Solutions Manual for STATISTICAL ANALYSIS FOR DECISION MAKING

THIRD EDITION

Morris Hamburg

Solutions Manual for
STATISTICAL ANALYSIS FOR DECISION MAKING
THIRD EDITION

Morris Hamburg

The Wharton School
University of Pennsylvania

Harcourt Brace Jovanovich, Inc.

New York San Diego Chicago San Francisco Atlanta
London Sydney Toronto

ISBN: 0-15-583451-7

Printed in the United States of America

CHAPTER 1

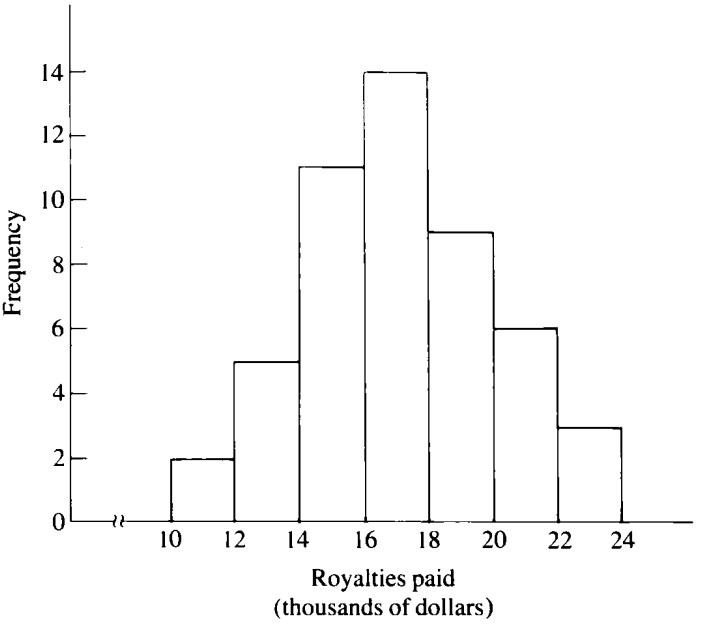
Section 1.6 (page 16)

1. a. For 50 numbers, we can reasonably use seven classes.

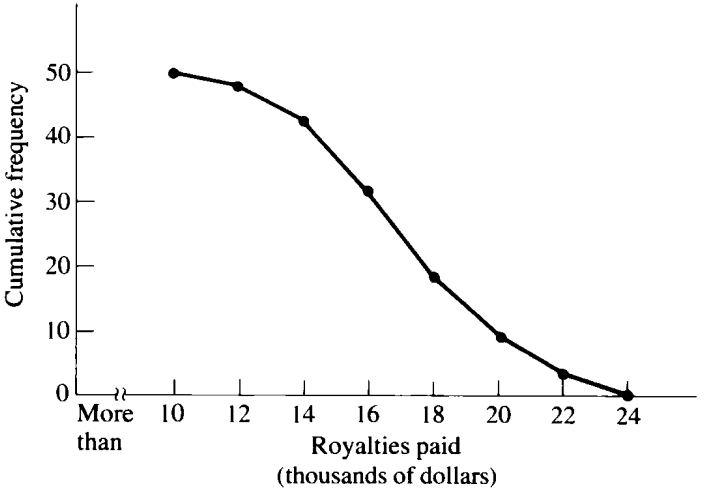
   b. Using seven classes, we obtain the following distribution for royalties paid by Champion House:

Royalties Paid (in dollars)	Frequency
\$10,000 and under \$12,000	2
12,000 and under 14,000	5
14,000 and under 16,000	11
16,000 and under 18,000	14
18,000 and under 20,000	9
20,000 and under 22,000	6
22,000 and under 24,000	3
Total	50

That distribution yields the following histogram for royalties paid:



