Chapter 3
Descriptive Statistics

LEARNING OBJECTIVES

The focus of Chapter 3 is the use of statistical techniques to describe data, thereby enabling you to:

1. Apply various measures of central tendency—including the mean, median, and mode—to a set of ungrouped data
2. Apply various measures of variability—including the range, interquartile range, mean absolute deviation, variance, and standard deviation (using the empirical rule and Chebyshev’s theorem)—to a set of ungrouped data
3. Compute the mean, mode, standard deviation, and variance of grouped data
4. Describe a data distribution statistically and graphically using skewness, kurtosis, and box-and-whisker plots
5. Use computer packages to compute various measures of central tendency, variation, and shape on a set of data, as well as to describe the data distribution graphically

CHAPTER OUTLINE

3.1 Measures of Central Tendency: Ungrouped Data
- Mode
- Median
- Mean
- Percentiles
- Quartiles

3.2 Measures of Variability - Ungrouped Data
- Range
- Interquartile Range
- Mean Absolute Deviation, Variance, and Standard Deviation
- Mean Absolute Deviation
- Variance
- Standard Deviation
- Meaning of Standard Deviation
  - Empirical Rule
  - Chebyshev's Theorem
  - Population Versus Sample Variance and Standard Deviation
  - Computational Formulas for Variance and Standard Deviation
- *z* Scores
- Coefficient of Variation

3.3 Measures of Central Tendency and Variability - Grouped Data
- Measures of Central Tendency
  - Mean
  - Median
  - Mode
- Measures of Variability

3.4 Measures of Shape
- Skewness
  - Skewness and the Relationship of the Mean, Median, and Mode
  - Coefficient of Skewness
- Kurtosis
- Box and Whisker Plot and Five-Number Summary

3.5 Descriptive Statistics on the Computer

KEY TERMS

Arithmetic Mean
Bimodal
Box and Whisker Plot
Chebyshev’s Theorem
Coefficient of Skewness
Coefficient of Variation (CV)
Deviation from the Mean
Empirical Rule
Five-Number Summary
Interquartile Range
Kurtosis
Leptokurtic
Mean Absolute Deviation (MAD)
Measures of Central Tendency
Measures of Shape
Measures of Variability
Median
Mesokurtic
Mode
Multimodal
Percentiles
Platykurtic
Quartiles
Range
Skewness
Standard Deviation
Sum of Squares of *x*
Variance
z Score

STUDY QUESTIONS

1. Statistical measures used to yield information about the center or middle part of a group of numbers are called ________________.

2. The "average" is the ________________.

3. The value occurring most often in a group of numbers is called ________________.

4. In a set of 110 numbers arranged in order, the median is located at the ________________ position.

5. If a set of data has an odd number of values arranged in ascending order, the median is the _________________ value.

   Consider the data: 5, 4, 6, 6, 4, 5, 3, 2, 6, 4, 6, 3, 5
   Answer questions 6-8 using this data.

6. The mode is ________________.

7. The median is ________________.

8. The mean is ________________.

9. If a set of values is a population, then the mean is denoted by _______________.

10. In computing a mean for grouped data, the _______________ is used to represent all data in a given class interval.

11. The mean for the data given below is _______________.

Class Interval	Frequency
50 - under 53	14
53 - under 56	17
56 - under 59	29
59 - under 62	31
62 - under 65	18

12. Measures of variability describe the _______________ of a set of data.

Use the following population data for Questions 13-17:

27 65 28 61 34 91 61 37 58 31
43 47 44 20 48 50 49 43 19 52

13. The range of the data is _______________.

14. The value of Q_1 is _____________, Q_2 is ____________, and Q_3 is ____________.

15. The interquartile range is _____________.

16. The value of the 34^{th} percentile is ________________.

17. The value of the Pearsonian coefficient of skewness for these data is ________________.

18. The Mean Absolute Deviation is computed by averaging the _______________ of deviations around the mean.

19. Subtracting each value of a set of data from the mean produces _______________ from the mean.

20. The sum of the deviations from the mean is always _______________.

21. The variance is the _______________ of the standard deviation.

22. The <u>population</u> variance is computed by using _______________ in the denominator. Whereas, the <u>sample</u> variance is computed by using _______________ in the denominator.

