DRAFT: DO NOT SHARE OR QUOTE

NOTE 2.1: Level of Measurement

Variables constitute scales of values. The variable, 'biological sex', is commonly (though not unproblematically) understood to take two values, female and male; another might be personal computer operating system that might take the values Mac OS, Windows, Linux. For these scales, the values stand in no particular order. Scales such as these are at a 'nominal' 'level of measurement'. Because they cannot be placed in order, no arithmetic operations should be carried out on them: it is not meaningful to refer to a 'mean' gender or a 'total' operating system. The Likert Scale used by Sanderson is a 5-point scale: strongly disagree; disagree; neither agree nor disagree; agree; strongly agree. There is clearly an order to these values. Where the values can be ordered, but the intervals between the values is not consistent, then the scale is at an 'ordinal' level of measurement and arithmetic operations still cannot be carried out on such scales. Now, if we think that the interval between, say, strongly disagree and disagree may be greater or smaller than the interval between disagree and neither agree nor disagree or that between agree and strongly agree, then the scale remains at the ordinal level and we are not permitted to conduct arithmetic operations on it. Since the whole point of designing a survey instrument on the basis of a Likert scale is to conduct statistical analysis on the results—and this certainly involves arithmetic operations—then this would seem to render the method unhelpful. Those employing the method, however, take the view that it is reasonable to assume that the intervals between the values of a Likert scale are equal so that statistical analysis of the responses is valid. This assumption raises the Likert scale to the 'interval' level of measurement. Sanderson, like many others, clearly takes this view and, as a first step, has assigned numerical values to the points of the scale so that they can be summed.

There is a limit to the range of arithmetic operations that can be used with interval scales and this limit excludes multiplication and division. This is because interval scales do not have an absolute zero. Scales that do have this property are 'ratio' scales. Age, pulse rate, temperature in Kelvin (or Absolute) are ratio scales; temperature in degrees Celsius or Fahrenheit are interval scale, because zero in these scales is not a true zero, but arbitrarily defined and it can (and often does) get colder; zero Kelvin means zero heat!

It is worth pointing out at this stage that many if not most of the categories that are of interest in social research are nominal or ordinal scales. An example drawn from participant observation is 'researcher role', which takes the values: participant; observer. Now here is an expression of a widely held view:

> The role the observer plays forms a continuum from completely removed to completely engaged with the participant. (Sauro, J. 2015; https://measuringu.com/observation-role/)

In contrast, I hold to the view that researcher role is a nominal category because there is no metric incorporated within the category that quantifies the amount of participation or the amount of observation. This being the case the formation of a continuum is not possible. Clearly, in an actual instance of fieldwork the researcher may sometimes or in some respects be acting as a participant and at other times or in other respects be acting as an observer. An analysis of their role, then, would involve coding their activity and aggregating the values either as a qualitative description or using an alternative quantifiable variable

DRAFT: DO NOT SHARE OR QUOTE

such as time. The latter, however, may well be seen to reduce the crucial differences between and within the roles.

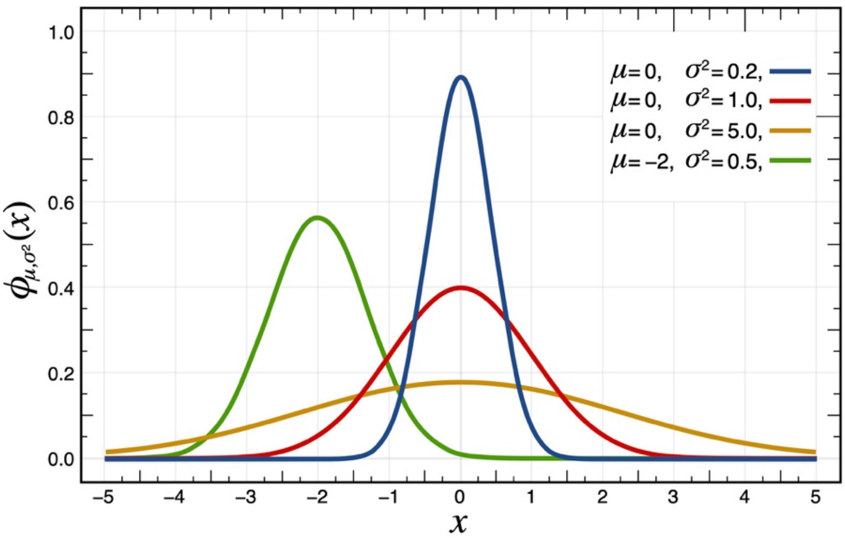
DRAFT: DO NOT SHARE OR QUOTE

NOTE 2.2: Frequency Distribution

The image below shows four *frequency distributions* that vary in respect of three parameters:

