WILEY

Solutions Manual to Accompany

MODERN ENGINEERING STATISTICS

THOMAS P. RYAN

Solutions Manual to Accompany

Modern Engineering Statistics

Thomas P. Ryan
Acworth, GA

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Wiley Bicentennical Logo: Richard J. Pacifico

Library of Congress Cataloging in Publication Data:

Ryan, Thomas P.
Solutions Manual to Accompany Modern Engineering Statistics

ISBN 978-0-470-09607-9 (paper)

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Solutions Manual to Accompany

Modern Engineering Statistics

1807 WILEY 2007
BICENTENNIAL

THE WILEY BICENTENNIAL–KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

Contents

1	**Methods of Collecting and Presenting Data**	1
2	**Measures of Location and Dispersion**	27
3	**Probability and Common Probability Distributions**	37
4	**Point Estimation**	51
5	**Confidence Intervals and Hypothesis Tests—One Sample**	59
6	**Confidence Intervals and Hypothesis Tests—Two Samples**	73
7	**Tolerance Intervals and Prediction Intervals**	83
8	**Simple Linear Regression, Correlation and Calibration**	87
9	**Multiple Regression**	101
10	**Mechanistic Models**	117
11	**Control Charts and Quality Improvement**	121
12	**Design and Analysis of Experiments**	139
13	**Measurement System Appraisal**	155
14	**Reliability Analysis and Life Testing**	157
15	**Analysis of Categorical Data**	163
16	**Distribution-Free Procedures**	171
17	**Tying It All Together**	175

Methods of Collecting and Presenting Data

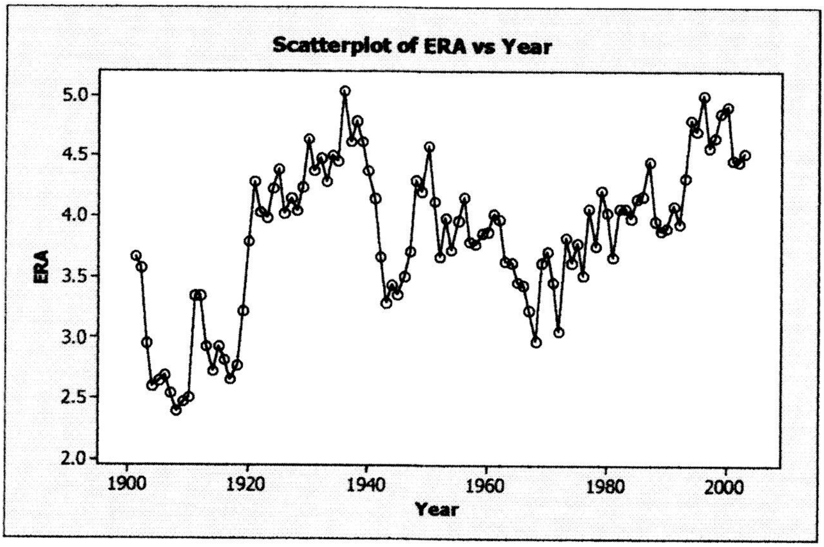
Note: The data in the following exercises, including data in MINITAB files (i.e., the files with the .MTW extension), can be found at the website for the text: ftp://ftp.wiley.com/public/sci_tech_med/engineering_statistics. This also applies to the other chapters in the text.

1.1. Given below are the earned run averages (ERAs) for the American League for 1901-2003 (in ERAMC.MTW), with the years 1916 and 1994 corrected from the source, *Total Baseball*, 8th edition, by John Thorn, Phil Birnbaum, and Bill Deane, since those two years were obviously in error. (The league started in 1901.)

