2. The U.S. Department of Energy provides fuel economy information for a variety of motor vehicles. A sample of 10 automobiles is shown in Table 1.6 (Fuel Economy website, February 22, 2008). Data show the size of the automobile (compact, midsize, or large), the number of cylinders in the engine, the city driving miles per gallon, the highway driving miles per gallon, and the recommended fuel (diesel, premium, or regular).
   a. How many elements are in this data set?
   b. How many variables are in this data set?
   c. Which variables are categorical and which variables are quantitative?
   d. What type of measurement scale is used for each of the variables?

TABLE 1.6 FUEL ECONOMY INFORMATION FOR 10 AUTOMOBILES

Car	Size	Cylinders	City MPG	Highway MPG	Fuel
Audi A8	Large	12	13	19	Premium
BMW 328Xi	Compact	6	17	25	Premium
Cadillac CTS	Midsize	6	16	25	Regular
Chrysler 300	Large	8	13	18	Premium
Ford Focus	Compact	4	24	33	Regular
Hyundai Elantra	Midsize	4	25	33	Regular
Jeep Grand Cherokee	Midsize	6	17	26	Diesel
Pontiac G6	Compact	6	15	22	Regular
Toyota Camry	Midsize	4	21	31	Regular
Volkswagen Jetta	Compact	5	21	29	Regular

Economy of fuel

Sample size =n=10

a. The number of elements= number of observations=number of records= same as sample size=10
b. Number of variables=Number of features= Number of columns=6
c. Categorical variables = Car, Size, Fuel. Quantitative variables=cylinders, city MPG, Highway MPG
d. Measurement scale, Car- Nominal, Size=Ordinal qualitative, Cylinders= **Ratio**, City Mpg=Ratio, Highway MPG=Ratio, Fuel=Ordinal qualitative.

3. Refer to Table 1.6.
   a. What is the average miles per gallon for city driving?
   b. On average, how much higher is the miles per gallon for highway driving as compared to city driving?

c. What percentage of the cars have four-cylinder engines?
d. What percentage of the cars use regular fuel?

TABLE 1.6 FUEL ECONOMY INFORMATION FOR 10 AUTOMOBILES

Car	Size	Cylinders	City MPG	Highway MPG	Fuel
Audi A8	Large	12	13	19	Premium
BMW 328Xi	Compact	6	17	25	Premium
Cadillac CTS	Midsize	6	16	25	Regular
Chrysler 300	Large	8	13	18	Premium
Ford Focus	Compact	4	24	33	Regular
Hyundai Elantra	Midsize	4	25	33	Regular
Jeep Grand Cherokee	Midsize	6	17	26	Diesel
Pontiac G6	Compact	6	15	22	Regular
Toyota Camry	Midsize	4	21	31	Regular
Volkswagen Jetta	Compact	5	21	29	Regular

a. Average miles per gallon for city driving =

$$\mu = \frac{(13+17+16+13+24+25+17+15+21+21+)}{10}$$

$$= \frac{182}{10} = 18.2$$

b. Average miles per gallon for high way=

$$\mu = \frac{(19+25+25+18+33+33+26+22+31+29)}{10}$$

$$\mu = \frac{261}{10}$$

$$\mu = 26.1$$

On Average 26.1-18.2=7.9 higher is miles per gallon for high way driving as compared to city driving.

c. Total number of cars having cylinder = 10

Number cars having 4 cylinders =3

$$= \frac{3x100}{10}$$

% of cars having 4 cylinders

=30%

d. Total number of cars using fuel= 10

Out of 10 Number of cars using regular fuel= 6

% or cars using regular fuel $= \frac{6x100}{10}$

= 60%

4. Table 1.7 shows data for seven colleges and universities. The endowment (in billions of dollars) and the percentage of applicants admitted are shown (*USA Today*, February 3, 2008). The state each school is located in, the campus setting, and the NCAA Division for varsity teams were obtained from the National Center of Education Statistics website, February 22, 2008.
   a. How many elements are in the data set?
   b. How many variables are in the data set?
   c. Which of the variables are categorical and which are quantitative?

TABLE 1.7 DATA FOR SEVEN COLLEGES AND UNIVERSITIES

School	State	Campus Setting	Endowment ($ billions)	% Applicants Admitted	NCAA Division
Amherst College	Massachusetts	Town: Fringe	1.7	18	III
Duke	North Carolina	City: Midsize	5.9	21	I-A
Harvard University	Massachusetts	City: Midsize	34.6	9	I-AA
Swarthmore College	Pennsylvania	Suburb: Large	1.4	18	III
University of Pennsylvania	Pennsylvania	City: Large	6.6	18	I-AA
Williams College	Massachusetts	Town: Fringe	1.9	18	III
Yale University	Connecticut	City: Midsize	22.5	9	I-AA

a. Number of elements= number of observations=sample size=number of records = 7
b. Number of variables= Number of features influencing the area of interest (schoo)= number of columns excluding area of interest variable = 5
c. Categorical Variable= State, Campus setting, NCAA Division. Quantitative=Endowment($ billions), %Applicants Admitted.

5. Consider the data set in Table 1.7
   a. Compute the average endowment for the sample.
   b. Compute the average percentage of applicants admitted.
   c. What percentage of the schools have NCAA Division III varsity teams?
   d. What percentage of the schools have a City: Midsize campus setting?

TABLE 1.7 DATA FOR SEVEN COLLEGES AND UNIVERSITIES

School	State	Campus Setting	Endowment ($ billions)	% Applicants Admitted	NCAA Division
Amherst College	Massachusetts	Town: Fringe	1.7	18	III
Duke	North Carolina	City: Midsize	5.9	21	I-A
Harvard University	Massachusetts	City: Midsize	34.6	9	I-AA
Swarthmore College	Pennsylvania	Suburb: Large	1.4	18	III
University of Pennsylvania	Pennsylvania	City: Large	6.6	18	I-AA
Williams College	Massachusetts	Town: Fringe	1.9	18	III
Yale University	Connecticut	City: Midsize	22.5	9	I-AA

a. Average Endowment of sample=

$$\mu = \frac{(1.7+5.9+34.6+1.4+6.6+1.9+22.5)}{7}$$

$$\mu = \frac{74.6}{7}$$

= 10.65

b. Average % of applicants admitted=

$$\mu = \frac{(18+21+9+18+18+18+9)}{7}$$

$$\mu = \frac{111}{7}$$

=15.87

c. Total number of schools having NCCA division= 7

Out of 7 number of schools, schools having III NCCAA division=3

$$= \frac{3}{7} x100$$

% of school having III NCCAA division =

= 42.85%

d. Total number of schools =7

Out of 7 number of schools having campus setting as midsize= 3

$$= \frac{3}{7} x100$$

% of schools having campus setting as midsize=

= 42.85%

6. *Foreign Affairs* magazine conducted a survey to develop a profile of its subscribers (Foreign Affairs website, February 23, 2008). The following questions were asked.
   a. How many nights have you stayed in a hotel in the past 12 months?
   b. Where do you purchase books? Three options were listed: Bookstore, Internet, and Book Club.
   c. Do you own or lease a luxury vehicle? (Yes or No)
   d. What is your age?
   e. For foreign trips taken in the past three years, what was your destination? Seven international destinations were listed.

   Comment on whether each question provides categorical or quantitative data.

a. Quantitative
b. Categorical
c. Categorical
d. Quantitative
e. Categorical

7. The Ritz-Carlton Hotel used a customer opinion questionnaire to obtain performance data about its dining and entertainment services (The Ritz-Carlton Hotel, Naples, Florida, February 2006). Customers were asked to rate six factors: Welcome, Service, Food, Menu Appeal, Atmosphere, and Overall Experience. Data were recorded for each factor with 1 for Fair, 2 for Average, 3 for Good, and 4 for Excellent.
   a. The customer responses provided data for six variables. Are the variables categorical or quantitative?
   b. What measurement scale is used?

a. Categorical

b. Ordinal scale of measurement is used.

8. The *FinancialTimes*/Harris Poll is a monthly online poll of adults from six countries in Europe and the United States. A January poll included 1015 adults in the United States. One of the questions asked was, "How would you rate the Federal Bank in handling the

credit problems in the financial markets?" Possible responses were Excellent, Good, Fair, Bad, and Terrible (Harris Interactive website, January 2008).

a. What was the sample size for this survey?

b. Are the data categorical or quantitative?

c. Would it make more sense to use averages or percentages as a summary of the data for this question?

d. Of the respondents in the United States, 10% said the Federal Bank is doing a good job. How many individuals provided this response?

a. Sample size= Number of observations = 1015

b. Categorical

c. No.

$$= \frac{x}{1015} x100$$

d. 10%=

$$x = \frac{10x1015}{100}$$

=101.5 individuals provided this response

9. The Commerce Department reported receiving the following applications for the Malcolm Baldrige National Quality Award: 23 from large manufacturing firms, 18 from large service firms, and 30 from small businesses.

a. Is type of business a categorical or quantitative variable?

b. What percentage of the applications came from small businesses?

a. Quantitative

$$= 23+18+30$$

b. Total number of applications came =

=71

Number of applications from small business = 30

% of applications from small business =

$$= \frac{30}{71} x100$$

=42.25%

10. *The Wall Street Journal (WSJ)* subscriber survey (October 13, 2003) asked 46 questions about subscriber characteristics and interests. State whether each of the following questions provided categorical or quantitative data and indicate the measurement scale appropriate for each.
    a. What is your age?
    b. Are you male or female?
    c. When did you first start reading the *WSJ*? High school, college, early career, mid-career, late career, or retirement?
    d. How long have you been in your present job or position?
    e. What type of vehicle are you considering for your next purchase? Nine response categories include sedan, sports car, SUV, minivan, and so on.

a. Quantitative- Ratio
b. Categorical – Nominal
c. Categorical-ordinal
d. Quantitative-Ratio
e. Categorical-Nominal

11. State whether each of the following variables is categorical or quantitative and indicate its measurement scale.
    a. Annual sales
    b. Soft drink size (small, medium, large)
    c. Employee classification (GS1 through GS18)
    d. Earnings per share
    e. Method of payment (cash, check, credit card)

a. Quantitative- Ratio
b. Categorical- Ordinal
c. Categorical- Ordinal
d. Quantitative- Ratio
e. Categorical-Nominal

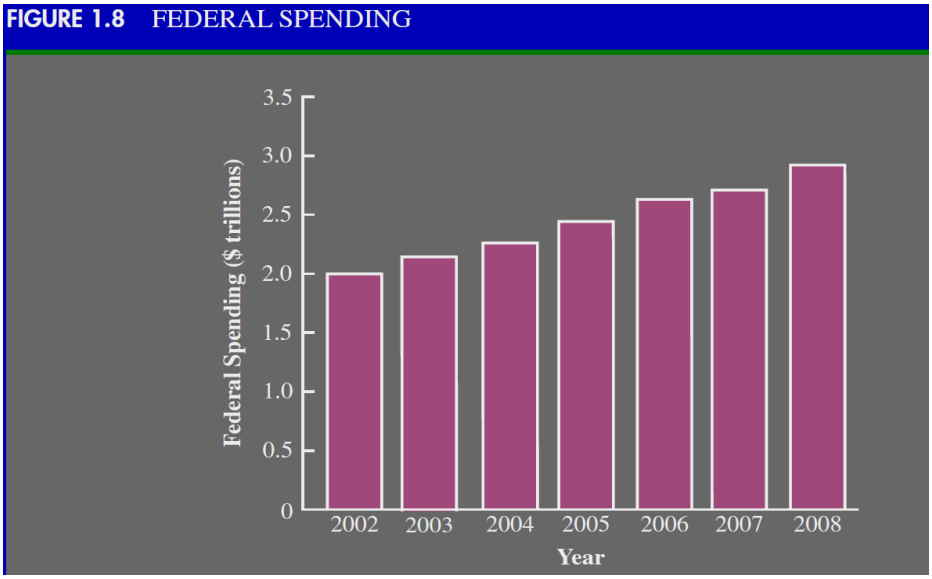
12. The Hawaii Visitors Bureau collects data on visitors to Hawaii. The following questions were among 16 asked in a questionnaire handed out to passengers during incoming airline flights in June 2003.
    - This trip to Hawaii is my: 1st, 2nd, 3rd, 4th, etc.
    - The primary reason for this trip is: (10 categories including vacation, convention, honeymoon)
    - Where I plan to stay: (11 categories including hotel, apartment, relatives, camping)
    - Total days in Hawaii

    a. What is the population being studied?
    b. Is the use of a questionnaire a good way to reach the population of passengers on incoming airline flights?
    c. Comment on each of the four questions in terms of whether it will provide categorical or quantitative data.

a. The population is Total number of Visitors to Hawaii in June 2003.
b. Yes.
c. Categorical- Ordinal, categorical-Nominal, Categorical-Nominal, Quantitative-Ratio.

13. Figure 1.8 provides a bar chart showing the amount of federal spending for the years 2002 to 2008 (*USA Today*, February 5, 2008).
    a. What is the variable of interest?
    b. Are the data categorical or quantitative?
    c. Are the data time series or cross-sectional?
    d. Comment on the trend in federal spending over time.

FIGURE 1.8 FEDERAL SPENDING



a. Federal Spending in $trillions
b. Quantitative
c. Time Series
d. Federal Spending is increasing over the period of time. It reached highest peak in 2008 and it was at its low in 2008.

14. CSM Worldwide forecasts global production for all automobile manufacturers. The following CSM data show the forecast of global auto production for General Motors, Ford, DaimlerChrysler, and Toyota for the years 2004 to 2007 (*USA Today*, December 21, 2005). Data are in millions of vehicles.

Manufacturer	2004	2005	2006	2007
General Motors	8.9	9.0	8.9	8.8
Ford	7.8	7.7	7.8	7.9
DaimlerChrysler	4.1	4.2	4.3	4.6
Toyota	7.8	8.3	9.1	9.6

a. Construct a time series graph for the years 2004 to 2007 showing the number of vehicles manufactured by each automotive company. Show the time series for all four manufacturers on the same graph.

b. General Motors has been the undisputed production leader of automobiles since 1931. What does the time series graph show about who is the world's biggest car company? Discuss.

c. Construct a bar graph showing vehicles produced by automobile manufacturer using the 2007 data. Is this graph based on cross-sectional or time series data?

a.

b.Toyota

d. Cross-sectional

15. The Food and Drug Administration (FDA) reported the number of new drugs approved over an eight-year period (*The Wall Street Journal*, January 12, 2004). Figure 1.9 provides a bar chart summarizing the number of new drugs approved each year.

a. Are the data categorical or quantitative?

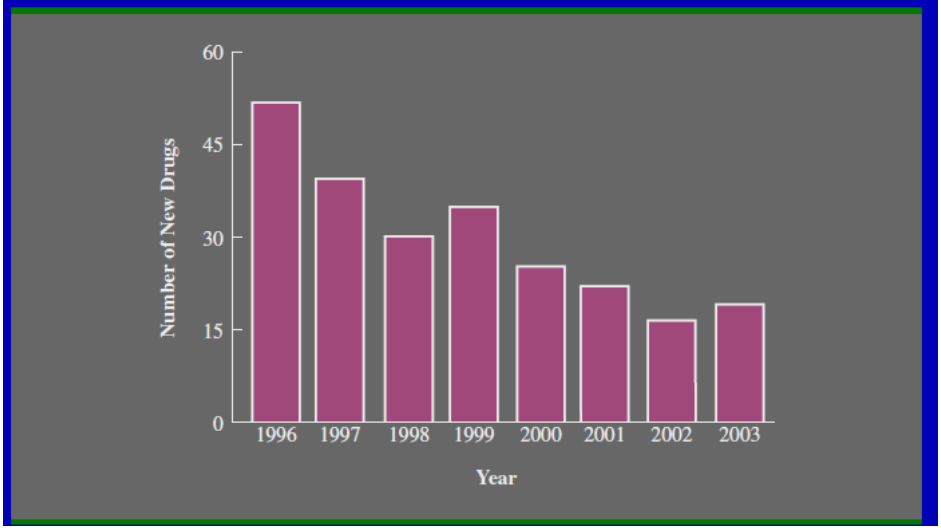
b. Are the data time series or cross-sectional?

c. How many new drugs were approved in 2003?

d. In what year were the fewest new drugs approved? How many?

e. Comment on the trend in the number of new drugs approved by the FDA over the eight-year period.

FIGURE 1.9 NUMBER OF NEW DRUGS APPROVED BY THE FOOD AND DRUG ADMINISTRATION



a. Quantitative
b. Time series
c. 17
d. 2002-15
e. Approval of new drug is gradually decreasing over the period of time.

16. The Energy Information Administration of the U.S. Department of Energy provided time series data for the U.S. average price per gallon of conventional regular gasoline between July 2006 and June 2009 (Energy Information Administration website, June 2009). Use the Internet to obtain the average price per gallon of conventional regular gasoline since June 2009.
    a. Extend the graph of the time series shown in Figure 1.1.
    b. What interpretations can you make about the average price per gallon of conventional regular gasoline since June 2009?
    c. Does the time series continue to show a summer increase in the average price per gallon? Explain.

Methods

1. The response to a question has three alternatives: A, B, and C. A sample of 120 responses provides 60 A, 24 B, and 36 C. Show the frequency and relative frequency distributions.

Answer: n=120

$$\text{Relative frequency of a class} = \frac{\text{Frequency of the class}}{n}$$

Class	Frequency	Relative Frequncy	Percent Frequency
A	60	0.5	50%
B	24	0.2	20%
C	36	0.3	30%
Total	120	1.0	100%

Frequency=Number of accorance of each categary of data.

$$\text{Relative Frequency} = \frac{\text{Number of occurance of one categary of data or frequency of class}}{\text{Total number of occurance of all data or number of observations}}$$

$$\text{Rl Freq of A} = \frac{60}{120} = 0.5$$

$$\text{Rl Freq of B} = \frac{24}{120} = 0.2$$

$$\text{Rl Freq of C} = \frac{36}{120} = 0.3$$

$$\text{Percent Frequency} = \frac{\text{Number of occurance of one category of data or frequency of class}}{\text{Total number of occurance of all data or number of observations}} \times 100$$

$$\text{Percent Freq of A} = \frac{60}{120} \times 100 = 50\%$$

$$\text{Percent Freq of B} = \frac{24}{120} \times 100 = 20\%$$

$$\text{Percent Freq of C} = \frac{36}{120} \times 100 = 30\%$$

2. A partial relative frequency distribution is given.

Class	Relative Frequency
A	.22
B	.18
C	.40
D	

a. What is the relative frequency of class D?
b. The total sample size is 200. What is the frequency of class D?
c. Show the frequency distribution.
d. Show the percent frequency distribution.

Class	Relative Frequency	Frequency	Percent Frequency
A	.22	44	22%
B	.18	36	18%
C	.40	80	40%
D	.20	40	20%
Total	1.0	200	100%

n=200

a) Total Relative Frequency is always = 1.0

∴ .22+.18+.40+D=1.0

∴ 0.8+D=1.0

∴ Relative freqeuncy of D=1.0-0.8=0.2

b)n=200

Relative frequency of D =0.2

$$\text{Relative freqncy of D} = \frac{\text{Frequency of D}}{\text{Total numbero fobservations}}$$

$$\therefore \quad 0.2 = \frac{\text{Frequency of D}}{200}$$

∴ Frequency of of D = 0.2 X 200 = 40

C) Frequency of A = Relative Freq of A X n= .22 X 200 = 44
Frequency of B = Relative Freq of B X n = .18 X 200 = 36
Frequency of C = Relative Freq of C X n = .40 X 200 = 80

D) Percent Freq of A = Relative Freq of A X 100= .22 X 100 = 22%
Percent Freq of B = Relative Freq of B X 100 = .18 X 100 = 18%
Percent Freq of C = Relative Freq of C X 100 = .40 X 100 = 40%
Percent Freq of D = Relatvie Freq of D X 100 = .20 X 100 = 20%

3. A questionnaire provides 58 Yes, 42 No, and 20 no-opinion answers.
   a. In the construction of a pie chart, how many degrees would be in the section of the pie showing the Yes answers?
   b. How many degrees would be in the section of the pie showing the No answers?
   c. Construct a pie chart.
   d. Construct a bar chart.

Class	Frequency	Relative Frequency
Yes	58	.4833
No	42	.35
No-Opinion	20	.167
Total	120	1.0

$$\text{Relative Freqyecy of Yes} = \frac{\text{Frequency of Yes}}{\text{Total number of observations}} = \frac{58}{120} = 0.4833$$

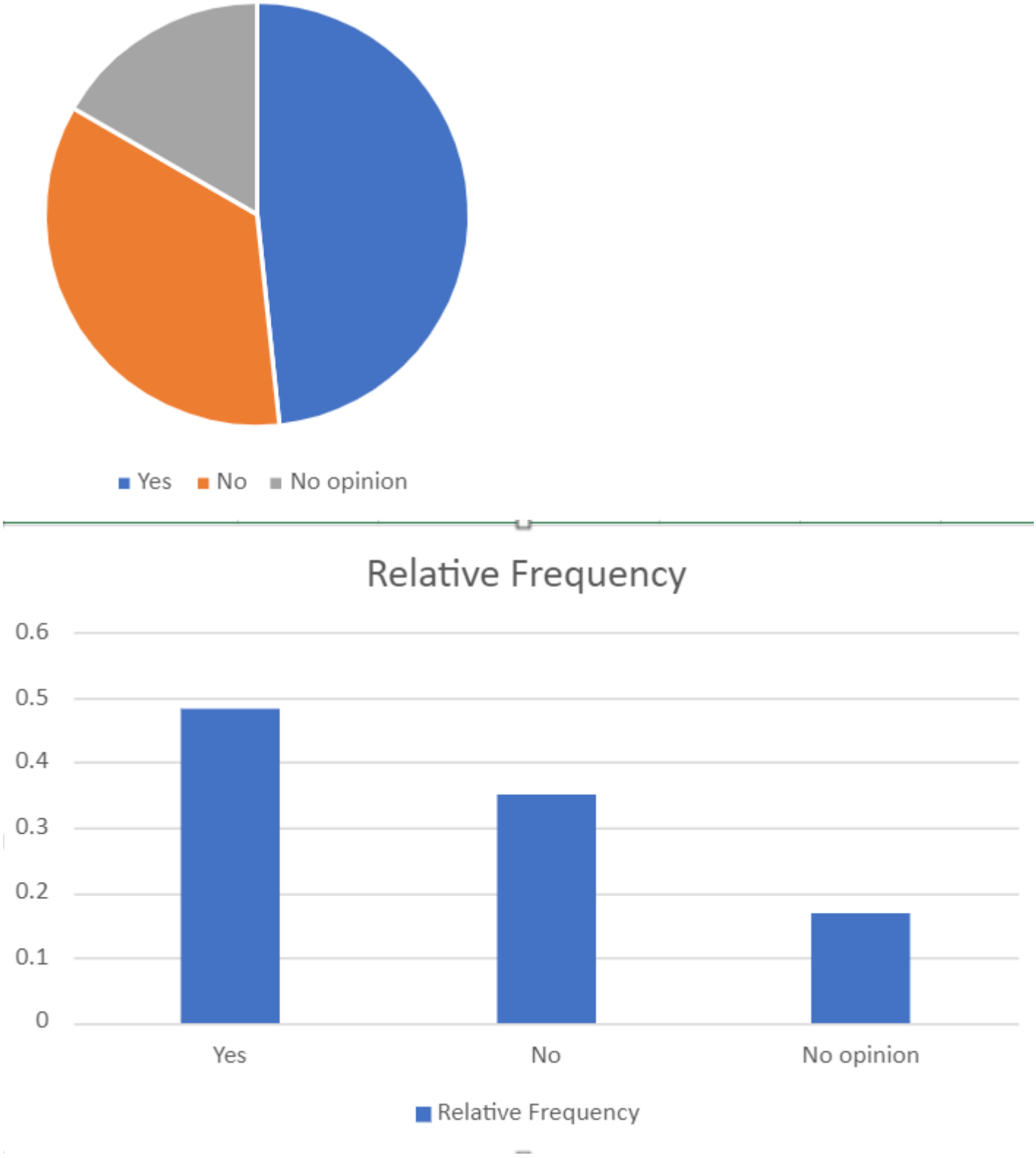
$$\text{Relative Freqyecy of Yes} = \frac{\text{Frequency of No}}{\text{Total number of observations}} = \frac{42}{120} = 0.35$$

$$\text{Relative Freqyecy of No Opinion} = \frac{\text{Frequency of Yes}}{\text{Total number of observations}} = \frac{20}{120} = 0.167$$

Pie Sector of Yes = Relative Frequency of Yes X 360 = .4833 X 360 = 173.98 = 174 degree

Pie Sector of No = Relative Frequency of No X 360 = .35 X 360 = 126 degree

Pie Sector of No opinion = Relative Frequency of No Opinion X 360 = 0.167 X 360 = 60.12 = 60 Degree



Applications

4. The top four prime-time television shows were *Law & Order, CSI, Without a Trace,* and *Desperate Housewives* (Nielsen Media Research, January 1, 2007). Data indicating the preferred shows for a sample of 50 viewers follow.

DH	CSI	DH	CSI	L&O
Trace	CSI	L&O	Trace	CSI
CSI	DH	Trace	CSI	DH
L&O	L&O	L&O	CSI	DH
CSI	DH	DH	L&O	CSI
DH	Trace	CSI	Trace	DH
DH	CSI	CSI	L&O	CSI
L&O	CSI	Trace	Trace	DH
L&O	CSI	CSI	CSI	DH
CSI	DH	Trace	Trace	L&O

a. Are these data categorical or quantitative?
b. Provide frequency and percent frequency distributions.
c. Construct a bar chart and a pie chart.
d. On the basis of the sample, which television show has the largest viewing audience? Which one is second?

a. These are ctegorical data.

b).

Class	Frequency	Relative Frequency	Percent Frequency
DH	13	.26	26%
L&O	10	.20	20%
Trace	9	.18	18%
CSI	18	.36	36%
Total	50	1.00	100%

$$\text{Relative Freq of DH} = \frac{\text{Frequency of DH}}{\text{Number of observations}} = \frac{13}{50} = .26$$

$$\text{Relative Freq of L\&O} = \frac{\text{Frequency of L\&O}}{\text{Number of observations}} = \frac{10}{50} = .2$$

$$\text{Relative Freq of Trace} = \frac{\text{Frequency of Trace}}{\text{Number of observations}} = \frac{9}{50} = .18$$

$$\text{Relative Freq of CSI} = \frac{\text{Frequency of CSI}}{\text{Number of observations}} = \frac{18}{50} = .36$$

Percent Frequency of DH = Relative Frequency of DH X 100 = .26 X 100 = 26 %
Percent Frequency of L&O = Relative Frequency of L&O X 100 = .2 X 100 = 20 %
Percent Frequency of Trace = Relative Frequency of Trace X 100 = .18 X 100 = 18 %
Percent Frequency of CSI = Relative Frequency of CSI X 100 = .36 X 100 = 36 %

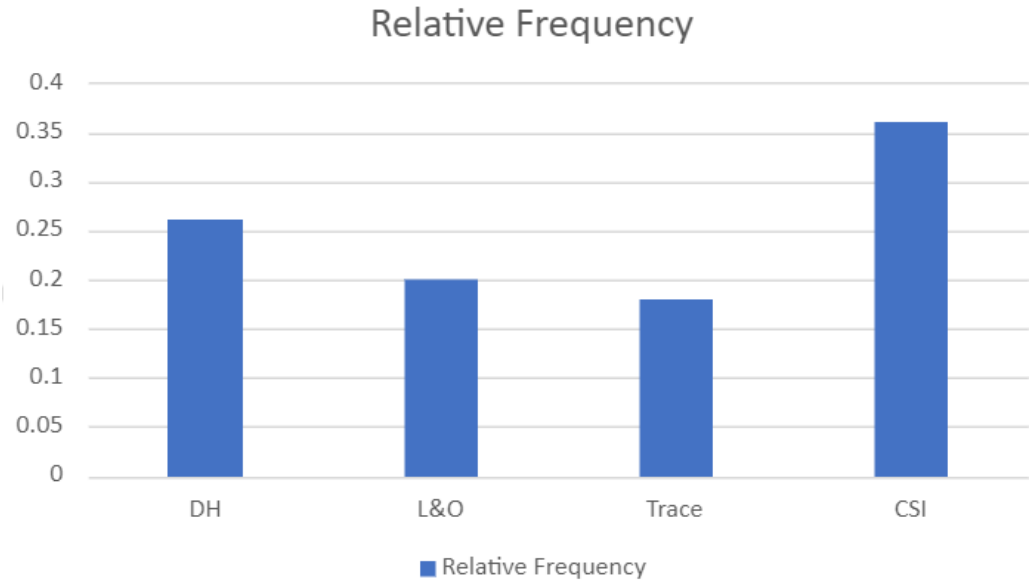
c).

Pie sector of DH=Relative Frequency of DH X 360 = .26 X 360 = 93.6 degreee

Pie Sector of L&O=Relative Frequency of L & O X 360 = .2 X 360 = 72 degree

Pie Sector of Trace =Relative Frequency of Trace X 360 = .18 X 360 = 64.8

Pie Sector of CSI =Relative Frequency of CSI X 360 = .36 X 360 = 129.6 degree



d). On the basis of given sample the CSI teliveison show has largest viewing audiance, and DH Telivision show has second largest viewing audiance.

5. In alphabetical order, the six most common last names in the United States are Brown, Davis, Johnson, Jones, Smith, and Williams (*The World Almanac,* 2006). Assume that a sample of 50 individuals with one of these last names provided the following data.

Brown	Williams	Williams	Williams	Brown
Smith	Jones	Smith	Johnson	Smith
Davis	Smith	Brown	Williams	Johnson
Johnson	Smith	Smith	Johnson	Brown
Williams	Davis	Johnson	Williams	Johnson
Williams	Johnson	Jones	Smith	Brown
Johnson	Smith	Smith	Brown	Jones
Jones	Jones	Smith	Smith	Davis
Davis	Jones	Williams	Davis	Smith
Jones	Johnson	Brown	Johnson	Davis

Summarize the data by constructing the following:

a. Relative and percent frequency distributions

b. A bar chart

c. A pie chart

d. Based on these data, what are the three most common last names?

a)

Class	Frequency	Relative Frequency	Percent Frequency
Brown	7	.14	14%
Smith	12	.24	24%
Davis	6	.12	12 %
Johnson	10	. 20	20%
Williams	8	.16	16%
Jones	7	.14	14%
Total	50	1.00	100%

Number of Observations n = 50

$$\text{Relative Frequency of Brown} = \frac{\text{Frequency of Brown}}{n} = \frac{7}{50} = .14$$

$$\text{Relative Frequency of Smith} = \frac{\text{Frequency of Smith}}{n} = \frac{12}{50} = .24$$

$$\text{Relative Frequency of Davis} = \frac{\text{Frequency of Davis}}{n} = \frac{6}{50} = .12$$

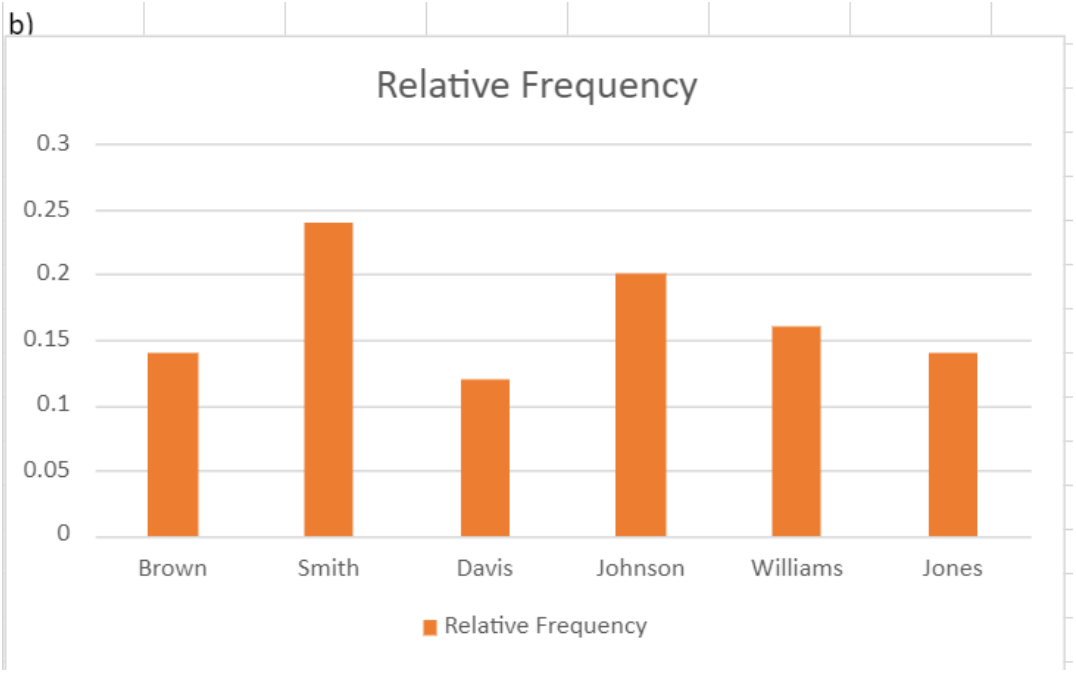
$$\text{Relative Frequency of Johnson} = \frac{\text{Frequency of Johnson}}{n} = \frac{10}{50} = .2$$

$$\text{Relative Frequency of Williams} = \frac{\text{Frequency of Williams}}{n} = \frac{8}{50} = .16$$

$$\text{Relative Frequency of Jones} = \frac{\text{Frequency of Jones}}{n} = \frac{7}{50} = .14$$

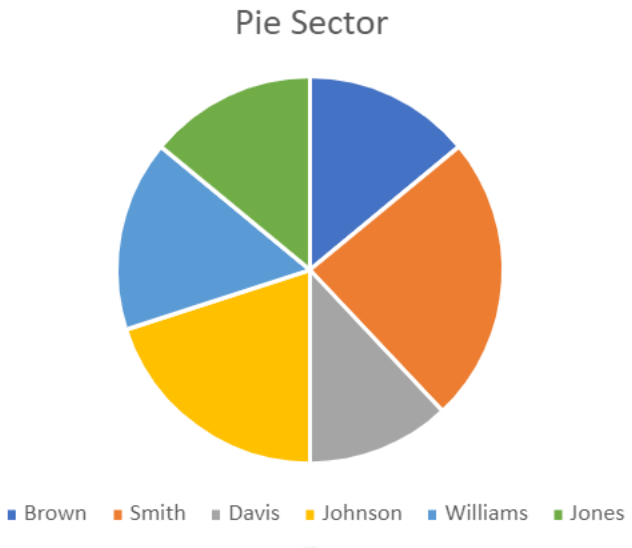
Percent Frequency of Brown = Relative Frequency of Brwon X 100 = .14 X 100 = 14%
Percent Frequency of Smith = Relative Frequency of Smith X 100 = .24 X 100 = 24 %
Percent Frequency of Davis = Relative Frequency of Davis X 100 = .12 X 100 = 12 %
Percent Frequency of Johnson = Relative Frequencyof Johnson X 100 = .2 X 100 = 20%
Percent Frequency of Williams = Relative Frequncy of Williams X 100 = .16 X 100 = 16%
Percent Frequency of Jones = Relative Frequency of Jones X 100 = .14 X 100 = 14 %

b)



c)

Pie sector of Brown = Relative Frequency of Brown X 360 = .14 X 360 = 50.4 degree
Pie sector of Smith = Relative Frequency of Smith X 360 = .24 X 360 = 86.4 degree
Pie Sector of Davis = Relative Frequency of Davis X 360 = .12 X 360 = 43.2 degree
Pie sector of Johnson = Relative Frequency of Johnsons X 360 = .2 X 360 = 72 degree
Pie Sector of Williams = Relative Frequency of Williams X 360 = .16 X 360 = 57.6 degree
Pie Sector of Jones = Relative Frequency of Jones X 360 = .14 X 360 = 50.4 degree



d) Based on the data Smith, Johnson and Williams are the most common last names.

6. The Nielsen Media Research television rating measures the percentage of television owners who are watching a particular television program. The highest-rated television program in television history was the *M*A*S*H Last Episode Special* shown on February 28, 1983. A 60.2 rating indicated that 60.2% of all television owners were watching this program. Nielsen Media Research provided the list of the 50 top-rated single shows in television history (*The New York Times Almanac,* 2006). The following data show the television network that produced each of these 50 top-rated shows.

ABC	ABC	ABC	NBC	CBS
ABC	CBS	ABC	ABC	NBC
NBC	NBC	CBS	ABC	NBC
CBS	ABC	CBS	NBC	ABC
CBS	NBC	NBC	CBS	NBC
CBS	CBS	CBS	NBC	NBC
FOX	CBS	CBS	ABC	NBC
ABC	ABC	CBS	NBC	NBC
NBC	CBS	NBC	CBS	CBS
ABC	CBS	ABC	NBC	ABC

a. Construct a frequency distribution, percent frequency distribution, and bar chart for the data.

b. Which network or networks have done the best in terms of presenting top-rated television shows? Compare the performance of ABC, CBS, and NBC.

a)

Class	Frequency	Relative Frequency	Percent Frequency
ABC	15	0.3	30%
NBC	17	0.34	34%
CBS	17	0.34	34%
FOX	1	0.02	2%
Total	50	1.00	100%

Number of observations n= 50

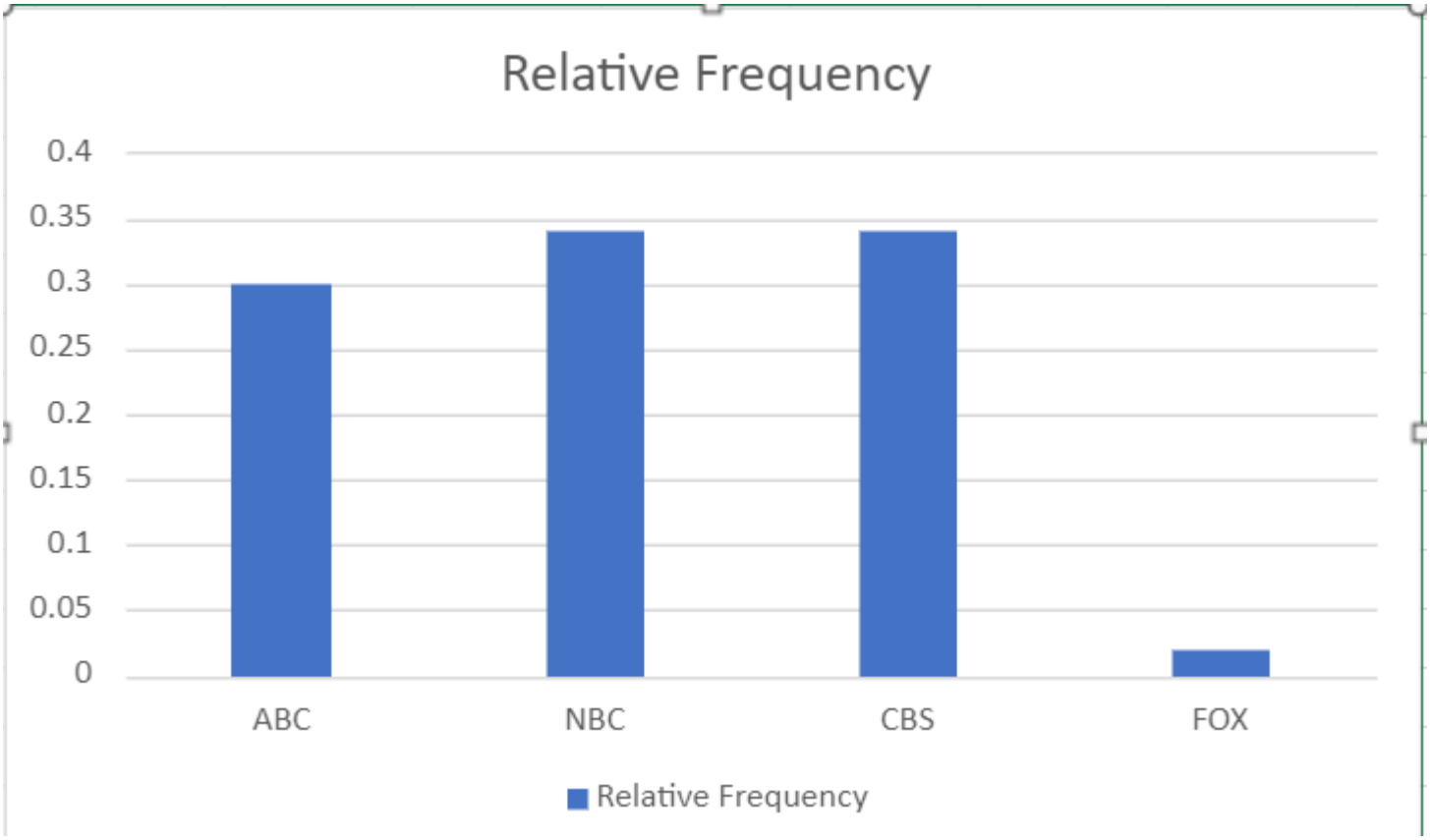
$$\text{Relative Frequency of ABC} = \frac{\text{Frequency of ABC}}{n} = \frac{15}{50} = 0.3$$

$$\text{Relative Frequency of NBC} = \frac{\text{Frequency of NBC}}{n} = \frac{17}{50} = 0.34$$

$$\text{Relative Frequency of CBS} = \frac{\text{Frequency of CBS}}{n} = \frac{17}{50} = 0.34$$

$$\text{Relative Frequency of FOX} = \frac{\text{Frequency of FOX}}{n} = \frac{1}{50} = 0.02$$

Percent Frequency of ABC = Relative frequency of ABC X 100 = .3 X 100 = 30%
Percent Frequency of NBC = Relative Frequency of NBC X 100 = .34 X 100 = 34%
Percent Frequency of CBS = Relative Frequency of CBS X 100 = .34 X 100 = 34%
Percent Frequency of FOX = Relative Frequency of FOX X 100 = 0.02 X 100 = 2%



b) Based on the data NBC and CBS networks have done best in presenting the top rated television show. NBC and CBC networks have equaly presented the top rated television show where as ABC presented slightly less than NBC and CBS networks.

7. Leverock's Waterfront Steakhouse in Maderia Beach, Florida, uses a questionnaire to ask customers how they rate the server, food quality, cocktails, prices, and atmosphere at the restaurant. Each characteristic is rated on a scale of outstanding (O), very good (V), good (G), average (A), and poor (P). Use descriptive statistics to summarize the following data collected on food quality. What is your feeling about the food quality ratings at the restaurant?

G	O	V	G	A	O	V	O	V	G	O	V	A
V	O	P	V	O	G	A	O	O	O	G	O	V
V	A	G	O	V	P	V	O	O	G	O	O	V
O	G	A	O	V	O	O	G	V	A	G		

a) Descriptive statistics summarisation of the given categorical data is as below

Class	Frequency	Relative Frequency	Percent Frequency
O	19	.38	38%
V	13	.26	26%
G	10	.2	20%
A	6	.12	12%
P	2	.04	4%
Total	50	1.00	100%

Number of observations n =50

$$\text{Relative Frequency of O} = \frac{\text{Frequency of O}}{n} = \frac{19}{50} = .38$$

$$\text{Relative Frequency of V} = \frac{\text{Frequency of V}}{n} = \frac{13}{50} = .26$$

$$\text{Relative Frequency of G} = \frac{\text{Frequency of G}}{n} = \frac{10}{50} = .2$$

$$\text{Relative Frequency of A} = \frac{\text{Frequency of A}}{n} = \frac{6}{50} = .12$$

$$\text{Relative Frequency of P} = \frac{\text{Frequency of P}}{n} = \frac{2}{50} = .04$$

Percent Frequency of O =Relative Frequency of O X 100 = .38 X 100 = 38%
Percent Frequency of V =Relative Frequency of V X 100 = .26 X 100 = 26%
Percent Frequency of G =Relative Frequency of G X 100 = .20 X 100n = 20%
Percent Frequency of A =Relative Frequency of A X 100 = .12 X 100 =12 %
Percent Frequency of P =Relative Frequency of P X 100 = .04 X 100 = 4%

b) Based on the data the food quality at the restorent is Very good and above as the 64% of the survey makes Very good and above. And 84% of surve says it is good and above. And only 16% survey says it is average and below.

8. Data for a sample of 55 members of the Baseball Hall of Fame in Cooperstown, New York, are shown here. Each observation indicates the primary position played by the Hall of Famers: pitcher (P), catcher (H), 1st base (1), 2nd base (2), 3rd base (3), shortstop (S), left field (L), center field (C), and right field (R).

L	P	C	H	2	P	R	1	S	S	1	L	P	R	P
P	P	P	R	C	S	L	R	P	C	C	P	P	R	P
2	3	P	H	L	P	1	C	P	P	P	S	1	L	R
R	1	2	H	S	3	H	2	L	P					

a. Use frequency and relative frequency distributions to summarize the data.
b. What position provides the most Hall of Famers?
c. What position provides the fewest Hall of Famers?
d. What outfield position (L, C, or R) provides the most Hall of Famers?
e. Compare infielders (1, 2, 3, and S) to outfielders (L, C, and R).

a)

Class	Frequency	Relative Frequency	Percent Frequency
P	17	.30	30%
H	4	.07	7%
1	5	.09	9%
2	4	.07	7%
3	2	.03	3%
S	5	.09	9%
L	6	.10	10%
C	5	.09	9%
R	7	.12	12%
Total	55	1.00	100%

Number of observations n = 55

Relative Freq of P = $\frac{17}{55}$ = .30

Relative Freq of H = $\frac{4}{55}$ = .07

Relative Freq of 1 = $\frac{55}{55}$ = .09

Relative Freq of 2 = $\frac{4}{55}$ = .07

Relative Freq of 3 = $\frac{2}{55}$ = .03

Relative Freq of S = $\frac{5}{55}$ = .09

Relative Freq of L = $\frac{6}{55}$ = .10

Relative Freq of C = $\frac{5}{55}$ = .09

Relative Freq of R = $\frac{7}{55}$ = .12

b) P poistion provieds the the most Hall of Famers

c) 2 poistion provides the fewest Hall of Famers

d) R poistion provides the most Hall of Famers among the out fields L, C, R.

e) Infielders(1,2,3 and S) provided 28% of the Hall of Famers where as 31% of Hall of famers provided by outfielders (L, C and R)

Methods

11. Consider the following data.

14	21	23	21	16
19	22	25	16	16
24	24	25	19	16
19	18	19	21	12
16	17	18	23	25
20	23	16	20	19
24	26	15	22	24
20	22	24	22	20

a. Develop a frequency distribution using classes of 12–14, 15–17, 18–20, 21–23, and 24–26.

b. Develop a relative frequency distribution and a percent frequency distribution using the classes in part (a).

Class	Frequency	Relative Frequency	Percent frequency
12-14	2	.05	5%
15-17	8	.2	20%
18-20	11	.275	27.5%
21-23	10	.25	25%
24-26	9	.225	22.5%
Total	40	1.00	100%

Total number of occurrences = 40

a) Frequency - Count the number of occrrences fall within the given class.

b)

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total number of occurrences}}$$

Percent Frequency = 100 X Relative Frequency

RF1= 2/40 = .05
RF2 = 8/40=.2
Rf3= 11/40=.275
RF4= 10/40 = .25
RF5 = 9/40 =.225

PF1 = RF1 X 100 = .05X100= 5%

.... PF5 = RF5 X 100 = .225 X100= 22.5%

12. Consider the following frequency distribution.

Class	Frequency
10–19	10
20–29	14
30–39	17
40–49	7
50–59	2

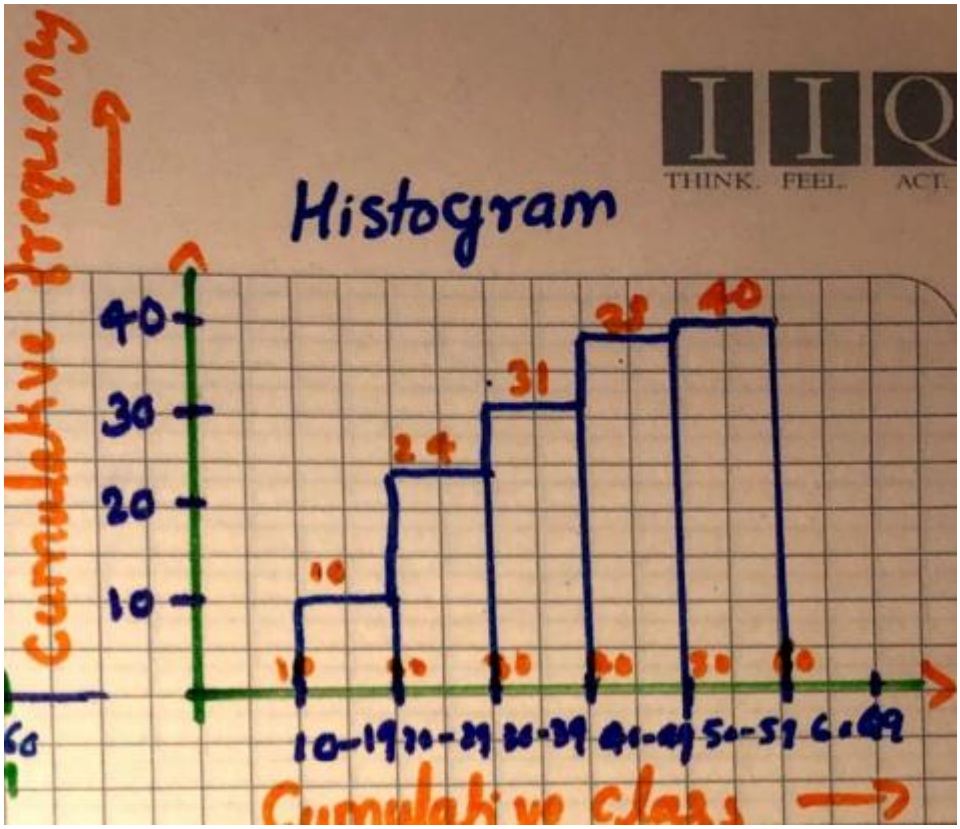
Construct a cumulative frequency distribution and a cumulative relative frequency distribution.

