regardless of which scale the object is being measured in (e.g. meters or yards). This is because there is a natural zero.

Ratio Data

- ordered, constant scale, natural zero
- e.g., height, weight, age, length

One can think of nominal, ordinal, interval, and ratio as being ranked in their relation to one another. Ratio is more sophisticated than interval, interval is more sophisticated than ordinal, and ordinal is more sophisticated than nominal.

8.5 Frequency Distribution

Frequency is how often something occurs. The frequency (**f**) of a particular observation is the number of times the observation occurs in the data.

Distribution

The *distribution* of a variable is the pattern of frequencies of the observation.

Frequency Distribution

It is a representation, either in a graphical or tabular format, which displays the number of observations within a given interval. Frequency distributions are usually used within a statistical context.

8.5.1 Frequency Distribution Tables

A frequency distribution table is one way you can organize data so that it makes more sense. Frequency distributions are also portrayed as frequency tables, histograms, orpolygons. Frequency distribution tables can be used for both categorical and numeric variables. The intervals of frequency table must be mutually exclusive and exhaustive. Continuous variables should only be used with class intervals. By counting frequencies, we can make a frequency distribution table. Following examples will figure out procedure of construction of frequency distribution table.

Example 1

For example, let's say you have a list of IQ scores for a gifted classroom in a particular elementary school. The IQ scores are: 118, 123, 124, 125, 127, 128, 129, 130, 130, 133, 136, 138, 141, 142, 149, 150, 154. That list doesn't tell you much about anything. You could draw a frequency distribution table, which will give a better picture of your data than a simple list.

Step 1:

- Figure out how many classes (categories) you need. There are no hard rules about how many classes to pick, but there are a couple of general guidelines:
- Pick between 5 and 20 classes. For the list of IQs above, we picked 5 classes.
- Make sure you have a few items in each category. For example, if you have 20 items, choose 5 classes (4 items per category), not 20 classes (which would give you only 1 item per category).

Step 2:

- Subtract the minimum data value from the maximum data value. For example, our the IQ list above had a minimum value of 118 and a maximum value of 154, so:
  154 – 118 = 36

Step 3:

- Divide your answer in Step 2 by the number of classes you chose in Step 1. 36 / 5 = 7.2

Step 4:

- Round the number from Step 3 up to a whole number to get the class width. Rounded up, 7.2 becomes 8.

Step 5:

- Write down your lowest value for your first minimum data value: The lowest value is 118

Step 6:

- Add the class width from Step 4 to Step 5 to get the next lower class limit: 118 + 8 = 126

Step 7:

- Repeat Step 6 for the other minimum data values (in other words, keep on adding your class width to your minimum data values) until you have created the number of classes you chose in Step 1. We chose 5 classes, so our 5 minimum data values are:
  118
  126 (118 + 8)
  134 (126 + 8)

142 (134 + 8)
150 (142 + 8)

Step 8:

- Write down the upper class limits. These are the highest values that can be in the category, so in most cases you can subtract 1 from class width and add that to the minimum data value. For example:
  118 + (8 – 1) = 125
  118 – 125
  126 – 133
  134 – 142
  143 – 149
  150 – 157

Step 9:

- Add a second column for the number of items in each class, and label the columns with appropriate headings:

IQ	Number
118 – 125	
126 – 133	
134 – 142	
143 – 149	
150 – 157	

Step 10:

- Count the number of items in each class, and put the total in the second column. The list of IQ scores are: 118, 123, 124, 125, 127, 128, 129, 130, 130, 133, 136, 138, 141, 142, 149, 150, 154.

IQ	**Number**
118 – 125	4
126 – 133	6
134 – 142	4
143 – 149	1
150 – 157	2

Example 2

A survey was taken in Lahore. In each of 20 homes, people were asked how many cars were registered to their households. The results were recorded as follows:

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0

Use the following steps to present this data in a frequency distribution table.

