PRELIMINARIES

Contents

If viewing on screen you can click on a contents item above to jump to the page the item is on

Right at the start it is probably worth reminding ourselves about types of distribution, measures of central tendency and some measures of dispersion.

The symbols you will come across in this section will probably be very well known to you, but if they are not you can consult the symbol glossary at the end of this document.

1. Types of Distributions

Distributions of test scores can, in theory be classified according to :

Modality

Kurtosis

Skew

Modality simply refers to the number of peaks a distribution has. The commonest form of a multi-modal distribution is probably a bi-modal one, i.e. it has two peaks.

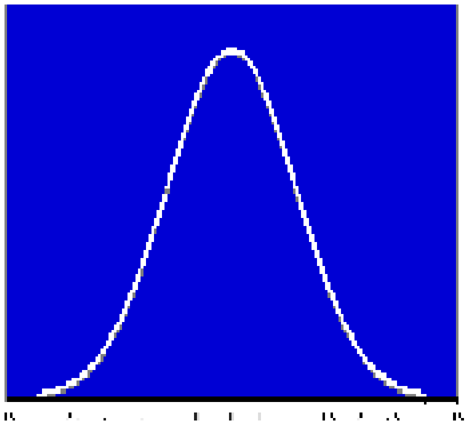
If such a distribution is obtained by giving a test to a sample of people, it is always wise to wonder if there might not be two very different sub-samples in the group tested.

Kurtosis

Kurtosis is the “peakedness” of the distribution.

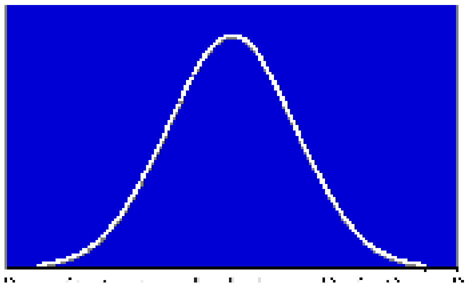
Tall pointy distributions are

leptokurtic



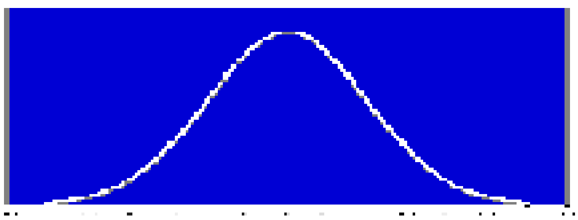
those in between are

mesokurtic.



and low flatter ones are

platykurtic



Clinical practitioners will very rarely (if ever) need to calculate kurtosis. But, for the record, and just in case, the formula for measuring degree of kurtosis is:

$$ku = \frac{\Sigma x^4 / N}{\sigma_x^4} - 3$$

In a normal distribution the value of *ku* is zero.

A positive value for the index means that the distribution is on the leptokurtic side.

A negative value means that it is on the platykurtic side.

Skew

Skew is the degree to which the distribution is asymmetrical. In a skewed distribution scores are mainly clumped together at on end, but with a long tail towards the opposite end.

If the tail is towards the higher scores the distribution is said to be positively skewed.

If the tail is towards the lower score, the distribution is said to be negatively skewed.

The formula for measuring skew is:

$$Sk = \frac{\Sigma x^3 / N}{\sigma_x^3}$$

If a distribution is normal, the value of *Sk* is zero. If the scores are negatively skewed, i.e., to the low scoring end, then the value will be negative, and if the scores are skewed to the high scoring end they will be positive. It is unlikely that you will ever need to use this formula, but you might see references to a distribution with a skew of a given value.

2. Measures of Central Tendency

The measures of central tendency are:

The mean

The median

The mode

Of these the most important and useful one for our purposes is the mean, whose well-known formula is of course

$$M_x = \frac{\Sigma X}{N}$$

The **median** is that point in a distribution above and below which 50 percent of scores fall

The **mode** is the most commonly occurring score

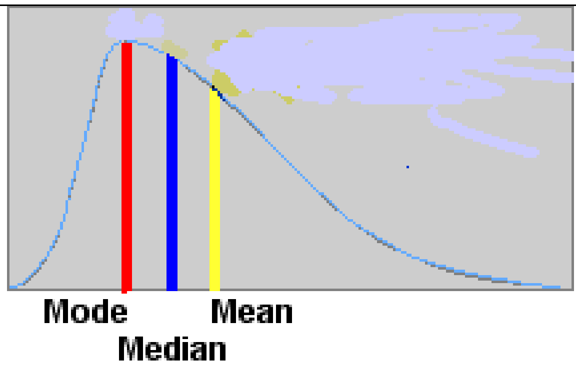
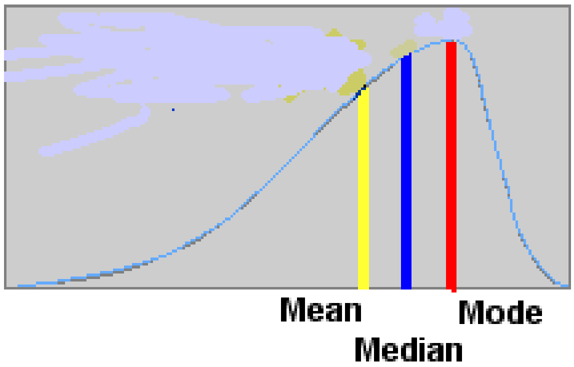
In symmetrical distributions (of which the normal distribution is one) the mean, median, and mode all have the same value, but in a skewed distribution (a distribution where most scores are bunched up towards one end with a tapering tail towards the other) this is not the case.