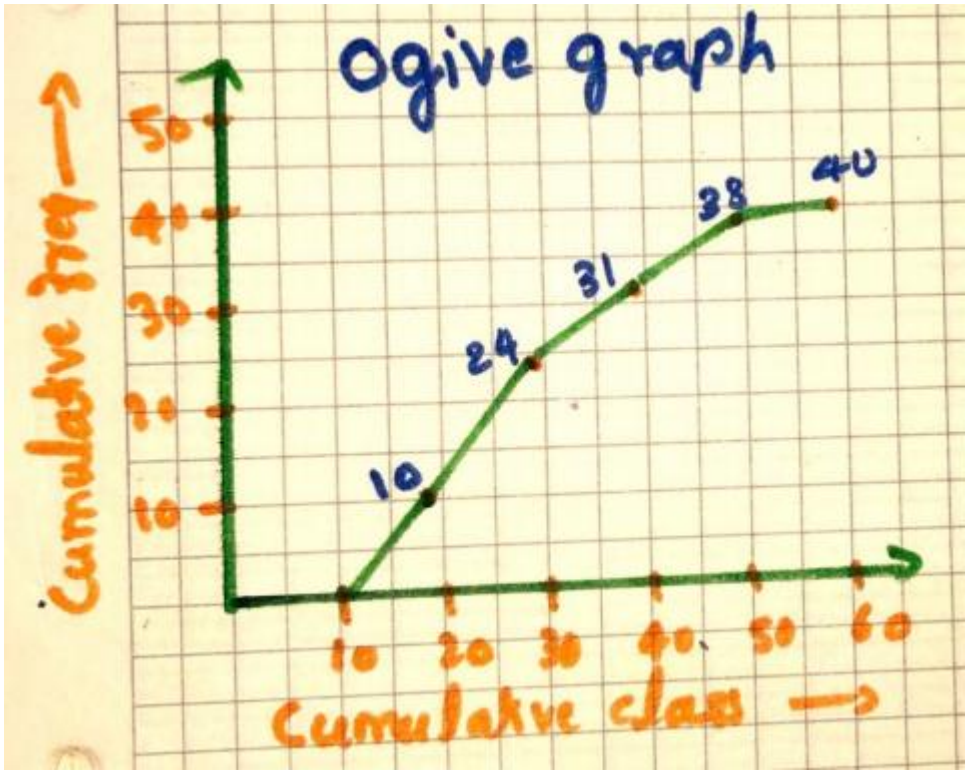
Cumulative Class	Cumulative Frequency	Cumulative Relative Frequency
less than or equal to 19	10	.25
Less than or equal to 29	24	.60
Less than or equal to 39	31	.775
Less than or equal to 49	38	.95
Less than or equal to 59	40	1.00

a)
Total number of occurrences = 40
Cumulative class = less than or equal to upper limit of Frequency distribution class
Cumulative frequency = Acumulated frequency at each cumulative class
CF1 = F1 = 10
CF2= F1+F2 = 10+14 = 24
CF3= F1+F2+F3= 10+14+17 = 31
CF4= F1+F2+F3+F4 = 10+14+17+7 = 38
CF5= F1+F2+F3+F4+F5=10+14+17+7+2=40

Cumulative Relative Frequency (CRF) = CF/Total number of occrrences
CRF1 = CF1/40= 10/40= .25
CRF2= CF2/40=24/40=.60
CRF3= CF3/40= 31/40=.775
CRF4= CF4/40=38/40=.95
CRF5= CF5/40= 40/40 =1.00

13. Construct a histogram and an ogive for the data in exercise 12.

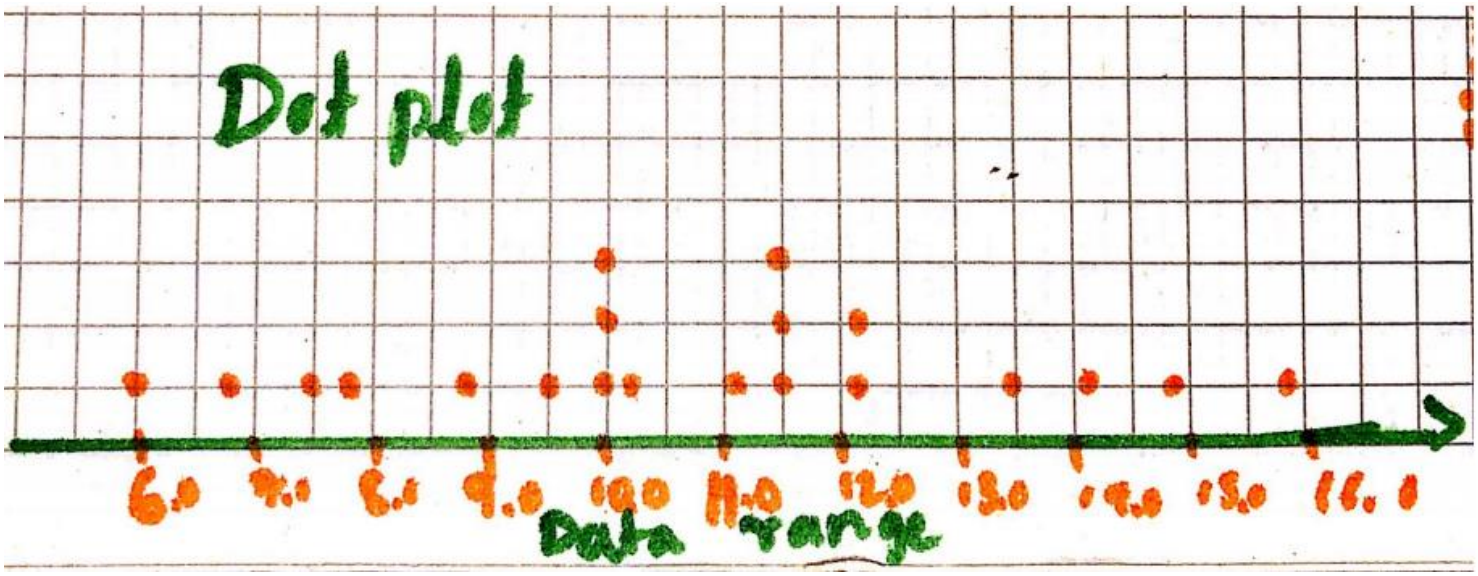

Histogram
THINK. FEEL. ACT.
Cumulative frequency
40
30
20
10
10
24
31
38
40
10-19
20-29
30-39
40-49
50-59
Cumulative class


Ogive graph
Cumulative freq
50
40
30
10
10
24
31
38
40
10
20
30
40
50
60
Cumulative class

14. Consider the following data.

8.9	10.2	11.5	7.8	10.0	12.2	13.5	14.1	10.0	12.2
6.8	9.5	11.5	11.2	14.9	7.5	10.0	6.0	15.8	11.5

a. Construct a dot plot.
b. Construct a frequency distribution.
c. Construct a percent frequency distribution.



Total number of occurrence = 20
Number of class we would like to have =5
min data = 6.0
max data=15.8
class width = max data - min data / No of class = 15.8-6.0 / 5 = 1.96 rounded to 2.0

class	Frequency	Relative frequency	percent frequency
6.0-7.9	4	.20	20%
8.0-9.9	2	.10	10 %
10.0-11.9	8	.40	40 %
12.0-13.9	3	.15	15%
14.0-15.9	3	.15	15%
Total	20	1.00	100%

Total number of occurrence =20

b)

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Total No of occurrence}}$$

c)

Percent Frequency = Relative Frequency X 100

RF1 = F1/20 =4/20=.20
RF2= F2/20 = 2/20=.10
RF3= F3/20 = 8/20 = .40
RF4= F4/20 = 3/20 = .15
RF5 = F5/20 = 3/20 = .15

PF1= RF1 X 100 = .20 X100 = 20%
PF2=RF2X100 = .10 X 100 = 10%
PF3= RF3 X 100 = .40 X 100 = 40 %
PF4 =RF4 X 100 = .15 X 100 = 15 %
PF5= RF5 X 100 = .15 X 100 = 15%

Applications

15. A doctor's office staff studied the waiting times for patients who arrive at the office with a request for emergency service. The following data with waiting times in minutes were collected over a one-month period.

2 5 10 12 4 4 5 17 11 8 9 8 12 21 6 8 7 13 18 3

Use classes of 0–4, 5–9, and so on in the following:

a. Show the frequency distribution.
b. Show the relative frequency distribution.
c. Show the cumulative frequency distribution.
d. Show the cumulative relative frequency distribution.
e. What proportion of patients needing emergency service wait 9 minutes or less?

a)

Class	Frequency	Relative Frequency
0-4	4	.20
5-9	8	.40
10-14	5	.25
15-19	2	.10
20-24	1	.05
Total	20	1.00

b)

Total number of occurrences n = 20

$$\text{Relative Frequency} = \frac{\text{Frequency}}{n}$$

RF1= F1/20= 4/20=.20
RF2=F2/20=8/20 =.40
RF3=F3/20=5/20=.25
RF4=F4/20=2/20=.10
RF5=F5/20=1/20=.05

c)

Cumulative class	Cumulative Frequency	Cumulative Relative Frequency
Less than or equal to 4	4	.20
Less than or equal to 9	12	.60
Less than or equal to 14	17	.85
Less than or equal to 19	19	.95
Less than or equal to 24	20	1.00

Cumulative Frequency CF1= F1 =4
CF2= F1+F2=4+8=12
CF3= F1+F2+F3=4+8+5=17
CF4=F1+F2+F3+F4=4+8+5+2=19
CF5=F1+F2+F3+F4+F5=4+8+5+2+1=20

d)

Cumulative Relatvie Frequency CRF1= RF1 or CF1/n= 4/20 = .20
CRF2= RF1+RF2 OR CF2/n= 12/20=.60
CRF3=RF1+RF2+RF3 OR CF3/20= 17/20=.85
CRF4= RF1+RF2+RF3+RF4 OR CF4/20=19/20 = .95
CRF5= RF1+RF2+RF3+RF4+RF5 OR CF5/20= 20/20= 1.00

e) .60 X 100 = 60% of patients needing emergency servies wait 9 minutes or less.

16. A shortage of candidates has required school districts to pay higher salaries and offer extras to attract and retain school district superintendents. The following data show the annual base salary ($1000s) for superintendents in 20 districts in the greater Rochester, New York, area (*The Rochester Democrat and Chronicle,* February 10, 2008).

187	184	174	185
175	172	202	197
165	208	215	164
162	172	182	156
172	175	170	183

Use classes of 150–159, 160–169, and so on in the following.

a. Show the frequency distribution.
b. Show the percent frequency distribution.
c. Show the cumulative percent frequency distribution.
d. Develop a histogram for the annual base salary.
e. Do the data appear to be skewed? Explain.
f. What percentage of the superintendents make more than $200,000?

a)

Class	Frequency	Relative Frequency	Percent Frequency
150-159	1	.05	5%
160-169	3	.15	15%
170-179	7	.35	35%
180-189	5	.25	25%
190-199	1	.05	5%
200-209	2	.10	10%
210-219	1	.05	5%
Total	20	1.00	100%

b)

Total number of observations n =20

Relative Frequency = Frequency/n

RF1= F1/n=1/20=.05
RF2= F2/N= 3/20 =.15
RF3= F3/N=7/20=.35
RF4 = F4/n = 5/20=.25
RF5=F5 / n = 1/20 = .05
RF6 = F6/n= 2/20 = .10
RF7= F7/n= 1/20 = .05

Percent Frequency = 100X Relative Frequency

PF1=RF1x100=.05 X100= 5%
PF2=RF2X100= .15 X 100= 15%
PF3 = RF3 X 100 = .35 X100=35%
PF4= RF4 X 100 = .25 X100=25%
PF5= RF5 X 100 = .05X 100 = 5%
PF6 = RF6 X 100 = .10 X 100 = 10%
PF7 = RF7 X 100 = .05 X 100 = 5%

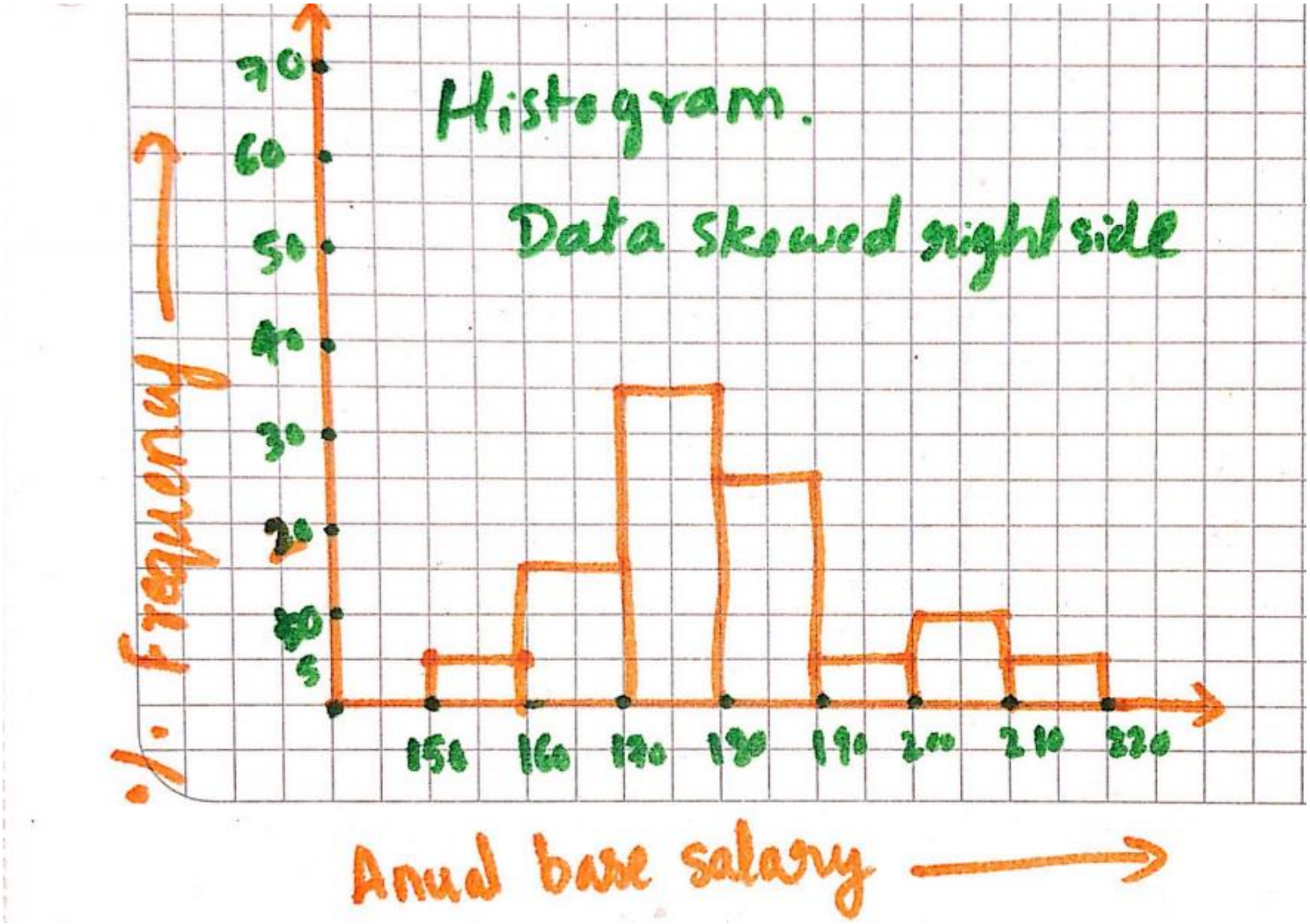
C)

Cumulative Class	Cumulative Frequency	Cumulative Relative Frequency	Cumulative Percent Freque
Less than or equal to 159	1	.05	5%
Less than or equal to 169	4	.20	20%
Less than or equal to 179	11	.55	55%
Less than or equal to 189	16	.80	80%
Less than or equal to 199	17	.85	85%
Less than or equal to 209	19	.95	95%
Less than or equal to 219	20	1.00	100%

Cumulative Frequencny CF1 = F1=1
CF2=F1 + F2 =1+3=4
CF3= F1+ F2+ F3= 1+3+7 = 11
CF4= F1+F2+F3+F4=1+3+7+5 = 16
CF5= F1+F2+F3+F4+F5 = 1+3+7+5+1= 17
CF6= F1+F2+F3+F4+F5+F6=1+3+7+5+1+2= 19
CF7=F1+F2+F3+F4+F5+F6+F7==1+3+7+5+1+2+1=20

Cumulative Relative Frequency CRF1=RF1 OR CF1/N = 1/20=.05
CRF2=RF1+RF2 OR CF2/N= 4/20= .20
CRF3=RF1+RF2+RF3 OR CF3/2N= 11/20=.55
CRF4=RF1+RF2+RF3+RF4 OR CF4/20 = 16/20=.80
CRF5 =RF1+RF2+RF3+RF4+RF5 OR CF5/N = 17/20=.85
CRF6=RF1+RF2+RF3+RF4+RF5+RF6 OR CF6/N= 19/20=.95
CRF7=RF1+RF2+RF3+RF4+RF5+RF6+RF7 OR CF7/N = 20/20= 1.00

Cumulative Percent Frequency CPF1 = CRF1 * 100 = .05 x 100= 5%
CPF2= CRF2 x 100 = .20 x 100 = 20%
CPF3=CRF3 x 100 = .55 X 100 = 55%
CPF4= CRF4 X 100 = .80 X 100 = 80%
CPF5= CRF5 X 100 = .85 X 100 = 85%
CPF6= CRF6 X 100= .95/100 = 95%
CPF7 = CRF7 X 100 = 1 X 100 = 100%



e) Data appear to be skewed at the right side as the only few superintendents have salary above 190,000.

f) % of Superintendents make more than $200,000 = PF6+PF7= 10%+5%= 15%

17. The Dow Jones Industrial Average (DJIA) underwent one of its infrequent reshufflings of companies when General Motors and Citigroup were replaced by Cisco Systems and Travelers (*The Wall Street Journal,* June 8, 2009). At the time, the prices per share for the 30 companies in the DJIA were as follows:

Company	$/Share	Company	$/Share
3M	61	IBM	107
Alcoa	11	Intel	16
American Express	25	J.P. Morgan Chase	35
AT&T	24	Johnson & Johnson	56
Bank of America	12	Kraft Foods	27
Boeing	52	McDonald's	59
Caterpillar	38	Merck	26
Chevron	69	Microsoft	22
Cisco Systems	20	Pfizer	14
Coca-Cola	49	Procter & Gamble	53
DuPont	27	Travelers	43
ExxonMobil	72	United Technologies	56
General Electric	14	Verizon	29
Hewlett-Packard	37	Wal-Mart Stores	51
Home Depot	24	Walt Disney	25

a. What is the highest price per share? What is the lowest price per share?
b. Using a class width of 10, develop a frequency distribution for the data.
c. Prepare a histogram. Interpret the histogram, including a discussion of the general shape of the histogram, the midprice range, and the most frequent price range.
d. Use the *The Wall Street Journal* or another newspaper to find the current price per share for these companies. Prepare a histogram of the data and discuss any changes since June 2009. What company has had the largest increase in the price per share? What company has had the largest decrease in the price per share?

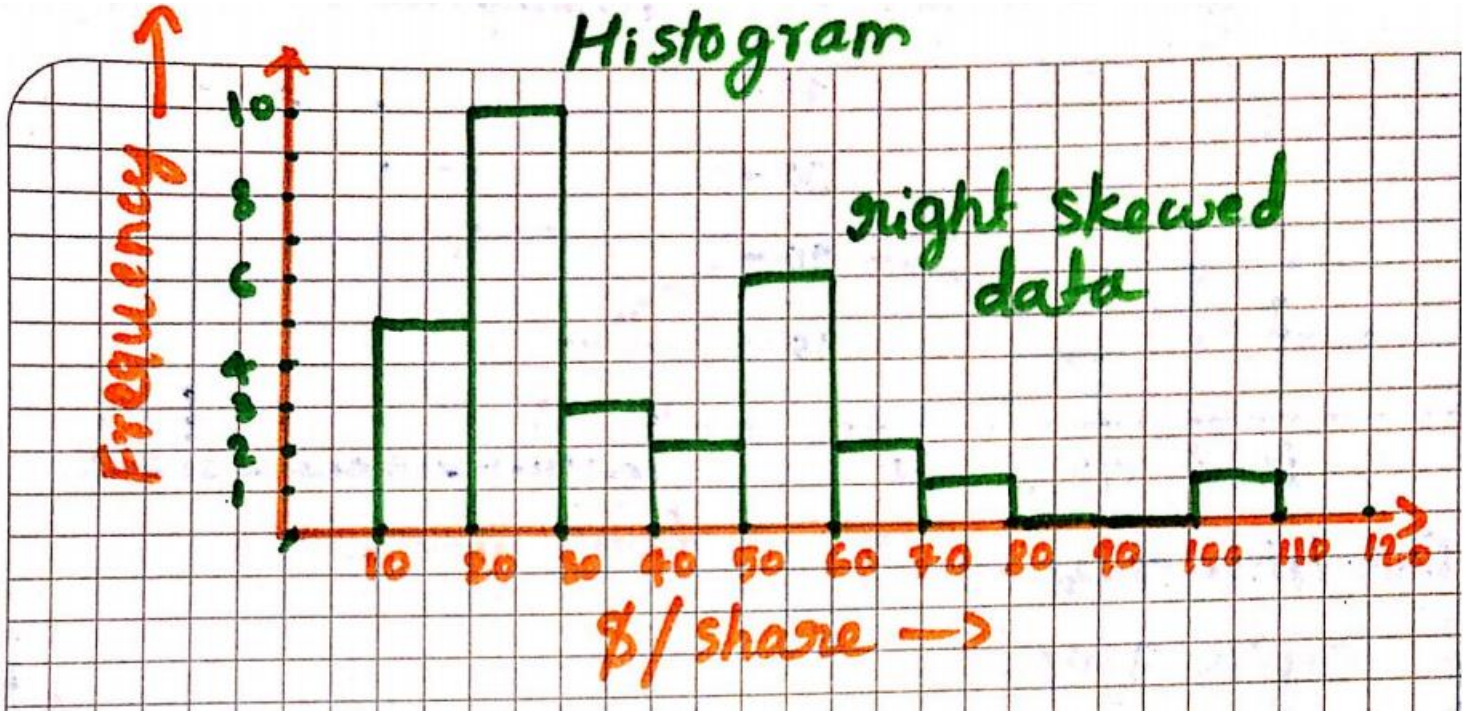
a)
IBM has highest per share $107/share
Alcoa has lowest per share $ 11/Share

b)

Class	Frequency	Relative Frequency	Percent Frequency
10-19	5	.167	16.7
20-29	10	.333	33.3
30-39	3	.10	10
40-49	2	.066	6.6
50-59	6	.20	20
69-69	2	.066	6.6
70-79	1	.033	3.3
80-89	0	.0	0
90-99	0	.0	0
100-109	1	.033	3.3
Total	30	1.00	100

Total number of occurrence n = 30

Relative frequency = Frequency/n Percent Frequency = Relative Frequency X 100



c) Out of 30 companies 10 companies have shares value as between 20 to 29 $/share.
only few companies have shares as above 70 $/share making the data as rightly skewed
15 out of 30 companies i.e. 50% of compalnies have share valueas between 10 to 30 $/share

18. NRF/BIG research provided results of a consumer holiday spending survey (*USA Today*, December 20, 2005). The following data provide the dollar amount of holiday spending for a sample of 25 consumers.

1200	850	740	590	340
450	890	260	610	350
1780	180	850	2050	770
800	1090	510	520	220
1450	280	1120	200	350

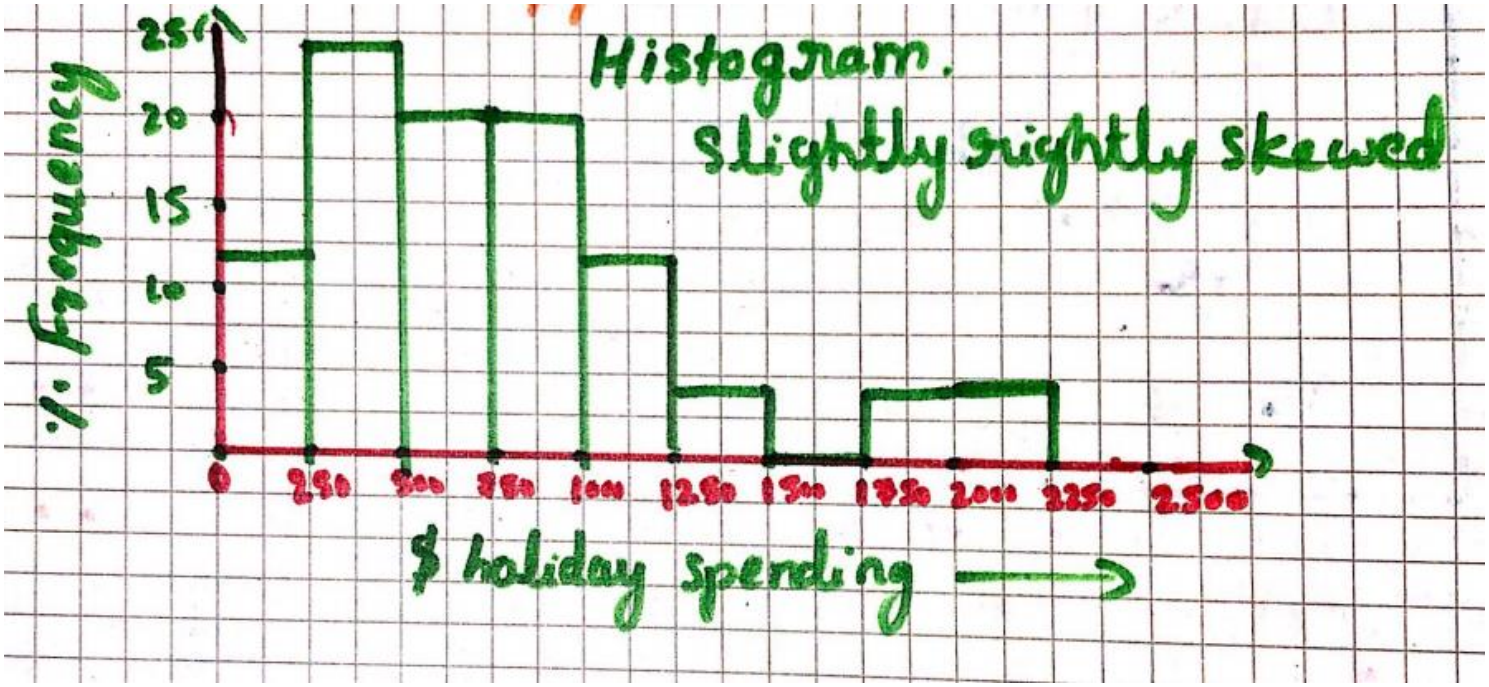
a. What is the lowest holiday spending? The highest?
b. Use a class width of $250 to prepare a frequency distribution and a percent frequency distribution for the data.
c. Prepare a histogram and comment on the shape of the distribution.
d. What observations can you make about holiday spending?

a) Lowest holiday spending is 180
Highest holiday spending is 2050

b)

Class	Frequency	Relative Frequency	Percent Frequency
0-249	3	.12	12%
250- 499	6	.24	24%
500- 749	5	.20	20%
750- 999	5	.20	20%
1000-1249	3	.12	12%
1250-1499	1	.04	4%
1500-1749	0	.00	0%
1750-1999	1	.04	4%
2000-2249	1	.04	4%
Total	25	1.00	100%

Total number of occurrence n= 25
Relative Frequency = Frequency/n
Percent Frequency = 100 X Relative Frequency



c. The shape of the distribution is slightly skewed at the right side.
d. the highest number of consumer 24% are spending amount between 250 to 499 $ on holidays

64 % of consumers are spending amount between $250 to 999 on holidays
only few consumers 12% are spending amount above 1250$ on holidays
88% of consumers are spending amount below 1250 $ on holidays

19. Sorting through unsolicited e-mail and spam affects the productivity of office workers. An InsightExpress survey monitored office workers to determine the unproductive time per day devoted to unsolicited e-mail and spam (*USA Today,* November 13, 2003). The following data show a sample of time in minutes devoted to this task.

2	4	8	4
8	1	2	32
12	1	5	7
5	5	3	4
24	19	4	14

Summarize the data by constructing the following:

a. A frequency distribution (classes 1–5, 6–10, 11–15, 16–20, and so on)
b. A relative frequency distribution
c. A cumulative frequency distribution
d. A cumulative relative frequency distribution
e. An ogive
f. What percentage of office workers spend 5 minutes or less on unsolicited e-mail and spam? What percentage of office workers spend more than 10 minutes a day on this task?

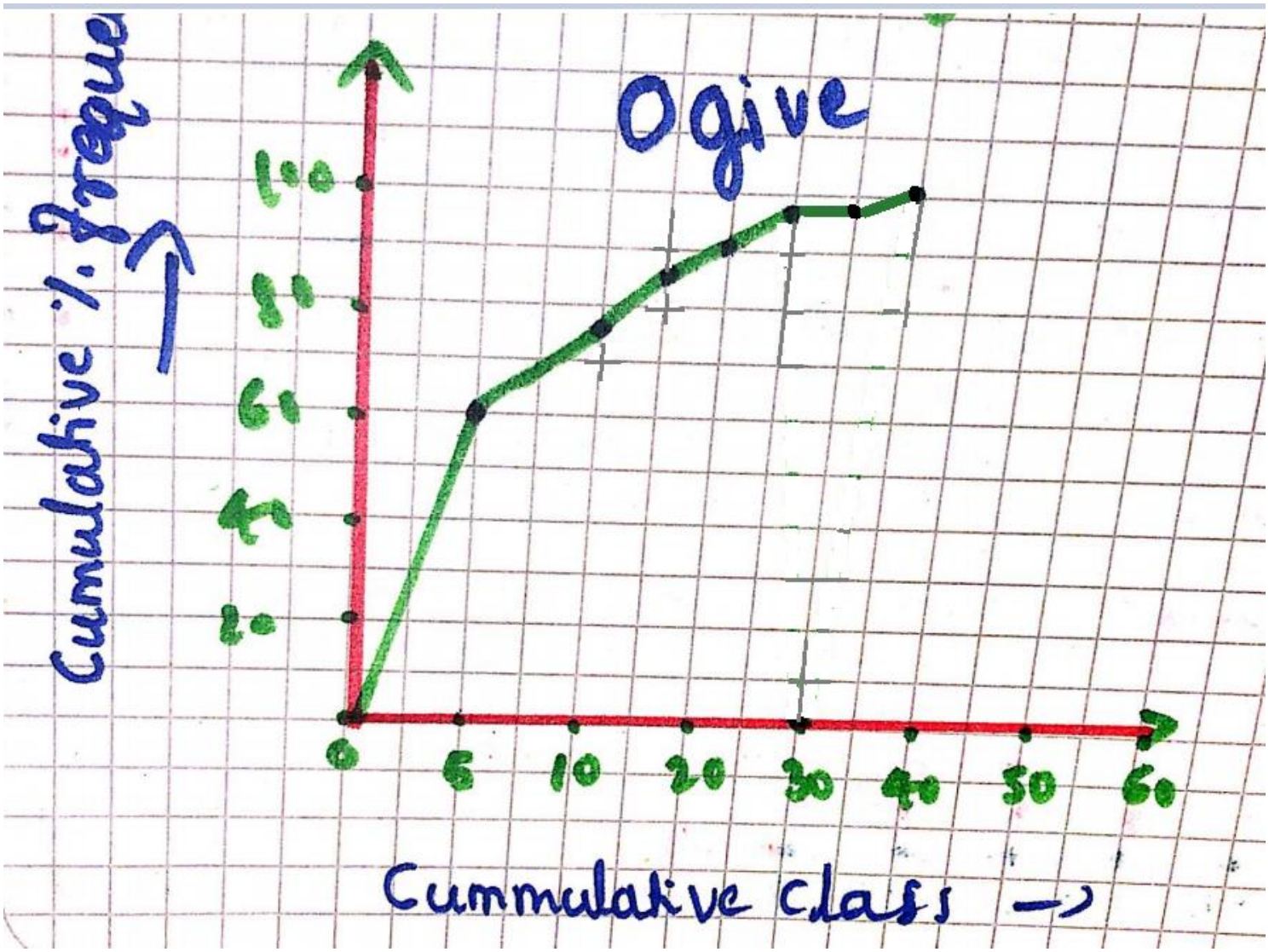
a)

class	Frequency	Relative Frequency	Percent Frequency
1-5	12	.60	60%
6-10	3	.15	15%
11-15	2	.10	10%
16-20	1	.05	5%
21-25	1	.05	5%
26-30	0	0	0%
31-35	1	.05	5%
Total	20	1.00	100%

b)
Total number of occurrences n = 20
Relative Frequency = Frequency/n Percent Frequency = 100 X Relative Frequency

Cumulative Class	Cumulative Frequency	Cumulative Relative Frequency
Less than or equal to 5	12	.60
Less than or equal to 10	15	.75
Less than or equal to 15	17	.85
Less than or equal to 20	18	.90
Less than or equal to 25	19	.95
Less than or equal to 30	19	.95
Less than or equal to 35	20	1.00

c) Cumulative frequncy CF1= F1
CF2=F1+F2
CF3= F1+F2+F3
.....
.....
d) Cumulative Relative frequency CRF= CF/n

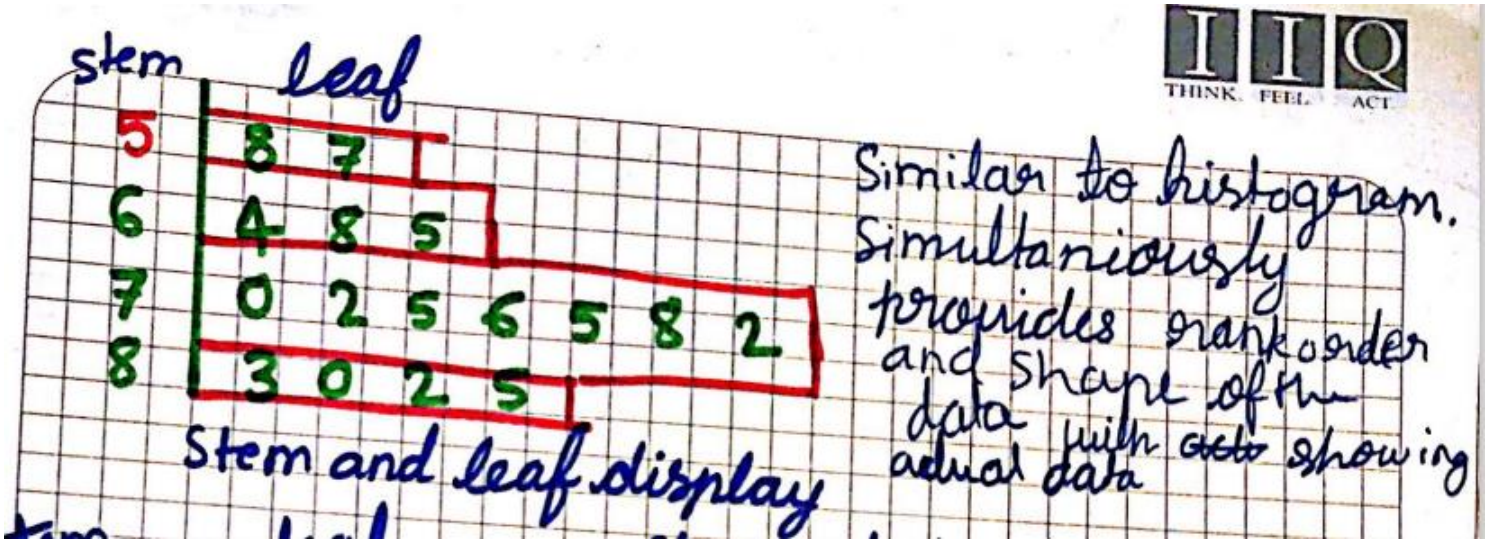


f) CRF1=60% of office workers spend 5 minutes or less on unsolicited email and spam
PF3+PF4+PF5+PF6+PF7 = 25% of office workers spend more than 10 minutes a day on this task.

Methods

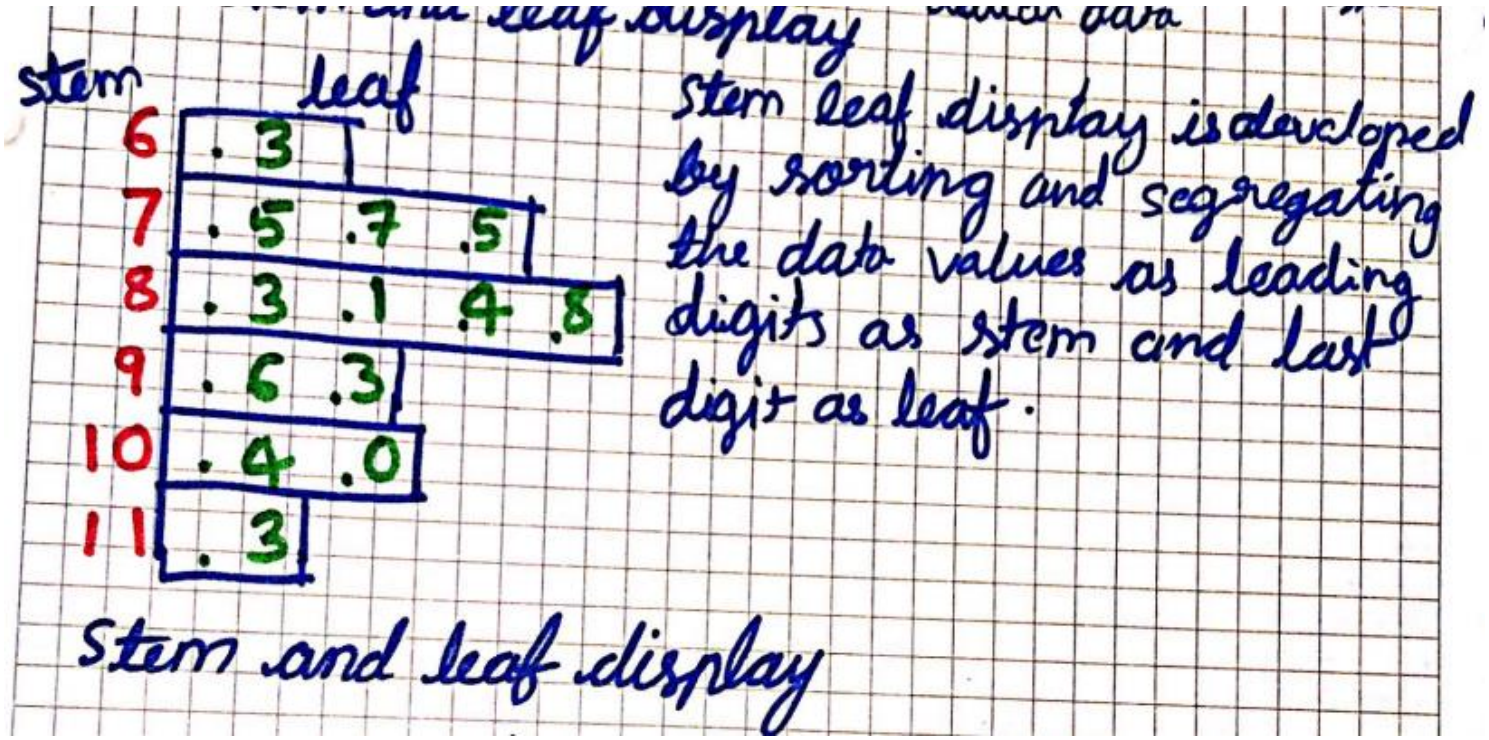
22. Construct a stem-and-leaf display for the following data.

70	72	75	64	58	83	80	82
76	75	68	65	57	78	85	72



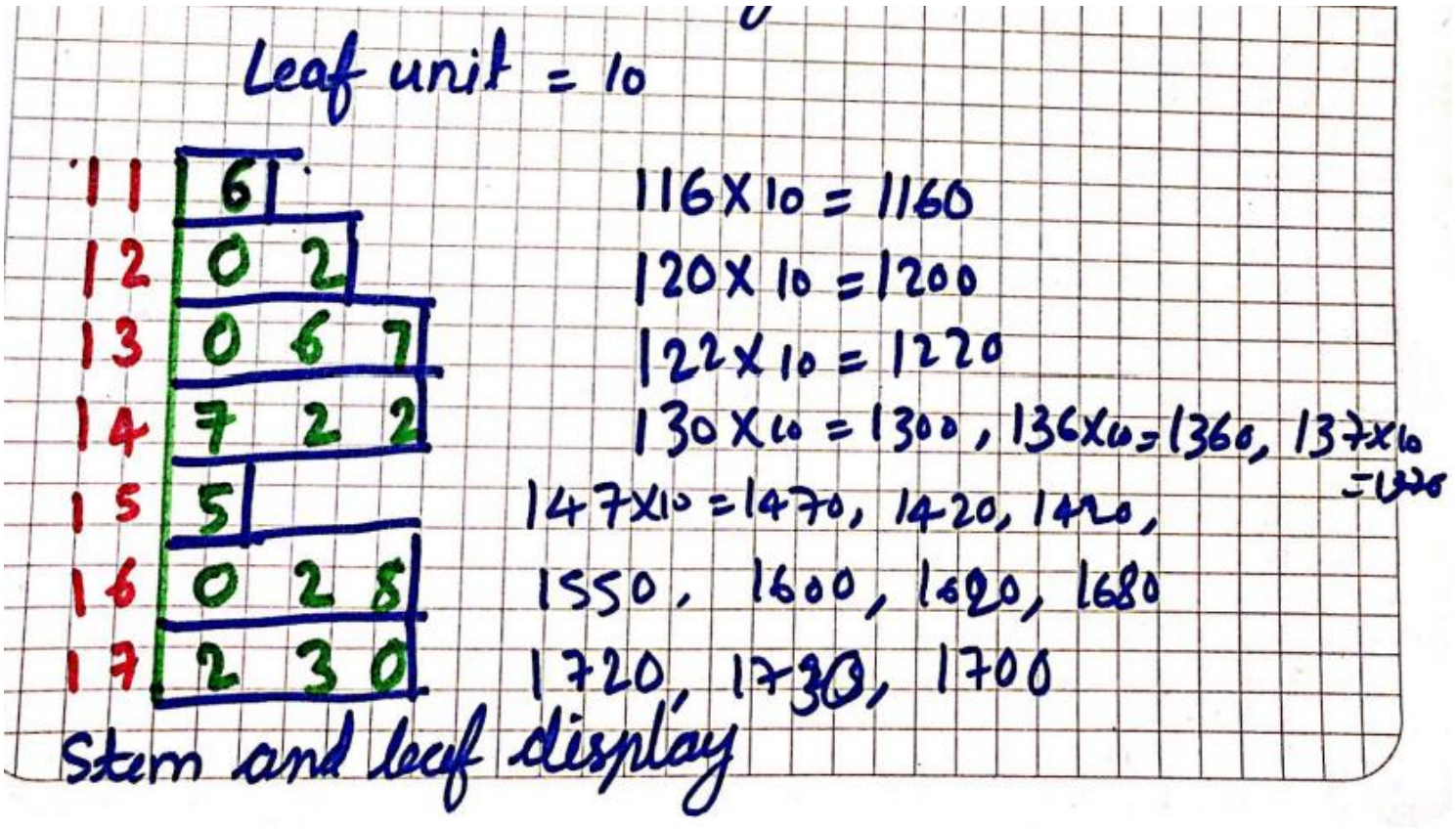
23. Construct a stem-and-leaf display for the following data.

11.3	9.6	10.4	7.5	8.3	10.5	10.0
9.3	8.1	7.7	7.5	8.4	6.3	8.8



24. Construct a stem-and-leaf display for the following data. Use a leaf unit of 10.

1161	1206	1478	1300	1604	1725	1361	1422
1221	1378	1623	1426	1557	1730	1706	1689

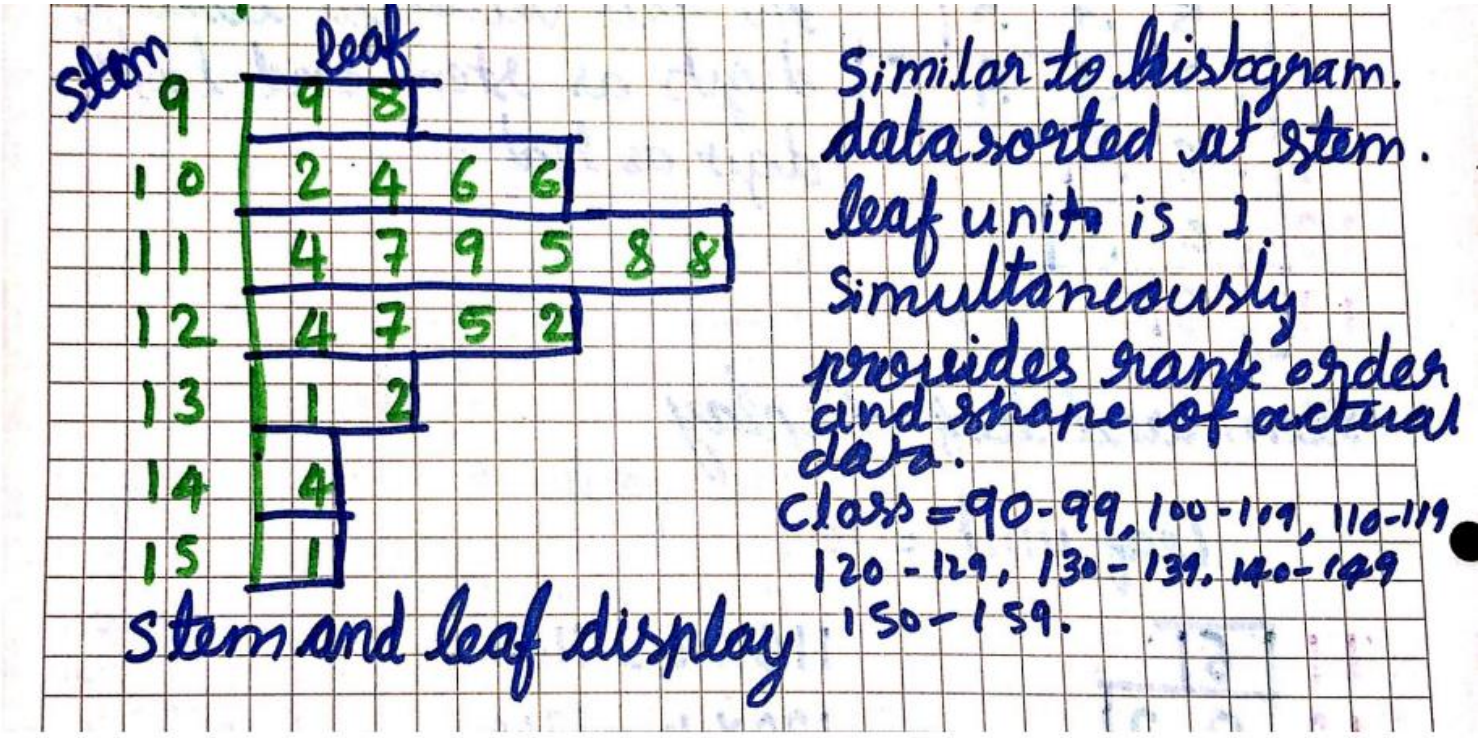


Applications

25. A psychologist developed a new test of adult intelligence. The test was administered to 20 individuals, and the following data were obtained.

114	99	131	124	117	102	106	127	119	115
98	104	144	151	132	106	125	122	118	118

Construct a stem-and-leaf display for the data.

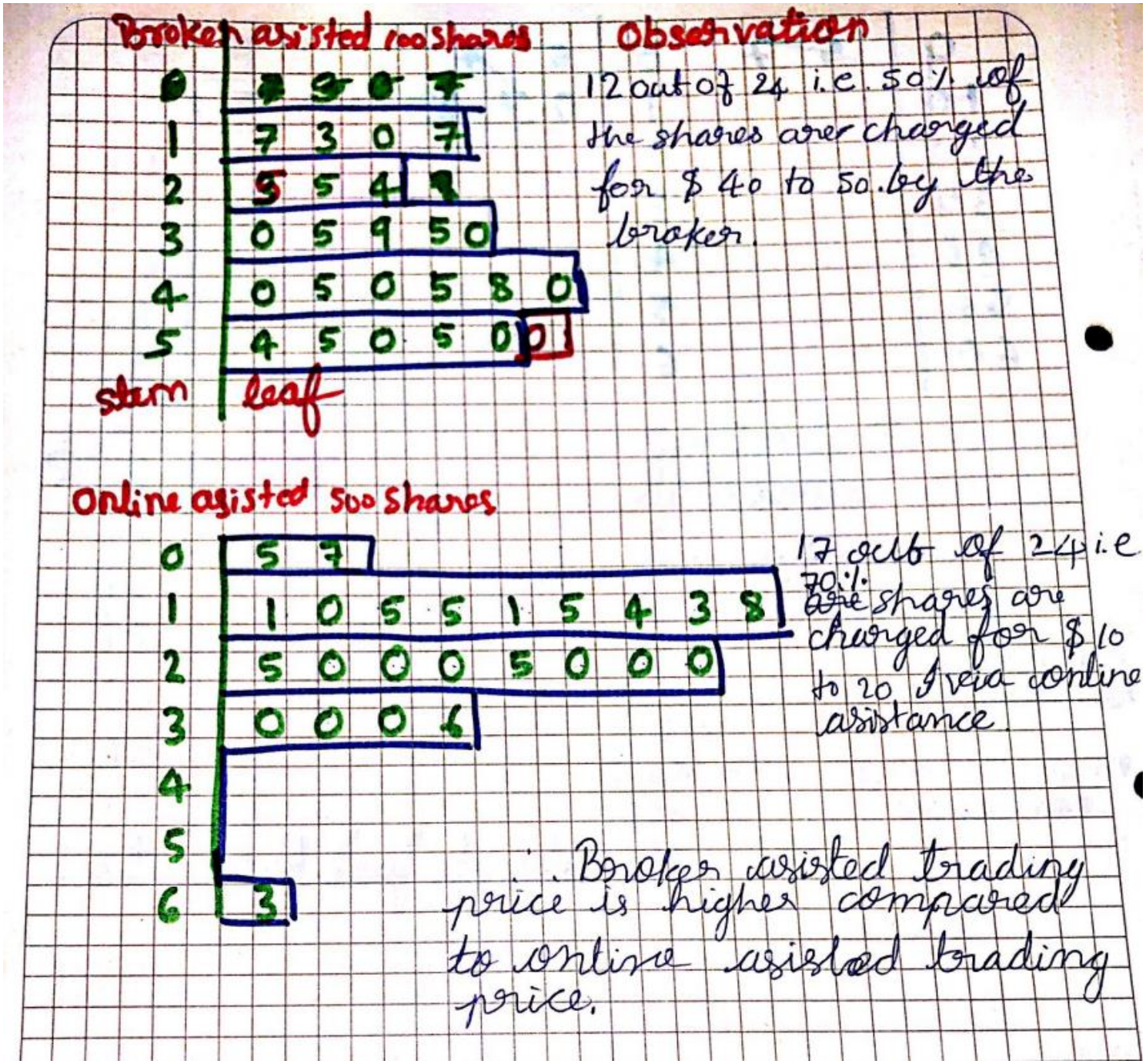


26. The American Association of Individual Investors conducts an annual survey of discount brokers. The following prices charged are from a sample of 24 discount brokers (*AAII Journal,* January 2003). The two types of trades are a broker-assisted trade of 100 shares at $50 per share and an online trade of 500 shares at $50 per share.