1. Divide the results (***x***) into intervals, and then count the number of results in each interval. In this case, the intervals would be the number of households with no car (0), one car (1), two cars (2) and so forth.
2. Make a table with separate columns for the interval numbers (the number of cars per household), the tallied results, and the frequency of results in each interval. Label these columns *Number of cars*, *Tally* and *Frequency*.
3. Read the list of data from left to right and place a tally mark in the appropriate row. For example, the first result is a 1, so place a tally mark in the row beside where 1 appears in the interval column (*Number of cars*). The next result is a 2, so place a tally mark in the row beside the 2, and so on. When you reach your fifth tally mark, draw a tally line through the preceding four marks to make your final frequency calculations easier to read.
4. Add up the number of tally marks in each row and record them in the final column entitled *Frequency*.

Your frequency distribution table for this exercise should look like this:

Table 1. Frequency table for the number of cars registered in each household

Number of cars (x)	Tally	Frequency (f)
0	𝍸	4
1	𝍸 I	6
2	𝍸	5
3	III	3
4	II	2

By looking at this frequency distribution table quickly, we can see that out of 20 households surveyed, 4 households had no cars, 6 households had 1 car, etc.

Relative frequency and percentage frequency

An analyst studying these data might want to know not only how long batteries last, but also what proportion of the batteries falls into each class interval of battery life.

This *relative frequency* of a particular observation or class interval is found by dividing the frequency (**f**) by the number of observations (**n**): that is, (**f ÷ n**). Thus:

Relative frequency = frequency ÷ number of observations

The *percentage frequency* is found by multiplying each relative frequency value by 100. Thus:

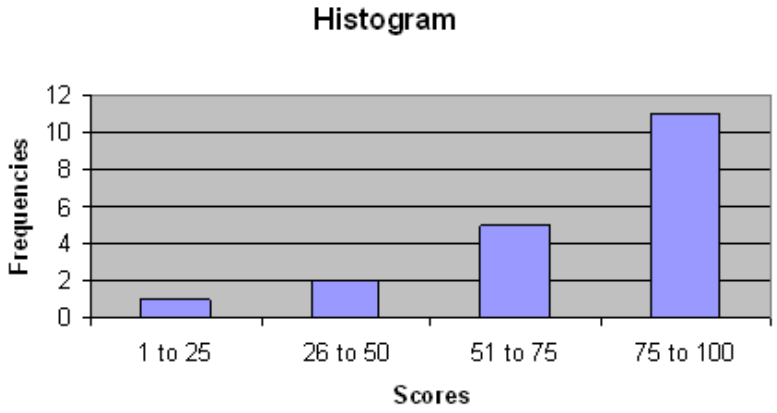
Percentage frequency = relative frequency X 100 = f ÷ n X 100

8.5 Interpreting Test Scores by Graphic Displays of Distributions

The data from a frequency table can be displayed graphically. A graph can provide a visual display of the distributions, which gives us another view of the summarized data. For example, the graphic representation of the relationship between two different test scores through the use of scatter plots. We learned that we could describe in general terms the direction and strength of the relationship between scores by visually examining the scores as they were arranged in a graph. Some other examples of these types of graphs include histograms and frequency polygons.

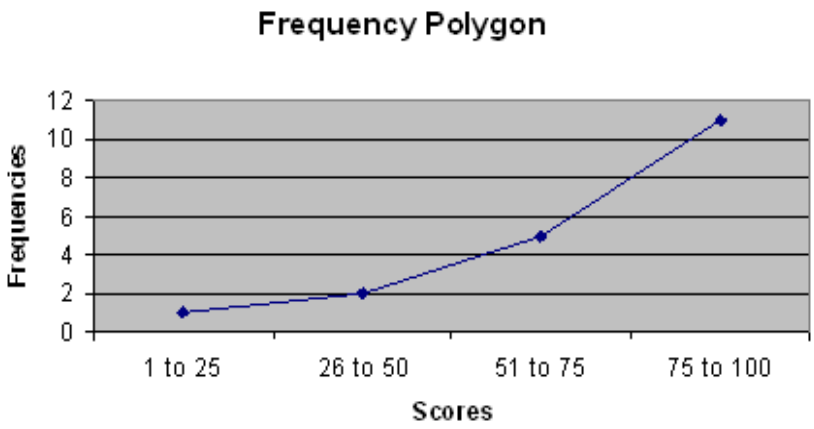
A **histogram** is a bar graph of scores from a frequency table. The horizontal x-axis represents the scores on the test, and the vertical y-axis represents the frequencies. The frequencies are plotted as bars.

