

Chapter 2 – Descriptive Statistics Organizing and Summarizing Data

Types of Variables:

Quantitative (Discrete & Continuous Variable) Numerically valued data.

Qualitative – a non-numerically valued data (arithmetic operations do not make sense).

Examples for the different types of variables:(weight, time, number of TV's in the house, number of m&m's in a bag, marital status, color of hair,)

Section 2.1 Organizing Qualitative Data

Tabular Summaries: A frequency distribution lists each category of data and the number of occurrences for each category of data.

The relative frequency is the proportion (or percent) of observations within a category and is found using the formula frequency/Total.

Multiply the relative frequency by 100 to turn it to a percent frequency.

Frequency and Relative or Percent Frequency Distributions

Example-1 : (Qualitative Variable) The party affiliation for 40 students in math 2620 is given below.

Democratic Republican Republican Democratic Democratic Republican Republican Republican
 Republican Republican Republican Republican Republican Republican Democratic Democratic
 Democratic Democratic Republican Republican Republican Republican Republican Republican
 Democratic Democratic Republican Democratic Democratic Democratic Democratic Other
 Other Other Other Other Other Other Other Other

Party	Frequency	Relative Freq.	Percent Freq.(%)
Democratic	13	$\frac{13}{40} = 0.325$	32.5
Republican	18	$\frac{18}{40} = 0.450$	45.0
Other	9	$\frac{9}{40} = 0.225$	22.5
Total	40	$\frac{40}{40} = 1$	100%

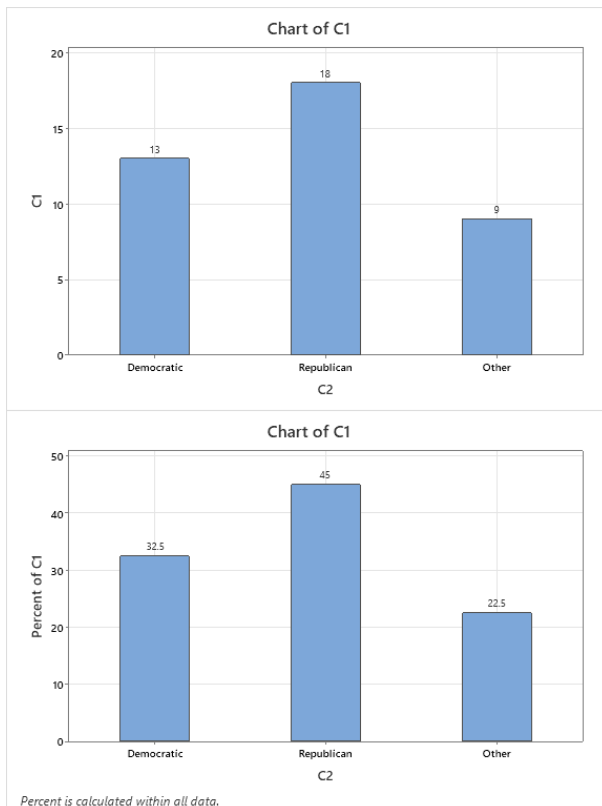
Graphical Summaries: Two methods for displaying qualitative data are Pie Charts and Bar Graphs.

Bar Graphs: A bar graph is constructed by labeling each category of data on either the horizontal or vertical axis and the frequency or relative frequency of the category on the other axis. Rectangles of equal width are drawn for each category. The height of each rectangle represents the category's frequency or relative frequency.

