# Statistical Methods AEMA-610

R. I. Cue

Department of Animal Science

**Pseudo-$R^2$ from a Mixed Model**

When we use PROC MIXED to estimate fixed and random effects we do not obtain an $R^2$ as we do under a fixed effects model. What if we are asked what was the $R^2$, or if we want to show how much of the variation was explained by the model, by individual terms in the model? Are we stuck? Well no; we can produce what is commonly called a 'Pseudo-$R^2$', something akin to the normal $R^2$ of a fixed effects model.

For a fixed effects model what does the $R^2$ represent?

$$R^2 = \frac{SSR_m}{CTSS} \qquad \text{or} \quad = \frac{SS}{CTSS}$$

where

$SSR_m$ = the Reduction Sums of Squares due to the Model over and above the Mean (as per S. Searle, Linear Models). This is what a SAS PROC GLM fixed effects model calls the 'Model', it is actually the Model over and above the Mean.

or, SS = the marginal Sums of Squares of an effect.

CTSS = the Corrected Total Sums of Squares, i.e. the Total Sums of Squares - the Correction Factor for the Mean.

If we fitted a PROC GLM model with nothing except the mean:

```
proc glm data=SASdataset;
model y = ;
run;
```

then the CTSS can be estimated from $(N - 1) * \sigma_e^2$, i.e. the residual degrees of freedom (since we only fit the mean this is the N-1) times the Residual Variance (= the Mean Square Error). This is actually redundant since PROC GLM computes the Total Sums of Squares and the Corrected Total Sums of Squares (CTSS), but it serves to illustrate the idea that we can 'back calculate' the CTSS ourselves.

Therefore this gives us the equivalent idea in PROC MIXED:

```
proc mixed data=SASdataset;
model y = ;
run;
```

This allows us to compute the CTSS, as simply the residual degrees of freedom (N-1) * $\sigma_e^2$.

NOTE: we must remember that $\text{SSR}_m + \text{SSE} = \text{CTSS}$. So, if we fit a model (using PROC MIXED) we can estimate SSE from the Residual variance * the Residual degrees of freedom. Then we can compute the $\text{SSR}_m$ as CTSS - SSE.

Suppose that we fit a model with 3 factors (A, B and C), call this model 1. Then

$\text{SSE}_1 = \sigma_{e_1}^2 * \text{dfe}_1$

Therefore $\text{SSR}_{m_1} = \text{CTSS} - \text{SSE}_1$

$$\text{R}_1^2 = \frac{\text{SSR}_{m_1}}{\text{CTSS}}$$

If we are asked what is the proportion of the variation (or Sums of Squares) explained by an effect (be it fixed or random), then we can use this approach, by dropping the desired term from the model and seeing the change in SSE.

Suppose that we are asked what it the proportion of the Sums of Squares (variation) explained by A?

Well. Fit a model without A in it, call this model 2.

$\text{SSE}_2 = \sigma^2_{e_2} * \text{dfe}_2$

Then $\text{SS}_A = \text{SSE}_2 - \text{SSE}_1 = \Delta_{SS_A}$

Then the R-squared for Factor A, i.e. the variation explained, is

$$R^2_A = \frac{SS_A}{CTSS}$$