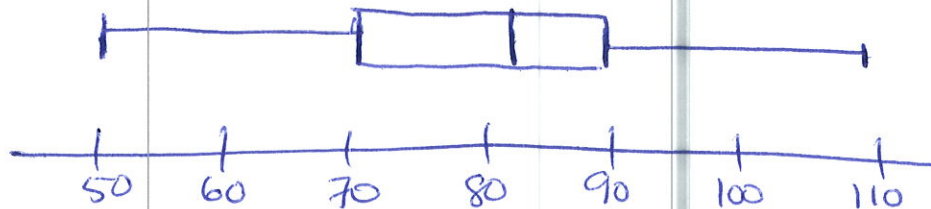
$$s_x = 13.44$$

$$\text{range} = 109 - 50 = 59$$

- c. Calculate the five-number summary for this data. (4 points)

$$\begin{array}{ll} \text{min} = 50 & Q_3 = 89.5 \\ Q_1 = 70 & \text{Max} = 109 \\ \text{Med} = 81 & \end{array}$$

- d. Use that information to construct a simple box plot. Be sure your graph is to scale. (6 points)

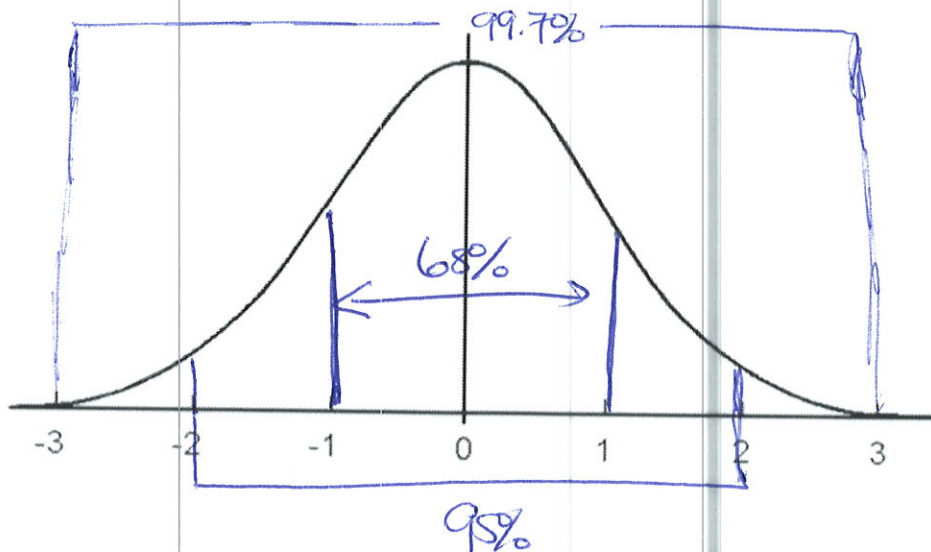


- e. Based on this information, what percentile is a 72-year-old retiree? (5 points)

$$\frac{17}{60} = 28.3\% \rightarrow \frac{20}{60} = 33.3$$

28th → 33rd percentile

9. Sketch the Empirical Rule on the graph below. Label the graph clearly. (7 points)



10. If the mean of a normal distribution is 72 points with a standard deviation of 5 points, find the following: (4 points each)

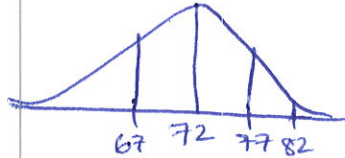
- a. The z-score of 88 points.

$$\frac{88 - 72}{5} = 3.2$$

b. What percentile 88 points represents.

$$\text{normalcdf}(-E99, 88, 72, 5) = 99.9\%$$

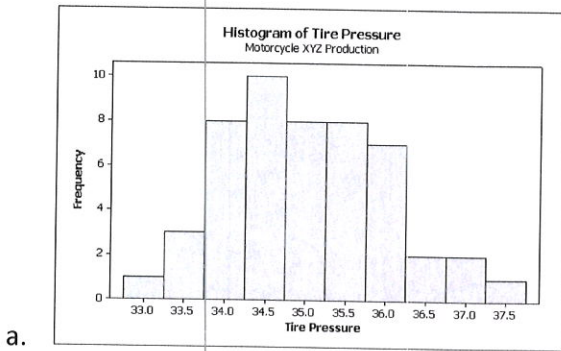
c. The area under the normal curve between 67 and 82 using the Empirical Rule.