23. If the sample standard deviation is 9, then the sample variance is _______________.

Consider the data below and answer questions 24-26 using the data:

2, 3, 6, 12

24. The mean absolute deviation for this data is ________________.

25. The sample variance for this data is ________________.

26. The population standard deviation for this data is ________________.

27. In estimating what proportion of values fall within so many standard deviations of the mean, a researcher should use ________________ if the shape of the distribution of numbers is unknown.

28. Suppose a distribution of numbers is mound shaped with a mean of 150 and a variance of 225. Approximately ________________ percent of the values fall between 120 and 180. Between ________________ and ________________ fall 99.7% of these values.

29. The shape of a distribution of numbers is unknown. The distribution has a mean of 275 and a standard deviation of 12. The value of *k* for 299 is ________________. At least ________________ percent of the values fall between 251 and 299.

30. Suppose data are normally distributed with a mean of 36 and a standard deviation of 4.8. The *z* score for 30 is ________________. The *z* score for 40 is ________________.

31. A normal distribution of values has a mean of 74 and a standard deviation of 21. The coefficient of variation for this distribution is ________________?

Consider the data below and use the data to answer questions 32-35.

Class Interval	Frequency
2- 4	5
4- 6	12
6- 8	14
8-10	15
10-12	8
12-14	4

32. The sample variance for the data above is ________________.

33. The population standard deviation for the data above is ________________.

34. The mode of the data is __________________________.

35. The median of the data is __________________________.

36. If a unimodal distribution has a mean of 50, a median of 48, and a mode of 47, the distribution is skewed ________________.

37. If the value of S_k is positive, then it may be said that the distribution is _________________________ skewed.

38. The peakedness of a distribution is called __________________________.

39. If a distribution is flat and spread out, then it is referred to as ________________________; if it is "normal" in shape, then it is referred to as __________________________; if it is high and thin, then it is referred to as ___________________________.

40. In a box plot, the inner fences are computed by _____________________ and ______________________. The outer fences are computed by ____________________ and ______________________.

41. Data values that lie outside the mainstream of values in a distribution are referred to as ___________________.

ANSWERS TO STUDY QUESTIONS

1. Measures of Central Tendency
2. Mean
3. Mode
4. 55.5th
5. Middle
6. 6
7. 5
8. 4.54
9. μ
10. Class Midpoint
11. 58.11
12. Spread or Dispersion
13. 72
14. $Q_1 = 32.5$, $Q_2 = 45.5$, $Q_3 = 55$
15. IQR = 22.5
16. $P_{34} = 37$
17. $S_k = -0.018$
18. Absolute Value
19. Deviations
20. Zero
21. Square
22. N, $n - 1$
23. 81
24. 3.25
25. 20.25
26. 3.897
27. Chebyshev's Theorem
28. 95, 105, and 195
29. 2, 75
30. –1.25, 0.83
31. 28.38%
32. 7.54
33. 2.72
34. 9
35. 7.7143
36. Right
37. Positively
38. Kurtosis
39. Platykurtic, Mesokurtic, Leptokurtic
40. Q_1 - 1.5 IQR and Q_3 + 1.5 IQR
    Q_1 - 3.0 IQR and Q_3 + 3.0 IQR
41. Outliers

SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 3

3.1 **Median**

Arrange in ascending order:
2, 2, 3, 3, 4, 4, 4, 4, 5, 6, 7, 8, 8, 8, 9
There are 15 terms.
Since there are an odd number of terms, the median is the middle number.
The median = **4**