Broker	Broker-Assisted 100 Shares at $50/Share	Online 500 Shares at $50/Share	Broker	Broker-Assisted 100 Shares at $50/Share	Online 500 Shares at $50/Share
Accutrade	30.00	29.95	Merrill Lynch Direct	50.00	29.95
Ameritrade	24.99	10.99	Muriel Siebert	45.00	14.95
Banc of America	54.00	24.95	NetVest	24.00	14.00
Brown & Co.	17.00	5.00	Recom Securities	35.00	12.95
Charles Schwab	55.00	29.95	Scottrade	17.00	7.00
CyberTrader	12.95	9.95	Sloan Securities	39.95	19.95
E*TRADE Securities	49.95	14.95	Strong Investments	55.00	24.95
First Discount	35.00	19.75	TD Waterhouse	45.00	17.95
Freedom Investments	25.00	15.00	T. Rowe Price	50.00	19.95
Harrisdirect	40.00	20.00	Vanguard	48.00	20.00
Investors National	39.00	62.50	Wall Street Discount	29.95	19.95
MB Trading	9.95	10.55	York Securities	40.00	36.00

a. Round the trading prices to the nearest dollar and develop a stem-and-leaf display for 100 shares at $50 per share. Comment on what you learned about broker-assisted trading prices.

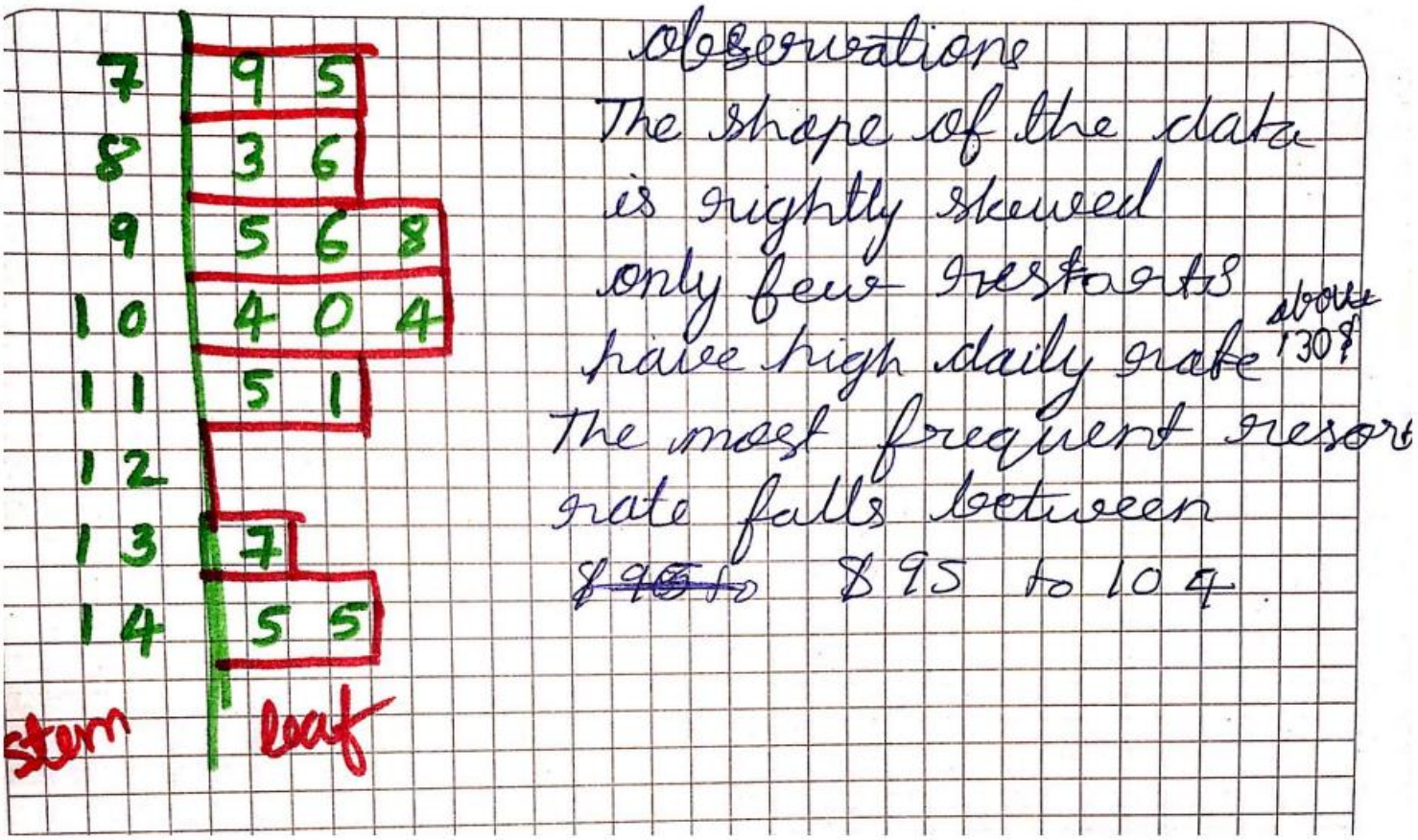
b. Round the trading prices to the nearest dollar and develop a stretched stem-and-leaf display for 500 shares online at $50 per share. Comment on what you learned about online trading prices.



27. Most major ski resorts offer family programs that provide ski and snowboarding instruction for children. The typical classes provide four to six hours on the snow with a certified instructor. The daily rate for a group lesson at 15 ski resorts follows (*The Wall Street Journal,* January 20, 2006).

Resort	Location	Daily Rate	Resort	Location	Daily Rate
Beaver Creek	Colorado	$137	Okemo	Vermont	$ 86
Deer Valley	Utah	115	Park City	Utah	145
Diamond Peak	California	95	Butternut	Massachusetts	75
Heavenly	California	145	Steamboat	Colorado	98
Hunter	New York	79	Stowe	Vermont	104
Mammoth	California	111	Sugar Bowl	California	100
Mount Sunapee	New Hampshire	96	Whistler-Blackcomb	British Columbia	104
Mount Bachelor	Oregon	83			

a. Develop a stem-and-leaf display for the data.
b. Interpret the stem-and-leaf display in terms of what it tells you about the daily rate for these ski and snowboarding instruction programs.



28. The 2004 Naples, Florida, minimarathon (13.1 miles) had 1228 registrants (*Naples Daily News,* January 17, 2004). Competition was held in six age groups. The following data show the ages for a sample of 40 individuals who participated in the marathon.

49	33	40	37	56
44	46	57	55	32
50	52	43	64	40
46	24	30	37	43
31	43	50	36	61
27	44	35	31	43
52	43	66	31	50
72	26	59	21	47

a. Show a stretched stem-and-leaf display.
b. What age group had the largest number of runners?
c. What age occurred most frequently?
d. A *Naples Daily News* feature article emphasized the number of runners who were "20-something." What percentage of the runners were in the 20-something age group? What do you suppose was the focus of the article?

Observations

Shape of the data is approximately symetric

b) The age group 40-45 had the largest number of runners.

c) 43 age occured most frequently

d) 4 out of 40 are belongs to age group 20-29. ~~i.e 1 of the~~

i.e only 10% of runners are belonging to this group.

e) The focus of the article was suppose to be on young yuth are not towards participating in marathon as compared to the middle age group between 30 to 49.

29. The following data are for 30 observations involving two qualitative variables, x and y. The categories for x are A, B, and C; the categories for y are 1 and 2.

Observation	x	y	Observation	x	y
1	A	1	16	B	2
2	B	1	17	C	1
3	B	1	18	B	1
4	C	2	19	C	1
5	B	1	20	B	1
6	C	2	21	C	2
7	B	1	22	B	1
8	C	2	23	C	2
9	A	1	24	A	1
10	B	1	25	B	1
11	A	1	26	C	2
12	B	1	27	C	2
13	C	2	28	A	1
14	C	2	29	B	1
15	C	2	30	B	2

a. Develop a crosstabulation for the data, with x as the row variable and y as the column variable.
b. Compute the row percentages.
c. Compute the column percentages.
d. What is the relationship, if any, between x and y?

a)

x \ y	1	2	Total
A	5	0	5
B	11	2	13
C	2	10	12
Total	18	12	30

Cross tabulation

b) Row Percentage

x \ y	1	2	Total
A	100%	0%	100%
B	84.6%	15.4%	100%
C	16.6%	83.4%	100%

c) Column Percentage

	y	
X	1	2
A	27.8%	0%
B	61.1%	16.7%
C	11.1%	83.3%
Total	100%	100%

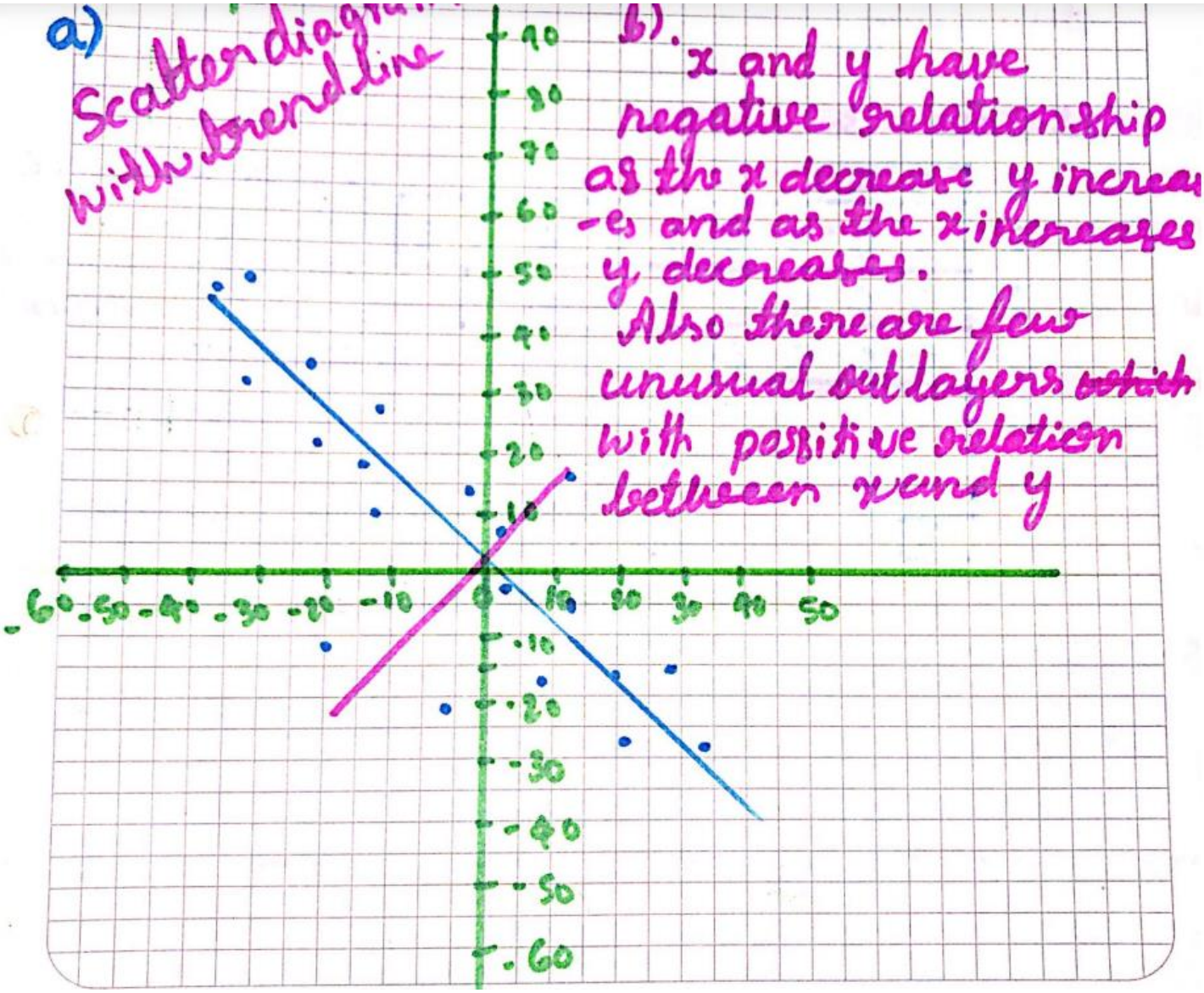
d) A always relates with y=1
B most of the time relates with y = 1
C most of the time relates with y = 2

30. The following 20 observations are for two quantitative variables, x and y.

Observation	x	y	Observation	x	y
1	−22	22	11	−37	48
2	−33	49	12	34	−29
3	2	8	13	9	−18
4	29	−16	14	−33	31
5	−13	10	15	20	−16
6	21	−28	16	−3	14
7	−13	27	17	−15	18
8	−23	35	18	12	17
9	14	−5	19	−20	−11
10	3	−3	20	−7	−22

a. Develop a scatter diagram for the relationship between x and y.
b. What is the relationship, if any, between x and y?

a) Scatter diagram with trend line



b). x and y have negative relationship as the x decrease y increas-es and as the x increases y decreases.
Also there are few unusual out layers which with positive relation between x and y

Applications

31. The following crosstabulation shows household income by educational level of the head of household (*Statistical Abstract of the United States: 2008*).

	Household Income ($1000s)					
Educational Level	Under 25	25.0–49.9	50.0–74.9	75.0–99.9	100 or more	Total
Not H.S. graduate	4207	3459	1389	539	367	9961
H.S. graduate	4917	6850	5027	2637	2668	22099
Some college	2807	5258	4678	3250	4074	20067
Bachelor's degree	885	2094	2848	2581	5379	13787
Beyond bach. deg.	290	829	1274	1241	4188	7822
Total	13106	18490	15216	10248	16676	73736

a. Compute the row percentages and identify the percent frequency distributions of income for households in which the head is a high school graduate and in which the head holds a bachelor's degree.

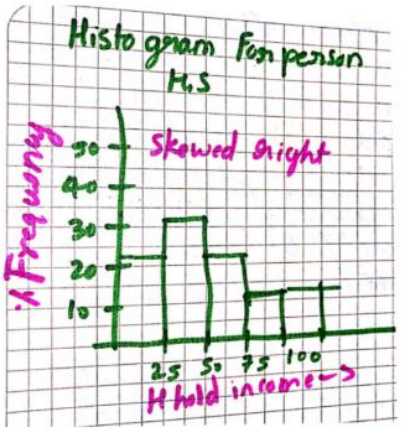
b. What percentage of households headed by high school graduates earn $75,000 or more? What percentage of households headed by bachelor's degree recipients earn $75,000 or more?

c. Construct percent frequency histograms of income for households headed by persons with a high school degree and for those headed by persons with a bachelor's degree. Is any relationship evident between household income and educational level?

a) Row Percentation with percent frequency

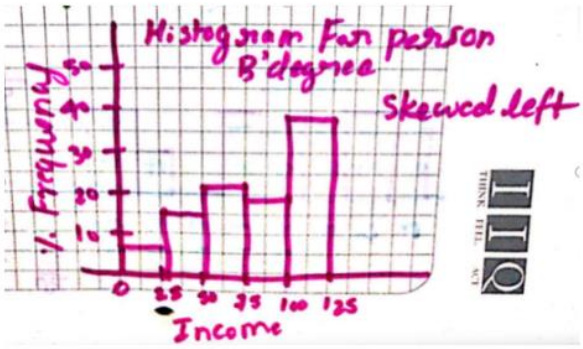
Edu Level	Household income ($1000s)					
	<25	25.0-49.9	50.0-74.9	75.0-99.9	>100	Total
Not H.S	42.3%	34.8%	14%	.5%	.4%	100%
H. S	22.2%	31%	22.7%	12%	12.1 %	100%
Some college	14%	26%	23%	16%	21%	100%
B'degree	6%	15%	21%	19%	39%	100%
Beyond B'Deg	3%	11%	16%	16%	54%	100%

b) 12% + 12.1 = 24.1 % percentage of house holds by high school graduates earn $75000 or more.
19%+19% = 38% percentage of households headed by bachelor's degree recipients earn $75000 or more



Histogram observations for Person with High school

Highest number of person with H.s has income in 25000 to 50000 range
Only few number of person with H.s has income more than 75000 causing skewness at right side.
76% of persons with H.s have income below 75000



Histogram observations for person with B'Degree

Highest number of person with B'degree have income above 100,000
68% of persons with B'degree have income more than 75000
only few number of person with B'degree have income less than 25000

We can observe that a positive relation between Education level and income as as the level of education increases the income level is also increases.

32. Refer again to the crosstabulation of household income by educational level shown in exercise 31.
   a. Compute column percentages and identify the percent frequency distributions displayed. What percentage of the heads of households did not graduate from high school?
   b. What percentage of the households earning $100,000 or more were headed by a person having schooling beyond a bachelor's degree? What percentage of the households headed by a person with schooling beyond a bachelor's degree earned over $100,000? Why are these two percentages different?
   c. Compare the percent frequency distributions for those households earning "Under 25," "100 or more," and for "Total." Comment on the relationship between household income and educational level of the head of household.

a) Column Percentation with percent frequency

Edu Level	Household income ($1000s)					
	<25	25.0-49.9	50.0-74.9	75.0-99.9	>100	Total
Not H.S	32%	18.7%	9%	5%	2%	
H. S	37.5%	37%	33%	26%	16%	
Some college	21.5%	28.4%	31%	32%	24%	
B'degree	6.8	11.4%	19%	25%	32%	
Beyond B'Deg	2.2%	4.5	8%	12%	25%	
Total	100%	100%	100%	100%	100%	

9961/73736 X100=13.5% percentage of heads of households did not graduate from high school

b) 25% of hosue holds earning %100000 or more were headed by person having schooling beyond a bacehlor's degree.

5379/13787 X100= 54% of persons with schooling beyond a bachelor's degree earned more than $100000

C. 32+37.5 = 69.5 % of <25000 earning person heads are of No H.s and H.s level education
only 2% of <25000 earning persons are of beyond B' Degree. Where as 57% of >100000 earnings are by person with B'Degree and Beyond B'degree and only 2% of >100000 earnings are by No. H.S persons.

There is a strong positive relationship between the Education levle of person with their capacity of earning. The people with higher education are earning higher income.

33. Recently, management at Oak Tree Golf Course received a few complaints about the condition of the greens. Several players complained that the greens are too fast. Rather than react to the comments of just a few, the Golf Association conducted a survey of 100 male and 100 female golfers. The survey results are summarized here.

Male Golfers

	Greens Condition	
Handicap	**Too Fast**	**Fine**
Under 15	10	40
15 or more	25	25

Female Golfers

	Greens Condition	
Handicap	**Too Fast**	**Fine**
Under 15	1	9
15 or more	39	51

a. Combine these two crosstabulations into one with Male and Female as the row labels and Too Fast and Fine as the column labels. Which group shows the highest percentage saying that the greens are too fast?

b. Refer to the initial crosstabulations. For those players with low handicaps (better players), which group (male or female) shows the highest percentage saying the greens are too fast?

c. Refer to the initial crosstabulations. For those players with higher handicaps, which group (male or female) shows the highest percentage saying the greens are too fast?

d. What conclusions can you draw about the preferences of men and women concerning the speed of the greens? Are the conclusions you draw from part (a) as compared with parts (b) and (c) consistent? Explain any apparent inconsistencies.

a. Summary cross tabulation from aggregates of two seperate cross tabulation

M/F	Greens Conditions		
	Too Fast	Fine	Total
Male	35	65	100
Female	40	60	100
Total	75	125	200

Female Group shows the highest percentage 53.3% saying that the greens are too fast.

b. For those players with low hadicaps (under 15) male group shows the highest percentage (10/35 = 28.57%)saying the greens are too fast.

c. For those players with higher hadicaps (15 or more) female group shows the higest percentage(39/40=97.5 %) saying the greens are too fast.

d. 125/200 = 62.5 % of both male and female are fine with the faster grwoing green. only 37.5 % of both male and female are not happy with the fast grwoing green

Row percentation

Male Handicap	Toofast	Fine	Total
Under 15	20%	80%	100%
15 or more	50%	50%	100%

Column percentation

Male Handicap	Toofast	Fine
Under 15	28.5%	61.5%
15 or more	71.5%	38.5%
Total	100%	100%

Row Percentation

Female Handicap	Too fast	Fine	Total
Under 15	10%	90%	100%
15 or more	43.3%	56.7%	100%

Column Percentation

Female Handicap	Toofast	Fine
Under 15	2.5%	15%
15 or more	97.5%	85%
Total	100%	100%

Row Percentation

M/F	Greens Conditions		
	Too Fast	Fine	Total
Male	35%	65%	100%
Female	40 %	60%	100%

Column Percentation

M/F	Greens Conditions	
	Too Fast	Fine
Male	46.7%	52%
Female	53.3%	48%
Total	100%	100%

35. Refer to the data in Table 2.13.
    a. Prepare a crosstabulation of the data on Fund Type (rows) and the expense ratio (columns). Use classes of .25–.49, .50–.74, .75–.99, 1.00–1.24, and 1.25–1.49 for Expense Ratio (%).
    b. Prepare a percent frequency distribution for Expense Ratio (%).
    c. What conclusions can you draw about fund type and the expense ratio?

TABLE 2.13 FINANCIAL DATA FOR A SAMPLE OF 45 MUTUAL FUNDS

Fund Name	Fund Type	Net Asset Value ($)	5-Year Average Return (%)	Expense Ratio (%)	Morningstar Rank
Amer Cent Inc & Growth Inv	DE	28.88	12.39	0.67	2-Star
American Century Intl. Disc	IE	14.37	30.53	1.41	3-Star
American Century Tax-Free Bond	FI	10.73	3.34	0.49	4-Star
American Century Ultra	DE	24.94	10.88	0.99	3-Star
Ariel	DE	46.39	11.32	1.03	2-Star
Artisan Intl Val	IE	25.52	24.95	1.23	3-Star
Artisan Small Cap	DE	16.92	15.67	1.18	3-Star
Baron Asset	DE	50.67	16.77	1.31	5-Star
Brandywine	DE	36.58	18.14	1.08	4-Star
Brown Cap Small	DE	35.73	15.85	1.20	4-Star
Buffalo Mid Cap	DE	15.29	17.25	1.02	3-Star
Delafield	DE	24.32	17.77	1.32	4-Star
DFA U.S. Micro Cap	DE	13.47	17.23	0.53	3-Star
Dodge & Cox Income	FI	12.51	4.31	0.44	4-Star
Fairholme	DE	31.86	18.23	1.00	5-Star
Fidelity Contrafund	DE	73.11	17.99	0.89	5-Star
Fidelity Municipal Income	FI	12.58	4.41	0.45	5-Star
Fidelity Overseas	IE	48.39	23.46	0.90	4-Star
Fidelity Sel Electronics	DE	45.60	13.50	0.89	3-Star
Fidelity Sh-Term Bond	FI	8.60	2.76	0.45	3-Star
Fidelity	DE	39.85	14.40	0.56	4-Star
FPA New Income	FI	10.95	4.63	0.62	3-Star
Gabelli Asset AAA	DE	49.81	16.70	1.36	4-Star
Greenspring	DE	23.59	12.46	1.07	3-Star
Janus	DE	32.26	12.81	0.90	3-Star

Janus Worldwide	IE	54.83	12.31	0.86	2-Star
Kalmar Gr Val Sm Cp	DE	15.30	15.31	1.32	3-Star
Managers Freemont Bond	FI	10.56	5.14	0.60	5-Star
Marsico 21st Century	DE	17.44	15.16	1.31	5-Star
Mathews Pacific Tiger	IE	27.86	32.70	1.16	3-Star
Meridan Value	DE	31.92	15.33	1.08	4-Star
Oakmark I	DE	40.37	9.51	1.05	2-Star
PIMCO Emerg Mkts Bd D	FI	10.68	13.57	1.25	3-Star
RS Value A	DE	26.27	23.68	1.36	4-Star
T. Rowe Price Latin Am.	IE	53.89	51.10	1.24	4-Star
T. Rowe Price Mid Val	DE	22.46	16.91	0.80	4-Star
Templeton Growth A	IE	24.07	15.91	1.01	3-Star
Thornburg Value A	DE	37.53	15.46	1.27	4-Star
USAA Income	FI	12.10	4.31	0.62	3-Star
Vanguard Equity-Inc	DE	24.42	13.41	0.29	4-Star
Vanguard Global Equity	IE	23.71	21.77	0.64	5-Star
Vanguard GNMA	FI	10.37	4.25	0.21	5-Star
Vanguard Sht-Tm TE	FI	15.68	2.37	0.16	3-Star
Vanguard Sm Cp Idx	DE	32.58	17.01	0.23	3-Star
Wasatch Sm Cp Growth	DE	35.41	13.98	1.19	4-Star

a)

	Expense Ration					
Fund tyepe	.25-.49	.50-.74	75-99	1.00-1.24	1.25- 1.49	T
DE	2	3	5	11	7	28
IE	0	1	1	4	1	7
FI	6	3	0	0	1	10
Total	8	7	6	15	9	45

B) Percent frequency distributon of expense ration

Class	Freqyency	Percent frquency
.25-.49	8	18
.50-.74	7	16
.75-.99	6	13
1.0-1.24	15	33
1.25-1.49	9	20
Total	45	100

C) Column Percentation

Fund type	.25-.49	.50-.74	.75-.99	1.00-1.24	1.25-1.49
DE	25%	43%	83%	73%	78%
IE	0	14%	17%	27%	11%
FI	75%	43%	00	00	11%
Total	100%	100%	100%	100%	100%

Conclusion

The expense ratin .25-.49 most often of Fi Fund type and some of DE type only

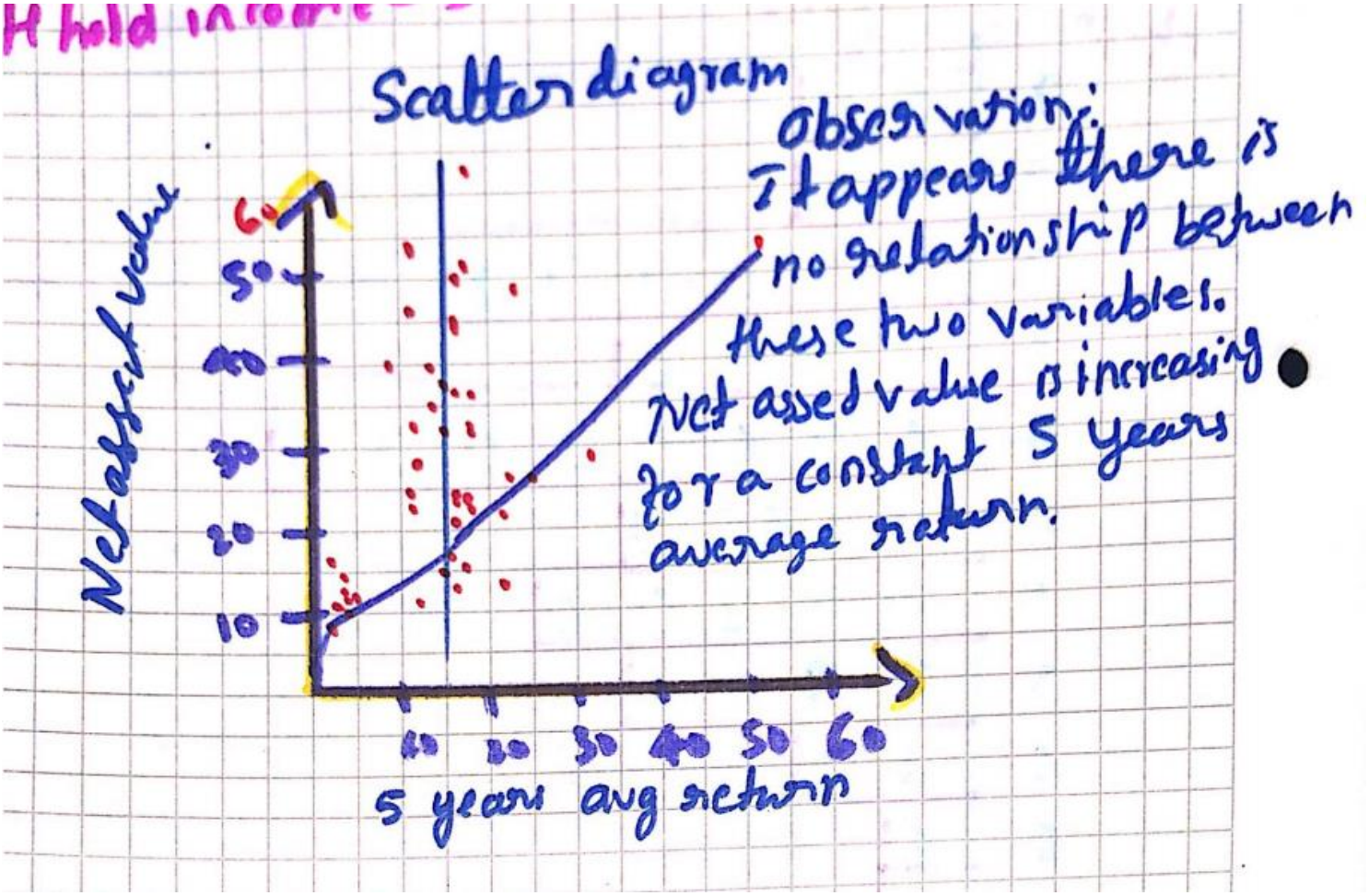
The expense ration range .50-.74 is most often of type either DE or FI and very less of type IE

The expnse ration range .75-.99 is most often of type DE and very less of type IE

The expnse ration range 1.00-1.24 is of DE fund type and some of IE type

The expense tarion range 1.25=1.49 is most often of DE type and very less of type IE and FI.

36. Refer to the data in Table 2.13.
    a. Prepare a scatter diagram with 5-Year Average Return (%) on the horizontal axis and Net Asset Value ($) on the vertical axis.
    b. Comment on the relationship, if any, between the variables.



34. Table 2.13 shows a data set containing information for 45 mutual funds that are part of the *Morningstar Funds500* for 2008. The data set includes the following five variables:

    Fund Type: The type of fund, labeled DE (Domestic Equity), IE (International Equity), and FI (Fixed Income)

    Net Asset Value ($): The closing price per share

    5-Year Average Return (%): The average annual return for the fund over the past 5 years

    Expense Ratio (%): The percentage of assets deducted each fiscal year for fund expenses

    Morningstar Rank: The risk adjusted star rating for each fund; Morningstar ranks go from a low of 1-Star to a high of 5-Stars

    a. Prepare a crosstabulation of the data on Fund Type (rows) and the average annual return over the past 5 years (columns). Use classes of 0–9.99, 10–19.99, 20–29.99, 30–39.99, 40–49.99, and 50–59.99 for the 5-Year Average Return (%).
    b. Prepare a frequency distribution for the data on Fund Type.
    c. Prepare a frequency distribution for the data on 5-Year Average Return (%).
    d. How has the crosstabulation helped in preparing the frequency distributions in parts (b) and (c)?
    e. What conclusions can you draw about the fund type and the average return over the past 5 years?

a)

Fund type	Average Annual return for 5 years						Total
	0-9.99	10-19.99	20-29.99	30-39.99	40-49.99	50-59.99	
DE	2 (20%)	25 (89%)	1 (25%)	0	0	0	28
IE	0	2 (7%)	3 (75%)	2 (100%)	0	1 (100%)	8
FI	8 (80%)	1 (4%)	0	0	0	0	9
Total	10	28	4	2	0	1	45

b) Frequency distribution of fund type

	Frequency	%
DE	28	62%
IE	8	18%
FI	9	20%
	45	100%

c) Frequency distribution of Average annual return

Average Return	%	Frequency
0-9.99	22%	10
10-19.99	62%	28
20-29.99	8%	4
30-39.99	4%	2
40-49.99		0
50-59.99	2%	1
	100%	45

e) 62% of average return are of type DE

62% of fund type have average annual return in between 10-19.99.

80% of 0-9.99 average return are of type FI

d) Using the cross tabulation we can directly get the Frequency distribution of fund type by using the row wise total in the cross tabulation and also frequency distribution of average annual return by using the column wise total in the cross tabulation.

37. The U.S. Department of Energy's Fuel Economy Guide provides fuel efficiency data for cars and trucks (Fuel Economy website, February 22, 2008). A portion of the data for 311 compact, midsize, and large cars is shown in Table 2.14. The data set contains the following variables:

Size: Compact, Midsize, and Large

Displacement: Engine size in liters

Cylinders: Number of cylinders in the engine

Drive: Front wheel (F), rear wheel (R), and four wheel (4)

Fuel Type: Premium (P) or regular (R) fuel

City MPG: Fuel efficiency rating for city driving in terms of miles per gallon

Hwy MPG: Fuel efficiency rating for highway driving in terms of miles per gallon

TABLE 2.14 FUEL EFFICIENCY DATA FOR 311 CARS

Car	Size	Displacement	Cylinders	Drive	Fuel Type	City MPG	Hwy MPG
1	Compact	3.1	6	4	P	15	25
2	Compact	3.1	6	4	P	17	25
3	Compact	3.0	6	4	P	17	25
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
161	Midsize	2.4	4	F	R	22	30
162	Midsize	2.0	4	F	P	19	29
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
310	Large	3.0	6	F	R	17	25
311	Large	3.0	6	F	R	18	25

The complete data set is contained in the file named FuelData08.

a. Prepare a crosstabulation of the data on Size (rows) and Hwy MPG (columns). Use classes of 15–19, 20–24, 25–29, 30–34, and 35–39 for Hwy MPG.

b. Comment on the relationship beween Size and Hwy MPG.

c. Prepare a crosstabulation of the data on Drive (rows) and City MPG (columns). Use classes of 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, and 35–39 for City MPG.
d. Comment on the relationship between Drive and City MPG.
e. Prepare a crosstabulation of the data on Fuel Type (rows) and City MPG (columns). Use classes of 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, and 35–39 for City MPG.
f. Comment on the relationship between Fuel Type and City MPG.

a) Cross tabulation of Hwy Mpg and size

Size	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	total
Compact	26	76	9	0	0	111
Midsize	0	0	85	46	4	135
Large	0	0	65	0	0	65
$\sum$	26	76	159	46	4	311

Row wise percentiation

Size	Hwy MPG					
	15-19	20-24	25-29	30-34	35-39	Total
Compact	24%	68%	8%	0%	0%	100%
Midsize	0%	0%	63%	34%	3%	100%
Large	0%	0%	100%	0%	0%	100%
%freq Hwy	8%	25%	51%	15%	1%	100%

Columnwise percentation

Size	Hwy MPG					
	15-19	20-24	25-29	30-34	35-39	%freq size
Compact	100%	100%	6%	0%	0%	36%
Midsize	0%	0%	53%	100%	100%	43%
Large	0%	0%	41%	0%	0%	21%
Total	100%	100%	100%	100%	100%	100%

b) 92% of compact size cars have less HwyMPG i.e. within 24 MPG
63% of midsize cars have Hwy MPG oin between 25-29 and 37 % of Middsize cars have above 30 Hwy MPG
All Large cars have Hwy MPg in between 25-29

All cars with Hwy MPG below 24 belongs to compact size cars
All cars with Hwy MPG above 30 belongs to Mid size cars
53% of cars with Hwy Mpg of 25-29 are belongs to Mid size cars
41% of cars with hwy MPg of 25-29 are belongs to Large size cars

51% of total cars are having Hwy MPg in between 25-29 range
only 16% of total cars are having Hwy MPG of above 30
33% of total cars are having less than 24 Hwy MPG

36% of total cars are of compact size
43% of total cars are having mid size
only 21% of total cars are having large size

c)

The rows in the table will represent the drive (front wheel(F), rear wheel(R), four wheel(4)) and the columns will represent CityMPG (fuel efficiency for city driving in terms of miles per gallon) separated in classes:

Size	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	total
F	0	2	98	74	9	1	1	185
R	1	23	50	1	0	0	0	75
4	0	10	33	8	0	0	0	51
$\sum$	1	35	181	83	9	1	1	311

d)

Row wise percentation

Drive	City MPG							
	5-9	10-14	15-19	20-24	25-29	30- 34	35- 39	Total
Front wheel	0%	1%	53%	40%	5%	.5%	.5%	100 %
Rear wheel	1%	31%	67%	1%	0%	0%	0%	100%
Four wheel	0%	20%	65%	15%	0%	0%	0%	100%
%Freq City MPG	0.3%	11.1%	58%	27%	3%	.3	.3	100%

Columnwise percentation

Drive	City MPG							
	5-9	10-14	15-19	20-24	25-29	30-34	35-39	% Freq of Drive
Front wheel	0%	6%	54%	89%	100%	100%	100%	59%
Rear wheel	100%	66%	28%	1%	0%	0%	0%	24
Four wheel	0%	28%	18%	10%	0%	0%	0%	17%
Total	100%	100%	100%	100%	100%	100%	100%	100%

54% of Front wheel cars are having below 19 city MPG
45% of front wheel cars are having above 20 city MPG
only 6% of Front wheel cars are having above 25 city MPG

67% of Rear wheel cars are having 15-19 city Mpg
99% of Rear wheel cars are having below 19 city MPG
None of the rear wheel cars are having above 24 city MPG

85% of Four wheel cars are having below 19 city MPG
15% of Four wheel cars are having city MPG between 20-24
None of the Four wheel cars are having city MPG above 25

only Front wheel cars are having above 25 city mpg

89% of city MPG range 20-24 are of Front wheel cars
10% of city MPG range 20-24 are of Four wheel cars
only 1% of MPG range 20-24 are of Rear wheel cars

54% of city MPG range 15-19 are of Front wheel cars
28% of city MPG range 15-19 are of Rear wheel cars
only 18% of MPG range 15-19 are of Four wheel cars

66% of City MPG range 10-14 are of Rear wheel cars
28% of city MPG range 10-14 are of Four wheel cars
only 6% of city MPG range 10-14 are of Front wheel cars

city MPG 5-9 are of only Rear wheel cars

59% of total cares are of Front wheel drive
24% of total cars are of Rear wheel drive
17% of total cars are of Four wheel drive

58% of totl cars are hving city MPG in range of 15-19
27% of total cars are having city MPG in range 20-24
only 11% of total cars are havign city MPG below 14
and only 3.6% of total cars are having city MPG above 25

OVer all the Front wheel drive has more fuel efficiency as 45% of Front wheel driving cars have above 20 City Mpg as compared to Four wheel and Rear wheel drive. And Over all 58% of total cars are having city MPG in range of 15-19 across all type of wheels so it appears there is a wheek relation between the Drive and city MPG

e)

The rows in the table will represent Fuel type (Premium (P) or Regular (R)) and the columns will represent CityMPG (fuel efficiency for city driving in terms of miles per gallon) separated in classes:

Size	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39	total
P	1	33	105	18	0	0	0	157
R	0	2	76	65	9	1	1	154
$\sum$	1	35	181	83	9	1	1	311

f) Row Percentation

Fuel Type	City MPG							
	5-9	10-14	15-19	20-24	25-29	30-34	35-39	Total
Premium	.6 %	20%	68%	11.4%	0 %	0%	0%	100%
Regular	0%	1%	49%	42%	7%	.5%	.5%	100%
% Freq of City MPG	0.3%	11%	58%	27%	3%	.3%	.4%	100%

Column Percentation

Fuel Type	city MPG							
	5-9	10-14	15-19	20-24	25-29	30-34	35-39	% Freq of Fuel Type
Premium	100%	94%	58%	22%	0%	0%	0%	50.5 %
Regular	0%	6%	42%	78%	100%	100%	100%	49.5%

68% of permium fuel car have city MPG between 15-19
88% of premium fule cars ara having city mpg below 19
only 11% of Premium fule cars are having city mpg 20-24
None of premium fule cars are having city mpg above 25

49% of regular fuel cars are having city mpg 15-19
50% of regular fuel cars are having city mpg 20 mpg
8% of regular fuel cars are hving city mpg above 25

Over all 50.5% of total cars are of premium fuel cars and 49.5% of total cars are of regular fuel cars

over all 58% of total cars are having city MPG 15-19
27% of total cars are hving city mpg 20-29
only 4% of cars having city mpg above 25

As 99% of regular fule cars having city mpg above 15 and only 79% of premium fule cars are having city mpg above 15, it shows that regular fule cars are having more efficiency than premium fule cars.
As 58% total cars exhibits that they are having fule range of 15-19 so we can under stand that there is a week relationship between Fuel and city MPG

38. Refer to exercise 37 and the data in the file named FuelData08.
   a. Prepare a crosstabulation of the data on Displacement (rows) and Hwy MPG (columns). Use classes of 1.0–2.9, 3.0–4.9, and 5.0–6.9 for Displacement. Use classes of 15–19, 20–24, 25–29, 30–34, and 35–39 for Hwy MPG.
   b. Comment on the relationship, if any, between Displacement and Hwy MPG.
   c. Develop a scatter diagram of the data on Displacement and Hwy MPG. Use the vertical axis for Hwy MPG.
   d. What does the scatter diagram developed in part (c) indicate about the relationship, if any, between Displacement and Hwy MPG?
   e. In investigating the relationship between Displacement and Hwy MPG you developed a tabular summary of the data (crosstabulation) and a graphical summary (scatter diagram). In this case which approach do you prefer? Explain.

a)

Displacement	Hwy mpg 15-19	20-24	25-29	30-34	35-39	total
1.0-2.9	0	6	72	46	4	128
3.0-4.9	3	56	86	0	0	145
5.0-6.9	23	14	1	0	0	38
total	26	76	159	46	4	311

Row percentation

Displacement	Hwy mpg 15-19	20-24	25-29	30-34	35-39	total % dsg Displacement
1.0-2.9	0%	4%	56%	36%	3%	100%
3.0-4.9	3%	38%	59%	0%	0%	100%
5.0-6.9	60%	37%	3%	0%	0%	100%
% dsg of Hwy mpg	8%	24%	51%	15%	2%	100%

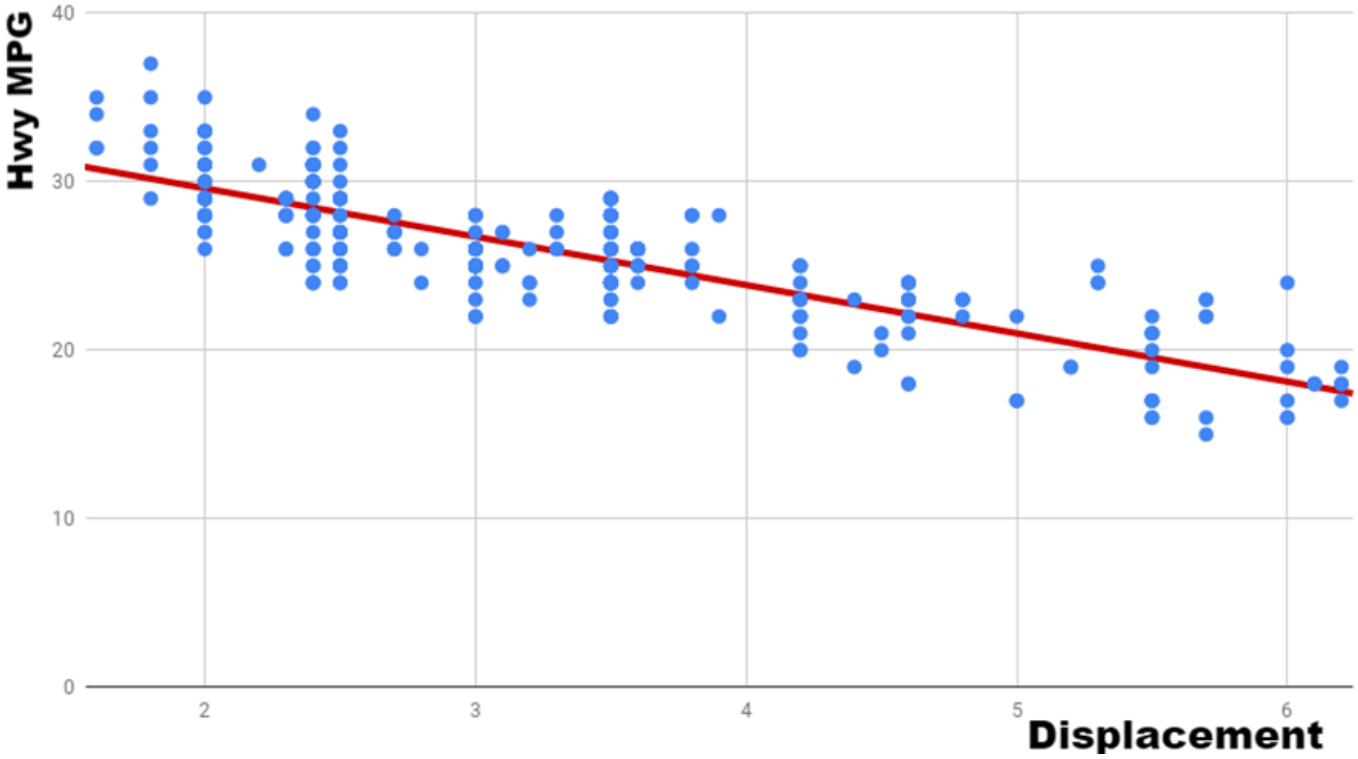
Column Percentation

Displacem	Hwy mpg 15-19	20-24	25-29	30-34	35-39	% dsg Disp
1.0-2.9	0%	8%	45%	100%	100%	41%
3.0-4.9	12%	74%	54%	0%	0	47%
5.0-6.9	88%	18%	1%	0%	0%	12%
Total	100%	100%	100%	100%	100%	100%

b) 96% of displacement 1.0-2.9 have above 25 Hwy mpg
59% of displacement 3.0-4.9 have above 25 Hwy mpg
only 3% of displacement 5.0-6.9 have above 25 Hwy mpg
Hence it shows that cars with lower displacement have higher fuel efficiency. Hence there is a strong relationship between Displacement and Hwy mpg.

c)

The scatter diagram is developed using raw data in file:



d)

Scatter diagram indicated on the existing relationship between displacement in vehicle and efficiency rating on highway driving in terms of miles per gallon. The relationship is negative because the line of trend have negative slope, which means that efficiency rating decreases with displacement increase.

e)

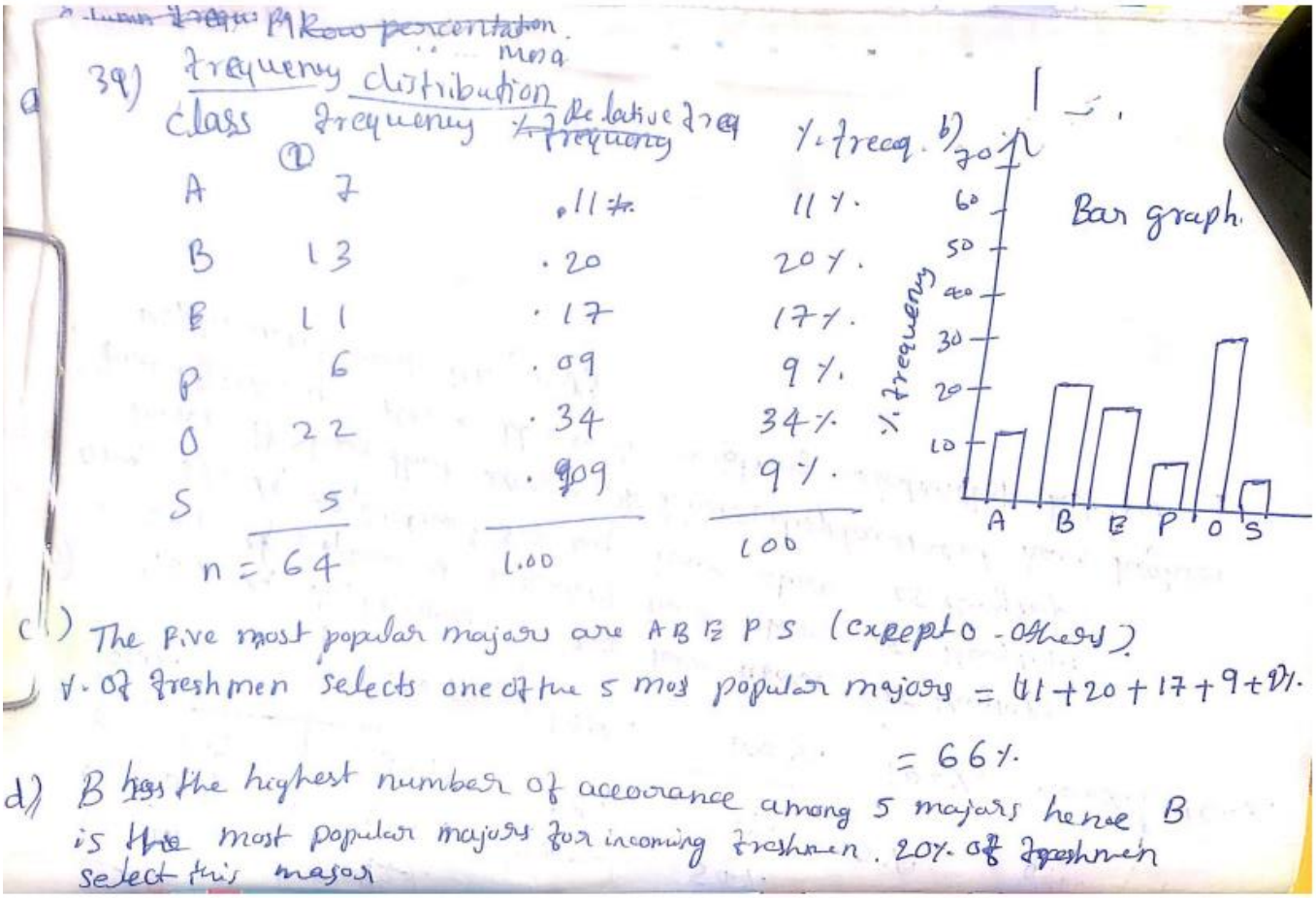
Crosstabulation method requires developing percentage distribution within data in case to estimate the relationship between data. Scatter diagram directly indicates on the nature of the relationship. Both methods are equally helping in estimation of relationship but graphic method does not require additional information which makes it a bit easier.