Histogram of Mid-Term Language Arts Exam



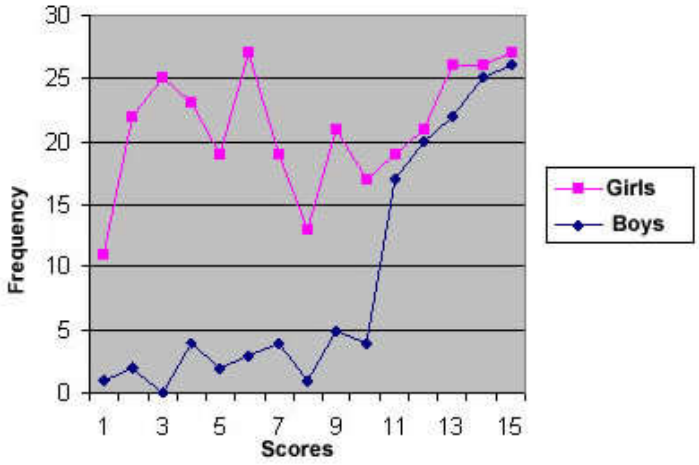
A **frequency polygon** is a line graph representation of a set of scores from a frequency table. The horizontal x-axis is represented by the scores on the scale and the vertical y-axis is represented by the frequencies.

Frequency Polygon of Mid-Term Language Arts Exam



A frequency polygon could also be used to compare two or more sets of data by representing each set of scores as a line graph with a different color or pattern. For example, you might be interested in looking at your students' scores by gender, or comparing students' performance on two tests (see Figure 9.4).

Frequency Polygon of Midterm by Gender



Frequency polygons are a graphical device for understanding the shapes of distributions. They serve the same purpose as histograms, but are especially helpful in comparing sets of data. Frequency polygons are also a good choice for displaying cumulative frequency distributions.

To create a frequency polygon, start just as for histograms, by choosing a class interval. Then draw an X-axis representing the values of the scores in your data. Mark the middle of each class interval with a tick mark, and label it with the middle value represented by

the class. Draw the Y-axis to indicate the frequency of each class. Place a point in the middle of each class interval at the height corresponding to its frequency. Finally, connect the points. You should include one class interval below the lowest value in your data and one above the highest value. The graph will then touch the X-axis on both sides.

A frequency polygon for 642 psychology test scores is shown in Figure 1. The first label on the X-axis is 35. This represents an interval extending from 29.5 to 39.5. Since the lowest test score is 46, this interval has a frequency of 0. The point labeled 45 represents the interval from 39.5 to 49.5. There are three scores in this interval. There are 150 scores in the interval that surrounds 85.

You can easily discern the shape of the distribution from Figure 1. Most of the scores are between 65 and 115. It is clear that the distribution is not symmetric inasmuch as good scores (to the right) trail off more gradually than poor scores (to the left). In the terminology of Chapter 3 (where we will study shapes of distributions more systematically), the distribution is **skewed**.

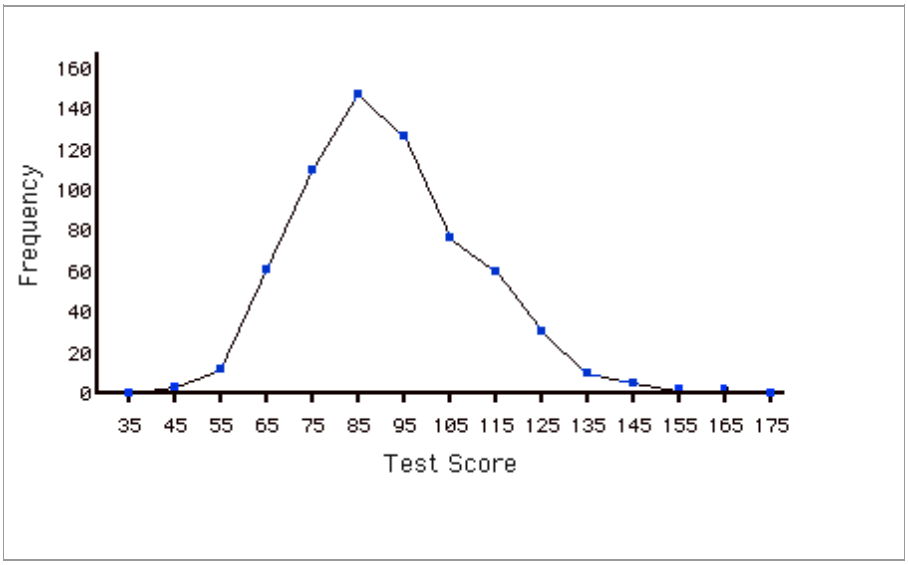


Figure 1: Frequency polygon for the psychology test scores.