Using the formula, the median is located
at the $\frac{n+1}{2}^{\text{th}}$ term = $\frac{15+1}{2}$ = 8$^{\text{th}}$ term
The 8$^{\text{th}}$ term = **4**

Mode

2, 2, 3, 3, 4, 4, 4, 4, 5, 6, 7, 8, 8, 8, 9
The mode = **4**
4 is the most frequently occurring value

3.3 Mean

17.3
44.5 $\mu = \Sigma x/N = (333.6)/8 =$ **41.7**
31.6
40.0
52.8 $\bar{x} = \Sigma x/n = (333.6)/8 =$ **41.7**
38.8
30.1
78.5 (It is not stated in the problem whether the
$\Sigma x = 333.6$ data represent as population or a sample).

3.5 Rearranging the data into ascending order:

11, 13, 16, 17, 18, 19, 20, 25, 27, 28, 29, 30, 32, 33, 34

$$i = \frac{35}{100}(15) = 5.25$$

P_{35} is located at the 5 + 1 = 6th term, P_{35} = **19**

$$i = \frac{55}{100}(15) = 8.25$$

P_{55} is located at the 8 + 1 = 9th term, P_{55} = **27**

$Q_1 = P_{25}$ but $i = \frac{25}{100}(15) = 3.75$

$Q_1 = P_{25}$ is located at the 3 + 1 = 4th term, Q_1 = **17**

Q_2 = Median but: The median is located at the $\left(\frac{15+1}{2}\right)^{th} = 8^{th}\, term$

Q_2 = **25**

$Q_3 = P_{75}$ but $i = \frac{75}{100}(15) = 11.25$

$Q_3 = P_{75}$ is located at the 11 + 1 = 12th term, Q_3 = **30**

3.7 Mean:

$$\mu = \frac{\Sigma x}{N} = \frac{413}{13} = 31.77$$

The median is located at the = (13+1)/2 = 7th term of ascending array

Median = **28**

Mode: There are three 21's. The mode is **21**.

3.9 The median is located at the

$$\left(\frac{10+1}{2}\right)^{th} = 5.5^{th}\ position$$

The median = (11,500 + 15,000)/2 = **13,250**

For $Q_3 = P_{75}$: $$i = \frac{75}{100}(10) = 7.5$$

P_{75} is located at the 8^{th} term.

$Q_3 = P_{75}$ Q_3 = **300,000**

For P_{20}: $$i = \frac{20}{100}(10) = 2.0$$

P_{20} is located halfway between the 2^{nd} and the 3^{rd} term

P_{20} = (8,800 + 9,000)/2 = **8,900**

For P_{60}: $$i = \frac{60}{100}(10) = 6.0$$

P_{60} is located halfway between the 6^{th} term and the 7^{th} term.

P_{60} = (15,000 + 122,000)/2 = **68,500**

For P_{80}: $$i = \frac{80}{100}(10) = 8.0$$

P_{80} is located halfway between the 8^{th} and 9^{th} terms.

P_{80} = (300,000 + 366,000)/2 = **333,000**

For P_{93}: $$i = \frac{93}{100}(10) = 9.3$$

P_{93} is located at the 10^{th} term P_{93} = **941,064**

3.11

x		$\lvert x-\mu \rvert$		$(x-\mu)^2$
6	6-4.2857 =	1.7143		2.9388
2		2.2857		5.2244
4		0.2857		0.0816
9		4.7143		22.2246
1		3.2857		10.7958
3		1.2857		1.6530
5		0.7143		0.5102
Σx = 30	$\Sigma \lvert x-\mu \rvert$ =	14.2857	$\Sigma(x-\mu)^2$ =	43.4284

$$\mu = \frac{\Sigma x}{N} = \frac{30}{7} = 4.2857$$

a.) Range = 9 - 1 = **8**

b.) M.A.D. = $\dfrac{\Sigma\lvert x-\mu\rvert}{N} = \dfrac{14.2857}{7}$ = **2.0408**