39. The Higher Education Research Institute at UCLA provides statistics on the most popular majors among incoming college freshmen. The five most popular majors are Arts and Humanities (A), Business Administration (B), Engineering (E), Professional (P), and Social Science (S) (*The New York Times Almanac,* 2006). A broad range of other (O) majors, including biological science, physical science, computer science, and education, are grouped together. The majors selected for a sample of 64 college freshmen follow.

S	P	P	O	B	E	O	E	P	O	O	B	O	O	O	A
O	E	E	B	S	O	B	O	A	O	E	O	E	O	B	P
B	A	S	O	E	A	B	O	S	S	O	O	E	B	O	B
A	E	B	E	A	A	P	O	O	E	O	B	B	O	P	B

a. Show a frequency distribution and percent frequency distribution.
b. Show a bar chart.

c. What percentage of freshmen select one of the five most popular majors?
d. What is the most popular major for incoming freshmen? What percentage of freshmen select this major?



39) Frequency distribution

class	frequency	relative freq	% freq.
A	7	.11	11%
B	13	.20	20%
E	11	.17	17%
P	6	.09	9%
O	22	.34	34%
S	5	.09	9%
	n = 64	1.00	100

b) Bar graph

c) The five most popular majors are A B E P S (except O - Others)
% of freshmen selects one of the 5 most popular majors = 11 + 20 + 17 + 9 + 9%.
= 66%.

d) B has the highest number of accorance among 5 majors hence B is most popular majors for incoming freshmen. 20% of freshmen select this major.

40. General Motors had a 23% share of the automobile industry with sales coming from eight divisions: Buick, Cadillac, Chevrolet, GMC, Hummer, Pontiac, Saab, and Saturn (*Forbes*, December 22, 2008). The data set GMSales shows the sales for a sample of 200 General Motors vehicles. The division for the vehicle is provided for each sale.

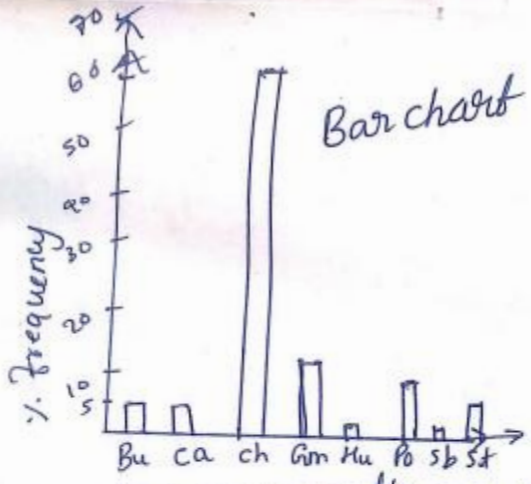
a. Show the frequency distribution and the percent frequency distribution of sales by division for General Motors.

b. Show a bar chart of the percent frequency distribution.

c. Which General Motors division was the company leader in sales? What was the percentage of sales for this division? Was this General Motors' most important division? Explain.

d. Due to the ongoing recession, high gasoline prices, and the decline in automobile sales, General Motors was facing bankruptcy in 2009. A government "bail-out" loan and a restructuring of the company were anticipated. Expectations were that General Motors could not continue to operate all eight divisions. Based on the percentage of sales, which of the eight divisions looked to be the best candidates for General Motors to discontinue? Which divisions looked to be the least likely candidates for General Motors to discontinue?

Division class	Frequency	% Frequency
Buick	10	5 %
Cadilac	10	5 %
Chevralet	122	61 %
Gmc	24	12 %
Hummer	2	1 %
Pontiac	18	9 %
Saab	2	1 %
Saturn	12	6 %
	200	100%



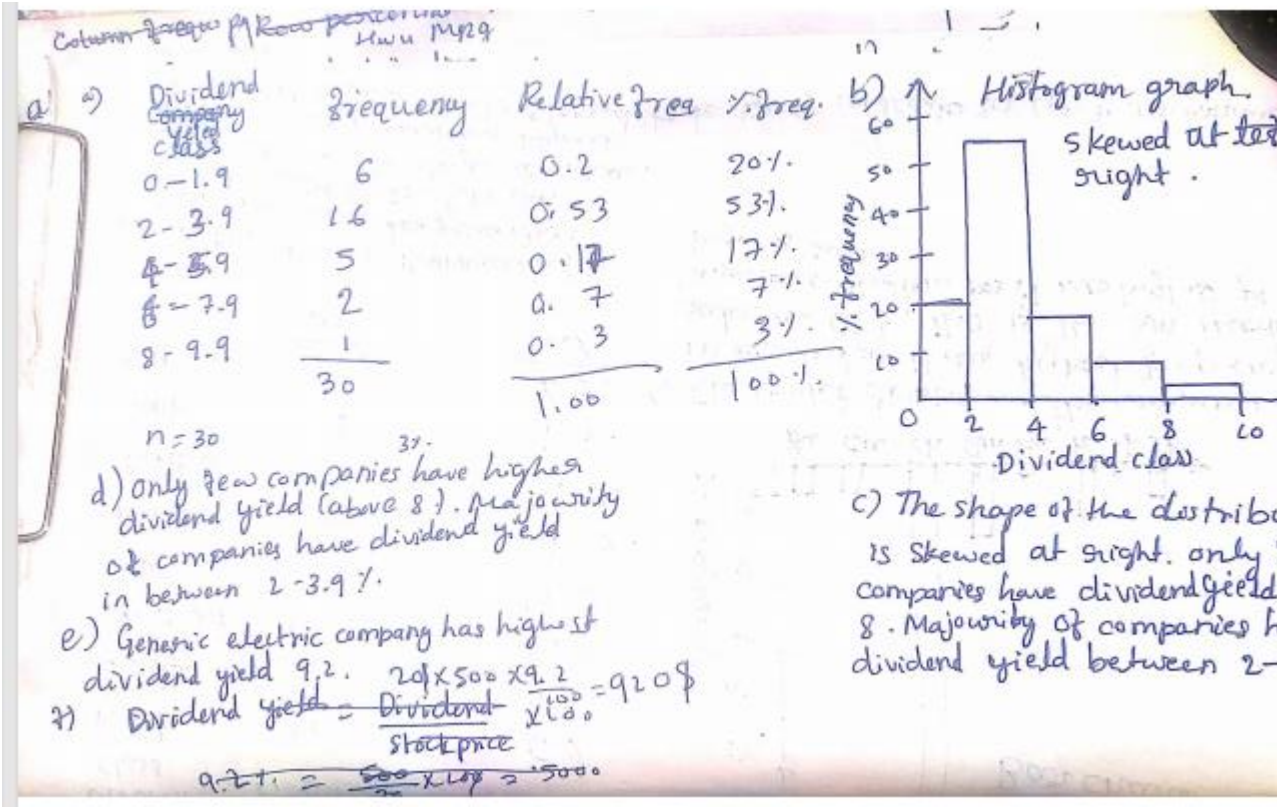
c) Chevralet division was the company lead in to sale has it has highest frequency of sales. as 61 %. This is the Gm most important division. as it has highest percentage of sales.

d) As the divisions Hummer and Saab have very less percentage of sale within 5%. GM can look into these division to discontinue. As the chevralet division has highest % of sales 61%. This divisions looked to be likely candidates for Gm to discontinue

41. Dividend yield is the annual dividend paid by a company expressed as a percentage of the price of the stock (Dividend/Stock Price × 100). The dividend yield for the Dow Jones Industrial Average companies is shown in Table 2.15 (*The Wall Street Journal*, June 8, 2009).
    a. Construct a frequency distribution and percent frequency distribution.
    b. Construct a histogram.
    c. Comment on the shape of the distribution.
    d. What do the tabular and graphical summaries tell about the dividend yields among the Dow Jones Industrial Average companies?
    e. What company has the highest dividend yield? If the stock for this company currently sells for $20 per share and you purchase 500 shares, how much dividend income will this investment generate in one year?

TABLE 2.15 DIVIDEND YIELD FOR DOW JONES INDUSTRIAL AVERAGE COMPANIES

Company	Dividend Yield %	Company	Dividend Yield %
3M	3.6	IBM	2.1
Alcoa	1.3	Intel	3.4
American Express	2.9	J.P. Morgan Chase	0.5
AT&T	6.6	Johnson & Johnson	3.6
Bank of America	0.4	Kraft Foods	4.4
Boeing	3.8	McDonald's	3.4
Caterpillar	4.7	Merck	5.5
Chevron	3.9	Microsoft	2.5
Cisco Systems	0.0	Pfizer	4.2
Coca-Cola	3.3	Procter & Gamble	3.4
DuPont	5.8	Travelers	3.0
ExxonMobil	2.4	United Technologies	2.9
General Electric	9.2	Verizon	6.3
Hewlett-Packard	0.9	Wal-Mart Stores	2.2
Home Depot	3.9	Walt Disney	1.5



42. Approximately 1.5 million high school students take the Scholastic Aptitude Test (SAT) each year and nearly 80% of the college and universities without open admissions policies use SAT scores in making admission decisions (College Board, March 2009). The current version of the SAT includes three parts: reading comprehension, mathematics, and writing. A perfect combined score for all three parts is 2400. A sample of SAT scores for the combined three-part SAT are as follows:

1665	1525	1355	1645	1780
1275	2135	1280	1060	1585
1650	1560	1150	1485	1990
1590	1880	1420	1755	1375
1475	1680	1440	1260	1730
1490	1560	940	1390	1175

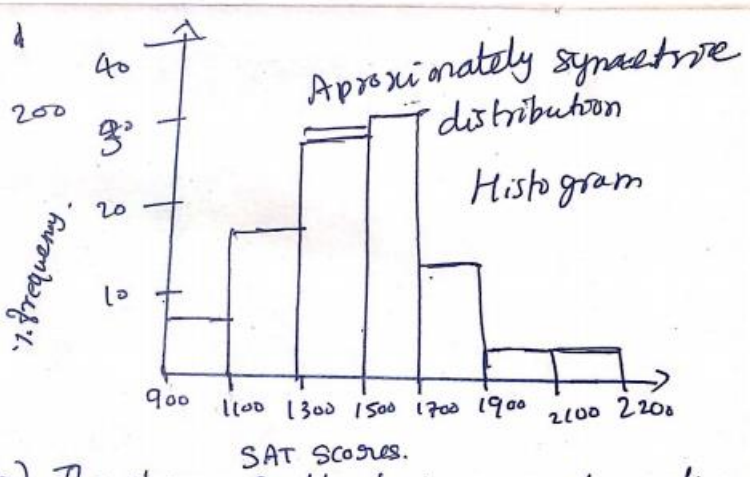
a. Show a frequency distribution and histogram. Begin with the first class starting at 800 and use a class width of 200.
b. Comment on the shape of the distribution.
c. What other observations can be made about the SAT scores based on the tabular and graphical summaries?

a) Low = 940 max = 2135

number of class = 5.

class width = $\frac{2135 - 940}{5} = 239 \approx 200$

class	freq.	% freq
900 – 1099	2	7%
1100 – 1299	5	17%
1300 – 1499	8	27%
1500 – 1699	9	30%
1700 – 1899	4	13
1900 – 2099	1	3
2100 – 2299	1	3
	n = 30	100%



b). The shape of the histogram shows the approximately symmetric distribution of the data.

c) 57% of scores are lying between 1300 – 1700 range: And only few scores are appearing at the lower and higher scores.

43. The Pittsburgh Steelers defeated the Arizona Cardinals 27 to 23 in professional football's 43rd Super Bowl. With this win, its sixth championship, the Pittsburgh Steelers became the team with the most wins in the 43-year history of the event (*Tampa Tribune,* February 2, 2009). The Super Bowl has been played in eight different states: Arizona (AZ), California (CA), Florida (FL), Georgia (GA), Louisiana (LA), Michigan (MI), Minnesota (MN), and Texas (TX). Data in the following table show the state where the Super Bowls were played and the point margin of victory for the winning team.

Super Bowl	State	Won By Points	Super Bowl	State	Won By Points	Super Bowl	State	Won By Points
1	CA	25	16	MI	5	31	LA	14
2	FL	19	17	CA	10	32	CA	7
3	FL	9	18	FL	19	33	FL	15
4	LA	16	19	CA	22	34	GA	7
5	FL	3	20	LA	36	35	FL	27
6	FL	21	21	CA	19	36	LA	3
7	CA	7	22	CA	32	37	CA	27
8	TX	17	23	FL	4	38	TX	3
9	LA	10	24	LA	45	39	FL	3
10	FL	4	25	FL	1	40	MI	11
11	CA	18	26	MN	13	41	FL	12
12	LA	17	27	CA	35	42	AZ	3
13	FL	4	28	GA	17	43	FL	4
14	CA	12	29	FL	23			
15	LA	17	30	AZ	10			

a. Show a frequency distribution and bar chart for the state where the Super Bowl was played.

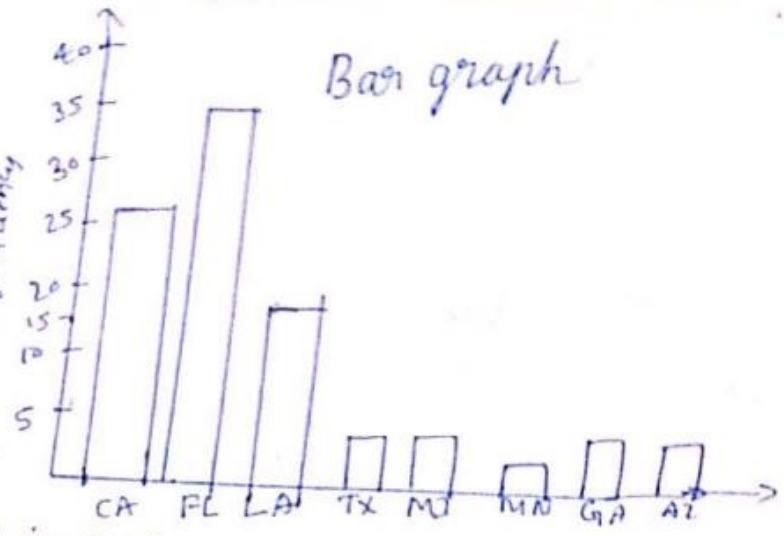
b. What conclusions can you draw from your summary in part (a)? What percentage of Super Bowls were played in the states of Florida or California? What percentage of Super Bowls were played in northern or cold-weather states?

c. Show a stretched stem-and-leaf display for the point margin of victory for the winning team. Show a histogram.

d. What conclusions can you draw from your summary in part (c)? What percentage of Super Bowls have been close games with the margin of victory less than 5 points? What percentage of Super Bowls have been won by 20 or more points?

e. The closest Super Bowl occurred when the New York Giants beat the Buffalo Bills. Where was this game played and what was the winning margin of victory? The biggest point margin in Super Bowl history occurred when the San Francisco 49ers beat the Denver Broncos. Where was this game played and what was the winning margin of victory?

state class	freq.	% freq
CA	11	26%
FL	15	35%
LA	8	17%
TX	2	5%
MI	2	5%
MN	1	2%
GA	2	5%
AZ	2	5%
n=	43	100%



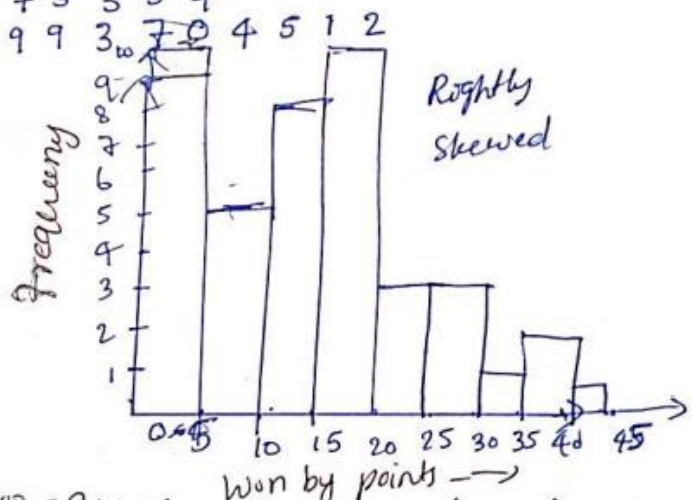
b) The Super Bowl played majority in FL & CA state. 26% of super bowls played in CA & 35% of super bowls played in LA = 26+35 = 61%.
MN & MI are the northern or cold weather states. 5+2 = 7% of super bowls played in these states.

c) Stem leaf

0	9 3 7 4 4 5 4 1 7 7 3 3 3 4
1	9 6 7 0 8 7 2 7 0 9 9 3 0 4 5 1 2
2	5 1 2 3 7 7
3	6 2 5
4	5

Streathed stem & leaf

0	3 4 4 4 1 3 3 3 4
0	9 7 5 7 7
1	0 2 0 3 0 4 1 2
1	9 6 7 8 7 7 9 9 7 5.
2	1 2 3
2	5 7 7
3	2
3	6 5.
4	
4	5



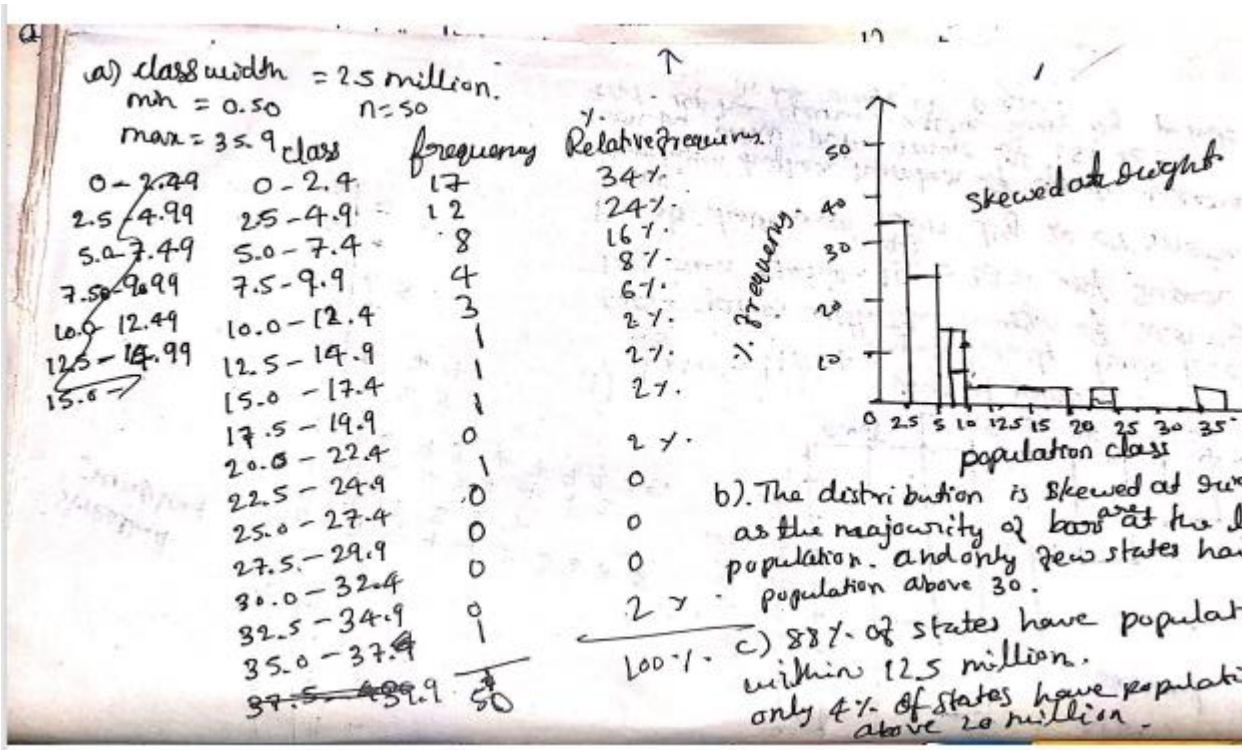
d) $\frac{9}{43} \times 100 = 21\%$ of Super bowls have been close games with the margin of victory less than 5 points. $\frac{10}{43} = 23\%$ of Super bowls have been won by 20 or more points.

Conclusion higher number of games were won by points in the range of 15-20.
74% of the games were won by points in the range of 0-20.

44. Data from the U.S. Census Bureau provides the population by state in millions of people (*The World Almanac,* 2006).

State	Population	State	Population	State	Population
Alabama	4.5	Louisiana	4.5	Ohio	11.5
Alaska	0.7	Maine	1.3	Oklahoma	3.5
Arizona	5.7	Maryland	5.6	Oregon	3.6
Arkansas	2.8	Massachusetts	6.4	Pennsylvania	12.4
California	35.9	Michigan	10.1	Rhode Island	1.1
Colorado	4.6	Minnesota	5.1	South Carolina	4.2
Connecticut	3.5	Mississippi	2.9	South Dakota	0.8
Delaware	0.8	Missouri	5.8	Tennessee	5.9
Florida	17.4	Montana	0.9	Texas	22.5
Georgia	8.8	Nebraska	1.7	Utah	2.4
Hawaii	1.3	Nevada	2.3	Vermont	0.6
Idaho	1.4	New Hampshire	1.3	Virginia	7.5
Illinois	12.7	New Jersey	8.7	Washington	6.2
Indiana	6.2	New Mexico	1.9	West Virginia	1.8
Iowa	3.0	New York	19.2	Wisconsin	5.5
Kansas	2.7	North Carolina	8.5	Wyoming	0.5
Kentucky	4.1	North Dakota	0.6		

a. Develop a frequency distribution, a percent frequency distribution, and a histogram. Use a class width of 2.5 million.
b. Discuss the skewness in the distribution.
c. What observations can you make about the population of the 50 states?



45. *Drug Store News* (September 2002) provided data on annual pharmacy sales for the leading pharmacy retailers in the United States. The following data are annual sales in millions.

Retailer	Sales	Retailer	Sales
Ahold USA	$ 1700	Medicine Shoppe	$ 1757
CVS	12700	Rite-Aid	8637
Eckerd	7739	Safeway	2150
Kmart	1863	Walgreens	11660
Kroger	3400	Wal-Mart	7250

a. Show a stem-and-leaf display.
b. Identify the annual sales levels for the smallest, medium, and largest drug retailers.
c. What are the two largest drug retailers?

a) n = 10

Stem	Leaf
0	
1	7 7 8
2	1
3	4
4	
5	
6	
7	7 2
8	6
9	
10	
11	6
12	7

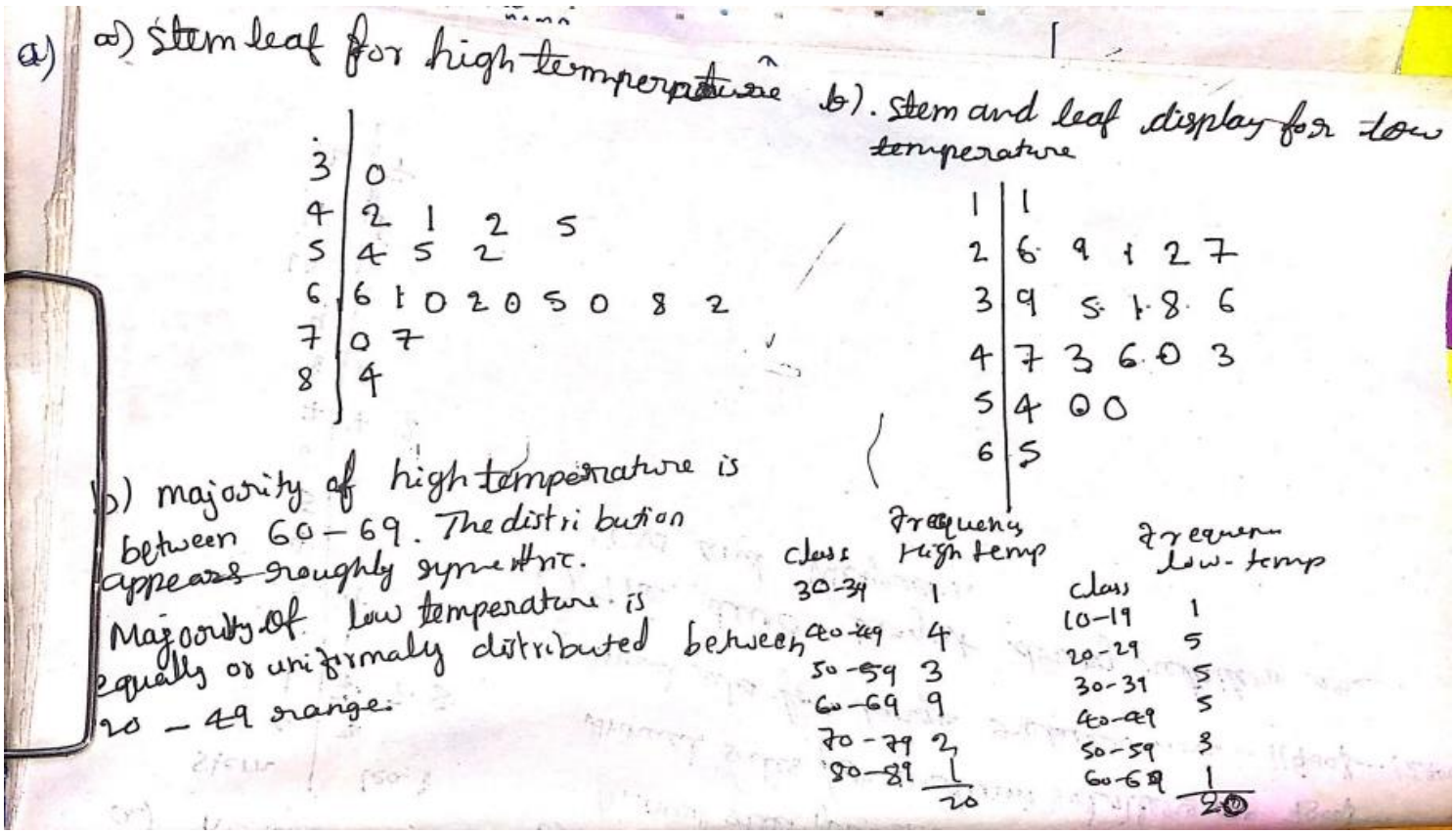
b). Annual sales for small retailers are 1700$ – 3400$
Annual sales for medium retailers are 7700$ – 8600$
Anual sales for large retailers are – 11600$ – 12700$

c) The two largest drug retailers are CVS and Walgreens.

46. The daily high and low temperatures for 20 cities follow (*USA Today*, March 3, 2006).

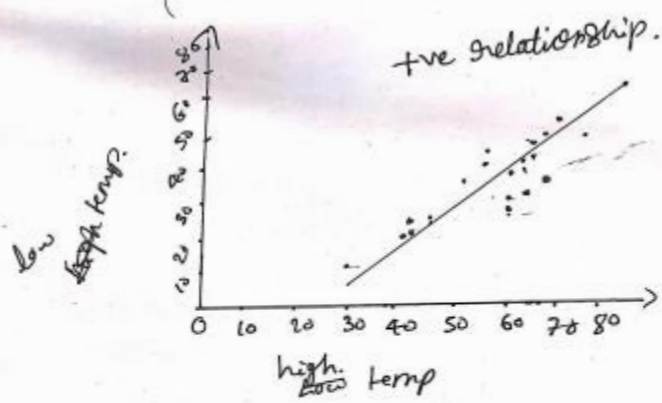
City	High	Low	City	High	Low
Albuquerque	66	39	Los Angeles	60	46
Atlanta	61	35	Miami	84	65
Baltimore	42	26	Minneapolis	30	11
Charlotte	60	29	New Orleans	68	50
Cincinnati	41	21	Oklahoma City	62	40
Dallas	62	47	Phoenix	77	50
Denver	60	31	Portland	54	38
Houston	70	54	St. Louis	45	27
Indianapolis	42	22	San Francisco	55	43
Las Vegas	65	43	Seattle	52	36

a. Prepare a stem-and-leaf display of the high temperatures.
b. Prepare a stem-and-leaf display of the low temperatures.
c. Compare the two stem-and-leaf displays and make comments about the difference between the high and low temperatures.
d. Provide a frequency distribution for both high and low temperatures.



47. Refer to the data set for high and low temperatures for 20 cities in exercise 46.
   a. Develop a scatter diagram to show the relationship between the two variables, high temperature and low temperature.
   b. Comment on the relationship between high and low temperatures.

a).



b) High and low temperature of the cities ~~are +vely~~ have +ve relationship. As the cities with higher ~~temperature~~ increases. correspondingly cities ~~lowest~~ temperature also increases.

48. One of the questions in a *Financial Times*/Harris Poll was, "How much do you favor or oppose a higher tax on higher carbon emission cars?" Possible responses were strongly favor, favor more than oppose, oppose more than favor, and strongly oppose. The following crosstabulation shows the responses obtained for 5372 adults surveyed in four countries in Europe and the United States (Harris Interactive website, February 27, 2008).

	Country					
Level of Support	**Great Britain**	**Italy**	**Spain**	**Germany**	**United States**	**Total**
Strongly favor	337	334	510	222	214	1617
Favor more than oppose	370	408	355	411	327	1871
Oppose more than favor	250	188	155	267	275	1135
Strongly oppose	130	115	89	211	204	749
Total	1087	1045	1109	1111	1020	5372

   a. Construct a percent frequency distribution for the level of support variable. Do you think the results show support for a higher tax on higher carbon emission cars?
   b. Construct a percent frequency distribution for the country variable.
   c. Does the level of support among adults in the European countries appear to be different than the level of support among adults in the United States? Explain.

a) Freque d

Class Level of support	Frequency	% frequency.
Strongly favor	1677	30%.
Favor	1871	35%.
oppose	1135	21%.
Strongly oppos op	749	14%.
	5372	100%.

b)

Class country	Freq
Britain	1087
Italy	1045
Spain	1109
German	1111
US	1020
	5372

c) Row percentages.

	GB	Italy	Spain	German	US	Total
SF	21%.	21%.	31%.	18%.	13%.	100%.
F	20%.	22%.	19%.	22%.	17%.	100%.
O	22%.	17%.	14%.	23%.	24%.	100%.
SO	17%.	16%.	12%.	28%.	27%.	100%.

Column %.

L S	GB	Italy	Spain	Germ.	US
S F	31%.	32%.	46%.	20%.	21%.
F	34%.	39%.	32%.	37%.	32%.
O	23%.	18%.	14%.	24%.	27%.
SO	12%.	11%.	8	20%.	20%.
Total	100%.	100%.	100%.	100%.	100%.

Majority of peoples in the European countries appear to be favour are in favour where as in US 50%. of are in ~~favour~~ and 50%. are in oppose.

49. Western University has only one women's softball scholarship remaining for the coming year. The final two players that Western is considering are Allison Fealey and Emily Janson. The coaching staff has concluded that the speed and defensive skills are virtually identical for the two players, and that the final decision will be based on which player has the best batting average. Crosstabulations of each player's batting performance in their junior and senior years of high school are as follows:

	Allison Fealey	
Outcome	**Junior**	**Senior**
Hit	15	75
No Hit	25	175
Total At-Bats	40	250

	Emily Janson	
Outcome	**Junior**	**Senior**
Hit	70	35
No Hit	130	85
Total At Bats	200	120

A player's batting average is computed by dividing the number of hits a player has by the total number of at-bats. Batting averages are represented as a decimal number with three places after the decimal.

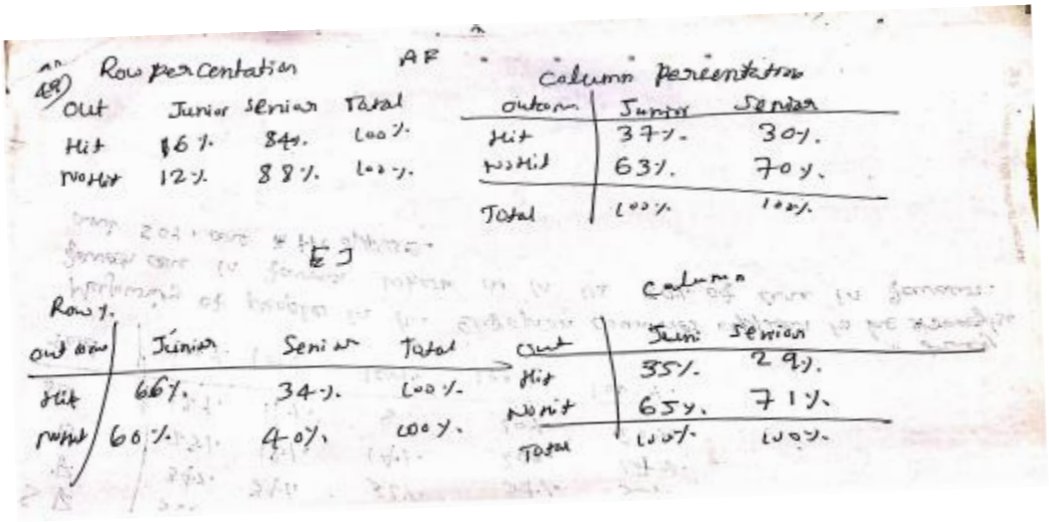
a. Calculate the batting average for each player in her junior year. Then calculate the batting average of each player in her senior year. Using this analysis, which player should be awarded the scholarship? Explain.

b. Combine or aggregate the data for the junior and senior years into one crosstabulation as follows:

	Player	
Outcome	**Fealey**	**Janson**
Hit		
No Hit		
Total At-Bats		

Calculate each player's batting average for the combined two years. Using this analysis, which player should be awarded the scholarship? Explain.

c. Are the recommendations you made in parts (a) and (b) consistent? Explain any apparent inconsistencies.



Row percentation AF

Out	Junior	Senior	Total
Hit	16%	84%	100%
NoHit	12%	88%	100%

Column percentation

outcom	Junior	Senior
Hit	37%	30%
NoHit	63%	70%
Total	100%	100%

EJ

Row%	Junior	Senior	Total
Hit	66%	34%	100%
NoHit	60%	40%	100%

Column

out	Junior	Senior
Hit	35%	29%
NoHit	65%	71%
Total	100%	100%

AF frequency outcome.

class outcome	frequency	% freq
Hit	90	31%
No Hit	200	69%
	290	100%

AF frequency type

class type	frequency	% freq
Junior	40	14%
Senior	250	86%
	290	100%

EJ outcome freq

class outcome	freq	% freq
Hit	105	33%
NoHit	215	67%
	320	100%

EJ

class type	
Junior	2
Senior	12
	32

a) batting average of AF player at Jun

$= \frac{15}{40} = .375$

batting average of AF player at Senior

$= \frac{75}{250} = 0.300$

batting average of EJ at Junior $= \frac{70}{200} = .3$

batting average of EJ at Senior $= \frac{35}{120} =$

Based on this the player AF should be a the scholarship as the average hit of AF at Junior and Senior level is better than EJ

b) Combined or aggregate cross tabulation

Outcome	Player AF	EJ
Hit	90	105
No Hit	200	215
total	290	320

Combined batting average of AF $= \frac{90}{290} = 0.310$

Combined batting average of EJ $= \frac{105}{320} = 0.328$

Based on these analysis. EJ player should be awarded the scholarship as the combined batting average of EJ is better than that of AF.

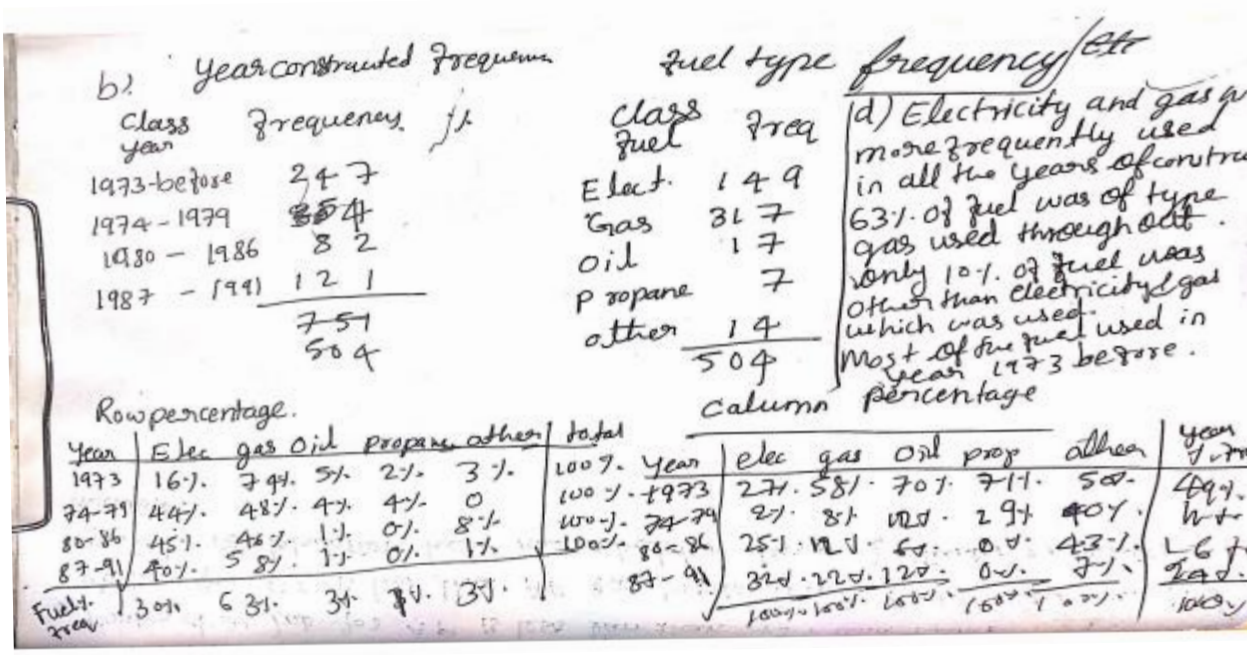
c) The recommendations we made in (a) and (b) are not consis
This inconsistancy is ~~due to~~ when the aggregated average says EJ play
better than the AF but when we looked at individual average at Junior
Senior level AF is doing better. This inconsistancy is due to the variat

in number of At-bats at each level for both players
Number of At bats for AF is less ~~than~~ than AJ. and at Junior lev
at bats for EJ is less than AF at Senior level. H
aggregated cross tabulation here has a hidden level of
variable.

50. A survey of commercial buildings served by the Cincinnati Gas & Electric Company asked what main heating fuel was used and what year the building was constructed. A partial crosstabulation of the findings follows.

Year Constructed	Fuel Type				
	Electricity	Natural Gas	Oil	Propane	Other
1973 or before	40	183	12	5	7
1974–1979	24	26	2	2	0
1980–1986	37	38	1	0	6
1987–1991	48	70	2	0	1

a. Complete the crosstabulation by showing the row totals and column totals.
b. Show the frequency distributions for year constructed and for fuel type.
c. Prepare a crosstabulation showing column percentages.
d. Prepare a crosstabulation showing row percentages.
e. Comment on the relationship between year constructed and fuel type.



51. Table 2.16 contains a portion of the data in the file named Fortune. Data on stockholders' equity, market value, and profits for a sample of 50 *Fortune* 500 companies are shown.

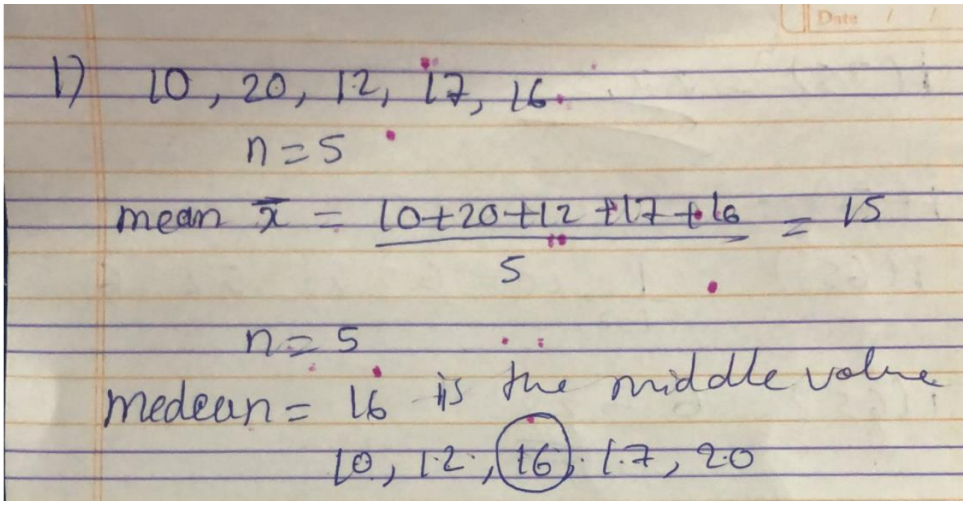
TABLE 2.16 DATA FOR A SAMPLE OF 50 *FORTUNE* 500 COMPANIES

Company	Stockholders' Equity ($1000s)	Market Value ($1000s)	Profit ($1000s)
AGCO	982.1	372.1	60.6
AMP	2698.0	12017.6	2.0
Apple Computer	1642.0	4605.0	309.0
Baxter International	2839.0	21743.0	315.0
Bergen Brunswick	629.1	2787.5	3.1
Best Buy	557.7	10376.5	94.5
Charles Schwab	1429.0	35340.6	348.5
.	.	.	.
.	.	.	.
.	.	.	.
Walgreen	2849.0	30324.7	511.0
Westvaco	2246.4	2225.6	132.0
Whirlpool	2001.0	3729.4	325.0
Xerox	5544.0	35603.7	395.0

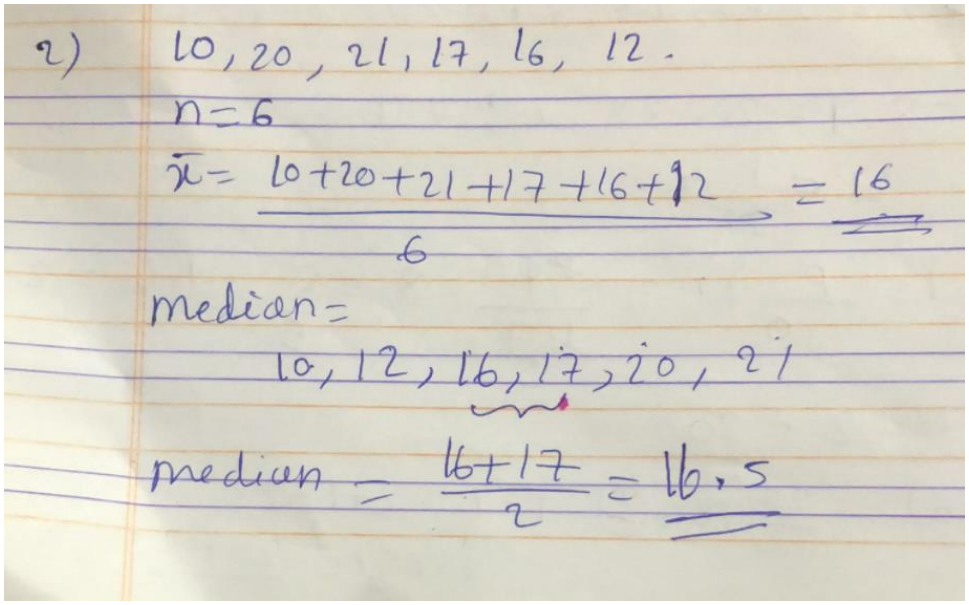
a. Prepare a crosstabulation for the variables Stockholders' Equity and Profit. Use classes of 0–200, 200–400, . . . , 1000–1200 for Profit, and classes of 0–1200, 1200–2400, . . . , 4800–6000 for Stockholders' Equity.
b. Compute the row percentages for your crosstabulation in part (a).
c. What relationship, if any, do you notice between Profit and Stockholders' Equity?

Methods

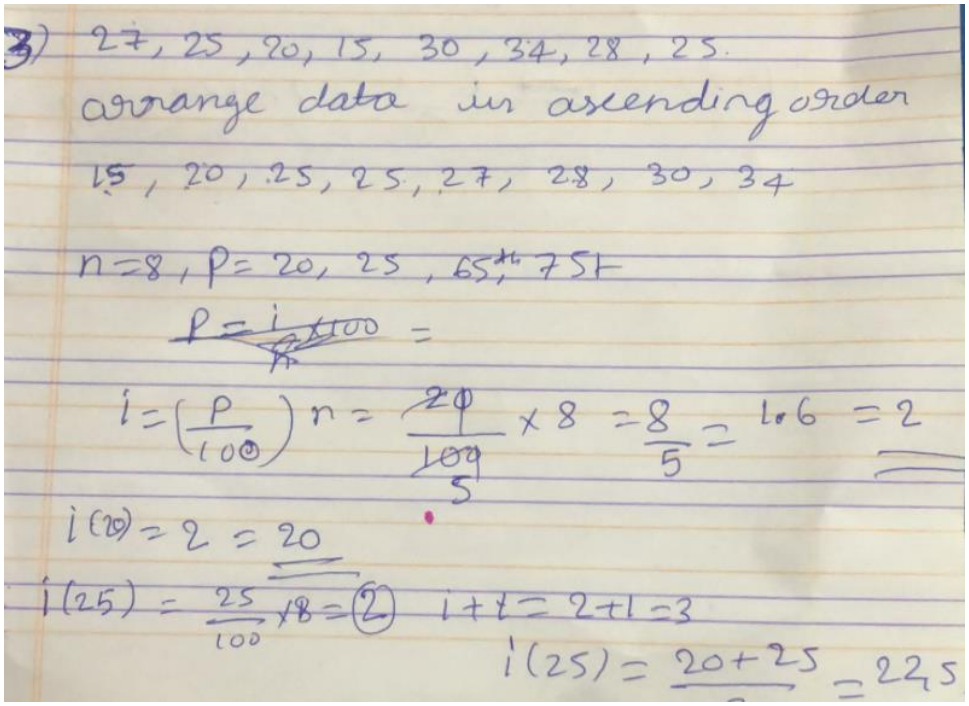
1. Consider a sample with data values of 10, 20, 12, 17, and 16. Compute the mean and median.

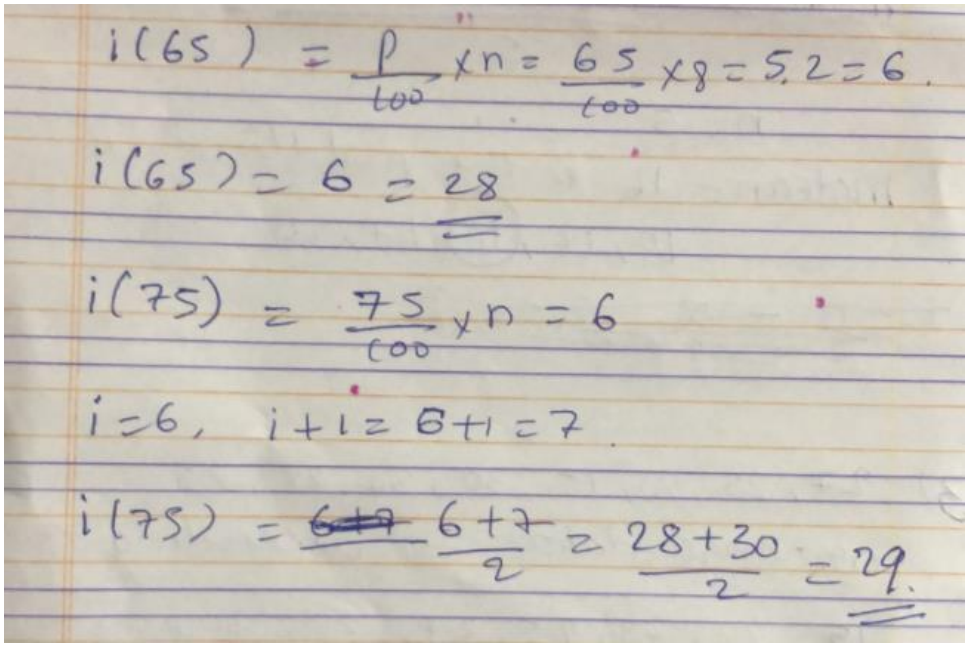


2. Consider a sample with data values of 10, 20, 21, 17, 16, and 12. Compute the mean and median.

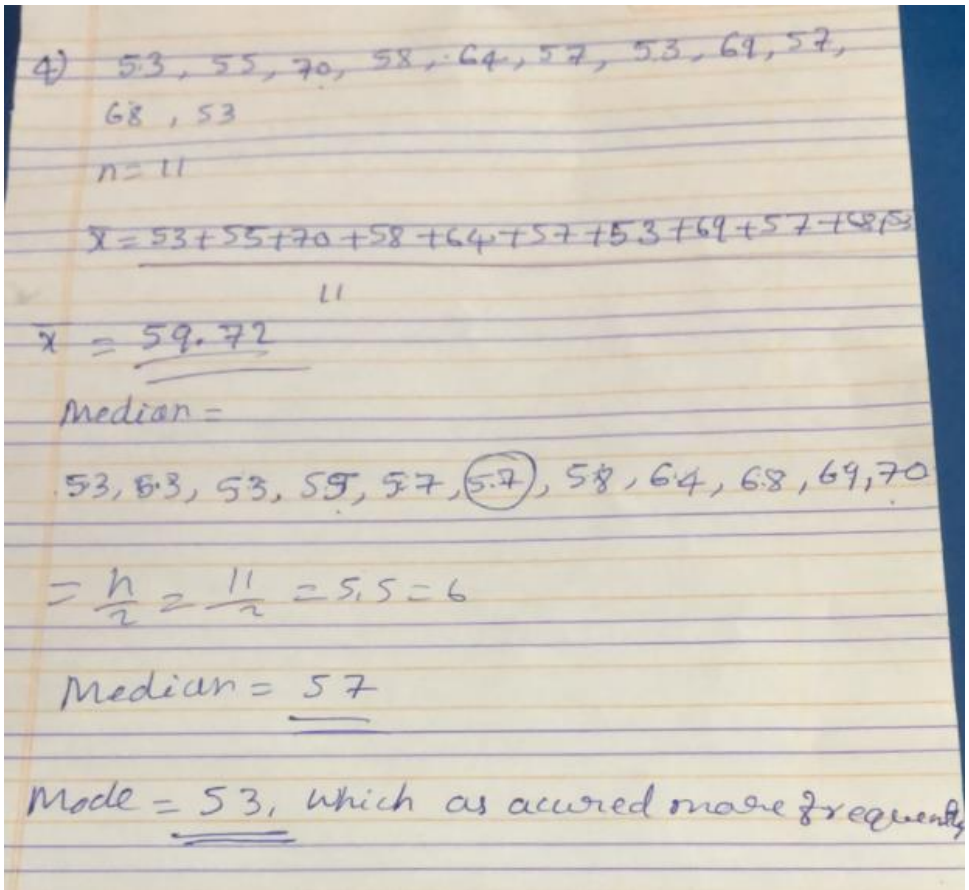


3. Consider a sample with data values of 27, 25, 20, 15, 30, 34, 28, and 25. Compute the 20th, 25th, 65th, and 75th percentiles.





