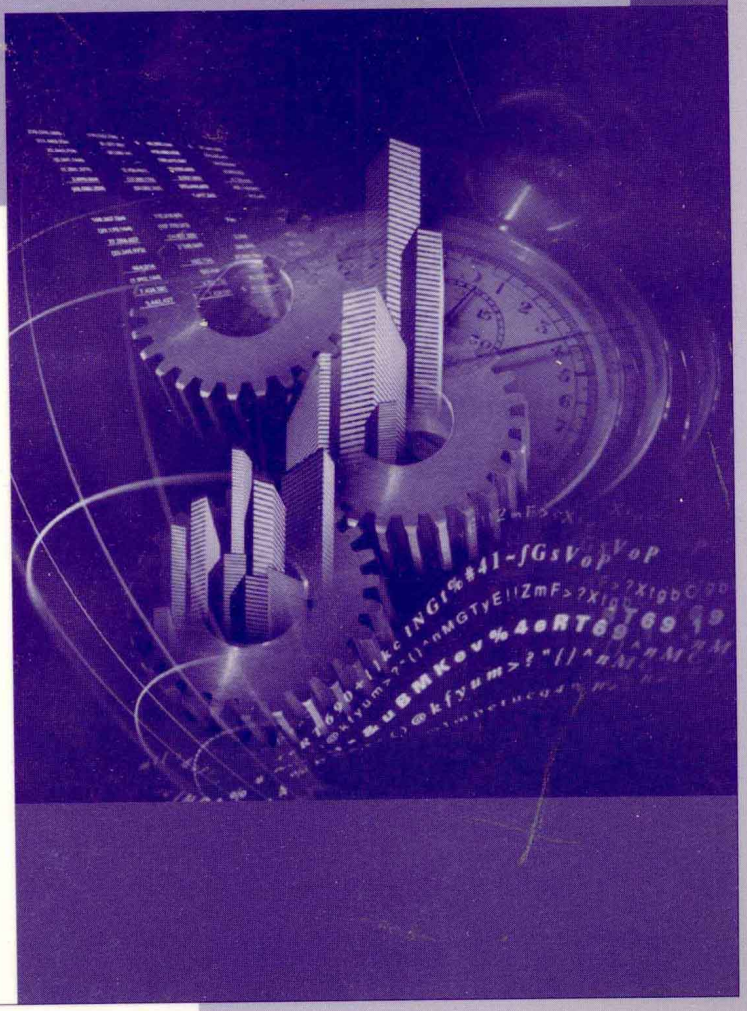
Study Guide

for use with

Second Edition

Business Statistics in Practice



BOWERMAN

O'CONNELL

HAND

Prepared by
Ron Barnes

Study Guide

for use with

Business Statistics in Practice

Second Edition

Bruce L. Bowerman
Miami University of Ohio

Richard T. O'Connell
Miami University of Ohio

Michael L. Hand
Willamette University

Prepared by
Ron Barnes
University of Houston

McGraw-Hill Irwin

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

McGraw-Hill Higher Education

A Division of The McGraw-Hill Companies

Study Guide for use with
BUSINESS STATISTICS IN PRACTICE, SECOND EDITION
Bowerman, O'Connell, and Hand

5 6 7 8 9 0 1BKM/1BKM 0 9 8 7 6 5 4 3

ISBN 0-07- 241534-7

www.mhhe.com

TO THE STUDENT

The study of statistics has been greatly enhanced by the efforts of Professors Bowerman, O'Connell, and Hand. The text makes extensive use of computer statistical software, including Minitab, Excel (and an extension of Excel-Megastat provided on the CD and included with the text at no extra cost) and SAS.

This study guide has been written to supplement the text and to help the student better understand the material presented in it. For each chapter of the text, this guide has a corresponding chapter consisting of: (1) key ideas; (2) A set of true-false questions; (3) a set of multiple-choice questions; and (4) a set of 1-7 problems illustrating the chapter material. The solutions of the true-false and multiple-choice problems are given in a format also listing text page reference and key idea for each answer. The additional large scale problems are also completely worked out, and these answers are also included at the end of each chapter.

A key feature of the study guide is that it contains over 60 Minitab, Excel, Excel Enhanced Megastat (provided free with text) and SAS Statistical Software Programs used to solve various problems including a number of exercises and case studies from the textbook. The Minitab (Release 13 for Windows) and Microsoft Excel 97 and the Excel "Megastat" Enhanced Version—all allow data entry into simple spreadsheets and the various programs and options can be executed by simple clicks on menu screens. This study guide not only includes the computer outputs (as does the text) but also the computer input steps necessary to run these programs. A few problems deal with the SAS commands needed to generate various SAS computer output given in text examples and exercises.

