**Quantitative Methods**

**List of Exercises N. 2**

**Selected Exercises from McClave (2014) – Chapter 2**

**2.4 Numerical Measures of Variability**

1. (66, HCOUGH). ***Is honey a cough remedy?*** Refer to the Archives of pediatrics and Adolescent Medicine (Dec. 2007) study of honey as a remedy of coughing. 105 ill children in the sample were randomly divided into 3 groups: those who received a dosage of an over-the-counter cough medicine (DM), those who received a dosage of honey (H), and those who received no dosage (control group). The coughing improvement scores (as determined by the children’s parents for the patients are reproduced in the table below:



1. Find the standard deviation of the improvement scores for the honey dosage group.
2. Find the standard deviation of the improvement scores for the DM dosage group.
3. Find the standard deviation of the improvement scores for the control group.
4. Based on the results, parts a-c, which group appears to have the most variability in coughing improvement scores? The least variability?

2. (67, CORSUS). ***Corporate sustainability of CPA firms***. Refer to the Business and Society (march 2011) study on the sustainability behaviors of CPA corporations. Numerical measures of variation for level of support for the 992 senior managers are showed in the accompanying.



1. Locate the range in the printout. Comment on the accuracy of the statement: “The difference between the largest and the smallest values of the level of support for the 992 senior managers is 155 points”.
2. Locate the variance on the printout. Comment on the accuracy of the following statement: “on average, the level of support for corporate sustainability for the 992 senior managers is 722 points.”
3. Locate the standard deviation on the printout. Does the distribution of support levels for the 992 senior managers have a more or less variation than another distribution with a standard deviation of 50? Explain.
4. Which measure of variation best describes the distribution of 992 support levels? Explain.

3. (69, NUKES). ***Active Nuclear power plants***. Refer to the US Energy Information Administration’s data on the number of nuclear power plants operating in each of the 20 states.

1. Find the range, variance and standard deviation of this data set.
2. Eliminate the largest value from the data set and repeat part a. What effect does dropping this measurement have on the measures of variation found in part a?
3. Eliminate the smallest and largest value from the data set and repeat part a. What effect does dropping both of these measurements have on the measures of variation found in part a?

**2.5 Using the Mean and standard deviation to describe Data**

4. (75). Given a data set with a largest value of 760 and a smallest value of 135, what would you estimate the standard deviation to be? Explain the logic behind the procedure you used to estimate the standard deviation. Suppose the standard deviation is reported to be 25. Is this feasible? Explain.

5. (78). ***Blogs for Fortune 500 firms***. Refer to the Journal of Relationship Marketing (Vol. 7, 2008) study of the prevalence of blogs and forums at fortune 500 firms with both English and Chinese Web sites. In a sample of firms that provide blogs and forums as marketing tools, the mean number of blogs/forums per site was 4.25, with a standard deviation of 12.02.

1. Provide an interval that is likely to contain the number of blogs/forums per site for at least 75% of the Fortune 500 firms in the sample.
2. Do you expect the distribution of the number of blogs/forums to be symmetric, skewed right or skewed left? Explain.

**2.6 Numerical Measures of Relative Standing**

6. (100, CORSUS). ***Corporate sustainability of CPA firms***. Refer to the Business and Society (march 2011) study on the sustainability behaviors of CPA corporations. Numerical descriptive measures for the level of support for corporate sustainability for the 992 senior managers are repeated in the accompanying printout. One of the managers reported a support level of 155 points. Would you consider this support level to be typical of the study sample? Explain.



7. (101, SPRFND). ***Hazardous waste cleanup in Arkansas***. The Superfund Act was passed by congress to encourage state participation in the implementation of a law relating to the release and cleanup of hazardous substances. Hazardous waste sites financed by the Superfund Act are called Superfund sites. A total of 393 Superfund sites are operated by waste management companies in Arkansas (Tabor and Stanwick, Arkansas Business and Economic Review, Summer 1995). The number of these Superfund sites in each of Arkansas’s 75 counties are shown in the next table.



1. Find the 10th percentile of the data set. Interpret the result.
2. Find the 95th percentile of the data set. Interpret the result.
3. Find the mean and standard deviation of the data. Then, use these values to calculate z-score for an Arkansas county with 48 Superfund sites.
4. Based on your answer c, would you classify 48 as an extreme number of Superfund sites?

