Chapter 1

Introduction to Statistics

LEARNING OBJECTIVES

The primary objective of chapter 1 is to introduce you to the world of statistics, enabling you to:

1. Define statistics.
2. Be aware of a wide range of applications of statistics in business.
3. Differentiate between descriptive and inferential statistics.
4. Classify numbers by level of data and understand why doing so is important.

CHAPTER TEACHING STRATEGY

In chapter 1 it is very important to motivate business students to study statistics by presenting them with many applications of statistics in business. The definition of statistics as a science dealing with the collection, analysis, interpretation, and presentation of numerical data is a very good place to start. Statistics is about dealing with data. Data are found in all areas of business. This is a time to have the students brainstorm on the wide variety of places in business where data are measured and gathered. It is important to define statistics for students because they bring so many preconceptions of the meaning of the term. For this reason, several perceptions of the word statistics are given in the chapter.

Chapter 1 sets up the paradigm of inferential statistics. The student will understand that while there are many useful applications of descriptive statistics in business, the strength of the application of statistics in the field of business is through inferential statistics. From this notion, we will later introduce probability, sampling, confidence intervals, and hypothesis testing. The process involves taking a sample from the population, computing a statistic on the sample data, and making an inference (decision or conclusion) back to the population from which the sample has been drawn.

In chapter 1, levels of data measurement are emphasized. Too many texts present data to the students with no comment or discussion of how the data were gathered or the level of data measurement. In chapter 7, there is a discussion of sampling techniques. However, in this chapter, four levels of data are discussed. It is important for students to understand that the statistician is often given data to analyze without input as to how it was gathered or the type of measurement. It is incumbent upon statisticians and researchers to ascertain the level of measurement that the data represent so that appropriate techniques can be used in analysis. All techniques presented in this text cannot be appropriately used to analyze all data.

CHAPTER OUTLINE

1.1 Statistics in Business

- Marketing
- Management
- Finance
- Economics
- Management Information Systems

1.2 Basic Statistical Concepts

1.3 Data Measurement

- Nominal Level

Ordinal Level

Interval Level

Ratio Level

Comparison of the Four Levels of Data

Statistical Analysis Using the Computer: Excel and MINITAB

KEY TERMS

Census	Ordinal Level Data
Descriptive Statistics	Parameter
Inferential Statistics	Parametric Statistics
Interval Level Data	Population
Metric Data	Ratio Level Data
Nominal Level Data	Sample
Non-metric Data	Statistic
Nonparametric Statistics	Statistics

SOLUTIONS TO PROBLEMS IN CHAPTER 1

1.1 Examples of data in functional areas:

accounting - cost of goods, salary expense, depreciation, utility costs, taxes, equipment inventory, etc.

finance - World bank bond rates, number of failed savings and loans, measured risk of common stocks, stock dividends, foreign exchange rate, liquidity rates for a single-family, etc.

human resources - salaries, size of engineering staff, years experience, age of employees, years of education, etc.

marketing - number of units sold, dollar sales volume, forecast sales, size of sales force, market share, measurement of consumer motivation, measurement of consumer frustration, measurement of brand preference, attitude measurement, measurement of consumer risk, etc.

information systems - CPU time, size of memory, number of work stations, storage capacity, percent of professionals who are connected to a computer network, dollar assets of company computing, number of “hits” on the Internet, time spent on the Internet per day, percentage of people who use the Internet, retail dollars spent in e-commerce, etc.

production - number of production runs per day, weight of a product; assembly time, number of defects per run, temperature in the plant, amount of inventory, turnaround time, etc.

management - measurement of union participation, measurement of employer support, measurement of tendency to control, number of subordinates reporting to a manager, measurement of leadership style, etc.

1.2 Examples of data in business industries:

manufacturing - size of punched hole, number of rejects, amount of inventory, amount of production, number of production workers, etc.

insurance - number of claims per month, average amount of life insurance per family head, life expectancy, cost of repairs for major auto collision, average medical costs incurred for a single female over 45 years of age, etc.

travel - cost of airfare, number of miles traveled for ground transported vacations, number of nights away from home, size of traveling party, amount spent per day on besides lodging, etc.

retailing - inventory turnover ratio, sales volume, size of sales force, number of competitors within 2 miles of retail outlet, area of store, number of sales people, etc.

communications - cost per minute, number of phones per office, miles of cable per customer headquarters, minutes per day of long distance usage, number of operators, time between calls, etc.

computing - age of company hardware, cost of software, number of CAD/CAM stations, age of computer operators, measure to evaluate competing software packages, size of data base, etc.

agriculture - number of farms per county, farm income, number of acres of corn per farm, wholesale price of a gallon of milk, number of livestock, grain storage capacity, etc.

banking - size of deposit, number of failed banks, amount loaned to foreign banks, number of tellers per drive-in facility, average amount of withdrawal from automatic teller machine, federal reserve discount rate, etc.

healthcare - number of patients per physician per day, average cost of hospital stay, average daily census of hospital, time spent waiting to see a physician, patient satisfaction, number of blood tests done per week.

1.3 Descriptive statistics in recorded music industry -

1) RCA total sales of compact discs this week, number of artists under contract to a company at a given time.

2) Total dollars spent on advertising last month to promote an album.

3) Number of units produced in a day.

4) Number of retail outlets selling the company's products.

Inferential statistics in recorded music industry -

1) Measure the amount spent per month on recorded music for a few consumers then use that figure to infer the amount for the population.

2) Determination of market share for rap music by randomly selecting a sample of 500 purchasers of recorded music.

3) Determination of top ten single records by sampling the number of requests at a few radio stations.

4) Estimation of the average length of a single recording by taking a sample of records and measuring them.

The difference between descriptive and inferential statistics lies mainly in the usage of the data. These descriptive examples all gather data from every item in the population about which the description is being made. For example, RCA measures the sales on all its compact discs for a week and reports the total.

In each of the inferential statistics examples, a sample of the population is taken and the population value is estimated or inferred from the sample. For example, it may be practically impossible to determine the proportion of buyers who prefer rap music. However, a random sample of buyers can be contacted and interviewed for music preference. The results can be inferred to population market share.

1.4 Descriptive statistics in manufacturing batteries to make better decisions -

1) Total number of worker hours per plant per week - help management understand labor costs, work allocation, productivity, etc.

2) Company sales volume of batteries in a year - help management decide if the product is profitable, how much to advertise in the coming year, compare to costs to determine profitability.

3) Total amount of sulfuric acid purchased per month for use in battery production. - can be used by management to study wasted inventory, scrap, etc.

Inferential Statistics in manufacturing batteries to make decisions -

1) Take a sample of batteries and test them to determine the average shelf life - use the sample average to reach conclusions about all batteries of this type. Management can then make labeling and advertising claims. They can compare these figures to the shelf-life of competing batteries.

2) Take a sample of battery consumers and determine how many batteries they purchase per year. Infer to the entire population - management can use this information to estimate market potential and penetration.

3) Interview a random sample of production workers to determine attitude towards company management - management can use this survey result to ascertain employee morale and to direct efforts towards creating a more positive working environment which, hopefully, results in greater productivity.

1.5 a) ratio

b) ratio

c) ordinal

d) nominal

e) ratio

f) ratio

g) nominal

h) ratio

1.6 a) ordinal

b) ratio

c) nominal

d) ratio

e) interval

f) interval

g) nominal
h) ordinal

1.7 a) The population for this study is the 900 electric contractors who purchased Rathburn wire.

b) The sample is the randomly chosen group of thirty-five contractors.

c) The statistic is the average satisfaction score for the sample of thirty-five contractors.

d) The parameter is the average satisfaction score for all 900 electric contractors in the population.

Chapter 2

Charts and Graphs

LEARNING OBJECTIVES

The overall objective of chapter 2 is for you to master several techniques for summarizing and depicting data, thereby enabling you to:

1. Recognize the difference between grouped and ungrouped data.
2. Construct a frequency distribution.
3. Construct a histogram, a frequency polygon, an ogive, a pie chart, a stem and leaf plot, a Pareto chart, and a scatter plot.

CHAPTER TEACHING STRATEGY

Chapter 1 brought to the attention of students the wide variety and amount of data available in the world of business. In chapter 2, we confront the problem of trying to begin to summarize and present the data in a meaningful manner. One mechanism for data summarization is the frequency distribution, which is essentially a way of organizing ungrouped or raw data into grouped data. It is important to realize that there is considerable art involved in constructing a frequency distribution. There are nearly as many possible frequency distributions for a problem as there are students in a class. Students should begin to think about the receiver or user of their statistical product. For example, what class widths and class endpoints would be most familiar and meaningful to the end user of the distribution? How can the data best be communicated and summarized using the frequency distribution?

The second part of chapter 2 presents various ways to depict data using graphs. The student should view these graphical techniques as tools for use in communicating characteristics of the data in an effective manner. Most business students will have some type of management opportunity in their field before their career ends. The ability to make effective presentations and communicate their ideas in succinct, clear ways is an asset. Through the use of graphics packages and such techniques as frequency polygons, ogives, histograms, and pie charts, the manager can enhance his/her personal image as a communicator and decision-maker. In addition, the manager can emphasize that the final product (the frequency polygon, etc.) is just the beginning. Students should be encouraged to study the graphical output to recognize business trends, highs, lows, etc. and realize that the ultimate goal for these tools is their usage in decision making.

CHAPTER OUTLINE

2.1 Frequency Distributions

- Class Midpoint
- Relative Frequency
- Cumulative Frequency

2.2 Graphic Depiction of Data

Histograms

Frequency Polygons

Ogives

Pie Charts

Stem and Leaf Plots

Pareto Charts

2.3 Graphical Depiction of Two-Variable Numerical Data: Scatter Plots

KEY TERMS

Class Mark

Class Midpoint

Cumulative Frequency

Frequency Distribution

Frequency Polygon

Grouped Data

Histogram

Ogive

Pareto Chart

Pie Chart

Range

Relative Frequency

Scatter Plot

Stem and Leaf Plot

Ungrouped Data

SOLUTIONS TO PROBLEMS IN CHAPTER 2

2.1

a) One possible 5 class frequency distribution:

Class Interval	Frequency
10 - under 25	9
25 - under 40	13
40 - under 55	11
55 - under 70	9
70 - under 85	8
	50

b) One possible 10 class frequency distribution:

Class Interval	Frequency
10 - under 18	7
18 - under 26	3
26 - under 34	5
34 - under 42	9
42 - under 50	7
50 - under 58	3
58 - under 66	6
66 - under 74	4
74 - under 82	4

82 - under 90	2

c) The ten class frequency distribution gives a more detailed breakdown of temperatures, pointing out the smaller frequencies for the higher temperature intervals. The five class distribution collapses the intervals into broader classes making it appear that there are nearly equal frequencies in each class.

2.2 One possible frequency distribution is the one below with 12 classes and class intervals of 2.

Class Interval	Frequency
39 - under 41	2
41 - under 43	1
43 - under 45	5
45 - under 47	10
47 - under 49	18
49 - under 51	13
51 - under 53	15
53 - under 55	15
55 - under 57	7
57 - under 59	9
59 - under 61	4
61 – under 63	1

The distribution reveals that only 13 of the 100 boxes of raisins contain 50 ± 1 raisin (49 -under 51). However, 71 of the 100 boxes of raisins contain between 45 and 55 raisins. It shows that there are five boxes that have 9 or more extra raisins (59-61 and 61-63) and two boxes that have 9-11 less raisins (39-41) than the boxes are supposed to contain.

2.3

Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency

0 - 5	6	2.5	6/86 = .0698	6
5 - 10	8	7.5	.0930	14
10 - 15	17	12.5	.1977	31
15 - 20	23	17.5	.2674	54
20 - 25	18	22.5	.2093	72
25 - 30	10	27.5	.1163	82
30 - 35	4	32.5	.0465	86
TOTAL	86		1.0000	

The relative frequency tells us that it is most probable that a customer is in the

15 - 20 category (.2674). Over two thirds (.6744) of the customers are between 10

and 25 years of age.

2.4

Class Interval	Frequency	Class Frequency	Relative Midpoint	Cumulative Frequency
0-2	218	1	.436	218
2-4	207	3	.414	425
4-6	56	5	.112	481
6-8	11	7	.022	492
8-10	8	9	.016	500
TOTAL	500		1.000	

2.5 Some examples of cumulative frequencies in business:

sales for the fiscal year,

costs for the fiscal year,

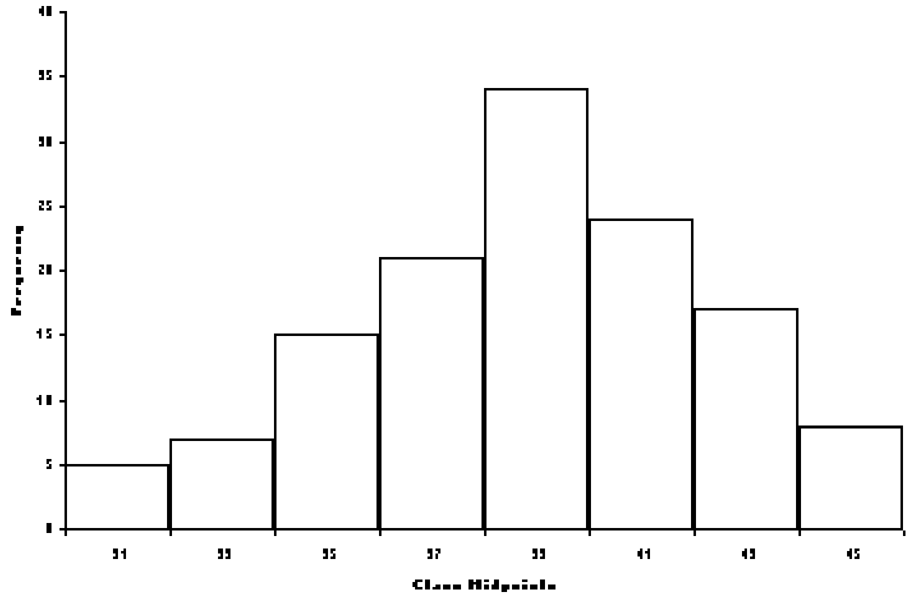
spending for the fiscal year,

inventory build-up,

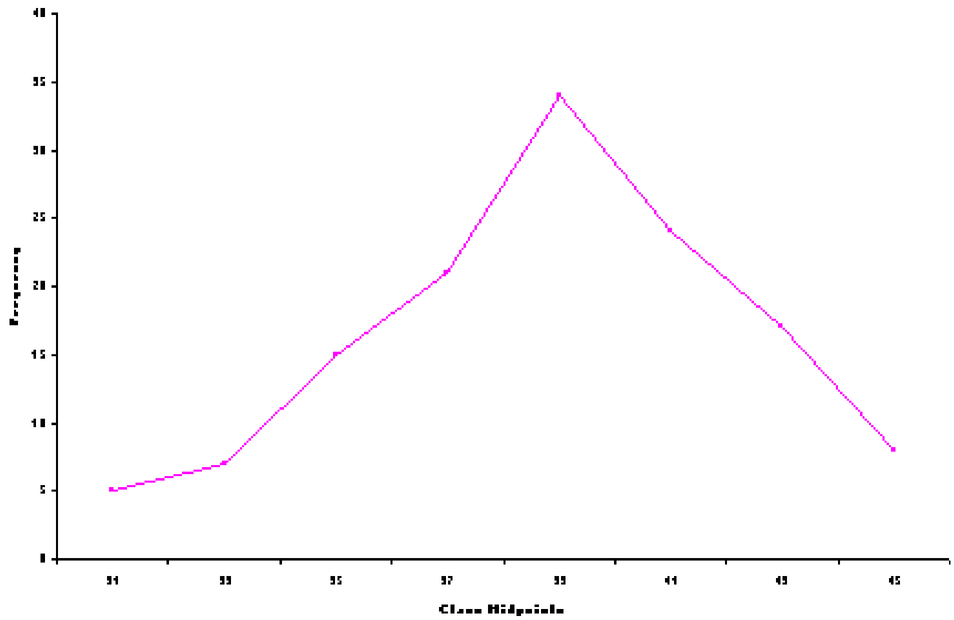
accumulation of workers during a hiring buildup,

production output over a time period.

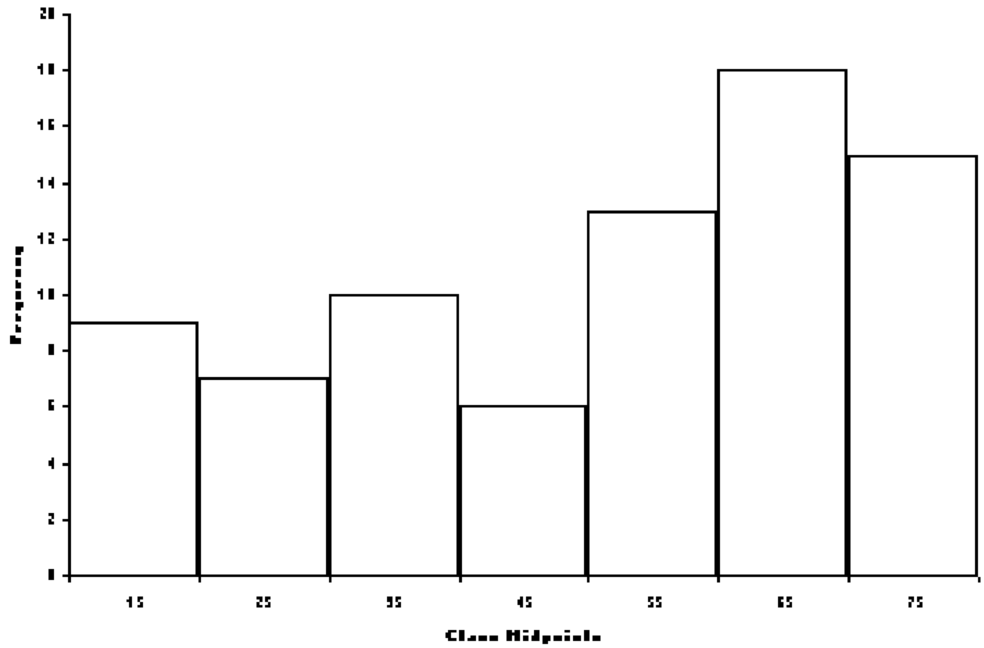
2.6 Histogram:



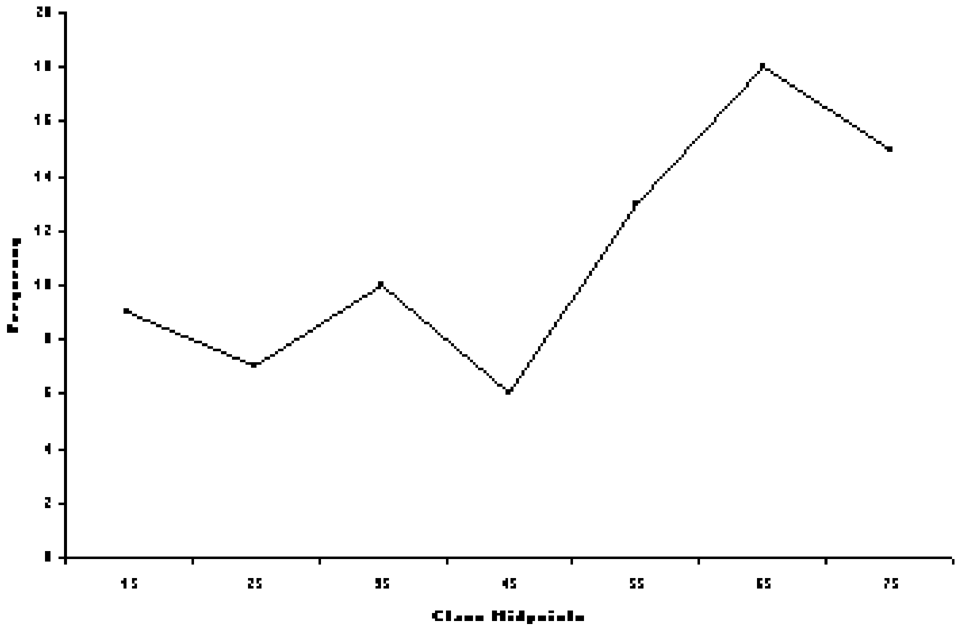
Frequency Polygon:



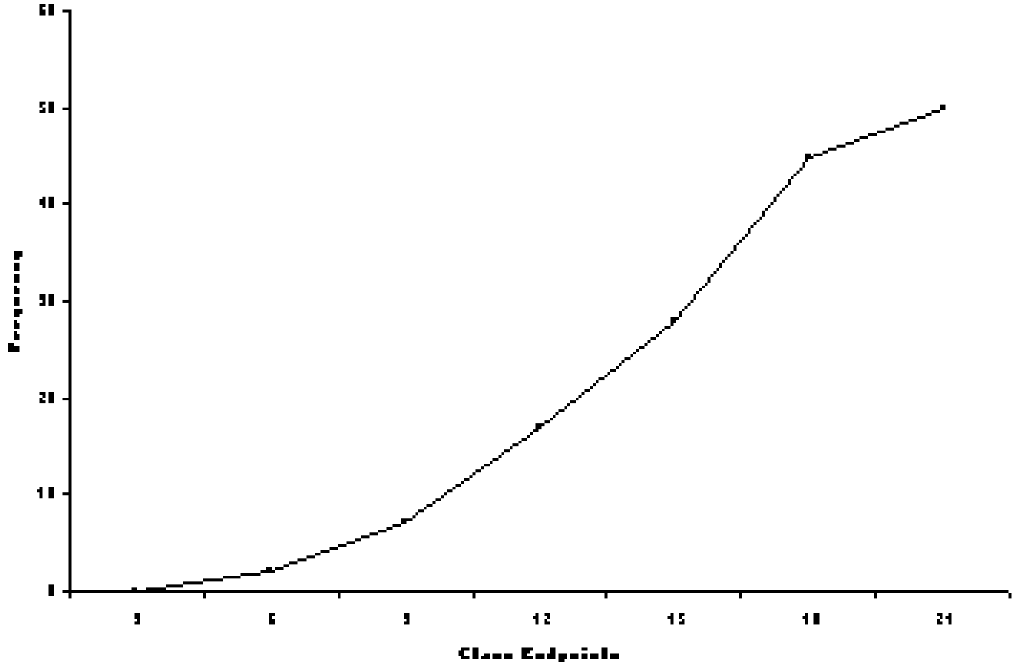
2.7 Histogram:


20
18
16
14
12
10
8
6
4
2
0
Frequency
15
25
35
45
55
65
75
Class Midpoints

Frequency Polygon:



2.8 Ogive:


60
50
40
30
20
10
0
3
6
9
12
15
18
21
Class Endpoints

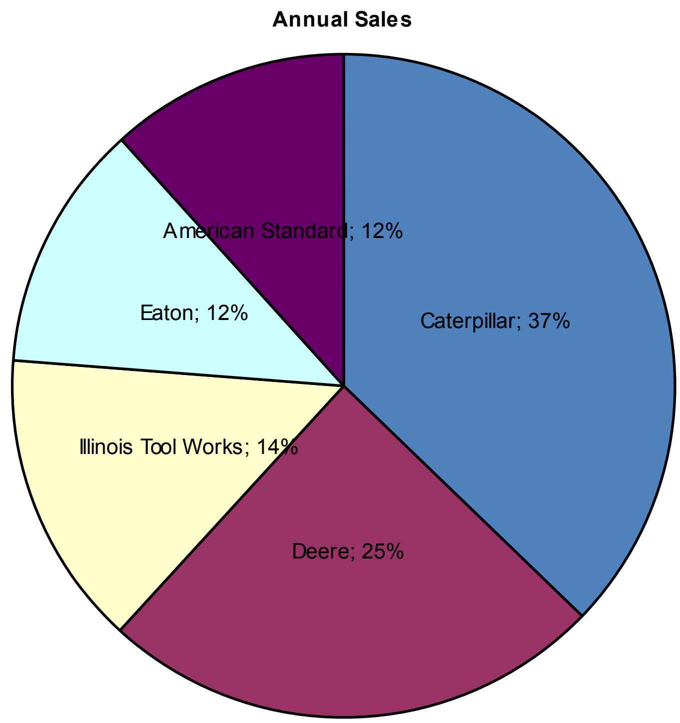
2.9

STEM	LEAF
21	2, 8, 8, 9
22	0, 1, 2, 4, 6, 6, 7, 9, 9
23	0, 0, 4, 5, 8, 8, 9, 9, 9, 9
24	0, 0, 3, 6, 9, 9, 9
25	0, 3, 4, 5, 5, 7, 7, 8, 9
26	0, 1, 1, 2, 3, 3, 5, 6
27	0, 1, 3

2.10

Firm	Proportion	Degrees
Caterpillar	.372	134
Deere	.246	89
Illinois Too Works	.144	52
Eaton	.121	44
American Standard	.117	42
TOTAL	1.000	361

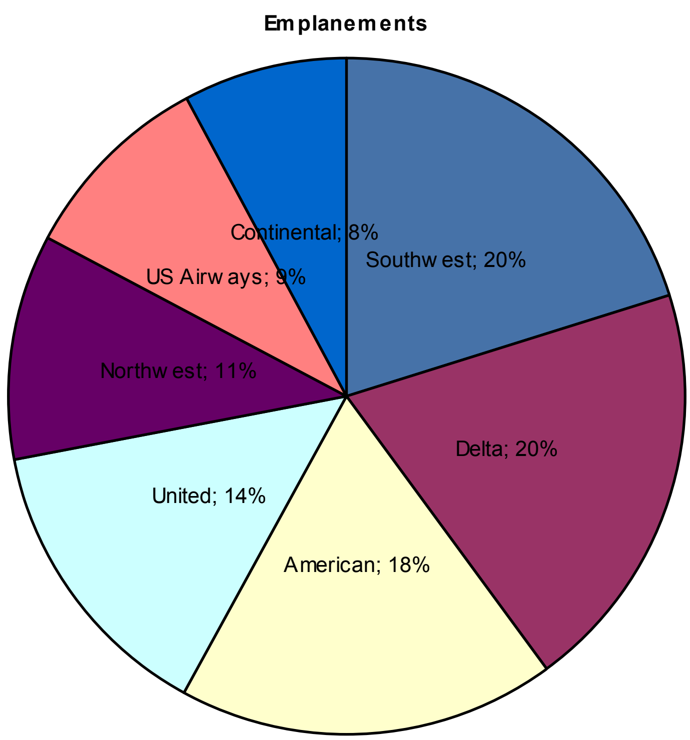
Pie Chart:


Annual Sales
American Standard; 12%
Eaton; 12%
Caterpillar; 37%
Illinois Tool Works; 14%
Deere; 25%

2\. 11

Company	Proportion	Degrees
Southwest	.202	73
Delta	.198	71
American	.181	65
United	.140	50
Northwest	.108	39
US Airways	.094	34
Continental	.078	28
TOTAL	1.001	360

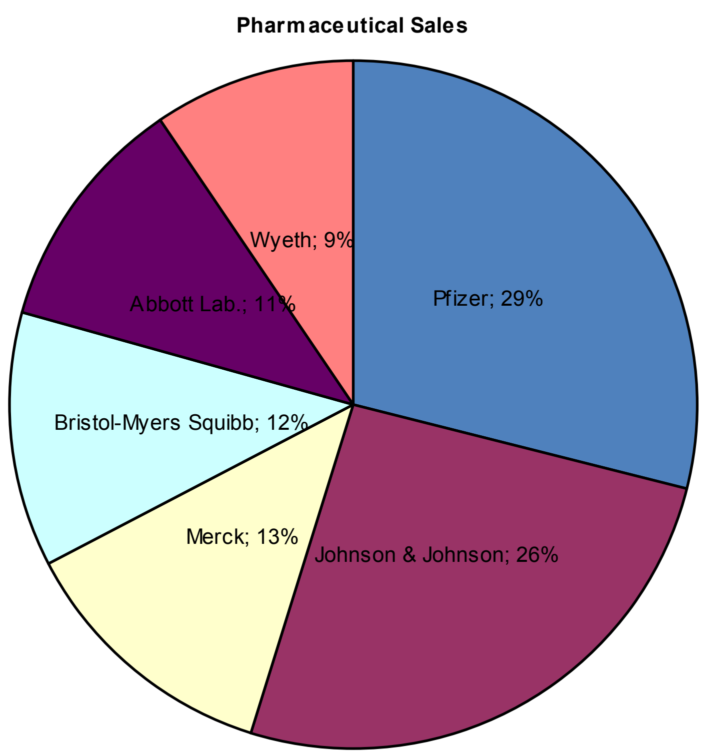
Pie Chart:



2.12

Brand	Proportion	Degrees
Pfizer	.289	104
Johnson & Johnson	.259	93
Merck	.125	45
Bristol-Myers Squibb	.120	43
Abbott Laboratories	.112	40
Wyeth	.095	34
TOTAL	1.000	359

Pie Chart:



2.13

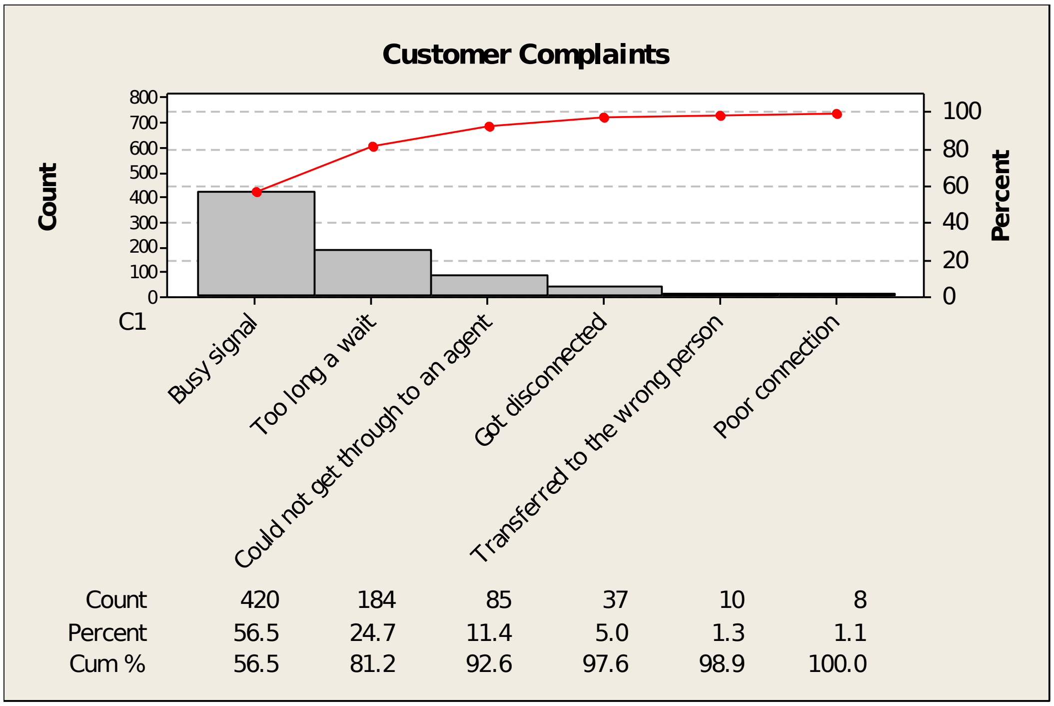
STEM	LEAF
1	3, 6, 7, 7, 7, 9, 9, 9
2	0, 3, 3, 5, 7, 8, 9, 9
3	2, 3, 4, 5, 7, 8, 8
4	1, 4, 5, 6, 6, 7, 7, 8, 8, 9
5	0, 1, 2, 2, 7, 8, 9
6	0, 1, 4, 5, 6, 7, 9
7	0, 7
8	0

The stem and leaf plot shows that the number of passengers per flight were relatively evenly distributed between the high teens through the sixties. Rarely was there a flight with at least 70 passengers. The category of 40's contained the most flights (10).

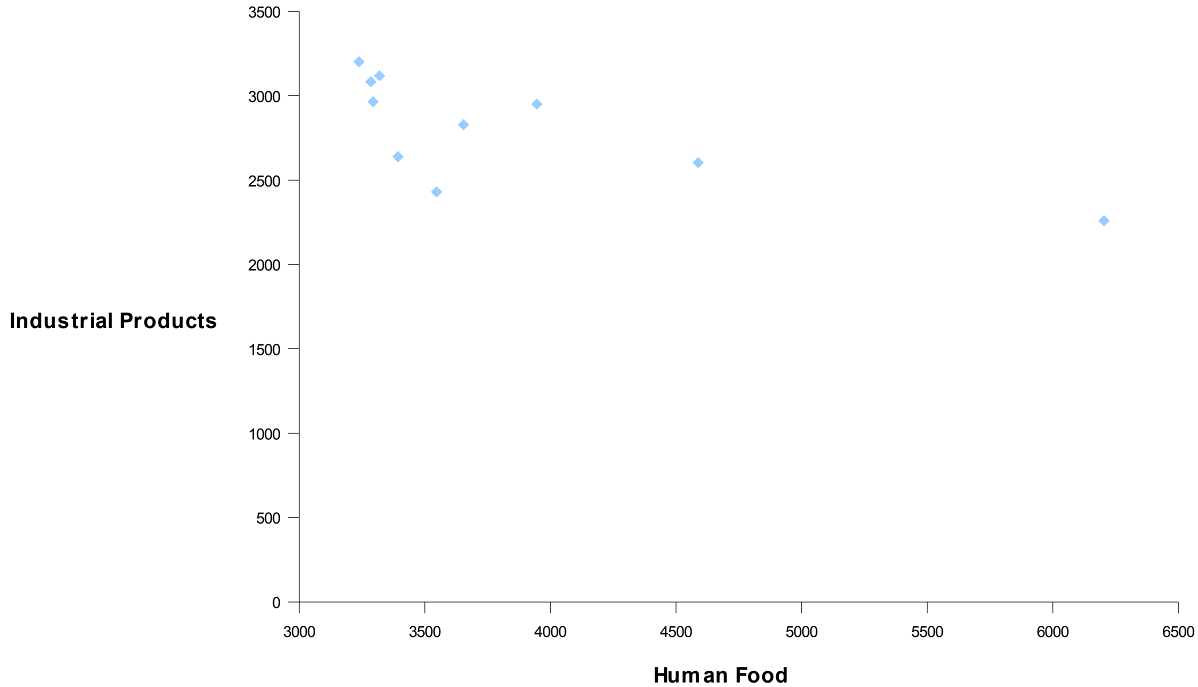
2.14

Complaint	Number	% of Total
Busy Signal	420	56.45
Too long a Wait	184	24.73
Could not get through	85	11.42
Got Disconnected	37	4.97
Transferred to the Wrong Person	10	1.34

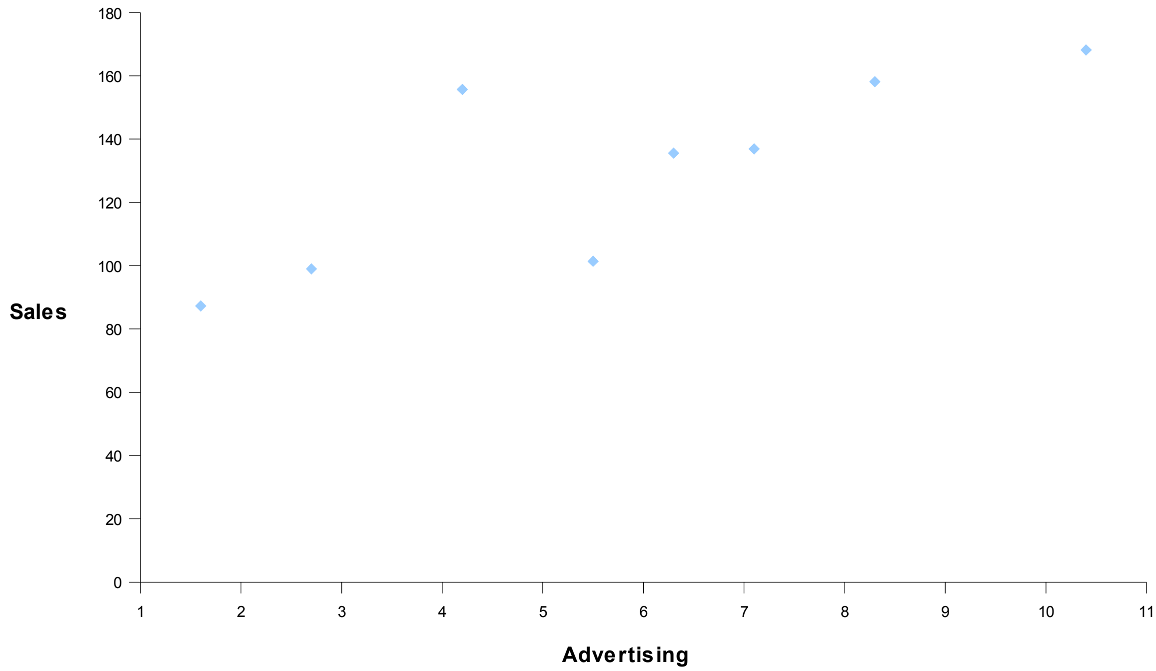
Poor Connection	8	1.08
Total	744	99.99



2.15


Industrial Products
3500
3000
2500
2000
1500
1000
500
0
3000
3500
4000
4500
5000
5500
6000
6500
Human Food

2.16



2.17

Class Interval	Frequencies
16 - under 23	6
23 - under 30	9
30 - under 37	4
37 - under 44	4
44 - under 51	4
51 - under 58	3
TOTAL	30

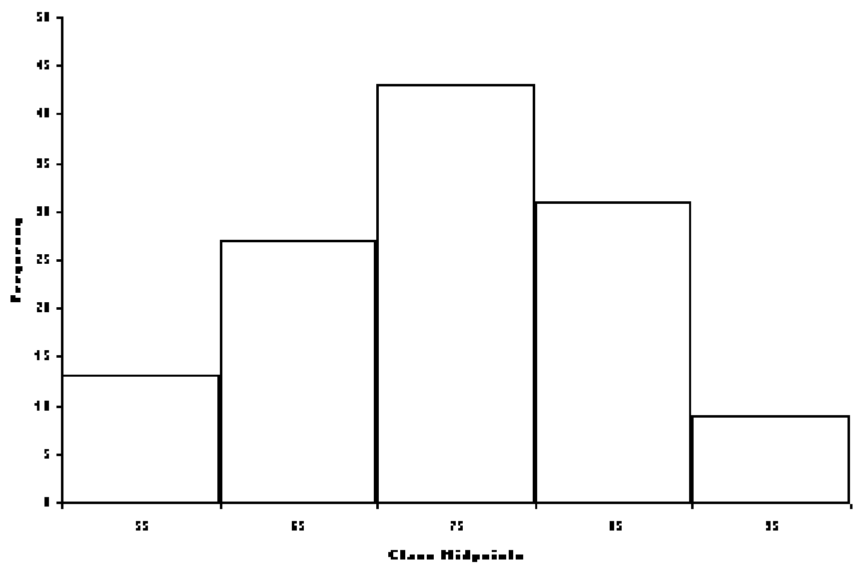
2.18

Class Interval	Frequency	Midpoint	Rel. Freq.	Cum. Freq.
20 - under 25	17	22.5	.207	17
25 - under 30	20	27.5	.244	37
30 - under 35	16	32.5	.195	53
35 - under 40	15	37.5	.183	68
40 - under 45	8	42.5	.098	76
45 - under 50	6	47.5	.073	82

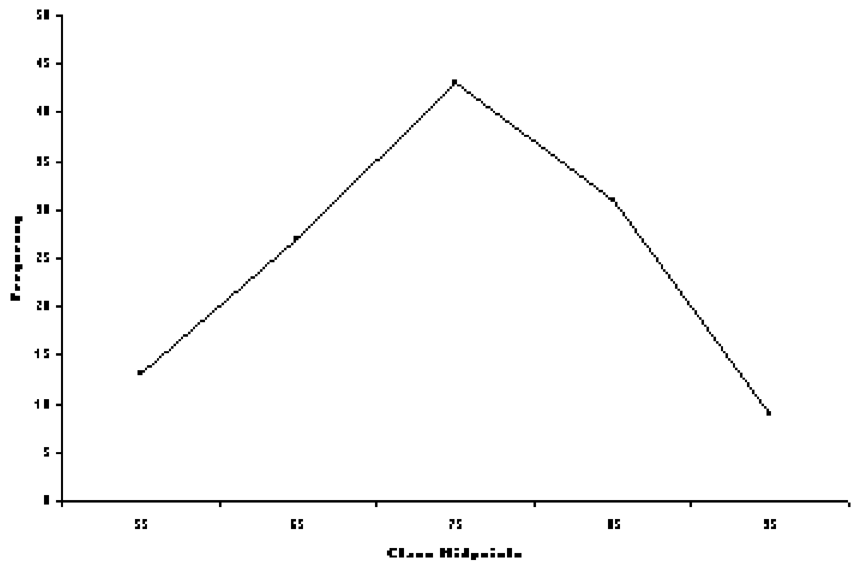
2.19 Class Interval Frequencies

Class Interval	Frequencies
50 - under 60	13
60 - under 70	27
70 - under 80	43
80 - under 90	31
90 - under 100	9
TOTAL	123

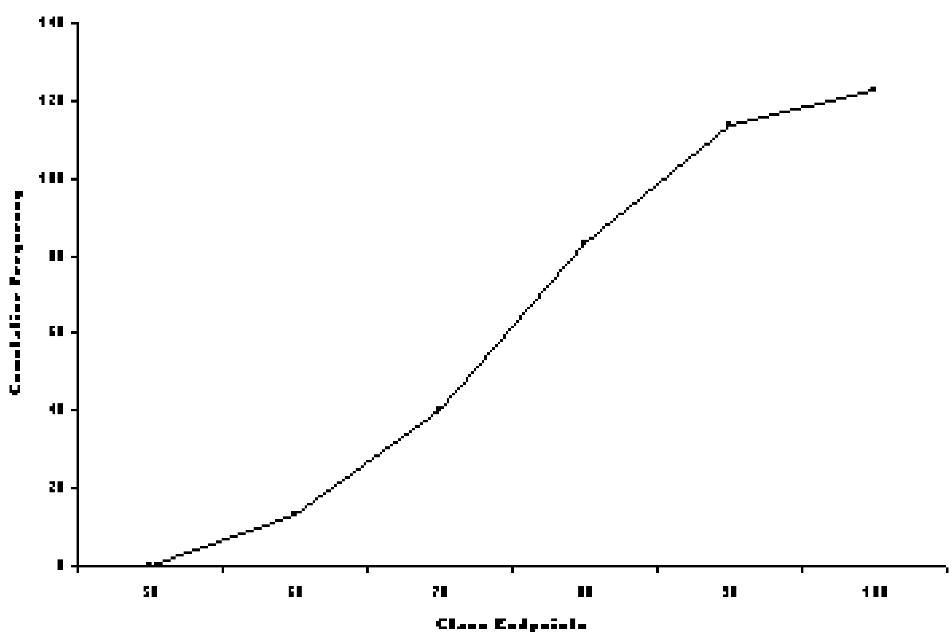
Histogram:



Frequency Polygon:



Ogive:

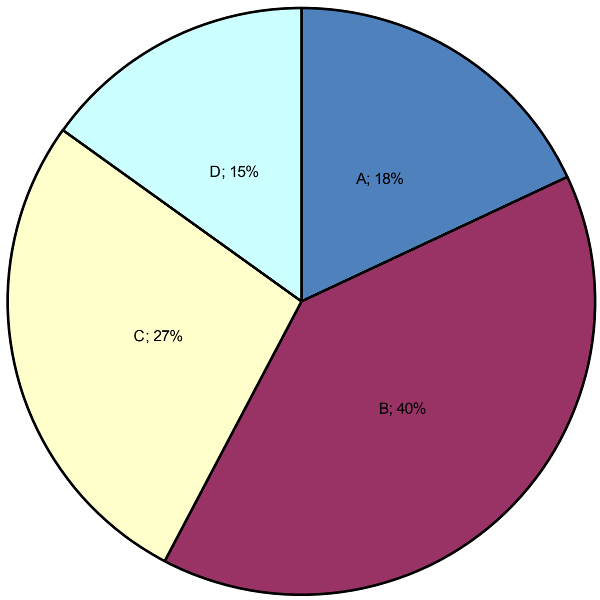


2.20

Label	Value	Proportion	Degrees
A	55	.180	65
B	121	.397	143

C	83	.272	98
D	46	.151	54
TOTAL	305	1.000	360

Pie Chart:



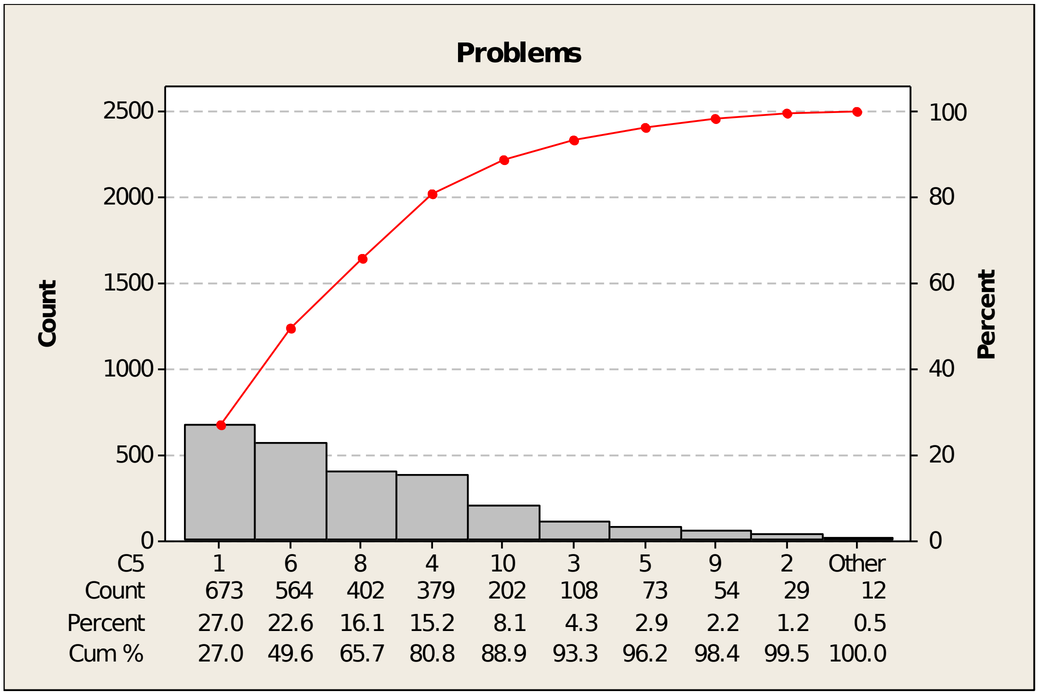
2.21

STEM	LEAF
28	4, 6, 9
29	0, 4, 8
30	1, 6, 8, 9
31	1, 2, 4, 6, 7, 7
32	4, 4, 6
33	5

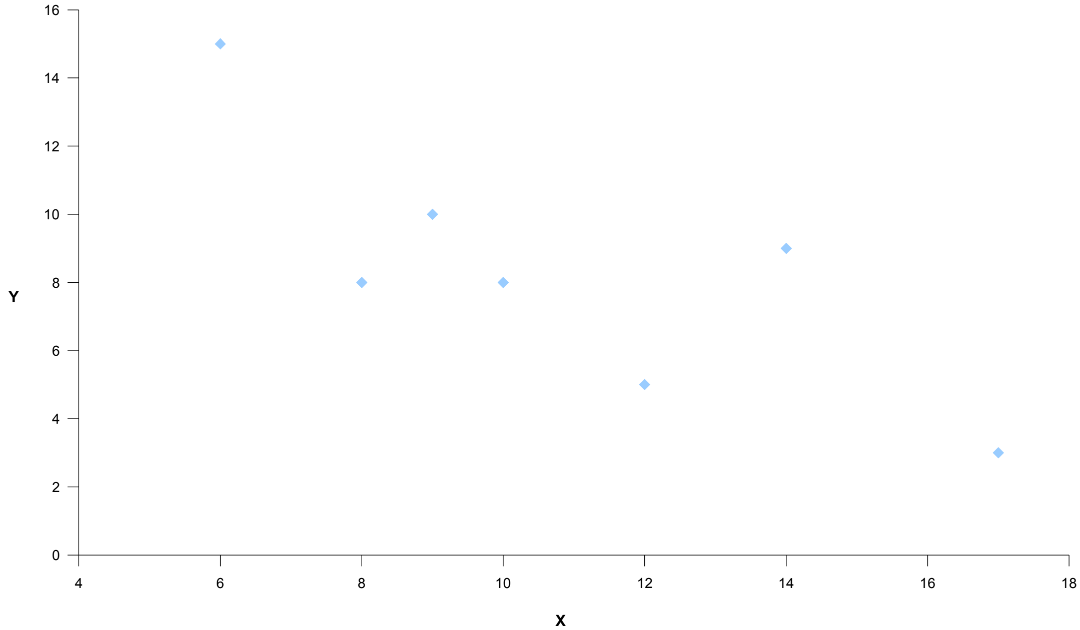
2.22

Problem	Frequency	Percent of Total
1	673	26.96
2	29	1.16
3	108	4.33
4	379	15.18
5	73	2.92
6	564	22.60
7	12	0.48
8	402	16.11
9	54	2.16
10	202	8.09
	2496	

Pareto Chart:


Problems
Count
Percent
2500
2000
1500
1000
500
0
100
80
60
40
20
0
C5 1 6 8 4 10 3 5 9 2 Other
Count 673 564 402 379 202 108 73 54 29 12
Percent 27.0 22.6 16.1 15.2 8.1 4.3 2.9 2.2 1.2 0.5
Cum % 27.0 49.6 65.7 80.8 88.9 93.3 96.2 98.4 99.5 100.0

2.23



2.24 Whitcomb Company

Class Interval	Frequency
32 - under 37	1
37 - under 42	4
42 - under 47	12
47 - under 52	11
52 - under 57	14
57 - under 62	5
62 - under 67	2

67 - under 72	1
TOTAL	50

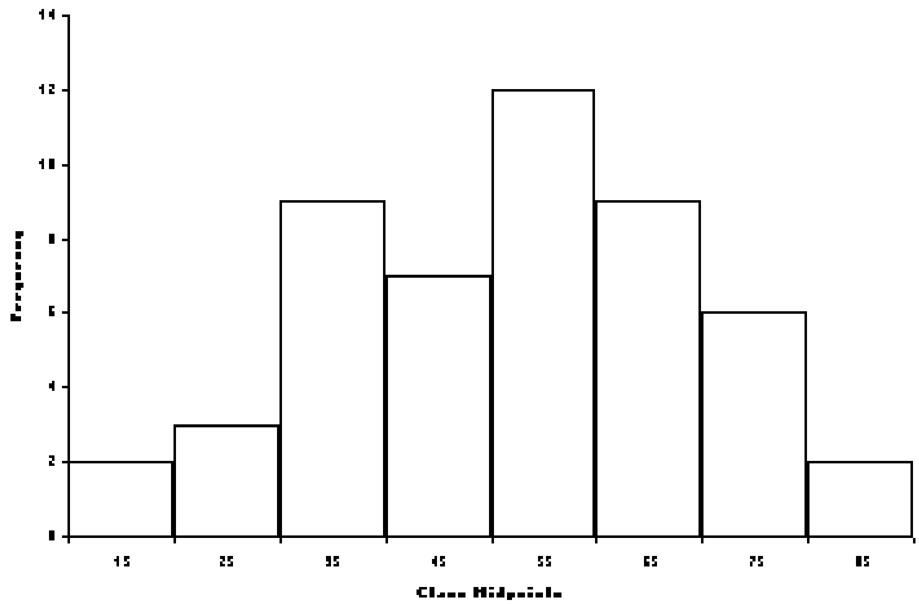
2.25

Class Interval	Frequency	Class Midpoint	Relative Frequency	Cumulative Frequency
20 – 25	8	22.5	8/53 = .1509	8
25 – 30	6	27.5	.1132	14
30 – 35	5	32.5	.0943	19
35 – 40	12	37.5	.2264	31
40 – 45	15	42.5	.2830	46
45 – 50	7	47.5	.1321	53
TOTAL	53		.9999	

2.26 Frequency Distribution:

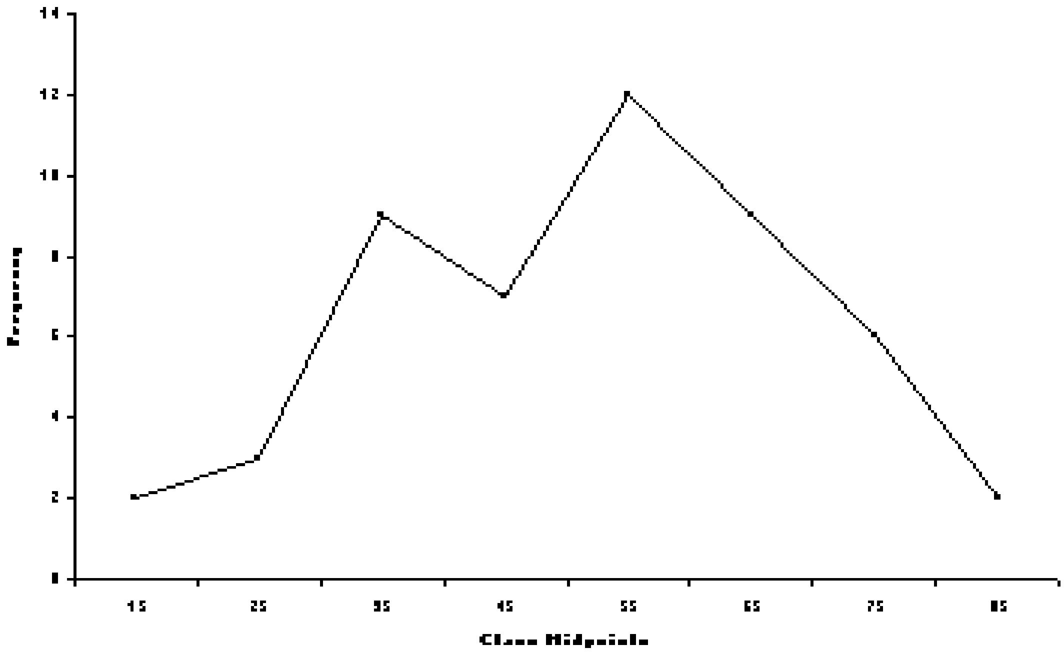
Class Interval	Frequency
10 - under 20	2
20 - under 30	3
30 - under 40	9
40 - under 50	7
50 - under 60	12
60 - under 70	9
70 - under 80	6
80 - under 90	2
	50

Histogram:



14
12
10
8
6
4
2
0
Frequency
15
25
35
45
55
65
75
85
Class Midpoints

Frequency Polygon:



The normal distribution appears to peak near the center and diminish towards the end intervals.

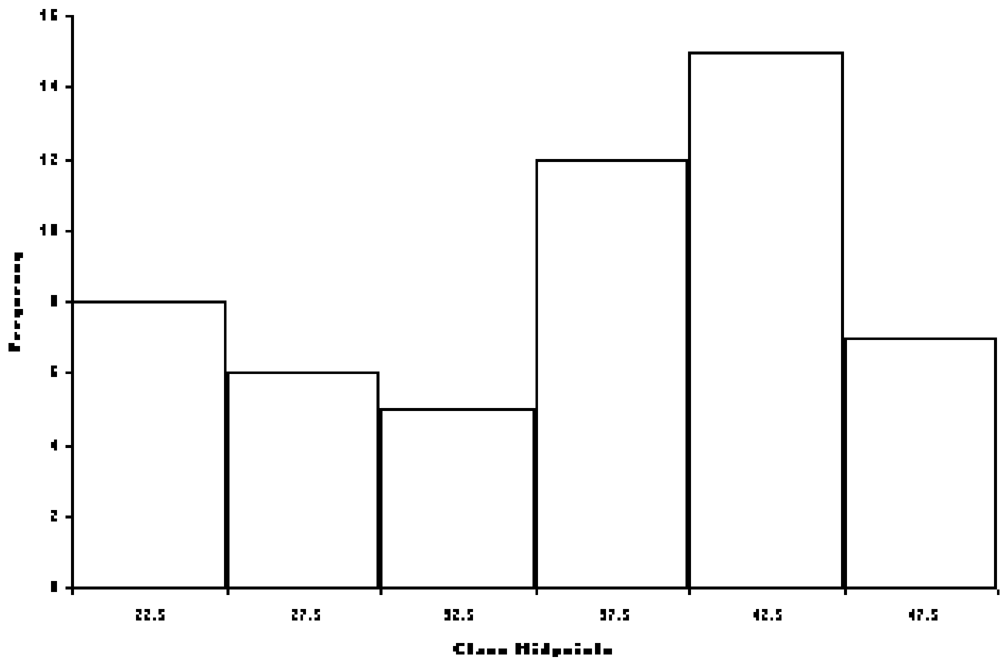
2.27

a. Histogram and a Frequency Polygon for Problem 2.25

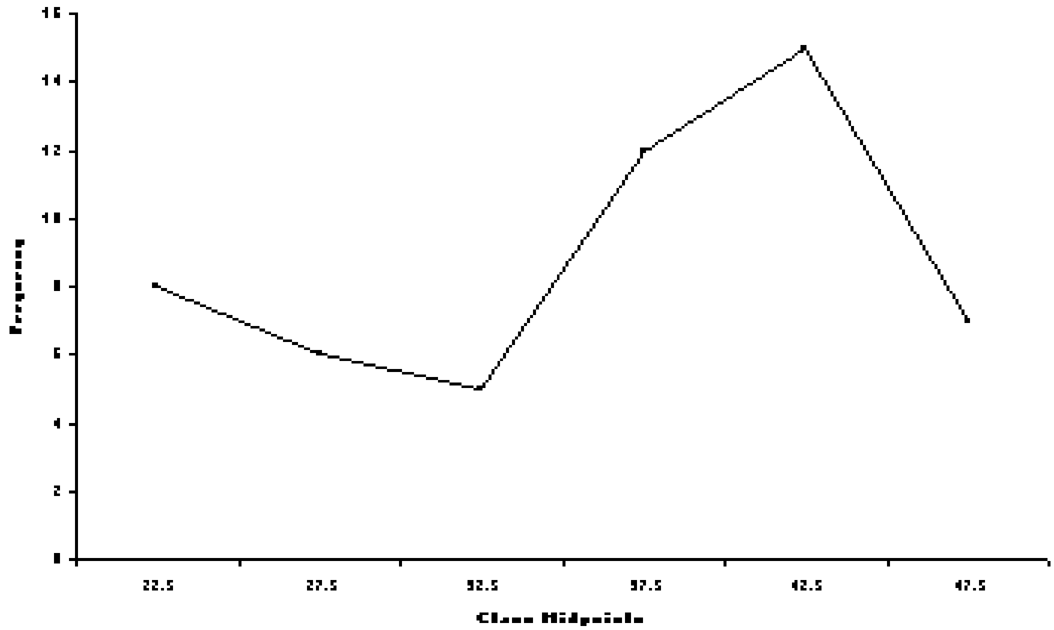
Class Interval	Frequency	Cumulative Frequency

20 – 25	8	8
25 – 30	6	14
30 – 35	5	19
35 – 40	12	31
40 – 45	15	46
45 – 50	7	53
TOTAL	53	

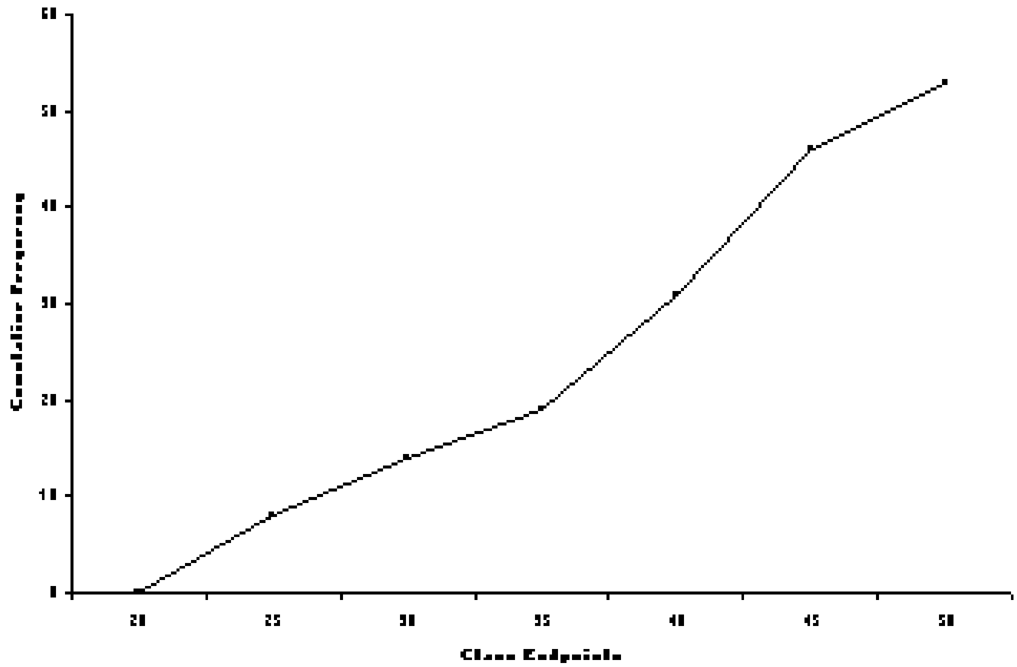
Histogram:



Frequency Polygon:


16
14
12
10
8
6
4
2
0
22.5
27.5
32.5
37.5
42.5
47.5
Class Midpoints

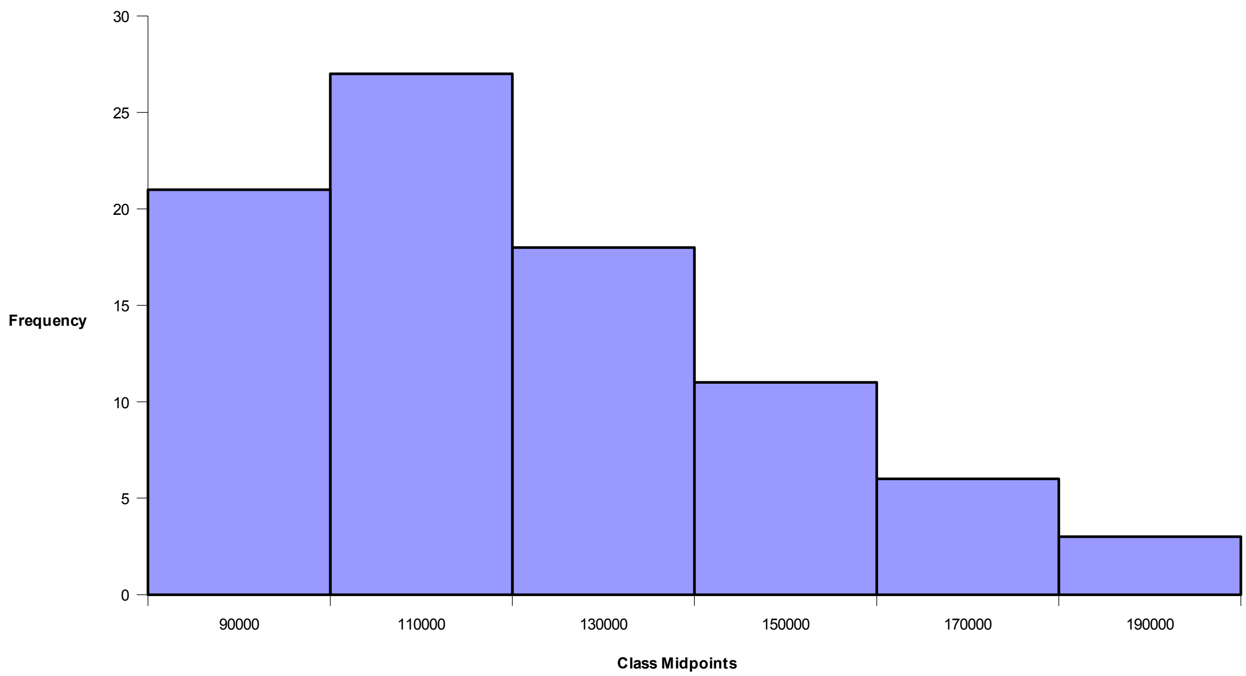
b. Ogive:



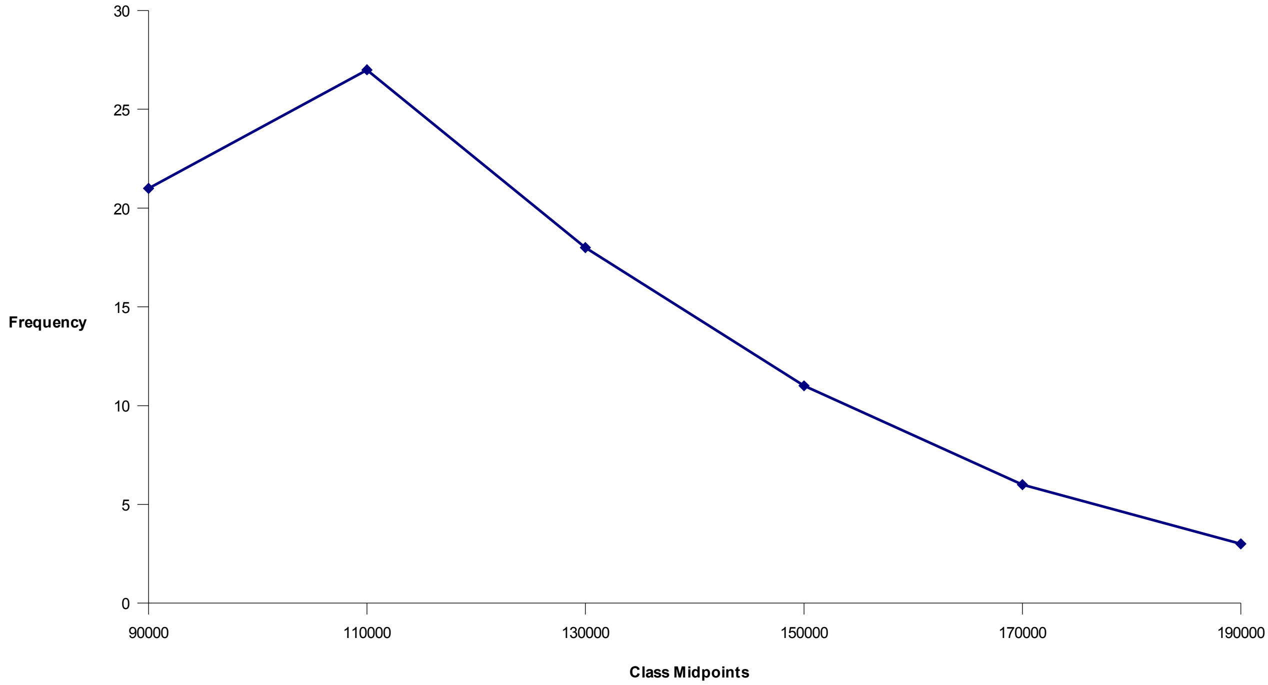
2.28 Cumulative

Asking Price	Frequency	Frequency
\$ 80,000 - under \$ 100,000	21	21
\$ 100,000 - under \$ 120,000	27	48
\$ 120,000 - under \$ 140,000	18	66
\$ 140,000 - under \$ 160,000	11	77
\$ 160,000 - under \$ 180,000	6	83
\$ 180,000 - under \$ 200,000	3	86
	86	

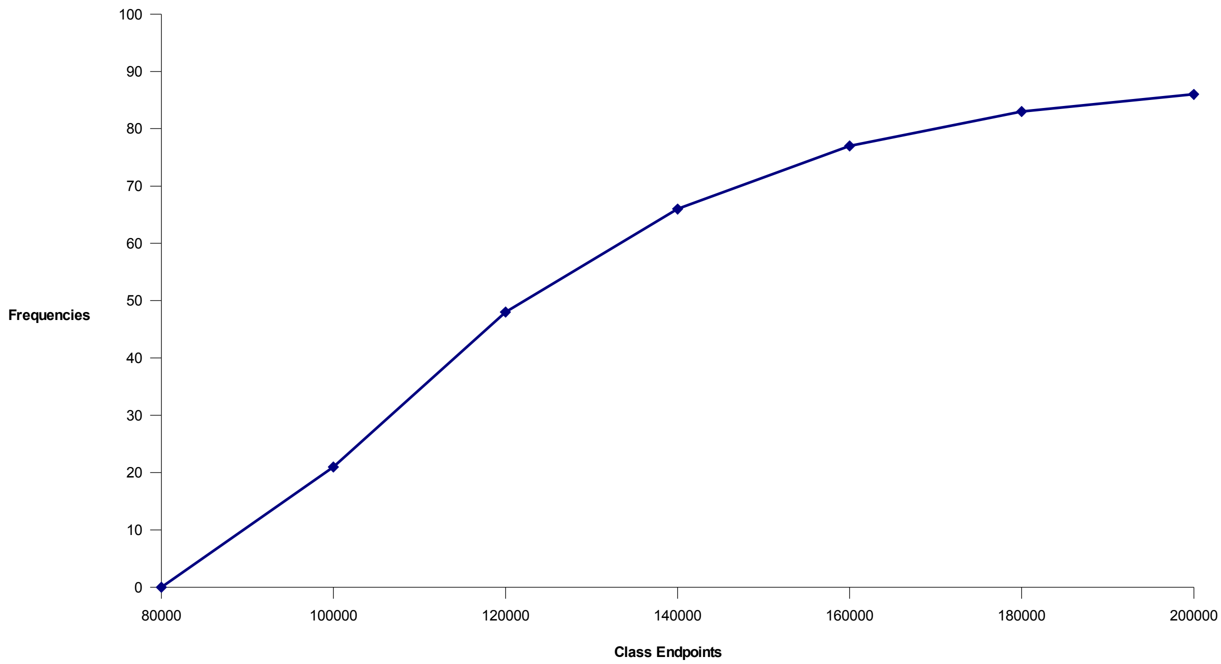
Histogram:



Frequency Polygon:


Frequency
30
25
20
15
10
5
0
90000
110000
130000
150000
170000
190000
Class Midpoints

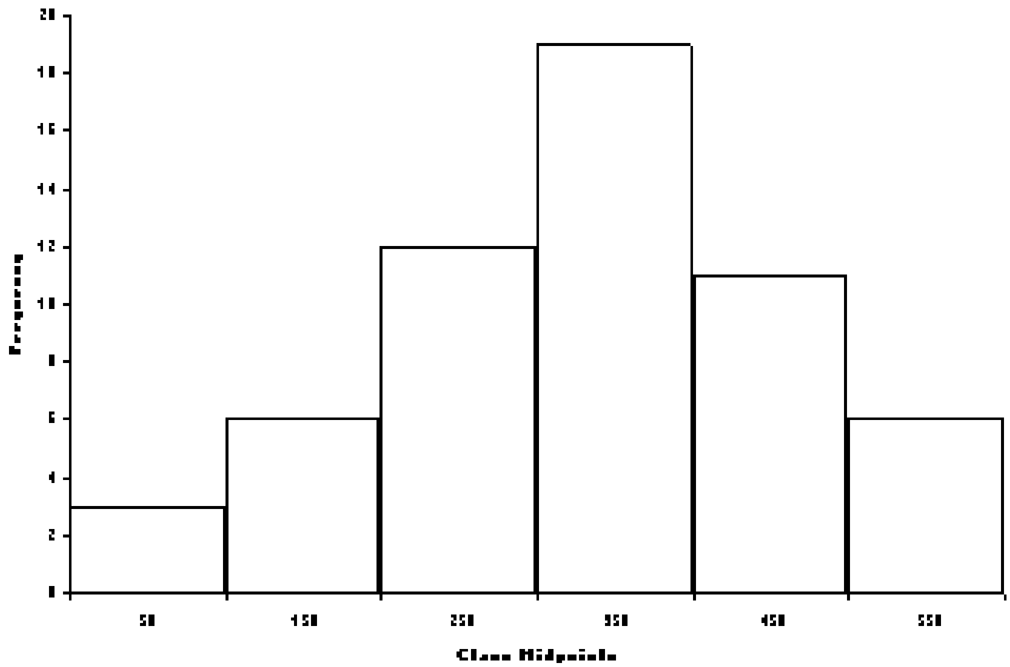
Ogive:



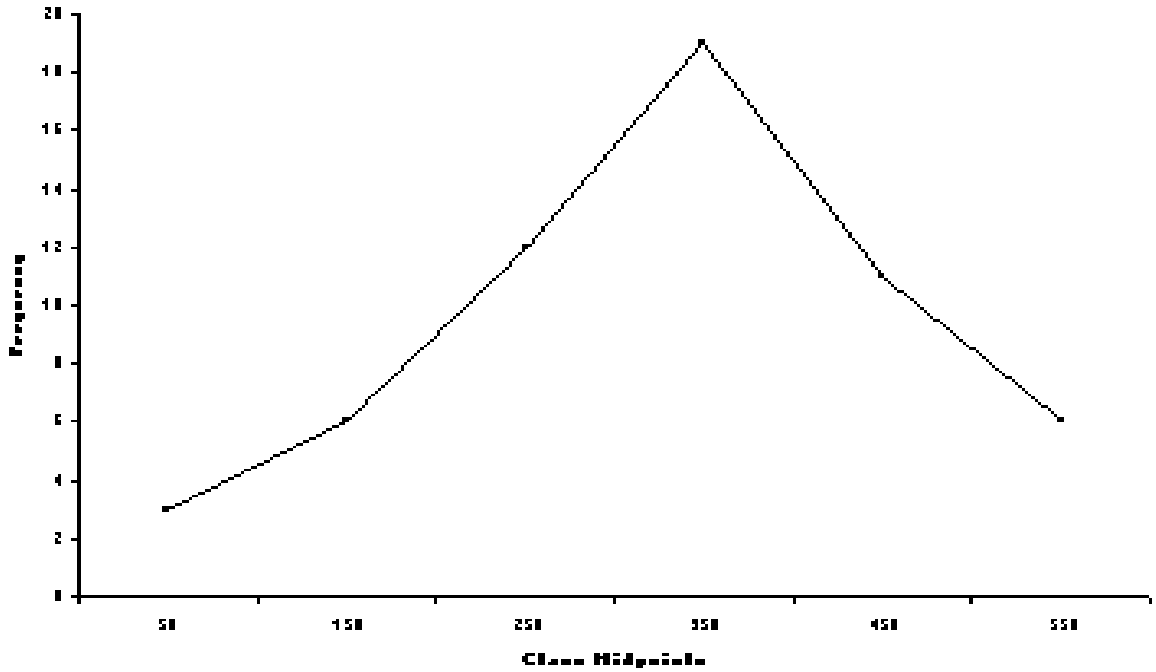
2.29

Amount Spent on Prenatal Care	Frequency	Cumulative Frequency
$ 0 - under $100	3	3
$100 - under $200	6	9
$200 - under $300	12	21
$300 - under $400	19	40
$400 - under $500	11	51
$500 - under $600	6	57
	57	

Histogram:

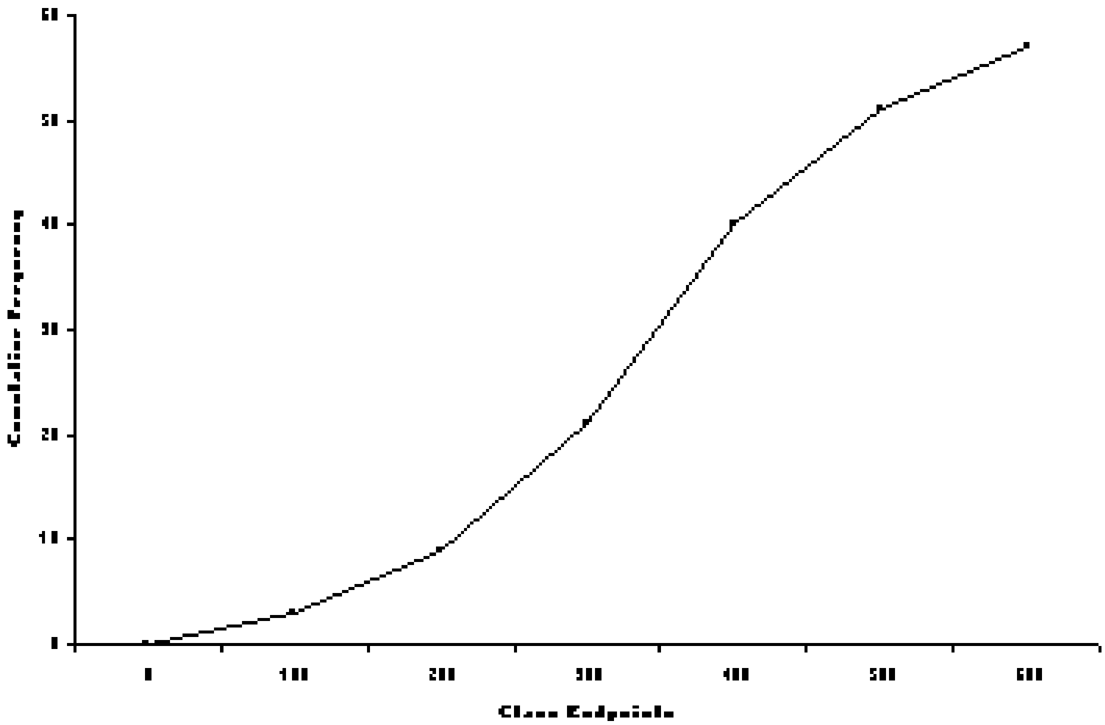


Frequency Polygon:



20
18
16
14
12
10
8
6
4
2
0
Frequency
50
150
250
350
450
550
Class Midpoints

Ogive:

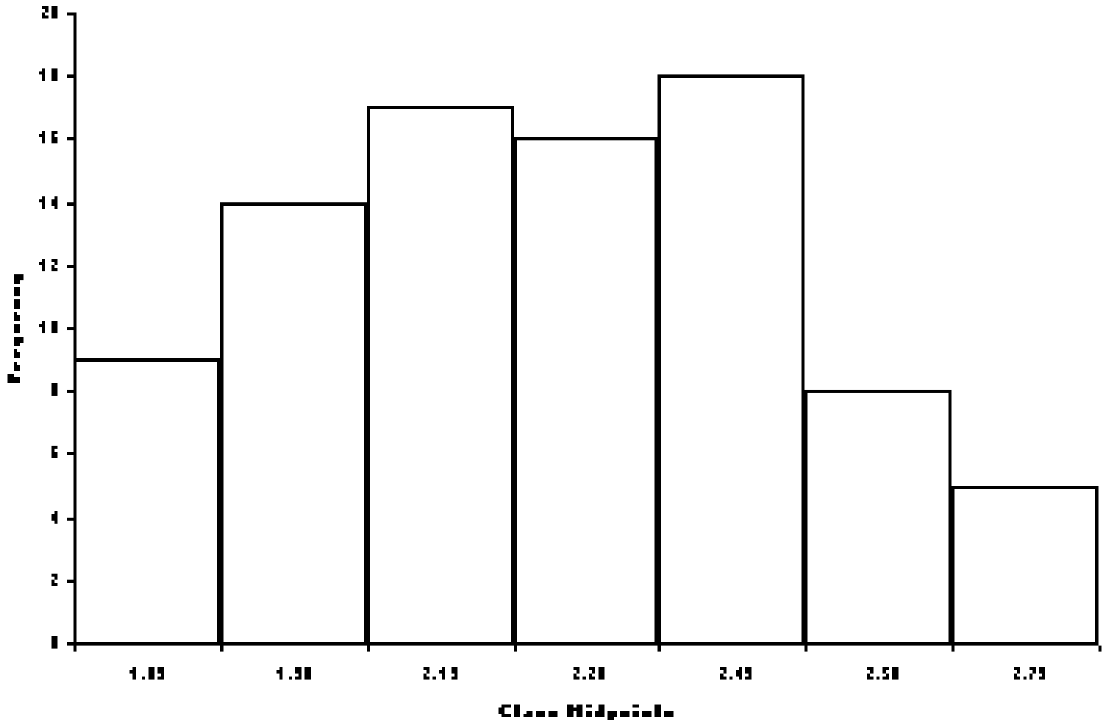


2.30

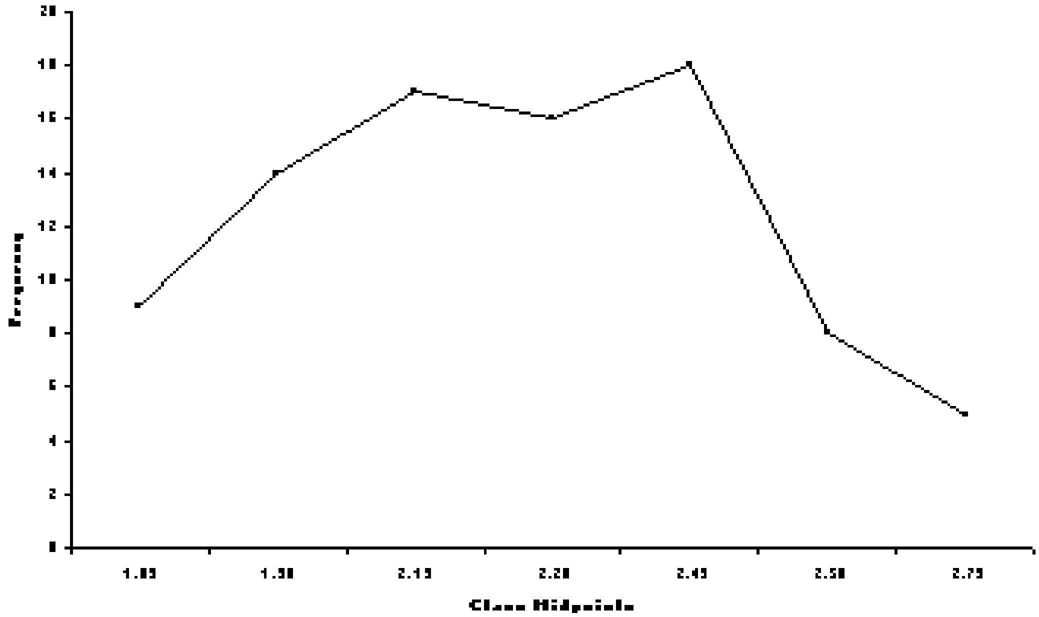
Price	Frequency	Cumulative Frequency
$1.75 - under $1.90	9	9
$1.90 - under $2.05	14	23
$2.05 - under $2.20	17	40

$2.20 - under $2.35	16	56
$2.35 - under $2.50	18	74
$2.50 - under $2.65	8	82
$2.65 - under $2.80	5	87
	87	

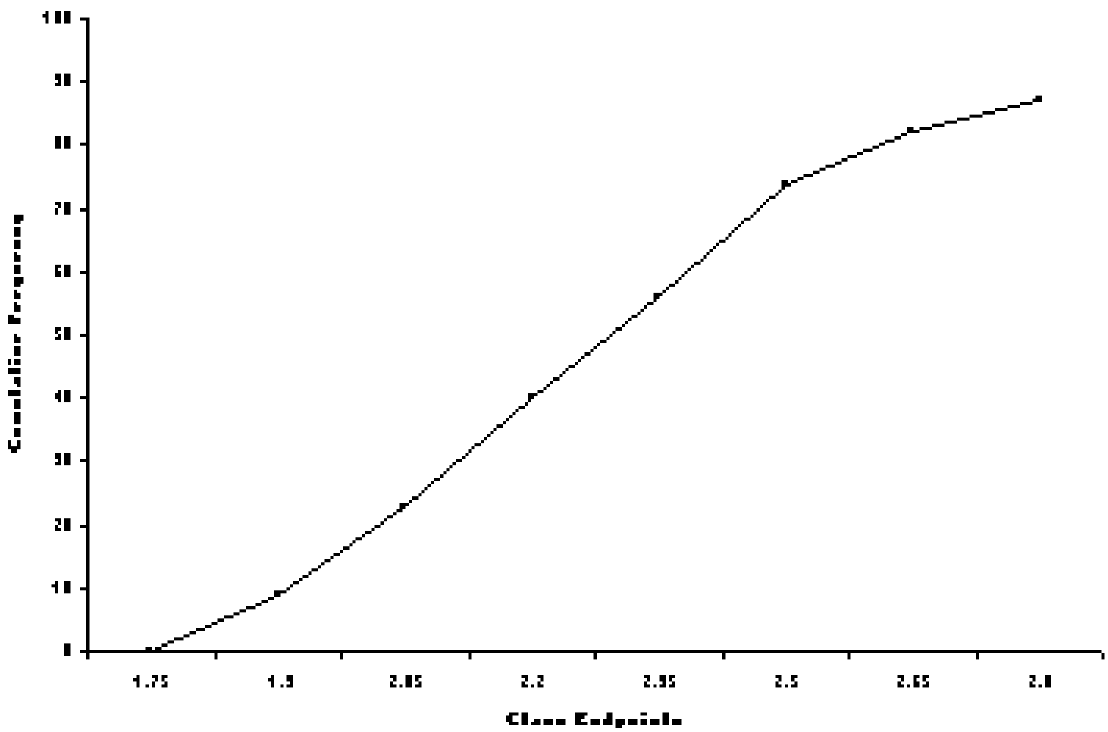
Histogram:



Frequency Polygon:



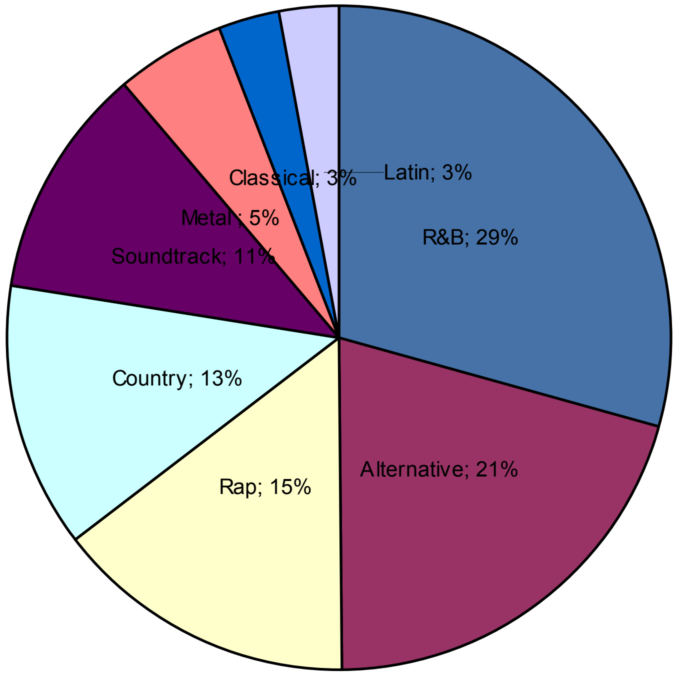
Ogive:



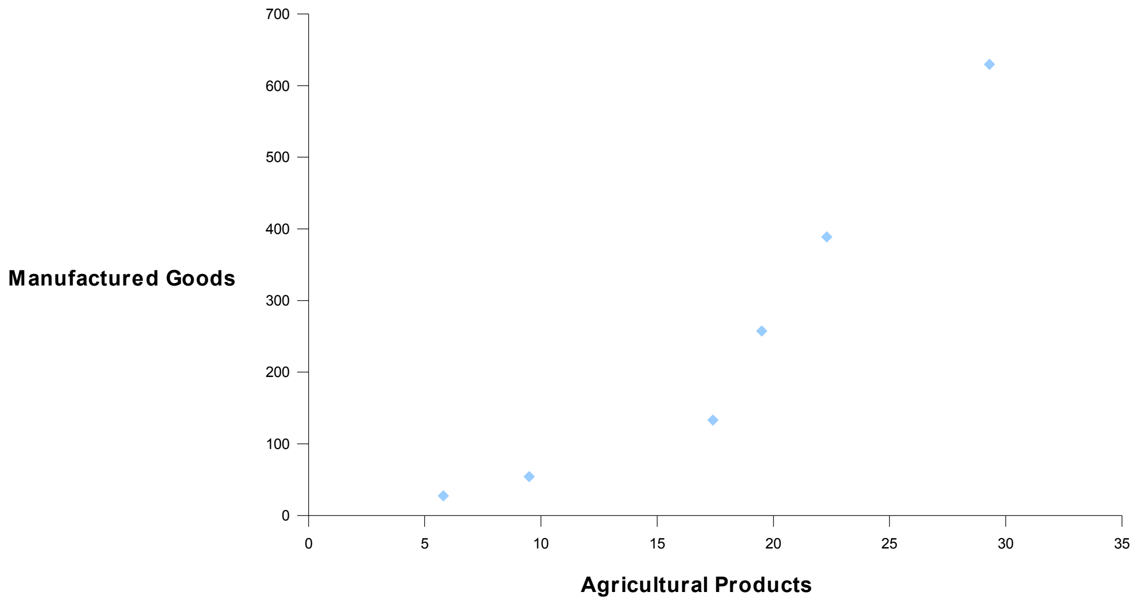
2.31

Genre	Albums Sold	Proportion	Degrees
R&B	146.4	.29	104
Alternative	102.6	.21	76
Rap	73.7	.15	54
Country	64.5	.13	47
Soundtrack	56.4	.11	40
Metal	26.6	.05	18
Classical	14.8	.03	11
Latin	14.5	.03	11
TOTAL		1.00	361

Pie Chart:


Latin; 3%
Classical; 3%
Metal; 5%
R&B; 29%
Soundtrack; 11%
Country; 13%
Alternative; 21%
Rap; 15%

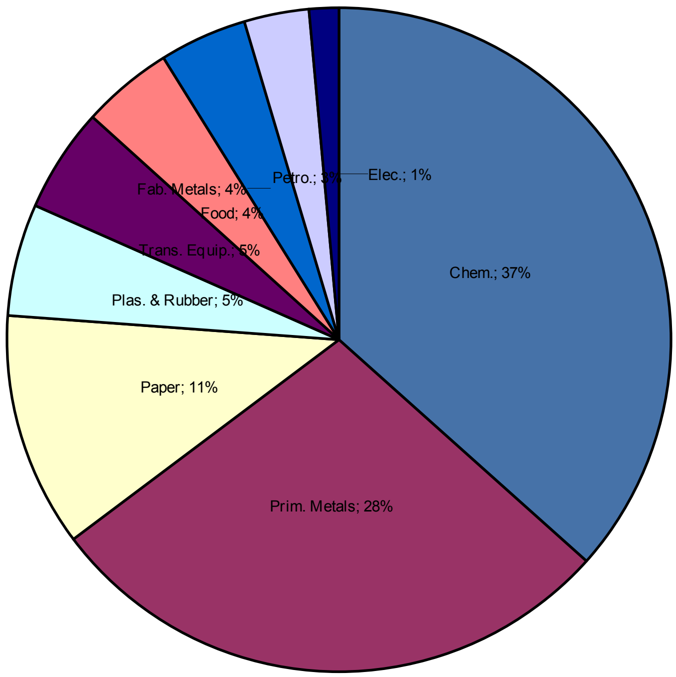
2.32


Manufactured Goods
Agricultural Products
700
600
500
400
300
200
100
0
0
5
10
15
20
25
30
35

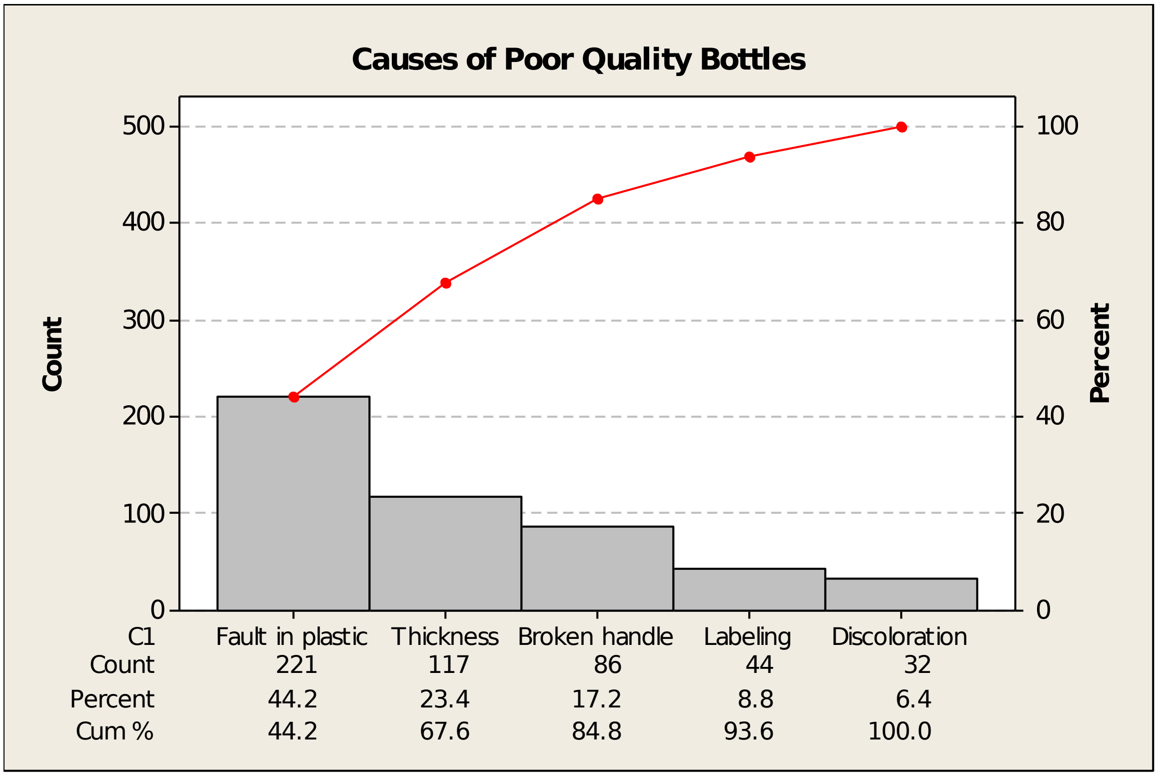
2.33

Industry	Total Release	Proportion	Degrees
Chemicals	737,100,000	.366	132
Primary metals	566,400,000	.281	103
Paper	229,900,000	.114	41
Plastics & Rubber	109,700,000	.054	19
Transportation Equipment	102,500,000	.051	18
Food	89,300,000	.044	16
Fabricated Metals	85,900,000	.043	15
Petroleum	63,300,000	.031	11
Electrical Equipment	29,100,000	.014	5
TOTAL		0.998	360

Pie Chart:


Elec.; 1%
Petro.; 3%
Fab. Metals; 4%
Food; 4%
Trans. Equip.; 5%
Plas. & Rubber; 5%
Paper; 11%
Prim. Metals; 28%
Chem.; 37%

2.34



2.35 STEM LEAF

STEM	LEAF
42	12, 16, 24, 32, 99, 99
43	04, 28, 39, 46, 61, 88
44	20, 40, 59

45	12
46	53, 54
47	30, 34, 58
48	22, 34, 66, 78
49	63
50	48, 49, 90
51	66
52	21, 54, 57, 63, 91
53	38, 66, 66
54	31, 78
55	56
56	69
57	37, 50
58	31, 32, 58, 73
59	19, 23

2.36

STEM	LEAF
22	00, 68
23	01, 37, 44, 75
24	05, 37, 48, 60, 68
25	24, 55
26	02, 56, 70, 77
27	42, 60, 64
28	14, 30
29	22, 61, 75, 76, 90, 96

30 02, 10

2.37 The distribution of household income is bell-shaped with an average of about $ 90,000 and a range of from $ 30,000 to $ 140,000.

2.38 Family practice is most prevalent with about 20% with pediatrics next at slightly less. A virtual tie exists between ob/gyn, general surgery, anesthesiology, and psychiatry at about 14% each.

2.39 The fewest number of audits is 12 and the most is 42. More companies (8) performed 27 audits than any other number. Thirty-five companies performed between 12 and 19 audits. Only 7 companies performed 40 or more audits.

2.40 There were relatively constant sales from January through August ($4 to 6 million).

Each month from September through December sales increased with December

having the sharpest increase ($15 million in sales in December).

Chapter 3

Descriptive Statistics

LEARNING OBJECTIVES

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

1. Distinguish between measures of central tendency, measures of variability, and measures of shape.
2. Understand conceptually the meanings of mean, median, mode, quartile, percentile, and range.
3. Compute mean, median, mode, percentile, quartile, range, variance, standard deviation, and mean absolute deviation on ungrouped data.
4. Differentiate between sample and population variance and standard deviation.
5. Understand the meaning of standard deviation as it is applied using the empirical rule and Chebyshev's theorem.
6. Compute the mean, median, standard deviation, and variance on grouped data.
7. Understand box and whisker plots, skewness, and kurtosis.
8. Compute a coefficient of correlation and interpret it.

CHAPTER TEACHING STRATEGY

In chapter 2, the students learned how to summarize data by constructing frequency distributions (grouping data) and by using graphical depictions. Much of the time, statisticians need to describe data by using single numerical measures. Chapter 3 presents a cadre of statistical measures for describing numerically sets of data.

It can be emphasized in this chapter that there are at least two major dimensions along which data can be described. One is the measure of central tendency with which statisticians attempt to describe the more central portions of the data. Included here are the mean, median, mode, percentiles, and quartiles. It is important to establish that the median is a useful device for reporting some business data, such as income and housing costs, because it tends to ignore the extremes. On the other hand, the mean utilizes every number of a data set in its computation. This makes the mean an attractive tool in statistical analysis.

A second major group of descriptive statistical techniques are the measures of variability. Students can understand that a measure of central tendency is often not enough to fully describe data, often giving information only about the center of the distribution or key milestones of the distribution. A measure of variability helps the researcher get a handle on the spread of the data. An attempt is made in this text to communicate to the student that through the use of the empirical rule and/or Chebyshev's Theorem, students can better understand the meaning of a standard deviation. The empirical rule will be referred to quite often throughout the course; and therefore, it is important to emphasize it as a rule of thumb. For example, in discussing control charts in chapter 18, the upper and lower control limits are established by using the range of $\pm$ 3 standard deviations of the statistic as limits within which 99.7% of the data values should fall if a process is in control.

In this section of chapter 3, z scores are presented mainly to bridge the gap between the discussion of means and standard deviations in chapter 3 and the normal curve of chapter 6. One application of the standard deviation in business is the use of it as a measure of risk in the financial world. For example, in tracking the price of a stock over a period of time, a financial analyst might determine that the larger the standard deviation, the greater the risk (because of “swings” in the price). However, because the size of a standard deviation is a function of the mean and a coefficient of variation conveys the size of a standard deviation relative to its mean, other financial researchers prefer the coefficient of variation as a measure of the risk. That is, it can be argued that a coefficient of variation takes into account the size of the mean (in the case of a stock, the investment) in determining the amount of risk as measured by a standard deviation.

It should be emphasized that the calculation of measures of central tendency and variability for grouped data is different than for ungrouped or raw data. While the principles are the same for the two types of data, implementation of the formulas is different. Computations of statistics from grouped data are based on class midpoints rather than raw values; and for this reason, students should be cautioned that group statistics are often just approximations.

Measures of shape are useful in helping the researcher describe a distribution of data. The Pearsonian coefficient of skewness is a handy tool for ascertaining the degree of skewness in the distribution. Box and Whisker plots can be used to determine the presence of skewness in a distribution and to locate outliers. The coefficient of correlation is introduced here instead of chapter 14 (regression chapter) so that the student can begin to think about two-variable relationships and analyses and view a correlation coefficient as a descriptive statistic. In addition, when the student studies simple regression in chapter 14, there will be a foundation upon which to build. All in all, chapter 3 is quite important because it presents some of the building blocks for many of the later chapters.

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

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

3.5 Measures of Association

Correlation

3.6 Descriptive Statistics on the Computer

KEY TERMS

Arithmetic Mean

Bimodal

Measures of Shape

Measures of Variability

Box and Whisker Plot
Chebyshev's Theorem
Coefficient of Correlation (*r*)
Coefficient of Skewness
Coefficient of Variation (CV)
Correlation
Deviation from the Mean
Empirical Rule
Interquartile Range
Kurtosis
Leptokurtic
Mean Absolute Deviation (MAD)
Measures of Central Tendency
Median
Mesokurtic
Mode
Multimodal
Percentiles
Platykurtic
Quartiles
Range
Skewness
Standard Deviation
Sum of Squares of *x*
Variance
z Score

SOLUTIONS TO PROBLEMS IN CHAPTER 3

3.1 **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

3.2 **Median** for values in 3.1

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 $\frac{n+1}{2}$th term = $\frac{15+1}{2}$ = 8th term

The 8th term = **4**

3.3 **Median**

Arrange terms in ascending order:

073, 167, 199, 213, 243, 345, 444, 524, 609, 682

There are 10 terms.

Since there are an even number of terms, the median is the average of the

middle two terms:

Median = $\frac{(243+345)}{2} = \frac{588}{2}$ = 294

Using the formula, the median is located at the $\frac{n+1}{2}^{th}$ term

n = 10 therefore $\frac{10+1}{2} = \frac{11}{2}$ = 5.5th term.

The median is located halfway between the 5th and 6th terms.

5th term = 243   6th term = 345

Halfway between 243 and 345 is the median = **294**

3.4 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

<u>78.5</u>   (It is not stated in the problem whether the

Σx = 333.6   data represent as population or a sample).

3.5 Mean

7

-2

5	$\mu = \Sigma x/N$ = -12/12 = **-1**
9	
0	
-3	
-6	$\bar{x} = \Sigma x/n$ = -12/12 = **-1**
-7	
-4	
-5	
2	
-8	(It is not stated in the problem whether the
Σx = -12	data represent a population or a sample).

3.6 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**

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

$Q_1 = P_{25}$ but

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

$$\left(\frac{15+1}{2}\right)^{th} = 8^{th} term$$

Q_2 = Median but: The median is located at the

Q_2 = **25**

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

$Q_3 = P_{75}$ but

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

3.7 Rearranging the data in ascending order:

80, 94, 97, 105, 107, 112, 116, 116, 118, 119, 120, 127,

128, 138, 138, 139, 142, 143, 144, 145, 150, 162, 171, 172

$n = 24$

$$i = \frac{20}{100}(24) = 4.8$$

For P_{20}:

Thus, P_{20} is located at the 4 + 1 = 5th term and P_{20} = **107**

$$i = \frac{47}{100}(24) = 11.28$$

For P_{47}:

Thus, P_{47} is located at the 11 + 1 = 12th term and P_{47} = **127**

$$i = \frac{83}{100}(24) = 19.92$$

For P_{83}:

Thus, P_{83} is located at the 19 + 1 = 20th term and P_{83} = **145**

$Q_1 = P_{25}$

For P_{25}: $i = \frac{25}{100}(24) = 6$

Thus, Q_1 is located at the 6.5th term and Q_1 = (112 + 116)/ 2 = **114**

Q_2 = Median

The median is located at the: $\left(\frac{24+1}{2}\right)^{th} = 12.5^{th}\,term$

Thus, Q_2 = (127 + 128)/ 2 = **127.5**

$Q_3 = P_{75}$

For P_{75}: $i = \frac{75}{100}(24) = 18$

Thus, Q_3 is located at the 18.5th term and Q_3 = (143 + 144)/ 2 = **143.5**

3.8 Mean = $\frac{\sum x}{N} = \frac{18,245}{15} = 1216.33$ The mean is **1216.33**.

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

Median = **1,233**

Q_2 = Median = **1,233**

$$i = \frac{63}{100}(15) = 9.45$$

For P_{63},

P_{63} is located at the 9 + 1 = 10th term, P_{63} = **1,277**

$$i = \frac{29}{100}(15) = 4.35$$

For P_{29},

P_{29} is located at the 4 + 1 = 5th term, P_{29} = **1,119**

3.9 The median is located at the $\left(\frac{12+1}{2}\right)^{th} = 6.5^{th}$ position

The median = (3.41 + 4.63)/2 = **4.02**

$$i = \frac{75}{100}(12) = 9$$

For $Q_3 = P_{75}$:

P_{75} is located halfway between the 9th and 10th terms.

$Q_3 = P_{75}$ $Q_3 = (5.70 + 7.88)/ 2 =$ **6.79**

For P_{20}: $$i = \frac{20}{100}(12) = 2.4$$

P_{20} is located at the 3rd term $P_{20} =$ **2.12**

For P_{60}: $$i = \frac{60}{100}(12) = 7.2$$

P_{60} is located at the 8th term $P_{60} =$ **5.10**

For P_{80}: $$i = \frac{80}{100}(12) = 9.6$$

P_{80} is located at the 10th term $P_{80} =$ **7.88**

For P_{93}: $$i = \frac{93}{100}(12) = 11.16$$

P_{93} is located at the 12th term $P_{93} =$ **8.97**

3.10 $n = 17$; Mean = $$\frac{\sum x}{N} = \frac{61}{17} = 3.588$$ The mean is **3.588**

The median is located at the $\left(\frac{17+1}{2}\right)^{th}$ = 9th term, Median = **4**

There are eight 4's, therefore the Mode = **4**

$Q_3 = P_{75}$: $i = \frac{75}{100}(17) = 12.75$

Q_3 is located at the 13th term and Q_3 = **4**

P_{11}: $i = \frac{11}{100}(17) = 1.87$

P_{11} is located at the 2nd term and P_{11} = **1**

P_{35}: $i = \frac{35}{100}(17) = 5.95$

P_{35} is located at the 6th term and P_{35} = **3**

P_{58}: $i = \frac{58}{100}(17) = 9.86$

P_{58} is located at the 10th term and P_{58} = **4**

P_{67}: i = $\frac{67}{100}(17) = 11.39$

P_{67} is located at the 12th term and P_{67} = **4**

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
$\Sigma x = 30$	$\Sigma\lvert x-\mu\rvert$ =	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. = $\frac{\Sigma|x-\mu|}{N} = \frac{14.2857}{7} =$ **2.0408**

c.) σ^2 = $\frac{\Sigma(x-\mu)^2}{N} = \frac{43.4284}{7}$ = **6.2041** d.) $\sigma = \sqrt{\dfrac{\Sigma(x-\mu)^2}{N}} = \sqrt{6.2041} =$ **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.) $z = \dfrac{6-4.2857}{2.4908} =$ **0.69**

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

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

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

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

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

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

3.12

$\underline{x}$	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$
4	0	0
3	1	1
0	4	16
5	1	1
2	2	4
9	5	25
4	0	0

5 1 1

$\Sigma x = 32$ $\Sigma|x-\bar{x}| = 14$ $\Sigma(x-\bar{x})^2 = 48$

$\bar{x} = \frac{\Sigma x}{n} = \frac{32}{8} = 4$

a) Range = 9 - 0 = **9**

b) M.A.D. = $\frac{\Sigma|x-\bar{x}|}{n} = \frac{14}{8}$ = **1.75**

c) s^2 = $\frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{48}{7}$ = **6.8571**

d) s = $\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{6.857}$ = **2.6186**

e) Numbers in order: 0, 2, 3, 4, 4, 5, 5, 9

$Q_1 = P_{25}$ i = $\frac{25}{100}(8)$ = 2

Q_1 is located at the average of the 2nd and 3rd terms, Q_1 = **2.5**

$Q_3 = P_{75}$ $i = \frac{75}{100}(8) = 6$

Q_3 is located at the average of the 6th and 7th terms, Q_3 = **5**

IQR = Q_3 - Q_1 = 5 - 2.5 = **2.5**

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
Σx = 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
Σx = 127	Σx^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.

3.14 $\boldsymbol{s^2 = 433.9267}$

$\boldsymbol{s = 20.8309}$

$\Sigma x = 1387$

$\Sigma x^2 = 87,365$

$n = 25$

$\bar{x} = 55.48$

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

$\boldsymbol{\sigma = 242.139}$

$\Sigma x = 6886$

$\Sigma x^2 = 3,901,664$

$n = 16$

$\mu = 430.375$

3.16

14, 15, 18, 19, 23, 24, 25, 27, 35, 37, 38, 39, 39, 40, 44,

46, 58, 59, 59, 70, 71, 73, 82, 84, 90

$Q_1 = P_{25}$ $\qquad i = \frac{25}{100}(25) = 6.25$

P_{25} is located at the 7th term, and therefore, $Q_1 = 25$

$Q_3 = P_{75}$ $\qquad i = \frac{75}{100}(25) = 18.75$

P_{75} is located at the 19th term, and therefore, $Q_3 = 59$

IQR = Q_3 - Q_1 = 59 - 25 = **34**

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**

3.18

Set 1:

$$\mu_1 = \frac{\Sigma x}{N} = \frac{262}{4} = 65.5$$

$$\sigma_1 = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{17{,}970 - \frac{(262)^2}{4}}{4}}$$

= **14.2215**

Set 2:

$$\mu_2 = \frac{\Sigma x}{N} = \frac{570}{4} = 142.5$$

$$\sigma_2 = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{82{,}070 - \frac{(570)^2}{4}}{4}}$$

= **14.5344**

CV_1 = $\frac{14.2215}{65.5}(100)$ = **21.71%**

$$CV_2 = \frac{14.5344}{142.5}(100) = \mathbf{10.20\%}$$

3.19

$\bar{x}$	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$
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 = $$\frac{\Sigma\lvert x-\bar{x}\rvert}{n} = \frac{32}{12}$$ = **2.667**

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

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

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

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

3.20 $n = 11$

x	$\lvert x-\mu \rvert$
768	475.64
429	136.64
323	30.64
306	13.64

286	6.36
262	30.36
215	77.36
172	120.36
162	130.36
148	144.36
145	147.36
$\Sigma x = 3216$	$\Sigma\lvert x-\mu\rvert = 1313.08$

$\mu = 292.36 \quad \Sigma x = 3216 \quad \Sigma x^2 = 1,267,252$

a.) Range = 768 - 145 = **623**

b.) MAD = $\dfrac{\Sigma\lvert x-\mu\rvert}{N} = \dfrac{1313.08}{11}$ = **119.37**

c.) σ^2 = $\dfrac{\Sigma x^2 - \dfrac{(\Sigma x)^2}{N}}{N} = \dfrac{1,267,252 - \dfrac{(3216)^2}{11}}{11}$ = **29,728.23**

d.) σ = $\sqrt{29,728.23}$ = **172.42**

e.) $Q_1 = P_{25}$: $\quad i = .25(11) = 2.75$

Q_1 is located at the 3rd term and $Q_1 = 162$

$Q_3 = P_{75}$: i = .75(11) = 8.25

Q_3 is located at the 9th term and Q_3 = 323

IQR = Q_3 - Q_1 = 323 - 162 = **161**

f.) x_{nestle} = 172

$$z = \frac{x-\mu}{\sigma} = \frac{172 - 292.36}{172.42} = \mathbf{-0.70}$$

g.) $$CV = \frac{\sigma}{\mu}(100) = \frac{172.42}{292.36}(100) = \mathbf{58.98\%}$$

3.21 μ = 125 σ = 12

68% of the values fall within:

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

between 113 and 137

95% of the values fall within:

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

between 101 and 149

99.7% of the values fall within:

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

between 89 and 161

3.22 μ = 38 σ = 6

between 26 and 50:

$x_1 - \mu$ = 50 - 38 = 12

$x_2 - \mu$ = 26 - 38 = -12

$$\frac{x_1 - \mu}{\sigma} = \frac{12}{6}$$
= 2

$$\frac{x_2 - \mu}{\sigma} = \frac{-12}{6}$$
= -2

-

k = 2, and since the distribution is not normal, use Chebyshev's theorem:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4}$$
= .75

at least 75% of the values will fall between 26 and 50

between 14 and 62? $\mu = 38$ $\sigma = 6$

$x_1 - \mu = 62 - 38 = 24$

$x_2 - \mu = 14 - 38 = -24$

$$\frac{x_1 - \mu}{\sigma} = \frac{24}{6} = 4$$

$$\frac{x_2 - \mu}{\sigma} = \frac{-24}{6} = -4$$

$k = 4$

$$1 - \frac{1}{k^2} = 1 - \frac{1}{4^2} = 1 - \frac{1}{16} = \frac{15}{16} = .9375$$

at least 93.75% of the values fall between 14 and 62

between what 2 values do at least 89% of the values fall?

$$1 - \frac{1}{k^2} = .89$$

$.11 = \frac{1}{k^2}$

$.11\,k^2 = 1$

$k^2 = \frac{1}{.11}$

$k^2 = 9.09$

$k = 3.015$

With $\mu = 38$, $\sigma = 6$ and $k = 3.015$ at least 89% of the values fall within:

$\mu \pm 3.015\sigma = 38 \pm 3.015\,(6) = 38 \pm 18.09$

Between **19.91 and 56.09**

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

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

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

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

2.236 standard deviations

3.24 $\mu = 43$. 68% of the values lie $\mu \pm 1\sigma$. Thus, between the mean, 43, and one of the values, 46, is one standard deviation. Therefore,

$1\sigma = 46 - 43 =$ **3**

within 99.7% of the values lie $\mu \pm 3\sigma$. Thus, between the mean, 43, and one of the values, 51, are three standard deviations. Therefore,

$3\sigma = 51 - 43 = 8$

$\sigma =$ **2.67**

$\mu = 28$ and 77% of the values lie between 24 and 32 or $\pm$ 4 from the mean:

$1 - \frac{1}{k^2} = .77$

Solving for *k*:

$.23 = \frac{1}{k^2}$ and therefore, $.23k^2 = 1$

$k^2 = 4.3478$

$k = 2.085$

$2.085\sigma = 4$

$\sigma = \frac{4}{2.085}$ = **1.918**

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

Between 21 and 37 days:

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

$\frac{x_2 - \mu}{\sigma} = \frac{37 - 29}{8} = \frac{8}{4}$ = 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:

$$\frac{x-\mu}{\sigma} = \frac{41-29}{4} = \frac{12}{4}$$

= 3 Standard deviations

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

29 ± 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: μ = 29 σ = 4

$$\frac{x-\mu}{\sigma}=\frac{25-29}{4}=\frac{-4}{4}$$

= -1 Standard Deviation

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

29 ± 1(4) = 29 ± 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**.

3.26

x
97
109
111
118
120
130
132
133
137
137

Σx = 1224 Σx^2 = 151,486 n = 10 $\bar{x}$ = 122.4 s = 13.615

Bordeaux: $x = 137$

$$z = \frac{137 - 122.4}{13.615} = \mathbf{1.07}$$

Montreal: $x = 130$

$$z = \frac{130 - 122.4}{13.615} = \mathbf{0.56}$$

Edmonton: $x = 111$

$$z = \frac{111 - 122.4}{13.615} = \mathbf{-0.84}$$

Hamilton: $x = 97$

$$z = \frac{97 - 122.4}{13.615} = \mathbf{-1.87}$$

3.27 Mean

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

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

Mode: The modal class is 0 – 2.

The midpoint of the modal class = the mode = **1**

3.28

Class	f	M	fM
1.2 - 1.6	220	1.4	308
1.6 - 2.0	150	1.8	270
2.0 - 2.4	90	2.2	198
2.4 - 2.8	110	2.6	286

2.8 - 3.2	<u>280</u>	3.0	<u>840</u>
	Σf=850		ΣfM=1902

Mean: $\mu = \frac{\Sigma fM}{\Sigma f} = \frac{1902}{850}$ = **2.24**

Mode: The modal class is 2.8 – 3.2.