The basic idea of the guide is to help you, the student to briefly review the topics covered in each chapter, then check your understanding with true-false and multiple-choice questions, and finally tie everything together with a few well chosen problems. Many of these problems use computer statistics programs which are included either in the problem itself or in the problem's solution contained at the end of the chapter. Computer input and output are included for each of these problems.

It is believed that the student can best understand and master the basic contents and tasks of statistics by seeing how they apply to realistic processes that arise in the business world. The text and this study guide were carefully developed to reflect and carry out this philosophy.

ACKNOWLEDGEMENTS

Many thanks to the staff of McGraw-Hill Irwin for help with this project, especially for the advice and encouragement of Ms. Wanda Zeman.

TABLE OF CONTENTS

CHAPTER 1
AN INTRODUCTION TO BUSINESS STATISTICS

KEY IDEAS

1. Statistics involves the study of how to collect, summarize and interpret data.
2. Statistical methodology enables us to draw conclusions from data. Such conclusions are important to the decision-making process of many organization including: analysis of economic data by the government for planning and policy decisions; politicians and pollsters; analysis of financial data for investment decisions; analysis of marketing, sales and production data for business and industry decisions; and the analysis of drug and health care data for improving health care delivery systems.
3. Business and others often utilize statistical analysis of data to improve their process. Such analyses may help to (1) demonstrate the need for improvements, (2) identify ways to make improvements, (3) assess the degree to which new improvements are successful, and (4) estimate the benefits of improved strategies.
4. Descriptive statistics is the science of describing the important aspects of a set of measurements.
5. When populations are too large to examine in their entirety, we select a sample or subset of the population. If the sample accurately represents the population (i.e. is randomly selected) then we can use statistics to make inferences about the underlying population. Such investigations are called enumerative studies.
6. When making statistical inferences about populations of all values that will or could be potentially generated from using a process, we are performing an analytical study.
7. A process is a sequence of operations that take inputs (labor, materials, methods, etc...) and turn them into outputs (products, services, etc...). Processes usually operate over time.
8. In order to make statistical inferences in an analytical study, the underlying process must be in statistical control, i.e. not exhibit any unusual process variation.
9. Statistical Process Control (SPC) is a systematic method for analyzing process data (quality characteristics) in which we monitor and study the process variation. The goal is to stabilize and reduce the effect of process variation.
10. A runs plot (a graph of a sample of individual process measurements, over time) is used to help decide whether a process is in control.
11. We must be very careful in deciding whether the statistical inference from a study can be extended to related populations. In general, an analytical study, in statistical control, (and expected to remain in control) can be extended to the infinite population of all values that will or could potentially result from the process.
12. Experiments are statistical studies where we wish to estimate and compare the effects of various conditions or treatments on a variable of interest (response variable).
13. This chapter introduces five applied case studies to illustrate the statistical methods and how they can be used to analyze data and aid in process improvement activities.
14. Statistical software packages such as Excel, Minitab, and SAS are often employed in the performance of various statistical calculations and procedures.

TRUE/FALSE QUESTIONS

1. Descriptive statistics is the science of using sample data to make generalizations about the corresponding aspects of the entire population.
2. Data values are always numerical values.
3. Business and others often utilize statistical analysis of data to improve their processes. Such analyses may help (1) document the need for improvements, (2) identify ways to make improvements, (3) assess the degree to which new improvements are successful, and (4) estimate the benefits of improvement strategies.
4. The three types of statistical studies considered in this chapter are enumerative, analytical and inferential.
5. If we wish to study the ages of students in a class of fifty students, we can get a random sample of ages of students by recording the ages of the fifth, tenth, fifteenth, and all succeeding multiples of five students listed on the alphabetical roll.
6. Random number tables are often used in the selection of random samples.
7. The variable describing the brands of automobiles of students in this class is a categorical variable.
8. The goal of statistical process control is to stabilize and reduce the amount of process variation.
9. In an experiment, we attempt to estimate the effect of various treatments on many different populations.
10. Computer statistical software packages are rarely used in statistical investigations.