Year	1901	1902	1903	1904	1905	1906	1907	1908
	1909	1910	1911	1912	1913	1914		
ERA	3.66	3.57	2.96	2.60	2.65	2.69	2.54	2.39
	2.47	2.51	3.34	3.34	2.93	2.73		
Year	1915	1916	1917	1918	1919	1920	1921	1922
	1923	1924	1925	1926	1927	1928		
ERA	2.93	2.82	2.66	2.77	3.22	3.79	4.28	4.03
	3.98	4.23	4.39	4.02	4.14	4.04		
Year	1929	1930	1931	1932	1933	1934	1935	1936
	1937	1938	1939	1940	1941	1942		
ERA	4.24	4.64	4.38	4.48	4.28	4.50	4.45	5.04
	4.62	4.79	4.62	4.38	4.15	3.66		
Year	1943	1944	1945	1946	1947	1948	1949	1950
	1951	1952	1953	1954	1955	1956		
ERA	3.29	3.43	3.36	3.50	3.71	4.29	4.20	4.58
	4.12	3.67	3.99	3.72	3.96	4.16		
Year	1957	1958	1959	1960	1961	1962	1963	1964
	1965	1966	1967	1968	1969	1970		
ERA	3.79	3.77	3.86	3.87	4.02	3.97	3.63	3.62
	3.46	3.43	3.23	2.98	3.62	3.71		
Year	1971	1972	1973	1974	1975	1976	1977	1978
	1979	1980	1981	1982	1983	1984		
ERA	3.46	3.06	3.82	3.62	3.78	3.52	4.06	3.76
	4.21	4.03	3.66	4.07	4.06	3.99		
Year	1985	1986	1987	1988	1989	1990	1991	1992
	1993	1994	1995	1996	1997	1998		
ERA	4.15	4.17	4.46	3.96	3.88	3.90	4.09	3.94
	4.32	4.80	4.71	5.00	4.57	4.65		
Year	1999	2000	2001	2002	2003			
ERA	4.86	4.91	4.47	4.46	4.52			

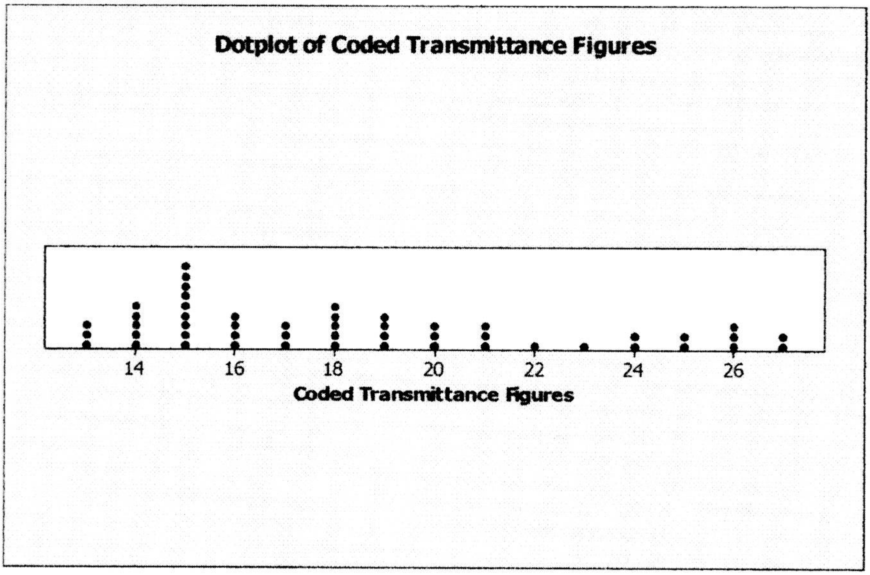
Construct a time sequence plot, either by hand or using software such as MINITAB, or equivalently a scatterplot with ERA plotted against Year. Does the plot reveal a random pattern about the overall average for these 103 years, or does the plot indicate nonrandomness and/or a change in the average?

Solution:
Here is the time sequence plot that is actually in the form of a scatterplot.



(a) There is considerable nonrandomness in this plot, especially the strong upward trend since about 1970 as well as the monotonicity during certain intervals of years (e.g., strictly decreasing from 1938-43.)

1.3. Construct a dotplot for the data in Example 1.2.
Solution:



1.5. Statistical literacy is important not only in engineering but also simply as a means of expression. There are many statistical guffaws that appear in lay publications. Some of these are given in the "Forsooth" section of *RSS News* (Royal Statistical Society News) each month. Others can be found online at *Chance News*, whose website is at the following URL: http://www.dartmouth.edu/~chance/chance_news/news.html. The following two statements can be found at the latter website. Explain what is wrong with each statement.

(a) Migraines affect approximately 14% of women and 7% of men, that's one-fifth the population (*Herbal Health Newsletter Issue 1*)

(b) Researchers at Cambridge University have found that supplementing with vitamin C may help reduce the risk of death by as much as 50% (*Higher Nature Health News* No. HN601, 2001).