The midpoint of the modal class is the mode = **3.0**

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

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

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

3.30

Class	f	M	fM	fM^2
5 - 9	20	7	140	980
9 - 13	18	11	198	2,178
13 - 17	8	15	120	1,800
17 - 21	6	19	114	2,166
21 - 25	2	23	46	1,058
	Σf=54		ΣfM= 618	ΣfM^2= 8,182

$$s^2 = \frac{\Sigma fM^2 - \frac{(\Sigma fM)^2}{n}}{n-1} = \frac{8182 - \frac{(618)^2}{54}}{53} = \frac{8182 - 7071.67}{53} = \mathbf{20.931}$$

$$s = \sqrt{s^2} = \sqrt{20.9} = \mathbf{4.575}$$

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
	$\Sigma f = 231$		$\Sigma fM = 10{,}371$	$\Sigma fM^2 = 508{,}671$

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

b.) Mode. The Modal Class = 36-42. The mode is the class midpoint = **39** c.) $s^2 = \dfrac{\Sigma fM^2 - \dfrac{(\Sigma fM)^2}{n}}{n-1} = \dfrac{508{,}671 - \dfrac{(10{,}371)^2}{231}}{230} = \dfrac{43{,}053.5065}{230}$ = **187.189**

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

3.32

Class	f	M	fM	fM^2
0 - 1	31	0.5	15.5	7.75
1 - 2	57	1.5	85.5	128.25
2 - 3	26	2.5	65.0	162.50
3 - 4	14	3.5	49.0	171.50
4 - 5	6	4.5	27.0	121.50
5 - 6	3	5.5	16.5	90.75
	Σf=137		ΣfM=258.5	ΣfM^2=682.25

a.) Mean

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{258.5}{137} = \mathbf{1.887}$$

b.) Mode: Modal Class = 1-2. Mode = **1.5**

c.) Variance:

$$\sigma^2 = \frac{\Sigma fM^2 - \frac{(\Sigma fM)^2}{N}}{N} = \frac{682.25 - \frac{(258.5)^2}{137}}{137} = \mathbf{1.4197}$$

d.) Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.4197} = \mathbf{1.1915}$$

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.) Variance:

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

d.) Standard Deviation:

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

3.34

No. of Farms	f	M	fM
0 - 20,000	16	10,000	160,000
20,000 - 40,000	11	30,000	330,000
40,000 - 60,000	10	50,000	500,000
60,000 - 80,000	6	70,000	420,000
80,000 - 100,000	5	90,000	450,000
100,000 - 120,000	1	110,000	110,000
	$\Sigma f = 49$		$\Sigma fM = 1{,}970{,}000$

$$\mu = \frac{\Sigma fM}{\Sigma f} = \frac{1{,}970{,}000}{49} = \mathbf{40{,}204}$$

The actual mean for the ungrouped data is 37,816. This computed group mean, 40,204, is really just an approximation based on using the class midpoints in the calculation. Apparently, the actual numbers of farms per state in some categories do not average to the class midpoint and in fact might be less than the class midpoint since the actual mean is less than the grouped data mean.

The value for ΣfM^2 is $1.185\text{x}10^{11}$

$$\sigma^2 = \frac{\Sigma fM^2 - \frac{(\Sigma fm)^2}{N}}{N} = \frac{1.185x10^{11} - \frac{(1{,}970{,}000)^2}{49}}{49} = \mathbf{801{,}999{,}167}$$

σ = **28,319.59**

The actual standard deviation was 29,341. The difference again is due to the grouping of the data and the use of class midpoints to represent the data. The class midpoints due not accurately reflect the raw data.

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.

3.36 mean = 51

median = 54

mode = 59

The distribution is skewed to the left. More people are older but the most

extreme ages are younger ages.

3.37 S_k = $\frac{3(\mu - M_d)}{\sigma} = \frac{3(5.51 - 3.19)}{9.59}$ = **0.726**

3.38 n = 25 $\sum x$ = 600

$\bar{x}$ = 24 s = 6.6521 M_d = 23

S_k = $\frac{3(\bar{x} - M_d)}{s} = \frac{3(24 - 23)}{6.6521}$ = **0.451**

There is a slight skewness to the right

3.39 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.

3.40 $n = 18$

Median: $\frac{(n+1)^{th}}{2} = \frac{(18+1)^{th}}{2} = \frac{19}{2}^{th}$ = 9.5th term

Median = **74**

$Q_1 = P_{25}$:

$i = \frac{25}{100}(18)$ = 4.5

Q_1 = 5th term = **66**

$Q_3 = P_{75}$:

$i = \frac{75}{100}(18)$ = 13.5

Q_3 = 14th term = **90**

Therefore, IQR = $Q_3 - Q_1$ = 90 - 66 = **24**

Inner Fences: Q_1 - 1.5 IQR = 66 - 1.5 (24) = **30**

Q_3 + 1.5 IQR = 90 + 1.5 (24) = **126**

Outer Fences: Q_1 - 3.0 IQR = 66 - 3.0 (24) = **-6**

Q_3 + 3.0 IQR = 90 + 3.0 (24) = **162**

There are no extreme outliers. The only mild outlier is 21. The distribution is positively skewed since the median is nearer to Q_1 than Q_3.

3.41 $\Sigma x = 80$ $\Sigma x^2 = 1{,}148$ $\Sigma y = 69$

$\Sigma y^2 = 815$ $\Sigma xy = 624$ $n = 7$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{624 - \frac{(80)(69)}{7}}{\sqrt{\left[1{,}148 - \frac{(80)^2}{7}\right]\left[815 - \frac{(69)^2}{7}\right]}} = \frac{-164.571}{\sqrt{(233.714)(134.857)}} =$$

$$r = \frac{-164.571}{177.533} = \mathbf{-0.927}$$

3.42 $\Sigma x = 1{,}087$ $\Sigma x^2 = 322{,}345$ $\Sigma y = 2{,}032$

$\Sigma y^2 = 878{,}686$ $\Sigma xy = 507{,}509$ $n = 5$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{507{,}509 - \frac{(1{,}087)(2{,}032)}{5}}{\sqrt{\left[322{,}345 - \frac{(1{,}087)^2}{5}\right]\left[878{,}686 - \frac{(2{,}032)^2}{5}\right]}} =$$

$$r = \frac{65{,}752.2}{\sqrt{(86{,}031.2)(52{,}881.2)}} = \frac{65{,}752.2}{67{,}449.5} = \mathbf{.975}$$

3.43

Delta (x)	SW (y)
47.6	15.1
46.3	15.4
50.6	15.9
52.6	15.6
52.4	16.4
52.7	18.1

$\Sigma x = 302.2$ $\Sigma y = 96.5$ $\Sigma xy = 4,870.11$

$\Sigma x^2 = 15,259.62$ $\Sigma y^2 = 1,557.91$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{4,870.11 - \frac{(302.2)(96.5)}{6}}{\sqrt{\left[15,259.62 - \frac{(302.2)^2}{6}\right]\left[1,557.91 - \frac{(96.5)^2}{6}\right]}} = \mathbf{.6445}$$

3.44 $\Sigma x = 6,087$ $\Sigma x^2 = 6,796,149$

$\Sigma y = 1,050$ $\Sigma y^2 = 194,526$

$\Sigma xy = 1,130,483$ $n = 9$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{1{,}130{,}483 - \frac{(6{,}087)(1{,}050)}{9}}{\sqrt{\left[6{,}796{,}149 - \frac{(6{,}087)^2}{9}\right]\left[194{,}526 - \frac{(1{,}050)^2}{9}\right]}} =$$

$$r = \frac{420{,}333}{\sqrt{(2{,}679{,}308)(72{,}026)}} = \frac{420{,}333}{439{,}294.705} = \mathbf{.957}$$

3.45 Correlation between Year 1 and Year 2:

$\Sigma x = 17.09$ $\Sigma x^2 = 58.7911$

$\Sigma y = 15.12$ $\Sigma y^2 = 41.7054$

$\Sigma xy = 48.97$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{48.97 - \frac{(17.09)(15.12)}{8}}{\sqrt{\left[58.7911 - \frac{(17.09)^2}{8}\right]\left[41.7054 - \frac{(15.12)^2}{8}\right]}} =$$

$$r = \frac{16.6699}{\sqrt{(22.28259)(13.1286)}} = \frac{16.6699}{17.1038} = \mathbf{.975}$$

Correlation between Year 2 and Year 3:

$\Sigma x = 15.12$ $\Sigma x^2 = 41.7054$

$\Sigma y = 15.86$ $\Sigma y^2 = 42.0396$

$\Sigma xy = 41.5934$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{41.5934 - \frac{(15.12)(15.86)}{8}}{\sqrt{\left[41.7054 - \frac{(15.12)^2}{8}\right]\left[42.0396 - \frac{(15.86)^2}{8}\right]}} =$$

$$r = \frac{11.618}{\sqrt{(13.1286)(10.59715)}} = \frac{11.618}{11.795} = \mathbf{.985}$$

Correlation between Year 1 and Year 3:

$\Sigma x = 17.09$ $\Sigma x^2 = 58.7911$

$\Sigma y = 15.86$ $\Sigma y^2 = 42.0396$

$\Sigma xy = 48.5827$ $n = 8$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{48.5827 - \frac{(17.09)(15.86)}{8}}{\sqrt{\left[58.7911 - \frac{(17.09)^2}{8}\right]\left[42.0396 - \frac{(15.86)^2}{8}\right]}}$$

$$r = \frac{14.702}{\sqrt{(22.2826)(10.5972)}} = \frac{14.702}{15.367} = \mathbf{.957}$$

The years 2 and 3 are the most correlated with r = .985.

3.46 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**

3.47 P_{10}: $i = \frac{10}{100}(40) = 4$

P_{10} = 4.5th term = **23**

P_{80}: $i = \frac{80}{100}(40) = 32$

P_{80} = 32.5th term = **49.5**

$Q_1 = P_{25}$: $i = \frac{25}{100}(40) = 10$

P_{25} = 10.5th term = **27.5**

$Q_3 = P_{75}$: $i = \frac{75}{100}(40) = 30$

P_{75} = 30.5th term = **47.5**

IQR = $Q_3 - Q_1$ = 47.5 - 27.5 = **20**

Range = 81 - 19 = **62**

3.48 $\mu = \frac{\Sigma x}{N} = \frac{126,904}{20} =$ **6345.2**

The median is located at the $\frac{n+1}{2}$th value = 21/2 = 10.5th value

The median is the average of 5414 and 5563 = **5488.5**

P_{30}: i = (.30)(20) = 6

P_{30} is located at the average of the 6th and 7th terms

P_{30} = (4507+4541)/2 = **4524**

P_{60}: i = (.60)(20) = 12

P_{60} is located at the average of the 12th and 13th terms

P_{60} = (6101+6498)/2 = **6299.5**

P_{90}: i = (.90)(20) = 18

P_{90} is located at the average of the 18th and 19th terms

P_{90} = (9863+11,019)/2 = **10,441**

$Q_1 = P_{25}$: $i = (.25)(20) = 5$

Q_1 is located at the average of the 5th and 6th terms

$Q_1 = (4464+4507)/2 =$ **4485.5**

$Q_3 = P_{75}$: $i = (.75)(20) = 15$

Q_3 is located at the average of the 15th and 16th terms

$Q_3 = (6796+8687)/2 =$ **7741.5**

Range = 11,388 - 3619 = **7769**

IQR = $Q_3 - Q_1$ = 7741.5 - 4485.5 = **3256**

3.49 $n = 10$ $\Sigma x = 87.95$ $\Sigma x^2 = 1130.9027$

$\mu = (\Sigma x)/N = 87.95/10 =$ **8.795**

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{1130.9027 - \frac{(87.95)^2}{10}}{10}} = \mathbf{5.978}$$

3.50 a.) $\mu = \frac{\Sigma x}{N}$ = 26,675/11 = **2425**

Median = **1965**

b.) Range = 6300 - 1092 = **5208**

Q_3 = 2867 Q_1 = 1532 IQR = Q_3 - Q_1 = **1335**

c.) Variance:

$$\sigma^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N} = \frac{86,942,873 - \frac{(26,675)^2}{11}}{11} = \mathbf{2,023,272.55}$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{2,023,272.55} = \mathbf{1422.42}$$

d.) Texaco:

$$z = \frac{x - \mu}{\sigma} = \frac{1532 - 2425}{1422.42} = \mathbf{-0.63}$$

Exxon Mobil:

$z = \dfrac{x-\mu}{\sigma} = \dfrac{6300-2425}{1422.42} =$ **2.72**

e.) Skewness:

$S_k = \dfrac{3(\mu - M_d)}{\sigma} = \dfrac{3(2425-1965)}{1422.42} =$ **0.97**

3.51 a.) Mean: $\mu = \dfrac{\Sigma x}{n} = \dfrac{32.95}{14}$ = **2.3536**

Median: $\dfrac{1.79+2.07}{2}$ = **1.93**

Mode: **No Mode**

b.) Range: 4.73 – 1.20 **= 3.53**

Q_1: $\dfrac{1}{4}(14) = 3.5$ Located at the 4th term. Q_1 = 1.68

Q_3: $\dfrac{3}{4}(14) = 10.5$ Located at the 11th term. Q_3 = 2.87

IQR = $Q_3 - Q_1$ = 2.87 – 1.68 = **1.19**

x	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$
4.73	2.3764	5.6473
3.64	1.2864	1.6548
3.53	1.1764	1.3839
2.87	0.5164	0.2667
2.61	0.2564	0.0657
2.59	0.2364	0.0559
2.07	0.2836	0.0804
1.79	0.5636	0.3176
1.77	0.5836	0.3406
1.69	0.6636	0.4404
1.68	0.6736	0.4537
1.41	0.9436	0.8904
1.37	0.9836	0.9675
1.20	1.1536	1.3308

$\sum\lvert x-\bar{x}\rvert$ = 11.6972 $\sum(x-\bar{x})^2$ = 13.8957

MAD = $\dfrac{\sum\lvert x-\bar{x}\rvert}{n} = \dfrac{11.6972}{14}$ = **0.8355**

s^2 = $\dfrac{\sum(x-\bar{x})^2}{n-1} = \dfrac{13.8957}{13}$ = **1.0689**

s = $\sqrt{s^2} = \sqrt{1.0689}$ = **1.0339**

c.) Pearson's Coefficient of Skewness:

S_k = $\dfrac{3(\bar{x}-M_d)}{s} = \dfrac{3(2.3536-1.93)}{1.0339}$ = **1.229**

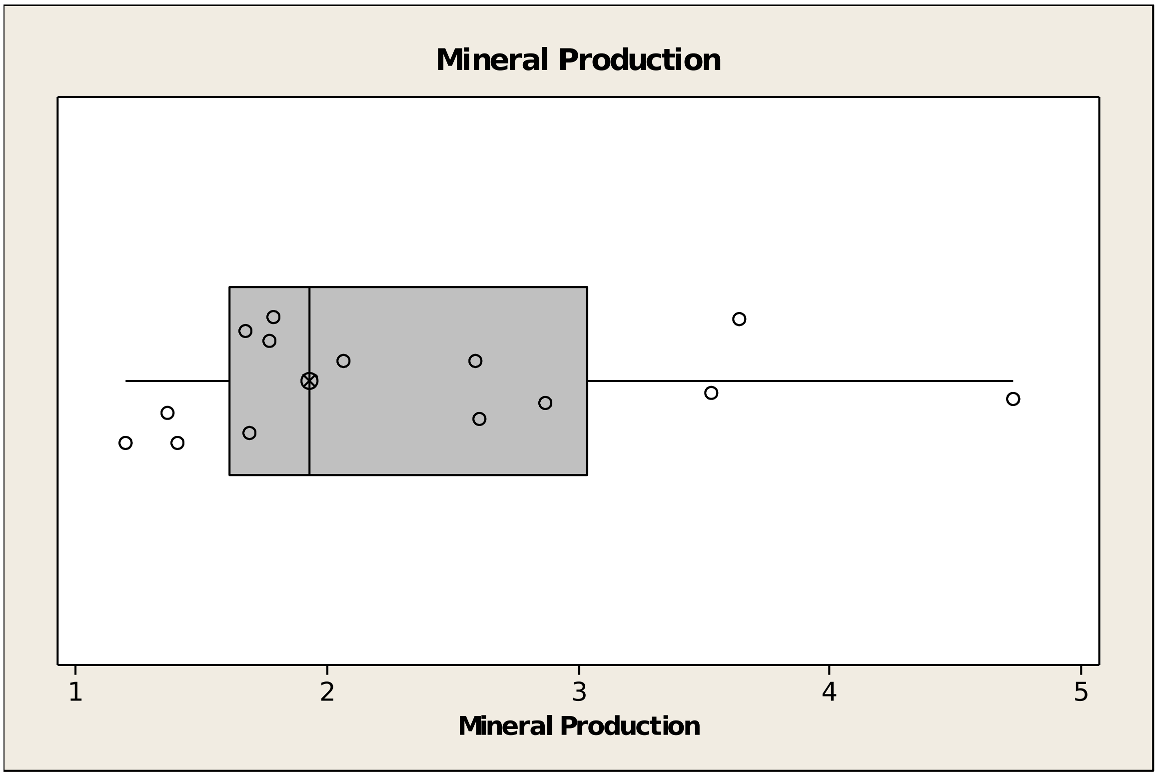
d.) Use Q_1 = 1.68, Q_2 = 1.93, Q_3 = 2.87, IQR = 1.19

Extreme Points: 1.20 and 4.73

Inner Fences: 1.68 – 1.5(1.19) = -0.105
1.93 + 1.5(1.19) = 3.715

Outer Fences: 1.68 + 3.0(1.19) = -1.890
1.93 + 3.0(1.19) = 5.500

There is one mild outlier. The 4.73 recorded for Arizona is outside the upper
inner fence.



3.52

Class	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:

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

Standard Deviation:

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

3.53

Class	f	M	fM	fM^2
0 - 20	32	10	320	3,200
20 - 40	16	30	480	14,400
40 - 60	13	50	650	32,500
60 - 80	10	70	700	49,000
80 - 100	19	90	1,710	153,900
	$\Sigma f = 90$		$\Sigma fM = 3,860$	$\Sigma fM^2 = 253,000$

a) Mean: $$\bar{x} = \frac{\Sigma fM}{n} = \frac{\Sigma fm}{\Sigma f} = \frac{3,860}{90}$$ = **42.89**

Mode: The Modal Class is 0-20. The midpoint of this class is the mode = **10**.

b) Sample Standard Deviation:

$$s = \sqrt{\frac{\Sigma fM^2 - \frac{(\Sigma fM)^2}{n}}{n-1}} = \sqrt{\frac{253,000 - \frac{(3860)^2}{90}}{89}} = \sqrt{\frac{253,000 - 165,551.1}{89}}$$

$$= \sqrt{\frac{87,448.9}{89}} = \sqrt{982.572}$$

= **31.346**

3.54 $\Sigma x = 36$ $\Sigma x^2 = 256$

$\Sigma y = 44$ $\Sigma y^2 = 300$

$\Sigma xy = 188$ $n = 7$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} = \frac{188 - \frac{(36)(44)}{7}}{\sqrt{\left[256 - \frac{(36)^2}{7}\right]\left[300 - \frac{(44)}{7}\right]}}$$

$$r = \frac{-38.2857}{\sqrt{(70.85714)(23.42857)}} = \frac{-38.2857}{40.7441} = \mathbf{-.940}$$

3.55 $CV_x = \frac{\sigma_x}{\mu_x}(100\%) = \frac{3.45}{32}(100\%) =$ **10.78%**

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

Stock X has a greater relative variability.

3.56 $\mu = 7.5$ Each of the numbers, 1 and 14, are 6.5 units away from the mean.

From the Empirical Rule: 99.7% of the values lie in $\mu \pm 3\sigma$

$3\sigma = 14 - 7.5 = 6.5$ Solving for $3\sigma = 6.5$ for σ: **σ = 2.167**

Suppose that $\mu = 7.5$, $\sigma = 1.7$:

95% lie within $\mu \pm 2\sigma = 7.5 \pm 2(1.7) = 7.5 \pm 3.4$

Between 4.1 and 10.9 lie 95% of the values.

3.57 $\mu = 419$, $\sigma = 27$

a.) 68%: $\mu \pm 1\sigma$ 419 $\pm$ 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 $x = 400$, $z = \frac{400-419}{27}$ = **-0.704**. This worker is in the lower half of

workers but within one standard deviation of the mean.

3.58 a.)

	x	x^2
Albania	4,900	24,010,000
Bulgaria	8,200	67,240,000
Croatia	11,200	125,440,000
Czech	16,800	282,240,000

Σx=41,100 Σx^2= 498,930,000

$$\mu = \frac{\Sigma x}{N} = \frac{41{,}100}{4} = \mathbf{10{,}275}$$

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{498{,}930{,}000 - \frac{(41{,}100)^2}{4}}{4}} = \mathbf{4376.86}$$

b.)

	x	x^2
Hungary	14,900	222,010,000
Poland	12,000	144,000,000
Romania	7,700	59,290,000
Bosnia/Herz	6,500	42,250,000

Σx=41,100 Σx^2=467,550,000

$$\mu = \frac{\Sigma x}{N} = \frac{41{,}100}{4} = \mathbf{10{,}275}$$

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}} = \sqrt{\frac{467{,}550{,}000 - \frac{(41{,}100)^2}{4}}{4}} = \mathbf{3363.31}$$

c.)

$$CV_1 = \frac{\sigma_1}{\mu_1}(100) = \frac{4376.86}{10,275}(100) = \mathbf{42.60\%}$$

$$CV_2 = \frac{\sigma_2}{\mu_2}(100) = \frac{3363.31}{10,275}(100) = \mathbf{32.73\%}$$

The first group has a larger coefficient of variation

3.59 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.

3.60 Σx = 36.62 Σx^2 = 217.137

Σy = 57.23 Σy^2 = 479.3231

Σxy = 314.9091 n = 8

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} =$$

$$r = \frac{314.9091 - \frac{(36.62)(57.23)}{8}}{\sqrt{\left[217.137 - \frac{(36.62)^2}{8}\right]\left[479.3231 - \frac{(57.23)^2}{8}\right]}} =$$

$$r = \frac{52.938775}{\sqrt{(49.50895)(69.91399)}} = \mathbf{.90}$$

There is a strong positive relationship between the inflation rate and the thirty-year treasury yield.

3.61 a.) $Q_1 = P_{25}$: $i = \frac{25}{100}(20) = 5$

Q_1 = 5.5th term = (48.3 + 49.9)/2 = 49.1

$Q_3 = P_{75}$: $i = \frac{75}{100}(20) = 15$

Q_3 = 15.5th term = (77.6+83.8)/2 = 80.7

Median: $\frac{n+1}{2}^{th} = \frac{20+1}{2}^{th}$ = 10.5th term

Median = (55.9 + 61.3)/2 = 58.6

IQR = Q_3 - Q_1 = 80.7 – 49.1 = 31.6

1.5 IQR = 47.4; 3.0 IQR = 94.8

Inner Fences:

Q_1 - 1.5 IQR = 49.1 – 47.4 = 1.7

Q_3 + 1.5 IQR = 80.7 + 47.4 = 128.1

Outer Fences:

Q_1 - 3.0 IQR = 49.1 – 94.8 = - 45.70

Q_3 + 3.0 IQR = 80.7 + 94.8 = 175.5

b.) and c.) There are no outliers in the lower end. There are two extreme

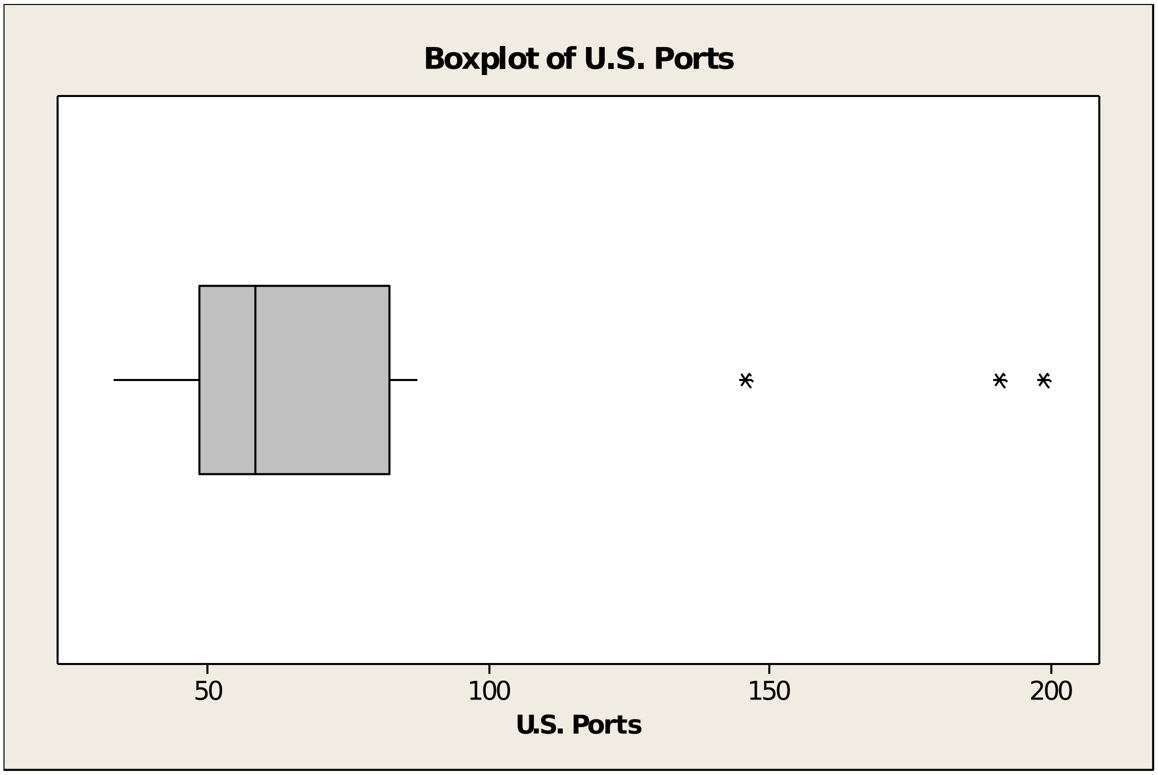
outliers in the upper end (South Louisiana, 198.8, and Houston,

190.9). There is one mild outlier at the upper end (New York, 145.9).

Since the median is nearer to Q_1, the distribution is positively skewed.

d.) There are three dominating, large ports

Displayed below is the MINITAB boxplot for this problem.



3.62 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**

Chapter 4

Probability

LEARNING OBJECTIVES

The main objective of Chapter 4 is to help you understand the basic principles of probability, specifically enabling you to

1. Comprehend the different ways of assigning probability.
2. Understand and apply marginal, union, joint, and conditional probabilities.
3. Select the appropriate law of probability to use in solving problems.

4. Solve problems using the laws of probability, including the law of addition, the law of multiplication , and the law of conditional probability.

5. Revise probabilities using Bayes' rule.

CHAPTER TEACHING STRATEGY

Students can be motivated to study probability by realizing that the field of probability has some stand-alone application in their lives in such applied areas human resource analysis, actuarial science, and gaming. In addition, students should understand that much of the rest of this course is based on probabilities even though they will not be directly applying many of these formulas in other chapters.

This chapter is frustrating for the learner because probability problems can be approached by using several different techniques. Whereas, in many chapters of this text, students will approach problems by using one standard technique, in chapter 4, different students will often use different approaches to the same problem. The text attempts to emphasize this point and underscore it by presenting several different ways to solve probability problems. The probability rules and laws presented in the chapter can virtually always be used in solving probability problems. However, it is sometimes easier to construct a probability matrix or a tree diagram or use the sample space to solve the problem. If the student is aware that what they have at their hands is an array of tools or techniques, they will be less overwhelmed in approaching a probability problem. An attempt has been made to differentiate the several types of probabilities so that students can sort out the various types of problems.

In teaching students how to construct a probability matrix, emphasize that it is usually best to place only one variable along each of the two dimensions of the matrix. (That is, place Mastercard with yes/no on one axis and Visa with yes/no on the other instead of trying to place Mastercard and Visa along the same axis).

This particular chapter is very amenable to the use of visual aids. Students enjoy rolling dice, tossing coins, and drawing cards as a part of the class experience. Of all the chapters in the book, it is most imperative that students work a lot of problems in this chapter. Probability problems are so varied and individualized that a significant portion of the learning comes in the doing. Experience is an important factor in working probability problems.

Section 4.8 on Bayes' theorem can be skipped in a one-semester course without losing any continuity. This section is a prerequisite to the chapter 19 presentation of "revising probabilities in light of sample information (section 19.4).

CHAPTER OUTLINE

4.1 Introduction to Probability

4.2 Methods of Assigning Probabilities
- Classical Method of Assigning Probabilities
- Relative Frequency of Occurrence
- Subjective Probability

4.3 Structure of Probability
- Experiment
- Event
- Elementary Events
- Sample Space
- Unions and Intersections
- Mutually Exclusive Events
- Independent Events
- Collectively Exhaustive Events
- Complimentary Events
- Counting the Possibilities
  - The *mn* Counting Rule
  - Sampling from a Population with Replacement
  - Combinations: Sampling from a Population Without Replacement

4.4 Marginal, Union, Joint, and Conditional Probabilities

4.5 Addition Laws

Probability Matrices

Complement of a Union

Special Law of Addition

4.6 Multiplication Laws

General Law of Multiplication

Special Law of Multiplication

4.7 Conditional Probability

Independent Events

4.8 Revision of Probabilities: Bayes' Rule

KEY TERMS

A Priori

Bayes' Rule

Classical Method of Assigning Probabilities

Collectively Exhaustive Events

Combinations

Complement of a Union

Intersection

Joint Probability

Marginal Probability

mn Counting Rule

Mutually Exclusive Events

Probability Matrix

Complement

Conditional Probability

Elementary Events

Event

Experiment

Independent Events

Relative Frequency of Occurrence

Sample Space

Set Notation

Subjective Probability

Union

Union Probability

SOLUTIONS TO PROBLEMS IN CHAPTER 4

4.1 Enumeration of the six parts: D_1, D_2, D_3, A_4, A_5, A_6

D = Defective part

A = Acceptable part

Sample Space:

$D_1 D_2$, $D_2 D_3$, $D_3 A_5$

$D_1 D_3$, $D_2 A_4$, $D_3 A_6$

$D_1 A_4$, $D_2 A_5$, $A_4 A_5$

$D_1 A_5$, $D_2 A_6$, $A_4 A_6$

$D_1 A_6$, $D_3 A_4$, $A_5 A_6$

There are **15** members of the sample space

The probability of selecting exactly one defect out of two is:

9/15 = **.60**

4.2 X = {1, 3, 5, 7, 8, 9}, Y = {2, 4, 7, 9}, and Z = {1, 2, 3, 4, 7,}

a) $X \cup Z =$ **{1, 2, 3, 4, 5, 7, 8, 9}**

b) $X \cap Y =$ **{7, 9}**

c) $X \cap Z =$ **{1, 3, 7}**

d) $X \cup Y \cup Z =$ **{1, 2, 3, 4, 5, 7, 8, 9}**

e) $X \cap Y \cap Z =$ **{7}**

f) $(X \cup Y) \cap Z$ = {1, 2, 3, 4, 5, 7, 8, 9} $\cap$ {1, 2, 3, 4, 7} = **{1, 2, 3, 4, 7}**

g) $(Y \cap Z) \cup (X \cap Y)$ = {2, 4, 7} $\cup$ {7, 9} = **{2, 4, 7, 9}**

h) X or Y = $X \cup Y$ = **{1, 2, 3, 4, 5, 7, 8, 9}**

i) Y and Z = $Y \cap Z$ = {**2, 4, 7**}

4.3 If A = {2, 6, 12, 24} and the population is the positive even numbers through 30,

A’ = **{4, 8, 10, 14, 16, 18, 20, 22, 26, 28, 30}**

4.4 6(4)(3)(3) = **216**

4.5 Enumeration of the six parts: D_1, D_2, A_1, A_2, A_3, A_4

D = Defective part

A = Acceptable part

Sample Space:

$D_1\ D_2\ A_1$, $D_1\ D_2\ A_2$, $D_1\ D_2\ A_3$,

$D_1\ D_2\ A_4$, $D_1\ A_1\ A_2$, $D_1\ A_1\ A_3$,

$D_1\ A_1\ A_4$, $D_1\ A_2\ A_3$, $D_1\ A_2\ A_4$,

$D_1\ A_3\ A_4$, $D_2\ A_1\ A_2$, $D_2\ A_1\ A_3$,

$D_2\ A_1\ A_4$, $D_2\ A_2\ A_3$, $D_2\ A_2\ A_4$,

$D_2\ A_3\ A_4$, $A_1\ A_2\ A_3$, $A_1\ A_2\ A_4$,

$A_1\ A_3\ A_4$, $A_2\ A_3\ A_4$

Combinations are used to counting the sample space because sampling is done without replacement.

$${}_6C_3 = \frac{6!}{3!3!} = \mathbf{20}$$

Probability that one of three is defective is:

12/20 = 3/5 **.60**

There are 20 members of the sample space and 12 of them have exactly

1 defective part.

4.6 10^7 = 10,000,000 different numbers

4.7 $_{20}C_6 = \dfrac{20!}{6!14!} =$ **38,760**

It is assumed here that 6 different (without replacement) employees are to be selected.

4.8 $P(A) = .10$, $P(B) = .12$, $P(C) = .21$, $P(A \cap C) = .05$ $P(B \cap C) = .03$

a) $P(A \cup C) = P(A) + P(C) - P(A \cap C) = .10 + .21 - .05 =$ **.26**

b) $P(B \cup C) = P(B) + P(C) - P(B \cap C) = .12 + .21 - .03 =$ **.30**

c) If A, B mutually exclusive, $P(A \cup B) = P(A) + P(B) = .10 + .12 =$ **.22**

4.9

	D	E	F	
A	5	8	12	25
B	10	6	4	20
C	8	2	5	15
	23	16	21	60

a) $P(A \cup D) = P(A) + P(D) - P(A \cap D) =$ 25/60 + 23/60 - 5/60 = 43/60 = **.7167**

b) $P(E \cup B) = P(E) + P(B) - P(E \cap B) =$ 16/60 + 20/60 - 6/60 = 30/60 = **.5000**

c) $P(D \cup E) = P(D) + P(E) =$ 23/60 + 16/60 = 39/60 = **.6500**

d) *P*(C ㅍ F) = *P*(C) + *P*(F) - *P*(C ㅎ F) = 15/60 + 21/60 - 5/60 = 31/60 = **.5167**

4.10

	E	F	
A	.10	.03	.13
B	.04	.12	.16
C	.27	.06	.33
D	.31	.07	.38
	.72	.28	1.00

a) $P(A \cup F) = P(A) + P(F) - P(A \cap F)$ = .13 + .28 - .03 = **.38**

b) $P(E \cup B) = P(E) + P(B) - P(E \cap B)$ = .72 + .16 - .04 = **.84**

c) $P(B \cup C) = P(B) + P(C)$ =.16 + .33 = **.49**

d) $P(E \cup F) = P(E) + P(F)$ = .72 + .28 = **1.000**

4.11 A = event of having flown in an airplane at least once

T = event of having ridden in a train at least once

$P(A) = .47$ $P(T) = .28$

P (ridden either a train or an airplane) =

$P(A \cup T) = P(A) + P(T) - P(A \cap T) = .47 + .28 - P(A \cap T)$

Cannot solve this problem without knowing the probability of the intersection.

We need to know the probability of the intersection of A and T, the proportion

who have ridden both or determine if these two events are mutually exclusive.

4.12 $P(L) = .75$ $P(M) = .78$ $P(M \cap L) = .61$

a) $P(M \cup L) = P(M) + P(L) - P(M \cap L) = .78 + .75 - .61 =$ **.92**

b) $P(M \cup L)$ but not both $= P(M \cup L) - P(M \cap L) = .92 - .61 =$ **.31**

c) $P(NM \cap NL) = 1 - P(M \cup L) = 1 - .92 =$ **.08**

Note: the neither/nor event is solved for here by taking the complement of the union.

4.13 Let C = have cable TV

Let T = have 2 or more TV sets

$P(\text{C}) = .67$, $P(\text{T}) = .74$, $P(\text{C} \cap \text{T}) = .55$

a) $P(\text{C} \cup \text{T}) = P(\text{C}) + P(\text{T}) - P(\text{C} \cap \text{T}) = .67 + .74 - .55 =$ **.86**
b) $P(\text{C} \cup \text{T}$ but not both$) = P(\text{C} \cup \text{T}) - P(\text{C} \cap \text{T}) = .86 - .55 =$ **.31**
c) $P(\text{NC} \cap \text{NT}) = 1 - P(\text{C} \cup \text{T}) = 1 - .86 =$ **.14**
d) The special law of addition does not apply because $P(\text{C} \cap \text{T})$ is not .0000. Possession of cable TV and 2 or more TV sets are not mutually exclusive.

4.14 Let T = review transcript

F = consider faculty references

$P(\text{T}) = .54$ $P(\text{F}) = .44$ $P(\text{T} \cap \text{F}) = .35$

a) $P(\text{F} \cup \text{T}) = P(\text{F}) + P(\text{T}) - P(\text{F} \cap \text{T}) = .44 + .54 - .35 =$ **.63**

b) $P(\text{F} \cup \text{T}) - P(\text{F} \cap \text{T}) = .63 - .35 =$ **.28**

c) $1 - P(\text{F} \cup \text{T}) = 1 - .63 =$ **.37**

d)

		Faculty References Y	N	
Transcript	Y	.35	.19	.54
	N	.09	.37	.46
		.44	.56	1.00

4.15

	C	D	E	F	
A	5	11	16	8	40
B	2	3	5	7	17
	7	14	21	15	57

a) $P(A \cap E)$ = 16/57 = **.2807**

b) $P(D \cap B)$ = 3/57 = **.0526**

c) $P(D \cap E)$ = **.0000**

d) $P(A \cap B)$ = **.0000**

4.16

	D	E	F	
A	.12	.13	.08	.33
B	.18	.09	.04	.31
C	.06	.24	.06	.36
	.36	.46	.18	1.00

a) $P(E \cap B)$ = **.09**

b) $P(C \cap F)$ = **.06**

c) $P(E \cap D)$ = **.00**

4.17 Let D = Defective part

a) (without replacement)

$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2 | D_1) = \frac{6}{50} \cdot \frac{5}{49} = \frac{30}{2450}$ = **.0122**

b) (with replacement)

$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2) = \frac{6}{50} \cdot \frac{6}{50} = \frac{36}{2500}$ = **.0144**

4.18 Let U = Urban

I = care for Ill relatives

$P(U) = .78$ $P(I) = .15$ $P(I|U) = .11$

a) $P(U \cap I) = P(U) \cdot P(I|U)$

$P(U \cap I) = (.78)(.11) =$ **.0858**

b) $P(U \cap NI) = P(U) \cdot P(NI|U)$ but $P(I|U) = .11$

So, $P(NI|U) = 1 - .11 = .89$ and

$P(U \cap NI) = P(U) \cdot P(NI|U) = (.78)(.89) =$ **.6942**

c)

		U: Yes	U: No	
I	Yes			.15
	No			.85

.78	.22

The answer to a) is found in the YES-YES cell. To compute this cell, take 11%

or .11 of the total (.78) people in urban areas. (.11)(.78) = .0858 which belongs in

the “YES-YES" cell. The answer to b) is found in the Yes for U and no for I cell.

It can be determined by taking the marginal, .78, less the answer for a), .0858.

d. $P(NU \cap I)$ is found in the no for U column and the yes for I row (1^{st} row and 2^{nd} column). Take the marginal, .15, minus the yes-yes cell, .0858, to get **.0642**.

4.19 Let S = stockholder Let C = college

$P(S) = .43$ $P(C) = .37$ $P(C|S) = .75$

a) $P(NS) = 1 - .43 =$ **.57**

b) $P(S \cap C) = P(S) \cdot P(C|S) = (.43)(.75) =$ **.3225**

c) $P(S \cup C) = P(S) + P(C) - P(S \cap C) = .43 + .37 - .3225 =$ **.4775**

d) $P(NS \cap NC) = 1 - P(S \cup C) = 1 - .4775 =$ **.5225**

e) $P(NS \cup NC) = P(NS) + P(NC) - P(NS \cap NC) = .57 + .63 - .5225 =$ **.6775**

f) $P(C \cap NS) = P(C) - P(C \cap S) = .37 - .3225 =$ **.0475**

The matrix:

		C		
		Yes	No	
	Yes	.3225	.1075	.43
S	No	.0475	.5225	.57
		.37	.63	1.00

4.20 Let F = fax machine Let P = personal computer

Given: $P(\text{F}) = .10$ $P(\text{P}) = .52$ $P(\text{P}|\text{F}) = .91$

a) $P(\text{F} \cap \text{P}) = P(\text{F}) \cdot P(\text{P}|\text{F}) = (.10)(.91) =$ **.091**

b) $P(\text{F} \cup \text{P}) = P(\text{F}) + P(\text{P}) - P(\text{F} \cap \text{P}) = .10 + .52 - .091 =$ **.529**

c) $P(\text{F} \cap \text{NP}) = P(\text{F}) \cdot P(\text{NP}|\text{F})$

Since $P(\text{P}|\text{F}) = .91$, $P(\text{NP}|\text{F}) = 1 - P(\text{P}|\text{F}) = 1 - .91 = .09$

$P(\text{F} \cap \text{NP}) = (.10)(.09) =$ **.009**

d) $P(\text{NF} \cap \text{NP}) = 1 - P(\text{F} \cup \text{P}) = 1 - .529 =$ **.471**

e) $P(\text{NF} \cap \text{P}) = P(\text{P}) - P(\text{F} \cap \text{P}) = .52 - .091 =$ **.429**

The matrix:

	P	NP	
F	.091	.009	.10
NF	.429	.471	.90
	.520	.480	1.00

4.21

Let S = safety Let A = age

P(S) = .30 *P*(A) = .39 *P*(A|S) = .87

a) *P*(S ∩ NA) = *P*(S)· *P*(NA|S)

but *P*(NA|S) = 1 - *P*(A|S) = 1 - .87 = .13

P(S ∩ NA) = (.30)(.13) = **.039**

b) *P*(NS ∩ NA) = 1 - *P*(S ∪ A) = 1 - [*P*(S) + *P*(A) - *P*(S ∩ A)]

but *P*(S ∩ A) = *P*(S) · *P*(A|S) = (.30)(.87) = .261

P(NS ∩ NA) = 1 - (.30 + .39 - .261) = **.571**

c) *P*(NS ∩ A) = *P*(NS) - *P*(NS ∩ NA)

but *P*(NS) = 1 - P(S) = 1 - .30 = .70

P(NS ∩ A) = .70 - 571 = **.129**

The matrix:

		A		
		Yes	No	
	Yes	.261	.039	.30
S	No	.129	.571	.70
		.39	.61	1.00

4.22 Let C = ceiling fans Let O = outdoor grill

$P(\text{C}) = .60$ $P(\text{O}) = .29$ $P(\text{C} \cap \text{O}) = .13$

a) $P(\text{C} \cup \text{O}) = P(\text{C}) + P(\text{O}) - P(\text{C} \cap \text{O}) = .60 + .29 - .13 =$ **.76**

b) $P(\text{NC} \cap \text{NO}) = 1 - P(\text{C} \cup \text{O}) = 1 - .76 =$ **.24**

c) $P(\text{NC} \cap \text{O}) = P(\text{O}) - P(\text{C} \cap \text{O}) = .29 - .13 =$ **.16**

d) $P(\text{C} \cap \text{NO}) = P(\text{C}) - P(\text{C} \cap \text{O}) = .60 - .13 =$ **.47**

The matrix:

		O		
		Yes	No	
C	Yes	.13	.47	.60
	No	.16	.24	.40
		.29	.71	1.00

4.23

	E	F	G	
A	15	12	8	35
B	11	17	19	47
C	21	32	27	80
D	18	13	12	43
	65	74	66	205

a) $P(G|A) = 8/35 =$ **.2286**

b) $P(B|F) = 17/74 =$ **.2297**

c) $P(C|E) = 21/65 =$ **.3231**

d) $P(E|G) =$ **.0000**

4.24

	C	D	
A	.36	.44	.80
B	.11	.09	.20
	.47	.53	1.00

a) $P(C|A) = .36/.80 =$ **.4500**

b) $P(B|D) = .09/.53 =$ **.1698**

c) $P(A|B) =$ **.0000**

4.25

		Calculator		
		Yes	No	
Computer	Yes	46	3	49
	No	11	15	26
		57	18	75

Select a category from each variable and test

$P(V_1 | V_2) = P(V_1)$.

For example, P(Yes Computer | Yes Calculator) = P(Yes Computer)?

$$\frac{46}{57} = \frac{49}{75}$$?

$.8070 \neq .6533$

Since this is one example that the conditional does not equal the marginal in

is matrix, the variable, computer, is not independent of the variable,

calculator.

4.26 Let C = construction Let S = South Atlantic

83,384 total failures

10,867 failures in construction

8,010 failures in South Atlantic

1,258 failures in construction and South Atlantic

a) *P*(S) = 8,010/83,384 = **.09606**

b) $P(C \cup S) = P(C) + P(S) - P(C \cap S) =$

10,867/83,384 + 8,010/83,384 - 1,258/83,384 = 17,619/83,384 = **.2113**

c) $P(C|S) = \dfrac{P(C \cap S)}{P(S)} = \dfrac{\frac{1258}{83,384}}{\frac{8010}{83,384}}$ = **.15705**

d) $P(S|C) = \dfrac{P(C \cap S)}{P(C)} = \dfrac{\frac{1258}{83,384}}{\frac{10,867}{83,384}}$ = **.11576**

e) $P(NS|NC) = \dfrac{P(NS \cap NC)}{P(NC)} = \dfrac{1 - P(C \cup S)}{P(NC)}$

but NC = 83,384 - 10,867 = 72,517

and P(NC) = 72,517/83,384 = .869675

Therefore, P(NS | NC) = (1 - .2113)/(.869675) = **.9069**

f) P(NS | C) = $$\frac{P(NS \cap C)}{P(C)} = \frac{P(C) - P(C \cap S)}{P(C)}$$

but P(C) = 10,867/83,384 = .1303

P(C ∩ S) = 1,258/83,384 = .0151

Therefore, P(NS | C) = (.1303 - .0151)/.1303 = **.8842**

4.27 Let E = Economy   Let Q = Qualified

$P(E) = .46$ $P(Q) = .37$ $P(E \cap Q) = .15$

a) $P(E|Q) = P(E \cap Q)/P(Q) = .15/.37 =$ **.4054**

b) $P(Q|E) = P(E \cap Q)/P(E) = .15/.46 =$ **.3261**

c) $P(Q|NE) = P(Q \cap NE)/P(NE)$

but $P(Q \cap NE) = P(Q) - P(Q \cap E) = .37 - .15 = .22$

$P(NE) = 1 - P(E) = 1 - .46 = .54$

$P(Q|NE) = .22/.54 =$ **.4074**

d) $P(NE \cap NQ) = 1 - P(E \cup Q) = 1 - [P(E) + P(Q) - P(E \cap Q)]$

$= 1 - [.46 + .37 - .15] = 1 - (.68) =$ **.32**

The matrix:

		Q: Yes	Q: No	
E	Yes	.15	.31	.46
	No	.22	.32	.54
		.37	.63	1.00

4.28 Let EM = email while on phone Let TD = “to-do” lists during meetings

P(EM) = .54 P(TD|EM) = .20

a) P(EM $\cap$ TD) = P(EM) · P(TD|EM) = (.54)(.20) = **.1080**

b) P(not TD|EM) = 1 - P(TD|EM) = 1 - .20 = **.80**

c) P(not TD $\cap$ EM) = P(EM) - P(EM $\cap$ TD) = .54 - .1080 = **.4320**

Could have been worked as: P(not TD $\cap$ EM) = P(EM)· P(not TD|EM) =

(.54)(.80) = .4320

The matrix:

		TD		
		Yes	No	
EM	Yes	.1080	.4320	.54
	No			.46
				1.00

4.29 Let H = hardware Let S = software

P(H) = .37 P(S) = .54 P(S|H) = .97

a) P(NS|H) = 1 - P(S|H) = 1 - .97 = **.03**

b) P(S|NH) = P(S ∩ NH)/P(NH)

but P(H ∩ S) = P(H) · P(S|H) = (.37)(.97) = .3589

so P(NH ∩ S) = P(S) - P(H ∩ S) = .54 - .3589 = .1811

P(NH) = 1 - P(H) = 1 - .37 = .63

P(S|NH) = (.1811)/(.63) = **.2875**

c) P(NH|S) = P(NH ∩ S)/P(S) = .1811//54 = **.3354**

d) P(NH|NS) = P(NH ∩ NS)/P(NS)

but P(NH ∩ NS) = P(NH) - P(NH ∩ S) = .63 - .1811 = .4489

and P(NS) = 1 - P(S) = 1 - .54 = .46

P(NH|NS) = .4489/.46 = **.9759**

The matrix:

	S		
	Yes	No	
Yes	.3589	.0111	.37

H	No	.1811	.4489	.63
		.54	.46	1.00

4.30 Let R = agreed or strongly agreed that lack of role models was a barrier

Let S = agreed or strongly agreed that gender-based stereotypes was a barrier

P(R) = .43 P(S) = .46 P(R|S) = .77 P(not S) = .54

a.) P(not R|S) = 1 - P(R|S) = 1 - .77 = **.23**

b.) P(not S|R) = P(not S $\cap$ R)/P(R)

but P(S $\cap$ R) = P(S) · P(R|S) = (.46)(.77) = .3542

so P(not S $\cap$ R) = P(R) - P(S $\cap$ R) = .43 - .3542 = .0758

Therefore, P(not S|R) = (.0758)/(.43) = **.1763**

c.) P(not R|not S) = P(not R $\cap$ not S)/P(not S)

but P(not R $\cap$ not S) = P(not S) - P(not S $\cap$ R) = .54 - .0758 = .4642

P(not R|not S) = .4642/.54 = **.8596**

The matrix:

		S		
		Yes	No	
R	Yes	.3542	.0758	.43
	No	.1058	.4642	.57

.46	.54	1.00

4.31 Let A = product produced on Machine A

B = product produces on Machine B

C = product produced on Machine C

D = defective product

P(A) = .10 P(B) = .40 P(C) = .50

$P(D|A)$ = .05 $P(D|B)$ = .12 $P(D|C)$ = .08

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(D\|E_i)$	$P(D \cap E_i)$	
A	.10	.05	.005	.005/.093=**.0538**
B	.40	.12	.048	.048/.093=**.5161**
C	.50	.08	.040	.040/.093=**.4301**
			P(D)=.093	

Revise: $P(A|D)$ = .005/.093 **= .0538**

$P(B|D)$ = .048/.093 = **.5161**

$P(C|D)$ = .040/.093 = **.4301**

4.32 Let A = Alex fills the order

B = Alicia fills the order

C = Juan fills the order

I = order filled incorrectly

K = order filled correctly

$P(\text{A}) = .30$ $P(\text{B}) = .45$ $P(\text{C}) = .25$

$P(\text{I}|\text{A}) = .20$ $P(\text{I}|\text{B}) = .12$ $P(\text{I}|\text{C}) = .05$

$P(\text{K}|\text{A}) = .80$ $P(\text{K}|\text{B}) = .88$ $P(\text{K}|\text{C}) = .95$

a) $P(\text{B}) = .45$

b) $P(\text{K}|\text{C}) = 1 - P(\text{I}|\text{C}) = 1 - .05 = .95$

c)

Event	Prior	Conditional	Joint	Revised
	$P(\text{E}_i)$	$P(\text{I}\|\text{E}_i)$	$P(\text{I} \cap \text{E}_i)$	$P(\text{E}_i\|\text{I})$
A	.30	.20	.0600	.0600/.1265=**.4743**
B	.45	.12	.0540	.0540/.1265=**.4269**
C	.25	.05	.0125	.0125/.1265=**.0988**
			$P(\text{I})$=.1265	

Revised: $P(\text{A}|\text{I}) = .0600/.1265 =$ **.4743**

$P(B|I) = .0540/.1265 =$ **.4269**

$P(C|I) = .0125/.1265 =$ **.0988**

d)

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(K\|E_i)$	$P(K \cap E_i)$	$P(E_i\|K)$
A	.30	.80	.2400	.2400/.8735=**.2748**
B	.45	.88	.3960	.3960/.8735=**.4533**
C	.25	.95	.2375	.2375/.8735=**.2719**
			$P(K)$=.8735	

4.33 Let T = lawn treated by Tri-state

G = lawn treated by Green Chem

V = very healthy lawn

N = not very healthy lawn

$P(T) = .72$ $P(G) = .28$ $P(V|T) = .30$ $P(V|G) = .20$

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(V\|E_i)$	$P(V \cap E_i)$	$P(E_i\|V)$
A	.72	.30	.216	.216/.272=**.7941**
B	.28	.20	.056	.056/.272=**.2059**
			$P(V)$=.272	

Revised: $P(T|V) = .216/.272 =$ **.7941**

$P(G|V) = .056/.272 =$ **.2059**

4.34 Let S = small Let L = large

The prior probabilities are: ***P*(S) = .70** ***P*(L) = .30**

$P(T|S) = .18$ $P(T|L) = .82$

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(T\|E_i)$	$P(T \cap E_i)$	$P(E_i\|T)$
S	.70	.18	.1260	.1260/.3720 = **.3387**
L	.30	.82	.2460	.2460/.3720 = **.6613**
			P(T)=**.3720**	

Revised: $P(S|T)$ = .1260/.3720 = **.3387**

$P(L|T)$ = .2460/.3720 = **.6613**

37.2% offer training since P(T) = .3720,.

4.35

		Variable 1		
		D	E	
	A	10	20	30
Variable 2	B	15	5	20

C	30	15	45
	55	40	95

a) $P(E) = 40/95 =$ **.42105**

b) $P(B \cup D) = P(B) + P(D) - P(B \cap D)$

$= 20/95 + 55/95 - 15/95 = 60/95 =$ **.63158**

c) $P(A \cap E) = 20/95 =$ **.21053**

d) $P(B|E) = 5/40 =$ **.1250**

e) $P(A \cup B) = P(A) + P(B) = 30/95 + 20/95 =$

$50/95 =$ **.52632**

f) $P(B \cap C) = .0000$ (mutually exclusive)

g) $P(D|C) = 30/45 =$ **.66667**

h) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{.0000}{20/95} =$ **.0000 (A and B are mutually exclusive)**

i) $P(A) = P(A|D)$??

Does $30/95 = 10/95$??

Since, $.31579 \neq .18182$, Variables 1 and 2 are <u>not</u> independent.

4.36

	D	E	F	G	
A	3	9	7	12	31
B	8	4	6	4	22
C	10	5	3	7	25
	21	18	16	23	78

a) $P(F \cap A)$ = 7/78 = **.08974**

b) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{.0000}{22/78}$ = **.0000 (A and B are mutually exclusive)**

c) $P(B)$ = 22/78 = **.28205**

d) $P(E \cap F)$ = **.0000 Mutually Exclusive**

e) $P(D|B)$ = 8/22 = **.36364**

f) $P(B|D)$ = 8/21 = **.38095**

g) $P(D \cup C)$ = 21/78 + 25/78 – 10/78 = 36/78 = **.4615**

h) $P(F)$ = 16/78 = **.20513**

4.37

		Age(years)					
		<35	35-44	45-54	55-64	>65	
	Male	.11	.20	.19	.12	.16	.78
Gender	Female	.07	.08	.04	.02	.01	.22
		.18	.28	.23	.14	.17	1.00

a) *P*(35-44) = **.28**

b) *P*(Woman ㅎ 45-54) = **.04**

c) *P*(Man ㅍ 35-44) = *P*(Man) + *P*(35-44) - *P*(Man ㅎ 35-44) = .78 + .28 - .20 = **.86**

d) *P*(<35 ㅍ 55-64) = *P*(<35) + *P*(55-64) = .18 + .14 = **.32**

e) *P*(Woman | 45-54) = *P*(Woman ㅎ 45-54)/*P*(45-54) = .04/.23= **.1739**

f) *P*(not W ㅎ not 55-64) = .11 + .20 + .19 + .16 = **.66**

4.38 Let T = thoroughness Let K = knowledge

$P(\text{T}) = .78$ $P(\text{K}) = .40$ $P(\text{T} \cap \text{K}) = .27$

a) $P(\text{T} \cup \text{K}) = P(\text{T}) + P(\text{K}) - P(\text{T} \cap \text{K}) =$

$.78 + .40 - .27 =$ **.91**

b) $P(\text{NT} \cap \text{NK}) = 1 - P(\text{T} \cup \text{K}) = 1 - .91 =$ **.09**

c) $P(\text{K}|\text{T}) = P(\text{K} \cap \text{T})/P(\text{T}) = .27/.78 =$ **.3462**

d) $P(\text{NT} \cap \text{K}) = P(\text{NT}) - P(\text{NT} \cap \text{NK})$

but $P(\text{NT}) = 1 - P(\text{T}) = .22$

$P(\text{NT} \cap \text{K}) = .22 - .09 =$ **.13**

The matrix:

		K: Yes	K: No	
T	Yes	.27	.51	.78
	No	.13	.09	.22
		.40	.60	1.00

4.39 Let R = retirement Let L = life insurance

$P(R) = .42$ $P(L) = .61$ $P(R \cap L) = .33$

a) $P(R|L) = P(R \cap L)/P(L) = .33/.61 =$ **.5410**

b) $P(L|R) = P(R \cap L)/P(R) = .33/.42 =$ **.7857**

c) $P(L \cup R) = P(L) + P(R) - P(L \cap R) = .61 + .42 - .33 =$ **.70**

d) $P(R \cap NL) = P(R) - P(R \cap L) = .42 - .33 =$.**09**

e) $P(NL|R) = P(NL \cap R)/P(R) = .09/.42 =$.**2143**

The matrix:

		L		
		Yes	No	
R	Yes	.33	.09	.42
	No	.28	.30	.58
		.61	.39	1.00

4.40 $P(T) = .16 \quad P(T|W) = .20 \quad P(T|NE) = .17$

$P(W) = .21 \quad P(NE) = .20$

a) $P(W \cap T) = P(W) \cdot P(T|W) = (.21)(.20) =$ **.042**

b) $P(NE \cap T) = P(NE) \cdot P(T|NE) = (.20)(.17) =$ **.034**

c) $P(W|T) = P(W \cap T)/P(T) = (.042)/(.16) =$ **.2625**

d) $P(NE|NT) = P(NE \cap NT)/P(NT) = \{P(NE) \cdot P(NT|NE)\}/P(NT)$

but $P(NT|NE) = 1 - P(T|NE) = 1 - .17 = .83$ and

$P(NT) = 1 - P(T) = 1 - .16 = .84$

Therefore, $P(NE|NT) = \{P(NE) \cdot P(NT|NE)\}/P(NT) =$

$\{(.20)(.83)\}/(.84) =$ **.1976**

e) $P(\text{not } W \cap \text{not } NE|T) = P(\text{not } W \cap \text{not } NE \cap T)/ P(T)$

but $P(\text{not } W \cap \text{not } NE \cap T) =$

$.16 - P(W \cap T) - P(NE \cap T) = .16 - .042 - .034 = .084$

$P(\text{not } W \cap \text{not } NE \cap T)/ P(T) = (.084)/(.16) =$ **.525**

4.41 Let M = MasterCard A = American Express V = Visa

P(M) = .30 P(A) = .20 P(V) = .25

P(M $\cap$ A) = .08 P(V $\cap$ M) = .12 P(A $\cap$ V) = .06

a) P(V $\cup$ A) = P(V) + P(A) - P(V $\cap$ A) = .25 + .20 - .06 = **.39**

b) P(V|M) = P(V $\cap$ M)/P(M) = .12/.30 = **.40**

c) P(M|V) = P(V $\cap$ M)/P(V) = .12/.25 = **.48**

d) P(V) = P(V|M)??

.25 $\neq$ .40

Possession of Visa is not independent of possession of MasterCard

e) American Express is not mutually exclusive of Visa

because P(A $\cap$ V) $\neq$ .0000

4.42 Let S = believe SS secure N = don't believe SS will be secure

<45 = under 45 years old ≥ 45 = 45 or more years old

$P(N) = .51$

Therefore, $P(S) = 1 - .51 = .49$

$P(<45) = .57$ $P(\geq 45) = .43$

$P(S|\geq 45) = .70$

Therefore, $P(N|\geq 45) = 1 - P(S|\geq 45) = 1 - .70 = .30$

a) $P(\geq 45) = 1 - P(<45) = 1 - .57 =$ **.43**

b) $P(<45 \cap S) = P(S) - P(\geq 45 \cap S) =$

but $P(\geq 45 \cap S) = P(\geq 45)\cdot P(S|\geq 45) = (.43)(.70) = .301$

$P(<45 \cap S) = P(S) - P(\geq 45 \cap S) = .49 - .301 =$ **.189**

c) $P(\geq 45|S) = P(\geq 45 \cap S)/P(S) = .189/.49 =$ **.6143**

d) $(<45 \cup N) = P(<45) + P(N) - P(<45 \cap N) =$

but $P(<45 \cap N) = P(<45) - P(<45 \cap S) = .57 - .189 = .381$

so $P(<45 \cup N) = .57 + .51 - .381 =$ **.699**

Probability Matrix Solution for Problem 4.42:

	S	N	
<45	.189	.381	.57
>45	.301	.129	.43
	.490	.510	1.00

4.43 Let M = expect to save more

R = expect to reduce debt

NM = don't expect to save more

NR = don't expect to reduce debt

$P(\text{M}) = .43$ $P(\text{R}) = .45$ $P(\text{R}|\text{M}) = .81$

$P(\text{NR}|\text{M}) = 1 - P(\text{R}|\text{M}) = 1 - .81 = .19$

$P(\text{NM}) = 1 - P(\text{M}) = 1 - .43 = .57$

$P(\text{NR}) = 1 - P(\text{R}) = 1 - .45 = .55$

a) $P(\text{M} \cap \text{R}) = P(\text{M})\cdot P(\text{R}|\text{M}) = (.43)(.81) =$ **.3483**

b) $P(\text{M} \cup \text{R}) = P(\text{M}) + P(\text{R}) - P(\text{M} \cap \text{R})$

$= .43 + .45 - .3483 =$ **.5317**

c) P(neither save nor reduce debt) =

$1 - P(\text{M} \cup \text{R}) = 1 - .5317 =$ **.4683**

d) $P(\text{M} \cap \text{NR}) = P(\text{M})\cdot P(\text{NR}|\text{M}) = (.43)(.19) =$ **.0817**

Probability matrix for problem 4.43:

		Reduce		
		Yes	No	
Save	Yes	.3483	.0817	.43
	No	.1017	.4683	.57
		.45	.55	1.00

4.44 Let R = read

Let B = checked in the with boss

P(R) = .40 P(B) = .34 P(B|R) = .78

a) $P(B \cap R) = P(R) \cdot P(B|R) = (.40)(.78) =$ **.312**

b) $P(NR \cap NB) = 1 - P(R \cup B)$

but $P(R \cup B) = P(R) + P(B) - P(R \cap B) =$

.40 + .34 - .312 = .428

$P(NR \cap NB) = 1 - .428 =$ **.572**

c) $P(R|B) = P(R \cap B)/P(B) = (.312)/(.34) =$ **.9176**

d) $P(NB|R) = 1 - P(B|R) = 1 - .78 =$ **.22**

e) $P(NB|NR) = P(NB \cap NR)/P(NR)$

but $P(NR) = 1 - P(R) = 1 - .40 = .60$

$P(NB|NR) = .572/.60 =$ **.9533**

f) Probability matrix for problem 4.44:

	B	NB	
R	.312	.088	.40

NR	.028	.572	.60
	.34	.66	1.00

4.45 Let Q = keep quiet when they see co-worker misconduct

Let C = call in sick when they are well

P(Q) = .35   P(NQ) = 1 - .35 = .65 P(C|Q) = .75 P(Q|C) = .40

a) $P(C \cap Q) = P(Q) \cdot P(C|Q)$ = (.35)(.75) = **.2625**

b) $P(Q \cup C) = P(Q) + P(C) - P(C \cap Q)$

but P(C) must be solved for:

$P(C \cap Q) = P(C) \cdot P(Q|C)$

.2625 = P(C) (.40)

Therefore, P(C) = .2625/.40 = .65625

and $P(Q \cup C)$ = .35 + .65626 - .2625 = **.74375**

c) $P(NQ|C) = P(NQ \cap C)/P(C)$

but $P(NQ \cap C) = P(C) - P(C \cap Q)$ = .65625 - .2625 = ..39375

Therefore, $P(NQ|C) = P(NQ \cap C)/P(C)$ = .39375/.65626 = **.60**

d) $P(NQ \cap NC) = 1 - P(Q \cup C)$ = 1 - .74375 = **.25625**

e) $P(Q \cap NC) = P(Q) - P(Q \cap C)$ = .35 - .2625 = **.0875**

Probability matrix for problem 4.45:

C

		Y	N	
Q	Y	.2625	.0875	.35
	N	.39375	.25625	.65
		.65625	.34375	1.00

4.46 Let: D = denial

I = inappropriate

C = customer

P = payment dispute

S = specialty

G = delays getting care

R = prescription drugs

P(D) = .17 P(I) = .14 P(C) = .14 P(P) = .11

P(S) = .10 P(G) = .08 P(R) = .07

a) $P(P \cup S) = P(P) + P(S)$ = .11 + .10 = **.21**

b) $P(R \cap C)$ = **.0000** (mutually exclusive)

c) $P(I|S) = P(I \cap S)/P(S)$ = .0000/.10 = **.0000**

d) $P(NG \cap NP) = 1 - P(G \cup P) = 1 - [P(G) + P(P)]$ =

1 – [.08 + .11] = 1 - .19 = **.81**

4.47 Let R = retention Let P = process improvement

P(R) = .56 P(P ㅎ R) = .36 P(R|P) = .90

a) P(R ㅎ NP) = P(R) - P(P ㅎ R) = .56 - .36 = **.20**

b) P(P|R) = P(P ㅎ R)/P(R) = .36/.56 = **.6429**

c) P(P) = ??

Solve P(R|P) = P(R ㅎ P)/P(P) for P(P):

P(P) = P(R ㅎ P)/P(R|P) = .36/.90 = **.40**

d) P(R ㅍ P) = P(R) + P(P) - P(R ㅎ P) =

.56 + .40 - .36 = **.60**

e) P(NR ㅎ NP) = 1 - P(R ㅍ P) = 1 - .60 = **.40**

f) P(R|NP) = P(R ㅎ NP)/P(NP)

but P(NP) = 1 - P(P) = 1 - .40 = .60

P(R|NP) = .20/.60 = **.3333**

		P: Y	P: N	
R	Y	.36	.20	.56
	N	.04	.40	.44
		.40	.60	1.00

Note: In constructing the matrix, we are given P(R) = .56, P(P $\cap$ R) = .36, and

P(R|P) = .90. That is, only one marginal probability is given.

From P(R), we can get P(NR) by taking 1 - .56 = .44.

However, only these two marginal values can be computed directly.

To solve for P(P), using what is given, since we know that 90% of P lies

in the intersection and that the intersection is .36, we can set up an

equation to solve for P:

.90P = .36

Solving for P = .40.

4.48 Let M = mail Let S = sales

P(M) = .38 *P*(M ㅎ S) = .0000 *P*(NM ㅎ NS) = .41

a) *P*(M ㅎ NS) = *P*(M) - *P*(M ㅎ S) = .38 - .00 = .**38**

b) Because *P*(M ㅎ S) = .0000, *P*(M ㅍ S) = *P*(M) + *P*(S)

Therefore, *P*(S) = *P*(M ㅍ S) - *P*(M)

but *P*(M ㅍ S) = 1 - *P*(NM ㅎ NS) = 1 - .41 = .59

Thus, *P*(S) = *P*(M ㅍ S) - *P*(M) = .59 - .38 = .**21**

c) *P*(S|M) = *P*(S ㅎ M)/*P*(M) = .0000/.38 = .**0000**

d) *P*(NM|NS) = *P*(NM ㅎ NS)/*P*(NS) = .41/.79 = .**5190**

where: *P*(NS) = 1 - *P*(S) = 1 - .21 = .79

Probability matrix for problem 4.48:

		S		
		Y	N	
M	Y	.0000	.38	.38
	N	.21	.41	.62
		.21	.79	1.00

4.49 Let F = Flexible Work Let V = Gives time off for Volunteerism

P(F) = .41 P(V|NF) = .10 P(V|F) = .60

from this, P(NF) = 1 - .41 = .59

a) $P(F \cup V) = P(F) + P(V) - P(F \cap V)$

P(F) = .41 and $P(F \cap V) = P(F)\cdot P(V|F)$ = (.41)(.60) = .246

Find P(V) by using $P(V) = P(F \cap V) + P(NF \cap V)$

but $P(NF \cap V) = P(NF)\cdot P(V|NF)$ = (.59)(.10) = .059

so, $P(V) = P(F \cap V) + P(NF \cap V)$ = .246 + .059 = .305

and $P(F \cup V) = P(F) + P(V) - P(F \cap V)$ = .41 + .305 - .246 = **.469**

b) $P(F \cap NV)$ = P(F) - $P(F \cap V)$ = .41 - .246 = **.164**

c) $P(F|NV) = P(F \cap NV)/P(NV)$

$P(F \cap NV)$ = .164

P(NV) = 1 – P(V) = 1 - .305 = .695.

$P(F|NV) = P(F \cap NV)/P(NV)$ = .164/.695 = **.2360**

d) $P(NF|V)$ = $P(NF \cap V)/P(V)$ = .059/.305 = **.1934**

e) $P(NF \cup NV) = P(NF) + P(NV) - P(NF \cap NV)$

P(NF) = .59 P(NV) = .695

Solve for $P(NF \cap NV)$ = P(NV) – $P(F \cap NV)$ = .695 - .164 = .531

$P(NF \cup NV) = P(NF) + P(NV) - P(NF \cap NV)$ = .59 + .695 - .531 = .
754

Probability matrix for problem 4.49:

V

		Y	N	
F	Y	.246	.164	.41
	N	.059	.531	.59
		.305	.695	1.00

4.50 Let S = Sarabia Let T = Tran Let J = Jackson Let B = blood test

P(S) = .41 P(T) = .32 P(J) = .27

$P(B|S)$ = .05 $P(B|T)$ = .08 $P(B|J)$ = .06

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(B\|E_i)$	$P(B \cap E_i)$	$P(B_i\|NS)$
S	.41	.05	.0205	**.3291**
T	.32	.08	.0256	.4109
J	.27	.06	.0162	.2600
			P(B) = **.0623**	

4.51 Let R = regulations T = tax burden

P(R) = .30 P(T) = .35 $P(T|R)$ = .71

a) $P(R \cap T) = P(R)\cdot P(T|R)$ = (.30)(.71) = .**2130**

b) $P(R \cup T) = P(R) + P(T) - P(R \cap T)$ =

.30 + .35 - .2130 = .**4370**

c) $P(R \cup T) - P(R \cap T)$ = .4370 - .2130 = .**2240**

d) $P(R|T) = P(R \cap T)/P(T)$ = .2130/.35 = .**6086**

e) P(NR|T) = 1 - *P*(R|T) = 1 - .6086 = **.3914**

f) *P*(NR|NT) = *P*(NR $\cap$ NT)/*P*(NT) = [1 - *P*(R $\cup$ T)]/*P*(NT) =

(1 - .4370)/.65 = **.8662**

Probability matrix for problem 4.51:

		T: Y	T: N	
R	Y	.213	.087	.30
	N	.137	.563	.70
		.35	.65	1.00

4.52

Event	Prior	Conditional	Joint	Revised
	$P(E_i)$	$P(RB \mid E_i)$	$P(RB \cap E_i)$	$P(E_i \mid RB)$
0-24	.353	.11	.03883	.03883/.25066 = **.15491**
25-34	.142	.24	.03408	.03408/.25066 = **.13596**
35-44	.160	.27	.04320	.04320/.25066 = **.17235**
$\geq$ 45	.345	.39	.13455	.13455/.25066 = **.53678**
			***P*(RB) = .25066**	

4.53 Let GH = Good health Let HM = Happy marriage Let FG = Faith in God

$P(GH) = .29$ $P(HM) = .21$ $P(FG) = .40$

a) $P(HM \cup FG) = P(HM) + P(FG) - P(HM \cap FG)$

but $P(HM \cap FG) = .0000$

$P(HM \cup FG) = P(HM) + P(FG) = .21 + .40 =$ **.61**

b) $P(HM \cup FG \cup GH) = P(HM) + P(FG) + P(GH) =$

.29 + .21 + .40 = **.9000**

c) $P(FG \cap GH)$ = **.0000**

The categories are mutually exclusive.

The respondent could not select more than one answer.

d) P(neither FG nor GH nor HM) = 1 - $P(HM \cup FG \cup GH)$ = 1 - .9000 = **.1000**

Chapter 5
Discrete Distributions

LEARNING OBJECTIVES

The overall learning objective of Chapter 5 is to help you understand a category of probability distributions that produces only discrete outcomes, thereby enabling you to:

1. Distinguish between discrete random variables and continuous random variables.
2. Know how to determine the mean and variance of a discrete distribution.
3. Identify the type of statistical experiments that can be described by the binomial distribution and know how to work such problems.
4. Decide when to use the Poisson distribution in analyzing statistical experiments and know how to work such problems.
5. Decide when binomial distribution problems can be approximated by the Poisson distribution and know how to work such problems.
6. Decide when to use the hypergeometric distribution and know how to work such problems

CHAPTER TEACHING STRATEGY

Chapters 5 and 6 introduce the student to several statistical distributions. It is important to differentiate between the discrete distributions of chapter 5 and the continuous distributions of chapter 6.

The approach taken in presenting the binomial distribution is to build on techniques presented in chapter 4. It can be helpful to take the time to apply the law of multiplication for independent events to a problem and demonstrate to students that sequence is important. From there, the student will more easily understand that by using combinations, one can more quickly determine the number of sequences and weigh the probability of obtaining a single sequence by that number. In a sense, we are developing the binomial formula through an inductive process. Thus, the binomial formula becomes more of a summary device than a statistical "trick". The binomial tables presented in this text are non cumulative. This makes it easier for the student to recognize that the table is but a listing of a series of binomial formula computations. In addition, it lends itself more readily to the graphing of a binomial distribution.

It is important to differentiate applications of the Poisson distribution from binomial distribution problems. It is often difficult for students to determine which type of distribution to apply to a problem. The Poisson distribution applies to rare occurrences over some interval. The parameters involved in the binomial distribution (n and p) are different from the parameter (Lambda) of a Poisson distribution.

It is sometimes difficult for students to know how to handle Poisson problems in which the interval for the problem is different than the stated interval for Lambda. Note that in such problems, it is always the value of Lambda that is adjusted not the value of x. Lambda is a

long-run average that can be appropriately adjusted for various intervals. For example, if a store is averaging λ customers in 5 minutes, then it will also be averaging 2λ customers in 10 minutes. On the other hand, x is a one-time observation and just because x customers arrive in 5 minutes does not mean that $2x$ customers will arrive in 10 minutes.

Solving for the mean and standard deviation of binomial distributions prepares the students for chapter 6 where the normal distribution is sometimes used to approximate binomial distribution problems. Graphing binomial and Poisson distributions affords the student the opportunity to visualize the meaning and impact of a particular set of parameters for a distribution. In addition, it is possible to visualize how the binomial distribution approaches the normal curve as *p* gets nearer to .50 and as *n* gets larger for other values of *p*. It can be useful to demonstrate this in class along with showing how the graphs of Poisson distributions also approach the normal curve as λ gets larger.

In this text (as in most) because of the number of variables used in its computation, only exact probabilities are determined for hypergeometric distribution. This, combined with the fact that there are no hypergeometric tables given in the text, makes it cumbersome to determine cumulative probabilities for the hypergeometric distribution. Thus, the hypergeometric distribution can be presented as a fall-back position to be used only when the binomial distribution should not be applied because of the non independence of trials and size of sample.

CHAPTER OUTLINE

5.1 Discrete Versus Continuous Distributions

5.2 Describing a Discrete Distribution

Mean, Variance, and Standard Deviation of Discrete Distributions

Mean or Expected Value

Variance and Standard Deviation of a Discrete Distribution

5.3 Binomial Distribution

Solving a Binomial Problem

Using the Binomial Table

Using the Computer to Produce a Binomial Distribution

Mean and Standard Deviation of the Binomial Distribution

Graphing Binomial Distributions

5.4 Poisson Distribution

Working Poisson Problems by Formula

Using the Poisson Tables

Mean and Standard Deviation of a Poisson Distribution

Graphing Poisson Distributions

Using the Computer to Generate Poisson Distributions

Approximating Binomial Problems by the Poisson Distribution

5.5 Hypergeometric Distribution

Using the Computer to Solve for Hypergeometric Distribution Probabilities

KEY TERMS

Binomial Distribution

Hypergeometric Distribution

Continuous Distributions

Lambda (λ)

Continuous Random Variables

Mean, or Expected Value

Discrete Distributions

Poisson Distribution

Discrete Random Variables

Random Variable

SOLUTIONS TO PROBLEMS IN CHAPTER 5

5.1

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
1	.238	.238	2.775556	0.6605823
2	.290	.580	0.443556	0.1286312
3	.177	.531	0.111556	0.0197454
4	.158	.632	1.779556	0.2811700
5	.137	.685	5.447556	0.7463152

$\mu = \sum[x \cdot P(x)] =$ **2.666** $\sigma^2 = \sum[(x-\mu)^2 \cdot P(x)] =$ **1.836444**

$\sigma = \sqrt{1.836444} =$ **1.355155**

5.2

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
0	.103	.000	7.573504	0.780071
1	.118	.118	3.069504	0.362201
2	.246	.492	0.565504	0.139114

3	.229	.687	0.061504	0.014084
4	.138	.552	1.557504	0.214936
5	.094	.470	5.053504	0.475029
6	.071	.426	10.549500	0.749015
7	.001	.007	18.045500	0.018046

$\mu = \sum [x \cdot P(x)] =$ **2.752** $\sigma^2 = \sum[(x-\mu)^2 \cdot P(x)] =$ **2.752496**

$\sigma = \sqrt{2.752496} =$ **1.6591**

5.3

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
0	.461	.000	0.913936	0.421324
1	.285	.285	0.001936	0.000552
2	.129	.258	1.089936	0.140602
3	.087	.261	4.177936	0.363480
4	.038	.152	9.265936	0.352106

$E(x) = \mu = \sum[x \cdot P(x)] =$ **0.956** $\sigma^2 = \sum[(x-\mu)^2 \cdot P(x)] =$ **1.278064**

$$\sigma = \sqrt{1.278064} = \mathbf{1.1305}$$

5.4

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$
0	.262	.000	1.4424	0.37791
1	.393	.393	0.0404	0.01588
2	.246	.492	0.6384	0.15705
3	.082	.246	3.2364	0.26538
4	.015	.060	7.8344	0.11752
5	.002	.010	14.4324	0.02886
6	.000	.000	23.0304	0.00000

$\mu = \sum [x \cdot P(x)] =$ **1.201** $\sigma^2 = \sum [(x-\mu)^2 \cdot P(x)] =$ **0.96260**

$\sigma = \sqrt{.96260} =$ **.98112**

5.5 a) $n = 4$ $p = .10$ $q = .90$

$P(x=3) = {}_4C_3(.10)^3(.90)^1 = 4(.001)(.90) =$ **.0036**

b) $n = 7$ $p = .80$ $q = .20$

$P(x=4) = {}_7C_4(.80)^4(.20)^3 = 35(.4096)(.008) =$ **.1147**

c) $n = 10$ $p = .60$ $q = .40$

$P(x \geq 7) = P(x{=}7) + P(x{=}8) + P(x{=}9) + P(x{=}10) =$

${}_{10}C_7(.60)^7(.40)^3 + {}_{10}C_8(.60)^8(.40)^2 + {}_{10}C_9(.60)^9(.40)^1 + {}_{10}C_{10}(.60)^{10}(.40)^0 =$

120(.0280)(.064) + 45(.0168)(.16) + 10(.0101)(.40) + 1(.0060)(1) =

.2150 + .1209 + .0403 + .0060 = **.3822**

d) $n = 12$ $p = .45$ $q = .55$

$P(5 \leq x \leq 7) = P(x{=}5) + P(x{=}6) + P(x{=}7) =$

${}_{12}C_5(.45)^5(.55)^7 + {}_{12}C_6(.45)^6(.55)^6 + {}_{12}C_7(.45)^7(.55)^5 =$

792(.0185)(.0152) + 924(.0083)(.0277) + 792(.0037)(.0503) =

.2225 + .2124 + .1489 = **.5838**

5.6 By Table A.2:

a) $n = 20$ $p = .50$

$P(x=12) =$ **.120**

b) $n = 20$ $p = .30$

$P(x > 8) = P(\text{x}=9) + P(x=10) + P(x=11) + ...+ P(x=20) =$

$.065 + .031 + .012 + .004 + .001 + .000 =$ **.113**

c) $n = 20$ $p = .70$

$P(x < 12) =$ $P(x=11) + P(x=10) + P(x=9) + ... + P(x=0)$
=

$.065 + .031 + .012 + .004 + .001 + .000 =$ **.113**

d) $n = 20$ $p = .90$

$P(x \leq 16) = P(x=16) + P(x=15) + P(x=14) + ...+ P(x=0) =$

$.090 + .032 + .009 + .002 + .000 =$ **.133**

e) $n = 15$ $p = .40$

$P(4 \leq x \leq 9) =$

$P(x{=}4) + P(x{=}5) + P(x{=}6) + P(x{=}7) + P(x{=}8) + P(x{=}9) =$

$.127 + .186 + .207 + .177 + .118 + .061 =$ **.876**

f) $n = 10$ $p = .60$

$P(x \geq 7) = P(x{=}7) + P(x{=}8) + P(x{=}9) + P(x{=}10) =$

$.215 + .121 + .040 + .006 =$ **.382**

5.7 a) $n = 20$
$p = .70$
$q = .30$

$\mu = n \cdot p = 20(.70) =$ **14**

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{20(.70)(.30)} = \sqrt{4.2} =$ **2.05**

b) $n = 70$ $p = .35$ $q = .65$

$\mu = n \cdot p = 70(.35) =$ **24.5**

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{70(.35)(.65)} = \sqrt{15.925} =$ **3.99**

c) $n = 100$ $p = .50$ $q = .50$

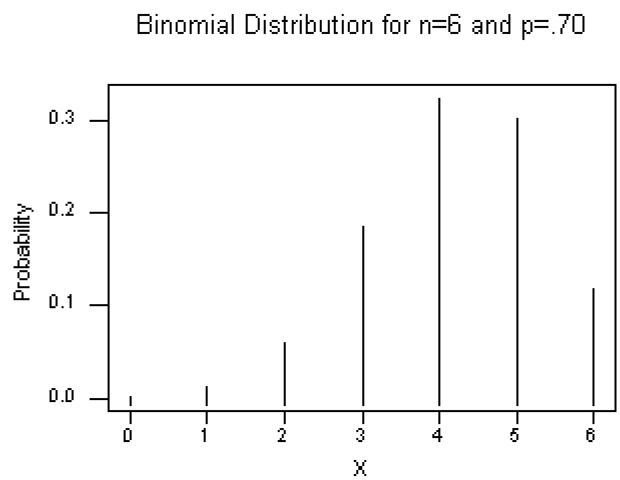
$\mu = n \cdot p = 100(.50) =$ **50**

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{100(.50)(.50)} = \sqrt{25} =$ **5**

5.8 a) $n = 6$ $p = .70$

x	Prob
0	.001

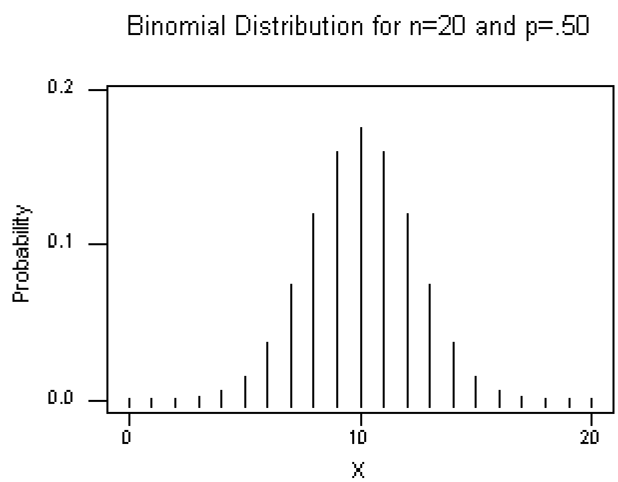
1	.010
2	.060
3	.185
4	.324
5	.303
6	.118



b) *n* = 20 *p* = .50

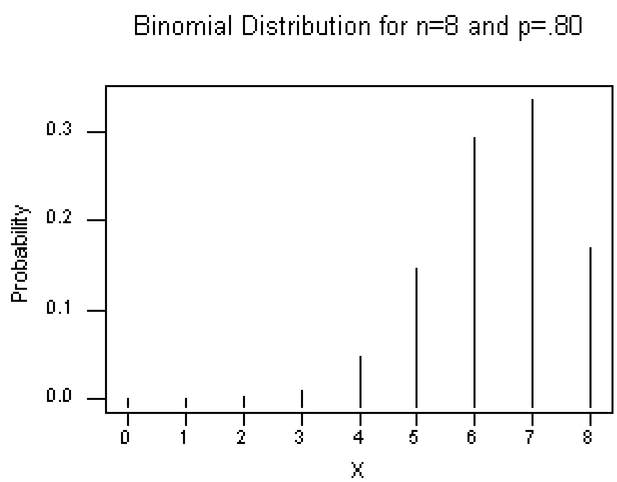
x	Prob
0	.000
1	.000
2	.000
3	.001
4	.005
5	.015
6	.037
7	.074
8	.120
9	.160

10	.176
11	.160
12	.120
13	.074
14	.037
15	.015
16	.005
17	.001
18	.000
19	.000
20	.000



c) $n = 8$ $p = .80$

x	Prob
0	.000
1	.000
2	.001
3	.009
4	.046
5	.147
6	.294
7	.336
8	.168



5.9 a) $n = 20$ $p = .78$ $x = 14$

$_{20}C_{14}\,(.78)^{14}(.22)^{6} = 38{,}760(.030855)(.00011338) = \mathbf{.1356}$

b) $n = 20$ $p = .75$ $x = 20$

${}_{20}C_{20}\ (.75)^{20}(.25)^{0}$ = (1)(.0031712)(1) = **.0032**

c) $n = 20$ $p = .70$ $x < 12$

Use table A.2:

$P(x=0) + P(x=1) + \ldots + P(x=11) =$

.000 + .000 + .000 + .000 + .000 + .000 + .000 +

.001 + .004 + .012 + .031 + .065 = .**113**

5.10 $n = 16$ $p = .40$

$P(x \geq 9)$: from Table A.2:

x	Prob
9	.084
10	.039
11	.014
12	.004
13	.001
	.142

$P(3 \leq x \leq 6)$:

x	Prob
3	.047
4	.101
5	.162
6	.198
	.508

$n = 13$ $p = .88$

$P(x = 10) = {}_{13}C_{10}(.88)^{10}(.12)^3 = 286(.278500976)(.001728) =$ **.1376**

$P(x = 13) = {}_{13}C_{13}(.88)^{13}(.12)^{0} = (1)(.1897906171)(1) =$ **.1898**

Expected Value = $\mu = n \cdot p = 13(.88) =$ **11.44**

5.11 $n = 25$ $p = .60$

a) $x \geq 15$

$P(x \geq 15) = P(x = 15) + P(x = 16) + \cdots + P(x = 25)$

Using Table A.2 $n = 25$, $p = .60$

x	Prob
15	.161
16	.151
17	.120
18	.080
19	.044
20	.020
21	.007
22	.002
	.585

b) $x > 20$

$P(x > 20) = P(x = 21) + P(x = 22) + P(x = 23) + P(x = 24) + P(x = 25) =$

Using Table A.2 $n = 25$, $p = .60$

.007 + .002 + .000 + .000 + .000 = **.009**

c) $P(x < 10)$

Using Table A.2 $n = 25$, $p = .60$ and $x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$

x	Prob.
9	.009
8	.003
7	.001
≤ 6	.000
	.013

5.12 $n = 16$ $p = .50$ $x > 10$

Using Table A.2, $n = 16$ and $p = .50$, $P(x=11) + P(x=12) + \ldots +$ $P(x=16) =$

x	Prob.
11	.067
12	.028
13	.009
14	.002
15	.000
16	.000
	.106

For $n = 10$ $p = .87$ $x = 6$

$_{10}C_6\,(.87)^6(.13)^4 = 210(.433626)(.00028561) =$ **.0260**

5.13 $n = 15$ $p = .20$

a) $P(x = 5) = {}_{15}C_5(.20)^5(.80)^{10} = 3003(.00032)(.1073742) =$ **.1032**

b) $P(x > 9)$: Using Table A.2

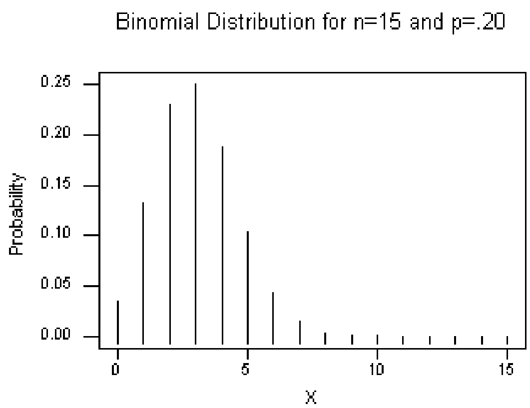
$P(x = 10) + P(x = 11) + \ldots + P(x = 15) = .000 + .000 + \ldots + .000 =$ **.000**

c) $P(x = 0) = {}_{15}C_0(.20)^0(.80)^{15} = (1)(1)(.035184) =$ **.0352**

d) $P(4 \leq x \leq 7)$: Using Table A.2

$P(x = 4) + P(x = 5) + P(x = 6) + P(x = 7) = .188 + .103 + .043 + .014 =$ **.348**

e)



5.14 $n = 18$

a) $p = .30$ $\mu = 18(.30) =$ **5.4**

$p = .34$ $\mu = 18(.34) =$ **6.12**

b) $P(x \geq 8)$ $n = 18$ $p = .30$

from Table A.2

x	Prob

8	.081
9	.039
10	.015
11	.005
12	.001
	.141

c) $n = 18$ $p = .34$

$P(2 \leq x \leq 4) = P(x = 2) + P(x = 3) + P(x = 4) =$

${}_{18}C_2(.34)^2(.66)^{16} + {}_{18}C_3(.34)^3(.66)^{15} + {}_{18}C_4(.34)^4(.66)^{14} =$

.0229 + .0630 + .1217 = **.2076**

d) $n = 18$ $p = .30$ $x = 0$

${}_{18}C_0(.30)^0(.70)^{18}$ = **.00163**

$n = 18$ $p = .34$ $x = 0$

${}_{18}C_0(.34)^0(.66)^{18}$ = **.00056**

Since only 30% (compared to 34%) fall in the $500,000 to $1,000,000 category, it is more likely that none of the CPA financial advisors would fall in this category.

5.15 a) $P(x=5|\lambda = 2.3) = \frac{2.3^5 \cdot e^{-2.3}}{5!} = \frac{(64.36343)(.100259)}{120}$ = **.0538**

b) $P(x=2|\lambda = 3.9) = \frac{3.9^2 \cdot e^{-3.9}}{2!} = \frac{(15.21)(.020242)}{2}$ = **.1539**

c) $P(x \leq 3|\lambda = 4.1) = P(x=3) + P(x=2) + P(x=1) + P(x=0) =$

$\frac{4.1^3 \cdot e^{-4.1}}{3!} = \frac{(68.921)(.016573)}{6}$ = .1904

$\frac{4.1^2 \cdot e^{-4.1}}{2!} = \frac{(16.81)(.016573)}{2}$ = .1393

$\frac{4.1^1 \cdot e^{-4.1}}{1!} = \frac{(4.1)(.016573)}{1}$ = .0679

$\frac{4.1^0 \cdot e^{-4.1}}{0!} = \frac{(1)(.016573)}{1}$ = .0166

.1904 + .1393 + .0679 + .0166 = **.4142**

d) $P(x=0|\lambda = 2.7) =$

$$\frac{2.7^0 \cdot e^{-2.7}}{0!} = \frac{(1)(.06721)}{1}$$

= **.0672**

e) $P(x=1 | \lambda = 5.4)=$

$$\frac{5.4^1 \cdot e^{-5.4}}{1!} = \frac{(5.4)(.0045166)}{1}$$

= **.0244**

f) $P(4 < x < 8 | \lambda = 4.4)$: $P(x=5|\lambda = 4.4) + P(x=6|\lambda = 4.4) + P(x=7|\lambda = 4.4)=$

$$\frac{4.4^5 \cdot e^{-4.4}}{5!} + \frac{4.4^6 \cdot e^{-4.4}}{6!} + \frac{4.4^7 \cdot e^{-4.4}}{7!} =$$

$$\frac{(1649.1622)(.01227734)}{120} + \frac{(7256.3139)(.01227734)}{720} + \frac{(31,927.781)(.01227734)}{5040}$$

= .1687 + .1237 + .0778 = **.3702**

5.16 a) $P(x=6|\lambda = 3.8)$ = **.0936**

b) $P(x>7|\lambda = 2.9)$:

x	Prob
8	.0068
9	.0022
10	.0006

11	.0002
12	.0000
	.0098

c) $P(3 \leq x \leq 9|\lambda = 4.2)=$

x	Prob
3	.1852
4	.1944
5	.1633
6	.1143
7	.0686
8	.0360
9	.0168
	.7786

d) $P(x=0|\lambda = 1.9)$ = **.1496**

e) $P(x \leq 6|\lambda = 2.9)=$

x	Prob
0	.0550
1	.1596
2	.2314
3	.2237
4	.1622
5	.0940

6	.0455
	.9714

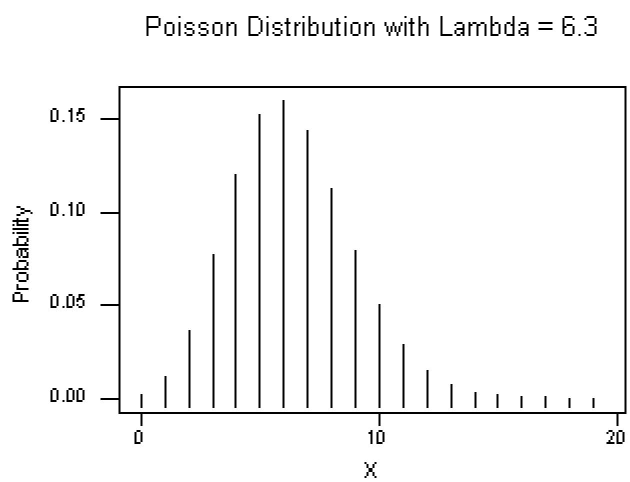
f) $P(5 < x \leq 8 \mid \lambda = 5.7) =$

x	Prob
6	.1594
7	.1298
8	.0925
	.3817

5.17 a) $\lambda = 6.3$ mean = **6.3** Standard deviation = $\sqrt{6.3}$ = **2.51**

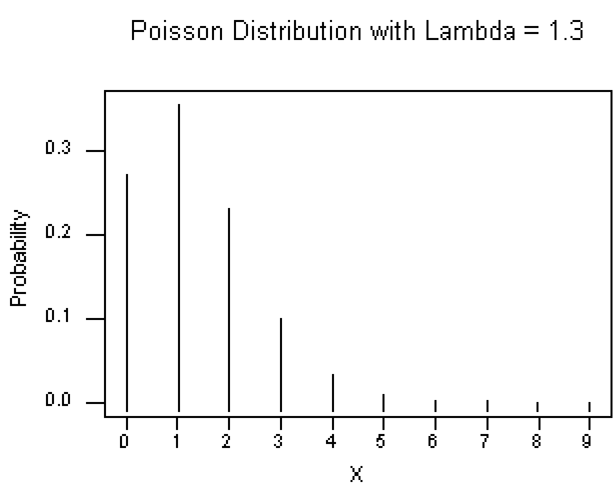
x	Prob
0	.0018
1	.0116
2	.0364
3	.0765
4	.1205
5	.1519
6	.1595
7	.1435
8	.1130
9	.0791
10	.0498

11	.0285
12	.0150
13	.0073
14	.0033
15	.0014
16	.0005
17	.0002
18	.0001
19	.0000



b) λ = 1.3 mean = **1.3** standard deviation = $\sqrt{1.3}$ = **1.14**

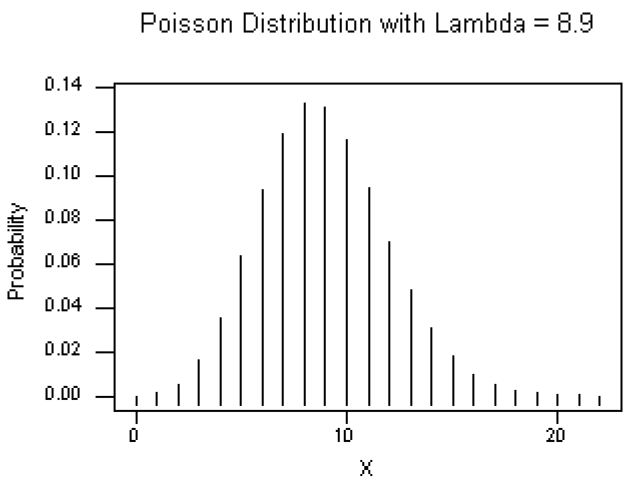
x	Prob
0	.2725
1	.3542
2	.2303
3	.0998
4	.0324
5	.0084
6	.0018
7	.0003
8	.0001
9	.0000



c) $\lambda = 8.9$ mean = **8.9** standard deviation = $\sqrt{8.9}$ = **2.98**

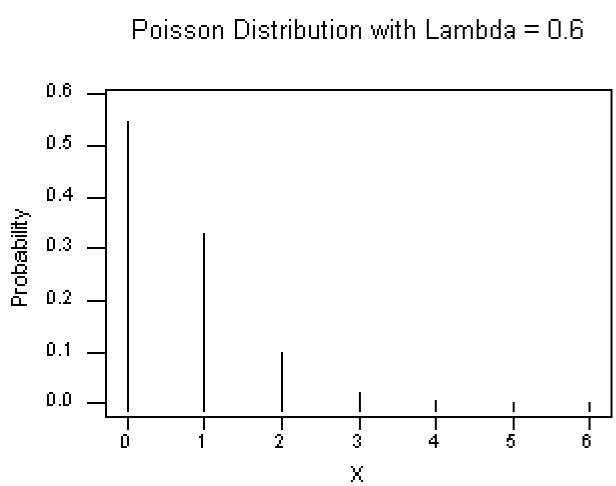
x	Prob
0	.0001
1	.0012
2	.0054
3	.0160
4	.0357
5	.0635
6	.0941
7	.1197
8	.1332
9	.1317
10	.1172
11	.0948
12	.0703
13	.0481
14	.0306
15	.0182
16	.0101
17	.0053
18	.0026
19	.0012
20	.0005
21	.0002

22	.0001


Poisson Distribution with Lambda = 8.9
Probability
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
10
20
X

d) $\lambda = 0.6$ mean = **0.6** standard deviation = $\sqrt{0.6}$ = **.775**

x	Prob
0	.5488
1	.3293
2	.0988
3	.0198
4	.0030
5	.0004
6	.0000



5.18 $\lambda = 2.8|4$ minutes

a) $P(x=6|\ \lambda = 2.8)$

from Table A.3 **.0407**

b) $P(x=0|\ \lambda = 2.8) =$

from Table A.3 **.0608**

c) Unable to meet demand if $x > 4|4$ minutes:

x	Prob.
5	.0872
6	.0407
7	.0163
8	.0057
9	.0018
10	.0005
11	.0001
	.1523

There is a **.1523** probability of being unable to meet the demand.

Probability of meeting the demand = 1 - (.1523) = **.8477**

15.23% of the time a second window will need to be opened.

d) λ = 2.8 arrivals | 4 minutes

$P(x=3)$ arrivals | 2 minutes = ??

Lambda must be changed to the same interval (½ the size)

New lambda=1.4 arrivals | 2 minutes

$P(x=3) | \lambda = 1.4)$ = from Table A.3 = **.1128**

$P(x \geq 5 |$ 8 minutes) = ??

Lambda must be changed to the same interval(twice the size):

New lambda = 5.6 arrivals | 8 minutes

$P(x \geq 5 \mid \lambda = 5.6)$:

From Table A.3:

x	Prob.
5	.1697
6	.1584
7	.1267
8	.0887
9	.0552
10	.0309
11	.0157
12	.0073
13	.0032
14	.0013
15	.0005
16	.0002
17	.0001
	.6579

5.19 $\lambda = \Sigma x/n = 126/36 =$ **3.5**

Using Table A.3

a) $P(x = 0) =$ **.0302**

b) $P(x \geq 6) = P(x = 6) + P(x = 7) + \ldots =$

.0771 + .0385 + .0169 + .0066 + .0023 +

$.0007 + .0002 + .0001 =$ **.1424**

c) $P(x < 4 \mid 10 \text{ minutes})$

Double Lambda to $\lambda = 7.0 \mid 10$ minutes

$P(x < 4) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) =$

$.0009 + .0064 + .0223 + .0521 =$ **.0817**

d) $P(3 \leq x \leq 6 \mid 10 \text{ minutes})$

$\lambda = 7.0 \mid 10$ minutes

$P(3 \leq x \leq 6) = P(x = 3) + P(x = 4) + P(x = 5) + P(x = 6)$

$= .0521 + .0912 + .1277 + .1490 =$ **.42**

e) $P(x = 8 \mid 15 \text{ minutes})$

Change Lambda for a 15 minute interval by multiplying the original Lambda by 3.

$\lambda = 10.5 \mid 15$ minutes

$$P(x = 8 \mid 15 \text{ minutes}) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{(10.5^8)(e^{-10.5})}{8!} = \mathbf{.1009}$$

5.20 λ = 5.6 days | 3 weeks

a) $P(x=0 \mid \lambda = 5.6)$:

from Table A.3 = **.0037**

b) $P(x=6 \mid \lambda = 5.6)$:

from Table A.3 = **.1584**

c) $P(x \geq 15 \mid \lambda = 5.6)$:

x	Prob.
15	.0005
16	.0002
17	.0001
	.0008

Because this probability is so low, if it actually occurred, the researcher would question the Lambda value as too low for this period. Perhaps the value of

Lambda has changed because of an overall increase in pollution.

5.21 λ = 0.6 trips | 1 year

a) $P(x=0 \mid \lambda = 0.6)$:

from Table A.3 = **.5488**

b) $P(x=1 \mid \lambda = 0.6)$:

from Table A.3 = **.3293**

c) $P(x \geq 2 \mid \lambda = 0.6)$:

from Table A.3

x	Prob.
2	.0988
3	.0198
4	.0030
5	.0004
6	.0000
	.1220

d) $P(x \leq 3 \mid 3 \text{ year period})$:

The interval length has been increased (3 times)

New Lambda = λ = 1.8 trips | 3 years

$P(x \leq 3 | \lambda = 1.8)$:

from Table A.3

x	Prob.
0	.1653
1	.2975
2	.2678
3	.1607
	.8913

e) $P(x=4 | 6 \text{ years})$:

The interval has been increased (6 times)

New Lambda = λ = 3.6 trips | 6 years

$P(x=4 | \lambda = 3.6)$:

from Table A.3 = **.1912**

5.22 λ = 1.2 collisions | 4 months

a) $P(x=0 | \lambda = 1.2)$:

from Table A.3 = **.3012**

b) $P(x=2 | 2 \text{ months})$:

The interval has been decreased (by ½)

New Lambda = λ = 0.6 collisions | 2 months

$P(x=2 | \lambda = 0.6)$:

from Table A.3 = **.0988**

c) $P(x \leq 1 \text{ collision} | 6 \text{ months})$:

The interval length has been increased (by 1.5)

New Lambda = λ = 1.8 collisions | 6 months

$P(x \leq 1 | \lambda = 1.8)$:

from Table A.3

x	Prob.
0	.1653
1	.2975
	.4628

The result is likely to happen almost half the time (46.26%). Ship channel and

weather conditions are about normal for this period. Safety awareness is

about normal for this period. There is no compelling reason to reject the

lambda value of 0.6 collisions per 4 months based on an outcome of 0 or 1

collisions per 6 months.

5.23 $\lambda = 1.2$ pens|carton

a) $P(x=0 \mid \lambda = 1.2)$:

from Table A.3 = **.3012**

b) $P(x \geq 8 \mid \lambda = 1.2)$:

from Table A.3 = **.0000**

c) $P(x > 3 \mid \lambda = 1.2)$:

from Table A.3

x	Prob.
4	.0260
5	.0062
6	.0012
7	.0002
8	.0000
	.0336

5.24 $n = 100{,}000$ $p = .00004$

$P(x \geq 7 \mid n = 100{,}000 \;\; p = .00004)$:

$\lambda = \mu = n \cdot p = 100{,}000(.00004) = 4.0$

Since $n > 20$ and $n \cdot p \leq 7$, the Poisson approximation to this binomial problem is

close enough.

$P(x \geq 7 \mid \lambda = 4)$:

Using Table A.3

x	Prob.
7	.0595
8	.0298
9	.0132
10	.0053
11	.0019
12	.0006
13	.0002
14	.0001
	.1106

$P(x > 10 \mid \lambda = 4)$:

Using Table A.3

x	Prob.
11	.0019

12	.0006	
13	.0002	
14	.0001	
		.0028

Since getting more than 10 is a rare occurrence, this particular geographic region appears to have a higher average rate than other regions. An investigation of particular characteristics of this region might be warranted.

5.25 $p = .009$ $n = 200$

Use the Poisson Distribution:

$\lambda = n \cdot p = 200(.009) = 1.8$

a) $P(x \geq 6)$ from Table A.3 =

$P(x = 6) + P(x = 7) + P(x = 8) + P(x = 9) + \ldots =$

.0078 + .0020 + .0005 + .0001 = **.0104**

b) $P(x > 10)$ = .**0000**

c) $P(x = 0)$ = **.1653**

d) $P(x < 5) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) =$

.1653 + .2975 + .2678 + .1607 + .0723 = **.9636**

5.26 If 99% see a doctor, then 1% do not see a doctor. Thus, p = .01 for this problem.

$n = 300$, $p = .01$, $\lambda = n(p) = 300(.01) = 3$

a) $P(x = 5)$:

Using $\lambda = 3$ and Table A.3 = **.1008**

b) $P(x < 4) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) =$

.0498 + .1494 + .2240 + .2240 = **.6472**

c) The expected number = $\mu = \lambda =$ **3**

5.27 a) $P(x = 3 \mid N = 11, A = 8, n = 4)$

$$\frac{{}_8C_3 \cdot {}_3C_1}{{}_{11}C_4} = \frac{(56)(3)}{330}$$

= **.5091**

b) $P(x < 2) \mid N = 15, A = 5, n = 6)$

$P(x = 1) + P\,(x = 0) =$

$$\frac{{}_5C_1 \cdot {}_{10}C_5}{{}_{15}C_6} + \frac{{}_5C_0 \cdot {}_{10}C_6}{{}_{15}C_6} = \frac{(5)(252)}{5005} + \frac{(1)(210)}{5005}$$

.2517 + .0420 = **.2937**

c) $P(x=0 \mid N = 9, A = 2, n = 3)$

$$\frac{{}_2C_0 \cdot {}_7C_3}{{}_9C_3} = \frac{(1)(35)}{84}$$

= **.4167**

d) $P(x > 4 \mid N = 20, A = 5, n = 7) =$

$P(x = 5) + P(x = 6) + P(x = 7) =$

$$\frac{{}_5C_5 \cdot {}_{15}C_2}{{}_{20}C_7} + \frac{{}_5C_6 \cdot {}_{15}C_1}{{}_{20}C_7} + \frac{{}_5C_7 \cdot {}_{15}C_0}{{}_{20}C_7} =$$

$\frac{(1)(105)}{77520}$ + ${}_5C_6$ (impossible) + ${}_5C_7$(impossible) = **.0014**

5.28 $N = 19$ $n = 6$

a) $P(x = 1 \text{ private})$ $A = 11$

$$\frac{{}_{11}C_1 \cdot {}_8C_5}{{}_{19}C_6} = \frac{(11)(56)}{27{,}132}$$

= **.0227**

b) $P(x = 4 \text{ private})$

$$\frac{{}_{11}C_4 \cdot {}_8C_2}{{}_{19}C_6} = \frac{(330)(28)}{27{,}132}$$

= **.3406**

c) $P(x = 6 \text{ private})$

$$\frac{{}_{11}C_6 \cdot {}_8C_0}{{}_{19}C_6} = \frac{(462)(1)}{27{,}132}$$

= **.0170**

d) $P(x = 0 \text{ private})$

$$\frac{{}_{11}C_0 \cdot {}_8C_6}{{}_{19}C_6} = \frac{(1)(28)}{27{,}132}$$

= **.0010**

5.29 $N = 17$ $A = 8$ $n = 4$

a) $P(x = 0) = \dfrac{{}_{8}C_{0} \cdot {}_{9}C_{4}}{{}_{17}C_{4}} = \dfrac{(1)(126)}{2380} =$ **.0529**

b) $P(x = 4) = \dfrac{{}_{8}C_{4} \cdot {}_{9}C_{0}}{{}_{17}C_{4}} = \dfrac{(70)(1)}{2380} =$ **.0294**

c) $P(x = 2 \text{ non computer}) = \dfrac{{}_{9}C_{2} \cdot {}_{8}C_{2}}{{}_{17}C_{4}} = \dfrac{(36)(28)}{2380} =$ **.4235**

5.30 $N = 20$ $A = 16$ white $N - A = 4$ red $n = 5$

a) $P(x = 4 \text{ white}) = \dfrac{{}_{16}C_{4} \cdot {}_{4}C_{1}}{{}_{20}C_{5}} = \dfrac{(1820)(4)}{15504} =$ **.4696**

b) $P(x = 4 \text{ red}) = \dfrac{{}_{4}C_{4} \cdot {}_{16}C_{1}}{{}_{20}C_{5}} = \dfrac{(1)(16)}{15504} =$ **.0010**

c) $P(x = 5 \text{ red}) = \dfrac{{}_{4}C_{5} \cdot {}_{16}C_{0}}{{}_{20}C_{5}}$ = **.0000** because ${}_{4}C_{5}$ is impossible to determine

The participant cannot draw 5 red beads if there are only 4 to draw from.

5.31 $N = 10$ $n = 4$

a) $A = 3$ $x = 2$

$P(x = 2) = \frac{{}_3C_2 \cdot {}_7C_2}{{}_{10}C_4} = \frac{(3)(21)}{210} =$ **.30**

b) $A = 5$ $x = 0$

$P(x = 0) = \frac{{}_5C_0 \cdot {}_5C_4}{{}_{10}C_4} = \frac{(1)(5)}{210} =$ **.0238**

c) $A = 5$ $x = 3$

$P(x = 3) = \frac{{}_5C_3 \cdot {}_5C_1}{{}_{10}C_4} = \frac{(10)(5)}{210} =$ **.2381**

5.32 $N = 16$ $A = 4$ defective $n = 3$

a) $P(x = 0) = \frac{{}_4C_0 \cdot {}_{12}C_3}{{}_{16}C_3} = \frac{(1)(220)}{560} =$ **.3929**

b) $P(x = 3) = \dfrac{{}_4C_3 \cdot {}_{12}C_0}{{}_{16}C_3} = \dfrac{(4)(1)}{560}$ = **.0071**

c) $P(x \geq 2)$ = $P(x=2) + P(x=3)$ = $\dfrac{{}_4C_2 \cdot {}_{12}C_1}{{}_{16}C_3}$ + .0071 (from part b.) =

$\dfrac{(6)(12)}{560}$ + .0071 = .1286 + .0071 = **.1357**

d) $P(x \leq 1) = P(x=1) + P(x=0)$ =

$\dfrac{{}_4C_1 \cdot {}_{12}C_2}{{}_{16}C_3}$ + .3929 (from part a.) = $\dfrac{(4)(66)}{560}$ + .3929 = .4714 + .3929 = **.8643**

5.33 $N = 18$ $A = 11$ Hispanic $n = 5$

$P(x \leq 1) = P(1) + P(0) =$

$$\frac{{}_{11}C_1 \cdot {}_7C_4}{{}_{18}C_5} + \frac{{}_{11}C_0 \cdot {}_7C_5}{{}_{18}C_5} = \frac{(11)(35)}{8568} + \frac{(1)(21)}{8568} = .0449 + .0025 = \mathbf{.0474}$$

It is fairly unlikely that these results occur by chance. A researcher might want to

further investigate this result to determine causes. Were officers selected based on

leadership, years of service, dedication, prejudice, or some other reason?

5.34 a) $P(x=4 \mid n = 11 \text{ and } p = .23)$

${}_{11}C_4(.23)^4(.77)^7 = 330(.0028)(.1605) =$ **.1482**

b) $P(x \geq 1 \mid n = 6 \text{ and } p = .50) =$

$1 - P(x < 1) = 1 - P(x = 0) =$

$1 – [{}_6C_0(.50)^0(.50)^6] = 1 – [(1)(1)(.0156)] =$ **.9844**

c) $P(x > 7 \mid n = 9 \text{ and } p = .85) = P(x = 8) + P(x = 9) =$

${}_9C_8(.85)^8(.15)^1 + {}_9C_9(.85)^9(.15)^0 =$

$(9)(.2725)(.15) + (1)(.2316)(1) = .3679 + .2316 =$ **.5995**

d) $P(x \leq 3 \mid n = 14 \text{ and } p = .70) =$

$P(x = 3) + P(x = 2) + P(x = 1) + P(x = 0) =$

${}_{14}C_3(.70)^3(.30)^{11} + {}_{14}C_2(.70)^2(.30)^{12} +$

${}_{14}C_1(.70)^1(.30)^{13} + {}_{14}C_0(.70)^0(.30)^{14} =$

$(364)(.3430)(.00000177) + (91)(.49)(.000000531)=$

$(14)(.70)(.00000016) + (1)(1)(.000000048) =$

$.0002 + .0000 + .0000 + .0000 =$ **.0002**

5.35 a) $P(x = 14 \mid n = 20 \text{ and } p = .60) =$ **.124**

b) $P(x < 5 \mid n = 10 \text{ and } p = .30) =$

$P(x = 4) + P(x = 3) + P(x = 2) + P(x = 1) + P(x=0) =$

x	Prob.
0	.028
1	.121
2	.233
3	.267
4	.200
	.849

c) $P(x \geq 12 \mid n = 15 \text{ and } p = .60) =$

$P(x = 12) + P(x = 13) + P(x = 14) + P(x = 15)$

x	Prob.
12	.063
13	.022
14	.005
15	.000
	.090

d) $P(x > 20 \mid n = 25 \text{ and } p = .40) = P(x = 21) + P(x = 22) +$

$P(x = 23) + P(x = 24) + P(x=25) =$

x	Prob.
21	.000
22	.000
23	.000
24	.000
25	.000
	.000

5.36

a) $P(x = 4 | \lambda = 1.25)$

$$\frac{(1.25^4)(e^{-1.25})}{4!} = \frac{(2.4414)(.2865)}{24}$$ $= \textbf{.0291}$

b) $P(x \leq 1 | \lambda = 6.37) = P(x = 1) + P(x = 0) =$

$$\frac{(6.37)^1(e^{-6.37})}{1!} + \frac{(6.37^0)(e^{-6.37})}{0!} = \frac{(6.37)(.0017)}{1} + \frac{(1)(.0017)}{1}$$

$.0109 + .0017 = \textbf{.0126}$

c) $P(x > 5 | \lambda = 2.4) = P(x = 6) + P(x = 7) + ... =$

$$\frac{(2.4^6)(e^{-2.4})}{6!} + \frac{(2.4^7)(e^{-2.4})}{7!} + \frac{(2.4^8)(e^{-2.4})}{8!} + \frac{(2.4^9)(e^{-2.4})}{9!} + \frac{(2.4^{10})(e^{-2.4})}{10!} + \cdots$$

$.0241 + .0083 + .0025 + .0007 + .0002 = \textbf{.0358}$

for values $x \geq 11$ the probabilities are each .0000 when rounded off to 4

decimal places.