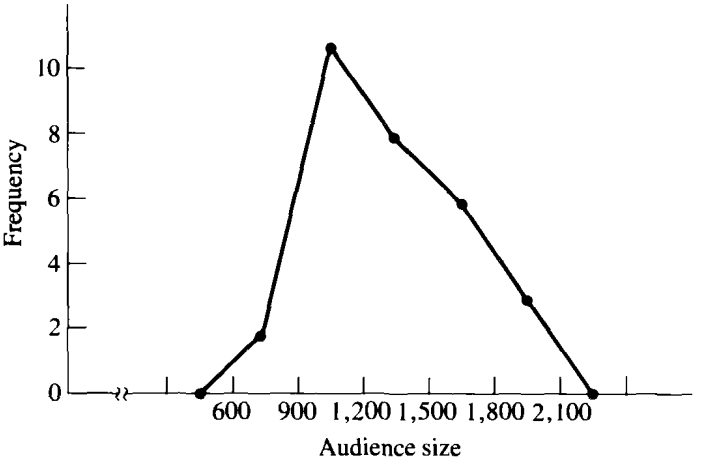
c. A "more than" ogive for royalties paid appears as follows:

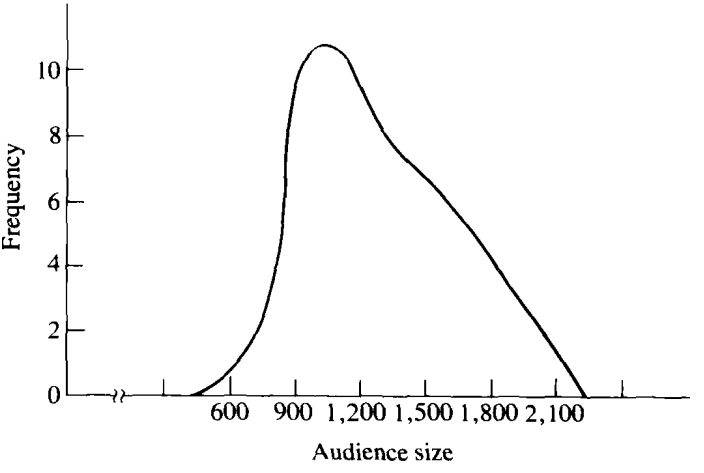


2. a. Using five classes we have the following table for audiences at Expo-Classic:

	Frequency
600 and under 900	2
900 and under 1,200	11
1,200 and under 1,500	8
1,500 and under 1,800	6
1,800 and under 2,100	3
Total	30

b. The frequency polygon and frequency curve, respectively, for audience sizes appear as follows:

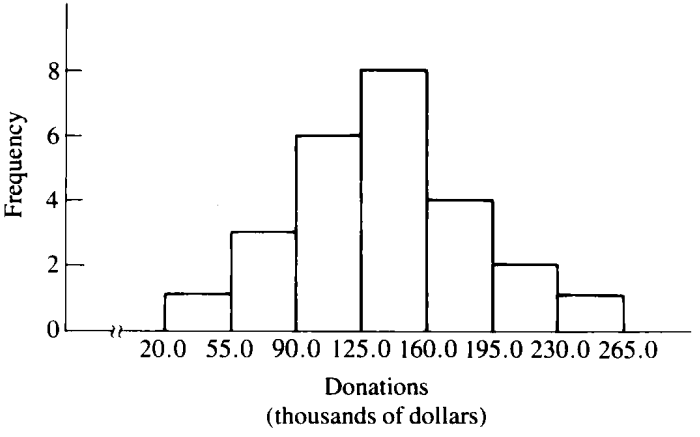


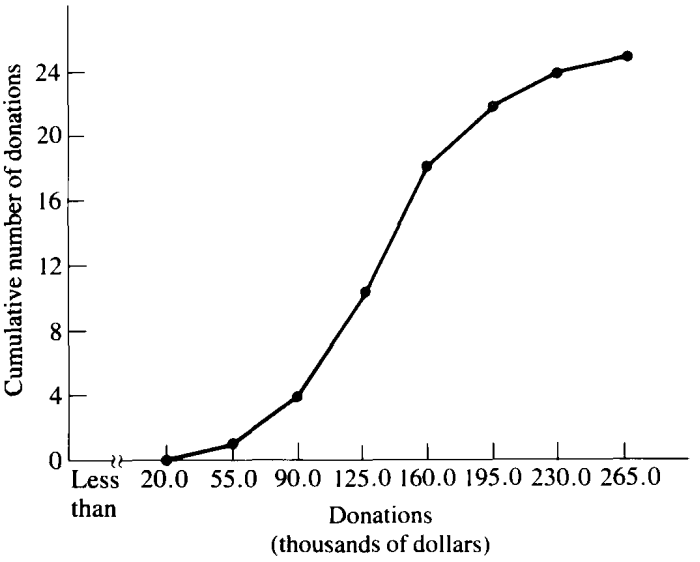


3. a. Using seven classes, we have the following distribution for donations received at Union-Path offices:

Donations (thousands of dollars)	Midpoints	Frequency
20.0 and under 55.0	37.5	1
55.0 and under 90.0	72.5	3
90.0 and under 125.0	107.5	6
125.0 and under 160.0	142.5	8
160.0 and under 195.0	177.5	4
195.0 and under 230.0	212.5	2
230.0 and under 265.0	247.5	1
Total		25

b. That distribution yields the following histogram and "less than" ogive, respectively.





4. The cumulative frequency distribution is as follows:

Policy Size	Number of Ordinary Life Insurance Policies in Force per 1,000 Policies
Less than $1,000	31
Less than $2,500	212
Less than $5,000	320
Less than $10,000	533
Less than maximum	1,000

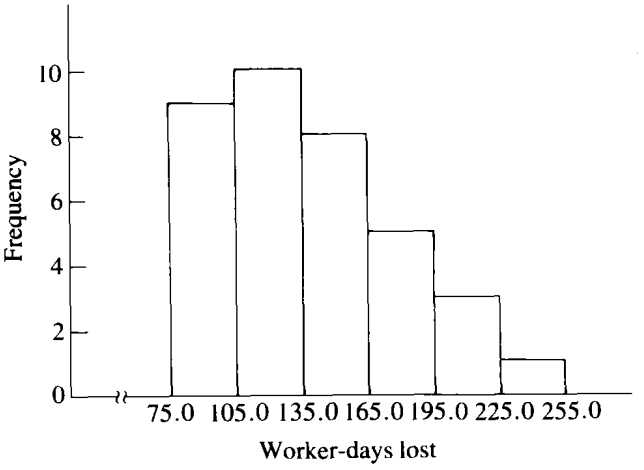
5. a. The class intervals are not mutually exclusive; the class limits are unclear and overlapping.

   b. There is a gap in values between classes; thus, the classes are not exhaustive.

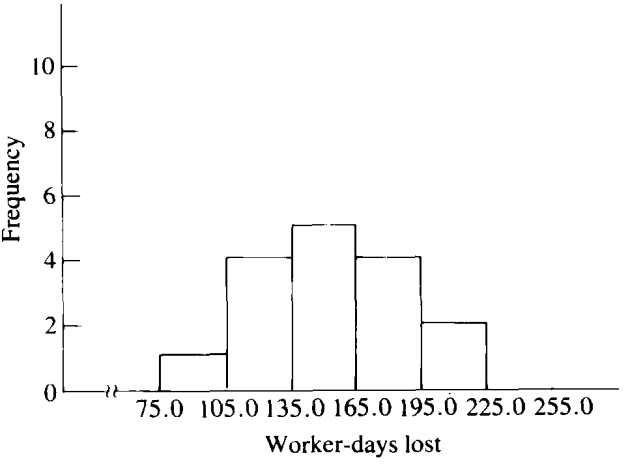
6. a. Using six classes, we have the following frequency distribution for worker-days lost last year at Ultra Corporation:

Worker-days Lost	Frequency		
	Total	Midwest	South
75.0 and under 105.0	9	8	1
105.0 and under 135.0	10	6	4
135.0 and under 165.0	8	3	5
165.0 and under 195.0	5	1	4
195.0 and under 225.0	3	1	2
225.0 and under 255.0	1	1	0
Total	36	20	16

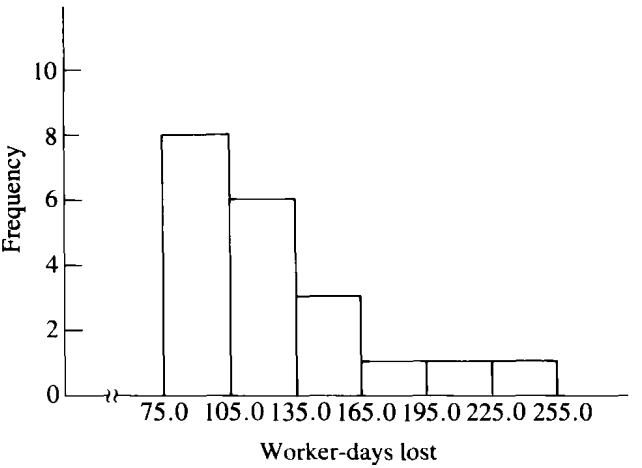
We obtain the following histogram for both regions combined:



b. Histogram for South alone:



Histogram for Midwest alone:



c. The South appears to have the bigger problem, because more of the area under its histogram is toward the right than for the Midwest. The histogram for the Midwest is concentrated toward the left.