(*Comment*: Although the errors in these two statements should be obvious, misstatements involving statistical techniques are often made, even in statistics books, that are not obvious unless one has a solid grasp of statistics.)

Solution:

(a) The percentages are not additive since neither is based on 100% of the population of men and women combined. More specifically, even if the number of women was the same as the number of men, the base would be doubled if the populations were combined, rather than staying at the common number, as would be necessary for the percentages to be additive.

(b) The risk will always be 100% ... regardless of the amount of Vitamin C that is ingested!

1.7. Consider Figure 1.5. The data are as follows (in 75-25PERCENTILES2002.MTW):

Row	SAT 75th percentile	acceptance rate	SAT 25th percentile
1	1540	12	1350
2	1580	11	1410
3	1550	16	1380
4	1580	13	1450
5	1560	16	1410
6	1560	13	1360
7	1490	23	1310
8	1500	26	1300
9	1510	13	1310
10	1520	21	1330
11	1490	44	1280
12	1470	33	1290
13	1510	23	1310
14	1450	30	1290
15	1490	16	1290
16	1480	32	1300
17	1460	45	1300
18	1460	31	1270
19	1430	34	1270
20	1450	26	1200
21	1410	39	1200
22	1400	55	1220
23	1460	36	1280
24	1400	29	1170
25	1380	49	1220
26	1410	26	1240
27	1340	37	1130
28	1410	41	1230
29	1370	38	1160
30	1450	22	1280
31	1420	29	1250
32	1420	48	1220
33	1400	34	1210
34	1410	50	1240
35	1390	32	1220
36	1450	71	1240
37	1365	46	1183
38	1420	57	1250
39	1290	63	1060
40	1275	57	1070
41	1330	79	1130
42	1270	78	1050
43	1290	48	1080
44	1350	36	1150
45	1370	73	1180
46	1290	66	1070
47	1290	47	1090
48	1310	62	1090
49	1390	73	1210

(a) Construct the graph of the acceptance rate against the 75th percentile SAT score with the latter on the horizontal axis. Is the slope exactly the same as the slope of Figure 1.5? Explain why the slope should or should not be the same.

(b) Construct the graph of the 25th percentile SAT score against the acceptance rate with the former on the vertical axis. (The data on the 25th percentile are in the third column in the file.) Does the point that corresponds to point #22 in Figure 1.5 also stand out in this graph?

(c) Compute the difference between the 75th percentile and 25th percentile for each school and plot those differences against the acceptance rate. Note that there are two extreme points on the plot, with differences of 250 and 130, respectively. One of these schools is for a prominent public university and the other is a private university, both in the same state. Which would you guess to be the public university?

Solution:

(a) No, the slopes (of a line fit through the points, for example) will not be the same because the axes are reversed.

(b) No, there are no points on the graph that stand out as being unusual.

(c) We would guess that the point with the higher acceptance rate would be the public university (which it is: University of California --Berkeley)

1.9. Consider different amounts of one-dimensional data. What graphical display would you recommend for each of the following numbers of observations: (a) 10, (b) 100, and (c) 1000?

Solution:

(a) dotplot (b) dotplot or histogram (c) histogram

1.11. The following numbers are the first 50 of 102 chemical data measurements of color from a leading chemical company that were given in Ryan (2000): 0.67, 0.63, 0.76, 0.66, 0.69, 0.71, 0.72, 0.71, 0.72, 0.72, 0.83, 0.87, 0.76, 0.79, 0.74, 0.81, 0.76, 0.77, 0.68, 0.68, 0.74, 0.68, 0.68, 0.74, 0.68, 0.69, 0.75, 0.80, 0.81, 0.86, 0.86, 0.79, 0.78, 0.77, 0.77, 0.80, 0.76, 0.67, 0.73, 0.69, 0.73, 0.73, 0.74, 0.71, 0.65, 0.67, 0.68, 0.71, 0.69, and 0.73.