A cumulative frequency polygon for the same test scores is shown in Figure 2. The graph is the same as before except that the Y value for each point is the number of students in the corresponding class interval plus all numbers in lower intervals. For example, there are no scores in the interval labeled "35," three in the interval "45,"and 10 in the interval "55."Therefore the Y value corresponding to "55" is 13. Since 642 students took the test, the cumulative frequency for the last interval is 642.

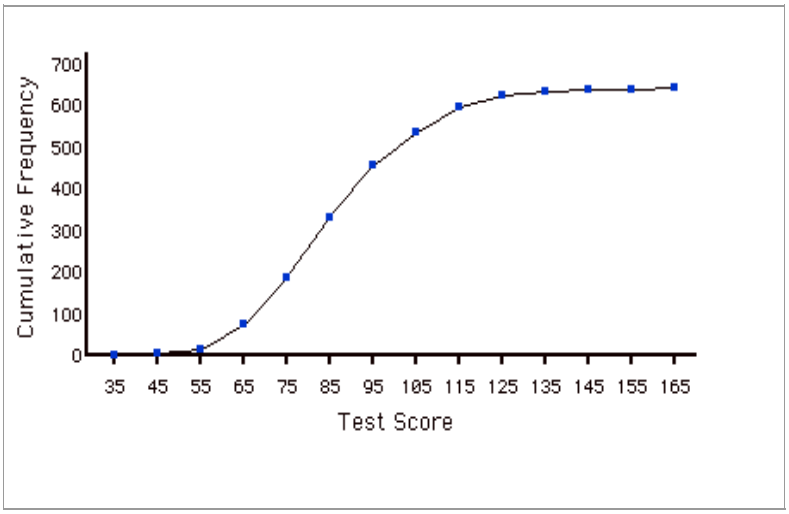


Figure 2: Cumulative frequency polygon for the psychology test scores.

Frequency polygons are useful for comparing distributions. This is achieved by overlaying the frequency polygons drawn for different data sets. Figure 3 provides an example. The data come from a task in which the goal is to move a computer mouse to a target on the screen as fast as possible. On 20 of the trials, the target was a small rectangle; on the other 20, the target was a large rectangle. Time to reach the target was recorded on each trial. The two distributions (one for each target) are plotted together in Figure 3. The figure shows that although there is some overlap in times, it generally took longer to move the mouse to the small target than to the large one.

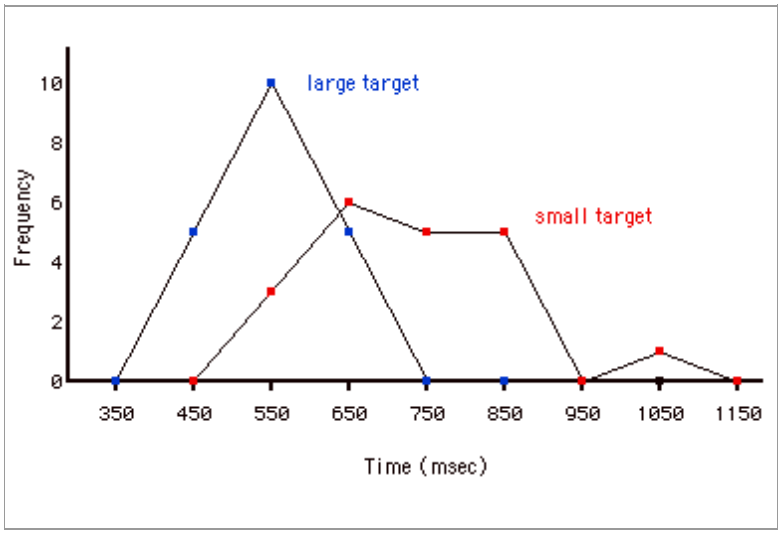


Figure 3: Overlaid frequency polygons.