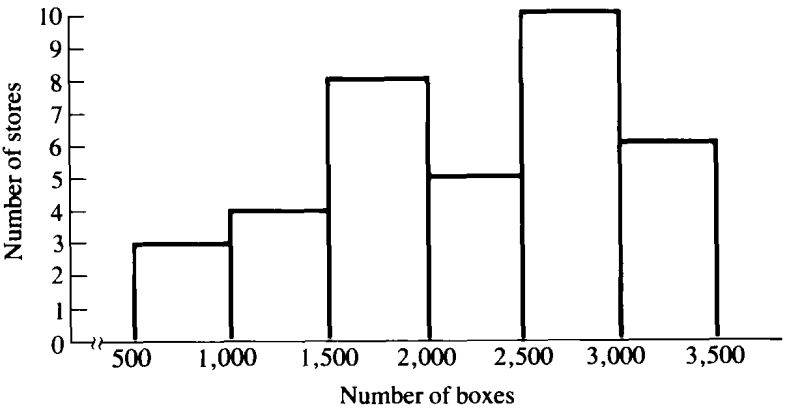
From part a we get an overall picture. Part b

shows us the differences between the regions and how those differences affect the overall distribution.

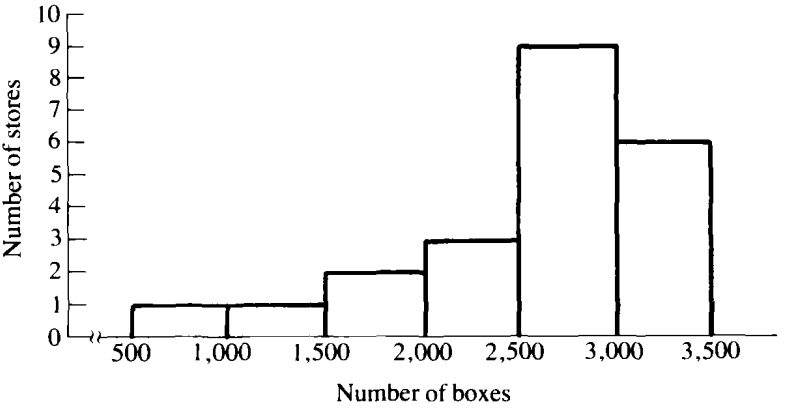
7. a and b. Number of boxes sold during the one-week period:

Number of Boxes	Number of Stores		
	Total	Philadelphia	Camden
500 and under 1,000	3	1	2
1,000 and under 1,500	4	1	3
1,500 and under 2,000	8	2	6
2,000 and under 2,500	5	3	2
2,500 and under 3,000	10	9	1
3,000 and under 3,500	6	6	0
Total	36	22	14

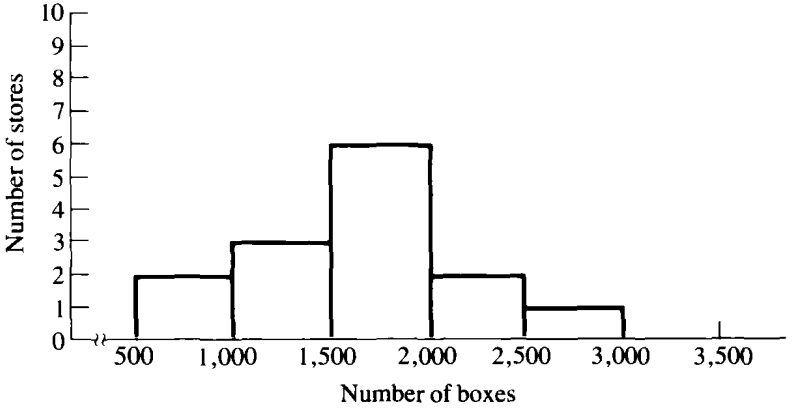
Histogram for the two cities combined:



Histogram for Philadelphia:



Histogram for Camden:



c. The answer in part b is preferable, because a bimodal distribution results if the data for the two market cities are merged.