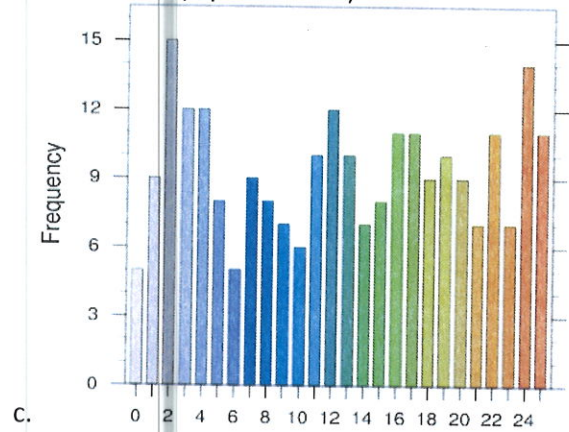
$$\frac{95-68}{2} = 13.5$$

$$68 + 13.5 = 81.5\%$$

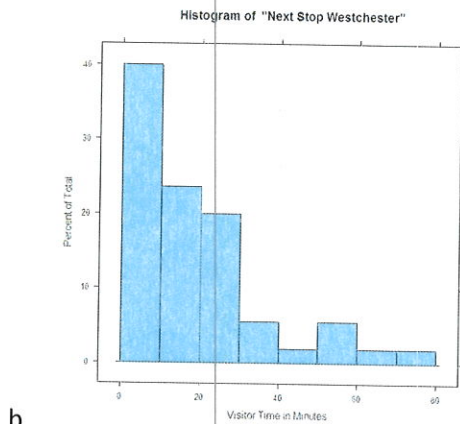
11. For each of the distributions shown below, determine if the distribution is roughly symmetric, left skewed, right skewed, roughly uniform, or none of these. (3 points each)



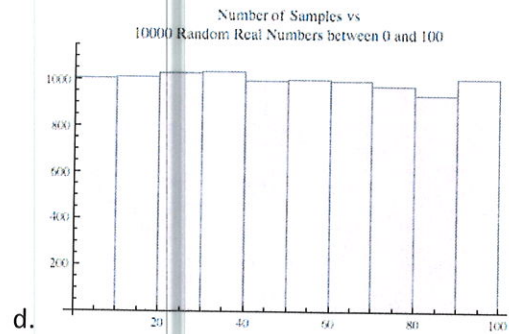
roughly symmetric



none of these



right skewed



uniform

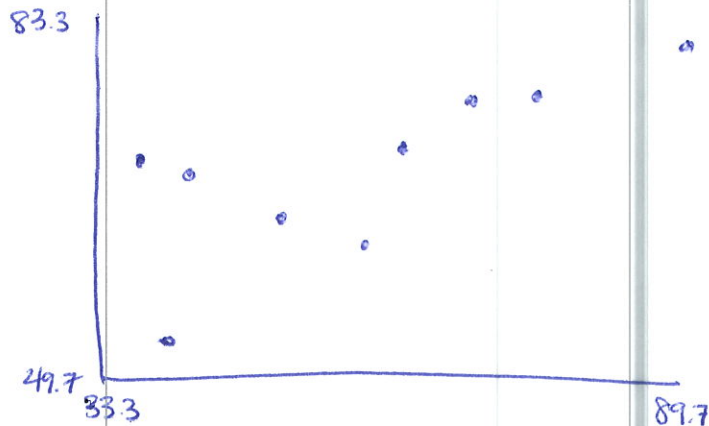
12. Use the data below to answer the questions that follow.

| Song Volume (db) | Age (years) | Weight (g) | Temperature (°F) |
|------------------|-------------|------------|------------------|
| 44 | 1.2 | 74 | 67 |
| 62 | 2.4 | 98 | 72 |
| 38 | 1 | 65 | 68 |
| 65 | 2.2 | 64 | 74 |
| 71 | 2.1 | 83 | 75 |
| 52 | 1.8 | 54 | 65 |
| 43 | 1.4 | 62 | 54 |
| 85 | 2.7 | 73 | 79 |
| 59 | 1.7 | 89 | 63 |

L1

L2

- a. Construct a scatterplot of song volume versus temperature. Sketch the result below. (5 points)



- b. Use the data of song volume and temperature to find linear regression line that models the data. Write the equation in slope-intercept form here. (8 points)

$$y = .376x + 46.86$$

- c. What is the correlation coefficient? Is the fit of the data to the line strong, moderate or weak? (5 points)

$$r = .756$$

fairly strong or strong-moderate

- d. What proportion of the change in song volume can be accounted for only by the change in temperature? (4 points)

about 57% (r^2 value is .572)