INSTRUCTOR'S SOLUTIONS MANUAL

NANCY S. BOUDREAU
Bowling Green State University

STATISTICS FOR BUSINESS AND ECONOMICS

TWELFTH EDITION

James T. McClave
Info Tech, Inc.
University of Florida

P. George Benson
College of Charleston

Terry Sincich
University of South Florida

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 75 Arlington Street, Boston, MA 02116.

ISBN-13: 978-0-321-83681-6
ISBN-10: 0-321-83681-2

www.pearsonhighered.com

PEARSON

Chapter 1
Statistics, Data, and Statistical Thinking

1.1 Statistics is a science that deals with the collection, classification, analysis, and interpretation of information or data. It is a meaningful, useful science with a broad, almost limitless scope of applications to business, government, and the physical and social sciences.

1.2 Descriptive statistics utilizes numerical and graphical methods to look for patterns, to summarize, and to present the information in a set of data. Inferential statistics utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.

1.3 The four elements of a descriptive statistics problem are:

1. The population or sample of interest. This is the collection of all the units upon which the variable is measured.
2. One or more variables that are to be investigated. These are the types of data that are to be collected.
3. Tables, graphs, or numerical summary tools. These are tools used to display the characteristic of the sample or population.
4. Identification of patterns in the data. These are conclusions drawn from what the summary tools revealed about the population or sample.

1.4 The five elements of an inferential statistical analysis are:

1. The population of interest. The population is a set of existing units.
2. One or more variables that are to be investigated. A variable is a characteristic or property of an individual population unit.
3. The sample of population units. A sample is a subset of the units of a population.
4. The inference about the population based on information contained in the sample. A statistical inference is an estimate, prediction, or generalization about a population based on information contained in a sample.
5. A measure of reliability for the inference. The reliability of an inference is how confident one is that the inference is correct.

1.5 The first major method of collecting data is from a published source. These data have already been collected by someone else and are available in a published source. The second method of collecting data is from a designed experiment. These data are collected by a researcher who exerts strict control over the experimental units in a study. These data are measured directly from the experimental units. The final method of collecting data is observational. These data are collected directly from experimental units by simply observing the experimental units in their natural environment and recording the values of the desired characteristics. The most common type of observational study is a survey.

1.6 Quantitative data are measurements that are recorded on a meaningful numerical scale. Qualitative data are measurements that are not numerical in nature; they can only be classified into one of a group of categories.

1.7 A population is a set of existing units such as people, objects, transactions, or events. A variable is a characteristic or property of an individual population unit such as height of a person, time of a reflex, amount of a transaction, etc.

1.8 A population is a set of existing units such as people, objects, transactions, or events. A sample is a subset of the units of a population.

1.9 A representative sample is a sample that exhibits characteristics similar to those possessed by the target population. A representative sample is essential if inferential statistics is to be applied. If a sample does not possess the same characteristics as the target population, then any inferences made using the sample will be unreliable.

1.10 An inference without a measure of reliability is nothing more than a guess. A measure of reliability separates statistical inference from fortune telling or guessing. Reliability gives a measure of how confident one is that the inference is correct.

1.11 A population is a set of existing units such as people, objects, transactions, or events. A process is a series of actions or operations that transform inputs to outputs. A process produces or generates output over time. Examples of processes are assembly lines, oil refineries, and stock prices.

1.12 Statistical thinking involves applying rational thought processes to critically assess data and inferences made from the data. It involves not taking all data and inferences presented at face value, but rather making sure the inferences and data are valid.

1.13 The data consisting of the classifications A, B, C, and D are qualitative. These data are nominal and thus are qualitative. After the data are input as 1, 2, 3, and 4, they are still nominal and thus qualitative. The only differences between the two data sets are the names of the categories. The numbers associated with the four groups are meaningless.

1.14 Answers will vary. First, number the elements of the population from 1 to 200,000. Using MINITAB, generate 10 numbers on the interval from 1 to 200,000, eliminating any duplicates.

The 10 numbers selected for the random sample are:

135075
89127
189226
83899
112367
191496
110021
44853
42091
198461

Elements with the above numbers are selected for the sample.

1.15 a. The experimental unit for this study is a single-family residential property in Arlington, Texas.

b. The variables measured are the sale price and the Zillow estimated value. Both of these variables are quantitative.

c. If these 2,045 properties were all the properties sold in Arlington, Texas in the past 6 months, then this would be considered the population.

d. If these 2,045 properties represent a sample, then the population would be all the single-family residential properties sold in the last 6 months in Arlington, Texas.

e. No. The real estate market across the United States varies greatly. The prices of single-family residential properties in this small area are probably not representative of all properties across the United States.