c.) σ^2 = $\dfrac{\Sigma(x-\mu)^2}{N} = \dfrac{43.4284}{7}$ = **6.2041**

d.) σ = $\sqrt{\dfrac{\Sigma(x-\mu)^2}{N}} = \sqrt{6.2041}$ = **2.4908**

e.) Arranging the data in order: 1, 2, 3, 4, 5, 6, 9

$Q_1 = P_{25}$ $\quad i = \dfrac{25}{100}(7) = 1.75$

Q_1 is located at the 2nd term, $Q_1 = 2$

$Q_3 = P_{75}$: $\quad i = \dfrac{75}{100}(7) = 5.25$

Q_3 is located at the 6th term, $Q_3 = 6$

IQR = $Q_3 - Q_1$ = 6 - 2 = **4**

f.) $z = \dfrac{6-4.2857}{2.4908} = \mathbf{0.69}$

$z = \dfrac{2-4.2857}{2.4908} = \mathbf{-0.92}$

$z = \dfrac{4-4.2857}{2.4908} = \mathbf{-0.11}$

$z = \dfrac{9-4.2857}{2.4908} = \mathbf{1.89}$

$z = \dfrac{1-4.2857}{2.4908} = \mathbf{-1.32}$

$z = \dfrac{3-4.2857}{2.4908} = \mathbf{-0.52}$

$z = \dfrac{5-4.2857}{2.4908} = \mathbf{0.29}$

3.13 a.)

x	$(x-\mu)$	$(x-\mu)^2$
12	12-21.167= -9.167	84.034
23	1.833	3.360
19	-2.167	4.696
26	4.833	23.358
24	2.833	8.026
23	1.833	3.360
$\Sigma x = 127$	$\Sigma(x-\mu) = -0.002$	$\Sigma(x-\mu)^2 = 126.834$

$$\mu = \frac{\Sigma x}{N} = \frac{127}{6} = 21.167$$

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} = \sqrt{\frac{126.834}{6}} = \sqrt{21.139} = \mathbf{4.598}$$ ORIGINAL FORMULA

b.)

x	x^2
12	144
23	529
19	361
26	676
24	576
23	529
$\Sigma x = 127$	$\Sigma x^2 = 2815$

$\sigma =$

$$\sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{2815 - \frac{(127)^2}{6}}{6}} = \sqrt{\frac{2815 - 2688.17}{6}} = \sqrt{\frac{126.83}{6}} = \sqrt{21.138}$$

= **4.598** SHORT-CUT FORMULA

The short-cut formula is faster, but the original formula gives insight into the meaning of a standard deviation.

3.15 $\boldsymbol{\sigma^2}$ **= 58,631.295**

$\boldsymbol{\sigma}$ **= 242.139**

$\Sigma x = 6886$ $\Sigma x^2 = 3{,}901{,}664$ $n = 16$

$\mu = 430.375$

3.17 a) $1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = .75$ **.75**

b) $1 - \frac{1}{2.5^2} = 1 - \frac{1}{6.25} = .84$ **.84**

c) $1 - \frac{1}{1.6^2} = 1 - \frac{1}{2.56} = .609$ **.609**

d) $1 - \frac{1}{3.2^2} = 1 - \frac{1}{10.24} = .902$ **.902**

3.19

x	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$
7	1.833	3.361
5	3.833	14.694
10	1.167	1.361
12	3.167	10.028
9	0.167	0.028
8	0.833	0.694
14	5.167	26.694
3	5.833	34.028
11	2.167	4.694
13	4.167	17.361
8	0.833	0.694
6	2.833	8.028
106	32.000	121.665