It is also possible to plot two cumulative frequency distributions in the same graph. This is illustrated in Figure 4 using the same data from the mouse task. The difference in distributions for the two targets is again evident.

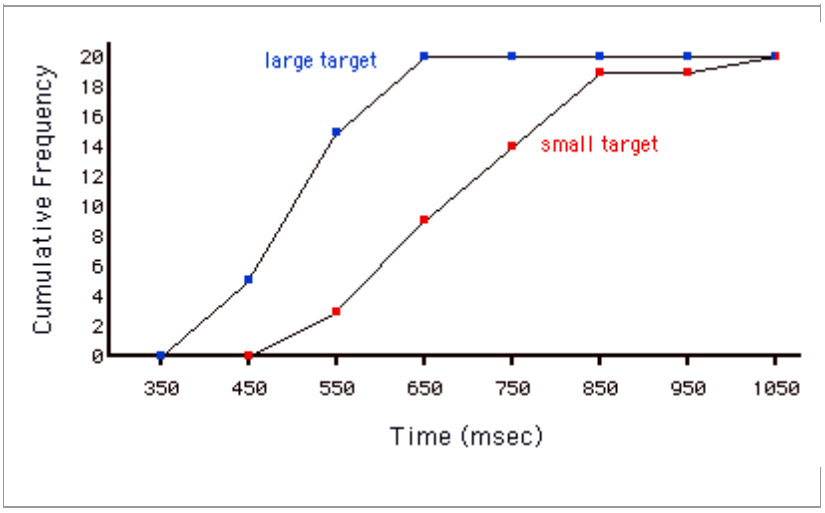


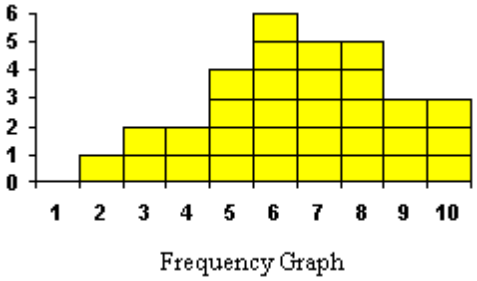
Figure 4: Overlaid cumulative frequency polygons.

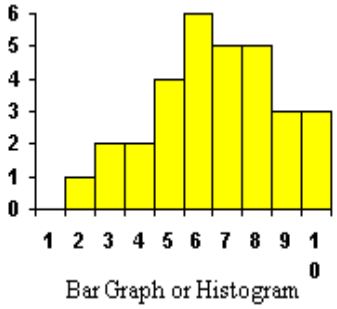
The raw scores for the 10 pt. quiz are:

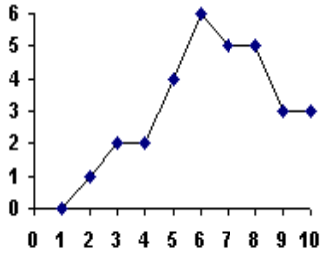
10 9 8 8 7 7 6 6 5 4 2 10 9 8 8 7 6 6 5 5 3 10 9 8 7 7 6 6 5 4 3

Draw frequency graph, bar graph, frequenvy polygone, and frequency curve

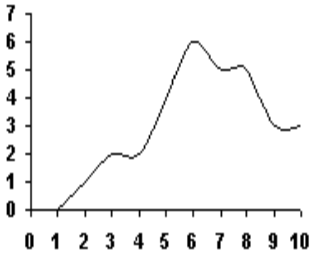
Solution







Frequency Polygon



Frequency Curve

8.7 Measures of Central Tendency

Suppose that a teacher gave the same test to two different classes and following results are obtained:

Class 1: 80%, 80%, 80%, 80%, 80%

Class 2: 60%, 70%, 80%, 90%, 100%

If you calculate the mean for both sets of scores, you get the same answer: 80%. But the data of two classes from which this mean was obtained was very different in the two cases. It is also possible that two different data sets may have same mean, median, and mode. For example:

Class A: 72 73 76 76 78

Class B: 67 76 76 78 80

Therefore class A and class B has same mean, mode, and median.

