

## Displaying data

It is nearly always useful to use graphical methods to illustrate your data. We shall describe in this section just a few of the methods available.

### Discrete data: frequency table and bar chart

Suppose that you have collected some discrete data. It will be difficult to get a "feel" for the distribution of the data just by looking at it in list form. It may be worthwhile constructing a frequency table or bar chart.

The frequency of a value is the number of observations taking that value.

A *frequency table* is a list of possible values and their frequencies.

A *bar chart consists of bars* corresponding to each of the possible values, whose heights are equal to the frequencies.

### Continuous data: histograms

When the variety is continuous, we do not look at the frequency of each value, but group the values into intervals. The plot of frequency against interval is called a histogram. Be careful to define the interval boundaries unambiguously.

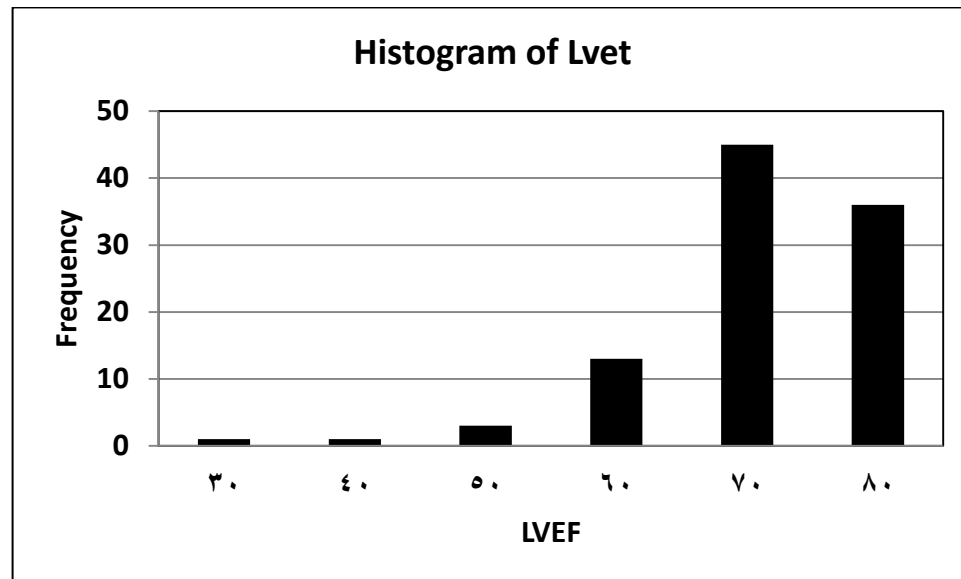
### Example

The following data are the left ventricular ejection fractions (LVEF) for a group of 99 heart transplant patients. Construct a frequency table and histogram.

62 64 63 70 63 69 65 74 67 77 65 72 65 77 71 79 75 78  
64 78 72 32 78 78 80 69 69 65 76 53 74 78 59 79 77 76  
72 76 70 76 76 74 67 65 79 63 71 70 84 65 78 66 72 55  
74 79 75 64 73 71 80 66 50 48 57 70 68 71 81 74 74 79  
79 73 77 80 69 78 73 78 78 66 70 36 79 75 73 72 57 69  
82 70 62 64 69 74 78 70 76.

## Frequency table

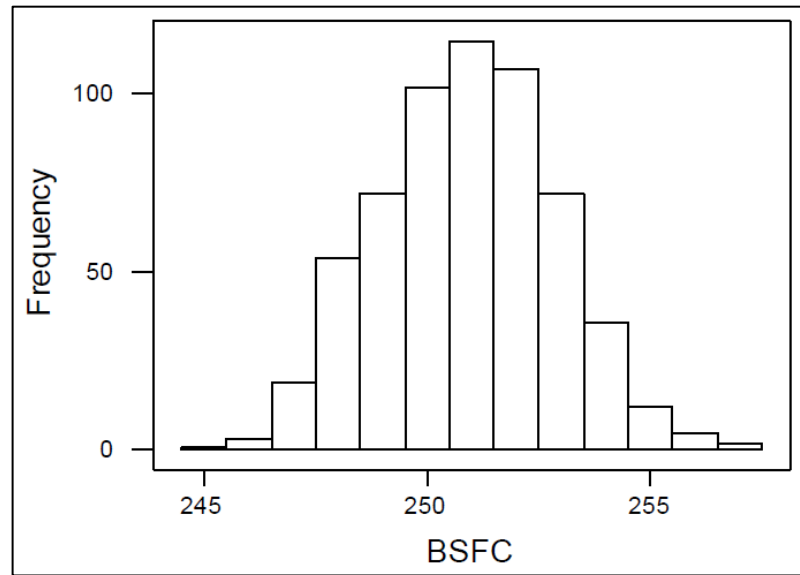
Frequency table LVEF	Tallies	Frequency
24.5 - 34.5		1
34.5 - 44.5		1
44.5 - 54.5		3
54.5 - 64.5		13
64.5 - 74.5	 	45
74.5 - 84.5	 	36



Bar charts and histograms provide an easily understood illustration of the distribution of the data. As well as showing where most observations lie and how variable the data are, they also indicate certain "danger signals" about the data.