$$\bar{x} = \frac{\Sigma x}{n} = \frac{106}{12} = 8.833$$

a) MAD = $\dfrac{\Sigma\lvert x-\bar{x}\rvert}{n} = \dfrac{32}{12}$ = **2.667**

b) s^2 = $\dfrac{\Sigma(x-\bar{x})^2}{n-1} = \dfrac{121.665}{11}$ = **11.06**

c) s = $\sqrt{s^2} = \sqrt{11.06}$ = **3.326**

d) Rearranging terms in order: 3 5 6 7 8 8 9 10 11 12 13 14

$Q_1 = P_{25}$: i = (.25)(12) = 3

Q_1 = the average of the 3rd and 4th terms: Q_1 = (6 + 7)/2 = 6.5

$Q_3 = P_{75}$: i = (.75)(12) = 9

Q_3 = the average of the 9th and 10th terms: Q_3 = (11 + 12)/2 = 11.5

IQR = Q_3 - Q_1 = 11.5 – 6.5 = **5**

e.) z = $\dfrac{6-8.833}{3.326}$ = **- 0.85**

f.) CV = $\dfrac{(3.326)(100)}{8.833}$ = **37.65%**

3.21 $\mu = 125$ $\sigma = 12$

68% of the values fall within:

$\mu \pm 1\sigma = 125 \pm 1(12) = 125 \pm 12$

between 113 and 137

95% of the values fall within:

$\mu \pm 2\sigma = 125 \pm 2(12) = 125 \pm 24$

between 101 and 149

99.7% of the values fall within:

$\mu \pm 3\sigma = 125 \pm 3(12) = 125 \pm 36$

between 89 and 161

3.23 $1 - \dfrac{1}{k^2} = .80$

$1 - .80 = \dfrac{1}{k^2}$

$.20 = \dfrac{1}{k^2}$ and $.20k^2 = 1$

$k^2 = 5$ and $k = 2.236$

2.236 standard deviations

3.25 $\mu = 29$ $\sigma = 4$

Between 21 and 37 days:

$$\frac{x_1 - \mu}{\sigma} = \frac{21 - 29}{4} = \frac{-8}{4} = \text{-2 Standard Deviations}$$

$$\frac{x_2 - \mu}{\sigma} = \frac{37 - 29}{8} = \frac{8}{4} = \text{2 Standard Deviations}$$

Since the distribution is normal, the empirical rule states that 95% of the values fall within $\mu \pm 2\sigma$.

Exceed 37 days:

Since 95% fall between 21 and 37 days, 5% fall outside this range. Since the normal distribution is symmetrical, 2½% fall below 21 and above 37.

Thus, 2½% lie above the value of 37.

Exceed 41 days:

$$\frac{x - \mu}{\sigma} = \frac{41 - 29}{4} = \frac{12}{4} = \text{3 Standard deviations}$$

The empirical rule states that 99.7% of the values fall within $\mu \pm 3\sigma = 29 \pm 3(4)$

$= 29 \pm 12$. That is, 99.7% of the values will fall between 17 and 41 days.

0.3% will fall outside this range and **half of this or .15% will lie above 41.**

Less than 25: $\mu = 29$ $\sigma = 4$

$$\frac{x - \mu}{\sigma} = \frac{25 - 29}{4} = \frac{-4}{4} = \text{-1 Standard Deviation}$$

According to the empirical rule, $\mu \pm 1\sigma$ contains 68% of the values.

$29 \pm 1(4) = 29 \pm 4$

Therefore, between 25 and 33 days, 68% of the values lie and 32% lie outside this range with ½(32%) = **16% less than 25**.