MULTIPLE-CHOICE QUESTIONS

1. Which of the following does not use statistics?
   a. Financial planners
   b. Manufacturers in the area of quality control
   c. Public opinion pollsters
   d. Marketing and sales researchers
   e. All of the above
   f. None of the above

2. Statistical analysis of data may help to improve a process by all of the following except
   a. Demonstrate the need for improvements
   b. Identify ways to make improvements
   c. Assess the success of improvements
   d. Increase the variation of the process
   e. Estimate the benefits of improved strategies

3. Describe which of the following variables are quantitative and which are qualitative.
   a. Brands of computers owned by students in this class
   b. Ages of student in this class
   c. Account balances in checking accounts
   d. Political affiliations of students
   e. Number of dependents declared on income tax returns

4. Match each of the following studies with the type of population it considers.

a. enumerative study	1. population of all values that would or could possibly result from using a process
b. analytical study	2. population that usually does not yet exist (uses new treatments)
c. experiment	3. population of existing units

PROBLEMS

1. For 8500 students enrolled at the college, we wish to construct a random sample of 12 students. If the students are listed in alphabetical order in the student directory, use the random number table on page 6 of your textbook and start with the entry in the 3rd column and 6th row of the table to generate a random sample of 12 students. Use the first 4 digits of each random number and assume student number 1 is the first student in the alphabetical listing and that student number 8500 is the last student in the alphabetic listing. Which students (numbers) are in your sample?

2. The weekly sales of a new fast food restaurant for its first 26 weeks of operation are given below:

Week #	Sales (in ($1000)	week #	Sales (in $1000)
1	5	14	24
2	4	15	28
3	7	16	33
4	8	17	45
5	13	18	43
6	12	19	51
7	14	20	52
8	18	21	50
9	16	22	43
10	21	23	51
11	23	24	56
12	24	25	49
13	27	26	58

   a. Construct a runs plot of this data.
   b. Do the 26 measurements appear to be in statistical control: Why or why not?

TRUE/FALSE ANSWERS

Question	Answer	Key Idea	Text Reference Page
1.	F	4	4
2.	F		3
3	T	3	3
4.	F		—
5.	F		5
6.	T		4
7.	T		3
8.	T	9	13
9.	F	12	—
10.	F	14	24

MULTIPLE-CHOICE ANSWERS

Question	Answer	Key Idea	Text Reference Page
1.	f	2	3
2.	d	9	13
3.	Quant.-b,c,e/Qual.-a,d		3
4.	matches a-3, b-1, c-2	5, 6, 12	—

SOLUTIONS TO PROBLEMS

1. Going down the column start with: 1630, 1988, 0414 = 414, 1451, 0669 = 669, 3016, 2530, 3800, 0025 = 25, (omit 9242 it is greater than 8500), 8265, 2084, 4002.

2. a. Minitab data input

	C1 week	C2 sales
1	1	5
2	2	4
3	3	7
4	4	8
5	5	13
6	6	12
7	7	14
8	8	18
9	9	16
10	10	21
11	11	23
12	12	24
13	13	27
14	14	24
15	15	28
16	16	33
17	17	45
18	18	43
19	19	51
20	20	52
21	21	50
22	22	43
23	23	51
24	24	56
25	25	49
26	26	58

To implement this runs program, first enter the data (as in the figure to the left) into the Minitab Spreadsheet (Grid).

Next choose

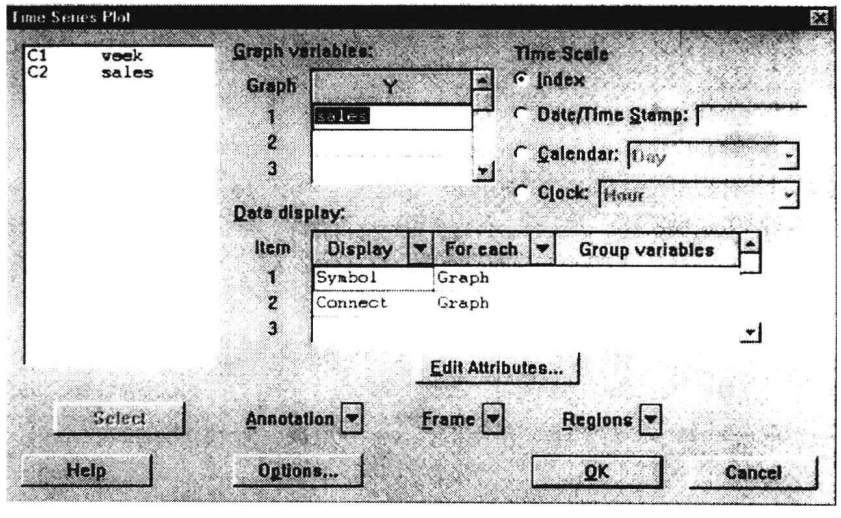
stat > time series > times series plot

After this, the following template will appear like in the figure below.

