Chapter 3 Numerical Descriptive Measures

Solutions

1. Sample mean: 

Median: The median is at the 3rd position in the arranged data; Median=10.

Mode: The mode is 12 since it appears twice in the dataset.

1. Sample mean: .

Median: The median is the average of the values at the 3rd and 4th positions; .



as it has the greatest frequency.



1. Population mean: .

Median: The median is located in the 5th position;



The distribution is bimodal: 55 and 110.

1. Population mean: .

Median: The median is located in the 5th position; .



The distribution is bimodal: 15 and 20.

* 1. Golf ball are most expensive in Frankfort with a price of $76.23. The least expensive golf balls are in Taipei with a price of $48.37.
  2. Mean Price: 

Median: The median is the average between the values in the 7th and 8th position; .



as it appears twice, both in Rome and in Paris.



* 1. Mean compensation = (in $ millions).



* 1. Median =.

The mean and the median differ significantly. The mean does not reflect the typical compensation because it is affected by outliers.

1. 1. MeanMkt = 164.10; MedianMkt = 167.50.
   2. Mean= 40.71%; Median = 6.05%.
   3. The mean and the median of the market capitalization are close in value. Since the values are close, the distribution is nearly symmetric. The mean and the median of the total return differ significantly. In this case, the mean is affected by outliers (extreme values), so the median is a better measure.
2. The mode is the most useful measure of central location. Massachusetts hospitals need information on diseases which cause the biggest number of deaths per day.

The data from the Massachusetts Department of Health indicate that cancer, unlike other diseases, causes the most daily deaths with 36.

1. 1. Mean= 4.70; Median= 4.68.
   2. Mean= 6.05; Median= 5.23.
   3. Based on the ERA, the Yankees are likely to have a better winning record because their mean and median ERA is lower.
2. 1. ; ; .



* 1. The mean and the median are close in value. However, the median is a better central location measure because it is not as sensitive to outliers (very low or very high selling prices) as the mean.

1. Thus the 20th percentile is located 60% of the distance between the first observation and the second observation or 0



 Thus the 50th percentile (also the median) is located at 5th position or 215.

Thus the 80th percentile is located 40% of the distance between the 6th observation and the 7th observation or



1.  The 20th percentile is: ‒362 + 0.6[(‒325) – (‒362)] = ‒339.8.

 The 40th percentile is: ‒300 + 0.2[(‒297)‒(‒300)] = ‒299.4.

. The 70th percentile is: ‒257 + 0.6[(‒255) – (‒257)] = ‒255.8.

1. 1.  The 25th percentile is 12 + 0.5(15‒12) = 13.5. Thus 25% of the observations are less than 13.5, and 75% of the observations are greater than 13.5.

 The 50th percentile is 22. Thus 50% of the observations are less than 22, and 50% are greater than 22.

 The 75th percentile is 32 + 0.5(35‒32) = 33.5. Thus 75% of the observations are less than 33.5, and 25% are greater than 33.5.



Therefore, there are no outliers.



1. 1.  The 25th percentile is ‒0.05.

Approximately 25 percent of the observations have values less than ‒0.05.

The 50th percentile is 0.04.

Approximately 50 percent of the observations have values less than 0.04.

 The 75th percentile is 0.10.

Approximately 75 percent of the observations have values less than 0.10.



Lower limit = Q1 ‒1.5×IQR=‒0.05‒0.225= ‒0.275.

Upper limit = Q3 + 1.5×IQR=0.1+ 0.225 = 0.325. Therefore, there are no outliers.

1. 1.  The 25th percentile is 65+0.25(70‒65) = 66.25.

Approximately 25 percent of the observations have values less than 66.25.

 The 50th percentile is 75.

Approximately 50 percent of the observations have values less than 75.

 The 75th percentile is 85 + 0.75(86‒85) = 85.75.

Approximately 75 percent of the observations have values less than 85.75.

* 1. IQR = 85.75‒66.25=19.50; 1.5× IQR = 29.25.

