

Instructions:

You must print out this homework assignment and complete all work in space provided. Do not turn in work completed on separate pieces of paper. Staple the pages together. This is to be turned in at the beginning of class.

You must show all work. Correct answers with no accompanying work showing how the solutions were obtained are not given credit. Draw a box around your final answer.

1. At 100 randomly selected points in the ocean, I measure the temperature and salinity of the water.

- How many variables do I have? 2
- Are they categorical or quantitative? quantitative
- To display all the information in my dataset, what visual display(s) can I use? Scatterplot

2. I select 75 adults and ask them whether or not they are registered to vote.

- How many variables do I have? 1
- Are they categorical or quantitative? categorical
- To display all the information in my dataset, what visual display(s) can I use? bar graph, pie chart

3. I record the gender and state of birth of 100 adults.

- How many variables do I have? 2
- Are they categorical or quantitative? categorical
- To display all the information in my dataset, what visual display(s) can I use? contingency table

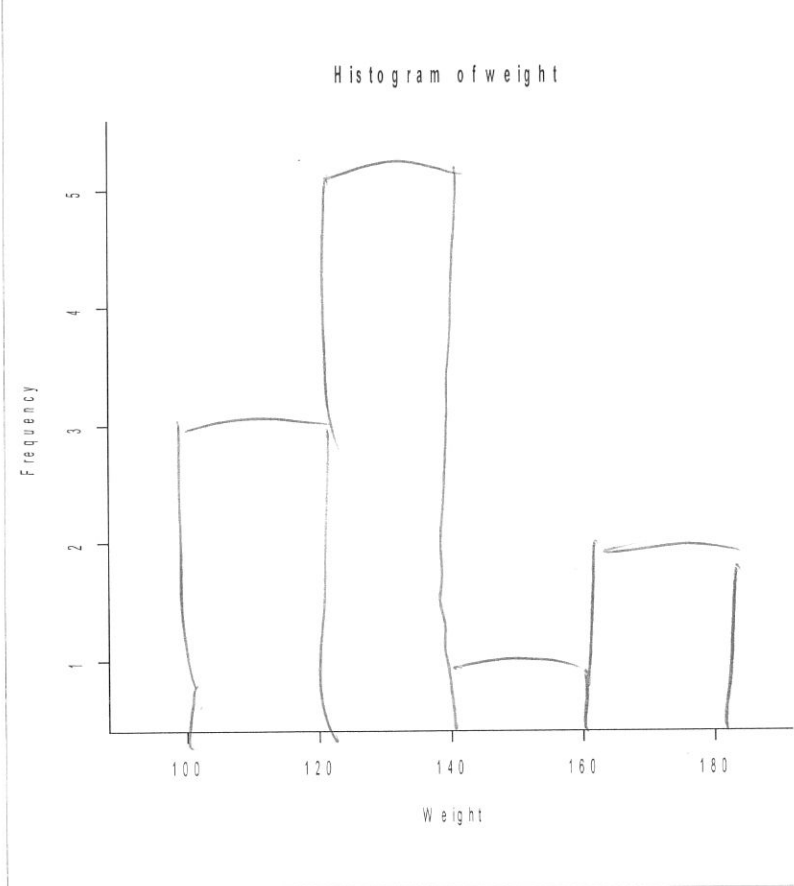
4. I record the weight of 80 bushels of rice.

- How many variables do I have? 1
- Are they categorical or quantitative? quantitative
- To display all the information in my dataset, what visual display(s) can I use? histogram, stem & leaf plot, boxplot

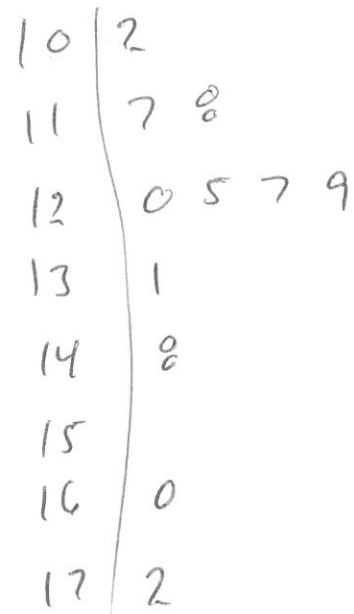
5. The following data represents the weight in lb in a random sample of adults