5.37 a) $P(x = 3|\lambda = 1.8) =$ **.1607**

b) $P(x < 5|\lambda = 3.3) =$

$P(x = 4) + P(x = 3) + P(x = 2) + P(x = 1) + P(x = 0) =$

x	Prob.
0	.0369
1	.1217
2	.2008
3	.2209
4	.1823
	.7626

c) $P(x \geq 3 | \lambda = 2.1) =$

x	Prob.
3	.1890
4	.0992
5	.0417
6	.0146
7	.0044
8	.0011
9	.0003
10	.0001
11	.0000
	.3504

d) $P(2 < x \leq 5 \mid \lambda = 4.2)$:

$P(x=3) + P(x=4) + P(x=5) =$

x	Prob.
3	.1852
4	.1944
5	.1633

.5429

5.38 a) $P(x = 3 \mid N = 6, n = 4, A = 5) = \dfrac{{}_5C_3 \cdot {}_1C_1}{{}_6C_4} = \dfrac{(10)(1)}{15} =$ **.6667**

b) $P(x \leq 1 \mid N = 10, n = 3, A = 5)$:

$P(x = 1) + P(x = 0) = \dfrac{{}_5C_1 \cdot {}_5C_2}{{}_{10}C_3} + \dfrac{{}_5C_0 \cdot {}_5C_3}{{}_{10}C_3} = \dfrac{(5)(10)}{120} + \dfrac{(1)(10)}{120}$

= .4167 + .0833 = **.5000**

c) $P(x \geq 2 \mid N = 13, n = 5, A = 3)$:

$P(x=2) + P(x=3)$ Note: only 3 x's in population

$\dfrac{{}_3C_2 \cdot {}_{10}C_3}{{}_{13}C_5} + \dfrac{{}_3C_3 \cdot {}_{10}C_2}{{}_{13}C_5} = \dfrac{(3)(120)}{1287} + \dfrac{(1)(45)}{1287}$ = .2797 + .0350 = **.3147**

5.39 $n = 25$ $p = .20$ retired

from Table A.2: $P(x = 7) =$ **.111**

$P(x \geq 10)$: $P(x = 10) + P(x = 11) + \ldots + P(x = 25) =$ .012 + .004 + .001 = **.017**

Expected Value = μ = $n{\cdot}p$ = 25(.20) = **5**

n = 20 p = .40 mutual funds

$P(x = 8)$ = .**180**

$P(x < 6) = P(x = 0) + P(x = 1) + . . . + P(x = 5)$ =

.000 + .000 + .003 +.012 + .035 + .075 = **.125**

$P(x = 0)$ = **.000**

$P(x \geq 12) = P(x = 12) + P(x = 13) + . . . + P(x = 20)$ = .035 + .015 + .005 + .001 = **.056**

x = **8**

Expected Number = μ = $n{\cdot} p$ = 20(.40) = **8**

5.40 λ = 3.2 cars|2 hours

a) $P(x$=3) cars per 1 hour) = ??

The interval has been decreased by ½.

The new λ = 1.6 cars | 1 hour.

$P(x = 3 | \lambda = 1.6)$ = (from Table A.3) **.1378**

b) $P(x = 0$ | cars per ½ hour) = ??

The interval has been decreased by ¼ the original amount.

The new λ = 0.8 cars | ½ hour.

$P(x = 0 | \lambda = 0.8)$ = (from Table A.3) **.4493**

c) $P(x \geq 5 | \lambda = 1.6)$ = (from Table A.3)

<u>x</u>	<u>Prob.</u>
5	.0176
6	.0047
7	.0011
8	<u>.0002</u>
	.0236

Either a rare event occurred or perhaps the long-run average, λ, has changed

(increased).

5.41 $N = 32$ A = 10 $n = 12$

a) $P(x = 3) = \dfrac{{}_{10}C_3 \cdot {}_{22}C_9}{{}_{32}C_{12}} = \dfrac{(120)(497,420)}{225,792,840} =$ **.2644**

b) $P(x = 6) = \dfrac{{}_{10}C_6 \cdot {}_{22}C_6}{{}_{32}C_{12}} = \dfrac{(210)(74,613)}{225,792,840} =$ **.0694**

c) $P(x = 0) = \dfrac{{}_{10}C_0 \cdot {}_{22}C_{12}}{{}_{32}C_{12}} = \dfrac{(1)(646,646)}{225,792,840} =$ **.0029**

d) $A = 22$

$$P(7 \leq x \leq 9) = \frac{{}_{22}C_7 \cdot {}_{10}C_5}{{}_{32}C_{12}} + \frac{{}_{22}C_8 \cdot {}_{10}C_4}{{}_{32}C_{12}} + \frac{{}_{22}C_9 \cdot {}_{10}C_3}{{}_{32}C_{12}}$$

$$= \frac{(170,544)(252)}{225,792,840} + \frac{(319,770)(210)}{225,792,840} + \frac{(497,420)(120)}{225,792,840}$$

= .1903 + .2974 + .2644 = **.7521**

5.42 $\lambda = 1.4$ defects$|$1 lot If $x > 3$, buyer rejects If $x \leq 3$, buyer accepts

$P(x \leq 3 | \lambda = 1.4)$ = (from Table A.3)

x	Prob.
0	.2466
1	.3452
2	.2417
3	.1128
	.9463

5.43 a) $n = 20$ and $p = .25$

The expected number = $\mu = n{\cdot}p = (20)(.25) =$ **5.00**

b) $P(x \leq 1 \mid n = 20 \text{ and } p = .25) =$

$P(x = 1) + P(x = 0) = {}_{20}C_1(.25)^1(.75)^{19} + {}_{20}C_0(.25)^0(.75)^{20}$

$= (20)(.25)(.00423) + (1)(1)(.0032) = .0212 +. 0032 =$ **.0244**

Since the probability is so low, the population of your state may have a lower

percentage of chronic heart conditions than those of other states.

5.44 a) $P(x > 7 \mid n = 10 \text{ and } p = .70)$ = (from Table A.2):

x	Prob.
8	.233
9	.121
10	.028
	.382

Expected number = $\mu = n \cdot p = 10(.70) =$ **7**

b) $n = 15$ $p = 1/3$ Expected number = $\mu = n \cdot p = 15(1/3) =$ **5**

$P(x=0 \mid n = 15 \text{ and } p = 1/3) =$

${}_{15}C_0(1/3)^0(2/3)^{15} =$ **.0023**

c) $n = 7$ $p = .53$

$P(x = 7 \mid n = 7 \text{ and } p = .53) = {}_7C_7(.53)^7(.47)^0 =$ **.0117**

Probably the 53% figure is too low for this population since the probability of

this occurrence is so low (.0117).

5.45 $n = 12$

a.) $P(x = 0$ long hours):

$p = .20$ ${}_{12}C_0(.20)^0(.80)^{12}$ = **.0687**

b.) $P(x \geq 6)$ long hours):

$p = .20$

Using Table A.2: .016 + .003 + .001 = **.020**

c) $P(x = 5$ good financing):

$p = .25$, ${}_{12}C_5(.25)^5(.75)^7$ = **.1032**

d.) $p = .19$ (good plan), expected number = $\mu = n(p) = 12(.19)$ = **2.28**

5.46 $n = 100,000$ $p = .000014$

Worked as a Poisson: $\lambda = n \cdot p = 100,000(.000014) = 1.4$

a) $P(x = 5)$:

from Table A.3 = **.0111**

b) $P(x = 0)$:

from Table A.3 = **.2466**

c) $P(x > 6)$: (from Table A.3)

x	Prob
7	.0005
8	.0001
	.0006

5.47 $P(x \leq 3) \mid n = 8$ and $p = .60$): From Table A.2:

$\underline{x}$	$\underline{\text{Prob.}}$
0	.001
1	.008
2	.041
3	$\underline{.124}$
	.174

17.4% of the time in a sample of eight, three or fewer customers are walk-ins by

chance. Other reasons for such a low number of walk-ins might be that she is

retaining more old customers than before or perhaps a new competitor is

attracting walk-ins away from her.

5.48 $n = 25$ $p = .20$

a) $P(x = 8 \mid n = 25$ and $p = .20)$ = (from Table A.2) **.062**

b) $P(x > 10) \mid n = 25$ and $p = .20)$ = (from Table A.2)

x	Prob.
11	.004
12	.001
13	.000
	.005

c) Since such a result would only occur 0.5% of the time by chance, it is likely

that the analyst's list was not representative of the entire state of Idaho or the

20% figure for the Idaho census is not correct.

5.49 λ = 0.6 flats | 2000 miles

$P(x = 0 \mid \lambda = 0.6)$ = (from Table A.3) **.5488**

$P(x \geq 3 \mid \lambda = 0.6)$ = (from Table A.3)

x	Prob.
3	.0198
4	.0030
5	.0004
	.0232

Assume one trip is independent of the other. Let F = flat tire and NF = no flat tire

$P(NF_1 \cap NF_2) = P(NF_1) \cdot P(NF_2)$

but $P(NF) = .5488$

$P(NF_1 \cap NF_2) = (.5488)(.5488) =$ **.3012**

5.50 $N = 25$ $n = 8$

a) $P(x = 1 \text{ in NY})$ $A = 4$

$$\frac{{}_4C_1 \cdot {}_{21}C_7}{{}_{25}C_8} = \frac{(4)(116,280)}{1,081,575} = \textbf{.4300}$$

b) $P(x = 4 \text{ in top } 10)$ $\quad A = 10$

$$\frac{{}_{10}C_4 \cdot {}_{15}C_4}{{}_{25}C_8} = \frac{(210(1365)}{1,081,575} = \textbf{.2650}$$

c) $P(x = 0 \text{ in California})$ $\quad A = 5$

$$\frac{{}_5C_0 \cdot {}_{20}C_8}{{}_{25}C_8} = \frac{(1)(125,970)}{1,081,575} = \textbf{.1165}$$

d) $P(x = 3 \text{ with M})$ $\quad A = 3$

$$\frac{{}_3C_3 \cdot {}_{22}C_5}{{}_{25}C_8} = \frac{(1)(26,334)}{1,081,575} = \textbf{.0243}$$

5.51 $\quad N = 24 \quad n = 6 \quad$ A = 8

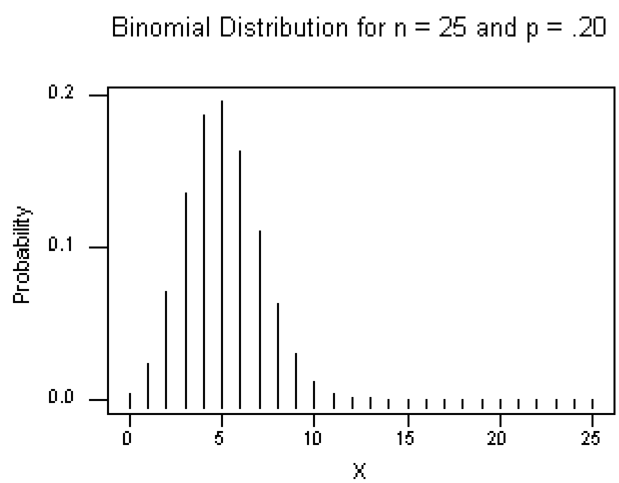
a) $P(x = 6) = \frac{{}_8C_6 \cdot {}_{16}C_0}{{}_{24}C_6} = \frac{(28)(1)}{134,596} = \textbf{.0002}$

b) $P(x = 0) = \dfrac{{}_{8}C_0 \cdot {}_{16}C_6}{{}_{24}C_6} = \dfrac{(1)(8008)}{134,596}$ = **.0595**

d) A = 16 East Side

$P(x = 3) = \dfrac{{}_{16}C_3 \cdot {}_{8}C_3}{{}_{24}C_6} = \dfrac{(560)(56)}{134,596}$ = **.2330**

5.52 $n = 25$ $p = .20$ Expected Value = $\mu = n \cdot p = 25(.20) =$ **5**



$\mu = 25(.20) =$ **5** $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{25(.20)(.80)}$ = **2**

$P(x > 12)$ = From Table A.2: $\underline{x}$

Prob

13	**.0000**

The values for $x > 12$ are so far away from the expected value that they are very

unlikely to occur.

$P(x = 14) = {}_{25}C_{14}(.20)^{14}(.80)^{11} =$ **.000063** which is very unlikely.

If this value ($x = 14$) actually occurred, one would doubt the validity of the

$p = .20$ figure or one would have experienced a very rare event.

5.53 $\lambda = 2.4$ calls | 1 minute

a) $P(x = 0|\lambda = 2.4) =$ (from Table A.3) **.0907**

b) Can handle $x \leq 5$ calls Cannot handle $x > 5$ calls

$P(x > 5|\lambda = 2.4) =$ (from Table A.3)

x	Prob.
6	.0241
7	.0083

8	.0025
9	.0007
10	.0002
11	.0000
	.0358

c) $P(x = 3 \text{ calls} | 2 \text{ minutes})$

The interval has been increased 2 times.

New Lambda: λ = 4.8 calls|2 minutes.

from Table A.3: **.1517**

d) $P(x \leq 1 \text{ calls} | 15 \text{ seconds})$:

The interval has been decreased by ¼.

New Lambda = λ = 0.6 calls|15 seconds.

$P(x \leq 1 | \lambda = 0.6)$ = (from Table A.3)

$P(x = 1)$ = .3293

$P(x = 0)$ = .5488

.8781

5.54 $n = 160$ $p = .01$

Working this problem as a Poisson problem:

a) Expected number = $\mu = n(p) = 160(.01) =$ **1.6**

b) $P(x \geq 8)$:

Using Table A.3:

x	Prob.
8	.0002
9	.0000
	.0002

c) $P(2 \leq x \leq 6)$:

Using Table A.3:

x	Prob.
2	.2584
3	.1378
4	.0551
5	.0176
6	.0047
	.4736

5.55 $p = .005$ $n = 1{,}000$

$\lambda = n \cdot p = (1{,}000)(.005) = 5$

a) $P(x < 4) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) =$

$.0067 + .0337 + .0842 + .1404 =$ **.265**

b) $P(x > 10) = P(x = 11) + P(x = 12) + \ldots =$

$.0082 + .0034 + .0013 + .0005 + .0002 =$ **.0136**

c) $P(x = 0) =$ **.0067**

5.56 $n = 8$ $p = .36$ $x = 0$ women

$_{8}C_{0}(.36)^{0}(.64)^{8} = (1)(1)(.0281475) =$ **.0281**

It is unlikely that a company would <u>randomly</u> hire 8 physicians from the U.S. pool

and none of them would be female. If this actually happened, figures similar to these

might be used as evidence in a lawsuit.

5.57 $N = 34$

a) $n = 5$ $x = 3$ $A = 13$

$$\frac{_{13}C_{3} \cdot {}_{21}C_{2}}{_{34}C_{5}} = \frac{(286)(210)}{278{,}256}$$

= **.2158**

b) $n = 8$ $x \le 2$ $A = 5$

$$\frac{_{5}C_{0} \cdot {}_{29}C_{8}}{_{34}C_{8}} + \frac{_{5}C_{1} \cdot {}_{29}C_{7}}{_{34}C_{8}} + \frac{_{5}C_{2} \cdot {}_{29}C_{6}}{_{34}C_{8}} =$$

$$\frac{(1)(4{,}292{,}145)}{18{,}156{,}204}+\frac{(5)(1{,}560{,}780)}{18{,}156{,}204}+\frac{(10)(475{,}020)}{18{,}156{,}204}$$

= .2364 + .4298 + .2616 = **.9278**

c) $n = 5$ $x = 2$ $A = 3$

${}_5C_2(3/34)^2(31/34)^3$ = (10)(.0077855)(.7579636) = **.0590**

5.58 $N = 14$ $n = 4$

a) $P(x = 4 | N = 14, n = 4, A = 10$ north side)

$$\frac{{}_{10}C_4 \cdot {}_4C_0}{{}_{14}C_4} = \frac{(210((1)}{1001}$$

= **.2098**

b) $P(x = 4 | N = 14, n = 4, A = 4$ west)

$$\frac{{}_4C_4 \cdot {}_{10}C_0}{{}_{14}C_4} = \frac{(1)(1)}{1001}$$

= **.0010**

c) $P(x = 2 | N = 14, n = 4, A = 4$ west)

$$\frac{{}_4C_2 \cdot {}_{10}C_2}{{}_{14}C_4} = \frac{(6)(45)}{1001}$$

= **.2697**

5.59 a) $\lambda = 3.84 | 1,000$

$$P(x = 0) = \frac{3.84^0 \cdot e^{-3.84}}{0!} = \mathbf{.0215}$$

b) $\lambda = 7.68 | 2{,}000$

$$P(x = 6) = \frac{7.68^6 \cdot e^{-7.68}}{6!} = \frac{(205{,}195.258)(.000461975)}{720} = \mathbf{.1317}$$

c) $\lambda = 1.6 | 1{,}000$ and $\lambda = 4.8 | 3{,}000$

from Table A.3:

$P(x < 7) = P(x = 0) + P(x = 1) + \ldots + P(x = 6) =$

.0082 + .0395 + .0948 + .1517 + .1820 + .1747 + .1398 = **.7907**

5.60 This is a binomial distribution with n = 15 and p = .36.

$\mu = n \cdot p = 15(.36) = 5.4$

$\sigma = \sqrt{15(.36)(.64)} = 1.86$

The most likely values are near the mean, 5.4. Note from the printout that the

most probable values are at x = 5 and x = 6 which are near the mean.

5.61 This printout contains the probabilities for various values of x from zero to eleven from a Poisson distribution with λ = 2.78. Note that the highest probabilities are at x = 2 and

x = 3 which are near the mean. The probability is slightly higher at x = 2 than at x = 3 even though x = 3 is nearer to the mean because of the "piling up" effect of x = 0.

5.62 This is a binomial distribution with n = 22 and p = .64.

The mean is $n \cdot p$ = 22(.64) = 14.08 and the standard deviation is:

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{22(.64)(.36)} = 2.25$$

The x value with the highest peak on the graph is at $x = 14$ followed by $x = 15$ and $x = 13$ which are nearest to the mean.

5.63 This is the graph of a Poisson Distribution with $\lambda = 1.784$. Note the high probabilities at $x = 1$ and $x = 2$ which are nearest to the mean. Note also that the probabilities for values of $x \geq 8$ are near to zero because they are so far away from the mean or expected value.

Chapter 6

Continuous Distributions

LEARNING OBJECTIVES

The primary objective of Chapter 6 is to help you understand continuous distributions, thereby enabling you to:

1. Understand concepts of the uniform distribution.
2. Appreciate the importance of the normal distribution.
3. Recognize normal distribution problems and know how to solve such problems.
4. Decide when to use the normal distribution to approximate binomial distribution problems and know how to work such problems.
5. Decide when to use the exponential distribution to solve problems in business and know how to work such problems.

CHAPTER TEACHING STRATEGY

Chapter 5 introduced the students to discrete distributions. This chapter introduces the students to three continuous distributions: the uniform distribution, the normal distribution and the exponential distribution. The normal distribution is probably the most widely known and used distribution. The text has been prepared with the notion that the student should be able to work many varied types of normal curve problems. Examples and practice problems are given wherein the student is asked to solve for virtually any of the four variables in the *z* equation. It is very helpful for the student to get into the habit of constructing a normal curve diagram, with a shaded portion for the desired area of concern for each problem using the normal distribution. Many students tend to be more visual learners than auditory and these diagrams will be of great assistance in problem demonstration and in problem solution.

This chapter contains a section dealing with the solution of binomial distribution problems by the normal curve. The correction for continuity is emphasized. In this text, the correction for continuity is always used whenever a binomial distribution problem is worked by the normal curve. Since this is often a stumbling block for students to comprehend, the chapter has included a table (Table 6.4) with rules of thumb as to how to apply the correction for continuity. It should be emphasized, however, that answers for this type of problem are still only approximations. For this reason and also in an effort to link chapters 5 & 6, the student is sometimes asked to work binomial problems both by methods in this chapter and also by using binomial tables (A.2). This also will allow the student to observe how good the approximation of the normal curve is to binomial problems.

The exponential distribution can be taught as a continuous distribution, which can be used in complement with the Poisson distribution of chapter 5 to solve inter-arrival time problems. The student can see that while the Poisson distribution is discrete because it describes the probabilities of whole number possibilities per some interval, the exponential distribution describes the probabilities associated with times that are continuously distributed.

CHAPTER OUTLINE

6.1 The Uniform Distribution

Determining Probabilities in a Uniform Distribution

Using the Computer to Solve for Uniform Distribution Probabilities

6.2 Normal Distribution

History of the Normal Distribution

Probability Density Function of the Normal Distribution

Standardized Normal Distribution

Solving Normal Curve Problems

Using the Computer to Solve for Normal Distribution Probabilities

6.3 Using the Normal Curve to Approximate Binomial Distribution Problems

Correcting for Continuity

6.4 Exponential Distribution

Probabilities of the Exponential Distribution

Using the Computer to Determine Exponential Distribution Probabilities

KEY TERMS

Correction for Continuity	Standardized Normal Distribution
Exponential Distribution	Uniform Distribution
Normal Distribution	*z* Distribution
Rectangular Distribution	*z* Score

SOLUTIONS TO PROBLEMS IN CHAPTER 6

6.1 a = 200 b = 240

a) $f(x) = \frac{1}{b-a} = \frac{1}{240-200} = \frac{1}{40}$ = **.025**

b) $\mu = \frac{a+b}{2} = \frac{200+240}{2}$ = **220**

$\sigma = \frac{b-a}{\sqrt{12}} = \frac{240-200}{\sqrt{12}} = \frac{40}{\sqrt{12}}$ = **11.547**

c) $P(x > 230) = \frac{240-230}{240-200} = \frac{10}{40}$ = **.250**

d) $P(205 \leq x \leq 220) = \frac{220-205}{240-200} = \frac{15}{40}$ = **.375**

e) $P(x \leq 225) = \frac{225-200}{240-200} = \frac{25}{40}$ = **.625**

6.2 $a = 8$ $b = 21$

a) $f(x) = \frac{1}{b-a} = \frac{1}{21-8} = \frac{1}{13}$ = **.0769**

b) $\mu = \frac{a+b}{2} = \frac{8+21}{2} = \frac{29}{2}$ = **14.5**

$\sigma = \frac{b-a}{\sqrt{12}} = \frac{21-8}{\sqrt{12}} = \frac{13}{\sqrt{12}}$ = **3.7528**

c) $P(10 \le x < 17) = \frac{17-10}{21-8} = \frac{7}{13}$ = **.5385**

d) $P(x > 22)$ = **.0000**

e) $P(x \ge 7)$ = **1.0000**

6.3 $a = 2.80$ $b = 3.14$

$\mu = \frac{a+b}{2} = \frac{2.80-3.14}{2}$ = **2.97**

$\sigma = \frac{b-a}{\sqrt{12}} = \frac{3.14-2.80}{\sqrt{12}}$ = **0.098**

$P(3.00 < x < 3.10) = \frac{3.10-3.00}{3.14-2.80}$ = **0.2941**

6.4 $a = 11.97$ $b = 12.03$

Height = $\frac{1}{b-a} = \frac{1}{12.03-11.97}$ = **16.667**

$P(x > 12.01) = \frac{12.03-12.01}{12.03-11.97}$ = **.3333**

$P(11.98 < x < 12.01) = \frac{12.01-11.98}{12.03-11.97}$ = **.5000**

6.5 $\mu = 2100$ $a = 400$ $b = 3800$

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{3800-400}{\sqrt{12}} = \mathbf{981.5}$$

$$\text{Height} = \frac{1}{b-a} = \frac{3800-400}{\sqrt{12}} = \mathbf{.000294}$$

$$P(x > 3000) = \frac{3800-3000}{3800-400} = \frac{800}{3400} = \mathbf{.2353}$$

$P(x > 4000)$ = **.0000**

$$P(700 < x < 1500) = \frac{1500-700}{3800-400} = \frac{800}{3400} = \mathbf{.2353}$$

6.6 a) $P(z \geq 1.96)$:

Table A.5 value for $z = 1.96$: .4750

$P(z \geq 1.96) = .5000 - .4750$ = **.0250**

b) $P\ (z < 0.73)$:

Table A.5 value for $z = 0.73$: .2673

$P(z < 0.73) = .5000 + .2673 =$ **.7673**

c) $P(-1.46 < z \leq 2.84)$:

Table A.5 value for $z = 2.84$: .4977

Table A.5 value for $z = 1.46$: .4279

$P(1.46 < z \leq 2.84) = .4977 + 4279 =$ **.9256**

d) $P(-2.67 < z \leq 1.08)$:

Table A.5 value for $z = -2.67$: .4962

Table A.5 value for $z = 1.08$: .3599

$P(-2.67 \leq z \leq 1.08) = .4962 + .3599 =$ **.8561**

e) $P\ (-2.05 < z \leq -.87)$:

Table A.5 value for $z = -2.05$: .4798

Table A.5 value for z = -0.87: .3078

$P(-2.05 < z \leq -.87) = .4798 - .3078 =$ **.1720**

6.7 a) $P(x \leq 635 | \mu = 604, \sigma = 56.8)$:

$$z = \frac{x-\mu}{\sigma} = \frac{635-604}{56.8} = 0.55$$

Table A.5 value for $z = 0.55$: .2088

$P(x \leq 635) = .2088 + .5000 =$ **.7088**

b) $P(x < 20 | \mu = 48, \sigma = 12)$:

$$z = \frac{x-\mu}{\sigma} = \frac{20-48}{12} = -2.33$$

Table A.5 value for $z = -2.33$: .4901

$P(x < 20) = .5000 - .4901 =$ **.0099**

c) $P(100 \leq x < 150 | \mu = 111, \sigma = 33.8)$:

$$z = \frac{x-\mu}{\sigma} = \frac{150-111}{33.8} = 1.15$$

Table A.5 value for $z = 1.15$: .3749

$z = \dfrac{x-\mu}{\sigma} = \dfrac{100-111}{33.8} = -0.33$

Table A.5 value for z = -0.33: .1293

$P(100 \le x < 150)$ = .3749 + .1293 = **.5042**

d) $P(250 < x < 255 \mid \mu = 264, \sigma = 10.9)$:

$z = \dfrac{x-\mu}{\sigma} = \dfrac{250-264}{10.9} = -1.28$

Table A.5 value for z = -1.28: .3997

$z = \dfrac{x-\mu}{\sigma} = \dfrac{255-264}{10.9} = -0.83$

Table A.5 value for z = -0.83: .2967

$P(250 < x < 255)$ = .3997 - .2967 = **.1030**

e) $P(x > 35 \mid \mu = 37, \sigma = 4.35)$:

$z = \dfrac{x-\mu}{\sigma} = \dfrac{35-37}{4.35} = -0.46$

Table A.5 value for z = -0.46: .1772

$P(x > 35)$ = .1772 + .5000 = **.6772**

f) $P(x \geq 170 \mid \mu = 156, \ \sigma = 11.4)$:

$$z = \frac{x-\mu}{\sigma} = \frac{170-156}{11.4} = 1.23$$

Table A.5 value for z = 1.23: .3907

$P(x \geq 170)$ = .5000 - .3907 = **.1093**

6.8 $\mu = 22$ $\sigma = 4$

a) $P(x > 17)$:

$$z = \frac{x-\mu}{\sigma} = \frac{17-22}{4} = -1.25$$

area between x = 17 and μ = 22 from table A.5 is .3944

$P(x > 17) = .3944 + .5000 = \mathbf{.8944}$

b) $P(x < 13)$:

$z = \frac{x-\mu}{\sigma} = \frac{13-22}{4} = -2.25$

from table A.5, area = .4878

$P(x < 13) = .5000 - .4878 = \mathbf{.0122}$

c) $P(25 \le x \le 31)$:

$$z = \frac{x-\mu}{\sigma} = \frac{31-22}{4} = 2.25$$

from table A.5, area = .4878

$$z = \frac{x-\mu}{\sigma} = \frac{25-22}{4} = 0.75$$

from table A.5, area = .2734

$P(25 \le x \le 31)$ = .4878 - .2734 = **.2144**

6.9 $\mu = 60$ $\sigma = 11.35$

a) $P(x > 85)$:

$$z = \frac{x-\mu}{\sigma} = \frac{85-60}{11.35} = 2.20$$

from Table A.5, the value for $z = 2.20$ is .4861

$P(x > 85) = .5000 - .4861 =$ **.0139**

b) $P(45 < x < 70)$:

$$z = \frac{x-\mu}{\sigma} = \frac{45-60}{11.35} = -1.32$$

$$z = \frac{x-\mu}{\sigma} = \frac{70-60}{11.35} = 0.88$$

from Table A.5, the value for $z = -1.32$ is .4066

and for $z = 0.88$ is .3106

$P(45 < x < 70) = .4066 + .3106 =$ **.7172**

c) $P(65 < x < 75)$:

$$z = \frac{x-\mu}{\sigma} = \frac{65-60}{11.35} = 0.44$$

$$z = \frac{x-\mu}{\sigma} = \frac{75-60}{11.35} = 1.32$$

from Table A.5, the value for z = 0.44 is .1700

from Table A.5, the value for z = 1.32 is .4066

$P(65 < x < 75)$ = .4066 - .1700 = **.2366**

d) $P(x \leq 40)$:

$$z = \frac{x-\mu}{\sigma} = \frac{40-60}{11.35} = -1.76$$

from Table A.5, the value for z = -1.76 is .4608

$P(x \leq 40)$ = .5000 - .4608 = **.0392**

6.10 μ = \$1332 σ = \$725

a) $P(x > \$2000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{2000-1332}{725} = 0.92$$

from Table A.5, the $z = 0.92$ yields: .3212

$P(x > \$2000) = .5000 - .3212 =$ **.1788**

b) $P(\text{owes money}) = P(x < 0)$:

$$z = \frac{x-\mu}{\sigma} = \frac{0-1332}{725} = -1.84$$

from Table A.5, the $z = -1.84$ yields: .4671

$P(x < 0) = .5000 - .4671 =$ **.0329**

c) $P(\$100 \le x \le \$700)$:

$$z = \frac{x-\mu}{\sigma} = \frac{100-1332}{725} = -1.70$$

from Table A.5, the $z = -1.70$ yields: .4554

$z = \dfrac{x-\mu}{\sigma} = \dfrac{700-1332}{725} = -0.87$

from Table A.5, the $z = -0.87$ yields: .3078

$P(\$100 \leq x \leq \$700) = .4554 - .3078 =$ **.1476**

6.11 $\mu = \$30{,}000$ $\sigma = \$9{,}000$

a) $P(\$15{,}000 \leq x \leq \$45{,}000)$:

$z = \dfrac{x-\mu}{\sigma} = \dfrac{45{,}000-30{,}000}{9{,}000} = 1.67$

From Table A.5, $z = 1.67$ yields: .4525

$z = \dfrac{x-\mu}{\sigma} = \dfrac{15{,}000-30{,}000}{9{,}000} = -1.67$

From Table A.5, $z = -1.67$ yields: .4525

$P(\$15{,}000 \leq x \leq \$45{,}000) = .4525 + .4525 =$ **.9050**

b) P(x > \$50,000):

$$z = \frac{x-\mu}{\sigma} = \frac{50{,}000-30{,}000}{9{,}000} = 2.22$$

From Table A.5, z = 2.22 yields: .4868

$P(x > \$50{,}000)$ = .5000 - .4868 = **.0132**

c) $P(\$5,000 \le x \le \$20,000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{5,000-30,000}{9,000} = -2.78$$

From Table A.5, z = -2.78 yields: .4973

$$z = \frac{x-\mu}{\sigma} = \frac{20,000-30,000}{9,000} = -1.11$$

From Table A.5, z = -1.11 yields .3665

$P(\$5,000 \le x \le \$20,000)$ = .4973 - .3665 = **.1308**

d) Since 90.82% of the values are greater than x = \$7,000, x = \$7,000 is in the

lower half of the distribution and .9082 - .5000 = .4082 lie between x and μ.

From Table A.5, z = -1.33 is associated with an area of .4082.

Solving for σ: $z = \dfrac{x-\mu}{\sigma}$

$$-1.33 = \frac{7{,}000-30{,}000}{\sigma}$$

σ = **17,293.23**

e) σ = \$9,000. If 79.95% of the costs are less than \$33,000, x = \$33,000 is in

the upper half of the distribution and .7995 - .5000 = .2995 of the values lie

between \$33,000 and the mean.

From Table A.5, an area of .2995 is associated with z = 0.84

Solving for μ: $z = \dfrac{x-\mu}{\sigma}$

$$0.84 = \frac{33{,}000-\mu}{9{,}000}$$

$\mu =$ **\$25,440**

6.12 μ = 200, σ = 47 Determine x

a) 60% of the values are greater than x:

Since 50% of the values are greater than the mean, μ = 200, 10% or .1000 lie

between x and the mean. From Table A.5, the z value associated with an area

of .1000 is z = -0.25. The z value is negative since x is below the mean.

Substituting z = -0.25, μ = 200, and σ = 47 into the formula and solving for x:

$$z = \frac{x-\mu}{\sigma}$$

$$-0.25 = \frac{x-200}{47}$$

x = **188.25**

b) x is less than 17% of the values.

Since x is only less than 17% of the values, 33% (.5000- .1700) or .3300 lie

between x and the mean. Table A.5 yields a z value of 0.95 for an area of

.3300. Using this z = 0.95, μ = 200, and σ = 47, x can be solved for:

$$z = \frac{x-\mu}{\sigma}$$

$$0.95 = \frac{x-200}{47}$$

x = **244.65**

c) 22% of the values are less than x.

Since 22% of the values lie below x, 28% lie between x and the mean

(.5000 - .2200). Table A.5 yields a z of -0.77 for an area of .2800. Using the z

value of -0.77, μ = 200, and σ = 47, x can be solved for:

$$z = \frac{x-\mu}{\sigma}$$

$$-0.77 = \frac{x-200}{47}$$

$x =$ **163.81**

d) x is greater than 55% of the values.

Since x is greater than 55% of the values, 5% (.0500) lie between x and the

mean. From Table A.5, a z value of 0.13 is associated with an area of .05.

Using $z = 0.13$, $\mu = 200$, and $\sigma = 47$, x can be solved for:

$$z = \frac{x-\mu}{\sigma}$$

$$0.13 = \frac{x-200}{47}$$

$x =$ **206.11**

6.13 σ = 625. If 73.89% of the values are greater than 1700, then 23.89% or .2389

lie between 1700 and the mean, μ. The z value associated with .2389 is -0.64

since the 1700 is below the mean.

Using z = -0.64, x = 1700, and σ = 625, μ can be solved for:

$$z = \frac{x-\mu}{\sigma}$$

$$-0.64 = \frac{1700-\mu}{625}$$

μ = **2100**

μ = 2258 and σ = 625. Since 31.56% are greater than x, 18.44% or .1844

(.5000 - .3156) lie between x and μ = 2258. From Table A.5, a z value of 0.48

is associated with .1844 area under the normal curve.

Using μ = 2258, σ = 625, and z = 0.48, x can be solved for:

$$z = \frac{x-\mu}{\sigma}$$

$0.48 = \dfrac{x - 2258}{625}$

x = **2558**

6.14 $\mu = 22$ σ = ??

Since 72.4% of the values are greater than 18.5, then 22.4% lie between 18.5 and μ. x = 18.5 is below the mean. From table A.5, z = - 0.59.

$-0.59 = \dfrac{18.5 - 22}{\sigma}$

$-0.59\sigma = -3.5$

$\sigma = \dfrac{3.5}{0.59}$ = **5.932**

6.15 $P(x < 20) = .2900$

x is less than μ because of the percentage. Between x and μ is .5000 - .2900 =

.2100 of the area. The z score associated with this area is -0.55.
Solving for μ:

$$z = \frac{x-\mu}{\sigma}$$

$$-0.55 = \frac{20-\mu}{4}$$

μ = **22.20**

6.16 $\mu = 9.7$ Since 22.45% are greater than 11.6, $x = 11.6$ is in the upper half of the distribution and .2755 (.5000 - .2245) lie between x and the mean. Table A.5 yields a $z = 0.76$ for an area of .2755.

Solving for σ:

$$z = \frac{x-\mu}{\sigma}$$

$$0.76 = \frac{11.6-9.7}{\sigma}$$

σ = **2.5**

6.17 a) $P(x \leq 16 \mid n = 30 \text{ and } p = .70)$

$\mu = n \cdot p = 30(.70) = 21$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{30(.70)(.30)} = 2.51$

$\boldsymbol{P(x \leq 16.5 \mid \mu = 21 \text{ and } \sigma = 2.51)}$

b) $P(10 < x \leq 20 \mid n = 25 \text{ and } p = .50)$

$\mu = n \cdot p = 25(.50) = 12.5$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{25(.50)(.50)} = 2.5$

$\boldsymbol{P(10.5 \leq x \leq 20.5 \mid \mu = 12.5 \text{ and } \sigma = 2.5)}$

c) $P(x = 22 \mid n = 40 \text{ and } p = .60)$

$\mu = n \cdot p = 40(.60) = 24$

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{40(.60)(.40)}$ = 3.10

***P*(21.5 $\leq$ *x* $\leq$ 22.5 | μ = 24 and σ = 3.10)**

d) *P*(*x* > 14 *n* = 16 and *p* = .45)

μ = *n*·*p* = 16(.45) = 7.2

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{16(.45)(.55)}$ = 1.99

***P*(*x* $\geq$ 14.5 | μ = 7.2 and σ = 1.99)**

6.18 a) n = 8 and p = .50 $\mu = n{\cdot}p = 8(.50) = 4$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{8(.50)(.50)} = 1.414$

$\mu \pm 3\sigma = 4 \pm 3(1.414) = 4 \pm 4.242$

(-0.242 to 8.242) does not lie between 0 and 8.

Do not use the normal distribution to approximate this problem.

b) n = 18 and p = .80 $\mu = n{\cdot}p = 18(.80) = 14.4$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{18(.80)(.20)} = 1.697$

$\mu \pm 3\sigma = 14.4 \pm 3(1.697) = 14.4 \pm 5.091$

(9.309 to 19.491) does not lie between 0 and 18.

Do not use the normal distribution to approximate this problem.

c) n = 12 and p = .30

$\mu = n{\cdot}p = 12(.30) = 3.6$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{12(.30)(.70)} = 1.587$

$\mu \pm 3\sigma = 3.6 \pm 3(1.587) = 3.6 \pm 4.761$

(-1.161 to 8.361) does <u>not</u> lie between 0 and 12.

Do <u>not</u> use the normal distribution to approximate this problem.

d) $n = 30$ and $p = .75$ $\mu = n \cdot p = 30(.75) = 22.5$

$\mu \pm 3\sigma = 22.5 \pm 3(2.37) = 22.5 \pm 7.11$

(15.39 to 29.61) does lie between 0 and 30.

The problem <u>can</u> be approximated by the normal curve.

e) $n = 14$ and $p = .50$ $\mu = n \cdot p = 14(.50) = 7$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{14(.50)(.50)} = 1.87$

$\mu \pm 3\sigma = 7 \pm 3(1.87) = 7 \pm 5.61$

(1.39 to 12.61) does lie between 0 and 14.

The problem <u>can</u> be approximated by the normal curve.

6.19 a) $P(x = 8 \mid n = 25 \text{ and } p = .40)$ $\mu = n \cdot p = 25(.40) = 10$

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{25(.40)(.60)}$ = 2.449

$\mu \pm 3\sigma$ = 10 ± 3(2.449) = 10 ± 7.347

(2.653 to 17.347) lies between 0 and 25.

Approximation by the normal curve is sufficient.

$P(7.5 \leq x \leq 8.5 | \mu = 10 \text{ and } \sigma = 2.449)$:

$z = \frac{7.5 - 10}{2.449} = -1.02$

From Table A.5, area = .3461

$z = \frac{8.5 - 10}{2.449} = -0.61$

From Table A.5, area = .2291

$P(7.5 \leq x \leq 8.5)$ = .3461 - .2291 = **.1170**

From Table A.2 (binomial tables) = **.120**

b) $P(x \geq 13 | n = 20 \text{ and } p = .60)$ $\mu = n \cdot p = 20(.60) = 12$

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{20(.60)(.40)}$ = 2.19

$\mu \pm 3\sigma$ = 12 ± 3(2.19) = 12 ± 6.57

(5.43 to 18.57) lies between 0 and 20.

Approximation by the normal curve is sufficient.

$P(x \geq 12.5 | \mu = 12 \text{ and } \sigma = 2.19)$:

z = $\frac{x-\mu}{\sigma} = \frac{12.5-12}{2.19}$ = 0.23

From Table A.5, area = .0910

$P(x \geq 12.5)$ = .5000 -.0910 = **.4090**

From Table A.2 (binomial tables) = **.415**

c) $P(x = 7 | n = 15 \text{ and } p = .50)$ $\mu = n \cdot p$ = 15(.50) = 7.5

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{15(.50)(.50)}$ = 1.9365

$\mu \pm 3\sigma$ = 7.5 ± 3(1.9365) = 7.5 ± 5.81

(1.69 to 13.31) lies between 0 and 15.

Approximation by the normal curve is sufficient.

$P(6.5 \le x \le 7.5 | \mu = 7.5 \text{ and } \sigma = 1.9365)$:

$z = \dfrac{x-\mu}{\sigma} = \dfrac{6.5-7.5}{1.9365} = -0.52$

From Table A.5, area = **.1985**

From Table A.2 (binomial tables) = **.196**

d) $P(x < 3 | n = 10 \text{ and } p = .70)$: $\mu = n \cdot p = 10(.70) = 7$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{10(.70)(.30)}$

$\mu \pm 3\sigma = 7 \pm 3(1.449) = 7 \pm 4.347$

(2.653 to 11.347) does not lie between 0 and 10.

The normal curve is not a good approximation to this problem.

6.20 $P(x < 40 | n = 120 \text{ and } p = .37)$: $\mu = n \cdot p = 120(.37) =$
44.4

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{120(.37)(.63)}$ = 5.29

$\mu \pm 3\sigma$ = 28.53 to 60.27 <u>does</u> lie between 0 and 120.

It is okay to use the normal distribution to approximate this problem

Correcting for continuity: x = 39.5

z = $\frac{39.5 - 44.4}{5.29}$ = -0.93

from Table A.5, the area of z = -0.93 is .3238

$P(x < 40)$ = .5000 - .3238 = .**1762**

6.21 n = 70, p = .59 $P(x < 35)$:

Converting to the normal dist.:

$\mu = n(p)$ = 70(.59) = 41.3 and σ = $\sqrt{n \cdot p \cdot q} = \sqrt{70(.59)(.41)}$ = 4.115

Test for normalcy:

$0 \le \mu \pm 3\sigma \le n$, $0 \le 41.3 \pm 3(4.115) \le 70$

0 < 28.955 to 53.645 < 70, passes the test

correction for continuity, use x = 34.5

z = $\frac{34.5 - 41.3}{4.115}$ = -1.65

from table A.5, area = .4505

$P(x < 35)$ = .5000 - .4505 = **.0495**

6.22 For parts a) and b), n = 300 p = .53

μ = 300(.53) = 159

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{300(.53)(.47)}$ = 8.645

Test: $\mu \pm 3\sigma$ = 159 $\pm$ 3(8.645) = 133.065 to 184.935

which lies between 0 and 300. It is okay to use the normal distribution as an approximation on parts a) and b).

a) $P(x > 175$ transmission)

correcting for continuity: x = 175.5

$z = \frac{175.5-159}{8.645} = 1.91$

from A.5, the area for z = 1.91 is .4719

$P(x > 175)$ = .5000 - .4719 = **.0281**

b) $P(165 \le x \le 170)$

correcting for continuity: x = 164.5; x = 170.5

z $= \frac{170.5-159}{8.645} = 1.33$ and $z = \frac{164.5-159}{8.645} = 0.64$

from A.5, the area for z = 1.33 is .4082

the area for z = 0.64 is .2389

$P(165 \le x \le 170)$ = .4082 - .2389 = .**1693**

For parts c) and d): n = 300 p = .60

$\mu = 300(.60) = 180$ $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{300(.60)(.40)} =$ 8.485

Test: $\mu \pm 3\sigma$ = 180 $\pm$ 3(8.485) = 180 $\pm$ 25.455

154.545 to 205.455 lies between 0 and 300

It is okay to use the normal distribution to approximate c) and d)

c) $P(155 \le x \le 170$ personnel):

correcting for continuity: x = 154.5; x = 170.5

$z = \frac{170.5-180}{8.485}$ = -1.12 and $z = \frac{154.5-180}{8.485}$ = -3.01

from A.5, the area for z = -1.12 is .3686

the area for z = -3.01 is .4987

$P(155 \le x \le 170)$ = .4987 - .3686 = **.1301**

d) $P(x < 200$ personnel):

correcting for continuity: x = 199.5

$z = \frac{199.5-180}{8.485}$ = 2.30

from A.5, the area for z = 2.30 is .4893

P(x < 200) = .5000 + .4893 = **.9893**

6.23 p = .25 n = 130

Conversion to normal dist.: $\mu = n(p) = 130(.25) = 32.5$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{130(.25)(.75)} = 4.94$

a) $P(x > 36)$:   Correct for continuity: $x = 36.5$

$z = \dfrac{36.5 - 32.5}{4.94} = 0.81$

from table A.5, area = .2910

$P(x > 20)$ = .5000 - .2910 = **.2090**

b) $P(26 \leq x \leq 35)$:   Correct for continuity: 25.5 to 35.5

$z = \dfrac{25.5 - 32.5}{4.94} = -1.42$ and $z = \dfrac{35.5 - 32.5}{4.94} = 0.61$

from table A.5, area for z = -1.42 is .4222

area for z = 0.61 is .2291

$P(26 \leq x \leq 35)$ = .4222 + .2291 = **.6513**

c) $P(x < 20)$: correct for continuity: $x = 19.5$

$$z = \frac{19.5 - 32.5}{4.94} = -2.63$$

from table A.5, area for $z = -2.63$ is .4957

$P(x < 20)$ = .5000 - .4957 = **.0043**

d) $P(x = 30)$: correct for continuity: 29.5 to 30.5

$$z = \frac{29.5 - 32.5}{4.94} = -0.61 \text{ and } z = \frac{30.5 - 32.5}{4.94} = -0.40$$

from table A.5, area for -0.61 = .2291

area for -0.40 = .1554

$P(x = 30)$ = .2291 - .1554 = **.0737**

6.24 $n = 95$

a) $P(44 \le x \le 52)$ agree with direct investments, $p = .52$

By the normal distribution: $\mu = n(p) = 95(.52) = 49.4$

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{95(.52)(.48)}$ = 4.87

test: $\mu \pm 3\sigma$ = 49.4 $\pm$ 3(4.87) = 49.4 $\pm$ 14.61

0 < 34.79 to 64.01 < 95 test passed

z = $\frac{43.5 - 49.4}{4.87}$ = -1.21

from table A.5, area = .3869

z = $\frac{52.5 - 49.4}{4.87}$ = 0.64

from table A.5, area = .2389

$P(44 \leq x \leq 52)$ = .3869 + .2389 = **.6258**

b) $P(x > 56)$:

correcting for continuity, x = 56.5

z = $\frac{56.5 - 49.4}{4.87}$ = 1.46

from table A.5, area = .4279

$P(x > 56)$ = .5000 - .4279 = **.0721**

c) Joint Venture:

p = .70, n = 95

By the normal dist.: $\mu = n(p) = 95(.70) = 66.5$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{95(.70)(.30)} = 4.47$

test for normalcy: $66.5 \pm 3(4.47) = 66.5 \pm 13.41$

0 < 53.09 to 79.91 < 95 test passed

$P(x < 60)$:

correcting for continuity: $x = 59.5$

$z = \dfrac{59.5 - 66.5}{4.47} = -1.57$

from table A.5, area = .4418

$P(x < 60) = .5000 - .4418 =$ **.0582**

d) $P(55 \leq x \leq 62)$:

correcting for continuity: 54.5 to 62.5

$$z = \frac{54.5 - 66.5}{4.47} = -2.68$$

from table A.5, area = .4963

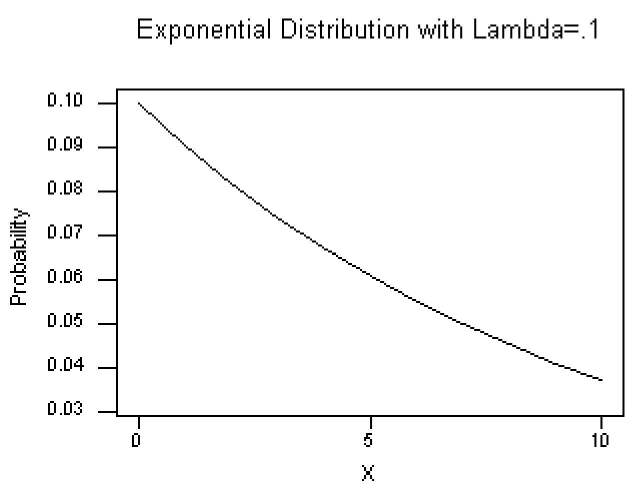
$$z = \frac{62.5 - 66.5}{4.47} = -0.89$$

from table A.5, area = .3133

$P(55 \leq x \leq 62) = .4963 - .3133 =$ **.1830**

6.25 a) λ = 0.1

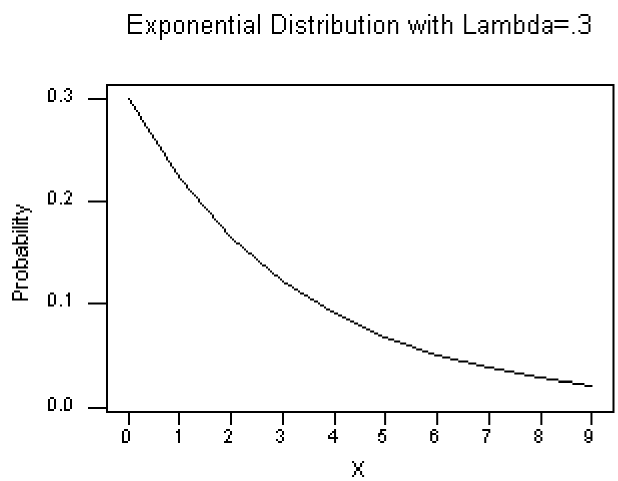
x_0	y
0	.1000
1	.0905
2	.0819
3	.0741
4	.0670
5	.0607
6	.0549
7	.0497
8	.0449
9	.0407
10	.0368



b) λ = 0.3

x_0	y
0	.3000

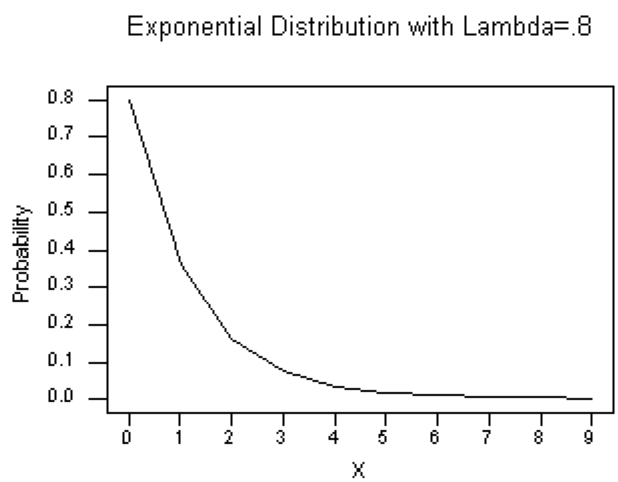
1	.2222
2	.1646
3	.1220
4	.0904
5	.0669
6	.0496
7	.0367
8	.0272
9	.0202



c) $\lambda = 0.8$

x_0	y
0	.8000
1	.3595
2	.1615
3	.0726

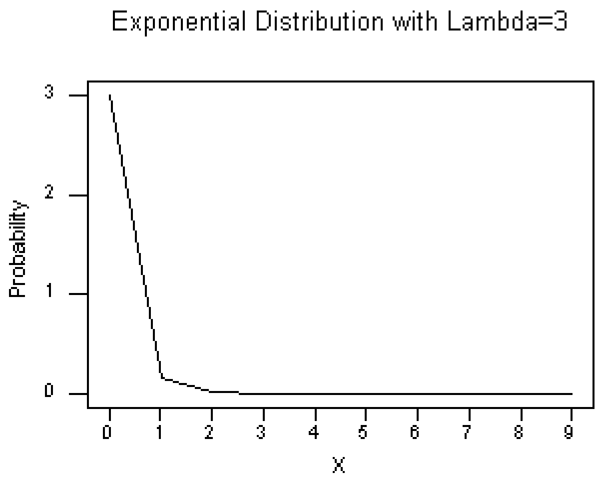
4	.0326
5	.0147
6	.0066
7	.0030
8	.0013
9	.0006



d) $\lambda = 3.0$

x_0	y
0	3.0000
1	.1494
2	.0074
3	.0004

4	.0000
5	.0000


Exponential Distribution with Lambda=3
Probability
X
0
1
2
3
4
5
6
7
8
9

6.26 a) λ = 3.25

$$\mu = \frac{1}{\lambda} = \frac{1}{3.25} = \mathbf{0.31}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{3.25} = \mathbf{0.31}$$

b) λ = 0.7

$$\mu = \frac{1}{\lambda} = \frac{1}{.007} = \mathbf{1.43}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{.007} = \mathbf{1.43}$$

c) λ = 1.1

$$\mu = \frac{1}{\lambda} = \frac{1}{1.1} = \mathbf{0.91}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{1.1} = \mathbf{0.91}$$

d) $\lambda = 6.0$

$$\mu = \frac{1}{\lambda} = \frac{1}{6} = \mathbf{0.17}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{6} = \mathbf{0.17}$$

6.27 a) $P(x \geq 5 \mid \lambda = 1.35) =$

for $x_0 = 5$: $P(x) = e^{-\lambda x} = e^{-1.35(5)} = e^{-6.75} =$ **.0012**

b) $P(x < 3 \mid \lambda = 0.68) = 1 - P(x \leq 3 \mid \lambda = .68) =$

for $x_0 = 3$: $1 – e^{-\lambda x} = 1 – e^{-0.68(3)} = 1 – e^{-2.04} = 1 - .1300 =$ **.8700**

c) $P(x > 4 \mid \lambda = 1.7) =$

for $x_0 = 4$: $P(x) = e^{-\lambda x} = e^{-1.7(4)} = e^{-6.8} =$ **.0011**

d) $P(x < 6 \mid \lambda = 0.80) = 1 - P(x \geq 6 \mid \lambda = 0.80) =$

for $x_0 = 6$: $P(x) = 1 – e^{-\lambda x} = 1 – e^{-0.80(6)} = 1 – e^{-4.8} = 1 - .0082$

$=$ **.9918**

6.28 $\mu = 23$ sec.

$\lambda = \frac{1}{\mu} = .0435$ per second

a) $P(x \geq 1 \text{ min} | \lambda = .0435/\text{sec.})$

Change λ to minutes: $\lambda = .0435(60) = 2.61$ min

$P(x \geq 1 \text{ min} | \lambda = 2.61/\text{min}) =$

for $x_0 = 1$: $P(x) = e^{-\lambda x} = e^{-2.61(1)} =$ **.0735**

b) $\lambda = .0435/\text{sec}$

Change λ to minutes: $\lambda = (.0435)(60) = 2.61$ min

$P(x \geq 3 | \lambda = 2.61/\text{min}) =$

for $x_0 = 3$: $P(x) = e^{-\lambda x} = e^{-2.61(3)} = e^{-7.83} =$ **.0004**

6.29 $\lambda = 2.44/\text{min.}$

a) $P(x \geq 10 \text{ min} | \lambda = 2.44/\text{min}) =$

Let $x_0 = 10$, $e^{-\lambda x} = e^{-2.44(10)} = e^{-24.4} =$ **.0000**

b) $P(x \geq 5 \text{ min} \mid \lambda = 2.44/\text{min}) =$

Let $x_0 = 5$, $e^{-\lambda x} = e^{-2.44(5)} = e^{-12.20} =$ **.0000**

c) $P(x \geq 1 \text{ min} \mid \lambda = 2.44/\text{min}) =$

Let $x_0 = 1$, $e^{-\lambda x} = e^{-2.44(1)} = e^{-2.44} =$ **.0872**

d) Expected time = $\mu = \frac{1}{\lambda} = \frac{1}{2.44}$ min. = **.41 min = 24.6 sec.**

6.30 $\lambda = 1.12$ planes/hr.

a) $\mu = \frac{1}{\lambda} = \frac{1}{1.12}$ = .89 hr. = **53.4 min**.

b) $P(x \geq 2 \text{ hrs} \mid \lambda = 1.12 \text{ planes/hr.}) =$

Let $x_0 = 2$, $e^{-\lambda x} = e^{-1.12(2)} = e^{-2.24} =$ **.1065**

c) $P(x < 10 \text{ min} \mid \lambda = 1.12/\text{hr.}) = 1 - P(x \geq 10 \text{ min} \mid \lambda = 1.12/\text{hr.})$

Change λ to 1.12/60 min. = .01867/min.

$1 - P(x \geq 10 \text{ min} \mid \lambda = .01867/\text{min}) =$

Let $x_0 = 10$, $1 - e^{-\lambda x} = 1 - e^{-.01867(10)} = 1 - e^{-.1867} = 1 - .8297 =$ **.1703**

6.31 λ = 3.39/ 1000 passengers

$\mu = \frac{1}{\lambda} = \frac{1}{3.39} = 0.295$

(0.295)(1,000) = **295**

$P(x > 500)$:

Let x_0 = 500/1,000 passengers = .5

$e^{-\lambda x} = e^{-3.39(.5)} = e^{-1.695} =$ **.1836**

$P(x < 200)$:

Let x_0 = 200/1,000 passengers = .2

$e^{-\lambda x} = e^{-3.39(.2)} = e^{-.678} = .5076$

$P(x < 200) = 1 - .5076 =$ **.4924**

6.32 μ = 20 years

$\lambda = \frac{1}{20} = .05$/year

x_0	Prob$(x > x_0)=e^{-\lambda x}$
1	.9512
2	.9048
3	.8607

If the foundation is guaranteed for 2 years, based on past history, 90.48% of the

foundations will last at least 2 years without major repair and only 9.52% will

require a major repair before 2 years.

6.33 λ = 2/month

Average number of time between rain = μ = $\frac{1}{\lambda} = \frac{1}{2}$ month = **15 days**

σ = μ = **15 days**

$P(x \leq$ 2 days$|$ λ = 2/month$)$:

Change λ to days: λ = $\frac{2}{30}$ = .067/day

$P(x \leq$ 2 days$|$ λ = .067/day$)$ =

1 – $P(x >$ 2 days$|$ λ = .067/day$)$

let x_0 = 2, 1 – $e^{-\lambda x}$ = 1 – $e^{-.067(2)}$ = 1 – .8746 = **.1254**

6.34 a = 6 b = 14

$f(x)$ = $\frac{1}{b-a} = \frac{1}{14-6} = \frac{1}{8}$ = **.125**

μ = $\frac{a+b}{2} = \frac{6+14}{2}$ = **10**

σ = $\frac{b-a}{\sqrt{12}} = \frac{14-6}{\sqrt{12}} = \frac{8}{\sqrt{12}}$ = **2.309**

$P(x > 11)$ = $\frac{14-11}{14-6} = \frac{3}{8}$ = **.375**

$P(7 < x < 12)$ = $\frac{12-7}{14-6} = \frac{5}{8}$ = **.625**

6.35 a) $P(x < 21 \mid \mu = 25$ and $\sigma = 4)$:

z = $\frac{x-\mu}{\sigma} = \frac{21-25}{4}$ = -1.00

From Table A.5, area = .3413

$P(x < 21)$ = .5000 -.3413 = **.1587**

b) $P(x \geq 77 \mid \mu = 50 \text{ and } \sigma = 9)$:

$$z = \frac{x-\mu}{\sigma} = \frac{77-50}{9} = 3.00$$

From Table A.5, area = .4987

$P(x \geq 77)$ = .5000 -.4987 = **.0013**

c) $P(x > 47 \mid \mu = 50 \text{ and } \sigma = 6)$:

$$z = \frac{x-\mu}{\sigma} = \frac{47-50}{6} = -0.50$$

From Table A.5, area = .1915

$P(x > 47)$ = .5000 + .1915 = **.6915**

d) $P(13 < x < 29 \mid \mu = 23 \text{ and } \sigma = 4)$:

$$z = \frac{x-\mu}{\sigma} = \frac{13-23}{4} = -2.50$$

From Table A.5, area = .4938

$z = \frac{x-\mu}{\sigma} = \frac{29-23}{4} = 1.50$

From Table A.5, area = .4332

$P(13 < x < 29)$ = .4938 + 4332 = **.9270**

e) $P(x \geq 105 | \mu = 90 \text{ and } \sigma = 2.86)$:

$z = \frac{x-\mu}{\sigma} = \frac{105-90}{2.86} = 5.24$

From Table A.5, area = .5000

$P(x \geq 105)$ = .5000 - .5000 = **.0000**

6.36 a) $P(x = 12 \mid n = 25 \text{ and } p = .60)$:

$\mu = n \cdot p = 25(.60) = 15$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{25(.60)(.40)} = 2.45$

$\mu \pm 3\sigma = 15 \pm 3(2.45) = 15 \pm 7.35$

(7.65 to 22.35) lies between 0 and 25.

The normal curve approximation is sufficient.

$P(11.5 \le x \le 12.5 \mid \mu = 15 \text{ and } \sigma = 2.45)$:

$z = \frac{x-\mu}{\sigma} = \frac{11.5-15}{2.45} = -1.43$ From Table A.5, area = .4236

$z = \frac{x-\mu}{\sigma} = \frac{12.5-15}{2.45} = -1.02$ From Table A.5, area = .3461

$P(11.5 \le x \le 12.5)$ = .4236 - .3461 = **.0775**

From Table A.2, $P(x = 12)$ = **.076**

b) $P(x > 5 \mid n = 15 \text{ and } p = .50)$:

μ = $n \cdot p$ = 15(.50) = 7.5

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{15(.50)(.50)}$ = 1.94

$\mu \pm 3\sigma$ = 7.5 ± 3(1.94) = 7.5 ± 5.82

(1.68 to 13.32) lies between 0 and 15.

The normal curve approximation is sufficient.

$P(x \geq 5.5 | \mu = 7.5 \text{ and } = \text{l}.94)$

z = $\frac{5.5 - 7.5}{1.94}$ = -1.03

From Table A.5, area = .3485

$P(x \geq 5.5)$ = .5000 + .3485 = **.8485**

Using table A.2, $P(x > 5)$ = **.849**

c) $P(x \leq 3 | n = 10 \text{ and } p = .50)$:

$\mu = n \cdot p = 10(.50) = 5$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{10(.50)(.50)} = 1.58$

$\mu \pm 3\sigma = 5 \pm 3(1.58) = 5 \pm 4.74$

(0.26 to 9.74) lies between 0 and 10.

The normal curve approximation is sufficient.

$P(x \leq 3.5 | \mu = 5 \text{ and } \sigma = \text{l}.58)$:

$z = \frac{3.5 - 5}{1.58} = -0.95$

From Table A.5, area = .3289

$P(x \leq 3.5) = .5000 - .3289 =$ **.1711**

d) $P(x \geq 8 | n = 15 \text{ and } p = .40)$:

$\mu = n \cdot p = 15(.40) = 6$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{15(.40)(.60)} = 1.90$

$\mu \pm 3\sigma = 6 \pm 3(1.90) = 6 \pm 5.7$

(0.3 to 11.7) lies between 0 and 15.

The normal curve approximation is sufficient.

$P(x \geq 7.5 \mid \mu = 6 \text{ and } \sigma = 1.9)$:

$z = \frac{7.5-6}{1.9} = 0.79$

From Table A.5, area = .2852

$P(x \geq 7.5)$ = .5000 - .2852 = **.2148**

6.37 a) $P(x \geq 3 | \lambda = 1.3)$:

let $x_0 = 3$

$P(x \geq 3 | \lambda = 1.3) = e^{-\lambda x} = e^{-1.3(3)} = e^{-3.9} =$ **.0202**

b) $P(x < 2 | \lambda = 2.0)$:

Let $x_0 = 2$

$P(x < 2 | \lambda = 2.0) = 1 - P(x \geq 2 | \lambda = 2.0) =$

$1 – e^{-\lambda x} = 1 – e^{-2(2)} = 1 – e^{-4} = 1 - .0183 =$ **.9817**

c) $P(1 \leq x \leq 3 | \lambda = 1.65)$:

$P(x \geq 1 | \lambda = 1.65)$:

Let $x_0 = 1$

$e^{-\lambda x} = e^{-1.65(1)} = e^{-1.65} = .1920$

$P(x \geq 3 | \lambda = 1.65)$:

Let $x_0 = 3$

$e^{-\lambda x} = e^{-1.65(3)} = e^{-4.95} = .0071$

$P(1 \le x \le 3) = P(x \ge 1) - P(x \ge 3) = .1920 - .0071 =$ **.1849**

d) $P(x > 2|\lambda = 0.405)$:

Let $x_0 = 2$

$e^{-\lambda x} = e^{-(.405)(2)} = e^{-.81} =$ **.4449**

6.38 $\mu = 43.4$

12% more than 48. $x = 48$

Area between x and μ is .50 - .12 = .38

z associated with an area of .3800 is $z = 1.175$

Solving for σ:

$$z = \frac{x-\mu}{\sigma}$$

$$1.175 = \frac{48-43.4}{\sigma}$$

$$\sigma = \frac{4.6}{1.175} = 3.915$$

6.39 $p = 1/5 = .20$ $n = 150$

$P(x > 50)$:

$\mu = 150(.20) = 30$

$\sigma = \sqrt{150(.20)(.80)} = 4.899$

$z = \frac{50.5 - 30}{4.899} = 4.18$

Area associated with $z = 4.18$ is .5000

$P(x > 50) = .5000 - .5000 =$ **.0000**

6.40 λ = 1 customer/20 minutes

$\mu = 1/\lambda = 1$

a) 1 hour interval

x_0 = 3 because 1 hour = 3(20 minute intervals)

$P(x \geq x_0) = e^{-\lambda x} = e^{-1(3)} = e^{-3} =$ **.0498**

b) 10 to 30 minutes

$x_0 = .5, \quad x_0 = 1.5$

$P(x \geq .5) = e^{-\lambda x} = e^{-1(.5)} = e^{-.5} = .6065$

$P(x \geq 1.5) = e^{-\lambda x} = e^{-1(1.5)} = e^{-1.5} = .2231$

P(10 to 30 minutes) = .6065 - .2231 = **.3834**

c) less than 5 minutes

$x_0 = 5/20 = .25$

$P(x \geq .25) = e^{-\lambda x} = e^{-1(.25)} = e^{-.25} = .7788$

$P(x < .25)$ = 1 - .7788 = **.2212**

6.41 μ = 90.28 σ = 8.53

$P(x < 80)$:

$z = \frac{80-90.28}{8.53} = -1.21$

from Table A.5, area for z = -1.21 is .3869

$P(x < 80)$ = .5000 - .3869 = .**1131**

$P(x > 95)$:

$z = \frac{95-90.28}{8.53} = 0.55$

from Table A.5, area for z = 0.55 is .2088

$P(x > 95)$ = .5000 - .2088 = **.2912**

$P(83 < x < 87)$:

$z = \frac{83 - 90.28}{8.53} = -0.85$

$z = \frac{87 - 90.28}{8.53} = -0.38$

from Table A.5, area for z = -0.85 is .3023

area for z = -0.38 is .1480

$P(83 < x < 87)$ = .3023 - .1480 = **.1543**

6.42 $\sigma = 83$

Since only 3% = .0300 of the values are greater than 2,655(million), x = 2655 lies in the upper tail of the distribution. .5000 - .0300 = .4700 of the values lie between 2655 and the mean.

Table A.5 yields a z = 1.88 for an area of .4700.

Using z = 1.88, x = 2655, σ = 83, μ can be solved for.

$z = \frac{x - \mu}{\sigma}$

$$1.88 = \frac{2655 - \mu}{83}$$

μ = **2498.96 million**

6.43 $a = 18$ $b = 65$

$P(25 < x < 50) = \frac{50-25}{65-18} = \frac{25}{47} =$.**5319**

$\mu = \frac{a+b}{2} = \frac{65+18}{2} =$ **41.5**

$f(x) = \frac{1}{b-a} = \frac{1}{65-18} = \frac{1}{47} =$ **.0213**

6.44 λ = 1.8 per 15 seconds

a) $\mu = \frac{1}{\lambda} = \frac{1}{1.8} =$ .5556(15 sec.) = **8.33 sec.**

b) For $x_0 > 25$ sec. use $x_0 = 25/15 = 1.67$

$P(x_0 > 25 \text{ sec.}) = e^{-1.6667(1.8)} =$ **.0498**

c) $x_0 < 5$ sec. = 1/3

$P(x_0 < 5 \text{ sec.}) = 1 - e^{-1/3(1.8)} = 1 - .5488 =$ **.4512**

d) $P(x_0 \geq 1 \text{ min.})$:

$x_0 = 1 \text{ min.} = 60/15 = 4$

$P(x_0 \geq 1 \text{ min.}) = e^{-4(1.8)} =$ **.0007**

6.45 $\mu = 951$ $\sigma = 96$

a) $P(x \geq 1000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{1000-951}{96} = 0.51$$

from Table A.5, the area for $z = 0.51$ is .1950

$P(x \geq 1000) = .5000 - .1950 =$.**3050**

b) $P(900 < x < 1100)$:

$$z = \frac{x-\mu}{\sigma} = \frac{900-951}{96} = -0.53$$

$$z = \frac{x-\mu}{\sigma} = \frac{1100-951}{96} = 1.55$$

from Table A.5, the area for $z = -0.53$ is .2019

the area for $z = 1.55$ is .4394

$P(900 < x < 1100) = .2019 + .4394 =$.**6413**

c) $P(825 < x < 925)$:

$$z = \frac{x-\mu}{\sigma} = \frac{825-951}{96} = -1.31$$

$$z = \frac{x-\mu}{\sigma} = \frac{925-951}{96} = -0.27$$

from Table A.5, the area for $z = -1.31$ is .4049

the area for $z = -0.27$ is .1064

$P(825 < x < 925) = .4049 - .1064 =$.**2985**

d) $P(x < 700)$:

$$z = \frac{x-\mu}{\sigma} = \frac{700-951}{96} = -2.61$$

from Table A.5, the area for $z = -2.61$ is .4955

$P(x < 700) = .5000 - .4955 =$ **.0045**

6.46 $n = 60$ $p = .24$

$\mu = n \cdot p = 60(.24) = 14.4$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{60(.24)(.76)} = 3.308$

test: $\mu \pm 3\sigma = 14.4 \pm 3(3.308) = 14.4 \pm 9.924 =$ 4.476 and 24.324

Since 4.476 to 24.324 lies between 0 and 60, the normal distribution can be used

to approximate this problem.

$P(x \geq 17)$:

correcting for continuity: $x = 16.5$

$z = \frac{x-\mu}{\sigma} = \frac{16.5-14.4}{3.308} = 0.63$

from Table A.5, the area for $z = 0.63$ is .2357

$P(x \geq 17) = .5000 - .2357 =$ **.2643**

$P(x > 22)$:

correcting for continuity: x = 22.5

$$z = \frac{x-\mu}{\sigma} = \frac{22.5-14.4}{3.308} = 2.45$$

from Table A.5, the area for z = 2.45 is .4929

$P(x > 22)$ = .5000 - .4929 = **.0071**

$P(8 \leq x \leq 12)$:

correcting for continuity: x = 7.5 and x = 12.5

$$z = \frac{x-\mu}{\sigma} = \frac{12.5-14.4}{3.308} = -0.57 \qquad z = \frac{x-\mu}{\sigma} = \frac{7.5-14.4}{3.308} = -2.09$$

from Table A.5, the area for z = -0.57 is .2157

the area for z = -2.09 is .4817

$P(8 \leq x \leq 12)$ = .4817 - .2157 = **.2660**

6.47 μ = 45,970 σ = 4,246

a) $P(x > 50{,}000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{50{,}000 - 45{,}970}{4{,}246} = 0.95$$

from Table A.5, the area for $z = 0.95$ is .3289

$P(x > 50{,}000) = .5000 - .3289 =$ **.1711**

b) $P(x < 40{,}000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{40{,}000 - 45{,}970}{4{,}246} = -1.41$$

from Table A.5, the area for $z = -1.41$ is .4207

$P(x < 40{,}000) = .5000 - .4207 =$ **.0793**

c) $P(x > 35{,}000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{35{,}000 - 45{,}970}{4{,}246} = -2.58$$

from Table A.5, the area for $z = -2.58$ is .4951

$P(x > 35{,}000) = .5000 + .4951 =$ **.9951**

d) $P(39{,}000 < x < 47{,}000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{39{,}000-45{,}970}{4{,}246} = -1.64$$

$$z = \frac{x-\mu}{\sigma} = \frac{47{,}000-45{,}970}{4{,}246} = 0.24$$

from Table A.5, the area for $z = -1.64$ is .4495

the area for $z = 0.24$ is .0948

$P(39{,}000 < x < 47{,}000) = .4495 + .0948 =$ **.5443**

6.48 μ = 9 minutes

$\lambda = 1/\mu$ = .1111/minute = .1111(60)/hour

λ = **6.67/hour**

$P(x \geq 5$ minutes$\mid \lambda$ = .1111/minute) =

1 - $P(x \geq 5$ minutes$\mid \lambda$ =.1111/minute):

Let $x_0 = 5$

$P(x \geq 5$ minutes$\mid \lambda$ = .1111/minute) =

$e^{-\lambda x} = e^{-.1111(5)} = e^{-.5555} = .5738$

$P(x < 5$ minutes) = 1 - $P(x \geq 5$ minutes) = 1 - .5738 = **.4262**

6.49 $\mu = 88$ $\sigma = 6.4$

a) $P(x < 70)$:

$$z = \frac{x-\mu}{\sigma} = \frac{70-88}{6.4} = -2.81$$

From Table A.5, area = .4975

$P(x < 70)$ = .5000 - .4975 = **.0025**

b) $P(x > 80)$:

$z = \frac{x-\mu}{\sigma} = \frac{80-88}{6.4} = -1.25$

From Table A.5, area = .3944

$P(x > 80)$ = .5000 + .3944 = **.8944**

c) $P(90 \leq x \leq 100)$:

$z = \frac{x-\mu}{\sigma} = \frac{100-88}{6.4} = 1.88$

From Table A.5, area = .4699

$z = \frac{x-\mu}{\sigma} = \frac{90-88}{6.4} = 0.31$

From Table A.5, area = .1217

$P(90 \le x \le 100)$ = .4699 - .1217 = **.3482**

6.50 n = 200, p = .81

expected number = μ = $n(p)$ = 200(.81) = **162**

μ = 162

σ = $\sqrt{n \cdot p \cdot q} = \sqrt{200(.81)(.19)}$ = 5.548

$\mu \pm 3\sigma$ = 162 $\pm$ 3(5.548) lie between 0 and 200, the normalcy test is passed

$P(150 < x < 155)$:

correction for continuity: 150.5 to 154.5

z = $\dfrac{150.5-162}{5.548}$ = -2.07

from table A.5, area = .4808

$z = \frac{154.5 - 162}{5.548} = -1.35$

from table A.5, area = .4115

$P(150 < x < 155) = .4808 - .4115 =$ **.0693**

$P(x > 158)$:

correcting for continuity, $x = 158.5$

$z = \frac{158.5 - 162}{5.548} = -0.63$

from table A.5, area = .2357

$P(x > 158) = .2357 + .5000 =$ **.7357**

$P(x < 144)$:

correcting for continuity, $x = 143.5$

$z = \frac{143.5 - 162}{5.548} = -3.33$

from table A.5, area = .4996

$P(x < 144)$ = .5000 - .4996 = **.0004**

6.51 $n = 150$ $p = .75$

$\mu = n \cdot p = 150(.75) = 112.5$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{150(.75)(.25)} = 5.3033$

a) $P(x < 105)$:

correcting for continuity: $x = 104.5$

$z = \dfrac{x-\mu}{\sigma} = \dfrac{104.5-112.5}{5.3033} = -1.51$

from Table A.5, the area for $z = -1.51$ is .4345

$P(x < 105)$ = .5000 - .4345 = **.0655**

b) $P(110 \le x \le 120)$:

correcting for continuity: $x = 109.5$, $x = 120.5$

$$z = \frac{109.5 - 112.5}{5.3033} = -0.57$$

$$z = \frac{120.5 - 112.5}{5.3033} = 1.51$$

from Table A.5, the area for $z = -0.57$ is .2157

the area for $z = 1.51$ is .4345

$P(110 \leq x \leq 120) = .2157 + .4345 =$ **.6502**

c) $P(x > 95)$:

correcting for continuity: $x = 95.5$

$$z = \frac{95.5 - 112.5}{5.3033} = -3.21$$

from Table A.5, the area for -3.21 is .4993

$P(x > 95) = .5000 + .4993 =$ **.9993**

6.52 $\mu = \frac{a+b}{2} = 2.165$

$a + b = 2(2.165) = 4.33$

$b = 4.33 - a$

Height $= \frac{1}{b-a} = 0.862$

$1 = 0.862b - 0.862a$

Substituting b from above, $1 = 0.862(4.33 - a) - 0.862a$

$1 = 3.73246 - 0.862a - 0.862a$

$1 = 3.73246 - 1.724a$

$1.724a = 2.73246$

$a =$ **1.585** and $b = 4.33 - 1.585 =$ **2.745**

6.53 $\mu = 85{,}200$

60% are between 75,600 and 94,800

94,800 –85,200 = 9,600

75,600 – 85,200 = 9,600

The 60% can be split into 30% and 30% because the two *x* values are equal distance from the mean.

The *z* value associated with .3000 area is 0.84

$z = \dfrac{x-\mu}{\sigma}$

$.84 = \dfrac{94{,}800-85{,}200}{\sigma}$

σ = **11,428.57**

6.54 n = 75 p = .81 prices p = .44 products

a) Expected Value: $\mu_1 = n \cdot p = 75(.81) = 60.75$ seeking price information

$\sigma_1 = \sqrt{n \cdot p \cdot q} = \sqrt{75(.81)(.19)} = 3.3974$

b) Expected Value: $\mu_2 = n \cdot p = 75(.44) = 33$

σ_2 = $\sqrt{n \cdot p \cdot q} = \sqrt{75(.44)(.56)}$ = 4.2988

Tests: $\mu \pm 3\sigma$ = 60.75 $\pm$ 3(3.397) = 60.75 $\pm$ 10.191

50.559 to 70.941 lies between 0 and 75. It is okay to use the normal

distribution to approximate this problem.

$\mu \pm 3\sigma$ = 33 $\pm$ 3(4.299) = 33 $\pm$ 12.897

20.103 to 45.897 lies between 0 and 75. It is okay to use the normal

distribution to approximate this problem.

c) $P(x \geq 67$ prices)

correcting for continuity: x = 66.5

z = $\frac{66.5 - 60.75}{3.3974}$ = 1.69

from Table A.5, the area for z = 1.69 is .4545

$P(x \geq 67$ prices) = .5000 - .4545 = **.0455**

d) $P(x < 23$ products):

correcting for continuity: $x = 22.5$

$$z = \frac{22.5-33}{4.2988} = -2.44$$

from Table A.5, the area for $z = -2.44$ is .4927

$P(x < 23) = .5000 - .4927 =$ **.0073**

6.55 $\lambda = 3$ hurricanes$|$5 months

$P(x \geq 1$ month$|\ \lambda = 3$ hurricanes per 5 months):

Since x and λ are for different intervals,

change Lambda = $\lambda = 3/$ 5 months = 0.6 month.

$P(x \geq$ month$|\ \lambda = 0.6$ per month):

Let $x_0 = 1$

$P(x \geq 1) = e^{-\lambda x} = e^{-0.6(1)} = e^{-0.6} =$ **.5488**

$P(x \leq 2 \text{ weeks})$: 2 weeks = 0.5 month.

$P(x \leq 0.5 \text{ month} \mid \lambda = 0.6 \text{ per month}) =$

$1 - P(x > 0.5 \text{ month} \mid \lambda = 0.6 \text{ per month})$

But $P(x > 0.5 \text{ month} \mid \lambda = 0.6 \text{ per month})$:

Let $x_0 = 0.5$

$P(x > 0.5) = e^{-\lambda x} = e^{-0.6(.5)} = e^{-0.30} = .7408$

$P(x \leq 0.5 \text{ month}) = 1 - P(x > 0.5 \text{ month}) = 1 - .7408 =$ **.2592**

Average time = Expected time = $\mu = 1/\lambda =$ **1.67 months**

6.56 $n = 50$ $p = .80$

$\mu = n \cdot p = 50(.80) = 40$

$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{50(.80)(.20)} = 2.828$

Test: $\mu \pm 3\sigma = 40 \pm 3(2.828) = 40 \pm 8.484$

31.516 to 48.484 lies between 0 and 50.

It is okay to use the normal distribution to approximate this binomial problem.

$P(x < 35)$: correcting for continuity: $x = 34.5$

$$z = \frac{34.5 - 40}{2.828} = -1.94$$

from Table A.5, the area for $z = -1.94$ is .4738

$P(x < 35) = .5000 - .4738 =$ **.0262**

The expected value = μ = **40**

$P(42 \le x \le 47)$:

correction for continuity: $x = 41.5$ $x = 47.5$

$$z = \frac{41.5 - 40}{2.828} = 0.53 \qquad z = \frac{47.5 - 40}{2.828} = 2.65$$

from Table A.5, the area for $z = 0.53$ is .2019

the area for $z = 2.65$ is .4960

$P(42 \leq x \leq 47) = .4960 - .2019 = .\mathbf{2941}$

6.57 $\mu = 2087$ $\sigma = 175$

If 20% are less, then 30% lie between x and μ.

$z_{.30} = -.84$

$$z = \frac{x-\mu}{\sigma}$$

$$-.84 = \frac{x-2087}{175}$$

x = **1940**

If 65% are more, then 15% lie between x and μ

$z_{.15} = -0.39$

$$z = \frac{x-\mu}{\sigma}$$

$$-.39 = \frac{x-2087}{175}$$

x = **2018.75**

If x is more than 85%, then 35% lie between x and μ.

$z_{.35} = 1.04$

$$z = \frac{x-\mu}{\sigma}$$

$$1.04 = \frac{x-2087}{175}$$

x = **2269**

6.58 $\lambda = 0.8$ person | minute

$P(x \geq 1 \text{ minute} \mid \lambda = 0.8 \text{ minute})$:

Let $x_0 = 1$

$P(x \geq 1) = e^{-\lambda x} = e^{-.8(1)} = e^{-.8} =$ **.4493**

$P(x \geq 2.5 \text{ Minutes} \mid \lambda = 0.8 \text{ per minute})$:

Let $x_0 = 2.5$

$P(x \geq 2.5) = e^{-\lambda x} = e^{-0.8(2.5)} = e^{-2} =$ **.1353**

6.59 $\mu = 2{,}106{,}774$ $\sigma = 50{,}940$

$P(x > 2{,}200{,}000)$:

$$z = \frac{2{,}200{,}000 - 2{,}106{,}774}{50{,}940} = 1.83$$

from table A.5 the area for $z = 1.83$ is .4664

$P(x > 2{,}200{,}000) = .5000 - .4664 =$ **.0336**

$P(x < 2{,}000{,}000)$:

$$z = \frac{2{,}000{,}000 - 2{,}106{,}774}{50{,}940} = -2.10$$

from table A.5 the area for $z = -2.10$ is .4821

$P(x < 2{,}000{,}000) = .5000 - .4821 =$ **.0179**

6.60 λ = 2.2 calls|30 secs.

Expected time between calls = μ = 1/λ = 1/(2.2) = .4545(30 sec.) = **13.64 sec.**

$P(x \geq 1$ min.| λ = 2.2 calls per 30 secs.):

Since Lambda and x are for different intervals,

Change Lambda to: λ = 4.4 calls/1 min.

$P(x \geq 1$ min| λ = 4.4 calls/1 min.):

For x_0 = 1: $e^{-\lambda x} = e^{-4.4(1)} =$ **.0123**

$P(x \geq 2$ min.| λ = 4.4 calls/1 min.):

For x_0 = 2: $e^{-\lambda x} = e^{-4.4(2)} = e^{-8.8} =$ **.0002**

6.61 This is a uniform distribution with a = 11 and b = 32.

The mean is (11 + 32)/2 = 21.5 and the standard deviation is

(32 - 11)/$\sqrt{12}$ = 6.06. Almost 81% of the time there are less than or equal to 28

sales associates working. One hundred percent of the time there are less than or

equal to 34 sales associates working and never more than 34. About 23.8% of

the time there are 16 or fewer sales associates working. There are 21 or fewer

sales associates working about 48% of the time.

6.62 The weight of the rods is normally distributed with a mean of 227 mg and a

standard deviation of 2.3 mg. The probability that a rod weighs less than or

equal to 220 mg is .0012, less than or equal to 225 mg is .1923, less than

or equal to 227 is .5000 (since 227 is the mean), less than 231 mg is .9590, and less than or equal to 238 mg is 1.000.

6.63 The lengths of cell phone calls are normally distributed with a mean of 2.35

minutes and a standard deviation of .11 minutes. Almost 99% of the calls are

less than or equal to 2.60 minutes, almost 82% are less than or equal to 2.45

minutes, over 32% are less than 2.3 minutes, and almost none are less than

2 minutes.

6.64 The exponential distribution has λ = 4.51 per 10 minutes and μ = 1/4.51 =

.22173 of 10 minutes or 2.2173 minutes. The probability that there is less than

.1 or 1 minute between arrivals is .3630. The probability that there is less than

.2 or 2 minutes between arrivals is .5942. The probability that there is .5 or 5 minutes or more between arrivals is .1049. The probability that there is more

than 1 or 10 minutes between arrivals is .0110. It is almost certain that there

will be less than 2.4 or 24 minutes between arrivals.

Chapter 7
Sampling and Sampling Distributions

LEARNING OBJECTIVES

The two main objectives for Chapter 7 are to give you an appreciation for the proper application of sampling techniques and an understanding of the sampling distributions of two statistics, thereby enabling you to:

1. Determine when to use sampling instead of a census.
2. Distinguish between random and nonrandom sampling.
3. Decide when and how to use various sampling techniques.
4. Be aware of the different types of errors that can occur in a study.
5. Understand the impact of the central limit theorem on statistical analysis.

6. Use the sampling distributions of $\bar{x}$ and $\hat{p}$.

CHAPTER TEACHING STRATEGY

Virtually every analysis discussed in this text deals with sample data. It is important, therefore, that students are exposed to the ways and means that samples are gathered. The first portion of chapter 7 deals with sampling. Reasons for sampling versus taking a census are given. Most of these reasons are tied to the fact that taking a census costs more than sampling if the same measurements are being gathered. Students are then exposed to the idea of random versus nonrandom sampling. Random sampling appeals to their concepts of fairness and equal opportunity. This text emphasizes that nonrandom samples are non probability samples and cannot be used in inferential analysis because levels of confidence and/or probability cannot be assigned. It should be emphasized throughout the discussion of sampling techniques that as future business managers (most students will end up as some sort of supervisor/manager) students should be aware of where and how data are gathered for studies. This will help to assure that they will not make poor decisions based on inaccurate and poorly gathered data.

The central limit theorem opens up opportunities to analyze data with a host of techniques using the normal curve. Section 7.2 begins by showing one population (randomly generated and presented in histogram form) that is uniformly distributed and one that is exponentially distributed. Histograms of the means for random samples of varying sizes are presented. Note that the distributions of means “pile up” in the middle and begin to approximate the normal curve shape as sample size increases. Note also by observing the values on the bottom axis that the dispersion of means gets smaller and smaller as sample size increases thus underscoring the formula for the standard error of the mean ($\frac{\sigma}{\sqrt{n}}$). As the student sees the central limit theorem unfold, he/she begins to see that if the sample size is large enough, sample means can be analyzed using the normal curve regardless of the shape of the population.

Chapter 7 presents formulas derived from the central limit theorem for both sample means and sample proportions. Taking the time to introduce these techniques in this chapter can expedite the presentation of material in chapters 8 and 9.

CHAPTER OUTLINE

7.1 Sampling

- Reasons for Sampling
- Reasons for Taking a Census
- Frame
- Random Versus Nonrandom Sampling
- Random Sampling Techniques
  - Simple Random Sampling
  - Stratified Random Sampling
  - Systematic Sampling
  - Cluster or Area Sampling
- Nonrandom Sampling
  - Convenience Sampling
  - Judgment Sampling
  - Quota Sampling
  - Snowball Sampling

Sampling Error

Nonsampling Errors

7.2 Sampling Distribution of $\bar{x}$

Sampling from a Finite Population

7.3 Sampling Distribution of $\hat{p}$

KEY TERMS

Central Limit Theorem
Cluster (or Area) Sampling
Convenience Sampling
Disproportionate Stratified Random Sampling
Finite Correction Factor
Frame
Judgment Sampling
Nonrandom Sampling
Nonrandom Sampling Techniques
Nonsampling Errors
Proportionate Stratified Random Sampling
Quota Sampling
Random Sampling
Sample Proportion
Sampling Error
Simple Random Sampling
Snowball Sampling
Standard Error of the Mean
Standard Error of the Proportion
Stratified Random Sampling
Systematic Sampling
Two-Stage Sampling

SOLUTIONS TO PROBLEMS IN CHAPTER 7

7.1 a) i. A union membership list for the company.

ii. A list of all employees of the company.

b) i. White pages of the telephone directory for Utica, New York.

ii. Utility company list of all customers.

c) i. Airline company list of phone and mail purchasers of tickets from the airline during the past six months.

ii. A list of frequent flyer club members for the airline.

d) i. List of boat manufacturer's employees.

ii. List of members of a boat owners association.

e) i. Cable company telephone directory.

ii. Membership list of cable management association.

7.4 a) Size of motel (rooms), age of motel, geographic location.

b) Gender, age, education, social class, ethnicity.

c) Size of operation (number of bottled drinks per month), number of employees, number of different types of drinks bottled at that location, geographic location.

d) Size of operation (sq.ft.), geographic location, age of facility, type of process used.

7.5 a) Under 21 years of age, 21 to 39 years of age, 40 to 55 years of age, over 55 years of age.

b) Under $1,000,000 sales per year, $1,000,000 to $4,999,999 sales per year, $5,000,000 to $19,999,999 sales per year, $20,000,000 to $49,000,000 per year, $50,000,000 to $99,999,999 per year, over $100,000,000 per year.

c) Less than 2,000 sq. ft., 2,000 to 4,999 sq. ft.,

5,000 to 9,999 sq. ft., over 10,000 sq. ft.

d) East, southeast, midwest, south, southwest, west, northwest.

e) Government worker, teacher, lawyer, physician, engineer, business person, police officer, fire fighter, computer worker.

f) Manufacturing, finance, communications, health care, retailing, chemical, transportation.

7.6 $n = N/k$ = 100,000/200 = **500**

7.7 $N = n \cdot k$ = **825**

7.8 $k = N/n$ = 3,500/175 = **20**

Start between 0 and 20. The human resource department probably has a list of company employees which can be used for the frame. Also, there might be a company phone directory available.

7.9 a) i. Counties

ii. Metropolitan areas

b) i. States (beside which the oil wells lie)

ii. Companies that own the wells

c) i. States

ii. Counties

7.10 Go to the district attorney's office and observe the apparent activity of various attorney's at work. Select some who are very busy and some who seem to be less active. Select some men and some women. Select some who appear to be older and some who are younger. Select attorneys with different ethnic backgrounds.

7.11 Go to a conference where some of the Fortune 500 executives attend. Approach those executives who appear to be friendly and approachable.

7.12 Suppose 40% of the sample is to be people who presently own a personal computer and 60% , people who do not. Go to a computer show at the city's conference center and start interviewing people. Suppose you get enough people who own personal computers but not enough interviews with those who do not. Go to a mall and start interviewing people. Screen out personal computer owners. Interview non personal computer owners until you meet the 60% quota.

7.13 $\mu = 50$, $\sigma = 10$, $n = 64$

a) $P(\bar{x} > 52)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52-50}{\frac{10}{\sqrt{64}}} = 1.6$$

from Table A.5, Prob. = .4452

$P(\bar{x} > 52) = .5000 - .4452 =$ **.0548**

b) $P(\bar{x} < 51)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{51-50}{\frac{10}{\sqrt{64}}} = 0.80$$

from Table A.5   prob. = .2881

$P(\bar{x} < 51) = .5000 + .2881 =$ **.7881**

c) $P(\bar{x} < 47)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{47-50}{\frac{10}{\sqrt{64}}} = -2.40$$

from Table A.5 prob. = .4918

$P(\bar{x} < 47) = .5000 - .4918 =$ **.0082**

d) $P(48.5 \leq \bar{x} \leq 52.4)$:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{48.5 - 50}{\frac{10}{\sqrt{64}}} = -1.20$$

from Table A.5 prob. = .3849

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52.4 - 50}{\frac{10}{\sqrt{64}}} = 1.92$$

from Table A.5 prob. = .4726

$P(48.5 \leq \bar{x} \leq 52.4) = .3849 + .4726 =$ **.8575**

e) $P(50.6 \le \bar{x} \le 51.3)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{50.6-50}{\frac{10}{\sqrt{64}}} = 0.48$$

from Table A.5, prob. = .1844

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{51.3-50}{\frac{10}{\sqrt{64}}} = 1.04$$

from Table A.5, prob. = .3508

$P(50.6 \le \bar{x} \le 51.3)$ = .3508 - .1844 = **.1664**

7.14 μ = 23.45 σ = 3.8

a) n = 10, $P(\bar{x} \ge 22)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{22-23.45}{\frac{3.8}{\sqrt{10}}} = -1.21$$

from Table A.5, prob. = .3869

$P(\bar{x} \geq 22)$ = .3869 + .5000 = **.8869**

b) n = 4, $P(\bar{x} > 26)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{26-23.45}{\frac{3.8}{\sqrt{4}}} = 1.34$$

from Table A.5, prob. = .4099

$P(\bar{x} > 26)$ = .5000 - .4099 = **.0901**

7.15 n = 36 μ = 278 $P(\bar{x} < 280)$ = .86

.3600 of the area lies between $\bar{x}$ = 280 and μ = 278. This probability is associated with z = 1.08 from Table A.5. Solving for σ:

$z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$1.08 = \dfrac{280 - 278}{\frac{\sigma}{\sqrt{36}}}$

$1.08 \dfrac{\sigma}{6} = 2$

$\sigma = \dfrac{12}{1.08} =$ **11.11**

7.16 $n = 81$ $\sigma = 12$ $P(\bar{x} > 300) = .18$

.5000 - .1800 = .3200 and from Table A.5, $z_{.3200} = 0.92$

Solving for μ:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$0.92 = \frac{300-\mu}{\frac{12}{\sqrt{81}}}$$

$$0.92\frac{12}{9} = 300 - \mu$$

$$1.2267 = 300 - \mu$$

$\mu = 300 - 1.2267 =$ **298.77**

7.17 a) $N = 1{,}000$ $n = 60$ $\mu = 75$ $\sigma = 6$

$P(\bar{x} < 76.5)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{76.5-75}{\frac{6}{\sqrt{60}}\sqrt{\frac{1000-60}{1000-1}}} = 2.00$$

from Table A.5, prob. = .4772

$P(\bar{x} < 76.5) = .4772 + .5000 =$ **.9772**

b) $N = 90$ $n = 36$ $\mu = 108$ $\sigma = 3.46$

$P(107 < \bar{x} < 107.7)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{107-108}{\frac{3.46}{\sqrt{36}}\sqrt{\frac{90-36}{90-1}}} = -2.23$$

from Table A.5, prob. = .4871

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{107.7-108}{\frac{3.46}{\sqrt{36}}\sqrt{\frac{90-36}{90-1}}} = -0.67$$

from Table A.5, prob. = .2486

$P(107 < \bar{x} < 107.7) = .4871 - .2486 =$ **.2385**

c) $N = 250$ $n = 100$ $\mu = 35.6$ $\sigma = 4.89$

$P(\bar{x} \geq 36)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{36-35.6}{\frac{4.89}{\sqrt{100}}\sqrt{\frac{250-100}{250-1}}} = 1.05$$

from Table A.5, prob. = .3531

$P(\bar{x} \geq 36)$ = .5000 - .3531 = **.1469**

d) N = 5000 n = 60 μ = 125 σ = 13.4

$P(\bar{x} \leq 123)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{123-125}{\frac{13.4}{\sqrt{60}}\sqrt{\frac{5000-60}{5000-1}}} = -1.16$$

from Table A.5, prob. = .3770

$P(\bar{x} \leq 123)$ = .5000 - .3770 = **.1230**

7.18 $\mu = 99.9$ $\sigma = 30$ $n = 38$

a) $P(\bar{x} < 90)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{90-99.9}{\frac{30}{\sqrt{38}}} = -2.03$$

from table A.5, area = .4788

$P(\bar{x} < 90) = .5000 - .4788 =$ **.0212**

b) $P(98 \le \bar{x} \le 105)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{105-99.9}{\frac{30}{\sqrt{38}}} = 1.05$$

from table A.5, area = .3531

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{98-99.9}{\frac{30}{\sqrt{38}}} = -0.39$$

from table A.5, area = .1517

$P(98 \leq \bar{x} \leq 105)$ = .3531 + .1517 = **.5048**

c) $P(\bar{x} < 112)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{112-99.9}{\frac{30}{\sqrt{38}}} = 2.49$$

from table A.5, area = .4936

$P(\bar{x} < 112)$ = .5000 + .4936 = **.9936**

d) $P(93 \leq \bar{x} \leq 96)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{93-99.9}{\frac{30}{\sqrt{38}}} = -1.42$$

from table A.5, area = .4222

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96-99.9}{\frac{30}{\sqrt{38}}} = -0.80$$

from table A.5, area = .2881

$P(93 \le \bar{x} \le 96)$ = .4222 - .2881 = **.1341**

7.19 N = 1500 n = 100 μ = 177,000 σ = 8,500

$P(\bar{X} >$ \$185,000):

$$z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{185{,}000-177{,}000}{\frac{8{,}500}{\sqrt{100}}\sqrt{\frac{1500-100}{1500-1}}} = 9.74$$

from Table A.5, prob. = .5000

$P(\bar{X} >$ \$185,000) = .5000 - .5000 = **.0000**

7.20 μ = \$65.12 σ = \$21.45 n = 45 $P(\bar{x} > \bar{x}_0)$ = .2300

Prob. $\bar{x}$ lies between $\bar{x}_0$ and μ = .5000 - .2300 = .2700

from Table A.5, $z_{.2700}$ = 0.74

Solving for $\bar{x}_0$:

$$z = \frac{\bar{x}_0 - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$0.74 = \frac{\bar{x}_0 - 65.12}{\frac{21.45}{\sqrt{45}}}$$

2.366 = $\bar{x}_0$ - 65.12 and $\bar{x}_0$ = 65.12 + 2.366 = **67.486**

7.21 μ = 50.4 σ = 11.8 n = 42

a) $P(\bar{x} > 52)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{52-50.4}{\frac{11.8}{\sqrt{42}}} = 0.88$$

from Table A.5, the area for z = 0.88 is .3106

$P(\bar{x} > 52) = .5000 - .3106 =$ **.1894**

b) $P(\bar{x} < 47.5)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{47.5-50.4}{\frac{11.8}{\sqrt{42}}} = -1.59$$

from Table A.5, the area for *z* = -1.59 is .4441

$P(\bar{x} < 47.5) = .5000 - .4441 =$ **.0559**

c) $P(\bar{x} < 40)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{40-50.4}{\frac{11.8}{\sqrt{42}}} = -5.71$$

from Table A.5, the area for z = -5.71 is .5000

$P(\bar{x} < 40)$ = .5000 - .5000 = **.0000**

d) 71% of the values are greater than 49. Therefore, 21% are between the sample mean of 49 and the population mean, μ = 50.4.

The *z* value associated with the 21% of the area is -0.55

$z_{.21}$ = -0.55

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-0.55 = \frac{49-50.4}{\frac{\sigma}{\sqrt{42}}}$$

σ = **16.4964**

7.22 $p = .25$

a) $n = 110$ $P(\hat{p} \le .21)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.21 - .25}{\sqrt{\frac{(.25)(.75)}{110}}} = -0.97$$

from Table A.5, prob. = .3340

$P(\hat{p} \le .21)$ = .5000 - .3340 = **.1660**

b) $n = 33$ $P(\hat{p} > .24)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.24 - .25}{\sqrt{\frac{(.25)(.75)}{33}}} = -0.13$$

from Table A.5, prob. = .0517

$P(\hat{p} > .24)$ = .5000 + .0517 = **.5517**

c) $n = 59$ $P(.24 \le \hat{p} \le .27)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.24 - .25}{\sqrt{\frac{(.25)(.75)}{59}}} = -0.18$$

from Table A.5, prob. = .0714

$$z = \frac{\hat{p} - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.27 - .25}{\sqrt{\frac{(.25)(.75)}{59}}} = 0.35$$

from Table A.5, prob. = .1368

$P(.24 \le \hat{p} \le .27)$ = .0714 + .1368 = **.2082**

d) $n = 80$ $P(\hat{p} > .30)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.30 - .25}{\sqrt{\frac{(.25)(.75)}{80}}} = 1.03$$

from Table A.5, prob. = .3485

$P(\hat{p} > .30)$ = .5000 - .3485 = **.1515**

e) $n = 800$ $P(\hat{p} > .30)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.30 - .25}{\sqrt{\frac{(.25)(.75)}{800}}} = 3.27$$

from Table A.5, prob. = .4995

$P(\hat{p} > .30)$ = .5000 - .4995 = **.0005**

7.23 $p = .58$ $n = 660$

a) $P(\hat{p} > .60)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.60 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = 1.04$$

from table A.5, area = .3508

$P(\hat{p} > .60)$ = .5000 - .3508 = **.1492**

b) $P(.55 < \hat{p} < .65)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.65 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = 3.64$$

from table A.5, area = .4998

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.55 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = 1.56$$

from table A.5, area = .4406

$P(.55 < \hat{p} < .65)$ = .4998 + .4406 = **.9404**

c) $P(\hat{p} > .57)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.57 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = -0.52$$

from table A.5, area = .1985

$P(\hat{p} > .57)$ = .1985 + .5000 = **.6985**

d) $P(.53 \le \hat{p} \le .56)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.56 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = 1.04 \qquad z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.53 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = 2.60$$

from table A.5, area for z = 1.04 is .3508

from table A.5, area for z = 2.60 is .4953

$P(.53 \le \hat{p} \le .56)$ = .4953 - .3508 = **.1445**

e) $P(\hat{p} < .48)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.48 - .58}{\sqrt{\frac{(.58)(.42)}{660}}} = -5.21$$

from table A.5, area = .5000

$P(\hat{p} < .48)$ = .5000 - .5000 = **.0000**

7.24 $p = .40$ $P(\hat{p} \geq .35) = .8000$

$P(.35 \leq \hat{p} \leq .40) = .8000 - .5000 = .3000$

from Table A.5, $z_{.3000} = -0.84$

Solving for n:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$-0.84 = \frac{.35 - .40}{\sqrt{\frac{(.40)(.60)}{n}}} = \frac{-.05}{\sqrt{\frac{.24}{n}}}$$

$$\frac{-0.84\sqrt{.24}}{-.05} = \sqrt{n}$$

$$8.23 = \sqrt{n}$$

$n = 67.73 \approx 68$

7.25 p = .28 n = 140 $P(\hat{p} < \hat{p}_0) = .3000$

$P(\hat{p} \le \hat{p}_0 \le .28) = .5000 - .3000 = .2000$

from Table A.5, $z_{.2000} = -0.52$

Solving for $\hat{p}_0$:

$$z = \frac{\hat{p}_0 - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$-0.52 = \frac{\hat{p}_0 - .28}{\sqrt{\frac{(.28)(.72)}{140}}}$$

$$-.02 = \hat{p}_0 - .28$$

$$\hat{p}_0 = .28 - .02 = \mathbf{.26}$$

7.26 $P(x > 150)$: n = 600 p = .21 x = 150

$$\hat{p} = \frac{150}{600} = .25$$

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.25-.21}{\sqrt{\dfrac{(.21)(.79)}{600}}} = 2.41$

from table A.5, area = .4920

$P(x > 150)$ = .5000 - .4920 = **.0080**

7.27 p = .48 n = 200

a) $P(x < 90)$:

$\hat{p} = \dfrac{90}{200} = .45$

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.45-.48}{\sqrt{\dfrac{(.48)(.52)}{200}}} = -0.85$

from Table A.5, the area for z = -0.85 is .3023

$P(x < 90)$ = .5000 - .3023 = **.1977**

b) $P(x > 100)$:

$\hat{p} = \dfrac{100}{200} = .50$

$z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}} = \dfrac{.50 - .48}{\sqrt{\dfrac{(.48)(.52)}{200}}} = 0.57$

from Table A.5, the area for *z* = 0.57 is .2157

P(*x* > 100) = .5000 - .2157 = .**2843**

c) $P(x > 80)$:

$\hat{p} = \frac{80}{200} = .40$

$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.40-.48}{\sqrt{\frac{(.48)(.52)}{200}}} = -2.26$

from Table A.5, the area for z = -2.26 is .4881

$P(x > 80)$ = .5000 + .4881 = **.9881**

7.28 p = .19 n = 950

a) $P(\hat{p} > .25)$:

$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.25-.19}{\sqrt{\frac{(.19)(.89)}{950}}} = 4.71$

from Table A.5, area = .5000

$P(\hat{p} > .25)$ = .5000 - .5000 = **.0000**

b) $P(.15 \le \hat{p} \le .20)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.15 - .19}{\sqrt{\frac{(.19)(.81)}{950}}} = -3.14$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.20 - .19}{\sqrt{\frac{(.19)(.89)}{950}}} = 0.79$$

from Table A.5, area for z = -3.14 is .4992

from Table A.5, area for z = 0.79 is .2852

$P(.15 \le \hat{p} \le .20)$ = .4992 + .2852 = **.7844**

c) $P(133 \le x \le 171)$:

$\hat{p}_1 = \frac{133}{950} = .14$ $\qquad$ $\hat{p}_2 = \frac{171}{950} = .18$

$P(.14 \le \hat{p} \le .18)$:

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.14-.19}{\sqrt{\dfrac{(.19)(.81)}{950}}} = -3.93$ $\qquad z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.18-.19}{\sqrt{\dfrac{(.19)(.81)}{950}}} = -0.79$

from Table A.5, the area for z = -3.93 is .49997

the area for z = -0.79 is .2852

$P(133 \le x \le 171)$ = .49997 - .2852 = **.21477**

7.29 $\mu = 76$, $\sigma = 14$

a) n = 35, $P(\bar{x} \ge 79)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{79-76}{\frac{14}{\sqrt{35}}} = 1.27$$

from table A.5, area = .3980

$P(\bar{x} \ge 79)$ = .5000 - .3980 = **.1020**

b) n = 140, $P(74 \leq \bar{x} \leq 77)$:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{74 - 76}{\frac{14}{\sqrt{140}}} = -1.69 \qquad z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{77 - 76}{\frac{14}{\sqrt{140}}} = 0.85$$

from table A.5, area for z = -1.69 is .4545

from table A.5, area for 0.85 is .3023

$P(74 \leq \bar{x} \leq 77)$ = .4545 + .3023 = **.7568**

c) n = 219, $P(\bar{x} < 76.5)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{76.5-76}{\frac{14}{\sqrt{219}}} = 0.53$$

from table A.5, area = .2019

$P(\bar{x} < 76.5)$ = .5000 + .2019 = **.7019**

7.30 p = .46

a) n = 60

$P(.41 < \hat{p} < .53)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.53-.46}{\sqrt{\frac{(.46)(.54)}{60}}} = 1.09$$

from table A.5, area = .3621

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.41-.46}{\sqrt{\frac{(.46)(.54)}{60}}} = -0.78$$

from table A.5, area = .2823

$P(.41 < \hat{p} < .53) = .3621 + .2823 =$ **.6444**

b) $n = 458$ $P(\hat{p} < .40)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.40-.46}{\sqrt{\frac{(.46)(.54)}{458}}} = -2.58$$

from table A.5, area = .4951

$P(\hat{p} < .40) = .5000 - .4951 =$ **.0049**

c) n = 1350 $P(\hat{p} > .49)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.49-.46}{\sqrt{\frac{(.46)(.54)}{1350}}} = 2.21$$

from table A.5, area = .4864

$P(\hat{p} > .49)$ = .5000 - .4864 = **.0136**

7.31	Under 18	250(.22) =	**55**
	18 – 25	250(.18) =	**45**
	26 – 50	250(.36) =	**90**
	51 – 65	250(.10) =	**25**
	over 65	250(.14) =	**35**
		n =	250

7.32 p = .55 n = 600 x = 298

$$\hat{p} = \frac{x}{n} = \frac{298}{600} = .497$$

$P(\hat{p} \leq .497)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.497 - .55}{\sqrt{\frac{(.55)(.45)}{600}}} = -2.61$$

from Table A.5, Prob. = .4955

$P(\hat{p} \leq .497)$ = .5000 - .4955 = **.0045**

No, the probability of obtaining these sample results by chance from a population that supports the candidate with 55% of the vote is extremely low (.0045). This is such an unlikely chance sample result that it would cause the researcher to probably reject her claim of 55% of the vote.

7.33 a) Roster of production employees secured from the human resources department of the company.

b) Alpha/Beta store records kept at the headquarters of their California division or merged files of store records from regional offices across the state.

c) Membership list of Maine lobster catchers association.

7.34 μ = \$ 17,755 σ = \$ 650 n = 30 N = 120

$P(\bar{x} < 17{,}500)$:

$$z = \frac{17{,}500 - 17{,}755}{\frac{650}{\sqrt{30}}\sqrt{\frac{120-30}{120-1}}} = -2.47$$

from Table A.5, the area for z = -2.47 is .4932

$P(\bar{x} < 17{,}500) = .5000 - .4932 =$ **.0068**

7.35 Number the employees from 0001 to 1250. Randomly sample from the random number table until 60 different usable numbers are obtained. You cannot use numbers from 1251 to 9999.

7.36 μ = \$125 n = 32 $\bar{x}$ = \$110 σ^2 = \$525

$P(\bar{x} \geq \$110)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{110-125}{\frac{\sqrt{525}}{\sqrt{32}}} = -3.70$$

from Table A.5, Prob.= .5000

$P(\bar{x} \geq \$110)$ = .5000 + .5000 = **1.0000**

$P(\bar{x} \geq \$135)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{135-125}{\frac{\sqrt{525}}{\sqrt{32}}} = 2.47$$

from Table A.5, Prob.= .4932

$P(\bar{x} \geq \$135)$ = .5000 - .4932 = **.0068**

$P(\$120 < \bar{x} < \$130)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{120-125}{\frac{\sqrt{525}}{\sqrt{32}}} = -1.23$$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{130-125}{\frac{\sqrt{525}}{\sqrt{32}}} = 1.23$$

from Table A.5, Prob.= .3907

$P(\$120 < \bar{x} < \$130) = .3907 + .3907$ = **.7814**

7.37 $n = 1100$

a) $x > 810$, $p = .73$

$\hat{p} = \dfrac{x}{n} = \dfrac{810}{1100}$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.7364-.73}{\sqrt{\frac{(.73)(.27)}{1100}}} = 0.48$$

from table A.5, area = .1844

$P(x > 810) = .5000 - .1844$ = **.3156**

b) $x < 1030$, $p = .96$,

$\hat{p} = \dfrac{x}{n} = \dfrac{1030}{1100} = .9364$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.9364 - .96}{\sqrt{\frac{(.96)(.04)}{1100}}} = -3.99$$

from table A.5, area = .49997

P(*x* < 1030) = .5000 - .49997 = **.00003**

c) *p* = .85

$P(.82 \le \hat{p} \le .84)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.82 - .85}{\sqrt{\frac{(.85)(.15)}{1100}}} = -2.79$$

from table A.5, area = .4974

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.84 - .85}{\sqrt{\frac{(.85)(.15)}{1100}}} = -0.93$$

from table A.5, area = .3238

$P(.82 \le \hat{p} \le .84)$ = .4974 - .3238 = **.1736**

7.38 1) The managers from some of the companies you are interested in studying do not belong to the American Managers Association.

2) The membership list of the American Managers Association is not up-to-date.

3) You are not interested in studying managers from some of the companies belonging to the American Management Association.

4) The wrong questions are asked.

5) The manager incorrectly interprets a question.

6) The assistant accidentally marks the wrong answer.

7) The wrong statistical test is used to analyze the data.

8) An error is made in statistical calculations.

9) The statistical results are misinterpreted.

7.39 Divide the factories into geographic regions and select a few factories to represent those regional areas of the country. Take a random sample of employees from each selected factory. Do the same for distribution centers and retail outlets. Divide the United States into regions of areas. Select a few areas. Take a random sample from each of the selected area distribution centers and retail outlets.

7.40 $N = 12{,}080 \quad n = 300$

$k = N/n = 12{,}080/300 = 40.27$

Select every **40th** outlet to assure $n \geq 300$ outlets.

Use a table of random numbers to select a value between 0 and 40 as a starting point.