Lower limit = Q1‒1.5×IQR = 66.25‒29.25 = 37.

Upper limit = Q3+1.5×IQR = 85.75+29.25 = 115.

The smallest value of 25 is an outlier because it is less than the lower limit of 37. The median of 75 is almost in the middle of 62.25 and 85.75. Minitab is used to plot the boxplot – it shows the distribution vertically instead of horizontally as in text. The tail of the distribution runs off to the left (bottom part of Minitab’s boxplot), thus the distribution is negatively skewed.



* 1. The 25t.h percentile is 3.12+0.75(4.07‒3.12) =3.83

Approximately 25 percent of the observations have values less than 3.83.

The 50th percentile is = 6.52+0.5(8.15‒6.52) = 7.34. Approximately 50 percent of the observations have values less than 7.34.

The 75th percentile is 9.44+0.25(18.62‒9.44)=11.74

Approximately 75 percent of the observations have values less than 11.74.

* 1. IQR = 11.74‒3.83 = 7.91; 1.5 × IQR = 11.87.

Lower limit = Q1‒1.5×IQR=3.83‒11.87 = ‒8.04.

Upper limit = Q3+1.5×IQR = 11.74+11.87 = 23.61. The outliers are ‒11.37 and 31.77.

* 1. The median of 7.34 is slightly closer to Q3 than to Q1. Further, Minitab is used to plot the boxplot (it shows the distribution vertically instead of horizontally as in text.) The tail of the distribution runs off slightly to the right (top part of Minitab’s boxplot), thus the distribution is positively skewed.



1. 1.  The 25th percentile is 12.

 The 50th percentile is 14.

 The 75th percentile is 17.25.

* 1. IQR= 17.25‒12=5.25; 1.5× IQR=7.875.

Lower limit = Q1‒1.5×IQR=12‒7.875 = 4.125.

Upper limit = Q3 + 1.5×IQR = 17.25+7.875=25.125.

The outliers are 33, 35 and 40. The existence of these three outliers in the right tail of the distribution indicates skewness to the right. Minitab is used

to plot the boxplot (it shows the distribution vertically instead of horizontally as in text). Note that the tail of the distribution runs off to the right (top part of Minitab’s boxplot).



1. ==0.0313, or 3.13%.
2. = ‒0.001, that is, ‒0.1%.
3. = ‒0.006, that is ‒0.6%.

1. , that is, 3.52%.

, that is, 6.47%.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Year 1 -Year 2 | Year 2-Year 3 | Year 3- Year 4 |
| Growth Rates | 0.2222 | 0.3636 | 0.0667 |

* 1. , that is, 20.91%.



|  |  |  |  |
| --- | --- | --- | --- |
|  | Year 1 -Year 2 | Year 2-Year 3 | Year 3- Year 4 |
| Growth Rates | 0.0667 | 0.0781 | 0.1014 |

* 1. , that is, 8.2%.

1. = 0.0305, that is, 3.05%.
2. 1. %. The average annual return between 2006 and 2009 is 8.61%.



* 1. , that is, 2.58%. Compared to the arithmetic mean, the geometric mean is less at 2.58%.
  2. By the end of 2009, the total accumulation will be:







The average annual return between 2005 and 2009 is



b. that is, 13.76%. Compared to the arithmetic mean, the annual geometric mean is less at 13.76%.

c. By the end of 2009, the total accumulation will be:

38,104.64







|  |  |  |
| --- | --- | --- |
|  | Home Depot | Lowe's |
| 2008-2009 | ‒0.0784 | ‒0.001 |
| 2009-2010 | ‒0.0717 | ‒0.0209 |

* 1.  = ‒0.075; = ‒0.011.