Data: 102 117 118 120 125 127 129 131 148 160 172

a) Use the following axes to draw a histogram using the above data

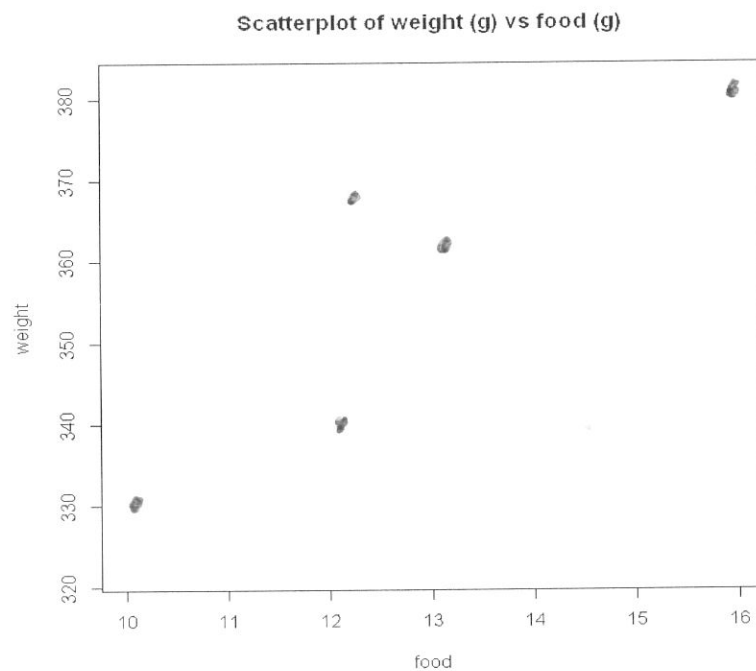


b) Draw a stemplot using the above data



6. The grams of food given to 5 rats on a daily basis was recorded, as well as their weight, resulting in the data below. Using the axes given create a scatterplot with the given data

Food (g)	Weight (g)
10	330
12	340
12	368
13	360
16	380



7. 24 elderly subjects who underwent invasive surgery were given either general anesthesia or spinal anesthesia. We recorded whether or not they had survived 1 week from surgery.

Observation Number	Anesthesia: 0 = General 1 = Spinal	Survived 0 = no 1 = yes
1	0	0
2	0	0
3	0	1
4	0	0
5	0	0
6	0	1
7	0	0
8	0	1
9	0	0
10	0	1
11	0	1
12	0	1
13	0	1
14	0	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	0
20	1	1
21	1	1
22	1	1
23	1	0
24	1	1

- How many variables are in this dataset? 2
- Are they categorical or quantitative?
- Use the data to fill in the contingency table below

		Anesthesia		
Survived		General	Spinal	Total
Yes		0	0	16
No		6	2	8
Total		14	10	24

- What proportion of subjects survived? $\frac{16}{24} = 0.67$
- What proportion of subjects were given general anesthesia? $\frac{14}{24} = 0.58$
- What proportion of subjects given spinal anesthesia died? $\frac{2}{10} = 0.20$
- What proportion of subjects that died were given spinal anesthesia? $\frac{2}{8} = 0.25$
- What proportion of subjects died and were given spinal anesthesia? $\frac{2}{24} = 0.08$

- Construct a conditional contingency table, showing the distribution of survival, by anesthesia.

$$\frac{8}{14} = 0.57$$

$$\frac{6}{14} = 0.43$$

	Gen	Sp	
Yes	0.57	0.0	
No	0.43	0.2	
T	1	1	

$$\frac{8}{10} = 0.8$$

$$\frac{2}{10} = 0.2$$

- Which group appears to have the highest risk of dying after surgery, those given general anesthesia or those given spinal anesthesia?

General

- Do survival and anesthesia appear to be independent or dependent variables?

dependent

8. Data: 14 17 24 29 | 31 32 35 36 (37) 37 39 39 41 | 45 58 61 64

- Find the 5-number summary $n=17$ Loc of $M = \frac{n+1}{2} = \frac{17+1}{2} = 9$ 8 obs in each half

$$\text{Loc of quartile} = \frac{8+1}{2} = 4.5$$

Min	Q_1	M	Q_3	Max
14	30	37	43	64

$$Q_1 = \frac{29+31}{2} = 30$$

$$Q_3 = \frac{41+45}{2} = 43$$

- Draw a boxplot using the axis below. Make sure to check for outliers using the appropriate formulas provided in class.