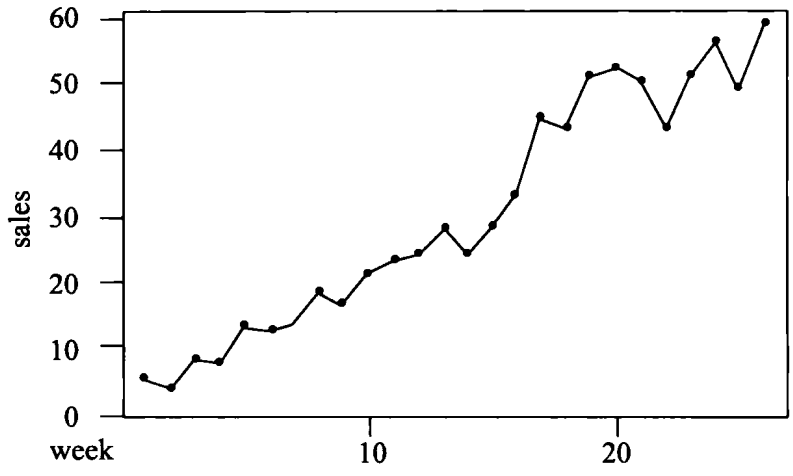
Type in

Sales under the Graph Y column and then press the OK bar.

To get your runs plot, see next page for plot.



Minitab Runs Plot



b. Does not appear to be in statistical control since its level is increasing.

CHAPTER 2
DESCRIPTIVE STATISTICS

KEY IDEAS

1. Descriptive statistics deals with the use of concepts, tools and methodology, including graphical displays, to describe and illustrate important aspects of a sample or population.
2. Dot plots, stem-and-leaf diagrams, histograms and bar charts are often used to portray data and illustrate its central tendency and variability. Dot plots (for small data sets) are the simplest to construct, just a number line with data points above numbers that corresponds to data values; followed by stem and-leaf and bar charts. Histograms require the construction of class intervals and tallying frequency or relative frequency of data occurrences within their classes. Computer software packages SAS, Excel, and Minitab all can be used to construct these descriptive statistics graphs.
3. Stem - leaf or histograms, often exhibit some well defined shapes including the symmetrical bell shape or normal curve which consists of most data centered near the average or middle of the data values with fewer data values on each end as you get further from the middle.
4. If the histogram curve is not symmetric the data may be skewed with a long tail to the right (skew to right) or a left tail (skewed to left). The height of a histogram bar (or the length of a stem and leaf line) represents the number of frequency data measurements in that interval (or with a relative frequency histogram the proportion of data measurements in that interval).
5. Other shapes of frequency curves can occur including ones with two or more maximum peaks.
6. If we have sampled a process, the process must be in statistical control in order to reliably use a stem-and-leaf display or a histogram to make inferences about the overall shape of the population. Similarly, if we have sampled a population and wish to use graphical representations of the sample measurements to make inferences about the related population, the sample data must be representative of the related population. If the process is not in statistical control, incorrect inferences may be drawn from sampled data as our textbook illustrates.
7. In measuring the central tendency of a data set, we employ three ideas - the mean, median, and mode. In general we prefer to use the mean (the usual average of the data values) unless the data exhibits some extremes (a few extremely large and/or extremely small data values), in which case we use the median (a middle value). In cases where the median also is unsatisfactory we resort to the mode or most frequently occurring measurement value. Categorical data such as the type of car owned by a random sample of students requires the mode to describe the brand of car owned by the typical or average student in the sample.
8. Population parameters, numbers or terms describing some aspects of a population are often estimated by their sample counterparts - sample statistics which are the corresponding numbers or terms calculated by using the sample measurements. Examples include sample means, sample variances and sample standard deviations.
9. When populations are symmetrical or not highly skewed, the population mean and population median are either equal or roughly equal and both are good measures of central tendency.
10. To estimate the variability in a population or sample data set we use the range (which measures the spread between the largest and smallest measurement values) or the variance and standard deviation. In estimating the population variances we use the sample variances which divide by n

- 1 instead of n to help correct for the tendency to underestimate the population variance when using only a sample of data values.