1. 1. =0.0367, or 3.67%; =0.0149, or 1.49%.
   2. Wal-Mart has a higher growth rate for the period 2008-2010.
2. 1. 0.0869, or 8.7%.
   2. =‒1 = 0.0869, or 8.7%. As expected, both ways yield the same result.
3. 1. Range =Max Value ‒ Min Value= 42‒10 = 32
   2. 
   3.  = 
   4. =12.74
4. 1. Range = 10‒ (‒8) = 18
   2. 
   3.  = 
   4. =6.07



* 1. Range = 52‒32=20
  2. 
  3. = 
  4. = 6.86

1. 1. Range = 12‒(‒10) = 22
   2. 
   3. = , = 8.82
   4. == 13.16
   5. Southwest Airlines received the least amount of complaints per million passengers at 1.82, while United Airlines fielded the most with 13.60. Thus, the range is 11.78 (13.60‒1.82).
   6.  = 8.34; Median = 8.84
   7. =; = 3.97
2. 1. For Starbucks,; = 1.6.

For Panera Bread Co, ; = 3.56

* 1. Relative to Panera Bread’s stock price, Starbucks’ stock price has a lower variability as indicated by a lower standard deviation.
  2. For Starbucks, 

For Panera, 

Starbucks has a greater relative dispersion since its sample coefficient of variation is higher.

1. 1. Monthly rent: 
   2. Square footage: 
   3. Monthly rent: 

Square Footage: 

Therefore, there is greater relative dispersion in square footage than in monthly rent.

1. 1. ; =; .
   2. ; ; .
   3. Total return has a higher coefficient of variation which translates in a greater relative dispersion.
2. 1. Investment B has a higher mean return compared to investment A. The higher standard deviation indicates that investment B has a higher risk as well. Investment A provides the least risk.
   2. ; . Investment A has a slightly higher Sharpe Ratio. Hence it provides a higher reward per unit of risk.
3. 1. Investment B provides a higher return. Investment A provides the least risk since it has a smaller standard deviation.
   2. ; . The Sharpe Ratio is higher for investment A; hence it provides a higher reward per unit of risk.
4. 1. , . The 2nd Investment provides a higher return since it has a higher mean.
   2. , . The 1st investment provides the least risk because it has a lower standard deviation.
   3. ; . The 2nd investment performs better because it offers more reward per risk.
5. 1. The Vanguard Energy Fund has a higher average return.
   2. The Vanguard Energy Fund has a higher standard deviation, hence it was riskier.
   3. , .

The Sharpe Ratio identifies the extra reward per unit of risk. The Vanguard Energy Fund has a higher Sharpe Ratio; hence it has a higher reward per unit of risk compared to the Vanguard Health Care Fund.

1. 1. , . The Fidelity Latin America Fund has a higher average return.
   2. , . The Fidelity Latin America Fund was riskier since it had a higher standard deviation compared to the Fidelity Canada Fund.
   3. , . The Fidelity Latin America Fund has a higher Sharpe Ratio; hence it has a higher reward per unit of risk.
3. The values 70 and 90 are two standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 75% of the scores fall within 70 and 90.
4. The values 65 and 95 are three standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 89% of the scores fall within 65 and 95.
6. The values 1300 and 1700 are two standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 75% of the scores fall within 1300 and 1700.
7. The values 1100 and 1900 are four standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 94% of the scores fall within 1100 and 1900.
8. We know that at least 75% of the observations fall within two standard deviations of the mean. We are given the mean and standard deviation of 500 and 25, respectively. Therefore, at least 75% of the observations fall within 450 and 550.
9. We know that at least 89% of the observations fall within three standard deviations of the mean. We are given the mean and standard deviation of 500 and 25, respectively. Therefore, at least 89% of the observations fall within 425 and 575.
   1. The interval [18, 22] is the interval  According to the empirical rule 68% of the observations fall within this interval.
   2. The interval [16, 24] is the interval  95% of the observations fall in this interval.
   3. 16 is two standard deviations below the mean. Since 95% of the observations fall within 2 standard deviations of the mean, we conclude that 5% fall outside this interval. Half of these observations fall below 16, hence 2.5% of the observations fall below 16.
   4. According to the empirical rule 68% of the observations are between 700 and 800. Hence, half of the remaining 32%, or 16%, of the observations are less than 700.
   5. 16% of 500 is 80 observations.
   6. According to the empirical rule 95% of the observations are between 17 and 33. Hence, 95% + 2.5% (half of 5%) = 97.5% of the observations are less than 33.
   7. 97.5% of 1000 is 975 observations.
10. 1. .
    2.  Given that 95% of the data falls within two standard deviations of the mean, 95%+2.5% = 97.5% of the observations are positive.
    3. of the observations are not positive.