The way that statisticians distinguish such cases as this is known as measuring the variability of the sample. As with measures of central tendency, there are a number of ways of measuring the variability of a sample.

Probably the simplest method is to find the range of the sample, that is, the difference between the largest and smallest observation. The range of measurements in Class 1 is 0, and the range in class 2 is 40%. Simply knowing that fact gives a much better understanding of the data obtained from the two classes. In class 1, the mean was 80%, and the range was 0, but in class 2, the mean was 80%, and the range was 40%.

Statisticians use summary measures to describe patterns of data. **Measures of central tendency** refer to the summary measures used to describe the most "typical" value in a set of values.

Here, we are interested in the typical, most representative score. There are three most common measures of central tendency are mean, mode, and median. A teacher should be familiar with these common measures of central tendencies.

8.7.1 Mean

The mean is simply the arithmetic average. It is sum of the scores divided by the number of scores. it is computed by adding all of the scores and dividing by the number of scores. When statisticians talk about the mean of a population, they use the Greek letter μ to refer to the mean score. When they talk about the mean of a sample, statisticians use the symbol $\overline{X}$ to refer to the mean score.

It is symbolized as: $$\overline{X} = \frac{\sum X}{N}$$

$\overline{X}$ (read as "X-Bar") when computed on a sample

Computation - Example: find the mean of 2,3,5, and 10.

$$\overline{X} = \frac{\sum X}{N} = \frac{2+3+5+10}{4} = \frac{20}{4} = 5$$

Since means are typically reported with one more digit of accuracy that is present in the data, I reported the mean as 5.0 rather than just 5.

Example 1

The marks of seven students in a mathematics test with a maximum possible mark of 20 are given below:

15 13 18 16 14 17 12

Find the mean of this set of data values.

Solution:

$$\text{Mean} = \frac{\text{Sum of all data values}}{\text{Number of data values}}$$

$$= \frac{15+13+18+16+14+17+12}{7}$$

$$= \frac{105}{7}$$

$$= 15$$

So, the mean mark is 15.

Symbolically, we can set out the solution as follows:

$$\bar{x} = \frac{\sum x}{n}$$
$$= \frac{15+13+18+16+14-17+12}{7}$$
$$= \frac{105}{7}$$
$$= 15$$

So, the mean mark is 15.

When working with grouped frequency distributions, we can use an approximation:

$$\bar{X} = \frac{\sum (Mdpt * f)}{N}$$

Where Mdpt. is midpoint of the group

For example:

Interval	Midpoint	f	Mid*f
95-99	97	1	97
90-94	92	3	276
85-89	87	5	435
80-84	82	6	492
75-79	77	4	308
70-74	72	3	216
65-69	67	1	67
60-64	62	2	124
		f=25=N	**Mid*f=2015**

:

$$\bar{X} = \frac{\sum (Mdpt * f)}{N}$$

$$\bar{X} = \frac{2015}{25} = 80.6$$

When computed on the raw data, we get:

$$\bar{X}=\frac{\sum X}{N}=\frac{2014}{25}=80.56$$

Thus the formula for computing the mean with grouped data gives us a good approximation of the actual mean. In fact, when we report the mean with one decimal more accuracy than what is in the data, the two techniques give the same result.

8.7.2 Median or M_d

The score that cuts the distribution into two equal halves (or the middle score in the distribution).

The **median** of a set of data values is the middle value of the data set when it has been arranged in ascending order. That is, from the smallest value to the highest value.

Example

The marks of nine students in a geography test that had a maximum possible mark of 50 are given below:

47 35 37 32 38 39 36 34 35

Find the median of this set of data values.

Solution:

Arrange the data values in order from the lowest value to the highest value:

32 34 35 35 36 37 38 39 47

The fifth data value, 36, is the middle value in this arrangement.

$\therefore$ Median = 36

In general:

Median = $\frac{1}{2}(N+1)$th value, where n is the number of data values in the sample.

If the number of values in the data set is even, then the **median** is the average of the two middle values.

Fortunately, there is a formula to take care of the more complicated situations, including computing the median for grouped frequency distributions.

$$M_d=L+\left(\frac{\frac{N}{2}-n_b}{n_w}\right)i$$