Pie Chart: A pie chart is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

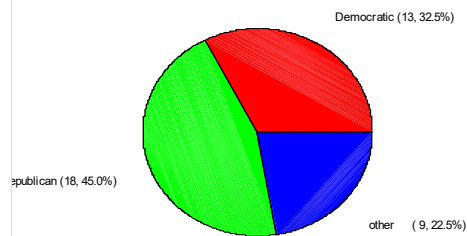
Example-2 (Reference-Example-1)

BAR GRAPH



PIE CHART

Political Party Affiliations

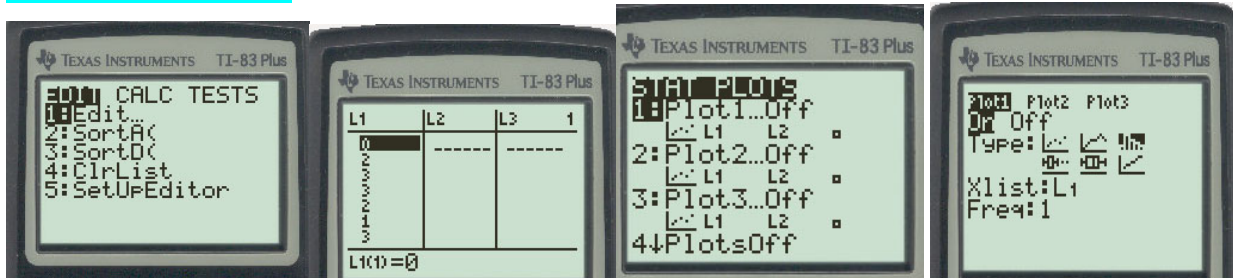


Below is the solution to # 8 on homework-“Section 2.1 and 2.2 Online – MyStatLab”.

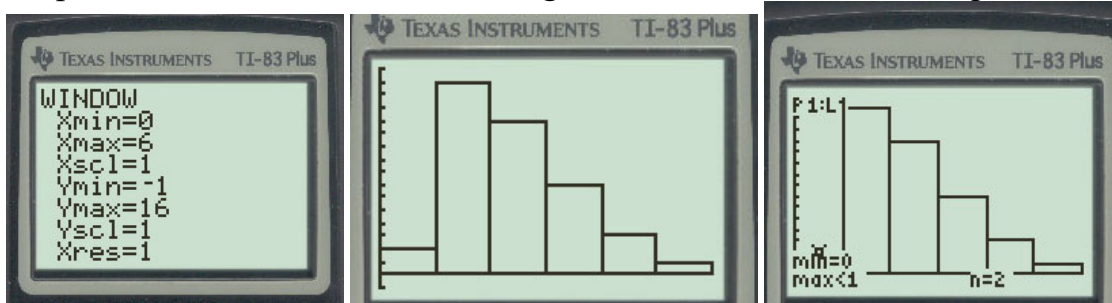
Using the TI 83/84 Calculator to do homework problem #8

Stat, 1:Edit and then enter the data in L1.

2nd and Y to access the STATPLOT, Chose 1 for Plot 1, put cursor on “ON” and enter to turn on Plot 1, move arrow down and chose the Histogram (newer calculators the graphs are in one row), then move the cursor to the XList: and do 2nd and 1 to choose L1,



then hit Window to adjust the data for the window, Xmin=0, Xmax=6, Xscl=1, Ymin=-1, Ymax=16, Yscl=1, Xres=1, then hit GRAPH, then TRACE to extract the frequencies, move the arrow to the right to trace the rest of the frequencies.



Televisions	Frequency
0	2
1	15
2	12
3	7
4	3
5	1

8. A researcher wanted to determine the number of televisions in households. He conducts a survey of 40 randomly selected households and obtains the data in the accompanying table. Complete parts (a) through (h) below.

¹ Click the icon to view the table of television counts.

(a) Are these data discrete or continuous? Explain.

- A. The given data are continuous because they can take on any real value.
- B. The given data are continuous because they can only have whole number values.
- C. The given data are discrete because they can only have whole number values.
- D. The given data are discrete because they can take on any real value.

(b) Construct a frequency distribution of the data.

Televisions	Frequency
0	2
1	15
2	12
3	7
4	3
5	1

0 2 2 3 5 4 2 2 1 1
 2 1 3 4 2 1 1 2 2 2
 3 3 1 2 2 1 1 1 2 1
 3 3 1 1 1 3 1 1 4 0
 3

(c) Construct a relative frequency distribution of the data.