1. 1. 
   2. 74 is 2 standard deviations above the mean. By the empirical rule, 95% of the observations fall in the interval  Therefore, 2.5% are more than 74; 2.5% of 250 is 6.25 observations or roughly 6 observations are greater than 74.
2. The salaries $66,000 and $78,000 are two standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 75% of the faculty earns at least $66,000 but no more than $78,000.
3. The salaries $63,000 and $81,000 are three standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . In other words, Chebyshev’s Theorem asserts that at least 89% of the faculty earns at least $63,000 but no more than $81,000.
   1. 68% of the observations fall in the interval [‒4, 20]. Thus, 16% (32%/2) of the observations are greater than 20 percent.
   2. 95% of the observations fall in the interval [‒16, 32]. Thus 2.5% (5%/2) of the observations are less than ‒16%.
4. 1. 68% of the scores are in the interval [84, 116].
   2. 95% of the scores are in the interval [68, 132]. 2.5% of the scores are less than 68.
   3. Using a., 16% of the scores are more than 116.
   4. 68% of returns are expected to be in the interval [2, 14]. The probability is 0.68.
   5. 32% of future returns fall out of the [2, 14] interval. 16% (half of 32%) of returns are greater than 14 percent. The probability is 0.16.
   6. 95% of future returns will fall in the interval [‒4, 20]. 5% of them will fall out of this interval. Hence 2.5% of returns will fall below ‒4%.
5. 1. Talk times 2.4 and 5.6 hours are two standard deviations below the mean and above the mean, respectively. Using Chebyshev’s Theorem and , we have . Therefore, at least 75% of cell phones will have talk time between 2.4 and 5.6 hours.
   2. Using the empirical rule we know that approximately 95% of cell phones will have talk time between 2.4 and 5.6 hours.
6. 1. 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class |  |  |  |  |
| 2 up to 4 | 20 | 3 | 60 | 193.44 |
| 4 up to 6 | 60 | 5 | 300 | 73.93 |
| 6 up to 8 | 80 | 7 | 560 | 63.37 |
| 8 up to 10 | 20 | 9 | 180 | 167.04 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Total | 180 |  | 1100 | 497.78 |

* 1. ; =1.67.

1. 1. 
   2. ; 
2. 1. 

, 

1. 1. 
   2. ; 
2. 1. 
   2. ; 
3. 1. 
   2. ; 
4. 
5. 1. 
   2. ; 
6. 1. 
   2. ; 
7. 1. 
   2. ; 



* 1. 
  2. 

1. 1. .
   2. . There is a strong positive relationship between *x* and *y*.
   3. .
   4. . There is a strong negative relationship between *x* and *y*.
2. 1. .
   2. . The correlation coefficient is close to 1. There is a strong positive relationship between the funds’ annual returns.
3. 1. . The covariance suggests that there is a positive linear relationship between the two variables.
   2. . The correlation coefficient indicates that the relationship is positive but moderately strong.
   3. . There is a positive relationship between GRE and subsequent GPA in graduate school. A higher GRE score tends to translate into a higher GPA as well.
   4. . The positive relationship between GRE and GPA is rather strong. GRE scores seem to be a good indicator for later performance in graduate school.
   5. . The relationship between education and salary is positive.
   6. . The correlation between the two variables is very strong. Education seems to be a good indicator of salary.
4. 1. 





