

Homework #5**DUE Thursday 6 December in class**

MRFIT was a randomized controlled trial on the primary prevention of CHD mortality among men ages 35 to 57 years who were at increased risk for, but without definitive clinical evidence of, CHD (coronary heart disease) at baseline. From 1973 to 1975, 22 clinical centers in 18 cities across the U.S. screened 361,662 men for participation, and 12,866 men were randomized into the study. We will use a random subset of the randomized men here.

The randomized intervention was comprised of dietary counseling to lower cholesterol, smoking cessation counseling for smokers, and hypertension medication if needed (called “special intervention,” `group=1`). The placebo group (called “usual care,” `group=2`) remained under the care of their primary physicians. On or about each anniversary of randomization for six years, participants returned to their clinical center for an examination by a MRFIT physician and to complete behavioral, medical history, and 24 hour dietary recall questionnaires. Lipids (total cholesterol, LDL-cholesterol, HDL-cholesterol) were measured from a fasting serum sample but only at baseline, visit 2 (24 months), visit 4 (48 months), and visit 6 (72 months).

The purpose of this homework is to carry out analyses of these data to answer the following scientific question: **Were the trajectories across time in LDL-cholesterol and in HDL-cholesterol different for the two randomized treatment groups? If so, how were they different?** Note that this is doubly-repeated measures (two cholesterol measures at each of four equally-spaced visits) and that the scales for LDL and HDL are different. You should NOT, however, rescale them; we will use a mean model that can accommodate the two scales. Note also that higher LDL is bad, but higher HDL is good.

Do NOT do a full analysis from beginning to end; just follow the questions below. Do NOT do any EDA. Do NOT fit any more covariance structures than those requested below. Do NOT do any mean model reduction. Do NOT do any diagnostics. The focus of this homework is on understanding how to fit and interpret a doubly-repeated measures model, not on the model selection process.

Answer the following questions.

1. We have repeated measures across time. Therefore, what is a cluster in this study? How big is the Σ_i for each cluster? Explain how you arrived at your answer.
2. These data could actually be considered triply-repeated measures. Explain.
3. Fit the following model:

```
proc mixed data=repdat method=reml;
  class type cvisit;
  model chol = type|si|visit agebl bmibl drinksbl
    fglubl cigsbl black incomebl sbpbl;
  random int / subject=id group=type v vcorr;
run;
```

(The variables `age`, `bmibl`, `drinksbl`, `fglubl`, `cigsbl`, `black`, `incomebl`, and `sbpbl` are being used here as adjusting variables. Keep them in the model, but you can ignore them in your SAS output. They are not needed to answer the questions. For simplicity for this HW, we are being a bit sloppy in our mean model here, because we are allowing the treatment effect (`group`) to be in effect at the baseline visit. We really should also be modeling time as piecewise linear (bent line), as in Weiss 7.6.)

What is the correlation among HDL values at different visits? Among LDL values at different visits? What is the correlation between LDL and HDL at the same visit? At different visits?

4. Examine the **R Correlation Matrix** output from the model. Describe how the correlation matrix is comprised of blocks, and explain what each of the blocks represents. (Note that the interpretation of these blocks depends entirely on how your data are sorted!)
5. Now fit the following model:

```
proc mixed data=repdat method=reml;
  class type cvisit;
  model chol = type|si|visit agebl bmibl drinksbl
    fglubl cigsbl black incomebl sbpbl;
  repeated type*cvisit / subject=id type=un r rcorr;
run;
```

What is the correlation among HDL values at different visits? Among LDL values at different visits? What is the correlation between LDL and HDL at the same visit? At different visits?

6. Decide which model provides the better covariance structure by carrying out an LRT (if possible) or by comparing AIC and BIC values.
7. Take the output from your chosen model and write a short paragraph which answers the scientific questions of interest.

```
OPTIONS LS=165 PS=MAX NOCENTER NODATE FORMDLIM='=';
data new;
  infile '/home/merganser/course_data/correlated.data/mrfit.subset.dat' firstobs=2;
  input agebl bmibl drinksbl fglubl cigsbl black incomebl id clinic $ si
    onmedsbl onmeds12 onmeds24 onmeds36 onmeds48 onmeds60 onmeds72
    sbpbl sbp12 sbp24 sbp36 sbp48 sbp60 sbp72
    dbpbl dbp12 dbp24 dbp36 dbp48 dbp60 dbp72
    pcholbl pchol24 pchol36 pchol48 pchol60 pchol72
    ptrigbl ptrig24 ptrig36 ptrig48 ptrig60 ptrig72
    hdlbl hdl24 hdl48 hdl72 ldlbl ldl24 ldl48 ldl72;
  keep agebl bmibl drinksbl fglubl cigsbl black incomebl id si sbpbl
    hdlbl hdl24 hdl48 hdl72 ldlbl ldl24 ldl48 ldl72;
run;
data repdat;
  set new;
  chol = ldlbl; type = "LDL"; visit = 0; output;
  chol = ldl24; type = "LDL"; visit = 2; output;
  chol = ldl48; type = "LDL"; visit = 4; output;
  chol = ldl72; type = "LDL"; visit = 6; output;
  chol = hdlbl; type = "HDL"; visit = 0; output;
  chol = hdl24; type = "HDL"; visit = 2; output;
  chol = hdl48; type = "HDL"; visit = 4; output;
  chol = hdl72; type = "HDL"; visit = 6; output;
  keep agebl bmibl drinksbl fglubl cigsbl black incomebl id si sbpbl
    chol type visit;
run;
data repdat;
```

```
set repdat;
cvisit = visit;
run;
proc sort data=repdat out=repdat;
  by id type visit;
proc print data=repdat;
  var id chol type visit si agebl bmibl drinksbl;
  where id=5328 or id=5455;
run;
```