8. (103, ECOPHD). ***Ranking PhD programs in economics***. Thousands of students apply for admission to graduate schools in economics each year with the intention of obtaining a PhD. The Southern Economic Journal (Apr. 2008) published a guide to graduate study in economics by ranking the PhD programs at 129 colleges and universities. Each program was evaluated according to the number of publications published by faculty teaching in the PhD program and by the quality of the publications. Data obtained from the Social Science Citation Index (SSCI) were used to calculate an overall productivity score for each PhD program. The mean and standard deviation of these 129 productivity scores were then used to compute a z-score for each economics program. Harvard University had the highest z-score (5.08) and, hence, was the top ranked school. Howard University was ranked last because it had the lowest z-score (-0.81). The data (z-scores) for all 129 economic programs are saved in the data file.

a) Interpret the z-score for Harvard University

b) Interpret the z-score for Howard University

c) The authors of the Scandinavian Economic Journal article note that “only 44 of the 129 school have positive z-scores, indicating that the distribution of overall productivity is skewed to the right.” Do you agree? Check your answer by constructing a histogram for the z-scores in the file.

**2.7 Methods for Detecting Outliers: Box Plots and z-Scores**

9. (115, BNKRPT). ***Time in Bankruptcy***. Refer to the Financial Management (Spring 1995) study of 49 firms filing for prepackaged bankruptcies. 3 types of firms exist: (1) Those who hold no prefiling vote, (2) those who vote their preference for a joint solution and (3) Those who vote their preference for a prepack.

1. Construct a box plot for the time in bankruptcy (months) for each type of firm.
2. Find the median bankruptcy times for the 3 types of firms.
3. How do the variabilities of the bankruptcy times compare for these 3 types of firms?
4. The standard deviation of the bankruptcy times are 2.47 for “none”, 1.72 for “joint” and 0.96 for “prepack” Do the standard deviations agree with the interquartile ranges with regard to the comparison of the variabilities of the bankruptcy times?
5. Is there evidence of outliers in any of the 3 distributions?

10. (117, SANIT). ***Sanitation Inspection of Cruise Ships***. Refer to the sanitation levels of passenger cruise ships.

1. Use the box plot method to detect any outliers in the data set.
2. Use the z-score method to detect any outliers in the data set.
3. Do the 2 methods agree? Explain why.

**2.8 Graphing Bivariate Relationships**

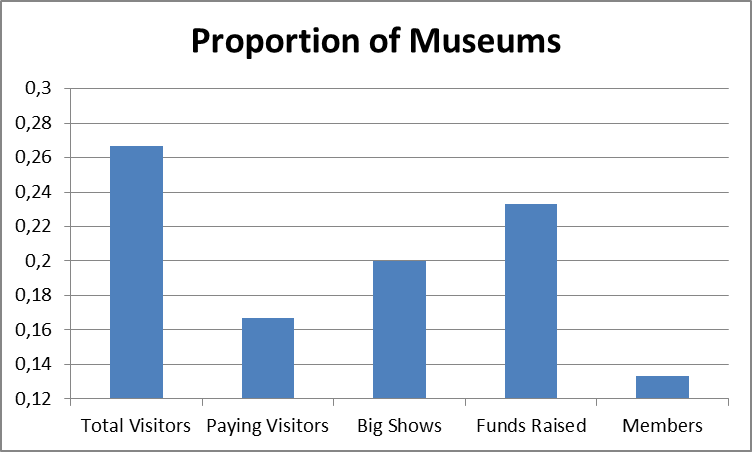
11. (129, PARS). ***Performance ratings of government agencies***. The U.S. Office of Management and Budget (OMB) requires government agencies to produce annual performance and accounting reports (PARS) each year. A research team at George Mason University evaluated the quality of the PARS for 24 government agencies (The Public Manager, Summer 2008). Evaluation scores ranged from 12 (lowest) to 60 (highest). The PARS evaluation scores for 2 consecutive years are shown in the table below:



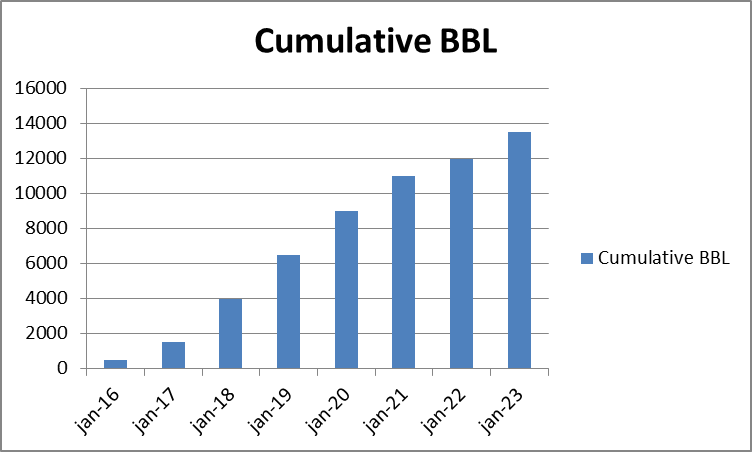
1. Construct a scatterplot for the data. DO you detect a trend in the data?
2. Based on the graph, identify one or two agencies that had greater than expected PARS evaluation scores for year 2.