7.41 p = .54 n = 565

a) $P(x \geq 339)$:

$$\hat{p} = \frac{x}{n} = \frac{339}{565} = .60$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.60-.54}{\sqrt{\frac{(.54)(.46)}{565}}} = 2.86$$

from Table A.5, the area for z = 2.86 is .4979

$P(x \geq 339)$ = .5000 - .4979 = **.0021**

b) $P(x \geq 288)$:

$$\hat{p} = \frac{x}{n} = \frac{288}{565} = .5097$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.5097-.54}{\sqrt{\frac{(.54)(.46)}{565}}} = -1.45$$

from Table A.5, the area for z = -1.45 is .4265

$P(x \geq 288)$ = .5000 + .4265 = **.9265**

c) $P(\hat{p} \leq .50)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.50-.54}{\sqrt{\frac{(.54)(.46)}{565}}} = -1.91$$

from Table A.5, the area for z = -1.91 is .4719

$P(\hat{p} \leq .50)$ = .5000 - .4719 = **.0281**

7.42 μ = \$550 n = 50 σ = \$100

$P(\bar{x} < \$530)$:

$z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{530-550}{\frac{100}{\sqrt{50}}} = -1.41$

from Table A.5, Prob.=.4207

$P(x < \$530) = .5000 - .4207 =$ **.0793**

7.43 $\mu = 56.8$ $n = 51$ $\sigma = 12.3$

a) $P(\bar{x} > 60)$:

$z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{60-56.8}{\frac{12.3}{\sqrt{51}}} = 1.86$

from Table A.5, Prob. = .4686

$P(\bar{x} > 60) = .5000 - .4686 =$ **.0314**

b) $P(\bar{x} > 58)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{58-56.8}{\frac{12.3}{\sqrt{51}}} = 0.70$$

from Table A.5, Prob.= .2580

$P(\bar{x} > 58)$ = .5000 - .2580 = **.2420**

c) $P(56 < \bar{x} < 57)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{56-56.8}{\frac{12.3}{\sqrt{51}}} = -0.46 \qquad z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{57-56.8}{\frac{12.3}{\sqrt{51}}} = 0.12$$

from Table A.5, Prob. for z = -0.46 is .1772

from Table A.5, Prob. for z = 0.12 is .0478

$P(56 < \bar{x} < 57)$ = .1772 + .0478 = **.2250**

d) $P(\bar{x} < 55)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{55-56.8}{\frac{12.3}{\sqrt{51}}} = -1.05$$

from Table A.5, Prob.= .3531

$P(\bar{x} < 55) = .5000 - .3531 =$ **.1469**

e) $P(\bar{x} < 50)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{50-56.8}{\frac{12.3}{\sqrt{51}}} = -3.95$$

from Table A.5, Prob.= .5000

$P(\bar{x} < 50) = .5000 - .5000 =$ **.0000**

7.45 $p = .73$ $n = 300$

a) $P(210 \leq x \leq 234)$:

$\hat{p}_1 = \frac{x}{n} = \frac{210}{300} = .70$ $\hat{p}_2 = \frac{x}{n} = \frac{234}{300} = .78$

z = $\frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.70-.73}{\sqrt{\frac{(.73)(.27)}{300}}}$ = -1.17

z = $\frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.78-.73}{\sqrt{\frac{(.73)(.27)}{300}}}$ = 1.95

from Table A.5, the area for z = -1.17 is .3790

the area for z = 1.95 is .4744

$P(210 \leq x \leq 234)$ = .3790 + .4744 = **.8534**

b) $P(\hat{p} \geq .78)$:

z = $\frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.78-.73}{\sqrt{\frac{(.73)(.27)}{300}}}$ = 1.95

from Table A.5, the area for z = 1.95 is .4744

$P(\hat{p} \geq .78)$ = .5000 - .4744 = **.0256**

c) p = .73 n = 800 $P(\hat{p} \geq .78)$:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.78 - .73}{\sqrt{\frac{(.73)(.27)}{800}}} = 3.19$$

from Table A.5, the area for z = 3.19 is .4993

$P(\hat{p} \geq .78)$ = .5000 - .4993 = **.0007**

7.46 n = 140 $P(x \geq 35)$:

$\hat{p} = \dfrac{35}{140} = .25$ $p = .22$

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.25-.22}{\sqrt{\dfrac{(.22)(.78)}{140}}} = 0.86$

from Table A.5, the area for z = 0.86 is .3051

$P(x \geq 35)$ = .5000 - .3051 = **.1949**

$P(x \leq 21)$:

$\hat{p} = \dfrac{21}{140} = .15$

$z = \dfrac{\hat{p}-p}{\sqrt{\dfrac{p\cdot q}{n}}} = \dfrac{.15-.22}{\sqrt{\dfrac{(.22)(.78)}{140}}} = -2.00$

from Table A.5, the area for z = 2.00 is .4772

$P(x \leq 21)$ = .5000 - .4772 = .**0228**

$n = 300$ $p = .20$

$P(.18 < \hat{p} < .25)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.18-.20}{\sqrt{\frac{(.20)(.80)}{300}}} = -0.87$$

from Table A.5, the area for $z = -0.87$ is .3078

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.25-.20}{\sqrt{\frac{(.20)(.80)}{300}}} = 2.17$$

from Table A.5, the area for $z = 2.17$ is .4850

$P(.18 < \hat{p} < .25) = .3078 + .4850 =$ **.7928**

7.47 By taking a sample, there is potential for obtaining more detailed information. More time can be spent with each employee. Probing questions can be asked. There is more time for trust to be built between employee and interviewer resulting in the potential for more honest, open answers.

With a census, data is usually more general and easier to analyze because it is in a more standard format. Decision-makers are sometimes more comfortable with a census because everyone is included and there is no sampling error. A census appears to be a better political device because the CEO can claim that everyone in the company has had input.

7.48 $p = .75$ $n = 150$ $x = 120$

$P(\hat{p} \geq .80)$:

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.80-.75}{\sqrt{\frac{(.75)(.25)}{150}}} = 1.41$$

from Table A.5, the area for $z = 1.41$ is .4207

$P(\hat{p} \geq .80) = .5000 - .4207 =$ **.0793**

7.49 Switzerland: $n = 40$ $\mu = \$\ 21.24$ $\sigma = \$\ 3$

$P(21 \leq \bar{x} \leq 22)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{21-21.24}{\frac{3}{\sqrt{40}}} = -0.51$$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{22-21.24}{\frac{3}{\sqrt{40}}} = 1.60$$

from Table A.5, the area for z = -0.51 is .1950

the area for z = 1.60 is .4452

$P(21 \leq \bar{x} \leq 22)$ = .1950 + .4452 = **.6402**

Japan: $n = 35$ $\mu = \$\ 22.00$ $\sigma = \$3$

$P(\bar{x} > 23)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{23-22}{\frac{3}{\sqrt{35}}} = 1.97$$

from Table A.5, the area for $z = 1.97$ is .4756

$P(\bar{x} > 23)$ = .5000 - .4756 = **.0244**

U.S.: $n = 50$ $\mu = \$\ 19.86$ $\sigma = \$\ 3$

$P(\bar{x} < 18.90)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{18.90-19.86}{\frac{3}{\sqrt{50}}} = -2.26$$

from Table A.5, the area for $z = -2.26$ is .4881

$P(\bar{x} < 18.90)$ = .5000 - .4881 = **.0119**

7.50 a) Age, Ethnicity, Religion, Geographic Region, Occupation, Urban-Suburban-Rural, Party Affiliation, Gender

b) Age, Ethnicity, Gender, Geographic Region, Economic Class

c) Age, Ethnicity, Gender, Economic Class, Education

d) Age, Ethnicity, Gender, Economic Class, Geographic Location

7.51 μ = \$281 n = 65 σ = \$47

$P(\bar{x} > \$273)$:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{273-281}{\frac{47}{\sqrt{65}}} = -1.37$$

from Table A.5 the area for z = -1.37 is .4147

$P(\bar{x} > \$273)$ = .5000 + .4147 = **.9147**

Chapter 8

Statistical Inference: Estimation for Single Populations

LEARNING OBJECTIVES

The overall learning objective of Chapter 8 is to help you understand estimating

parameters of single populations, thereby enabling you to:

1. Know the difference between point and interval estimation.
2. Estimate a population mean from a sample mean when σ is known.
3. Estimate a population mean from a sample mean when σ is unknown.
4. Estimate a population proportion from a sample proportion.
5. Estimate the population variance from a sample variance.
6. Estimate the minimum sample size necessary to achieve given statistical goals.

CHAPTER TEACHING STRATEGY

Chapter 8 is the student's introduction to interval estimation and estimation of sample size. In this chapter, the concept of point estimate is discussed along with the notion that as each sample changes in all likelihood so will the point estimate. From this, the student can see that an interval estimate may be more usable as a one-time proposition than the point estimate. The confidence interval formulas for large sample means and proportions can be presented as mere algebraic manipulations of formulas developed in chapter 7 from the Central Limit Theorem.

It is very important that students begin to understand the difference between mean and proportions. Means can be generated by averaging some sort of measurable item such as age, sales, volume, test score, etc. Proportions are computed by counting the number of items containing a characteristic of interest out of the total number of items. Examples might be proportion of people carrying a VISA card, proportion of items that are defective, proportion of market purchasing brand A. In addition, students can begin to see that sometimes single samples are taken and analyzed; but that other times, two samples are taken in order to compare two brands, two techniques, two conditions, male/female, etc.

In an effort to understand the impact of variables on confidence intervals, it may be useful to ask the students what would happen to a confidence interval if the sample size is varied or the confidence is increased or decreased. Such consideration helps the student see in a different light the items that make up a confidence interval. The student can see that increasing the sample size reduces the width of the confidence interval, all other things being constant, or that it increases confidence if other things are held constant. Business students probably understand that increasing sample size costs more and thus there are trade-offs in the research set-up.

In addition, it is probably worthwhile to have some discussion with students regarding the meaning of confidence, say 95%. The idea is presented in the chapter that if 100 samples are randomly taken from a population and 95% confidence intervals are computed on each sample, that 95%(100) or 95 intervals should contain the parameter of estimation and approximately 5 will not. In most cases, only one confidence interval is computed, not 100, so the 95% confidence puts the odds in the researcher's favor. It should be pointed out, however, that the confidence interval computed may not contain the parameter of interest.

This chapter introduces the student to the *t* distribution for

estimating population means when σ is unknown. Emphasize that this applies only when the population is normally distributed because it is an assumption underlying the *t* test that the population is normally distributed, albeit that this assumption is robust. The student will observe that the *t* formula is essentially the same as the *z* formula and that it is the table that is different. When the population is normally distributed and σ is known, the *z* formula can be used even for small samples.

A formula is given in chapter 8 for estimating the population variance; and

it is here that the student is introduced to the chi-square distribution. An

assumption underlying the use of this technique is that the population is normally

distributed. The use of the chi-square statistic to estimate the population variance

is extremely sensitive to violations of this assumption. For this reason, extreme

caution should be exercised in using this technique. Because of this, some

statisticians omit this technique from consideration presentation and usage.

Lastly, this chapter contains a section on the estimation of sample size.

One of the more common questions asked of statisticians is: "How large of a

sample size should I take?" In this section, it should be emphasized that sample

size estimation gives the researcher a "ball park" figure as to how many to sample.

The “error of estimation “ is a measure of the sampling error. It is also equal to

the $\pm$ error of the interval shown earlier in the chapter.

CHAPTER OUTLINE

KEY WORDS

Bounds	Point Estimate
Chi-square Distribution	Robust
Degrees of Freedom(df)	Sample-Size Estimation
Error of Estimation	*t* Distribution
Interval Estimate	*t* Value

SOLUTIONS TO PROBLEMS IN CHAPTER 8

8.1 a) $\bar{x}$ = 25 σ = 3.5 n = 60

95% Confidence $z_{.025}$ = 1.96

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ = 25 $\pm$ 1.96 $\frac{3.5}{\sqrt{60}}$ = 25 $\pm$ 0.89 = **24.11 $\le$ μ $\le$ 25.89**

b) $\bar{x}$ = 119.6 σ = 23.89 n = 75

98% Confidence $z_{.01}$ = 2.33

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ = 119.6 $\pm$ 2.33 $\frac{23.89}{\sqrt{75}}$ = 119.6 ± 6.43 = **113.17 $\le$ μ $\le$ 126.03**

c) $\bar{x}$ = 3.419 σ = 0.974 n = 32

90% C.I. $z_{.05}$ = 1.645

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 3.419 \pm 1.645\frac{0.974}{\sqrt{32}} = 3.419 \pm .283 =$ **3.136 ≤ μ ≤ 3.702**

d) $\bar{x} = 56.7$ $\sigma = 12.1$ $N = 500$ $n = 47$

80% C.I. $z_{.10} = 1.28$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 56.7 \pm 1.28\frac{12.1}{\sqrt{47}}\sqrt{\frac{500-47}{500-1}} =$

$56.7 \pm 2.15 =$ **54.55 ≤ μ ≤ 58.85**

8.2 $n = 36$ $\bar{x} = 211$ $\sigma = 23$

95% C.I. $z_{.025} = 1.96$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 211 \pm 1.96\frac{23}{\sqrt{36}} = 211 \pm 7.51 =$ **$203.49 \leq \mu \leq 218.51$**

8.3 $n = 81$ $\bar{x} = 47$ $\sigma = 5.89$

90% C.I. $z_{.05} = 1.645$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 47 \pm 1.645\frac{5.89}{\sqrt{81}} = 47 \pm 1.08 =$ **$45.92 \leq \mu \leq 48.08$**

8.4 $n = 70$ $\sigma^2 = 49$ $\bar{x} = 90.4$

$\bar{x} = 90.4$ Point Estimate

94% C.I. $z_{.03} = 1.88$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 90.4 \pm 1.88\frac{\sqrt{49}}{\sqrt{70}} = 90.4 \pm 1.57 =$ **$88.83 \leq \mu \leq 91.97$**

8.5 $n = 39$ $N = 200$ $\bar{x} = 66$ $\sigma = 11$

96% C.I. $z_{.02} = 2.05$

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 66 \pm 2.05\frac{11}{\sqrt{39}}\sqrt{\frac{200-39}{200-1}} =$$

66 ± 3.25 = **62.75 ≤ μ ≤ 69.25**

$\bar{x}$ **= 66 Point Estimate**

8.6 $n = 120$ $\bar{x} = 18.72$ $\sigma = 0.8735$

99% C.I. $z_{.005} = 2.575$

$\bar{x}$ **= 18.72 Point Estimate**

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 18.72 \pm 2.575\frac{0.8735}{\sqrt{120}} = 8.72 \pm .21 =$$ **18.51 ≤ μ ≤ 18.93**

8.7 $N = 1500$ $n = 187$ $\bar{x} = 5.3$ years $\sigma = 1.28$ years

95% C.I. $z_{.025} = 1.96$

$\bar{x}$ **= 5.3 years Point Estimate**

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ = 5.3 ± 1.96 $\frac{1.28}{\sqrt{187}}\sqrt{\frac{1500-187}{1500-1}}$ =

5.3 ± .17 = **5.13 ≤ *μ* ≤ 5.47**

8.8 $n = 24$ $\bar{x}$ = 5.625 σ = 3.23

90% C.I. $z_{.05}$ = 1.645

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ = 5.625 ± 1.645 $\frac{3.23}{\sqrt{24}}$ = 5.625 ± 1.085 = **4.540 ≤ *μ* ≤ 6.710**

8.9 $n = 36$ $\bar{x}$ = 3.306 σ = 1.17

98% C.I. $z_{.01}$ = 2.33

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ = 3.306 ± 2.33 $\frac{1.17}{\sqrt{36}}$ = 3.306 ± .454 = **2.852 ≤ *μ* ≤ 3.760**

8.10 $n = 36$ $\bar{x}$ = 2.139 σ = .113

$\bar{x}$ **= 2.139   Point Estimate**

90% C.I. $z_{.05} = 1.645$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 2.139 \pm 1.645\frac{(.113)}{\sqrt{36}} = 2.139 \pm .031 = \mathbf{2.108 \le \mu \le 2.170}$

8.11 95% confidence interval $n = 45$

$\bar{x} = 24.533$ $\sigma = 5.124$ $z = \pm 1.96$

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 24.533 \pm 1.96\frac{5.124}{\sqrt{45}} =$

$24.533 \pm 1.497 =$ $\mathbf{23.036 \le \mu \le 26.030}$

8.12 The point estimate is **0.5765**. $n = 41$

The assumed standard deviation is **0.14**

95% level of confidence: $z = \pm$ **1.96**

Confidence interval: $\mathbf{0.533647 \le \mu \le 0.619353}$

Error of the estimate: 0.619353 - 0.5765 = **0.042853**

8.13 $n = 13$ $\bar{x} = 45.62$ $s = 5.694$ df = 13 – 1 = 12

95% Confidence Interval and $\alpha/2=.025$

$t_{.025,12}$ = 2.179

$\bar{x} \pm t\dfrac{s}{\sqrt{n}}$ = 45.62 ± 2.179 $\dfrac{5.694}{\sqrt{13}}$ = 45.62 ± 3.44 = **42.18 $\le \mu \le$ 49.06**

8.14 n = 12 $\bar{x}$ = 319.17 s = 9.104 df = 12 - 1 = 11

90% confidence interval

$\alpha/2$ = .05 $t_{.05,11}$ = 1.796

$\bar{x} \pm t\dfrac{s}{\sqrt{n}}$ = 319.17 ± (1.796) $\dfrac{9.104}{\sqrt{12}}$ = 319.17 ± 4.72 = **314.45 $\le \mu \le$ 323.89**

8.15 n = 41 $\bar{x}$ = 128.4 s = 20.6 df = 41 – 1 = 40

98% Confidence Interval

$\alpha/2$ = .01

$t_{.01,40}$ = 2.423

$\bar{x} \pm t \frac{s}{\sqrt{n}}$ = 128.4 ± 2.423 $\frac{20.6}{\sqrt{41}}$ = 128.4 ± 7.80 = **120.6 ≤ μ ≤ 136.2**

$\bar{x}$ **= 128.4 Point Estimate**

8.16 n = 15 $\bar{x}$ = 2.364 s^2 = 0.81 df = 15 – 1 = 14

90% Confidence interval

$\alpha/2$ = .05

$t_{.05,14}$ = 1.761

$\bar{x} \pm t \frac{s}{\sqrt{n}}$ = 2.364 ± 1.761 $\frac{\sqrt{0.81}}{\sqrt{15}}$ = 2.364 ± .409 = **1.955 ≤ μ ≤ 2.773**

8.17 n = 25 $\bar{x}$ = 16.088 s = .817 df = 25 – 1 = 24

99% Confidence Interval

$\alpha/2$ = .005

$t_{.005,24}$ = 2.797

$\bar{x} \pm t \frac{s}{\sqrt{n}}$ = 16.088 ± 2.797 $\frac{(.817)}{\sqrt{25}}$ = 16.088 ± .457 = **15.631 ≤ μ ≤ 16.545**

$\bar{x}$ **= 16.088 Point Estimate**

8.18 $n = 22$ $\bar{x} = 1{,}192$ $s = 279$ df = n - 1 = 21

98% CI and $\alpha/2 = .01$ $t_{.01,21} = 2.518$

$\bar{x} \pm t\frac{s}{\sqrt{n}} = 1{,}192 \pm 2.518\frac{279}{\sqrt{22}} = 1{,}192 \pm 149.78$ = **$1{,}042.22 \le \mu \le 1{,}341.78$**

The figure given by Runzheimer International falls within the confidence interval. Therefore, there is no reason to reject the Runzheimer figure as different from what we are getting based on this sample.

8.19 $n = 20$ df = 19 95% CI $t_{.025,19} = 2.093$

$\bar{x} = 2.36116$ $s = 0.19721$

$2.36116 \pm 2.093\frac{0.1972}{\sqrt{20}} = 2.36116 \pm 0.0923$ = **$2.26886 \le \mu \le 2.45346$**

Point Estimate = **2.36116**

Error = **0.0923**

8.20 $n = 28$ $\bar{x} = 5.335$ $s = 2.016$ df = 28 – 1 = 27

90% Confidence Interval $\alpha/2 = .05$

$t_{.05,27} = 1.703$

$\bar{x} \pm t\frac{s}{\sqrt{n}} = 5.335 \pm 1.703\frac{2.016}{\sqrt{28}} = 5.335 \pm .649 =$ $\mathbf{4.686 \le \mu \le 5.984}$

8.21 $n = 10$ $\bar{x} = 49.8$ $s = 18.22$ df = 10 – 1 = 9

95% Confidence $\alpha/2 = .025$ $t_{.025,9} = 2.262$

$\bar{x} \pm t\frac{s}{\sqrt{n}} = 49.8 \pm 2.262\frac{18.22}{\sqrt{10}} = 49.8 \pm 13.03 =$ $\mathbf{36.77 \le \mu \le 62.83}$

8.22 $n = 14$ 98% confidence $\alpha/2 = .01$ df = 13

$t_{.01,13} = 2.650$

from data: $\bar{x}$ = 152.16 s = 14.42

confidence interval: $\bar{x} \pm t\frac{s}{\sqrt{n}}$ = 152.16 $\pm$ 2.65 $\frac{14.42}{\sqrt{14}}$ =

152.16 $\pm$ 10.21 = **141.95 $\le$ μ $\le$ 162.37**

The point estimate is **152.16**

8.23 n = 17 df = 17 – 1 = 16 99% confidence $\alpha/2$ = .005

$t_{.005,16}$ = 2.921

from data: $\bar{x}$ = 8.06 s = 5.07

confidence interval: $\bar{x} \pm t\frac{s}{\sqrt{n}}$ = 8.06 $\pm$ 2.921 $\frac{5.07}{\sqrt{17}}$ =

8.06 $\pm$ 3.59 = **4.47 $\le$ μ $\le$ 11.65**

8.24 The point estimate is $\bar{x}$ which is **25.4134** hours. The sample size is **26** skiffs. The confidence level is **98%**. The confidence interval is:

$$\bar{x} \pm t\frac{s}{\sqrt{n}} \le \mu \le \bar{x} \pm t\frac{s}{\sqrt{n}}$$

$$= \mathbf{22.8124 \le \mu \le 28.0145}$$

The error of the confidence interval is **2.6011**.

8.25 a) $n = 44$ $\hat{p} = .51$ 99% C.I. $z_{.005} = 2.575$

$\hat{p} \pm z\sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} = .51 \pm 2.575\sqrt{\dfrac{(.51)(.49)}{44}} = .51 \pm .194 =$ **$.316 \le p \le .704$**

b) $n = 300$ $\hat{p} = .82$ 95% C.I. $z_{.025} = 1.96$

$\hat{p} \pm z\sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} = .82 \pm 1.96\sqrt{\dfrac{(.82)(.18)}{300}} = .82 \pm .043 =$ **$.777 \le p \le .863$**

c) $n = 1150$ $\hat{p} = .48$ 90% C.I. $z_{.05} = 1.645$

$\hat{p} \pm z\sqrt{\dfrac{\hat{p}\cdot\hat{q}}{n}} = .48 \pm 1.645\sqrt{\dfrac{(.48)(.52)}{1150}} = .48 \pm .024 =$ **$.456 \le p \le .504$**

d) $n = 95$ $\hat{p} = .32$ 88% C.I. $z_{.06} = 1.555$

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .32 \pm 1.555\sqrt{\frac{(.32)(.68)}{95}} = .32 \pm .074 =$ **$.246 \le p \le .394$**

8.26 a) $n = 116$ $x = 57$ 99% C.I. $z_{.005} = 2.575$

$\hat{p} = \frac{x}{n} = \frac{57}{116} = .49$

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .49 \pm 2.575\sqrt{\frac{(.49)(.51)}{116}} = .49 \pm .12 =$ **$.37 \le p \le .61$**

b) $n = 800$ $x = 479$ 97% C.I. $z_{.015} = 2.17$

$\hat{p} = \frac{x}{n} = \frac{479}{800} = .60$

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$ = .60 ± 2.17 $\sqrt{\frac{(.60)(.40)}{800}}$ = .60 ± .038 = **.562 ≤ *p* ≤ .638**

c) n = 240 x = 106 85% C.I. $z_{.075}$ = 1.44

$\hat{p}$ = $\frac{x}{n} = \frac{106}{240}$ = .44

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$ = .44 ± 1.44 $\sqrt{\frac{(.44)(.56)}{240}}$ = .44 ± .046 = **.394 ≤ *p* ≤ .486**

d) n = 60 x = 21 90% C.I. $z_{.05}$ = 1.645

$\hat{p}$ = $\frac{x}{n} = \frac{21}{60}$ = .35

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$ = .35 ± 1.645 $\sqrt{\frac{(.35)(.65)}{60}}$ = .35 ± .10 = **.25 ≤ *p* ≤ .45**

8.27 n = 85 x = 40 90% C.I. $z_{.05}$ = 1.645

$\hat{p} = \frac{x}{n} = \frac{40}{85} = .47$

$\hat{p} \pm z\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = .47 \pm 1.645\sqrt{\frac{(.47)(.53)}{85}} = .47 \pm .09 =$ **$.38 \le p \le .56$**

95% C.I. $z_{.025} = 1.96$

$\hat{p} \pm z\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = .47 \pm 1.96\sqrt{\frac{(.47)(.53)}{85}} = .47 \pm .11 =$ **$.36 \le p \le .58$**

99% C.I. $z_{.005} = 2.575$

$\hat{p} \pm z\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} = .47 \pm 2.575\sqrt{\frac{(.47)(.53)}{85}} = .47 \pm .14 =$ **$.33 \le p \le .61$**

All other things being constant, as the confidence increased, the width of the interval increased.

8.28 $n = 1003$ $\hat{p} = .255$ 99% CI $z_{.005} = 2.575$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .255 \pm 2.575\sqrt{\frac{(.255)(.745)}{1003}} = .255 \pm .035 = \mathbf{.220 \le p \le .290}$$

$n = 10{,}000$ $\hat{p} = .255$ 99% CI $z_{.005} = 2.575$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .255 \pm 2.575\sqrt{\frac{(.255)(.745)}{10{,}000}} = .255 \pm .011 = \mathbf{.244 \le p \le .266}$$

The confidence interval constructed using n = 1003 is wider than the confidence interval constructed using n = 10,000. One might conclude that, all other things being constant, increasing the sample size reduces the width of the confidence interval.

8.29 $n = 560$ $\hat{p} = .47$ 95% CI $z_{.025} = 1.96$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .47 \pm 1.96\sqrt{\frac{(.47)(.53)}{560}} = .47 \pm .0413 = \mathbf{.4287 \le p \le .5113}$$

$n = 560$ $\hat{p} = .28$ 90% CI $z_{.05} = 1.645$

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .28 \pm 1.645\sqrt{\frac{(.28)(.72)}{560}} = .28 \pm .0312 =$ $\mathbf{.2488 \le p \le .3112}$

8.30 $n = 1250$ $x = 997$ 98% C.I. $z_{.01} = 2.33$

$\hat{p} = \frac{x}{n} = \frac{997}{1250} = .80$

$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .80 \pm 2.33\sqrt{\frac{(.80)(.20)}{1250}} = .80 \pm .026 =$ $\mathbf{.774 \le p \le .826}$

8.31 $n = 3481$ $x = 927$

$$\hat{p} = \frac{x}{n} = \frac{927}{3481} = .266$$

a) $\hat{p}$ = **.266 Point Estimate**

b) 99% C.I. $z_{.005} = 2.575$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .266 \pm 2.575\sqrt{\frac{(.266)(.734)}{3481}} = .266 \pm .019 =$$

$$\mathbf{.247 \le p \le .285}$$

8.32 $n = 89$ $x = 48$ 85% C.I. $z_{.075} = 1.44$

$$\hat{p} = \frac{x}{n} = \frac{48}{89} = .54$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .54 \pm 1.44\sqrt{\frac{(.54)(.46)}{89}} = .54 \pm .076 = \mathbf{.464 \le p \le .616}$$

8.33 $\hat{p} = .63$ $n = 672$ 95% Confidence $z_{.025} = \pm 1.96$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .63 \pm 1.96\sqrt{\frac{(.63)(.37)}{672}} = .63 \pm .0365 = \mathbf{.5935 \le p \le .6665}$$

8.34 $n = 275$ $x = 121$ 98% confidence $z_{.01} = 2.33$

$$\hat{p} = \frac{x}{n} = \frac{121}{275} = .44$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .44 \pm 2.33\sqrt{\frac{(.44)(.56)}{275}} = .44 \pm .07 = \mathbf{.37 \le p \le .51}$$

8.35 a) $n = 12$ $\bar{x} = 28.4$ $s^2 = 44.9$ 99% C.I. df = 12 – 1 = 11

$\chi^2_{.995,11} = 2.60320$ $\chi^2_{.005,11} = 26.7569$

$$\frac{(12-1)(44.9)}{26.7569} \le \sigma^2 \le \frac{(12-1)(44.9)}{2.60320}$$

18.46 ≤ σ^2 ≤ 189.73

b) $n = 7$ $\bar{x} = 4.37$ $s = 1.24$ $s^2 = 1.5376$ 95% C.I. df = 7 – 1 = 6

$\chi^2_{.975,6} = 1.23734$ $\chi^2_{.025,6} = 14.4494$

$$\frac{(7-1)(1.5376)}{14.4494} \le \sigma^2 \le \frac{(7-1)(1.5376)}{1.23734}$$

0.64 ≤ σ^2 ≤ 7.46

c) $n = 20$ $\bar{x} = 105$ $s = 32$ $s^2 = 1024$ 90% C.I. df = 20 – 1 = 19

$\chi^2_{.95,19} = 10.11701$ $\chi^2_{.05,19} = 30.1435$

$$\frac{(20-1)(1024)}{30.1435} \leq \sigma^2 \leq \frac{(20-1)(1024)}{10.11701}$$

645.45 ≤ σ^2 ≤ 1923.10

d) $n = 17$ $s^2 = 18.56$ 80% C.I. df = 17 – 1 = 16

$\chi^2_{.90,16} = 9.31224$ $\chi^2_{.10,16} = 23.5418$

$$\frac{(17-1)(18.56)}{23.5418} \leq \sigma^2 \leq \frac{(17-1)(18.56)}{9.31224}$$

12.61 ≤ σ^2 ≤ 31.89

8.36 $n = 16$ $s^2 = 37.1833$ 98% C.I. df = 16-1 = 15

$\chi^2_{.99,15} = 5.22936$ $\chi^2_{.01,15} = 30.5780$

$$\frac{(16-1)(37.1833)}{30.5780} \le \sigma^2 \le \frac{(16-1)(37.1833)}{5.22936}$$

18.24 $\le \sigma^2 \le$ 106.66

8.37 $n = 20$ $s = 4.3$ $s^2 = 18.49$ 98% C.I. df = 20 – 1 = 19

$\chi^2_{.99,19} = 7.63270$ $\chi^2_{.01,19} = 36.1908$

$$\frac{(20-1)(18.49)}{36.1908} \le \sigma^2 \le \frac{(20-1)(18.49)}{7.63270}$$

9.71 $\le \sigma^2 \le$ 46.03

Point Estimate = s^2 = 18.49

8.38 $n = 15$ $s^2 = 3.067$ 99% C.I. df = 15 – 1 = 14

$\chi^2_{.995,14} = 4.07466$ $\chi^2_{.005,14} = 31.3194$

$$\frac{(15-1)(3.067)}{31.3194} \leq \sigma^2 \leq \frac{(15-1)(3.067)}{4.07466}$$

1.37 ≤ σ^2 ≤ 10.54

8.39 n = 14 s^2 = 26,798,241.76 95% C.I. df = 14 – 1 = 13

Point Estimate = s^2 = 26,798,241.76

$\chi^2_{.975,13}$ = 5.00874 $\chi^2_{.025,13}$ = 24.7356

$$\frac{(14-1)(26{,}798{,}241.76)}{24.7356} \leq \sigma^2 \leq \frac{(14-1)(26{,}798{,}241.76)}{5.00874}$$

14,084,038.51 ≤ σ^2 ≤ 69,553,848.45

8.40 a) σ = 36 E = 5 95% Confidence $z_{.025}$ = 1.96

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.96)^2(36)^2}{5^2} = 199.15$$

Sample 200

b) $\sigma = 4.13$ $E = 1$ 99% Confidence $z_{.005} = 2.575$

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(2.575)^2(4.13)^2}{1^2} = 113.1$$

Sample 114

c) $E = 10$ Range = 500 - 80 = 420

1/4 Range = (.25)(420) = 105

90% Confidence $z_{.05} = 1.645$

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.645)^2(105)^2}{10^2} = 298.3$$

Sample 299

d) $E = 3$ Range = 108 - 50 = 58

1/4 Range = (.25)(58) = 14.5

88% Confidence $z_{.06} = 1.555$

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.555)^2(14.5)^2}{3^2} = 56.5$$

Sample 57

8.41 a) E = .02 p = .40 96% Confidence $z_{.02}$ = 2.05

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(2.05)^2 (.40)(.60)}{(.02)^2} = 2521.5$$

Sample 2522

b) E = .04 p = .50 95% Confidence $z_{.025}$ = 1.96

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.96)^2 (.50)(.50)}{(.04)^2} = 600.25$$

Sample 601

c) E = .05 p = .55 90% Confidence $z_{.05}$ = 1.645

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.645)^2 (.55)(.45)}{(.05)^2} = 267.9$$

Sample 268

d) E =.01 p = .50 99% Confidence $z_{.005}$ = 2.575

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(2.575)^2(.50)(.50)}{(.01)^2} = 16,576.6$$

Sample 16,577

8.42 E = \$200 σ = \$1,000 99% Confidence $z_{.005}$ = 2.575

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(2.575)^2(1000)^2}{200^2} = 165.77$$

Sample 166

8.43 E = \$2 σ = \$12.50 90% Confidence $z_{.05}$ = 1.645

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.645)^2(12.50)^2}{2^2} = 105.7$$

Sample 106

8.44 E = \$100 Range = \$2,500 - \$600 = \$1,900

$\sigma \approx$ 1/4 Range = (.25)(\$1,900) = \$475

90% Confidence $z_{.05}$ = 1.645

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.645)^2(475)^2}{100^2} = 61.05$$

Sample 62

8.45 p = .20 q = .80 E = .02

90% Confidence, $z_{.05}$ = 1.645

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.645)^2(.20)(.80)}{(.02)^2} = 1082.41$$

Sample 1083

8.46 $p = .50$ $q = .50$ $E = .05$

95% Confidence, $z_{.025} = 1.96$

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.96)^2(.50)(.50)}{(.05)^2} = 384.16$$

Sample 385

8.47 $E = .10$ $p = .50$ $q = .50$

95% Confidence, $z_{.025} = 1.96$

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.96)^2(.50)(.50)}{(.10)^2} = 96.04$$

Sample 97

8.48 $\bar{x}$ = 45.6 σ = 7.75 n = 35

80% confidence $z_{.10}$ = 1.28

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 45.6 \pm 1.28\frac{7.75}{\sqrt{35}}$$

= 45.6 $\pm$ 1.68

43.92 $\leq$ μ $\leq$ 47.28

94% confidence $z_{.03}$ = 1.88

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 45.6 \pm 1.88\frac{7.75}{\sqrt{35}}$$

= 45.6 $\pm$ 2.46

43.14 $\leq$ μ $\leq$ 48.06

98% confidence $z_{.01}$ = 2.33

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 45.6 \pm 2.33\frac{7.75}{\sqrt{35}}$$

$$= 45.6 \pm 3.05$$

$$\mathbf{42.55 \le \mu \le 48.65}$$

8.49 $\bar{x}$ = **12.03** (point estimate) s = .4373 n = 10 df = 9

For 90% confidence: $\alpha/2$ = .05 $t_{.05,9}$= 1.833

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 12.03 \pm 1.833\frac{(.4373)}{\sqrt{10}}$$

= 12.03 $\pm$ .25

11.78 $\le$ μ $\le$ 12.28

For 95% confidence: $\alpha/2$ = .025 $t_{.025,9}$ = 2.262

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 12.03 \pm 2.262\frac{(.4373)}{\sqrt{10}}$$

= 12.03 $\pm$ .31

11.72 $\le$ μ $\le$ 12.34

For 99% confidence: $\alpha/2$ = .005 $t_{.005,9}$ = 3.25

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 12.03 \pm 3.25\frac{(.4373)}{\sqrt{10}}$$

= 12.03 $\pm$ .45

11.58 $\le$ μ $\le$ 12.48

8.50 a) $n = 715$ $x = 329$ 95% confidence $z_{.025} = 1.96$

$$\hat{p} = \frac{329}{715} = .46$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .46 \pm 1.96\sqrt{\frac{(.46)(.54)}{715}} = .46 \pm .0365$$

$.4235 \le p \le .4965$

b) $n = 284$ $\hat{p} = .71$ 90% confidence $z_{.05} = 1.645$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .71 \pm 1.645\sqrt{\frac{(.71)(.29)}{284}} = .71 \pm .0443$$

$.6657 \le p \le .7543$

c) $n = 1250$ $\hat{p} = .48$ 95% confidence $z_{.025} = 1.96$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .48 \pm 1.96\sqrt{\frac{(.48)(.52)}{1250}} = .48 \pm .0277$$

$.4523 \le p \le .5077$

d) n = 457 x = 270 98% confidence $z_{.01}$ = 2.33

$$\hat{p} = \frac{270}{457}$$ = .591

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .591 \pm 2.33\sqrt{\frac{(.591)(.409)}{457}}$$ = .591 $\pm$ .0536

.5374 $\le$ *p* $\le$ .6446

8.51 n = 10 s = 7.40045 s^2 = 54.7667 df = 10 – 1 = 9

90% confidence, $\alpha/2$ = .05 1 - $\alpha/2$ = .95

$\chi^2_{.95,9}$ = 3.32512 $\chi^2_{.05,9}$ = 16.9190

$$\frac{(10-1)(54.7667)}{16.9190} \le \sigma^2 \le \frac{(10-1)(54.7667)}{3.32512}$$

29.133 $\le$ σ^2 $\le$ 148.235

95% confidence, $\alpha/2 = .025$ $1 - \alpha/2 = .975$

$\chi^2_{.975,9} = 2.70039$ $\chi^2_{.025,9} = 19.0228$

$$\frac{(10-1)(54.7667)}{19.0228} \leq \sigma^2 \leq \frac{(10-1)(54.7667)}{2.70039}$$

$25.911 \leq \sigma^2 \leq 182.529$

8.52 a) σ = 44 E = 3 95% confidence $z_{.025}$ = 1.96

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.96)^2(44)^2}{3^2} = 826.4$$

Sample 827

b) E = 2 Range = 88 - 20 = 68

use σ = 1/4(range) = (.25)(68) = 17

90% confidence $z_{.05}$ = 1.645

$$\frac{z^2\sigma^2}{E^2} = \frac{(1.645)^2(17)^2}{2^2} = 195.5$$

Sample 196

c) E = .04 p = .50 q = .50

98% confidence $z_{.01}$ = 2.33

$$\frac{z^2 p\cdot q}{E^2} = \frac{(2.33)^2(.50)(.50)}{(.04)^2} = 848.3$$

Sample 849

d) $E = .03$ $p = .70$ $q = .30$

95% confidence $z_{.025} = 1.96$

$$\frac{z^2 p \cdot q}{E^2} = \frac{(1.96)^2(.70)(.30)}{(.03)^2}$$

$= 896.4$

Sample 897

8.53 $n = 17$ $\bar{x} = 10.765$ $s = 2.223$ df $= 17 - 1 = 16$

99% confidence $\alpha/2 = .005$ $t_{.005,16} = 2.921$

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 10.765 \pm 2.921\frac{2.223}{\sqrt{17}} = 10.765 \pm 1.575$$

9.19 $\leq$ *μ* $\leq$ 12.34

8.54 $p = .40$ $E = .03$ 90% Confidence $z_{.05} = 1.645$

$$n = \frac{z^2 p \cdot q}{E^2} = \frac{(1.645)^2(.40)(.60)}{(.03)^2} = 721.61$$

Sample 722

8.55 $n = 17$ $s^2 = 4.941$ 99% C.I. df $= 17 - 1 = 16$

$\chi^2_{.995,16} = 5.14216$ $\chi^2_{.005,16} = 34.2671$

$$\frac{(17-1)(4.941)}{34.2671} \leq \sigma^2 \leq \frac{(17-1)(4.941)}{5.14216}$$

$$2.307 \le \sigma^2 \le 15.374$$

8.56 $n = 45$ $\bar{x} = 213$ $\sigma = 48$

98% Confidence $z_{.01} = 2.33$

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 213 \pm 2.33\frac{48}{\sqrt{45}}$$

$= 213 \pm 16.67$

$$196.33 \le \mu \le 229.67$$

8.57 $n = 39$ $\bar{x} = 37.256$ $\sigma = 3.891$

90% confidence $z_{.05} = 1.645$

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 37.256 \pm 1.645\frac{3.891}{\sqrt{39}} = 37.256 \pm 1.025$$

36.231 $\leq$ μ $\leq$ 38.281

8.58 $\sigma = 6$ $E = 1$ 98% Confidence $z_{.98} = 2.33$

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(2.33)^2(6)^2}{1^2} = 195.44$$

Sample 196

8.59 $n = 1,255$ $x = 714$ 95% Confidence $z_{.025} = 1.96$

$$\hat{p} = \frac{714}{1255} = .569$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .569 \pm 1.96\sqrt{\frac{(.569)(.431)}{1{,}255}}$$

$$= \quad .569 \pm .027$$

$$\mathbf{.542 \leq \boldsymbol{p} \leq .596}$$

8.60 n = 41 s = 21 $\bar{x} = 128$ 98% C.I. df = 41 – 1 = 40

$t_{.01,40}$ = 2.423

Point Estimate = $128

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 128 \pm 2.423\frac{21}{\sqrt{41}}$$ = 128 $\pm$ 7.947

120.053 $\leq$ μ $\leq$ 135.947

Interval Width = 135.947 – 120.053 = 15.894

8.61 n = 60 $\bar{x}$ = 6.717 σ = 3.06 N = 300

98% Confidence $z_{.01}$ = 2.33

$$\bar{x} \pm z\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 6.717 \pm 2.33\frac{3.06}{\sqrt{60}}\sqrt{\frac{300-60}{300-1}}$$ =

6.717 ± 0.825

$5.892 \le \mu \le 7.542$

8.62 E = \$20 Range = \$600 - \$30 = \$570

1/4 Range = (.25)(\$570) = \$142.50

95% Confidence $z_{.025} = 1.96$

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(1.96)^2(142.50)^2}{20^2} = 195.02$$

Sample 196

8.63 n = 245 x = 189 90% Confidence $z_{.05}$= 1.645

$$\hat{p} = \frac{x}{n} = \frac{189}{245}$$

= .77

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .77 \pm 1.645\sqrt{\frac{(.77)(.23)}{245}}$$

= .77 ± .044

.726 $\leq$ *p* $\leq$ .814

8.64 n = 90 x = 30 95% Confidence $z_{.025}$ = 1.96

$$\hat{p} = \frac{x}{n} = \frac{30}{90}$$

= .33

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .33 \pm 1.96\sqrt{\frac{(.33)(.67)}{90}}$$

= .33 ± .097

.233 $\leq$ *p* $\leq$ .427

8.65 n = 12 $\bar{x}$ = 43.7 s^2 = 228 df = 12 – 1 = 11 95% C.I.

$t_{.025,11} = 2.201$

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 43.7 \pm 2.201\frac{\sqrt{228}}{\sqrt{12}}$$

$= 43.7 \pm 9.59$

$34.11 \leq \mu \leq 53.29$

$\chi^2_{.99,11} = 3.05350$ $\chi^2_{.01,11} = 24.7250$

$$\frac{(12-1)(228)}{24.7250} \leq \sigma^2 \leq \frac{(12-1)(228)}{3.05350}$$

$101.44 \leq \sigma^2 \leq 821.35$

8.66 $n = 27$ $\bar{x} = 4.82$ $s = 0.37$ df = 26

95% CI: $t_{.025,26} = 2.056$

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 4.82 \pm 2.056\frac{0.37}{\sqrt{27}}$$

$= 4.82 \pm .1464$

$4.6736 \leq \mu \leq 4.9664$

Since 4.50 is not in the interval, we are 95% confident that μ does not equal 4.50.

8.67 $n = 77$ $\bar{x} = 2.48$ $\sigma = 12$

95% Confidence $z_{.025} = 1.96$

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 2.48 \pm 1.96\frac{12}{\sqrt{77}}$$

$= 2.48 \pm 2.68$

$-0.20 \leq \mu \leq 5.16$

The point estimate is 2.48

The interval is inconclusive. It says that we are 95% confident that the average arrival time is somewhere between .20 of a minute (12 seconds) early and 5.16 minutes late. Since zero is in the interval, there is a possibility that, on average, the flights are on time.

8.68 $n = 560$ $\hat{p} = .33$

99% Confidence $z_{.005} = 2.575$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\cdot\hat{q}}{n}} = .33 \pm 2.575\sqrt{\frac{(.33)(.67)}{560}}$$

$= .33 \pm .05$

$.28 \leq p \leq .38$

8.69 $p = .50$ $E = .05$ 98% Confidence $z_{.01} = 2.33$

$$\frac{z^2 p \cdot q}{E^2} = \frac{(2.33)^2(.50)(.50)}{(.05)^2}$$

$= 542.89$

Sample 543

8.70 $n = 27$ $\bar{x} = 2.10$ $s = 0.86$ df = 27 - 1 = 26

98% confidence $\alpha/2 = .01$ $t_{.01,26} = 2.479$

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 2.10 \pm 2.479\frac{0.86}{\sqrt{27}}$$

$= 2.10 \pm 0.41$

$1.69 \leq \mu \leq 2.51$

8.71 $n = 23$ df = 23 – 1 = 22 $s = .0631455$ 90% C.I.

$\chi^2_{.95,22} = 12.33801$ $\chi^2_{.05,22} = 33.9245$

$$\frac{(23-1)(.0631455)^2}{33.9245} \le \sigma^2 \le \frac{(23-1)(.0631455)^2}{12.33801}$$

.0026 ≤ σ^2 ≤ .0071

8.72 n = 39 $\bar{x}$ = 1.294 σ = 0.205 99% Confidence

$z_{.005}$ = 2.575

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} = 1.294 \pm 2.575\frac{0.205}{\sqrt{39}}$$

= 1.294 ± .085

1.209 ≤ μ ≤ 1.379

8.73 The sample mean fill for the 58 cans is 11.9788 oz. with a standard deviation of .0536 oz. The 99% confidence interval for the population fill is 11.9607 oz. to 11.9969 oz. which does not include 12 oz. We are 99% confident that the population mean is not 12 oz., indicating that the machine may be under filling the cans.

8.74 The point estimate for the average length of burn of the new bulb is 2198.217 hours. Eighty-four bulbs were included in this study. A 90% confidence interval can be constructed from the information given. The error of the confidence interval is $\pm$ 27.76691. Combining this with the point estimate yields the 90% confidence interval of 2198.217 $\pm$ 27.76691 = 2170.450 $\leq \mu \leq$ 2225.984.

8.75 The point estimate for the average age of a first time buyer is 27.63 years. The sample of 21 buyers produces a standard deviation of 6.54 years. We are 98% confident that the actual population mean age of a first-time home buyer is between 24.0222 years and 31.2378 years.

8.76 A poll of 781 American workers was taken. Of these, 506 drive their cars to work. Thus, the point estimate for the population proportion is 506/781 = .647887. A 95% confidence interval to estimate the population proportion shows that we are 95% confident that the actual value lies between .61324 and .681413. The error of this interval is $\pm$ .0340865.

Chapter 9

Statistical Inference:

Hypothesis Testing for Single Populations

LEARNING OBJECTIVES

The main objective of Chapter 9 is to help you to learn how to test hypotheses on single populations, thereby enabling you to:

1. Understand the logic of hypothesis testing and know how to establish null and alternate hypotheses.
2. Understand Type I and Type II errors and know how to solve for Type II errors.
3. Know how to implement the HTAB system to test hypotheses.
4. Test hypotheses about a single population mean when σ is known.
5. Test hypotheses about a single population mean when σ is unknown.
6. Test hypotheses about a single population proportion.
7. Test hypotheses about a single population variance.

CHAPTER TEACHING STRATEGY

For some instructors, this chapter is the cornerstone of the first statistics course.

Hypothesis testing presents the logic in which ideas, theories, etc., are scientifically

examined. The student can be made aware that much of the development of concepts to this point including sampling, level of data measurement, descriptive tools such as mean and standard deviation, probability, and distributions pave the way for testing hypotheses. Often students (and instructors) will say "Why do we need to test this hypothesis when we can make a decision by examining the data?" Sometimes it is true that examining the data could allow hypothesis decisions to be made. However, by using the methodology and structure of hypothesis testing even in "obvious" situations, the researcher has added credibility and rigor to his/her findings. Some statisticians actually report findings in a court of law as an expert witness. Others report their findings in a journal, to the public, to the corporate board, to a client, or to their manager. In each case, by using the hypothesis testing method rather than a "seat of the pants" judgment, the researcher stands on a much firmer foundation by using the principles of hypothesis testing and random sampling. Chapter 9 brings together many of the tools developed to this point and formalizes a procedure for testing hypotheses.

The statistical hypotheses are set up as to contain all possible decisions. The

two-tailed test always has = and ≠ in the null and alternative hypothesis. One-tailed tests are presented with = in the null hypothesis and either > or < in the alternative hypothesis. If in doubt, the researcher should use a two-tailed test. Chapter 9 begins with a two-tailed test example. Often that which the researcher wants to demonstrate true or prove true is set up as the alternative hypothesis. The null hypothesis is that the new theory or idea is not true, the status quo is still true, or that there is no difference. The null hypothesis is assumed to be true before the process begins. Some researchers liken this procedure to a court of law where the defendant is presumed innocent (assume null is true - nothing has happened). Evidence is brought before the judge or jury. If enough evidence is presented, then the null hypothesis (defendant innocent) can no longer be accepted or assumed true. The null hypothesis is rejected as not true and the alternate hypothesis is accepted as true by default. Emphasize that the researcher needs to make a decision after examining the observed statistic.

Some of the key concepts in this chapter are one-tailed and two-tailed test and Type I and Type II error. In order for a one-tailed test to be conducted, the problem must include some suggestion of a direction to be tested. If the student sees such words as greater, less than, more than, higher, younger, etc., then he/she knows to use a one-tail test. If no direction is given (test to determine if there is a "difference"), then a two-tailed test is called for. Ultimately, students will see that the only effect of using a one-tailed test versus a two-tailed test is on the critical table value. A one-tailed test uses all of the value of alpha in one tail. A two-tailed test splits alpha and uses alpha/2 in each tail thus creating a critical value that is further out in the distribution. The result is that (all things being the same) it is more difficult to reject the null hypothesis with a two-tailed test. Many computer packages such as MINITAB include in the results a *p*-value. If you

designate that the hypothesis test is a two-tailed test, the computer will double the *p*-value so that it can be compared directly to alpha.

In discussing Type I and Type II errors, there are a few things to consider. Once a decision is made regarding the null hypothesis, there is a possibility that the decision is correct or that an error has been made. Since the researcher virtually never knows for certain whether the null hypothesis was actually true or not, a probability of committing one of these errors can be computed. Emphasize with the students that a researcher can never commit a Type I error and a Type II error at the same time. This is so because a Type I error can only be committed when the null hypothesis is rejected and a Type II error can only be committed when the decision is to not reject the null hypothesis. Type I and Type II errors are important concepts for managerial students to understand even beyond the realm of statistical hypothesis testing. For example, if a manager decides to fire or not fire an employee based on some evidence collected, he/she could be committing a Type I or a Type II error depending on the decision. If the production manager decides to stop the production line because of evidence of faulty raw materials, he/she might be committing a Type I error. On the other hand, if the manager fails to shut the production line down even when faced with evidence of faulty raw materials, he/she might be committing a Type II error.

The student can be told that there are some widely accepted values for alpha (probability of committing a Type I error) in the research world and that a value is usually selected before the research begins. On the other hand, since the value of Beta (probability of committing a Type II error) varies with every possible alternate value of the parameter being tested, Beta is usually examined and computed over a range of possible values of that parameter. As you can see, the concepts of hypothesis testing are difficult and represent higher levels of learning (logic, transfer, etc.). Student understanding of these concepts will improve as you work your way through the techniques in this chapter and in chapter 10.

CHAPTER OUTLINE

9.1 Introduction to Hypothesis Testing

Types of Hypotheses

Research Hypotheses

Statistical Hypotheses

Substantive Hypotheses

Using the HTAB System to Test Hypotheses

Rejection and Non-rejection Regions

Type I and Type II errors

9.2 Testing Hypotheses About a Population Mean Using the *z* Statistic (σ known)

Testing the Mean with a Finite Population

Using the *p*-Value Method to Test Hypotheses

Using the Critical Value Method to Test Hypotheses

Using the Computer to Test Hypotheses about a Population Mean Using

KEY TERMS

Alpha(α)	One-tailed Test
Alternative Hypothesis	Operating-Characteristic Curve (OC)
Beta(β)	*p*-Value Method
Critical Value	Power
Critical Value Method	Power Curve
Hypothesis	Rejection Region
Hypothesis Testing	Research Hypothesis
Level of Significance	Statistical Hypothesis
Nonrejection Region	Substantive Result
Null Hypothesis	Two-Tailed Test
Observed Significance Level	Type I Error
	Observed Value
	Type II Error

SOLUTIONS TO PROBLEMS IN CHAPTER 9

9.1 a) H_o: $\mu = 25$

H_a: $\mu \neq 25$

$\bar{x}$ = 28.1 n = 57 σ = 8.46 α = .01

For two-tail, $\alpha/2$ = .005 z_c = 2.575

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{28.1-25}{\frac{8.46}{\sqrt{57}}} = \mathbf{2.77}$$

observed z = 2.77 > z_c = 2.575

Reject the null hypothesis

b) from Table A.5, inside area between z = 0 and z = 2.77 is .4972

p-value = .5000 - .4972 = **.0028**

Since the p-value of .0028 is less than $\alpha/2$ = .005, the decision is to:

Reject the null hypothesis

c) critical mean values:

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\pm 2.575 = \frac{\bar{x}_c - 25}{\frac{8.46}{\sqrt{57}}}$$

$$\bar{x}_c = 25 \pm 2.885$$

$\bar{x}_c$ = **27.885 (upper value)**

$\bar{x}_c$ = **22.115 (lower value)**

9.2 H_o: $\mu = 7.48$

H_a: $\mu < 7.48$

$\bar{x} = 6.91$ $n = 24$ $\sigma = 1.21$ $\alpha = .01$

For one-tail, $\alpha = .01$ $z_c = -2.33$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{6.91 - 7.48}{\frac{1.21}{\sqrt{24}}} = \mathbf{-2.31}$$

observed $z = -2.31 > z_c = -2.33$

Fail to reject the null hypothesis

9.3 a) H_o: $\mu = 1{,}200$

H_a: $\mu > 1{,}200$

$\bar{x} = 1{,}215$ $n = 113$ $\sigma = 100$ $\alpha = .10$

For one-tail, $\alpha = .10$ $z_c = 1.28$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1{,}215-1{,}200}{\frac{100}{\sqrt{113}}} = \mathbf{1.59}$$

observed $z = 1.59 > z_c = 1.28$

Reject the null hypothesis

b) Probability > observed $z = 1.59$ is .5000 - .4441 = **.0559** (the *p*-value) which is

less than $\alpha = .10$.

Reject the null hypothesis.

c) Critical mean value:

$$z_c = \frac{\bar{x}_c-\mu}{\frac{\sigma}{\sqrt{n}}}$$

$$1.28 = \frac{\bar{x}_c - 1{,}200}{\frac{100}{\sqrt{113}}}$$

$\bar{x}_c = 1{,}200 + 12.04$

Since the observed $\bar{x} = 1{,}215$ is greater than the critical $\bar{x} = 1212.04$, the decision is to reject the null hypothesis.

9.4 H$_o$: $\mu = 82$

H$_a$: $\mu < 82$

$\bar{x} = 78.125$ $n = 32$ $\sigma = 9.184$ $\alpha = .01$

$z_{.01} = -2.33$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{78.125-82}{\frac{9.184}{\sqrt{32}}} = \mathbf{-2.39}$$

Since observed $z = -2.39 < z_{.01} = -2.33$

Reject the null hypothesis

Statistically, we can conclude that urban air soot is significantly lower. From a business and community point-of-view, assuming that the sample result is representative of how the air actually is now; is a reduction of suspended particles from 82 to 78.125 really an *important* reduction in air pollution (is it substantive)? Certainly it marks an important first step and perhaps a significant start. Whether or not it would really make a difference in the quality of life for people in the city of St. Louis remains to be seen. Most likely, politicians and city chamber of commerce folks would jump on such results as indications of improvement in city conditions.

9.5 H$_0$: μ = \$424.20

H_a: $\mu \neq \$424.20$

$\bar{x} = \$432.69$ $n = 54$ $\sigma = \$33.90$ $\alpha = .05$

2-tailed test, $\alpha/2 = .025$ $z_{.025} = \underline{+}\ 1.96$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{432.69-424.20}{\frac{33.90}{\sqrt{54}}} = \mathbf{1.84}$$

Since the observed $z = 1.85 < z_{.025} = 1.96$, the decision is to **fail** to reject the null hypothesis.

9.6 H_0: μ = \$62,600

H_a: μ < \$62,600

$\bar{x}$ = \$58,974 n = 18 σ = \$7,810 α = .01

1-tailed test, α = .01 $z_{.01}$ = -2.33

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{58{,}974 - 62{,}600}{\frac{7{,}810}{\sqrt{18}}} = \mathbf{-1.97}$$

Since the observed z = -1.97 > $z_{.01}$ = -2.33, the decision is to **fail** to reject the null hypothesis.

9.7 H0: ⌙ = 5

H_a: $\mu \neq 5$

$\bar{x}$ = 5.0611 n = 42 N = 650 σ = 0.2803 α = .10

2-tailed test, $\alpha/2$ = .05 $z_{.05} = \pm 1.645$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} = \frac{5.0611-5}{\frac{0.2803}{\sqrt{42}}\sqrt{\frac{650-42}{650-1}}} = \mathbf{1.46}$$

Since the observed $z = 1.46 < z_{.05} = 1.645$, the decision is to **fail to reject** the

null hypothesis.

9.8 H_o: $\mu = 18.2$

H_a: $\mu < 18.2$

$\bar{x} = 15.6$ $n = 32$ $\sigma = 2.3$ $\alpha = .10$

For one-tail, $\alpha = .10$, $z_{.10} = -1.28$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{15.6-18.2}{\frac{2.3}{\sqrt{32}}} = \mathbf{-6.39}$$

Since the observed $z = -6.39 < z_{.10} = -1.28$, the decision is to

Reject the null hypothesis

9.9 H_o: $\mu = \$4,292$

H_a: $\mu < \$4,292$

$\bar{x}$ = \$4,008 n = 55 σ = \$386 α = .01

For one-tailed test, α = .01, $z_{.01}$ = -2.33

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\$4{,}008 - \$4{,}292}{\frac{\$386}{\sqrt{55}}} = \mathbf{-5.46}$$

Since the observed z = -5.46 < $z_{.01}$ = -2.33, the decision is to

Reject the null hypothesis

The CEO could use this information as a way of discrediting the Runzheimer study and using her own figures in recruiting people and in discussing relocation options. In such a case, this could be a substantive finding. However, one must ask if the difference between \$4,292 and \$4,008 is really an important difference in monthly rental expense. Certainly, Paris is expensive either way. However, an almost \$300 difference in monthly rental cost is a nontrivial amount for most people and therefore might be considered substantive.

9.10 H_o: μ = 123

H_a: μ > 123

α = .05 n = 40 40 people were sampled

$\bar{x}$ = 132.36 s = 27.68

This is a one-tailed test. Since the p-value = **.016**, we

reject the null hypothesis at α = .05.

The average water usage per person is greater than 123 gallons.

9.11 $n = 20$ $\bar{x} = 16.45$ $s = 3.59$ df = 20 - 1 = 19 $\alpha = .05$

H_o: $\mu = 16$

H_a: $\mu \neq 16$

For two-tail test, $\alpha/2 = .025$, critical $t_{.025,19} = \pm 2.093$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{16.45-16}{\frac{3.59}{\sqrt{20}}} = \mathbf{0.56}$$

Observed $t = 0.56 < t_{.025,19} = 2.093$

The decision is to **Fail to reject the null hypothesis**

9.12 $n = 51$ $\bar{x} = 58.42$ $s^2 = 25.68$ df = 51 - 1 = 50 $\alpha = .01$

H_o: $\mu = 60$

H_a: $\mu < 60$

For one-tail test, $\alpha = .01$ critical $t_{.01,50} = -2.403$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{58.42 - 60}{\frac{\sqrt{25.68}}{\sqrt{51}}} = \mathbf{-2.23}$$

Observed $t = -2.23 > t_{.01,7} = -2.403$

The decision is to **Fail to reject the null hypothesis**

9.13 $n = 11$ $\bar{x} = 1,235.36$ $s = 103.81$ df = 11 - 1 = 10 $\alpha = .05$

H$_o$: $\mu = 1,160$

H$_a$: $\mu > 1,160$

or one-tail test, $\alpha = .05$ critical $t_{.05,10} = 1.812$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{1,236.36-1,160}{\frac{103.81}{\sqrt{11}}} = \mathbf{2.44}$$

Observed $t = 2.44 > t_{.05,10} = 1.812$

The decision is to **Reject the null hypothesis**

9.14 $n = 20$ $\bar{x} = 8.37$ $s = .1895$ df = 20-1 = 19 $\alpha = .01$

H$_o$: $\mu = 8.3$

H$_a$: $\mu \neq 8.3$

For two-tail test, $\alpha/2 = .005$ critical $t_{.005,19} = \pm 2.861$

$t = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \dfrac{8.37-8.3}{\frac{.1895}{\sqrt{20}}} =$ **1.65**

Observed $t = 1.65 < t_{.005,19} = 2.861$

The decision is to **Fail to reject the null hypothesis**

9.15 $n = 12$ $\bar{x} = 1.85083$ $s = .02353$ df = 12 - 1 = 11 $\alpha = .10$

H$_0$: $\mu = 1.84$

H$_a$: $\mu \neq 1.84$

For a two-tailed test, $\alpha/2 = .05$ critical $t_{.05,11} = 1.796$

$t = \dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \dfrac{1.85083-1.84}{\frac{.02353}{\sqrt{12}}} =$ **1.59**

Since $t = 1.59 < t_{11,.05} = 1.796$,

The decision is to **fail to reject the null hypothesis**.

9.16 $n = 25$ $\bar{x} = 3.1948$ $s = .0889$ df = 25 - 1 = 24 $\alpha = .01$

H_o: μ = \$3.16

H_a: μ > \$3.16

For one-tail test, = .01 Critical $t_{.01,24}$ = 2.492

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3.1948-3.16}{\frac{.0889}{\sqrt{25}}} = \mathbf{1.96}$$

Observed t = 1.96 < $t_{.01,24}$ = 2.492

The decision is to **Fail to reject the null hypothesis**

9.17 n = 19 $\bar{x}$ = \$31.67 s = \$1.29 df = 19 – 1 = 18

α = .05

H_0: μ = \$32.28

H_a: $\mu \neq$ \$32.28

Two-tailed test, $\alpha/2$ = .025 $t_{.025,18} = \pm$ 2.101

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{31.67-32.28}{\frac{1.29}{\sqrt{19}}} = \mathbf{-2.06}$$

The observed t = -2.06 > $t_{.025,18}$ = -2.101,

The decision is to **fail to reject the null hypothesis**

9.18 n = 61 $\bar{x}$ = 3.72 s = 0.65 df = 61 – 1 = 60
α = .01

H_0: μ = 3.51

H_a: μ > 3.51

One-tailed test, α = .01 $t_{.01,60}$ = 2.390

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3.72-3.51}{\frac{0.65}{\sqrt{61}}} = \mathbf{2.52}$$

The observed t = 2.52 > $t_{.01,60}$ = 2.390,

The decision is to **reject the null hypothesis**

9.19 $n = 22$ $\bar{x} = 1031.32$ $s = 240.37$ df = 22 – 1 = 21
$\alpha = .05$

H$_0$: $\mu = 1135$

H$_a$: $\mu \neq 1135$

Two-tailed test, $\alpha/2 = .025$ $t_{.025,21} = \pm 2.080$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{1031.32-1135}{\frac{240.37}{\sqrt{22}}} = \mathbf{-2.02}$$

The observed $t = -2.02 > t_{.025,21} = -2.080$,

The decision is to **fail to reject the null hypothesis**

9.20 $n = 12$ $\bar{x} = 42.167$ $s = 9.124$ df = 12 – 1 = 11
$\alpha = .01$

H$_0$: $\mu = 46$

H$_a$: $\mu < 46$

One-tailed test, $\alpha = .01$ $t_{.01,11} = -2.718$

$$t = \frac{\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{42.167-46}{\frac{9.124}{\sqrt{12}}}} = \mathbf{-1.46}$$

The observed t = -1.46 > $t_{.01,11}$ = -2.718,

The decision is to **fail to reject the null hypothesis**

9.21 n = 26 $\bar{x}$ = 19.534 minutes s = 4.100 minutes α = .05

H_0: μ = 19

H_a: $\mu \neq$ 19

Two-tailed test, α/2 = .025, critical t value = $\pm$ 2.06

Observed t value = 0.66. Since the observed t = 0.66 < critical t value = 2.06,

The decision is to **fail to reject the null hypothesis**.

Since the Excel p-value = .256 > α/2 = .025 and MINITAB p-value =.513 > .05, the decision is to **fail to reject the null hypothesis.**

She would <u>not</u> conclude that her city is any different from the ones in the

national survey.

9.22 H$_o$: p = .45

H$_a$: p > .45

n = 310 $\hat{p}$ = .465 α = .05

For one-tail, α = .05 $z_{.05}$ = 1.645

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.465-.45}{\sqrt{\frac{(.45)(.55)}{310}}} = \mathbf{0.53}$$

observed z = 0.53 < $z_{.05}$ = 1.645

The decision is to **Fail to reject the null hypothesis**

9.23 H$_o$: p = 0.63

H$_a$: p < 0.63

n = 100 x = 55 $\hat{p} = \frac{x}{n} = \frac{55}{100}$ = .55

For one-tail, α = .01 $z_{.01}$ = -2.33

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.55 - .63}{\sqrt{\frac{(.63)(.37)}{100}}} = \mathbf{-1.66}$$

observed z = -1.66 > z_c = -2.33

The decision is to **Fail to reject the null hypothesis**

9.24 H_o: $p = .29$

H_a: $p \neq .29$

$n = 740$ $x = 207$ $\hat{p} = \frac{x}{n} = \frac{207}{740} = .28$ $\alpha = .05$

For two-tail, $\alpha/2 = .025$ $z_{.025} = \pm 1.96$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.28 - .29}{\sqrt{\frac{(.29)(.71)}{740}}} = \mathbf{-0.60}$$

observed $z = -0.60 > z_c = -1.96$

The decision is to **Fail to reject the null hypothesis**

<u>*p*-Value Method</u>:

$z = -0.60$

from Table A.5, area = .2257

Area in tail = .5000 - .2257 = **.2743** which is the *p*-value

Since the p-value = .2743 > $\alpha/2$ = .025, the decision is to **Fail to reject the null hypothesis**

<u>Solving for critical values</u>:

$$z = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$\pm 1.96 = \frac{\hat{p}_c - .29}{\sqrt{\frac{(.29)(.71)}{740}}}$$

$$\hat{p}_c = .29 \pm .033$$

.257 and .323 are the critical values

Since $\hat{p}$ = .28 is not outside critical values in tails, the decision is to **Fail to reject the null hypothesis**

9.25 H_o: $p = .48$

H_a: $p \neq .48$

$n = 380$ $x = 164$ $\alpha = .01$ $\alpha/2 = .005$ $z_{.005} = \pm 2.575$

$$\hat{p} = \frac{x}{n} = \frac{164}{380} = .4316$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.4316 - .48}{\sqrt{\frac{(.48)(.52)}{380}}} = \mathbf{-1.89}$$

Since the observed $z = -1.89$ is greater than $z_{.005} = -2.575$, The decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that the proportion is any different than .48.

9.26 H_o: $p = .79$

H_a: $p < .79$

$n = 415$ $x = 303$ $\alpha = .01$ $z_{.01} = -2.33$

$$\hat{p} = \frac{x}{n} = \frac{303}{415} = .7301$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{7301-.79}{\sqrt{\frac{(.79)(.21)}{415}}} = \mathbf{-3.00}$$

Since the observed z = -3.00 is less than $z_{.01}$= -2.33, The decision is to **reject the null hypothesis**.

9.27 H_o: $p = .31$

H_a: $p \neq .31$

$n = 600$ $x = 200$ $\alpha = .10$ $\alpha/2 = .05$ $z_{.005} = \pm 1.645$

$$\hat{p} = \frac{x}{n} = \frac{200}{600} = .3333$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.3333 - .31}{\sqrt{\frac{(.31)(.69)}{600}}} = \mathbf{1.23}$$

Since the observed $z = 1.23$ is less than $z_{.005} = 1.645$, The decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that the proportion is any different than .31.

H_o: $p = .24$

H_a: $p < .24$

$n = 600$ $x = 130$ $\alpha = .05$ $z_{.05} = -1.645$

$$\hat{p} = \frac{x}{n} = \frac{130}{600} = .2167$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.2167 - .24}{\sqrt{\frac{(.24)(.76)}{600}}} = \mathbf{-1.34}$$

Since the observed z = -1.34 is greater than $z_{.05}$= -1.645, The decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that the proportion is less than .24.

9.28 H_o: p = .18

H_a: p > .18

n = 376 $\hat{p}$ = .22 α = .01

one-tailed test, $z_{.01}$ = 2.33

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.22-.18}{\sqrt{\frac{(.18)(.82)}{376}}} = \mathbf{2.02}$$

Since the observed z = 2.02 is less than $z_{.01}$= 2.33, The decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that the proportion is greater than .18.

9.29 H_o: p = .32

H_a: p < .32

n = 118 x = 22 $\hat{p} = \frac{x}{n} = \frac{22}{118}$ = .1864 α = .05

For one-tailed test, $z_{.05}$ = -1.645

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.1864 - .32}{\sqrt{\frac{(.32)(.68)}{118}}} = \mathbf{-3.11}$$

Observed z = -3.11 < $z_{.05}$ –1.645

Since the observed z = -3.11 is less than $z_{.05}$= -1.645, the decision is to **reject the null hypothesis**.

9.30 H_o: p = .47

H_a: $p \neq .47$

n = 67 x = 40 α = .05 $\alpha/2$ = .025

For a two-tailed test, $z_{.025} = \pm 1.96$

$$\hat{p} = \frac{x}{n} = \frac{40}{67} = .597$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.597-.47}{\sqrt{\frac{(.47)(.53)}{67}}} = \mathbf{2.08}$$

Since the observed z = 2.08 is greater than $z_{.025}$= 1.96, The decision is to **reject the null hypothesis**.

9.31 a) H_0: $\sigma^2 = 20$ α = .05 n = 15 df = 15 – 1 = 14 s^2 = 32

H_a: $\sigma^2 > 20$

$\chi^2_{.05,14} = 23.6848$

$$\chi^2 = \frac{(15-1)(32)}{20} = \mathbf{22.4}$$

Since $\chi^2 = 22.4 < \chi^2_{.05,14} = 23.6848$, the decision is to **fail to reject the null**

hypothesis.

b) H_0: $\sigma^2 = 8.5$ $\alpha = .10$ $\alpha/2 = .05$ $n = 22$ df = n-1 = 21 $s^2 = 17$

H_a: $\sigma^2 \neq 8.5$

$\chi^2_{.05,21} = 32.6706$

$$\chi^2 = \frac{(22-1)(17)}{8.5} = \mathbf{42}$$

Since $\chi^2 = 42 > \chi^2_{.05,21} = 32.6706$, the decision is to **reject the null hypothesis**.

c) H_0: $\sigma^2 = 45$ $\alpha = .01$ $n = 8$ df = n – 1 = 7 $s = 4.12$

H_a: $\sigma^2 < 45$

$\chi^2_{.01,7} = 18.4753$

$\chi^2 = \dfrac{(8-1)(4.12)^2}{45} =$ **2.64**

Since $\chi^2 = 2.64 < \chi^2_{.01,7} = 18.4753$, the decision is to **fail to reject the null hypothesis**.

d) H_0: $\sigma^2 = 5$ $\alpha = .05$ $\alpha/2 = .025$ $n = 11$ df = 11 – 1 = 10 $s^2 = 1.2$

H_a: $\sigma^2 \neq 5$

$\chi^2_{.025,10} = 20.4832$ $\chi^2_{.975,10} = 3.24696$

$\chi^2 = \dfrac{(11-1)(1.2)}{5} =$ **2.4**

Since $\chi^2 = 2.4 < \chi^2_{.975,10} = 3.24696$, the decision is to **reject the null hypothesis**.

9.32 H_0: $\sigma^2 = 14$ $\alpha = .05$ $\alpha/2 = .025$ $n = 12$ df = 12 – 1 = 11 $s^2 = 30.0833$

H_a: $\sigma^2 \neq 14$

$\chi^2_{.025,11} = 21.9200$ $\chi^2_{.975,11} = 3.81574$

$$\chi^2 = \frac{(12-1)(30.0833)}{14} = \mathbf{23.64}$$

Since $\chi^2 = 23.64 > \chi^2_{.025,11} = 21.9200$, the decision is to **reject the null hypothesis**.

9.33 H$_0$: $\sigma^2 = .001$ $\alpha = .01$ $n = 16$ df = 16 – 1 = 15 $s^2 = .00144667$

H$_a$: $\sigma^2 > .001$

$\chi^2_{.01,15} = 30.5780$

$$\chi^2 = \frac{(16-1)(.00144667)}{.001} = \mathbf{21.7}$$

Since $\chi^2 = 21.7 < \chi^2_{.01,15} = 30.5780$, the decision is to **fail to reject the null hypothesis**.

9.34 H$_0$: σ^2 = 199,996,164 α = .10 α/2 = .05 n = 13 df =13 - 1 = 12

H$_a$: $\sigma^2 \neq$ 199,996,164 s^2 = 832,089,743.7

$\chi^2_{.05,12}$ = 21.0261 $\chi^2_{.95,12}$ = 5.22603

$$\chi^2 = \frac{(13-1)(832{,}089{,}743.6)}{199{,}996{,}164} = \mathbf{49.93}$$

Since χ^2 = 49.93 > $\chi^2_{.05,12}$ = 21.0261, the decision is to **reject the null hypothesis**. The variance has changed.

9.35 H$_0$: σ^2 = .04 α = .01 n = 7 df = 7 – 1 = 6 s = .34 s^2 = .1156

H$_a$: σ^2 > .04

$\chi^2_{.01,6}$ = 16.8119

$$\chi^2 = \frac{(7-1)(.1156)}{.04} = \mathbf{17.34}$$

Since χ^2 = 17.34 > $\chi^2_{.01,6}$ = 16.8119, the decision is to **reject the null hypothesis**

9.36 H$_0$: μ = 100

H$_a$: μ < 100

n = 48 μ = 99 σ = 14

a) α = .10 $z_{.10}$ = -1.28

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-1.28 = \frac{\bar{x}_c - 100}{\frac{14}{\sqrt{48}}}$$

$\bar{x}_c$ = 97.4

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{97.4 - 99}{\frac{14}{\sqrt{48}}} = -0.79$$

from Table A.5, area for z = -0.79 is .2852

β = .2852 + .5000 = **.7852**

b) α = .05 $z_{.05}$ = -1.645

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-1.645 = \frac{\bar{x}_c - 100}{\frac{14}{\sqrt{48}}}$$

$\bar{x}_c$ = 96.68

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96.68 - 99}{\frac{14}{\sqrt{48}}} = -1.15$$

from Table A.5, area for z = -1.15 is .3749

β = .3749 + .5000 = **.8749**

c) $\alpha = .01$ $z_{.01} = -2.33$

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-2.33 = \frac{\bar{x}_c - 100}{\frac{14}{\sqrt{48}}}$$

$$\bar{x}_c = 95.29$$

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{95.29 - 99}{\frac{14}{\sqrt{48}}} = -1.84$$

from Table A.5, area for $z = -1.84$ is .4671

β = .4671 + .5000 = **.9671**

d) As α gets smaller (other variables remaining constant), β gets larger. Decreasing the probability of committing a Type I error increases the probability of committing a Type II error if other variables are held constant.

9.37 $\alpha = .05$ $\mu = 100$ $n = 48$ $\sigma = 14$

a) $\mu_a = 98.5$ $z_c = -1.645$

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-1.645 = \frac{\bar{x}_c - 100}{\frac{14}{\sqrt{48}}}$$

$\bar{x}_c = 96.68$

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96.68 - 98.5}{\frac{14}{\sqrt{48}}} = -0.90$$

from Table A.5, area for $z = -0.90$ is .3159

β = .3159 + .5000 = **.8159**

b) $\mu_a = 98$ $z_c = -1.645$

$\bar{x}_c = 96.68$

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96.68 - 98}{\frac{14}{\sqrt{48}}} = -0.65$$

from Table A.5, area for $z = -0.65$ is .2422

β = .2422 + .5000 = **.7422**

c) $\mu_a = 97$ $z_{.05} = -1.645$

$\bar{x}_c = 96.68$

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96.68 - 97}{\frac{14}{\sqrt{48}}} = -0.16$$

from Table A.5, area for $z = -0.16$ is .0636

β = .0636 + .5000 = **.5636**

d) $\mu_a = 96$ $z_{.05} = -1.645$

$\bar{x}_c = 96.68$

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{96.68 - 96}{\frac{14}{\sqrt{48}}} = 0.34$$

from Table A.5, area for $z = 0.34$ is .1331

β = .5000 - .1331 = **.3669**

e) As the alternative value gets farther from the null hypothesized value, the probability of committing a Type II error reduces (all other variables being held constant).

9.38 H_o: $\mu = 50$

H_a: $\mu \neq 50$

$\mu_a = 53$ $n = 35$ $\sigma = 7$ $\alpha = .01$

Since this is two-tailed, $\alpha/2 = .005$ $z_{.005} = \pm 2.575$

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\pm 2.575 = \frac{\bar{x}_c - 50}{\frac{7}{\sqrt{35}}}$$

$$\bar{x}_c = 50 \pm 3.05$$

46.95 and 53.05

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{53.05 - 53}{\frac{7}{\sqrt{35}}} = 0.04$$

from Table A.5 for z = 0.04, area = .0160

Other end:

$$z = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{46.95 - 53}{\frac{7}{\sqrt{35}}} = -5.11$$

Area associated with z = -5.11 is .5000

β = .5000 + .0160 = **.5160**

9.39 a) H_o: $p = .65$

H_a: $p < .65$

$n = 360$ $\alpha = .05$ $p_a = .60$ $z_{.05} = -1.645$

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$-1.645 = \frac{\hat{p}_c - .65}{\sqrt{\frac{(.65)(.35)}{360}}}$$

$\hat{p}_c = .65 - .041 = .609$

$$z = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.609 - .60}{\sqrt{\frac{(.60)(.40)}{360}}} = 0.35$$

from Table A.5, area for $z = -0.35$ is .1368

$\beta = .5000 - .1368 =$ **.3632**

b) $p_a = .55$ $z_{.05} = -1.645$ $\hat{p}_c = .609$

$$z = \frac{\hat{p}_c - P}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.609 - .55}{\sqrt{\frac{(.55)(.45)}{360}}} = 2.25$$

from Table A.5, area for z = -2.25 is .4878

β = .5000 - .4878 = **.0122**

c) p_a = .50 $z_{.05}$ = -1.645 $\hat{p}_c$ = .609

$$z = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.609 - .50}{\sqrt{\frac{(.50)(.50)}{360}}} = -4.14$$

from Table A.5, the area for z = -4.14 is .5000

β = .5000 - .5000 = **.0000**

9.40 n = 58 $\bar{x}$ = 45.1 σ = 8.7 α = .05 $\alpha/2$ = .025

H_0: $\mu = 44$

H_a: $\mu \neq 44$ $z_{.025} = \pm$ 1.96

$$z = \frac{45.1 - 44}{\frac{8.7}{\sqrt{58}}} = 0.96$$

Since $z = 0.96 < z_c = 1.96$, the decision is to fail to reject the null hypothesis.

$$\pm 1.96 = \frac{\bar{x}_c - 44}{\frac{8.7}{\sqrt{58}}}$$

$$\pm 2.239 = \bar{x}_c - 44$$

$$\bar{x}_c = 46.239 \text{ and } 41.761$$

For 45 years:

$$z = \frac{46.239 - 45}{\frac{8.7}{\sqrt{58}}} = 1.08$$

from Table A.5, the area for $z = 1.08$ is .3599

β = .5000 + .3599 = **.8599**

Power = 1 - β = 1 - .8599 = .1401

For 46 years:

$$z = \frac{46.239 - 46}{\frac{8.7}{\sqrt{58}}} = 0.21$$

From Table A.5, the area for z = 0.21 is .0832

β = .5000 + .0832 = .**5832**

Power = 1 - β = 1 - .5832 = .4168

For 47 years:

$$z = \frac{46.239-47}{\frac{8.7}{\sqrt{58}}} = -0.67$$

From Table A.5, the area for z = -0.67 is .2486

β = .5000 - .2486 = .**2514**

Power = 1 - β = 1 - .2514 = .7486

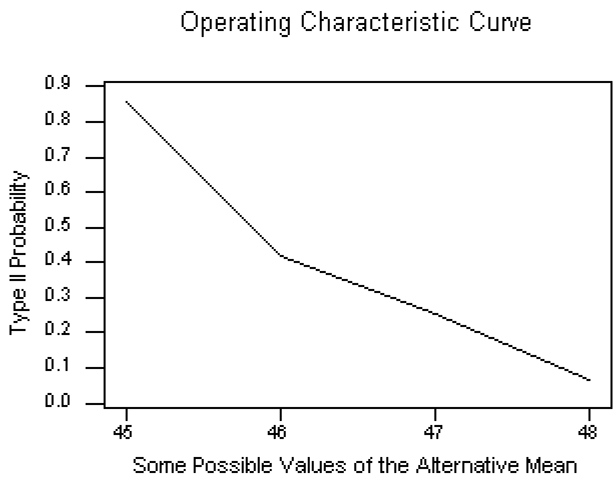
For 48 years:

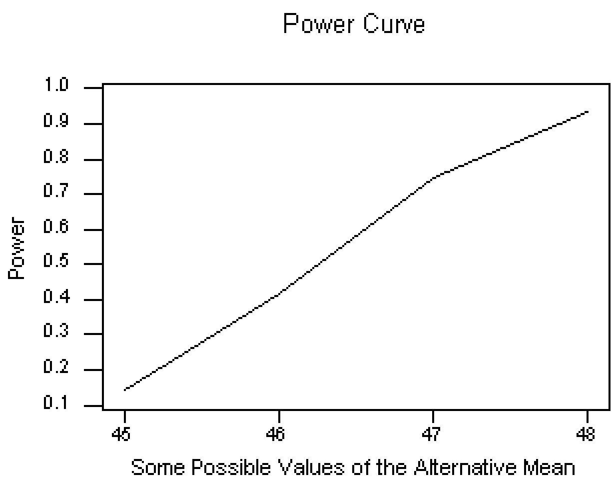
$$z = \frac{46.239-48}{\frac{8.7}{\sqrt{58}}} = 1.54$$

From Table A.5, the area for z = 1.54 is .4382

β = .5000 - .4382 = **.0618**

Power = 1 - β = 1 - .0618 = .9382





9.41 H_0: $p = .71$

H_a: $p < .71$

$n = 463$ $x = 324$ $\hat{p} = \frac{324}{463} = .6998$ $\alpha = .10$

$z_{.10}$ = -1.28

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.6998 - .71}{\sqrt{\frac{(.71)(.29)}{463}}} = \mathbf{-0.48}$$

Since the observed z = -0.48 > $z_{.10}$ = -1.28, the decision is to **fail to reject the null hypothesis**.

Type II error:

Solving for the critical proportion, $\hat{p}_c$:

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$-1.28 = \frac{\hat{p}_c - .71}{\sqrt{\frac{(.71)(.29)}{463}}}$$

$$\hat{p} = .683$$

For p_a = .69

$$z = \frac{.683 - .69}{\sqrt{\frac{(.69)(.31)}{463}}} = -0.33$$

From Table A.5, the area for z = -0.33 is .1293

The probability of committing a Type II error = .1293 + .5000 = **.6293**

For p_a = .66

$$z = \frac{.683 - .66}{\sqrt{\frac{(.66)(.34)}{463}}} = 1.04$$

From Table A.5, the area for z = 1.04 is .3508

The probability of committing a Type II error = .5000 - .3508 = **.1492**

For p_a = .60

$$z = \frac{.683 - .60}{\sqrt{\frac{(.60)(.40)}{463}}} = 3.65$$

From Table A.5, the area for z = 3.65 is essentially, .5000

The probability of committing a Type II error = .5000 - .5000 = **.0000**

9.42 HTAB steps:

1) H_o: $\mu = 36$

H_a: $\mu \neq 36$

2) $z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

3) $\alpha = .01$

4) two-tailed test, $\alpha/2 = .005$, $z_{.005} = \pm\ 2.575$

If the observed value of z is greater than 2.575 or less than -2.575, the decision will be to reject the null hypothesis.

5) $n = 63$, $\bar{x} = 38.4$, $\sigma = 5.93$

6) $z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \dfrac{38.4-36}{\frac{5.93}{\sqrt{63}}} =$ **3.21**

7) Since the observed value of $z = 3.21$ is greater than $z_{.005} = 2.575$, the decision is

to **reject the null hypothesis.**

8) The mean is likely to be greater than 36.

9.43 HTAB steps:

1) H_o: $\mu = 7.82$

H_a: $\mu < 7.82$

2) The test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

3) $\alpha = .05$

4) df = n - 1 = 16, $t_{.05,16}$ = -1.746. If the observed value of t is less than -1.746, then the decision will be to reject the null hypothesis.

5) $n = 17$ $\bar{x} = 7.01$ $s = 1.69$

6) $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \dfrac{7.01 - 7.82}{\frac{1.69}{\sqrt{17}}} =$ **-1.98**

7) Since the observed t = -1.98 is less than the table value of t = -1.746, the decision

is to **reject the null hypothesis**.

8) The population mean is significantly less than 7.82.

9.44 HTAB steps:

a. 1) H_o: $p = .28$
H_a: $p > .28$

2) $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}}$

3) $\alpha = .10$

4) This is a one-tailed test, $z_{.10} = 1.28$. If the observed value of z is greater than 1.28, the decision will be to reject the null hypothesis.

5) $n = 783 \qquad x = 230$

$\hat{p} = \dfrac{230}{783} = .2937$

6) $z = \dfrac{.2937 - .28}{\sqrt{\dfrac{(.28)(.72)}{783}}} = \mathbf{0.85}$

7) Since $z = 0.85$ is less than $z_{.10} = 1.28$, the decision is to **fail to reject the null hypothesis**.

8) There is not enough evidence to declare that $p > .28$.

b. 1) H_o: $p = .61$
H_a: $p \neq .61$

2) $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}}$

3) $\alpha = .05$

4) This is a two-tailed test, $z_{.025} = \pm$ 1.96. If the observed value of z is greater than 1.96 or less than -1.96, then the decision will be to reject the null hypothesis.

5) n = 401 $\hat{p}$ = .56

6) $z = \dfrac{.56 - .61}{\sqrt{\dfrac{(.61)(.39)}{401}}}$ = **-2.05**

7) Since z = -2.05 is less than $z_{.025}$ = -1.96, the decision is to **reject the null hypothesis.**

8) The population proportion is not likely to be .61.

9.45 HTAB steps:

1) H_0: $\sigma^2 = 15.4$

H_a: $\sigma^2 > 15.4$

2) $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$

3) $\alpha = .01$

4) $n = 18$, df = 17, one-tailed test

$\chi^2_{.01,17} = 33.4087$

5) $s^2 = 29.6$

6) $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{(17)(29.6)}{15.4} =$ **32.675**

7) Since the observed $\chi^2 = 32.675$ is less than 33.4087, the decision is to **fail to reject the null hypothesis**.

8) The population variance is not significantly more than 15.4.

9.46 a) H_0: $\mu = 130$

H_a: $\mu > 130$

$n = 75$ $\sigma = 12$ $\alpha = .01$ $z_{.01} = 2.33$ $\mu_a = 135$

Solving for $\bar{x}_c$:

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$2.33 = \frac{\bar{x}_c - 130}{\frac{12}{\sqrt{75}}}$$

$$\bar{x}_c = 133.23$$

$$z = \frac{133.23 - 135}{\frac{12}{\sqrt{75}}} = -1.28$$

from table A.5, area for $z = -1.28$ is .3997

$\beta = .5000 - .3997 =$ **.1003**

b) H_0: $p = .44$
H_a: $p < .44$

n = 1095 α = .05 p_a = .42 $z_{.05}$ = -1.645

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$-1.645 = \frac{\hat{p}_c - .44}{\sqrt{\frac{(.44)(.56)}{1095}}}$$

$$\hat{p}_c = .4153$$

$$z = \frac{.4153 - .42}{\sqrt{\frac{(.42)(.58)}{1095}}} = \mathbf{-0.32}$$

from table A.5, area for z = -0.32 is .1255

β = .5000 + .1255 = **.6255**

9.47 H_0: p = .32

H_a: p > .32

n = 80 α = .01 $\hat{p}$ = .39

$z_{.01}$ = 2.33

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.39 - .32}{\sqrt{\frac{(.32)(.68)}{80}}} = \mathbf{1.34}$$

Since the observed z = 1.34 < $z_{.01}$ = 2.33, the decision is to **fail to reject the null hypothesis**.

9.48 $\bar{x}$ = 3.45 n = 64 σ^2 = 1.31 α = .05

H_o: μ = 3.3

H_a: $\mu \neq$ 3.3

For two-tail, $\alpha/2$ = .025 z_c = ±1.96

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.45 - 3.3}{\frac{\sqrt{1.31}}{\sqrt{64}}} = \mathbf{1.05}$$

Since the observed z = 1.05 < z_c = 1.96, the decision is to **Fail to reject the null hypothesis**.

9.49 $n = 210$ $x = 93$ $\alpha = .10$ $\hat{p} = \frac{x}{n} = \frac{93}{210}$ = .443

H_o: $p = .57$

H_a: $p < .57$

For one-tail, $\alpha = .10$ $z_c = -1.28$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.443 - .57}{\sqrt{\frac{(.57)(.43)}{210}}} = \mathbf{-3.72}$$

Since the observed $z = -3.72 < z_c = -1.28$, the decision is to **reject the null hypothesis**.

9.50 H$_0$: $\sigma^2 = 16$ $n = 12$ $\alpha = .05$ df = 12 - 1 = 11

H$_a$: $\sigma^2 > 16$

s = 0.4987864 ft. = 5.98544 in.

$\chi^2_{.05,11} = 19.6752$

$$\chi^2 = \frac{(12-1)(5.98544)^2}{16} = \mathbf{24.63}$$

Since $\chi^2 = 24.63 > \chi^2_{.05,11} = 19.6752$, the decision is to **reject the null hypothesis.**

9.51 H$_0$: $\mu = 8.4$ $\alpha = .01$ $\alpha/2 = .005$ $n = 7$ df = 7 – 1 = 6 $s = 1.3$

H$_a$: $\mu \neq 8.4$

$\bar{x} = 5.6$ $t_{.005,6} = \pm 3.707$

$$t = \frac{5.6-8.4}{\frac{1.3}{\sqrt{7}}} = \mathbf{-5.70}$$

Since the observed t = - 5.70 < $t_{.005,6}$ = -3.707, the decision is to **reject the null hypothesis**.

9.52 $\bar{x}$ = $26,650 n = 100 σ = $12,000

a) H_o: μ = $25,000

H_a: μ > $25,000 α = .05

For one-tail, α = .05 $z_{.05}$ = 1.645

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{26{,}650-25{,}000}{\frac{12{,}000}{\sqrt{100}}} = \mathbf{1.38}$$

Since the observed z = 1.38 < $z_{.05}$ = 1.645, the decision is to **fail to reject the null hypothesis**.

b) μ_a = $30,000 z_c = 1.645

Solving for $\bar{x}_c$:

$$z_c = \frac{\bar{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$1.645 = \frac{(\bar{x}_c - 25{,}000)}{\frac{12{,}000}{\sqrt{100}}}$$

$\bar{x}_c$ = 25,000 + 1,974 = 26,974

$$z = \frac{26{,}974 - 30{,}000}{\frac{12{,}000}{\sqrt{100}}} = \mathbf{-2.52}$$

from Table A.5, the area for z = -2.52 is .4941

β = .5000 - .4941 = **.0059**

9.53 H$_0$: $\sigma^2 = 4$ $n = 8$ $s = 7.80$ $\alpha = .10$ df = 8 – 1 = 7

H$_a$: $\sigma^2 > 4$

$\chi^2_{.10,7} = 12.0170$

$$\chi^2 = \frac{(8-1)(7.80)^2}{4} = \mathbf{106.47}$$

Since observed $\chi^2 = 106.47 > \chi^2_{.10,7} = 12.017$, the decision is to **reject the null hypothesis**.

9.54 H$_0$: $p = .46$

H$_a$: $p > .46$

$n = 125$ $x = 66$ $\alpha = .05$ $\hat{p} = \frac{x}{n} = \frac{66}{125} = .528$

Using a one-tailed test, $z_{.05} = 1.645$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.528 - .46}{\sqrt{\frac{(.46)(.54)}{125}}} = \mathbf{1.53}$$

Since the observed value of $z = 1.53 < z_{.05} = 1.645$, the decision is to **fail to reject the null hypothesis**.

Solving for $\hat{p}_c$:

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$1.645 = \frac{\hat{p}_c - .46}{\sqrt{\frac{(.46)(.54)}{125}}}$$ and therefore, $\hat{p}_c$ = .533

$$z = \frac{\hat{p}_c - p_a}{\sqrt{\frac{p_a \cdot q_a}{n}}} = \frac{.533 - .50}{\sqrt{\frac{(.50)(.50)}{125}}} = \mathbf{0.74}$$

from Table A.5, the area for z = 0.74 is .2704

β = .5000 + .2704 = **.7704**

9.55 n = 16 $\bar{x}$ = 175 s = 14.28286 df = 16 - 1 = 15 α = .05

H_0: μ = 185

H_a: μ < 185

$t_{.05,15}$ = - 1.753

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{175-185}{\frac{14.28286}{\sqrt{16}}} = \mathbf{-2.80}$$

Since observed t = - 2.80 < $t_{.05,15}$ = - 1.753, the decision is to **reject the null hypothesis**.

9.56 H$_0$: p = .182

H$_a$: p > .182

n = 428 x = 84 α = .01 $\hat{p} = \frac{x}{n} = \frac{84}{428}$ = .1963

For a one-tailed test, $z_{.01}$ = 2.33

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}} = \frac{.1963-.182}{\sqrt{\frac{(.182)(.818)}{428}}} = \mathbf{0.77}$$

Since the observed z = 0.77 < $z_{.01}$ = 2.33, the decision is to **fail to reject the null hypothesis**.

The probability of committing a Type I error is .**01**.

Solving for $\hat{p}_c$:

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$2.33 = \frac{.\hat{p}_c - .182}{\sqrt{\frac{(.182)(.818)}{428}}}$$

$$\hat{p}_c = .2255$$

$$z = \frac{\hat{p}_c - p_a}{\sqrt{\frac{p_a \cdot q_a}{n}}} = \frac{.2255 - .21}{\sqrt{\frac{(.21)(.79)}{428}}} = \mathbf{0.79}$$

from Table A.5, the area for z = 0.79 is .2852

β = .5000 + .2852 = **.7852**

9.57 H_o: μ = \$15

H_a: μ > \$15

$\bar{x}$ = \$19.34 n = 35 σ = \$4.52 α = .10

For one-tail and α = .10 z_c = 1.28

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{19.34-15}{\frac{4.52}{\sqrt{35}}} = \mathbf{5.68}$$

Since the observed z = 5.68 > z_c = 1.28, the decision is to **reject the null hypothesis**.

9.58 H_0: σ^2 = 16 n = 22 df = 22 –1 = 21 s = 6 α = .05

H_a: σ^2 > 16

$\chi^2_{.05,21}$ = 32.6706

$$\chi^2 = \frac{(22-1)(6)^2}{16} = \mathbf{47.25}$$

Since the observed χ^2 = 47.25 > $\chi^2_{.05,21}$ = 32.6706, the decision is to **reject the null hypothesis**.

9.59 H$_0$: μ = 2.5 $\bar{x}$ = 3.4 s = 0.6 α = .01 n = 9 df = 9 – 1 = 8

H$_a$: μ > 2.5

$t_{.01,8}$ = 2.896

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3.4-2.5}{\frac{0.6}{\sqrt{9}}} = \mathbf{4.50}$$

Since the observed t = 4.50 > $t_{.01,8}$ = 2.896, the decision is to **reject the null hypothesis**.

9.60 a) H$_o$: μ = 23.58

H$_a$: $\mu \neq$ 23.58

n = 95 $\bar{x}$ = 22.83 σ = 5.11 α = .05

Since this is a two-tailed test and using $\alpha/2$ = .025: $z_{.025}$ = $\pm$ 1.96

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{22.83-23.58}{\frac{5.11}{\sqrt{95}}} = \mathbf{-1.43}$$

Since the observed z = -1.43 > $z_{.025}$ = -1.96, the decision is to **fail to reject the null hypothesis**.

b) $$z_c = \frac{\overline{x}_c - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\pm 1.96 = \frac{\overline{x}_c - 23.58}{\frac{5.11}{\sqrt{95}}}$$

$$\overline{x}_c = 23.58 \pm 1.03$$

$$\overline{x}_c = 22.55,\ 24.61$$

for H_a: $\mu = 22.30$

$$z = \frac{\overline{x}_c - \mu_a}{\frac{\sigma}{\sqrt{n}}} = \frac{22.55 - 22.30}{\frac{5.11}{\sqrt{95}}} = \mathbf{0.48}$$

$$z = \frac{\bar{x}_c - \mu_a}{\frac{\sigma}{\sqrt{n}}} = \frac{24.61 - 22.30}{\frac{5.11}{\sqrt{95}}} = \mathbf{4.41}$$

from Table A.5, the areas for $z = 0.48$ and $z = 4.41$ are .1844 and .5000

β = .5000 - .1844 = **.3156**

The upper tail has no effect on β.

9.61 $n = 12$ $\bar{x} = 12.333$ $s^2 = 10.424$

H_0: $\sigma^2 = 2.5$

H_a: $\sigma^2 \neq 2.5$

$\alpha = .05$ df = 11 two-tailed test, $\alpha/2 = .025$

$\chi^2_{.025,11} = 21.9200$

$\chi^2_{..975,11} = 3.81574$

If the observed χ^2 is greater than 21.9200 or less than 3.81574, the decision is to reject the null hypothesis.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{11(10.424)}{2.5} = \mathbf{45.866}$$

Since the observed $\chi^2 = 45.866$ is greater than $\chi^2_{.025,11} = 21.92$, the decision is to **reject the null hypothesis**. The population variance is significantly more than 2.5.

9.62 H$_0$: $\mu = 23$ $\bar{x} = 18.5$ $s = 6.91$ $\alpha = .10$ $n = 16$ df = 16 – 1 = 15

H$_a$: $\mu < 23$

$t_{.10,15} = -1.341$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{18.5-23}{\frac{6.91}{\sqrt{16}}} = \mathbf{-2.60}$$

Since the observed $t = -2.60 < t_{.10,15} = -1.341$, the decision is to **reject the null hypothesis**.

9.63 The sample size is 22. $\bar{x}$ is 3.969 $s = 0.866$ df = 21

The test statistic is:

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

The observed $t = -2.33$. The p-value is .015.

The results are statistical significant at $\alpha = .05$.

The decision is to reject the null hypothesis.

9.64 H_0: $p = .25$

H_a: $p \neq .25$

This is a two-tailed test with $\alpha = .05$. $n = 384$.

Since the p-value = .045 < α = .05, the decision is to **reject the null hypothesis**.

The sample proportion, $\hat{p}$ = .205729 which is less than the hypothesized p = .25.

One conclusion is that the population proportion is lower than .25.

9.65 H$_0$: $\mu = 2.51$

H$_a$: $\mu > 2.51$

This is a one-tailed test. The sample mean is 2.55 which is more than the hypothesized value. The observed t value is 1.51 with an associated

p-value of .072 for a one-tailed test. Because the p-value is greater than

$\alpha = .05$, the decision is to fail to reject the null hypothesis.

There is not enough evidence to conclude that beef prices are higher.

9.66 H$_0$: $\mu = 2747$

H$_a$: $\mu < 2747$

This is a one-tailed test. Sixty-seven households were included in this study.

The sample average amount spent on home-improvement projects was 2,349.

Since $z = -2.09 < z_{.05} = -1.645$, the decision is to reject the null hypothesis at

$\alpha = .05$. This is underscored by the p-value of .018 which is less than $\alpha = .05$.

However, the p-value of .018 also indicates that we would not reject the null

hypothesis at $\alpha = .01$.

Chapter 10

Statistical Inferences about Two Populations

LEARNING OBJECTIVES

The general focus of Chapter 10 is on testing hypotheses and constructing confidence intervals about parameters from two populations, thereby enabling you to

1. Test hypotheses and construct confidence intervals about the difference in two population means using the z statistic.

2. Test hypotheses and establish confidence intervals about the difference in two population means using the t statistic.

3. Test hypotheses and construct confidence intervals about the

difference in two related populations.

4. Test hypotheses and construct confidence intervals about the difference in two population proportions.

5. Test hypotheses and construct confidence intervals about two population variances.

CHAPTER TEACHING STRATEGY

The major emphasis of chapter 10 is on analyzing data from two samples. The student should be ready to deal with this topic given that he/she has tested hypotheses and computed confidence intervals in previous chapters on single sample data.

In this chapter, the approach as to whether to use a *z* statistic or a *t* statistic for analyzing the differences in two sample means is the same as that used in chapters 8

and 9. When the population variances are known, the *z* statistic can be used. However, if the population variances are unknown and sample variances are being used, then the

t test is the appropriate statistic for the analysis. It is always an assumption underlying the use of the *t* statistic that the populations are normally distributed. If sample sizes are small and the population variances are known, the *z* statistic can be used if the populations are normally distributed.

In conducting a t test for the difference of two means from independent populations, there are two different formulas given in the chapter. One version of this test uses a "pooled" estimate of the population variance and assumes that the population variances are equal. The other version does not assume equal population variances and is simpler to compute. In doing hand calculations, it is generally easier to use the “pooled” variance formula because the degrees of freedom formula for the unequal variance formula is quite complex. However, it is good to expose students to both formulas since computer software packages often give you the option of using the “pooled” that assumes equal population variances or the formula for unequal variances.

A *t* test is also included for related (non independent) samples. It is important that the student be able to recognize when two samples are related and when they are independent. The first portion of section 10.3 addresses this issue. To underscore the potential difference in the outcome of the two techniques, it is sometimes valuable to analyze some related measures data with both techniques and demonstrate that the results and conclusions are usually quite different. You can have your students work problems like this using both techniques to help them understand the differences between the two tests (independent and dependent *t* tests) and the different outcomes they will obtain.

A *z* test of proportions for two samples is presented here along with an *F* test for two population variances. This is a good place to introduce the student to the *F* distribution in preparation for analysis of variance in Chapter 11. The student will begin to understand that the *F* values have two different degrees of freedom. The *F* distribution tables are upper tailed only. For this reason, formula 10.14 is given in the chapter to be used to compute lower tailed *F* values for two-tailed tests.

CHAPTER OUTLINE

10.1 Hypothesis Testing and Confidence Intervals about the Difference in Two Means using the *z* Statistic (Population Variances Known)

Hypothesis Testing

Confidence Intervals

Using the Computer to Test Hypotheses about the Difference in Two

Population Means Using the *z* Test

10.2 Hypothesis Testing and Confidence Intervals about the Difference in Two Means: Independent Samples and Population Variances Unknown

Hypothesis Testing

Using the Computer to Test Hypotheses and Construct Confidence

Intervals about the Difference in Two Population Means Using the *t*

Test

Confidence Intervals

10.3 Statistical Inferences For Two Related Populations

Hypothesis Testing

Using the Computer to Make Statistical Inferences about Two Related

Populations

Confidence Intervals

10.4 Statistical Inferences About Two Population Proportions, $p_1 - p_2$

Hypothesis Testing

Confidence Intervals

Using the Computer to Analyze the Difference in Two Proportions

10.5 Testing Hypotheses About Two Population Variances

Using the Computer to Test Hypotheses about Two Population Variances

KEY TERMS

Dependent Samples	Independent Samples
F Distribution	Matched-Pairs Test
F Value	Related Measures

SOLUTIONS TO PROBLEMS IN CHAPTER 10

10.1 | | Sample 1 | Sample 2
|---|---|---|
| | $\bar{x}_1 = 51.3$ | $\bar{x}_2 = 53.2$ |
| | $s_1^2 = 52$ | $s_2^2 = 60$ |
| | $n_1 = 31$ | $n_2 = 32$ |

a) H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 < 0$

For one-tail test, $\alpha = .10$ $z_{.10} = -1.28$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(51.3 - 53.2) - (0)}{\sqrt{\frac{52}{31} + \frac{60}{32}}} = \mathbf{-1.01}$$

Since the observed $z = -1.01 > z_c = -1.28$, the decision is to **fail to reject the null hypothesis**.

b) Critical value method:

$$z_c = \frac{(\bar{x}_1 - \bar{x}_2)_c - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$-1.28 = \frac{(\bar{x}_1 - \bar{x}_2)_c - (0)}{\sqrt{\frac{52}{31} + \frac{60}{32}}}$$

$(\bar{x}_1 - \bar{x}_2)_c$ = **-2.41**

c) The area for z = -1.01 using Table A.5 is .3438.

The p-value is .5000 - .3438 = **.1562**

10.2 Sample 1 Sample 2

$n_1 = 32$ $n_2 = 31$

$\bar{x}_1 = 70.4$ $\bar{x}_2 = 68.7$

$\sigma_1 = 5.76$ $\sigma_2 = 6.1$

For a 90% C.I., $z_{.05} = 1.645$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

$$(70.4) – 68.7) \pm 1.645\sqrt{\frac{5.76^2}{32} + \frac{6.1^2}{31}}$$

1.7 ± 2.46

$-.76 \le \mu_1 - \mu_2 \le 4.16$

10.3 a) Sample 1 Sample 2

$\bar{x}_1 = 88.23$ $\bar{x}_2 = 81.2$

$\sigma_1^2 = 22.74$ $\sigma_2^2 = 26.65$

$n_1 = 30$ $n_2 = 30$

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

For two-tail test, use $\alpha/2 = .01$ $z_{.01} = \pm 2.33$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(88.23 - 81.2) - (0)}{\sqrt{\frac{22.74}{30} + \frac{26.65}{30}}} = \mathbf{5.48}$$

Since the observed $z = 5.48 > z_{.01} = 2.33$, the decision is to **reject the null hypothesis**.

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

b)

$$(88.23 - 81.2) \pm 2.33\sqrt{\frac{22.74}{30} + \frac{26.65}{30}}$$

7.03 ± 2.99

$\mathbf{4.04 \le \mu \le 10.02}$

This supports the decision made in a) to reject the null hypothesis because zero is not in the interval.

10.4 <u>Computers/electronics</u> <u>Food/Beverage</u>

$\bar{x}_1 = 1.96$	$\bar{x}_2 = 3.02$
$\sigma_1^2 = 1.0188$	$\sigma_2^2 = .9180$
$n_1 = 50$	$n_2 = 50$

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

For two-tail test, $\alpha/2 = .005$ $z_{.005} = \pm 2.575$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}} = \frac{(1.96 - 3.02) - (0)}{\sqrt{\frac{1.0188}{50} + \frac{0.9180}{50}}} = \mathbf{-5.39}$$

Since the observed $z = -5.39 < z_c = -2.575$, the decision is to **reject the null hypothesis**.

10.5 A B

$n_1 = 40$ $n_2 = 37$

$\bar{x}_1 = 5.3$ $\bar{x}_2 = 6.5$

$\sigma_1^2 = 1.99$ $\sigma_2^2 = 2.36$

For a 95% C.I., $z_{.025} = 1.96$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

$$(5.3 - 6.5) \pm 1.96\sqrt{\frac{1.99}{40} + \frac{2.36}{37}}$$

$-1.2 \pm .66$ **$-1.86 \le \mu \le -.54$**

The results indicate that we are 95% confident that, on average, Plumber B does between 0.54 and 1.86 more jobs per day than Plumber A. Since zero does not lie in this interval, we are confident that there is a difference between Plumber A and Plumber B.

10.6 Managers Specialty

$n_1 = 35$ $n_2 = 41$

$\bar{x}_1 = 1.84$ $\bar{x}_2 = 1.99$

$\sigma_1 = .38$ $\sigma_2 = .51$

for a 98% C.I., $z_{.01} = 2.33$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

$$(1.84 - 1.99) \pm 2.33\sqrt{\frac{.38^2}{35} + \frac{.51^2}{41}}$$

$-.15 \pm .2384$

$-.3884 \le \mu_1 - \mu_2 \le .0884$

Point Estimate = -.15

Hypothesis Test:

1) H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

2) $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^{\ 2}}{n_1} + \dfrac{\sigma_2^{\ 2}}{n_2}}}$

3) α = .02

4) For a two-tailed test, $z_{.01}$ = $\pm$ 2.33. If the observed z value is greater than 2.33

or less than -2.33, then the decision will be to reject the null hypothesis.

5) Data given above

6) $z = \dfrac{(1.84 - 1.99) - (0)}{\sqrt{\dfrac{(.38)^2}{35} + \dfrac{(.51)^2}{41}}}$ = **-1.47**

7) Since z = -1.47 > $z_{.01}$ = -2.33, the decision is to **fail to reject the null**

hypothesis.

8) There is no significant difference in the hourly rates of the two groups.

10.7 <u>1996</u> <u>2006</u>

$\bar{x}_1 = 190$ $\bar{x}_2 = 198$

$\sigma_1 = 18.50$ $\sigma_2 = 15.60$

$n_1 = 51$ $n_2 = 47$ $\alpha = .01$

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 < 0$

For a one-tailed test, $z_{.01} = -2.33$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(190 - 198) - (0)}{\sqrt{\frac{(18.50)^2}{51} + \frac{(15.60)^2}{47}}} = \mathbf{-2.32}$$

Since the observed $z = -2.32 > z_{.01} = -2.33$, the decision is to **fail to reject the null hypothesis**.

10.8 Seattle Atlanta

$n_1 = 31$ $n_2 = 31$

$\bar{x}_1 = 2.64$ $\bar{x}_2 = 2.36$

$\sigma_1^2 = .03$ $\sigma_2^2 = .015$

For a 99% C.I., $z_{.005} = 2.575$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}$$

$$(2.64\text{-}2.36) \pm 2.575\sqrt{\frac{.03}{31} + \frac{.015}{31}}$$

.28 ± .10 **.18 $\leq$ μ $\leq$ .38**

Between \$ .18 and \$ .38 difference with Seattle being more expensive.

10.9 <u>Canon</u> <u>Pioneer</u>

$\bar{x}_1 = 5.8$ $\bar{x}_2 = 5.0$

$\sigma_1 = 1.7$ $\sigma_2 = 1.4$

$n_1 = 36$ $n_2 = 45$

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

For two-tail test, $\alpha/2 = .025$ $z_{.025} = \pm 1.96$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} = \frac{(5.8 - 5.0) - (0)}{\sqrt{\frac{(1.7)^2}{36} + \frac{(1.4)}{45}}} = \mathbf{2.27}$$

Since the observed $z = 2.27 > z_c = 1.96$, the decision is to **reject the null hypothesis**.

10.10 A B

$\bar{x}_1 = 8.05$ $\bar{x}_2 = 7.26$

$\sigma_1 = 1.36$ $\sigma_2 = 1.06$

$n_1 = 50$ $n_2 = 38$

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

For one-tail test, $\alpha = .10$ $z_{.10} = 1.28$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(8.05 - 7.26) - (0)}{\sqrt{\frac{(1.36)^2}{50} + \frac{(1.06)^2}{38}}} = \mathbf{3.06}$$

Since the observed $z = 3.06 > z_c = 1.28$, the decision is to **reject the null hypothesis**.

10.11 H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .01$

H_a: $\mu_1 - \mu_2 < 0$ df = 8 + 11 - 2 = 17

Sample 1 Sample 2

$n_1 = 8$ $n_2 = 11$

$\bar{x}_1 = 24.56$ $\bar{x}_2 = 26.42$

$s_1^2 = 12.4$ $s_2^2 = 15.8$

For one-tail test, $\alpha = .01$ Critical $t_{.01,17} = -2.567$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(24.56 - 26.42) - (0)}{\sqrt{\frac{12.4(7) + 15.8(10)}{8 + 11 - 2}}\sqrt{\frac{1}{8} + \frac{1}{11}}} = \mathbf{-1.05}$$

Since the observed t = -1.05 > $t_{.01,19}$ = -2.567, the decision is to **fail to reject the null hypothesis**.

10.12 a) H$_o$: $\mu_1 - \mu_2 = 0$ $\alpha = .10$

H$_a$: $\mu_1 - \mu_2 \neq 0$ df = 20 + 20 - 2 = 38

<u>Sample 1</u>	<u>Sample 2</u>
$n_1 = 20$	$n_2 = 20$
$\bar{x}_1 = 118$	$\bar{x}_2 = 113$
$s_1 = 23.9$	$s_2 = 21.6$

For two-tail test, $\alpha/2 = .05$ Critical $t_{.05,38} = 1.697$ (used df=30)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} =$$

$$t = \frac{(118-113)-(0)}{\sqrt{\frac{(23.9)^2(19)+(21.6)^2(19)}{20+20-2}}\sqrt{\frac{1}{20}+\frac{1}{20}}} = \mathbf{0.69}$$

Since the observed $t = 0.69 < t_{.05,38} = 1.697$, the decision is to **fail to reject the null hypothesis**.

b) $$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1{}^2(n_1-1)+s_2{}^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} =$$

$$(118 - 113) \pm 1.697\sqrt{\frac{(23.9)^2(19)+(21.6)^2(19)}{20+20-2}}\sqrt{\frac{1}{20}+\frac{1}{20}}$$

5 ± 12.224

$-7.224 \leq \mu_1 - \mu_2 \leq 17.224$

10.13 H$_o$: μ_1 - μ_2 = 0 α = .05

H$_a$: μ_1 - μ_2 > 0 df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18

Sample 1	Sample 2
n_1 = 10	n_2 = 10
$\bar{x}_1$ = 45.38	$\bar{x}_2$ = 40.49
s_1 = 2.357	s_2 = 2.355

For one-tail test, α = .05 Critical $t_{.05,18}$ = 1.734

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$t = \frac{(45.38 - 40.49) - (0)}{\sqrt{\frac{(2.357)^2(9) + (2.355)^2(9)}{10 + 10 - 2}}\sqrt{\frac{1}{10} + \frac{1}{10}}} = \mathbf{4.64}$$

Since the observed t = 4.64 > $t_{.05,18}$ = 1.734, the decision is to **reject the null hypothesis**.