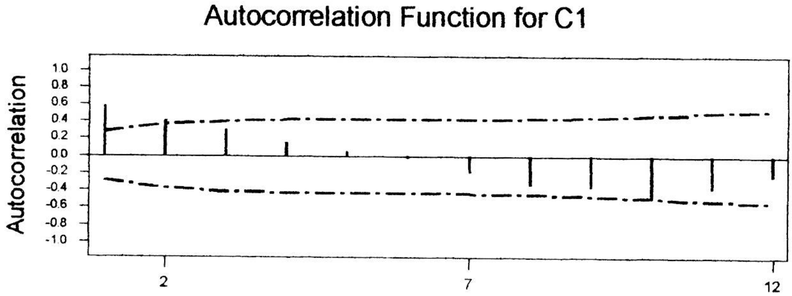
(a) What graphical display would you suggest if it was suspected that there may be some relationship between consecutive measurements (which would violate one of the assumptions of the statistical methods presented in later chapters)?

(b) Construct the display that you suggested in part (a). Do consecutive observations appear to be related?

Solution:

(a) Time sequence plot or autocorrelation plot, preferably the latter.

(b) The autocorrelation function plot (the correlation between units one unit apart, two units apart ...) is given below. There does appear to be autocorrelation, especially of observations one or two units apart since the autocorrelations for the first two lags are above the upper decision line (95% confidence interval) and the *t*-statistics exceed 2, as can be seen from the numbers below the graph.)



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.58	4.11	17.95	8	-0.34	-1.55	43.89
2	0.42	2.30	27.56	9	-0.36	-1.57	52.01
3	0.31	1.53	32.79	10	-0.46	-1.94	65.86
4	0.17	0.81	34.43	11	-0.37	-1.46	75.10
5	0.06	0.30	34.66	12	-0.23	-0.86	78.66
6	-0.02	-0.11	34.69				
7	-0.19	-0.89	36.88				

1.13. This exercise illustrates how the choice of the number of intervals greatly influences the shape of a histogram. Construct a histogram of the first 100 positive integers for the following numbers of classes: 3, 4, 6, 7, 10, and 12. (The number of classes can be specified in MINITAB, for example, by using the NINT subcommand with the HIST command, and the sequence MTB>SET C1, DATA>1:100, DATA>END will place the first 100 integers in the first column of the worksheet.) We know that the distribution of numbers is uniform over the integers 1-100 because we have one of each. We also have the same number of observations in the intervals 1-9, 10-19, 20-29, and so on. Therefore, the histograms should theoretically be perfectly flat. Are any of the histograms flat? In particular, what is the shape when only three classes are used? Explain why this shape results. What does this exercise tell you about relying on a histogram to draw inferences about the shape of the population of values from which the sample was obtained?

Solution:

(simulation exercise by student which shows that histograms are not reliable indicators for the shape of population distributions)

1.15. Explain why consecutive observations that are correlated will be apparent from a digidot plot but not from a dotplot, histogram, stem-and-leaf display, scatter plot, or boxplot. Is there another plot that you would recommend for detecting this type of correlation? Explain.

Solution:
There is no time order involved in a dotplot, histogram, stem-and-leaf display, scatter plot, or boxplot. A time sequence plot would be another possibility.

1.17. Construct a box plot of your driving times from the previous problem. Do any of your times show as an outlier? If the box doesn't exhibit approximate symmetry, try to provide an explanation for the asymmetry.

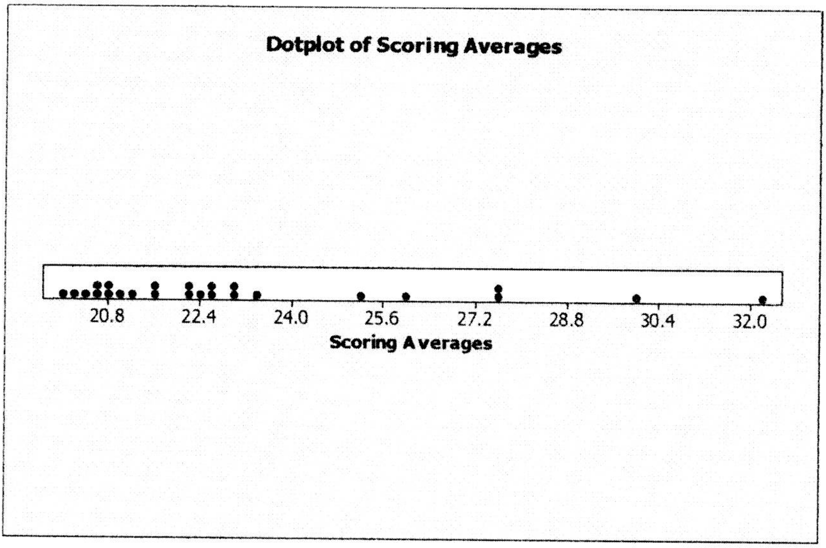
Solution:
(box plot to be individually constructed using the student driving data from Exercise 1.16)

1.19. Given in file NBA2003.MTW are the scoring averages for the top 25 scorers in the National Basketball Association (NBA) in 2002. The data are given below.

