

Exercises for Section 1.2

10. A sample of 100 adult women was taken, and each was asked how many children she had. The results were as follows:

Children	0	1	2	3	4	5
Number of Women	27	22	30	12	7	2

- Find the sample mean number of children.
- Find the sample standard deviation of the number of children.
- Find the sample median of the number of children.
- What is the first quartile of the number of children?
- What proportion of the women had more than the mean number of children?
- For what proportion of the women was the number of children more than one standard deviation greater than the mean?
- For what proportion of the women was the number of children within one standard deviation of the mean?

Solution:

- $$\sum x = 27(0) + 22(1) + 30(2) + 12(3) + 7(4) + 2(5) = 156$$

$$\therefore \bar{x} = \sum x/n = 156/100 = 1.56$$
- $$\sum x^2 = 27(0^2) + 22(1^2) + 30(2^2) + 12(3^2) + 7(4^2) + 2(5^2) = 742 + 252 = 412$$

$$s^2 = \frac{1}{n-1} (\sum x^2 - n\bar{x}^2)$$

$$= \frac{1}{99} [412 - 100(1.56^2)]$$

$$= 1.7034$$

$$s = \sqrt{s^2} = \sqrt{1.7034} = 1.3052$$
- Position for median = $(n+1)(.5) = (101)(.5) = 50.5$
The sample median is the average of the 50th and 51st value.
$$x_{50} = 2; x_{51} = 2; \therefore \text{Median} = 2$$
- Position for 1st quartile = $(n+1)(.25) = (101)(.25) = 25.25$
The sample median is the average of the 25th and 26th value.
$$x_{25} = 0; x_{26} = 0; \therefore \text{1st quartile} = 0$$
- $(30+12+7+2)$ women had more than the mean number of children.
$$\therefore \text{proportion} = 51/100 = 51\%$$
- $\bar{x} + s = 1.56 + 1.3052 = 2.8652$
 $(12+7+2)$ women had more than 2.86 children.
$$\therefore \text{proportion} = 21/100 = 21\%$$
- $\bar{x} - s = 1.56 - 1.3052 = 0.2548$
 $(22+30)$ women had between 0.2548 and more than 2.86 children. $\therefore \text{proportion} = 52/100 = 52\%$

14. There are 10 employees in a particular division of a company. Their salaries have a mean of \$70,000, a

median of \$55,000, and a standard deviation of \$60,000. The largest number on the list is \$100,000. By accident, this number is changed to \$1,000,000.

- What is the value of the mean after the change?
- What is the value of the median after the change?
- What is the value of the standard deviation after the change?

Solution:

- $$\text{original } \bar{x} = 70000 = \sum x/n$$

$$\therefore \text{original } \sum x = 70000(10) = 700000$$

$$\text{modified } \sum x = 700000 - 100000 + 1000000$$

$$= 1,600,000$$

$$\text{modified } \bar{x} = 1600000/10 = 160,000$$
- Median remains the same.
- $$\text{original } s^2 = (60000)^2 = \frac{1}{n-1} [\sum x^2 - n\bar{x}^2]$$

$$\therefore \text{orig. } \sum x^2 = (10-1)(60000)^2 + (10)(70000)^2$$

$$= 81,400,000,000$$

$$\text{mod. } \sum x^2 = 81400000000 - 100000^2 + 1000000^2$$

$$= 1.0714E12$$

$$\text{mod. } s^2 = \frac{1}{n-1} [\sum x^2 - n\bar{x}^2]$$

$$= \frac{1}{9} [1.0714E12 - (10)160000^2]$$

$$= 9.06E11$$

$$\text{modified } s = \sqrt{s^2} = \sqrt{9.06E11} = 300998.34$$

16. In each of the following data sets, tell whether the outlier seems certain to be due to an error or whether it could conceivably be correct.

- A rock is weighed five times. The readings in grams are 48.5, 47.2, 4.91, 49.5, 46.3.
- A sociologist samples five families in a certain town and records their annual income. The incomes are \$34,000, \$57,000, \$13,000, \$1,200,000, \$62,000.

Solution:

- It seems to be an error.
- It could happen.

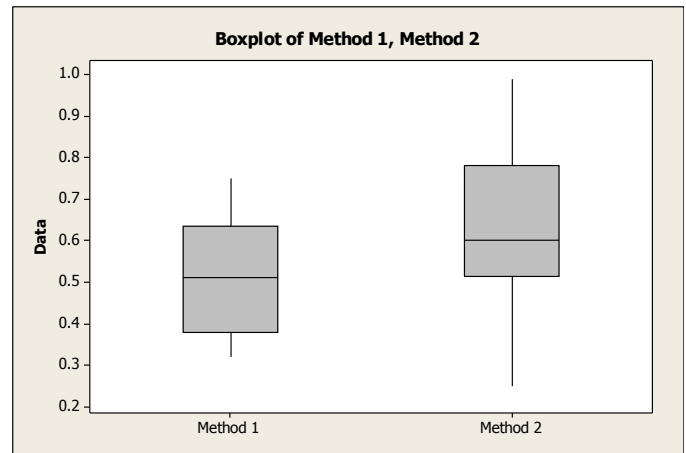
Exercises for Section 1.3

4. Two methods were studied for the recovery of protein. Thirteen runs were made using each method, and the fraction of protein recovered was recorded for each run. The results are as follows:

Method 1	Method 2
0.32	0.25
0.35	0.40
0.37	0.48

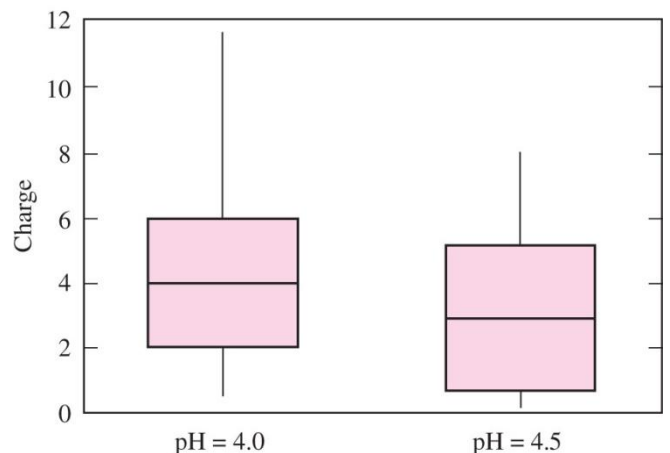
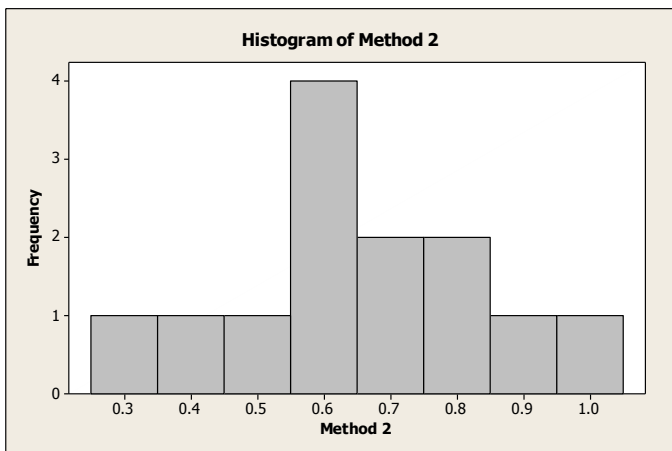
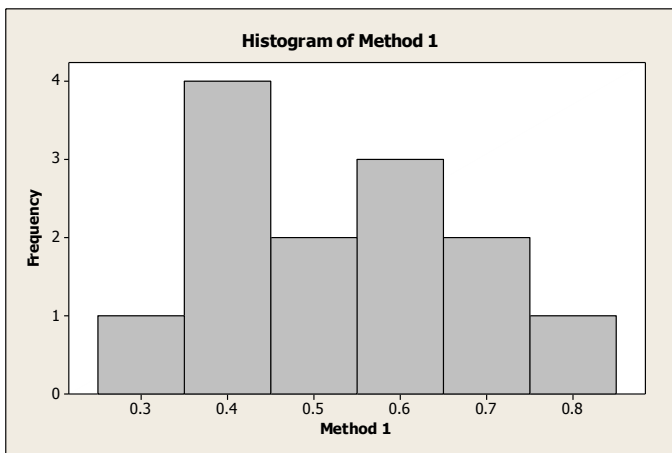
0.39	0.55
0.42	0.56
0.47	0.58
0.51	0.60
0.58	0.65
0.60	0.70
0.62	0.76
0.65	0.80
0.68	0.91
0.75	0.99

- Construct a histogram for the results of each method.
- Construct comparative boxplots for the two methods.
- Using the boxplots, what differences can be seen between the results of the two methods?



12. Following are boxplots comparing the charge (in coulombs per mole $[C/mol] \times 10^{-25}$) at pH 4.0 and pH 4.5 for a collection of proteins (from the article "Optimal Synthesis of Protein Purification Processes," E. Vasquez-Alvarez, M. Leinqueo, and J. Pinto, *Biotechnology Progress* 2001:685-695). True or false:

- The median charge for the pH of 4.0 is greater than the 75th percentile of charge for the pH of 4.5.
- Approximately 25% of the charges for pH 4.5 are less than the smallest charge at pH 4.0.
- About half the sample values for pH 4.0 are between 2 and 4.
- There is a greater proportion of values outside the box for pH 4.0 than for pH 4.5.
- Both samples are skewed to the right.
- Both samples contain outliers.



Answer:

- False
- True
- False
- False
- True
- False