11. If the underlying population is reasonably normal then the empirical rule allows us to estimate tolerance interval formulas that contain approximately 68, 95, and 99.7% of the measurement values of a population.

12. Getting a process into statistical control does not guarantee that the process is capable of meeting product specifications. Control implies that the process level or mean is remaining relatively constant and the amount of variation about the mean is relatively constant. Control does not say anything about how much variation there is. If there is too much variation the product may be outside the specification lines and therefore the process is not capable of meeting product requirements or specifications. For example if a process fills 12oz. cola cans (which can actually hold up to 14 ounces) with a average of 11.9 ounces per fill but the standard deviation is 8 ounces then most cans will be filled to overflowing wasting cola or grossly under filled causing customer complaints. Tolerance intervals are often used in quality and process improvement projects that evaluate the capability of meeting product specification requirements.

13. If the population is not normally distributed, then the empirical rule is not always accurate in estimating tolerance intervals. Chebyshev's theorem can be used to estimate tolerance intervals for any population distribution, however it is only of practical use for non normal populations that are not extremely skewed [some radical skewing will grossly inflate the estimate for the standard deviation].

14. When a data set is highly skewed it is best to use percentiles and quantiles to measure variation. A box and whiskers plot (or box plot) is a graphical display of data that employs quantiles calculations to describe the data.

15. Bar charts and pie charts are graphical methods that can be used to display quantitative or categorical data. Sample proportions calculated from sample data sets are often used to estimate the proportion of participants that fall into a given category.

16. Pareto charts are bar charts that summarize data concerning the occurrence of various types of defects (non conforming) possessed by a product produced by a process. They are very helpful in pinpointing areas to be addressed that will most effectively bring about quality improvement.

17. A simple way to explore the relationship between a variable of interest (dependent variable) and another variable x (independent variable) which we suspect affects y, is to make a scatter diagram and scatter plot. Such plots often suggest linear or quadratic relationships that may occur between the variables.

18. Another graphical way to study relationships between variables is to construct runs plots, which can be studied to see if the variables tend to move together over time.

19. Statistical graphs and charts can be manipulated in order to distort the truth. Also statistical assumptions and inferences drawn from data should be subject to scrutiny.

20. All the descriptive statistics and techniques in this chapter are accessible in the Minitab, Excel and SAS statistical software packages.

TRUE/FALSE QUESTIONS

1. Descriptive statistics deals with the use of sample measurements to make generalizations about important aspects of a population, while inferential statistics deals with describing the important characteristics of a population or sample using various concepts, tools and graphical displays.

2. The concept of central tendency deals with the average or middle of a collection of measurements, while variation deals with how the measurement values tend to spread out.
3. Categorical measurement values usually require the use of the median as the appropriate measure of central tendency.
4. If the variance in salaries at a company is zero then all employees make the same amount of money.
5. The standard deviation is always smaller than the variance.
6. The heights of American citizens exhibit a normal or bell shaped symmetric frequency distribution curve.
7. An interval that contains a specified percentage of the individual measurements is a population is called a tolerance interval.
8. The empirical rule enables us to estimate tolerance intervals even when the underlying population is highly skewed.
9. Stem-and-leaf and histograms often exhibit similarity to some well defined shapes such as the normal curve or skewed distribution curves.
10. When a data set is highly skewed it is best to use percentiles and quantiles and tools like box-and-whiskers plots to measure variation.
11. If we have sampled a process, the process must be in statistical control in order to reliably use a stem-and-leaf or histogram to make inferences about the overall shape of the population distribution.
12. Scatter plots may often suggest linear or quadratic relationships that occur between variables.