10.14 H$_o$: μ_1 - μ_2 = 0 α =.01

H_a: $\mu_1 - \mu_2 \neq 0$ df = 18 + 18 - 2 = 34

Sample 1	Sample 2
$n_1 = 18$	$n_2 = 18$
$\bar{x}_1 = 5.333$	$\bar{x}_2 = 9.444$
$s_1^2 = 12$	$s_2^2 = 2.026$

For two-tail test, $\alpha/2 = .005$ Critical $t_{.005,34} = \pm 2.75$ (used df=30)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$t = \frac{(5.333 - 9.444) - (0)}{\sqrt{\frac{12(17) + (2.026)17}{18 + 18 - 2}}\sqrt{\frac{1}{18} + \frac{1}{18}}} = \mathbf{-4.66}$$

Since the observed $t = -4.66 < t_{.005,34} = -2.75$, **reject the null hypothesis.**

b) For 98% confidence, $t_{.01,\,30} = 2.457$

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^{\,2}(n_1-1)+s_2^{\,2}(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} =$$

$$(5.333 - 9.444) \pm 2.457\sqrt{\frac{(12)(17)+(2.026)(17)}{18+18-2}}\sqrt{\frac{1}{18}+\frac{1}{18}}$$

-4.111 ± 2.1689

$\mathbf{-6.2799 \le \mu_1 - \mu_2 \le -1.9421}$

10.15 Peoria Evansville

Peoria	Evansville	
$n_1 = 21$	$n_2 = 26$	
$\bar{x}_1 = 116{,}900$	$\bar{x}_2 = 114{,}000$	
$s1 = 2{,}300$	$s2 = 1{,}750$	df = 21 + 26 – 2

90% level of confidence, $\alpha/2 = .05$ $t_{.05,45} = 1.684$ (used df = 40)

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^{\,2}(n_1-1)+s_2^{\,2}(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} =$$

$$= (116{,}900 - 114{,}000) \pm 1.684 \sqrt{\frac{(2300)^2(20)+(1750)^2(25)}{21+26-2}}\sqrt{\frac{1}{21}+\frac{1}{26}}$$

$2{,}900 \pm 994.62$

$1905.38 \leq \mu_1 - \mu_2 \leq 3894.62$

10.16 H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .10$

H_a: $\mu_1 - \mu_2 \neq 0$ df = 12 + 12 - 2 = 22

Co-op	Interns
$n_1 = 12$	$n_2 = 12$
$\bar{x}_1 = \$15.645$	$\bar{x}_2 = \$15.439$
$s_1 = \$1.093$	$s_2 = \$0.958$

For two-tail test, $\alpha/2 = .05$

Critical $t_{.05,22} = \pm\ 1.717$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$t = \frac{(15.645 - 15.439) - (0)}{\sqrt{\frac{(1.093)^2(11) + (0.958)^2(11)}{12 + 12 - 2}}\sqrt{\frac{1}{12} + \frac{1}{12}}} = \mathbf{0.49}$$

Since the observed $t = 0.49 < t_{.05,22} = 1.717$, the decision is to **fail reject the null hypothesis**.

<u>90% Confidence Interval</u>: $t_{.05,22} = \pm 1.717$

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1{}^2(n_1-1)+s_2{}^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} =$$

$$(15.645 - 15.439) \pm 1.717\sqrt{\frac{(1.093)^2(11)+(0.958)^2(11)}{12+12-2}}\sqrt{\frac{1}{12}+\frac{1}{12}} =$$

0.206 ± 0.7204

$$\mathbf{-0.5144 \leq \mu_1 - \mu_2 \leq 0.9264}$$

10.17 Let Boston be group 1

1) H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

2) $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$

3) $\alpha = .01$

4) For a one-tailed test and df = 8 + 9 - 2 = 15, $t_{.01,15} = 2.602$. If the observed

value of t is greater than 2.602, the decision is to reject the null hypothesis.

5) <u>Boston</u> <u>Dallas</u>

$n_1 = 8$ $n_2 = 9$

$\bar{x}_1 = 47$ $\bar{x}_2 = 44$

$s_1 = 3$ $s_2 = 3$

6) $\boldsymbol{t} = \dfrac{(47-44)-(0)}{\sqrt{\dfrac{7(3)^2+8(3)^2}{15}}\sqrt{\dfrac{1}{8}+\dfrac{1}{9}}} = \mathbf{2.06}$

7) Since $t = 2.06 < t_{.01,15} = 2.602$, the decision is to fail to reject the null

hypothesis.

8) There is no significant difference in rental rates between Boston and Dallas.

10.18 $n_m = 22$ $n_{no} = 20$

$\bar{x}_m = 112$ $\bar{x}_{no} = 122$

$s_m = 11$ $s_{no} = 12$

df = $n_m + n_{no} - 2 = 22 + 20 - 2 = 40$

For a 98% Confidence Interval, $\alpha/2 = .01$ and $t_{.01,40} = 2.423$

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} =$$

$$(112 - 122) \pm 2.423\sqrt{\frac{(11)^2(21)+(12)^2(19)}{22+20-2}}\sqrt{\frac{1}{22}+\frac{1}{20}}$$

-10 ± 8.60

$-\$18.60 \le \mu_1 - \mu_2 \le -\$1.40$

Point Estimate = -$10

10.19 H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

df = $n_1 + n_2 - 2 = 11 + 11 - 2 = 20$

<u>Toronto</u>	<u>Mexico City</u>
$n_1 = 11$	$n_2 = 11$
$\bar{x}_1$ = \$67,381.82	$\bar{x}_2$ = \$63,481.82
s_1 = \$2,067.28	s_2 = \$1,594.25

For a two-tail test, $\alpha/2 = .005$ Critical $t_{.005,20} = \pm 2.845$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$t = \frac{(67{,}381.82 - 63{,}481.82) - (0)}{\sqrt{\frac{(2{,}067.28)^2(10) + (1{,}594.25)^2(10)}{11 + 11 - 2}}\sqrt{\frac{1}{11} + \frac{1}{11}}} = \mathbf{4.95}$$

Since the observed t = 4.95 > $t_{.005,20}$ = 2.845, the decision is to **Reject the null hypothesis**.

10.20 H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

df = $n_1 + n_2 - 2 = 9 + 10 - 2 = 17$

<u>Men</u>	<u>Women</u>
$n_1 = 9$	$n_2 = 10$
$\bar{x}_1$ = \$110.92	$\bar{x}_2$ = \$75.48
s_1 = \$28.79	s_2 = \$30.51

This is a one-tail test, α = .01   Critical $t_{.01,17}$ = 2.567

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$t = \frac{(110.92 - 75.48) - (0)}{\sqrt{\frac{(28.79)^2(8) + (30.51)^2(9)}{9 + 10 - 2}}\sqrt{\frac{1}{9} + \frac{1}{10}}} = \mathbf{2.60}$$

Since the observed t = 2.60 > $t_{.01,17}$ = 2.567, the decision is to **Reject the null**

hypothesis.

10.21 H_o: D = 0

H_a: D > 0

Sample 1	Sample 2	*d*
38	22	16
27	28	-1
30	21	9
41	38	3
36	38	-2
38	26	12
33	19	14
35	31	4
44	35	9

n = 9 $\bar{d}$ =7.11 s_d=6.45 α = .01

df = n - 1 = 9 - 1 = 8

For one-tail test and α = .01, the critical $t_{.01,8}$ = ±2.896

$$t = \frac{\overline{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{7.11 - 0}{\frac{6.45}{\sqrt{9}}} = \mathbf{3.31}$$

Since the observed t = 3.31 > $t_{.01,8}$ = 2.896, the decision is to **reject the null hypothesis**.

10.22 H_o: $D = 0$

H_a: $D \neq 0$

Before	After	d
107	102	5
99	98	1
110	100	10
113	108	5
96	89	7
98	101	-3
100	99	1
102	102	0
107	105	2
109	110	-1
104	102	2
99	96	3
101	100	1

$n = 13$ $\bar{d} = 2.5385$ $s_d = 3.4789$ $\alpha = .05$

df = $n - 1 = 13 - 1 = 12$

For a two-tail test and $\alpha/2 = .025$ Critical $t_{.025,12} = \pm 2.179$

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{2.5385 - 0}{\frac{3.4789}{\sqrt{13}}} = \mathbf{2.63}$$

Since the observed $t = 2.63 > t_{.025,12} = 2.179$, the decision is to **reject the null hypothesis**.

10.23 $n = 22$ $\bar{d} = 40.56$ $s_d = 26.58$

For a 98% Level of Confidence, $\alpha/2 = .01$, and df = $n - 1 = 22 - 1 = 21$

$t_{.01,21} = 2.518$

$$\bar{d} \pm t\frac{s_d}{\sqrt{n}}$$

$$40.56 \pm (2.518)\frac{26.58}{\sqrt{22}}$$

40.56 ± 14.27

$26.29 \le D \le 54.83$

10.24

Before	After	d
32	40	-8
28	25	3
35	36	-1
32	32	0
26	29	-3

25	31	-6
37	39	-2
16	30	-14
35	31	4

$n = 9$ $\qquad \overline{d} = -3$ $\qquad s_d = 5.6347$ $\qquad \alpha = .025$

df = $n - 1 = 9 - 1 = 8$

For 90% level of confidence and $\alpha/2 = .05$, $\qquad t_{.05,8} = 1.86$

$$t = \overline{d} \pm t\frac{s_d}{\sqrt{n}}$$

$$t = -3 \pm (1.86)\frac{5.6347}{\sqrt{9}} = -3 \pm 3.49$$

$-6.49 \le D \le 0.49$

10.25

City	Cost	Resale	d
Atlanta		20427 25163	-4736
Boston		27255 24625	2630
Des Moines	22115	12600	9515

Kansas City	23256	24588	-1332
Louisville	21887	19267	2620
Portland	24255	20150	4105
Raleigh-Durham	19852	22500	-2648
Reno	23624	16667	6957
Ridgewood	25885	26875	- 990
San Francisco	28999	35333	-6334
Tulsa	20836	16292	4544

$\bar{d}$ = 1302.82 $\quad s_d$ = 4938.22 $\quad n$ = 11, $\quad$ df = 10

α = .01 $\quad \alpha/2$ = .005 $\quad t_{.005,10}$= 3.169

$\bar{d} \pm t\frac{s_d}{\sqrt{n}}$ = 1302.82 $\pm$ 3.169 $\frac{4938.22}{\sqrt{11}}$ = 1302.82 $\pm$ 4718.42

-3415.6 $\le$ *D* $\le$ 6021.2

10.26 H_o: $D = 0$

H_a: $D < 0$

Before	After	d
2	4	-2
4	5	-1
1	3	-2
3	3	0
4	3	1
2	5	-3
2	6	-4
3	4	-1
1	5	-4

$n = 9$ $\quad \bar{d} = -1.778$ $\quad s_d = 1.716$ $\quad \alpha = .05$ $\quad$ df = $n - 1 = 9 - 1 = 8$

For a one-tail test and $\alpha = .05$, the critical $t_{.05,8} = -1.86$

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{-1.778 - 0}{\frac{1.716}{\sqrt{9}}} = \mathbf{-3.11}$$

Since the observed $t = -3.11 < t_{.05,8} = -1.86$, the decision is to **reject the null hypothesis**.

10.27

Before	After	d
255	197	58
230	225	5
290	215	75
242	215	27
300	240	60
250	235	15
215	190	25
230	240	-10
225	200	25
219	203	16
236	223	13

$n = 11$ $\bar{d} = 28.09$ $s_d = 25.813$ df = $n - 1 = 11 - 1 = 10$

For a 98% level of confidence and $\alpha/2 = .01$, $t_{.01,10} = 2.764$

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

$$28.09 \pm (2.764) \frac{25.813}{\sqrt{11}} = 28.09 \pm 21.51$$

$$\mathbf{6.58 \leq \mathit{D} \leq 49.60}$$

10.28 H_0: $D = 0$

H_a: $D > 0$ $n = 27$ df = 27 – 1 = 26 $\overline{d} = 3.17$ $s_d = 5$

Since $\alpha = .01$, the critical $t_{.01,26} = 2.479$

$$t = \frac{\overline{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{3.71 - 0}{\frac{5}{\sqrt{27}}} = \mathbf{3.86}$$

Since the observed $t = 3.86 > t_{.01,26} = 2.479$, the decision is to **reject the null hypothesis**.

10.29 $n = 21$ $\overline{d} = 75$ $s_d = 30$ df = 21 - 1 = 20

For a 90% confidence level, $\alpha/2 = .05$ and $t_{.05,20} = 1.725$

$$\overline{d} \pm t\frac{s_d}{\sqrt{n}}$$

$$75 \pm 1.725\frac{30}{\sqrt{21}} = 75 \pm 11.29$$

63.71 $\le$ *D* $\le$ 86.29

10.30 H_o: $D = 0$

H_a: $D \neq 0$

$n = 15$ $\bar{d} = -2.85$ $s_d = 1.9$ $\alpha = .01$ df = 15 - 1 = 14

For a two-tail test, $\alpha/2 = .005$ and the critical $t_{.005,14} = \pm 2.977$

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{-2.85 - 0}{\frac{1.9}{\sqrt{15}}} = \mathbf{-5.81}$$

Since the observed $t = -5.81 < t_{.005,14} = -2.977$, the decision is to **reject the null hypothesis**.

10.31 a) <u>Sample 1</u> <u>Sample 2</u>

$n_1 = 368$ $n_2 = 405$

$x_1 = 175$ $x_2 = 182$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{175}{368}$$
= .476

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{182}{405}$$
= .449

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{175 + 182}{368 + 405} = \frac{357}{773}$$
= .462

H_o: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 \neq 0$

For two-tail, $\alpha/2 = .025$ and $z_{.025} = \pm 1.96$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p} \cdot \overline{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.476 - .449) - (0)}{\sqrt{(.462)(.538)\left(\frac{1}{368} + \frac{1}{405}\right)}}$$
= **0.75**

Since the observed $z = 0.75 < z_c = 1.96$, the decision is to **fail to reject the null**

hypothesis.

b) Sample 1 / Sample 2

Sample 1	Sample 2
$\hat{p}_1 = .38$	$\hat{p}_2 = .25$
$n_1 = 649$	$n_2 = 558$

$$\overline{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{649(.38) + 558(.25)}{649 + 558}$$

= .32

H_o: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 > 0$

For a one-tail test and $\alpha = .10$, $z_{.10} = 1.28$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p} \cdot \overline{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.38 - .25) - (0)}{\sqrt{(.32)(.68)\left(\frac{1}{649} + \frac{1}{558}\right)}}$$

= **4.83**

Since the observed $z = 4.83 > z_c = 1.28$, the decision is to **reject the null hypothesis**.

10.32 a) $n_1 = 85$ $n_2 = 90$ $\hat{p}_1 = .75$ $\hat{p}_2 = .67$

For a 90% Confidence Level, $z_{.05} = 1.645$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.75 - .67) \pm 1.645\sqrt{\frac{(.75)(.25)}{85}+\frac{(.67)(.33)}{90}} = .08 \pm .11$$

$-.03 \le p_1 - p_2 \le .19$

b) $n_1 = 1100$ $n_2 = 1300$ $\hat{p}_1 = .19$ $\hat{p}_2 = .17$

For a 95% Confidence Level, $\alpha/2 = .025$ and $z_{.025} = 1.96$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.19 - .17) \pm 1.96\sqrt{\frac{(.19)(.81)}{1100}+\frac{(.17)(.83)}{1300}} = .02 \pm .03$$

$-.01 \le p_1 - p_2 \le .05$

c) $n_1 = 430$ $n_2 = 399$ $x_1 = 275$ $x_2 = 275$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{275}{430} = .64 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{275}{399} = .69$$

For an 85% Confidence Level, $\alpha/2$ = .075 and $z_{.075}$ = 1.44

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.64 - .69) \pm 1.44\sqrt{\frac{(.64)(.36)}{430} + \frac{(.69)(.31)}{399}} = -.05 \pm .047$$

$-.097 \le p_1 - p_2 \le -.003$

d) $n_1 = 1500$ $n_2 = 1500$ $x_1 = 1050$ $x_2 = 1100$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{1050}{1500} = .70 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{1100}{1500} = .733$$

For an 80% Confidence Level, $\alpha/2 = .10$ and $z_{.10} = 1.28$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.70 - .733) \pm 1.28\sqrt{\frac{(.70)(.30)}{1500} + \frac{(.733)(.267)}{1500}} = -.033 \pm .02$$

$$\mathbf{-.053 \leq p_1 - p_2 \leq -.013}$$

10.33 H_0: $p_m - p_w = 0$

H_a: $p_m - p_w < 0$ $n_m = 374$ $n_w = 481$ $\hat{p}_m = .59$ $\hat{p}_w = .70$

For a one-tailed test and $\alpha = .05$, $z_{.05} = -1.645$

$$\bar{p} = \frac{n_m \hat{p}_m + n_w \hat{p}_w}{n_m + n_w} = \frac{374(.59) + 481(.70)}{374 + 481}$$

= .652

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p} \cdot \bar{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.59 - .70) - (0)}{\sqrt{(.652)(.348)\left(\frac{1}{374} + \frac{1}{481}\right)}}$$

= **-3.35**

Since the observed z = -3.35 < $z_{.05}$ = -1.645, the decision is to **reject the null**

hypothesis.

10.34 $n_1 = 210$ $n_2 = 176$ $\hat{p}_1 = .24$ $\hat{p}_2 = .35$

For a 90% Confidence Level, $\alpha/2 = .05$ and $z_{.05} = \pm 1.645$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.24 - .35) \pm 1.645\sqrt{\frac{(.24)(.76)}{210} + \frac{(.35)(.65)}{176}} = -.11 \pm .0765$$

$$\mathbf{-.1865 \le p_1 - p_2 \le -.0335}$$

10.35 <u>Computer Firms</u> <u>Banks</u>

$\hat{p}_1 = .48$ $\hat{p}_2 = .56$

$n_1 = 56$ $n_2 = 89$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{56(.48) + 89(.56)}{56 + 89} = .529$$

H_o: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 \neq 0$

For two-tail test, $\alpha/2 = .10$ and $z_c = \pm 1.28$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p} \cdot \bar{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.48 - .56) - (0)}{\sqrt{(.529)(.471)\left(\frac{1}{56} + \frac{1}{89}\right)}}$$

$=$ **-0.94**

Since the observed $z = -0.94 > z_c = -1.28$, the decision is to **fail to reject the null hypothesis**.

10.36 A B

$n_1 = 35$ $n_2 = 35$

$x_1 = 5$ $x_2 = 7$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{5}{35} = .14 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{7}{35} = .20$$

For a 98% Confidence Level, $\alpha/2 = .01$ and $z_{.01} = 2.33$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.14 - .20) \pm 2.33\sqrt{\frac{(.14)(.86)}{35} + \frac{(.20)(.80)}{35}} = -.06 \pm .21$$

$$\mathbf{-.27 \leq p_1 - p_2 \leq .15}$$

10.37 H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 \neq 0$

$\alpha = .10$ $\hat{p}_1 = .09$ $\hat{p}_2 = .06$ $n_1 = 780$ $n_2 = 915$

For a two-tailed test, $\alpha/2 = .05$ and $z_{.05} = \pm 1.645$

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{780(.09) + 915(.06)}{780 + 915}$$

$= .0738$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\cdot\bar{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.09 - .06) - (0)}{\sqrt{(.0738)(.9262)\left(\frac{1}{780} + \frac{1}{915}\right)}}$$

$=$ **2.35**

Since the observed $z = 2.35 > z_{.05} = 1.645$, the decision is to **reject the null hypothesis**.

10.38 $n_1 = 850$ $n_2 = 910$ $\hat{p}_1 = .60$ $\hat{p}_2 = .52$

For a 95% Confidence Level, $\alpha/2 = .025$ and $z_{.025} = \pm 1.96$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.60 - .52) \pm 1.96\sqrt{\frac{(.60)(.40)}{850} + \frac{(.52)(.48)}{910}} = .08 \pm .046$$

$\mathbf{.034 \leq p_1 - p_2 \leq .126}$

10.39 H_0: $\sigma_1^2 = \sigma_2^2$ $\alpha = .01$ $n_1 = 10$ $s_1^2 = 562$

H_a: $\sigma_1^2 < \sigma_2^2$ $n_2 = 12$ $s_2^2 = 1013$

$df_{num} = 12 - 1 = 11$ $df_{denom} = 10 - 1 = 9$

Table $F_{.01,10,9} = 5.26$

$F = \dfrac{s_2^{\,2}}{s_1^{\,2}} = \dfrac{1013}{562} =$ **1.80**

Since the observed $F = 1.80 < F_{.01,10,9} = 5.26$, the decision is to **fail to reject the null hypothesis**.

10.40 H$_0$: $\sigma_1^2 = \sigma_2^2$ $\alpha = .05$ $n_1 = 5$ $s_1 = 4.68$

H$_a$: $\sigma_1^2 \neq \sigma_2^2$ $n_2 = 19$ $s_2 = 2.78$

$df_{num} = 5 - 1 = 4$ $df_{denom} = 19 - 1 = 18$

The critical table *F* values are: $F_{.025,4,18} = 3.61$ $F_{.95,18,4} = .277$

$F = \dfrac{s_1^{\,2}}{s_2^{\,2}} = \dfrac{(4.68)^2}{(2.78)^2} =$ **2.83**

Since the observed $F = 2.83 < F_{.025,4,18} = 3.61$, the decision is to **fail to reject the null hypothesis**.

10.41 <u>City 1</u> <u>City 2</u>

3.43	3.33
3.40	3.42
3.39	3.39
3.32	3.30
3.39	3.46
3.38	3.39
3.34	3.36
3.38	3.44
3.38	3.37
3.28	3.38

$n_1 = 10$ $\quad df_1 = 9$ $\quad n_2 = 10$ $\quad df_2 = 9$

$s_1^2 = .0018989$ $\quad s_2^2 = .0023378$

H_0: $\sigma_1^2 = \sigma_2^2$ $\quad \alpha = .10$ $\quad \alpha/2 = .05$

H_a: $\sigma_1^2 \neq \sigma_2^2$

Upper tail critical F value = $F_{.05,9,9} = 3.18$

Lower tail critical F value = $F_{.95,9,9} = 0.314$

$$F = \frac{s_1^2}{s_2^2} = \frac{.0018989}{.0023378} = \mathbf{0.81}$$

Since the observed *F* = 0.81 is greater than the lower tail critical value of 0.314 and less than the upper tail critical value of 3.18, the decision is to **fail**

to reject the null hypothesis.

10.42 Let Houston = group 1 and Chicago = group 2

1) H_0: $\sigma_1^2 = \sigma_2^2$

H_a: $\sigma_1^2 \neq \sigma_2^2$

2) $F = \dfrac{s_1^2}{s_2^2}$

3) $\alpha = .01$

4) $df_1 = 12$ $df_2 = 10$ This is a two-tailed test

The critical table F values are: $F_{.005,12,10} = 5.66$ $F_{.995,10,12} = .177$

If the observed value is greater than 5.66 or less than .177, the decision will be to reject the null hypothesis.

5) $s_1^2 = 393.4$ $s_2^2 = 702.7$

6) $F = \dfrac{393.4}{702.7}$ = **0.56**

7) Since $F = 0.56$ is greater than .177 and less than 5.66,

the decision is to **fail to reject the null hypothesis**.

8) There is no significant difference in the variances of

number of days between Houston and Chicago.

10.43 H_0: $\sigma_1^2 = \sigma_2^2$ $\alpha = .05$ $n_1 = 12$ $s_1 = 7.52$

H_a: $\sigma_1^2 > \sigma_2^2$ $n_2 = 15$ $s_2 = 6.08$

$df_{num} = 12 - 1 = 11$ $df_{denom} = 15 - 1 = 14$

The critical table F value is $F_{.05,10,14} = 2.60$

$$F = \frac{s_1^2}{s_2^2} = \frac{(7.52)^2}{(6.08)^2} = \mathbf{1.53}$$

Since the observed $F = 1.53 < F_{.05,10,14} = 2.60$, the decision is to **fail to reject the null hypothesis**.

10.44 H_0: $\sigma_1^2 = \sigma_2^2$ $\alpha = .01$ $n_1 = 15$ $s_1^2 = 91.5$

H_a: $\sigma_1^2 \neq \sigma_2^2$ $n_2 = 15$ $s_2^2 = 67.3$

$df_{num} = 15 - 1 = 14$ $df_{denom} = 15 - 1 = 14$

The critical table F values are: $F_{.005,12,14} = 4.43$ $F_{.995,14,12} = .226$

$$F = \frac{s_1^2}{s_2^2} = \frac{91.5}{67.3} = \mathbf{1.36}$$

Since the observed $F = 1.36 < F_{.005,12,14} = 4.43$ and $> F_{.995,14,12} = .226$, the decision is to **fail to reject the null hypothesis**.

10.45 H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

For $\alpha = .10$ and a two-tailed test, $\alpha/2 = .05$ and $z_{.05} = \pm 1.645$

<u>Sample 1</u> <u>Sample 2</u>

$\bar{x}_1 = 138.4$ $\bar{x}_2 = 142.5$

$\sigma_1 = 6.71$ $\sigma_2 = 8.92$

$n_1 = 48$ $n_2 = 39$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}} = \frac{(138.4 - 142.5) - (0)}{\sqrt{\frac{(6.71)^2}{48} + \frac{(8.92)}{39}}} = \mathbf{-2.38}$$

Since the observed value of $z = -2.38$ is less than the critical value of $z = -1.645$, the decision is to **reject the null hypothesis**. There is a significant difference in the means of the two populations.

10.46 Sample 1 Sample 2

$\bar{x}_1 = 34.9$ $\bar{x}_2 = 27.6$

$\sigma_1^2 = 2.97$ $\sigma_2^2 = 3.50$

$n_1 = 34$ $n_2 = 31$

For 98% Confidence Level, $z_{.01} = 2.33$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(34.9 - 27.6) \pm 2.33\sqrt{\frac{2.97}{34} + \frac{3.50}{31}} = 7.3 \pm 1.04$$

$6.26 \leq \mu_1 - \mu_2 \leq 8.34$

10.47 H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

Sample 1 Sample 2

$\bar{x}_1 = 2.06$ $\bar{x}_2 = 1.93$

$s_1^2 = .176$ $s_2^2 = .143$

$n_1 = 12$ $n_2 = 15$ $\alpha = .05$

This is a one-tailed test with df = 12 + 15 - 2 = 25. The critical value is

$t_{.05,25} = 1.708$. If the observed value is greater than 1.708, the decision will be to reject the null hypothesis.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(2.06 - 1.93) - (0)}{\sqrt{\frac{(.176)(11) + (.143)(14)}{25}}\sqrt{\frac{1}{12} + \frac{1}{15}}} = \mathbf{0.85}$$

Since the observed value of $t = 0.85$ is less than the critical value of $t = 1.708$, the decision is to **fail to reject the null hypothesis**. The mean for population one is not significantly greater than the mean for population two.

10.48 Sample 1 Sample 2

$\bar{x}_1 = 74.6$ $\bar{x}_2 = 70.9$

$s_1^2 = 10.5$ $s_2^2 = 11.4$

$n_1 = 18$ $n_2 = 19$

For 95% confidence, $\alpha/2 = .025$.

Using df = 18 + 19 - 2 = 35, $t_{30,.025} = 2.042$

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^{\,2}(n_1 - 1) + s_2^{\,2}(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(74.6 - 70.9) \pm 2.042\sqrt{\frac{(10.5)(17) + (11.4)(18)}{18 + 19 - 2}}\sqrt{\frac{1}{18} + \frac{1}{19}}$$

3.7 ± 2.22

$1.48 \le \mu_1 - \mu_2 \le 5.92$

10.49 H_o: $D = 0$ $\alpha = .01$

H_a: $D < 0$

$n = 21$ df = 20 $\bar{d} = -1.16$ $s_d = 1.01$

The critical $t_{.01,20} = -2.528$. If the observed t is less than -2.528, then the decision will be to reject the null hypothesis.

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{-1.16 - 0}{\frac{1.01}{\sqrt{21}}} = \mathbf{-5.26}$$

Since the observed value of *t* = -5.26 is less than the critical t value of -2.528, the decision is to **reject the null hypothesis**. The population difference is less

than zero.

10.50

Respondent	Before	After	d
1	47	63	-16
2	33	35	- 2
3	38	36	2
4	50	56	- 6
5	39	44	- 5
6	27	29	- 2
7	35	32	3
8	46	54	- 8
9	41	47	- 6

$\bar{d}$ = -4.44 s_d = 5.703 df = 8

For a 99% Confidence Level, $\alpha/2$ = .005 and $t_{8,.005}$ = 3.355

$$\bar{d} \pm t\frac{s_d}{\sqrt{n}} = -4.44 \pm 3.355\frac{5.703}{\sqrt{9}} = -4.44 \pm 6.38$$

$-10.82 \leq D \leq 1.94$

10.51 H_o: $p_1 - p_2 = 0$ $\alpha = .05$ $\alpha/2 = .025$

H_a: $p_1 - p_2 \neq 0$ $z_{.025} = \pm 1.96$

If the observed value of z is greater than 1.96 or less than -1.96, then the decision will be to reject the null hypothesis.

<u>Sample 1</u>   <u>Sample 2</u>

$x_1 = 345$ $x_2 = 421$

$n_1 = 783$ $n_2 = 896$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{345 + 421}{783 + 896}$$

$= .4562$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{345}{783}$$

$= .4406$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{421}{896}$$

$= .4699$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p} \cdot \bar{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.4406 - .4699) - (0)}{\sqrt{(.4562)(.5438)\left(\frac{1}{783} + \frac{1}{896}\right)}}$$

$=$ **-1.20**

Since the observed value of z = -1.20 is greater than -1.96, the decision is to **fail to reject the null hypothesis**. There is no significant difference.

10.52 <u>Sample 1</u>   <u>Sample 2</u>

$n_1 = 409$ $n_2 = 378$

$\hat{p}_1 = .71$ $\hat{p}_2 = .67$

For a 99% Confidence Level, $\alpha/2 = .005$ and $z_{.005} = 2.575$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.71 - .67) \pm 2.575\sqrt{\frac{(.71)(.29)}{409} + \frac{(.67)(.33)}{378}} = .04 \pm .085$$

$-.045 \le p_1 - p_2 \le .125$

10.53 H_0: $\sigma_1^2 = \sigma_2^2$ $\alpha = .05$ $n_1 = 8$ $s_1^2 = 46$

H_a: $\sigma_1^2 \neq \sigma_2^2$ $n_2 = 10$ $s_2^2 = 37$

$df_{num} = 8 - 1 = 7$ $df_{denom} = 10 - 1 = 9$

The critical F values are: $F_{.025,7,9} = 4.20$ $F_{.975,9,7} = .238$

If the observed value of F is greater than 4.20 or less than .238, then the decision will be to reject the null hypothesis.

$$F = \frac{s_1^2}{s_2^2} = \frac{46}{37} = \mathbf{1.24}$$

Since the observed F = 1.24 is less than $F_{.025,7,9}$ =4.20 and greater than

$F_{.975,9,7}$ = .238, the decision is to **fail to reject the null hypothesis**. There is no significant difference in the variances of the two populations.

10.54 Term Whole Life

$\bar{x}_t$ = \$75,000 $\bar{x}_w$ = \$45,000

s_t = \$22,000 s_w = \$15,500

n_t = 27 n_w = 29

df = 27 + 29 - 2 = 54

For a 95% Confidence Level, $\alpha/2$ = .025 and $t_{.025,50}$ = 2.009 (used df=50)

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^{\,2}(n_1-1)+s_2^{\,2}(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(75{,}000 - 45{,}000) \pm 2.009\sqrt{\frac{(22{,}000)^2(26)+(15{,}500)^2(28)}{27+29-2}}\sqrt{\frac{1}{27}+\frac{1}{29}}$$

30,000 ± 10,160.11

19,839.89 $\le \mu_1 - \mu_2 \le$ 40,160.11

10.55

Morning	Afternoon	*d*
43	41	2

51	49	2
37	44	-7
24	32	-8
47	46	1
44	42	2
50	47	3
55	51	4
46	49	-3

$n = 9$ $\qquad \bar{d} = -0.444$ $\qquad s_d = 4.447$ $\qquad df = 9 - 1 = 8$

For a 90% Confidence Level: $\alpha/2 = .05$ and $t_{.05,8} = 1.86$

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

$$-0.444 \pm (1.86)\frac{4.447}{\sqrt{9}} = -0.444 \pm 2.757$$

$-3.201 \le D \le 2.313$

10.56 Marketing Accountants

$n_1 = 400$ $\qquad n_2 = 450$

$x_1 = 220$ $\qquad x_2 = 216$

H_o: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 > 0$ $\alpha = .01$

The critical table z value is: $z_{.01} = 2.33$

$$\hat{p}_1 = \frac{220}{400} = .55 \qquad \hat{p}_2 = \frac{216}{450} = .48$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{220 + 216}{400 + 450} = .513$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p} \cdot \bar{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.55 - .48) - (0)}{\sqrt{(.513)(.487)\left(\frac{1}{400} + \frac{1}{450}\right)}} = \mathbf{2.04}$$

Since the observed $z = 2.04$ is less than $z_{.01} = 2.33$, the decision is to **fail to reject the null hypothesis**. There is no significant difference between marketing managers and accountants in the proportion who keep track of obligations "in their head".

10.57 <u>Accounting</u> <u>Data Entry</u>

$n_1 = 16$ $n_2 = 14$

$\bar{x}_1 = 26{,}400$ $\bar{x}_2 = 25{,}800$

$s_1 = 1{,}200$ $s_2 = 1{,}050$

H_0: $\sigma_1^2 = \sigma_2^2$

H_a: $\sigma_1^2 \neq \sigma_2^2$ $\alpha = .05$ and $\alpha/2 = .025$

$df_{num} = 16 - 1 = 15$ $df_{denom} = 14 - 1 = 13$

The critical F values are: $F_{.025,15,13} = 3.05$ $F_{.975,15,13} = 0.33$

$F = \frac{s_1^2}{s_2^2} = \frac{1{,}440{,}000}{1{,}102{,}500} =$ **1.31**

Since the observed $F = 1.31$ is less than $F_{.025,15,13} = 3.05$ and greater than

$F_{.975,15,13} = 0.33$, the decision is to **fail to reject the null hypothesis**.

10.58 Men Women

$n_1 = 60$ $n_2 = 41$

$\bar{x}_1 = 631$ $\bar{x}_2 = 848$

$\sigma_1 = 100$ $\sigma_2 = 100$

For a 95% Confidence Level, $\alpha/2 = .025$ and $z_{.025} = 1.96$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

$$(631 - 848) \pm 1.96\sqrt{\frac{100^2}{60} + \frac{100^2}{41}} = -217 \pm 39.7$$

$$\mathbf{-256.7 \le \mu_1 - \mu_2 \le -177.3}$$

10.59 H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .01$

H_a: $\mu_1 - \mu_2 \neq 0$ df = 20 + 24 - 2 = 42

Detroit Charlotte

n_1 = 20 n_2 = 24

$\bar{x}_1$ = 17.53 $\bar{x}_2$ = 14.89

s_1 = 3.2 s_2 = 2.7

For two-tail test, $\alpha/2$ = .005 and the critical $t_{.005,40}$ = ±2.704 (used df=40)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$t = \frac{(17.53-14.89)-(0)}{\sqrt{\frac{(3.2)^2(19)+(2.7)^2(23)}{42}}\sqrt{\frac{1}{20}+\frac{1}{24}}} = \mathbf{2.97}$$

Since the observed t = 2.97 > $t_{.005,40}$ = 2.704, the decision is to **reject the null hypothesis**.

10.60 <u>With Fertilizer</u> <u>Without Fertilizer</u>

$\bar{x}_1$ = 38.4 $\bar{x}_2$ = 23.1

σ_1 = 9.8 σ_2 = 7.4

$n_1 = 35$ $n_2 = 35$

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

For one-tail test, $\alpha = .01$ and $z_{.01} = 2.33$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(38.4 - 23.1) - (0)}{\sqrt{\frac{(9.8)^2}{35} + \frac{(7.4)^2}{35}}} = \mathbf{7.37}$$

Since the observed $z = 7.37 > z_{.01} = 2.33$, the decision is to **reject the null hypothesis**.

10.61 Specialty Discount

$n_1 = 350$ $n_2 = 500$

$\hat{p}_1 = .75$ $\hat{p}_2 = .52$

For a 90% Confidence Level, $\alpha/2 = .05$ and $z_{.05} = 1.645$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.75 - .52) \pm 1.645\sqrt{\frac{(.75)(.25)}{350} + \frac{(.52)(.48)}{500}} = .23 \pm .053$$

$$\mathbf{.177 \le p_1 - p_2 \le .283}$$

10.62 H_0: $\sigma_1^2 = \sigma_2^2$ $\alpha = .01$ $n_1 = 8$ $n_2 = 7$

H_a: $\sigma_1^2 \neq \sigma_2^2$ $s_1^2 = 72{,}909$ $s_2^2 = 129{,}569$

$df_{num} = 6$ $df_{denom} = 7$

The critical F values are: $F_{.005,6,7} = 9.16$ $F_{.995,7,6} = .11$

$$F = \frac{s_1^{\,2}}{s_2^{\,2}} = \frac{129{,}569}{72{,}909} = \mathbf{1.78}$$

Since $F = 1.78 < F_{.005,6,7} = 9.16$ but also $> F_{.995,7,6} = .11$, the decision is to **fail to reject the null hypothesis**. There is no difference in the variances of the shifts.

10.63	Name Brand	Store Brand	d
	54	49	5
	55	50	5
	59	52	7
	53	51	2
	54	50	4
	61	56	5
	51	47	4
	53	49	4

$n = 8$ $\quad \bar{d} = 4.5$ $\quad s_d = 1.414$ $\quad$ df = 8 - 1 = 7

For a 90% Confidence Level, $\alpha/2 = .05$ and $t_{.05,7} = 1.895$

$$\bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

$$4.5 \pm 1.895 \frac{1.414}{\sqrt{8}} = 4.5 \pm .947$$

$3.553 \le D \le 5.447$

10.64 H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .01$

H_a: $\mu_1 - \mu_2 < 0$ df = 23 + 19 - 2 = 40

Wisconsin	Tennessee
$n_1 = 23$	$n_2 = 19$
$\bar{x}_1 = 69.652$	$\bar{x}_2 = 71.7368$
$s_1^2 = 9.9644$	$s_2^2 = 4.6491$

For one-tail test, $\alpha = .01$ and the critical $t_{.01,40} = -2.423$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(69.652 - 71.7368) - (0)}{\sqrt{\frac{(9.9644)(22) + (4.6491)(18)}{40}}\sqrt{\frac{1}{23} + \frac{1}{19}}} = \mathbf{-2.44}$$

Since the observed $t = -2.44 < t_{.01,40} = -2.423$, the decision is to **reject the null hypothesis**.

10.65

Wednesday	Friday	d
71	53	18
56	47	9
75	52	23
68	55	13
74	58	16

$n = 5$ $\bar{d} = 15.8$ $s_d = 5.263$ df = 5 - 1 = 4

H_o: $D = 0$ $\alpha = .05$

H_a: $D > 0$

For one-tail test, $\alpha = .05$ and the critical $t_{.05,4} = 2.132$

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{15.8 - 0}{\frac{5.263}{\sqrt{5}}} = \mathbf{6.71}$$

Since the observed $t = 6.71 > t_{.05,4} = 2.132$, the decision is to **reject the null hypothesis**.

10.66 H_o: $p_1 - p_2 = 0$ $\alpha = .05$

H_a: $p_1 - p_2 \neq 0$

Machine 1 | Machine 2

$x_1 = 38$ | $x_2 = 21$

$n_1 = 191$ | $n_2 = 202$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{38}{191} = .199$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{21}{202} = .104$$

$$\overline{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{(.199)(191) + (.104)(202)}{191 + 202} = .15$$

For two-tail, $\alpha/2 = .025$ and the critical z values are: $z_{.025} = \pm 1.96$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\overline{p}\cdot\overline{q}\left(\frac{1}{n_1} + \frac{1}{n}\right)}} = \frac{(.199 - .104) - (0)}{\sqrt{(.15)(.85)\left(\frac{1}{191} + \frac{1}{202}\right)}} = \mathbf{2.64}$$

Since the observed $z = 2.64 > z_c = 1.96$, the decision is to **reject the null hypothesis**.

10.67 Construction | Telephone Repair

$n_1 = 338$ | $n_2 = 281$

$x_1 = 297$ | $x_2 = 192$

$\hat{p}_1 = \frac{x_1}{n_1} = \frac{297}{338}$ $= .879$ $\hat{p}_2 = \frac{x_2}{n_2} = \frac{192}{281}$ $= .683$

For a 90% Confidence Level, $\alpha/2 = .05$ and $z_{.05} = 1.645$

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(.879 - .683) \pm 1.645\sqrt{\frac{(.879)(.121)}{338} + \frac{(.683)(.317)}{281}} = .196 \pm .054$$

$$\mathbf{.142 \le p_1 - p_2 \le .250}$$

10.68 Aerospace Automobile

Aerospace	Automobile
$n_1 = 33$	$n_2 = 35$
$\bar{x}_1 = 12.4$	$\bar{x}_2 = 4.6$
$\sigma_1 = 2.9$	$\sigma_2 = 1.8$

For a 99% Confidence Level, $\alpha/2 = .005$ and $z_{.005} = 2.575$

$$(\bar{x}_1 - \bar{x}_2) \pm z\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}$$

$(12.4 – 4.6)$ $\pm$ $2.575 \sqrt{\frac{(2.9)^2}{33} + \frac{(1.8)^2}{35}}$ = 7.8 ± 1.52

6.28 $\le \mu_1 - \mu_2 \le$ 9.32

10.69 Discount Specialty

Discount	Specialty
$\bar{x}_1$ = \$47.20	$\bar{x}_2$ = \$27.40
σ_1 = \$12.45	σ_2 = \$9.82
n_1 = 60	n_2 = 40

H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .01$

H_a: $\mu_1 - \mu_2 \neq 0$

For two-tail test, $\alpha/2 = .005$ and $z_c = \pm 2.575$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1} + \frac{\sigma_2^{\,2}}{n_2}}} = \frac{(47.20 - 27.40) - (0)}{\sqrt{\frac{(12.45)^2}{60} + \frac{(9.82)^2}{40}}} = \mathbf{8.86}$$

Since the observed $z = 8.86 > z_c = 2.575$, the decision is to **reject the null**

hypothesis.

10.70

Before	After	d
12	8	4
7	3	4
10	8	2
16	9	7
8	5	3

$n = 5$ $\bar{d} = 4.0$ $s_d = 1.8708$ df = 5 - 1 = 4

H_o: $D = 0$ $\alpha = .01$

H_a: $D > 0$

For one-tail test, $\alpha = .01$ and the critical $t_{.01,4} = 3.747$

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}} = \frac{4.0 - 0}{\frac{1.8708}{\sqrt{5}}} = \mathbf{4.78}$$

Since the observed $t = 4.78 > t_{.01,4} = 3.747$, the decision is to **reject the null hypothesis**.

10.71 H_o: $\mu_1 - \mu_2 = 0$ $\alpha = .01$

H_a: $\mu_1 - \mu_2 \neq 0$ df = 10 + 6 - 2 = 14

A	B
$n_1 = 10$	$n_2 = 6$
$\bar{x}_1 = 18.3$	$\bar{x}_2 = 9.667$
$s_1^2 = 17.122$	$s_2^2 = 7.467$

For two-tail test, $\alpha/2 = .005$ and the critical $t_{.005,14} = \pm 2.977$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(18.3 - 9.667) - (0)}{\sqrt{\frac{(17.122)(9) + (7.467)(5)}{14}}\sqrt{\frac{1}{10} + \frac{1}{6}}} = \mathbf{4.52}$$

Since the observed $t = 4.52 > t_{.005,14} = 2.977$, the decision is to **reject the null hypothesis**.

10.72 A t test was used to test to determine if Hong Kong has significantly different rates than Mumbai. Let group 1 be Hong Kong.

H_o: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 \neq 0$

$n_1 = 19$ $n_2 = 23$ $\bar{x}_1 = 130.4$ $\bar{x}_2 = 128.4$

$s_1 = 12.9$ $s_2 = 13.9$ 98% C.I. and $\alpha/2 = .01$

t = 0.48 with a *p*-value of .634 which is not significant at of .05. There is not

enough evidence in these data to declare that there is a difference in the average

rental rates of the two cities.

10.73 H_0: $D = 0$

H_a: $D \neq 0$

This is a related measures before and after study. Fourteen people were involved in the study. Before the treatment, the sample mean was 4.357 and after the treatment, the mean was 5.214. The higher number after the treatment indicates that subjects were more likely to "blow the whistle" after having been through the treatment. The observed *t* value was –3.12 which was more extreme than two-tailed table *t* value of $\underline{+}$ 2.16 and as a result, the researcher rejects the null hypothesis. This is underscored by a *p*-value of .0081 which is less than $\alpha = .05$. The study concludes that there is a significantly higher likelihood of "blowing the whistle" after the treatment.

10.74 The point estimates from the sample data indicate that in the northern city the market share is .31078 and in the southern city the market share is .27013. The point estimate for the difference in the two proportions of market share are .04065. Since the 99% confidence interval ranges from -.03936 to +.12067 and zero is in the interval, any hypothesis testing decision based on this interval would result in failure to reject the null hypothesis. Alpha is .01 with a two-tailed test. This is underscored by an observed z value of 1.31 which has an associated p-value of .191 which, of course, is not significant for any of the usual values of α.

10.75 A test of differences of the variances of the populations of the two machines is being computed. The hypotheses are:

H_0: $\sigma_1^2 = \sigma_2^2$

H_a: $\sigma_1^2 > \sigma_2^2$

Twenty-six pipes were measured for sample one and twenty-eight pipes were measured for sample two. The observed $F = 2.0575$ is significant at $\alpha = .05$ for a one-tailed test since the associated p-value is .034787. The variance of pipe lengths for machine 1 is significantly greater than the variance of pipe lengths for machine 2.

Chapter 11
Analysis of Variance and Design of Experiments

LEARNING OBJECTIVES

The focus of this chapter is learning about the design of experiments and the analysis of variance thereby enabling you to:

1. Understand the differences between various experiment designs and when to use them.
2. Compute and interpret the results of a one-way ANOVA.
3. Compute and interpret the results of a random block design.
4. Compute and interpret the results of a two-way ANOVA.
5. Understand and interpret interaction.
6. Know when and how to use multiple comparison techniques.

CHAPTER TEACHING STRATEGY

This important chapter opens the door for students to a broader view of statistics than they have seen to this time. Through the topic of experimental designs, the student begins to understand how they can scientifically set up controlled experiments in which to test certain hypotheses. They learn about independent and dependent variables. With the completely randomized design, the student can see how the *t* test for two independent samples can be expanded to include three or more samples by using analysis of variance. This is something that some of the more curious students were probably wondering about in chapter 10. Through the randomized block design and the factorial designs, the student can understand how we can analyze not only multiple categories of one variable, but we can simultaneously analyze multiple variables with several categories each. Thus, this chapter affords the instructor an opportunity to help the student develop a structure for statistical analysis.

In this chapter, we emphasize that the total sum of squares in a given problem do not change. In the completely randomized design, the total sums of squares are parceled into between treatments sum of squares and error sum of squares. By using a blocking design when there is significant blocking, the blocking effects are removed from the error effect, which reduces the size of the mean square error and can potentially create a more powerful test of the treatment. A similar thing happens in the two-way factorial design when one significant treatment variable siphons off the sum of squares from the error term that reduces the mean square error and creates the potential for a more powerful test of the other treatment variable.

In presenting the random block design in this chapter, the emphasis is on determining if the *F* value for the *treatment* variable is significant or not. There is a de- emphasis on examining the *F* value of the blocking effects. However, if the blocking effects are not significant, the random block design may be a less powerful analysis of the treatment effects. If the blocking effects are not significant, even though the error sum of squares is reduced, the mean square error might increase because the blocking effects may reduce the degrees of freedom error in a proportionaly greater amount. This might result in a smaller treatment *F* value than would occur in a completely randomized design. The repeated-measures design is shown in the chapter as a special case of the random block design.

In factorial designs, if there are multiple values in the cells, it is possible to analyze interaction effects. Random block designs do not have multiple values in cells and therefore interaction effects cannot be calculated. It is emphasized in this chapter that if significant interaction occurs, then the main effects analysis are confounded and should not be analyzed in the usual manner. There are various philosophies about how to handle significant interaction but are beyond the scope of this chapter. The main factorial example problem in the chapter was created to have no significant interaction so that the student can learn how to analyze main effects. The demonstration problem has significant interaction and these interactions are displayed graphically for the student to see. You might consider taking this same problem and graphing the interactions using row effects along the *x* axis and graphing the column means for the student to see.

There are a number of multiple comparison tests available. In this text, one of the more well-known tests, Tukey's HSD, is featured in the case of equal sample sizes. When sample sizes are unequal, a variation on Tukey’s HSD, the Tukey-Kramer test, is used. MINITAB uses the Tukey test as one of its options under multiple comparisons and uses the Tukey-Kramer test for unequal sample sizes. Tukey's HSD is one of the more powerful multiple comparison tests but protects less against Type I errors than some of the other tests.

CHAPTER OUTLINE

Using a Computer to Do a Two-Way ANOVA

KEY TERMS

a posteriori

a priori

Analysis of Variance (ANOVA)

Blocking Variable

Classification Variables

Classifications

Completely Randomized Design

Concomitant Variables

Confounding Variables

Dependent Variable

Experimental Design

F Distribution

F Value

Factorial Design

Factors

Independent Variable

Interaction

Levels

Multiple Comparisons

One-way Analysis of Variance

Post-hoc

Randomized Block Design

Repeated Measures Design

Treatment Variable

Tukey-Kramer Procedure

Tukey's HSD Test

Two-way Analysis of Variance

SOLUTIONS TO PROBLEMS IN CHAPTER 11

11.1 a) Time Period, Market Condition, Day of the Week, Season of the Year

b) Time Period - 4 P.M. to 5 P.M. and 5 P.M. to 6 P.M.

Market Condition - Bull Market and Bear Market

Day of the Week - Monday, Tuesday, Wednesday, Thursday, Friday

Season of the Year – Summer, Winter, Fall, Spring

c) Volume, Value of the Dow Jones Average, Earnings of Investment Houses

11.2 a) Type of 737, Age of the plane, Number of Landings per Week of the plane,

City that the plane is based

b) Type of 737 - Type I, Type II, Type III

Age of plane - 0-2 y, 3-5 y, 6-10 y, over 10 y

Number of Flights per Week - 0-5, 6-10, over 10

City - Dallas, Houston, Phoenix, Detroit

c) Average annual maintenance costs, Number of annual hours spent on

maintenance

11.3 a) Type of Card, Age of User, Economic Class of Cardholder, Geographic Region

b) Type of Card - Mastercard, Visa, Discover, American Express

Age of User - 21-25 y, 26-32 y, 33-40 y, 41-50 y, over 50

Economic Class - Lower, Middle, Upper

Geographic Region - NE, South, MW, West

c) Average number of card usages per person per month,

Average balance due on the card, Average per expenditure per person,

Number of cards possessed per person

11.4 Average dollar expenditure per day/night, Age of adult registering the family, Number of days stay (consecutive)

11.5

Source	df	SS	MS	F
Treatment	2	22.20	11.10	11.07
Error	14	14.03	1.00	
Total	16	36.24		

$\alpha = .05$ Critical $F_{.05,2,14} = 3.74$

Since the observed $F = 11.07 > F_{.05,2,14} = 3.74$, the decision is to **reject the null hypothesis**.

11.6

Source	df	SS	MS	F
Treatment	4	93.77	23.44	15.82
Error	18	26.67	1.48	
Total	22	120.43		

$\alpha = .01$ Critical $F_{.01,4,18} = 4.58$

Since the observed $F = 15.82 > F_{.01,4,18} = 4.58$, the decision is to **reject the null hypothesis**.

11.7

Source	**df**	**SS**	**MS**	***F***
Treatment	3	544.2	181.4	13.00
Error	12	167.5	14.0	
Total	15	711.8		

$\alpha = .01$ Critical $F_{.01,3,12} = 5.95$

Since the observed $F = 13.00 > F_{.01,3,12} = 5.95$, the decision is to **reject the null hypothesis**.

11.8

Source	**df**	**SS**	**MS**	***F***
Treatment	1	64.29	64.29	17.76
Error	12	43.43	3.62	
Total	13	107.71		

$\alpha = .05$ Critical $F_{.05,1,12} = 4.75$

Since the observed $F = 17.76 > F_{.05,1,12} = 4.75$, the decision is to **reject the null hypothesis**.

Observed t value using t test:

1	2
$n_1 = 7$	$n_2 = 7$
$\bar{x}_1 = 29$	$\bar{x}_2 = 24.71429$
$s_1^2 = 3$	$s_2^2 = 4.238095$

$$t = \frac{(29 - 24.71429) - (0)}{\sqrt{\frac{3(6) + (4.238095)(6)}{7 + 7 - 2}}\sqrt{\frac{1}{7} + \frac{1}{7}}} = \mathbf{4.214}$$

Also, $t = \sqrt{F} = \sqrt{17.76} = 4.214$

11.9

Source	SS	df	MS	F
Treatment	583.39	**4**	**145.8475**	**7.50**
Error	972.18	**50**	**19.4436**	
Total	1,555.57	**54**		

11.10

Source	SS	df	MS	F
Treatment	29.64	2	14.820	3.03
Error	68.42	14	4.887	
Total	98.06	16		

$F_{.05,2,14} = 3.74$

Since the observed $F = 3.03 < F_{.05,2,14} = 3.74$, the decision is to **fail to reject**

the null hypothesis

11.11

Source	df	SS	MS	F

Treatment	3	.007076	.002359	10.10
Error	15	.003503	.000234	
Total	18	.010579		

$\alpha = .01$ Critical $F_{.01,3,15} = 5.42$

Since the observed $F = 10.10 > F_{.01,3,15} = 5.42$, the decision is to **reject the null hypothesis**.

11.12

Source	**df**	**SS**	**MS**	***F***
Treatment	2	180700000	90350000	92.67
Error	12	11699999	975000	
Total	14	192400000		

$\alpha = .01$ Critical $F_{.01,2,12} = 6.93$

Since the observed $F = 92.67 > F_{.01,2,12} = 6.93$, the decision is to **reject the null hypothesis**.

11.13

Source	**df**	**SS**	**MS**	***F***
Treatment	2	29.61	14.80	11.76
Error	15	18.89	1.26	
Total	17	48.50		

$\alpha = .05$ Critical $F_{.05,2,15} = 3.68$

Since the observed $F = 11.76 > F_{.05,2,15} = 3.68$, the decison is to **reject the null hypothesis**.

11.14

Source	df	SS	MS	*F*
Treatment	3	456630	152210	11.03
Error	16	220770	13798	
Total	19	677400		

$\alpha = .05$ Critical $F_{.05,3,16} = 3.24$

Since the observed $F = 11.03 > F_{.05,3,16} = 3.24$, the decision is to **reject the null hypothesis**.

11.15 There are **4 treatment levels**. The sample sizes are **18, 15, 21, and 11.** The *F* value is **2.95** with a *p*-value of **.04.** There is an overall significant difference at alpha of .05. The means are **226.73, 238.79, 232.58, and 239.82.**

11.16 The independent variable for this study was *plant* with five classification levels (the five plants). There were a total of 43 workers who participated in the study. The dependent variable was *number of hours worked per* *week*. An observed F value of **3.10** was obtained with an associated p-value of **.026595**. With an alpha of .05, there was a **significant overall difference** in the average number of hours worked per week by plant. A cursory glance at the plant averages revealed that workers at plant 3 averaged 61.47 hours per week (highest number) while workers at plant 4 averaged 49.20 (lowest number).

11.17 $C = 6$ MSE = .3352 $\alpha = .05$ $N = 46$

$q_{.05,6,40} = 4.23$ $n_3 = 8$ $n_6 = 7$ $\bar{x}_3 = 15.85$ $\bar{x}_6 = 17.2$

$$\text{HSD} = 4.23\sqrt{\frac{.3352}{2}\left(\frac{1}{8}+\frac{1}{7}\right)} = 0.896$$

$$\left|\bar{x}_3 - \bar{x}_6\right| = \left|15.85 - 17.21\right| = 1.36$$

Since 1.36 > 0.896, **there is a significant difference between the means of groups 3 and 6**.

11.18 $C = 4$ $n = 6$ $N = 24$ $df_{error} = N - C = 24 - 4 = 20$ $\alpha = .05$

MSE = 2.389 $q_{.05,4,20}$ = 3.96

HSD = $q\sqrt{\frac{MSE}{n}}$ = $(3.96)\sqrt{\frac{2.389}{6}}$ = **2.50**

11.19 C = 3 MSE = 1.002381 α = .05 N = 17 $N - C$ = 14

$q_{.05,3,14}$ = 3.70 n_1 = 6 n_2 = 5 $\bar{x}_1$ = 2 $\bar{x}_2$ = 4.6

HSD = $3.70\sqrt{\frac{1.002381}{2}\left(\frac{1}{6}+\frac{1}{5}\right)}$ = 1.586

$\left|\bar{x}_1 - \bar{x}_2\right| = \left|2 - 4.6\right|$ = 2.6

Since 2.6 > 1.586, **there is a significant difference between the means of groups 1 and 2**.

11.20 From problem 11.6, MSE = 1.481481 C = 5 N = 23 N – C = 18

n_2 = 5 n_4 = 5 α = .01 $q_{.01,5,18}$ = 5.38

HSD = 5.38 $\sqrt{\frac{1.481481}{2}\left(\frac{1}{5}+\frac{1}{5}\right)}$ = 2.93

$\bar{x}_2$ = 10 $\bar{x}_4$ = 16

$\left|\bar{x}_3 - \bar{x}_6\right| = |10-16|$ = 6

Since 6 > 2.93, **there is a significant difference in the means of groups 2 and 4**.

11.21 $N = 16$ $n = 4$ $C = 4$ $N - C = 12$ MSE = 13.95833
$q_{.01,4,12} = 5.50$

$$\text{HSD} = q\sqrt{\frac{MSE}{n}} = 5.50\sqrt{\frac{13.95833}{4}} = 10.27$$

$\bar{x}_1 = 115.25$ $\bar{x}_2 = 125.25$ $\bar{x}_3 = 131.5$ $\bar{x}_4 = 122.5$

$\bar{x}_1$ and $\bar{x}_3$ are the only pair that are significantly different using the HSD test.

11.22 $n = 7$ $C = 2$ MSE = 3.619048 $N = 14$ $N - C = 14 - 2 = 12$

$\alpha = .05$ $q_{.05,2,12} = 3.08$

$$\text{HSD} = q\sqrt{\frac{MSE}{n}} = 3.08\sqrt{\frac{3.619048}{7}} = 2.215$$

$\bar{x}_1 = 29$ and $\bar{x}_2 = 24.71429$

Since $\bar{x}_1 - \bar{x}_2 = 4.28571 > \text{HSD} = 2.215$, the decision is to **reject the null hypothesis**.

11.23 $C = 4$ MSE = .000234 $\alpha = .01$ $N = 19$ $N - C = 15$

$q_{.01,4,15} = 5.25$ $n_1 = 4$ $n_2 = 6$ $n_3 = 5$ $n_4 = 4$

$\bar{x}_1 = 4.03$, $\bar{x}_2 = 4.001667$, $\bar{x}_3 = 3.974$, $\bar{x}_4 = 4.005$

$$\text{HSD}_{1,2} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{4}+\frac{1}{6}\right)} = .0367$$

$$\text{HSD}_{1,3} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{4}+\frac{1}{5}\right)} = .0381$$

$$\text{HSD}_{1,4} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{4}+\frac{1}{4}\right)} = .0402$$

$$\text{HSD}_{2,3} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{6}+\frac{1}{5}\right)} = .0344$$

$$\text{HSD}_{2,4} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{6}+\frac{1}{4}\right)} = .0367$$

$$\mathrm{HSD}_{3,4} = 5.25\sqrt{\frac{.000234}{2}\left(\frac{1}{5}+\frac{1}{4}\right)} = .0381$$

$$\left|\bar{x}_1 - \bar{x}_3\right|$$

= .056

This is the only pair of means that are significantly different.

11.24 $\alpha = .01$ $C = 3$ $n = 5$ $N = 15$ $N - C = 12$

MSE = 975,000 $q_{.01,3,12} = 5.04$

$$\text{HSD} = q\sqrt{\frac{MSE}{n}} = 5.04\sqrt{\frac{975{,}000}{5}} = 2{,}225.6$$

$\bar{x}_1 = 40{,}900$ $\bar{x}_2 = 49{,}400$ $\bar{x}_3 = 45{,}300$

$\left|\bar{x}_1 - \bar{x}_2\right| = \mathbf{8{,}500}$

$\left|\bar{x}_1 - \bar{x}_3\right| = \mathbf{4{,}400}$

$\left|\bar{x}_2 - \bar{x}_3\right| = \mathbf{4{,}100}$

Using Tukey's HSD, **all three pairwise comparisons are significantly different**.

11.25 $\alpha = .05$ $C = 3$ $N = 18$ $N - C = 15$ MSE = 1.259365

$q_{.05,3,15} = 3.67$ $n_1 = 5$ $n_2 = 7$ $n_3 = 6$

$\bar{x}_1 = 7.6$ $\bar{x}_2 = 8.8571$ $\bar{x}_3 = 5.8333$

$$HSD_{1,2} = 3.67\sqrt{\frac{1.259365}{2}\left(\frac{1}{5}+\frac{1}{7}\right)} = 1.705$$

$$HSD_{1,3} = 3.67\sqrt{\frac{1.259365}{2}\left(\frac{1}{5}+\frac{1}{6}\right)} = 1.763$$

$$HSD_{2,3} = 3.67\sqrt{\frac{1.259365}{2}\left(\frac{1}{7}+\frac{1}{6}\right)} = 1.620$$

$\left|\bar{x}_1 - \bar{x}_3\right|$ = **1.767 (is significant)**

$\left|\bar{x}_2 - \bar{x}_3\right|$ = **3.024 (is significant)**

11.26 α = .05 n = 5 C = 4 N = 20 N - C = 16 MSE = 13,798.13

$\bar{x}_1$ = 591 $\bar{x}_2$ = 350 $\bar{x}_3$ = 776 $\bar{x}_4$ = 563

HSD = $q\sqrt{\frac{MSE}{n}}$ = 4.05$\sqrt{\frac{13,798.13}{5}}$ = **212.76**

$\left|\bar{x}_1-\bar{x}_2\right|$ = 241 $\left|\bar{x}_1-\bar{x}_3\right|$ = 185 $\left|\bar{x}_1-\bar{x}_4\right|$ = 28

$\left|\bar{x}_2-\bar{x}_3\right|$ = 426 $\left|\bar{x}_2-\bar{x}_4\right|$ = 213 $\left|\bar{x}_3-\bar{x}_4\right|$ = 213

Using Tukey's HSD = 212.76, means 1 and 2, means 2 and 3, means 2 and 4,

and means 3 and 4 are significantly different.

11.27 α = .05. There were five plants and ten pairwise comparisons. The MINITAB

output reveals that the only significant pairwise difference is between plant 2 and plant 3 where the reported confidence interval (0.180 to 22.460) contains the same sign throughout indicating that 0 is not in the interval. Since zero is not in the interval, then we are 95% confident that there is a pairwise difference significantly different from zero. The lower and upper values for all other confidence intervals have different signs indicating that zero is included in the interval. This indicates that the difference in the means for these pairs might be zero.

11.28 H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: At least one treatment mean is different from the others

Source	**df**	**SS**	**MS**	***F***
Treatment	3	62.95	20.9833	5.56
Blocks	4	257.50	64.3750	17.05
Error	12	45.30	3.7750	
Total	19	365.75		

$\alpha = .05$ Critical $F_{.05,3,12} = 3.49$ for treatments

For treatments, the observed $F = 5.56 > F_{.05,3,12} = 3.49$, the decision is to

reject the null hypothesis.

11.29 H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: At least one treatment mean is different from the others

Source	**df**	**SS**	**MS**	***F***
Treatment	2	.001717	.000858	1.48
Blocks	3	.076867	.025622	44.13
Error	6	.003483	.000581	
Total	11	.082067		

$\alpha = .01$ Critical $F_{.01,2,6} = 10.92$ for treatments

For treatments, the observed $F = 1.48 < F_{.01,2,6} = 10.92$ and the decision is to

fail to reject the null hypothesis.

11.30

Source	**df**	**SS**	**MS**	***F***
Treatment	5	2477.53	495.506	1.91
Blocks	9	3180.48	353.387	1.36
Error	45	11661.38	259.142	
Total	59	17319.39		

$\alpha = .05$ Critical $F_{.05,5,45} = 2.45$ for treatments

For treatments, the observed $F = 1.91 < F_{.05,5,45} = 2.45$ and decision is to **fail to**

reject the null hypothesis.

11.31

Source	**df**	**SS**	**MS**	***F***
Treatment	3	199.48	66.493	3.90
Blocks	6	265.24	44.207	2.60
Error	18	306.59	17.033	

Total	27	771.31

$\alpha = .01$ Critical $F_{.01,3,18} = 5.09$ for treatments

For treatments, the observed $F = 3.90 < F_{.01,3,18} = 5.09$ and the decision is to

fail **to reject the null hypothesis**.

11.32

Source	**df**	**SS**	**MS**	***F***
Treatment	3	2302.5	767.5000	15.67
Blocks	9	5402.5	600.2778	12.26
Error	27	1322.5	48.9815	
Total	39	9027.5		

$\alpha = .05$ Critical $F_{.05,3,27} = 2.96$ for treatments

For treatments, the observed $F = 15.67 > F_{.05,3,27} = 2.96$ and the decision is to

reject the null hypothesis.

11.33

Source	df	SS	MS	F
Treatment	2	64.5333	32.2667	15.37
Blocks	4	137.6000	34.4000	16.38
Error	8	16.8000	2.1000	
Total	14	218.9300		

$\alpha = .01$ Critical $F_{.01,2,8} = 8.65$ for treatments

For treatments, the observed $F = 15.37 > F_{.01,2,8} = 8.65$ and the decision is to

reject the null hypothesis.

11.34 This is a randomized block design with 3 treatments (machines) and 5 block levels (operators). The *F* for treatments is 6.72 with a *p*-value of .019. There is a significant difference in machines at α = .05. The *F* for blocking effects is 0.22 with a *p*-value of .807. There are no significant blocking effects. The blocking effects reduced the power of the treatment effects since the blocking effects were not significant.

11.35 The *p* value for Phone Type, .00018, indicates that there is an overall significant difference in treatment means at alpha .001. The lengths of calls differ according to type of telephone used. The *p*-value for managers, .00028, indicates that there is an overall difference in block means at alpha .001. The lengths of calls differ according to Manager. The significant blocking effects have improved the power of the *F* test for treatments.

11.36 This is a two-way factorial design with two independent variables and one dependent variable. It is 2x4 in that there are two row treatment levels and four column treatment levels. Since there are three measurements per cell, interaction can be analyzed.

$df_{\text{row treatment}} = 1$ $df_{\text{column treatment}} = 3$ $df_{\text{interaction}} = 3$ $df_{\text{error}} = 16$ $df_{\text{total}} = 23$

11.37 This is a two-way factorial design with two independent variables and one dependent variable. It is 4x3 in that there are four treatment levels and three column treatment levels. Since there are two measurements per cell, interaction can be analyzed.

$df_{\text{row treatment}} = 3$ $df_{\text{column treatment}} = 2$ $df_{\text{interaction}} = 6$ $df_{\text{error}} = 12$ $df_{\text{total}} = 23$

11.38

Source	**df**	**SS**	**MS**	***F***
Row	3	126.98	42.327	3.46
Column	4	37.49	9.373	0.77
Interaction	12	380.82	31.735	2.60
Error	60	733.65	12.228	
Total	79	1278.94		

$\alpha = .05$

Critical $F_{.05,3,60} = 2.76$ for rows. For rows, the observed $F = 3.46 > F_{.05,3,60} = 2.76$

and the decision is to **reject the null hypothesis**.

Critical $F_{.05,4,60} = 2.53$ for columns. For columns, the observed $F = 0.77 <$

$F_{.05,4,60} = 2.53$ and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,12,60} = 1.92$ for interaction. For interaction, the observed $F = 2.60 >$

$F_{.05,12,60}$ = 1.92 and the decision is to **reject the null hypothesis**.

Since there is significant interaction, the researcher should exercise extreme

caution in analyzing the "significant" row effects.

11.39

Source	df	SS	MS	F
Row	1	1.047	1.047	2.40
Column	3	3.844	1.281	2.94
Interaction	3	0.773	0.258	0.59
Error	16	6.968	0.436	
Total	23	12.632		

α = .05

Critical $F_{.05,1,16}$ = 4.49 for rows. For rows, the observed F = 2.40 < $F_{.05,1,16}$ = 4.49

and decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,3,16}$ = 3.24 for columns. For columns, the observed F = 2.94 <

$F_{.05,3,16}$ = 3.24 and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,3,16}$ = 3.24 for interaction. For interaction, the observed F = 0.59 <

$F_{.05,3,16}$ = 3.24 and the decision is to **fail to reject the null hypothesis**.

11.40

Source	df	SS	MS	F
Row	1	60.750	60.750	38.37
Column	2	14.000	7.000	4.42
Interaction	2	2.000	1.000	0.63
Error	6	9.500	1.583	
Total	11	86.250		

$\alpha = .01$

Critical $F_{.01,1,6}$ = 13.75 for rows. For rows, the observed F = 38.37 > $F_{.01,1,6}$ = 13.75 and the decision is to **reject the null hypothesis**.

Critical $F_{.01,2,6}$ = 10.92 for columns. For columns, the observed F = 4.42 < $F_{.01,2,6}$ = 10.92 and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.01,2,6}$ = 10.92 for interaction. For interaction, the observed F = 0.63 < $F_{.01,2,6}$ = 10.92 and the decision is to **fail to reject the null hypothesis**.

11.41

Source	df	SS	MS	F
Treatment 1	1	1.24031	1.24031	63.67
Treatment 2	3	5.09844	1.69948	87.25
Interaction	3	0.12094	0.04031	2.07

Error	24	0.46750	0.01948
Total	31	6.92719	

$\alpha = .05$

Critical $F_{.05,1,24} = 4.26$ for treatment 1. For treatment 1, the observed $F = 63.67 >$

$F_{.05,1,24} = 4.26$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,3,24} = 3.01$ for treatment 2. For treatment 2, the observed $F = 87.25 >$

$F_{.05,3,24} = 3.01$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,3,24} = 3.01$ for interaction. For interaction, the observed $F = 2.07 <$

$F_{.05,3,24} = 3.01$ and the decision is to **fail to reject the null hypothesis**.

11.42

Source	**df**	**SS**	**MS**	***F***
Age	3	42.4583	14.1528	14.77
No. Children	2	49.0833	24.5417	25.61
Interaction	6	4.9167	0.8194	0.86
Error	12	11.5000	0.9583	
Total	23	107.9583		

$\alpha = .05$

Critical $F_{.05,3,12} = 3.49$ for Age. For Age, the observed $F = 14.77 >$

$F_{.05,3,12} = 3.49$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,2,12} = 3.89$ for No. Children. For No. Children, the observed

$F = 25.61 > F_{.05,2,12} = 3.89$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,6,12} = 3.00$ for interaction. For interaction, the observed $F =$ 0.86 <

$F_{.05,6,12} = 3.00$ and **fail to reject the null hypothesis**.

11.43

Source	df	SS	MS	*F*
Location	2	1736.22	868.11	34.31
Competitors	3	1078.33	359.44	14.20
Interaction	6	503.33	83.89	3.32
Error	24	607.33	25.31	
Total	35	3925.22		

$\alpha = .05$

Critical $F_{.05,2,24} = 3.40$ for rows. For rows, the observed $F = 34.31 >$ $F_{.05,2,24} = 3.40$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,3,24} = 3.01$ for columns. For columns, the observed $F = 14.20 >$ $F_{.05,3,24} = 3.01$ and decision is to **reject the null hypothesis**.

Critical $F_{.05,6,24} = 2.51$ for interaction. For interaction, the observed $F = 3.32 >$ $F_{.05,6,24} = 2.51$ and the decision is to **reject the null hypothesis**.

Note: There is significant interaction in this study. This may confound the interpretation of the main effects, Location and Number of Competitors.

11.44 This two-way design has 3 row treatments and 5 column treatments. There are 45 total observations with 3 in each cell.

F_R = $\dfrac{MS_R}{MS_E} = \dfrac{46.16}{3.49}$ = **13.23**

p-value = .000 and the decision is to **reject the null hypothesis for rows**.

F_C = $\dfrac{MS_C}{MS_E} = \dfrac{249.70}{3.49}$ = **71.57**

p-value = .000 and the decision is to **reject the null hypothesis for columns**.

F_I = $\dfrac{MS_I}{MS_E} = \dfrac{55.27}{3.49}$ = **15.84**

p-value = .000 and the decision is to reject the null hypothesis for interaction.

Because there is significant interaction, **the analysis of main effects is confounded**. The graph of means displays the crossing patterns of the line segments indicating the presence of interaction.

11.45 The null hypotheses are that there are no interaction effects, that there are no significant differences in the means of the valve openings by machine, and that there are no significant differences in the means of the valve openings by shift. Since the *p*-value for interaction effects is .876, there are no significant interaction effects and that is good since significant interaction effects would confound that study. The *p*-value for columns (shifts) is .008 indicating that column effects are significant at alpha of .01. There is a significant difference in the mean valve opening according to shift. No multiple comparisons are given in the output. However, an examination of the shift means indicates that the mean valve opening on shift one was the largest at 6.47 followed by shift three with 6.3 and shift two with 6.25. The *p*-value for rows (machines) is .937 and that is not significant.

11.46 This two-way factorial design has 3 rows and 3 columns with three observations per cell. The observed *F* value for rows is 0.19, for columns is 1.19, and for interaction is 1.40. Using an alpha of .05, the critical *F* value for rows and columns (same df) is $F_{2,18,.05}$ = 3.55. Neither the observed *F* value for rows nor the observed *F* value for columns is significant. The critical *F* value for interaction is $F_{4,18,.05}$ = 2.93. There is no significant *interaction*.

11.47

Source	**df**	**SS**	**MS**	***F***
Treatment	3	66.69	22.23	8.82
Error	12	30.25	2.52	
Total	15	96.94		

α = .05 Critical $F_{.05,3,12}$ = 3.49

Since the treatment F = 8.82 > $F_{.05,3,12}$ = 3.49, the decision is to **reject the null hypothesis**.

For Tukey's HSD:

MSE = 2.52 n = 4 N = 16 C = 4 N - C = 12

$q_{.05,4,12}$ = 4.20

HSD = $q\sqrt{\frac{MSE}{n}}$ = $(4.20)\sqrt{\frac{2.52}{4}}$ = **3.33**

$\bar{x}_1$ = 12 $\bar{x}_2$ = 7.75 $\bar{x}_3$ = 13.25 $\bar{x}_4$ = 11.25

Using HSD of 3.33, there are significant pairwise differences between means 1 and 2, means 2 and 3, and means 2 and 4.

11.48

Source	**df**	**SS**	**MS**	***F***
Treatment	6	68.19	11.365	0.87
Error	19	249.61	13.137	
Total	25	317.80		

11.49

Source	df	SS	MS	F
Treatment	5	210	42.000	2.31
Error	36	655	18.194	
Total	41	865		

11.50

Source	df	SS	MS	F
Treatment	2	150.91	75.46	16.19
Error	22	102.53	4.66	
Total	24	253.44		

$\alpha = .01$ Critical $F_{.01,2,22} = 5.72$

Since the observed $F = 16.19 > F_{.01,2,22} = 5.72$, the decision is to **reject the null hypothesis**.

$\bar{x}_1 = 9.200$ $\bar{x}_2 = 14.250$ $\bar{x}_3 = 8.714286$

$n_1 = 10$ $n_2 = 8$ $n_3 = 7$

MSE = 4.66 C = 3 N = 25 N - C = 22

α = .01 $q_{.01,3,22}$ = 4.64

$$\text{HSD}_{1,2} = 4.64\sqrt{\frac{4.66}{2}\left(\frac{1}{10}+\frac{1}{8}\right)} = 3.36$$

$$\text{HSD}_{1,3} = 4.64\sqrt{\frac{4.66}{2}\left(\frac{1}{10}+\frac{1}{7}\right)} = 3.49$$

$$\text{HSD}_{2,3} = 4.64\sqrt{\frac{4.66}{2}\left(\frac{1}{8}+\frac{1}{7}\right)} = 3.67$$

$\left|\bar{x}_1 - \bar{x}_2\right|$ = **5.05** and $\left|\bar{x}_2 - \bar{x}_3\right|$ = **5.5357** are **significantly different** at α = .01

11.51 This design is a repeated-measures type random block design. There is one

treatment variable with three levels. There is one blocking variable with six

people in it (six levels). The degrees of freedom treatment are two. The degrees

of freedom block are five. The error degrees of freedom are ten. The total

degrees of freedom are seventeen. There is one dependent variable.

11.52

Source	**df**	**SS**	**MS**	***F***
Treatment	3	20,994	6998.00	5.58
Blocks	9	16,453	1828.11	1.46
Error	27	33,891	1255.22	
Total	39	71,338		

$\alpha = .05$ Critical $F_{.05,3,27} = 2.96$ for treatments

Since the observed $F = 5.58 > F_{.05,3,27} = 2.96$ for treatments, the decision is to

reject the null hypothesis.

11.53

Source	df	SS	MS	*F*
Treatment	3	240.125	80.042	31.51
Blocks	5	548.708	109.742	43.20
Error	15	38.125	2.542	
Total	23			

α = .05 Critical $F_{.05,3,15}$ = 3.29 for treatments

Since for treatments the observed F = 31.51 > $F_{.05,3,15}$ = 3.29, the decision is to

reject the null hypothesis.

For Tukey's HSD:

Ignoring the blocking effects, the sum of squares blocking and sum of squares error are combined together for a new SS_{error} = 548.708 + 38.125 = 586.833. Combining the degrees of freedom error and blocking yields a new df_{error} = 20. Using these new figures, we compute a new mean square error, MSE = (586.833/20) = 29.34165.

$n = 6$ $C = 4$ $N = 24$ $N - C = 20$ $q_{.05,4,20} = 3.96$

$$\text{HSD} = q\sqrt{\frac{MSE}{n}} = (3.96)\sqrt{\frac{29.34165}{6}} = \mathbf{8.757}$$

$\bar{x}_1 = 16.667$ $\bar{x}_2 = 12.333$ $\bar{x}_3 = 12.333$ $\bar{x}_4 = 19.833$

None of the pairs of means are significantly different using Tukey's HSD = 8.757.

This may be due in part to the fact that we compared means by folding the blocking effects back into error and the blocking effects were highly significant.

11.54

Source	**df**	**SS**	**MS**	***F***
Treatment 1	4	29.13	7.2825	1.98
Treatment 2	1	12.67	12.6700	3.44
Interaction	4	73.49	18.3725	4.99
Error	30	110.38	3.6793	
Total	39	225.67		

$\alpha = .05$

Critical $F_{.05,4,30}$ = 2.69 for treatment 1. For treatment 1, the observed F = 1.98 < $F_{.05,4,30}$ = 2.69 and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,1,30}$ = 4.17 for treatment 2. For treatment 2 observed F = 3.44 < $F_{.05,1,30}$ = 4.17 and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,4,30}$ = 2.69 for interaction. For interaction, the observed F = 4.99 > $F_{.05,4,30}$ = 2.69 and the decision is to **reject the null hypothesis**.

Since there are significant interaction effects, examination of the main effects should not be done in the usual manner. However, in this case, there are no significant treatment effects anyway.

11.55

Source	**df**	**SS**	**MS**	***F***
Treatment 2	3	257.889	85.963	38.21
Treatment 1	2	1.056	0.528	0.23
Interaction	6	17.611	2.935	1.30
Error	24	54.000	2.250	
Total	35	330.556		

$\alpha = .01$

Critical $F_{.01,3,24} = 4.72$ for treatment 2. For the treatment 2 effects, the observed

$F = 38.21 > F_{.01,3,24} = 4.72$ and the decision is to **reject the null hypothesis**.

Critical $F_{.01,2,24} = 5.61$ for Treatment 1. For the treatment 1 effects, the observed

$F = 0.23 < F_{.01,2,24} = 5.61$ and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.01,6,24} = 3.67$ for interaction. For the interaction effects, the observed

$F = 1.30 < F_{.01,6,24} = 3.67$ and the decision is to **fail to reject the null hypothesis**.

11.56

Source	df	SS	MS	F
Age	2	49.3889	24.6944	38.65
Column	3	1.2222	0.4074	0.64
Interaction	6	1.2778	0.2130	0.33
Error	24	15.3333	0.6389	
Total	35	67.2222		

$\alpha = .05$

Critical $F_{.05,2,24}$ = 3.40 for Age. For the age effects, the observed F = 38.65 > $F_{.05,2,24}$ = 3.40 and the decision is to **reject the null hypothesis**.

Critical $F_{.05,3,24}$ = 3.01 for Region. For the region effects, the observed F = 0.64 < $F_{.05,3,24}$ = 3.01 and the decision is to **fail to reject the null hypothesis**.

Critical $F_{.05,6,24}$ = 2.51 for interaction. For interaction effects, the observed F = 0.33 < $F_{.05,6,24}$ = 2.51 and the decision is to **fail to reject the null hypothesis**.

There are no significant interaction effects. Only the Age effects are significant.

Computing Tukey's HSD for Age:

$\bar{x}_1 = 2.667$ $\bar{x}_2 = 4.917$ $\bar{x}_3 = 2.250$

$n = 12$ $C = 3$ $N = 36$ $N - C = 33$

MSE is recomputed by folding together the interaction and column sum of

squares and degrees of freedom with previous error terms:

MSE = (1.2222 + 1.2778 + 15.3333)/(3 + 6 + 24) = 0.5404

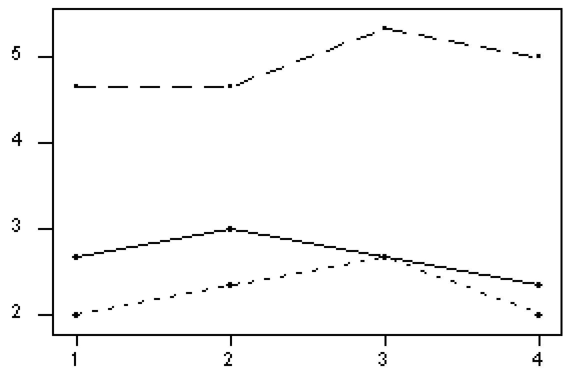
$q_{.05,3,33} = 3.49$

$$\text{HSD} = q\sqrt{\frac{MSE}{n}} = (3.49)\sqrt{\frac{.5404}{12}} = 0.7406$$

Using HSD, there are significant pairwise differences between means 1 and 2 and

between means 2 and 3.

Shown below is a graph of the interaction using the cell means by Age.


5
4
3
2
1
2
3
4

11.57

Source	df	SS	MS	F
Treatment	3	90477679	30159226	7.38
Error	20	81761905	4088095	
Total	23	172000000		

$\alpha = .05$ Critical $F_{.05,3,20} = 3.10$

The treatment $F = 7.38 > F_{.05,3,20} = 3.10$ and the decision is to **reject the null hypothesis**.

11.58

Source	df	SS	MS	F
Treatment	2	460,353	230,176	103.70
Blocks	5	33,524	6,705	3.02
Error	10	22,197	2,220	
Total	17	516,074		

$\alpha = .01$ Critical $F_{.01,2,10} = 7.56$ for treatments

Since the treatment observed $F = 103.70 > F_{.01,2,10} = 7.56$, the decision is to **reject the null hypothesis**.

11.59

Source	df	SS	MS	*F*
Treatment	2	9.555	4.777	0.46
Error	18	185.1337	10.285	
Total	20	194.6885		

$\alpha = .05$ Critical $F_{.05,2,18} = 3.55$

Since the treatment $F = 0.46 > F_{.05,2,18} = 3.55$, the decision is to **fail to reject the null hypothesis**.

Since there are no significant treatment effects, it would make no sense to compute Tukey-Kramer values and do pairwise comparisons.

11.60

Source	df	SS	MS	F
Years	2	4.875	2.437	5.16
Size	3	17.083	5.694	12.06
Interaction	6	2.292	0.382	0.81
Error	36	17.000	0.472	
Total	47	41.250		

$\alpha = .05$

Critical $F_{.05,2,36} = 3.32$ for Years. For Years, the observed $F = 5.16 >$ $F_{.05,2,36} = 3.32$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,3,36} = 2.92$ for Size. For Size, the observed $F = 12.06 >$ $F_{.05,3,36} = 2.92$ and the decision is to **reject the null hypothesis**.

Critical $F_{.05,6,36} = 2.42$ for interaction. For interaction, the observed $F = 0.81 <$ $F_{.05,6,36} = 2.42$ and the decision is to **fail to reject the null hypothesis**.

There are no significant interaction effects. There are significant row and column effects at $\alpha = .05$.

11.61

Source	df	SS	MS	F
Treatment	4	53.400	13.350	13.64
Blocks	7	17.100	2.443	2.50
Error	28	27.400	0.979	
Total	39	97.900		

$\alpha = .05$ Critical $F_{.05,4,28} = 2.71$ for treatments

For treatments, the observed $F = 13.64 > F_{.05,4,28} = 2.71$ and the decision is to

reject the null hypothesis.

11.62 This is a one-way ANOVA with four treatment levels. There are 36 observations in the study. The p-value of .045 indicates that there is a significant overall difference in the means at $\alpha = .05$. An examination of the mean analysis shows that the sample sizes are different with sizes of 8, 7, 11, and 10, respectively. No multiple comparison technique was used here to conduct pairwise comparisons. However, a study of sample means shows that the two most extreme means are from levels one and four. These two means would be the most likely candidates for multiple comparison tests. Note that the confidence intervals for means one and four (shown in the graphical output) are seemingly non-overlapping indicating a potentially significant difference.

11.63 Excel reports that this is a two-factor design without replication indicating that this is a random block design. Neither the row nor the column p-values are less than .05 indicating that there are no significant treatment or blocking effects in this study. Also displayed in the output to underscore this conclusion are the observed and critical F values for both treatments and blocking. In both cases, the observed value is less than the critical value.

11.64 This is a two-way ANOVA with 5 rows and 2 columns. There are 2 observations per cell. For rows, F_R = 0.98 with a *p*-value of .461 which is not significant. For columns, F_C = 2.67 with a *p*-value of .134 which is not significant. For interaction, F_I = 4.65 with a *p*-value of .022 which is significant at α = .05. Thus, there are significant interaction effects and the row and column effects are confounded. An examination of the interaction plot reveals that most of the lines cross verifying the finding of significant interaction.

11.65 This is a two-way ANOVA with 4 rows and 3 columns. There are 3 observations per cell. F_R = 4.30 with a *p*-value of .014 is significant at α = .05. The null hypothesis is rejected for rows. F_C = 0.53 with a *p*-value of .594 is not significant. We fail to reject the null hypothesis for columns. F_I = 0.99 with a *p*-value of .453 for interaction is not significant. We fail to reject the null hypothesis for interaction effects.

11.66 This was a random block design with 5 treatment levels and 5 blocking levels.

For both treatment and blocking effects, the critical value is $F_{.05,4,16}$ = 3.01. The

observed F value for treatment effects is MS_C / MS_E = 35.98 / 7.36 = 4.89 which

is greater than the critical value. The null hypothesis for treatments is rejected,

and we conclude that there is a significant different in treatment means. No

multiple comparisons have been computed in the output. The observed F value

for blocking effects is MS_R / MS_E = 10.36 /7.36 = 1.41 which is less than the

critical value. There are no significant blocking effects. Using random block

design on this experiment might have cost a loss of power.

11.67 This one-way ANOVA has 4 treatment levels and 24 observations. The F = 3.51

yields a p-value of .034 indicating significance at α = .05. Since the sample sizes

are equal, Tukey's HSD is used to make multiple comparisons. The computer output shows that means 1 and 3 are the only pairs that are significantly different

(same signs in confidence interval). Observe on the graph that the confidence intervals for means 1 and 3 barely overlap.

Chapter 12

Analysis of Categorical Data

LEARNING OBJECTIVES

This chapter presents several nonparametric statistics that can be used to analyze data enabling you to:

1. Understand the chi-square goodness-of-fit test and how to use it.

2. Analyze data using the chi-square test of independence.

CHAPTER TEACHING STRATEGY

Chapter 12 is a chapter containing the two most prevalent chi-square tests: chi-square goodness-of-fit and chi-square test of independence. These two techniques are important because they give the statistician a tool that is particularly useful for analyzing nominal data (even though independent variable categories can sometimes have ordinal or higher categories). It should be emphasized that there are many instances in business research where the resulting data gathered are merely categorical identification. For example, in segmenting the market place (consumers or industrial users), information is gathered regarding gender, income level, geographical location, political affiliation, religious preference, ethnicity, occupation, size of company, type of industry, etc. On these variables, the measurement is often a tallying of the frequency of occurrence of individuals, items, or companies in each category. The subject of the research is given no "score" or "measurement" other than a 0/1 for being a member or not of a given category. These two chi-square tests are perfectly tailored to analyze such data.

The chi-square goodness-of-fit test examines the categories of one variable to determine if the distribution of observed occurrences matches some expected or theoretical distribution of occurrences. It can be used to determine if some standard or previously known distribution of proportions is the same as some observed distribution of proportions. It can also be used to validate the theoretical distribution of occurrences of phenomena such as random arrivals that are often assumed to be Poisson distributed. You will note that the degrees of freedom, *k* - 1 for a given set of expected values or for the uniform distribution, change to *k* - 2 for an expected Poisson distribution and to *k* - 3 for an expected normal distribution. To conduct a chi-square goodness-of-fit test to analyze an expected Poisson distribution, the value of lambda must be estimated from the observed data. This causes the loss of an additional degree of freedom. With the normal distribution, both the mean and standard deviation of the expected distribution are estimated from the observed values causing the loss of two additional degrees of freedom from the *k* - 1 value.

The chi-square test of independence is used to compare the observed frequencies along the categories of <u>two</u> independent variables to expected values to determine if the two variables are independent or not. Of course, if the variables are not independent, they are dependent or related. This allows business researchers to reach some conclusions about such questions as: is smoking independent of gender or is type of housing preferred independent of geographic region? The chi-square test of independence is often used as a tool for preliminary analysis of data gathered in exploratory research where the researcher has little idea of what variables seem to be related to what variables, and the data are nominal. This test is particularly useful with demographic type data.

A word of warning is appropriate here. When an expected frequency is small, the observed chi-square value can be inordinately large thus yielding an increased possibility of committing a Type I error. The research on this problem has yielded varying results with some authors indicating that expected values as low as two or three are acceptable and other researchers demanding that expected values be ten or more. In this text, we have settled on the fairly widespread accepted criterion of five or more.

CHAPTER OUTLINE

12.1 Chi-Square Goodness-of-Fit Test

Testing a Population Proportion Using the Chi-square Goodness-of-Fit

Test as an Alternative Technique to the z Test

12.2 Contingency Analysis: Chi-Square Test of Independence

KEY TERMS

Categorical Data	Chi-Square Test of Independence
Chi-Square Distribution	Contingency Analysis
Chi-Square Goodness-of-Fit Test	Contingency Table

SOLUTIONS TO THE ODD-NUMBERED PROBLEMS IN CHAPTER 12

12.1

f_0	f_e	$\frac{(f_o - f_e)^2}{f_e}$
53	68	3.309
37	42	0.595
32	33	0.030
28	22	1.636
18	10	6.400
15	8	6.125

H_o: The observed distribution is the same as the expected distribution.

H_a: The observed distribution is not the same as the expected distribution.

Observed $\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ = **18.095**

df = k - 1 = 6 - 1 = 5, α = .05

$\chi^2_{.05,5}$ = 11.0705

Since the observed $\chi^2 = 18.095 > \chi^2_{.05,5} = 11.0705$, the decision is to **reject the**

null hypothesis.

The observed frequencies are not distributed the same as the expected frequencies.

12.2	f_0	f_e	$\frac{(f_o - f_e)^2}{f_e}$
	19	18	0.056
	17	18	0.056
	14	18	0.889
	18	18	0.000
	19	18	0.056
	21	18	0.500
	18	18	0.000
	18	18	0.000
	$\Sigma f_o =$ 144	$\Sigma f_e =$ 144	1.557

H_o: The observed frequencies are uniformly distributed.

H_a: The observed frequencies are not uniformly distributed.

$$\bar{x} = \frac{\sum f_0}{k} = \frac{144}{8}$$

$= 18$

In this uniform distribution, each $f_e = 18$

df = k – 1 = 8 – 1 = 7, $\alpha = .01$

$\chi^2_{.01,7} = 18.4753$

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

Observed = **1.557**

Since the observed $\chi^2 = 1.557 < \chi^2_{.01,7} = 18.4753$, the decision is to **fail to reject**

the null hypothesis

There is no reason to conclude that the frequencies are not uniformly

distributed.

12.3

Number	f_0	(Number)(f_0)
0	28	0
1	17	17
2	11	22
3	5	15

61

54

H_o: The frequency distribution is Poisson.

H_a: The frequency distribution is not Poisson.

$\lambda = \frac{54}{61} = 0.9$

Number	Expected Probability	Expected Frequency
0	.4066	24.803
1	.3659	22.320
2	.1647	10.047
$\geq$ 3	.0628	3.831

Since f_e for $\geq$ 3 is less than 5, collapse categories 2 and $\geq$3:

Number	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
0	28	24.803	0.412

1	17	22.320	1.268
$\geq$2	16	13.878	0.324
	61	60.993	2.004

df = k - 2 = 3 - 2 = 1, α = .05

$\chi^2_{.05,1}$ = 3.8415

Observed $\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ = **2.001**

Since the observed χ^2 = 2.001 < $\chi^2_{.05,1}$ = 3.8415, the decision is to **fail to reject**

the null hypothesis.

There is insufficient evidence to reject the distribution as Poisson distributed. The conclusion is that the distribution is Poisson distributed.

12.4

Category	f(observed)	Midpt.	fm	fm^2
10-20	6	15	90	1,350
20-30	14	25	350	8,750
30-40	29	35	1,015	35,525
40-50	38	45	1,710	76,950
50-60	25	55	1,375	75,625
60-70	10	65	650	42,250
70-80	7	75	525	39,375

$n = \Sigma f = 129$ $\Sigma fm = 5{,}715$ $\Sigma fm^2 = 279{,}825$

$$\bar{x} = \frac{\sum fm}{\sum f} = \frac{5{,}715}{129} = 44.3$$

$$s = \sqrt{\frac{\sum fM^2 - \frac{(\sum fM)^2}{n}}{n-1}} = \sqrt{\frac{279{,}825 - \frac{(5{,}715)^2}{129}}{128}} = 14.43$$

H_o: The observed frequencies are normally distributed.

H_a: The observed frequencies are not normally distributed.

<u>For Category 10 - 20 Prob</u>

$z = \dfrac{10 - 44.3}{14.43} = -2.38$ .4913

$z = \dfrac{20 - 44.3}{14.43} = -1.68$ - .4535

Expected prob.: .0378

-

For Category 20-30 Prob

for x = 20, z = -1.68 .4535

$z = \dfrac{30 - 44.3}{14.43} = -0.99$ -.3389

Expected prob: .1146

For Category 30 - 40 Prob

for x = 30, z = -0.99 .3389

$z = \frac{40-44.3}{14.43} = -0.30$ -.1179

Expected prob: .2210

For Category 40 - 50 Prob

for x = 40, z = -0.30 .1179

$z = \frac{50-44.3}{14.43} = 0.40$ +.1554

Expected prob: .2733

For Category 50 - 60 Prob

$z = \frac{60-44.3}{14.43} = 1.09$ .3621

for x = 50, z = 0.40 -.1554

Expected prob: .2067

For Category 60 - 70 Prob

$z = \frac{70 - 44.3}{14.43} = 1.78$ .4625

for $x = 60$, $z = 1.09$ -.3621

Expected prob: .1004

For Category 70 - 80 Prob

$z = \frac{80 - 44.3}{14.43} = 2.47$ .4932

for $x = 70$, $z = 1.78$ -.4625

Expected prob: .0307

For $x < 10$:

Probability between 10 and the mean, 44.3, = (.0378 + .1145 + .2210 + .1179) = .4913. Probability < 10 = .5000 - .4912 = .0087

For $x > 80$:

Probability between 80 and the mean, 44.3, = (.0307 + .1004 + .2067 + .1554) = .4932. Probability > 80 = .5000 - .4932 = .0068

Category	Prob	expected frequency
< 10	.0087	.0087(129) = 1.12
10-20	.0378	.0378(129) = 4.88
20-30	.1146	14.78
30-40	.2210	28.51
40-50	.2733	35.26
50-60	.2067	26.66
60-70	.1004	12.95
70-80	.0307	3.96
> 80	.0068	0.88

Due to the small sizes of expected frequencies, category < 10 is folded into 10-20

and >80 into 70-80.

Category	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
10-20	6	6.00	.000
20-30	14	14.78	.041
30-40	29	28.51	.008
40-50	38	35.26	.213
50-60	25	26.66	.103
60-70	10	12.95	.672
70-80	7	4.84	.964

2.001

Calculated $\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ = **2.001**

df = k - 3 = 7 - 3 = 4, α = .05

$\chi^2_{.05,4}$ = 9.4877

Since the observed χ^2 = 2.004 < $\chi^2_{.05,4}$ = 9.4877, the decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that the observed frequencies are not normally distributed.

12.5

Definition	f_o	Exp.Prop.	f_e	$\frac{(f_o - f_e)^2}{f_e}$
Happiness	42	.39	227(.39)= 88.53	24.46
Sales/Profit	95	.12	227(.12)= 27.24	168.55
Helping Others	27	.18	40.86	4.70
Achievement/				

Challenge	63	.31	70.37	0.77
	227			198.48

H_o: The observed frequencies are distributed the same as the expected frequencies.

H_a: The observed frequencies are not distributed the same as the expected frequencies.

Observed χ^2 = **198.48**

df = k – 1 = 4 – 1 = 3, α = .05

$\chi^2_{.05,3}$ = 7.8147

Since the observed χ^2 = 198.48 > $\chi^2_{.05,3}$ = 7.8147, the decision is to **reject the null hypothesis**.

The observed frequencies for men are not distributed the same as the expected frequencies which are based on the responses of women.

12.6 Age	f_o	Prop. from survey	f_e	$\frac{(f_o - f_e)^2}{f_e}$
10-14	22	.09	(.09)(212)=19.08	0.45
15-19	50	.23	(.23)(212)=48.76	0.03
20-24	43	.22	46.64	0.28
25-29	29	.14	29.68	0.02
30-34	19	.10	21.20	0.23
$\geq$ 35	49	.22	46.64	0.12
	212			1.13

H_o: The distribution of observed frequencies is the same as the distribution of

expected frequencies.

H_a: The distribution of observed frequencies is not the same as the distribution of

expected frequencies.

$\alpha = .01$, df = k - 1 = 6 - 1 = 5

$\chi^2_{.01,5} = 15.0863$

The observed χ^2 = **1.13**

Since the observed $\chi^2 = 1.13 < \chi^2_{.01,5} = 15.0863$, the decision is to **fail to reject**

the null hypothesis.

There is not enough evidence to declare that the distribution of observed frequencies is different from the distribution of expected frequencies.

12.7

Age	f_o	m	fm	fm^2
10-20	16	15	240	3,600
20-30	44	25	1,100	27,500
30-40	61	35	2,135	74,725
40-50	56	45	2,520	113,400
50-60	35	55	1,925	105,875
60-70	19	65	1,235	80,275
	231		Σfm = 9,155	Σfm^2 = 405,375

$$\bar{x} = \frac{\sum fM}{n} = \frac{9{,}155}{231} = 39.63$$

$$s = \sqrt{\frac{\sum fM^2 - \frac{(\sum fM)^2}{n}}{n-1}} = \sqrt{\frac{405{,}375 - \frac{(9{,}155)^2}{231}}{230}} = 13.6$$

H_o: The observed frequencies are normally distributed.

H_a: The observed frequencies are not normally distributed.

For Category 10-20	Prob
$z = \frac{10-39.63}{13.6} = -2.18$	.4854
$z = \frac{20-39.63}{13.6} = -1.44$	-.4251

Expected prob. .0603

For Category 20-30	Prob
for x = 20, z = -1.44	.4251
$z = \frac{30-39.63}{13.6} = -0.71$	-.2611
Expected prob.	.1640

For Category 30-40	Prob
for x = 30, z = -0.71	.2611
$z = \frac{40-39.63}{13.6} = 0.03$	+.0120
Expected prob.	.2731

For Category 40-50	Prob
$z = \frac{50-39.63}{13.6} = 0.76$	.2764

for $x = 40$, $z = 0.03$ <u>-.0120</u>

Expected prob. .2644

<u>For Category 50-60 Prob</u>

$$z = \frac{60 - 39.63}{13.6} = 1.50 \qquad .4332$$

for $x = 50$, $z = 0.76$ <u>-.2764</u>

Expected prob. .1568

<u>For Category 60-70 Prob</u>

$$z = \frac{70 - 39.63}{13.6} = 2.23 \qquad .4871$$

for $x = 60$, $z = 1.50$ <u>-.4332</u>

Expected prob. .0539

For < 10:

Probability between 10 and the mean = .0603 + .1640 + .2611 = .4854

Probability < 10 = .5000 - .4854 = .0146

For > 70:

Probability between 70 and the mean = .0120 + .2644 + .1568 + .0539 = .4871

Probability > 70 = .5000 - .4871 = .0129

Age	Probability		f_e
< 10	.0146	(.0146)(231) =	3.37
10-20	.0603	(.0603)(231) =	13.93
20-30	.1640		37.88
30-40	.2731		63.09
40-50	.2644		61.08
50-60	.1568		36.22
60-70	.0539		12.45
> 70	.0129		2.98

Categories < 10 and > 70 are less than 5.

Collapse the < 10 into 10-20 and > 70 into 60-70.

Age	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
10-20	16	17.30	0.10
20-30	44	37.88	0.99
30-40	61	63.09	0.07
40-50	56	61.08	0.42
50-60	35	36.22	0.04
60-70	19	15.43	0.83
			2.45

df = k - 3 = 6 - 3 = 3, α = .05

$\chi^2_{.05,3} = 7.8147$

Observed χ^2 = **2.45**

Since the observed $\chi^2 < \chi^2_{.05,3} = 7.8147$, the decision is to **fail to reject the null hypothesis**.

There is no reason to reject that the observed frequencies are normally distributed.

12.8

Number	f	$(f)\cdot$ (number)
0	18	0
1	28	28
2	47	94
3	21	63
4	16	64
5	11	55
6 or more	9	54

$\Sigma f = 150$ $\Sigma f\cdot(\text{number}) = 358$

$$\lambda = \frac{\sum f \cdot number}{\sum f} = \frac{358}{150} = 2.4$$

H_o: The observed frequencies are Poisson distributed.

H_a: The observed frequencies are not Poisson distributed.

Number	Probability		f_e
0	.0907	(.0907)(150) =	13.61
1	.2177	(.2177)(150) =	32.66
2	.2613		39.20
3	.2090		31.35
4	.1254		18.81
5	.0602		9.03
6 or more	.0358		5.36

f_o	f_e	$\frac{(f_0 - f_e)^2}{f_0}$
18	13.61	1.42
28	32.66	0.66
47	39.20	1.55
21	31.35	3.42
16	18.81	0.42
11	9.03	0.43
9	5.36	2.47
		10.37

The observed χ^2 = **10.37**

α = .01, df = k – 2 = 7 – 2 = 5, $\chi^2_{.01,5}$ = 15.0863

Since the observed χ^2 = 10.37 < $\chi^2_{.01,5}$ = 15.0863, the decision is to **fail to reject**

the null hypothesis.

There is not enough evidence to reject the claim that the observed

frequencies are Poisson distributed.

12.9 H_0: p = .28 n = 270 x = 62

H_a: $p \neq .28$

	f_o		f_e	$\frac{(f_o - f_e)^2}{f_e}$
Spend More	62	270(.28) =	75.6	2.44656
Don't Spend More	208	270(.72) =	194.4	0.95144
Total	270		270.0	3.39800

The observed value of χ^2 is **3.398**

α = .05 and $\alpha/2$ = .025 df = k - 1 = 2 - 1 = 1

$\chi^2_{.025,1}$ = 5.02389

Since the observed χ^2 = 3.398 < $\chi^2_{.025,1}$ = 5.02389, the decision is to **fail to reject the null hypothesis**.

12.10 H_0: p = .30 n = 180 x= 42

H_a: p < .30

	f_0		f_e	$\frac{(f_o - f_e)^2}{f_e}$
Provide	42	180(.30) =	54	2.6666
Don't Provide	138	180(.70) =	126	1.1429
Total	180		180	3.8095

The observed value of χ^2 is **3.8095**

α = .05 df = k - 1 = 2 - 1 = 1

$\chi^2_{.05,1}$ = 3.8415

Since the observed χ^2 = 3.8095 < $\chi^2_{.05,1}$ = 3.8415, the decision is to **fail to reject the null hypothesis**.

12.11

	Variable Two		
Variable One	203	326	529
	68	110	178
	271	436	707

H_o: Variable One is independent of Variable Two.

H_a: Variable One is not independent of Variable Two.

$$e_{11} = \frac{(529)(271)}{707} = 202.77 \qquad e_{12} = \frac{(529)(436)}{707} = 326.23$$

$$e_{21} = \frac{(271)(178)}{707} = 68.23 \qquad e_{22} = \frac{(436)(178)}{707} = 109.77$$

	Variable Two		
Variable One	(202.77) 203	(326.23) 326	529
	(68.23) 68	(109.77) 110	178

	271 436	707

$$\chi^2 = \frac{(203-202.77)^2}{202.77} + \frac{(326-326.23)^2}{326.23} + \frac{(68-6.23)^2}{68.23} + \frac{(110-109.77)^2}{109.77} =$$

$$.00 + .00 + .00 + .00 = \mathbf{0.00}$$

α = .01, df = (c-1)(r-1) = (2-1)(2-1) = 1

$\chi^2_{.01,1}$ = 6.6349

Since the observed χ^2 = 0.00 < $\chi^2_{.01,1}$ = 6.6349, the decision is to **fail to reject the null hypothesis**.

Variable One is independent of Variable Two.

12.12

	Variable Two				
Variable One	24	13	47	58	142
	93	59	187	244	583
	117	72	234	302	725

H_o: Variable One is independent of Variable Two.

H_a: Variable One is not independent of Variable Two.

$$e_{11} = \frac{(142)(117)}{725} = 22.92 \qquad e_{12} = \frac{(142)(72)}{725} = 14.10$$

$$e_{13} = \frac{(142)(234)}{725} = 45.83 \qquad e_{14} = \frac{(142)(302)}{725} = 59.15$$

$$e_{21} = \frac{(583)(117)}{725} = 94.08 \qquad e_{22} = \frac{(583)(72)}{725} = 57.90$$

$$e_{23} = \frac{(583)(234)}{725} = 188.17 \qquad e_{24} = \frac{(583)(302)}{725} = 242.85$$

	Variable Two				
Variable One	(22.92) 24	(14.10) 13	(45.83) 47	(59.15) 58	142
	(94.08) 93	(57.90) 59	(188.17) 187	(242.85) 244	583
	117	72	234	302	725

$$\chi^2 = \frac{(24-22.92)^2}{22.92} + \frac{(13-14.10)^2}{14.10} + \frac{(47-45.83)^2}{45.83} + \frac{(58-59.15)^2}{59.15} +$$

$$\frac{(93-94.08)^2}{94.08} + \frac{(59-57.90)^2}{57.90} + \frac{(188-188.17)^2}{188.17} + \frac{(244-242.85)^2}{242.85} =$$

$$.05 + .09 + .03 + .02 + .01 + .02 + .01 + .01 = \mathbf{0.24}$$

$\alpha = .01$, df = $(c-1)(r-1) = (4-1)(2-1) = 3$, $\chi^2_{.01,3} = 11.3449$

Since the observed $\chi^2 = 0.24 < \chi^2_{.01,3} = 11.3449$, the decision is to **fail to reject the null hypothesis**.

Variable One is independent of Variable Two.

12.13

		Social Class			
		Lower Upper		Middle	
Number of Children	0	7	18	6	31
	1	9	38	23	70
	2 or 3	34	97	58	189
	>3	47	31	30	108
		97 117		184	398

H_o: Social Class is independent of Number of Children.

H_a: Social Class is not independent of Number of Children.

$e_{11} = \dfrac{(31)(97)}{398} = 7.56$ $\quad$ $e_{31} = \dfrac{(189)(97)}{398} = 46.06$

$e_{12} = \dfrac{(31)(184)}{398} = 14.3$ $\quad$ $e_{32} = \dfrac{(189)(184)}{398} = 87.38$

$e_{13} = \dfrac{(31)(117)}{398} = 9.11$ $\quad$ $e_{33} = \dfrac{(189)(117)}{398} = 55.56$

$$e_{21} = \frac{(70)(97)}{398} = 17.06 \qquad e_{41} = \frac{(108)(97)}{398} = 26.32$$

$$e_{22} = \frac{(70)(184)}{398} = 32.36 \qquad e_{42} = \frac{(108)(184)}{398} = 49.93$$

$$e_{23} = \frac{(70)(117)}{398} = 20.58 \qquad e_{43} = \frac{(108)(117)}{398} = 31.75$$

		Social Class			
Number of Children		Lower Upper		Middle	
	0	(7.56) 7	(14.33) 18	(9.11) 6	31
	1	(17.06) 9	(32.36) 38	(20.58) 23	70
	2 or 3	(46.06) 34	(87.38) 97	(55.56) 58	189
	>3	(26.32) 47	(49.93) 31	(31.75) 30	108
		97 117		184	398

$$\chi^2 = \frac{(7-7.56)^2}{7.56} + \frac{(18-14.33)^2}{14.33} + \frac{(6-9.11)^2}{9.11} + \frac{(9-17.06)^2}{17.06} +$$

$$\frac{(38-32.36)^2}{32.36} + \frac{(23-20.58)^2}{20.58} + \frac{(34-46.06)^2}{46.06} + \frac{(97-87.38)^2}{87.38} +$$

$$\frac{(58-55.56)^2}{55.56} + \frac{(47-26.32)^2}{26.32} + \frac{(31-49.93)^2}{49.93} + \frac{(30-31.75)^2}{31.75} =$$

.04 + .94 + 1.06 + 3.81 + .98 + .28 + 3.16 + 1.06 + .11 + 16.25 +

7.18 + .10 = **34.97**

α = .05, df = (c-1)(r-1) = (3-1)(4-1) = 6

$\chi^2_{.05,6}$ = 12.5916

Since the observed χ^2 = 34.97 > $\chi^2_{.05,6}$ = 12.5916, the decision is to **reject the null hypothesis**.

Number of children is not independent of social class.

12.14

	Type of Music Preferred					
		Rock	R&B	Coun	Clssic	
Region	NE	140	32	5	18	195
	S	134	41	52	8	235
	W	154	27	8	13	202
						632
		428	100	65	39	

H_o: Type of music preferred is independent of region.

H_a: Type of music preferred is not independent of region.

$$e_{11} = \frac{(195)(428)}{632} = 132.6 \qquad e_{23} = \frac{(235)(65)}{632} = 24.17$$

$$e_{12} = \frac{(195)(100)}{632} = 30.85 \qquad e_{24} = \frac{(235)(39)}{632} = 14.50$$

$$e_{13} = \frac{(195)(65)}{632} = 20.06 \qquad e_{31} = \frac{(202)(428)}{632} = 136.80$$

$$e_{14} = \frac{(195)(39)}{632} = 12.03 \qquad e_{32} = \frac{(202)(100)}{632} = 31.96$$

$$e_{21} = \frac{(235)(428)}{632} = 159.15 \qquad e_{33} = \frac{(202)(65)}{632} = 20.78$$

$$e_{22} = \frac{(235)(100)}{632} = 37.18 \qquad e_{34} = \frac{(202)(39)}{632} = 12.47$$

	Type of Music Preferred					
		Rock	R&B	Coun	Clssic	
Region	NE	(132.06) 140	(30.85) 32	(20.06) 5	(12.03) 18	195
	S	(159.15) 134	(37.18) 41	(24.17) 52	(14.50) 8	235
	W	(136.80) 154	(31.96) 27	(20.78) 8	(12.47) 13	202 632

	39	428	100	65	

$$\chi^2 = \frac{(141-132.06)^2}{132.06} + \frac{(32-30.85)^2}{30.85} + \frac{(5-20.06)^2}{20.06} + \frac{(18-12.03)^2}{12.03} +$$

$$\frac{(134-159.15)^2}{159.15} + \frac{(41-37.18)^2}{37.18} + \frac{(52-24.17)^2}{24.17} + \frac{(8-14.50)^2}{14.50} +$$

$$\frac{(154-136.80)^2}{136.80} + \frac{(27-31.96)^2}{31.96} + \frac{(8-20.78)^2}{20.78} + \frac{(13-12.47)^2}{12.47} =$$

.48 + .04 + 11.31 + 2.96 + 3.97 + .39 + 32.04 + 2.91 + 2.16 + .77 +

7.86 + .02 = **64.91**

α = .01, df = (c-1)(r-1) = (4-1)(3-1) = 6

$\chi^2_{.01,6}$ = 16.8119

Since the observed χ^2 = 64.91 > $\chi^2_{.01,6}$ = 16.8119, the decision is to **reject the null hypothesis**.

Type of music preferred is not independent of region of the country.

12.15

	Transportation Mode				
Industry		Air	Train	Truck	
	Publishing	32	12	41	85
	Comp.Hard.	5	6	24	35
		37	18	65	120

H_0: Transportation Mode is independent of Industry.

H_a: Transportation Mode is not independent of Industry.

$e_{11} = \dfrac{(85)(37)}{120} = 26.21$ $\qquad$ $e_{21} = \dfrac{(35)(37)}{120} = 10.79$

$e_{12} = \dfrac{(85)(18)}{120} = 12.75$ $\qquad$ $e_{22} = \dfrac{(35)(18)}{120} = 5.25$

$e_{13} = \dfrac{(85)(65)}{120} = 46.04$ $\qquad$ $e_{23} = \dfrac{(35)(65)}{120} = 18.96$

	Transportation Mode				
Industry		Air	Train	Truck	
	Publishing	(26.21) 32	(12.75) 12	(46.04) 41	85

	Comp.Hard.	(10.79) 5	(5.25) 6	(18.96) 24	35
		37	18	65	120

$\chi^2 = \frac{(32-26.21)^2}{26.21} + \frac{(12-12.75)^2}{12.75} + \frac{(41-46.04)^2}{46.04} +$

$\frac{(5-10.79)^2}{10.79} + \frac{(6-5.25)^2}{5.25} + \frac{(24-18.96)^2}{18.96} =$

1.28 + .04 + .55 + 3.11 + .11 + 1.34 = **6.43**

α = .05, df = (c-1)(r-1) = (3-1)(2-1) = 2

$\chi^2_{.05,2}$ = 5.9915

Since the observed χ^2 = 6.43 > $\chi^2_{.05,2}$ = 5.9915, the decision is to **reject the null hypothesis**.