4. Consider a sample with data values of 53, 55, 70, 58, 64, 57, 53, 69, 57, 68, and 53. Compute the mean, median, and mode.

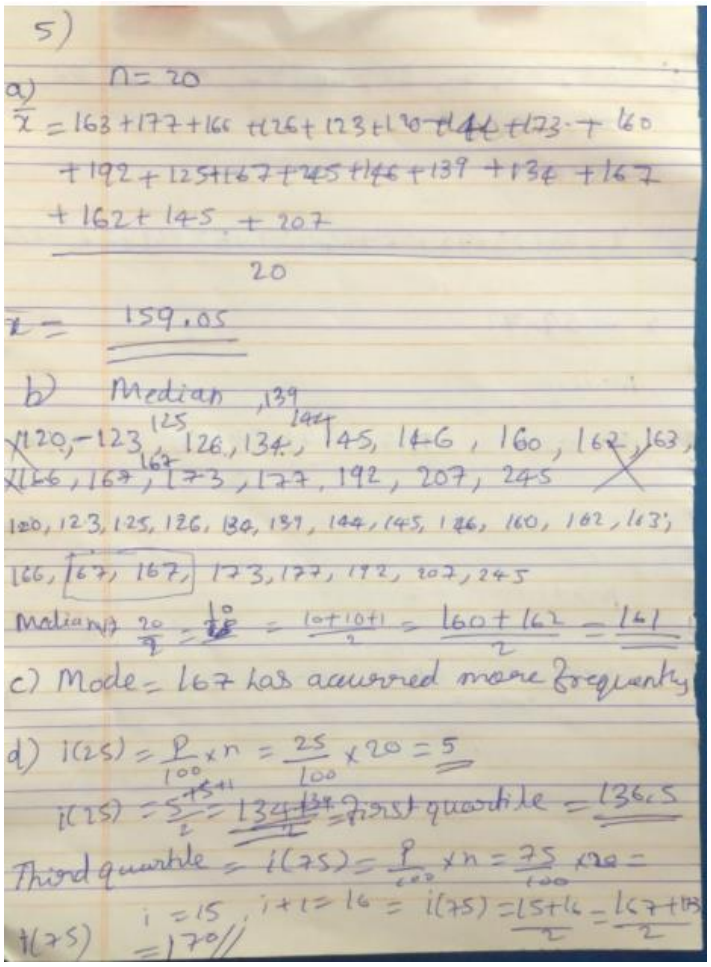


Applications

5. The Dow Jones Travel Index reported what business travelers pay for hotel rooms per night in major U.S. cities (*The Wall Street Journal*, January 16, 2004). The average hotel room rates for 20 cities are as follows:

Atlanta	$163	Minneapolis	$125
Boston	177	New Orleans	167
Chicago	166	New York	245
Cleveland	126	Orlando	146
Dallas	123	Phoenix	139
Denver	120	Pittsburgh	134
Detroit	144	San Francisco	167
Houston	173	Seattle	162
Los Angeles	160	St. Louis	145
Miami	192	Washington, D.C.	207

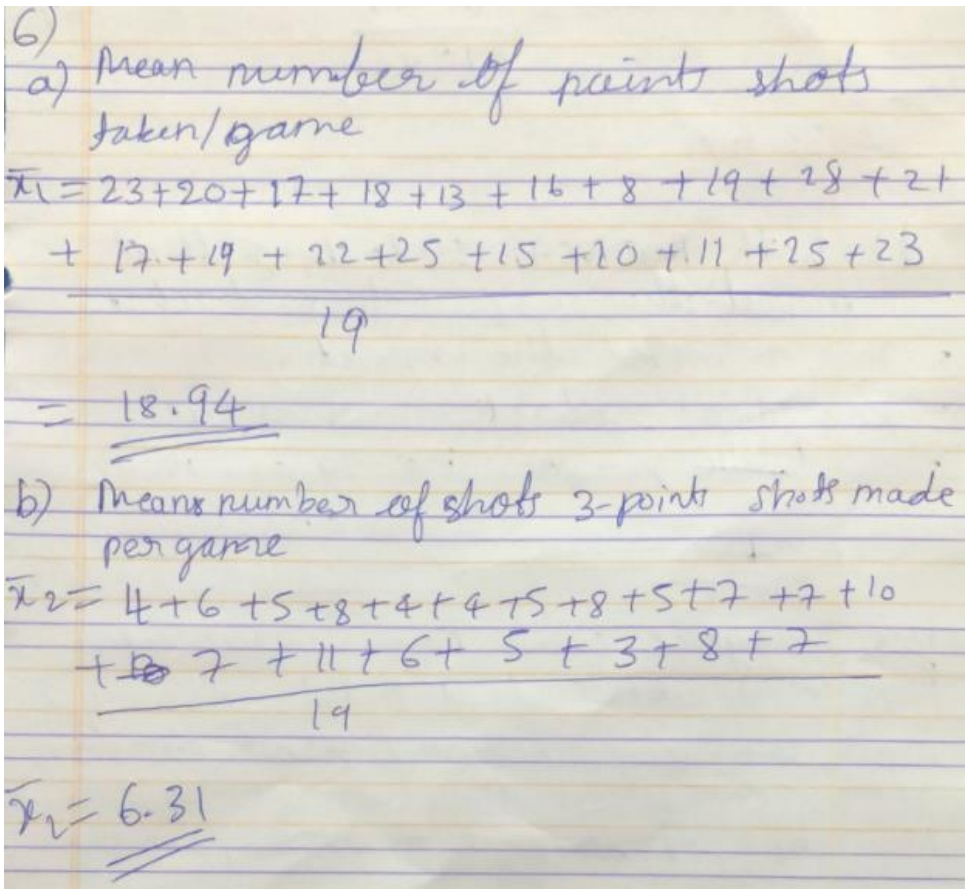
a. What is the mean hotel room rate?
b. What is the median hotel room rate?
c. What is the mode?
d. What is the first quartile?
e. What is the third quartile?

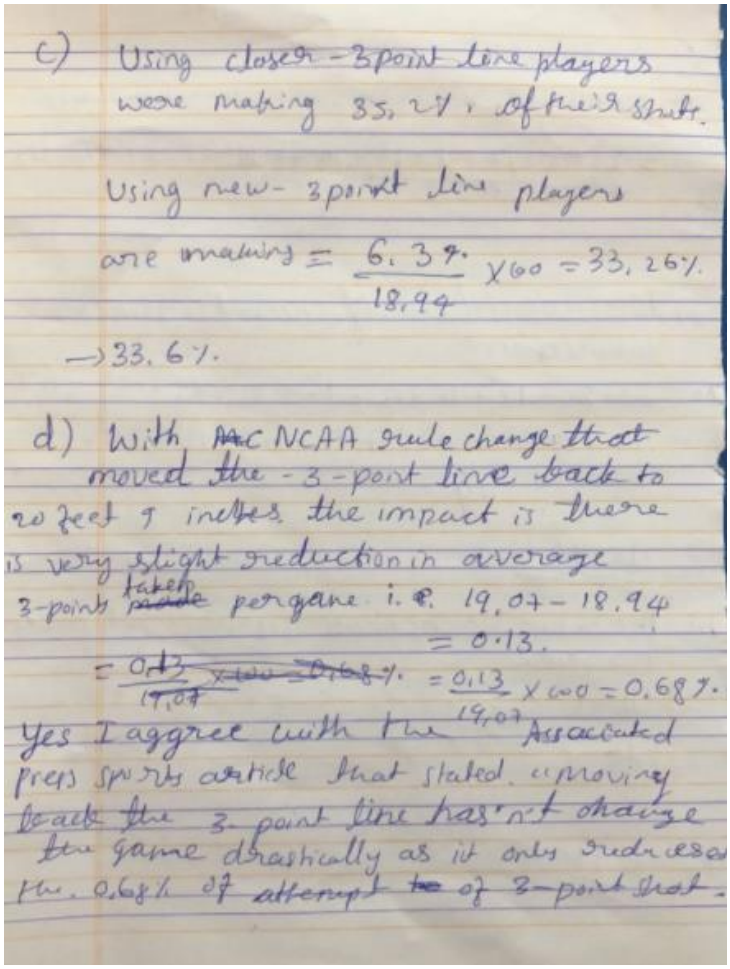


6. During the 2007–2008 NCAA college basketball season, men's basketball teams attempted an all-time high number of 3-point shots, averaging 19.07 shots per game (Associated Press Sports, January 24, 2009). In an attempt to discourage so many 3-point shots and encourage more inside play, the NCAA rules committee moved the 3-point line back from 19 feet, 9 inches to 20 feet, 9 inches at the beginning of the 2008–2009 basketball season. Shown in the following table are the 3-point shots taken and the 3-point shots made for a sample of 19 NCAA basketball games during the 2008–2009 season.

3-Point Shots	Shots Made	3-Point Shots	Shots Made
23	4	17	7
20	6	19	10
17	5	22	7
18	8	25	11
13	4	15	6
16	4	10	5
8	5	11	3
19	8	25	8
28	5	23	7
21	7		

a. What is the mean number of 3-point shots taken per game?
b. What is the mean number of 3-point shots made per game?
c. Using the closer 3-point line, players were making 35.2% of their shots. What percentage of shots were players making from the new 3-point line?
d. What was the impact of the NCAA rules change that moved the 3-point line back to 20 feet, 9 inches for the 2008–2009 season? Would you agree with the Associated Press Sports article that stated, "Moving back the 3-point line hasn't changed the game dramatically"? Explain.


6)
a) Mean number of point shots taken/game
$\bar{x}_1 = 23+20+17+18+13+16+8+19+28+21+17+19+22+25+15+20+11+25+23$
$= 18.94$
b) Means number of shots 3-point shots made per game
$\bar{x}_2 = 4+6+5+8+4+4+5+8+5+7+7+10+7+11+6+5+3+8+7$
$\bar{x}_2 = 6.31$



7. Endowment income is a critical part of the annual budgets at colleges and universities. A study by the National Association of College and University Business Officers reported that the 435 colleges and universities surveyed held a total of $413 billion in endowments. The 10 wealthiest universities are shown below (*The Wall Street Journal*, January 27, 2009). Amounts are in billion of dollars.

University	Endowment ($billion)	University	Endowment ($billion)
Columbia	7.2	Princeton	16.4
Harvard	36.6	Stanford	17.2
M.I.T.	10.1	Texas	16.1
Michigan	7.6	Texas A&M	6.7
Northwestern	7.2	Yale	22.9

a. What is the mean endowment for these universities?
b. What is the median endowment?
c. What is the mode endowment?
d. Compute the first and third quartiles?

e. What is the total endowment at these 10 universities? These universities represent 2.3% of the 435 colleges and universities surveyed. What percentage of the total $413 billion in endowments is held by these 10 universities?
f. *The Wall Street Journal* reported that over a recent five-month period, a downturn in the economy has caused endowments to decline 23%. What is the estimate of the dollar amount of the decline in the total endowments held by these 10 universities? Given this situation, what are some of the steps you would expect university administrators to be considering?

7)

a) $\bar{x} = \frac{72 + 36.6 + 19.1 + 7.6 + 7.2 + 16.4 + 17.2 + 16.1 + 6.7 + 22.9}{10}$

$= \frac{148}{10} = 14.8$

b) Median

$= 6.7, 7.2, 7.2, 7.6, 10.1, 16.1, 16.4, 17.2, 22.9, 36.6$

$= \frac{10.1 + 16.1}{2} = 13.1$

c) Mode = 7.2 which has occurred more frequently

d) $i(25) = \frac{P}{100} \times n = \frac{25}{100} \times 10 = 2.5 = 3$

$i(25) = 3 = 7.2$

$i(75) = \frac{P}{100} \times n = \frac{75}{100} \times 10 = 7.5 = 8$

$i(75) = 8 = 17.2$

e) Total endowment of these 10 universities = 148 billion

% of total $413 held by these 10 universities

$= \frac{148}{413} \times 100 = 35.83\%$

f. total endowment to declined by 23% =

$= \frac{23 \times 413}{100} = 94.99$

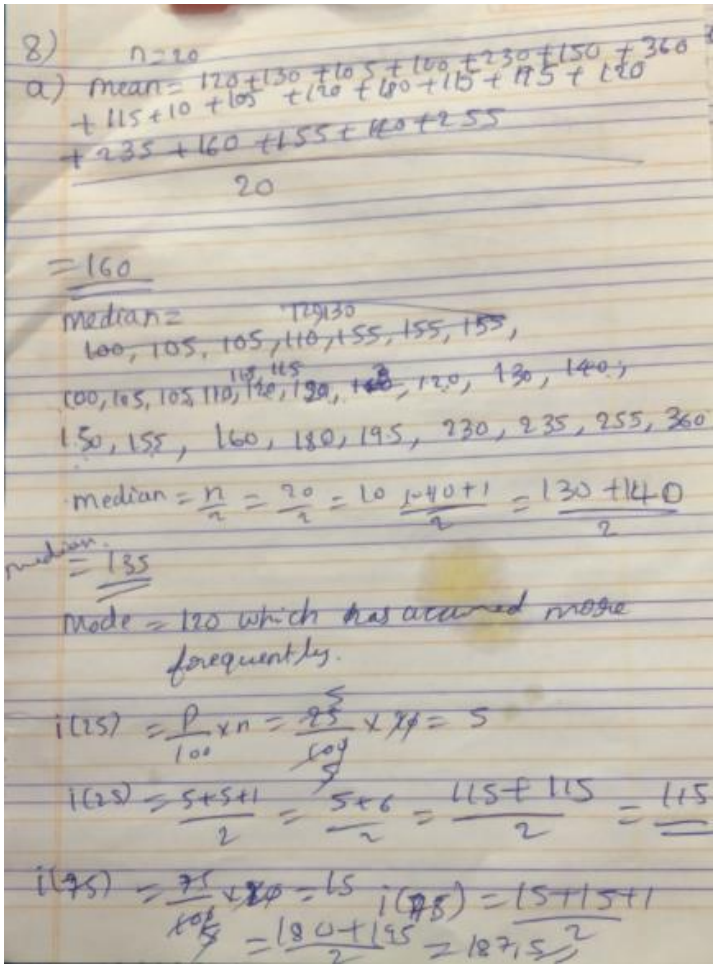
$= 413 - 94.99 = 318.01$

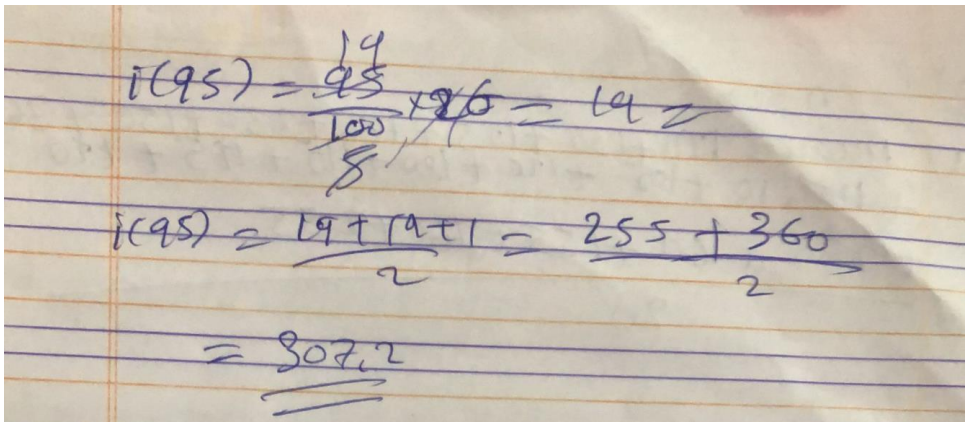
Estimate dollar amount decline in the total endowment held by these 10 universities $= \frac{23 \times 148}{100} = 34.04$

8. The cost of consumer purchases such as single-family housing, gasoline, Internet services, tax preparation, and hospitalization were provided in *The Wall-Street Journal* (January 2, 2007). Sample data typical of the cost of tax-return preparation by services such as H&R Block are shown below.

120	230	110	115	160
130	150	105	195	155
105	360	120	120	140
100	115	180	235	255

a. Compute the mean, median, and mode.
b. Compute the first and third quartiles.
c. Compute and interpret the 90th percentile.

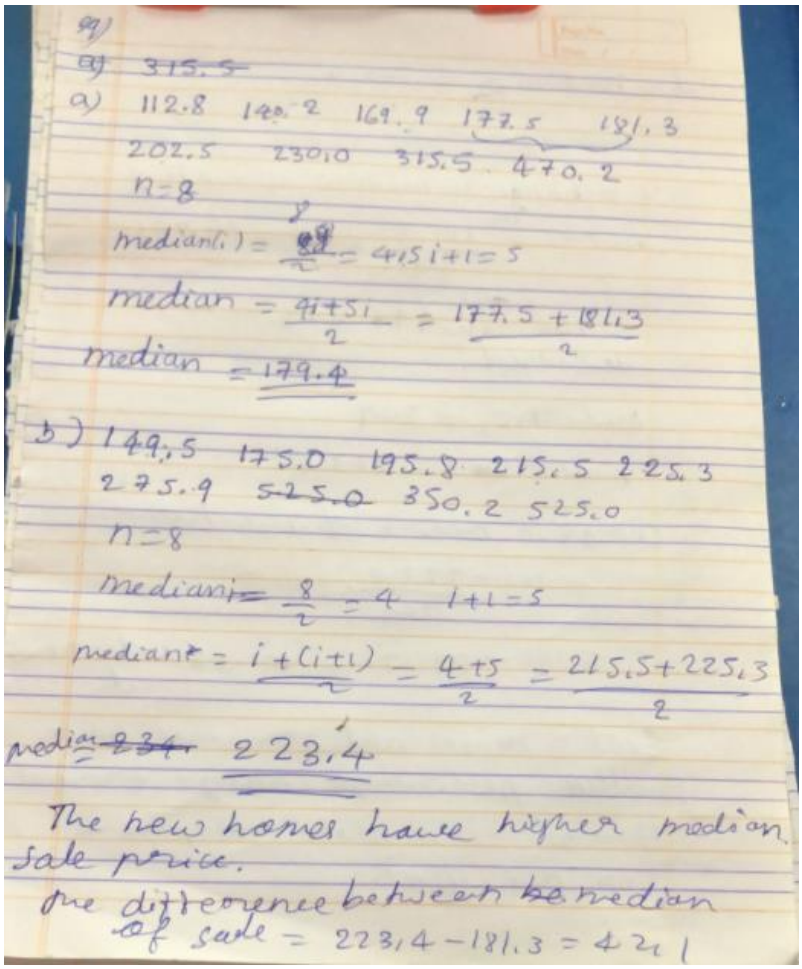


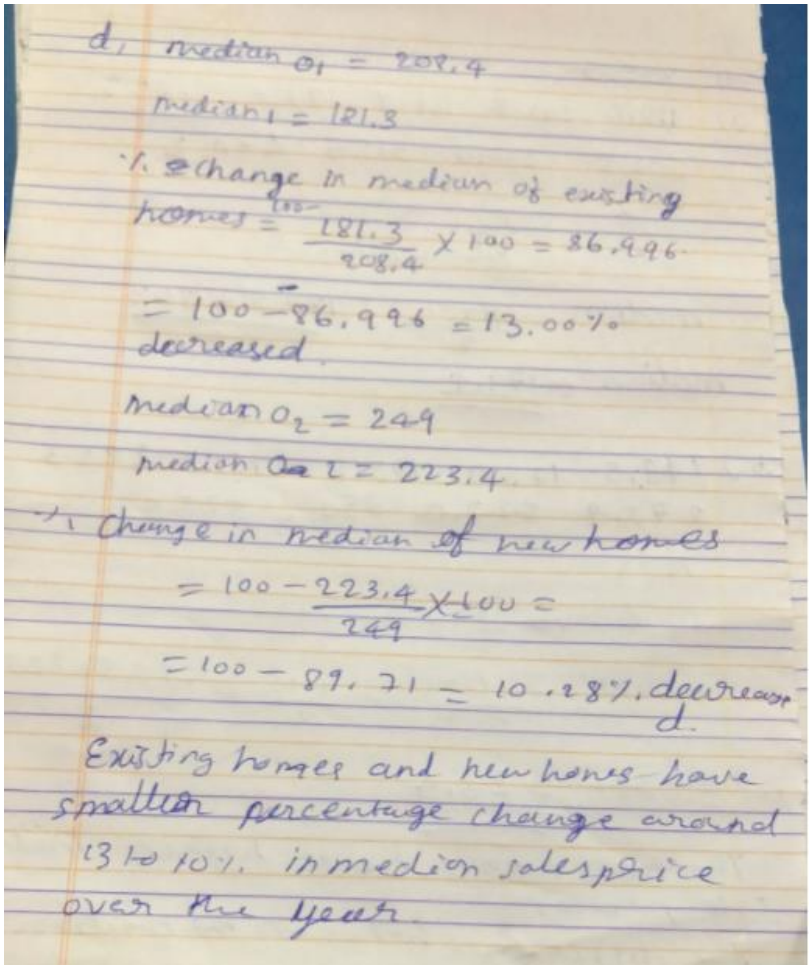


9. The National Association of Realtors provided data showing that home sales were the slowest in 10 years (Associated Press, December 24, 2008). Sample data with representative sales prices for existing homes and new homes follow. Data are in thousands of dollars:

Existing Homes	315.5	202.5	140.2	181.3	470.2	169.9	112.8	230.0	177.5
New Homes	275.9	350.2	195.8	525.0	225.3	215.5	175.0	149.5	

a. What is the median sales price for existing homes?
b. What is the median sales price for new homes?
c. Do existing homes or new homes have the higher median sales price? What is the difference between the median sales prices?
d. A year earlier the median sales price for existing homes was $208.4 thousand and the median sales price for new homes was $249 thousand. Compute the percentage change in the median sales price of existing and new homes over the one-year period. Did existing homes or new homes have the larger percentage change in median sales price?





10. A panel of economists provided forecasts of the U.S. economy for the first six months of 2007 (*The Wall Street Journal,* January 2, 2007). The percent changes in the gross domestic product (GDP) forecasted by 30 economists are as follows.

2.6	3.1	2.3	2.7	3.4	0.9	2.6	2.8	2.0	2.4
2.7	2.7	2.7	2.9	3.1	2.8	1.7	2.3	2.8	3.5
0.4	2.5	2.2	1.9	1.8	1.1	2.0	2.1	2.5	0.5

a. What is the minimum forecast for the percent change in the GDP? What is the maximum?
b. Compute the mean, median, and mode.
c. Compute the first and third quartiles.
d. Did the economists provide an optimistic or pessimistic outlook for the U.S. economy? Discuss.

10

a) minimum forecast for the % change in GDP = 0.4

Maximum forecast % change in GDP = 3.5

b) mean = $\frac{2.6 + 2.7 + 0.4 + \ldots\ldots + 2.5}{30}$

$\bar{x} = \frac{64.1}{30} = 2.13$

median =

0.4, 0.5, 0.9, 1.1, 1.7, 1.8, 1.9, 2.0, 2.0, 2.1, 2.2, 2.3, 2.3, 2.4, 2.5, 2.5, 2.6, 2.6 2.7, 2.7, 2.7, 2.7, 2.8, 2.8, 2.8, 2.9 3.1, 3.1, 3.4, 3.5

$i = \frac{n}{2} = \frac{30}{2} = 15, \quad 1+i = 16.$

median = $\frac{15i + 16i}{2} = \frac{2.5 + 2.5}{2} = 2.5$

mode = 2.7, which has occurred more number of times,

1st quartile (i) = $\frac{P}{100} \times n = \frac{25}{100} \times 30 = 7.5 = 8$

1st quartile (25) = 8i = 2.0

3rd quartile (75) = $\frac{P}{100} \times n = \frac{75}{100} \times 30 = 22.5 = 23$

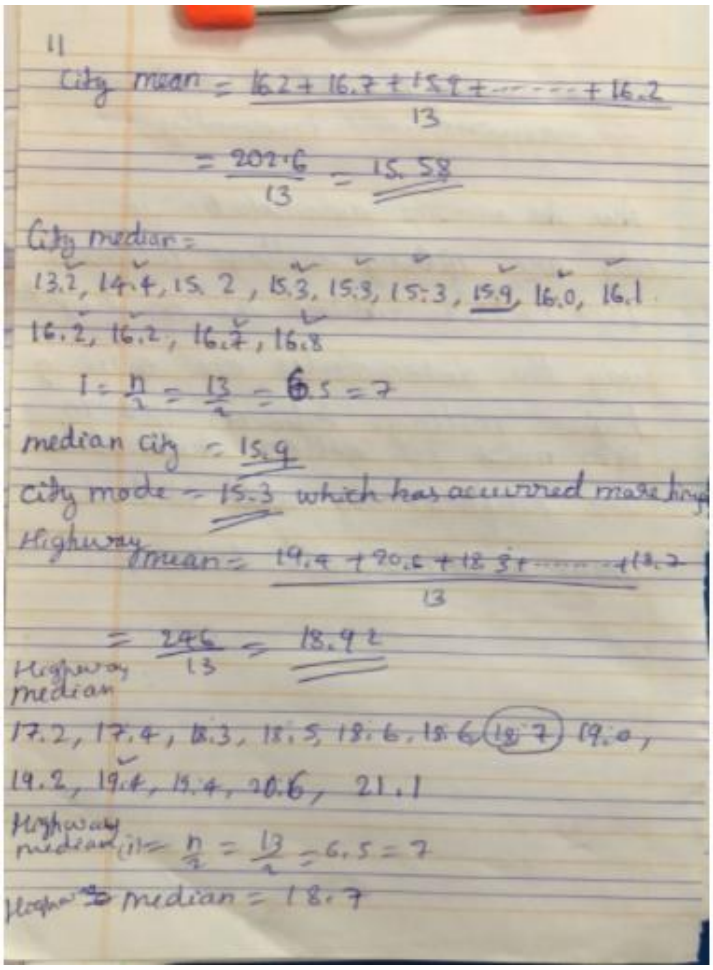
3rd quartile (75) = 23i = 2.8

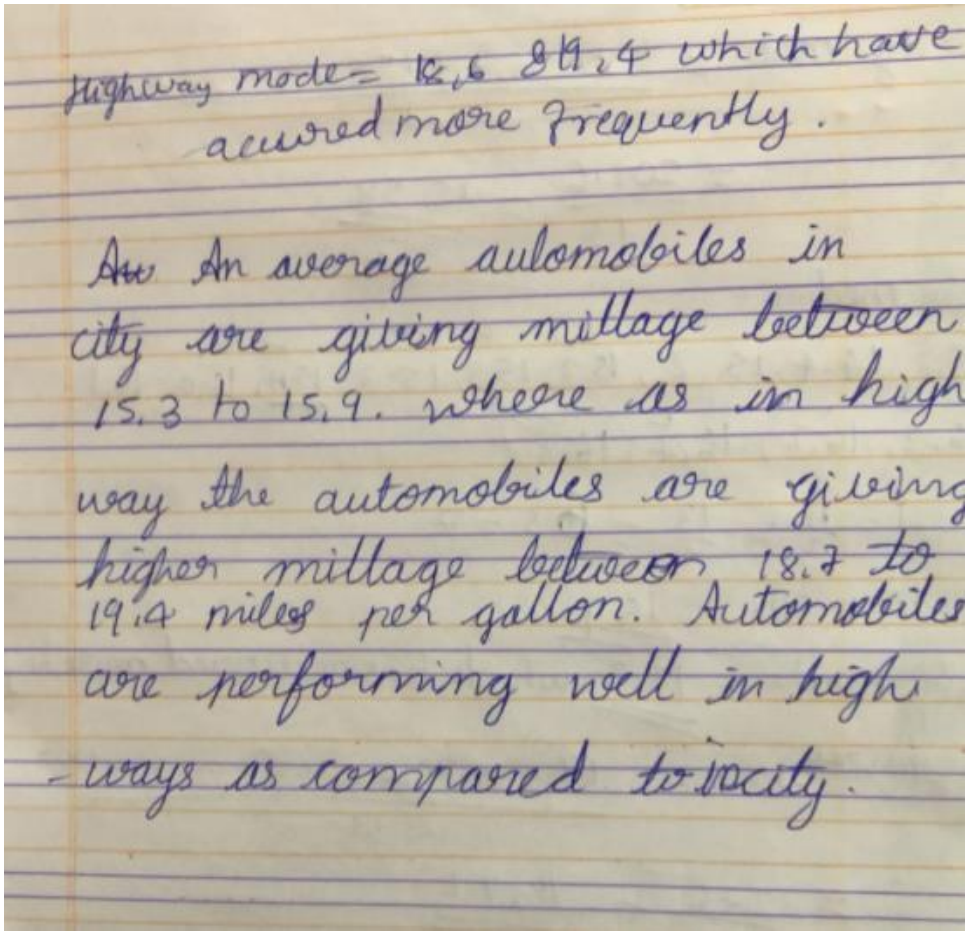
d) Optimistic, because all percentage are positive. which indicates an increase in the GDP. Negative precentage would correspond with a decrease in the GDP (Gross Domestic Product)

11. In automobile mileage and gasoline-consumption testing, 13 automobiles were road tested for 300 miles in both city and highway driving conditions. The following data were recorded for miles-per-gallon performance.

City:	16.2	16.7	15.9	14.4	13.2	15.3	16.8	16.0	16.1	15.3	15.2	15.3	16.2
Highway:	19.4	20.6	18.3	18.6	19.2	17.4	17.2	18.6	19.0	21.1	19.4	18.5	18.7

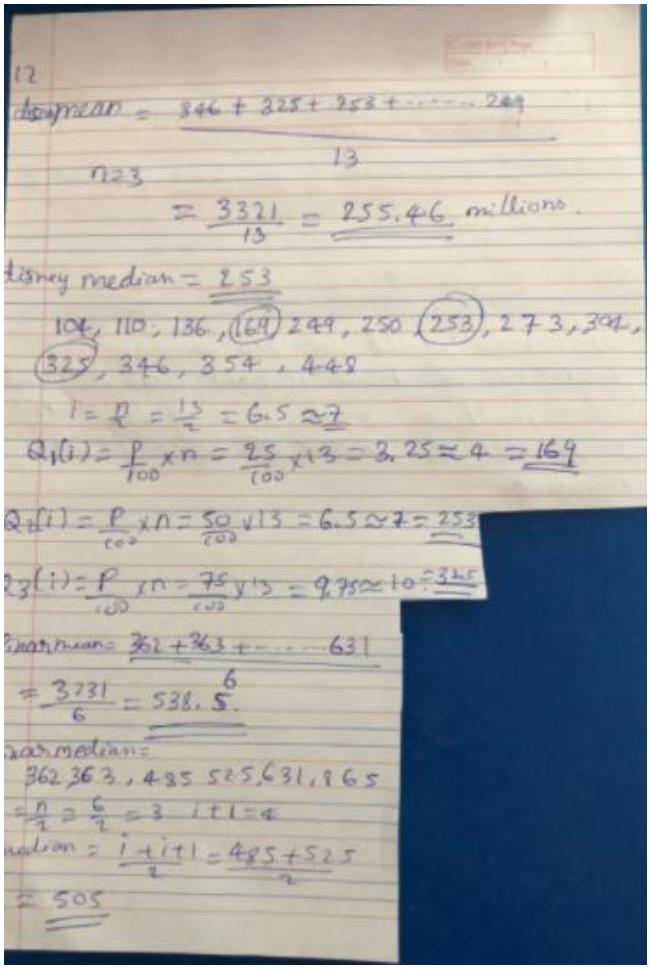
Use the mean, median, and mode to make a statement about the difference in performance for city and highway driving.





12. Walt Disney Company bought Pixar Animation Studios, Inc., in a deal worth $7.4 billion (CNN Money website, January 24, 2006). The animated movies produced by Disney and Pixar during the previous 10 years are listed in the following table. The box office revenues are in millions of dollars. Compute the total revenue, the mean, the median, and the quartiles to compare the box office success of the movies produced by both companies. Do the statistics suggest at least one of the reasons Disney was interested in buying Pixar? Discuss.

Disney Movies	Revenue ($millions)	Pixar Movies	Revenue ($millions)
Pocahontas	346	*Toy Story*	362
Hunchback of Notre Dame	325	*A Bug's Life*	363
Hercules	253	*Toy Story 2*	485
Mulan	304	*Monsters, Inc.*	525
Tarzan	448	*Finding Nemo*	865
Dinosaur	354	*The Incredibles*	631
The Emperor's New Groove	169		
Lilo & Stitch	273		
Treasure Planet	110		
The Jungle Book 2	136		
Brother Bear	250		
Home on the Range	104		
Chicken Little	249		



12

disney mean = $\frac{346 + 325 + 253 + \cdots + 249}{13}$

$n = 13$

$= \frac{3321}{13} = 255.46$ millions.

disney median = 253

104, 110, 136, (169) 249, 250 (253), 273, 304, (325), 346, 354, 448

$i = \frac{n}{2} = \frac{13}{2} = 6.5 \simeq 7$

$Q_1(i) = \frac{P}{100} \times n = \frac{25}{100} \times 13 = 3.25 \simeq 4 = 169$

$Q_2(i) = \frac{P}{100} \times n = \frac{50}{100} \times 13 = 6.5 \simeq 7 = 253$

$Q_3(i) = \frac{P}{100} \times n = \frac{75}{100} \times 13 = 9.75 \simeq 10 = 325$

pixar mean = $\frac{362 + 363 + \cdots 631}{6}$

$= \frac{3731}{6} = 538.5$

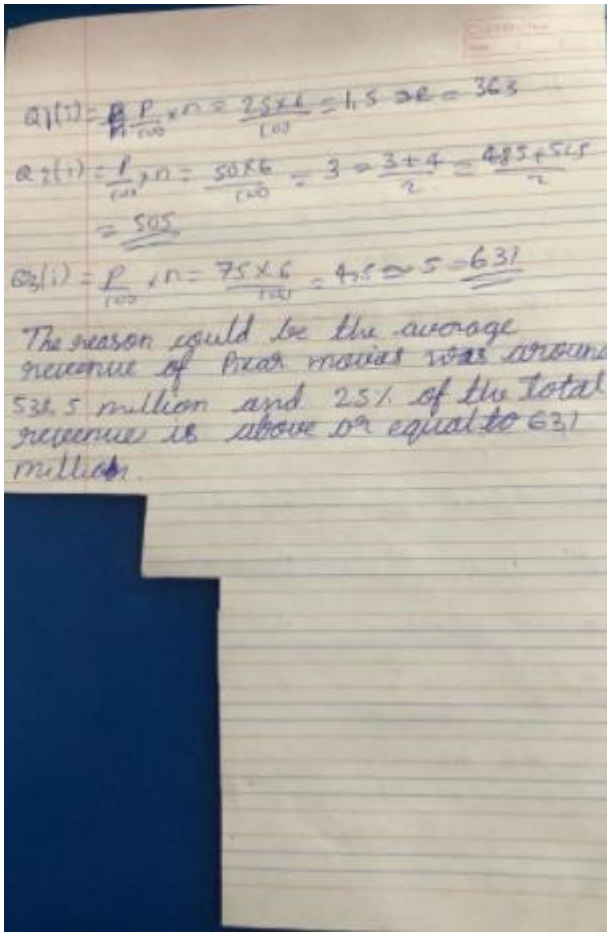
pixar median =

362, 363, 485, 525, 631, 865

$= \frac{n}{2} = \frac{6}{2} = 3 \quad i+1 = 4$

median = $\frac{i + i+1}{2} = \frac{485 + 525}{2}$

$= 505$



$Q_1(i) = \frac{P}{100} \times n = \frac{25 \times 6}{100} = 1.5 \simeq 2 = 363$

$Q_2(i) = \frac{P}{100} \times n = \frac{50 \times 6}{100} = 3 = \frac{3+4}{2} = \frac{485 + 525}{2}$

$= 505$

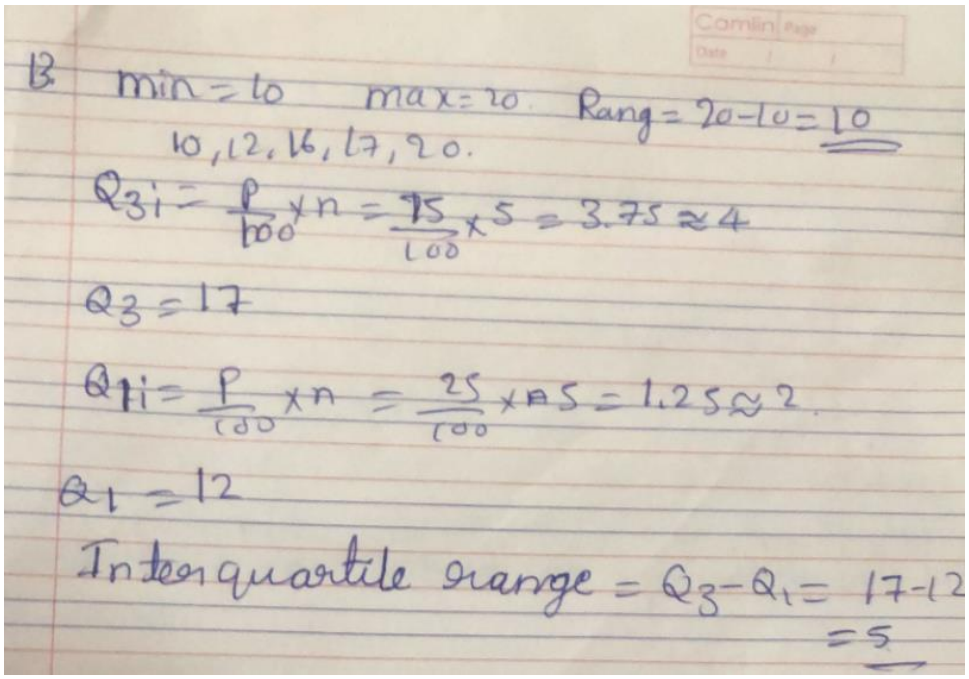
$Q_3(i) = \frac{P}{100} \times n = \frac{75 \times 6}{100} = 4.5 \simeq 5 = 631$

The reason could be the average revenue of Pixar movies was around 538.5 million and 25% of the total revenues is above or equal to 631 million.

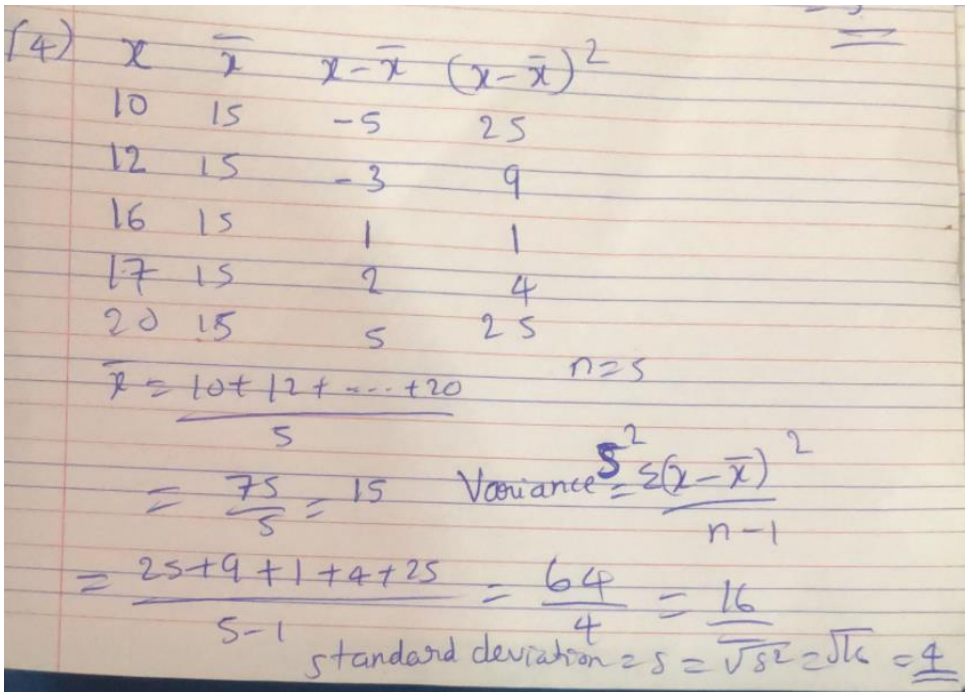
Exercises

Methods

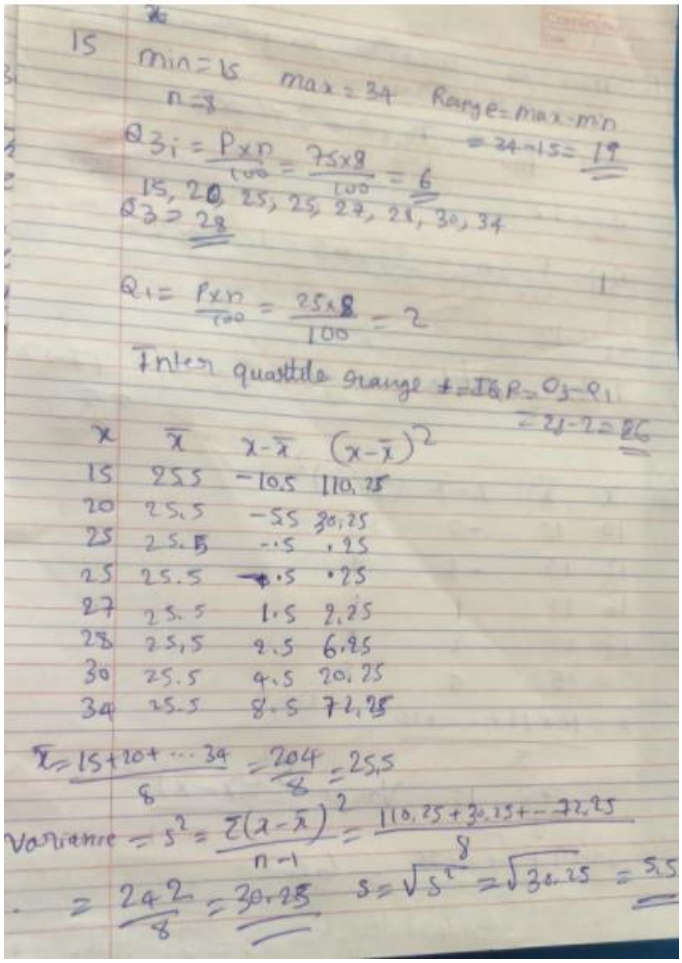
13. Consider a sample with data values of 10, 20, 12, 17, and 16. Compute the range and interquartile range.



14. Consider a sample with data values of 10, 20, 12, 17, and 16. Compute the variance and standard deviation.



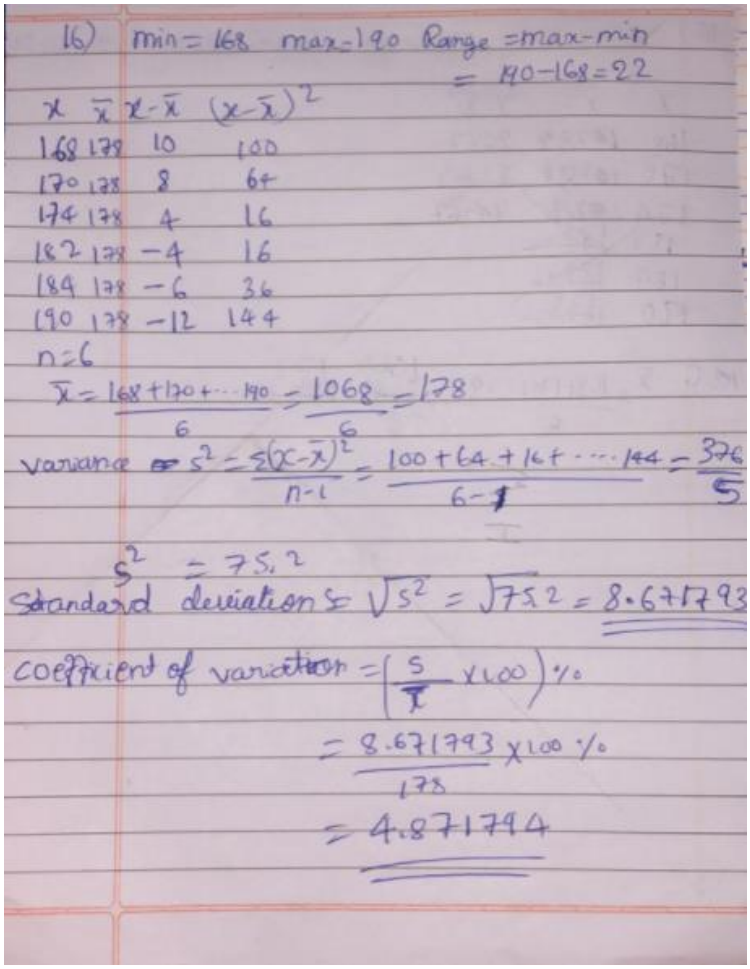
15. Consider a sample with data values of 27, 25, 20, 15, 30, 34, 28, and 25. Compute the range, interquartile range, variance, and standard deviation.



x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
15	25.5	−10.5	110.25
20	25.5	−5.5	30.25
25	25.5	−.5	.25
25	25.5	−.5	.25
27	25.5	1.5	2.25
28	25.5	2.5	6.25
30	25.5	4.5	20.25
34	25.5	8.5	72.25

Applications

16. A bowler's scores for six games were 182, 168, 184, 190, 170, and 174. Using these data as a sample, compute the following descriptive statistics:
    a. Range
    b. Variance
    c. Standard deviation
    d. Coefficient of variation



17. A home theater in a box is the easiest and cheapest way to provide surround sound for a home entertainment center. A sample of prices is shown here (*Consumer Reports Buying Guide,* 2004). The prices are for models with a DVD player and for models without a DVD player.

Models with DVD Player	Price	Models without DVD Player	Price
Sony HT-1800DP	$450	Pioneer HTP-230	$300
Pioneer HTD-330DV	300	Sony HT-DDW750	300
Sony HT-C800DP	400	Kenwood HTB-306	360
Panasonic SC-HT900	500	RCA RT-2600	290
Panasonic SC-MTI	400	Kenwood HTB-206	300

a. Compute the mean price for models with a DVD player and the mean price for models without a DVD player. What is the additional price paid to have a DVD player included in a home theater unit?

b. Compute the range, variance, and standard deviation for the two samples. What does this information tell you about the prices for models with and without a DVD player?