	Name		Scoring Average
1	Tracy	McGrady	32.1
2	Kobe	Bryant	30.0
3	Allen	Iverson	27.6
4	Shaquille	O'Neal	27.5
5	Paul	Pierce	25.9
6	Dirk	Nowitzki	25.1
7	Tim	Duncan	23.3
8	Chris	Webber	23.0
9	Kevin	Garnett	23.0
10	Ray	Allen	22.5
11	Allan	Houston	22.5
12	Stephon	Marbury	22.3
13	Antawn	Jamison	22.2
14	Jalen	Rose	22.1
15	Jamal	Mashburn	21.6
16	Jerry	Stackhouse	21.5
17	Shawn	Marion	21.2
18	Steve	Francis	21.0
19	Glenn	Robinson	20.8
20	Jermaine	O'Neal	20.8
21	Ricky	Davis	20.6
22	Karl	Malone	20.6
23	Gary	Payton	20.4
24	Antoine	Walker	20.1
25	Michael	Jordan	20.0

What type of graphical display would you recommend for displaying the data? Construct the display, but before doing so, would you expect the averages to exhibit asymmetry? Why or why not?

Solution:
A dotplot, given below, would be a reasonable way to display the data. We would expect the data to display right skewness as we would expect to see more scoring averages below the median of the top 25 than above the median.



1.21. With a conventional scatter plot, two variables are displayed --- one on the vertical axis and one on the horizontal axis. How many variables were displayed in the scatter plot in Figure 1.5? Can you think of how additional variables might be displayed?

Solution:
Three variables were displayed in Figure 1.5. A fourth variable could be displayed by having a separate graph for each value of that variable. And so on.

1.23. A data set contains 25 observations. The median is equal to 26.8, the range is 62, $Q_1 = 16.7$, and $Q_3 = 39.8$. What is the numerical value of the interquartile range?

Solution:
The interquartile range is $Q_3 - Q_1 = 39.8 - 16.7 = 23.1$

1.25. Would a histogram of the data given in Exercise1.1 be a meaningful display? Why or why not?

Solution:
No, a histogram would not be particularly useful. Since the data are obtained over time, a display that incorporates time should be used.

1.27. What graphical display discussed in this chapter would be best suited for showing the breakdown of the number of Nobel Prize winners by country for a specified time period?

Solution:
A Pareto chart could be a good choice, depending on how many countries are represented. A bar chart would be another possibility, and for political reasons might be the preferred alternative.

1.29. Toss a coin twice and record the number of tails; then do this nine more times. Does the string of numbers appear to be random?

Solution:
(simulation exercise to be performed by student)

1.31. In an article in the Winter, 2001 issue of *Chance* magazine, the author, Derek Briggs found that SAT and ACT preparation courses had a limited impact on students' test results, contrary to what companies that offer these courses have claimed. Read this article and write a report explaining how an experiment would have to be conducted before any claim of usefulness of these courses could be made. (source: http://www.public.iastate.edu/~chance99/141.briggs.pdf)

Solution:
(Student exercise; write a report of the article)

1.33. The following data are frequency distributions of weights of cars and trucks sold in the United States in 1975 and 1990. (*Source*: U.S. Environmental Protection Agency, *Automotive Technology and Full Economic Trends through 1991*, EPA/AA/CTAB/91-02, 1991.)