Transportation mode is not independent of industry.

12.16

	Number of Bedrooms				
		≤ 2	3	≥ 4	
Number of Stories	1	116	101	57	274
	2	90	325	160	575
		206	426	217	849

H_0: Number of Stories is independent of number of bedrooms.

H_a: Number of Stories is not independent of number of bedrooms.

$e_{11} = \dfrac{(274)(206)}{849} = 66.48$ $\qquad e_{21} = \dfrac{(575)(206)}{849} = 139.52$

$e_{12} = \dfrac{(274)(426)}{849} = 137.48$ $\qquad e_{22} = \dfrac{(575)(426)}{849} = 288.52$

$e_{13} = \dfrac{(274)(217)}{849} = 70.03$ $\qquad e_{23} = \dfrac{(575)(217)}{849} = 146.97$

$$\chi^2 = \frac{(90-139.52)^2}{139.52} + \frac{(101-137.48)^2}{137.48} + \frac{(57-70.03)^2}{70.03} + \frac{(90-139.52)^2}{139.52}$$

$+$

$$\frac{(325-288.52)^2}{288.52} + \frac{(160-146.97)^2}{146.97} =$$

χ^2 = 36.89 + 9.68 + 2.42 + 17.58 + 4.61 + 1.16 = **72.34**

α = .10 df = (c-1)(r-1) = (3-1)(2-1) = 2

$\chi^2_{.10,2}$ = 4.6052

Since the observed χ^2 = 72.34 > $\chi^2_{.10,2}$ = 4.6052, the decision is to **reject the null hypothesis**.

Number of stories is not independent of number of bedrooms.

12.17

	Mexican Citizens			
		Yes	No	
Type of Store	Dept.	24	17	41
	Disc.	20	15	35
	Hard.	11	19	30
	Shoe	32	28	60
		87	79	166

H_o: Citizenship is independent of store type

H_a: Citizenship is not independent of store type

$$e_{11} = \frac{(41)(87)}{166} = 21.49 \qquad e_{31} = \frac{(30)(87)}{166} = 15.72$$

$$e_{12} = \frac{(41)(79)}{166} = 19.51 \qquad e_{32} = \frac{(30)(79)}{166} = 14.28$$

$$e_{21} = \frac{(35)(87)}{166} = 18.34 \qquad e_{41} = \frac{(60)(87)}{166} = 31.45$$

$$e_{22} = \frac{(35)(79)}{166} = 16.66 \qquad e_{42} = \frac{(60)(79)}{166} = 28.55$$

	Mexican Citizens			
		Yes	No	
Type of Store	Dept.	(21.49) 24	(19.51) 17	41
	Disc.	(18.34) 20	(16.66) 15	35
	Hard.	(15.72) 11	(14.28) 19	30
	Shoe	(31.45) 32	(28.55) 28	60
		87	79	166

$$\chi^2 = \frac{(24-21.49)^2}{21.49} + \frac{(17-19.51)^2}{19.51} + \frac{(20-18.34)^2}{18.34} + \frac{(15-16.66)^2}{16.66} +$$

$$\frac{(11-15.72)^2}{15.72} + \frac{(19-14.28)^2}{14.28} + \frac{(32-31.45)^2}{31.45} + \frac{(28-28.55)^2}{28.55} =$$

.29 + .32 + .15 + .17 + 1.42 + 1.56 + .01 + .01 = **3.93**

α = .05, df = (c-1)(r-1) = (2-1)(4-1) = 3

$\chi^2_{.05,3}$ = 7.8147

Since the observed χ^2 = 3.93 < $\chi^2_{.05,3}$ = 7.8147, the decision is to **fail to**

reject the null hypothesis.

Citizenship is independent of type of store.

12.18 α = .01, k = 7, df = 6

H_0: The observed distribution is the same as the expected distribution

H_a: The observed distribution is not the same as the expected distribution

Use:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

critical $\chi^2_{.01,6}$ = 16.8119

f_o	f_e	$(f_0-f_e)^2$	$\frac{(f_o-f_e)^2}{f_e}$
214	206	64	0.311
235	232	9	0.039
279	268	121	0.451
281	284	9	0.032
264	268	16	0.060
254	232	484	2.086
211	206	25	0.121
			3.100

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

= **3.100**

Since the observed value of $\chi^2 = 3.1 < \chi^2_{.01,6} = 16.8119$, the decision is to **fail to reject the null hypothesis**. The observed distribution is not different from the expected distribution.

12.19

	Variable 2			
	12	23	21	56

Variable 1	8	17	20	45
	7	11	18	36
	27	51	59	137

e_{11} = 11.04 e_{12} = 20.85 e_{13} = 24.12

e_{21} = 8.87 e_{22} = 16.75 e_{23} = 19.38

e_{31} = 7.09 e_{32} = 13.40 e_{33} = 15.50

$$\chi^2 = \frac{(12-11.04)^2}{11.04} + \frac{(23-20.85)^2}{20.85} + \frac{(21-24.12)^2}{24.12} + \frac{(8-8.87)^2}{8.87} +$$

$$\frac{(17-16.75)^2}{16.75} + \frac{(20-19.38)^2}{19.38} + \frac{(7-7.09)^2}{7.09} + \frac{(11-13.40)^2}{13.40} +$$

$$\frac{(18-15.50)^2}{15.50} =$$

.084 + .222 + .403 + .085 + .004 + .020 + .001 + .430 + .402 = **1.652**

df = (c-1)(r-1) = (2)(2) = 4 α = .05

$\chi^2_{.05,4}$ = 9.4877

Since the observed value of $\chi^2 = 1.652 < \chi^2_{.05,4} = 9.4877$, the decision is to **fail**

to reject the null hypothesis.

12.20

		Location			
		NE	W	S	
Customer	Industrial	230	115	68	413
	Retail	185	143	89	417
		415	258	157	830

$e_{11} = \frac{(413)(415)}{830} = 206.5$ $e_{21} = \frac{(417)(415)}{830} = 208.5$

$e_{12} = \frac{(413)(258)}{830} = 128.38$ $e_{22} = \frac{(417)(258)}{830} = 129.62$

$e_{13} = \frac{(413)(157)}{830} = 78.12$ $e_{23} = \frac{(417)(157)}{830} = 78.88$

		Location			
		NE	W	S	
Customer	Industrial	(206.5) 230	(128.38) 115	(78.12) 68	413

	Retail	(208.5) 185	(129.62) 143	(78.88) 89	417
		415	258	157	830

$$\chi^2 = \frac{(230-206.5)^2}{206.5} + \frac{(115-128.38)^2}{128.38} + \frac{(68-78.12)^2}{78.12} +$$

$$\frac{(185-208.5)^2}{208.5} + \frac{(143-129.62)^2}{129.62} + \frac{(89-78.88)^2}{78.88} =$$

2.67 + 1.39 + 1.31 + 2.65 + 1.38 + 1.30 = **10.70**

α = .10 and df = (c - 1)(r - 1) = (3 - 1)(2 - 1) = 2

$\chi^2_{.10,2}$ = 4.6052

Since the observed χ^2 = 10.70 > $\chi^2_{.10,2}$ = 4.6052, the decision is to **reject the null hypothesis**.

Type of customer is not independent of geographic region.

12.21 <u>Cookie Type f_o</u>

Chocolate Chip 189

Peanut Butter	168
Cheese Cracker	155
Lemon Flavored	161
Chocolate Mint	216
Vanilla Filled	165
	$\Sigma f_o = 1{,}054$

H_o: Cookie Sales is uniformly distributed across kind of cookie.

H_a: Cookie Sales is not uniformly distributed across kind of cookie.

If cookie sales are uniformly distributed, then $f_e = \dfrac{\sum f_0}{no.kinds} = \dfrac{1{,}054}{6} = 175.67$

f_o	f_e	$\dfrac{(f_o - f_e)^2}{f_e}$
189	175.67	1.01
168	175.67	0.33
155	175.67	2.43
161	175.67	1.23
216	175.67	9.26
165	175.67	0.65
		14.91

The observed χ^2 = **14.91**

$\alpha = .05$ df = k - 1 = 6 - 1 = 5

$\chi^2_{.05,5} = 11.0705$

Since the observed $\chi^2 = 14.91 > \chi^2_{.05,5} = 11.0705$, the decision is to **reject the**

null hypothesis.

Cookie Sales is not uniformly distributed by kind of cookie.

12.22

		Gender		
		M	F	
Bought Car	Y	207	65	272
	N	811	984	1,795
		1,018	1,049	2,067

H_o: Purchasing a car or not is independent of gender.

H_a: Purchasing a car or not is <u>not</u> independent of gender.

$$e_{11} = \frac{(272)(1,018)}{2,067} = 133.96 \qquad e_{12} = \frac{(27)(1,049)}{2,067} = 138.04$$

$$e_{21} = \frac{(1,795)(1,018)}{2,067} = 884.04 \qquad e_{22} = \frac{(1,795)(1,049)}{2,067} = 910.96$$

		Gender		
		M	F	
Bought Car	Y	(133.96) 207	(138.04) 65	272
	N	(884.04) 811	(910.96) 984	1,795

	1,018	1,049	2,067

$$\chi^2 = \frac{(207-133.96)^2}{133.96} + \frac{(65-138.04)^2}{138.04} + \frac{(811-884.04)^2}{884.04} +$$

$$\frac{(984-910.96)^2}{910.96} = 39.82 + 38.65 + 6.03 + 5.86 = \mathbf{90.36}$$

$\alpha = .05$ df = (c-1)(r-1) = (2-1)(2-1) = 1

$\chi^2_{.05,1} = 3.8415$

Since the observed $\chi^2 = 90.36 > \chi^2_{.05,1} = 3.8415$, the decision is to **reject the null hypothesis**.

Purchasing a car is not independent of gender.

12.23

Arrivals	f_o	(f_o)(Arrivals)
0	26	0
1	40	40
2	57	114
3	32	96
4	17	68
5	12	60
6	8	48
	$\Sigma f_o = 192$	$\Sigma(f_o)$(arrivals) = 426

$$\lambda = \frac{\sum (f_0)(arrivals)}{\sum f_0} = \frac{426}{192} = 2.2$$

H_o: The observed frequencies are Poisson distributed.

H_a: The observed frequencies are not Poisson distributed.

Arrivals	Probability	f_e
0	.1108	(.1108)(192) = 21.27
1	.2438	(.2438)(192) = 46.81
2	.2681	51.48
3	.1966	37.75
4	.1082	20.77
5	.0476	9.14
6	.0249	4.78

f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
26	21.27	1.05
40	46.81	0.99
57	51.48	0.59
32	37.75	0.88
17	20.77	0.68
12	9.14	0.89
8	4.78	2.17
		7.25

Observed χ^2 = **7.25**

α = .05 df = k - 2 = 7 - 2 = 5

$\chi^2_{.05,5}$ = 11.0705

Since the observed χ^2 = 7.25 < $\chi^2_{.05,5}$ = 11.0705, the decision is to **fail to reject**

the null hypothesis. There is not enough evidence to reject the claim that the

observed frequency of arrivals is Poisson distributed.

12.24 H_o: The distribution of observed frequencies is the same as the distribution of expected frequencies.

H_a: The distribution of observed frequencies is not the same as the distribution of expected frequencies.

Soft Drink	f_o	proportions		f_e	$\frac{(f_o - f_e)^2}{f_e}$
Classic Coke	314	.179	(.179)(1726) =	308.95	0.08
Pepsi	219	.115	(.115)(1726) =	198.49	2.12
Diet Coke	212	.097		167.42	11.87
Mt. Dew	121	.063		108.74	1.38
Diet Pepsi	98	.061		105.29	0.50
Sprite	93	.057		98.32	0.29
Dr. Pepper	88	.056		96.66	0.78
Others	581	.372		642.07	5.81
	$\sum f_o$ = 1,726				22.83

Observed χ^2 = **22.83**

α = .05 df = k - 1 = 8 - 1 = 7

$\chi^2_{.05,7}$ = 14.0671

Since the observed χ^2 = 22.83 > $\chi^2_{.05,6}$ = 14.0671, the decision is to **reject the**

null hypothesis.

The observed frequencies are not distributed the same as the expected frequencies

from the national poll.

12.25

		Position				
		Manager	Programmer	Operator	Systems Analyst	
Years	0-3	6	37	11	13	67
	4-8	28	16	23	24	91
	> 8	47	10	12	19	88
		81	63	46	56	246

$e_{11} = \frac{(67)(81)}{246} = 22.06$

$e_{23} = \frac{(91)(46)}{246} = 17.02$

$e_{12} = \frac{(67)(63)}{246} = 17.16$

$e_{24} = \frac{(91)(56)}{246} = 20.72$

$e_{13} = \frac{(67)(46)}{246} = 12.53$

$e_{31} = \frac{(88)(81)}{246} = 28.98$

$e_{14} = \frac{(67)(56)}{246} = 15.25$

$e_{32} = \frac{(88)(63)}{246} = 22.54$

$$e_{21} = \frac{(91)(81)}{246} = 29.96$$

$$e_{33} = \frac{(88)(46)}{246} = 16.46$$

$$e_{22} = \frac{(91)(63)}{246} = 23.30$$

$$e_{34} = \frac{(88)(56)}{246} = 20.03$$

		Position				
		Manager	Programmer	Operator	Systems Analyst	
Years	0-3	(22.06) 6	(17.16) 37	(12.53) 11	(15.25) 13	67
	4-8	(29.96) 28	(23.30) 16	(17.02) 23	(20.72) 24	91
	> 8	(28.98) 47	(22.54) 10	(16.46) 12	(20.03) 19	88
		81	63	46	56	246

$$\chi^2 = \frac{(6-22.06)^2}{22.06} + \frac{(37-17.16)^2}{17.16} + \frac{(11-12.53)^2}{12.53} + \frac{(13-15.25)^2}{15.25} +$$

$$\frac{(28-29.96)^2}{29.96} + \frac{(16-23.30)^2}{23.30} + \frac{(23-17.02)^2}{17.02} + \frac{(24-20.72)^2}{20.72} +$$

$$\frac{(47-28.98)^2}{28.98} + \frac{(10-22.54)^2}{22.54} + \frac{(12-16.46)^2}{16.46} + \frac{(19-20.03)^2}{20.03} =$$

11.69 + 22.94 + .19 + .33 + .13 + 2.29 + 2.1 + .52 + 11.2 + 6.98 +

1.21 + .05 = **59.63**

α = .01 df = (c-1)(r-1) = (4-1)(3-1) = 6

$\chi^2_{.01,6}$ = 16.8119

Since the observed χ^2 = 59.63 > $\chi^2_{.01,6}$ = 16.8119, the decision is to **reject the null hypothesis**. Position is not independent of number of years of experience.

12.26 H_0: $p = .43$ $n = 315$ $\alpha = .05$

H_a: $p \neq .43$ $x = 120$ $\alpha/2 = .025$

	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
More Work, More Business	120	(.43)(315) = 135.45	1.76
Others	195	(.57)(315) = 179.55	1.33
Total	315	315.00	3.09

The observed value of χ^2 is **3.09**

$\alpha = .05$ and $\alpha/2 = .025$ df = k - 1 = 2 - 1 = 1

$\chi^2_{.025,1} = 5.0239$

Since $\chi^2 = 3.09 < \chi^2_{.025,1} = 5.0239$, the decision is to **fail to reject the null hypothesis**.

12.27

		Type of College or University			
		Community College	Large University	Small College	
Number of Children	0	25	178	31	234
	1	49	141	12	202
	2	31	54	8	93
	≥ 3	22	14	6	42
		127	387	57	571

H_o: Number of Children is independent of Type of College or University.

H_a: Number of Children is not independent of Type of College or University.

$$e_{11} = \frac{(234)(127)}{571} = 52.05 \qquad e_{31} = \frac{(93)(127)}{571} = 20.68$$

$$e_{12} = \frac{(234)(387)}{571} = 158.60 \qquad e_{32} = \frac{(193)(387)}{571} = 63.03$$

$$e_{13} = \frac{(234)(57)}{571} = 23.36 \qquad e_{33} = \frac{(93)(57)}{571} = 9.28$$

$$e_{21} = \frac{(202)(127)}{571} = 44.93 \qquad e_{41} = \frac{(42)(127)}{571} = 9.34$$

$$e_{22} = \frac{(202)(387)}{571} = 136.91 \qquad e_{42} = \frac{(42)(387)}{571} = 28.47$$

$$e_{23} = \frac{(202)(57)}{571} = 20.16 \qquad e_{43} = \frac{(42)(57)}{571} = 4.19$$

		Type of College or University			
		Community College	Large University	Small College	
Number of Children	0	(52.05) 25	(158.60) 178	(23.36) 31	234
	1	(44.93) 49	(136.91) 141	(20.16) 12	202
	2	(20.68) 31	(63.03) 54	(9.28) 8	93
	≥3	(9.34) 22	(28.47) 14	(4.19) 6	42
		127	387	57	571

$$\chi^2 = \frac{(25-52.05)^2}{52.05} + \frac{(178-158.6)^2}{158.6} + \frac{(31-23.36)^2}{23.36} + \frac{(49-44.93)^2}{44.93} +$$

$$\frac{(141-136.91)^2}{136.91} + \frac{(12-20.16)^2}{20.16} + \frac{(31-20.68)^2}{20.68} + \frac{(54-63.03)^2}{63.03} +$$

$$\frac{(8-9.28)^2}{9.28} + \frac{(22-9.34)^2}{9.34} + \frac{(14-28.47)^2}{28.47} + \frac{(6-4.19)^2}{4.19} =$$

14.06 + 2.37 + 2.50 + 0.37 + 0.12 + 3.30 + 5.15 + 1.29 + 0.18 +

17.16 + 7.35 + 0.78 = **54.63**

α = .05, df= (c - 1)(r - 1) = (3 - 1)(4 - 1) = 6

$\chi^2_{.05,6}$ = 12.5916

Since the observed χ^2 = 54.63 > $\chi^2_{.05,6}$ = 12.5916, the decision is to **reject the null hypothesis**.

Number of children is not independent of type of College or University.

12.28 The observed chi-square is 30.18 with a *p*-value of .0000043. The chi-square
goodness-of-fit test indicates that there is a significant difference between the observed frequencies and the expected frequencies. The distribution of responses to the question is not the same for adults between 21 and 30 years of age as they are for others. Marketing and sales people might reorient their 21 to 30 year old efforts away from home improvement and pay more attention to leisure travel/vacation, clothing, and home entertainment.

12.29 The observed chi-square value for this test of independence is 5.366. The associated *p*-value of .252 indicates failure to reject the null hypothesis. There is not enough evidence here to say that color choice is dependent upon gender. Automobile marketing people do not have to worry about which colors especially appeal to men or to women because car color is independent of gender. In addition, design and production people can determine car color quotas based on other variables.

Chapter 13

Nonparametric Statistics

LEARNING OBJECTIVES

This chapter presents several nonparametric statistics that can be used to analyze data enabling you to:

1. Recognize the advantages and disadvantages of nonparametric statistics.
2. Understand how to use the runs test to test for randomness.
3. Know when and how to use the Mann-Whitney *U* Test, the Wilcoxon matched-pairs signed rank test, the Kruskal-Wallis test, and the Friedman test.
4. Learn when and how to measure correlation using Spearman's rank correlation measurement.

CHAPTER TEACHING STRATEGY

Chapter 13 contains new six techniques for analysis. Only the first technique, the runs test, is conceptually a different idea for the student to consider than anything presented in the text to this point. The runs test is a mechanism for testing to determine if a string of data are random. There is a runs test for small samples that uses Table A.12 in the appendix and a test for large samples, which utilizes a z test.

The main portion of chapter 13 (middle part) contains nonparametric alternatives to parametric tests presented earlier in the book. The Mann-Whitney U test is a nonparametric alternative to the t test for independent means. The Wilcoxon matched-pairs signed ranks test is an alternative to the t test for matched-pairs. The Kruskal-Wallis is a nonparametric alternative to the one-way analysis of variance test. The Friedman test is a nonparametric alternative to the randomized block design presented in chapter 11. Each of these four tests utilizes rank analysis.

The last part of the chapter is a section on Spearman's rank correlation. This correlation coefficient can be presented as a nonparametric alternative to the Pearson product-moment correlation coefficient of chapter 3. Spearman's rank correlation uses either ranked data or data that is converted to ranks. The interpretation of Spearman's rank correlation is similar to Pearson's product-moment correlation coefficient.

CHAPTER OUTLINE

13.1 Runs Test

Small-Sample Runs Test
Large-Sample Runs Test

13.2 Mann-Whitney *U* Test

Small-Sample Case
Large-Sample Case

13.3 Wilcoxon Matched-Pairs Signed Rank Test

Small-Sample Case ($n \leq 15$)

Large-Sample Case ($n > 15$)

13.4 Kruskal-Wallis Test

13.5 Friedman Test

13.6 Spearman's Rank Correlation

KEY TERMS

Friedman Test	Parametric Statistics
Kruskal-Wallis Test	Runs Test
Mann-Whitney *U* Test	Spearman's Rank Correlation

Nonparametric Statistics
Wilcoxon Matched-Pairs Signed Rank Test

SOLUTIONS TO CHAPTER 13

13.1 H_o: The observations in the sample are randomly generated.

H_a: The observations in the sample are not randomly generated.

This is a small sample runs test since n_1, $n_2 \le 20$

α = .05, The lower tail critical value is 6 and the upper tail critical value is 16

$n_1 = 10 \quad n_2 = 10$

***R* = 11**

Since R = 11 is between the two critical values, the decision is to **fail to reject the null hypothesis.**

The data are random.

13.2 H_o: The observations in the sample are randomly generated.

H_a: The observations in the sample are not randomly generated.

α = .05, $\alpha/2$ = .025, $z_{.025} = \pm 1.96$

n_1 = 26 n_2 = 21 n = 47

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2(26)(21)}{26+21}+1$$

= 24.234

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(26)(21)[2(26)(21)-26-21]}{(26+21)^2(26+21-1)}}$$

= 3.351

***R* = 9**

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{9-24.234}{3.351}$$

= **-4.55**

Since the observed value of z = -4.55 < $z_{.025}$ = -1.96, the decision is to **reject the**

null hypothesis. The data are not randomly generated.

13.3 $n_1 = 8$ $n_2 = 52$ $\alpha = .05$

This is a two-tailed test and $\alpha/2 = .025$. **The *p*-value from the printout is .0264**.

Since the *p*-value is the lowest value of "alpha" for which the null hypothesis can

be rejected, the decision is to **fail to reject the null hypothesis**

(*p*-value = .0264 > .025). There is not enough evidence to reject that the data are randomly generated.

13.4 The observed number of runs is 18. The mean or expected number of runs

is 14.333. **The *p*-value for this test is .1452**. Thus, the test is not significant at alpha of .05 or .025 for a two-tailed test. The decision is to **fail to reject the**

null hypothesis. There is not enough evidence to declare that the data are not random. Therefore, we must conclude that the data a randomly generated.

13.5 H_o: The observations in the sample are randomly generated.

H_a: The observations in the sample are not randomly generated.

Since n_1, n_2 > 20, use large sample runs test

α = .05 Since this is a two-tailed test, $\alpha/2$ = .025, $z_{.025} = \pm 1.96$. If the

observed value of z is greater than 1.96 or less than -1.96, the decision is to reject the null hypothesis.

***R* = 27** n_1 = 40 n_2 = 24

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2(40)(24)}{64}+1$$

= 31

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(40)(24)[2(40)(24)-40-24]}{(64)^2(63)}}$$

= 3.716

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{27-31}{3.716}$$

= **-1.08**

Since the observed z of -1.08 is greater than the critical lower tail z value

of -1.96, the decision is to **fail to reject the null hypothesis**. The data are randomly generated.

13.6 H_o: The observations in the sample are randomly generated.

H_a: The observations in the sample are not randomly generated.

$n_1 = 5$ $n_2 = 8$ $n = 13$ $\alpha = .05$

Since this is a two-tailed test, $\alpha/2 = .025$

From Table A.11, the lower critical value is 3

From Table A.11, the upper critical value is 11

***R* = 4**

Since $R = 4 >$ than the lower critical value of 3 and less than the upper critical

value of 11, the decision is to **fail to reject the null hypothesis**. The data are randomly generated.

13.7 H_o: Group 1 is identical to Group 2

H_a: Group 1 is not identical to Group 2

Use the small sample Mann-Whitney *U* test since both $n_1, n_2 \leq 10$, $\alpha = .05$. Since this is a two-tailed test, $\alpha/2 = .025$. The *p*-value is obtained using Table A.13.

Value	Rank	Group
11	1	1

13	2.5		1
13	2.5	2	
14	4		2
15	5		1
17	6	1	
18	7.5		1
18	7.5		2
21	9.5		1
21	9.5		2
22	11		1
23	12.5		2
23	12.5		2
24	14		2
26	15		1
29	16		1

$n_1 = 8$

$n_2 = 8$

$W_1 = 1 + 2.5 + 5 + 6 + 7.5 + 9.5 + 15 + 16 = 62.5$

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (8)(8) + \frac{(8)(9)}{2} - 62.5$$

$$= 37.5$$

$$U' = n_1 \cdot n_2 - U$$

$$= 64 - 37.5 = 26.5$$

We use the small *U* which is 26.5

From Table A.13, the *p*-value for *U* = 27 is .3227(2) = **.6454**

Since this *p*-value is greater than $\alpha/2$ = .025, the decision is to **fail to reject the**

null hypothesis.

13.8 H_o: Population 1 has values that are no greater than population 2

H_a: Population 1 has values that are greater than population 2

Value	Rank	Group
203	1	2
208	2	2
209	3	2
211	4	2
214	5	2
216	6	1
217	7	1
218	8	2
219	9	2
222	10	1
223	11	2
224	12	1

227	13	2
229	14	2
230	15.5	2
230	15.5	2
231	17	1
236	18	2
240	19	1
241	20	1
248	21	1
255	22	1
256	23	1
283	24	1

$n_1 = 11$

$n_2 = 13$

$W_1 = 6 + 7 + 10 + 12 + 17 + 19 + 20 + 21 + 22 + 23 + 24 =$

$W_1 = 181$

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(11)(13)}{2}$$

$= 71.5$

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(11)(13)(25)}{12}}$$

$= 17.26$

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (11)(13 + \frac{(11)(12)}{2} - 181$$

= 28

$$z = \frac{U-\mu}{\sigma} = \frac{28-71.5}{17.26}$$

= **-2.52**

α = .01, $z_{.01}$ = 2.33

Since $z = \left|-2.52\right|$ = 2.52 > z = 2.33, the decision is to **reject the null hypothesis**.

13.9

Contacts	Rank	Group
6	1	1
8	2	1
9	3.5	1
9	3.5	2
10	5	2
11	6.5	1
11	6.5	1

12	8.5	1
12	8.5	2
13	11	1
13	11	2
13	11	2
14	13	2
15	14	2
16	15	2
17	16	2

W_1 = 39

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (7)(9) + \frac{(7)(8)}{2} - 39$$

= 52

$$U_2 = n_1 \cdot n_2 - U_1$$

= (7)(9) – 52 = 11

U = 11

From Table A.13, the p-value = **.0156**. Since this p-value is greater than α = .01,

the decision is to **fail to reject the null hypothesis**.

13.10 H_o: Urban and rural spend the same

H_a: Urban and rural spend different amounts

Expenditure	Rank	Group
1950	1	U
2050	2	R
2075	3	R
2110	4	U
2175	5	U
2200	6	U
2480	7	U
2490	8	R
2540	9	U
2585	10	R
2630	11	U
2655	12	U
2685	13	R
2710	14	U
2750	15	U
2770	16	R
2790	17	R
2800	18	R
2850	19.5	U
2850	19.5	U
2975	21	R
2995	22.5	R

2995	22.5	R
3100	24	R

$n_1 = 12$

$n_2 = 12$

$W_1 = 1 + 4 + 5 + 6 + 7 + 9 + 11 + 12 + 14 + 15 + 19.5 + 19.5 = 123$

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(12)(12)}{2} = 72$$

$$\sigma = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(12)(12)(25)}{12}} = 17.32$$

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (12)(12) + \frac{(12)(13)}{2} - 123 = 99$$

$$z = \frac{U - \mu}{\sigma} = \frac{99 - 72}{17.32} = \mathbf{1.56}$$

$\alpha = .05$ $\quad\quad$ $\alpha/2 = .025$

$z_{.025} = \pm 1.96$

Since the observed $z = 1.56 < z_{.025} = 1.96$, the decision is to **fail to reject the null hypothesis**.

13.11 H_o: Males do not earn more than females

H_a: Males do earn more than females

Earnings	Rank	Gender
$28,900	1	F
31,400	2	F
36,600	3	F
40,000	4	F
40,500	5	F
41,200	6	F
42,300	7	F
42,500	8	F
44,500	9	F
45,000	10	M
47,500	11	F
47,800	12.5	F
47,800	12.5	M
48,000	14	F
50,100	15	M
51,000	16	M
51,500	17.5	M
51,500	17.5	M
53,850	19	M
55,000	20	M
57,800	21	M

61,100	22	M
63,900	23	M

$n_1 = 11$ $n_2 = 12$

$W_1 = 10 + 12.5 + 15 + 16 + 17.5 + 17.5 + 19 + 20 + 21 + 22 + 23 = 193.5$

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(11)(12)}{2}$$ = 66 and $$\sigma = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(11)(12)(24)}{12}}$$ = 16.25

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (11)(12) + \frac{(11)(12)}{2} - 193.5$$ = 4.5

$$z = \frac{U - \mu}{\sigma} = \frac{4.5 - 66}{16.25}$$ = **-3.78**

$\alpha = .01$, $z_{.01} = 2.33$

Since the observed $z = 3.78 > z_{.01} = 2.33$, the decision is to **reject the null hypothesis**.

13.12 H_0: There is no difference in the price of a single-family home in Denver and

Hartford

H_a: There is a difference in the price of a single-family home in Denver and

Hartford

Price	Rank	City
132,405	1	D
134,127	2	H
134,157	3	D
134,514	4	H
135,062	5	D
135,238	6	H
135,940	7	D
136,333	8	H
136,419	9	H
136,981	10	D
137,016	11	D
137,359	12	H
137,741	13	H
137,867	14	H
138,057	15	D
139,114	16	H
139,638	17	D
140,031	18	H
140,102	19	D
140,479	20	D

141,408	21	D
141,730	22	D
141,861	23	D
142,012	24	H
142,136	25	H
143,947	26	H
143,968	27	H
144,500	28	H

$n_1 = 13$

$n_2 = 15$

$W_1 = 1 + 3 + 5 + 7 + 10 + 11 + 15 + 17 + 19 + 20 + 21 + 22 + 23 = 174$

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (13)(15) + \frac{(13)(14)}{2} - 174$$

$= 112$

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(13)(15)}{2}$$

$= 97.5$

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(13)(15)(29)}{12}}$$

$= 21.708$

$$z = \frac{U - \mu}{\sigma} = \frac{112 - 97.5}{21.708}$$

= **0.67**

For α = .05 and a two-tailed test, α/2 = .025 and

$z_{.025} = \pm 1.96$.

Since the observed z = 0.67 < $z_{.025}$ = 1.96, the decision is to **fail to reject the null hypothesis**. There is not enough evidence to declare that there is a price difference for single family homes in Denver and Hartford.

13.13 H_o: The population differences = 0

H_a: The population differences ≠ 0

1	2	*d*	Rank
212	179	33	15
234	184	50	16
219	213	6	7.5
199	167	32	13.5
194	189	5	6
206	200	6	7.5

234	212	22	11
225	221	4	5
220	223	-3	- 3.5
218	217	1	1
234	208	26	12
212	215	-3	-3.5
219	187	32	13.5
196	198	-2	-2
178	189	-11	-9
213	201	12	10

n = 16

T_- = 3.5 + 3.5 + 2 + 9 = 18

$$\mu = \frac{(n)(n+1)}{4} = \frac{(16)(17)}{4} = 68$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{16(17)(33)}{24}} = 19.34$$

$$z = \frac{T-\mu}{\sigma} = \frac{18-68}{19.34} = \mathbf{-2.59}$$

$\alpha = .10$ $\alpha/2 = .05$ $z_{.05} = \pm 1.645$

Since the observed $z = -2.59 < z_{.05} = -1.645$, the decision is to **reject the null**

hypothesis.

13.14 H_o: $M_d = 0$

H_a: $M_d \neq 0$

Before	After	d	Rank
49	43	6	+ 9
41	29	12	+12
47	30	17	+14
39	38	1	+ 1.5
53	40	13	+13
51	43	8	+10
51	46	5	+ 7.5
49	40	9	+11
38	42	-4	- 5.5
54	50	4	+ 5.5
46	47	-1	- 1.5
50	47	3	+ 4
44	39	5	+ 7.5
49	49	0	
45	47	-2	- 3

n = 15 but after dropping the zero difference, n = 14

α = .05, for two-tailed $\alpha/2$ = .025, and from Table A.14, $T_{.025,14}$= 21

T_{+} = 9 + 12+ 14 + 1.5 + 13 + 10 + 7.5 + 11 + 5.5 + 4 + 7.5 = 95

T_{-} = 5.5 + 1.5 + 3 = 10

T = min(T_{+},T_{-}) = min(95, 10) = 10

Since the observed value of T = 10 < $T_{.025,\ 14}$ = 21, the decision is to **reject the**

null hypothesis. There is a significant difference in before and after.

13.15 H_o: The population differences ≥ 0

H_a: The population differences < 0

Before	After	*d*	Rank
10,500	12,600	-2,100	-11
8,870	10,660	-1,790	-9
12,300	11,890	410	3
10,510	14,630	-4,120	-17
5,570	8,580	-3,010	-15
9,150	10,115	-965	-7
11,980	14,320	-2,370	-12
6,740	6,900	-160	-2
7,340	8,890	-1,550	-8
13,400	16,540	-3,140	-16
12,200	11,300	900	6
10,570	13,330	-2,760	-13
9,880	9,990	-110	-1
12,100	14,050	-1,950	-10
9,000	9,500	-500	-4
11,800	12,450	-650	-5
10,500	13,450	-2,950	-14

Since $n = 17$, use the large sample test

$T+ = 3 + 6 = 9$

$T = 9$

$$\mu = \frac{(n)(n+1)}{4} = \frac{(17)(18)}{4}$$
= 76.5

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{17(18)(35)}{24}}$$
= 21.12

$$z = \frac{T-\mu}{\sigma} = \frac{9-76.5}{21.12}$$
= **-3.20**

$\alpha = .05$ $z_{.05} = -1.645$

Since the observed $z = -3.20 < z_{.05} = -1.645$, the decision is to **reject the null hypothesis**.

13.16 $H_o: M_d = 0$

$H_a: M_d < 0$

Manual	Scanner	*d*	Rank
426	473	-47	-11

387	446	-59	-13
410	421	-11	-5.5
506	510	-4	-2
411	465	-54	-12
398	409	-11	-5.5
427	414	13	7
449	459	-10	-4
407	502	-95	-14
438	439	-1	-1
418	456	-38	-10
482	499	-17	-8
512	517	-5	-3
402	437	-35	-9

$n = 14$

$T_+ = (+7)$

$T_- = (11 + 13 + 5.5 + 3 + 12 + 5.5 + 4 + 14 + 1 + 10 + 8 + 3 + 9) = 98$

$T = \min(T_+, T_-) = \min(7, 98) = 7$

from Table A.14 with $\alpha = .05$, $n = 14$, $T_{.05,14} = 26$

Since the observed $T = 7 < T_{.05,14} = 26$, the decision is to reject the null hypothesis.

The differences are significantly less than zero and the after scores are significantly higher.

13.17 H_o: The population differences   0

H_a: The population differences < 0

1999	2006	*d*	Rank
49	54	-5	-7.5
27	38	-11	-15
39	38	1	2
75	80	-5	-7.5
59	53	6	11
67	68	-1	-2
22	43	-21	-20
61	67	-6	-11
58	73	-15	-18
60	55	5	7.5
72	58	14	16.5
62	57	5	7.5
49	63	-14	-16.5
48	49	-1	-2
19	39	-20	-19
32	34	-2	-4.5
60	66	-6	-11
80	90	-10	-13.5
55	57	-2	-4.5
68	58	10	13.5

n = 20

T+ = 2 + 11 + 7.5 + 16.5 + 7.5 + 13.5 = 58

T = 58

$$\mu = \frac{(n)(n+1)}{4} = \frac{(20)(21)}{4}$$

= 105

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{20(21)(41)}{24}}$$

= 26.79

$$z = \frac{T-\mu}{\sigma} = \frac{58-105}{26.79}$$

= **-1.75**

For α = .10, $z_{.10}$ = -1.28

Since the observed z = -1.75 < $z_{.10}$ = -1.28, the decision is to **reject the null hypothesis**.

13.18 H_o: The population differences $\leq$ 0

H_a: The population differences > 0

April April

2002	2006	d	Rank
63.1	57.1	5.7	16
67.1	66.4	0.7	3.5
65.5	61.8	3.7	12
68.0	65.3	2.7	8.5
66.6	63.5	3.1	10.5
65.7	66.4	-0.7	-3.5
69.2	64.9	4.3	14
67.0	65.2	1.8	6.5
65.2	65.1	0.1	1.5
60.7	62.2	-1.5	-5
63.4	60.3	3.1	10.5
59.2	57.4	1.8	6.5
62.9	58.2	4.7	15
69.4	65.3	4.1	13
67.3	67.2	0.1	1.5
66.8	64.1	2.7	8.5

$n = 16$

$T- = 8.5$

$T = 8.5$

$$\mu = \frac{(n)(n+1)}{4} = \frac{(16)(17)}{4}$$

$= 68$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{16(17)(33)}{24}}$$

= 19.339

$$z = \frac{T-\mu}{\sigma} = \frac{8.5-68}{19.339}$$

= **-3.08**

For α = .05, $z_{.05}$ = 1.645

Since the observed z = $|-3.08|$ > $z_{.05}$ = 1.645, the decision is to **reject the null hypothesis**.

13.19 H_o: The 5 populations are identical

H_a: At least one of the 5 populations is different

1	2	3	4	5
157	165	219	286	197
188	197	257	243	215
175	204	243	259	235
174	214	231	250	217
201	183	217	279	240
203		203		233
				213

BY RANKS

	1	2	3	4	5
	1	2	18	29	7.5
	6	7.5	26	23.5	15
	4	12	23.5	27	21
	3	14	19	25	16.5
	9	5	16.5	28	22
	10.5		10.5		20
					13
T_j		33.5	40.5	113.5	132.5 115
n_j	6	5	6	5	7

$$\sum \frac{T_j^{\,2}}{n_j} = \frac{(33.5)^2}{6} + \frac{(40.5)^2}{5} + \frac{(113.5)^2}{6} + \frac{(132.5)}{5} + \frac{(115)^2}{7}$$

= 8,062.67

n = 29

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{29(30)}(8{,}062.67) - 3(30)$$

= **21.21**

α = .01   df = c - 1 = 5 - 1 = 4

$\chi^2_{.01,4}$ = 13.2767

Since the observed K = 21.21 > $\chi^2_{.01,4}$ = 13.2767, the decision is to **reject the null hypothesis**.

13.20 H_o: The 3 populations are identical

H_a: At least one of the 3 populations is different

Group 1	Group 2	Group 4
19	30	39
21	38	32
29	35	41
22	24	44
37	29	30
42		27
		33

By Ranks

	Group 1	Group 2	Group 3
	1	8.5	15
	2	14	10
	6.5	12	16
	3	4	18
	13	6.5	8.5
	17		5
			11
T_j	42.5	45	83.5
n_j	6	5	7

$$\sum \frac{T_j^{\,2}}{n_j} = \frac{(42.5)^2}{6} + \frac{(45)^2}{5} + \frac{(83.5)^2}{7}$$

= 1,702.08

n = 18

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{18(19)}(1{,}702.08) - 3(19)$$

= **2.72**

α = .05, df = *c* - 1 = 3 - 1 = 2

$\chi^2_{.05,2}$ = 5.9915

Since the observed *K* = 2.72 < $\chi^2_{.05,2}$ = 5.9915, the decision is to **fail to reject**

the null hypothesis.

13.21 H_o: The 4 populations are identical

H_a: At least one of the 4 populations is different

Region 1	Region 2	Region 3	Region 4
$1,200	$225	$ 675	$1,075
450	950	500	1,050
110	100	1,100	750
800	350	310	180
375	275	660	330
200			680
			425

By Ranks

	Region 1	Region 2	Region 3	Region 4
	23	5	15	21
	12	19	13	20
	2	1	22	17
	18	9	7	3
	10	6	14	8
	4			16
				11
T_j	69	40	71	96
n_j	6	5	5	7

$$\sum \frac{T_j^{\,2}}{n_j} = \frac{(69)^2}{6} + \frac{(40)^2}{5} + \frac{(71)^2}{5} + \frac{(96)^2}{7}$$

= 3,438.27

n = 23

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{23(24)}(3{,}428.27) - 3(24)$$

= **2.75**

α = .05 df = c - 1 = 4 - 1 = 3

$\chi^2_{.05,3}$ = 7.8147

Since the observed K = 2.75 < $\chi^2_{.05,3}$ = 7.8147, the decision is to **fail to reject the null hypothesis**.

13.22 H_o: The 3 populations are identical

H_a: At least one of the 3 populations is different

Small Town	City	Suburb
$21,800	$22,300	$22,000

22,500	21,900	22,600
21,750	21,900	22,800
22,200	22,650	22,050
21,600	21,800	21,250
		22,550

By Ranks

	Small Town	City	Suburb
	4.5	11	8
	12	6.5	14
	3	6.5	16
	10	15	9
	2	4.5	1
			13
T_j	31.5	43.5	61
n_j	5	5	6

$$\sum \frac{T_j^2}{n_j} = \frac{(31.5)^2}{5} + \frac{(43.5)^2}{5} + \frac{(61)^2}{6}$$

$$= 1{,}197.07$$

$n = 16$

$$K = \frac{12}{n(n+1)}\sum\frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{16(17)}(1{,}197.07) - 3(17)$$

= **1.81**

α = .05 df = c - 1 = 3 - 1 = 2

$\chi^2_{.05,2}$ = 5.9915

Since the observed K = 1.81 < $\chi^2_{.05,2}$ = 5.9915, the decision is to **fail to reject**

the null hypothesis.

13.23 H_o: The 4 populations are identical

H_a: At least one of the 4 populations is different

Amusement Parks	Lake Area	City	National Park
0	3	2	2
1	2	2	4
1	3	3	3
0	5	2	4
2	4	3	3
1	4	2	5
0	3	3	4
	5	3	4
	2	1	
		3	

By Ranks

Amusement Parks	Lake Area	City	National Park
2	20.5	11.5	11.5
5.5	11.5	11.5	28.5
5.5	20.5	20.5	20.5
2	33	11.5	28.5
11.5	28.5	20.5	20.5
5.5	28.5	11.5	33
2	20.5	20.5	28.5

		33	20.5	28.5
		11.5	5.5	
			20.5	
T_j	34	207.5	154.0	199.5
n_j	7	9	10	8

$$\sum \frac{T_j^2}{n_j} = \frac{(34)^2}{7} + \frac{(207.5)^2}{9} + \frac{(154)^2}{10} + \frac{(199.5)}{8}$$

= 12,295.80

n = 34

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^2}{n_j} - 3(n+1) = \frac{12}{34(35)}(12{,}295.80) - 3(35)$$

= **18.99**

α = .05 df = c - 1 = 4 - 1 = 3

$\chi^2_{.05,3}$ = 7.8147

Since the observed K = 18.99 > $\chi^2_{.05,3}$ = 7.8147, the decision is to **reject the null hypothesis**.

13.24 H_o: The 3 populations are identical

H_a: At least one of the 3 populations is different

Day Shift	Swing Shift	Graveyard Shift
52	45	41
57	48	46
53	44	39
56	51	49
55	48	42
50	54	35
51	49	52
	43	

By Ranks

	Day Shift	Swing Shift	Graveyard Shift
	16.5	7	3
	22	9.5	8
	18	6	2
	21	14.5	11.5
	20	9.5	4
	13	19	1
	14.5	11.5	16.5
		5	
T_j	125	82	46
n_j	7	8	7

$$\sum \frac{T_j^2}{n_j} = \frac{(125)^2}{7} + \frac{(82)^2}{8} + \frac{(46)^2}{7}$$

$$= 3{,}374.93$$

$n = 22$

$$K = \frac{12}{n(n+1)}\sum\frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{22(23)}(3{,}374.93) - 3(23)$$

= **11.04**

α = .05 df = c - 1 = 3 - 1 = 2

$\chi^2_{.05,2} = 5.9915$

Since the observed K = 11.04 > $\chi^2_{.05,2}$ = 5.9915, the decision is to **reject the**

null hypothesis.

13.25 H_o: The treatment populations are equal

H_a: At least one of the treatment populations yields larger values than at least one

other treatment population.

Use the Friedman test with $\alpha = .05$

$c = 5$, $b = 5$, df = $c - 1 = 4$, $\chi^2_{.05,4} = 9.4877$

If the observed value of $\chi^2 > 9.4877$, then the decision will be to reject the null

hypothesis.

Shown below are the data ranked by blocks:

	1	2	3	4	5
1	1	4	3	5	2
2	1	3	4	5	2
3	2.5	1	4	5	2.5
4	3	2	4	5	1
5	4	2	3	5	1
R_j	11.5	12	18	25	8.5

R_j^2 132.25 144 324 625
72.25

ΣR_j^2 = 1,297.5

$$\chi_r^{\,2} = \frac{12}{bc(c+1)}\sum R_j^{\,2} - 3b(c+1) = \frac{12}{(5)(5)(6)}(1{,}297.5) - 3(5)(6)$$

= **13.8**

Since the observed value of χ_r^2 = 13.8 > $\chi_{4,.05}^2$ = 9.4877, the decision is to

reject the null hypothesis. At least one treatment population yields larger values than at least one other treatment population.

13.26 H_o: The treatment populations are equal

H_a: At least one of the treatment populations yields larger values than at least one

other treatment population.

Use the Friedman test with α = .05

c = 6, b = 9, df = c - 1 = 5, $\chi^2_{.05,5}$ = 11.0705

If the observed value of χ^2 > 11.0705, then the decision will be to reject the null hypothesis.

Shown below are the data ranked by blocks:

	1	2	3	4	5	6
1	1	3	2	6	5	4
2	3	5	1	6	4	2
3	1	3	2	6	5	4
4	1	3	4	6	5	2
5	3	1	2	4	6	5
6	1	3	2	6	5	4
7	1	2	4	6	5	3
8	3	1	2	6	5	4
9	1	2	3	6	5	4
R_j	15	25	25	56	50	38

R_j^2	225	625	625	3136	2500	1444

$\Sigma R_j^2 = 8,555.5$

$$\chi_r^{\ 2} = \frac{12}{bc(c+1)}\sum R_j^{\ 2} - 3b(c+1) = \frac{12}{(9)(6)(7)}(8,555) - 3(9)(7)$$

$= $ **82.59**

Since the observed value of $\chi_r^2 = 82.59 > \chi_{5,.05}^2 = 11.0705$, the decision is to

reject the null hypothesis. At least one treatment population yields larger values

than at least one other treatment population.

13.27 H_o: The treatment populations are equal

H_a: At least one of the treatment populations yields larger values than at least one

other treatment population.

Use the Friedman test with $\alpha = .01$

$c = 4$, $b = 6$, df = $c - 1 = 3$, $\chi^2_{.01,3} = 11.3449$

If the observed value of $\chi^2 > 11.3449$, then the decision will be to reject the null hypothesis.

Shown below are the data ranked by blocks:

	1	2	3	4
1	1	4	3	2
2	2	3	4	1
3	1	4	3	2
4	1	3	4	2
5	1	3	4	2
6	2	3	4	1
R_j	8	20	22	10
R_j^2	64	400	484	100

ΣR_j^2 = 1,048

$$\chi_r^{\ 2} = \frac{12}{bc(c+1)} \sum R_j^{\ 2} - 3b(c+1) = \frac{12}{(6)(4)(5)}(1{,}048) - 3(6)(5)$$

= **14.8**

Since the observed value of χ_r^2 = 14.8 > $\chi^2_{3,.01}$ = 11.3449, the decision is to

reject the null hypothesis. At least one treatment population yields larger values

than at least one other treatment population.

13.28 H_o: The treatment populations are equal

H_a: At least one of the treatment populations yields larger values than at least one

other treatment population.

Use the Friedman test with $\alpha = .05$

$c = 3$, $b = 10$, df = $c - 1 = 2$, $\chi^2_{.05,2} = 5.9915$

If the observed value of $\chi^2 > 5.9915$, then the decision will be to reject the null

hypothesis.

Shown below are the data ranked by blocks:

Worker	5-day	4-day	3.5 day	
1	3	2	1	
2	3	2	1	
3		1		3
2				
4	3	2	1	
5	2	3	1	
6	3	2	1	
7	3	1	2	
8	3	2	1	
9	3	2	1	

10	3	1	2
R_j	29	18	13
R_j^2	841	324	169

$\Sigma R_j^2 = 1{,}334$

$$\chi_r^{\ 2} = \frac{12}{bc(c+1)}\sum R_j^{\ 2} - 3b(c+1) = \frac{12}{(10)(3)(4)}(1{,}334) - 3(10)(4)$$

$= \mathbf{13.4}$

Since the observed value of $\chi_r^2 = 13.4 > \chi^2_{.05,2} = 5.9915$, the decision is to

reject the null hypothesis. At least one treatment population yields larger values

than at least one other treatment population.

13.29 c = 4 treatments b = 5 blocks

$S = \chi_r^2 = 2.04$ with a p-value of .564.

Since the p-value of .564 > α = .10, .05, or .01, the decision is to **fail to reject**

the null hypothesis. There is no significant difference in treatments.

13.30 The experimental design is a random block design that has been analyzed using a Friedman test. There are five treatment levels and seven blocks. Thus, the degrees of freedom are four. The observed value of S = 13.71 is the equivalent of χ_r^2. The p-value is .009 indicating that this test is significant at alpha .01. The null hypothesis is rejected. That is, at least one population yields larger values than at least one other population. An examination of estimated medians shows that treatment 1 has the lowest value and treatment 3 has the highest value.

13.31

x	y	x Ranked	y Ranked	d	d^2
23	201	3	2	1	1
41	259	10.5	11	-.5	0.25
37	234	8	7	1	1
29	240	6	8	-2	4

25	231	4	6	-2	4
17	209	1	3	-2	4
33	229	7	5	2	4
41	246	10.5	9	1.5	2.25
40	248	9	10	-1	1
28	227	5	4	1	1
19	200	2	1	1	1
					$\Sigma d^2 = 23.5$

$n = 11$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(23.5)}{11(120)}$$

$= \mathbf{.893}$

13.32

x	y	d	d^2
4	6	-2	4
5	8	-3	9
8	7	1	1
11	10	1	1
10	9	1	1
7	5	2	4
3	2	1	1
1	3	-2	4
2	1	1	1
9	11	-2	4
6	4	2	4

$\Sigma d^2 = 34$

$n = 11$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(34)}{11(120)}$$

$= \mathbf{.845}$

13.33

x	y	x Ranked	y Ranked	d	d^2

99	108	8	2	6	36
67	139	4	5	-1	1
82	117	6	3	3	9
46	168	1	8	-7	49
80	124	5	4	1	1
57	162	3	7	-4	16
49	145	2	6	-4	16
91	102	7	1	6	36

$\Sigma d^2 = 164$

$n = 8$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(164)}{8(63)}$$

$= \mathbf{-.95}$

13.34

x	y	x Ranked	y Ranked	d	d^2
92	9.3	8	9	-1	1
96	9.0	9	8	1	1
91	8.5	6.5	7	-.5	.25
89	8.0	5	3	2	4
91	8.3	6.5	5	1.5	2.25
88	8.4	4	6	-2	4
84	8.1	3	4	-1	1
81	7.9	1	2	-1	1
83	7.2	2	1	1	1

$\Sigma d^2 = 15.5$

$n = 9$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(15.5)}{9(80)}$$

= **.871**

13.35

Bank Credit Card	Home Equity Loan	Bank Cr. Cd. Rank	Home Eq. Loan Rank	d	d^2
2.51	2.07	12	1	11	121
2.86	1.95	6.5	2	4.5	20.25
2.33	1.66	13	6	7	49

2.54	1.77	10	3	7	49
2.54	1.51	10	7.5	2.5	6.25
2.18	1.47	14	10	4	16
3.34	1.75	3	4	-1	1
2.86	1.73	6.5	5	1.5	2.25
2.74	1.48	8	9	-1	1
2.54	1.51	10	7.5	2.5	6.25
3.18	1.25	4	14	-10	100
3.53	1.44	1	11	-10	100
3.51	1.38	2	12	-10	100
3.11	1.30	5	13	-8	64

$\Sigma d^2 = 636$

$n = 14$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(636)}{14(14^2-1)}$$

= **-.398**

There is a very modest negative correlation between overdue payments for bank

credit cards and home equity loans.

13.36

Year	Iron Rank	Steel Rank	d	d^2
1	12	12	0	0

2	11	10	1	1
3	3	5	-2	4
4	2	7	-5	25
5	4	6	-2	4
6	10	11	-1	1
7	9	9	0	0
8	8	8	0	0
9	7	4	3	9
10	1	3	-2	4
11	6	2	4	16
12	5	1	4	16

$\Sigma d^2 = 80$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(80)}{12(144-1)}$$

$= \mathbf{0.72}$

13.37

No. Co. on NYSE	No. Eq. Is. on AMEX	Rank NYSE	Rank AMEX	d	d^2
1774	1063	11	1	10	100
1885	1055	10	2	8	64
2088	943	9 5	4	16	
2361	1005	8	3	5	25

2570	981	7	4	3	9
2675	936	6	6	0	0
2907	896	4	7	-3	9
3047	893	2	8	-6	36

3114 862 1 9 -8 64
3025 769 3 10 -7 49
2862 765 5 11 -6
36 $\Sigma d^2 = 162$

$$\sum d^2 = 408$$

$n = 11$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(408)}{11(11^2-1)}$$

= **-0.855**

There is a strong negative correlation between the number of companies listed on the NYSE and the number of equity issues on the American Stock Exchange.

13.38 $\alpha = .05$

H_0: The observations in the sample are randomly generated

H_a: The observations in the sample are not randomly generated

n_1 = 13, n_2 = 21

R = 10

Since this is a two-tailed test, use $\alpha/2$ = .025. The critical value is: $z_{.025}$ = $\pm$ 1.96

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2(13)(21)}{13+21}+1$$

= 17.06

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(13)(21)[2(13)(21)-13-21]}{(13+21)^2(13+21-1)}}$$

= 2.707

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{10-17.06}{2.707}$$

= **-2.61**

Since the observed z = - 2.61 < $z_{.025}$ = - 1.96, the decision is to **reject the null hypothesis**. The observations in the sample are not randomly generated.

13.39

Sample 1	Sample 2
573	547
532	566
544	551
565	538
540	557
548	560
536	557
523	547

α = .01 Since n_1 = 8, n_2 = 8 $\leq$ 10, use the small sample Mann-Whitney U test.

x	Rank	Group
523	1	1
532	2	1
536	3	1
538	4	2
540	5	1
544	6	1
547	7.5	2
547	7.5	2
548	9	1
551	10	2
557	11.5	2
557	11.5	2
560	13	2
565	14	1
566	15	2
573	16	1

$W_1 = 1 + 2 + 3 + 5 + 6 + 9 + 14 + 16 = 56$

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (8)(8) + \frac{(8)(9)}{2} - 56$$

$= 44$

$U_2 = n_1 \cdot n_2 - U_1$

$= 8(8) - 44 = 20$

Take the smaller value of U_1, U_2 = **20**

From Table A.13, the *p*-value (1-tailed) is .1172, for 2-tailed, the *p*-value is **.2344**. Since the *p*-value is > α = .05, the decision is to **fail to reject the null hypothesis**.

13.40 $\alpha = .05$, $n = 9$

H_0: $M_d = 0$

H_a: $M_d \neq 0$

Group 1	Group 2	d	Rank
5.6	6.4	-0.8	-8.5
1.3	1.5	-0.2	-4.0
4.7	4.6	0.1	2.0
3.8	4.3	-0.5	-6.5
2.4	2.1	0.3	5.0
5.5	6.0	-0.5	-6.5
5.1	5.2	-0.1	-2.0
4.6	4.5	0.1	2.0
3.7	4.5	-0.8	-8.5

Since $n = 9$, from Table A.14 (2-tailed test), $T_{.025} = 6$

$T_+ = 2 + 5 + 2 = 9$

$T_- = 8.5 + 4 + 6.5 + 6.5 + 2 + 8.5 = 36$

$T = \min(T_+, T_-) = \mathbf{9}$

Since the observed value of T = 9 > $T_{.025}$ = 6, the decision is to **fail to reject the**

null hypothesis. There is not enough evidence to declare that there is a difference between the two groups.

13.41 $n_j = 7$, $n = 28$, $c = 4$, df = 3

Group 1	Group 2	Group 3	Group 4
6	4	3	1
11	13	7	4
8	6	7	5
10	8	5	6
13	12	10	9
7	9	8	6
10	8	5	7

By Ranks:

	Group 1	Group 2	Group 3	Group 4
	9.5	3.5	2	1
	25	27.5	13.5	3.5
	17.5	9.5	13.5	6
	23	17.5	6	9.5
	27.5	26	23	20.5
	13.5	20.5	17.5	9.5
	23	17.5	6	13.5
T_j	139	122	81.5	63.5

$$\sum \frac{T_j^{\ 2}}{n_j}$$

$= 2760.14 + 2126.29 + 948.89 + 576.04 = 6411.36$

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^{\ 2}}{n_j} - 3(n+1) = \frac{12}{28(29)}(6411.36) - 3(29)$$

= **7.75**

The critical value of chi-square is: $\chi^2_{3,.01} = 11.3449$.

Since $K = 7.75 < \chi^2_{3,.01} = 11.3449$, the decision is to **fail to reject the null**

hypothesis.

13.42 $\alpha = .05$, $b = 7$, $c = 4$, df = 3

$\chi^2_{.05,3} = 7.8147$

H_0: The treatment populations are equal

H_a: At least one treatment population yields larger values than at least one other treatment population

Blocks	Group 1	Group 2	Group 3	Group 4
1	16	14	15	17
2	8	6	5	9
3	19	17	13	9
4	24	26	25	21
5	13	10	9	11
6	19	11	18	13
7	21	16	14	15

By Ranks:

Blocks	Group 1	Group 2	Group 3	Group 4
1	3	1	2	4
2	3	2	1	4
3	4	3	2	1
4	2	4	3	1

5	4	2	1	3
6	4	1	3	2
7	4	3	1	2
R_j	24	16	13	17
R_j^2	576	256	169	289

R_j^2 = 567 + 256 + 169 + 289 = 1290

$$\chi_r^2 = \frac{12}{bC(C+1)}\sum R_j^2 - 3b(C+1) = \frac{12}{(7)(4)(5)}(1{,}290) - 3(7)(5)$$

= **5.57**

Since χ_r^2 = 5.57 < $\chi^2_{.05,3}$ = 7.8147, the decision is to **fail to reject the null hypothesis**. The treatment population means are equal.

13.43 Ranks

1	2	1	2	*d*	d^2
101	87	1	7	-6	36
129	89	2	8	-6	36
133	84	3	6	-3	9
147	79	4	5	-1	1
156	70	5	3	2	4
179	64	6	1	5	25
183	67	7	2	5	25
190	71	8	4	4	16

$\Sigma d^2 = 152$

$n = 8$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(152)}{8(63)}$$

= **-.81**

13.44 H_o: The 3 populations are identical

H_a: At least one of the 3 populations is different

1 Gal.	5 Gal.	10 Gal.
1.1	2.9	3.1

1.4	2.5	2.4
1.7	2.6	3.0
1.3	2.2	2.3
1.9	2.1	2.9
1.4	2.0	1.9
2.1	2.7	

By Ranks

	1 Gal.	5 Gal.	10 Gal.
	1	17.5	20
	3.5	14	13
	5	15	19
	2	11	12
	6.5	9.5	17.5
	3.5	8	6.5
	9.5	16	
T_j	31	91	88
n_j	7	7	6

$$\sum \frac{T_j^2}{n_j} = \frac{(31)^2}{7} + \frac{(91)^2}{7} + \frac{(88)^2}{6}$$

$$= 2{,}610.95$$

$n = 20$

$$K = \frac{12}{n(n+1)}\sum\frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{20(21)}(2{,}610.95) - 3(21)$$

= **11.60**

α = .01 df = c - 1 = 3 - 1 = 2

$\chi^2_{.01,2}$ = 9.2104

Since the observed K = 11.60 > $\chi^2_{.01,2}$ = 9.2104, the decision is to **reject the null**

hypothesis.

13.45 N = 40 n_1 = 24 n_2 = 16 α = .05

Use the large sample runs test since both n_1, n_2 are not less than 20.

H_0: The observations are randomly generated

H_a: The observations are not randomly generated

With a two-tailed test, $\alpha/2$ = .025, $z_{.025}$ = $\pm$ 1.96. If the observed z > . 196

or < -1.96, the decision will be to reject the null hypothesis.

R = 19

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2(24)(16)}{24+16}+1$$

$$= 20.2$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(24)(16)\left[2(24)(16)-24-16\right]}{(40)^2(39)}}$$

$$= 2.993$$

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{19-20.2}{2.993}$$

$$= \mathbf{-0.40}$$

Since z = -0.40 > $z_{.025}$ = -1.96, the decision is to **fail to reject the null hypothesis**.

13.46 Use the Friedman test. Let α = .05

H_0: The treatment populations are equal

H_a: The treatment populations are not equal

c = 3 and b = 7

Operator	Machine 1	Machine 2	Machine 3
1	231	229	234

2	233	232	231
3	229	233	230
4	232	235	231
5	235	228	232
6	234	237	231
7	236	233	230

By ranks:

Operator	Machine 1	Machine 2	Machine 3
1	2	1	1
3	1	3	2
4	2	3	1
5	3	1	2
6	2	3	1
7	3	2	1
R_j	16	15	11
R_j^2	256	225	121

df = c - 1 = 2 $\chi^2_{.05,2}$ = 5.99147.

If the observed χ^2_r > 5.99147, the decision will be to reject the null hypothesis.

$\Sigma R_j^2 = 256 + 225 + 121 = 602$

$$\chi_r^{\ 2} = \frac{12}{bc(c+1)}\sum R_j^{\ 2} - 3b(c+1) = \frac{12}{(7)(3)(4)}(602) - 3(7)(4)$$

$= \mathbf{2}$

Since $\chi^2_r = 2 < \chi^2_{.05,2} = 5.99147$, the decision is to **fail to reject the null**

hypothesis.

13.47 H_o: EMS workers are not older

H_a: EMS workers are older

Age	Rank	Group
21	1	1
23	2	1
24	3	1
25	4	1
27	6	1
27	6	2
27	6	2
28	9	1
28	9	2
28	9	2
29	11	2
30	13	2
30	13	2
30	13	2
32	15	1
33	16.5	2
33	16.5	2
36	18.5	1
36	18.5	2
37	20	1
39	21	2

41 22 1

n_1 = 10 n_2 = 12

W_1 = 1 + 2 + 3 + 4 + 6 + 9 + 15 + 18.5 + 20 + 22 = 100.5

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(10)(12)}{2}$$

= 60

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(10)(12)(23)}{12}}$$

= 15.17

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (10)(12) + \frac{(10)(11)}{2} - 100.5$$

= 74.5

$$z = \frac{U - \mu}{\sigma} = \frac{74.5 - 60}{15.17}$$

= **0.96** with α = .05, $z_{.05}$ = -1.645

Since the observed z = 0.96 < $z_{.05}$ = $|-1.645|$, the decision is to **fail to reject the null hypothesis**.

13.48 H_o: The population differences = 0

H_a: The population differences ≠ 0

With	Without	*d*	Rank
1180	1209	-29	-6
874	902	-28	-5
1071	862	209	18
668	503	165	15
889	974	-85	-12.5
724	675	49	9
880	821	59	10
482	567	-85	-12.5
796	602	194	16
1207	1097	110	14
968	962	6	1
1027	1045	-18	-4
1158	896	262	20
670	708	-38	-8
849	642	207	17
559	327	232	19
449	483	-34	-7
992	978	14	3
1046	973	73	11
852	841	11	2

$n = 20$

T_{-} = 6 + 5 + 12.5 + 12.5 + 4 + 8 + 7 = 55

T = 55

$$\mu = \frac{(n)(n+1)}{4} = \frac{(20)(21)}{4}$$
= 105

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{20(21)(41)}{24}}$$
= 26.79

$$z = \frac{T-\mu}{\sigma} = \frac{55-105}{26.79}$$
= **-1.87**

α = .01, $\alpha/2$ = .005 $z_{.005}$ = ±2.575

Since the observed z = -1.87 > $z_{.005}$ = -2.575, the decision is to **fail to reject the**

null hypothesis.

13.49 H_0: There is no difference between March and June

H_a: There is a difference between March and June

GMAT	Rank	Month
350	1	J

430	2	M
460	3	J
470	4	J
490	5	M
500	6	M
510	7	M
520	8	J
530	9.5	M
530	9.5	J
540	11	M
550	12.5	M
550	12.5	J
560	14	M
570	15.5	M
570	15.5	J
590	17	J
600	18	M
610	19	J
630	20	J

$n_1 = 10 \quad n_2 = 10$

$W_1 = 1 + 3 + 4 + 8 + 9.5 + 12.5 + 15.5 + 17 + 19 + 20 = 109.5$

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (10)(10) + \frac{(10)(11)}{2} - 109.5$$

= 45.5

$$U_2 = n_1 \cdot n_2 - U_1$$

= (10)(10) - 45.5 = 54.5

From Table A.13, the *p*-value for U = 45 is .3980 and for 44 is .3697. For a two-tailed test, double the *p*-value to at least .739. Using α = .10, the decision is to **fail to reject the null hypothesis**.

13.50 Use the Friedman test. b = 6, c = 4, df = 3, α = .05

H_0: The treatment populations are equal

H_a: At least one treatment population yields larger values than at least on other

treatment population

The critical value is: $\chi^2_{.05,3}$ = 7.8147

Location

Brand	1	2	3	4
A	176	58	111	120
B	156	62	98	117
C	203	89	117	105
D	183	73	118	113
E	147	46	101	114
F	190	83	113	115

By ranks:

Location

Brand	1	2	3	4
A	4	1	2	3
B	4	1	2	3
C	4	1	3	2
D	4	1	3	2

E	4	1	2	3
F	4	1	2	3
R_j	24	6	14	16
R_j^2	576	36	196	256

$\Sigma R_j^2 = 1{,}064$

$$\chi_r^2 = \frac{12}{bc(c+1)}\sum R_j^2 - 3b(c+1) = \frac{12}{(6)(4)(5)}(1{,}064) - 3(6)(5)$$

$$= \mathbf{16.4}$$

Since $\chi_r^2 = 16.4 > \chi^2_{.05,3} = 7.8147$, the decision is to reject the null hypothesis.

At least one treatment population yields larger values than at least one other

treatment population. An examination of the data shows that location one produced the highest sales for all brands and location two produced the lowest sales of gum for all brands.

13.51 H_o: The population differences = 0

H_a: The population differences ≠ 0

Box	No Box	*d*	Rank
185	170	15	11
109	112	-3	-3
92	90	2	2
105	87	18	13.5
60	51	9	7
45	49	-4	-4.5
25	11	14	10
58	40	18	13.5
161	165	-4	-4.5
108	82	26	15.5
89	94	-5	-6
123	139	-16	-12
34	21	13	8.5
68	55	13	8.5
59	60	-1	-1
78	52	26	15.5

$n = 16$

$T_- = 3 + 4.5 + 4.5 + 6 + 12 + 1 = 31$

$T = 31$

$$\mu = \frac{(n)(n+1)}{4} = \frac{(16)(17)}{4}$$

= 68

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{16(17)(33)}{24}}$$

= 19.34

$$z = \frac{T-\mu}{\sigma} = \frac{31-68}{19.34}$$

= **-1.91**

α = .05, α/2 = .025 $z_{.025}$ = ±1.96

Since the observed z = -1.91 > $z_{.025}$ = -1.96, the decision is to **fail to reject the**

null hypothesis.

13.52

Cups	Stress	Ranked Cups	Ranked Stress	*d*	d^2
25	80	6	8	-2	4
41	85	9	9	0	0
16	35	4	3	1	1
0	45	1	5	-4	16
11	30	3	2	1	1
28	50	7	6	1	1
34	65	8	7	1	1
18	40	5	4	1	1
5	20	2	1	1	1

$\Sigma d^2 = 26$

$n = 9$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(26)}{9(80)}$$

$= \mathbf{.783}$

13.53 $n_1 = 15$, $n_2 = 15$ Use the small sample Runs test

$\alpha = .05$, $\alpha/.025$

H_0: The observations in the sample were randomly generated.

H_a: The observations in the sample were not randomly generated

From Table A.11, lower tail critical value = 10

From Table A.12, upper tail critical value = 22

***R* = 21**

Since *R* = 21 between the two critical values, the decision is to **fail to reject the null hypothesis**. The observations were randomly generated.

13.54 H$_o$: The population differences $\geq$ 0

H$_a$: The population differences < 0

Before	After	*d*	Rank
430	465	-35	-11
485	475	10	5.5
520	535	-15	- 8.5
360	410	-50	-12
440	425	15	8.5
500	505	-5	-2
425	450	-25	-10
470	480	-10	-5.5
515	520	-5	-2
430	430	0	OMIT
450	460	-10	-5.5
495	500	-5	-2
540	530	10	5.5

n = 12

T+ = 5.5 + 8.5 + 5.5 = 19.5 T = **19.5**

From Table A.14, using n = 12, the critical T for α = .01, one-tailed, is 10.

Since $T = 19.5$ is not less than or equal to the critical $T = 10$, the decision is to **fail to reject the null hypothesis**.

13.55 H_o: With ties have no higher scores
H_a: With ties have higher scores

Rating	Rank	Group
16	1	2
17	2	2
19	3.5	2
19	3.5	2
20	5	2
21	6.5	2
21	6.5	1
22	9	1
22	9	1
22	9	2
23	11.5	1
23	11.5	2
24	13	2
25	15.5	1
25	15.5	1
25	15.5	1
25	15.5	2
26	19	1
26	19	1
26	19	2
27	21	1
28	22	1

$n_1 = 11$ $n_2 = 11$

$W_1 = 6.5 + 9 + 9 + 11.5 + 15.5 + 15.5 + 15.5 + 19 + 19 + 21 + 22 =$
163.5

$$\mu = \frac{n_1 \cdot n_2}{2} = \frac{(11)(11)}{2}$$
= 60.5

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(11)(11)(23)}{12}}$$
= 15.23

$$U = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (11)(11) + \frac{(11)(12)}{2} - 163.5$$
= 23.5

$$z = \frac{U - \mu}{\sigma} = \frac{23.5 - 60.5}{15.23}$$
= **-2.43** For $\alpha = .05$, $z_{.05} = 1.645$

Since the observed $z = |-2.43| > z_{.05} = 1.645$, the decision is to **reject the null hypothesis**.

13.56 H_o: Automatic no more productive

H_a: Automatic more productive

Sales	Rank	Type of Dispenser
92	1	M
105	2	M
106	3	M
110	4	A
114	5	M
117	6	M
118	7.5	A
118	7.5	M
125	9	M
126	10	M
128	11	A
129	12	M
137	13	A
143	14	A
144	15	A
152	16	A
153	17	A
168	18	A

$n_1 = 9 \quad n_2 = 9$

$W_1 = 4 + 7.5 + 11 + 13 + 14 + 15 + 16 + 17 + 18 = 115.5$

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - W_1 = (9)(9) + \frac{(9)(10)}{2} - 115.5$$

= 10.5

$$U_2 = n_1 \cdot n_2 - U_1$$

= 81 – 10.5 = 70.5

The smaller of the two is U_1 = 10.5

α = .01

From Table A.13, the *p*-value = **.0039**. The decision is to **reject the null**

hypothesis since the *p*-value is less than .01.

13.57 H_o: The 4 populations are identical

H_a: At least one of the 4 populations is different

45	55	70	85
216	228	219	218
215	224	220	216
218	225	221	217
216	222	223	221
219	226	224	218
214	225		217

By Ranks:

	45	55	70	85	
	4	23	11.5	9	
	2	18.5	13	4	
	9	20.5	14.5	6.5	
	4	16	17	14.5	
	11.5	22	18.5	9	
	1	20.5		6.5	
T_j		31.5	120.5	74.5	49.5
n_j	6	6	5	6	

$$\sum \frac{T_j^{\,2}}{n_j} = \frac{(31.5)^2}{6} + \frac{(120.5)^2}{6} + \frac{(74.5)^2}{5} + \frac{(49.5)}{6}$$

= 4,103.84

n = 23

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^{\,2}}{n_j} - 3(n+1) = \frac{12}{23(24)}(4{,}103.84) - 3(24)$$

= **17.21**

α = .01 df = c - 1 = 4 - 1 = 3

$\chi^2_{.01,3}$ = 11.3449

Since the observed K = 17.21 > $\chi^2_{.01,3}$ = 11.3449, the decision is to **reject the**

null hypothesis.

13.58

Sales	Miles	Ranks Sales	Ranks Miles	*d*	d^2
150,000	1,500	1	1	0	0
210,000	2,100	2	2	0	0
285,000	3,200	3	7	-4	16
301,000	2,400	4	4	0	0
335,000	2,200	5	3	2	4
390,000	2,500	6	5	1	1
400,000	3,300	7	8	-1	1
425,000	3,100	8	6	2	4
440,000	3,600	9	9	0	0

$\Sigma d^2 = 26$

$n = 9$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(26)}{9(80)}$$

= **.783**

13.59 H_o: The 3 populations are identical

H_a: At least one of the 3 populations is different

3-day	Quality	Mgmt. Inv.
9	27	16
11	38	21
17	25	18
10	40	28
22	31	29
15	19	20
6	35	31

By Ranks:

	3-day	Quality	Mgmt. Inv.
	2	14	6
	4	20	11
	7	13	8
	3	21	15
	12	17.5	16
	5	9	10
	1	19	17.5
T_j	34	113.5	83.5
n_j	7	7	7

$$\sum \frac{T_j^2}{n_j} = \frac{(34)^2}{7} + \frac{(113.5)^2}{7} + \frac{(83.5)^2}{7}$$

$= 3{,}001.5$

$n = 21$

$$K = \frac{12}{n(n+1)} \sum \frac{T_j^2}{n_j} - 3(n+1) = \frac{12}{21(22)}(3{,}001.5) - 3(22)$$

$=$ **11.96**

$\alpha = .10$ df = c - 1 = 3 - 1 = 2

$\chi^2_{.10,2} = 4.6052$

Since the observed $K = 11.96 > \chi^2_{.10,2} = 4.6052$, the decision is to **reject the**

null hypothesis.

13.60 H_o: The population differences ≥ 0

H_a: The population differences < 0

Husbands	Wives	d	Rank
27	35	-8	-12
22	29	-7	-11

28	30	-2	-6.5
19	20	-1	-2.5
28	27	1	2.5
29	31	-2	-6.5
18	22	-4	-9.5
21	19	2 6.5	
25	29	-4	-9.5
18	28	-10	-13.5
20	21	-1	-2.5
24	22	2	6.5
23	33	-10	-13.5
25	38	-13	-16.5
22	34	-12	-15
16	31	-15	-18
23	36	-13	-16.5
30	31	-1	-2.5

n = 18

T+ = 2.5 + 6.5 + 6.5 = 15.5

T = 15.51

$$\mu = \frac{(n)(n+1)}{4} = \frac{(18)(19)}{4}$$

= 85.5

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{18(19)(37)}{24}}$$
= 22.96

$$z = \frac{T-\mu}{\sigma} = \frac{15.5-85.5}{22.96}$$
= **-3.05**

α = .01 $z_{.01}$ = -2.33

Since the observed z = -3.05 < $z_{.01}$ = -2.33, the decision is to **reject the null hypothesis**.

13.61 This problem uses a random block design, which is analyzed by the Friedman nonparametric test. There are 4 treatments and 10 blocks. The value of the observed χ_r^2 (shown as *S*) is 12.16 (adjusted for ties) and has an associated *p*-value of .007 that is significant at α = .01. At least one treatment population yields larger values than at least one other treatment population. Examining the treatment medians, treatment one has an estimated median of 20.125 and treatment two has a treatment median of 25.875. These two are the farthest apart.

13.62 This is a Runs test for randomness. $n_1 = 21$, $n_2 = 29$. Because of the size of the

n's, this is a large sample Runs test. There are 28 runs, ***R* = 28.**

$\mu_R = 25.36$ $\sigma_R = 3.34$

$$z = \frac{28 - 25.36}{3.34}$$

= **0.79**

The *p*-value for this statistic is .4387 for a two-tailed test. The decision is to **fail**

to reject the null hypothesis at $\alpha = .05$.

13.63 A large sample Mann-Whitney *U* test is being computed. There are 16 observations in each group. The null hypothesis is that the two populations are identical. The alternate hypothesis is that the two populations are not identical. The value of *W* is 191.5. The *p*-value for the test is .0066. The test is significant at $\alpha = .01$. The decision is to reject the null hypothesis. The two populations are not identical. An examination of medians shows that the median for group two (46.5) is larger than the median for group one (37.0).

13.64 A Kruskal-Wallis test has been used to analyze the data. The null hypothesis is that the four populations are identical; and the alternate hypothesis is that at least one of the four populations is different. The H statistic (same as the *K* statistic) is 11.28 when adjusted for ties. The *p*-value for this H value is .010 which indicates that there is a significant difference in the four groups at $\alpha = .05$ and marginally so for $\alpha = .01$. An examination of the medians reveals that all group medians are the same (35) except for group 2 that has a median of 25.50. It is likely that it is group 2 that differs from the other groups.

Chapter 14

Simple Regression Analysis

LEARNING OBJECTIVES

The overall objective of this chapter is to give you an understanding of bivariate regression analysis, thereby enabling you to:

1. Compute the equation of a simple regression line from a sample of data and interpret the slope and intercept of the equation.
2. Understand the usefulness of residual analysis in examining the fit of the regression line to the data and in testing the assumptions underlying regression analysis.
3. Compute a standard error of the estimate and interpret its meaning.
4. Compute a coefficient of determination and interpret it.
5. Test hypotheses about the slope of the regression model and interpret the results.
6. Estimate values of *y* using the regression model.
7. Develop a linear trend line and use it to forecast.

CHAPTER TEACHING STRATEGY

This chapter is about all aspects of simple (bivariate, linear) regression. Early in the chapter through scatter plots, the student begins to understand that the object of simple regression is to fit a line through the points. Fairly soon in the process, the student learns how to solve for slope and *y* intercept and develop the equation of the regression line. Most of the remaining material on simple regression is to determine how good the fit of the line is and if assumptions underlying the process are met.

The student begins to understand that by entering values of the independent variable into the regression model, predicted values can be determined. The question then becomes: Are the predicted values good estimates of the actual dependent values? One rule to emphasize is that the regression model should not be used to predict for independent variable values that are outside the range of values used to construct the model. MINITAB issues a warning for such activity when attempted. There are many instances where the relationship between *x* and *y* are linear over a given interval but outside the interval the relationship becomes curvilinear or unpredictable. Of course, with this caution having been given, many forecasters use such regression models to extrapolate to values of *x* outside the domain of those used to construct the model. Such forecasts are introduced in section 14.8, “Using Regression to Develop a Forecasting Trend Line”. Whether the forecasts obtained under such conditions are any better than "seat of the pants" or "crystal ball" estimates remains to be seen.

The concept of residual analysis is a good one to show graphically and numerically how the model relates to the data and the fact that it more closely fits some points than others, etc. A graphical or numerical analysis of residuals demonstrates that the regression line fits the data in a manner analogous to the way a mean fits a set of numbers. The regression model passes through the points such that the vertical distances from the actual *y* values to the predicted values will sum to zero. The fact that the residuals sum to zero points out the need to square the errors (residuals) in order to get a handle on total error. This leads to the sum of squares error and then on to the standard error of the estimate. In addition, students can learn why the process is called least squares analysis (the slope and intercept formulas are derived by calculus such that the sum of squares of error is minimized - hence "least squares"). Students can learn that by examining the values of s_e, the residuals, r^2, and the t ratio to test the slope they can begin to make a judgment about the fit of the model to the data. Many of the chapter problems ask the student to comment on these items (s_e, r^2, etc.).

It is my view that for many of these students, the most important facet of this chapter lies in understanding the "buzz" words of regression such as standard error of the estimate, coefficient of determination, etc. because they may only interface regression again as some type of computer printout to be deciphered. The concepts then may be more important than the calculations.

CHAPTER OUTLINE

14.1 Introduction to Simple Regression Analysis

14.2 Determining the Equation of the Regression Line

14.3 Residual Analysis

Using Residuals to Test the Assumptions of the Regression Model

Using the Computer for Residual Analysis

14.4 Standard Error of the Estimate

14.5 Coefficient of Determination

Relationship Between r and r^2

14.6 Hypothesis Tests for the Slope of the Regression Model and Testing the Overall

Model

Testing the Slope

Testing the Overall Model

14.7 Estimation

Confidence Intervals to Estimate the Conditional Mean of y: $\mu_{y/x}$

Prediction Intervals to Estimate a Single Value of y

14.8 Using Regression to Develop a Forecasting Trend Line

Determining the Equation of the Trend Line

Forecasting Using the Equation of the Trend Line

Alternate Coding for Time Periods

14.8 Interpreting Computer Output

KEY TERMS

Coefficient of Determination (r^2)

Confidence Interval

Dependent Variable

Deterministic Model

Heteroscedasticity

Homoscedasticity

Independent Variable

Least Squares Analysis

Outliers

Prediction Interval

Probabilistic Model

Regression Analysis

Residual

Residual Plot

Scatter Plot

Simple Regression

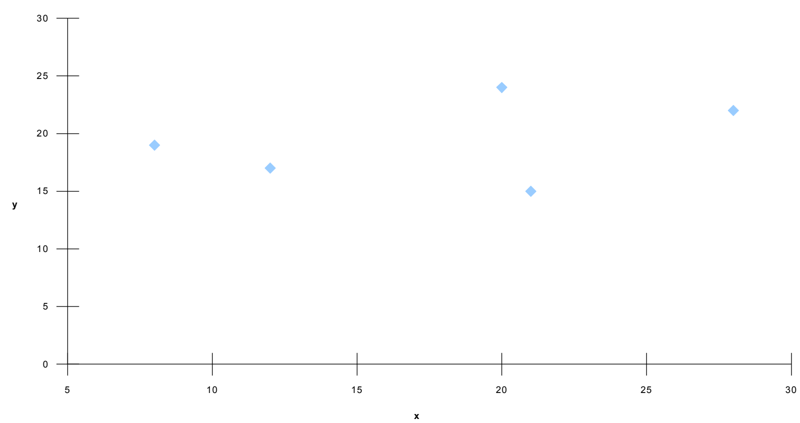
Standard Error of the Estimate (s_e)

Sum of Squares of Error (SSE)

SOLUTIONS TO CHAPTER 14

14.1

x	x
12	17
21	15
28	22
8	19
20	24



$\Sigma x = 89$ $\Sigma y = 97$ $\Sigma xy = 1{,}767$

$\Sigma x^2 = 1{,}833$ $\Sigma y^2 = 1{,}935$ $n = 5$

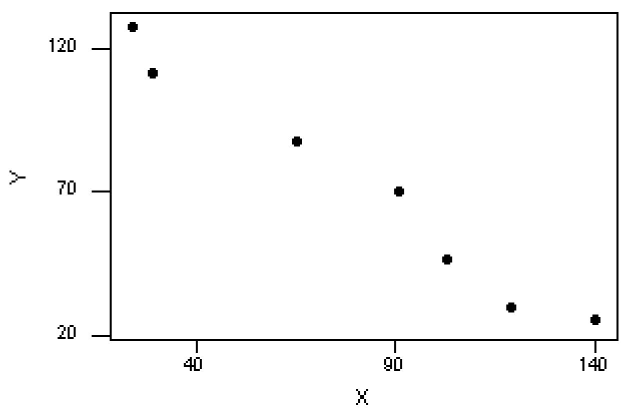
b_1 = $$\frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$ = $$\frac{1{,}767 - \frac{(89)(97)}{5}}{1{,}833 - \frac{(89)^2}{5}}$$ = **0.162**

b_0 = $$\frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{97}{5} - 0.162\frac{89}{5}$$ = **16.51**

$\hat{y}$ = **16.51 + 0.162 *x***

14.2

x	y
140	25
119	29
103	46
91	70
65	88
29	112
24	128



$\Sigma x = 571$ $\Sigma y = 498$ $\Sigma xy = 30{,}099$

$\Sigma x^2 = 58{,}293$ $\Sigma y^2 = 45{,}154$ $n = 7$

b_1 = $\dfrac{SS_{xy}}{SS_x} = \dfrac{\sum xy - \dfrac{\sum x \sum y}{n}}{\sum x^2 - \dfrac{(\sum x)^2}{n}}$ = $\dfrac{30{,}099 - \dfrac{(571)(498)}{7}}{58{,}293 - \dfrac{(571)^2}{7}}$ = **-0.898**

b_0 = $\dfrac{\sum y}{n} - b_1 \dfrac{\sum x}{n} = \dfrac{498}{7} - (-0.898)\dfrac{571}{7}$ = **144.414**

$\hat{y}$ = **144.414 – 0.898 *x***

14.3

(Advertising) x	(Sales) y
12.5	148
3.7	55
21.6	338
60.0	994
37.6	541
6.1	89
16.8	126
41.2	379

$\Sigma x = 199.5$ $\Sigma y = 2,670$ $\Sigma xy = 107,610.4$

$\Sigma x^2 = 7,667.15$ $\Sigma y^2 = 1,587,328$ $n = 8$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{107,610.4 - \frac{(199.5)(2,670)}{8}}{7,667.15 - \frac{(199.5)^2}{8}} = \mathbf{15.240}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{2,670}{8} - 15.24 \frac{199.5}{8} = \mathbf{-46.292}$$

$$\hat{y} = \mathbf{-46.292 + 15.240\, x}$$

14.4

(Prime) x	(Bond) y
16	5
6	12
8	9
4	15
7	7

$\Sigma x = 41$ $\Sigma y = 48$ $\Sigma xy = 333$

$\Sigma x^2 = 421$ $\Sigma y^2 = 524$ $n = 5$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{333 - \frac{(41)(48)}{5}}{421 - \frac{(41)^2}{5}} = \mathbf{-0.715}$$

$$b_0 = \frac{\sum y}{n} - b_1\frac{\sum x}{n} = \frac{48}{5} - (-0.715)\frac{41}{5} = \mathbf{15.460}$$

$$\hat{y} = \mathbf{15.460 - 0.715}\ \boldsymbol{x}$$

14.5

Bankruptcies(y)	Firm Births(x)
34.3	58.1

35.0	55.4
38.5	57.0
40.1	58.5
35.5	57.4
37.9	58.0

$\Sigma x = 344.4$ $\Sigma y = 221.3$ $\Sigma x^2 = 19{,}774.78$

$\Sigma y^2 = 8188.41$ $\Sigma xy = 12{,}708.08$ $n = 6$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{12{,}708.08 - \frac{(344.4)(221.3)}{6}}{19{,}774.78 - \frac{(344.4)^2}{6}} =$$

b_1 = **0.878**

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{221.3}{6} - (0.878)\frac{344.4}{6} = \mathbf{-13.503}$$

$\hat{y}$ = **-13.503 + 0.878 *x***

14.6	No. of Farms (x)	Avg. Size (y)
	5.65	213
	4.65	258
	3.96	297
	3.36	340
	2.95	374
	2.52	420
	2.44	426
	2.29	441
	2.15	460
	2.07	469
	2.17	434
	2.10	444

$\Sigma x = 36.31$ $\Sigma y = 4{,}576$ $\Sigma x^2 = 124.7931$

$\Sigma y^2 = 1{,}825{,}028$ $\Sigma xy = 12{,}766.71$ $n = 12$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{12{,}766.71 - \frac{(36.31)(4{,}576)}{12}}{124.7931 - \frac{(36.31)^2}{12}} = \mathbf{-72.328}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{4{,}576}{12} - (-72.328)\frac{36.31}{12} = \mathbf{600.186}$$

$$\hat{y} = \mathbf{600.186 - 72.3281}\ \boldsymbol{x}$$

14.7

Steel	New Orders
99.9	2.74
97.9	2.87
98.9	2.93
87.9	2.87
92.9	2.98
97.9	3.09
100.6	3.36
104.9	3.61
105.3	3.75
108.6	3.95

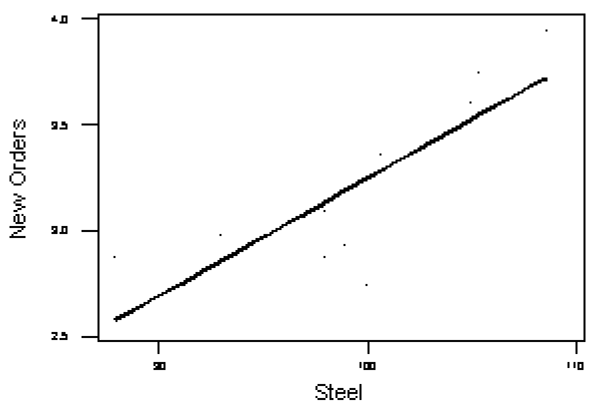
$\Sigma x = 994.8$ $\Sigma y = 32.15$ $\Sigma x^2 = 99{,}293.28$

$\Sigma y^2 = 104.9815$ $\Sigma xy = 3{,}216.652$ $n = 10$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{3{,}216.652 - \frac{(994.8)(32.15)}{10}}{99{,}293.28 - \frac{(994.8)^2}{10}} = \mathbf{0.05557}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{32.15}{10} - (0.05557)\frac{994.8}{10} = \mathbf{-2.31307}$$

$$\hat{y} = \mathbf{-2.31307 + 0.05557}\ \boldsymbol{x}$$



14.8

x	y
15	47
8	36
19	56
12	44
5	21

$\hat{y}$ = 13.625 + 2.303 x

Residuals:

x	y	$\hat{y}$	Residuals ($y-\hat{y}$)
15	47	48.1694	-1.1694
8	36	32.0489	3.9511
19	56	57.3811	-1.3811
12	44	41.2606	2.7394
5	21	25.1401	-4.1401

14.9

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
12	17	18.4582	-1.4582

21	15	19.9196	-4.9196
28	22	21.0563	0.9437
8	19	17.8087	1.1913
20	24	19.7572	4.2428

$\hat{y}$ = 16.51 + 0.162 x

14.10

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
140	25	18.6597	6.3403
119	29	37.5229	-8.5229
103	46	51.8948	-5.8948
91	70	62.6737	7.3263
65	88	86.0281	1.9720
29	112	118.3648	-6.3648
24	128	122.8561	5.1439

$\hat{y}$ = 144.414 - 0.898 x

14.11

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
12.5	148	144.2053	3.7947
3.7	55	10.0954	44.9047
21.6	338	282.8873	55.1127
60.0	994	868.0945	125.9055
37.6	541	526.7236	14.2764
6.1	89	46.6708	42.3292
16.8	126	209.7364	-83.7364
41.2	379	581.5868	-202.5868

$\hat{y}$ = -46.292 + 15.240x

14.12

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
16	5	4.0259	0.9741
6	12	11.1722	0.8278
8	9	9.7429	-0.7429
4	15	12.6014	2.3986
7	7	10.4576	-3.4575

$$\hat{y} = 15.460 - 0.715\,x$$

14.13

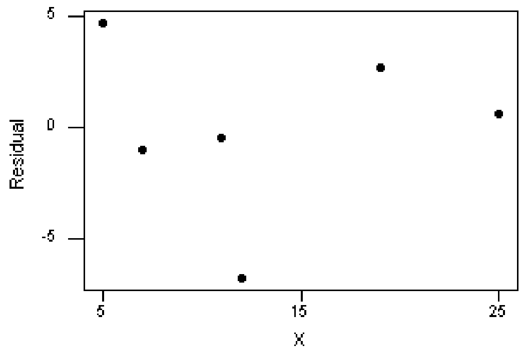
x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
58.1	34.3	37.4978	-3.1978
55.4	35.0	35.1277	-0.1277
57.0	38.5	36.5322	1.9678
58.5	40.1	37.8489	2.2511
57.4	35.5	36.8833	-1.3833
58.0	37.9	37.4100	0.4900

The residual for x = 58.1 is relatively large, but the residual for x = 55.4 is quite small.

14.14

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)
5	47	42.2756	4.7244
7	38	38.9836	-0.9836
11	32	32.3997	-0.3996
12	24	30.7537	-6.7537
19	22	19.2317	2.7683
25	10	9.3558	0.6442

$\hat{y} = 50.506 - 1.646\,x$



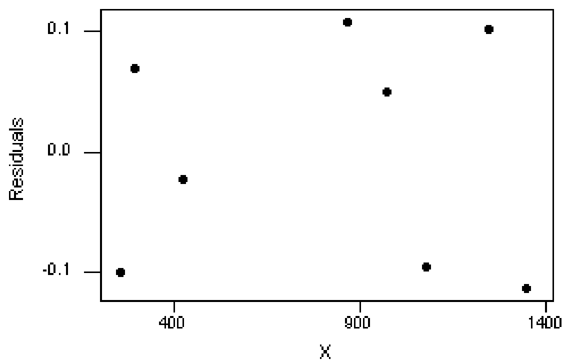
No apparent violation of assumptions

14.15

Miles (x)	Cost y	($\hat{y}$)	($y-\hat{y}$)
1,245	2.64	2.5376	.1024
425	2.31	2.3322	-.0222
1,346	2.45	2.5629	-.1128
973	2.52	2.4694	.0506

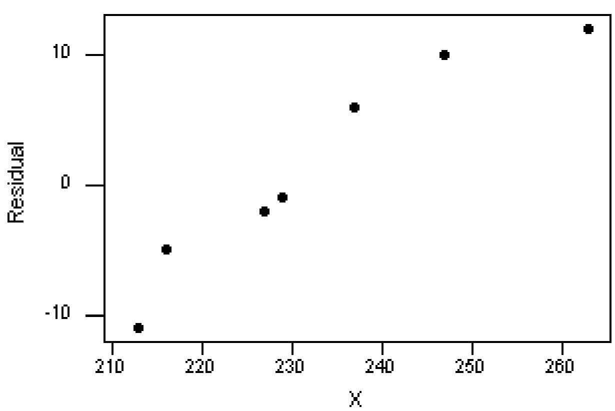
255	2.19	2.2896	-.0996
865	2.55	2.4424	.1076
1,080	2.40	2.4962	-.0962
296	2.37	2.2998	.0702

$\hat{y}$ = 2.2257 – 0.00025 x



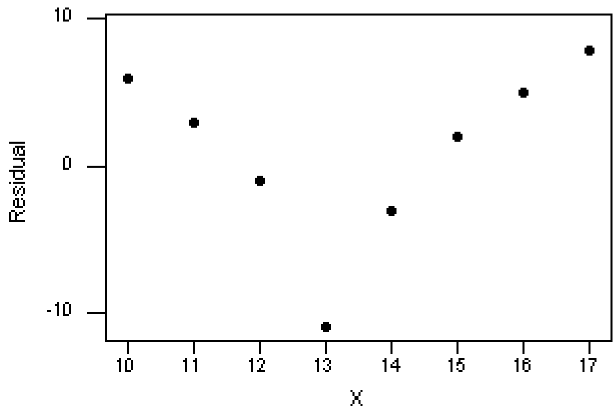
No apparent violation of assumptions

14.16



Error terms appear to be non independent

14.17



There appears to be nonlinear regression

14.18 The MINITAB Residuals vs. Fits graphic is strongly indicative of a violation of the homoscedasticity assumption of regression. Because the residuals are very close together for small values of *x*, there is little variability in the residuals at the left end of the graph. On the other hand, for larger values of *x*, the graph flares out indicating a much greater variability at the upper end. Thus, there is a lack of homogeneity of error across the values of the independent variable.

14.19 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$ = 1,935 - (16.51)(97) - 0.1624(1767) = 46.5692

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{46.5692}{3}}$$

= **3.94**

Approximately 68% of the residuals should fall within ±1s_e.

3 out of 5 or 60% of the actual residuals fell within ± 1s_e.

14.20 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$ = 45,154 - 144.414(498) - (-.89824)(30,099) =

SSE = 272.0

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{272.0}{5}}$$

= **7.376**

6 out of 7 = 85.7% fall within $\pm$ 1s_e

7 out of 7 = 100% fall within $\pm$ 2s_e

14.21 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$ = 1,587,328 - (-46.29)(2,670) - 15.24(107,610.4) =

SSE = 70,940

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{70,940}{6}}$$

= **108.7**

Six out of eight (75%) of the sales estimates are within $108.7 million.

14.22 SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$ = 524 - 15.46(48) - (-0.71462)(333) **= 19.8885**

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{19.8885}{3}}$$

= **2.575**

Four out of five (80%) of the estimates are within 2.575 of the actual rate for

bonds. This amount of error is probably not acceptable to financial analysts.

14.23

x	y	Predicted ($\hat{y}$)	Residuals ($y-\hat{y}$)	$(y-\hat{y})^2$
58.1	34.3	37.4978	-3.1978	10.2259
55.4	35.0	35.1277	-0.1277	0.0163
57.0	38.5	36.5322	1.9678	3.8722
58.5	40.1	37.8489	2.2511	5.0675
57.4	35.5	36.8833	-1.3833	1.9135
58.0	37.9	37.4100	0.4900	0.2401

$\sum(y-\hat{y})^2 = 21.3355$

SSE $= \sum(y-\hat{y})^2 = 21.3355$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{21.3355}{4}} = \mathbf{2.3095}$$

This standard error of the estimate indicates that the regression model is with $\pm$ 2.3095(1,000) bankruptcies about 68% of the time. In this particular problem, 5/6 or 83.3% of the residuals are within this standard error of the estimate.

14.24	$(y-\hat{y})$	$(y-\hat{y})^2$
	4.7244	22.3200
	-0.9836	.9675
	-0.3996	.1597
	-6.7537	45.6125
	2.7683	7.6635
	0.6442	.4150

$\Sigma(y-\hat{y})^2 = 77.1382$

SSE $= \sum(y-\hat{y})^2 = 77.1382$

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{77.1382}{4}}$$

$=$ **4.391**

14.25 $(y-\hat{y})$	$(y-\hat{y})^2$
.1024	.0105
-.0222	.0005
-.1129	.0127
.0506	.0026
-.0996	.0099
.1076	.0116
-.0962	.0093
.0702	.0049

$\Sigma(y-\hat{y})^2 = .0620$ SSE = $\sum (y-\hat{y})^2$ = **.0620**

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{.0620}{6}}$$

= **.1017**

The model produces estimates that are ±.1017 or within about 10 cents 68% of the time. However, the range of milk costs is only 45 cents for this data.

14.26	Volume (x)	Sales (y)
	728.6	10.5
	497.9	48.1
	439.1	64.8

377.9	20.1
375.5	11.4
363.8	123.8
276.3	89.0

$n = 7$ $\Sigma x = 3059.1$ $\Sigma y = 367.7$

$\Sigma x^2 = 1,464,071.97$ $\Sigma y^2 = 30,404.31$ $\Sigma xy = 141,558.6$

$b_1 = -.1504$ $b_0 = 118.257$

$\hat{y} = 118.257 - .1504x$

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$

= 30,404.31 - (118.257)(367.7) - (-0.1504)(141,558.6) = 8211.6245

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{8211.6245}{5}} = \mathbf{40.526}$$

This is a relatively large standard error of the estimate given the sales values

(ranging from 10.5 to 123.8).

14.27 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{46.6399}{1{,}935 - \dfrac{(97)^2}{5}}$ = **.123**

This is a low value of r^2

14.28 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{272.121}{45{,}154 - \dfrac{(498)^2}{7}}$ = **.972**

This is a high value of r^2

14.29 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{70{,}940}{1{,}587{,}328 - \dfrac{(2{,}670)^2}{8}}$ = **.898**

This value of r^2 is relatively high

14.30 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{19.8885}{524 - \dfrac{(48)^2}{5}}$ = **.685**

This value of r^2 is a modest value.

68.5% of the variation of *y* is accounted for by *x* but 31.5% is unaccounted for.

14.31 $r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{21.33547}{8{,}188.41 - \dfrac{(221.3)^2}{6}}$ = **.183**

This value is a low value of r^2.

Only 18.3% of the variability of *y* is accounted for by the *x* values and 81.7% are

unaccounted for.

14.32

CCI	Median Income
116.8	37.415
91.5	36.770
68.5	35.501
61.6	35.047
65.9	34.700
90.6	34.942
100.0	35.887

104.6	36.306
125.4	37.005

$\Sigma x = 323.573$ $\Sigma y = 824.9$ $\Sigma x^2 = 11{,}640.93413$

$\Sigma y^2 = 79{,}718.79$ $\Sigma xy = 29{,}804.4505$ $n = 9$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{29{,}804.4505 - \frac{(323.573)(824.9)}{9}}{11{,}640.93413 - \frac{(323.573)^2}{9}} =$$

b_1 = **19.2204**

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{824.9}{9} - (19.2204)\frac{323.573}{9} = \mathbf{-599.3674}$$

$\hat{y}$ = **-599.3674 + 19.2204 *x***

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma XY$ =

79,718.79 – (-599.3674)(824.9) – 19.2204(29,804.4505) = 1283.13435

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{1283.13435}{7}} = \mathbf{13.539}$$

$r^2 = 1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1 - \dfrac{1283.13435}{79{,}718.79 - \dfrac{(824.9)^2}{9}}$ = **.688**

14.33 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{3.94}{\sqrt{1.833 - \dfrac{(89)^2}{5}}}$ = .2498

b_1 = 0.162

H_o: $\beta = 0$ $\alpha = .05$

H_a: $\beta \neq 0$

This is a two-tail test, $\alpha/2 = .025$ df = n - 2 = 5 - 2 = 3

$t_{.025,3} = \pm 3.182$

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{0.162 - 0}{.2498}$ = **0.65**

Since the observed t = 0.65 < $t_{.025,3}$ = 3.182, the decision is to **fail to reject the null hypothesis**.

14.34 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{7.376}{\sqrt{58{,}293 - \dfrac{(571)^2}{7}}}$ = .068145

b_1 = -0.898

H_o: $\beta = 0$ α = .01

H_a: $\beta \neq 0$

Two-tail test, $\alpha/2$ = .005 df = n - 2 = 7 - 2 = 5

$t_{.005,5}$ = ±4.032

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{-0.898 - 0}{.068145}$ = **-13.18**

Since the observed t = -13.18 < $t_{.005,5}$ = -4.032, the decision is to **reject the null hypothesis**.

14.35 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{108.7}{\sqrt{7,667.15 - \dfrac{(199.5)^2}{8}}} = 2.095$

$b_1 = 15.240$

H_o: $\beta = 0$ $\alpha = .10$

H_a: $\beta \neq 0$

For a two-tail test, $\alpha/2 = .05$ df = n - 2 = 8 - 2 = 6

$t_{.05,6} = 1.943$

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{15,240 - 0}{2.095} =$ **7.27**

Since the observed $t = 7.27 > t_{.05,6} = 1.943$, the decision is to **reject the null hypothesis**.

14.36 $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{2.575}{\sqrt{421 - \dfrac{(41)^2}{5}}} = .27963$

$b_1 = -0.715$

H_o: $\beta = 0$ $\alpha = .05$

H_a: $\beta \neq 0$

For a two-tail test, $\alpha/2 = .025$ df = $n - 2 = 5 - 2 = 3$

$t_{.025,3} = \pm 3.182$

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{-0.715 - 0}{.27963} =$ **-2.56**

Since the observed $t = -2.56 > t_{.025,3} = -3.182$, the decision is to **fail to reject the**

null hypothesis.

14.37 s_b = $\dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{2.3095}{\sqrt{19{,}774.78 - \dfrac{(344.4)^2}{6}}}$ = 0.926025

b_1 = 0.878

H_o: $\beta = 0$ α = .05

H_a: $\beta \neq 0$

For a two-tail test, $\alpha/2$ = .025 df = n - 2 = 6 - 2 = 4

$t_{.025,4}$ = ±2.776

t = $\dfrac{b_1 - \beta_1}{s_b} = \dfrac{0.878 - 0}{.926025}$ = **0.948**

Since the observed t = 0.948 < $t_{.025,4}$ = 2.776, the decision is to **fail to reject the null hypothesis**.

14.38 F = 8.26 with a p-value of .021. The overall model is significant at α = .05 but not at $\alpha = .01$. For simple regression,

$$t = \sqrt{F} = 2.874$$

$t_{.05,8} = 1.86$ but $t_{.01,8} = 2.896$. The slope is significant at $\alpha = .05$ but not at

$\alpha = .01$.

14.39 $x_0 = 25$

95% confidence $\alpha/2 = .025$

df = $n - 2 = 5 - 2 = 3$ $t_{.025,3} = \pm 3.182$

$$\bar{x} = \frac{\sum x}{n} = \frac{89}{5} = 17.8$$

$\Sigma x = 89$ $\Sigma x^2 = 1{,}833$

$s_e = 3.94$

$$\hat{y} = 16.5 + 0.162(25) = 20.55$$

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$20.55 \pm 3.182(3.94)\sqrt{\frac{1}{5} + \frac{(25 - 17.8)^2}{1{,}833 - \frac{(89)^2}{5}}} = 20.55 \pm 3.182(3.94)(.63903) =$$

20.55 ± 8.01

$$12.54 \leq E(y_{25}) \leq 28.56$$

14.40 $x_0 = 100$ For 90% confidence, $\alpha/2 = .05$

df = n - 2 = 7 - 2 = 5 $t_{.05,5} = \pm 2.015$

$$\bar{x} = \frac{\sum x}{n} = \frac{571}{7}$$
= 81.57143

$\Sigma x = 571$ $\Sigma x^2 = 58{,}293$ $s_e = 7.377$

$\hat{y}$ = 144.414 - .0898(100) = 54.614

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x^2-\frac{(\sum x)^2}{n}}} =$$

$$54.614 \pm 2.015(7.377)\sqrt{1+\frac{1}{7}+\frac{(100-81.57143)^2}{58{,}293-\frac{(571)^2}{7}}} =$$

54.614 ± 2.015(7.377)(1.08252) = 54.614 ± 16.091

38.523 ≤ y ≤ 70.705

For x_0 = 130, $\hat{y}$ = 144.414 - 0.898(130) = 27.674

$$y \pm t_{/2,n-2}\, s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum x^2-\frac{(\sum x)^2}{n}}} =$$

$$27.674 \pm 2.015(7.377)\sqrt{1+\frac{1}{7}+\frac{(130-81.57143)^2}{58{,}293-\frac{(571)^2}{7}}} =$$

27.674 ± 2.015(7.377)(1.1589) = 27.674 ± 17.227

10.447 ≤ *y* ≤ 44.901

The width of this confidence interval of *y* for x_0 = 130 is <u>wider</u> that the confidence interval of *y* for x_0 = 100 because x_0 = 100 is nearer to the value of

x = 81.57 than is x_0 = 130.

14.41 x_0 = 20 For 98% confidence, $\alpha/2$ = .01

df = *n* - 2 = 8 - 2 = 6 $t_{.01,6}$ = 3.143

$$\bar{x}=\frac{\sum x}{n}=\frac{199.5}{8} = 24.9375$$

$\Sigma x = 199.5$ $\Sigma x^2 = 7{,}667.15$ $s_e = 108.8$

$\hat{y} = -46.29 + 15.24(20) = 258.51$

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$258.51 \pm (3.143)(108.8)\sqrt{\frac{1}{8} + \frac{(20 - 24.9375)^2}{7{,}667.15 - \frac{(199.5)^2}{8}}}$$

$258.51 \pm (3.143)(108.8)(0.36614) = 258.51 \pm 125.20$

$133.31 \le E(y_{20}) \le 383.71$

For single *y* value:

$$\hat{y} \pm t_{/2,n-2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$258.51 \pm (3.143)(108.8)\sqrt{1+\frac{1}{8}+\frac{(20-24.9375)^2}{7{,}667.15-\frac{(199.5)^2}{8}}}$$

258.51 ± (3.143)(108.8)(1.06492) = 258.51 ± 364.16

-105.65 ≤ *y* ≤ 622.67

The confidence interval for the single value of *y* is wider than the confidence interval for the average value of *y* because the average is more towards the middle and individual values of *y* can vary more than values of the average.

14.42 $x_0 = 10$ For 99% confidence $\alpha/2 = .005$

df = n - 2 = 5 - 2 = 3 $t_{.005,3} = 5.841$

$$\bar{x} = \frac{\sum x}{n} = \frac{41}{5}$$ = 8.20

$\Sigma x = 41$ $\Sigma x^2 = 421$ $s_e = 2.575$

$\hat{y}$ = 15.46 - 0.715(10) = 8.31

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$8.31 \pm 5.841(2.575)\sqrt{\frac{1}{5} + \frac{(10 - 8.2)^2}{421 - \frac{(41)^2}{5}}} =$$

8.31 ± 5.841(2.575)(.488065) = 8.31 ± 7.34

0.97 ≤ $E(y_{10})$ ≤ 15.65

If the prime interest rate is 10%, we are 99% confident that the average bond rate

is between 0.97% and 15.65%.

14.43

Year	Fertilizer
2001	11.9
2002	17.9
2003	22.0
2004	21.8
2005	26.0

$\Sigma x = 10{,}015$ $\Sigma y = 99.6$ $\Sigma xy = 199{,}530.9$

$\Sigma x^2 = 20{,}060{,}055$ $\Sigma y^2 = 2097.26$ $n = 5$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{199{,}530.9 - \frac{(10{,}015)(99.6)}{5}}{20{,}060{,}055 - \frac{(10{,}015)^2}{5}} = 3.21$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{99.6}{5} - 3.21\frac{10{,}015}{5} = -6{,}409.71$$

$\hat{y}$ = **-6,409.71 + 3.21 *x***

$\hat{y}$(2008) = -6,409.71 + 3.21(2008) = **35.97**

14.44

Year	Fertilizer
1998	5860
1999	6632
2000	7125
2001	6000
2002	4380
2003	3326
2004	2642

$\Sigma x = 14{,}007$ $\Sigma y = 35{,}965$ $\Sigma xy = 71{,}946{,}954$

$\Sigma x^2 = 28{,}028{,}035$ $\Sigma y^2 = 202{,}315{,}489$ $n = 7$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{71{,}946{,}954 - \frac{(14{,}007)(35{,}965)}{7}}{28{,}028{,}035 - \frac{(14{,}007)^2}{7}} = -678.9643$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{35{,}965}{7} - -678.9643 \frac{14{,}007}{7} = 1{,}363{,}745.39$$

$\hat{y}$ = **1,363,745.39 + -678.9643 *x***

$\hat{y}$(2007) = 1,363,745.39 + -678.9643(2007) = **1,064.04**

14.45

Year	Quarter	Cum. Quarter(x)	Sales(y)
2003	1	1	11.93
	2	2	12.46
	3	3	13.28
	4	4	15.08
2004	1	5	16.08
	2	6	16.82
	3	7	17.60
	4	8	18.66
2005	1	9	19.73
	2	10	21.11
	3	11	22.21
	4	12	22.94

Use the cumulative quarters as the predictor variable, x, to predict sales, y.

$\Sigma x = 78$ $\Sigma y = 207.9$ $\Sigma xy = 1{,}499.07$

$\Sigma x^2 = 650$ $\Sigma y^2 = 3{,}755.2084$ $n = 12$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{1{,}499.07 - \frac{(78)(207.9)}{12}}{650 - \frac{(78)^2}{12}} = 1.033$$

$$b_0 = \frac{\sum y}{n} - b_1\frac{\sum x}{n} = \frac{207.9}{12} - 1.033\frac{78}{12} = 10.6105$$

$$\hat{y} = \mathbf{10.6105 + 1.033\,x}$$

Remember, this trend line was constructed using cumulative quarters. To forecast

sales for the third quarter of year 2007, we must convert this time frame to

cumulative quarters. The third quarter of year 2007 is quarter number 19 in our

scheme.

$\hat{y}(19) = 10.6105 + 1.033(19) = \mathbf{30.2375}$

14.46

x	y
5	8
7	9
3	11
16	27
12	15
9	13

$\Sigma x = 52$ $\Sigma x^2 = 564$

$\Sigma y = 83$ $\Sigma y^2 = 1{,}389$ $b_1 = 1.2853$

$\Sigma xy = 865$ $n = 6$ $b_0 = 2.6941$

a) $\hat{y}$ **= 2.6941 + 1.2853 *x***

b)

$\hat{y}$ (Predicted Values)	(y-$\hat{y}$) residuals
9.1206	**-1.1206**
11.6912	**-2.6912**
6.5500	**4.4500**
23.2588	**3.7412**
18.1177	**-3.1176**
14.2618	**-1.2618**

c) $(y-\hat{y})^2$

1.2557
7.2426
19.8025
13.9966
9.7194
1.5921

SSE = 53.6089

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{53.6089}{4}}$$ = **3.661**

d) $r^2 = 1-\dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}} = 1-\dfrac{53.6089}{1{,}389-\dfrac{(83)^2}{6}}$ = **.777**

e) H$_o$: $\beta = 0$ $\alpha = .01$

H$_a$: $\beta \neq 0$

Two-tailed test, $\alpha/2 = .005$ df = n - 2 = 6 - 2 = 4

$t_{.005,4} = \pm 4.604$

$$s_b = \frac{s_e}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \frac{3.661}{\sqrt{564 - \frac{(52)^2}{6}}} = .34389$$

$$t = \frac{b_1 - \beta_1}{s_b} = \frac{1.2853 - 0}{.34389} = \mathbf{3.74}$$

Since the observed t = 3.74 < $t_{.005,4}$ = 4.604, the decision is to **fail to reject the null hypothesis**.

f) The r^2 = 77.74% is modest. There appears to be some prediction with this model. The slope of the regression line is not significantly different from zero using α = .01. However, for α = .05, the null hypothesis of a zero

slope is rejected. The standard error of the estimate, s_e = 3.661 is not

particularly small given the range of values for *y* (11 - 3 = 8).

