Online Resource

# Chapter 4: Relationships between variables

## Logistic regression

The concept of linear regression and estimating the best fitting line by finding the least-squares fit (where the sum of the squared residuals between each data point and the line are smallest) has been well established for well over a century. It works well with one proviso: the DV must be an interval variable. If the DV is categorical, then the process of least squares does not work well at all. Given the simplicity of the concept and the success with interval DVs, it is not surprising that eventually a version of regression for categorical DVs was invented.

In linear regression, the regression line shows the expected value of the DV and is found by comparing the expected values with the observed values of the DV.

The version of regression for categorical DVs has only one real conceptual difference compared to linear regression: instead of the observed values of the categorical DV, we use the probability that the categorical variable is a particular category. We find a regression line that works so that the values of the IV predict the probability that the value of the DV will be a particular category. After historically using various different versions of probability, it is now normal to use the logarithm of the odds that the value will be a particular category. We’ve got a little bit more information about logarithms with Chapter 9, and online here. The odds that the value will be X category is the probability that it *will* be X, divided by the probability that it *won’t* be X. The logit function is that logarithm of this.

The reason that the logit function has become the commonest is largely because it produces an S-shaped function, which typically matches what is found. The figure shows an example of the effect of an interval IV (RiskTaking) on a categorical DV (Exam Pass?):



Alongside that conceptual difference of how to treat the DV, there is another important procedural difference. Linear regression minimizes the sum of squared differences (residuals). Logistic regression uses a more general quantity: the deviance between the line of expected outcomes, and the observed data. For logistic regression, deviance is usually defined in terms of expected probabilities and observed probabilities.

## Standard deviations: Groups and residuals

In much of the remainder of this book, we will be encountering comparisons between the **variance** in a DV that is explained by the IV or IVs, and the variance left over, which is the variance of the **residuals**. In this section, we will look specifically at a case when the IV is a categorical variable and the DV is an interval variable.

There is an interesting difference between observational and experimental IVs in how the variance (and therefore standard deviation, or SD) of these two types behave. It is easy to think of an experimental IV as creating new additional variance. If we invent an intervention that increases everyone’s exam grade by 15 points but only apply it to half the class, then the spread of exam grades is wider than would otherwise be the case. The next few paragraphs explain this further.

Variance has a very useful property: if two sources of variance are independent, then the overall variance is just their sum. Variances from independent sources can be added together. This is what allows us to write:

var(DV) = var(group means) + var(residuals)

This makes it easier to talk about the corresponding standard deviation.

*Observational IV*: when the IV is an observational variable, then the overall standard deviation of the DV is fixed. This means that the standard deviation of the residuals cannot be fixed. If we look at the effect of RiskTaker? on ExamGrade we have this situation:

1. Overall standard deviation of Exam Grade is fixed

2. Standard deviation of Exam Grade in each group is therefore reduced

*Experimental IV*: when the IV is an experimental variable, then the opposite situation is found. The standard deviation of the DV within the groups is unchanged, which means that the overall standard deviation is increased. If we look at the effect of Mindfulness on ExamGrade, then doing the experiment increases the range of exam grade we will see. We have this situation:

1. Standard deviation of Exam Grade in each group is fixed.

2. Overall standard deviation of Exam Grade is therefore increased

## Interpreting standardized effects

In some areas of psychology, notably those where research typically involve a categorical IV with 2 categories, it is typical to use Cohen’s *d* (a standardized effect size) to compare between different studies. It is often the case that the categorical IV is an experimental variable – an intervention of manipulation – and it is particularly useful to be able to compare different treatments of the same DV. An impressive sustained example of this is found in Wampold and Imel (2015) who show how to combine the results from many different studies examining different treatments into a common framework by the use of standardized effect sizes.

Just to provide an example, and without considering the merit scientifically of the argument that they make, by converting effect sizes into this standardized form, they are able to claim that the effectiveness of psychotherapies is very much the same regardless of specific therapy details or specific client condition diagnosis. The effect, they argue, of CBT on depression is similar in size to the effect of EMDR on post-traumatic stress disorder. Given the wide, wide range of different DVs and DV measures, such a claim is a remarkable benefit of the use of standardized effect sizes.

In this book, we have chosen to use normalized effect sizes rather than Cohen’s *d* for immediate three reasons and one bit of looking ahead:

1. We find that a measure of effect strength that ranges from 0 to ±1 is easier to grasp than one that ranges up to infinity. It is hard to know how close a Cohen’s *d* value of 2 is to infinity; it is easier to know how close an *r* value of 0.6 is to +1.

2. Strictly speaking Cohen’s *d*, which is by far the commonest version of a standardized effect size, only applies to cases where the IV has two categories or groups. Something exactly equivalent to it can be calculated for other situations, but they are then difficult to explain.

3. The *r*-family have a simple interpretation in terms of the proportion of variance explained. A modification of Cohen’s *d* (called Cohen’s *f*) makes it the ratio of variance explained to variance unexplained. Since variance unexplained is zero at a perfect correlation, standardized effect sizes go to infinity (anything divided by zero is infinite).

And looking ahead:

4. The Fisher Z-transform applied to *r*-family effect sizes produces some very useful properties, such as standard errors being independent of effect size.

Cohen’s *d* was devised by J. Cohen (a central figure in the development of useful statistics). He proposed the following subjective scale for assessing the importance of values of *d* (Cohen, 1972):

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cohen’s proposal** | **Equivalent value for *r*-family\*** | **Cohen’s description** |
| Small | *d* = 0.2 | *r* = 0.1 | Cohen suggested that this might be expected when variables are studied outside of laboratory conditions where many extraneous variables affect the measurements. He also gave it as the difference in height between 15-yr-old and 16-yr-old girls. More contentiously\*\* he also gave as an example, the difference in IQ between twins and non-twins. |
| Medium | *d* = 0.5 | *r* = 0.24 | Cohen described this as an effect that is large enough to be ‘visible to the naked eye’. He also gave it as the difference in height between 14-yr-old and 18-yr-old girls. Or the difference in IQ between clerical and semi-skilled workers. |
| Large | *d* = 0.8 | *r* = 0.37 | This is described as ‘grossly perceptible’. He also gave it as the difference in height between 13-yr-old and 18-yr-old girls. Or the difference in IQ between PhD holders and college freshmen. |

\* Cohen is clear that these values might not be suitable when thinking of effect sizes as correlation coefficients in some areas of research. It is often overlooked that just what counts as small, medium or large is very much context dependent.

\*\* IQ differences might be differently understood nearly 50 years later. These examples have an uncomfortable trace of one the important applications of psychological measurements in times past: IQ testing to filter people into ‘suitable’ employment. Perhaps the principle is not yet consigned to the past. If this interests you, read: R. Roberts (2015) *Psychology and Capitalism: The Manipulation of Mind*. Zero Books.

**Reference**

Wampold B.E. and Imel Z.E. (2015). *The Great Psychotherapy Debate*. 2nd Edition. Routledge