1.16 a. The experimental unit for this study is an NFL quarterback.

b. The variables measured in this study include draft position, NFL winning ratio, and QB production score. Since the draft position was put into 3 categories, it is a qualitative variable. The NFL winning ratio and the QB production score are both quantitative.

c. Since we want to project the performance of future NFL QBs, this would be an application of inferential statistics.

1.17 a. The population of interest is all citizens of the United States.

b. The variable of interest is the view of each citizen as to whether the president is doing a good or bad job. It is qualitative.

c. The sample is the 2000 individuals selected for the poll.

d. The inference of interest is to estimate the proportion of all U.S. citizens who believe the president is doing a good job.

e. The method of data collection is a survey.

f. It is not very likely that the sample will be representative of the population of all citizens of the United States. By selecting phone numbers at random, the sample will be limited to only those people who have telephones. Also, many people share the same phone number, so each person would not have an equal chance of being contacted. Another possible problem is the time of day the calls are made. If the calls are made in the evening, those people who work in the evening would not be represented.

1.18 a. High school GPA is a number usually between 0.0 and 4.0. Therefore, it is quantitative.

b. Honors/awards would have responses that name things. Therefore, it would be qualitative.

c. The scores on the SAT's are numbers between 200 and 800. Therefore, it is quantitative.

d. Gender is either male or female. Therefore, it is qualitative.

e. Parent's income is a number: $25,000, $45,000, etc. Therefore, it is quantitative.

f. Age is a number: 17, 18, etc. Therefore, it is quantitative.

1.19 I. Qualitative; the possible responses are "yes" or "no," which are non-numerical.

II. Quantitative; age is measured on a numerical scale, such as 15, 32, etc.

III. Qualitative; the possible responses are “yes” or “no,” which are non-numerical.

IV. Qualitative; the possible responses are "laser printer" or "another type of printer," which are non-numerical.

c. For the physicians who would use ethics consultation in the future, the standard deviation is 8.95 years.

d. The variation in the length of time in practice for the physicians who would refuse to use ethics consultation in the future is greater than that for the physicians who would use ethics consultation in the future.

2.69 a. The range is the largest observation minus the smallest observation or 11 – 1 = 10.

The variance is: $s^2 = \dfrac{\sum_i x_i^2 - \dfrac{\left(\sum_i x_i\right)^2}{n}}{n-1} = \dfrac{450 - \dfrac{78^2}{20}}{20-1} = 7.6737$

The standard deviation is: $s = \sqrt{s^2} = \sqrt{7.6737} = 2.77$

b. The largest observation is 11. It is deleted from the data set. The new range is: 9 – 1 = 8.

The variance is: $s^2 = \dfrac{\sum_i x_i^2 - \dfrac{\left(\sum_i x_i\right)^2}{n}}{n-1} = \dfrac{329 - \dfrac{67^2}{19}}{19-1} = 5.1520$

The standard deviation is: $s = \sqrt{s^2} = \sqrt{5.1520} = 2.27$

When the largest observation is deleted, the range, variance and standard deviation decrease.

c. The largest observation is 11 and the smallest is 1. When these two observations are deleted from the data set, the new range is: 9 – 1 = 8.

The variance is: $s^2 = \dfrac{\sum_i x_i^2 - \dfrac{\left(\sum_i x_i\right)^2}{n}}{n-1} = \dfrac{328 - \dfrac{66^2}{18}}{18-1} = 5.0588$

The standard deviation is: $s = \sqrt{s^2} = \sqrt{5.0588} = 2.25$

When the largest and smallest observations are deleted, the range, variance and standard deviation decrease.

2.70 a. A worker's overall time to complete the operation under study is determined by adding the subtask-time averages.

Worker A

The average for subtask 1 is: $\bar{x} = \dfrac{\sum x}{n} = \dfrac{211}{7} = 30.14$

The average for subtask 2 is: $\bar{x} = \dfrac{\sum x}{n} = \dfrac{21}{7} = 3$

Worker A's overall time is 30.14 + 3 = 33.14.