Televisions	Relative Frequency
0	0.050
1	0.375
2	0.300
3	0.175
4	0.075
5	0.025

$2/40 = 0.050$
 $15/40 = 0.375$
 $12/40 = 0.300$
 $7/40 = 0.175$
 $3/40 = 0.075$
 $1/40 = 0.025$

(Type integers or decimals. Do not round.)

(d) What percentage of households in the survey have three televisions?

17.5 %

(Type an integer or a decimal. Do not round.)

(e) What percentage of households in the survey have four or more televisions?

10 %

$0.075 + 0.025 = 0.10$

(Type an integer or a decimal. Do not round.)

Homework-Section 2.1 and 2.2 Online - MyStatLab

Section 2.2 Organizing Quantitative Data

Frequency and Relative Frequency Distributions:

Example-3 (Quantitative-Discrete Variable)

A nursery school offers programs for 4-year olds ranging from 1-day-a-week program to 5-day-a-week program. To help in planning, the school's director surveyed parents regarding the type of program they prefer. The following data, which represents the number of days, were obtained.

2 3 3 4 5 5 3 3 4 4 1 2 1
1 2 2 3 2 2 2 1 3 2 5 2 2

Construct the frequency and relative frequency distributions and answer the following.

- (a) What percentage of parents prefer 2-day-a week program?
(b) What percentage of parents prefers 4-day-a-week or 5-day-a-week program?

Solution

In the data set there are 4-1s, 10-2s, 6-3s, 3-4s, and 3-5s with 26 total data items. For example, the relative frequency of the data value 1 is $(4/26)100 = 15.4\%$.

Number of Days	Frequency	Relative Freq.	Percent Freq.(%)
1	4	0.154	15.4%
2	10	0.385	38.5%
3	6	0.231	23.1%
4	3	0.115	11.5%
5	3	0.115	11.5%
Total	26	1	100%

- (a) The percentage of parents that prefer 2-day-a-week program is 38.5%.
(b) The percentage of parents that prefer 4-day-a-week program or 5 day-a- week program is $11.5 + 11.5 = 23\%$.

Guidelines of Constructing a Frequency distribution and Histogram

Guidelines for Determining the Lower Class Limit of the First Class and Class Width

Choosing the Lower Class Limit of the First Class

Choose the smallest observation in the data set or a convenient number slightly lower than the smallest observation in the data set. For example, in Table 12, the smallest observation is 8.28. A convenient lower class limit of the first class is 8.

Determining the Class Width

- Decide on the number of classes. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer classes you should have. For example, we might choose 12 classes for the data in Table 12.
- Determine the class width by computing

$$\text{Class width} \approx \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$$

Round this value *up* to a convenient number. For example, using the data in Table 12, we obtain class width $\approx \frac{19.43 - 8.28}{12} = 0.929$. We round this up to 1 because this is an easy number to work with. Rounding up may result in fewer classes than were originally intended.

Applying these guidelines, to the five-year rate of return data, we would end up with the frequency distribution shown in Table 13.

Graphical Summaries

NOTE: Questions-What are we looking for when we look at data?

- a. The shape of the distribution of the data.
- b. The symmetry or skewness of the data.
- c. The center of the data.
- d. The spread of the data.

