BASIC ANALYSIS FOR TRIAL DATA PART FOUR

DAN W. JOYCE

Interpreting the Results

Now that we've established how the estimated model from a statistics package corresponds to an intuitive understanding in terms of means, we turn to understanding the output. The question we want to answer is – "At week 6 (after treatment), is their a difference in mean BPRS between patients treated with clozapine and those treated with chlorpromazine ?". Additionally, we usually want to know which of the medications was *superior*. To do so, we need to formulate and specify a hypothesis in terms of the variables of our model (equation 1 in Part Three) and it's estimated fit (Table 3):

- the intercept in the model (β_0) tells us the mean BPRS for the chlorpromazine group at baseline/before treatment. Testing a hypothesis on β_0 would only be able to inform us about the mean BPRS scores at baseline for the chlorpromazine group.
- if we only vary the Drug term (D_{ij}), and ignore Time (T_{it}) – i.e. inspecting β_1 – then we are 'collapsing' the BPRS scores for the chlorpromazine and clozapine groups separately *across* time points. Practically, this tells us the difference in mean BPRS scores between chlorpromazine and clozapine pre-treatment; if the randomisation of patients to drug was fair and unbiased, then we would expect β_1 to reflect this.
- if we instead vary the Time variable alone (from baseline to post-treatment at week 6), *without* differentiating between drugs, we will be 'collapsing' together the BPRS scores of both drugs pre- *and* post-treatment. We can then see if treatment with *either* drug works because we only examine the effect of time (pre- to post-treatment). This corresponds to examining the estimated coefficient β_2.
- if we vary Time *and* Drug (the so-called *interaction* term) then we can see if patients assigned clozapine and chlorpromazine *differed* in mean BPRS post-treatment. This term let's us test if one medication differs from the other, and potentially superiority of one over the other, and corresponds to examining estimated coefficient β_3

Before continuing, we note that it is convention to formulate our hypotheses in terms of the **null hypothesis** that, perhaps counter-intuitively, states an expectation of there being *no effect* of any variable/term on the mean BPRS.

The ongoing debate and history of null hypotheses in inferential statistical tests is beyond our scope, but (Sterne and Smith, 2001) provides a readable overview that complements our discussion

We are interested in the mean BPRS difference between clozapine and chlorpromazine ($D_{i1} = 1$ versus $D_{i0} = 0$) after treatment ($T_{i1} = 1$ versus $T_{i0} = 0$) so our focus will be on testing β_3 – the interaction of Time and Drug. If there is no effect of medication moving from pre- and post-treatment, then we expect β_3 to be close to zero because then the term $D_{i1} \times T_{i1} = 1$ does not contribute any additional weight to the sum in equation 1 beyond that already captured by β_0, β_1 and β_2:

$$Y_{ijt} = \beta_0 + \beta_1 D_{ij} + \beta_2 T_{it} + \beta_3 D_{ij} T_{it} + \epsilon_{ijt} \quad (1)$$

We'll describe our null hypothesis as H_0: there is no interaction between Time and Drug on mean BPRS scores. Our *alternative* hypothesis is then H_1: there is an interaction between Time and Drug on mean BPRS scores. Formally,

$$H_0 : \beta_3 = 0 \quad (2)$$

And we will make a decision to *reject* the null hypothesis H_0 in favour of H_1 according to:

$$\text{Reject } H_0 \text{ if } \beta_3 \neq 0 \text{ with } p < 0.05 \quad (3)$$

The symbol $\neq$ is read 'not equal to'

A brief reminder on p values – they represent the probability of obtaining an effect at least as extreme as the one in your data, assuming H_0 is true. The significance level is often referred to as α and is chosen rather arbitrarily by convention as $\alpha = 0.05$ which equates to there being a 5% chance of Type I error – discussed further in (Sterne and Smith, 2001). Finally, if $\alpha = 0.05$ then the 'confidence level' is said to be $1 - \alpha = 1 - 0.05 = 0.95$ or 95%.

It is good practice to report confidence intervals, and not just the estimated β and the p value

Examining Table 3 (from Part Three), we find $\beta_3 = -10.90$, with $p < 0.05$. So we have evidence to reject H_0 in favour of H_1. We should remind ourselves that if something is "significant" with $p < 0.05$, what we really mean is that we are allowing for a 5% chance that we have *incorrectly* rejected H_0.

To report our result more convincingly, we should not rely on just the point estimate . In this example, β_3 is a point estimate of the interaction term $D_{ij} \times T_{it}$, with a 95% **confidence interval** of $[-14.33, -7.47]$. We can now report that:

A **point estimate** results from using a sample of data to calculate a single value which is representative of an unknown parameter. A simple example: in a population of 65 million, there will be an "average height" – which is the unknown parameter. With a random sample of 1000 people from this population, we can compute the **sample mean** and use this as an **point estimate** of the population's unknown parameter "average height"

1. there is an effect of Drug and Time at the $p < 0.05$ level of significance (i.e. we reject the null hypothesis $H_0 : \beta_3 = 0$)

2. the mean change in BPRS scores for patients treated with clozapine is -10.90 lower than for patients treated with chlorpromazine, but at the 95% level of confidence, this could range from -14.33 to -7.47

3. further, note how for β_3 at the $p < 0.05$ confidence level, the 95% confidence interval *does not contain* the value given in the null hypothesis: $H_0 : \beta_3 = 0$.

Before we finish, we might want to convince ourselves that the randomisation was unbiased at baseline, so that we are sure the patients assigned chlorpromazine where not more ill or well than those in the clozapine group. This is easy – if the groups were the same at baseline (i.e. randomisation robust and unbiased), then the estimate β_1 would be zero, or more formally, the 95% confidence interval would *include* zero:

- the null hypothesis is that the pre-treatment mean BPRS scores in patients assigned chlorpromazine *were not* different to those in the clozapine group
- this equates to:

$$\begin{gathered} H_0 : \beta_1 = 0 \\ \text{Reject } H_0 \text{ if } \beta_1 \neq 0 \text{ with } p < 0.05 \end{gathered} \tag{4}$$

- inspecting Table 3 (again, from Part Three) we find that $\beta_1 = 1.90$ but with $p > 0.05$ and a confidence interval $[-0.53, 4.32]$ – so the result does not reach the significance level, and the confidence interval *includes* the null hypothesis of $\beta_1 = 0$
- so we *accept* the null hypothesis at the 95% confidence level and infer that there is not enough evidence to conclude the groups were different on BPRS scores at baseline.

Questions and Exercises

If you are familiar with statistics and experimental design, you might be asking why we did not conduct a *t*-test for differences in pre-treatment BPRS scores for the chlorpromazine and clozapine groups. Conduct a two-tailed independent samples *t*-test using the simulated data; you should be able to confirm that indeed, there is no difference and get a result along the lines of $t(261) = -1.54, p = 0.126$ with the mean BPRS for chlorpromazine $M = 62.23$ and for clozapine $M = 64.13$.

References

Sterne, J. A. and Smith, G. D. (2001). Sifting the evidence – what's wrong with significance tests? *British Medical Journal*, 322:226–231.