In such distributions the order going from the tapered end to the bulky end will be

Mean

Median

Mode

Thus in the case of positive skew:	
and in the case of negative skew:	

When it comes to psychological tests, skewed distributions are rare, but they do occur sometimes, e.g., in diagnostic tests. When they do occur they can be, and often are, transformed into more normal distributions, as we shall see.

In tests of ability, one of the main factors affecting skew is item difficulty. The harder the items, the greater will be the proportion of people getting low scores and thus the distribution will be positively skewed. The reverse will be true if the items are too easy. The bright folks will be banging their heads against the test's ceiling.

Test yourself

Here are three sets of scores:

A. 1, 1, 2, 3

B. 1, 2, 2, 3

C. 1, 2, 3, 3

1. Which set of scores is symmetrical?

2. Which set of scores is positively skewed?

3. Which set of scores is negatively skewed?

4. Calculate the mean, mode and median for sets A, B, and C

Answers

1. B

2. A

3. C

4.

	Set A	Set B	Set C
Mean	1.75	2	2.25
Median	1.5	2	2.5
Mode	1	2	3

3. Measures of dispersion

These include

The range

The inter-quartile range

The standard deviation

The range is simply the difference between the highest and lowest scores. For most purposes this measure is of little use or interest.

The inter-quartile range (sometimes reported when the median is used as a measure of central tendency) states the range in which the middle 50 percent of cases lie. It is also the range used to define the average range of intelligence on many intelligence tests.

The standard deviation is a much more useful measure of dispersion. It is the square root of the variance, and its formula is:

$$\sigma_X = \sqrt{\frac{\Sigma\left(X - M_X\right)^2}{N}}$$

As in the field of psychological testing we are interested in population values rather than the strictly sample ones, it is usual, especially if samples are small, to divide by $(N - 1)$ rather than N in the above equation.

Putting some of this together

Let's calculate a standard deviation, and some other values too.

What are the mean, the median and the mode of the following six scores?

1, 2 , 3, 3, 4, 5.

The histogram of this distribution of scores looks like this

We know from what was said earlier that, as this is a symmetric distribution, the mean, mode and median will have the same value.

Mean = 3

Mode = 3

Median = 3

Recalling that positive skew should make the mean move towards the skewed end, let's now change the score of 1 to a score of 7. That will skew the distribution to the higher end, and the sum of the scores will equal 24. So the mean will rise from 3 to 4. The mode will remain the same and the median will rise slightly to half way between 3 and 4.

4. Transforming Distributions

Let's go back a step or two and look at some ways of transforming distributions.

In general, skewed distributions can be normalised by using one of the following transformations.

In general, skewed distributions can be normalised by using one of the following transformations.

Square root transformation

Logarithmic transformation (using base 10; 2; or e)

Inverse transformation (i.e. using 1/*X* – this increases the size of small numbers and decreases the size of big ones).

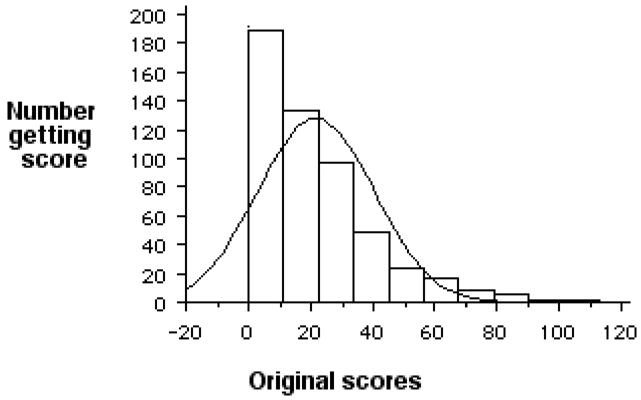
These three transformations are useful in dealing with positive skew.

To deal with negative skew the distribution is first reflected (ie., multiplied by minus 1), then a constant is added to the distribution of negative scores to make the smallest score equal plus 1. Then the transformation is applied.

