

LAB 1: Notes

Analyze the fecal fat data (fecfat) using SAS:

- 1) Calculate estimates of the mean value for each of the patients.
- 2) Fit a model using SAS Proc GLM with outcome of fecal fat and the predictors being pill type and subject.
- 3) Fit the same model using SAS Proc MIXED (declare subject to be a random factor).
- 4) How do the analyses compare with respect to both results and fitting methods?

Fitting methods are different. GLM uses OLS, MIXED uses REML. GLM gives an F-test of $F=6.26$, $df=3, 15$. MIXED gives the same test.

- 5) Obtain the predicted values for each of the patients using both GLM and MIXED. (Use an ESTIMATE statement and specify the intercept as well). How do these compare to each other and the values calculated in 1)? What is the explanation?

GLM gives the averages as the estimated values. MIXED gives shrinkage estimators (note they are all closer to the overall mean than the raw means)

- 6) Now add sex to the model and analyze using GLM and MIXED. How do they compare?

Because sex is a between persons predictor, GLM is unable to give a test without special code. MIXED generates an $F=0.98$, $df=1, 15$. By specifying an error term of subjects nested within sex, a proper test can be obtained in GLM.

Analyze the fecal fat zero data (fecfatzero) using SAS. This is a dataset artificially created to have little subject to subject variation:

- 7) Rerun the GLM and MIXED analyses. How do the predicted values compare?

As before, GLM gives the simple averages. MIXED reports the overall average since there is no evidence of differences between subjects.

- 8) What does the log for MIXED indicate?

MIXED gives the warning note that "Estimated G matrix is not positive definite." In this case it merely means that the variance component is estimated to be zero. This message will occur whenever one or more variance components is zero. Also, in a random slopes and intercepts model (with an unstructured var-cov structure), SAS does not constrain the var-cov matrix to be positive semi-definite. If this occurs you can force the var-cov matrix to be positive semi-definite by using the factor analysis structure (type=FA0).

- 9) Run the MIXED analysis without the random statement.

Without the RANDOM statement, the F for pilltypes is unchanged, but the d.f. change to 3 and 20.

- 10) Run the MIXED analysis using the DDFM=KENROG option on the model statement.

- 11) How do the tests for pilltypes compare across the analyses? What is the explanation?

GLM gives $F=6.27$ with $d.f.=3,15$. MIXED default gives $F=8.22$ with $d.f.=3,15$. Dropping the RANDOM statement gives $F=8.22$ with $d.f.=3,20$. MIXED with DDFM=KENROG and a RANDOM statement gives $F=8.22$ with $d.f.=3,20$. So the Kenward-Roger adjustment recognizes the estimated variance component of zero and adjusts the d.f. accordingly. In some analyses in the default mode, the p-values of tests can change drastically depending on whether you drop or keep a random effect, making it more important to do model reduction on the random effects before performing the fixed effects analysis. The Kenward-Roger adjustment minimizes this effect.