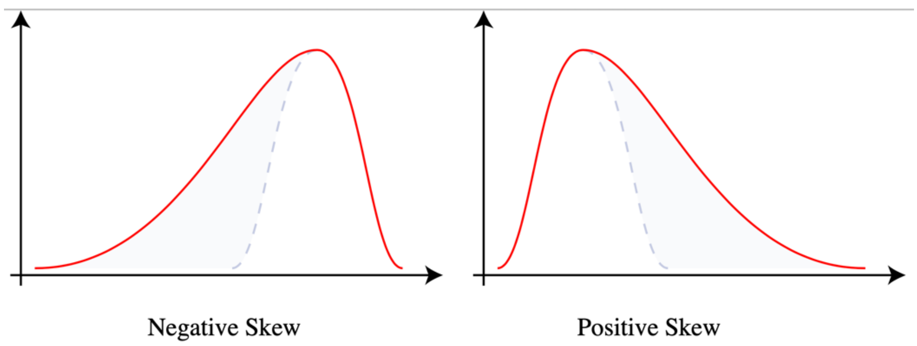
> The *mean* value (μ) of the variable *x*, which is 0 in three cases and -2 in one;
> The maximum value of the *frequency* (φ), where the value of *x* is the *mode*;
> And the width or spread of the distribution, which is measured by the *variance*, σ^2.

σ^2, incidentally, is the square of the *standard deviation*, *σ.*



Original diagram at https://en.wikipedia.org/wiki/Normal_distribution#/media/File:Normal_Distribution_PDF.svg. Copyright free.

The distributions in the diagram are all *normal distributions*, which is to say, they are each symmetrical about their respective means, so the mean is equal to the mode. A distribution where the mode is shifted to the right is *negatively skewed*, one that is shifted to the left is *positively skewed* as illustrated below.



Rodolfo Hermans (Godot) at en.wikipedia. [CC BY-SA 3.0], via Wikimedia Commons

The shaded areas are the regions to the left or right of the normal distribution.

Other shapes of distributions are possible including *multimodal* distributions that have more than one 'hump'.

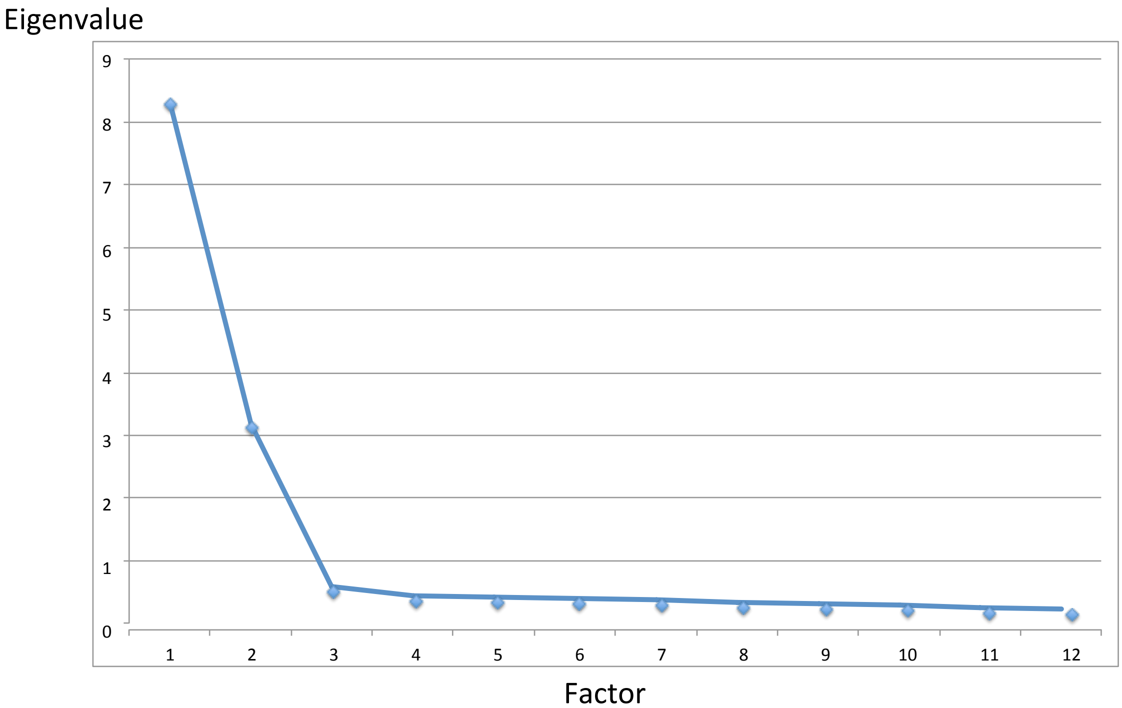
DRAFT: DO NOT SHARE OR QUOTE

Note 2.3: How many factors to retain

The diagram shows the *scree plot* of *eigenvalues* for each of twelve *common factors* that have been extracted. Two of the rules for deciding how many factors to retain are:

> The *Kaiser rule*, which tells you to retain those factors having eigenvalues greater than 1
>
> The *scree plot* rule, which tells you to retain those factors to the left of the elbow in the graph.