3.27

Class	f	M	fM
0 - 2	39	1	39
2 - 4	27	3	81
4 - 6	16	5	80
6 - 8	15	7	105
8 - 10	10	9	90
10 - 12	8	11	88
12 - 14	6	13	78
	Σf=121		ΣfM=561

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{561}{121} = \mathbf{4.64}$$

$$Median = L + \frac{\frac{N}{2} - cf_p}{f_{med}}(W) =$$

$$2 + \frac{60.5 - 39}{27}(2) = 2 + \frac{21.5}{27}(2) = 2 + 1.59 = 3.59$$

Mode: The modal class is 0 – 2.
The midpoint of the modal class = the mode = **1**

3.29

Class	f	M	fM
20-30	7	25	175
30-40	11	35	385
40-50	18	45	810
50-60	13	55	715
60-70	6	65	390
70-80	4	75	300
Total	59		2775

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{2775}{59} = 47.034$$

$M - \mu$	$(M - \mu)^2$	$f(M - \mu)^2$
-22.0339	485.4927	3398.449
-12.0339	144.8147	1592.962
- 2.0339	4.1367	74.462
7.9661	63.4588	824.964
17.9661	322.7808	1936.685
27.9661	782.1028	3128.411
	Total	10,955.933

$$\sigma^2 = \frac{\Sigma f(M - \mu)^2}{\Sigma f} = \frac{10{,}955.93}{59} = \mathbf{185.694}$$

$$\sigma = \sqrt{185.694} = \mathbf{13.627}$$

3.31

Class	f	M	fM	fM^2
18 - 24	17	21	357	7,497
24 - 30	22	27	594	16,038
30 - 36	26	33	858	28,314
36 - 42	35	39	1,365	53,235
42 - 48	33	45	1,485	66,825
48 - 54	30	51	1,530	78,030
54 - 60	32	57	1,824	103,968
60 - 66	21	63	1,323	83,349
66 - 72	15	69	1,035	71,415
	Σf= 231		ΣfM= 10,371	ΣfM^2= 508,671

a.) Mean: $\bar{x} = \dfrac{\Sigma fM}{n} = \dfrac{\Sigma fM}{\Sigma f} = \dfrac{10{,}371}{231}$ = **44.896**

b.) Mode. The Modal Class = 36-42. The mode is the class midpoint = **39**

c.) Median:

$$Median = L + \frac{\frac{N}{2} - cf_p}{f_{med}}(W) =$$

$$42 + \frac{115.5-100}{33}(6) = 42 + \frac{15.5}{33}(6) = 42 + 2.818 = 44.818$$

d.) $s^2 = \dfrac{\Sigma fM^2 - \dfrac{(\Sigma fM)^2}{n}}{n-1} = \dfrac{508{,}671 - \dfrac{(10{,}371)^2}{231}}{230} = \dfrac{43{,}053.5065}{230}$ = **187.189**

e.) $s = \sqrt{187.2}$ = **13.682**

3.33

	f	M	fM	fM^2
20-30	8	25	200	5000
30-40	7	35	245	8575
40-50	1	45	45	2025
50-60	0	55	0	0
60-70	3	65	195	12675
70-80	1	75	75	5625
	$\Sigma f = 20$		$\Sigma fM = 760$	$\Sigma fM^2 = 33900$

a.) Mean:

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{760}{20} = \mathbf{38}$$

b.) Mode. The Modal Class = 20-30.
The mode is the midpoint of this class = **25**.

c.) Median:

$$Median = L + \frac{\frac{N}{2} - cf_p}{f_{med}}(W) =$$

$$30 + \frac{10-8}{7}(10) = 30 + \frac{2}{7}(10) = 30 + 2.857 = 32.857$$

d.) Variance:

$$\sigma^2 = \frac{\Sigma fM^2 - \frac{(\Sigma fM)^2}{N}}{N} = \frac{33{,}900 - \frac{(760)^2}{20}}{20} = \mathbf{251}$$

e.) Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{251} = \mathbf{15.843}$$

3.35 mean = \$35
median = \$33
mode = \$21

The stock prices are skewed to the right. While many of the stock prices are at the cheaper end, a few extreme prices at the higher end pull the mean.