Section 1.10 (page 26)

1. Mean from the original data is calculated as follows:

$$\bar{X} = \frac{\$10,214 + \$11,394 + \cdots + \$23,854}{50} = \frac{\$862,068}{50}$$

$$= \$17,241.36$$

From the grouped data we have the following:

Class Midpoint m	Frequency f	fm
$11,000	2	$ 22,000
13,000	5	65,000
15,000	11	165,000
17,000	14	238,000
19,000	9	171,000
21,000	6	126,000
23,000	3	69,000
Total	50	$856,000

$$\bar{X} = \frac{\Sigma fm}{n} = \frac{\$856,000}{50} = \$17,120$$

a. The difference arises from the assumption made in the frequency distribution that the midpoints are the means of the observations in each class interval. Generally this is not so, and in the absence of any compensating factors, we would expect a difference.

b. The mean computed from grouped data would equal the mean computed from original data under either of the following conditions:

   (1) If the means of the items within classes are equal to the midpoints of the classes. The original data may or may not be evenly distributed within the classes.

   (2) If some class means are greater and some are less than the midpoints and the differences compensate exactly.

2. No, this is not enough information for us to make a decision. Because we do not know the proportions of the ingredients in the two spreads, we cannot say that Brandex has a lower overall Oxy-toxin content. For example, if the proportions are as follows, we could conclude that Brandex actually has a higher Oxy-toxin level:

	Proportion		Weighted Oxy-toxin Level (%)	
Ingredient	Brandex	Leading Spread	Brandex	Leading Spread
Moonflower oil	0.4	0.1	0.48	0.18
Processed butter-glop	0.3	0.4	0.18	0.32
Flavoring	0.1	0.4	0.03	0.12
Coloring	0.2	0.1	0.10	0.06
Total	1.0	1.0	0.79	0.68

From this table, we see an overall level of 0.79% for Brandex compared with an overall level of 0.68% for the leading spread. Therefore, we need to know the proportions of ingredients before we can make a decision. We also need to know about any ingredients that have not been advertised and the levels of Oxy-toxin in those ingredients as well.

3. a. For the first strategy, average cost of United Aerodynamics is ($5,275/125) = $42.20; average cost of Mitton Industries is ($4,675/125) = $37.40. For the second, average cost of United Aerodynamics is ($5,039/125) = $40.31; average cost of Mitton Industries is ($4,959/141) = $35.17.

   b. The second strategy achieved the lower average cost for both stocks.

   c. For the second strategy, since the number of shares bought each time varied inversely with the stock price, higher weights were placed on the lower stock prices in the calculation of the average cost per share. This resulted in a lower overall average for each stock than in the first strategy, in which the average cost per share was the unweighted mean of the five share prices. That is, if equal weights are used in computing a weighted arithmetic mean, the resulting figure is the same as the unweighted mean.

4. a. The overall percentage of defectives for the last 50 runs is

$$\bar{X} = [(0.25)(3) + (0.75)(12) + (1.25)(24) + (1.75)(9) + (2.25)(2)]/50 = 1.2\%$$

b. This method would not necessarily give the same answer, because the answer in part a is computed from a grouped data table using class midpoints, whereas the other is a mean computed from raw data.

5. a. Weighted average of pollutant level in the three tanks is

$$\bar{X} = \frac{(4.5)(261,432) + (15.0)(118,300) + (21.3)(287,456)}{667,188}$$

$$= 13.6 \text{ parts per } 10,000$$

b. Unweighted average is

$$\bar{X} = \frac{4.5 + 15.0 + 21.3}{3} = 13.6 \text{ parts per } 10,000$$

c. Yes, the answer in part b shows an unexpected correspondence to that in part a.

