Statistics

Raw Data is information obtained by observing values of a variable. This Raw Data falls into two categories: Qualitative Data and Quantitative Data.

Qualitative data is the data obtained by observing values of a qualitative variable. It is descriptive and conceptual. For example, the Math Concepts class is: 1. Easy

2. It is given at Vanier
3. It course number is
4. It's given on Mondays and Wednesdays.

Quantitative data is the data obtained by observing a quantitative variable. It is statistical. It is more concise and close-ended. It can be generated through tests, experiments, surveys and market reports.

We will only be concerned with Quantitative Data in this course.

The following examples will all be about quantitative data.

Frequency Distribution for Quantitative Data

Example

Suppose the following table gives a frequency distribution of the Stanford-Binet (IQ) intelligence test scores for 75 adults.

IQ Scores	Frequency
80 - 94	8
95-109	14
110-124	24
125-139	16
140-154	13

Sum of the numbers in the Frequency column is 75.

According to this table, eight of the individuals observes have IQ scores between 80 and 94, fourteen have scores between 95 and 109, twenty four individuals have IQ scores between 110 and 124 and so on and so forth.

The IQ score is the quantitative variable.

In the above example, we have five classes of the study being made. In each class, there are 15 IQ scores. For example, the class 80 – 94 includes the scores 80,81,82,…,92,93,94. These are 15 scores in total.

Example

Group the following weights of 21 individuals into classes 100 to under 125, 125 to under 150, 150 to under 175, and so on and so forth.

111 120 127 129 130 145 145 150 153 155 160 161 167 170 171 174 175 177 179 180 180

So the frequency table would look as follows:

Weight	Frequency
100 to under 125	2
125 to under 150	5
150 to under 175	9
175 ti under 200	5

Again the sum of the numbers is 21 which is the number of individuals being observed.

Example

The price for 500 aspirin tablets is determined for each twenty randomly selected stores as part of a larger consumer study. The prices are as follows:

2.50 2.95 2.65 3.10 **3.15** 3.05 3.05 2.60 2.70 2.75
2.80 2.85 2.80 3.00 3.00 2.90 2.90 2.85 2.85 2.50

If we want to group the data into 7 classes, then what we have to do is the following:
Maximum price – Minimum price = 3.15 – 2.50 = 0.65 = 65 cents

65/7 ≅ 9 So the classes would be the following: 2.50 – 2.59
2.60 – 2.69
2.70 – 2.79
2.80 – 2.89
2.90 – 2.99
3.00 – 3.09
3.10 – 3.20

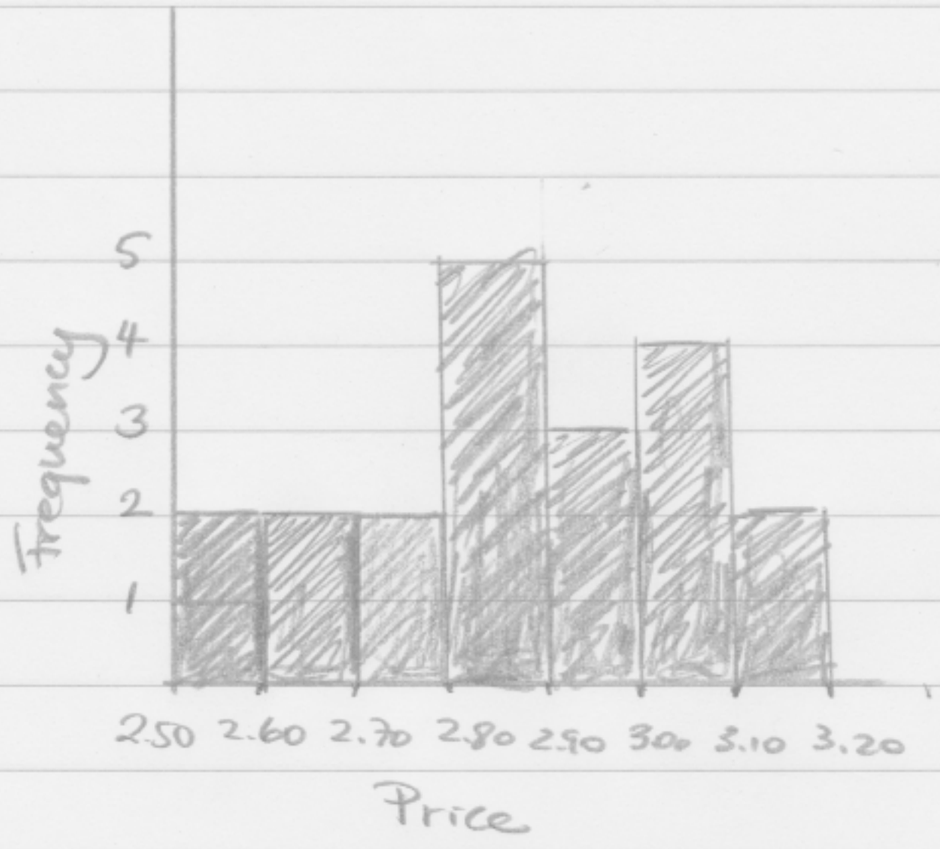
And the Frequency Distribution table would be the following:

Price	Frequrncy
2.50-2.59	2
2.60-2.69	2
2.70-2.79	2
2.80-2.89	5
2.90-2.99	3
3.00-3.09	4
3.10-3.20	2

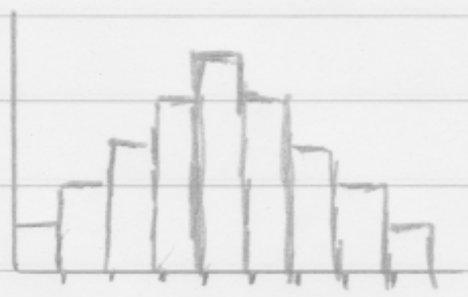
Again, the sum of the numbers in this column is 20.

Histograms

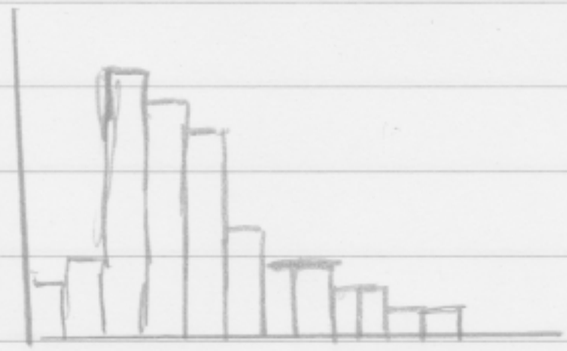
A histogram is a graph that displays the classes on the horizontal axis and the frequencies on the vertical axis. The frequency of each class is represented by a vertical bar whose height is equal to the frequency of the class.



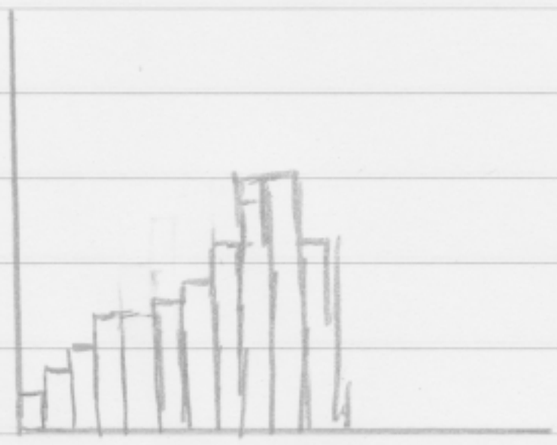
A symmetric Histogram is of the shape:



Skewed to the right histogram has a longer tail on the right



Skewed to the left histogram has a longer tail on the left.



Relative Frequency Distribution

Example

The following table shows the frequency distribution of the contents in milliliters of a sample of 25 one-liter bottles of soda.

Content	Frequency	Relative Frequency
970 to less than 990	5	5/25 = 0.20
990 to less than 1010	10	10/25=0.40
1010 to less than 1030	5	5/25=0.20
1030 to less than 1050	3	3/25=0.12
1050 to less than 1070	2	2/25=0.08
	25 observations	The sum of the numbers in this column is 1.00

Relative Frequency = Frequency/number of observations