3.37 Q_1 = 500. Median = 558.5. Q_3 = 589.

IQR = 589 - 500 = 89

Inner Fences: Q_1 - 1.5 IQR = 500 - 1.5 (89) = 366.5

and Q_3 + 1.5 IQR = 589 + 1.5 (89) = 722.5

Outer Fences: Q_1 - 3.0 IQR = 500 - 3 (89) = 233

and Q_3 + 3.0 IQR = 589 + 3 (89) = 856

The distribution is negatively skewed. There are no mild or extreme outliers.

3.39 Arranging the values in an ordered array:

1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
3, 3, 3, 3, 3, 3, 4, 4, 5, 6, 8

Mean: $\bar{x} = \frac{\Sigma x}{n} = \frac{75}{30}$ = **2.5**

Mode = **2** (There are eleven 2's)

Median: There are n = 30 terms.

The median is located at $\frac{n+1}{2}^{th} = \frac{30+1}{2} = \frac{31}{2}$ = 15.5th position.

Median is the average of the 15th and 16th value.

However, since these are both 2, the median is **2**.

Range = 8 - 1 = **7**

$Q_1 = P_{25}$: i = $\frac{25}{100}(30)$ = 7.5

Q_1 is the 8th term = **1**

$Q_3 = P_{75}$: i = $\frac{75}{100}(30)$ = 22.5

Q_3 is the 23rd term = **3**

IQR = Q_3 - Q_1 = 3 - 1 = **2**

3.41

$\mu = \frac{\sum x}{N} = \frac{7199.5}{15} =$ **479.97**

The median is located at the $\frac{n+1}{2}$th value = 16/2 = 8th value of an ascending array. The median is **290.9**

P_{30}: i = (.30)(15) = 4.5

P_{30} is located at the 5th term and equals **153.2**

P_{60}: i = (.60)(15) = 9

P_{60} is located at the average of the 9th and 10th terms

P_{60} = (379.6 + 390.4)/2 = **385.0**

P_{90}: i = (.90)(15) = 13.5

P_{90} is located at the 14th term

P_{90} = **914.7**

$Q_1 = P_{25}$: i = (.25)(15) = 3.75

Q_1 is located at the 4th term.

Q_1 = **147.3**

$Q_3 = P_{75}$: i = (.75)(15) = 11.25

Q_3 is located at the 12th term.

Q_3 = **457.0**

Range = 2,407 – 110.7 = **2,296.3**

IQR = Q_3 - Q_1 = 457.0 – 147.3 = **309.7**

3.43 a.) $\mu = \frac{\sum x}{N} = \frac{62.2}{12} =$ **5.183**

The median is located at the $\frac{n+1}{2}$th value = 13/2 = 6.5th value of an ascending array. It is the average of the 6th term and the 7th term:

Median = $\frac{3.9+4.1}{2} = \mathbf{4.0}$

b.) Range = 12.5 – 2.7 = **9.8**

Q_3 = 5.85 Q_1 = 3.35 IQR = Q_3 - Q_1 = **2.5**

c.) Variance:

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} = \frac{421.92 - \frac{62.2^2}{12}}{12} = \mathbf{8.293}$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{(8.293)} = \mathbf{2.880}$$

d.) ADNOC:

$$z = \frac{x - \mu}{\sigma} = \frac{2.9 - 5.183}{2.880} = -0.79$$

ExxonMobil:

$$z = \frac{x - \mu}{\sigma} = \frac{5.3 - 5.183}{2.880} = 0.04$$

3.45

	f	M	fM	fM^2
15-20	9	17.5	157.5	2756.25
20-25	16	22.5	360.0	8100.00
25-30	27	27.5	742.5	20418.75
30-35	44	32.5	1430.0	46475.00
35-40	42	37.5	1575.0	59062.50
40-45	23	42.5	977.5	41543.75
45-50	7	47.5	332.5	15793.75
50-55	2	52.5	105.0	5512.50
	$\Sigma f = 170$		$\Sigma fM = 5680.0$	$\Sigma fM^2 = 199662.50$

a.) Mean:

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{5680}{170} = \mathbf{33.412}$$