Chapter 3
Probability

3.1 a. Since the probabilities must sum to 1,

$$P(E_3) = 1 - P(E_1) - P(E_2) - P(E_4) - P(E_5) = 1 - .1 - .2 - .1 - .1 = .5$$

b. $$P(E_3) = 1 - P(E_1) - P(E_2) - P(E_4) - P(E_5) = 1 - P(E_3) - P(E_2) - P(E_4) - P(E_5)$$
$$\Rightarrow 2P(E_3) = 1 - .1 - .2 - .1 \Rightarrow 2P(E_3) = .6 \Rightarrow P(E_3) = .3$$

c. $$P(E_3) = 1 - P(E_1) - P(E_2) - P(E_4) - P(E_5) = 1 - .1 - .1 - .1 - .1 = .6$$

3.2 a. This is a Venn Diagram.

b. If the sample points are equally likely, then

$$P(1) = P(2) = P(3) = \cdots = P(10) = \frac{1}{10}$$

Therefore,

$$P(A) = P(4) + P(5) + P(6) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10} = .3$$

$$P(B) = P(6) + P(7) = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = .2$$

c. $$P(A) = P(4) + P(5) + P(6) = \frac{1}{20} + \frac{1}{20} + \frac{3}{20} = \frac{5}{20} = .25$$

$$P(B) = P(6) + P(7) = \frac{3}{20} + \frac{3}{20} = \frac{6}{20} = .3$$

3.3 $$P(A) = P(1) + P(2) + P(3) = .05 + .20 + .30 = .55$$

$$P(B) = P(1) + P(3) + P(5) = .05 + .30 + .15 = .50$$

$$P(C) = P(1) + P(2) + P(3) + P(5) = .05 + .20 + .30 + .15 = .70$$

3.4 a. $$\binom{9}{4} = \frac{9!}{4!(9-4)!} = \frac{9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1\cdot5\cdot4\cdot3\cdot2\cdot1} = 126$$

b. $$\binom{7}{2} = \frac{7!}{2!(7-2)!} = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot5\cdot4\cdot3\cdot2\cdot1} = 21$$

c. $$\binom{4}{4} = \frac{4!}{4!(4-4)!} = \frac{4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1\cdot1} = 1$$

d. $$\binom{5}{0} = \frac{5!}{0!(5-0)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{1\cdot5\cdot4\cdot3\cdot2\cdot1} = 1$$

4.99 a. Using MINITAB with $\mu = 59$ and $\sigma = 5$, the probability is:

Cumulative Distribution Function

```
Normal with mean = 59 and standard deviation = 5

 x  P( X <= x )
60     0.579260
```

$P(x > 60) = 1 - P(x \le 60) = 1 - .57926 = .42074$

b. Using MINITAB with $\mu = 43$ and $\sigma = 5$, the probability is:

Cumulative Distribution Function

```
Normal with mean = 43 and standard deviation = 5

 x  P( X <= x )
60     0.999663
```

$P(x > 60) = 1 - P(x \le 60) = 1 - .999663 = .000337$

4.100 a. Using Table II, Appendix D,

$$P(x > 0) = P\left(z > \frac{0 - 5.26}{10}\right) = P(z > -0.526)$$
$$= .5 + P(-0.53 < z < 0) = .5 + .2019 = .7019$$

b. $$P(5 < x < 15) = P\left(\frac{5 - 5.26}{10} < z < \frac{15 - 5.26}{10}\right) = P(-0.026 < z < 0.974)$$
$$= P(-.03 < z < 0) + P(0 < z < .97) = .0120 + .3340 = .3460$$

c. $$P(x < 1) = P\left(z < \frac{1 - 5.26}{10}\right) = P(z < -0.426)$$
$$= .5 - P(-0.43 < z < 0) = .5 - .1664 = .3336$$

d. $$P(x \le -25) = P\left(z \le \frac{-25 - 5.26}{10}\right) = P(z \le -3.026)$$
$$= .5 - P(-3.03 \le z < 0) = .5 - .4988 = .0012$$

Since the probability of seeing a win percentage of -25% or anything more unusual is so small (p = .0012), we would conclude that the average casino win percentage is not 5.26%.

4.101 a. Let x = buy-side analyst's forecast error. Then x has an approximate normal distribution with $\mu = .85$ and $\sigma = 1.93$. Using Table II, Appendix D,

$$P(x > 2.00) = P\left(z > \frac{2.00 - .85}{1.93}\right) = P(z > .60) = .5 - .2257 = .2743$$

d. The observed significance level is p-value $= p = P(t \le -1.89) + P(t \ge 1.89)$. Since we did not reject H_0 in part c, we know that the p-value must be greater than .05. Using MINITAB,

Cumulative Distribution Function

```
Student's t distribution with 6 DF
    x  P( X <= x )
-1.89    0.0538261
```

Thus, the p-value is $p = 2(.0538261) = .1076522$.