* 1. Hasbro has the highest average growth rate at 6.3%, but also the higher variability at 8.61% against 6.56% for Mattel.

1. . The median best reflects the typical sales as the value 3,300 is clearly an outlier that pulls the mean up.







|  |  |  |
| --- | --- | --- |
|  | Electronic | Utilities |
| Mean | 9.92 | 8.44 |
| Variance | 2215.22 | 668.68 |
| Standard Deviation | 47.07 | 25.86 |
| Coefficient of Variation | 4.74 | 3.06 |

* 1. The Electronic fund has a higher average return, compared to the Utilities fund (9.92 against 8.44)
  2. A higher standard deviation is an indication of a more risky asset: the Electronic fund is riskier compared to the Utilities fund. Furthermore, the coefficient of variation of Electronic is 4.74 compared to 3.06 for Utilities.
  3. ; . The Utilities fund has a higher Sharpe Ratio. Hence it provides a higher reward per unit of risk relative to the Electronic fund.



* 1. American Eagle growth is less negative as compared to The Gap over this period.



|  |  |  |
| --- | --- | --- |
|  | Firm A | Firm B |
| Mean | 31.33 | 23.50 |
| Variance | 8.67 | 6.70 |
| Standard Deviation | 2.94 | 2.59 |
| Coefficient of Variation | 0.09 | 0.11 |

* 1. Firm A has the higher average stock price over this period (31.33 versus 23.50 for Firm B).
  2. The stock prices of Firm A showed a greater standard deviation (2.94 versus 2.59 for Firm B). However, Firm B displays greater relative dispersion with a coefficient of variation of 0.11 against 0.09 for Firm A.

1. 1. 
   2. , 
2. 1. 
   2. , 
3. 1. 
   2. . The correlation coefficient is close to 1. There is a strong positive relationship between the funds’ annual returns.







* 1. . There is a positive correlation Advertising and Sales.
  2. . The correlation coefficient is close to 1. There is a strong positive relationship between Advertising and Sales.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Income | Unemployment | Debt |
| Mean | 74.05 | 9.765385 | 983.4615 |
| Median | 71.5 | 9.3 | 987.5 |
| Mode | 78.3 | 9.3 | 1133 |
| Standard Deviation | 10.35325 | 2.225838 | 124.6086 |
| Range | 44.1 | 9.4 | 522 |
| Minimum | 59.4 | 6.3 | 763 |
| Maximum | 103.5 | 15.7 | 1285 |

85.

|  |  |
| --- | --- |
|  | Income |
| Mean | 8.86 |
| Median | 7.96 |
| Standard Deviation | 6.71 |
| Range | 24.93 |

There is a great deal of variability in salaries. The difference between the highest and lowest paid quarterback is $24.93 million. The sample standard deviation of salaries is $6.71 million.

**Case Study 3.1**

|  |  |  |
| --- | --- | --- |
|  | American League | National League |
| Mean | 95820948 | 86016975 |
| Median | 84061250 | 84325333 |
| Standard Deviation | 43803002 | 33403170 |
| Skewness | 1.497845 | 0.43003 |
| Range | 1.55E+08 | 1.12E+08 |

|  |  |  |
| --- | --- | --- |
| Minimum | 51654900 | 34943000 |
| Maximum | 2.06E+08 | 1.47E+08 |
| Count | 14 | 16 |

* The mean opening-day salaries are higher for AL as compared to NL. The high mean for AL is because of extremely large salaries of New York Yankees and Boston Red Socks.
* The median opening-day salaries are slightly lower for AL as compared to NL.
* Both distributions are skewed to the right since the skewness coefficient is positive. Given the outliers, the median is a more appropriate measure of central location than the mean, especially for the AL.
* Both range and standard deviation suggest that there is more dispersion of salaries in the AL than the NL.