Mode: The Modal Class is 30-35. The class midpoint is the mode = **32.5**.

b.) Variance:

$$s^2 = \frac{\Sigma fM^2 - \frac{(\Sigma fM)^2}{n}}{n-1} = \frac{199{,}662.5 - \frac{(5680)^2}{170}}{169} = \mathbf{58.483}$$

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{58.483} = \mathbf{7.647}$$

3.47 $CV_x = \frac{\sigma_x}{\mu_x}(100\%) = \frac{3.45}{32}(100\%) = \mathbf{10.78\%}$

$CV_Y = \frac{\sigma_y}{\mu_y}(100\%) = \frac{5.40}{84}(100\%) = \mathbf{6.43\%}$

Stock X has a greater relative variability.

3.49 $\mu = 419$, $\sigma = 27$

a.) 68%: $\mu \pm 1\sigma$ 419 ± 27 **392 to 446**

95%: $\mu \pm 2\sigma$ $419 \pm 2(27)$ **365 to 473**

99.7%: $\mu \pm 3\sigma$ $419 \pm 3(27)$ **338 to 500**

b.) Use Chebyshev's:

Each of the points, 359 and 479 is a distance of 60 from the mean, $\mu = 419$.

k = (distance from the mean)/σ = 60/27 = 2.22

Proportion = $1 - 1/k^2 = 1 - 1/(2.22)^2$ = .797 = **79.7%**

c.) Since $x = 400$, $z = \dfrac{400-419}{27}$ = **-0.704**. This worker is in the lower half of workers but within one standard deviation of the mean.

3.51 Mean \$35,748 Median \$31,369 Mode \$29,500

Since these three measures are not equal, the distribution is skewed. The distribution is skewed to the right because the mean is greater than the median. Often, the median is preferred in reporting income data because it yields information about the middle of the data while ignoring extremes.

3.53 Paris: Since $1 - 1/k^2 = .53$, solving for k: $k = 1.459$

The distance from $\mu = 349$ to $x = 381$ is 32

$1.459\sigma = 32$ $\sigma =$ **21.93**

Moscow: Since $1 - 1/k^2 = .83$, solving for k: $k = 2.425$

The distance from $\mu = 415$ to $x = 459$ is 44

$2.425\sigma = 44$ $\sigma =$ **18.14**

3.55 Looking at the graph associated with Minitab's Graphical Summary, it is obvious that the distribution is skewed to the right. This is further underscored by the fact that the measure of skewness is 2.34450, a rather large positive measure. The skewness indicates that the top 50 advertisers in the Hispanic market are probably dominated by a few very large advertisers. The mean advertising figure is $ 41.853 million. The standard deviation is $ 31.415 million showing that there is considerable variability in the advertising spending of these top 50 companies. The minimum figure for these companies is $ 18.079 million, the first quartile is $ 21.617 million, the median is $ 31.167, the third quartile is $ 51.154 million, and the largest advertiser is $ 162,695 million.

3.57 The N indicates that there are 25 companies in this database. The mean advertising expenditure is $ 27.24 million while the median is $ 24.00 million. The gap between the mean and the median indicates that there is probably some positive skewness. While no skewness figure is given in the MINITAB output, the boxplot shows extreme outliers on the right or positive side of the data supporting the conclusion that there is positive skewness. The standard deviation is $ 14.19 million. Using this figure, it is possible to see that there is less than one standard deviation between the mean and the minimum ($ 16.20 million). If we apply the empirical rule to these data using the mean and the standard deviation, we can see that the data are not normally distributed (because there should be at least three standard deviations of data on either side of the mean if the data are approximately normally distributed) and are likely to be positively skewed. The values of the first quartile ($ 21.25 million) and the third quartile ($ 27.50 million) are also given and can be used to compute the interquartile deviation (middle 50% of the data), construct the boxplot, and indicate advertising expenditures for both the bottom and top 25 percent of companies.