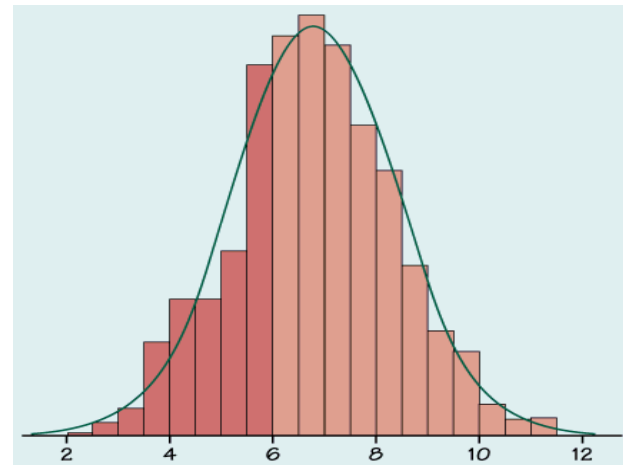


Table 12**Five-Year Rate of Return of Mutual Funds**

10.94	14.60	12.80	16.00	11.93	15.68	9.03	13.40
10.53	13.98	13.86	12.36	13.54	9.94	13.93	13.63
14.12	14.88	14.77	13.13	8.28	19.43	12.98	13.16
12.26	14.20	14.80	13.26	13.67	10.08	14.86	8.71
12.17	10.26	15.22	13.36	13.55	13.90	15.64	12.80

Source: Morningstar.com

Approach To construct a frequency distribution, first create classes of equal width. Table 12 has 40 observations that range from 8.28 to 19.43, so we decide to create the classes such that the lower class limit of the first class is 8 (a little smaller than the smallest data value) and the class width is 1. There is nothing magical about the choice of 1 as a class width. We could have selected a class width of 3 or any other class width. We choose a class width that we think will nicely summarize the data. If our choice doesn't accomplish this, we can always try another. The second class has a lower class limit $8 + 1 = 9$. The classes cannot overlap, so the upper class limit of the first class is 8.99. Continuing in this fashion, we obtain the following classes:

$$\begin{array}{c} 8-8.99 \\ 9-9.99 \\ \vdots \\ 19-19.99 \end{array}$$

This gives us 12 classes. Tally the number of observations in each class, count the tallies, and create the frequency distribution. Divide the frequency of each class by 40, the number of observations, to obtain the relative frequency.

Solution Tally the data as shown in the second column of Table 13. The third column shows the frequency of each class. From the frequency distribution, we conclude that a five-year rate of return between 13% and 13.99% occurs with the most frequency. The fourth column shows the relative frequency of each class. So, 32.5% of the large-blended mutual funds had a five-year rate of return between 13% and 13.99%.

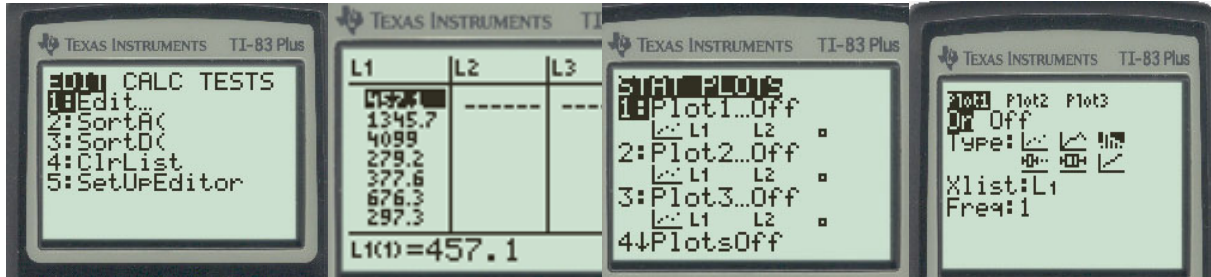
Table 13

Class (Five-year rate of return)	Tally	Frequency	Relative Frequency
8-8.99		2	$2/40 = 0.05$
9-9.99		2	$2/40 = 0.05$
10-10.99		4	$4/40 = 0.10$
11-11.99		1	$1/40 = 0.025$
12-12.99		6	$6/40 = 0.15$
13-13.99		13	$13/40 = 0.325$
14-14.99		7	$7/40 = 0.175$
15-15.99		3	$3/40 = 0.075$
16-16.99		1	$1/40 = 0.025$
17-17.99		0	$0/40 = 0$
18-18.99		0	$0/40 = 0$
19-19.99		1	$1/40 = 0.025$