MULTIPLE-CHOICE QUESTIONS

1. Which of the following is not a descriptive statistical tool?
   a. Bar Chart
   b. Histogram
   c. Mean
   d. Quality Control Chart
   e. Box Plot
   f. Scatter Plot

2. Which would be the more appropriate descriptive tool for each of the following – Histogram, Bar Chart?
   a. The distribution of different majors in this class of 20 students
   b. The weeking sales data for 50 Mickey D's restaurants in the metropolitan area
   c. The distribution of reasons why customers changed their long distance telephone company

3. Describe which is the appropriate measure of central tendency – (i.e. mean, median or mode for each of the following):
   a. Ages of students in the class which range from 18 to 35
   b. Salaries of students in this class of 20 student if 19 are unemployed, i.e. make $0 salary and the other made $20 million last year
   c. The car driven by a typical student in this class if 8 drive Hondas, 4 drive Fords, 6 drive VW's and the remaining 2 take the bus

4. Which of each pair of descriptive tools is preferred if the underlying population is to be normal and symmetric?
   a. Chebyshev's Theorem or Empirical Rule
   b. Variance and standard deviation or percentiles and box plots

5. Which of the following populations would you expect to be (1) Normal symmetric (2) skewed to right or (3) skewed to left?
   a. Weight of all U.S. citizens (including all infants and children)
   b. Weight of all U.S. adults over 21 years of age
   c. The final grade distribution of A's, B's, C's, D's, F's, I's, and W's by all students enrolled in freshman English
   d. The grade distribution of all students enrolled in the Senior Seminar course for Accounting Honors majors

6. For each of the populations in the previous example (#5) determine which descriptive tools (1) variance, standard deviation and tolerance interval estimates or (2) percentiles, quantiles and box plots would be more appropriate to describe the variability of the population.

PROBLEM

1. The Fresh Detergent Case Predicting Demand

   Enterprise Industries produces and markets Fresh, a brand of liquid detergent. The company wishes to study how price and advertising expenditure affect the demand for Fresh. To this end, the company has gathered data concerning demand for Fresh over the last 30 sales periods (each sales period is defined to be a four-week period). The demand data is presented for each sales period.

   y = the demand for the large-size bottle of Fresh (in hundreds of thousands of bottles) in the sales period

   x_1 = the price (in dollars) of Fresh as offered by Enterprise Industries in the sales period

   x_2 = the average industry price (in dollars) of competitors' similar detergents in the sales period

   x_3 = Enterprise Industries' advertising expenditure (in hundreds of thousands of dollars) to promote Fresh in the sales period

   $x_4 = x_2 - x_1$ = the *price difference* in the sales period.

Data Concerning Demand for Fresh Detergent

Sales Period	Price for Fresh, x_1 (dollars)	Average Industry Price, x_2 (dollars)	Price Difference, $x_4 = x_2 - x_1$ (dollars)	Advertising Expenditure for Fresh, x_3 (hundreds of thousands of dollars)	Demand for Fresh, y (hundreds of thousands of dollars)
1	3.85	3.80	-.05	5.50	7.38
2	3.75	4.00	.25	6.75	8.51
3	3.70	4.30	.60	7.25	9.52
4	3.70	3.70	0	5.50	7.50
5	3.60	3.85	.25	7.00	9.33
6	3.60	3.80	.20	6.50	8.28
7	3.60	3.75	.15	6.75	8.75
8	3.80	3.85	.05	5.25	7.87
9	3.80	3.65	-.15	5.25	7.10
10	3.85	4.00	.15	6.00	8.00
11	3.90	4.10	.20	6.50	7.89
12	3.90	4.00	.10	6.25	8.15
13	3.70	4.10	.40	7.00	9.10
14	3.75	4.20	.45	6.90	8.86
15	3.75	4.10	.35	6.80	8.90
16	3.80	4.10	.30	6.80	8.87
17	3.70	4.20	.50	7.10	9.26
18	3.80	4.30	.50	7.00	9.00
19	3.70	4.10	.40	6.80	8.75
20	3.80	3.75	-.05	6.50	7.95
21	3.80	3.75	-.05	6.25	7.65
22	3.75	3.65	-.10	6.00	7.27
23	3.70	3.90	.20	6.50	8.00
24	3.55	3.65	.10	7.00	8.50
25	3.60	4.10	.50	6.80	8.75
26	3.65	4.25	.60	6.80	9.21
27	3.70	3.65	-.05	6.50	8.27
28	3.75	3.75	0	5.75	7.67
29	3.80	3.85	.05	5.80	7.93
30	3.70	4.25	.55	6.80	9.26

Use this data and a Minitab program to solve each of the following:

a. Construct a run plot of demand (sales) versus sales period (time). Does the graph suggest that the trend of sales is in general increasing, decreasing, remaining the same over time, or none of these.

b. Construct a histogram of the demand (sales) data.

c. Do the tallest frequency boxes of your histogram appear to reflect any particular values of x_1, x_2, x_3, or x_4?

d. Calculate the mean, range, variance, and standard deviation of this demand (sales) data.