17 $n_1 = 5$ $n_2 = 5$

$\bar{x}_1 = \frac{450 + 300 + \cdots 500}{5}$ $\bar{x}_2 = \frac{300 + 360 + \cdots 300}{5}$

$\bar{x}_1 = \frac{2050}{5} = 410$ $\bar{x}_2 = \frac{1550}{5} = 310$

x_1	$\bar{x}_1$	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	x_2	$\bar{x}_2$	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$
300	410	-110	12100	290	310	-20	400
400	410	-10	100	300	310	-10	100
400	410	-10	100	300	310	-10	100
450	410	40	1600	300	310	-10	100
500	410	90	8100	360	310	50	2500

Variance $S_1^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2}{n-1}$ Variance $S_2^2 = \frac{\Sigma (x_2 - \bar{x}_2)^2}{n-1}$

$= \frac{12100 + 100 + \cdots 8100}{5-1}$ $= \frac{400 + 100 + \cdots 2500}{5-1}$

$S_1^2 = \frac{22000}{4} = 5500$ $S_2^2 = \frac{3200}{4} = 800$

Standard deviation $S_1 = \sqrt{S_1^2} = \sqrt{5500}$ $S_2 = \sqrt{S_2^2} = \sqrt{800} =$

$S_1 = 74.16$ $S_2 = 28.28$

Coefficient of variance 1 coefficient of variance 2

$= \frac{S_1}{\bar{x}_1} \times 100\%$ $= \frac{S_2}{\bar{x}_2} \times 100\%$

$= \frac{74.16}{400} \times 100 = 18.03\%$ $= \frac{28.28}{310} \times 100 = 9.1239$

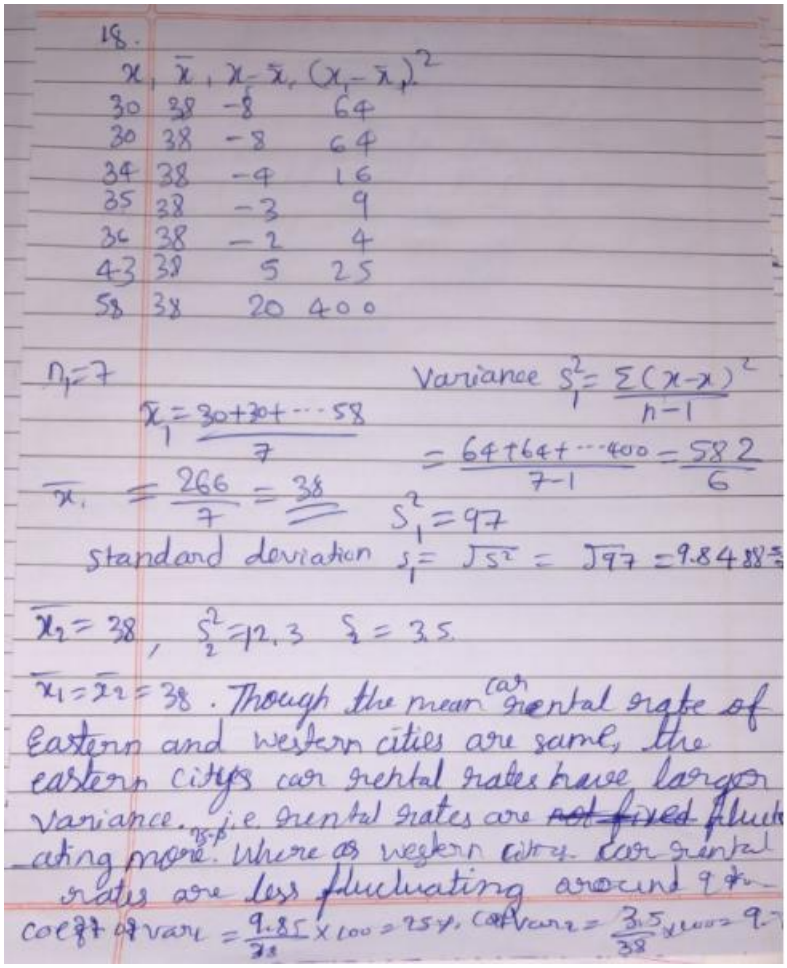
Page No.
Date / /

The prices of the models with DVD players have large variance to its mean or average price around 18 %. of standard deviation to its mean. While the prices of the model with DVD player have very less variance its mean price around 9% of standard deviation to its mean.

18. Car rental rates per day for a sample of seven Eastern U.S. cities are as follows (*The Wall Street Journal*, January 16, 2004).

City	Daily Rate
Boston	$43
Atlanta	35
Miami	34
New York	58
Orlando	30
Pittsburgh	30
Washington, D.C.	36

a. Compute the mean, variance, and standard deviation for the car rental rates.
b. A similar sample of seven Western U.S. cities showed a sample mean car rental rate of $38 per day. The variance and standard deviation were 12.3 and 3.5, respectively. Discuss any difference between the car rental rates in Eastern and Western U.S. cities.



19. The *Los Angeles Times* regularly reports the air quality index for various areas of Southern California. A sample of air quality index values for Pomona provided the following data: 28, 42, 58, 48, 45, 55, 60, 49, and 50.
a. Compute the range and interquartile range.
b. Compute the sample variance and sample standard deviation.
c. A sample of air quality index readings for Anaheim provided a sample mean of 48.5, a sample variance of 136, and a sample standard deviation of 11.66. What comparisons can you make between the air quality in Pomona and that in Anaheim on the basis of these descriptive statistics?

19

a) min = 28 max = 60 Range = max − min = 60 − 28 = 32

n = 9

$Q_1 = \frac{P}{100} \times n = \frac{25}{100} \times 9 = 2.25 \approx 3$

28, 42, 45, 48, 49, 50, 55, 58, 60.

$Q_1 = 45$

$Q_3 = \frac{P}{100} \times n = \frac{75}{100} \times 9 = 6.75 \approx 7 = 55$

x_i	$\bar{x}_1$	$x_i - \bar{x}_1$	$(x_i - \bar{x}_1)^2$
28	48.33	−20.33	413.30
42	48.33	−6.33	40.06
45	48.33	−3.33	11.08
48	48.33	−0.33	0.10
49	48.33	0.67	0.44
50	48.33	1.67	2.78
55	48.33	6.67	44.48
58	48.33	9.67	93.50
60	48.33	11.67	136.18

n = 9 $\bar{x}_1 = \frac{28 + 42 + \cdots + 60}{9}$

$= \frac{435}{9} = 48.33$

b) Variance $s_1^2 = \frac{\sum (x_i - \bar{x}_1)^2}{n-1}$

$= \frac{413.30 + 40.06 + \cdots + 136.18}{9-1}$

$s_1^2 = \frac{741.92}{8} = 92.74$

b) va

Standard deviation $s_1 = \sqrt{s^2} = \sqrt{92.74} = 9.63$

c) $\bar{x}_2 = 48.5$, $S_2^2 = 136$. $S_2 = 11.66$.

$\bar{x}_1 = 48.33$ $S_1^2 = 92.74$ $S_1 = 9.63$

Coefficient of variance$_1$ $= \frac{S_1}{\bar{x}_1} \times 100 = \frac{9.63}{48.3} \times 100\%$

Coeff Var = 19.93 %

coefficient of variance$_2$ $= \frac{S_2}{\bar{x}_2} \times 100 = \frac{11.66}{48.5} \times 100\%$

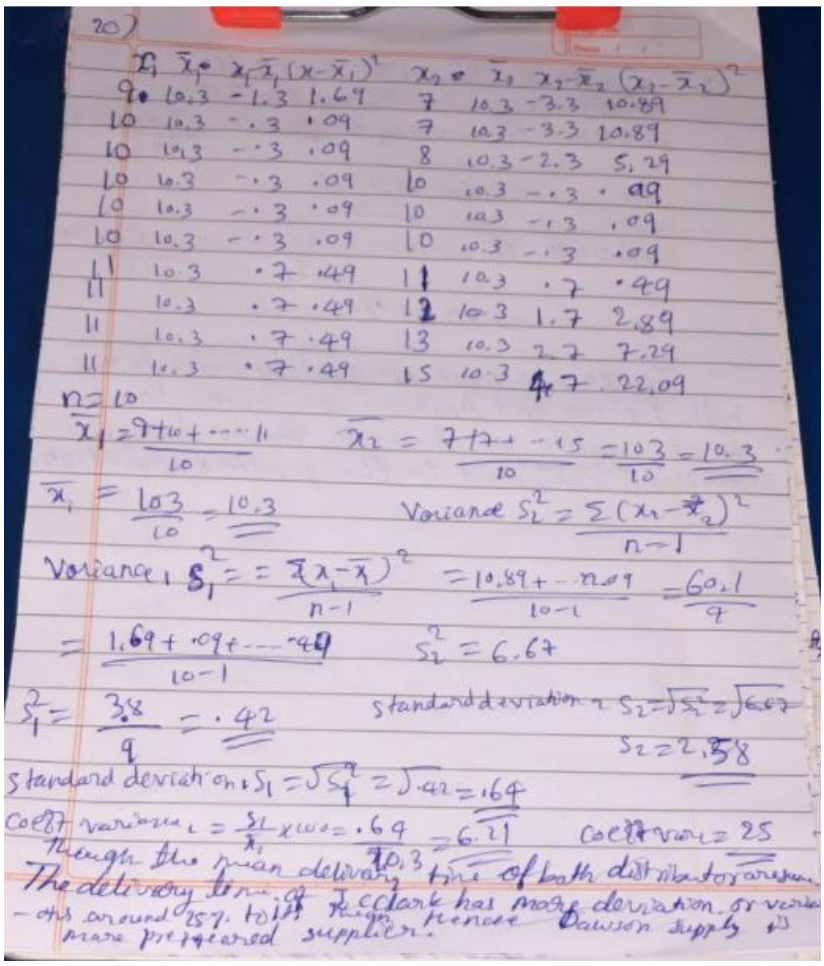
$= 24.04\%$

Though the mean air quality index for both Pomona and Anaheim are nearly same. the air quality of Anaheim has large variance compared to the pona around (19.9% through 24.1%

20. The following data were used to construct the histograms of the number of days required to fill orders for Dawson Supply, Inc., and J.C. Clark Distributors (see Figure 3.2).

Dawson Supply Days for Delivery:	11	10	9	10	11	11	10	11	10	10
Clark Distributors Days for Delivery:	8	10	13	7	10	11	10	7	15	12

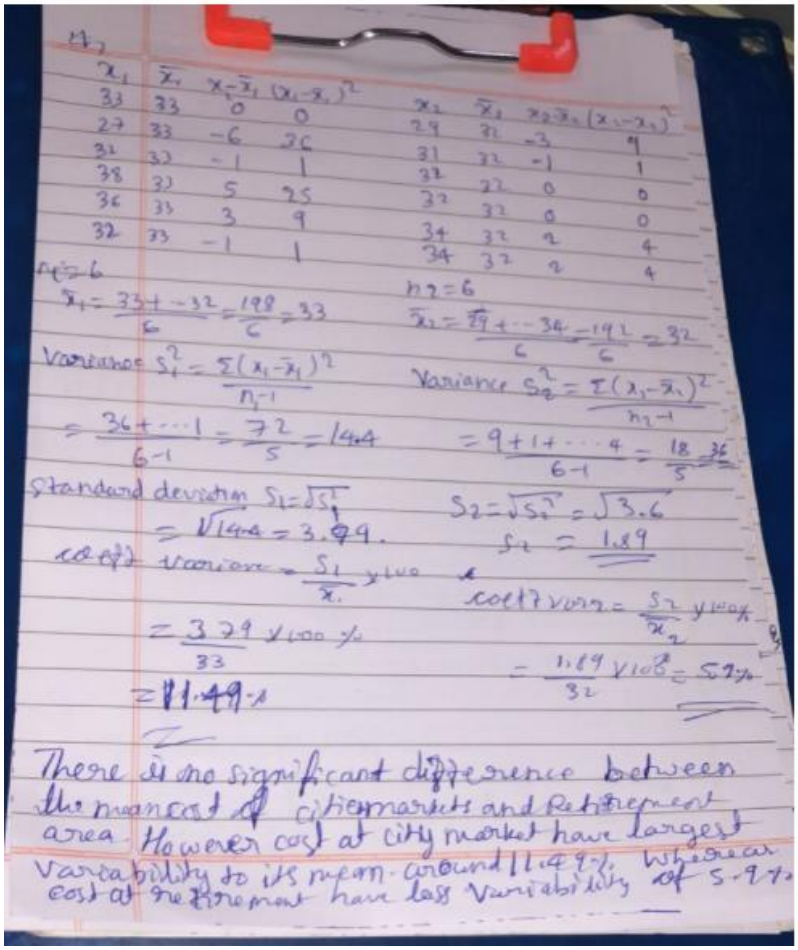
Use the range and standard deviation to support the previous observation that Dawson Supply provides the more consistent and reliable delivery times.



21. How do grocery costs compare across the country? Using a market basket of 10 items including meat, milk, bread, eggs, coffee, potatoes, cereal, and orange juice, *Where to Retire* magazine calculated the cost of the market basket in six cities and in six retirement areas across the country (*Where to Retire,* November/December 2003). The data with market basket cost to the nearest dollar are as follows:

City	Cost	Retirement Area	Cost
Buffalo, NY	$33	Biloxi-Gulfport, MS	$29
Des Moines, IA	27	Asheville, NC	32
Hartford, CT	32	Flagstaff, AZ	32
Los Angeles, CA	38	Hilton Head, SC	34
Miami, FL	36	Fort Myers, FL	34
Pittsburgh, PA	32	Santa Fe, NM	31

a. Compute the mean, variance, and standard deviation for the sample of cities and the sample of retirement areas.
b. What observations can be made based on the two samples?

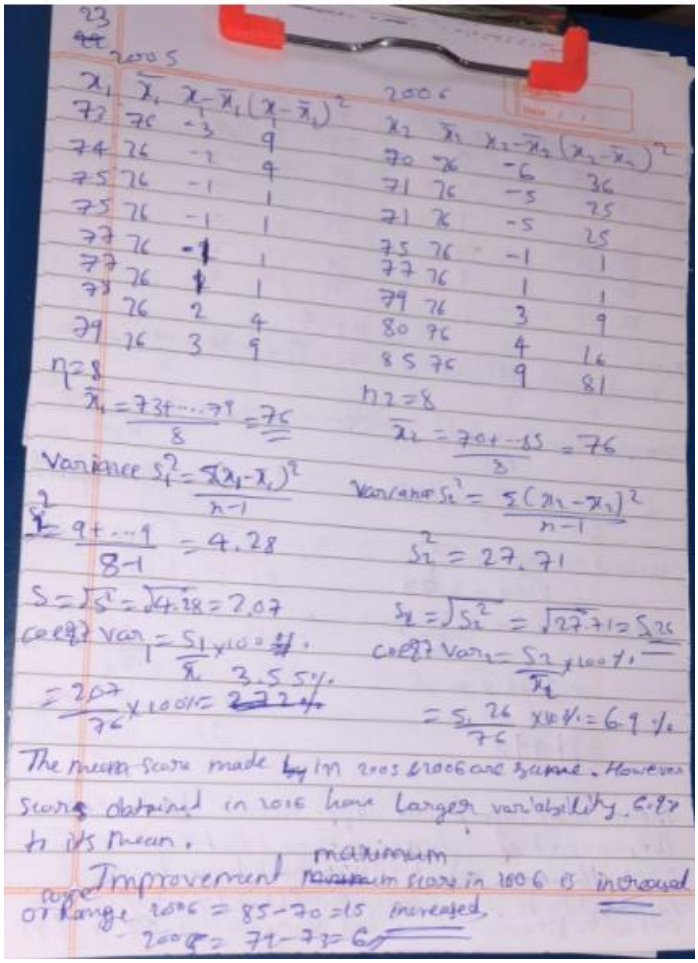


23. Scores turned in by an amateur golfer at the Bonita Fairways Golf Course in Bonita Springs, Florida, during 2005 and 2006 are as follows:

2005 Season:	74	78	79	77	75	73	75	77
2006 Season:	71	70	75	77	85	80	71	79

a. Use the mean and standard deviation to evaluate the golfer's performance over the two-year period.

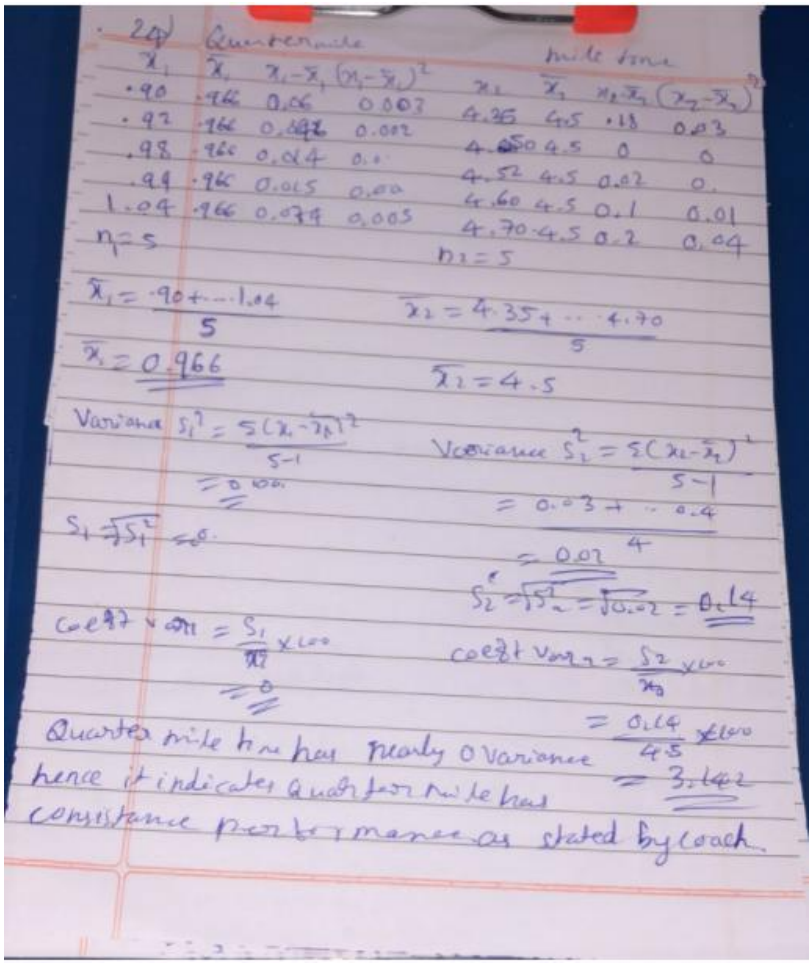
b. What is the primary difference in performance between 2005 and 2006? What improvement, if any, can be seen in the 2006 scores?



24. The following times were recorded by the quarter-mile and mile runners of a university track team (times are in minutes).

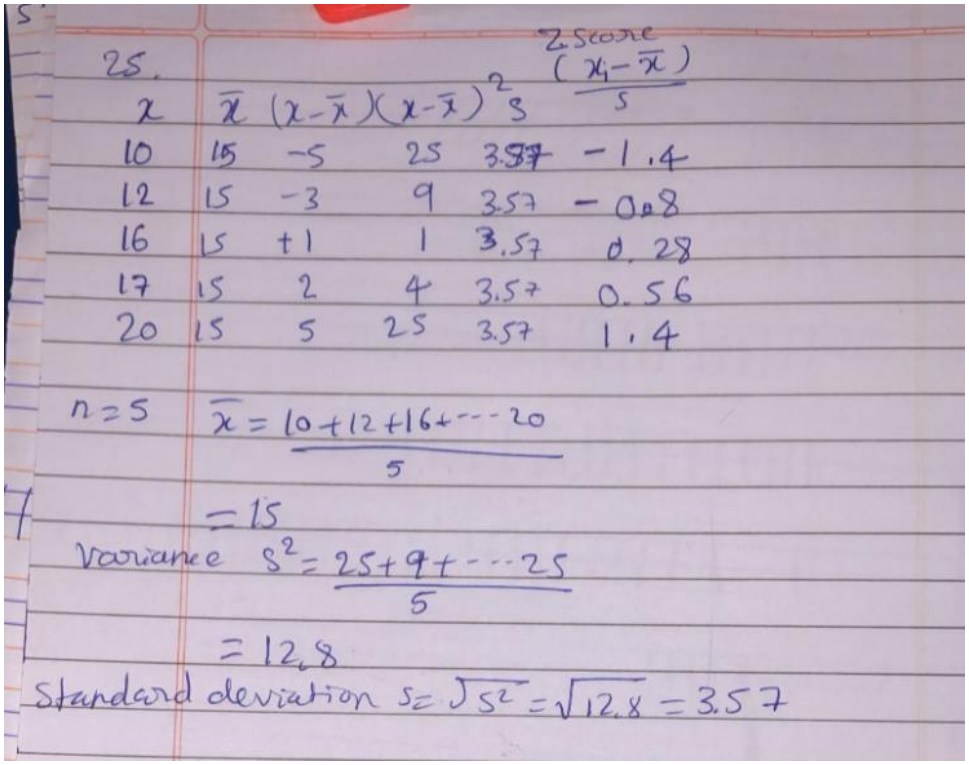
Quarter-Mile Times:	.92	.98	1.04	.90	.99
Mile Times:	4.52	4.35	4.60	4.70	4.50

After viewing this sample of running times, one of the coaches commented that the quarter-milers turned in the more consistent times. Use the standard deviation and the coefficient of variation to summarize the variability in the data. Does the use of the coefficient of variation indicate that the coach's statement should be qualified?

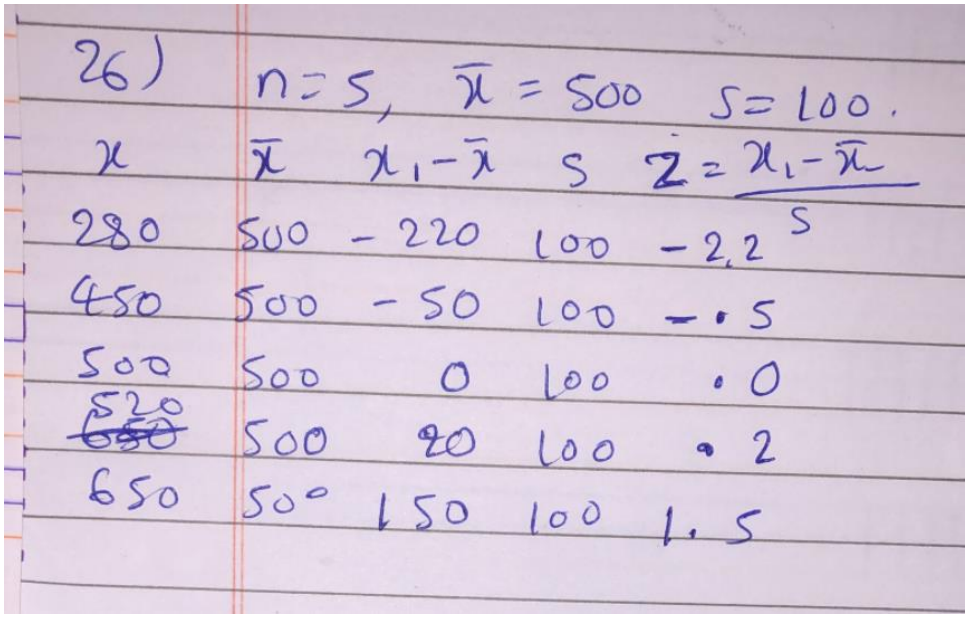


Methods

25. Consider a sample with data values of 10, 20, 12, 17, and 16. Compute the z-score for each of the five observations.



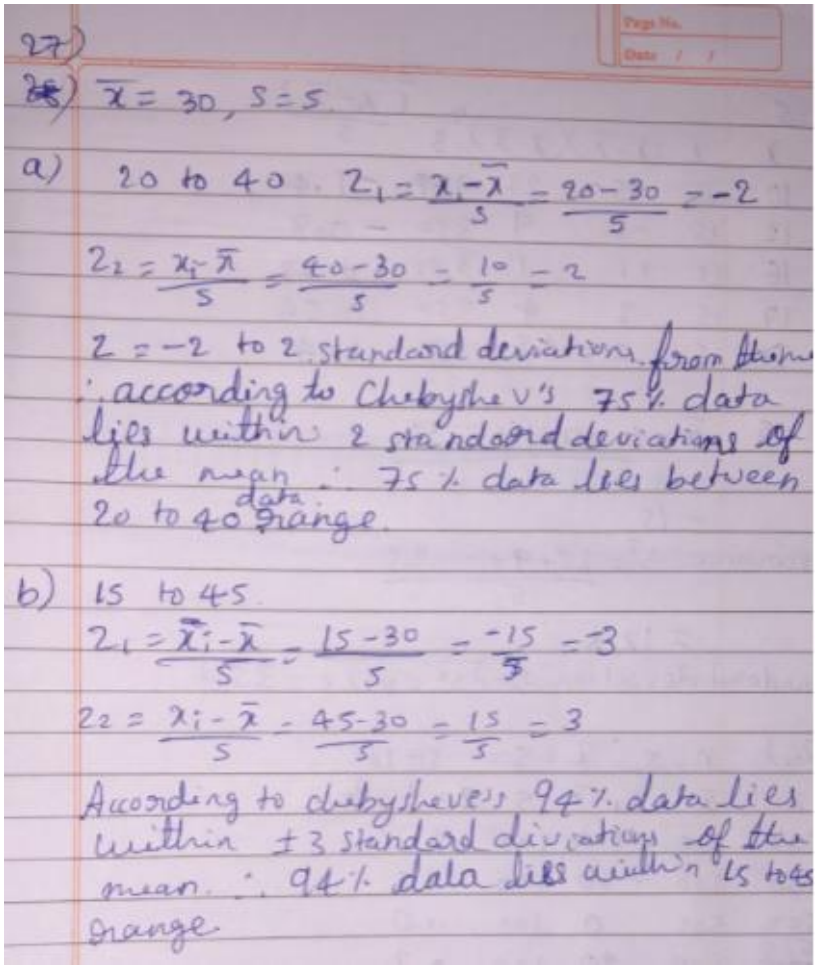
26. Consider a sample with a mean of 500 and a standard deviation of 100. What are the z-scores for the following data values: 520, 650, 500, 450, and 280?



26) $n = 5$, $\bar{x} = 500$ $s = 100$.

x	$\bar{x}$	$x_i - \bar{x}$	s	$z = \frac{x_i - \bar{x}}{s}$
280	500	−220	100	−2.2
450	500	−50	100	−.5
500	500	0	100	.0
~~520~~ ~~650~~	500	20	100	.2
650	500	150	100	1.5

27. Consider a sample with a mean of 30 and a standard deviation of 5. Use Chebyshev's theorem to determine the percentage of the data within each of the following ranges:
    a. 20 to 40
    b. 15 to 45
    c. 22 to 38
    d. 18 to 42
    e. 12 to 48



27)

28) $\bar{x} = 30$, $s = 5$

a) 20 to 40 $z_1 = \frac{x_i - \bar{x}}{s} = \frac{20 - 30}{5} = -2$

$z_2 = \frac{x_i - \bar{x}}{s} = \frac{40 - 30}{5} = \frac{10}{5} = 2$

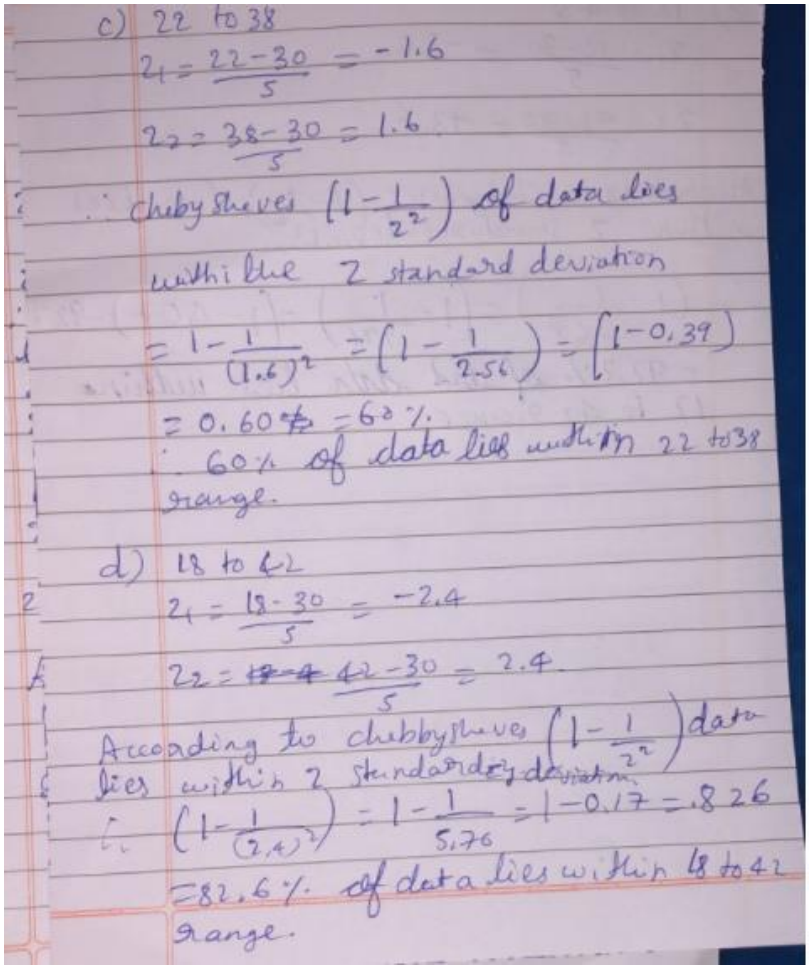
$z = -2$ to 2 standard deviations from the mean. ∴ according to Chebyshev's 75% data lies within 2 standard deviations of the mean ∴ 75% data lies between 20 to 40 data range.

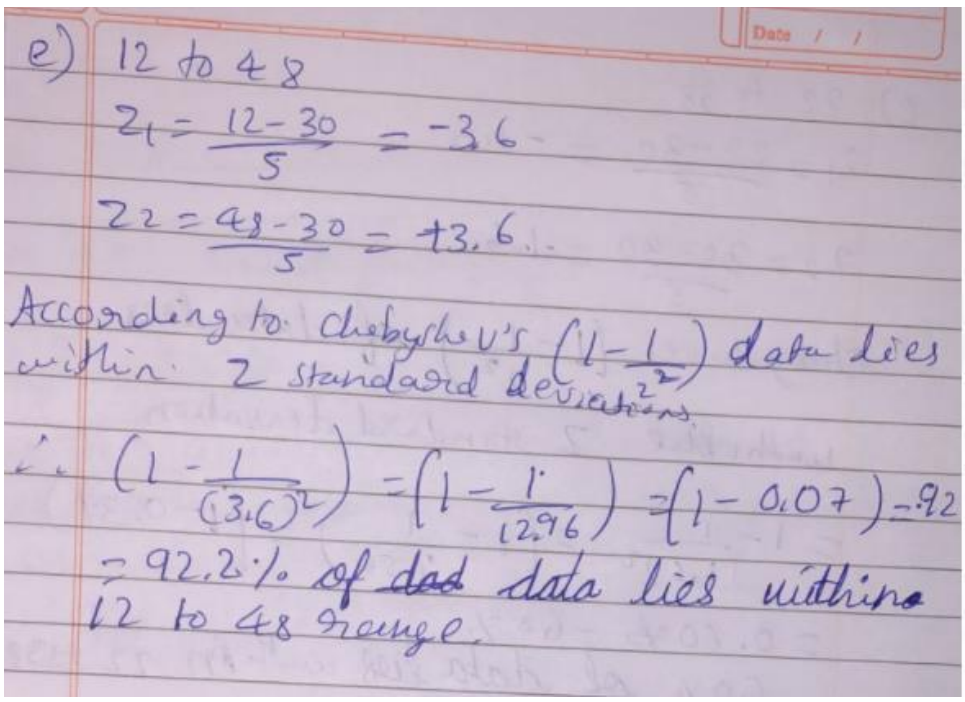
b) 15 to 45

$z_1 = \frac{\bar{x}_i - \bar{x}}{s} = \frac{15 - 30}{5} = \frac{-15}{5} = -3$

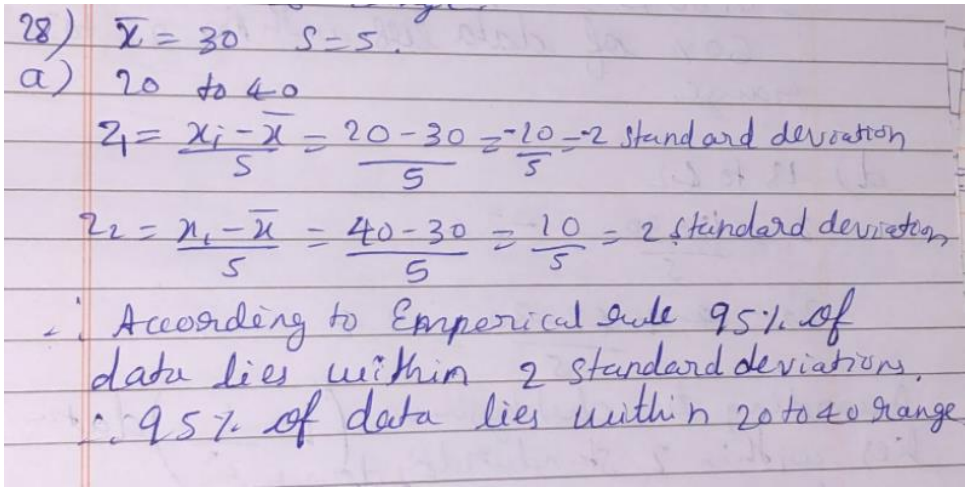
$z_2 = \frac{x_i - \bar{x}}{s} = \frac{45 - 30}{5} = \frac{15}{5} = 3$

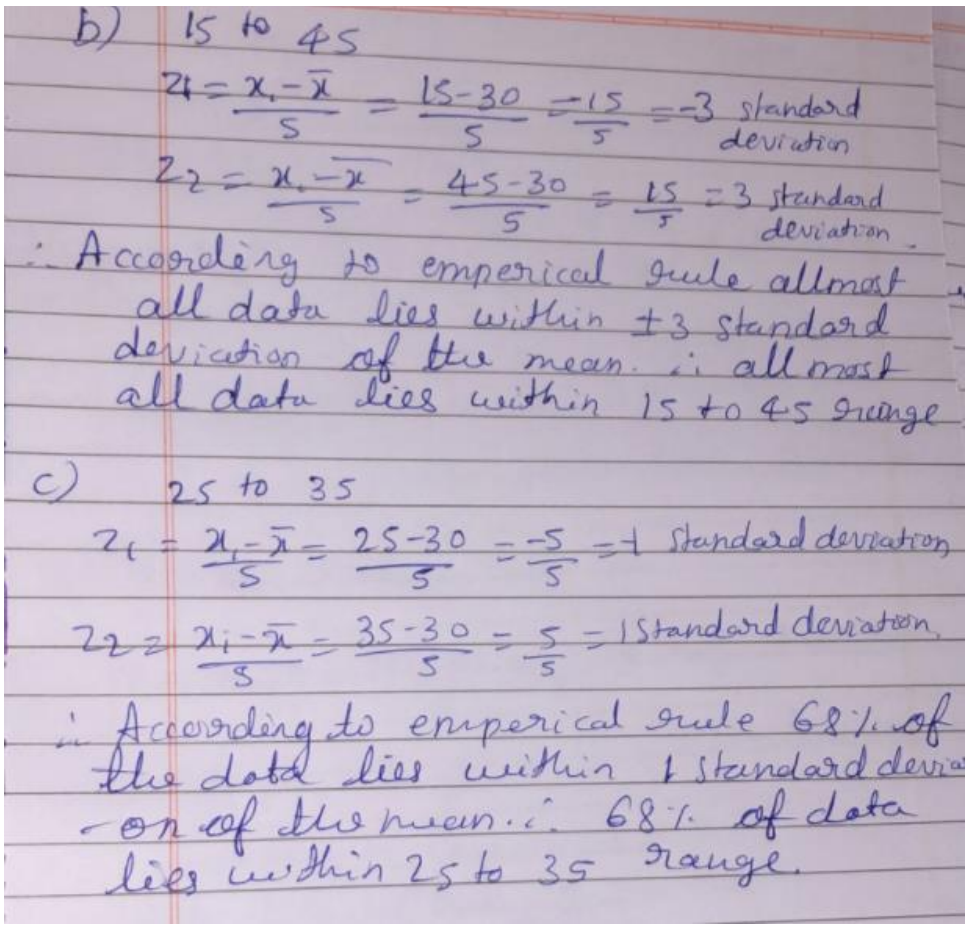
According to chebyshev's 94% data lies within ±3 standard deviations of the mean. ∴ 94% data lies within 15 to 45 range.





28. Suppose the data have a bell-shaped distribution with a mean of 30 and a standard deviation of 5. Use the empirical rule to determine the percentage of data within each of the following ranges:
    a. 20 to 40
    b. 15 to 45
    c. 25 to 35





Applications

29. The results of a national survey showed that on average, adults sleep 6.9 hours per night. Suppose that the standard deviation is 1.2 hours.
    a. Use Chebyshev's theorem to calculate the percentage of individuals who sleep between 4.5 and 9.3 hours.
    b. Use Chebyshev's theorem to calculate the percentage of individuals who sleep between 3.9 and 9.9 hours.
    c. Assume that the number of hours of sleep follows a bell-shaped distribution. Use the empirical rule to calculate the percentage of individuals who sleep between 4.5 and 9.3 hours per day. How does this result compare to the value that you obtained using Chebyshev's theorem in part (a)?

29. $\bar{x} = 6.9 \quad s = 1.2$.

a) 4.5 to 9.3 hours

$z_1 = \frac{x_i - \bar{x}}{s} = \frac{4.5 - 6.9}{1.2} = -2$ standard deviation

$z_2 = \frac{9.3 - 6.9}{1.2} = 2$ standard deviation

∴ According to chebby'shevs 75% data lies within ±2 standard deviation of the mean.

∴ 75% of individuals sleep between 4.5 to and 9.3 hours.

b) 3.9 to 9.9.

$z_1 = \frac{x_i - \bar{x}}{s} = \frac{3.9 - 6.9}{1.2} = -2.5$ standard deviation

$z_2 = \frac{9.9 - 6.9}{1.2} = +2.5$ standard deviation.

According to chebyshevs $(1 - \frac{1}{2^2})$ % of data lies within 2 standard deviation

$\therefore (1 - \frac{1}{(2.5)^2}) = (1 - \frac{1}{6.25}) = (1 - 0.16) = 184$

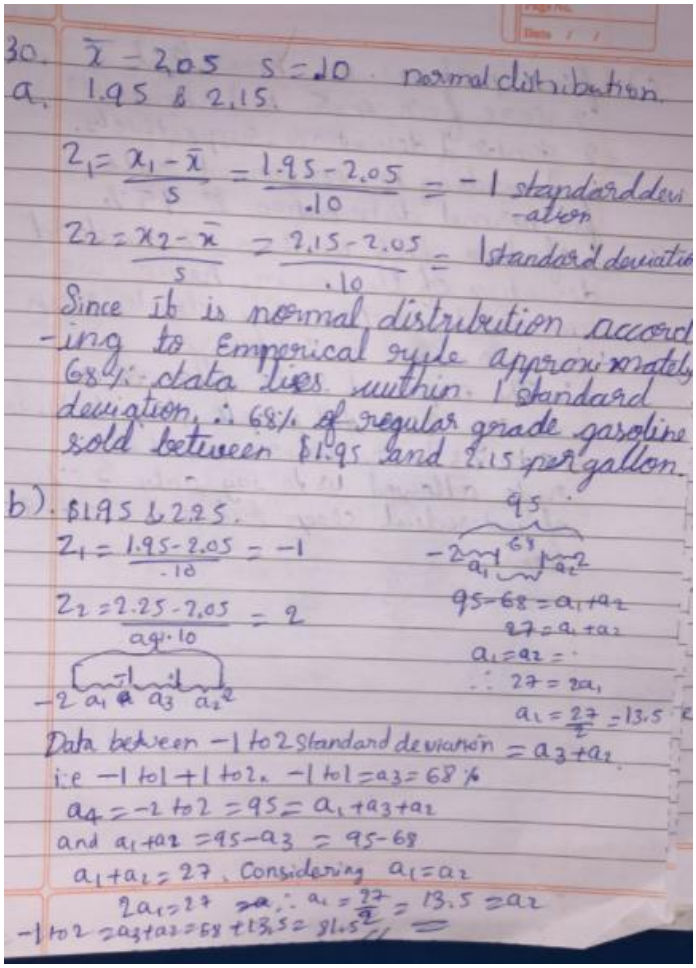
=84% of people sleep between 3.9 and 9.9 hours.

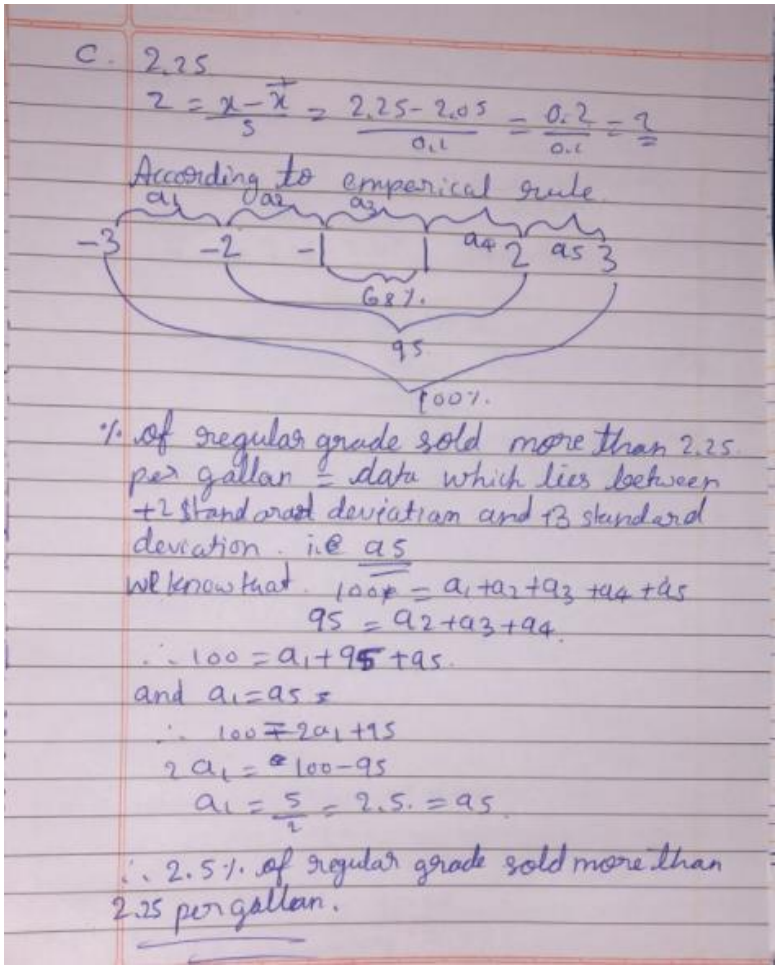
c) from part (a) we know that 2 store for 4-5 to 9.3 hours as ±2 standard deviation. respectively.

∴ According to emperical rule for normal distribution 95% of data lies within ±2 standard deviations of the mean hence we can say that 95% individuals sleep between 4-5 to 9.3 hours.

∴ The emperical rule allowed us to say 95% people sleep between 4.5 to 9.3 hours where as cheby's rule allowed us to say only 75% of individual sleep between 4.5 to 9. hours.

30. The Energy Information Administration reported that the mean retail price per gallon of regular grade gasoline was $2.05 (Energy Information Administration, May 2009). Suppose that the standard deviation was $.10 and that the retail price per gallon has a bell-shaped distribution.
    a. What percentage of regular grade gasoline sold between $1.95 and $2.15 per gallon?
    b. What percentage of regular grade gasoline sold between $1.95 and $2.25 per gallon?
    c. What percentage of regular grade gasoline sold for more than $2.25 per gallon?





31. The national average for the math portion of the College Board's Scholastic Aptitude Test (SAT) is 515 (*The World Almanac*, 2009). The College Board periodically rescales the test scores such that the standard deviation is approximately 100. Answer the following questions using a bell-shaped distribution and the empirical rule for the verbal test scores.

a. What percentage of students have an SAT verbal score greater than 615?
b. What percentage of students have an SAT verbal score greater than 715?
c. What percentage of students have an SAT verbal score between 415 and 515?
d. What percentage of students have an SAT verbal score between 315 and 615?

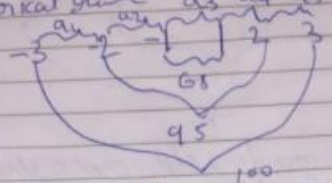
3) $s = 100$ normal distribution, $\bar{x} = 515$

$\bar{x} = 515$ $s = 100$

a) 615

$z = \frac{x - \bar{x}}{s} = \frac{615 - 515}{100} = \frac{100}{100} = 1$ standard deviation

empirical rule a_3 a_4 a_5



∴ % of student have an SAT verbal scored greater than 615 = the data lies between 1 to 3 standard deviation = $a_4 + a_5 = 16\%$

$a_3 = 68$, $a_2 + a_3 + a_4 = 95$, $a_1 + a_2 + a_3 + a_4 + a_5 = 100$

∴ $95 = a_2 + a_3 + a_4 = a_2 + 68 + a_4$

$a_2 = a_4$

∴ $95 = a_4 + 68 + a_4$

$2a_4 = 95 - 68 = 27$

$a_4 = \frac{27}{2} = 13.5$

$100 = a_1 + a_2 + a_3 + a_4 + a_5$

$= a_1 + 95 + a_5$ $a_1 = a_5$

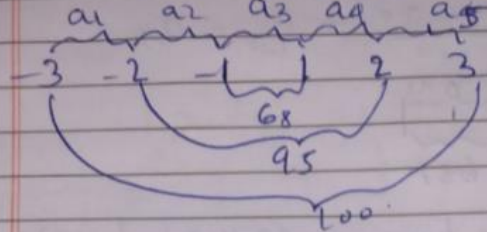
$100 = a_5 + 95 + a_5$

$2a_5 = 100 - 95$ ∴ Answer $a_4 + a_5 = 13.5 + 2.5$

$a_5 = \frac{5}{2} = 2.5$ $= 16\%$

b 715

$z = \frac{x - \bar{x}}{s} = \frac{715 - 515}{100} = \frac{200}{100} = 2$ standard deviation



According to empirical rule.

% of students have an SAT verbal score greater than 715 = data lies between 2 and 3 standard deviation. $= a_5 = 2.5\%$

$100 = a_1 + a_2 + a_3 + a_4 + a_5$

$a_1 = a_5$ & $a_2 + a_3 + a_4 = 95$

$100 = a_5 + 95 + a_5$

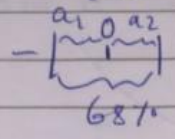
$2a_5 = 100 - 95$

$a_5 = \frac{5}{2} = 2.5\%$

c) 415 & 515.

$z_1 = \frac{x-\bar{x}}{s} = \frac{415-515}{100} = \frac{100}{100} = -1$

$z_2 = \frac{x-\bar{x}}{s} = \frac{515-515}{100} = \frac{0}{100} = 0$



% students have an SAT verbal score between 415 and 515 = % data between −1 to 0 standard deviation = a_1 = 34%.

According to empirical rule $a_1 + a_2 = 68\%$.

$a_1 = a_2$ ∴ ~~$a_1 = 68$~~ $a_1 + a_1 = 68\%$.

$2a_1 = 68\%$.

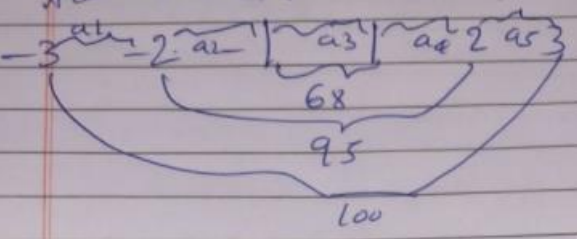
$a_1 = \frac{68}{2} = 34\%$.

d) 315 & 615

$z_1 = \frac{x-\bar{x}}{s} = \frac{315-515}{100} = \frac{-200}{100} = -2$.

$z_2 = \frac{x-\bar{x}}{s} = \frac{615-515}{100} = \frac{100}{100} = 1$

According to empirical rule.



% students have an SAT verbal between 315 and 615 = % data lies between −2 to 1 standard deviation = $a_2 + a_3 = 81.5\%$

$95 = a_2 + a_3 + a_4$ $a_3 = 68\%$. $a_2 = a_4$

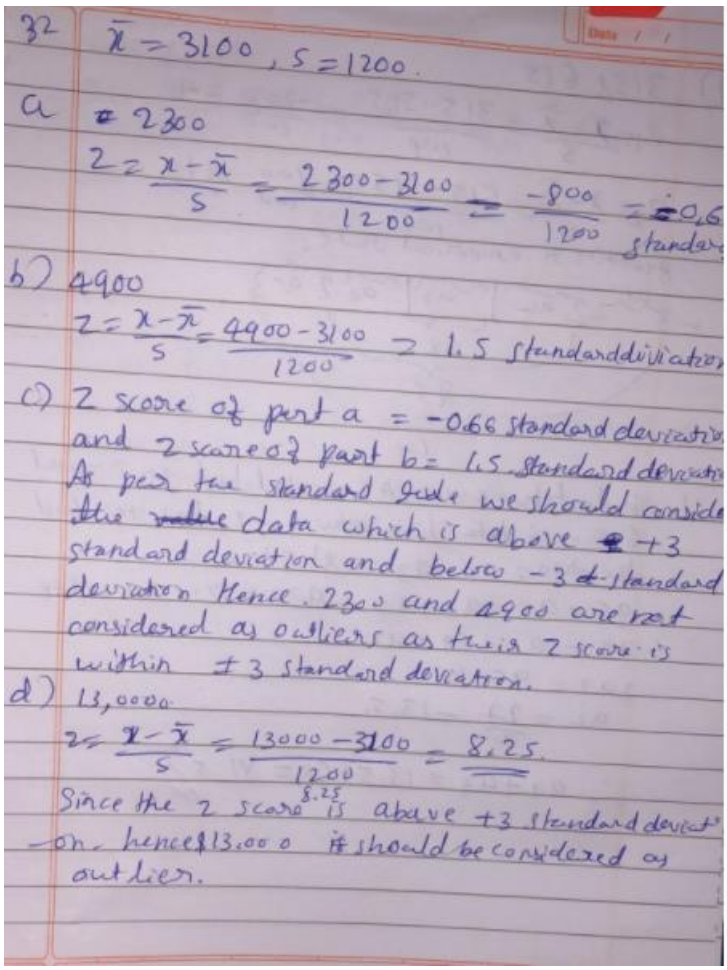
$95 = a_2 + 68 + a_2$

$2a_2 = 95 - 68$

$a_2 = \frac{27}{2} = 13.5$.

∴ $a_2 + a_3 = 13.5 + 68 = 81.5\%$

32. The high costs in the California real estate market have caused families who cannot afford to buy bigger homes to consider backyard sheds as an alternative form of housing expansion. Many are using the backyard structures for home offices, art studios, and hobby areas as well as for additional storage. The mean price of a customized wooden, shingled backyard structure is $3100 (*Newsweek,* September 29, 2003). Assume that the standard deviation is $1200.
    a. What is the z-score for a backyard structure costing $2300?
    b. What is the z-score for a backyard structure costing $4900?
    c. Interpret the z-scores in parts (a) and (b). Comment on whether either should be considered an outlier.
    d. The *Newsweek* article described a backyard shed-office combination built in Albany, California, for $13,000. Should this structure be considered an outlier? Explain.

