**BMI Dataset – Including Baseline in HLMs**

This data contains information on the BMIs of various people over 5 time periods. This is a repeated measures problem, and so HLMs are ideal to use in this scenario.

Below I have produced tables which outlines the results of all of the models. It is much easier to see in this format. If an entry is left blank for the fixed effects, that parameter is not included in the model.



The consideration in the first five models – will including the baseline have favorable results on model prediction?

Since zj is already contained in Yij, this could cause some issues. Since Y1j = zj, in this case we have an exact solution to the problem, with the baseline coefficient equal to 1 and all other coefficients equal to zero. There is no randomness here, assuming we already know the baseline. Thus, including this in the model does not account for the fact that we know that Y1j must equal zj for all j. The model is treating them as though they are two separate entities, when in fact they are directly correlated when i = 1.

If we know the baseline BMI from the beginning, there is no need to predict the BMI at time 1, as this is presumably contained in the data itself. The models as they are do not treat zj in this way, they simply treat it as another random variable by which the intercept can be determined.

If we look at the predictions from mod1 for example, we see that the predictions for Y1j for j=1,2,3,4,5 is somewhat different then the baseline. But why would our model predict for something that is already known? For instance, we know that if the baseline BMI is 29, then Y1j is 29. There is no need to estimate it! We only are interested in estimating the time intervals after this, as this is what our model was built to predict. Including zj in this way does not account for the fact that we already *know* the value for Y1j.

Models 4 and 5 consider the differences from the baseline and uses them as the dependent variable.

Model 1 and Model 4 are directly comparable, Model 4 is exactly Model 1 when fixing the baseline coefficient to 1. We can see that the coefficient in the new model is close to zero, which makes sense since the Yij are now close to zero in a process similar to standardizing, but not quite, as zj is only the first entry, and not the mean.

The standard error for the intercept fixed effect in this appended model is larger than that of the original model, which makes sense – we have transferred the variability from the free coefficient of the baseline to the intercept, as we have fixed it to 1. Indeed, using the regression model predicted a baseline coefficient of .69, which is far from zero. Setting this to 1 then reduces the effectiveness of the model as is and still does not consider the fact that zj = y1j, besides the fact that the first terms are zeros. However, neither model has very significant intercept coefficients, as the standard errors are relatively large in both cases. The fixed effect for time is very similar in both models.

This is not the wanted remedy to the earlier problem, as the variance in the intercept random effect has increased while leaving the variance of the error terms exactly the same. This means we have only introduced less confidence in the intercept term, without changing the resulting errors from the predictions very much. Thus the ICC has increased, but this can be misleading.

Similarly, Model 5 and Model 2 are directly comparable. As there is no intercept term in either model, the effects are transferred into the fixed effect for time. In these two models, the number we set the fixed effect to (1) and the originally predicted with a free coefficient (1.02) are similar, so the resulting differences are not as pronounced. Model 5 in fact has a lower standard error for the random effect in the intercept, while having a slightly lower standard error for the level 1 random effect.

Using the models with the baseline removed, we see a better result – much of the variance is within the groups rather than between them. The ICCs for all of these new models is considerably higher. Since we know the baseline of each, we can set each prediction of Y1 j to zj, and use these models to predict the following times. Although this model contains less observations, it is more accurate, since we now consider the special properties of the baseline – that is, that it is the first observation (time =1). Thus the variance of the level 2 terms are higher (3.07,3.67, etc.) but the variance of the level one terms are lower (1.31).

Using the sum squared error, we can see that the new set of models has less error than the first set of models, since we are now considering some more of the information that we have. This is included in the R output, but this makes sense – since we know undoubtedly that Y1j = zj, it makes little sense to predict it using zj as a predictor variable. If we remove this from the model and assume for all predictions where i=1 are equal to the baseline of that group, we get a better result, as there is a relationship between the baseline and the later times, and including the baseline prediction skewed the results. Unequivocally, each model performed better after removing the baseline results.

The residual plots show an interesting result. As each of these models have fixed effects for the slopes, only the intercepts will vary. If we view the intercepts then as a combination of determined effects and random effects, it makes sense that the results would be very similar in the residuals. Indeed, we can see that, for instance, for the first five models and last five models the first few residuals are identical in relation to one another. This makes sense, as changing the intercept of a model will only shift it up or down vertically. However, it is not a perfect transformation of course, as there is also a random effect which varies the residuals along this “determined” transformation. For each prediction we are then adding a white noise term which accounts for the variation in the random effect. This is why we see very similar, but not exactly transformed, residuals in each of the cases. The type of model has an effect on this as well.

The residuals in the last five models are more standard, as they are varying around zero, while the first five models seem to be skewed in the positive direction. Thus, the residuals are more likely to be positive than negative for these models, which violated a normality assumption in the residuals. Overall, the last five models proved to perform much better than the first five, so this finding coincides with my earlier findings about the sum squared errors.