Section 1.13 (page 37)

1. $\bar{X} = \left(\frac{\$30 + \$45 + \cdots + \$41 + \$21}{30}\right) = \$38.97$

To find the median we have to rank the data:

$15	$30	$33	$41	$46	$49
21	30	33	42	47	50
26	30	39	45	48	51
28	32	40	45	48	51
29	33	41	45	49	52

Since we have 30 values, the median is between the fifteenth and sixteenth entries.

$$Md = \left(\frac{\$41 + \$41}{2}\right) = \$41$$

Three values occur with equal frequency--30, 33, and 45. Thus, there is no unique mode.

2. Using five classes, we get the following distribution:

Closing Prices (in dollars)	Midpoints	Frequency	
$15 and under $23	19	2	
23 and under 31	27	6	
31 and under 39	35	4	$\Sigma f_p = 12$
39 and under 47	43	9	
47 and under 55	51	9	
Total		30	

$$\bar{X} = \frac{\Sigma fm}{n} = \frac{1,186}{30} = \$39.53$$

$$Md = 39 + \left(\frac{\$15 - \$12}{9}\right)(8) = \$41.67$$

The differences between these answers and the answers in exercise 1 arise because of the assumption of even distribution of the values.

3. In all three cases, it is quite logical to assume that extreme values would appear at the upper end of the scale and that such values would tend to make the arithmetic mean larger than the median.

4. a. The median class and the modal class are the same--$20.0 million and under $25.0 million.

   b. The median is obtained by interpolation.

$$Md = 20.0 + \left(\frac{\$50 - \$47}{26}\right)(5) = \$20.6 \text{ million}$$

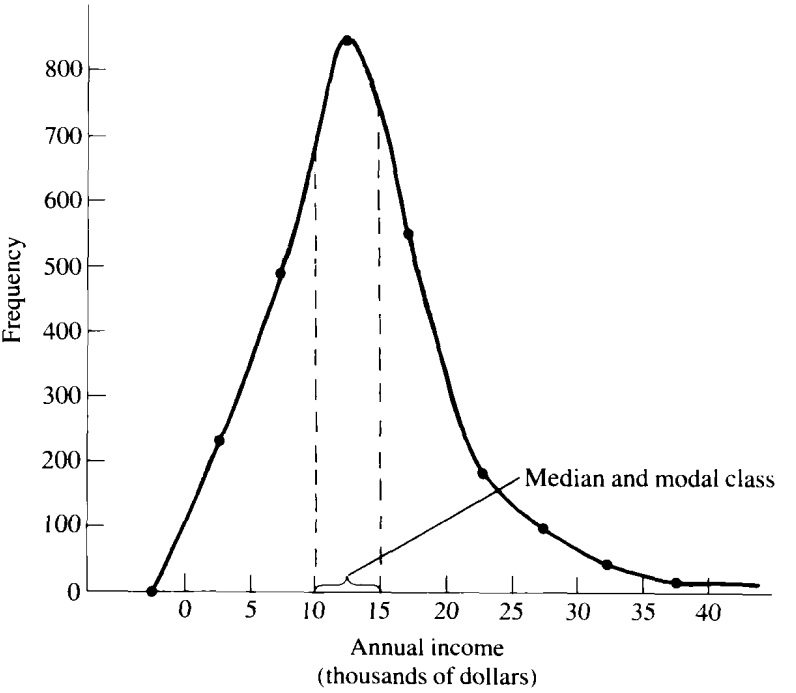
5. a. Yes, it is appropriate. First, the number of observations in this class is quite small compared with the size of the sample. Second, using closed classes to include these few outlying values would be impractical and would tend to distort the major portion of the distribution.

   b. The modal class and the median class are the same--"$10,000 and under $15,000." The median is obtained by interpolation.

$$Md = \$10,000 + \left(\frac{1,250 - 733}{851}\right)5,000 = \$13,037.60$$

$$Mode = \text{Midpoint of modal class} = \$12,500$$

The frequency curve for annual income distribution appears below:



Note that the curve is left open at the upper end. The open-ended class at the upper end indicates that we could expect a skew to the right in this distribution.

c. "More than" cumulative frequency distribution of annual income data:

Annual Income (dollars)	Cumulative Frequency
More than $40,000	8
More than 35,000	29
More than 30,000	76
More than 25,000	178
More than 20,000	361
More than 15,000	916
More than 10,000	1,767
More than 5,000	2,265
More than 0	2,500