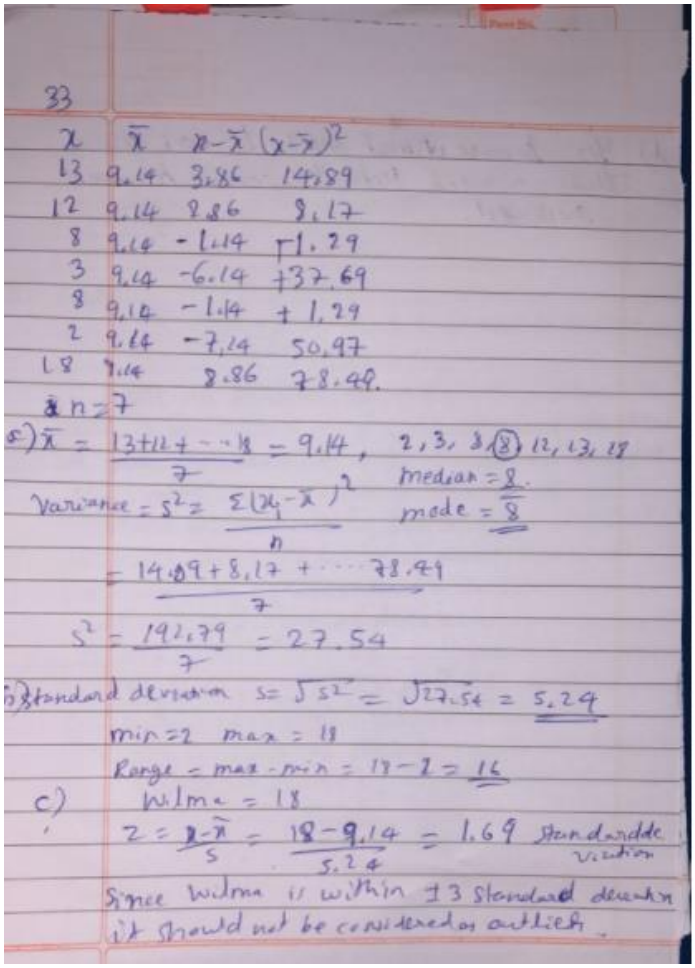


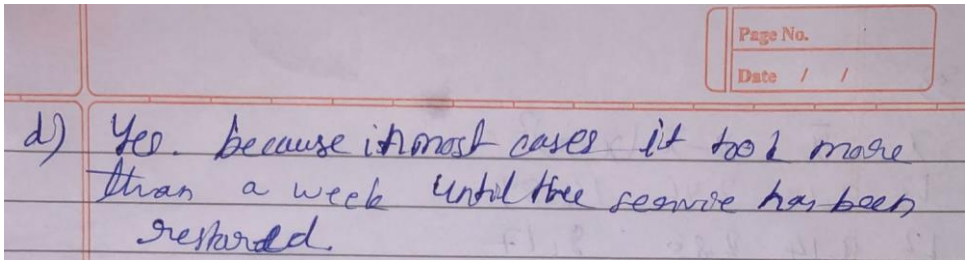
33. Florida Power & Light (FP&L) Company has enjoyed a reputation for quickly fixing its electric system after storms. However, during the hurricane seasons of 2004 and 2005, a new reality was that the company's historical approach to emergency electric system repairs was no longer good enough (*The Wall Street Journal,* January 16, 2006). Data showing the days required to restore electric service after seven hurricanes during 2004 and 2005 follow.

Hurricane	Days to Restore Service
Charley	13
Frances	12
Jeanne	8
Dennis	3
Katrina	8
Rita	2
Wilma	18

Based on this sample of seven, compute the following descriptive statistics:

a. Mean, median, and mode
b. Range and standard deviation
c. Should Wilma be considered an outlier in terms of the days required to restore electric service?
d. The seven hurricanes resulted in 10 million service interruptions to customers. Do the statistics show that FP&L should consider updating its approach to emergency electric system repairs? Discuss.



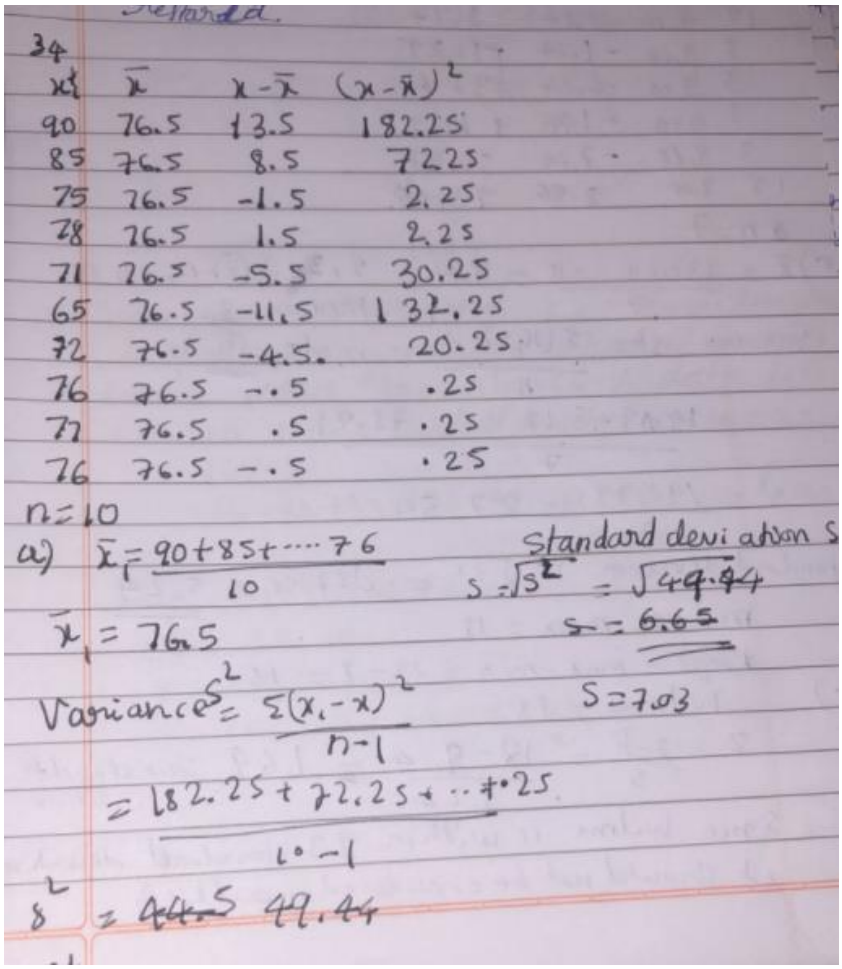


34. A sample of 10 NCAA college basketball game scores provided the following data (*USA Today*, January 26, 2004).

Winning Team	Points	Losing Team	Points	Winning Margin
Arizona	90	Oregon	66	24
Duke	85	Georgetown	66	19
Florida State	75	Wake Forest	70	5
Kansas	78	Colorado	57	21
Kentucky	71	Notre Dame	63	8
Louisville	65	Tennessee	62	3
Oklahoma State	72	Texas	66	6

Winning Team	Points	Losing Team	Points	Winning Margin
Purdue	76	Michigan State	70	6
Stanford	77	Southern Cal	67	10
Wisconsin	76	Illinois	56	20

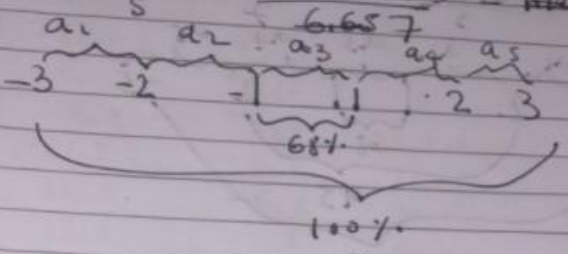
a. Compute the mean and standard deviation for the points scored by the winning team.
b. Assume that the points scored by the winning teams for all NCAA games follow a bell-shaped distribution. Using the mean and standard deviation found in part (a), estimate the percentage of all NCAA games in which the winning team scores 84 or more points. Estimate the percentage of NCAA games in which the winning team scores more than 90 points.
c. Compute the mean and standard deviation for the winning margin. Do the data contain outliers? Explain.



b) $\bar{x} = 76.5$ $s = 6.65$ ∴ normal distribution

84

i) $z_1 = \frac{(x_1 - \bar{x})}{s} = \frac{(84 - 76.5)}{6.657} \approx 1.07$



Since 84 is at around 1 standard deviation

∴ % of all NCAA games in which winning team scores 84 or above = % data lies between 1 to 3 standard deviation = $a_4 + a_5$

$= 16\%$ //

$100 = a_1 + a_2 + a_3 + a_4 + a_5$. $a_1 = a_5$, $a_2 = a_4$

$a_3 = 68\%$

$100 = a_1 + a_2 + 68 + a_1 + a_2$

$100 - 68 = 2a_1 + 2a_2$

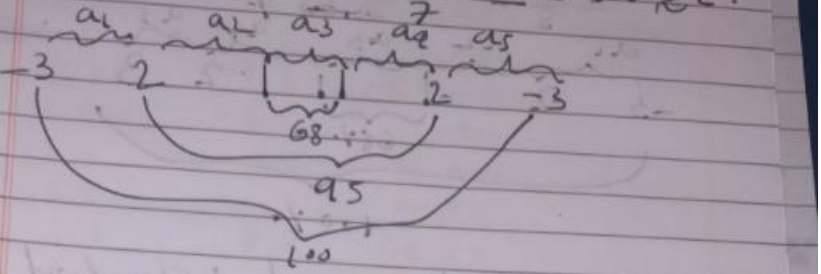
$32 = 2(a_1 + a_2)$

$a_1 + a_2 = a_4 + a_5 = \frac{32}{2} \Rightarrow 16\%$

$\bar{x} = 76.5$ $s = 7$

90

$z = \frac{(x - \bar{x})}{s} = \frac{90 - 76.5}{7} = 1.9 \approx 2$.

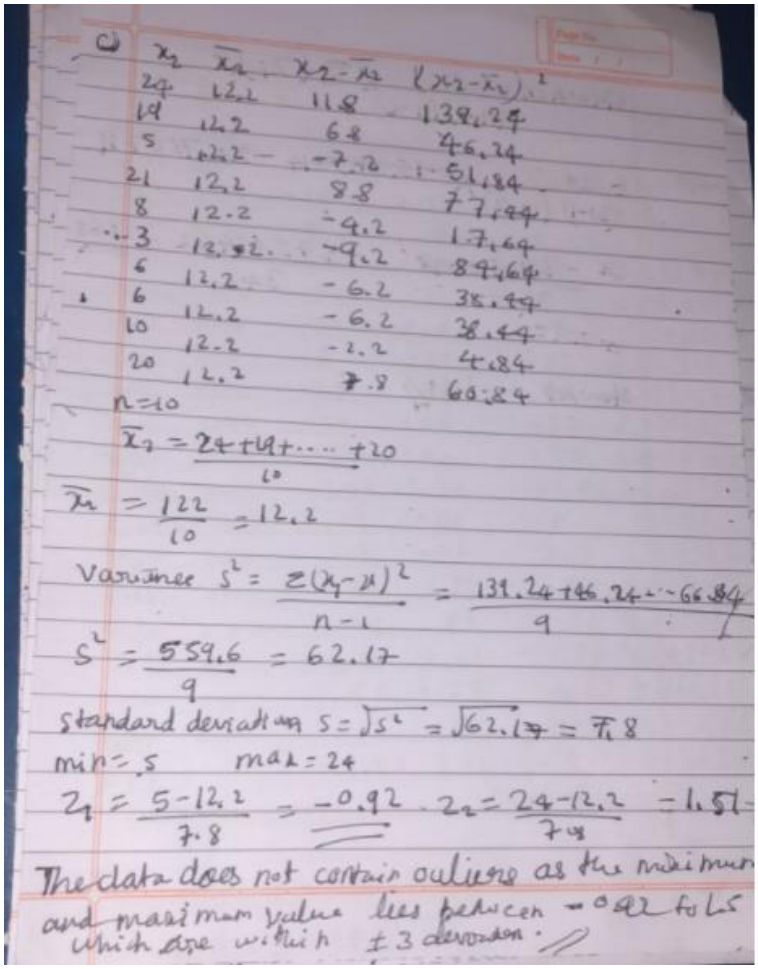


% of NCAA games in which the winning team scores more than 90 points = % data lies between 2 to 3 standard deviation $= a_5 = 2.5\%$

$100 = a_1 + a_2 + a_3 + a_4 + a_5$, $a_2 + a_3 + a_4 = 95$, $a_1 = a_5$

$100 = 95 + 95 + a_5$

$a_5 = \frac{100 - 95}{2} = 2.5\%$



35. *Consumer Reports* posts reviews and ratings of a variety of products on its website. The following is a sample of 20 speaker systems and their ratings. The ratings are on a scale of 1 to 5, with 5 being best.

Speaker	Rating	Speaker	Rating
Infinity Kappa 6.1	4.00	ACI Sapphire III	4.67
Allison One	4.12	Bose 501 Series	2.14
Cambridge Ensemble II	3.82	DCM KX-212	4.09
Dynaudio Contour 1.3	4.00	Eosone RSF1000	4.17
Hsu Rsch. HRSW12V	4.56	Joseph Audio RM7si	4.88
Legacy Audio Focus	4.32	Martin Logan Aerius	4.26
Mission 73li	4.33	Omni Audio SA 12.3	2.32
PSB 400i	4.50	Polk Audio RT12	4.50
Snell Acoustics D IV	4.64	Sunfire True Subwoofer	4.17
Thiel CS1.5	4.20	Yamaha NS-A636	2.17

a. Compute the mean and the median.
b. Compute the first and third quartiles.
c. Compute the standard deviation.
d. The skewness of this data is −1.67. Comment on the shape of the distribution.
e. What are the z-scores associated with Allison One and Omni Audio?
f. Do the data contain any outliers? Explain.

35

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	$Z=\frac{x-\bar{x}}{s}$	Z^3
4.00	4.02	0	0	0	0
4.12	4.02	.12	.0144	0.13	0.002197
3.82	4.02	-1.8	3.24	-2	-8
4.00	4.02	0	0	0	0
4.56	4.02	.56	.31	.62	0.238328
4.32	4.02	.32	.10	.35	0.042875
4.33	4.02	.33	.10	.36	0.046656
4.50	4.02	.5	.25	.55	0.166375
4.66	4.02	.66	.43	.73	0.389017
4.20	4.02	.20	.04	.22	0.010648
4.67	4.02	.67	.44	.74	0.405224
2.14	4.02	-1.83	3.34	-2.03	-8.365447
4.09	4.02	.09	.00	0.1	0.001
4.17	4.02	.17	.02	0.18	0.005832
4.88	4.02	.88	.77	.97	0.917673
4.26	4.02	.26	.06	.28	0.021952
2.32	4.02	-1.68	2.82	-1.86	-6.434856
4.50	4.02	.50	.25	0.55	0.166375
4.17	4.02	.17	.02	0.18	0.005832
2.17		-1.83	3.34	-2.03	-8.365447

$\bar{x}=\frac{\sum x_i}{n}=\frac{4.00+4.12+\cdots+2.17}{20}=\frac{80.74}{20}$

Variance $s^2=\frac{\sum(x-\bar{x})^2}{n-1}=\frac{.0144+3.24+\cdots}{19}$

$z=\frac{16}{19}=0.842$

Standard deviation $=s=\sqrt{s^2}=\sqrt{.84}$

$=.92$

a). mean $\bar{x}=4.02$

2.14, 2.17, 2.32, 3.82, 4.00, 4.00, 4.09, 4.12, 4.17, 4.17, 4.20, 4.26, 4.32, 4.33, 4.50, 4.50, 4.56, 4.64, 4.67, 4.88

median $=\frac{n}{2}=\frac{20}{2}=10$

median $=\frac{i+i+1}{2}=\frac{4.17+4.20}{2}=4.18$

b). $Q_1=25\%$

$i=\frac{n\times p}{100}=\frac{20\times 25}{100}=5$

$Q(25)=\frac{i+(i+1)}{2}=\frac{5+6}{2}=\frac{3.82+4.00}{2}=3.91$

$Q(25)=3.91$

$Q3=75\%$

$i=\frac{p\times n}{100}=\frac{75\times 20}{100}=15$

$Q(75)=\frac{i+(i+1)}{2}=\frac{4.50+4.50}{2}=4.50$

c) Standard deviation $= s = \sqrt{s^2} = \sqrt{.84} = .92$

d) Since the skewness is −1.67. ~~negatively skewed~~ it indicates the data is negatively skewed or skewed towards left.

e) Allison = 4.2, Omni = 2.32, $\bar{x} = 4.0$, $s = .92$

$$z_1 = \frac{x - \bar{x}}{s} = \frac{4.2 - 4.0}{.92} = .2$$

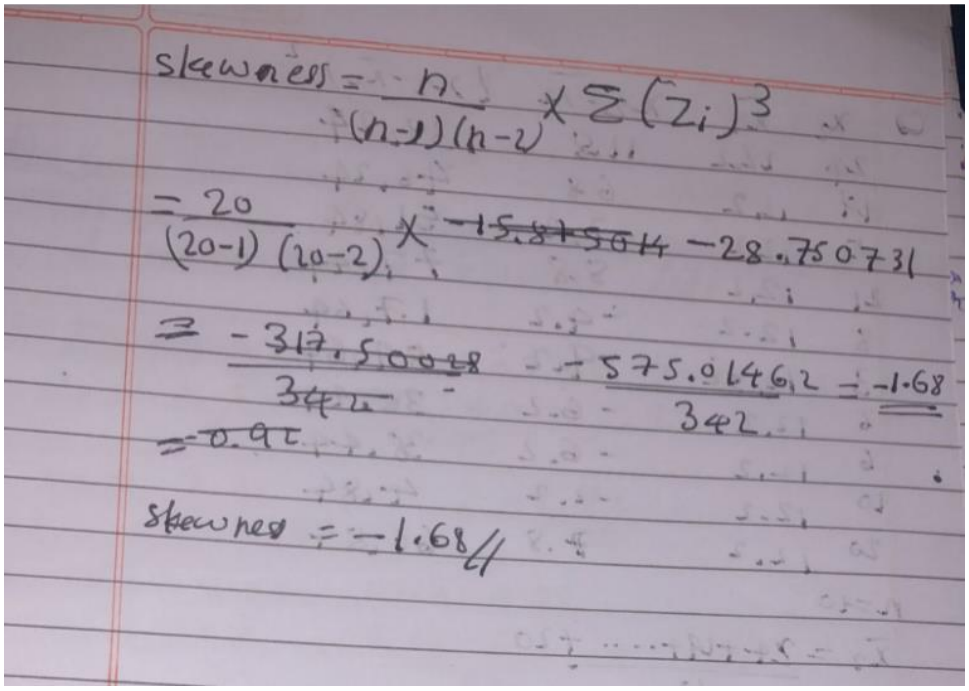
$$z_2 = \frac{2.32 - 4.0}{.92} = -1.82$$

f) min = 2.64 max = 4.88

$$z_1 = \frac{x - \bar{x}}{s} = \frac{2.64 - 4.0}{.92} = -2.02$$

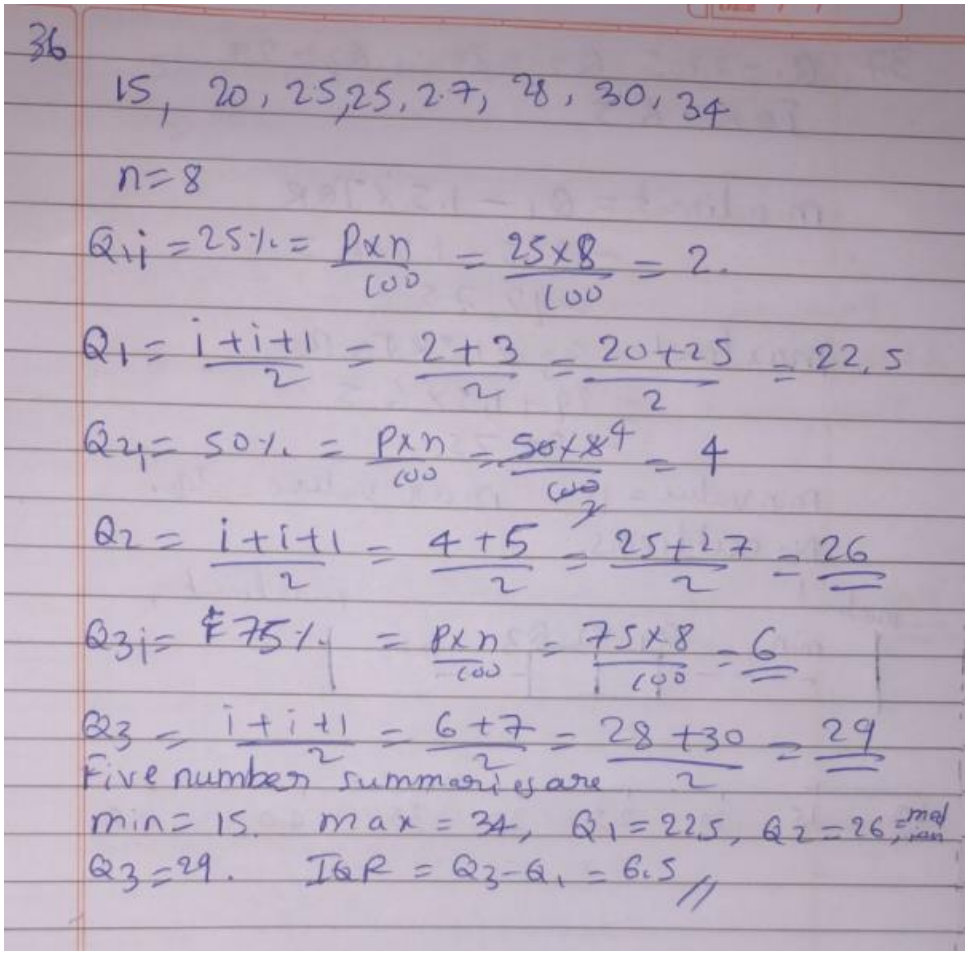
$$z_2 = \frac{x - \bar{x}}{s} = \frac{4.88 - 4.0}{.92} = .95.$$

Since the 2 scores of the minimum and maximum values in the data are −2.02 and .95 which are lying within ±3 standard deviation there are no outliers in the given data.

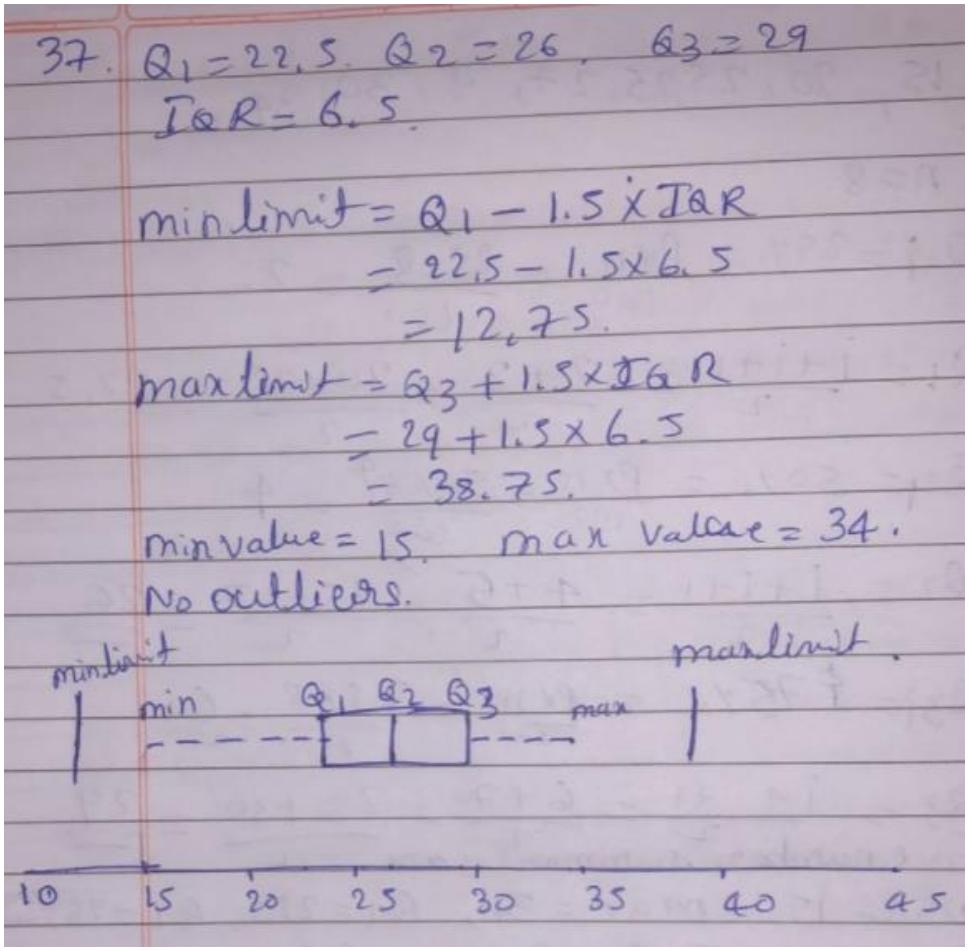


Methods

36. Consider a sample with data values of 27, 25, 20, 15, 30, 34, 28, and 25. Provide the five-number summary for the data.



37. Show the box plot for the data in exercise 36.



38. Show the five-number summary and the box plot for the following data: 5, 15, 18, 10, 8, 12, 16, 10, 6.

38) 5, 6, 8, 10, 10, 12, 15, 16, 18

$n = 9$

Five Number Summary

① min = 5 2) max = 18

$Q_1 = \frac{25 \times 9}{100} = 2.25 \approx 3 = 8$

$Q_2 = \frac{50 \times 9}{100} = 4.5 \approx 5 = 10$

$Q_3 = \frac{75 \times 9}{100} = 6.75 \approx 7 = 15$

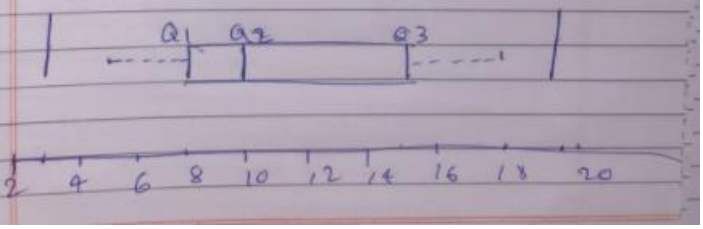
Box plot $Q_1 = 8$, $Q_2 = 10$, $Q_3 = 15$,

$IQR = Q_3 - Q_1 = 15 - 8 = 3$

min limit $= Q_1 - 1.5 \times IQR = 8 - 1.5 \times 3 = 3.5$

max limit $= Q_3 + 1.5 \times IQR = 15 + 1.5 \times 3 = 19.5$

min value = 5 max value = 18. No outliers



39. A data set has a first quartile of 42 and a third quartile of 50. Compute the lower and upper limits for the corresponding box plot. Should a data value of 65 be considered an outlier?

34 $Q_1 = 42$, $Q_3 = 50$. $IQR = 50 - 42 = 8$

min limit $= Q_1 - 1.5 \times IQR = 42 - 1.5 \times 8$

$= 30$

max limit $= Q_3 + 1.5 \times IQR = 50 + 1.5 \times 8$

$= 62$

The data 65 should be considered as outlier as it is above max limit 62

Applications

40. Naples, Florida, hosts a half-marathon (13.1-mile race) in January each year. The event attracts top runners from throughout the United States as well as from around the world. In January 2009, 22 men and 31 women entered the 19–24 age class. Finish times in minutes are as follows (*Naples Daily News,* January 19, 2009). Times are shown in order of finish.

Finish	Men	Women	Finish	Men	Women	Finish	Men	Women
1	65.30	109.03	11	109.05	123.88	21	143.83	136.75
2	66.27	111.22	12	110.23	125.78	22	148.70	138.20
3	66.52	111.65	13	112.90	129.52	23		139.00
4	66.85	111.93	14	113.52	129.87	24		147.18
5	70.87	114.38	15	120.95	130.72	25		147.35
6	87.18	118.33	16	127.98	131.67	26		147.50
7	96.45	121.25	17	128.40	132.03	27		147.75
8	98.52	122.08	18	130.90	133.20	28		153.88
9	100.52	122.48	19	131.80	133.50	29		154.83
10	108.18	122.62	20	138.63	136.57	30		189.27
						31		189.28

a. George Towett of Marietta, Georgia, finished in first place for the men and Lauren Wald of Gainesville, Florida, finished in first place for the women. Compare the first-place finish times for men and women. If the 53 men and women runners had competed as one group, in what place would Lauren have finished?

b. What is the median time for men and women runners? Compare men and women runners based on their median times.

c. Provide a five-number summary for both the men and the women.

d. Are there outliers in either group?

e. Show the box plots for the two groups. Did men or women have the most variation in finish times? Explain.

(40)

c) Five point summary of men. $n_1 = 22$

$min_1 = 65.30$ $max_1 = 148.70$

$Q_1 = \frac{25 \times 22}{100} = 5.5 \approx 6. = 87.18$

$Q_2 = \frac{50 \times 22}{100} = 11 = \frac{11+12}{2} = \frac{109.05 + 110.23}{2}$

$= 109.64$

$Q_3 = \frac{75 \times 22}{100} = 16.5 \approx 17 = 128.40$

Five point summary of women $n_2 = 31$

$min = 101.03$. $max = 189.78$.

$Q_1 = \frac{25 \times 31}{100} = 7.75 \approx 8 = 122.08$

$Q_2 = \frac{50 \times 31}{100} = 15.5 \approx 16 = 127.98$

$Q_3 = \frac{75 \times 31}{100} = 23.25 \approx 24 = 147.18$

men

IQR $Q_3 - Q_1 = 128.40 - 87.18 = 41.22$

min limit $= Q_1 - 1.5 \times IQR = 87.18 - 41.22 \times 1.5$

$= 25.35$

max limit $= Q_3 + 1.5 \times IQR = 128.40 + 1.5 \times 41.22$

$= 190.23$

As the min value 65.30 and max value 148.70 are within the min limit and max limit there are no outliers in men.

women.

$IQR = Q_3 - Q_1 = 147.18 - 122.08 = 25.1$

min limit $= Q_1 - 1.5 \times IQR = 122.08 - 1.5 \times 25.1$

$= 84.43$

max limit $= Q_3 - 1.5 \times IQR = 147.18 - 1.5 \times 25.1$

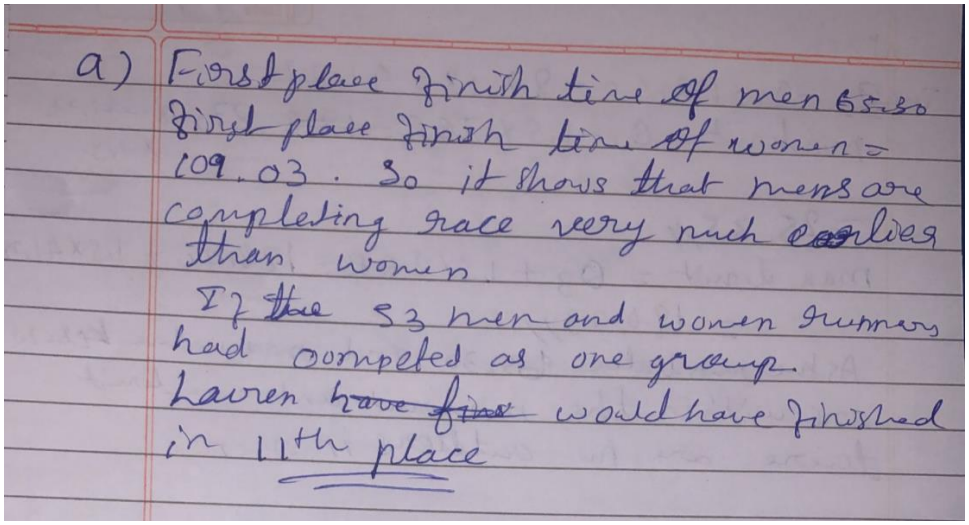
$= 184.83$

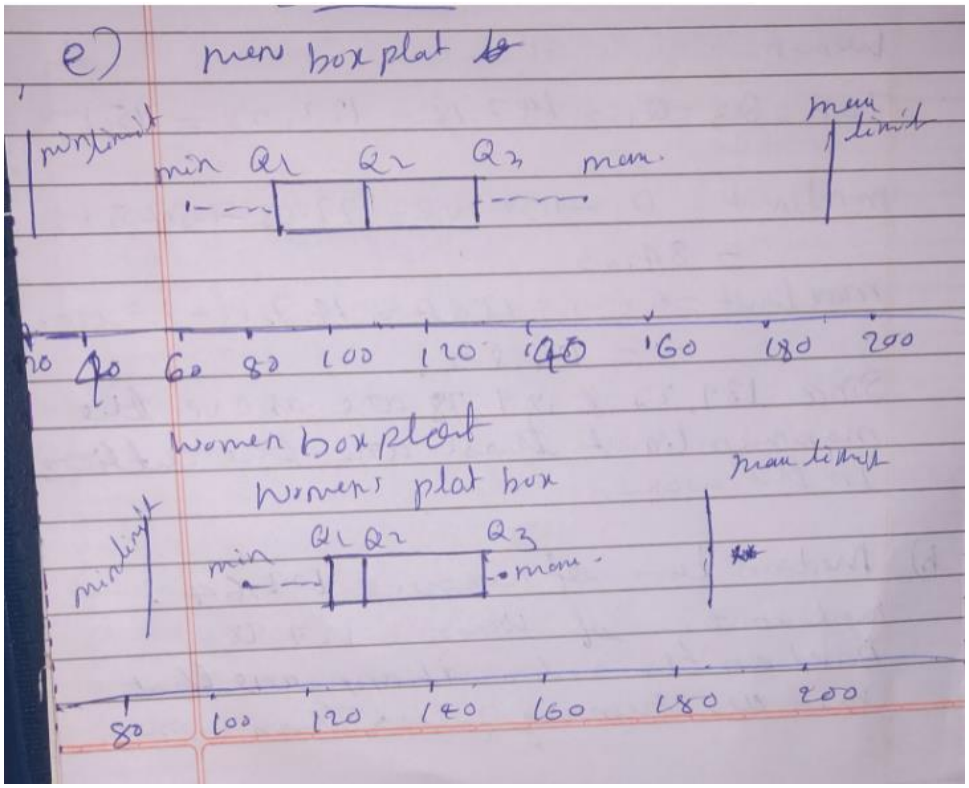
Since 189.27 & 189.78 are above the maximum limit these are the outliers in the women

b) Median time of men $= Q_2 = 109.64$.

median time of women $= 127.98$

Based on the median it appears that men are running faster than women





41. Annual sales, in millions of dollars, for 21 pharmaceutical companies follow.

8408	1374	1872	8879	2459	11413
608	14138	6452	1850	2818	1356
10498	7478	4019	4341	739	2127
3653	5794	8305			

a. Provide a five-number summary.
b. Compute the lower and upper limits.
c. Do the data contain any outliers?
d. Johnson & Johnson's sales are the largest on the list at $14,138 million. Suppose a data entry error (a transposition) had been made and the sales had been entered as $41,138 million. Would the method of detecting outliers in part (c) identify this problem and allow for correction of the data entry error?
e. Show a box plot.

4) n=21

608, 739, 1356, 1374, 1850, 1872, 2127, 2489, 2818, 3653, 4019, 4341, 5794, 6452, 7478, 8305, 8408, 8827, 10498, 11413, 14138.

a) Five-number summary

min = 608, max = 14138

$Q_1 = \frac{25 \times 21}{100} = 5.25 \simeq 6 = 1872$

$Q_2 = \frac{50 \times 21}{100} = 10.5 \simeq 11 = 4019$

$Q_3 = \frac{75 \times 21}{100} = 15.75 \simeq 16 = 8305.$

b) minimum limit = $Q_1 - IQR \times 1.5$

$IQR = Q_3 - Q_1 = 8305 - 1872 = 6433$

minimum limit = $1872 - 6433 \times 1.5$

$= -7777$

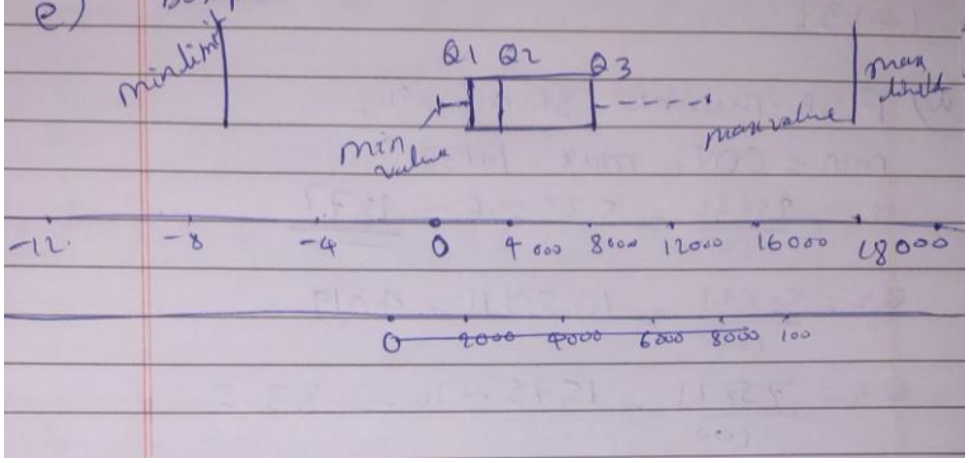
max limit = $Q_3 + IQR \times 1.5$

$= 8305 + 6433 \times 1.5 = 17954.5$

c) Since minimum 608 & maximum 14138 are within the minimum limit & maximum limit there are no outliers

d) Yes, 14138 will be identified as outlier as it is above the maximum limit 17954.5.

e) Boxplot

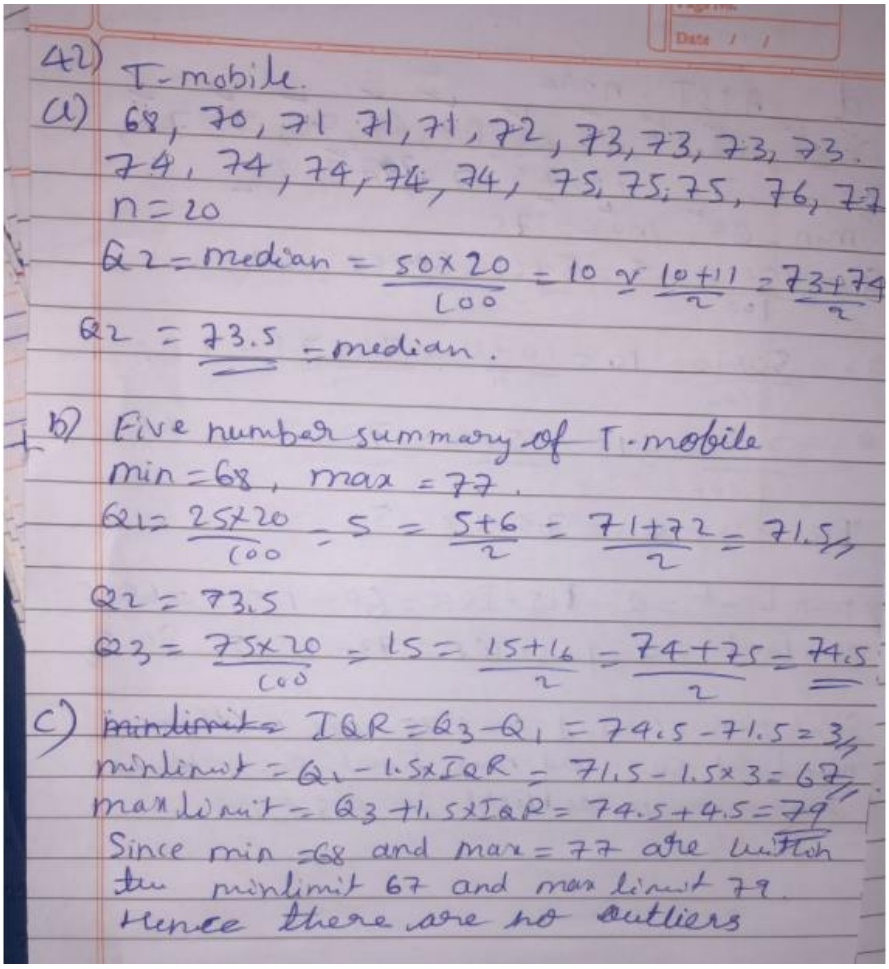


e. Show a box plot.

42. *Consumer Reports* provided overall customer satisfaction scores for AT&T, Sprint, T-Mobile, and Verizon cell-phone services in major metropolitan areas throughout the United States. The rating for each service reflects the overall customer satisfaction considering a variety of factors such as cost, connectivity problems, dropped calls, static interference, and customer support. A satisfaction scale from 0 to 100 was used with 0 indicating completely dissatisfied and 100 indicating completely satisfied. The ratings for the four cell-phone services in 20 metropolitan areas are as shown (*Consumer Reports*, January 2009).

Metropolitan Area	AT&T	Sprint	T-Mobile	Verizon
Atlanta	70	66	71	79
Boston	69	64	74	76
Chicago	71	65	70	77
Dallas	75	65	74	78
Denver	71	67	73	77
Detroit	73	65	77	79
Jacksonville	73	64	75	81
Las Vegas	72	68	74	81
Los Angeles	66	65	68	78
Miami	68	69	73	80
Minneapolis	68	66	75	77
Philadelphia	72	66	71	78
Phoenix	68	66	76	81
San Antonio	75	65	75	80
San Diego	69	68	72	79
San Francisco	66	69	73	75
Seattle	68	67	74	77
St. Louis	74	66	74	79
Tampa	73	63	73	79
Washington	72	68	71	76

a. Consider T-Mobile first. What is the median rating?
b. Develop a five-number summary for the T-Mobile service.
c. Are there outliers for T-Mobile? Explain.
d. Repeat parts (b) and (c) for the other three cell-phone services.



d AT&T. n=20

66, 66, 68, 68, 68, 68, 69, 69, 70, 71, 71, 72, 72, 72, 73, 73, 73, 74, 75, 75

min = 66, max = 75

$Q_1 = \frac{25 \times 20}{100} = 5 = \frac{5+6}{2} = \frac{68+68}{2} = 68$

$Q_2 = \frac{50 \times 20}{100} = 10 = \frac{10+11}{2} = \frac{71+71}{2} = 71$

$Q_3 = \frac{75 \times 20}{100} = 15 = \frac{15+16}{2} = \frac{73+73}{2} = 73$

$IQR = Q_3 - Q_1 = 73 - 68 = 5$

min limit $= Q_1 - 1.5 \times IQR = 68 - 1.5 \times 5 = 60.5$

max limit $= Q_3 + 1.5 \times IQR = 73 - 1.5 \times 5 = 80.5$

Since min 66 and max 75 are within min limit 60.5 and max limit 80.5. Hence there are no outliers in AT&T

e) Sprint. n=20

63, 64, 64, 65, 65, 65, 65, 65, 66, 66, 66, 66, 66, 67, 67, 68, 68, 68, 69, 69.

min = 63 max = 69.

$Q_1 = \frac{25 \times 20}{100} = 5 = \frac{5+6}{2} = \frac{65+65}{2} = 65$

$Q_2 = \frac{50 \times 20}{100} = 10 = \frac{10+11}{2} = \frac{66+66}{2} = 66$

$Q_3 = \frac{75 \times 20}{100} = 15 = \frac{15+16}{2} = \frac{67+68}{2} = 67.5$

min limit $= Q_1 - IQR \times 1.5 = 65 - 1.5 \times 2.5 = 61.25$

$IQR = Q_3 - Q_1 = 67.5 - 65 = 2.5$

max limit $= Q_3 + 1.5 \times IQR = 67.5 + 1.5 \times 2.5 = 71.25$

Since the min = 63 & max = 69 are within the min limit 61.25 and max limit 71.25 Hence there are no outliers.

d) Verizon $n = 20$

75, 76, 76, 77, 77, 77, 77, 78, 78, 78, 79, 79, 79, 79, 80, 80, 81, 81, 81

min = 75, max = 81

$Q_1 = \frac{25 \times 20}{100} = 5 = \frac{5+6}{2} = \frac{77+77}{2} = 77$

$Q_2 = \frac{50 \times 20}{100} = 10 = \frac{10+11}{2} = \frac{79+79}{2} = 79$

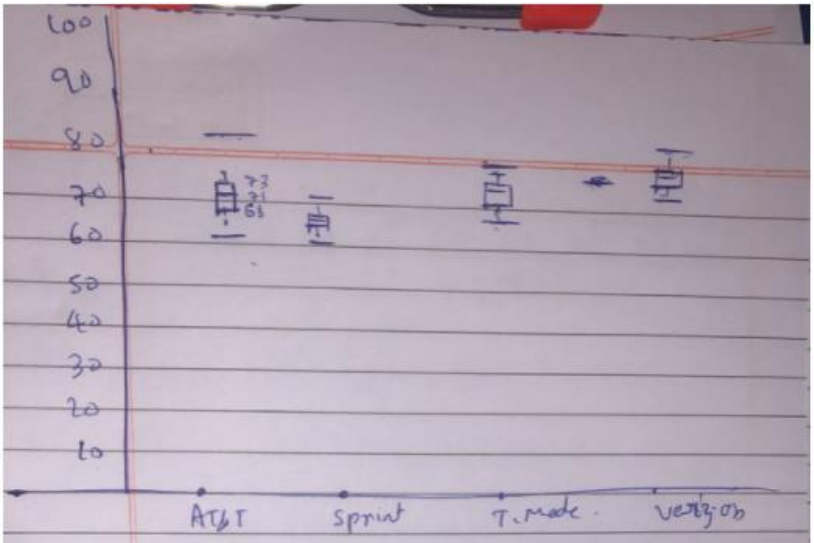
$Q_3 = \frac{75 \times 20}{100} = 15 = \frac{15+16}{2} = \frac{79+80}{2} = 79.5$

$IQR = Q_3 - Q_1 = 79.5 - 77 = 2.5$

min limit $= Q_1 - 1.5 \times IQR = 77 - 1.5 \times 2.5 = 73.25$

max limit $= Q_3 + 1.5 \times IQR = 79.5 + 1.5 \times 2.5 = 83.25$.

Since min = 75 & max = 81 are within the min limit 73.25 and max limit 83.25 Hence there are no outliers.



Observations

→ higher ratings are in verizon. and lower ratings are in sprint.

→ In terms of median higher ratings are in verizon. and then in T mode.

→ Sprint & verizon have less variance. where as AT&T has most variance.

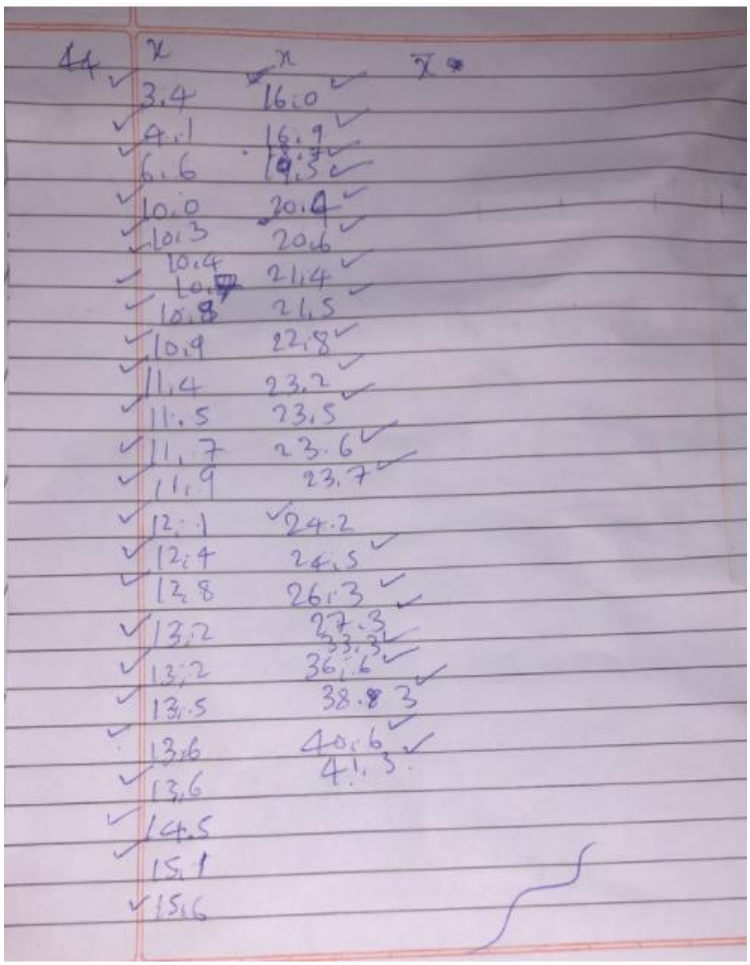
→ there are no outliers in any of the categories.