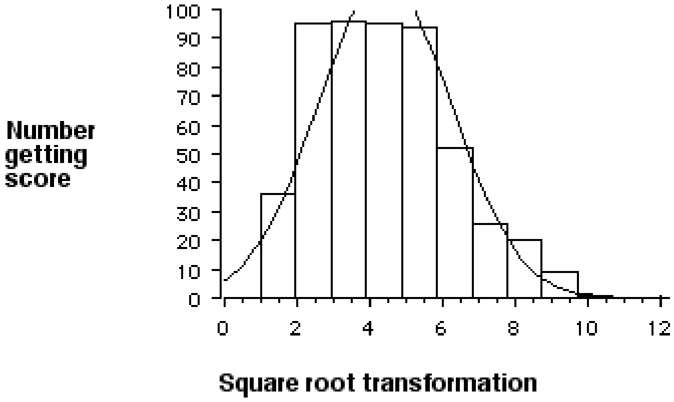
Unfortunately at this stage, because we have multiplied everything by minus 1 to reflect it, the normalised scores are in the wrong order. We have made the first into the last. The people who scored highest now get the lowest scores. So the final stage is to reflect the transformed distribution again so that the first shall be first again.

To give some idea of what happens when data are transformed the distributions of scores from a large clinical sample to whom a test of Dissociation was administered are shown below (data courtesy of Dr J T Quinn).

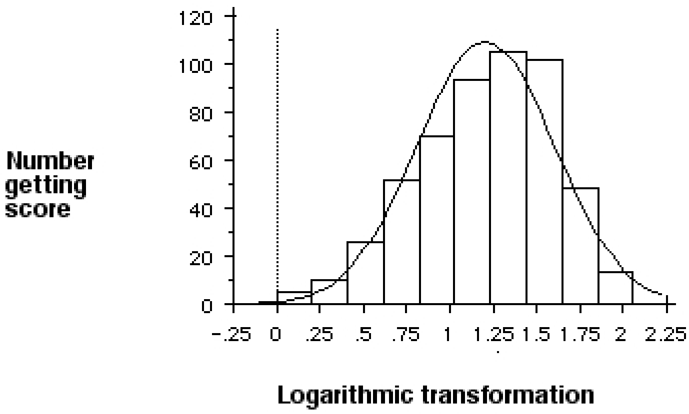
First , let's look at the distribution of original scores . As you can see it is positively skewed. That makes sense as we would not expect most people to have many dissociative symptoms . (In the graphs there is a super-imposed normal curve so that you can more easily see the departures from normality.)



Now compare the original distribution with the distribution of these scores after they have undergone a square root transformation



Finally see the effect of a logarithmic transformation



Both transformations normalise the distribution of scores to quite a large extent. The table on the right summarises the relevant statistics Note that both transformations bring mean, mode, and median closer to the same value, and reduce both kurtosis and skew. Transformations have lead to the distribution becoming more
normal

Summary statistics for original and transformed distributions			
	Original	**Square root transformation**	**Log(10) transformation**
Mean	**21.8**	**4.4**	**1.2**
Median	**17.3**	**4.3**	**1.3**
Mode	**16**	**4.1**	**1.2**
Kurtosis	**2.4**	**- 0.1**	**- 0.1**
Skew	**1.5**	**0.5**	**- 0.5**
Standard deviation	**18.5**	**1.8**	**0.4**

It is unlikely that you will ever need to bother with these transformations outside a research context, and, indeed there is a further transformation which is easier to understand and more often employed by test makers.

This will be described in detail in the Module 2.

A final word here, though, is that if ever a speed test whose scoring involves the time taken to solve each individual item individual items is administered, (for example, the Nufferno Speed Tests, or reaction time tests), the log of the time in seconds to complete the item is commonly used. This is because speed scores of this sort are notoriously positively skewed. As the score on such tests involves mean time to complete an item, mean log time is going to be a much better indicator of average speed than straight mean time in seconds would be.

5. Need more help with Basic Statistics?

If you feel you would like to revise your general statistical knowledge there are several very good FREE online texts to help

Most contain or give links to online demonstrations and calculators

Without prejudice to any of the others, we list the following.

The first is an excellent brief refresher/introduction.

The second goes into more detail and covers a wider range of topics

The third uses a ‘dictionary’ type approach.

http://davidmlane.com/hyperstat/

http://faculty.vassar.edu/lowry/webtext.html

http://www.sportsci.org/resource/stats/index.html

Finally, for a more detailed on-line statistics text book Statsoft’s text is hard to beat.

http://www.statsoft.com/textbook/stathome.html

6. Symbol Glossary

Symbols used in this module

***M* = the mean**

***N* = the number of scores**

ΣX ***= the sum all the X values***

$x = (X - M_x)$

σ ***= the standard deviation***

σ^2 ***= the variance***

sk = the skew index value