1) Loading and working with data in SAS. This example, courtesy of Bryan Matlen, is very typical of real data you will encounter, but the tasks listed below have a moderate degree of difficulty. I will send you the data rather than post it. All of the tasks listed require only the material in the SAS workshop handout. The file cvsdata.dat is a plain text file in the “tall format”, with one line per observation. The repeated measure is redundantly coded as both “session” and “sessionNum”.
   a. Examine the data with, e.g., Wordpad. Note that there are 5 conditions and 8 session types. Load cvsdata.dat into a permanent SAS data set. Print the first 10 lines, and make frequency tables for condition and session. Identify the problem (if you were very smart there will be no problem).
   b. Modify your DATA step to correct the problem. Repeat the frequency tables and add a cross-tabulation of condition with session.
   c. Obtain standard statistics (mean, variance, etc.) for the score, and make a histogram and a boxplot. You can make the plots with “VBOX score;” and with “HISTOGRAM score;” as the sole statements in two “PROC SGPLOT;” blocks.
   d. Using PROC GLM, perform a regression (ANCOVA) of score on sessionNum with separate (fixed) intercepts for each condition, i.e., with sessionNum and condition on the right side of the MODEL “=” and with a class statement for condition. (This fits a straight line across sessions for each condition, which makes no sense in Bryan’s project, so just consider this an exercise analogous to what would make sense in another context.) Include a residual vs. fit plot and a quantile normal plot of the residuals.
   e. Perform ANOVA of score on id (subject number) to verify the need for a random intercept.

2) Fit Bryan’s data using SAS PROC MIXED with and without a random intercept and including both condition and session as categorical. Use the MODEL option “/ SOLUTION” so that you can get “Solution for Fixed Effects”. Also add the MODEL option “OUTPRED=myDatasetName” to get a new dataset with residuals (called RESID) and predicted values (called PRED) for each model. Compare the “Dimensions” sections. Compare the “Covariance Parameter Estimates”. What tests are automatically made in the “Solutions for Fixed Effects” section? What tests are made in the “Type 3 Tests of Fixed Effects” section? Compare the BIC values for the two models. Which is better?

Use this code to make a residual vs. fit plot:
```
PROC SGPLOT DATA=yourOutPredDatasetName;
   SCATTER X=pred Y=resid;
   REFLINE 0 / AXIS=Y;
RUN;
```
For the random intercept model, use PROC UNIVARIATE to get the standard deviation of the residuals. Then use this code to make a quantile normal plot of the residuals (using the s.d. value to set SIGMA for the NORMAL subcommand):

```plaintext
PROC UNIVARIATE DATA=brianDiag NOPRINT;
   VAR resid;
   QQPLOT / NORMAL (MU=0 SIGMA=0.34488 COLOR=red);
RUN;
```

Interpret the residual plots.