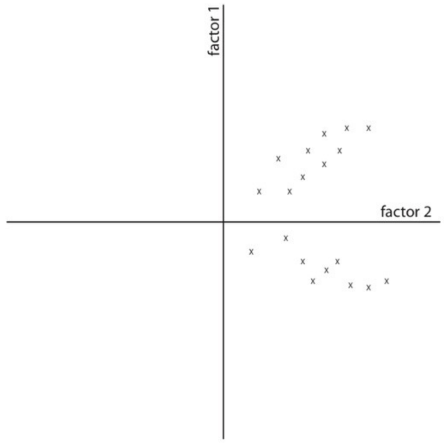
In this (fictitious) case both rules retain two factors.



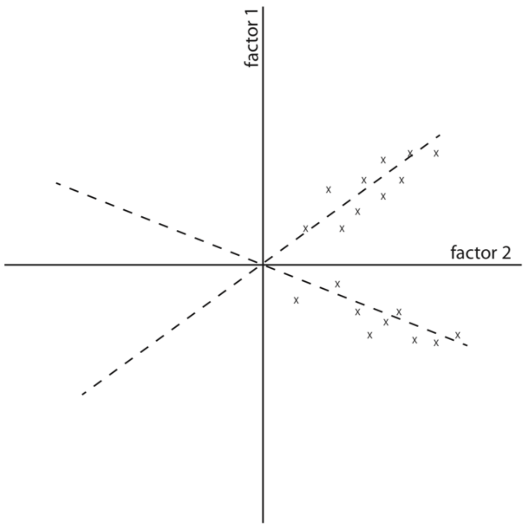
DRAFT: DO NOT SHARE OR QUOTE

NOTE 2.4: Rotation of Axes

The first diagram shows the (fictitious) plot of a 2-factor solution. The loading of factor 1 on each measured variable is plotted on the *x*-axis, the loadings of factor 2 on the *y*-axis. You can see that the measured variables form clusters (the result of the iterations), but the clusters are not grouped on either factor, so there is no clear interpretation of the meanings of the factors.



The second diagram, shows the result of an *oblique rotation* of the axes, which now align quite closely with the respective clusters of measured variables. An inspection of the measured variables in each cluster will enable an interpretation of each rotated factor. An alternative would have been to perform an *orthogonal rotation*, which would retain the orthogonality—the mutual independence—of the factors. This would not have resulted in as clear an analysis in this case. The mathematics for orthogonal rotation is simpler, but, since the computation is nowadays handled by computer, Osborne (2015) suggests that there is no compelling reason to use orthogonal rotation: if there is no correlation between the factors oblique rotation will not force a correlation and will produce much the same outcome as orthogonal rotation.



DRAFT: DO NOT SHARE OR QUOTE

Note 2.5: Random Sample

A *simple random sample* is on in which each member of a population has an equal probability of being selected for the sample.

If I want a 10 percent *simple random* sample of a population, then I should ideally pick the names out of a hat or get a computer to simulate this process, but I could achieve a 10 percent *systematic random sample* from a list of the population, first choosing a number between 1 and 10 (for example, by sticking a pin in a page of random numbers), if this turns out to be, say, 3, then the 3rd member of the population is selected and every tenth member thereafter.

In practice it is rarely possible to generate an accurate listing of the population—the population of England, for example—so an approximation, such as the electoral register for England is used. The sample is then drawn from this *sampling frame*. I could ensure that particular categories—for example genders or age ranges—were approximately equally represented in the sample by *stratifying* the sampling frame by these categories (listing all females then all males and/or listing each age range separately) to generate a *stratefied random sample*.

If one wanted to interview sample of UCL undergraduates, then it might be convenient first to take a random sample of undergraduate courses and interview everyone in the resulting *cluster sample* of courses. In a larger population—say the UK electorate—one might use a *multistage sampling method* by first taking a random sample of constituencies, then a random sample of postcodes within this sample, and finally a sample (simple random or even 100 percent sample) of households within the sampled postcodes.

The rationale for using random sampling methods entails the assumption that, if the sample is random and not motivated in some way, then its characteristics will match those of the population. This principle is somewhat weakened by the compromises in the above methods. The UK electoral register, which may be used as a sampling frame, is not the same as the UK population, for example: it lists only those who are eligible and who have registered to vote and there are reasons why some people choose not to register. Each deviation from a simple random sample of the population introduces biases into the sample and possible biases should be taken into account when interpreting the findings of a survey or experiment.