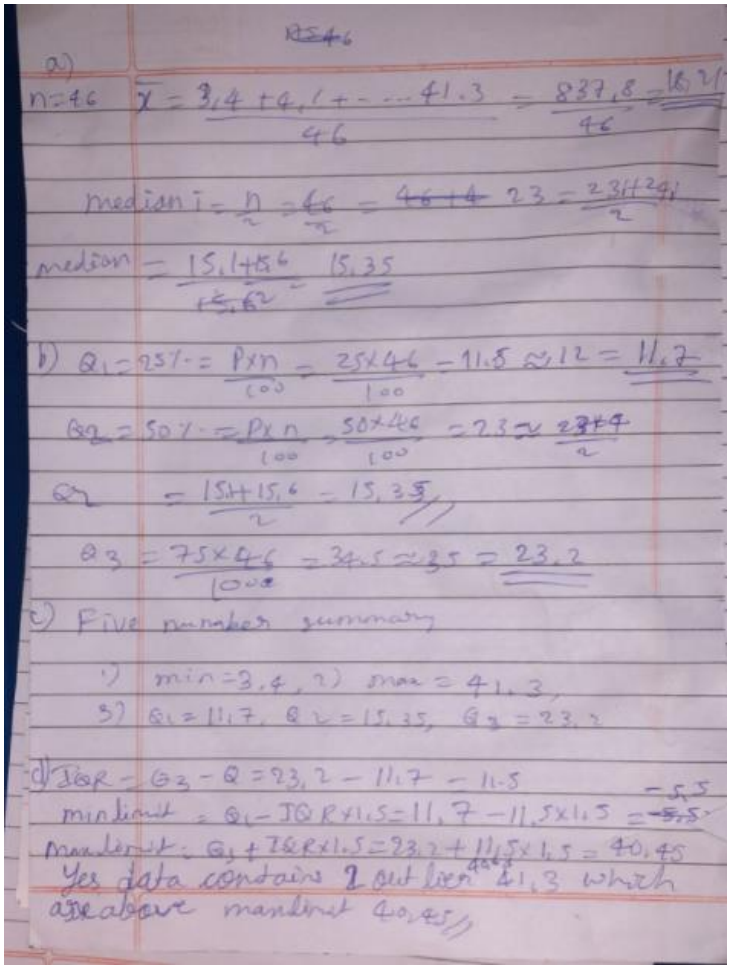
WEB file
Mutual

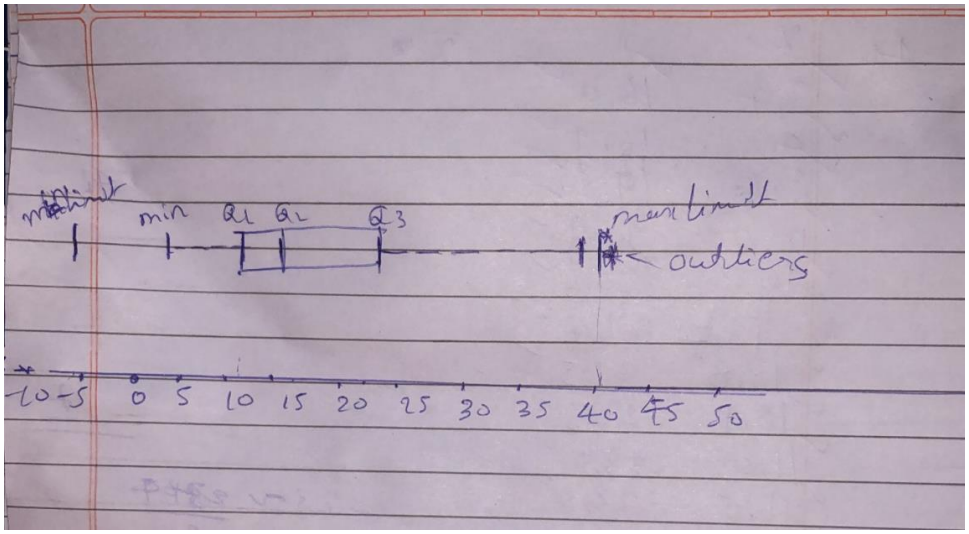
44. A listing of 46 mutual funds and their 12-month total return percentage is shown in Table 3.5 (*Smart Money*, February 2004).
    a. What are the mean and median return percentages for these mutual funds?
    b. What are the first and third quartiles?
    c. Provide a five-number summary.
    d. Do the data contain any outliers? Show a box plot.

TABLE 3.5 TWELVE-MONTH RETURN FOR MUTUAL FUNDS

Mutual Fund	Return (%)	Mutual Fund	Return (%)
Alger Capital Appreciation	23.5	Nations Small Company	21.4
Alger LargeCap Growth	22.8	Nations SmallCap Index	24.5
Alger MidCap Growth	38.3	Nations Strategic Growth	10.4
Alger SmallCap	41.3	Nations Value Inv	10.8
AllianceBernstein Technology	40.6	One Group Diversified Equity	10.0
Federated American Leaders	15.6	One Group Diversified Int'l	10.9
Federated Capital Appreciation	12.4	One Group Diversified Mid Cap	15.1
Federated Equity-Income	11.5	One Group Equity Income	6.6
Federated Kaufmann	33.3	One Group Int'l Equity Index	13.2
Federated Max-Cap Index	16.0	One Group Large Cap Growth	13.6
Federated Stock	16.9	One Group Large Cap Value	12.8
Janus Adviser Int'l Growth	10.3	One Group Mid Cap Growth	18.7
Janus Adviser Worldwide	3.4	One Group Mid Cap Value	11.4
Janus Enterprise	24.2	One Group Small Cap Growth	23.6
Janus High-Yield	12.1	PBHG Growth	27.3
Janus Mercury	20.6	Putnam Europe Equity	20.4
Janus Overseas	11.9	Putnam Int'l Capital Opportunity	36.6
Janus Worldwide	4.1	Putnam International Equity	21.5
Nations Convertible Securities	13.6	Putnam Int'l New Opportunity	26.3
Nations Int'l Equity	10.7	Strong Advisor Mid Cap Growth	23.7
Nations LargeCap Enhd. Core	13.2	Strong Growth 20	11.7
Nations LargeCap Index	13.5	Strong Growth Inv	23.2
Nation MidCap Index	19.5	Strong Large Cap Growth	14.5







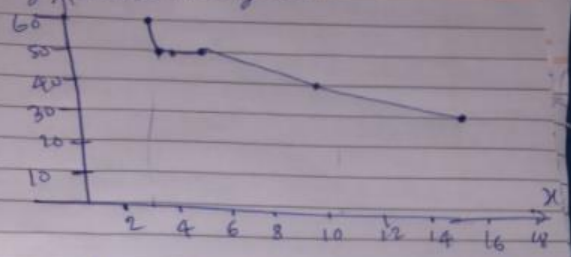
Methods

45. Five observations taken for two variables follow.

x_i	4	6	11	3	16
y_i	50	50	40	60	30

a. Develop a scatter diagram with x on the horizontal axis.
b. What does the scatter diagram developed in part (a) indicate about the relationship between the two variables?
c. Compute and interpret the sample covariance.
d. Compute and interpret the sample correlation coefficient.

4.5.
a)



b) The scatter diagram indicates that x and y are negatively related to eachother.

c) Sample covariance $S_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n-1}$

$\bar{x} = \frac{4+6+\cdots+16}{5} = 8$ // $\bar{y} = \frac{50+50+\cdots+30+46}{5}$

$\Sigma(x_i-\bar{x}) = -4-2+3-5+8 = 0$

$\Sigma(y-\bar{y}) = 4+4-6+14-16 = 0$

$S_{xy} = \frac{0}{4} = 0$ //

Sample covariance = 0 //

It indicates that there is no relation exists between x and y.

d) Since sample covariance = 0, sample correlation indicating no relation

Variance $= \frac{\Sigma(x_i-\bar{x})^2}{n-1} = \frac{16+4+9+25+64}{4} = \frac{29.5}{5.43}$

Standard deviation $S_x = \sqrt{29.5} =$

Variance $S_y^2 = \frac{\Sigma(y-\bar{y})^2}{n-1} = \frac{16+16+36+196+256}{4}$

$= \frac{520}{4} = 130$ // $S_y = \sqrt{130} = 11.40$

Correlation coefficience $= \frac{S_{xy}}{S_x S_y} = \frac{-60}{5.43 \times 130}$

$= \frac{-60}{616.2} = -0.097$

$= -0.09 \quad \frac{60}{706.01} = 0.084$

correlation coefficient $= \frac{S_{xy}}{S_x S_y} = \frac{-60}{5.43 \times 11.40}$

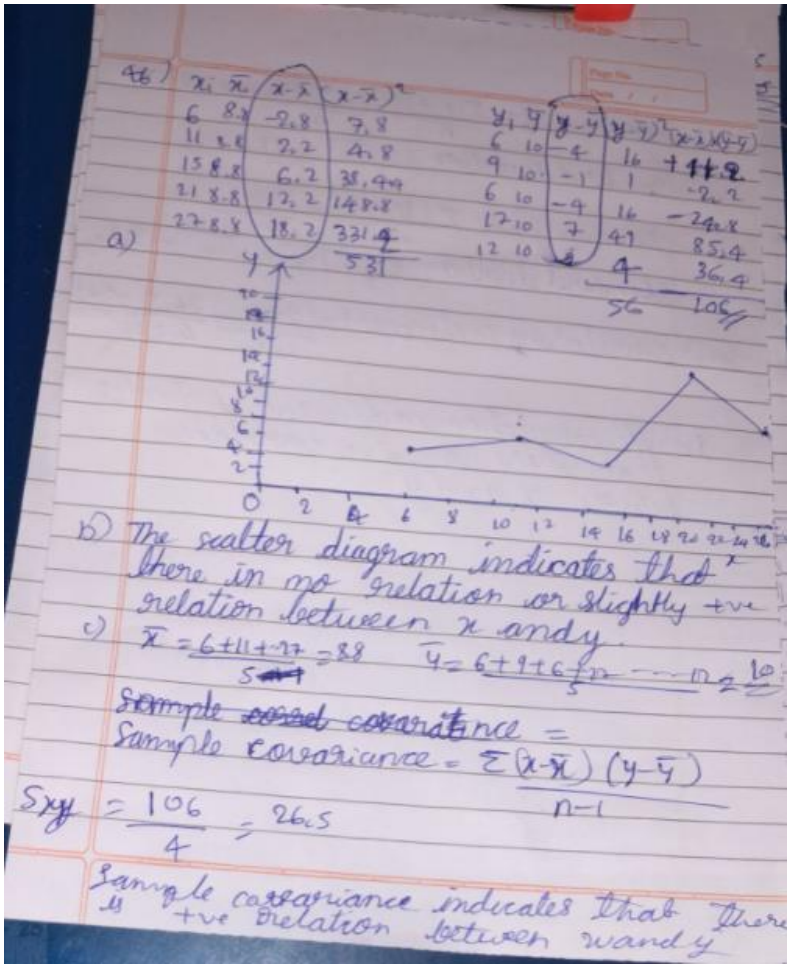
$= \frac{-60}{61.902} = 0.96 \approx -1$

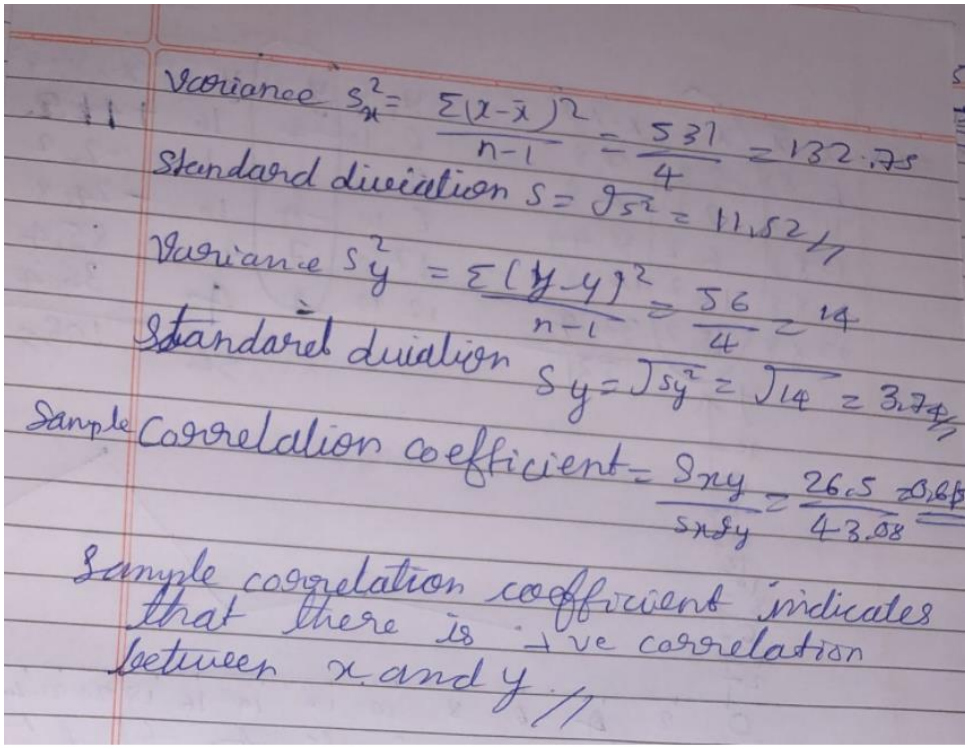
∴ x and This indicates that x and y are strongly negatively correlated

46. Five observations taken for two variables follow.

x_i	6	11	15	21	27
y_i	6	9	6	17	12

a. Develop a scatter diagram for these data.
b. What does the scatter diagram indicate about a relationship between x and y?
c. Compute and interpret the sample covariance.
d. Compute and interpret the sample correlation coefficient.





Applications

47. Nielsen Media Research provides two measures of the television viewing audience: a television program *rating*, which is the percentage of households with televisions watching a program, and a television program *share*, which is the percentage of households watching a program among those with televisions in use. The following data show the Nielsen television ratings and share data for the Major League Baseball World Series over a nine-year period (Associated Press, October 27, 2003).

Rating	19	17	17	14	16	12	15	12	13
Share	32	28	29	24	26	20	24	20	22

a. Develop a scatter diagram with rating on the horizontal axis.
b. What is the relationship between rating and share? Explain.
c. Compute and interpret the sample covariance.
d. Compute the sample correlation coefficient. What does this value tell us about the relationship between rating and share?

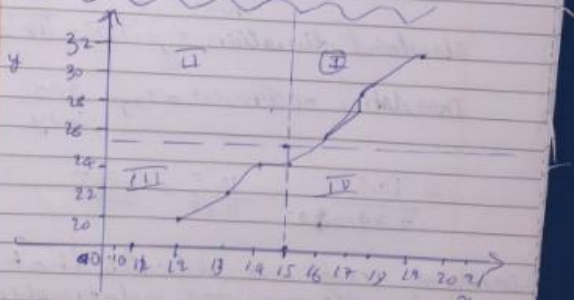
47)

Rating

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	y	$\bar{y}$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
19	15	4	16	32	25	7	49	$4\times7=28$
17	15	2	4	28	25	3	9	$2\times3=6$
17	15	2	4	29	25	4	16	$2\times4=8$
14	15	-1	1	24	25	-1	1	$-1\times-1=1$
16	15	-1	1	26	25	1	1	$1\times1=1$
12	15	-3	9	20	25	-5	25	$-3\times-5=15$
15	15	0	0	24	25	-1	1	$0\times-1=0$
12	15	-3	9	20	25	-5	25	$-3\times-5=15$
13	15	-2	2	22	25	-3	9	$-2\times-5=10$
			46				136	84

$\bar{x} = \frac{19+\dots+13}{9} = 15$ $\quad \bar{y} = \frac{32+\dots+22}{9} = 25$

a)



b) Scatter diagram indicates that there is a strong positive relationship between Rating and Share.

c) Sample covariance $S_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n-1}$

$S_{xy} = \frac{84}{8} = 10.5$

Sample covariance indicates there is +ve relationship between rating and share.

d) Variance $S_x^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{46}{8} = 5.75$

standard diviation $s = \sqrt{S_x^2} = \sqrt{5.75} = 2.39$

Variance $S_y^2 = \frac{\Sigma(y-\bar{y})^2}{n-1} = \frac{136}{8} = 17$

standard diviation $S_y = \sqrt{S_y^2} = \sqrt{17} = 4.12$

Correlation coefficient $= r_{xy} = \frac{S_{xy}}{S_x S_y}$

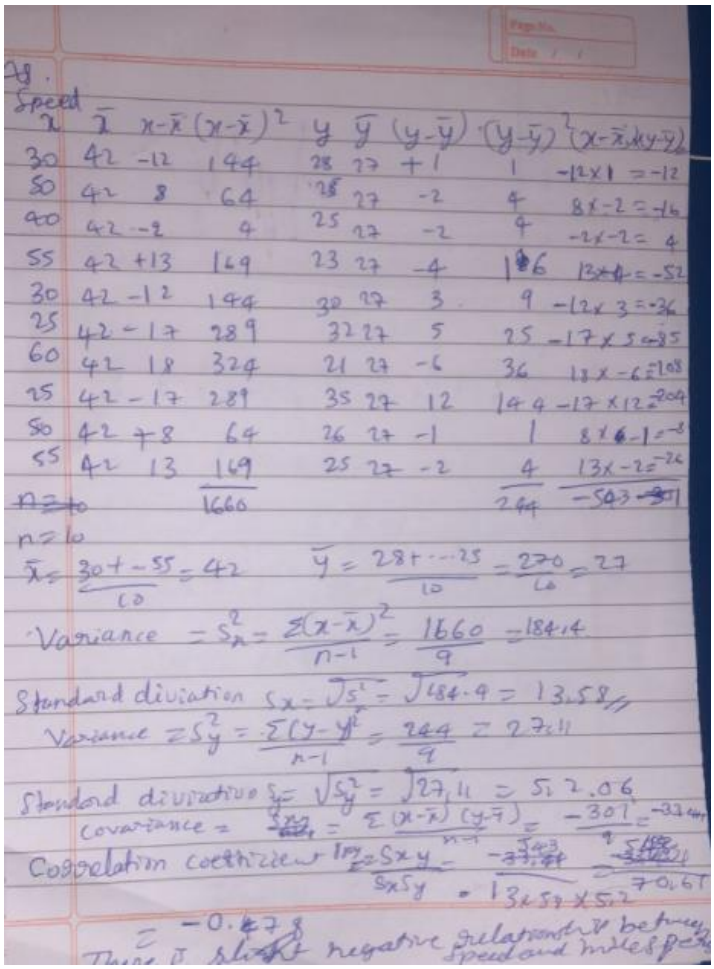
$= \frac{10.5}{2.39 \times 4.12} = \frac{10.5}{9.85} = 1.065$

correlation coefficient indicates that there is strong +ve relationship between share and ratings

48. A department of transportation's study on driving speed and miles per gallon for midsize automobiles resulted in the following data:

Speed (Miles per Hour)	30	50	40	55	30	25	60	25	50	55
Miles per Gallon	28	25	25	23	30	32	21	35	26	25

Compute and interpret the sample correlation coefficient.

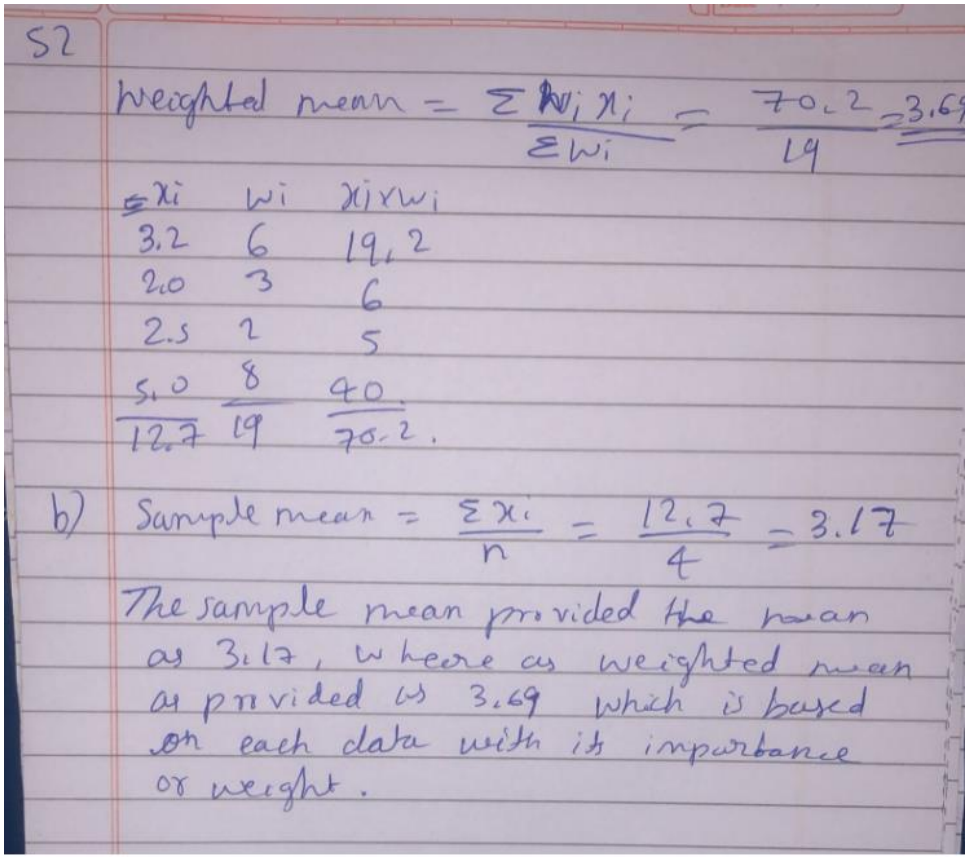


Methods

52. Consider the following data and corresponding weights.

x_i	Weight (w_i)
3.2	6
2.0	3
2.5	2
5.0	8

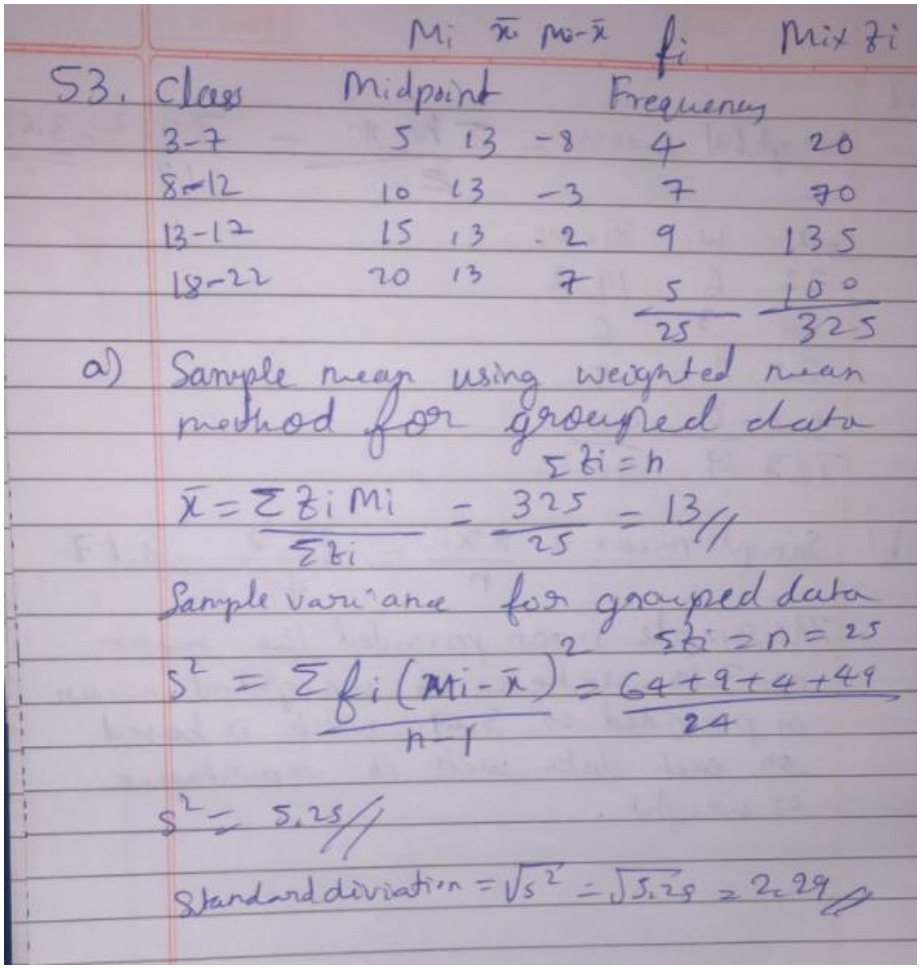
a. Compute the weighted mean.
b. Compute the sample mean of the four data values without weighting. Note the difference in the results provided by the two computations.



53. Consider the sample data in the following frequency distribution.

Class	Midpoint	Frequency
3–7	5	4
8–12	10	7
13–17	15	9
18–22	20	5

a. Compute the sample mean.
b. Compute the sample variance and sample standard deviation.



53.

Class	Midpoint M_i	$\bar{x}$	$M_i - \bar{x}$	Frequency f_i	$M_i \times f_i$
3–7	5	13	−8	4	20
8–12	10	13	−3	7	70
13–17	15	13	2	9	135
18–22	20	13	7	5	100
				25	325

a) Sample mean using weighted mean method for grouped data

$\sum f_i = n$

$\bar{x} = \frac{\sum f_i M_i}{\sum f_i} = \frac{325}{25} = 13$

Sample variance for grouped data

$\sum f_i = n = 25$

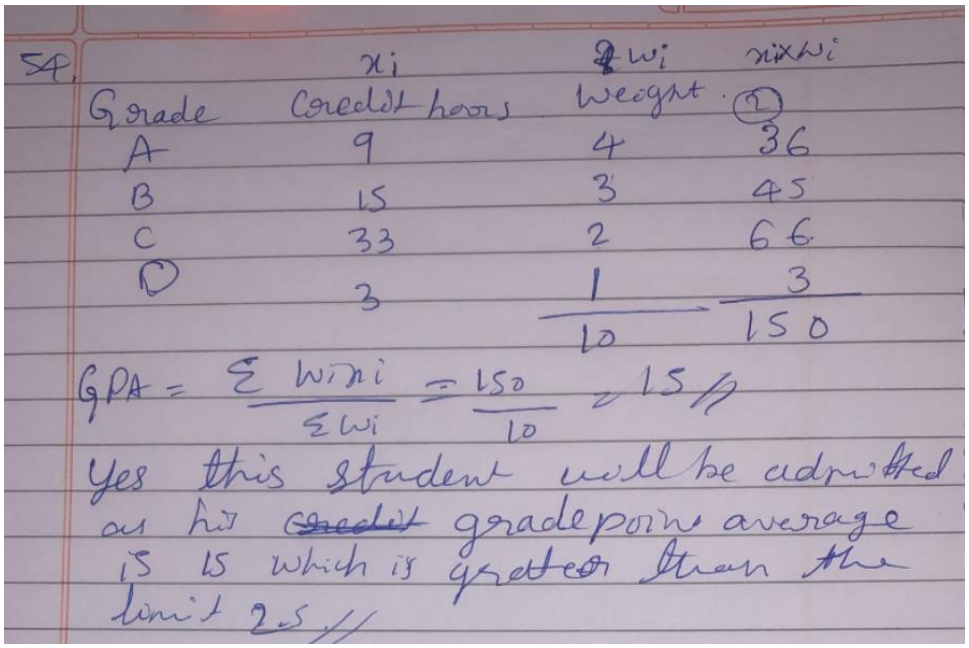
$s^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n-1} = \frac{64+9+4+49}{24}$

$s^2 = 5.25$

Standard deviation $= \sqrt{s^2} = \sqrt{5.25} = 2.29$

Applications

54. The grade point average for college students is based on a weighted mean computation. For most colleges, the grades are given the following data values: A (4), B (3), C (2), D (1), and F (0). After 60 credit hours of course work, a student at State University earned 9 credit hours of A, 15 credit hours of B, 33 credit hours of C, and 3 credit hours of D.
    a. Compute the student's grade point average.
    b. Students at State University must maintain a 2.5 grade point average for their first 60 credit hours of course work in order to be admitted to the business college. Will this student be admitted?



54.

Grade	x_i Credit hours	w_i Weight (1)	$x_i \times w_i$
A	9	4	36
B	15	3	45
C	33	2	66
D	3	1	3
		10	150

$GPA = \frac{\sum w_i x_i}{\sum w_i} = \frac{150}{10} = 15$

Yes this student will be admitted as his ~~Credit~~ grade point average is 15 which is greater than the limit 2.5

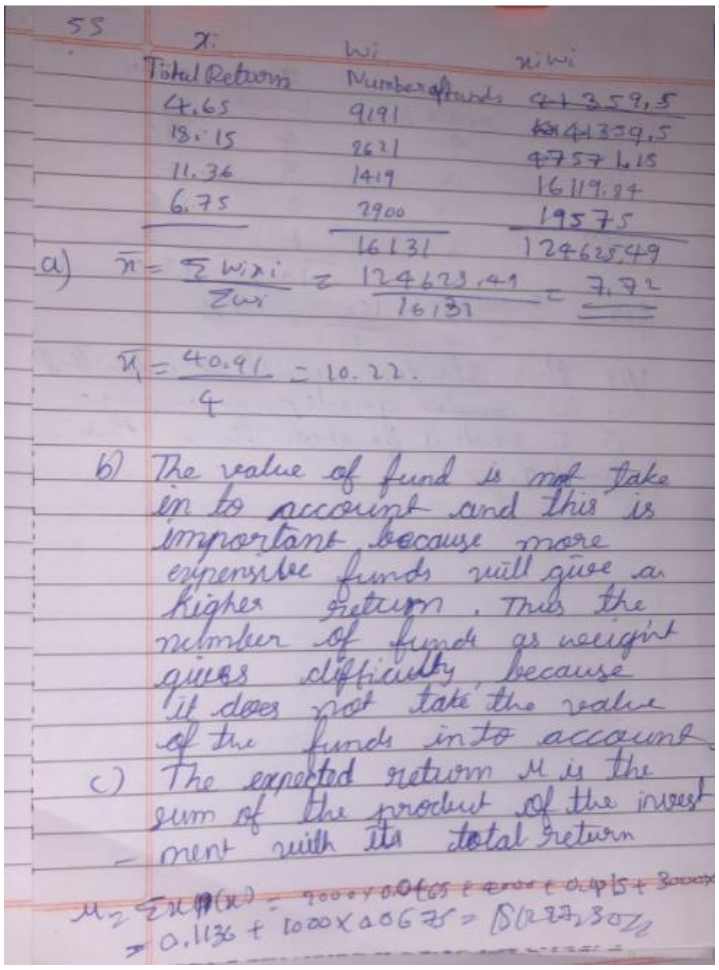
55. Morningstar tracks the total return for a large number of mutual funds. The following table shows the total return and the number of funds for four categories of mutual funds (*Morningstar Funds500*, 2008).

Type of Fund	Number of Funds	Total Return (%)
Domestic Equity	9191	4.65
International Equity	2621	18.15
Specialty Stock	1419	11.36
Hybrid	2900	6.75

a. Using the number of funds as weights, compute the weighted average total return for the mutual funds covered by Morningstar.

b. Is there any difficulty associated with using the "number of funds" as the weights in computing the weighted average total return for Morningstar in part (a)? Discuss. What else might be used for weights?

c. Suppose you had invested $10,000 in mutual funds at the beginning of 2007 and diversified the investment by placing $2000 in Domestic Equity funds, $4000 in International Equity funds, $3000 in Specialty Stock funds, and $1000 in Hybrid funds. What is the expected return on the portfolio?


55

x_i Total Return	w_i Number of funds	$x_i w_i$
4.65	9191	42739.5
18.15	2621	47576.15
11.36	1419	16119.84
6.75	2900	19575
	16131	124623.49

a) $\bar{x} = \frac{\sum w_i x_i}{\sum w_i} = \frac{124623.49}{16131} = 7.72$

$\bar{x}_1 = \frac{40.91}{4} = 10.22$

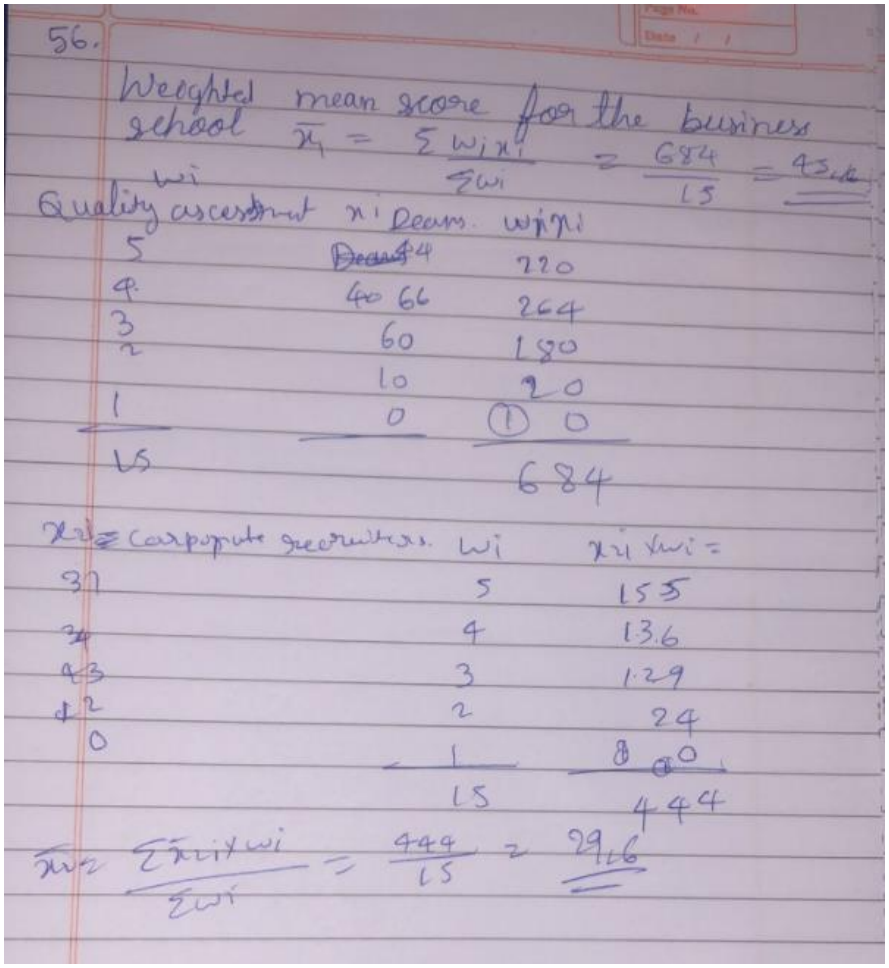
b) The value of fund is not take in to account and this is important because more expensive funds will give as higher return. Thus the number of funds as weight gives difficulty, because it does not take the value of the funds into account.

c) The expected return μ is the sum of the product of the investment with its total return

$\mu = \sum x_i p(x_i) = 10000 \times 0.0465 + 4000 \times 0.1815 + 3000 \times 0.1136 + 1000 \times 0.0675 = \$1223.307$

56. Based on a survey of 425 master's programs in business administration, the *U. S. News & World Report* ranked the Indiana University Kelley Business School as the 20th best business program in the country (*America's Best Graduate Schools,* 2009). The ranking was based in part on surveys of business school deans and corporate recruiters. Each survey respondent was asked to rate the overall academic quality of the master's program on a scale from 1 "marginal" to 5 "outstanding." Use the sample of responses shown below to compute the weighted mean score for the business school deans and the corporate recruiters. Discuss.

Quality Assessment	Business School Deans	Corporate Recruiters
5	44	31
4	66	34
3	60	43
2	10	12
1	0	0

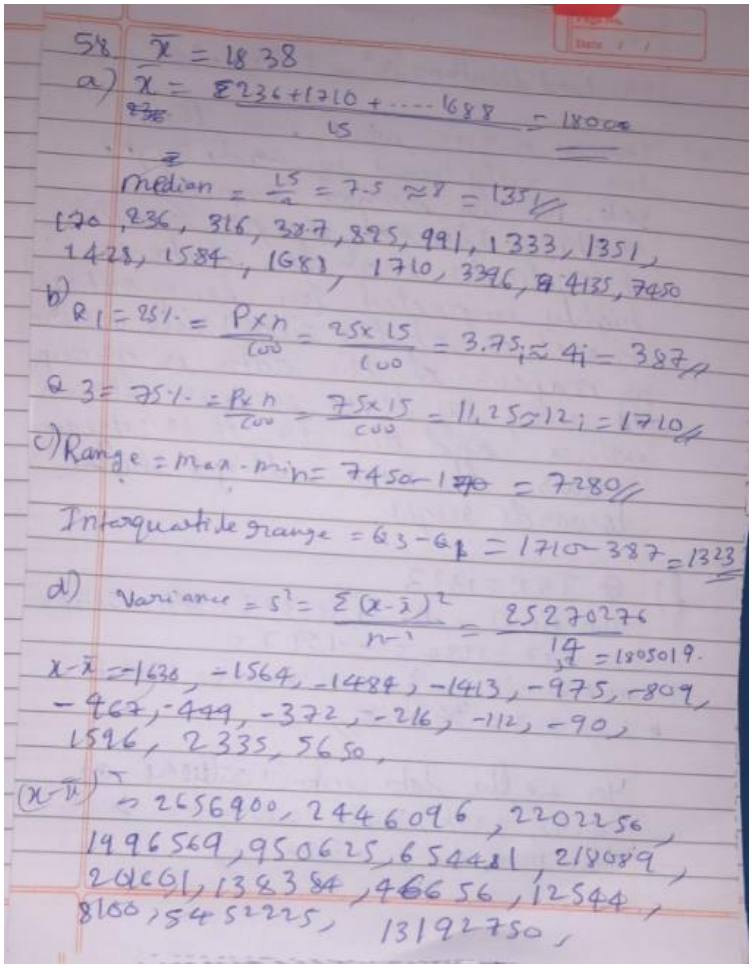


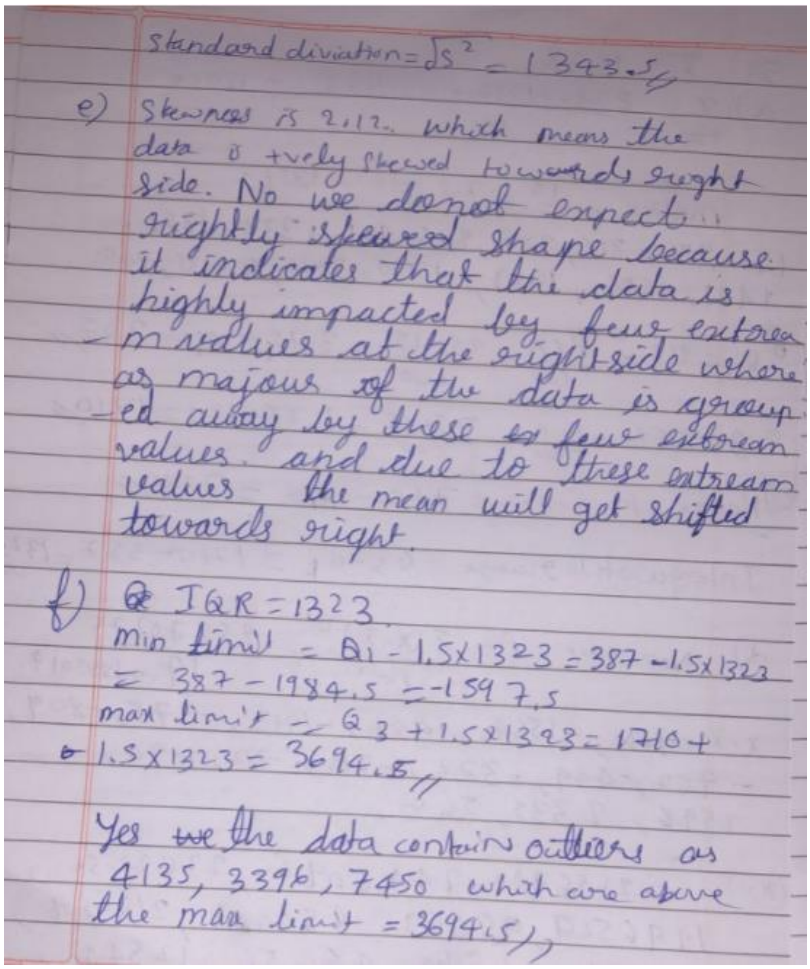
Supplementary Exercises

58. According to an annual consumer spending survey, the average monthly Bank of America Visa credit card charge was $1838 (*U.S. Airways Attaché Magazine,* December 2003). A sample of monthly credit card charges provides the following data.

236	1710	1351	825	7450
316	4135	1333	1584	387
991	3396	170	1428	1688

a. Compute the mean and median.
b. Compute the first and third quartiles.
c. Compute the range and interquartile range.
d. Compute the variance and standard deviation.
e. The skewness measure for these data is 2.12. Comment on the shape of this distribution. Is it the shape you would expect? Why or why not?
f. Do the data contain outliers?

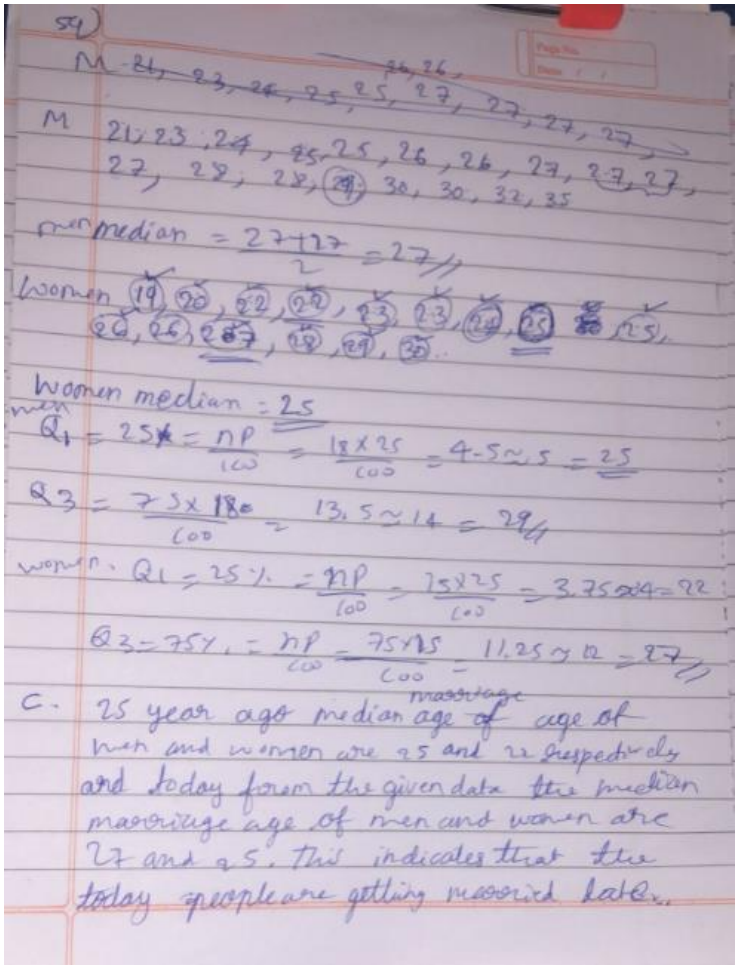




59. The U.S. Census Bureau provides statistics on family life in the United States, including the age at the time of first marriage, current marital status, and size of household (U.S. Census Bureau website, March 20, 2006). The following data show the age at the time of first marriage for a sample of men and a sample of women.

Men	26	23	28	25	27	30	26	35	28
	21	24	27	29	30	27	32	27	25
Women	20	28	23	30	24	29	26	25	
	22	22	25	23	27	26	19		

a. Determine the median age at the time of first marriage for men and women.
b. Compute the first and third quartiles for both men and women.
c. Twenty-five years ago the median age at the time of first marriage was 25 for men and 22 for women. What insight does this information provide about the decision of when to marry among young people today?



60. Dividend yield is the annual dividend per share a company pays divided by the current market price per share expressed as a percentage. A sample of 10 large companies provided the following dividend yield data (*The Wall Street Journal,* January 16, 2004).

Company	Yield %	Company	Yield %
Altria Group	5.0	General Motors	3.7
American Express	0.8	JPMorgan Chase	3.5
Caterpillar	1.8	McDonald's	1.6
Eastman Kodak	1.9	United Technology	1.5
ExxonMobil	2.5	Wal-Mart Stores	0.7

a. What are the mean and median dividend yields?
b. What are the variance and standard deviation?
c. Which company provides the highest dividend yield?
d. What is the z-score for McDonald's? Interpret this z-score.
e. What is the z-score for General Motors? Interpret this z-score.
f. Based on z-scores, do the data contain any outliers?

Q9 a) $\bar{x} = \frac{5.0 + \cdots 0.7}{10} = \frac{23}{10} = 2.3$ $n=10$

median

0.7, 0.8, 1.5, 1.6, 1.8, 1.9, 2.5, 3.5, 3.7, 5.0

median $= \frac{1.8+1.9}{2} = 1.85$

b)

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
5.0	2.3	2.7	7.29
0.8	2.3	-1.5	2.25
1.8	2.3	-0.5	0.25
1.9	2.3	-0.4	0.16
2.5	2.3	.2	0.04
3.7	2.3	1.4	1.96
3.5	2.3	1.2	1.44
1.6	2.3	-0.7	0.49
1.5	2.3	-0.8	0.64
0.7	2.3	-1.6	2.56
			17.08

Variance $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{17.08}{9} = 1.89$

Standard deviation $s = \sqrt{s^2} = 1.37$

c) Altria Group company provides highest dividend yield.

d) 1.6

$z_i = \frac{(x_i - \bar{x})}{s} = \frac{(1.6-2.3)}{1.37} = -0.51$

% yield of Mcdonald's is 0.5 less away from the mean % yield

e) GM = 3.7

$z_i = \frac{(3.7-2.3)}{1.37} = 1.02$

% yield of GM is 1.02 more away from its mean % yield.

f) 5.0

$z_i = \frac{(5.0-2.3)}{1.37} = 1.9$

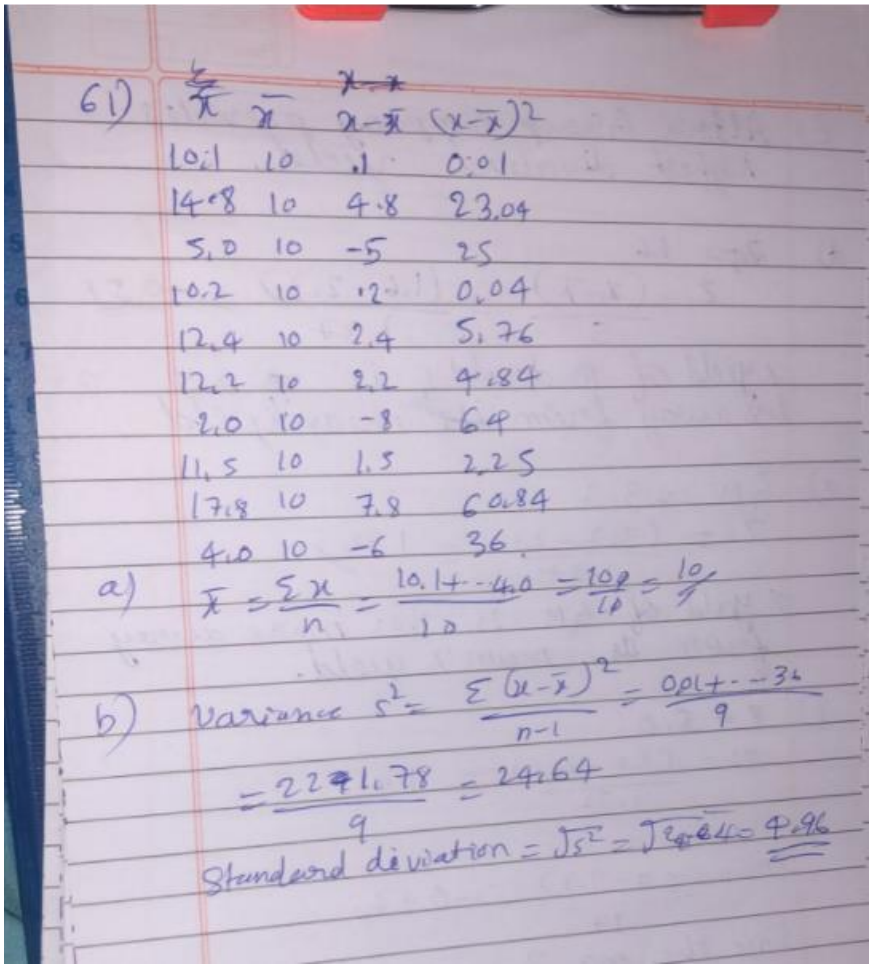
$z_{min} = \frac{(0.7-2.3)}{1.37} = -0.94$

Since the z-scores for maximum and minimum are 1.9 and -0.94 which are within the ± 3 standard deviations here there are no outliers.

61. The U.S. Department of Education reports that about 50% of all college students use a student loan to help cover college expenses (National Center for Educational Studies, January 2006). A sample of students who graduated with student loan debt is shown here. The data, in thousands of dollars, show typical amounts of debt upon graduation.

10.1	14.8	5.0	10.2	12.4	12.2	2.0	11.5	17.8	4.0

a. For those students who use a student loan, what is the mean loan debt upon graduation?
b. What is the variance? Standard deviation?



62. Small business owners often look to payroll service companies to handle their employee payroll. Reasons are that small business owners face complicated tax regulations and penalties for employment tax errors are costly. According to the Internal Revenue Service, 26% of all small business employment tax returns contained errors that resulted in a tax penalty to the owner (*The Wall Street Journal,* January 30, 2006). The tax penalty for a sample of 20 small business owners follows:

820	270	450	1010	890	700	1350	350	300	1200
390	730	2040	230	640	350	420	270	370	620

a. What is the mean tax penalty for improperly filed employment tax returns?
b. What is the standard deviation?
c. Is the highest penalty, $2040, an outlier?
d. What are some of the advantages of a small business owner hiring a payroll service company to handle employee payroll services, including the employment tax returns?

c)

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$
820	670	150	22500
390	670	-280	78400
270	670	-400	160000
730	670	60	3600
450	670	-220	48400
2040	670	1370	1876900
1010	670	340	115600
230	670	-440	193600
890	670	220	48400
640	670	-30	900
700	670	30	900
350	670	-320	102400
1350	670	680	462400
420	670	-250	62500
350	670	-320	102400
270	670	-400	160000
30	670	-370	136900
370	670	-300	90000
1200	670	530	280900
620	670	-50	2500
			3949200

$\bar{x} = \frac{\sum x}{n} = \frac{820 + \cdots 620}{20} = \frac{13400}{20} = 670$

b) Variance $= s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{3949200}{19}$ $(n = 20)$

$s^2 = 207852.63$

Standard deviation $s = \sqrt{s^2} = \sqrt{207852.63}$

$= 455.90$ //

c) 2040

$z_i = \frac{x-\bar{x}}{s} = \frac{1370}{455.90} = 3$ standard deviation

Since the 2040 within ± 3 standard deviation so it is not an outleer.

d) The advantage of a small business owner hiring a payroll service companies to handle employee pay-roll services, including the employment tax returns are they need not to face complicated tax regulations, and the accuracy of payroll service companies will be high, and it also saves time and penalty cost.