7.126 a. Let p = proportion of shoppers using cents-off coupons. To determine if the proportion of shoppers using cents-off coupons exceeds .65, we test:

$H_0: p = .65$
$H_a: p > .65$

The test statistic is $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \dfrac{.72 - .65}{\sqrt{\dfrac{.65(.35)}{1{,}000}}} = 4.64$

The rejection region requires $\alpha = .05$ in the upper tail of the z-distribution. From Table II, Appendix D, $z_{.05} = 1.645$. The rejection region is $z > 1.645$.

Since the observed value of the test statistic falls in the rejection region $(z = 4.64 > 1.645)$, H_0 is rejected. There is sufficient evidence to indicate the proportion of shoppers using cents-off coupons exceeds .65 at $\alpha = .05$.

b. The sample size is large enough if the $np_0 \ge 15$ and $nq_0 \ge 15$.

$np_0 = 1000(.65) = 650$ $nq_0 = 1000(.35) = 350$

Since both $np_0 \ge 15$ and $nq_0 \ge 15$, the normal distribution will be adequate.

c. The p-value is $p = P(z \ge 4.64) = (.5 - .5) \approx .0$. (Using Table II, Appendix D.) Since the p-value is smaller than $\alpha = .05$, H_0 is rejected. There is sufficient evidence to indicate the proportion of shoppers using cents-off coupons exceeds .65 at $\alpha = .05$.

7.127 a. The hypotheses would be:

H_0: Individual does not have the disease
H_a: Individual does have the disease

b. A Type I error would be: Conclude the individual has the disease when in fact he/she does not. This would be a false positive test.

A Type II error would be: Conclude the individual does not have the disease when in fact he/she does. This would be a false negative test.

The test statistic is $F = 9.97$.

The rejection region requires $\alpha = .01$ in the upper tail of the F-distribution with $\nu_1 = k - 1 = 3 - 1 = 2$ and $\nu_2 = n - k = 76 - 3 = 73$. From Table VIII, Appendix D, $F_{.01} = 4.92$. The rejection region is $F > 4.92$.

Since the observed value of the test statistic falls in the rejection region $(F = 9.97 > 4.92)$, H_0 is rejected. There is sufficient evidence to indicate a difference in the mean drops in anxiety levels among the three groups at $\alpha = .01$.

e. The assumption of constant variance is satisfied since the three sample variances are all very similar $(7.6^2 = 57.76,\ 7.5^2 = 56.25,\ \text{and}\ 7.5^2 = 56.25)$.

We are unable to check the normality assumption since we need the individual drops in anxiety levels to create a histogram or stem-and-leaf plot.

9.35 The number of pairwise comparisons is equal to $k(k-1)/2$.

a. For $k = 3$, the number of comparisons is $3(3-1)/2 = 3$.

b. For $k = 5$, the number of comparisons is $5(5-1)/2 = 10$.

c. For $k = 4$, the number of comparisons is $4(4-1)/2 = 6$.

d. For $k = 10$, the number of comparisons is $10(10-1)/2 = 45$.

9.36 The experimentwise error rate is the probability of making a Type I error for at least one of all of the comparisons made. If the experimentwise error rate is $\alpha = .05$, then each individual comparison is made at a value of α which is less than .05.

9.37 A comparisonwise error rate is the error rate (or the probability of declaring the means different when, in fact, they are not different, which is also the probability of a Type I error) for each individual comparison. That is, if each comparison is run using $\alpha = .05$, then the comparisonwise error rate is .05.

9.38 a. From the diagram, the following pairs of treatments are significantly different because they are not connected by a line: A and E, A and B, A and D, C and E, C and B, C and D, and E and D. All other pairs of means are not significantly different because they are connected by lines.

b. From the diagram, the following pairs of treatments are significantly different because they are not connected by a line: A and B, A and D, C and B, C and D, E and B, E and D, and B and D. All other pairs of means are not significantly different because they are connected by lines.

c. From the diagram, the following pairs of treatments are significantly different because they are not connected by a line: A and E, A and B, and A and D. All other pairs of means are not significantly different because they are connected by lines.

d. From the diagram, the following pairs of treatments are significantly different because they are not connected by a line: A and E, A and B, A and D, C and E, C and B, C and D, E and D, and B and D. All other pairs of means are not significantly different because they are connected by lines.