DRAFT: DO NOT SHARE OR QUOTE

Note 2.6: Statistical Significance

My colleague Andrew Brown examined diaries written by parents in two London schools, the diaries being reports by parents on their children's homework activities as part of a parental involvement project (Brown, 1999). Each parent was categorised as being either a localiser or a generaliser on the basis of their diary entries. The results are summarised in Table 2.1.

Table 2.1: Generalisers and localisers amongst samples of parents in two London schools

	East Wood	*Chambers*	*total*
localiser	28	48	76
generaliser	28	6	34
Total	56	54	110

We would like to know whether these findings reveal any difference between the parents at the two schools. The difference in the localiser:generaliser ratios in the two schools suggests that there is a substantial difference. However, if the samples were drawn at random from each school, then it is possible that this result could have arisen on this occasion and on another occasion the results might have been very different, even reversing the apparent contrast. We would like to know the *probability* that this result would be achieved *if in fact there was no difference* in the distribution of localisers and generalisers in the two schools. The proposition that *there is no difference* is called the *null hypothesis*. We can calculate the probability of arriving at this result using an Excel spreadsheet, selecting the CHISQ.TEST (*chi-square test*) function and inputing the shaded region of Table 2.1 as the 'Actual range'. To complete the formula we must first calculate the 'Expected range'. If the null hypothesis holds, then the ratio of localisers to generalisers would be the same in each school, that is, 76/110 for localisers and 34/110 for generalisers. We multiply these ratios by the number of parents in each school to find the *expected* values, that is, the values that we expect if the null hypothesis holds. The 'Expected range' for the formula in Excel is shown in the shaded cells in Table 2.2. The contents of the cells are in the format to be inserted in the Excel table (pressing return after each entry).

Table 2.2: Expected frequencies under the null hypothesis for Brown's data

	East Wood	*Chambers*	*total*
localiser	=76*56/110	=76*54/110	76
generaliser	=34*56/110	=34*54/110	34
Total	56	54	110

Entering these ranges in the CHISQ.TEST yields a *p-value* of 1.02271E-05, which is the 'exponential' form of 0.00002271, which is the *probability* of obtaining the results in Table 2.1 *if the null hypothesis holds*. This is a very small probability, in other words, it is very unlikely indeed that Brown would have obtained the results that he did if there had been no

DRAFT: DO NOT SHARE OR QUOTE

association between the ratio of localisers to generalisers and the two schools. We say that the result is *statistically significant*.

Conventionally, the researcher can claim *statistical significance* if the *p-value* is below the 5 percent level (0.05) or—a stronger claim—below the 1 percent level (0.01).

It is important to note that *statistical significance* is not the same as substantive significance. In general, if the sample size is large enough, even very small substantive differences may prove to be statistically significant. This distinction would be particularly relevant where the test is measuring the effectiveness of a drug in the treatment of a health problem: the drug may yield statistically significant results, but if the substantive benefits are very small can the potential cost of its use be justified?

All statistical tests involve assumptions about the data and it is important that these assumptions are valid. Recall that it is a moot point whether a Likert scale produces interval rather than ordinal level data as required by Exploratory Analysis (Note 2.1). Chi-square requires that the data in the cells are frequencies (ie not percentages) and that the expected frequencies are all at least 5; these assumptions hold for Brown's data (complete the calculations in Table 2.2 to check the second assumption). Also, the categories should be independent, which is the case here: each subject is either a localiser or a generaliser and is associated either with East Wood or Chambers school.

Now the chi-square test result tells us that there is an association between the two variables (the columns and rows of Table 2.1). This does not, however, say anything about the *strength* of the association, which we can estimate using *Pearson's phi coefficient*, which, in the case of a 2x2 contingency table (as we have here) is the square root of the chi square statistic divided by the number of observations. We have the *p-value*, here, but not yet the chi square statistic itself. To calculate this (the calculation is not directly available in Excel) for each cell we square the difference between the actual and expected frequencies and divide this by the expected frequency, then sum these values. For these results the chi square value is 19.47 (rounded up), which gives us a value of 0.177 for *phi squared* and 0.421 for *phi*. Conventionally, a coefficient of correlation (which this is) value of greater than 0.4 indicates a "relatively" strong correlation.

Note that **we cannot infer causality from a coefficient of correlation**. In this case it seems clear that we would not expect a change of school necessarily to transform a parent from a localiser to a generaliser, but there are many instances where this is not as obvious and people do make the mistake of inferring causality from a correlation: *this is incorrect*. If there is a correlation between two variables this entails that as one variable increases so does the other (a positive correlation) or as one variable increases the other decreases (a negative correlation). This does not necessarily mean that the increase in the first variable causes the increase or decrease in the other, there may, for example, be other factors that cause the changes in both.