### Normally distributed data

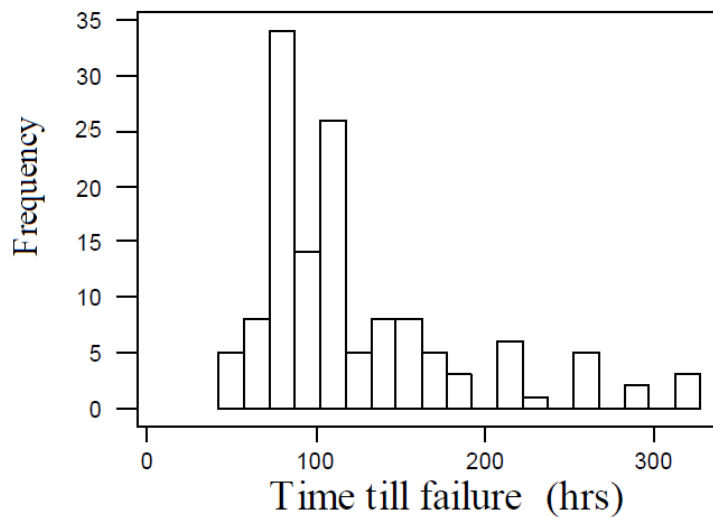
The histogram is bell-shaped, like the probability density function of a Normal distribution. It appears, therefore, that the data can be modelled by a Normal distribution. (Other methods for checking this assumption are available.) Similarly, the histogram can be used to see whether data look as if they are from an Exponential or Uniform distribution.



### Very skew data

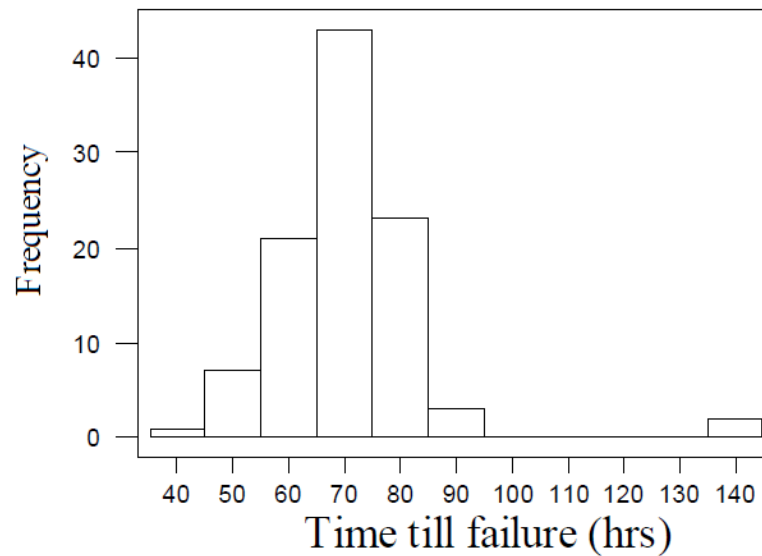
The relatively few large observations can have an undue influence when comparing two or more sets of data. It might be worthwhile using a transformation e.g. taking logarithms.

**Bimodality** This may indicate the presence of two sub-populations with different characteristics. If the subpopulations can be identified it might be better to analyse them separately.



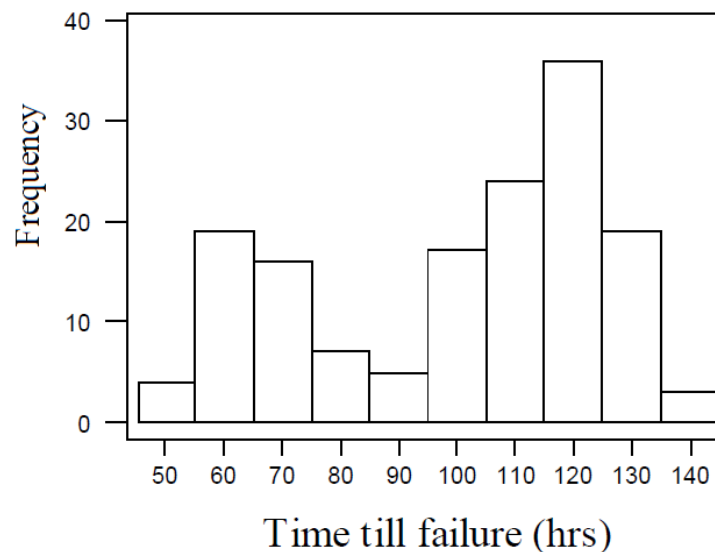
## Outliers

The data appear to follow a pattern with the exception of one or two values. You need to decide whether the strange values are simply mistakes, are to be expected or whether they are correct but unexpected. The outliers may have the most interesting story to tell.