14.47

x	*y*
53	5
47	5
41	7
50	4
58	10
62	12
45	3
60	11

Σx = 416 Σx^2 = 22,032

Σy = 57 Σy^2 = 489 b_1 = 0.355

Σxy = 3,106 n = 8 b_0 = -11.335

a) $\hat{y}$ **= -11.335 + 0.355 *x***

b)

$\hat{y}$ (Predicted Values)	$(y-\hat{y})$ residuals
7.48	**-2.48**
5.35	**-0.35**
3.22	**3.78**
6.415	**-2.415**
9.255	**0.745**
10.675	**1.325**
4.64	**-1.64**
9.965	**1.035**

c)

$(y-\hat{y})^2$
6.1504
0.1225
14.2884
5.8322
0.5550
1.7556
2.6896
1.0712

SSE = **32.4649**

d) $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{32.4649}{6}}$ = **2.3261**

e) $r^2 = \dfrac{1 - \dfrac{SSE}{\sum y^2 - \dfrac{(\sum y)^2}{n}}}{} = 1 - \dfrac{32.4649}{489 - \dfrac{(57)^2}{8}}$ = **.608**

f) H_o: $\beta = 0$ $\alpha = .05$

H_a: $\beta \neq 0$

Two-tailed test, $\alpha/2 = .025$ df = n - 2 = 8 - 2 = 6

$t_{.025,6} = \pm 2.447$

$s_b = \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}} = \dfrac{2.3261}{\sqrt{22{,}032 - \dfrac{(416)^2}{8}}}$ = 0.116305

$t = \dfrac{b_1 - \beta_1}{s_b} = \dfrac{0.3555 - 0}{.116305}$ = **3.05**

Since the observed $t = 3.05 > t_{.025,6} = 2.447$, the decision is to **reject the null hypothesis**.

The population slope is different from zero.

g) This model produces only a modest r^2 = .608. Almost 40% of the

variance of y is unaccounted for by x. The range of y values is 12 - 3 = 9

and the standard error of the estimate is 2.33. Given this small range, the

s_e is not small.

14.48 Σx = 1,263 Σx^2 = 268,295

Σy = 417 Σy^2 = 29,135

Σxy = 88,288 n = 6

b_0 = 25.42778 b_1 = 0.209369

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ =

29,135 - (25.42778)(417) - (0.209369)(88,288) = 46.845468

$$r^2 = 1-\frac{SSE}{\sum y^2-\frac{(\sum y)^2}{n}} = 1-\frac{46.845468}{153.5} = \mathbf{.695}$$

Coefficient of determination = r^2 = **.695**

14.49a) $x_0 = 60$

$\Sigma x = 524$ $\Sigma x^2 = 36,224$

$\Sigma y = 215$ $\Sigma y^2 = 6,411$ $b_1 = .5481$

$\Sigma xy = 15,125$ $n = 8$ $b_0 =$ -9.026

$s_e = 3.201$ 95% Confidence Interval $\alpha/2 = .025$

df = n - 2 = 8 - 2 = 6

$t_{.025,6} = \pm 2.447$

$\hat{y}$ = -9.026 + 0.5481(60) = 23.86

$\bar{x} = \frac{\sum x}{n} = \frac{524}{8}$ = 65.5

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$23.86 \pm 2.447(3.201)\sqrt{\frac{1}{8} + \frac{(60-65.5)^2}{36{,}224 - \frac{(524)^2}{8}}}$$

$23.86 \pm 2.447(3.201)(.375372) = 23.86 \pm 2.94$

$20.92 \le E(y_{60}) \le 26.8$

b) $x_0 = 70$

$\hat{y}_{70} = -9.026 + 0.5481(70) = 29.341$

$$\hat{y} \pm t_{\alpha/2,n-2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$29.341 \pm 2.447(3.201)\sqrt{1 + \frac{1}{8} + \frac{(70-65.5)^2}{36{,}224 - \frac{(524)^2}{8}}}$$

$29.341 \pm 2.447(3.201)(1.06567) = 29.341 \pm 8.347$

$\mathbf{20.994 \le y \le 37.688}$

c) The confidence interval for (b) is much wider because part (b) is for a single value of *y* which produces a much greater possible variation. In actuality, $x_0 = 70$ in part (b) is slightly closer to the mean (x) than $x_0 = 60$. However, the width of the single interval is much greater than that of the average or expected *y* value in part (a).

14.50

Year	Cost
1	56
2	54
3	49
4	46
5	45

$\Sigma x = 15$ $\Sigma y = 250$ $\Sigma xy = 720$

$\Sigma x^2 = 55$ $\Sigma y^2 = 12{,}594$ $n = 5$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{720 - \frac{(15)(250)}{5}}{55 - \frac{(15)^2}{5}} = -3$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{250}{5} - (-3)\frac{15}{5} = 59$$

$\hat{y}$ = **59 - 3 *x***

$\hat{y}(7)$ = 59 - 3(7) = **38**

14.51 $\Sigma y = 267$ $\Sigma y^2 = 15,971$

$\Sigma x = 21$ $\Sigma x^2 = 101$

$\Sigma xy = 1,256$ $n = 5$

$b_0 = 9.234375$ $b_1 = 10.515625$

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ =

15,971 - (9.234375)(267) - (10.515625)(1,256) = 297.7969

$$r^2 = 1-\frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1-\frac{297.7969}{1,713.2} = \mathbf{.826}$$

If a regression model would have been developed to predict number of cars sold by the number of sales people, the model would have had an r^2 of 82.6%. The same would hold true for a model to predict number of sales people by the number of cars sold.

14.52 $n = 12$ $\Sigma x = 548$ $\Sigma x^2 = 26,592$

$\Sigma y = 5940$ $\Sigma y^2 = 3,211,546$ $\Sigma xy = 287,908$

b_1 = 10.626383 b_0 = 9.728511

$\hat{y}$ **= 9.728511 + 10.626383 *x***

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ =

3,211,546 - (9.728511)(5940) - (10.626383)(287,908) = 94337.9762

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{94,337.9762}{10}}$$ = **97.1277**

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{94,337.9762}{271,246}$$ = **.652**

$$t = \frac{10.626383 - 0}{\frac{97.1277}{\sqrt{26,592 - \frac{(548)^2}{12}}}}$$ = **4.33**

If α = .01, then $t_{.005,10}$ = 3.169. Since the observed t = 4.33 > $t_{.005,10}$ = 3.169, the

decision is to **reject the null hypothesis**.

14.53

Sales(y)	Number of Units(x)
17.1	12.4
7.9	7.5
4.8	6.8
4.7	8.7
4.6	4.6
4.0	5.1
2.9	11.2
2.7	5.1
2.7	2.9

$\Sigma y = 51.4$ $\Sigma y^2 = 460.1$ $\Sigma x = 64.3$

$\Sigma x^2 = 538.97$ $\Sigma xy = 440.46$ $n = 9$

$b_1 = 0.92025$ $b_0 = -0.863565$

$$\hat{y} = \mathbf{-0.863565 + 0.92025}\ x$$

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy =$

$460.1 - (-0.863565)(51.4) - (0.92025)(440.46) = 99.153926$

$$r^2 = 1-\frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1-\frac{99.153926}{166.55} = \mathbf{.405}$$

14.54

Year	Total Employment
1995	11,152
1996	10,935
1997	11,050
1998	10,845
1999	10,776
2000	10,764
2001	10,697
2002	9,234
2003	9,223
2004	9,158

$\Sigma x = 19{,}995$ $\Sigma y = 103{,}834$ $\Sigma xy = 207{,}596{,}350$

$\Sigma x^2 = 39{,}980{,}085$ $\Sigma y^2 = 1{,}084{,}268{,}984$ $n = 7$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{207{,}596{,}350 - \frac{(19{,}995)(103{,}834)}{10}}{39{,}980{,}085 - \frac{(19{,}995)^2}{10}} = -239.188$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{103{,}834}{10} - (-239.188)\frac{19{,}995}{10} = 488{,}639.564$$

$\hat{y}$ **= 488,639.564 + -239.188 *x***

$\hat{y}$(2008) = 488,639.564 + -239.188(2008) = **8,350.30**

14.55

	1977	2003
	581	666
	213	214
	668	496
	345	204
	1476	1600
	1776	6278

$\Sigma x = 5059$ $\quad \Sigma y = 9458$ $\quad \Sigma x^2 = 6{,}280{,}931$

$\Sigma y^2 = 42{,}750{,}268$ $\quad \Sigma xy = 14{,}345{,}564$ $\quad n = 6$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{13{,}593{,}272 - \frac{(5059)(9358)}{6}}{6{,}280{,}931 - \frac{(5059)^2}{6}} = 3.1612$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{9458}{6} - (3.1612)\frac{5059}{6} = -1089.0712$$

$$\hat{y} = \mathbf{-1089.0712 + 3.1612}\ \boldsymbol{x}$$

for x = 700:

$\hat{y}$ = 1076.6044

$$\hat{y} \pm t_{\alpha/2,n-2} s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

α = .05, $t_{.025,4}$ = 2.776

x_0 = 700, n = 6

$\bar{x}$ = 843.167

SSE = $\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy$ =

42,750,268 – (-1089.0712)(9458) – (3.1612)(14,345,564) = 7,701,506.49

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{7,701,506.49}{4}} = 1387.58$$

Confidence Interval =

$$1123.757 \pm (2.776)(1387.58)\sqrt{\frac{1}{6}+\frac{(700-843.167)^2}{6{,}280{,}931-\frac{(5059)^2}{6}}} =$$

1123.757 ± 1619.81

-496.05 to 2743.57

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

$\alpha = .05$ df = 4

Table $t_{.025,4} = 2.776$

$$t = \frac{b_1 - 0}{s_b} = \frac{3.1612 - 0}{\frac{1387.58}{\sqrt{2{,}015{,}350.833}}} = \frac{2.9736}{.8231614} = \mathbf{3.234}$$

Since the observed $t = 3.234 > t_{.025,4} = 2.776$, the decision is to **reject the null hypothesis**.

14.56 $\Sigma x = 11.902$ $\Sigma x^2 = 25.1215$

$\Sigma y = 516.8$ $\Sigma y^2 = 61{,}899.06$ $b_1 = 66.36277$

Σxy = 1,202.867 n = 7 b_0 = -39.0071

$\hat{y}$ = -39.0071 + 66.36277 x

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$

SSE = 61,899.06 - (-39.0071)(516.8) - (66.36277)(1,202.867) = 2,232.343

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{2,232.343}{5}}$$ = **21.13**

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{2,232.343}{61,899.06 - \frac{(516.8)^2}{7}}$$ = 1 - .094 = **.906**

14.57 Σx = 44,754 Σy = 17,314 Σx^2 = 167,540,610

Σy^2 = 24,646,062 n = 13 Σxy = 59,852,571

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{59,852,571 - \frac{(44,754)(17,314)}{13}}{167,540,610 - \frac{(44,754)^2}{13}}$$ = .01835

$$b_0 = \frac{\sum y}{n} - b_1\frac{\sum x}{n} = \frac{17{,}314}{13} - (.01835)\frac{44{,}754}{13} = 1268.685$$

$\hat{y}$ = **1268.685 + .01835 *x***

r^2 for this model is .002858. There is **no predictability** in this model.

Test for slope: t = 0.18 with a p-value of 0.8623. Not significant

Time-Series Trend Line:

$\Sigma x = 91$ $\Sigma y = 44{,}754$ $\Sigma xy = 304{,}797$

$\Sigma x^2 = 819$ $\Sigma y^2 = 167{,}540{,}610$ $n = 13$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{304{,}797 - \frac{(91)(44{,}754)}{13}}{819 - \frac{(91)^2}{13}} = -46.5989$$

$$b_0 = \frac{\sum y}{n} - b_1\frac{\sum x}{n} = \frac{44{,}754}{13} - (-46.5989)\frac{91}{13} = 3{,}768.81$$

$\hat{y}$ = **3,768.81 – 46.5989 *x***

$\hat{y}(2007) = 3{,}768.81 - 46.5989(15) =$ **3,069.83**

14.58 $\Sigma x = 323.3$ $\Sigma y = 6765.8$

$\Sigma x^2 = 29,629.13$ $\Sigma y^2 = 7,583,144.64$

$\Sigma xy = 339,342.76$ $n = 7$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{339,342.76 - \frac{(323.3)(6765.8)}{7}}{29,629.13 - \frac{(323.3)^2}{7}} = 1.82751$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{6765.8}{7} - (1.82751)\frac{323.3}{7} = 882.138$$

$\hat{y}$ = **882.138 + 1.82751 *x***

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$

= 7,583,144.64 –(882.138)(6765.8) –(1.82751)(339,342.76) = 994,623.07

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{994,623.07}{5}}$$ = **446.01**

r^2 = $1-\dfrac{SSE}{\sum y^2-\dfrac{(\sum y)^2}{n}}=1-\dfrac{994,623.07}{7,583,144.64-\dfrac{(6765.8)^2}{7}}$ = 1 - .953 = **.047**

H$_0$: $\beta = 0$

H$_a$: $\beta \neq 0$ α = .05 $t_{.025,5}$ = 2.571

SS$_{xx}$ = $\sum x^2-\dfrac{\left(\sum x\right)^2}{n}=29,629.13-\dfrac{(323.3)^2}{7}$ = 14,697.29

t = $\dfrac{b_1-0}{\dfrac{s_e}{\sqrt{SS_{xx}}}}=\dfrac{1.82751-0}{\dfrac{446.01}{\sqrt{14,697.29}}}$ = **0.50**

Since the observed t = 0.50 < $t_{.025,5}$ = 2.571, the decision is to **fail to reject the**

null hypothesis.

14.59 Let Water use = y and Temperature = x

Σx = 608 Σx^2 = 49,584

Σy = 1,025 Σy^2 = 152,711 b_1 = 2.40107

Σxy = 86,006 n = 8 b_0 = -54.35604

$\hat{y}$ = **-54.35604 + 2.40107 *x***

$\hat{y}_{100}$ = -54.35604 + 2.40107(100) = **185.751**

SSE = $\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$

SSE = 152,711 - (-54.35604)(1,025) - (2.40107)(86,006) = 1919.5146

$$s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{1,919.5146}{6}}$$ = **17.886**

$$r^2 = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}} = 1 - \frac{1,919.5145}{152,711 - \frac{(1025)^2}{8}}$$ = 1 - .09 = **.91**

Testing the slope:

H_o: $\beta = 0$

H_a: $\beta \neq 0$ $\alpha = .01$

Since this is a two-tailed test, $\alpha/2 = .005$

df = n - 2 = 8 - 2 = 6

$t_{.005,6} = \pm 3.707$

$$s_b = \frac{s_e}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \frac{17.886}{\sqrt{49{,}584 - \frac{(608)^2}{8}}} = .30783$$

$$t = \frac{b_1 - \beta_1}{s_b} = \frac{2.40107 - 0}{.30783} = \mathbf{7.80}$$

Since the observed t = 7.80 < $t_{.005,6}$ = 3.707, the decision is to **reject the null**

hypothesis.

14.60 a) The regression equation is: $\hat{y}$ = **67.2 – 0.0565 *x***

b) For every unit of increase in the value of x, the predicted value of y will decrease by -.0565.

c) The t ratio for the slope is –5.50 with an associated p-value of .000. This is significant at α = .10. The t ratio negative because the slope is negative and the numerator of the t ratio formula equals the slope minus zero.

d) r^2 is .627 or 62.7% of the variability of y is accounted for by x. This is only a modest proportion of predictability. The standard error of the estimate is 10.32. This is best interpreted in light of the data and the magnitude of the data.

e) The F value which tests the overall predictability of the model is 30.25. For simple regression analysis, this equals the value of t^2 which is $(-5.50)^2$.

f) The negative is not a surprise because the slope of the regression line is also negative indicating an inverse relationship between x and y. In addition, taking the square root of r^2 which is .627 yields .7906 which is the magnitude of the value of r considering rounding error.

14.61 The F value for overall predictability is 7.12 with an associated p-value of .0205 which is significant at α = .05. It is not significant at alpha of .01. The coefficient of determination is .372 with an adjusted r^2 of .32. This represents very modest predictability. The standard error of the estimate is 982.219, which in units of 1,000 laborers means that about 68% of the predictions are within 982,219 of the actual figures. The regression model is:

Number of Union Members = 22,348.97 - 0.0524 Labor Force. For a labor force of 100,000 (thousand, actually 100 million), substitute x = 100,000 and get a predicted value of 17,108.97 (thousand) which is actually 17,108,970 union members.

14.62 The Residual Model Diagnostics from MINITAB indicate a relatively healthy set of residuals. The Histogram indicates that the error terms are generally normally distributed. This is somewhat confirmed by the semi straight line Normal Plot of Residuals. However, the Residuals vs. Fits graph indicates that there may be some heteroscedasticity with greater error variance for small x values.

Chapter 15

Multiple Regression Analysis

LEARNING OBJECTIVES

This chapter presents the potential of multiple regression analysis as a tool in business decision making and its applications, thereby enabling you to:

1. Develop a multiple regression model.
2. Understand and apply significance tests of the regression model and its coefficients.

3. Compute and interpret residuals, the standard error of the estimate, and the coefficient of determination.

4. Interpret multiple regression computer output.

CHAPTER TEACHING STRATEGY

In chapter 14 using simple regression, the groundwork was prepared for chapter 15 by presenting the regression model along with mechanisms for testing the strength of the model such as s_e, r^2, a t test of the slope, and the residuals. In this chapter, multiple regression is presented as an extension of the simple linear regression case. It is initially pointed out that any model that has at least one interaction term or a variable that represents a power of two or more is considered a multiple regression model. Multiple regression opens up the possibilities of predicting by multiple independent variables and nonlinear relationships. It is emphasized in the chapter that with both simple and multiple regression models there is only one dependent variable. Where simple regression utilizes only one independent variable, multiple regression can utilize more than one independent variable.

Presented early in chapter 15 are the simultaneous equations that need to be solved to develop a first-order multiple regression model using two predictors. This should help the student to see that there are three equations with three unknowns to be solved. In addition, there are eight values that need to be determined before solving the simultaneous equations (x_1, x_2, y, x_1^2, . . .) Suppose there are five predictors. Six simultaneous equations must be solved and the number of sums needed as constants in the equations become overwhelming. At this point, the student will begin to realize that most researchers do not want to take the time nor the effort to solve for multiple regression models by hand. For this reason, much of the chapter is presented using computer printouts. The assumption is that the use of multiple regression analysis is largely from computer analysis.

Topics included in this chapter are similar to the ones in chapter 14 including tests of the slope, R^2, and s_e. In addition, an adjusted R^2 is introduced in chapter 15. The adjusted R^2 takes into account the degrees of freedom error and total degrees of freedom whereas R^2 does not. If there is a significant discrepancy between adjusted R^2 and R^2, then the regression model may not be as strong as it appears to be with the R^2. The gap between R^2 and adjusted R^2 tends to increase as non significant independent variables are added to the regression model and decreases with increased sample size.

CHAPTER OUTLINE

KEY TERMS

Adjusted R^2

Coefficient of Multiple Determination (R^2)

Dependent Variable

Independent Variable

Least Squares Analysis

Multiple Regression

Outliers

Partial Regression Coefficient

R^2

Residual

Response Plane

Response Surface

Response Variable

Standard Error of the Estimate

Sum of Squares of Error

SOLUTIONS TO PROBLEMS IN CHAPTER 15

15.1 The regression model is:

$$\hat{y} = 25.03 - 0.0497\, x_1 + 1.928\, x_2$$

Predicted value of y for $x_1 = 200$ and $x_2 = 7$ is:

$\hat{y}$ = 25.03 - 0.0497(200) + 1.928(7) = **28.586**

15.2 The regression model is:

$\hat{y}$ **= 118.56 - 0.0794 x_1 - 0.88428 x_2 + 0.3769 x_3**

Predicted value of y for $x_1 = 33$, $x_2 = 29$, and $x_3 = 13$ is:

$\hat{y}$ = 118.56 - 0.0794(33) - 0.88428(29) + 0.3769(13) = **95.19538**

15.3 The regression model is:

$$\hat{y} = 121.62 - 0.174\, x_1 + 6.02\, x_2 + 0.00026\, x_3 + 0.0041\, x_4$$

There are **four** independent variables. If x_2, x_3, and x_4 are held constant, the predicted y will decrease by - 0.174 for every unit increase in x_1. Predicted y will increase by 6.02 for every unit increase in x_2 as x_1, x_3, and x_4 are held constant. Predicted y will increase by 0.00026 for every unit increase in x_3 holding x_1, x_2, and x_4 constant. If x_4 is increased by one unit, the predicted y will increase by 0.0041 if x_1, x_2, and x_3 are held constant.

15.4 The regression model is:

$$\hat{y} = 31{,}409.5 + 0.08425\, x_1 + 289.62\, x_2 - 0.0947\, x_3$$

For every unit increase in x_1, the predicted y increases by 0.08425 if x_2 and x_3 are held constant. The predicted y will increase by 289.62 for every unit increase in x_2 if x_1 and x_3 are held constant. The predicted y will decrease by 0.0947 for every unit increase in x_3 if x_1 and x_2 are held constant.

15.5 The regression model is:

Per Capita = -7,629.627 + 116.2549 Paper Consumption

- 120.0904 Fish Consumption + 45.73328 Gasoline Consumption.

For every unit increase in paper consumption, the predicted per capita consumption increases by 116.2549 if fish and gasoline consumption are held constant. For every unit increase in fish consumption, the predicted per capita consumption decreases by 120.0904 if paper and gasoline consumption are held constant. For every unit increase in gasoline consumption, the predicted per capita consumption increases by 45.73328 if paper and fish consumption are held constant.

15.6 The regression model is:

Insider Ownership =

17.68 - 0.0594 Debt Ratio - 0.118 Dividend Payout

The coefficients mean that for every unit of increase in debt ratio there is a predicted decrease of - 0.0594 in insider ownership if dividend payout is held constant. On the other hand, if dividend payout is increased by one unit, then there is a predicted drop of insider ownership by 0.118 with debt ratio is held constant.

15.7 There are 9 predictors in this model. The F test for overall significance of the model is 1.99 with a probability of .0825. This model is not significant at α = .05. Only one of the t values is statistically significant. Predictor x_1 has a t of 2.73 which has an associated probability of .011 and this is significant at α = .05.

15.8 This model contains three predictors. The F test is significant at α = .05 but not at α = .01. The t values indicate that only one of the three predictors is significant. Predictor x_1 yields a t value of 3.41 with an associated probability of .005. The recommendation is to rerun the model using only x_1 and then search for other variables besides x_2 and x_3 to include in future models.

15.9 The regression model is:

Per Capita Consumption = -7,629.627 + 116.2549 Paper Consumption

- 120.0904 Fish Consumption + 45.73328 Gasoline Consumption

This model yields an F = 14.319 with p-value = .0023. Thus, there is overall significance at α = .01. One of the three predictors is significant. Gasoline Consumption has a t = 2.67 with p-value of .032 which is statistically significant at α = .05. The p-values of the t statistics for the other two predictors are insignificant indicating that a model with just

Gasoline Consumption as a single predictor might be nearly as strong.

15.10 The regression model is:

Insider Ownership =

17.68 - 0.0594 Debt Ratio - 0.118 Dividend Payout

The overall value of *F* is only 0.02 with *p*-value of .982. This model is not significant. Neither of the *t* values are significant (t_{Debt} = -0.19 with a *p*-value of .855 and $t_{Dividend}$ = -0.11 with a *p*-value of .913).

15.11 The regression model is:

$$\hat{y} = \mathbf{3.981 + 0.07322}\ \boldsymbol{x_1} - \mathbf{0.03232}\ \boldsymbol{x_2} - \mathbf{0.003886}\ \boldsymbol{x_3}$$

The overall *F* for this model is 100.47 with is significant at $\alpha = .001$. Only one of the predictors, x_1, has a significant *t* value ($t = 3.50$, *p*-value of .005). The other independent variables have non significant *t* values

(x_2: $t = -1.55$, *p*-value of .15 and x_3: $t = -1.01$, *p*-value of .332). Since x_2 and x_3 are non significant predictors, the researcher should consider the using a simple regression model with only x_1 as a predictor. The R^2 would drop some but the model would be much more parsimonious.

15.12 The regression equation for the model using both x_1 and x_2 is:

$\hat{y}$

= 243.44 - 16.608 x_1 - 0.0732 x_2

The overall F = 156.89 with a p-value of .000. x_1 is a significant predictor of y as indicated by t = - 16.10 and a p-value of .000.

For x_2, t = -0.39 with a p-value of .702. x_2 is not a significant predictor of y when included with x_1. Since x_2 is not a significant predictor, the researcher might want to rerun the model using just x_1 as a predictor.

The regression model using only x_1 as a predictor is:

$\hat{\hat{Y}}$

= 235.143 - 16.7678 x_1

There is very little change in the coefficient of x_1 from model one

(2 predictors) to this model. The overall F = 335.47 with a p-value of .000 is highly significant. By using the one-predictor model, we get virtually the same predictability as with the two predictor model and it is more parsimonious.

15.13 There are **3 predictors in this model and 15 observations**.

The regression equation is:

$\hat{y}$

= 657.053 + 5.7103 x_1 – 0.4169 x_2 –3.4715 x_3

F = 8.96 with a p-value of .0027

x_1 is significant at $\alpha = .01$ (t = 3.19, p-value of .0087)

x_3 is significant at $\alpha = .05$ (t = - 2.41, p-value of .0349)

The model is significant overall.

15.14 The standard error of the estimate is 3.503. R^2 is .408 and the adjusted R^2 is only .203. This indicates that there are a lot of insignificant predictors in the model. That is underscored by the fact that eight of the nine predictors have non significant t values.

15.15 s_e = 9.722, R^2 = .515 but the adjusted R^2 is only .404. The difference in the two is due to the fact that two of the three predictors in the model are non-significant. The model fits the data only modestly. The adjusted R^2 indicates that 40.4% of the variance of y is accounted for by this model and 59.6% is unaccounted for by the model.

15.16 The standard error of the estimate of 14,660.57 indicates that this model predicts Per Capita Personal Consumption to within $\pm$ 14,660.57 about 68% of the time. The entire range of Personal Per Capita for the data is slightly less than 110,000. Relative to this range, the standard error of the estimate is modest. R^2 = .85988 and the adjusted value of R^2 is .799828 indicating that there are potentially some non significant variables in the model. An examination of the t statistics reveals that two of the three predictors are not significant. The model has relatively good predictability.

15.17 $s_e = 6.544$. $R^2 = .005$. This model has no predictability.

15.18 The value of s_e = 0.2331, R^2 = .965, and adjusted R^2 = .955. This is a

very strong regression model. However, since x_2 and x_3 are not significant predictors, the researcher should consider the using a simple regression model with only x_1 as a predictor. The R^2 would drop some but the model would be much more parsimonious.

15.19 For the regression equation for the model using both x_1 and x_2, s_e = 6.333,

R^2 = .963 and adjusted R^2 = .957. Overall, this is a very strong model. For the regression model using only x_1 as a predictor, the standard error of the estimate is 6.124, R^2 = .963 and the adjusted R^2 = .960. The value of R^2 is the same as it was with the two predictors. However, the adjusted R^2 is slightly higher with the one-predictor model because the non-significant variable has been removed. In conclusion, by using the one predictor model, we get virtually the same predictability as with the two predictor model and it is more parsimonious.

15.20 R^2 = .710, adjusted R^2 = .630, s_e = 109.43. The model is significant

overall. The R^2 is higher but the adjusted R^2 by 8%. The model is

moderately strong.

15.21 The Histogram indicates that there may be some problem with the error

terms being normally distributed. The Residuals vs. Fits plot reveals that there may be some lack of homogeneity of error variance.

15.22 There are four predictors. The equation of the regression model is:

$\hat{y}$ = -55.9 + 0.0105 x_1 – 0.107 x_2 + 0.579 x_3 – 0.870 x_4

The test for overall significance yields an F = 55.52 with a p-value of .000

which is significant at α = .001. Three of the t tests for regression coefficients are significant at α = .01 including the coefficients for

x_2, x_3, and x_4. The R^2 value of 80.2% indicates strong predictability for the model. The value of the adjusted R^2 (78.8%) is close to R^2 and s_e is 9.025.

15.23 There are two predictors in this model. The equation of the regression model is:

$\hat{y}$ = 203.3937 + 1.1151 x_1 – 2.2115 x_2

The F test for overall significance yields a value of 24.55 with an

associated p-value of .0000013 which is significant at α = .00001. Both variables yield t values that are significant at a 5% level of significance.

x_2 is significant at $\alpha = .001$. The R^2 is a rather modest 66.3% and the

standard error of the estimate is 51.761.

15.24 The regression model is:

$$\hat{y} = 137.27 + 0.0025\, x_1 + 29.206\, x_2$$

$F = 10.89$ with $p = .005$, $s_e = 9.401$, $R^2 = .731$, adjusted $R^2 = .664$. For x_1, $t = 0.01$ with $p = .99$ and for x_2, $t = 4.47$ with $p = .002$. This model has good predictability. The gap between R^2 and adjusted R^2 indicates that there may be a non-significant predictor in the model. The t values show x_1 has virtually no predictability and x_2 is a significant predictor of y.

15.25 The regression model is:

$$\hat{Y} = 362.3054 - 4.745518\, x_1 - 13.89972\, x_2 + 1.874297\, x_3$$

F = 16.05 with p = .001, s_e = 37.07, R^2 = .858, adjusted R^2 = .804. For x_1, t = -4.35 with p = .002; for x_2, t = -0.73 with p = .483, for x_3, t = 1.96 with p = .086. Thus, only one of the three predictors, x_1, is a significant predictor in this model. This model has very good predictability (R^2 = .858). The gap between R^2 and adjusted R^2 underscores the fact that there are two non-significant predictors in this model.

15.26 The overall F for this model was 12.19 with a p-value of .002 which is significant at $\alpha = .01$. The t test for Silver is significant at $\alpha = .01$ (t = 4.94, p = .001). The t test for Aluminum yields a t = 3.03 with a p-value of .016 which is significant at $\alpha = .05$. The t test for Copper was insignificant with a p-value of .939. The value of R^2 was 82.1% compared

to an adjusted R^2 of 75.3%. The gap between the two indicates the presence of some insignificant predictors (Copper). The standard error of the estimate is 53.44.

15.27 The regression model was:

Employment = 71.03 + 0.4620 NavalVessels + 0.02082 Commercial

F = 1.22 with p = .386 (not significant)

R^2 = .379 and adjusted R^2 = .068

The low value of adjusted R^2 indicates that the model has very low predictability. Both t values are not significant ($t_{\text{NavalVessels}}$ = 0.67 with

p = .541 and $t_{\text{Commercial}}$ = 1.07 with p = .345). Neither predictor is a significant predictor of employment.

15.28 The regression model was:

All = -1.06 + 0.475 Food + 0.250 Shelter – 0.008 Apparel +

0.272 Fuel Oil

F = 97.98 with a p-value of .000

s_e = 0.7472, R^2 = .963 and adjusted R^2 = .953

One of the predictor variables, Food, produces a t value that is significant at α = .001. Two others are significant at α = .05: Shelter (t = 2.48, p-value of .025 and Fuel Oil (t = 2.36 with a p-value of .032).

15.29 The regression model was:

Corn = -2718 + 6.26 Soybeans – 0.77 Wheat

F = 14.25 with a p-value of .003 which is significant at α = .01

s_e = 862.4, R^2 = 80.3%, adjusted R^2 = 74.6%

One of the two predictors, Soybeans, yielded a t value that was significant at α = .01 while the other predictor, Wheat was not significant (t = -0.75 with a p-value of .476).

15.30 The regression model was:

Grocery = 76.23 + 0.08592 Housing + 0.16767 Utility

+ 0.0284 Transportation - 0.0659 Healthcare

F = 2.29 with p = .095 which is not significant at α = .05.

s_e = 4.416, R^2 = .315, and adjusted R^2 = .177.

Only one of the four predictors has a significant t ratio and that is Utility with t = 2.57 and p = .018. The ratios and their respective probabilities are:

$t_{housing}$ = 1.68 with p = .109, $t_{transportation}$ = 0.17 with p = .87, and

$t_{\text{healthcare}}$ = - 0.64 with p = .53.

This model is very weak. Only the predictor, Utility, shows much promise in accounting for the grocery variability.

15.31 The regression equation is:

$$\hat{y} = 87.89 – 0.256\,x_1 – 2.714\,x_2 + 0.0706\,x_3$$

F = 47.57 with a p-value of .000 significant at α = .001.

s_e = 0.8503, R^2 = .941, adjusted R^2 = .921.

All three predictors produced significant t tests with two of them

(x_2 and x_3) significant at .01 and the other, x_1 significant at α = .05. This is

a very strong model.

15.32 Two of the diagnostic charts indicate that there may be a problem with the error terms being normally distributed. The histogram indicates that the error term distribution might be skewed to the right and the normal probability plot is somewhat nonlinear. In addition, the residuals vs. fits chart indicates a potential heteroscadasticity problem with residuals for middle values of x producing more variability that those for lower and higher values of x.

Chapter 16

Building Multiple Regression Models

LEARNING OBJECTIVES

This chapter presents several advanced topics in multiple regression analysis enabling you to:

1. Analyze and interpret nonlinear variables in multiple regression analysis.
2. Understand the role of qualitative variables and how to use them in multiple regression analysis.
3. Learn how to build and evaluate multiple regression models.
4. Detect influential observations in regression analysis.

CHAPTER TEACHING STRATEGY

In chapter 15, the groundwork was prepared for chapter 16 by presenting multiple regression models along with mechanisms for testing the strength of the models such as s_e, R^2, t tests of the regression coefficients, and the residuals.

The early portion of this chapter is devoted to nonlinear regression models and search procedures. There are other exotic types of regression models that can be explored. It is hoped that by studying section 16.1, the student will be somewhat prepared to explore other nonlinear models on his/her own. Tukey's Ladder of Transformations can be useful in steering the research towards particular recoding schemes that will result in better fits for the data.

Dummy or indicator variables can be useful in multiple regression analysis. Remember to emphasize that only one dummy variable is used to represent two categories (yes/no, male/female, union/nonunion, etc.). For *c* categories of a qualitative variable, only *c*-1 indicator variables should be included in the multiple regression model.

Several search procedures have been discussed in the chapter including stepwise regression, forward selection, backward elimination, and all possible regressions. All possible regressions is presented mainly to demonstrate to the student the large number of possible models that can be examined. Most of the effort is spent on stepwise regression because of its common usage. Forward selection is presented as the same as stepwise regression except that forward selection procedures do not go back and examine variables that have been in the model at each new step. That is, with forward selection, once a variable is in the model, it stays in the model. Backward elimination begins with a "full" model of all predictors. Sometimes there may not be enough observations to justify such a model.

CHAPTER OUTLINE

KEY TERMS

All Possible Regressions

Backward Elimination

Dummy Variable

Forward Selection

Indicator Variable

Multicollinearity

Quadratic Model

Qualitative Variable

Search Procedures

Stepwise Regression

Tukey's Four-quadrant Approach

Tukey's Ladder of Transformations

Variance Inflation Factor

SOLUTIONS TO PROBLEMS IN CHAPTER 16

16.1 Simple Regression Model:

$$\hat{y} = -147.27 + 27.128\,x$$

F = 229.67 with p = .000, s_e = 27.27, R^2 = .97, adjusted R^2 = .966, and

t = 15.15 with p = .000. This is a very strong simple regression model.

<u>Quadratic Model (Using both x and x^2)</u>:

$\hat{y}$ **= - 22.01 + 3.385 X + 0.9373 x^2**

F = 578.76 with p = .000, s_e = 12.3, R^2 = .995, adjusted R^2 = .993, for x:

t = 0.75 with p = .483, and for x^2: t = 5.33 with p = .002. The quadratic model is also very strong with an even higher R^2 value. However, in this model only the x^2 term is a significant predictor.

16.2 The model is:

$$\hat{y} = b_0 b_1^x$$

Using logs: $\log y = \log b_0 + x \log b_1$

The regression model is solved for in the computer using the values of x and the values of $\log y$. The resulting regression equation is:

$\log y = 0.5797 + 0.82096\ x$

$F = 68.83$ with $p = .000$, $s_e = 0.1261$, $R^2 = .852$, and adjusted $R^2 = .839$. This model has relatively strong predictability.

16.3 Simple regression model:

$\hat{\hat{Y}}$ **= - 1456.6 + 71.017 *x***

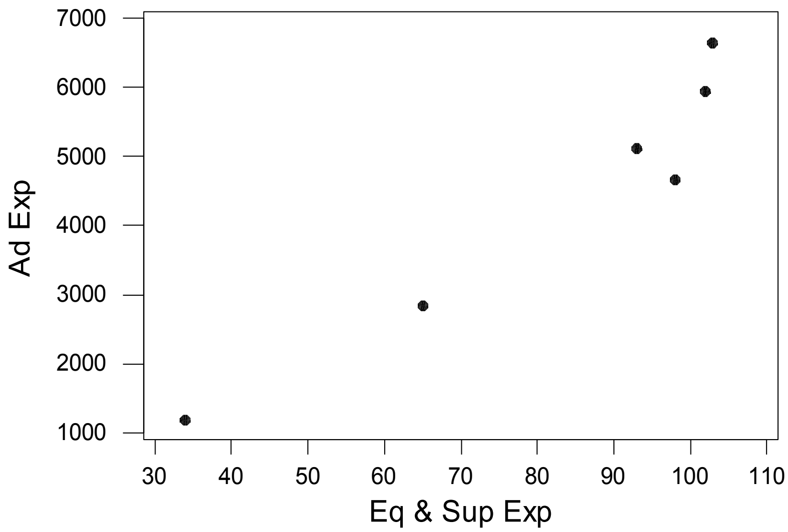
R^2 = .928 and adjusted R^2 = .910. t = 7.17 with p = .002.

Quadratic regression model:

$\hat{y}$ **= 1012 - 14.06 *x* + 0.6115 x^2**

R^2 = .947 but adjusted R^2 = .911. The t ratio for the x term is t = -0.17 with p = .876. The t ratio for the x^2 term is t = 1.03 with p = .377

Neither predictor is significant in the quadratic model. Also, the adjusted R^2 for this model is virtually identical to the simple regression model. The quadratic model adds virtually no predictability that the simple regression model does not already have. The scatter plot of the data follows:



16.4 The model is:

$$\hat{y} = b_0 b_1^x$$

Using logs: $\log y = \log b_0 + x \log b_1$

The regression model is solved for in the computer using the values of *x* and the values of log *y* where *x* is failures and *y* is liabilities. The resulting regression equation is:

log liabilities = 3.1256 + 0.012846 failures

F = 19.98 with p = .001, s_e = 0.2862, R^2 = .666, and adjusted R^2 = .633. This model has modest predictability.

16.5 The regression model is:

$\hat{y}$ **= - 28.61 - 2.68 x_1 + 18.25 x_2 - 0.2135 x_1^2 - 1.533 x_2^2 + 1.226 $x_1 * x_2$**

F = 63.43 with p = .000 significant at α = .001

s_e = 4.669, R^2 = .958, and adjusted R^2 = .943

None of the t ratios for this model are significant. They are $t(x_1)$ = - 0.25 with p = .805, $t(x_2)$ = 0.91 with p = .378, $t(x_1^2)$ = - 0.33 with .745,

$t(x_2^2)$ = - 0.68 with .506, and $t(x_1 * x_2)$ = 0.52 with p = .613. This model has a high R^2 yet none of the predictors are individually significant.

The same thing occurs when the interaction term is not in the model. None of the t tests are significant. The R^2 remains high at .957 indicating

that the loss of the interaction term was insignificant.

16.6 The F value shows very strong overall significance with a p-value of .00000073. This is reinforced by the high R^2 of .910 and adjusted R^2 of .878. An examination of the t values reveals that only one of the regression coefficients is significant at α = 05 and that is the interaction term with a p-value of .039. Thus, this model with both variables, the square of both variables, and the interaction term contains only one significant t test and that is for interaction.

Without interaction, the R^2 drops to .877 and adjusted R^2 to .844. With the interaction term removed, both variable x_2 and x_2^2 are significant at

$\alpha = .01$.

16.7 The regression equation is:

$\hat{y}$ **= 13.619 - 0.01201 x_1 + 2.998 x_2**

The overall $F = 8.43$ is significant at $\alpha = .01$ ($p = .009$).

$s_e = 1.245$, $R^2 = .652$, adjusted $R^2 = .575$

The t ratio for the x_1 variable is only $t = -0.14$ with $p = .893$. However the t ratio for the dummy variable, x_2 is $t = 3.88$ with $p = .004$. The indicator variable is the significant predictor in this regression model that has some predictability (adjusted $R^2 = .575$).

16.8 The indicator variable has $c = 4$ categories as shown by the $c - 1 = 3$ categories of the predictors (x_2, x_3, x_4).

The regression equation is:

$\hat{y}$ **= 7.909 + 0.581 x_1 + 1.458 x_2 - 5.881 x_3 - 4.108 x_4**

Overall $F = 13.54$, $p = .000$ significant at $\alpha = .001$

s_e = 1.733, R^2 = .806, and adjusted R^2 = .747

For the predictors, t = 0.56 with p = .585 for the x_1 variable (not significant), t = 1.32 with p = .208 for the first indicator variable (x_2) and is non significant, t = -5.32 with p = .000 for x_3 the second indicator variable and this is significant at α = .001, t = -3.21 with p = .007 for the third indicator variable (x_4) which is significant at α = .01. This model has strong predictability and the only significant predictor variables are the two dummy variables, x_3 and x_4.

16.9 This regression model has relatively strong predictability as indicated by R^2 = .795. Of the three predictor variables, only x_1 and x_2 have significant t ratios (using α = .05). x_3 (a non indicator variable) is not a significant predictor. x_1, the indicator variable, plays a significant role in this model along with x_2.

16.10 The regression model is:

$$\hat{y} = \mathbf{41.225 + 1.081\, x_1 - 18.404\, x_2}$$

$F = 8.23$ with $p = .0017$ which is significant at $\alpha = .01$. $s_e = 11.744$,

$R^2 = .388$ and the adjusted $R^2 = .341$.

The t-ratio for x_2 (the dummy variable) is –4.05 which has an associated

p-value of .0004 and is significant at $\alpha = .001$. The t-ratio of 0.80 for x_1

is not significant (p-value = .4316). With $x_2 = 0$, the regression model

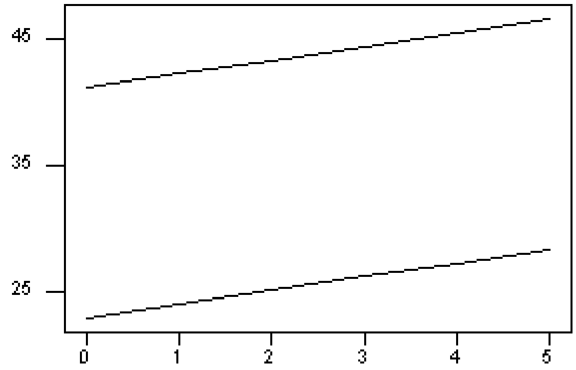
becomes $\hat{y} = 41.225 + 1.081x_1$. With $x_2 = 1$, the regression model

becomes $\hat{y} = 22.821 + 1.081x_1$. The presence of x_2 causes the y

intercept to drop by 18.404. The graph of each of these models (without

the dummy variable and with the dummy variable equal to one) is shown

below:



16.11 The regression equation is:

Price = 7.066 - 0.0855 Hours + 9.614 ProbSeat + 10.507 FQ

The overall F = 6.80 with p = .009 which is significant at α = .01. s_e = 4.02, R^2 = .671, and adjusted R^2 = .573. The difference between R^2 and adjusted R^2 indicates that there are some non-significant predictors in the model. The t ratios, t = - 0.56 with p = .587 and t = 1.37 with p = .202, of Hours and Probability of Being Seated are non-significant at α = .05. The only significant predictor is the dummy variable, French Quarter or not, which has a t ratio of 3.97 with p = .003 which is significant at α = .01. The positive coefficient on this variable indicates that being in the French Quarter adds to the price of a meal.

16.12 There will be six predictor variables in the regression analysis:

three for occupation, two for industry, and one for marital status. The dependent variable is job satisfaction. In total, there will be seven variables in this analysis.

16.13 Stepwise Regression:

Step 1: x_2 enters the model, t = - 7.35 and R^2 = .794

The model is = 36.15 - 0.146 x_2

Step 2: x_3 enters the model and x_2 remains in the model.

t for x_2 is -4.60, t for x_3 is 2.93. R^2 = .876.

The model is $\hat{y}$ = 26.40 - 0.101 x_2 + 0.116 x_3

The variable, x_1, never enters the procedure.

16.14 Stepwise Regression:

Step 1: x_4 enters the model, t = - 4.20 and R^2 = .525

The model is $\hat{y}$ = 133.53 - 0.78 x_4

Step 2: x_2 enters the model and x_4 remains in the model.

t for x_4 is - 3.22 and t for x_2 is 2.15. R^2 = .637

The model is $\hat{y}$ = 91.01 - 0.60 x_4 + 0.51 x_2

The variables, x_1 and x_3 never enter the procedure.

16.15 The output shows that the final model had four predictor variables, x_4, x_2, x_5, and x_7. The variables, x_3 and x_6 did not enter the stepwise analysis. The procedure took four steps. The final model was:

y_1 = - 5.00 x_4 + 3.22 x_2 + 1.78 x_5 + 1.56 x_7

The R^2 for this model was .5929, and s_e was 3.36. The t ratios were:

t_{x4} = 3.07, t_{x2} = 2.05, t_{x5} = 2.02, and t_{x7} = 1.98.

16.16 The output indicates that the stepwise process only went two steps. Variable x_3 entered at step one. However, at step two, x_3 dropped out of the analysis and x_2 and x_4 entered as the predictors. x_1 was the dependent variable. x_5 never entered the procedure and was not included in the final model as x_3 was not. The final regression model was:

$$\hat{Y} = 22.30 + 12.38\, x_2 + 0.0047\, x_4.$$

R^2 = .682 and s_e = 9.47. t_{x2} = 2.64 and t_{x4} = 2.01.

16.17 The output indicates that the procedure went through two steps. At step 1, dividends entered the process yielding an r^2 of .833 by itself. The t value was 6.69 and the model was $\hat{y}$ = - 11.062 + 61.1 x_1. At step 2, net income entered the procedure and dividends remained in the model. The R^2 for this two-predictor model was .897 which is a modest increase from the simple regression model shown in step one. The step 2 model was:

Premiums earned = - 3.726 + 45.2 dividends + 3.6 net income

For step 2, $t_{\text{dividends}}$ = 4.36 (p-value = .002)

and $t_{\text{net income}}$ = 2.24 (p-value = .056).

correlation matrix

	Premiums	*Income*	*Dividends*	*Gain/Loss*
Premiums	1			
Income	0.808236	1		

Dividends	0.912515	0.682321	1	
Gain/Loss	-0.40984	0.0924	-0.52241	1

16.18 This stepwise regression procedure only went one step. The only significant predictor was natural gas. No other predictors entered the model. The regression model is:

Electricity = 1.748 + 0.994 Natural Gas

For this model, R^2 = .9295 and s_e = 0.490. The t value for natural gas was 11.48.

Chapter 18

Statistical Quality Control

LEARNING OBJECTIVES

Chapter 18 presents basic concepts in quality control, with a particular emphasis on statistical quality control techniques, thereby enabling you to:

1. Understand the concepts of quality, quality control, and total quality management.
2. Understand the importance of statistical quality control in total quality management.
3. Learn about process analysis and some process analysis tools.
4. Learn how to construct $\bar{x}$ charts, *R* charts, *p* charts, and *c* charts.
5. Understand the theory and application of acceptance sampling.

CHAPTER TEACHING STRATEGY

The objective of this chapter is to present the major concepts of statistical quality control including control charts and acceptance sampling in a context of total quality management. Too many texts focus only on a few statistical quality control procedures and fail to provide the student with a managerial, decision-making context within which to use quality control statistical techniques. In this text, the concepts of total quality management along with some varying definitions of quality and some of the major theories in quality are presented. From this, the student can formulate a backdrop for the statistical techniques presented. Some statistics' instructors argue that students are exposed to some of this material in other courses. However, the background material on quality control is relatively brief; and at the very least, students should be required to read over these pages before beginning the study of control charts.

The background material helps the students understand that everyone does not agree on what a quality product is. After all, if there is no agreement on what is quality, then it is very difficult to ascertain or measure if it is being accomplished. The notion of in-process quality control helps the student understand why we generate the data that we use to construct control charts. Once the student is in the work world, it will be incumbent upon him/her to determine what measurements should be taken and monitored. A discussion on what types of measurements can be garnered in a particular business setting might be worthy of some class time. For example, if a hospital lab wants to improve quality, how would they go about it? What measurements might be useful? How about a production line of computer chips?

The chapter contains "some important quality concepts". The attempt is to familiarize, if only in passing, the student with some of the more well-known quality concepts. Included in the chapter are such things as team-building, benchmarking, just-in-time, reengineering, FMEA, Six Sigma, and Poka-Yoke all of which can effect the types of measurements being taken and the statistical techniques being used. It is a disservice to send students into the business world armed with statistical techniques such as acceptance sampling and control charts but with their heads in the sand about how the techniques fit into the total quality picture.

Chapter 18 contains a section on process analysis. Improving quality usually involves an investigation of the process from which the product emerges. The most obvious example of a process is a manufacturing assembly line. However, even in most service industries such insurance, banking, or healthcare there are processes. A useful class activity might be to brainstorm about what kind of process is involved in a person buying gasoline for their car, checking in to a hospital, or purchasing a health club membership. Think about it from a company's perspective. What activities must occur in order for a person to get their car filled up?

In analyzing process, we first discuss the construction of flowcharts. Flowcharting can be very beneficial in identifying activities and flows that need to be studied for quality improvement. One very important outcome of a flowchart is the identification of bottlenecks. You may find out that all applications for employment, for example, must pass across a clerk's desk where they sit for several days. This backs up the system and prevents flow. Other process techniques include fishbone diagrams, Pareto analysis, and control charts.

In this chapter, four types of control charts are presented. Two of the charts, the *x* bar chart and the *R* chart, deal with measurements of product attributes such as weight, length, temperature and others. The other two charts deal with whether or not items are in compliance with specifications (*p* chart) or the number of noncompliances per item (c chart). The *c* chart is less widely known and used than the other three. As part of the material on control charts, a discussion on variation is presented. Variation is one of the main concerns of quality control. A discussion on various types of variation that can occur in a business setting can be profitable in helping the student understand why particular measurements are charted and controlled.

CHAPTER OUTLINE

18.1 Introduction to Quality Control

What is Quality Control?

Total Quality Management

Some Important Quality Concepts

Benchmarking

Just-in-Time Systems

Reengineering

Failure Mode and Effects Analysis (FMEA)

Poka-Yoke

Six Sigma

Design for Six Sigma

Lean Manufacturing

Team Building

18.2 Process Analysis

Flowcharts

Pareto Analysis

Cause-and-Effect (Fishbone) Diagrams

Control Charts

Check Sheets

Histogram

Scatter Chart

18.3 Control Charts

Variation

Types of Control Charts

$\bar{x}$ Chart

R Charts

p Charts

KEY TERMS

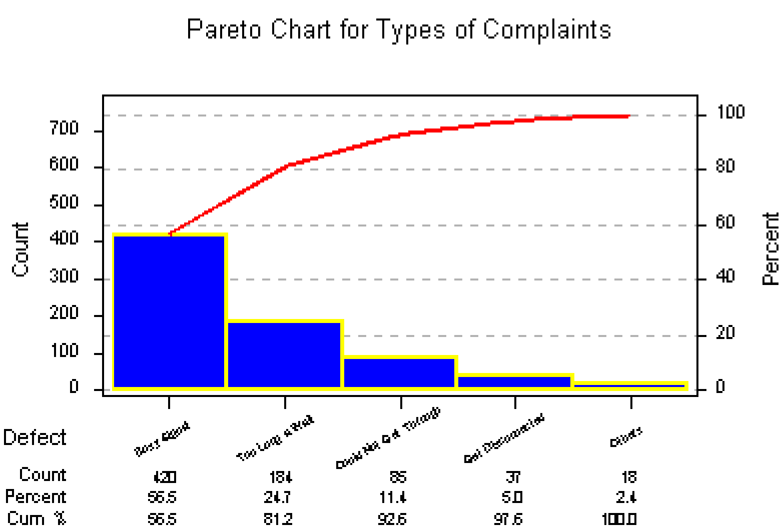
Acceptance Sampling
After-Process Quality Control
Benchmarking
c Chart
Cause-and-Effect Diagram
Centerline
Check Sheet
Consumer's Risk
Control Chart
Design for Six Sigma
Double-Sample Plan
Failure Mode and Effects Analysis
Fishbone Diagram
Flowchart
Histogram
In-Process Quality Control
Ishikawa Diagram
Just-in-Time Inventory Systems
Lean Manufacturing
Lower Control Limit (LCL)
Manufacturing Quality
Multiple-Sample Plan
p Chart
Pareto Analysis
Pareto Chart
Poka-Yoke
Process
Producer’s Risk
Product Quality
Quality
Quality Circle
Quality Control
R Chart
Reengineering
Scatter Chart
Single-Sample Plan
Six Sigma
Team Building
Total Quality Management
Transcendent Quality
Upper Control Limit (UCL)
User Quality
Value Quality
$\bar{x}$ Chart

Operating Characteristic (OC) Curve

SOLUTIONS TO PROBLEMS IN CHAPTER 18

18.2

Complaint	Number	% of Total
Busy Signal	420	56.45
Too long a Wait	184	24.73
Could not get through	85	11.42
Get Disconnected	37	4.97
Transferred to the Wrong Person	10	1.34
Poor Connection	8	1.08
Total	744	99.99



18.4 $\bar{x}_1 = 27.00$, $\bar{x}_2 = 24.29$, $\bar{x}_3 = 25.29$, $\bar{x}_4 = 27.71$, $\bar{x}_5 = 25.86$

$R_1 = 8$, $R_2 = 8$, $R_3 = 9$, $R_4 = 7$, $R_5 = 6$

$\bar{\bar{x}} = 26.03$ $\bar{R} = 7.6$

For $\bar{x}$ Chart: Since $n = 7$, $A_2 = 0.419$

Centerline: $\bar{\bar{x}} = 26.03$

UCL: $\bar{\bar{x}} + A_2\bar{R} = 26.03 + (0.419)(7.6) = 29.21$

LCL: $\bar{\bar{x}} - A_2\bar{R} = 26.03 – (0.419)(7.6) = 22.85$

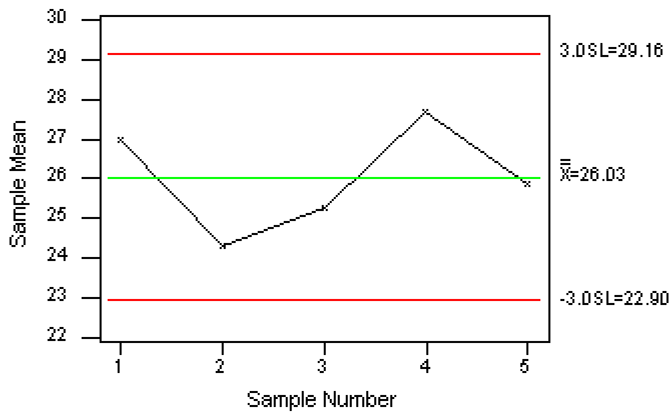
For R Chart: Since $n = 7$, $D_3 = 0.076$ $D_4 = 1.924$

Centerline: $\bar{R} = 7.6$

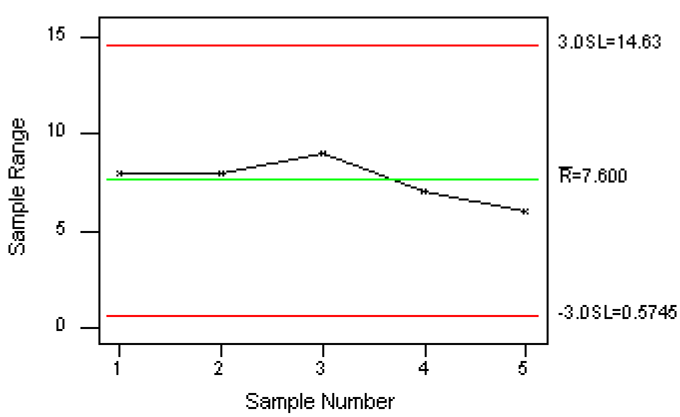
UCL: $D_4\overline{R} = (1.924)(7.6) = 14.62$

LCL: $D_3\overline{R} = (0.076)(7.6) = 0.58$

$\overline{x}$ Chart:



R Chart:



18.5 $\bar{x}_1 = 4.55$, $\bar{x}_2 = 4.10$, $\bar{x}_3 = 4.80$, $\bar{x}_4 = 4.70$,

$\bar{x}_5 = 4.30$, $\bar{x}_6 = 4.73$, $\bar{x}_7 = 4.38$

$R_1 = 1.3$, $R_2 = 1.0$, $R_3 = 1.3$, $R_4 = 0.2$, $R_5 = 1.1$, $R_6 = 0.8$, $R_7 = 0.6$

$\bar{\bar{x}} = 4.51$ $\bar{R} = 0.90$

For $\bar{x}$ Chart: Since $n = 4$, $A_2 = 0.729$

Centerline: $\bar{\bar{x}} = 4.51$

UCL: $\bar{\bar{x}} + A_2\bar{R} = 4.51 + (0.729)(0.90) = 5.17$

LCL: $\bar{\bar{x}} - A_2\bar{R} = 4.51 – (0.729)(0.90) = 3.85$

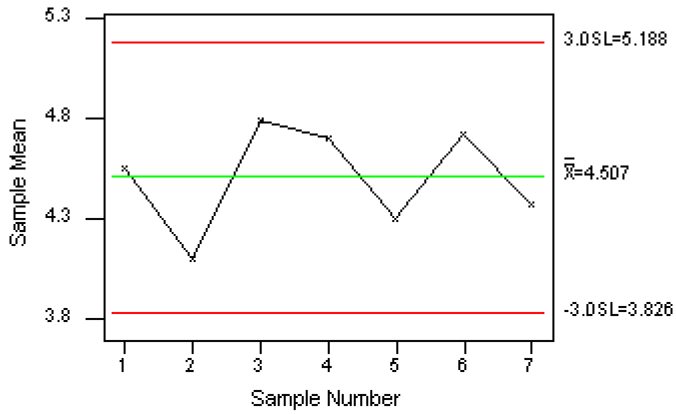
For R Chart: Since $n = 4$, $D_3 = 0$ $D_4 = 2.282$

Centerline: $\bar{R} = 0.90$

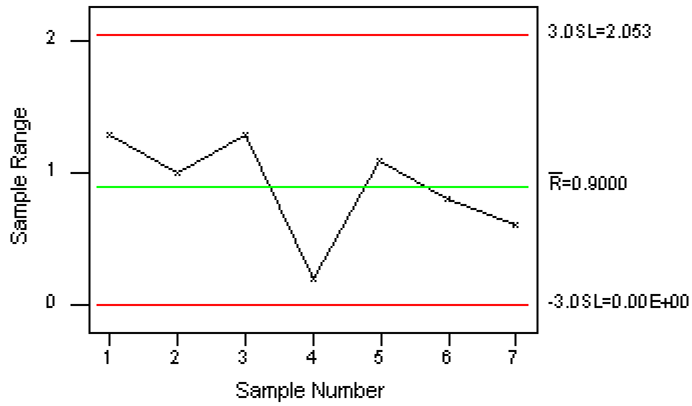
UCL: $D_4\overline{R} = (2.282)(0.90) = 2.05$

LCL: $D_3\overline{R} = 0$

$\overline{x}$ Chart:



R Chart:



18.6 $\hat{p}_1 = .02$, $\hat{p}_2 = .07$, $\hat{p}_3 = .04$, $\hat{p}_4 = .03$, $\hat{p}_5 = .03$

$\hat{p}_6 = .05$, $\hat{p}_7 = .02$, $\hat{p}_8 = .00$, $\hat{p}_9 = .01$, $\hat{p}_{10} = .06$

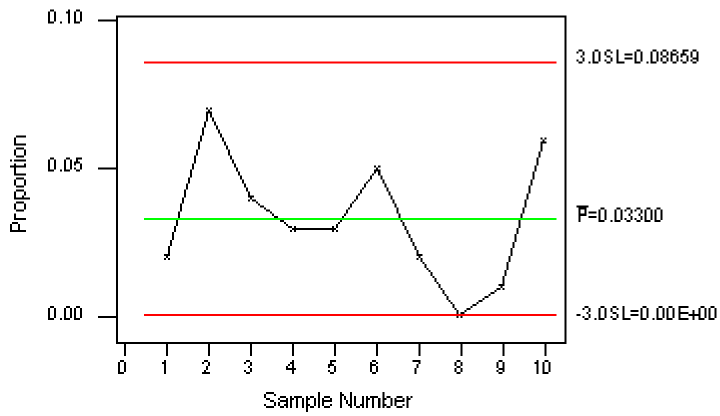
$p = .033$

Centerline: $p = .033$

UCL: $.033 + 3\sqrt{\dfrac{(.033)(.967)}{100}} = .033 + .054 = .087$

LCL: $.033 - 3\sqrt{\frac{(.033)(.967)}{100}} = .033 - .054 = .000$

*p*Chart:



18.7 $\hat{p}_1 = .025$, $\hat{p}_2 = .000$, $\hat{p}_3 = .025$, $\hat{p}_4 = .075$,

$\hat{p}_5 = .05$, $\hat{p}_6 = .125$, $\hat{p}_7 = .05$

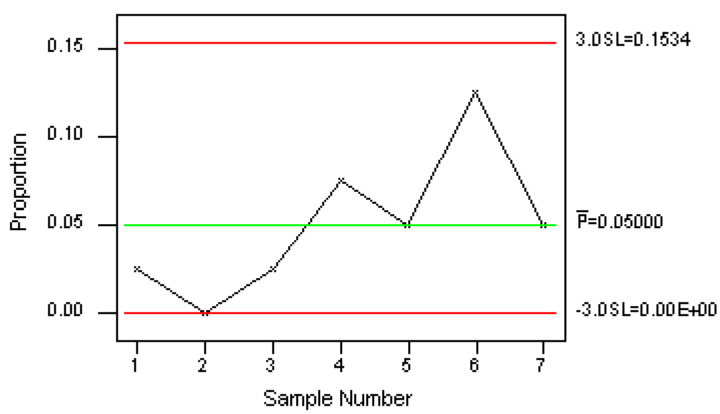
$p = .050$

Centerline: $p = .050$

UCL: $.05 + 3\sqrt{\dfrac{(.05)(.95)}{40}} = .05 + .1034 = .1534$

LCL: $.05 - 3\sqrt{\dfrac{(.05)(.95)}{40}} = .05 - .1034 = .000$

p Chart:



18.8 $\bar{c}$ = $\frac{22}{35}$ = 0.62857

Centerline: $\bar{c}$ = 0.62857

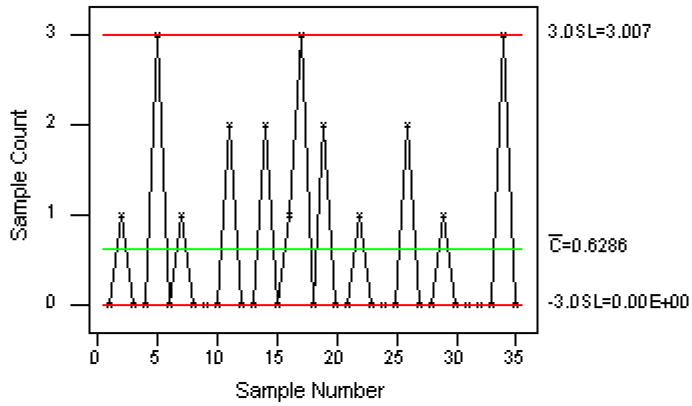
UCL: $\bar{c}+3\sqrt{\bar{c}}$ = 0.62857 + 3 $\sqrt{0.62857}$ =

0.62857 + 2.37847 = 3.00704

LCL: $\bar{c}+3\sqrt{\bar{c}}$ = 0.62857 - 3 $\sqrt{0.62857}$ =

0.62857 – 2.37847 = .000

c Chart:



18.9 $\bar{c} = \dfrac{43}{32}$ = 1.34375

Centerline: $\bar{c}$ = 1.34375

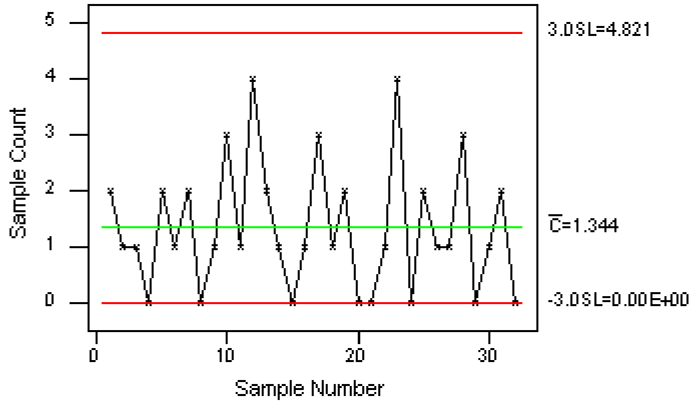
UCL: $\bar{c}+3\sqrt{\bar{c}}$ = 1.34375 + 3 $\sqrt{1.34375}$ =

1.34375 + 3.47761 = 4.82136

LCL: $\bar{c}+3\sqrt{\bar{c}}$ = 1.34375 - 3 $\sqrt{1.34375}$ =

1.34375 - 3.47761 = 0.000

c Chart:



18.10 a.) Six or more consecutive points are decreasing. Two of three consecutive points are in the outer one-third (near LCL). Four out of five points are in the outer two-thirds (near LCL).

b.) This is a relatively healthy control chart with no obvious rule violations.

c.) One point is above the UCL. Two out of three consecutive points are in the outer one-third (both near LCL and near UCL). There are six consecutive increasing points.

18.11 While there are no points outside the limits, the first chart exhibits some problems. The chart ends with 9 consecutive points below the centerline.

Of these 9 consecutive points, there are at least 4 out of 5 in the outer 2/3 of the lower region. The second control chart contains no points outside the control limit. However, near the end, there are 8 consecutive points above the centerline. The p chart contains no points outside the upper control limit. Three times, the chart contains two out of three points in the outer third. However, this occurs in the lower third where the proportion of noncompliance items approaches zero and is probably not a problem to be concerned about. Overall, this seems to display a process that is in control. One concern might be the wide swings in the proportions at samples 15, 16 and 22 and 23.

18.12 For the first sample:

If $x_1 > 4$ then reject

If $x_1 < 2$ then accept

If $2 \leq x_1 \leq 4$ then take a second sample

For the second sample, $c_2 = 3$:

If $x_1 + x_2 \leq c_2$ then accept

If $x_1 + x_2 > c_2$ then reject

But $x_1 = 2$ and $x_2 = 2$ so $x_1 + x_2 = 4 > 3$

Reject the lot because $x_1 + x_2 = 4 > c_2 = 3$

This is a double sample acceptance plan

18.13 $n = 10$ $c = 0$ $p_0 = .05$

$P(x = 0) = {}_{10}C_0(.05)^0(.95)^{10} = .5987$

$1 - P(x = 0) = 1 - .5987 = .4013$

The producer's risk is **.4013**

$p_1 = .14$

$P(x = 0) = {}_{15}C_0(.14)^0(.86)^{10} =$ **.2213**

The consumer's risk is .2213

18.14 $n = 12$ $c = 1$ $p_0 = .04$

Producer's Risk = $1 - [P(x = 0) + P(x = 1)] =$

$1 - [{}_{12}C_0(.04)^0(.96)^{12} + {}_{12}C_1(.04)^1(.96)^{11}] =$

$1 - [.6127 + .30635] = 1 - .91905 =$ **.08095**

$p_1 = .15$

Consumer's Risk = $P(x = 0) + P(x = 1) = {}_{12}C_0(.15)^0(.85)^{12} +$

${}_{12}C_1(.15)^1(.85)^{11} = .14224 + .30122 =$ **.44346**

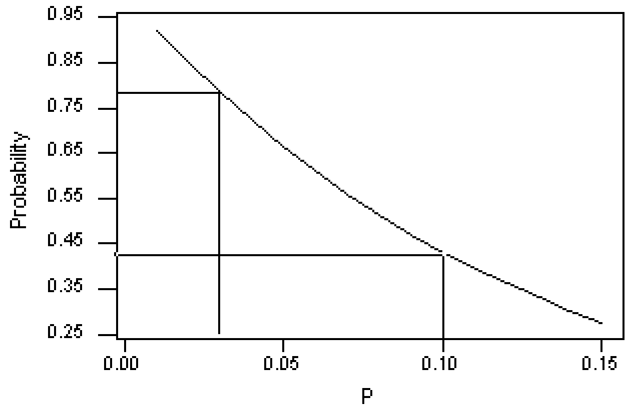
18.15 $n = 8$ $c = 0$ $p_0 = .03$ $p_1 = .1$

p	Probability
.01	.9227
.02	.8506
.03	.7837
.04	.7214
.05	.6634
.06	.6096
.07	.5596
.08	.5132
.09	.4703
.10	.4305
.11	.3937
.12	.3596
.13	.3282
.14	.2992
.15	.2725

Producer's Risk for ($p_0 = .03$) = 1 - .7837 = **.2163**

Consumer's Risk for ($p_1 = .10$) = **.4305**

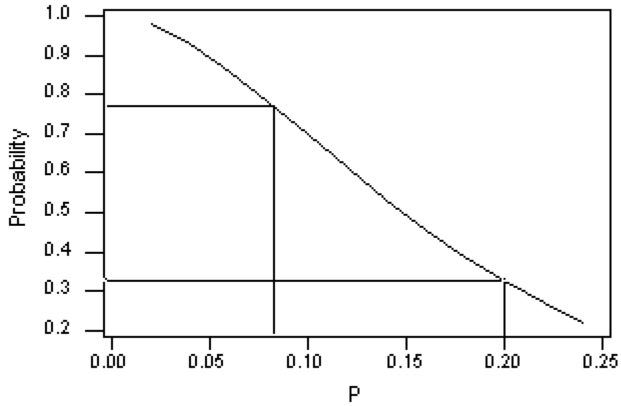
OC Chart:


0.95
0.85
0.75
0.65
0.55
0.45
0.35
0.25
Probability
0.00
0.05
0.10
0.15
P

18.16 $n = 11$ $c = 1$ $p_0 = .08$ $p_1 = .20$

p	Probability	
.02	.9805	
.04	.9308	
.06	.8618	
.08	.7819	Producer's Risk for (p_0 = .08) = 1 - .7819 = **.2181**
.10	.6974	
.12	.6127	
.14	.5311	
.16	.4547	Consumer's Risk for (p_1 = 20) = **.3221**
.18	.3849	
.20	.3221	
.22	.2667	
.24	.2186	

OC Chart:


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Probability
0.00
0.05
0.10
0.15
0.20
0.25
P

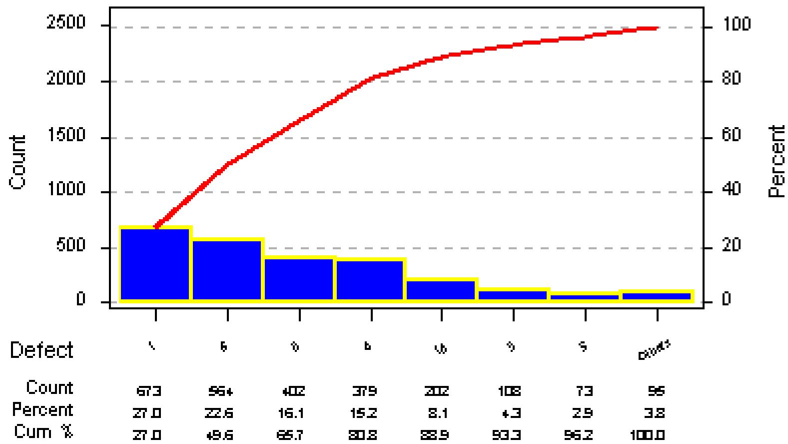
18.17

```
                    Stop
                    ↑
                     N
                    ↑
                    (no)
       D  →  K →   L  →   M (yes) → Stop
       ↑
       ↑                        Stop
       ↑                          ↑
      (no)                      (no)
Start →  A →  B (yes)→ C →  E →   F  →   G
                                (yes)
                                  ↓
                                 H(no)→  J  →  Stop
                                (yes)                ↑
                                  ↓                  ↑
                                  I  →  →  →  →  ↑
```

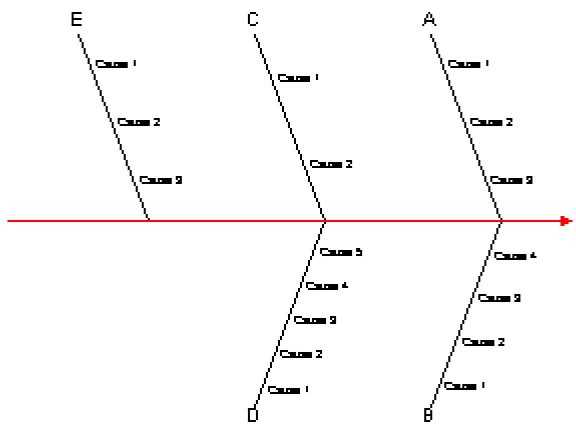
18.18

Problem	Frequency	Percent of Total
1	673	26.96
2	29	1.16
3	108	4.33
4	379	15.18
5	73	2.92
6	564	22.60
7	12	0.48
8	402	16.11
9	54	2.16
10	202	8.09
	2496	

Pareto Chart:



18.19 Fishbone Diagram:



18.20 a) $n = 13$, $c = 1$, $p_0 = .05$, $p_1 = .12$

Under $p_0 = .05$, the probability of acceptance is:

${}_{13}C_0(.05)^0(.95)^{13} + {}_{13}C_1(.05)^1(.95)^{12} =$

(1)(1)(.51334) + (13)(.05)(.54036) =

.51334 + .35123 = **.86457**

The probability of being rejected = 1 - .86457 = **.13543**

Under $p_1 = .12$, the probability of acceptance is:

${}_{13}C_0(.12)^0(.88)^{13} + {}_{13}C_1(.12)^1(.88)^{12} =$

(1)(1)(.18979) + (13)(.12)(.21567) =

.18979 + .33645 = **.52624**

The probability of being rejected = 1 - .52624 = **.47376**

b) $n = 20$, $c = 2$, $p_0 = .03$

The probability of acceptance is:

${}_{20}C_0(.03)^0(.97)^{19} + {}_{20}C_1(.03)^1(.97)^{19} + {}_{20}C_2(.03)^2(.97)^{18} =$

(1)(1)(.54379) + (20)(.03)(.56061) + (190)(.0009)(.57795) =

.54379 + .33637 + .09883 = **.97899**

The probability of being rejected = 1 - .97899 = **.02101**

which is the producer's risk

18.21 $\hat{p}_1$ = .06, $\hat{p}_2$ = .22, $\hat{p}_3$ = .14, $\hat{p}_4$ = .04, $\hat{p}_5$ = .10,

$\hat{p}_6$ = .16, $\hat{p}_7$ = .00, $\hat{p}_8$ = .18, $\hat{p}_9$ = .02, $\hat{p}_{10}$ = .12

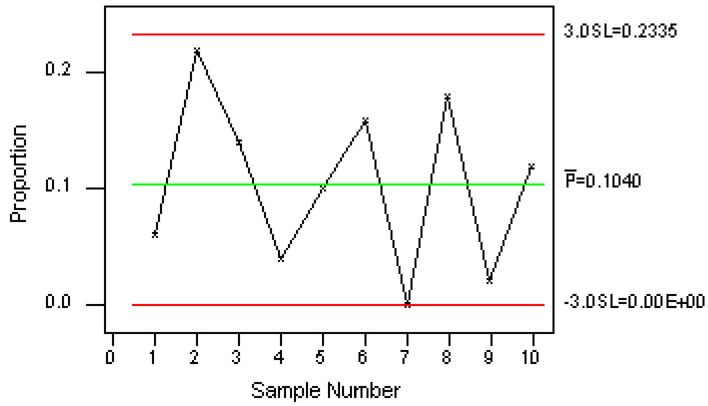
$p = \frac{52}{500}$ = .104

Centerline: p = .104

UCL: $.104 + 3\sqrt{\frac{(.104)(.896)}{50}}$ = .104 + .130 = .234

LCL: $.104 - 3\sqrt{\frac{(.104)(.896)}{50}}$ = .104 - .130 = .000

p Chart:



18.22 $\bar{x}_1 = 24.022$, $\bar{x}_2 = 24.048$, $\bar{x}_3 = 23.996$, $\bar{x}_4 = 24.000$,

$\bar{x}_5 = 23.998$, $\bar{x}_6 = 24.018$, $\bar{x}_7 = 24.000$, $\bar{x}_8 = 24.034$,

$\bar{x}_9 = 24.014$, $\bar{x}_{10} = 24.002$, $\bar{x}_{11} = 24.012$, $\bar{x}_{12} = 24.022$

$R_1 = .06$, $R_2 = .09$, $R_3 = .08$, $R_4 = .03$, $R_5 = .05$, $R_6 = .05$,

$R_7 = .05$, $R_8 = .08$, $R_9 = .03$, $R_{10} = .01$, $R_{11} = .04$, $R_{12} = .05$

$\bar{\bar{x}} = 24.01383$ $\quad\quad$ $\bar{R} = 0.05167$

For $\bar{x}$ Chart: Since $n = 12$, $A_2 = .266$

Centerline: $\bar{\bar{x}} = 24.01383$

UCL: $\bar{\bar{x}} + A_2\bar{R} = 24.01383 + (0.266)(.05167) =$

$24.01383 + .01374 = 24.02757$

LCL: $\bar{\bar{x}} - A_2\bar{R} = 24.01383 - (0.266)(.05167) =$

$24.01383 - .01374 = 24.00009$

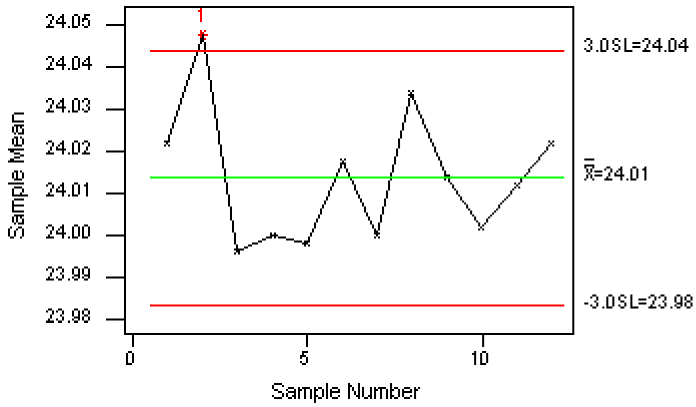
For R Chart: Since $n = 12$, $D_3 = .284$ $\quad D_4 = 1.716$

Centerline: $\bar{R} = .05167$

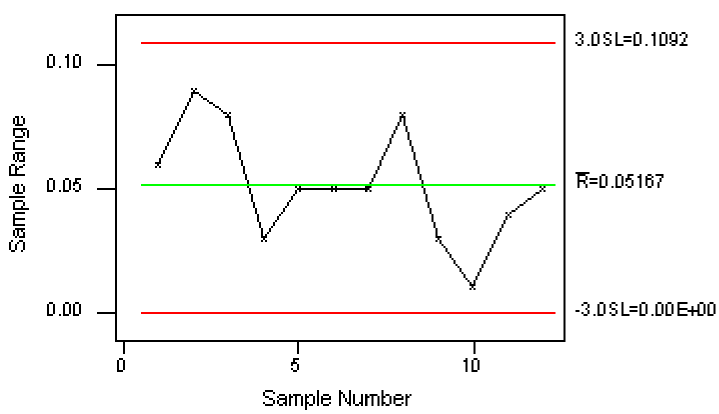
UCL: $D_4\bar{R} = (1.716)(.05167) = .08866$

LCL: $D_3\bar{R} = (.284)(.05167) = .01467$

$\overline{\overline{x}}$ Chart:



R Chart:



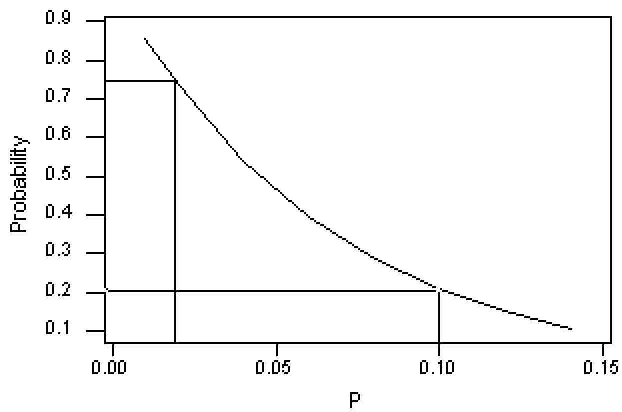
18.23 $n = 15$, $c = 0$, $p_0 = .02$, $p_1 = .10$

p	Probability
.01	.8601
.02	.7386
.04	.5421
.06	.3953
.08	.2863
.10	.2059
.12	.1470
.14	.1041

Producer's Risk for (p_0 = .02) = 1 - .7386 = **.2614**

Consumer's Risk for (p_1 = .10) = **.2059**

OC Curve:


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Probability
0.00
0.05
0.10
0.15
P

18.24 $\bar{c} = \frac{77}{36} = 2.13889$

Centerline: $\bar{c} = 2.13889$

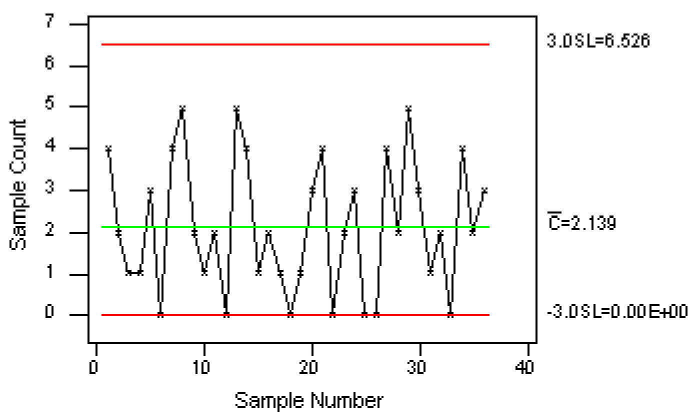
UCL: $\bar{c} + 3\sqrt{\bar{c}} = 2.13889 + 3\sqrt{2.13889} =$

2.13889 + 4.38748 = 6.52637

LCL: $\bar{c} + 3\sqrt{\bar{c}} = 2.13889 - 3\sqrt{2.13889} =$

2.13889 – 4.38748 = .00000

c Chart:



18.25 $\bar{x}_1 = 1.2100$, $\bar{x}_2 = 1.2050$, $\bar{x}_3 = 1.1900$, $\bar{x}_4 = 1.1725$,

$\bar{x}_5 = 1.2075$, $\bar{x}_6 = 1.2025$, $\bar{x}_7 = 1.1950$, $\bar{x}_8 = 1.1950$,

$\bar{x}_9 = 1.1850$

$R_1 = .04$, $R_2 = .02$, $R_3 = .04$, $R_4 = .04$, $R_5 = .06$, $R_6 = .02$,

$R_7 = .07$, $R_8 = .07$, $R_9 = .06$,

$\bar{\bar{x}} = 1.19583$ $\bar{R} = 0.04667$

For $\bar{x}$ Chart: Since $n = 4$, $A_2 = .729$

Centerline: $\bar{\bar{x}} = 1.19583$

UCL: $\bar{\bar{x}} + A_2\bar{R} = 1.19583 + .729(.04667) =$

$1.19583 + .03402 = 1.22985$

LCL: $\bar{\bar{x}} - A_2 \bar{R}$ = 1.19583 - .729(.04667) =

1.19583 - .03402 = 1.16181

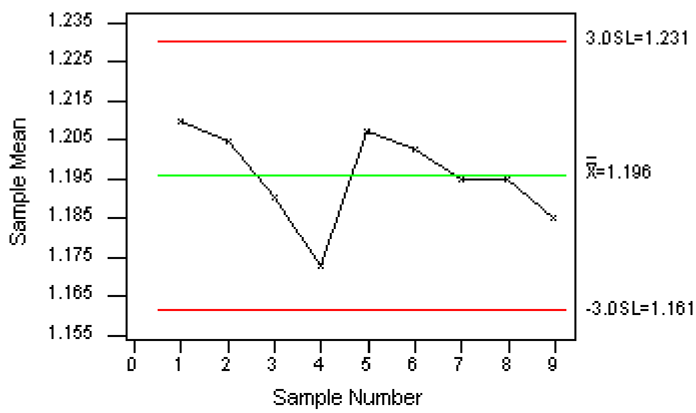
For R Chart: Since n = 9, D_3 = .184 D_4 = 1.816

Centerline: $\bar{R}$ = .04667

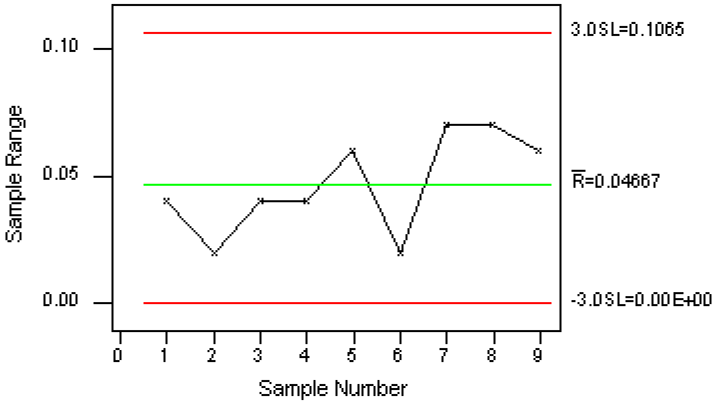
UCL: $D_4 \bar{R}$ = (1.816)(.04667) = .08475

LCL: $D_3 \bar{R}$ = (.184)(.04667) = .00859

$\bar{x}$ Chart:



R chart:



18.26 $\bar{x}_1 = 14.99333$, $\bar{x}_2 = 15.00000$, $\bar{x}_3 = 14.97833$, $\bar{x}_4 = 14.97833$,

$\bar{x}_5 = 15.01333$, $\bar{x}_6 = 15.00000$, $\bar{x}_7 = 15.01667$, $\bar{x}_8 = 14.99667$,

$R_1 = .03$, $R_2 = .07$, $R_3 = .05$, $R_4 = .05$,

$R_5 = .04$, $R_6 = .05$, $R_7 = .05$, $R_8 = .06$

$\bar{\bar{x}} = 14.99854$ $\bar{R} = 0.05$

For $\bar{x}$ Chart: Since $n = 6$, $A_2 = .483$

Centerline: $\bar{\bar{x}}$ = 14.99854

UCL: $\bar{\bar{x}}$ + $A_2 \bar{R}$ = 14.99854 + .483(.05) =

14.00854 + .02415 = 15.02269

LCL: $\bar{\bar{x}}$ - $A_2 \bar{R}$ = 14.99854 - .483(.05) =

14.00854 - .02415 = 14.97439

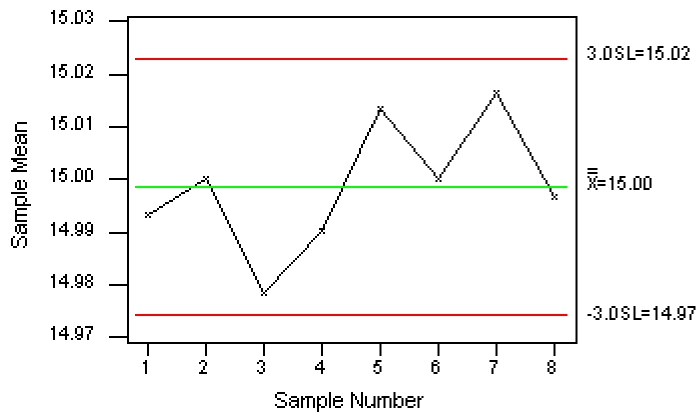
For *R* Chart: Since $n = 6$, $D_3 = 0$ $D_4 = 2.004$

Centerline: $\bar{R}$ = .05

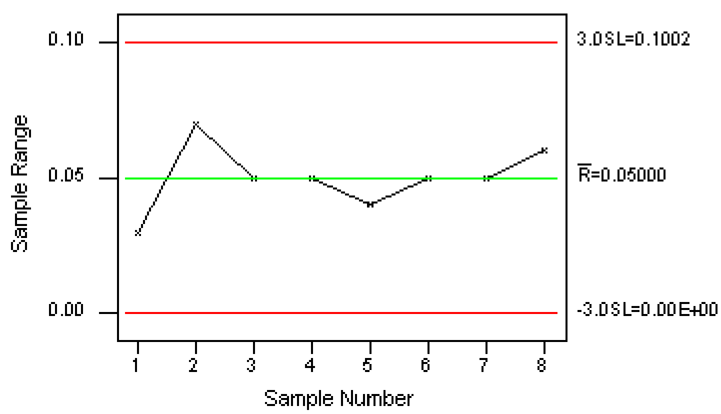
UCL: $D_4 \bar{R}$ = 2.004(.05) = .1002

LCL: $D_3 \bar{R}$ = 0(.05) = .0000

$\bar{x}$ Chart:



R chart:



18.27 $\hat{p}_1 = .12$, $\hat{p}_2 = .04$, $\hat{p}_3 = .00$, $\hat{p}_4 = .02667$,

$\hat{p}_5 = .09333$, $\hat{p}_6 = .18667$, $\hat{p}_7 = .14667$, $\hat{p}_8 = .10667$,

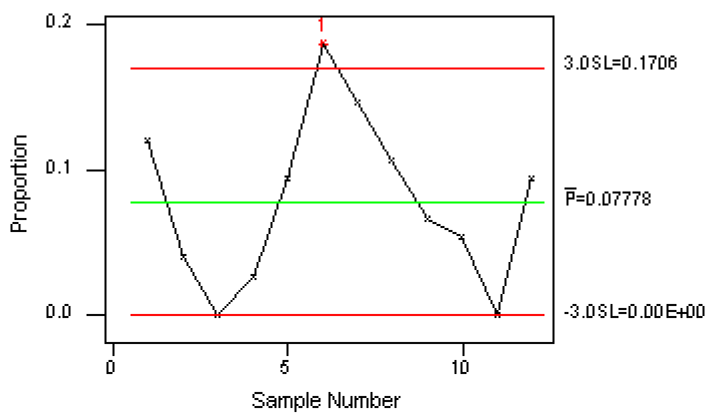
$\hat{p}_9 = .06667$, $\hat{p}_{10} = .05333$, $\hat{p}_{11} = .0000$, $\hat{p}_{12} = .09333$

$p = \frac{70}{900} = .07778$ Centerline: $p = .07778$

UCL: $.07778 + 3\sqrt{\frac{(.07778)(.92222)}{75}} = .07778 + .09278 = .17056$

LCL: $.07778 - 3\sqrt{\frac{(.07778)(.92222)}{75}} = .07778 - .09278 = .00000$

p Chart:



18.28 $\bar{c} = \frac{16}{25} = 0.64$

Centerline: $\bar{c} = 0.64$

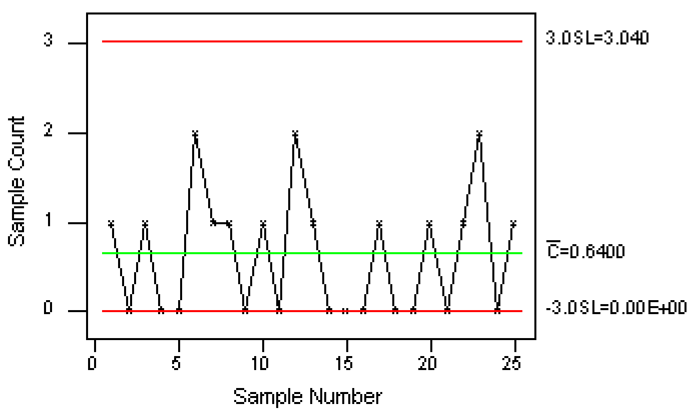
UCL: $\bar{c} + 3\sqrt{\bar{c}}$ = 0.64 + 3 $\sqrt{0.64}$ =

0.64 + 2.4 = 3.04

LCL: $\bar{c} + 3\sqrt{\bar{c}}$ = 0.64 - 3 $\sqrt{0.64}$ =

0.64 – 2.4 = .00000

c Chart:

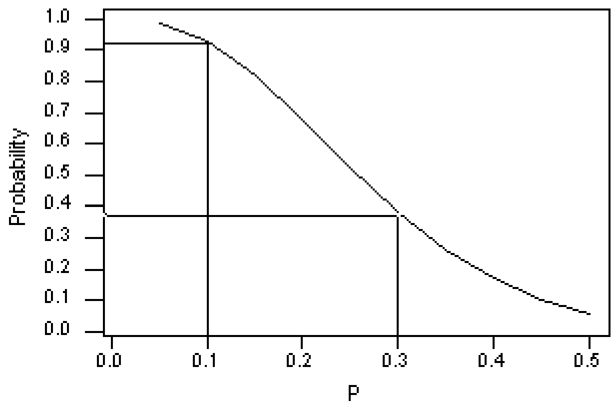


18.29 $n = 10$ $c = 2$ $p_0 = .10$ $p_1 = .30$

p	Probability
.05	.9885
.10	.9298
.15	.8202
.20	.6778
.25	.5256
.30	.3828
.35	.2616
.40	.1673
.45	.0996
.50	.0547

Producer's Risk for ($p_0 = .10$) = 1 - .9298 = **.0702**

Consumer's Risk for ($p_1 = .30$) = **.3828**


Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
0.5
P

18.30 $\bar{c} = \frac{81}{40} = 2.025$

Centerline: $\bar{c} = 2.025$

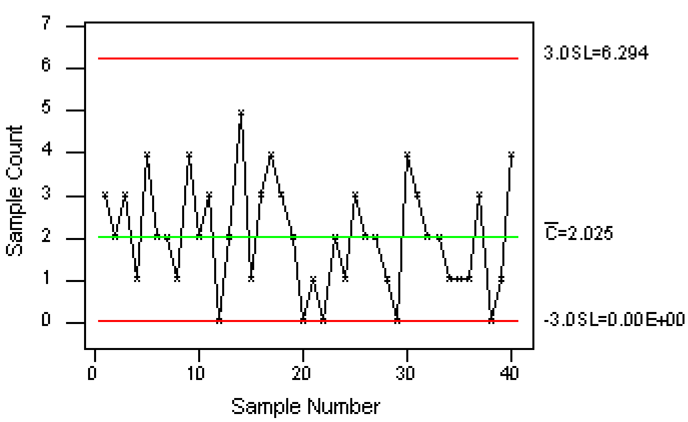
UCL: $\bar{c} + 3\sqrt{\bar{c}} = 2.025 + 3\sqrt{2.025} =$

2.025 + 4.26907 = 6.29407

LCL: $\bar{c} + 3\sqrt{\bar{c}} = 2.025 - 3\sqrt{2.025} =$

2.025 – 4.26907 = .00000

c Chart:



18.31 $\hat{p}_1 = .05$, $\hat{p}_2 = .00$, $\hat{p}_3 = .15$, $\hat{p}_4 = .075$,

$\hat{p}_5 = .025$, $\hat{p}_6 = .025$, $\hat{p}_7 = .125$, $\hat{p}_8 = .00$,

$\hat{p}_9 = .10$, $\hat{p}_{10} = .075$, $\hat{p}_{11} = .05$, $\hat{p}_{12} = .05$,

$\hat{p}_{13} = .15$, $\hat{p}_{14} = .025$, $\hat{p}_{15} = .000$

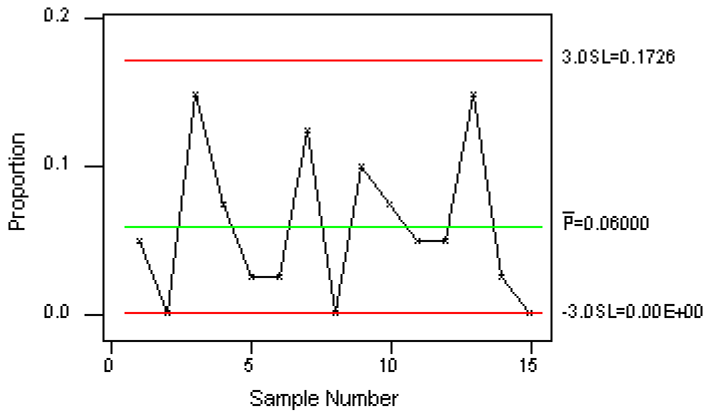
$$p = \frac{36}{600} = .06$$

Centerline: $p = .06$

$$\text{UCL: } .06 + 3\sqrt{\frac{(.06)(.94)}{40}} = .06 + .11265 = .17265$$

LCL: $.06 - 3\sqrt{\frac{(.06)(.94)}{40}} = .06 - .112658 = .00000$

p Chart:



18.32 The process appears to be in control. Only 1 sample mean is beyond the outer limits (97% of the means are within the limits). There are no more than four means in a row on one side of the centerline. There are no more than five consecutive points decreasing or three consecutive points increasing. About 2/3 (67%) of the points are within the inner 1/3 of the confidence bands ($\pm$ 1 $\sigma_{\bar{x}}$).

18.33 There are some items to be concerned about with this chart. Only one sample range is above the upper control limit. However, near the beginning of the chart there are eight sample ranges in a row below the centerline. Later in the run, there are nine sample ranges in a row above the centerline. The quality manager or operator might want to determine if there is some systematic reason why there is a string of ranges below the centerline and, perhaps more importantly, why there are a string of ranges above the centerline.

18.34 This *p* chart reveals that two of the sixty samples (about 3%) produce proportions that are too large. Nine of the sixty samples (15%) produce proportions large enough to be greater than 1 $\sigma_{\hat{p}}$ above the centerline. In general, this chart indicates a process that is under control.

18.35 The centerline of the *c* chart indicates that the process is averaging 0.74 nonconformances per part. Twenty-five of the fifty sampled items have zero nonconformances. None of the samples exceed the upper control limit for nonconformances. However, the upper control limit is 3.321 nonconformances, which in and of itself, may be too many. Indeed, three of the fifty (6%) samples actually had three nonconformances. An additional six samples (12%) had two nonconformances. One matter of concern may be that there is a run of ten samples in which nine of the samples exceed the centerline (samples 12 through 21). The question raised by this phenomenon is whether or not there is a systematic flaw in the process that produces strings of nonconforming items.

Chapter 19

Decision Analysis

LEARNING OBJECTIVES

Chapter 19 describes how to use decision analysis to improve management decisions, thereby enabling you to:

1. Learn about decision making under certainty, under uncertainty, and under risk.

2. Learn several strategies for decision-making under uncertainty, including expected payoff, expected opportunity loss, maximin, maximax, and minimax regret.

3. Learn how to construct and analyze decision trees.

4. Understand aspects of utility theory.

5. Learn how to revise probabilities with sample information.

CHAPTER TEACHING STRATEGY

The notion of contemporary decision making is built into the title of the text as a statement of the importance of recognizing that statistical analysis is primarily done as a decision-making tool. For the vast majority of students, statistics take on importance only in as much as they aid decision-makers in weighing various alternative pathways and helping the manager make the best possible determination. It has been an underlying theme from chapter 1 that the techniques presented should be considered in a decision-making context. This chapter focuses on analyzing the decision-making situation and presents several alternative techniques for analyzing decisions under varying conditions.

Early in the chapter, the concepts of decision alternatives, the states of nature, and the payoffs are presented. It is important that decision makers spend time brainstorming about possible decision alternatives that might be available to them. Sometimes the best alternatives are not obvious and are not immediately considered. The international focus on foreign companies investing in the U.S. presents a scenario in which there are several possible alternatives available. By using cases such as the Fletcher-Terry case at the chapter's end, students can practice enumerating possible decision alternatives.

States of nature are possible environments within which the outcomes will occur over which we have no control. These include such things as the economy, the weather, health of the CEO, wildcat strikes, competition, change in consumer demand, etc. While the text presents problems with only a few states of nature in order to keep the length of solution reasonable, students should learn to consider as many states of nature as possible in decision making. Determining payoffs is relatively difficult but essential in the analysis of decision alternatives.

Decision-making under uncertainty is the situation in which the outcomes are not known and there are no probabilities given as to the likelihood of them occurring. With these techniques, the emphasis is whether or not the approach is optimistic, pessimistic, or weighted somewhere in between.

In making decisions under risk, the probabilities of each state of nature occurring are known or are estimated. Decision trees are introduced as an alternative mechanism for displaying the problem. The idea of an expected monetary value is that if this decision process were to continue with the same parameters for a long time, what would the long-run average outcome be? Some decisions lend themselves to long-run average analysis such as gambling outcomes or insurance actuary analysis. Other decisions such as building a dome stadium downtown or drilling one oil well tend to be more one time activities and may not lend themselves as nicely to expected value analysis. It is important that the student understand that expected value outcomes are long-run averages and probably will not occur in single instance decisions.

Utility is introduced more as a concept than an analytic technique. The idea here is to aid the decision-maker in determining if he/she tends to be more of a risk-taker, an EMV'r, or risk-averse. The answer might be that it depends on the matter over which the decision is being made. One might be a risk-taker on attempting to employ a more diverse work force and at the same time be more risk-averse in investing the company's retirement fund.

CHAPTER OUTLINE

19.1 The Decision Table and Decision Making Under Certainty

Decision Table

Decision-Making Under Certainty

19.2 Decision Making Under Uncertainty

Maximax Criterion

Maximin Criterion

Hurwicz Criterion

Minimax Regret

19.3 Decision Making Under Risk

Decision Trees

Expected Monetary Value (EMV)

Expected Value of Perfect Information

Utility

19.4 Revising Probabilities in Light of Sample Information

Expected Value of Sample Information

KEY TERMS

Decision Alternatives

Decision Analysis

Decision Making Under Certainty

Decision Making Under Risk

Decision Making Under Uncertainty

Decision Table

Decision Trees

EMV'er

Expected Monetary Value (EMV)

Expected Value of Perfect Information

Expected Value of Sample Information

Hurwicz Criterion

Maximax Criterion

Maximin Criterion

Minimax Regret

Opportunity Loss Table

Payoffs

Payoff Table

Risk-Avoider

Risk-Taker

States of Nature

Utility

SOLUTIONS TO PROBLEMS IN CHAPTER 19

19.1

	S_1	S_2	S_3	Max	Min
d_1	250	175	-25	250	-25
d_2	110	100	70	110	70
d_3	390	140	-80	390	-80

a.) Max {250, 110, 390} = **390** decision: Select d_3

b.) Max {-25, 70, -80} = **70** decision: Select d_2

c.) For α = .3

d_1: .3(250) + .7(-25) = 57.5

d_2: .3(110) + .7(70) = **82**

d_3: .3(390) + .7(-80) = 61

decision: Select d_2

For $\alpha = .8$

d_1: .8(250) + .2(-25) = 195

d_2: .8(110) + .2(70) = 102

d_3: .8(390) + .2(-80) = **296**

decision: Select d_3

Comparing the results for the two different values of alpha, with a more pessimist point-of-view ($\alpha = .3$), the decision is to select d_2 and the payoff is 82. Selecting by using a more optimistic point-of-view ($\alpha = .8$) results in choosing d_3 with a higher payoff of 296.

d.) The opportunity loss table is:

	S_1	S_2	S_3	Max
d_1	140	0	95	140

d_2	280	75	0	280
d_3	0	35	150	150

The minimax regret = min {140, 280, 150} = **140**

Decision: Select d_1 to minimize the regret.

19.2

	S_1	S_2	S_3	S_4	Max	Min
d_1	50	70	120	110	120	50
d_2	80	20	75	100	100	20
d_3	20	45	30	60	60	20
d_4	100	85	-30	-20	100	-30
d_5	0	-10	65	80	80	-10

a.) Maximax = Max {120, 100, 60, 100, 80} = **120**

Decision: Select d_1

b.) Maximin = Max {50, 20, 20, -30, -10} = **50**

Decision: Select d_1

c.) $\alpha = .5$

Max {[.5(120)+.5(50)], [.5(100)+.5(20)],

[.5(60)+.5(20)], [.5(100)+.5(-30)], [.5(80)+.5(-10)]}=

Max { 85, 60, 40, 35, 35 } = **85**

Decision: Select d_1

d.) Opportunity Loss Table:November 8, 1996

	S_1	S_2	S_3	S_4	Max
d_1	50	15	0	0	50
d_2	20	65	45	10	65
d_3	80	40	90	50	90
d_4	0	0	150	130	150
d_5	100	95	55	30	100

Min {50, 65, 90, 150, 100} = **50**

Decision: Select d_1

19.3

	R	D	I	Max	Min
A	60	15	-25	60	-25
B	10	25	30	30	10
C	-10	40	15	40	-10
D	20	25	5	25	5

Maximax = Max {60, 30, 40, 25} = **60**

Decision: Select A

Maximin = Max {-25, 10, -10, 5} = **10**

Decision: Select B

19.4

	Not	Somewhat	Very	Max	Min
None	-50	-50	-50	-50	-50
Few	-200	300	400	400	-200
Many	-600	100	1000	1000	-600

a.) For Hurwicz criterion using $\alpha = .6$:

Max {[.6(-50) + .4(-50)], [.6(400) + .4(-200)],

[.6(1000) + .4(-600)]} = {-50, -160, 360}= **360**

Decision: Select "Many"

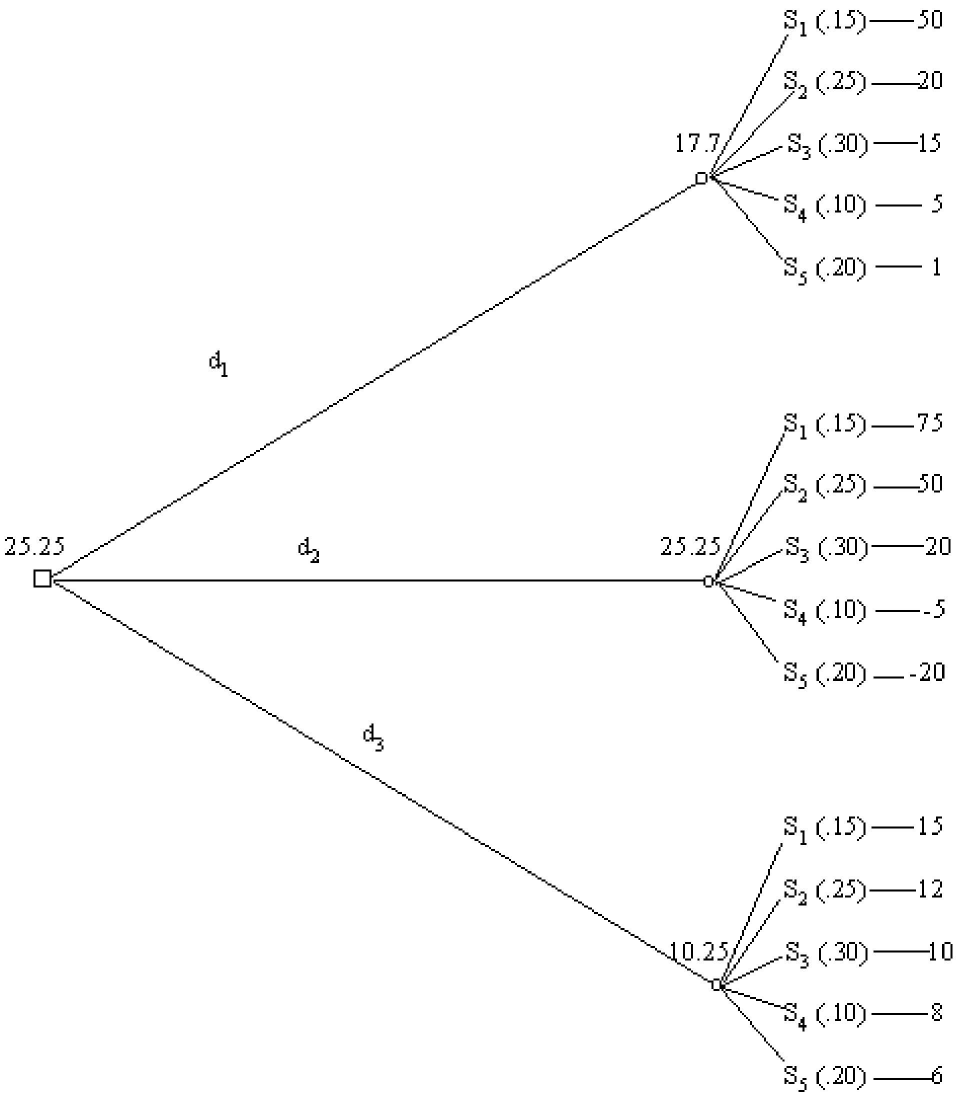
b.) Opportunity Loss Table:

	Not	Somewhat	Very	Max
None	0	350	1050	1050
Few	150	0	600	600
Many	550	200	0	550

Minimax regret = Min {1050, 600, 550} = **550**

Decision: Select "Many"

19.5, 19.6



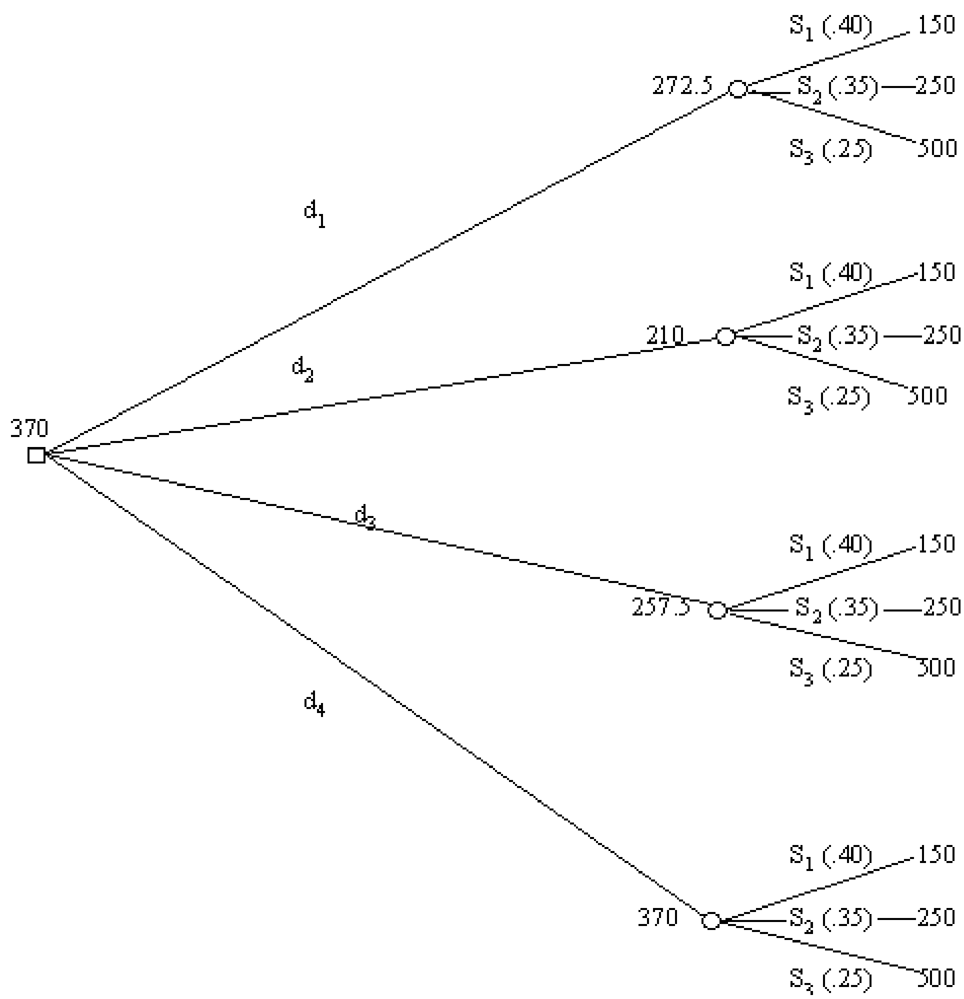
25.25
d1
d2
d3
17.7
S1 (.15) 50
S2 (.25) 20
S3 (.30) 15
S4 (.10) 5
S5 (.20) 1
25.25
S1 (.15) 75
S2 (.25) 50
S3 (.30) 20
S4 (.10) -5
S5 (.20) -20
10.25
S1 (.15) 15
S2 (.25) 12
S3 (.30) 10
S4 (.10) 8
S5 (.20) 6

19.7 Expected Payoff with Perfect Information =

5(.15) + 50(.25) + 20(.30) + 8(.10) + 6(.20) = **31.75**

Expected Value of Perfect Information = 31.25 - 25.25 = **6.50**

19.8 a.) & b.)



c.) Expected Payoff with Perfect Information =

150(40) + 450(.35) + 700(.25) = **392.5**

Expected Value of Perfect Information = 392.5 - 370 = **22.50**

19.9	Down(.30)	Up(.65)	No Change(.05)	EMV
Lock-In	-150	200	0	85
No	175	-250	0	-110

Decision: Based on the highest EMV)**(85)**, "Lock-In"

Expected Payoff with Perfect Information =

175(.30) + 200(.65) + 0(.05) = **182.5**

Expected Value of Perfect Information = 182.5 - 85 = **97.5**

19.10	EMV
No Layoff	-960
Layoff 1000	-320

Layoff 5000   **400**

Decision: Based on maximum EMV (400), Layoff 5000

Expected Payoff with Perfect Information =

100(.10) + 300(.40) + 600(.50) = **430**

Expected Value of Perfect Information = 430 - 400 = **30**

19.11   a.) EMV = 200,000(.5) + (-50,000)(.5) = **75,000**

b.) Risk Avoider because the EMV is more than the investment (75,000 > 50,000)

c.) You would have to offer more than 75,000 which is the expected value.

19.12 a.)