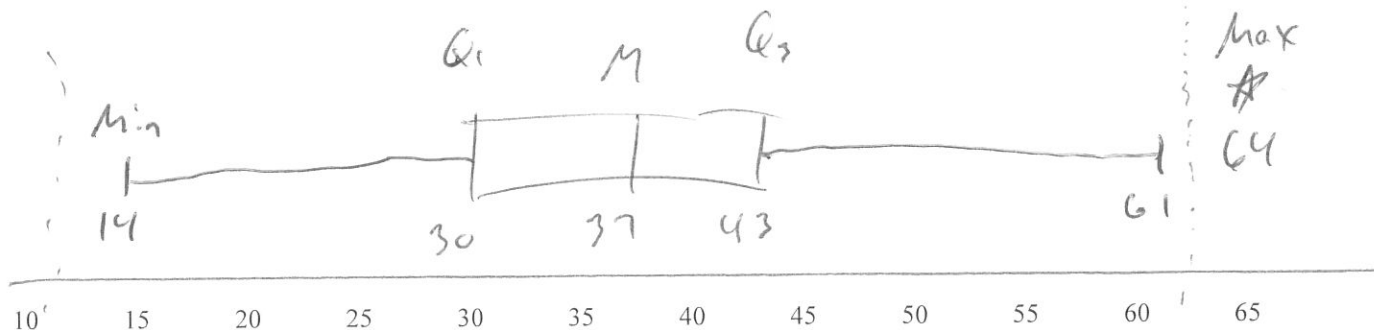
$$IQR = Q_3 - Q_1 = 43 - 30 = 13$$

$$\text{lower fence} = Q_1 - 1.5(IQR) = 30 - 1.5(13) = 10.5$$

$$\text{upper fence} = Q_3 + 1.5(IQR) = 43 + 1.5(13) = 62.5$$

10.5

62.5



9. Using the following data, calculate the standard deviation by hand (show all work!). You should check your answer using your calculator.

Data: 9 14 10 15

$$\bar{X} = \frac{9+14+10+15}{4} = 12$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(9-12)^2 + (14-12)^2 + (10-12)^2 + (15-12)^2}{4-1}} = \sqrt{\frac{(-3)^2 + (2)^2 + (-2)^2 + (3)^2}{3}}$$

$$= \sqrt{\frac{26}{3}} = \boxed{2.94}$$

10. Using the data in the histogram below, find the bins in which are located Q1, M, and Q3.

$$n = 3+4+4+6+7+8+3+1 = 36$$

$$\text{Loc of } M = \frac{n+1}{2} = \frac{36+1}{2} = 18.5$$

18 obs in each half

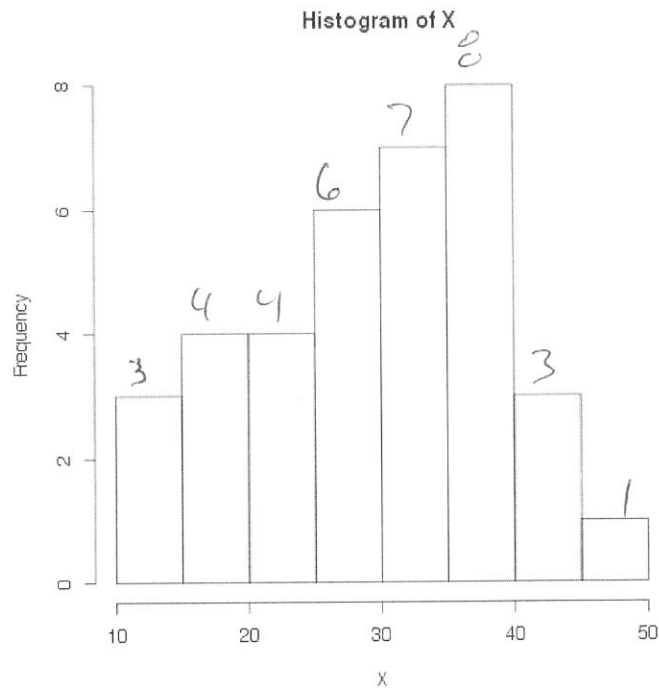
Loc of quartiles:

$$\frac{18+1}{2} = 9.5$$

$$Q_1: (20-25)$$

$$M: (30-35)$$

$$Q_3: (35-40)$$



11. At the company at which Susan works, Susan is doing data entry, entering the hours worked for all employees into a spreadsheet. However, Susan is not very good at data entry, and often makes mistakes, entering 40 as 400, for example. You would like to have some numerical measure of the center and variability of the number of hours worked for all employees at this company. Being aware of Susan's poor data entry skills, state the appropriate measure of:

center: Median

variability: IQR

data contains outliers,
use robust statistics

12. Write down the definition of robust statistics (hint: find the definition of robust statistics in your notes, and write that definition down verbatim for the correct answer).

Statistics which are not significantly affected by outliers or skewed data

13. Read the following: For this coming quiz and all remaining quizzes, remember to review your notes before the quiz, not just the homework. I do not guarantee homework will exhaustively cover everything in your notes, though you are still expected to know everything in your notes for a quiz. For example, not seen in this homework assignment are questions relating to the shape of a histogram (is it unimodal, bimodal, multimodal, and is it skewed left or right). (hint: I will ask you a question regarding the shape of a histogram on quiz 1.)