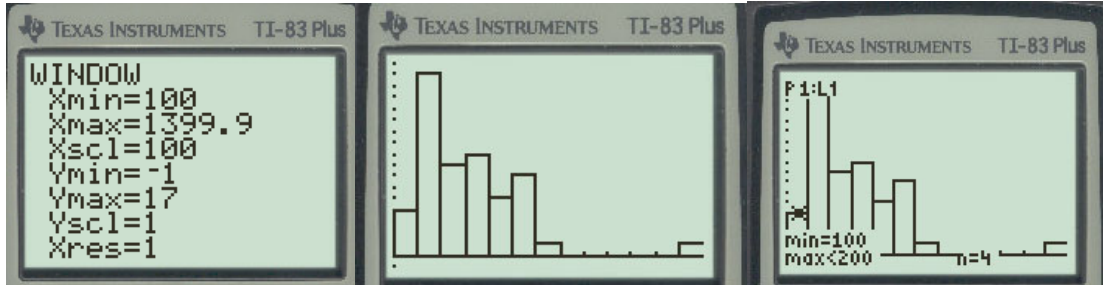
Using the TI 83/84 Calculator to do homework problem #9

Stat, 1:Edit and then enter the data in L1.

2nd and Y to access the STATPLOT, Chose 1 for Plot 1, put cursor on "ON" and enter to turn on Plot 1, move arrow down and chose the Histogram (newer calculators the graphs are in one row), then move the cursor to the XList: and do 2nd and 1 to choose L1,



then hit Window to adjust the data for the window, Xmin=100, Xmax=1399.9, Xscl=100, Ymin=-1, Ymax=17, Yscl=1, Xres=1, then hit GRAPH, then TRACE to extract the frequencies, move the arrow to the right to trace the rest of the frequencies.



Violent Crimes	Frequency
100-199.9	4
200-299.9	16
300-399.9	8
400-499.9	9
500-599.9	5
600-699.9	7
700-799.9	1

Violent Crimes	Frequency
800-899.9	0
900-999.9	0
1000-1099.9	0
1100-1199.9	0
1200-1299.9	0
1300-1399.9	1

9. Violent crimes include murder, forcible rape, robbery, and aggravated assault. The data to the right represent the violent-crime rate (crimes per 100,000 population) by region of a certain country. Complete parts (a) through (f) below.

end at 1399.9

457.1	1345.7	409.9	279.2	377.6	676.3	297.3
644.5	611.7	257.1	488.4	403.9	177.3	261.1
417.2	419.9	616.7	261.8	202.9	665.9	229.9
524.3	277.5	119.6	287.8	338.8	490.2	
465.1	235.3	589.8	704.6	506.8	200.7	
333.7	508.6	467.7	158.9	245.4	120.7	
308.3	326.7	509.7	314.9	383.2	234.3	
635.7	287.2	254.5	622.6	256.7	335.7	

start at 100

(a) If thirteen classes are to be formed, choose an appropriate lower class limit for the first class and a class width

An appropriate lower class limit for the first class is (1) 100 with a class width of (2) 100.

(b) Construct a frequency distribution.

Violent Crimes	Frequency	Violent Crimes	Frequency
100-199.9	<u>4</u>	800-899.9	<u>0</u>
200-299.9	<u>16</u>	900-999.9	<u>0</u>
300-399.9	<u>8</u>	1000-1099.9	<u>0</u>
400-499.9	<u>9</u>	1100-1199.9	<u>0</u>
500-599.9	<u>5</u>	1200-1299.9	<u>0</u>
600-699.9	<u>7</u>	1300-1399.9	<u>1</u>
700-799.9	<u>1</u>		

Below are some videos how to create a histogram on TI 83 or 84.

https://mediaplayer.pearsoncmg.com/assets/sst5e_2_2_4_Histogram_cont_data_by_hand_TI84

https://mediaplayer.pearsoncmg.com/assets/02_02_EX4_ti

Homework-Section 2.1 and 2.2 Online - MyStatLab