**2.9 Distorting the Truth with Descriptive Techniques**

12. (133, MMC). ***Museum Management***. Refer to the Museum Management and Curatorship (June 2010) study of how museums evaluate their performance. Managers of 30 museums of contemporary art identified the performance measure used most often. A summary of the results are reproduced in the table. Consider the bar graph shown. Identify two ways in which the bar graph might mislead the viewer by overemphasizing the importance of one of the performance measures.



13. (136, BPOIL) ***BP oil leak***. In the summer of 2010, an explosion on the Deepwater Horizon oil drilling rig caused a leak in one of British Petroleum (BP) Oil Company’s wells in the Gulf of Mexico. Crude oil rushed unabated for 3 straight months into the Gulf until BP could fix the leak. During the disaster, BP used suction tubes to capture some of the gushing oil. In May of 2011, in an effort to demonstrate the daily improvement in the process, a BP representative presented a graphic on the daily number of 42-gallon barrels (bbl) of oil collected by the suctioning process. A graphic similar to the one used by BP is shown below:



1. Note that the vertical axis represents the cumulative number of barrels collected per day. This is calculated by adding the amounts of the previous days’ oil collection to the current days’ oil collection. Explain why this graph is misleading.
2. Estimates of the actual number of barrels of oil collected per day for each of the 8 days are listed in the table below. Construct a graph for this data that accurately depicts BP’s progress in its daily collection of oil. What conclusions can you draw with your graphic?



S**upplementary Exercises on Methods for Describing Sets of Data**

14. (137).Construct a relative frequency histogram for the data summarized in the following table:



15. (142). For each of the following data sets, compute the mean, the variance and the standard deviation. If appropriate, specify the units in which your answers are expressed.

1. 4, 6, 6, 5, 6, 7.
2. -1, 4, -3, 0, -3, -6.
3. 60%, 80%, 40%, 20%, 6.25%.
4. Calculate the range of each data set in parts a-c.

16. (150). ***Crash tests on new cars***. One quantitative variable recorded by the NHTSA is driver’s severity of head injury (measured on a scale from 0 to 1500). The mean and standard deviation for the 98 driver head-injury ratings are displayed in the printout below.



a) Give a practical interpretation of the mean.

b) Use the mean and the standard deviation to make a statement about where most of the head-injury ratings fall.

c) Find the z-score for a driver head-injury rating of 408. Interpret the result.

17. (159, HULL). ***Hull failures of oil tankers***. Owing to several major ocean oil spills by tank vessels, Congress passed the 1990 Oil Pollution Act, which requires all tankers to be designed with ticker hulls. Further improvements in the structural design of a tank vessel have been proposed since then, each with the objective of reducing the likelihood of oil spill and decreasing the amount of outflow in the event of a hull puncture. To aid in this development, Marine Technology (Jan. 1995) reported on the spillage amount (in thousands of metric tons) and cause of puncture for 42 major oil spills from tankers and carriers. Cause of puncture is classified as either collision (C), fire / explosion (FE), hull failure (HE) or grounding (G). The data are saved in the accompanying file.

1. Use a graphical method to describe the cause of oil spillage for the 42 tankers. Does the graph suggest that any one cause is more likely to occur than any other? How is this information of value to the design engineers?
2. Find and interpret descriptive statistics for the 42 spillage amounts. Use this information to form an interval that can be used to predict the spillage amount of the next major oil spill.

18. (162, WHEELS). ***Time to develop price quotes***. A manufacturer of industrial wheels is losing many profitable orders because of the long time it takes the firm’s marketing, engineering, and accounting departments to develop price quotes for potential customers. To remedy this problem, the firm’s management would like to set guidelines for the length of time each department should spend developing price quotes. To help develop these guidelines, 50 requests for price quotes were randomly selected from the set of price quotes made last year: the processing time (in days) was determined for each price quote for each department. Several observations are displayed in the table below. The price quotes are also classified by whether or not they were “lost” (i.e., whether or not the customer placed an order after receiving the price quote).



1. Construct a stem-and-leaf display for the total processing time for each department. Shade the leaves that correspond to “lost” orders in each of the displays, and interpret each of the displays.
2. Using your results from a, develop “maximum processing time” guidelines for each department that, if followed, will help the firm reduce the number of lost orders.
3. Generate summary statistics for the processing times. Interpret the results.
4. Calculate the z-score corresponding to the maximum processing time guideline you developed in part b for each department, and for the total processing time.
5. Calculate the maximum processing time corresponding to a z-score of 3 for each of the departments. What percentage of the orders exceed this guidelines. How does this agree with Chebyshev’s Rule and the Empirical Rule?
6. Repeat part e using a z-score of 2.
7. Compare the percentage of “lost” orders with corresponding times that exceed at least one of the guidelines in part e to the same percentage using the guidelines in part f. Which set of guidelines would you recommend to be adopted? Why?