## Bimodality

This may indicate the presence of two sub-populations with different characteristics. If the subpopulations can be identified it might be better to analyse them separately.



**A frequency:** is the number of times a given datum occurs in a data set.

**A relative frequency ( rf ):** To find the relative frequencies, divide each frequency by the total number of frequencies. For example in the sample below we have 20 cases. Relative frequencies can be written as fractions, percent's, or decimals.

Frequency Table of Student Work Hours w/ Relative Frequency

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15
3	5	$\frac{5}{20}$ or 0.25
4	3	$\frac{3}{20}$ or 0.15
5	6	$\frac{6}{20}$ or 0.30
6	2	$\frac{2}{20}$ or 0.10
7	1	$\frac{1}{20}$ or 0.05
	<b>Total= 20</b>	

The sum of the relative frequency column is  $20/20$ , or 1.

**Cumulative relative frequency** is the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row.

Frequency Table of Student Work Hours w/ Relative and Cumulative Relative Frequency

DATA	Frequency	Relative frequency	Cumulative Relative Frequency
2	3	$\frac{3}{20}$ or 0.15	0.15
3	5	$\frac{5}{20}$ or 0.25	0.15+0.25 = 0.40
4	3	$\frac{3}{20}$ or 0.15	0.40+0.15 = 0.55
5	6	$\frac{6}{20}$ or 0.30	0.55+0.30 = 0.85
6	2	$\frac{2}{20}$ or 0.10	0.85+0.10 = 0.95
7	1	$\frac{1}{20}$ or 0.05	0.95+0.05 = 1.00

The following table represents the precipitation amount, in millimeters, Frequency

Precipitation (mm)	Frequency	Relative frequency	Cumulative Relative Frequency
59.95 - 61.95	5	$\frac{5}{100} = 0.05$	0.05
61.95 - 63.95	3	$\frac{3}{100} = 0.03$	0.05 + 0.03 = 0.08
63.95 - 65.95	15	$\frac{15}{100} = 0.15$	0.08 + 0.15 = 0.23
65.95 - 67.95	40	$\frac{40}{100} = 0.40$	0.23 + 0.40 = 0.63
67.95 - 69.95	17	$\frac{17}{100} = 0.17$	0.63 + 0.17 = 0.80
69.95 - 71.95	12	$\frac{12}{100} = 0.12$	0.80 + 0.12 = 0.92
71.95 - 73.95	7	$\frac{7}{100} = 0.07$	0.92 + 0.07 = 0.99
73.95 - 75.95	1	$\frac{1}{100} = 0.01$	0.99 + 0.01 = 1.00
	<b>Total = 100</b>	<b>Total = 1.00</b>	

While the relative *frequency percentage* (rf%) computed by multiplying (rf) by 100.

Precipitation (mm)	Frequency	Relative frequency (rf)	Relative frequency percentage (rf %)
0- 5	2	0.0667	0.0667 *100 = 6.67 %
6-10	5	0.1667	0.1667*100 = 16.67 %
11-15	5	0.1667	0.1667*100 = 16.67 %
16-20	11	0.3667	0.3667*100 = 36.67 %
21-25	4	0.1333	0.1333*100 = 13.33 %
26-30	3	0.1000	0.1000*100 = 10.00 %
	<b>n = 30</b>	<b>Total = 1.0001 ≈ 1.0</b>	<b>Total = 100.01 ≈ 100</b>

From this frequency table we can quickly identify information such as 11 of precipitation events (36.67% of all precipitation events) are in the 16-20 (mm) height range, and there are 2 of precipitation events (6.67% of all precipitation events ) are in the 0-5 (mm) less range.

**Construct the Increasing and Decreasing cumulative frequency table.**

The **Increasing cumulative frequency table** (ICF) is constructed by copying the first frequency and adding it to the next frequency. The result is to be added to the next frequency until you reach the last cumulative frequency which is equal to the total number of observations.

The **Decreasing cumulative frequency table** (DCF) follows the same procedure, it's just that it starts from the last frequency and accumulates upward and the first greater than cumulative frequency is equal to the total number of observations.

<b>Classes</b>	<b>Frequency</b>	<b>(ICF)</b>	<b>(DCF)</b>
60-62	5	5	100
63-65	18	23	95
66-68	42	65	77
69-71	27	92	35
72-74	8	100	8
	Total= 100		