

Correlation	Population
Correlation coefficient (Pearson product moment correlation coefficient)	Process capability index
Covariance	Proportion
Dispersion	Range
Empirical rules	Return to risk
Interquartile range (IRQ, or midspread)	Sample
Kurtosis	Sample correlation coefficient
Median	Skewness
Midrange	Standard deviation
Mode	Standardized value (z-score)
Outlier	Statistical thinking
	Unimodal
	Variance

Problems and Exercises

- Data obtained from a county auditor in the Excel file *Home Market Value* provide information about the age, square footage, and current market value of houses along one street in a particular subdivision. Considering these data as a population of homeowners on this street, compute the mean, variance, and standard deviation for each of these variables using a spreadsheet and formulas (4.1), (4.4), and (4.6). Verify your calculations using the appropriate Excel function.
- In the Excel file *Facebook Survey*, find the average and median hours online/week and number of friends in the sample using the appropriate Excel functions. Compute the midrange and compare all measures of location.
- For the Excel file *Tablet Computer Sales*, find the average number, standard deviation, and interquartile range of units sold per week. Show that Chebyshev's theorem holds for the data and determine how accurate the empirical rules are.
- The Excel file *Atlanta Airline Data* provides arrival and taxi-in time statistics for one day at Atlanta Hartsfield International airport. Find the average and standard deviation of the difference between the scheduled and actual arrival times and the taxi-in time to the gate. Compute the z-scores for each of these variables.
- Data obtained from a county auditor in the Excel file *Home Market Value* provides information about the age, square footage, and current market value of houses along one street in a particular subdivision.
 - Considering these data as a sample of homeowners on this street, compute the mean, variance, and standard deviation for each of these variables using formulas (4.2), (4.5), and (4.7). Verify your calculations using the appropriate Excel function.
 - Compute the coefficient of variation for each variable. Which has the least and greatest relative dispersion?
- Find 30 days of stock prices for three companies in different industries. The average stock prices should have a wide range of values. Using the data, compute and interpret the coefficient of variation.
- Compute descriptive statistics for liberal arts colleges and research universities in the Excel file *Colleges and Universities*. Compare the two types of colleges. What can you conclude?
- Use the *Descriptive Statistics* tool to summarize the mean, median, variance, and standard deviation of the prices of shares in the Excel file *Coffee Shares Data*.
- The worksheet *Data* in the Excel file *Airport Service Times* lists a large sample of the times in seconds to process customers at a ticket counter. The second worksheet shows a frequency distribution and histogram of the data.
 - Summarize the data using the *Descriptive Statistics* tool. What can you say about the shape of the distribution of times?
 - Find the 90th percentile.
 - How might the airline use these results to manage its ticketing counter operations?

10. The data in the Excel file *Church Contributions* were reported on annual giving for a church. Estimate the mean and standard deviation of the annual contributions of all parishioners by implementing formulas (4.13) and (4.15) on a spreadsheet, assuming these data represent the entire population of parishioners. Second, estimate the mean contribution of families with children in the parish school. How does this compare with all parishioners?
11. The average monthly wages and standard deviations for the two garments manufacturing factories X and Y are given below:
 - Factory X: the average monthly wage is \$4600, the standard deviation of the wage is \$500, and the number of wage-earners is 100
 - Factory Y: the average monthly wage is \$4900, standard deviation is \$400, and the number of wage-earners is 80
 - a. Which factory pays the larger amount as monthly wages?
 - b. Which factory shows greater variability in the distribution of wages?
12. Consider the Excel file *Mobiles Usage*, which shows the number of people using different kinds of mobile phones in the northern region. Find the proportion of BlackBerry and Android usage in that region.
13. In the Excel file *Bicycle Inventory*, find the proportion of bicycle models that sell for less than \$200.
14. In the *Sales Transactions* database, find the proportion of customers who used PayPal and the proportion of customers who used credit cards. Also, find the proportion that purchased a book and the proportion that purchased a DVD.
15. In the Excel file *Economic Poll*, find the proportions of each categorical variable.
16. In the Excel file *Facebook Survey*, use a PivotTable to find the average and standard deviation of hours online/week and number of friends for females and males in the sample.
17. In the Excel file *Cell Phone Survey*, use PivotTables to find the average for each of the numerical variables for different cell phone carriers and gender of respondents.
18. Using PivotTables, find the average and standard deviation of sales in the *Sales Transactions* database.

