Online Resource

# Chapter 14: Which model is best?

## Fitting complex model

Every model has a set of coefficients that we must estimate from our data. We have seen that the principle of least squares can be used for this in all the cases we have considered so far. Under a set of fairly safe assumptions, the set of coefficients that minimize the sum of squared residuals are also the set of coefficients that produce the model that has the maximum likelihood. When we turn to situations where there isn’t a single DV, or the distinction between DVs and IVs is broken, then it becomes much more difficult to estimate coefficients in this way.

There is a nice theorem in statistics, called the equivalence relation, that shows that these two things contain the same information:

1. the set of coefficients of a model

2. the pattern of covariances and variances between all the variables in the model

They are equivalent descriptions of a model. If we know the coefficients, then we can work out what pattern of covariances we should see. This means that we can find the best fitting set of coefficients for a model by examining the pattern of variances and covariances. The technical term for this pattern is the covariance matrix (matrix in this context denotes a table).

Because of this equivalence relation, we can fit models of arbitrary complexity, including those with multiple DVs. We take the covariance matrix for the data we have itself. Then we take a set of possible coefficients for the model. With those coefficients, we can work out what the model covariance matrix would look like. Now we just have to keep on doing that until we find the model coefficients that produce the best possible fit between the model covariance matrix and the data one.

We saw a simple version of this in Intermezzo 1, where we worked out that the slope of a simple regression line is the covariance of the IV and DV, divided by the variance of the IV. If we have only 2 variables, then the covariance matrix has their individual variances and their covariance: it has everything we need to calculate the coefficient which is the slope of the regression line.

## AIC

The AIC measure that we have introduced in this chapter is a good way to balance the competing requirements of any model: good fit to the data vs simplicity. That is as much as we need to understand to be able to use it successfully. However, it does have a slightly deeper, more satisfying meaning.

Suppose that we are comparing a model with 2 coefficients with another that has 3 coefficients. We can see that the second model includes some information that the first model lacks: the third coefficient in the second model is telling us more about the model. The number we get for AIC is related to the amount of information that a model has lost compared to an ideal. Models with larger AIC are losing more information.

If we take the AIC values for any two models and we combine them in this specific way:



then the quantity that we have calculated is an estimate of how much more probable the second model is than the first to have minimized the information lost. Usually, in this calculation the second model has the higher AIC and the quantity calculated is less than 1.

In our usage of probability (for future events that haven’t happened) and likelihood (for past events that have happened) we would prefer to say “how much more likely…”. We can think of this quantity by its technical term – the relative likelihood of model 2 compared to model 1.

The table gives some differences in AIC and the corresponding relative likelihoods. Suppose we have two models. The table takes the model with the lowest AIC and then compares it with a model with a higher AIC: the table shows how much smaller the better model is, so the first row is for a case where the better model has an AIC value that is smaller by 1. The difference in AIC allows us to calculate the relative likelihood of the better model compared to the worse model. The values shown in the table are how much more probable it is that the better model should be preferred.

|  |  |
| --- | --- |
| **Difference in AIC** |  **of better model compared to worse model** |
| –1 | 1.64 times more likely |
| –10 | 1429 times more likely |
| –100 | massively better |

Notice that a difference of 1 doesn’t tell us how likely either model is, just how more likely the first model is than the other.