**Case Study 3.2**

|  |  |  |
| --- | --- | --- |
|  | Janus Balanced Fund | Janus Overseas Fund |
| Mean (Arith.) | 4.62 | 12.25 |
| Median | 8.23 | 23.17 |
| Sample Variance | 135.55 | 1620.09 |
| Standard Deviation | 11.64 | 40.25 |
| Coef. Of Variation | 2.52 | 3.29 |
| Range | 39.5 | 130.87 |
| Minimum | ‒15.22 | ‒52.75 |
| Maximum | 24.28 | 78.12 |

* Compared to Balanced Fund, the Overseas Fund has a higher average return. However the Overseas Fund has a higher standard deviation and coefficient of variation, indicating higher risk and dispersion relatively to Balanced Fund.
* The coefficient of correlation between the two funds is positive and very strong at 0.98. This indicates that the two funds are linearly related.

**Case Study 3.3**

|  |  |  |
| --- | --- | --- |
|  | Revenues Growth Rates | |
|  | Nike | Adidas |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Asia | Latin America | Asia | Latin America |
| Mean (Arith.) | 15.22 | 16.91 | 15.13 | 34.17 |
| Mean (Geom.) | 15.04 | 16.57 | 14.46 | 33.26 |
| Median | 13.40 | 15.39 | 14.84 | 33.79 |
| Sample Variance | 57.22 | 109.90 | 204.83 | 322.54 |
| Standard Deviation | 7.56 | 10.48 | 14.31 | 17.96 |
| Coef. Of Variation | 0.50 | 0.62 | 0.95 | 0.53 |
| Range | 17.51 | 23.18 | 34.44 | 43.77 |
| Minimum | 8.28 | 6.85 | ‒1.80 | 12.65 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Maximum | 25.78 | 30.03 | 32.63 | 56.43 |

* For Nike, the average growth rate is higher in Latin America compared to Asia (16.57 against 15.04 for geometric means). However, a greater standard deviation in Latin America implies a greater variability in growth rates as compared to Asia. This is also indicated by a relatively higher range in Latin America (23.18 versus 17.51).
* For Adidas, the average growth rate is higher as well in Latin America (33.26 against 14.46 for geometric means). On the other hand, a higher standard deviation in Latin America implies more variability in growth rates in Latin America as compared to Asia.
* Although both companies are growing more in Latin America than in Asia, it is clearly the case that Adidas’ average growth rate isurpasses that of Nike in Latin

America. However, the average growth rates of Nike and Adidas are fairly the similar in Asia (15.04 for Nike versus 14.46 for Adidas).

**Case Study 3.4**

|  |  |  |
| --- | --- | --- |
|  | 2007 Home Sale Prices in Fort Myers, FL | |
|  | January | July |

|  |  |  |
| --- | --- | --- |
| Mean | 231.08 | 182.72 |
| Median | 205.00 | 180.00 |
| Mode | 220.00 | 210.00 |
| Sample Variance | 12926.08 | 4813.88 |
| Standard Deviation | 113.69 | 69.38 |
| Coef. of Variation | 0.49 | 0.38 |
| Range | 510.00 | 314.00 |
| Minimum | 100.00 | 86.00 |
| Maximum | 610.00 | 400.00 |

|  |  |  |
| --- | --- | --- |
| 1st Quartile | 158.00 | 133.00 |
| 3rd Quartile | 256.00 | 215.00 |

* The average home price in January of 2007 was $231, 080 versus $182, 720 in July of the same year. That is about a 21% drop in the average home price. Also in January, half of the homes sold for more than $205,000, versus only $180,000 in July (see the median). Since the mean is more effected by outliers (in this case, a few relatively high prices), the median is an appropriate measure of central location.
* The higher standard deviation in January also translates a higher volatility in home prices in January. This adds to a relatively dispersed home price data in January compared to July as measured by the coefficient of variation (0.49 versus 0.38).
* While 25% of the houses were sold at the price of $158, 000 or less in January, the first quartile for July is only $133, 000. The third quartile is also larger in January.
* All indicators support the fact that the crisis had a negative impact on home prices in Florida.