WT	WT(L)	WT (U)	CA75	TR75	CA90	TR90
1750	1625	1875	0	0	1	0
2000	1875	2125	105	0	109	0
2250	2125	2375	375	0	107	0
2500	2375	2625	406	0	1183	34
2750	2625	2875	281	204	999	45
3000	2875	3250	828	60	3071	428
3500	3250	3750	1029	55	2877	784
4000	3750	4250	1089	1021	1217	1260
4500	4250	4750	1791	386	71	797
5000	4750	5250	1505	201	0	457
5500	5250	5750	828	59	1	46
6000	5750	6250	0	1	0	32

Variable Names:

WT: Weight in pounds, class midpoint
WT(L): Weight in pounds, class lower limit
WT(U): Weight in pounds, class upper limit
CA75: Cars sold, 1975 (thousands)
TR75: Trucks sold, 1975 (thousands)
CA90: Cars sold, 1990 (thousands)
TR90: Trucks sold, 1990 (thousands)

(a) Compare the distributions of CA75 and CA90 by constructing a histogram of each. Comment on the comparison. In particular, does there appear to have been a significant change in the distribution from 1975 to 1990? If so, what is the change? (In MINITAB, the histograms can be constructed using the CHART command with the C1*C2 option; that is, CHART C1 C2 with C1 containing the data and C2 being a category variable, and these two column numbers being arbitrary designations.)

(b) Construct the histograms for TR75 and TR90 and answer the same questions as in part (a).

(c) Having constructed these four histograms, is there any problem posed by the fact that the intervals are not of equal width? In particular, does it create a problem relative to the 1975 and 1990 comparisons? If so, how would you correct for the unequal widths? If necessary, make the appropriate correction. Does this affect the comparison?

(d) In view of the small number of observations, would it be better to use another type of graphical display for the comparison? If so, use that display and repeat the comparisons.

Solution:

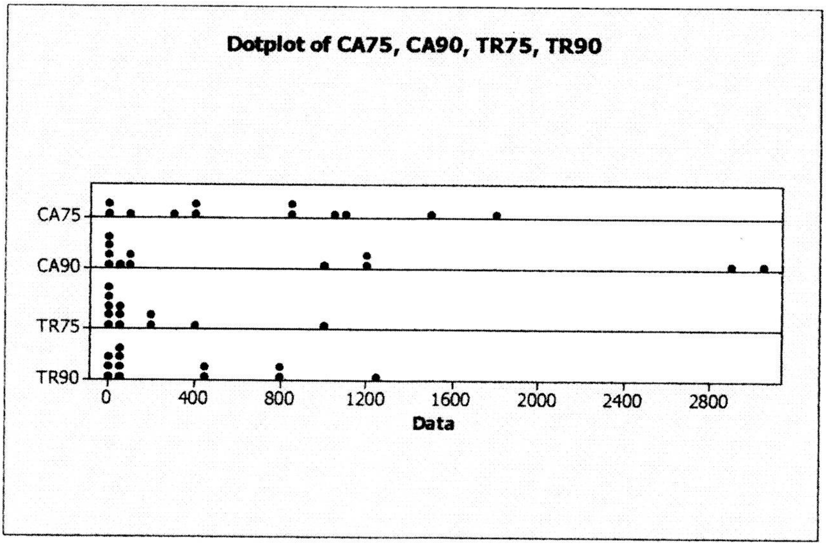
(a) The histograms are similar in that they both exhibit extreme right skewness. Beyond that, a finer comparison would not be practical since each histogram is constructed for only 12 observations, so a very large histogram variation would be observed in repeated sampling.

(b) Although both histograms exhibit right skewness, the two large values for TR90 coupled with the fact that there are only 12 observations make it impossible to construct a meaningful histogram.

(c) The unequal widths are a problem with the comparison of TR 75 and TR 90, with the difference in the widths primarily due to the two large values of TR 90.

(d) It would be more meaningful to construct multiple dotplots with the same scale used for each comparison. These effectively show the gaps between the

CA90 and TR90 values, especially in regard to the much smaller differences in the corresponding CA75 and TR75 values, respectively.



1.35. Use appropriate software, such as MINITAB, or Table A in the back of the book to generate three samples of size 20 from the first 50 positive integers. Compare the three samples. Is there much variability between the three samples? If so, would you have anticipated this amount of variability?

Solution:
(student simulation exercise)

1.37. Consider the Lighthall (1991) article that was discussed at the beginning of the chapter. If you are presently taking engineering courses, can you think of data that should be collected and analyzed on some aspect in an engineering discipline, but that are usually not collected and analyzed? Explain.

Solution:
(student exercise)

1.39. Given the following stem-and-leaf display,

3| 1 2 2 4 5 7
4| 1 3 5 7 7 9
5| 2 4 5 6 8 9 9
6| 1 3 3 4 7 8

determine the median.