Also, find the average sales by source (Web or e-mail). Do you think this information could be useful in advertising? Explain how and why or why not.
19. For the Excel file *Travel Expenses*, use a PivotTable to find the average and standard deviation of expenses for each sales rep.
20. Using PivotTables, compute the mean and standard deviation for each metric by year in the Excel file *Freshman College Data*. Are any differences apparent from year to year?
21. The Excel file *Freshman College Data* shows data for 4 years at a large urban university. Use PivotTables to examine differences in student high school performance and first-year retention among different colleges at this university. What conclusions do you reach?
22. The Excel file *Cell Phone Survey* reports opinions of a sample of consumers regarding the signal strength, value for the dollar, and customer service for their cell phone carriers. Use PivotTables to find the following:
 - a. the average signal strength by type of carrier
 - b. average value for the dollar by type of carrier and usage level
 - c. variance of perception of customer service by carrier and gender

What conclusions might you reach from this information?
23. Call centers have high turnover rates because of the stressful environment. The national average is approximately 50%. The director of human resources for a large bank has compiled data about 70 former employees at one of the bank's call centers (see the Excel file *Call Center Data*). Use PivotTables to find these statistics:
 - a. the average length of service for males and females in the sample
 - b. the average length of service for individuals with and without a college degree
 - c. the average length of service for males and females with and without prior call center experience
24. In the Excel file *Weddings*, determine the correlation between the wedding costs and attendance.
25. For the data in the Excel file *Rin's Gym*, find the covariances and correlations among height, weight, and BMI calculation.

26. For the Excel file *Test Scores and Sales* made by nine salesmen during the past year, compute the coefficient of correlation between the test scores and sales using Excel's CORREL function.
27. The Excel file *Beverage Sales* lists a sample of weekday sales at a convenience store, along with the daily high temperature. Compute the covariance and correlation between temperature and sales.
28. For the Excel file *Credit Risk Data*, compute the correlation between age and months employed, age and combined checking and savings account balance, and the number of months as a customer and amount of money in the bank. Interpret your results.
29. In the Excel file *Call Center Data*, how strongly is length of service correlated with starting age?
30. A national homebuilder builds single-family homes and condominium-style townhouses. The Excel file *House Sales* provides information on the selling price, lot cost, type of home, and region of the country (M = Midwest, S = South) for closings during 1 month. Use PivotTables to find the average selling price and lot cost for each type of home in each region of the market. What conclusions might you reach from this information?
31. The Excel file *Auto Survey* contains a sample of data about vehicles owned, whether they were purchased new or used, and other types of data. Use the *Descriptive Statistics* tool to summarize the numerical data, find the correlations among each of the numerical variables, and construct PivotTables to find the average miles/gallon for each type of vehicle, and also the average miles/gallon and average age for each type of new and used vehicle. Summarize the observations that you can make from these results.
32. Compute the z -scores for the data in the Excel file *Airport Service Times*. How many observations fall farther than three standard deviations from the mean? Would you consider these as outliers? Why or why not?
33. Use the *Manufacturing Measurements* data to compute sample averages, assuming that each row in the data file represents a sample from the manufacturing process. Plot the sample averages on a line chart, add the control limits, and interpret your results.
34. Find the mean and variance of a deck of 52 cards, where an ace is counted as 11 and a picture card as 10. Construct a frequency distribution and histogram of the card values. Shuffle the deck and deal two samples of 20 cards (starting with a full deck each time); compute the mean and variance and construct a histogram. How does the sample data differ from the population data? Repeat this experiment for samples of 5 cards and summarize your conclusions.
35. Examine the z -scores you computed in Problem 4 for the *Atlanta Airline Data*. Do they suggest any outliers in the data?
36. In the Excel file *Weddings*, find the averages and median wedding cost and the sample standard deviation. What would you tell a newly engaged couple about what cost to expect? Consider the effect of possible outliers in the data.
37. A producer of computer-aided design software for the aerospace industry receives numerous calls for technical support. Tracking software is used to monitor response and resolution times. In addition, the company surveys customers who request support using the following scale:
 - 0—did not exceed expectations
 - 1—marginally met expectations
 - 2—met expectations
 - 3—exceeded expectations
 - 4—greatly exceeded expectations

The questions are as follows:

 - Q1: Did the support representative explain the process for resolving your problem?
 - Q2: Did the support representative keep you informed about the status of progress in resolving your problem?
 - Q3: Was the support representative courteous and professional?
 - Q4: Was your problem resolved?
 - Q5: Was your problem resolved in an acceptable amount of time?
 - Q6: Overall, how did you find the service provided by our technical support department?

A final question asks the customer to rate the overall quality of the product using this scale:

 - 0—very poor
 - 1—poor
 - 2—good
 - 3—very good
 - 4—excellent

A sample of survey responses and associated resolution and response data are provided in the Excel