	$S_1(.30)$	$S_2(.70)$	EMV
d_1	350	-100	35
d_2	-200	325	167.5

Decision: Based on EMV,

maximum {35, 167.5} = **167.5**

b. & c.) For Forecast S_1:

	Prior	Cond.	Joint	Revised
S_1	.30	.90	.27	.6067
S_2	.70	.25	<u>.175</u>	.3933

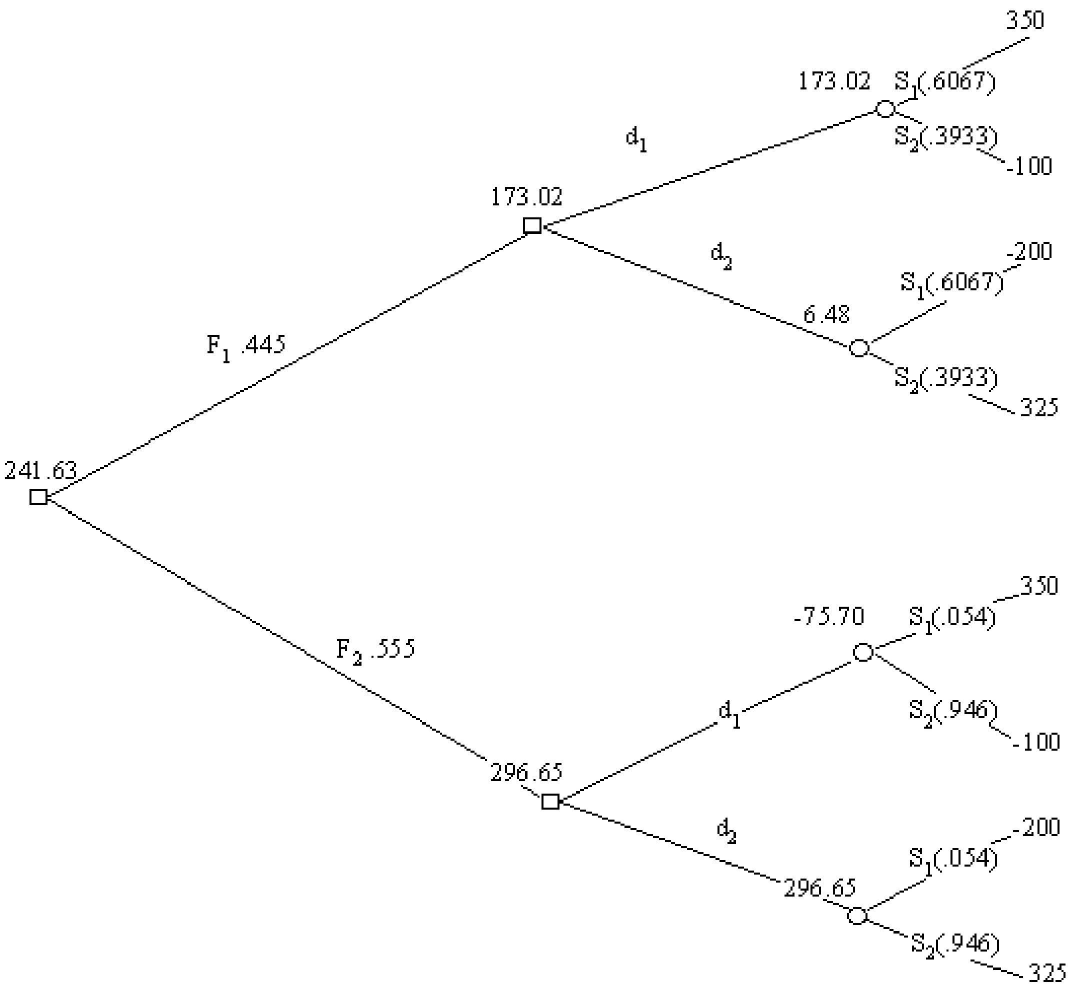
$F(S_1) = .445$

For Forecast S_2:

	Prior	Cond.	Joint	Revised

S_1	.30	.10	.030	.054
S_2	.70	.75	<u>.525</u>	.946

$F(S_2) = .555$



EMV with Sample Information = **241.63** d.) Value of Sample Information = 241.63 - 167.5 = **74.13**

19.13

	Dec(.60)	Inc(.40)	EMV
S	-225	425	35
M	125	-150	15
L	350	-400	50

Decision: Based on EMV = Maximum {35, 15, 50} = **50**

For Forecast (Decrease):

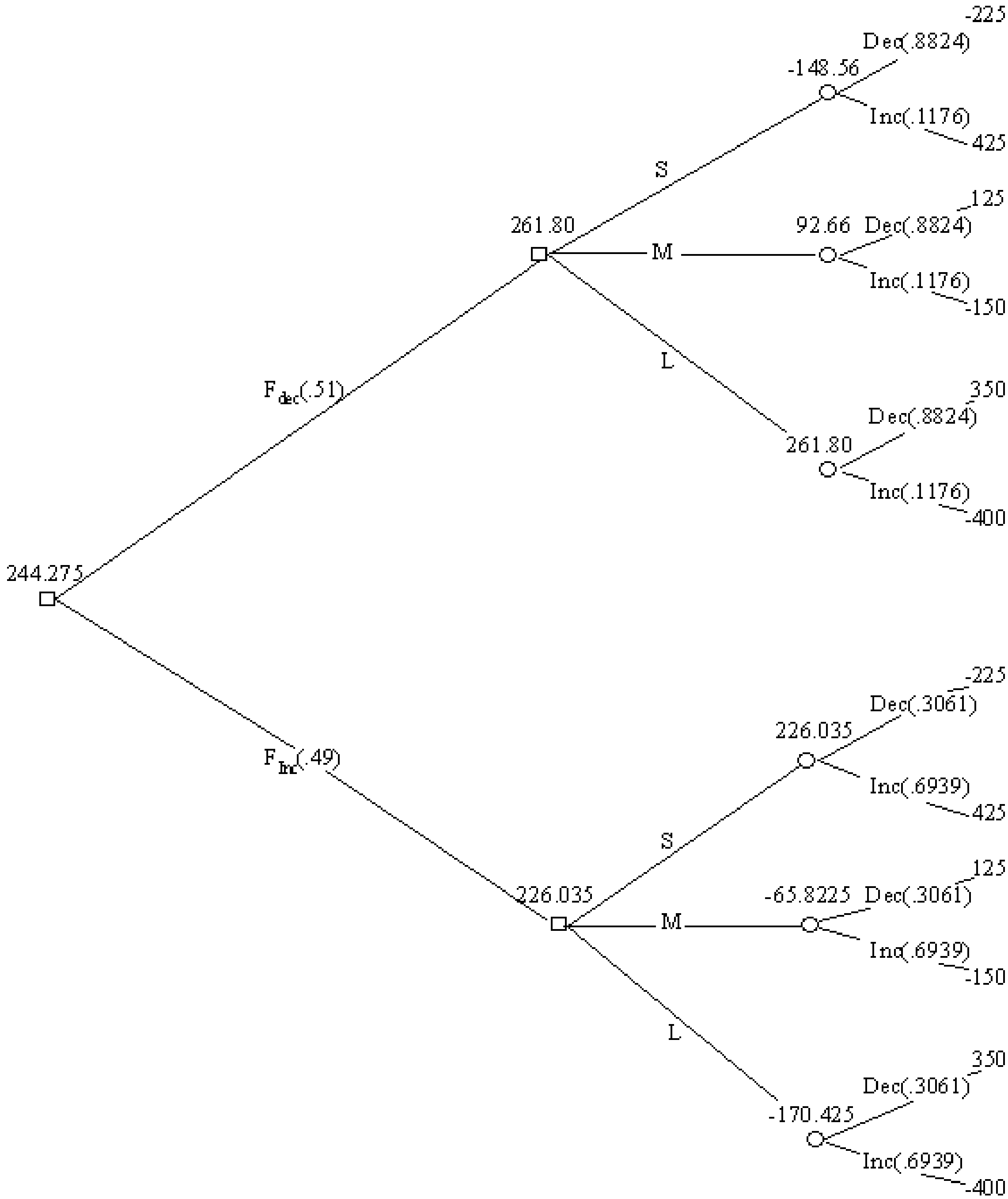
	Prior	Cond.	Joint	Revised
Decrease	.60	.75	.45	.8824
Increase	.40	.15	.06	.1176

F(Dec) = .51

For Forecast (Increase):

	Prior	Cond.	Joint	Revised
Decrease	.60	.25	.15	.3061
Increase	.40	.85	.34	.6939

F(Inc) = .49



The expected value with sampling is **244.275**

EVSI = EVWS - EMV = 244.275 - 50 = **194.275**

19.14

	Decline(.20)	Same(.30)	Increase(.50)	EMV
Don't Plant	20	0	-40	-16
Small	-90	10	175	72.5
Large	-600	-150	800	235

Decision: Based on Maximum EMV =

Max {-16, 72.5, 235} = **235**, plant a large tree farm

For forecast decrease:

Prior	Cond.	Joint	Revised
.20	.70	.140	.8974
.30	.02	.006	.0385
.50	.02	.010	.0641

$P(F_{dec}) = .156$

For forecast same:

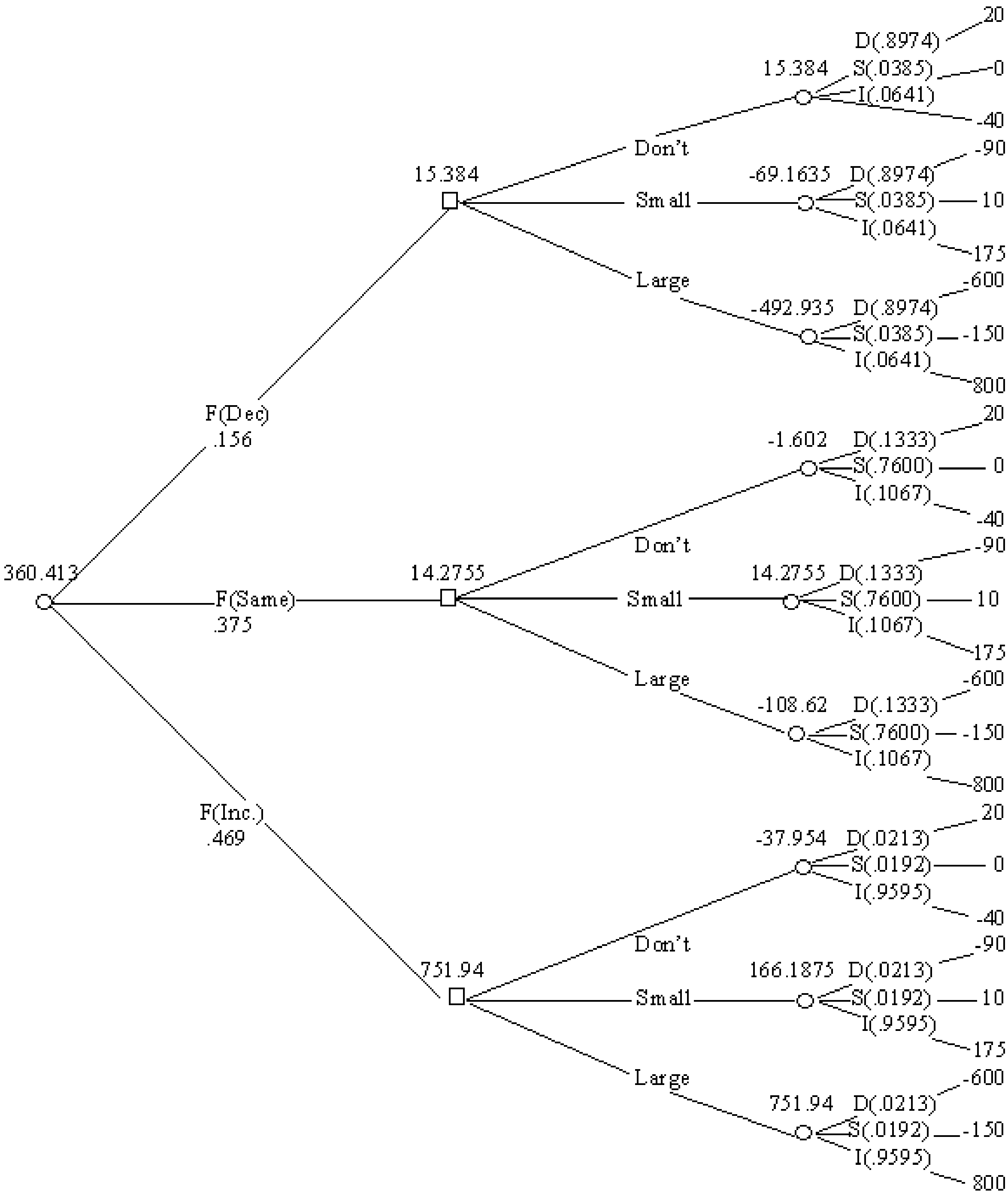
Prior	Cond.	Joint	Revised
.20	.25	.05	.1333
.30	.95	.285	.7600
.50	.08	.040	.1067

$P(F_{same}) = .375$

For forecast increase:

Prior	Cond.	Joint	Revised
.20	.05	.01	.0213
.30	.03	.009	.0192
.50	.90	.45	.9595

$P(F_{inc}) = .469$


360.413
F(Dec)
.156
F(Same)
.375
F(Inc.)
.469
15.384
14.2755
751.94
Don't
Small
Large
15.384
D(.8974)
S(.0385)
I(.0641)
-69.1635
-492.935
-1.602
D(.1333)
S(.7600)
I(.1067)
14.2755
-108.62
-37.954
D(.0213)
S(.0192)
I(.9595)
166.1875
751.94
20
0
-40
-90
10
175
-600
-150
800

The Expected Value with Sampling Information is **360.413**

EVSI = EVWSI - EMV = 360.413 - 235 = **125.413**

19.15

	Oil(.11)	No Oil(.89)	EMV
Drill	1,000,000	-100,000	21,000
Don't Drill	0	0	0

Decision: The EMV for this problem is Max {21,000, 0} = 21,000.

The decision is to Drill.

		Actual	
		Oil	No Oil
Forecast	Oil	.20	.10
	No Oil	.80	.90

Forecast Oil:

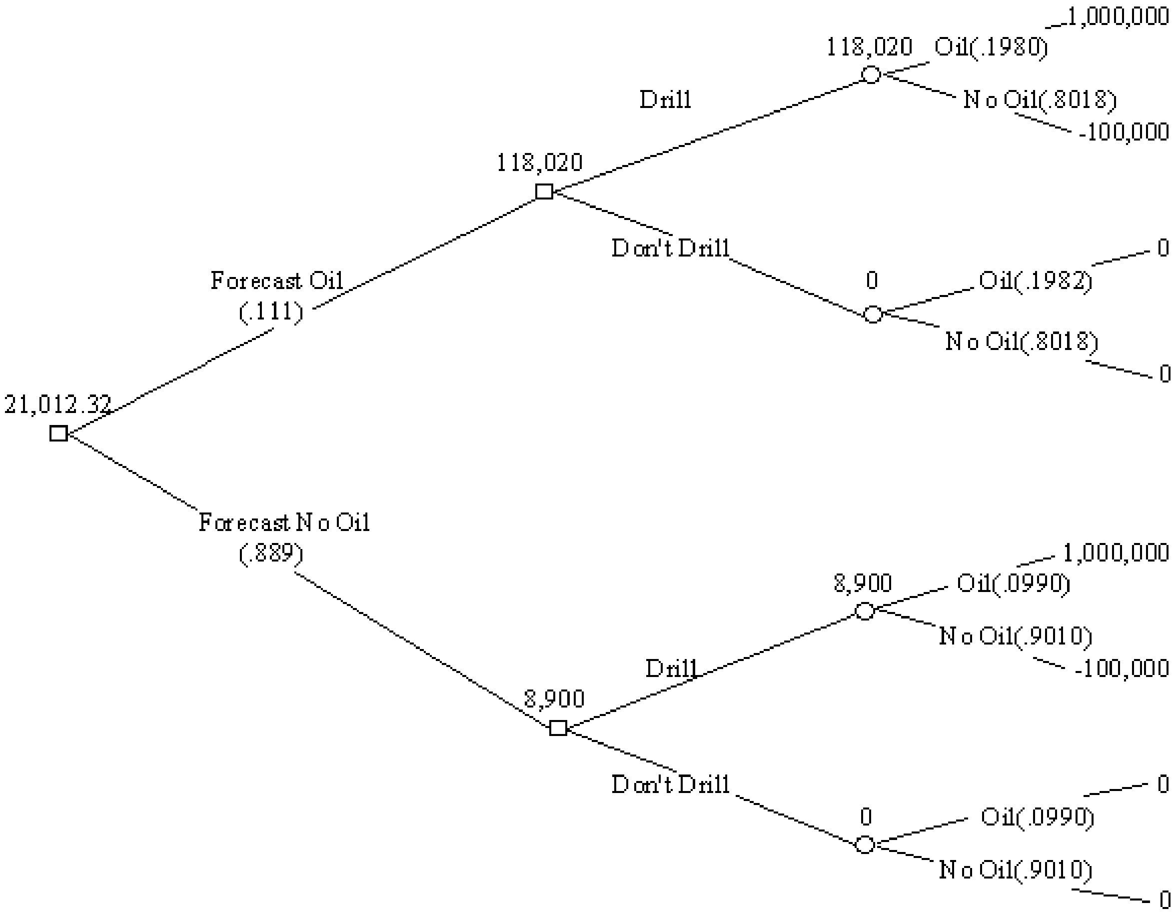
State	Prior	Cond.	Joint	Revised
Oil	.11	.20	.022	.1982
No Oil	.89	.10	.089	.8018

$P(F_{Oil}) = .111$

Forecast No Oil:

State	Prior	Cond.	Joint	Revised
Oil	.11	.80	.088	.0990
No Oil	.89	.90	.801	.9010

$P(F_{No\ Oil}) = .889$



The Expected Value With Sampling Information is **21,012.32**

EVSI = EVWSI - EMV = 21,000 - 21,012.32 = **12.32**

19.16

	S_1	S_2	Max.	Min.
d_1	50	100	100	50
d_2	-75	200	200	-75
d_3	25	40	40	25
d_4	75	10	75	10

a.) Maximax: Max {100, 200, 40, 75} = **200**

Decision: Select d_2

b.) Maximin: Max {50, -75, 25, 10} = **50**

Decision: Select d_1

c.) Hurwicz with α = .6

d_1: 100(.6) + 50(.4) = 80

d_2: 200(.6) + (-75)(.4) = 90

d_3: 40(.6) + 25(.4) = 34

d_4: 75(.6) + 10(.4) = 49

Max {80, 90, 34, 49} = **90**

Decision: Select d_2

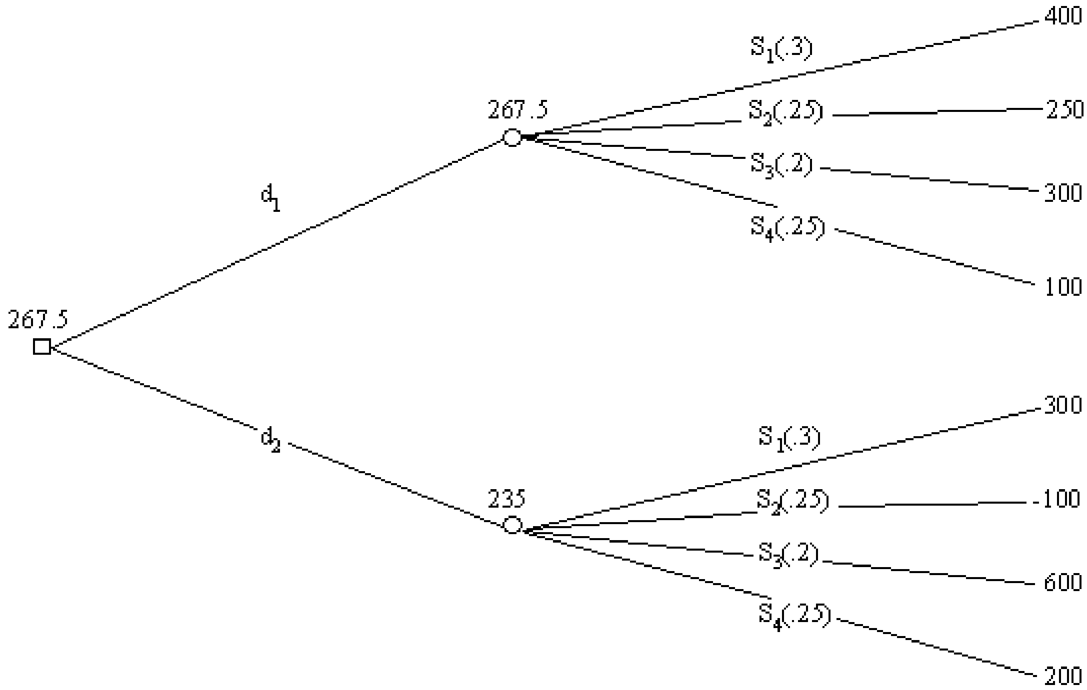
d.) Opportunity Loss Table:

	S_1	S_2	Maximum
d_1	25	100	100
d_2	150	0	150
d_3	50	160	160
d_4	0	190	190

Min {100, 150, 160, 190} = **100**

Decision: Select d_1

19.17



b.) d_1: 400(.3) + 250(.25) + 300(.2) + 100(.25) = **267.5**

d_2: 300(.3) + (-100)(.25) + 600(.2) + 200(.25) = 235

Decision: Select d_1

c.) Expected Payoff of Perfect Information:

400(.3) + 250(.25) + 600(.2) + 200(.25) = **352.5**

Value of Perfect Information = 352.5 - 267.5 = **85**

19.18

	$S_1(.40)$	$S_2(.60)$	EMV
d_1	200	150	170
d_2	-75	450	240
d_3	175	125	145

Decision: Based on Maximum EMV =

Max {170, 240, 145} = **240**

Select d_2

Forecast S_1:

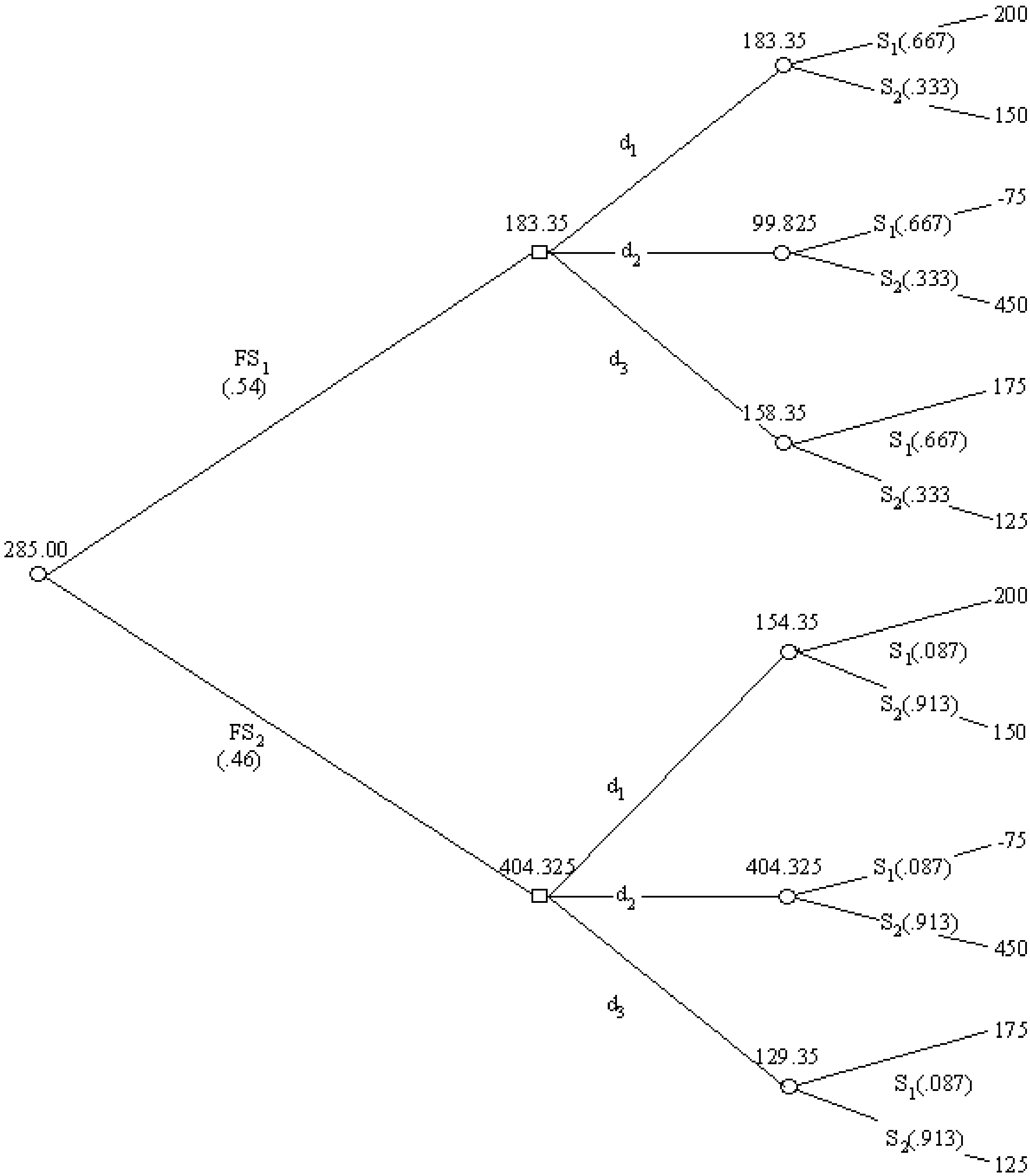
State	Prior	Cond.	Joint	Revised
S_1	.4	.9	.36	.667
S_2	.6	.3	.18	.333

$P(F_{S1}) = .54$

Forecast S_2:

State	Prior	Cond.	Joint	Revised
S_1	.4	.1	.04	.087
S_2	.6	.7	.42	.913

$$P(F_{S2}) = .46$$



The Expected Value With Sample Information is **285.00**

EVSI = EVWSI - EMV = 285 - 240 = **45**

19.19	Small	Moderate	Large	Min	Max
Small	200	250	300	200	300
Modest	100	300	600	100	600
Large	-300	400	2000	-300	2000

a.) Maximax: Max {300, 600, 2000} = **2000**

Decision: Large Number

Minimax: Max {200, 100, -300} = **200**

Decision: Small Number

b.) Opportunity Loss:

	Small	Moderate	Large	Max
Small	0	150	1700	1700
Modest	100	100	1400	1400
Large	500	0	0	500

Min {1700, 1400, 500} = **500**

Decision: Large Number

c.) Minimax regret criteria leads to the same decision as Maximax.

19.20	No	Low	Fast	Max	Min
Low	-700	-400	1200	1200	-700
Medium	-300	-100	550	550	-300
High	100	125	150	150	100

a.) α= .1:

Low: 1200(.1) + (-700)(.9) = -510

Medium: 550(.1) + (-300)(.9) = -215

High: 150(.1) + 100(.9) = 105

Decision: Price High **(105)**

b.) α = .5:

Low: 1200(.5) + (-700)(.5) = 250

Medium: 550(.5) + (-300)(.5) = 125

High: 150(.5) + 100(.5) = 125

Decision: Price Low **(250)**

c.) α = .8:

Low: 1200(.8) + (-700)(.2) = 820

Medium: 550(.8) + (-300)(.2) = 380

High: 150(.8) + 100(.2) = 140

Decision: Price Low **(820)**

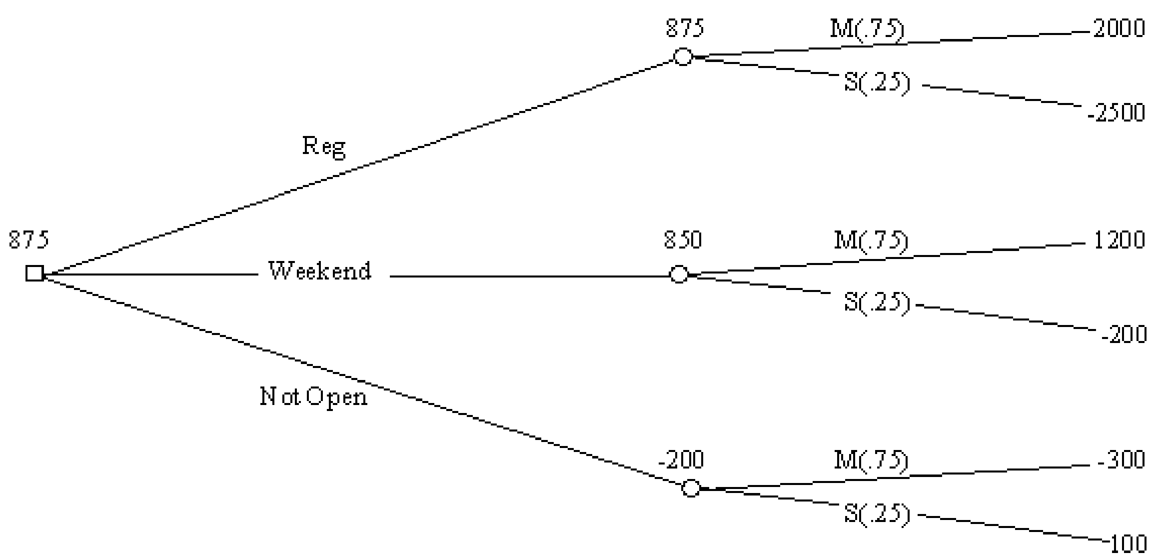
d.) Two of the three alpha values (.5 and .8) lead to a decision of pricing low.

Alpha of .1 suggests pricing high as a strategy. For optimists (high alphas), pricing low is a better strategy; but for more pessimistic people, pricing high may be the best strategy.

19.21	Mild(.75)	Severe(.25)	EMV
Reg.	2000	-2500	875
Weekend	1200	-200	850
Not Open	-300	100	-200

Decision: Based on Max EMV =

Max{875, 850, -200} = **875**, open regular hours.



Expected Value with Perfect Information =

2000(.75) + 100(.25) = **1525**

Value of Perfect Information = 1525 - 875 = **650**

19.22

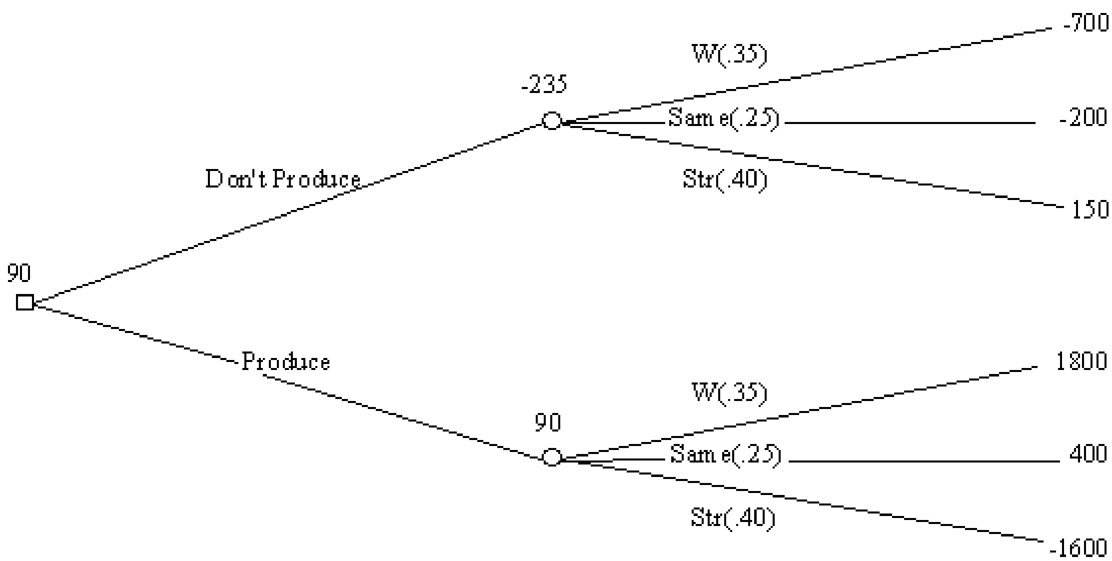
	Weaker(.35)	Same(.25)	Stronger(.40)	EMV
Don't Produce	-700	-200	150	-235
Produce	1800	400	-1600	90

Decision: Based on Max EMV = Max {-235, 90} = **90**, select Produce.

Expected Payoff With Perfect Information =

1800(.35) + 400(.25) + 150(.40) = **790**

Value of Perfect Information = 790 - 90 = **700**



19.23

	Red.(.15)	Con.(.35)	Inc.(.50)	EMV
Automate	-40,000	-15,000	60,000	18,750
Do Not	5,000	10,000	-30,000	-10,750

Decision: Based on Max EMV =

Max {18750, -10750} = **18,750**, Select Automate

Forecast Reduction:

State	Prior	Cond.	Joint	Revised
R	.15	.60	.09	.60
C	.35	.10	.035	.2333
I	.50	.05	.025	.1667

$P(F_{Red}) = .150$

Forecast Constant:

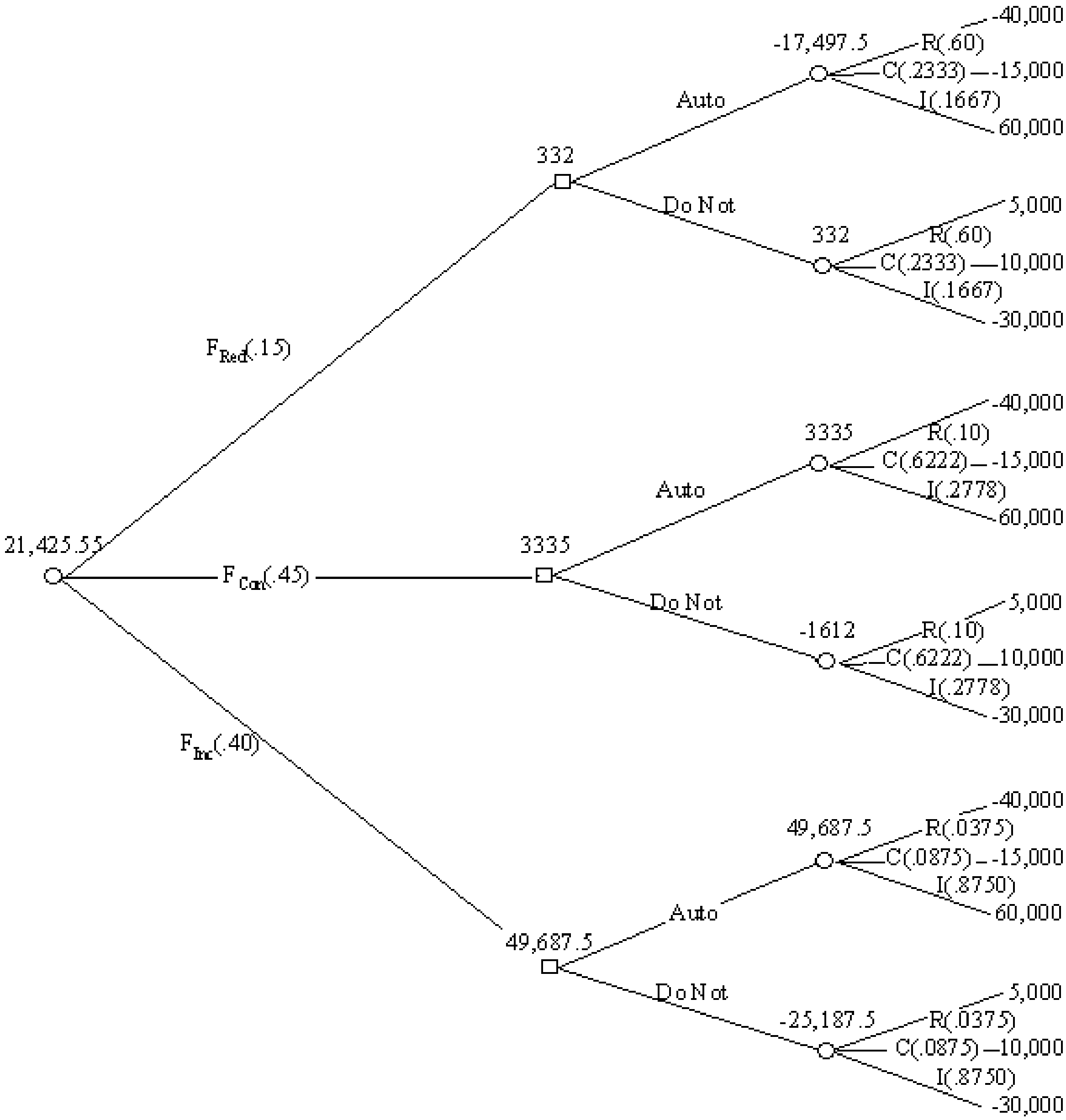
State	Prior	Cond.	Joint	Revised
R	.15	.30	.045	.10
C	.35	.80	.280	.6222
I	.50	.25	.125	.2778

$P(F_{Cons}) = .450$

Forecast Increase:

State	Prior	Cond.	Joint	Revised
R	.15	.10	.015	.0375
C	.35	.10	.035	.0875
I	.50	.70	<u>.350</u>	.8750

$P(F_{Inc}) = .400$



Expected Value With Sample Information = **21,425.55**

EVSI = EVWSI - EMV = 21,425.55 - 18,750 = **2,675.55**

19.24

	Chosen(.20)	Not Chosen(.80)	EMV
Build	12,000	-8,000	-4,000
Don't	-1,000	2,000	1,400

Decision: Based on Max EMV = Max {-4000, 1400} = **1,400**, choose "Don't Build" as a strategy.

Forecast Chosen:

State	Prior	Cond.	Joint	Revised
Chosen	.20	.45	.090	.2195
Not Chosen	.80	.40	.320	.7805

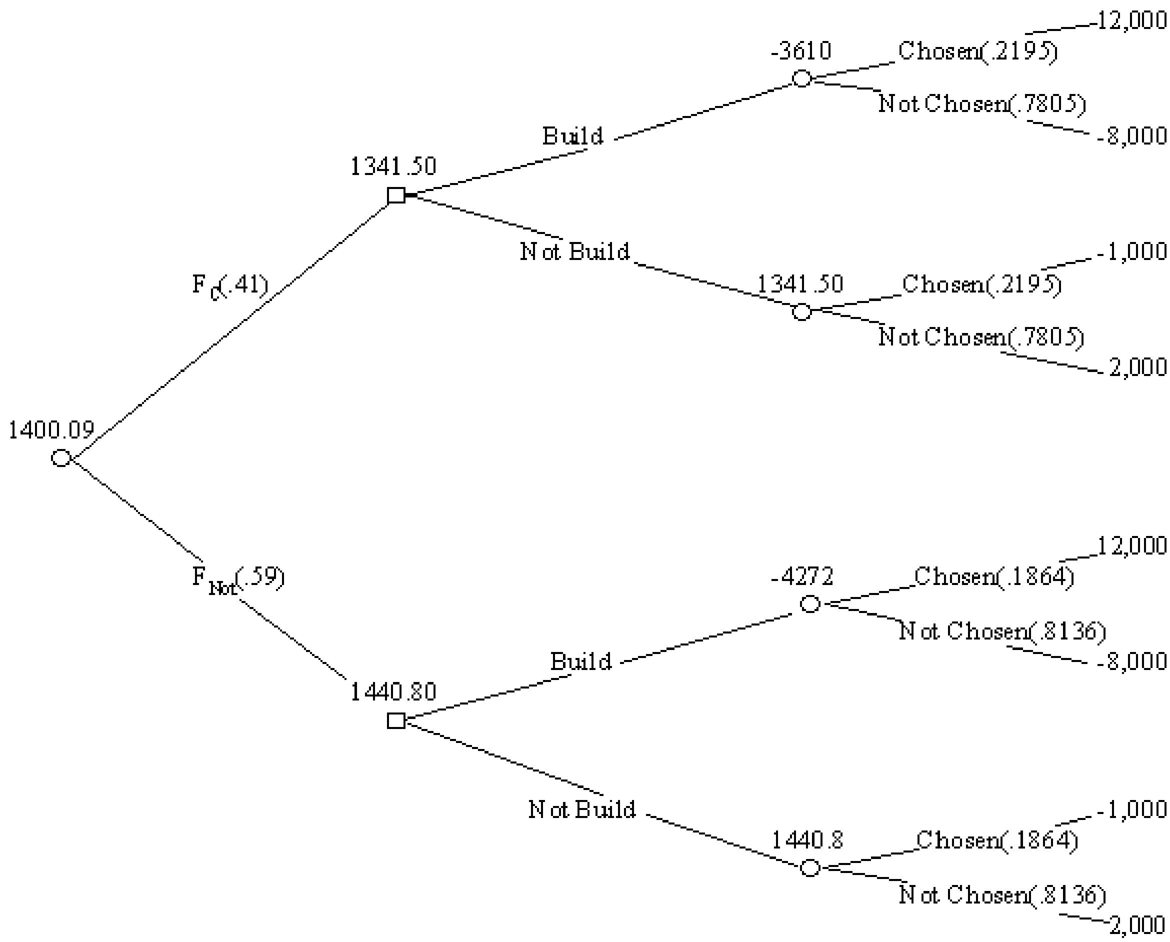
$P(F_C) = .410$

Forecast Not Chosen:

State	Prior	Cond.	Joint	Revised
Chosen	.20	.55	.110	.1864

Not Chosen	.80	.60	.480	.8136

$P(F_C) = .590$



Expected Value With Sample Information = **1,400.09**

EVSI = EVWSI - EMV = 1,400.09 - 1,400 = **.09**

16.19

	y	x_1	x_2	x_3
y	-	-.653	-.891	.821
x_1	-.653	-	.650	-.615
x_2	-.891	.650	-	-.688
x_3	.821	-.615	-.688	-

There appears to be some correlation between all pairs of the predictor variables, x_1, x_2, and x_3. All pairwise correlations between independent variables are in the .600 to .700 range.

16.20

	y	x_1	x_2	x_3	x_4
y	-	-.241	.621	.278	-.724
x_1	-.241	-	-.359	-.161	.325
x_2	.621	-.359	-	.243	-.442
x_3	.278	-.161	.243	-	-.278
x_4	-.724	.325	-.442	-.278	-

An examination of the intercorrelations of the predictor variables reveals that the highest pairwise correlation exists between variables x_2 and

x_4 (-.442). Other correlations between independent variables are less than .400. Multicollinearity may not be a serious problem in this regression analysis.

16.21 The stepwise regression analysis of problem 16.17 resulted in two of the three predictor variables being included in the model. The simple regression model yielded an R^2 of .833 jumping to .897 with the two predictors. The predictor intercorrelations are:

	Net Income	Dividends	Gain/Loss
Net Income	-	.682	.092
Dividends	.682	-	-.522
Gain/Loss	.092	-.522	-

An examination of the predictor intercorrelations reveals that Gain/Loss and Net Income have very little correlation, but Net Income and Dividends have a correlation of .682 and Dividends and Gain/Loss have a correlation of -.522. These correlations might suggest multicollinearity.

16.22 The intercorrelations of the predictor variables are:

Natural	Fuel

	Gas	Oil	Gasoline
Natural Gas	-	.570	.701
Fuel Oil	.570	-	.934
Gasoline	.701	.934	-

Each of these intercorrelations is not small. Of particular concern is the correlation between fuel oil and gasoline, which is .934. These two variables seem to be adding about the same predictability to the model. In the stepwise regression analysis only natural gas entered the procedure. Perhaps the overlapping information between natural gas and fuel oil and gasoline was such that fuel oil and gasoline did not have significant unique variance to add to the prediction.

16.23 The regression model is:

$$\hat{y} = 564 - 27.99\, x_1 - 6.155\, x_2 - 15.90\, x_3$$

$F = 11.32$ with $p = .003$, $s_e = 42.88$, $R^2 = .809$, adjusted $R^2 = .738$. For x_1, $t = -0.92$ with $p = .384$, for x_2, $t = -4.34$ with $p = .002$, for x_3, $t = -0.71$ with $p = .497$. Thus, only one of the three predictors, x_2, is a significant predictor in this model. This model has very good predictability ($R^2 = .809$). The gap between R^2 and adjusted R^2 underscores the fact that there are two nonsignificant predictors in this model. x_1 is a nonsignificant indicator variable.

16.24 The stepwise regression process included two steps. At step 1, x_1 entered the procedure producing the model:

$$\hat{y} = 1540 + 48.2\, x_1.$$

The R^2 at this step is .9112 and the t ratio is 11.55. At step 2, x_1^2 entered the procedure and x_1 remained in the analysis. The stepwise regression procedure stopped at this step and did not proceed. The final model was:

$$\hat{y} = 1237 + 136.1\, x_1 - 5.9\, x_1^2.$$

The R^2 at this step was .9723, the t ratio for x_1 was 7.89, and the t ratio for x_1^2 was - 5.14.

16.25 In this model with x_1 and the log of x_1 as predictors, only the log x_1 was a significant predictor of y. The stepwise procedure only went to step 1. The regression model was:

$\hat{y}$ = - 13.20 + 11.64 Log x_1. R^2 = .9617 and the t ratio of Log x_1 was 17.36. This model has very strong predictability using only the log of the x_1 variable.

16.26 The regression model is:

Grain = - 4.675 + 0.4732 Oilseed + 1.18 Livestock

The value of R^2 was .901 and adjusted R^2 = .877.

s_e = 1.761. F = 36.55 with p = .000.

t_{oilseed} = 3.74 with p = .006 and $t_{\text{livestock}}$ = 3.78 with p = .005. Both predictors are significant at α = .01. This is a model with strong predictability.

16.27 The stepwise regression procedure only used two steps. At step 1, Silver was the lone predictor. The value of R^2 was .5244. At step 2, Aluminum entered the model and Silver remained in the model. However, the R^2 jumped to .8204. The final model at step 2 was:

Gold = - 50.19 + 18.9 Silver +3.59 Aluminum.

The t values were: t_{Silver} = 5.43 and t_{Aluminum} = 3.85.

Copper did not enter into the process at all.

16.28 The regression model was:

Employment = 71.03 + 0.4620 NavalVessels + 0.02082 Commercial

$F = 1.22$ with $p = .386$ (not significant)

$R^2 = .379$ and adjusted $R^2 = .068$

The low value of adjusted R^2 indicates that the model has very low predictability. Both t values are not significant ($t_{\text{NavalVessels}} = 0.67$ with

$p = .541$ and $t_{\text{Commercial}} = 1.07$ with $p = .345$). Neither predictor is a significant predictor of employment.

16.29 There were four predictor variables. The stepwise regression procedure went three steps. The predictor, apparel, never entered in the stepwise process. At step 1, food entered the procedure producing a model with an R^2 of .84. At step 2, fuel oil entered and food remained. The R^2 increased to .95. At step 3, shelter entered the procedure and both fuel oil and food remained in the model. The R^2 at this step was .96. The final model was:

All = -1.0615 + 0.474 Food + 0.269 Fuel Oil + 0.249 Shelter

The t ratios were: t_{food} = 8.32, $t_{\text{fuel oil}}$ = 2.81, t_{shelter} = 2.56.

16.30 The stepwise regression process with these two independent variables only went one step. At step 1, Soybeans entered in producing the model,

Corn = - 2,962 + 5.4 Soybeans. The R^2 for this model was .7868.

The t ratio for Soybeans was 5.43. Wheat did not enter in to the analysis.

16.31 The regression model was:

Grocery = 76.23 + 0.08592 Housing + 0.16767 Utility

+ 0.0284 Transportation - 0.0659 Healthcare

F = 2.29 with p = .095 which is not significant at α = .05.

s_e = 4.416, R^2 = .315, and adjusted R^2 = .177.

Only one of the four predictors has a significant t ratio and that is Utility with t = 2.57 and p = .018. The ratios and their respective probabilities are:

t_{housing} = 1.68 with p = .109, $t_{\text{transportation}}$ = 0.17 with p = .87, and

$t_{\text{healthcare}}$ = - 0.64 with p = .53.

This model is very weak. Only the predictor, Utility, shows much promise in accounting for the grocery variability.

16.32 The output suggests that the procedure only went two steps.

At step 1, x_1 entered the model yielding an R^2 of .7539. At step 2,

x_2 entered the model and x_1 remained. The procedure stopped here with a final model of:

$$\hat{y} = 124.5 - 43.4\,x_1 + 1.36\,x_2$$

The R^2 for this model was .8059 indicating relatively strong predictability with two independent variables. Since there were four predictor variables, two of the variables did not enter the stepwise process.

16.33 Of the three predictors, x_2 is an indicator variable. An examination of the stepwise regression output reveals that there were three steps and that all three predictors end up in the final model. Variable x_3 is the strongest individual predictor of y and entered at step one resulting in an R^2 of .8124. At step 2, x_2 entered the process and variable x_3 remained in the model. The R^2 at this step was .8782. At step 3, variable x_1 entered the procedure. Variables x_3 and x_2 remained in the model. The final R^2 was .9407. The final model was:

$$\hat{y} = 87.89 + 0.071\,x_3 - 2.71\,x_2 - 0.256\,x_1$$

16.34 The R^2 for the full model is .321. After dropping out variable, x_3, the R^2 is still .321. Variable x_3 added virtually no information to the model. This is underscored by the fact that the *p*-value for the *t* test of the slope for x_3 is .878 indicating that there is no significance. The standard error of the estimate actually drops slightly after x_3 is removed from the model.

Chapter 17

Time-Series Forecasting and Index Numbers

LEARNING OBJECTIVES

This chapter discusses the general use of forecasting in business, several tools that are available for making business forecasts, and the nature of time series data, thereby enabling you to:

1. Gain a general understanding time series forecasting techniques.

2. Understand the four possible components of time-series data.

3. Understand stationary forecasting techniques.

4. Understand how to use regression models for trend analysis.

5. Learn how to decompose time-series data into their various elements and to

   forecast by using decomposition techniques

6. Understand the nature of autocorrelation and how to test for it.

7. Understand autoregression in forecasting.

CHAPTER TEACHING STRATEGY

Time series analysis attempts to determine if there is something inherent in the history of a variable that can be captured in a way that will help business analysts forecast the future values for the variable.

The first section of the chapter contains a general discussion about the various possible components of time-series data. It creates the setting against which the chapter later proceeds into trend analysis and seasonal effects. In addition, two measurements of forecasting error are presented so that students can measure the error of forecasts produced by the various techniques and begin to compare the merits of each.

A full gamut of time series forecasting techniques has been presented beginning with the most naïve models and progressing through averaging models and exponential smoothing. An attempt is made in the section on exponential smoothing to show the student, through algebra, why it is called by that name. Using the derived equations and a few selected values for alpha, the student is shown how past values and forecasts are smoothed in the prediction of future values. The more advanced smoothing techniques are briefly introduced in later sections but are explained in much greater detail on WileyPLUS.

Trend is solved for next using the time periods as the predictor variable. In this chapter both linear and quadratic trends are explored and compared. There is a brief introduction to Holt's two-parameter exponential smoothing method that includes trend. A more detailed explanation of Holt's method is available on WileyPLUS. The trend analysis section is placed earlier in the chapter than seasonal effects because finding seasonal effects makes more sense when there are no trend effects in the data or the trend effect has been removed.

Section 17.4 includes a rather classic presentation of time series decomposition only it is done on a smaller set of data so as not to lose the reader. It was felt that there may be a significant number of instructors who want to show how a time series of data can be broken down into the components of trend, cycle, and seasonality. This text assumes a multiplicative model rather than an additive model. The main example used throughout this section is a database of 20 quarters of actual data on Household Appliances. A graph of these data is presented both before and after deseasonalization so that the student can visualize what happens when the seasonal effects are removed. First, 4-quarter centered moving averages are computed which dampen out the seasonal and irregular effects leaving trend and cycle. By dividing the original data by these 4-quarter centered moving averages (trend·cycle), the researcher is left with seasonal effects and irregular effects. By casting out the high and low values and averaging the seasonal effects for each quarter, the irregular effects are removed.

In regression analysis involving data over time, autocorrelation can be a problem. Because of this, section 17.5 contains a discussion on autocorrelation and autoregression. The Durbin-Watson test is presented as a mechanism for testing for the presence of autocorrelation. Several possible ways of overcoming the autocorrelation problem are presented such as the addition of independent variables, transforming variables, and autoregressive models.

The last section in this chapter is a classic presentation of Index Numbers. This

section is essentially a shortened version of an entire chapter on Index Numbers. It includes most of the traditional topics of simple index numbers, unweighted aggregate price index numbers, weighted price index numbers, Laspeyres price indexes, and Paasche price indexes.

CHAPTER OUTLINE

17.1 Introduction to Forecasting

Time Series Components

The Measurement of Forecasting Error

Error

Mean Absolute Deviation (MAD)

Mean Square Error (MSE)

17.2 Smoothing Techniques

Naïve Forecasting Models

Averaging Models

Simple Averages

Moving Averages

Weighted Moving Averages

Exponential Smoothing

17.3 Trend Analysis

Linear Regression Trend Analysis

Regression Trend Analysis Using Quadratic Models

Holt’s Two-Parameter Exponential Smoothing Method

17.4 Seasonal Effects

Decomposition

Finding Seasonal Effects with the Computer

KEY TERMS

Autocorrelation
Autoregression
Averaging Models
Cycles
Cyclical Effects
Decomposition
Deseasonalized Data
Durbin-Watson Test
Error of an Individual Forecast
Exponential Smoothing
First-Difference Approach
Forecasting
Forecasting Error
Index Number
Irregular Fluctuations
Laspeyres Price Index
Mean Absolute Deviation (MAD)
Mean Squared Error (MSE)
Moving Average
Naïve Forecasting Methods
Paasche Price Index
Seasonal Effects
Serial Correlation
Simple Average
Simple Average Model
Simple Index Number
Smoothing Techniques
Stationary
Time-Series Data
Trend
Unweighted Aggregate Price Index Number
Weighted Aggregate Price Index Number
Weighted Moving Average

SOLUTIONS TO PROBLEMS IN CHAPTER 17

17.1

Period	e	$\lvert e \rvert$	e^2
1	2.30	2.30	5.29
2	1.60	1.60	2.56
3	-1.40	1.40	1.96
4	1.10	1.10	1.21
5	0.30	0.30	0.09
6	-0.90	0.90	0.81
7	-1.90	1.90	3.61
8	-2.10	2.10	4.41
9	0.70	0.70	0.49
Total	-0.30	12.30	20.43

MAD = $\dfrac{\sum \lvert e \rvert}{no.forecasts} = \dfrac{12.30}{9}$ = **1.367**

MSE = $\dfrac{\sum e^2}{no.forecasts} = \dfrac{20.43}{9}$ = **2.27**

17.2

Period	Value	F	e	$\lvert e \rvert$	e^2
1	202	-			
2	191	202	-11	11	121
3	173	192	-19	19	361
4	169	181	-12	12	144
5	171	174	- 3	3	9
6	175	172	3	3	9
7	182	174	8	8	64
8	196	179	17	17	289
9	204	189	15	15	225
10	219	198	21	21	441
11	227	211	16	16	256
	Total		35	125	1919

$$\text{MAD} = \frac{\sum \lvert e \rvert}{no.forecasts} = \frac{125.00}{10} = \mathbf{12.5}$$

$$\text{MSE} = \frac{\sum e^2}{no.forecasts} = \frac{1{,}919}{10} = \mathbf{191.9}$$

17.3

Period	Value	F	e	$\lvert e\rvert$	e^2
1	19.4	16.6	2.8	2.8	7.84
2	23.6	19.1	4.5	4.5	20.25
3	24.0	22.0	2.0	2.0	4.00
4	26.8	24.8	2.0	2.0	4.00
5	29.2	25.9	3.3	3.3	10.89
6	35.5	28.6	6.9	6.9	47.61
	Total		21.5	21.5	94.59

$$\text{MAD} = \frac{\sum |e|}{\text{No. Forecasts}} = \frac{21.5}{6} = \mathbf{3.583}$$

$$\text{MSE} = \frac{\sum e^2}{\text{No. Forecasts}} = \frac{94.59}{6} = \mathbf{15.765}$$

17.4

Year	Acres	Forecast	e	$\lvert e\rvert$	e^2
1	140,000	-	-	-	-
2	141,730	140,000	1730	1730	2,992,900
3	134,590	141,038	-6448	6448	41,576,704
4	131,710	137,169	-5459	5459	29,800,681

5	131,910	133,894	-1984	1984	3,936,256
6	134,250	132,704	1546	1546	2,390,116
7	135,220	133,632	1588	1588	2,521,744
8	131,020	134,585	-3565	3565	12,709,225
9	120,640	132,446	-11806	11806	139,381,636
10	115,190	125,362	-10172	10172	103,469,584
11	114,510	119,259	-4749	4749	22,553,001
	Total		-39,319	49047	361,331,847

$$\text{MAD} = \frac{\sum |e|}{\text{No. Forecasts}} = \frac{49{,}047}{10} = \mathbf{4{,}904.7}$$

$$\text{MSE} = \frac{\sum e^2}{\text{No. Forecasts}} = \frac{361{,}331{,}847}{10} = \mathbf{36{,}133{,}184.7}$$

17.5 a.)

<u>4-mo. mov. avg.</u>	<u>error</u>
44.75	14.25
52.75	13.25
61.50	9.50
64.75	21.25

70.50	30.50
81.00	16.00

b.)

4-mo. wt. mov. avg.	error
53.25	5.75
56.375	9.625
62.875	8.125
67.25	18.75
76.375	24.625
89.125	7.875

c.) difference in errors

14.25 - 5.75 = 8.5

3.626

1.375

2.5

5.875

8.125

In each time period, the four-month moving average produces greater errors of forecast than the four-month weighted moving average.

17.6

Period	Value	$F(\alpha = .1)$	Error	$F(\alpha = .8)$	Error	Difference
1	211					
2	228	211		211		
3	236	213	23	225	11	12
4	241	215	26	234	7	19
5	242	218	24	240	2	22
6	227	220	7	242	-15	22
7	217	221	-4	230	-13	9
8	203	221	-18	220	-17	0

Using alpha of .1 produced forecasting errors that were larger than those using alpha = .8 for the first three forecasts. For the next two forecasts (periods 6

and 7), the forecasts using alpha = .1 produced smaller errors. Each exponential smoothing model produced nearly the same amount of error in forecasting the value for period 8. There is no strong argument in favor of either model.

17.7

Period	Value	α =.3	Error	α =.7	Error	3-mo.avg.	Error
1	9.4						
2	8.2	9.4	-1.2	9.4	-1.2		
3	7.9	9.0	-1.1	8.6	-0.7		
4	9.0	8.7	0.3	8.1	0.9	8.5	0.5
5	9.8	8.8	1.0	8.7	1.1	8.4	1.4
6	11.0	9.1	1.9	9.5	1.5	8.9	1.1
7	10.3	9.7	0.6	10.6	-0.3	9.9	0.4
8	9.5	9.9	-0.4	10.4	-0.9	10.4	-0.9
9	9.1	9.8	-0.7	9.8	-0.7	10.3	-1.2

17.8

Year	Orders	(a) $F(a)$	(c) $e(a)$	(b) $F(b)$	(c) $e(b)$
1	2512.7				
2	2739.9				
3	2874.9				
4	2934.1				
5	2865.7				
6	2978.5	2785.46	193.04	2852.36	126.14
7	3092.4	2878.62	213.78	2915.49	176.91
8	3356.8	2949.12	407.68	3000.63	356.17
9	3607.6	3045.50	562.10	3161.94	445.66
10	3749.3	3180.20	569.10	3364.41	384.89
11	3952.0	3356.92	595.08	3550.76	401.24
12	3949.0	3551.62	397.38	3740.97	208.03
13	4137.0	3722.94	414.06	3854.64	282.36

17.9

Year	No.Issues	$F(\alpha=.2)$	$\lvert e \rvert$	$F(\alpha=.9)$	$\lvert e \rvert$
1	332	-			
2	694	332.0	362.0	332.0	362.0
3	518	404.4	113.6	657.8	139.8
4	222	427.1	205.1	532.0	310.0

5	209	386.1	177.1	253.0	44.0
6	172	350.7	178.7	213.4	41.4
7	366	315.0	51.0	176.1	189.9
8	512	325.2	186.8	347.0	165.0
9	667	362.6	304.4	495.5	171.5
10	571	423.5	147.5	649.9	78.9
11	575	453.0	122.0	578.9	3.9
12	865	477.4	387.6	575.4	289.6
13	609	554.9	54.1	836.0	227.0

$\sum|e| = 2289.9$ $\qquad$ $\sum|e| = 2023.0$

For $\alpha = .2$, MAD = $\frac{2289.9}{12}$ = **190.8**

For $\alpha = .9$, MAD = $\frac{2023.0}{12}$ = **168.6**

α = .9 produces a smaller mean average error.

17.10 Simple Regression Trend Model:

$\hat{y}$ = 37,969 + 9899.1 Period

F = 1603.11 (p = .000), R^2 = .988, adjusted R^2 = .988,

s_e = 6,861, t = 40.04 (p = .000)

Quadratic Regression Trend Model:

$\hat{y}$ = 35,769 + 10,473 Period - 26.08 Period^2

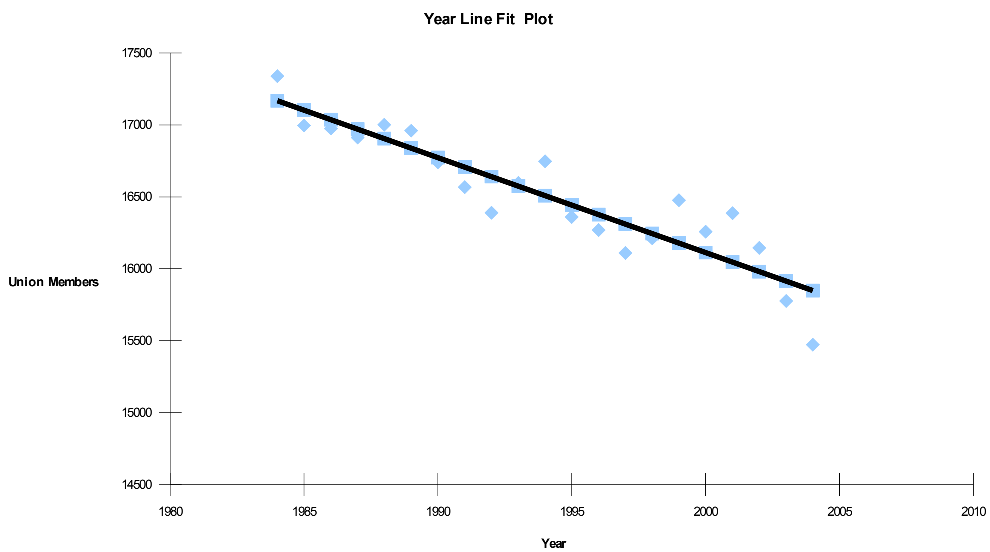
F = 772.71 (p = .000), R^2 = .988, adjusted R^2 = .987

s_e = 6,988, t_{period} = 9.91 (p = .000), $t_{periodsq}$ = -0.56 (p = .583)

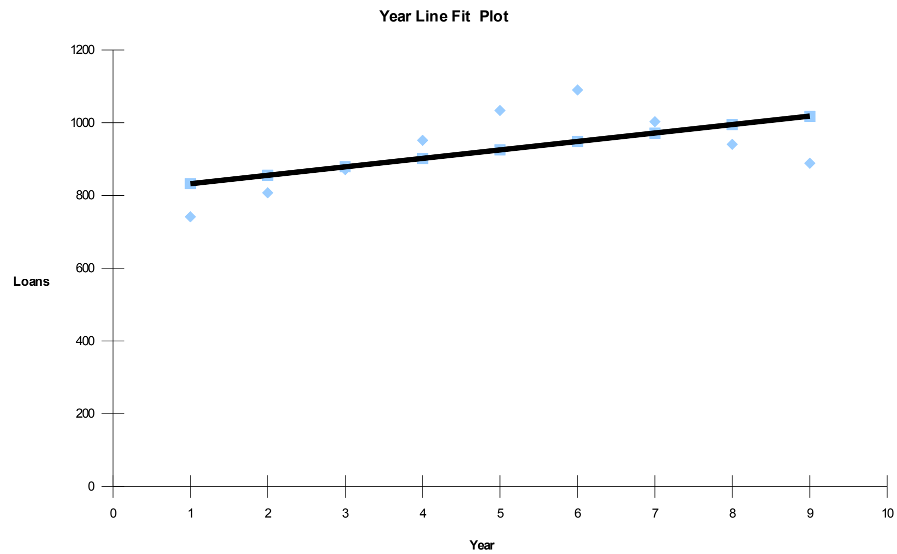
The simple linear regression trend model is superior, the period^2 variable is not a significant addition to the model.

17.11 Trend line: Members = 148,141.8 – 66.0143 Year

R^2 = 83.0% s_e = 190.1374 F = 92.82, reject the null hypothesis.


Year Line Fit Plot
17500
17000
16500
16000
15500
15000
14500
Union Members
1980
1985
1990
1995
2000
2005
2010
Year

17.12



Trend Model:

Loans = 809.2611 + 23.18333 Year

R^2 = 33.0 adjusted R^2 = 23.4 s_e = 96.80

t = 1.86 (p = .106) F = 3.44 (p = .1.06)

Quadratic Model:

Loans = 561.1738 + 158.5037 Year – 13.5320 Year^2

R^2 = 90.6 adjusted R^2 = 87.5 s_e = 39.13

t_{year} = 6.93 (p = .0004)

t_{yearsq} = -6.07 (p = .000)

$F = 28.95$ ($p = .0008$)

The graph indicates a quadratic fit rather than a linear fit. The quadratic model produced an $R^2 = 90.6$ compared to $R^2 = 33.0$ for linear trend indicating a much better fit for the quadratic model. In addition, the standard error of the estimate drops from 96.80 to 39.13 with the quadratic model along with the *t* values becoming significant.

17.13

Month	Broccoli	12-Mo. Mov.Tot.	2-Yr.Tot.	TC	SI
Jan.(yr. 1)	132.5				
Feb.	164.8				
Mar.	141.2				
Apr.	133.8				
May	138.4				
June	150.9				
		1655.2			
July	146.6		3282.8	136.78	93.30
		1627.6			
Aug.	146.9		3189.7	132.90	90.47
		1562.1			
Sept.	138.7		3085.0	128.54	92.67
		1522.9			
Oct.	128.0		3034.4	126.43	98.77
		1511.5			
Nov.	112.4		2996.7	124.86	111.09
		1485.2			
Dec.	121.0		2927.9	122.00	100.83

		1442.7			
Jan.(yr. 2)	104.9		2857.8	119.08	113.52
		1415.1			
Feb.	99.3		2802.3	116.76	117.58
		1387.2			
Mar.	102.0		2750.6	114.61	112.36
		1363.4			
Apr.	122.4		2704.8	112.70	92.08
		1341.4			
May	112.1		2682.1	111.75	99.69
		1340.7			
June	108.4		2672.7	111.36	102.73
		1332.0			
July	119.0				
Aug.	119.0				
Sept.	114.9				
Oct.	106.0				
Nov.	111.7				
Dec.	112.3				

17.14

Month	Ship	12m tot	2yr tot	TC	SI	TCI	T	C
Jan(Yr1)	1891					1952.50	2042.72	
Feb	1986					1975.73	2049.87	
Mar	1987					1973.78	2057.02	
Apr	1987					1972.40	2064.17	
May	2000					1976.87	2071.32	
June	2082					1982.67	2078.46	
		23822						
July	1878		47689	1987.04	94.51	1970.62	2085.61	94.49
		23867						
Aug	2074		47852	1993.83	104.02	2011.83	2092.76	96.13
		23985						
Sept	2086		48109	2004.54	104.06	2008.47	2099.91	95.65
		24124						
Oct	2045		48392	2016.33	101.42	1969.76	2107.06	93.48

24268

Nov 1945 48699 2029.13 95.85 2024.57
2114.20 95.76

24431

Dec 1861 49126 2046.92 90.92 2002.80
2121.35 94.41

24695

Jan(Yr2) 1936 49621 2067.54 93.64 1998.97 2128.50
93.91

24926

Feb 2104 49989 2082.88 101.01 2093.12
2135.65 98.01

25063

Mar 2126 50308 2096.17 101.42 2111.85
2142.80 98.56

25245

Apr 2131 50730 2113.75 100.82 2115.35 2149.94
98.39

25485

May 2163 51132 2130.50 101.53 2137.99 2157.09
99.11

25647

June 2346 51510 2146.25 109.31 2234.07 2164.24
103.23

25863

July 2109 51973 2165.54 97.39 2213.01 2171.39
101.92

26110

Aug 2211 52346 2181.08 101.37 2144.73
2178.54 98.45

26236

Sept 2268 52568 2190.33 103.55 2183.71
2185.68 99.91

26332

Oct 2285 52852 2202.17 103.76 2200.93 2192.83
100.37

26520

Nov 2107 53246 2218.58 94.97 2193.19
2199.98 99.69

26726

Dec 2077 53635 2234.79 92.94 2235.26
2207.13 101.27 26909

Jan(Yr3) 2183 53976 2249.00 97.07 2254.00 2214.28
101.79 27067

Feb 2230 54380 2265.83 98.42 2218.46
2221.42 99.87 27313

Mar 2222 54882 2286.75 97.17 2207.21
2228.57 99.04 27569

Apr 2319 55355 2306.46 100.54 2301.97 2235.72
102.96 27786

May 2369 55779 2324.13 101.93 2341.60 2242.87
104.40 27993

June 2529 56186 2341.08 108.03 2408.34
2250.02 107.04

28193

July 2267 56539 2355.79 96.23 2378.80 2257.17
105.39

28346

Aug	2457		56936	2372.33	103.57	2383.35	2264.31	105.26
		28590						
Sept	2524		57504	2396.00	105.34	2430.19	2271.46	106.99
		28914						
Oct	2502		58075	2419.79	103.40	2409.94	2278.61	105.76
		29161						
Nov	2314		58426	2434.42	95.05	2408.66	2285.76	105.38
		29265						
Dec	2277		58573	2440.54	93.30	2450.50	2292.91	106.87
		29308						
Jan(Yr4)	2336		58685	2445.21	95.53	2411.98	2300.05	104.87
		29377						
Feb	2474		58815	2450.63	100.95	2461.20	2307.20	106.67
		29438						
Mar	2546		58806	2450.25	103.91	2529.06	2314.35	109.28
		29368						
Apr	2566		58793	2449.71	104.75	2547.15	2321.50	109.72
		29425						
May	2473		58920	2455.00	100.73	2444.40	2328.65	104.97

29495

June 2572 59018 2459.08 104.59 2449.29 2335.79
104.86

29523

July 2336 59099 2462.46 94.86 2451.21 2342.94
104.62

29576

Aug 2518 59141 2464.21 102.18 2442.53
2350.09 103.93

29565

Sept 2454 59106 2462.75 99.64 2362.80
2357.24 100.24

29541

Oct 2559 58933 2455.54 104.21 2464.84
2364.39 104.25

29392

Nov 2384 58779 2449.13 97.34 2481.52
2371.53 104.64

29387

Dec 2305 58694 2445.58 94.25 2480.63
2378.68 104.29

29307

Jan(Yr5) 2389 58582 2440.92 97.87 2466.70 2385.83
103.39

29275

Feb 2463 58543 2439.29 100.97 2450.26
2392.98 102.39

29268

Mar 2522 58576 2440.67 103.33 2505.22
2400.13 104.38

29308

Apr 2417 58587 2441.13 99.01 2399.25 2407.27
99.67

29279

May 2468 58555 2439.79 101.16 2439.46 2414.42
101.04

29276

June 2492 58458 2435.75 102.31 2373.11
2421.57 98.00

29182

July 2304 58352 2431.33 94.76 2417.63 2428.72
99.54

29170

Aug 2511 58258 2427.42 103.44 2435.74
2435.87 99.99

29088

Sept 2494 57922 2413.42 103.34 2401.31
2443.01 98.29

28834

Oct 2530 57658 2402.42 105.31 2436.91
2450.16 99.46

28824

Nov 2381 57547 2397.79 99.30 2478.40
2457.31 100.86

28723

Dec 2211 57400 2391.67 92.45 2379.47 2464.46
96.55

28677

Jan(Yr6) 2377 57391 2391.29 99.40 2454.31 2471.61
99.30

28714

Feb 2381 57408 2392.00 99.54 2368.68
2478.76 95.56

28694

Mar 2268 57346 2389.42 94.92 2252.91
2485.90 90.63

28652

Apr 2407 57335 2388.96 100.76 2389.32 2493.05
95.84

28683

May 2367 57362 2390.08 99.03 2339.63 2500.20
93.58

28679

June 2446 57424 2392.67 102.23 2329.30 2507.35
92.90

28745

July 2341

Aug 2491

Sept 2452

Oct 2561

Nov 2377

Dec 2277

Seasonal Indexing:

Month Year1 Year2 Year3 Year4 Year5 Year6 Index

Jan		93.64	97.07	95.53	97.87	99.40	96.82
Feb		101.01	98.42	100.95	100.97	99.54	100.49
Mar		101.42	97.17	103.91	103.33	94.92	100.64
Apr		100.82	100.54	104.75	99.01	100.76	100.71
May		101.53	101.93	100.73	101.16	99.03	101.14
June		109.31	108.03	104.59	102.31	102.23	104.98
July	94.51	97.39	96.23	94.86	94.76		95.28
Aug	104.02	101.37	103.57	102.18	103.44		103.06
Sept	104.60	103.55	105.34	99.64	103.34		103.83
Oct	101.42	103.76	103.40	104.21	105.31		103.79
Nov	95.85	94.97	95.05	97.24	99.30		96.05
Dec	90.92	92.94	93.30	94.25	92.45		92.90
Total							1199.69

Adjust each seasonal index by 1.0002584

Final Seasonal Indexes:

Month	Index
Jan	96.85
Feb	100.52
Mar	100.67
Apr	100.74
May	101.17

June	105.01
July	95.30
Aug	103.09
Sept	103.86
Oct	103.82
Nov	96.07
Dec	92.92

Regression Output for Trend Line: $\hat{Y}$ = 2035.58 + 7.1481 X

R^2 = .682, s_e = 102.9

Note: Trend Line was determined after seasonal effects were removed (based on TCI column).

17.15 Regression Analysis

The regression equation is: Food = 1.4228 + 0.4925 Housing

Predictor	Coef	*t*-ratio	*p*
Constant	1.4228	4.57	0.0001
Shelter	0.4925	7.99	0.0000

s = 0.939 R-sq = 72.7% R-sq(adj) = 71.5%

Food	Housing	$\hat{Y}$	e	e^2
8.5	15.7	9.1555	-0.6555	0.4296

7.8	11.5	7.0868	0.7132	0.5086
4.1	7.2	4.9690	-0.8690	0.7551
2.3	2.7	2.7526	-0.4526	0.2048
3.7	4.1	3.4421	0.2579	0.0665
2.3	4.0	3.3929	-1.0929	1.1944
3.3	3.0	2.9004	0.3996	0.1597
4.0	3.0	2.9004	1.0996	1.2092
4.1	3.8	3.2944	0.8056	0.6490
5.7	3.8	3.2944	2.4056	5.7870
5.8	4.5	3.6391	2.1609	4.6693
3.6	4.0	3.3929	0.2071	0.0429
1.4	2.9	2.8511	-1.4511	2.1057
2.1	2.7	2.7526	-0.6526	0.4259
2.3	2.5	2.6541	-0.3541	0.1254
2.8	2.6	2.7033	0.0967	0.0093
3.2	2.9	2.8511	0.3489	0.1217
2.6	2.6	2.7033	-0.1033	0.0107
2.2	2.3	2.5556	-0.3556	0.1264
2.2	2.2	2.5063	-0.3063	0.0938
2.3	3.5	3.1466	-0.8466	0.7168
3.1	4.0	3.3929	-0.2929	0.0858
1.8	2.2	2.5063	-0.7063	0.4989
2.1	2.5	2.6541	-0.5541	0.3070
3.4	2.5	2.6541	0.7459	0.5564
2.5	3.3	3.0481	-0.5481	0.3004
			Total	21.1603

$\sum(e_t - e_{t-1})^2$ = 1.873 + 2.503 + 0.173 + 0.505 + 1.825 + 2.228 + 0.490 +

0.0864 + 2.560 + 0.060 + 3.817 + 2.750 + 0.638 + 0.089 +

0.203 + 0.064 + 0.205 + 0.064 + 0.205 + 0.064 + 0.002 +

0.292 + 0.307 + 0.171 + 0.023 + 1.690 + 1.674 = 24.561

$\sum e^2$ = 21.160

$$D = \frac{\sum(e_t - e_{t-1})^2}{\sum e^2} = \frac{24.561}{21.160} = \mathbf{1.16}$$

Critical values of *D*: Using 1 independent variable, *n* = 26, and α = .05,

d_L = 1.30 and d_U = 1.46

Since *D* = 1.16 is less than d_L, the decision is to reject the null hypothesis.

There is significant autocorrelation.

17.16 Regression Analysis

The regression equation is: Food = 3.1286 - 0.2003 First Diff in Housing

Predictor	Coef	*t*-ratio	*p*
Constant	3.1286	10.30	0.000
First Diff	-.2003	-1.09	0.287

s = 1.44854 R-sq = 4.9% R-sq(adj) = 0.8%

Food	Housing	First Diff in Housing
8.5	15.7	-
7.8	11.5	-4.2
4.1	7.2	-4.3
2.3	2.7	-4.5
3.7	4.1	1.4
2.3	4.0	-0.1
3.3	3.0	-1.0
4.0	3.0	0.0
4.1	3.8	0.8
5.7	3.8	0.0
5.8	4.5	0.7
3.6	4.0	-0.5
1.4	2.9	-1.1
2.1	2.7	-0.2
2.3	2.5	-0.2
2.8	2.6	0.1
3.2	2.9	0.3
2.6	2.6	-0.3
2.2	2.3	-0.3
2.2	2.2	-0.1
2.3	3.5	1.3
3.1	4.0	0.5

1.8	2.2	-1.8
2.1	2.5	0.3
3.4	2.5	0.0
2.5	3.3	0.8

Critical values of D: Using 1 independent variable, n = 25, and α = .05,

d_L = 1.29 and d_U = 1.46

Since D = 1.04 is less than d_L, the decision is to reject the null hypothesis.

There is significant autocorrelation.

17.17 The regression equation is:

Failed Bank Assets = 1,379 + 136.68 Number of Failures

for x= 150: $\hat{y}$ = 21,881 (million $)

R^2 = 37.9% adjusted R^2 = 34.1% s_e = 13,833 F = 9.78, p = .006

The Durbin Watson statistic for this model is:

D = 2.49

The critical table values for k = 1 and n = 18 are d_L = 1.16 and d_U = 1.39. Since the observed value of D = 2.49 is above d_U, the decision is to fail to reject the null hypothesis. There is no significant autocorrelation.

Failed Bank Assets	Number of Failures	$\hat{y}$	e	e^2
8,189	11	2,882.8	5,306.2	28,155,356
104	7	2,336.1	-2,232.1	4,982,296
1,862	34	6,026.5	-4,164.5	17,343,453
4,137	45	7,530.1	-3,393.1	11,512,859
36,394	79	12,177.3	24,216.7	586,449,390
3,034	118	17,507.9	-14,473.9	209,494,371
7,609	144	21,061.7	-13,452.7	180,974,565
7,538	201	28,852.6	-21,314.6	454,312,622
56,620	221	31,586.3	25,033.7	626,687,597
28,507	206	29,536.0	- 1,029.0	1,058,894
10,739	159	23,111.9	-12,372.9	153,089,247
43,552	108	16,141.1	27,410.9	751,357,974

16,915	100	15,047.6	1,867.4	3,487,085
2,588	42	7,120.0	- 4,532.0	20,539,127
825	11	2,882.8	- 2,057.8	4,234,697
753	6	2,199.4	- 1,446.4	2,092,139
186	5	2,062.7	- 1,876.7	3,522,152
27	1	1,516.0	- 1,489.0	2,217,144

17.18

Failed Bank Assets	Number of Failures	First Diff Failures
8,189	11	-
104	7	-4
1,862	34	27
4,137	45	11
36,394	79	34
3,034	118	39
7,609	144	26
7,538	201	57
56,620	221	20
28,507	206	-15
10,739	159	-47
43,552	108	-51
16,915	100	-8
2,588	42	-58
825	11	-31
753	6	-5
186	5	-1
27	1	-4

The regression equation is:

Failed Bank Assets = 13,019 –7.3 First Diff Failures

R^2 = 0.0%   adjusted R^2 = 0.0% s_e = 18,091.7 F = 0.00, p = .958

The Durbin Watson statistic for this model is:

D = 1.57

The critical table values for k = 1 and n = 17 are d_L = 1.13 and d_U = 1.38. Since the observed value of D = 1.57 is above d_U, the decision is to fail to reject the null hypothesis. There is no significant autocorrelation.

17.19

Starts	lag1	lag2
333.0	*	*

270.4	333.0	*
281.1	270.4	333.0
443.0	281.1	270.4
432.3	443.0	281.1
428.9	432.3	443.0
443.2	428.9	432.3
413.1	443.2	428.9
391.6	413.1	443.2
361.5	391.6	413.1
318.1	361.5	391.6
308.4	318.1	361.5
382.2	308.4	318.1
419.5	382.2	308.4
453.0	419.5	382.2
430.3	453.0	419.5
468.5	430.3	453.0
464.2	468.5	430.3
521.9	464.2	468.5
550.4	521.9	464.2
529.7	550.4	521.9
556.9	529.7	550.4
606.5	556.9	529.7
670.1	606.5	556.9
745.5	670.1	606.5
756.1	745.5	670.1

826.8 756.1 745.5

The model with 1 lag:

Housing Starts = -8.87 + 1.06 lag 1

F = 198.67 p = .000 R^2 = 89.2% adjusted R^2 = 88.8% s_e = 48.52

The model with 2 lags:

Housing Starts = 13.66 + 1.0569 lag 2

F = 72.36 p = .000 R^2 = 75.9% adjusted R^2 = 74.8% S_e = 70.84

The model with 1 lag is the best model with a strong R^2 = 89.2%. The model

with 2 lags is relatively strong also.

17.20 The autoregression model is: Juice = 552 + 0.645 Juicelagged2

The F value for this model is 27.0 which is significant at alpha = .001.

The value of R^2 is 56.2% which denotes modest predictability. The adjusted R^2 is 54.2%. The standard error of the estimate is 216.6. The Durbin-Watson statistic is 1.70 indicating that there is no significant autocorrelation in this model.

17.21

Year	Price	a.) Index_{1950}	b.) Index_{1980}
1950	22.45	100.0	32.2
1955	31.40	139.9	45.0
1960	32.33	144.0	46.4
1965	36.50	162.6	52.3
1970	44.90	200.0	64.4
1975	61.24	272.8	87.8
1980	69.75	310.7	100.0
1985	73.44	327.1	105.3
1990	80.05	356.6	114.8
1995	84.61	376.9	121.3
2000	87.28	388.8	125.1
2005	89.56	398.9	128.4

17.22

Year	Patents	Index
1980	66.2	66.8
1981	71.1	71.7
1982	63.3	63.9
1983	62.0	62.6
1984	72.7	73.4
1985	77.2	77.9
1986	76.9	77.6
1987	89.4	90.2
1988	84.3	85.1
1989	102.5	103.4
1990	99.1	100.0
1991	106.7	107.7
1992	107.4	108.4
1993	109.7	110.7
1994	113.6	114.8
1995	113.8	115.3
1996	121.7	122.8
1997	124.1	125.2
1998	163.1	164.6
1999	169.1	170.6
2000	176.0	177.6
2001	184.0	185.7

2002	184.4	186.1
2003	187.0	188.7
2004	181.3	182.9

17.23 Year

	1993	1999	2005
	1.53	1.40	2.17
	2.21	2.15	2.51
	1.92	2.68	2.60
	3.38	3.10	4.00
Totals	9.04	9.33	11.28

$$\text{Index}_{1993} = \frac{9.04}{9.04}(100) = \mathbf{100.0}$$

$$\text{Index}_{1999} = \frac{9.33}{9.04}(100) = \mathbf{103.2}$$

$$\text{Index}_{2005} = \frac{11.28}{9.04}(100) = \mathbf{124.8}$$

17.24 Year

	1998	1999	2000	2001	2002	2003	2004	2005	2006
	1.10	1.16	1.23	1.23	1.08	1.56	1.85	2.59	2.89
	1.58	1.61	1.78	1.77	1.61	1.71	1.90	2.05	2.08
	1.80	1.82	1.98	1.96	1.94	1.90	1.92	1.94	1.96
	7.95	7.96	8.24	8.21	8.19	8.05	8.12	8.10	8.24
Totals	12.43	12.55	13.23	13.17	12.82	13.22	13.79	14.68	15.17

$$\text{Index}_{1998} = \frac{12.43}{13.23}(100) = 94.0$$

$$\text{Index}_{1999} = \frac{12.55}{13.23}(100) = 94.9$$

$$\text{Index}_{2000} = \frac{13.23}{13.23}(100) = 100.0$$

$$\text{Index}_{2001} = \frac{13.17}{13.23}(100) = 99.5$$

$$\text{Index}_{2002} = \frac{12.82}{13.23}(100) = 100.0$$

$$\text{Index}_{2003} = \frac{13.22}{13.23}(100) = 101.0$$

$$\text{Index}_{2004} = \frac{13.79}{13.23}(100) = 106.4$$

$$\text{Index}_{2005} = \frac{14.68}{13.23}(100) = 111.0$$

$$\text{Index}_{2006} = \frac{15.17}{13.23}(100) = 114.7$$

17.25

Item	Quantity 2000	Price 2000	Price 2004	Price 2005	Price 2006
1	21	0.50	0.67	0.68	0.71
2	6	1.23	1.85	1.90	1.91
3	17	0.84	0.75	0.75	0.80
4	43	0.15	0.21	0.25	0.25

	$P_{2000}Q_{2000}$	$P_{2004}Q_{2000}$	$P_{2005}Q_{2000}$	$P_{2006}Q_{2000}$
	10.50	14.07	14.28	14.91
	7.38	11.10	11.40	11.46
	14.28	12.75	12.75	13.60
	6.45	9.03	10.75	10.75
Totals	38.61	46.95	49.18	50.72

$$\text{Index}_{2000} = \frac{\sum P_{2004}Q_{2000}}{\sum P_{2000}Q_{2000}} = \frac{46.95}{38.61}(100) = \mathbf{121.6}$$

$$\text{Index}_{2001} = \frac{\sum P_{2005}Q_{2000}}{\sum P_{2000}Q_{2000}} = \frac{49.18}{38.61}(100) = \mathbf{127.4}$$

$$\text{Index}_{2002} = \frac{\sum P_{2006}Q_{2000}}{\sum P_{2000}Q_{2000}} = \frac{50.72}{38.61}(100) = \mathbf{131.4}$$

17.26

Item	Price 2000	Price 2005	Quantity 2005	Price 2006	Quantity 2006

1	22.50	27.80	13	28.11	12
2	10.90	13.10	5	13.25	8
3	1.85	2.25	41	2.35	44

	$P_{2000}Q_{2005}$	$P_{2000}Q_{2006}$	$P_{2005}Q_{2005}$	$P_{2006}Q_{2006}$
	292.50	270.00	361.40	337.32
	54.50	87.20	65.50	106.00
	75.85	81.40	92.25	103.40
Totals	422.85	438.60	519.15	546.72

$$\text{Index}_{2005} = \frac{\sum P_{2005}Q_{2005}}{\sum P_{2000}Q_{2005}}(100) = \frac{519.15}{422.85}(100) = \mathbf{122.8}$$

$$\text{Index}_{2006} = \frac{\sum P_{2006}Q_{2006}}{\sum P_{2000}Q_{2006}}(100) = \frac{546.72}{438.60}(100) = \mathbf{124.7}$$

17.27 a) The <u>linear model</u>: Yield = 9.96 - 0.14 Month

$F = 219.24$ $p = .000$ $R^2 = 90.9$ $s_e = .3212$

The quadratic model: Yield = 10.4 - 0.252 Month + .00445 Month^2

F = 176.21 p = .000 R^2 = 94.4% s_e = .2582

In the quadratic model, both *t* ratios are significant,

for x: t = - 7.93, p = .000 and for x^2d: t = 3.61, p = .002

The linear model is a strong model. The quadratic term adds some predictability but has a smaller *t* ratio than does the linear term.

b)

x	*F*	*e*
10.08	-	-
10.05	-	-
9.24		-
-		
9.23	-	-
9.69	9.65	.04
9.55	9.55	.00
9.37	9.43	.06
8.55	9.46	.91
8.36	9.29	.93
8.59	8.96	.37
7.99	8.72	.73
8.12	8.37	.25
7.91	8.27	.36
7.73	8.15	.42
7.39	7.94	.55
7.48	7.79	.31
7.52	7.63	.11
7.48	7.53	.05
7.35	7.47	.12
7.04	7.46	.42
6.88	7.35	.47
6.88	7.19	.31
7.17	7.04	.13
7.22	6.99	.23

$$\sum|e| = 6.77$$

$$\text{MAD} = \frac{6.77}{20} = \mathbf{.3385}$$

c)

	$\alpha = .3$		$\alpha = .7$	
x	*F*	$\lvert e\rvert$	*F*	$\lvert e\rvert$
10.08	-	-	-	-
10.05	10.08	.03	10.08	.03
9.24	10.07	.83	10.06	.82
9.23	9.82	.59	9.49	.26
9.69	9.64	.05	9.31	.38
9.55	9.66	.11	9.58	.03
9.37	9.63	.26	9.56	.19
8.55	9.55	1.00	9.43	.88
8.36	9.25	.89	8.81	.45
8.59	8.98	.39	8.50	.09
7.99	8.86	.87	8.56	.57
8.12	8.60	.48	8.16	.04
7.91	8.46	.55	8.13	.22
7.73	8.30	.57	7.98	.25
7.39	8.13	.74	7.81	.42

7.48	7.91	.43	7.52	.04
7.52	7.78	.26	7.49	.03
7.48	7.70	.22	7.51	.03
7.35	7.63	.28	7.49	.14
7.04	7.55	.51	7.39	.35
6.88	7.40	.52	7.15	.27
6.88	7.24	.36	6.96	.08
7.17	7.13	.04	6.90	.27
7.22	7.14	.08	7.09	.13

$\sum|e| = 10.06$ $\quad\sum|e| = 5.97$

$\text{MAD}_{\alpha=.3} = \frac{10.06}{23} = \mathbf{.4374}$ $\qquad \text{MAD}_{\alpha=.7} = \frac{5.97}{23} = \mathbf{.2596}$

α = .7 produces better forecasts based on MAD.

d). MAD for b) .3385, c) .4374 and .2596. Exponential smoothing with α = .7 produces the lowest error (.2596 from part c).

e)

TCSI	4 period moving tots	8 period moving tots	TC	SI
10.08				
10.05				
	38.60			

9.24		76.81	9.60	96.25
	38.21			
9.23		75.92	9.49	97.26
	37.71			
9.69		75.55	9.44	102.65
	37.84			
9.55		75.00	9.38	101.81
	37.16			
9.37		72.99	9.12	102.74
	35.83			
8.55		70.70	8.84	96.72
	34.87			
8.36		68.36	8.55	97.78
	33.49			
8.59		66.55	8.32	103.25
	33.06			
7.99		65.67	8.21	97.32
	32.61			
8.12		64.36	8.05	100.87
	31.75			
7.91		62.90	7.86	100.64
	31.15			
7.73		61.66	7.71	100.26
	30.51			
7.39		60.63	7.58	97.49

	30.12			
7.48		59.99	7.50	99.73
	29.87			
7.52		59.70	7.46	100.80
	29.83			
7.48		59.22	7.40	101.08
	29.39			
7.35		58.14	7.27	101.10
	28.75			
7.04		56.90	7.11	99.02
	28.15			
6.88		56.12	7.02	98.01
	27.97			
6.88		56.12	7.02	98.01
	28.15			
7.17				
7.22				

1st Period	102.65	97.78	100.64	100.80	98.01
2nd Period	101.81	103.25	100.26	101.08	98.01
3rd Period	96.25	102.74	97.32	97.49	101.10
4th Period	97.26	96.72	100.87	99.73	99.02

The highs and lows of each period (underlined) are eliminated and the others are averaged resulting in:

Seasonal Indexes: 1st 99.82

2nd	101.05
3rd	98.64
4th	98.67
total	398.18

Since the total is not 400, adjust each seasonal index by multiplying by $\frac{400}{398.18}$ = 1.004571 resulting in the final seasonal indexes of:

1st	100.28
2nd	101.51
3rd	99.09
4th	99.12

17.28

Year	Quantity	Index Number
1992	2073	100.0
1993	2290	110.5
1994	2349	113.3
1995	2313	111.6
1996	2456	118.5
1997	2508	121.1
1998	2463	118.8
1999	2499	120.5
2000	2520	121.6
2001	2529	122.0

2002	2483	119.8
2003	2467	119.0
2004	2397	115.6
2005	2351	113.4
2006	2308	111.3

17.29

Item	2002	2003	2004	2005	2006
1	3.21	3.37	3.80	3.73	3.65
2	0.51	0.55	0.68	0.62	0.59
3	0.83	0.90	0.91	1.02	1.06
4	1.30	1.32	1.33	1.32	1.30
5	1.67	1.72	1.90	1.99	1.98
6	0.62	0.67	0.70	0.72	0.71
Totals	8.14	8.53	9.32	9.40	9.29

$$\text{Index}_{2002} = \frac{\sum P_{2002}}{\sum P_{2002}}(100) = \frac{8.14}{8.14}(100) = \mathbf{100.0}$$

$$\text{Index}_{2003} = \frac{\sum P_{2003}}{\sum P_{2002}}(100) = \frac{8.53}{8.14}(100) = \mathbf{104.8}$$

Index$_{2004}$ = $\frac{\sum P_{2004}}{\sum P_{2002}}(100) = \frac{9.32}{8.14}(100)$ = **114.5**

Index$_{2005}$ = $\frac{\sum P_{2005}}{\sum P_{2002}}(100) = \frac{9.40}{8.14}(100)$ = **115.5**

Index$_{2006}$ = $\frac{\sum P_{2006}}{\sum P_{2002}}(100) = \frac{9.29}{8.14}(100)$ = **114.1**

17.30

	2003		2004		2005		2006	
Item	P	Q	P	Q	P	Q	P	Q
1	2.75	12	2.98	9	3.10	9	3.21	11
2	0.85	47	0.89	52	0.95	61	0.98	66
3	1.33	20	1.32	28	1.36	25	1.40	32

Laspeyres:	$P_{2003}Q_{2003}$	$P_{2006}Q_{2003}$
	33.00	38.52
	39.95	46.06
	26.60	28.00
Totals	99.55	112.58

$$\text{Laspeyres Index}_{2006} = \frac{\sum P_{2006}Q_{2003}}{\sum P_{2003}Q_{2003}}(100) = \frac{112.58}{99.55}(100) = \mathbf{113.1}$$

Paasche$_{2005}$:	$P_{2003}Q_{2005}$	$P_{2005}Q_{2005}$
	24.75	27.90
	51.85	57.95
	33.25	34.00
Totals	109.85	119.85

$$\text{Paasche Index}_{2005} = \frac{\sum P_{2005}Q_{2005}}{\sum P_{2003}Q_{2005}}(100) = \frac{119.85}{109.85}(100) = \mathbf{109.1}$$

17.31 a) moving average b) α = .2

Year	Quantity	F	$\lvert e \rvert$	F	$\lvert e \rvert$
1980	6559				
1981	6022			6022.00	
1982	6439			6022.00	
1983	6396	6340.00	56.00	6105.40	290.60
1984	6405	6285.67	119.33	6163.52	241.48
1985	6391	6413.33	22.33	6211.82	179.18
1986	6152	6397.33	245.33	6247.65	95.65
1987	7034	6316.00	718.00	6228.52	805.48
1988	7400	6525.67	874.33	6389.62	1010.38
1989	8761	6862.00	1899.00	6591.69	2169.31
1990	9842	7731.67	2110.33	7025.56	2816.45
1991	10065	8667.67	1397.33	7588.84	2476.16
1992	10298	9556.00	742.00	8084.08	2213.93
1993	10209	10068.33	140.67	8526.86	1682.14
1994	10500	10190.67	309.33	8863.29	1636.71
1995	9913	10335.67	422.67	9190.63	722.37
1996	9644	10207.33	563.33	9335.10	308.90

1997	9952	10019.00	67.00	9396.88	555.12
1998	9333	9836.33	503.33	9507.91	174.91
1999	9409	9643.00	234.00	9472.93	63.93
2000	9143	9564.67	421.67	9460.14	317.14
2001	9512	9295.00	217.00	9396.71	115.29
2002	9430	9354.67	75.33	9419.77	10.23
2003	9513	9361.67	151.33	9421.82	91.18
2004	10085	9485.00	600.00	9440.05	644.95

$\sum|e| = 11{,}889.67$ $\qquad$ $\sum|e| = 18{,}621.46$

MAD$_{\text{moving average}}$ $= \dfrac{\sum|e|}{number forecasts} = \dfrac{11{,}889.67}{22} =$ **540.44**

MAD$_{\alpha=.2}$ $= \dfrac{\sum|e|}{number forecasts} = \dfrac{18{,}621.46}{22} =$ **846.43**

c) The three-year moving average produced a smaller MAD (540.44) than did exponential smoothing with $\alpha = .2$ (MAD = 846.43). Using MAD as the criterion, the three-year moving average was a better forecasting tool than the exponential smoothing with $\alpha = .2$.

17.32-17.34

Month Chem 12m tot 2yr tot TC SI TCI T

Jan(1) 23.701

Feb 24.189

Mar 24.200

Apr 24.971

May 24.560

June 24.992

288.00

July 22.566 575.65 23.985

94.08 23.872 23.917

287.65

Aug 24.037 575.23 23.968 100.29

24.134 23.919

287.58

Sept 25.047 576.24 24.010 104.32

24.047 23.921

288.66

Oct 24.115 577.78 24.074 100.17

24.851 23.924

289.12

Nov 23.034 578.86 24.119 95.50

24.056 23.926

		289.74					
Dec	22.590			580.98	24.208		93.32
	23.731	23.928					
		291.24					
Jan(2)	23.347		584.00		24.333		95.95
	24.486	23.931					
		292.76					
Feb	24.122		586.15		24.423		98.77
	24.197	23.933					
		293.39					
Mar	25.282		587.81		24.492	103.23	
	23.683	23.936					
		294.42					
Apr	25.426		589.05	24.544	103.59		24.450
	23.938						
		294.63					
May	25.185		590.05		24.585	102.44	
	24.938	23.940					
		295.42					
June	26.486		592.63		24.693	107.26	
	24.763	23.943					
		297.21					
July	24.088		595.28		24.803		97.12
	25.482	23.945					
		298.07					
Aug	24.672		597.79		24.908		99.05
	24.771	23.947					
		299.72					

Sept	26.072		601.75	25.073	103.98		25.031
	23.950						
		302.03					
Oct	24.328		605.59		25.233		96.41
	25.070	23.952					
		303.56					
Nov	23.826		607.85		25.327		94.07
	24.884	23.955					
		304.29					
Dec	24.373		610.56		25.440		95.81
	25.605	23.957					
		306.27					
Jan(3)	24.207		613.27		25.553		94.73
	25.388	23.959					
		307.00					
Feb	25.772		614.89		25.620	100.59	
	25.852	23.962					
		307.89					
Mar	27.591		616.92		25.705	107.34	
	25.846	23.964					
		309.03					
Apr	26.958		619.39		25.808	104.46	
	25.924	23.966					
		310.36					
May	25.920		622.48		25.937	99.93	
	25.666		23.969				312.12
June	28.460		625.24		26.052	109.24	
	26.608	23.971					

		313.12					
July	24.821			627.35	26.140		94.95
	26.257	23.974				314.23	
Aug	25.560		629.12		26.213		97.51
	25.663	23.976					
		314.89					
Sept	27.218		631.53		26.314	103.44	
	26.131	23.978					
		316.64					
Oct	25.650		635.31		26.471		96.90
	26.432	23.981					
		318.67					
Nov	25.589		639.84		26.660		95.98
	26.725	23.983					
		321.17					
Dec	25.370		644.03		26.835		94.54
	26.652	23.985					
		322.86					
Jan(4)	25.316		647.65		26.985		93.82
	26.551	23.988					
		324.79					
Feb	26.435		652.98		27.208		97.16
	26.517	23.990					
		328.19					
Mar	29.346		659.95		27.498	106.72	
	27.490	23.992					
		331.76					
Apr	28.983		666.46		27.769	104.37	
	27.871	23.995					

		334.70					
May	28.424		672.57		28.024	101.43	
	28.145	23.997					
		337.87					
June	30.149			679.39		28.308	106.50
	28.187	24.000					
		341.52					
July	26.746		686.66		28.611		93.48
	28.294	24.002					
		345.14					
Aug	28.966		694.30		28.929	100.13	
	29.082	24.004					
		349.16					
Sept	30.783		701.34		29.223	105.34	
	29.554	24.007					
		352.18					
Oct	28.594		706.29		29.429		97.16
	29.466	24.009					
		354.11					
Nov	28.762		710.54	29.606		97.14	
	30.039	24.011					
		356.43					
Dec	29.018		715.50		29.813		97.33
	30.484	24.014					
		359.07					
Jan(5)	28.931		720.74		30.031		96.34
	30.342	24.016					
		361.67					

Feb	30.456		725.14	30.214	100.80
	30.551	24.019			
		363.47			
Mar	32.372		727.79	30.325	106.75
	30.325	24.021			
		364.32			
Apr	30.905		730.25	30.427	101.57
	29.719	24.023			365.93
May	30.743		733.94	30.581	100.53
	30.442	24.026			
		368.01			
June	32.794		738.09	30.754	106.63
	30.660	24.028			
		370.08			
July	29.342				
Aug	30.765				
Sept	31.637				
Oct	30.206				
Nov	30.842				
Dec	31.090				

Seasonal Indexing:

Month	Year1	Year2	Year3	Year4	Year5	Index
Jan		95.95	94.73	93.82	96.34	95.34
Feb		98.77	100.59	97.16	100.80	99.68
Mar		103.23	107.34	106.72	106.75	106.74
Apr		103.59	104.46	104.37	101.57	103.98

May		102.44	99.93	101.43	100.53	100.98
June		107.26	109.24	106.50	106.63	106.96
July	94.08	97.12	94.95	93.48		94.52
Aug	100.29	99.05	97.51	100.13		99.59
Sept	104.32	103.98	103.44	105.34		104.15
Oct	100.17	96.41	96.90	97.16		97.03
Nov	95.50	94.07	95.98	97.14		95.74
Dec	93.32	95.81	94.54	97.33		95.18
Total						1199.88

Adjust each seasonal index by 1200/1199.88 = 1.0001

Final Seasonal Indexes:

Month	Index
Jan	95.35
Feb	99.69
Mar	106.75
Apr	103.99
May	100.99
June	106.96
July	94.53
Aug	99.60
Sept	104.16
Oct	97.04
Nov	95.75
Dec	95.19

Regression Output for Trend Line:

$$\hat{y} = 22.4233 + 0.144974\, x$$

$R^2 = .913$

Regression Output for Quadratic Trend:

$$\hat{y} = 23.8158 + 0.01554\,x + .000247\,x^2$$

$R^2 = .964$

In this model, the linear term yields a $t = 0.66$ with $p = .513$ but the squared term predictor yields a $t = 8.94$ with $p = .000$.

Regression Output when using only the squared predictor term:

$$\hat{y} = 23.9339 + 0.00236647\,x^2$$

$R^2 = .964$

Note: The trend model derived using only the squared predictor was used in computing T (trend) in the decomposition process.

17.35

Item	2004 Price	2004 Quantity	2005 Price	2005 Quantity	2006 Price	2006 Quantity
Margarine (lb.)	1.26	21	1.32	23	1.39	22
Shortening (lb.)	0.94	5	0.97	3	1.12	4
Milk (1/2 gal.)	1.43	70	1.56	68	1.62	65
Cola (2 liters)	1.05	12	1.02	13	1.25	11
Potato Chips (12 oz.)	2.81	27	2.86	29	2.99	28
Total	7.49		7.73		8.37	

$$\text{Index}_{2004} = \frac{\sum P_{2004}}{\sum P_{2004}}(100) = \frac{7.49}{7.49}(100) = \mathbf{100.0}$$

$$\text{Index}_{2005} = \frac{\sum P_{2005}}{\sum P_{2004}}(100) = \frac{7.73}{7.49}(100) = \mathbf{103.2}$$

$$\text{Index}_{2006} = \frac{\sum P_{2006}}{\sum P_{2004}}(100) = \frac{8.37}{7.49}(100) = \mathbf{111.8}$$

$P_{2004}Q_{2004}$ $P_{2005}Q_{2004}$ $P_{2006}Q_{2004}$

	26.46	27.72	29.19
	4.70	4.85	5.60
	100.10	109.20	113.40
	12.60	12.24	15.00
	75.87	77.22	80.73
Totals	219.73	231.23	243.92

$$\text{Index}_{\text{Laspeyres2005}} = \frac{\sum P_{2005}Q_{2004}}{\sum P_{2004}Q_{2004}}(100) = \frac{231.23}{219.73}(100) = \mathbf{105.2}$$

$$\text{Index}_{\text{Laspeyres2006}} = \frac{\sum P_{2006}Q_{2004}}{\sum P_{2004}Q_{2004}}(100) = \frac{243.92}{219.73}(100) = \mathbf{111.0}$$

	$P_{2004}Q_{2005}$	$P_{2004}Q_{2006}$	$P_{2005}Q_{2005}$	$P_{2006}Q_{2006}$
	28.98	27.726	30.36	30.58
	2.82	3.76	2.91	4.48
	97.24	92.95	106.08	105.30
	13.65	11.55	13.26	13.75
	81.49	78.68	82.94	83.72
Total	224.18	214.66	235.55	237.83

$$\text{Index}_{\text{Paasche2005}} = \frac{\sum P_{2005}Q_{2005}}{\sum P_{2004}Q_{2005}}(100) = \frac{235.55}{224.18}(100) = \mathbf{105.1}$$

$$\text{Index}_{\text{Paasche2006}} = \frac{\sum P_{2006}Q_{2006}}{\sum P_{2004}Q_{2006}}(100) = \frac{237.83}{214.66}(100) = \mathbf{110.8}$$

17.36 $\hat{y} = 9.5382 - 0.2716\,x$

$\hat{y}(7) = 7.637$

$R^2 = 40.2\%$ $F = 12.78$, $p = .002$

$s_e = 0.264862$

Durbin-Watson:

$n = 21$ $k = 1$ $\alpha = .05$

$\boldsymbol{D = 0.44}$

$d_L = 1.22$ and $d_U = 1.42$

Since $D = 0.44 < d_L = 1.22$, the decision is to **reject the null hypothesis**.

There is significant autocorrelation.

17.37

Year	x	F_{ma}	F_{wma}	SE_{MA}	SE_{WMA}
1988	118.5				
1989	123.0				
1990	128.5				
1991	133.6				
1992	137.5	125.9	128.4	134.56	82.08
1993	141.2	130.7	133.1	111.30	65.93
1994	144.8	135.2	137.3	92.16	56.25
1995	148.5	139.3	141.1	85.10	54.17
1996	152.8	143.0	144.8	96.04	63.52
1997	156.8	146.8	148.8	99.50	64.80
1998	160.4	150.7	152.7	93.61	58.68
1999	163.9	154.6	156.6	86.03	53.14
2000	169.6	158.5	160.3	123.77	86.12
2001	176.4	162.7	164.8	188.38	135.26
2002	180.3	167.6	170.3	161.93	100.80
2003	184.8	172.6	175.4	150.06	89.30
2004	189.5	177.8	180.3	137.48	85.56
2005	195.7	182.8	184.9	167.70	115.78
			SE =	1,727.60	1,111.40

$$MSE_{ma} = \frac{SE}{No.Forecasts} = \frac{1727.60}{14} = \mathbf{123.4}$$

$$MSE_{wma} = \frac{SE}{No.Forecasts} = \frac{1111.4}{14} = \mathbf{79.39}$$

The weighted moving average does a better job of forecasting the data using MSE as the criterion.

17.38 The regression model with one-month lag is:

Cotton Prices = - 61.24 + 1.1035 LAG1

F = 130.46 (p = .000), R^2 = .839, adjusted R^2 = .833,

s_e = 17.57, t = 11.42 (p = .000).

The regression model with four-month lag is:

Cotton Prices = 303.9 + 0.4316 LAG4

F = 1.24 (p = .278), R^2 .053, adjusted R^2 = .010,

s_e = 44.22, t = 1.11 (p = .278).

The model with the four-month lag does not have overall significance and has an adjusted R^2 of 1%. This model has virtually no predictability. The model with the one-month lag has relatively strong predictability with adjusted R^2 of 83.3%. In addition, the F value is significant at α = .001 and the standard error of the estimate is less than 40% as large as the standard error for the four-month lag model.

17.39

	Qtr	TSCI	4qrtot	8qrtot	TC	SI	TCI	T
Year1	1	54.019						
	2	56.495						
			213.574					
	3	50.169		425.044	53.131	94.43	51.699	53.722
			211.470					
	4	52.891		421.546	52.693	100.38	52.341	55.945
			210.076					
Year2	1	51.915		423.402	52.925	98.09	52.937	58.274
			213.326					
	2	55.101		430.997	53.875	102.28	53.063	60.709
			217.671					
	3	53.419		440.490	55.061	97.02	55.048	63.249
			222.819					
	4	57.236		453.025	56.628	101.07	56.641	65.895
			230.206					
Year3	1	57.063		467.366	58.421	97.68	58.186	68.646
			237.160					
	2	62.488		480.418	60.052	104.06	60.177	71.503
			243.258					
	3	60.373		492.176	61.522	98.13	62.215	74.466
			248.918					
	4	63.334		503.728	62.966	100.58	62.676	77.534
			254.810					

Year4	1	62.723	257.693	512.503	64.063	97.91	63.957	80.708
	2	68.380	260.805	518.498	64.812	105.51	65.851	83.988
	3	63.256	263.527	524.332	65.542	96.51	65.185	87.373
	4	66.446	263.158	526.685	65.836	100.93	65.756	90.864
Year5	1	65.445	263.147	526.305	65.788	99.48	66.733	94.461
	2	68.011	263.573	526.720	65.840	103.30	65.496	98.163
	3	63.245	257.842	521.415	65.177	97.04	65.174	101.971
	4	66.872	253.421	511.263	63.908	104.64	66.177	105.885
Year6	1	59.714	248.264	501.685	62.711	95.22	60.889	109.904
	2	63.590		491.099	61.387	103.59	61.238	114.029
	3	58.088						
	4	61.443						

Quarter	Year1	Year2	Year3	Year4	Year5	Year6	Index

1		98.09	97.68	97.91	99.48	95.22	97.89
2		102.28	104.06	105.51	103.30	103.59	103.65
3	94.43	97.02	98.13	96.51	97.04		96.86
4	100.38	101.07	100.58	100.93	104.64		100.86
Total							399.26

Adjust the seasonal indexes by: $\frac{400}{399.26} = 1.00185343$

Adjusted Seasonal Indexes:

Quarter	Index
1	98.07
2	103.84
3	97.04
4	101.05
Total	400.00

17.40

Time Period	Deseasonalized Data
Q1(yr1)	55.082
Q2	54.406
Q3	51.699
Q4	52.341
Q1(yr2)	52.937
Q2	53.063
Q3	55.048
Q4	56.641
Q1(yr3)	58.186
Q2	60.177
Q3	62.215
Q4	62.676
Q1(yr4)	63.957
Q2	65.851
Q3	65.185
Q4	65.756
Q1(yr5)	66.733
Q2	65.496
Q3	65.174
Q4	66.177
Q1(yr6)	60.889
Q2	61.238
Q3	59.860
Q4	60.805

17.41 Linear Model: $\hat{y}$ = 53.41032 + 0.532488 x

R^2 = 55.7% F = 27.65 with p = .000

s_e = 3.43

Quadratic Model: $\hat{y}$ = 47.68663 + 1.853339 x –0.052834 x^2

R^2 = 76.6% F = 34.37 with p = .000

s_e = 2.55

In the quadratic regression model, both the linear and squared terms have significant t statistics at alpha .001 indicating that both are contributing. In addition, the R^2 for the quadratic model is considerably higher than the R^2 for the linear model. Also, s_e is smaller for the quadratic model. All of these indicate that the quadratic model is a stronger model.

17.42 R^2 = 55.8% F = 8.83 with p = .021

s_e = 50.18

This model with a lag of one year has modest predictability. The overall F is

significant at α = .05 but not at α = .01.

17.43 The regression equation is:

Equity Funds = -359.1 + 2.0898 Money Market Funds

R^2 = 88.2% s_e = 582.685

D = **0.84**

For n = 26 and α = .01, d_L = 1.07 and d_U = 1.22.

Since $D = 0.84 < d_L = 1.07$, the null hypothesis is rejected. **There is significant autocorrelation in this model.**

17.44

Year	PurPwr	$\alpha = .1$ F	$\lvert e \rvert$	$\alpha = .5$ F	$\lvert e \rvert$	$\alpha = .8$ F	$\lvert e \rvert$
1	6.04						
2	5.92	6.04	.12	6.04	.12	6.04	.12
3	5.57	6.03	.46	5.98	.41	5.94	.37
4	5.40	5.98	.58	5.78	.38	5.64	.24
5	5.17	5.92	.75	5.59	.42	5.45	.28
6	5.00	5.85	.85	5.38	.38	5.23	.23
7	4.91	5.77	.86	5.19	.28	5.05	.14
8	4.73	5.68	.95	5.05	.32	4.94	.21
9	4.55	5.59	1.04	4.89	.34	4.77	.22
10	4.34	5.49	1.15	4.72	.38	4.59	.25
11	4.67	5.38	.71	4.53	.14	4.39	.28
12	5.01	5.31	.30	4.60	.41	4.61	.40
13	4.86	5.28	.42	4.81	.05	4.93	.07
14	4.72	5.24	.52	4.84	.12	4.87	.15
15	4.60	5.19	.59	4.78	.18	4.75	.15
16	4.48	5.13	.65	4.69	.21	4.63	.15
17	4.86	5.07	.21	4.59	.27	4.51	.35

18	5.15	5.05	.10	4.73	.42	4.79	.36

$\sum|e| = 10.26$. $\sum|e| = 4.83$ $\sum|e| = 3.97$

$$\text{MAD}_1 = \frac{\sum|e|}{N} = \frac{10.26}{17} = \mathbf{.60}$$

$$\text{MAD}_2 = \frac{\sum|e|}{N} = \frac{4.83}{17} = \mathbf{.28}$$

$$\text{MAD}_3 = \frac{\sum|e|}{N} = \frac{3.97}{17} = \mathbf{.23}$$

The smallest mean absolute deviation error is produced using $\alpha = .8$.

The forecast for year 19 is: $F(19) = (.8)(5.15) + (.2)(4.79) =$ **5.08**

17.45 The model is: Bankruptcies = 75,532.436 – 0.016 Year

Since R^2 = .28 and the adjusted R^2 = .23, this is a weak model.

e_t	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
- 1,338.58			1,791,796
- 8,588.28	- 7,249.7	52,558,150	73,758,553
- 7,050.61	1,537.7	2,364,521	49,711,101
1,115.01	8,165.6	66,677,023	1,243,247
12,772.28	11,657.3	135,892,643	163,131,136
14,712.75	1,940.5	3,765,540	216,465,013
- 3,029.45	-17,742.2	314,785,661	9,177,567
- 2,599.05	430.4	185,244	6,755,061
622.39	3,221.4	10,377,418	387,369
9,747.30	9,124.9	83,263,800	95,009,857
9,288.84	- 458.5	210,222	86,282,549
- 434.76	- 9,723.6	94,548,397	189,016
-10,875.36	-10,440.6	109,006,128	118,273,455
- 9,808.01	1,067.4	1,139,343	96,197.060
- 4,277.69	5,530.3	30,584,218	18,298,632
- 256.80	4,020.9	16,167,637	65,946

$$\sum (e_t - e_{t-1})^2 = 921,525,945 \qquad \sum e_t^2 = 936,737,358$$

$$D = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^{\ 2}} = \frac{921{,}525{,}945}{936{,}737{,}358} = \mathbf{0.98}$$

For $n = 16$, $\alpha = .05$, $d_L = 1.10$ and $d_U = 1.37$

Since $D = 0.98 < d_L = 1.10$, the decision is to **reject the null hypothesis and conclude that there is significant autocorrelation**.