**Lab 9: Residual Analysis for Random Effects Models**

**Overview**

**In this lab, we calculate residuals to diagnose our random effects models. In the pain data, we fit the random intercept model with treatment by coping style interaction. We calculate fitted values and random effects and plot the residuals and random effects as functions of the predictors coping style and treatment. We look at**

**(1) Histograms of the estimated random effect by coping style, and**

**(2) Profile plots of the residuals Y\_{ij} - x\_{ij}'\hat{\alpha} - \beta\_i**

**(a) overall**

**(b) separately for attenders and distracters**

**(c) 6 plots separately by coping style \* tmt.**

**For the weight loss data, we build up to a more complicated model starting from a simple model. Each successor model is motivated by the information gleaned from the residual plot of the previous model. In particular, we fit 4 models**

**1. Random intercept**

**2. Fixed intercept and slope + random intercept**

**3. Unstructured mean + random intercept**

**4. Unstructured mean + random intercept and slope**

**For each model, we plot a**

**(a) profile plot of the fitted values, and**

**(b) profile plot of the residuals**

**Y\_{ij} - x\_{ij}'\hat{\alpha} - \beta\_i or**

**Y\_{ij} - x\_{ij}'\hat{\alpha} - \beta\_{i1} - t\_{ij} \beta\_{i2}**

**and for model 4, we plot a scatterplot of the intercepts versus slopes.**

Acknowledgement: Some code was adapted from the UCLA Academic Technology Services website at <https://stats.idre.ucla.edu/other/examples/mld-weiss/> chapter 10.

**Lab Tasks**

TODO: For each residual plot, determine what information you glean from the plot, and decide what features the next model should have. While we don’t fit a ‘next model’ to the pain data, there are at least three changes that could be made to the pain data models to improve the models.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

Assign appropriate working directory. Also, set up output file to save output (ods rtf file command). Last command at the very end of the lab closes the file. Download the pain and weightloss1 data sets from the web site. Do a little data management.

%let workdir = C:\Documents and Settings\Leannica\Desktop\SCHOOL\B236 Winter 2008;

\*%let workdir = C:\rob\Courses\rm\SAShandouts\lab x residuals;

libname b236 "&workdir";

ods rtf file = "&workdir\newlaboutput &SYSDATE..doc";

**data** b236.pain;

set b236.pain;

if trial < **4** then tmt = "zbaseline";

if trial= **4** then tmt=treatment;

**run**;

\*Random Intercept Model for Pain Data Set;

**proc** **mixed** data=b236.pain;

 class id cs tmt;

 model l2paintol= cs cs\*tmt/s outp=outp(keep=id cs tmt treatment resid trial); /\*outp gives us predicted values\*/

 random intercept/s subject=id ; /\*s gives us solution for the random effects\*/

 ods output solutionr=sr; /\*outputs solutions to random effects into a data set\*/

 title 'Random Intercept Model for Pain Data';

**run**;

\*GET ID's and coping styles;

**proc** **freq** data=b236.pain noprint;

 tables id\*cs/out=idcs;

**run**;

\*Sort Data Sets;

**proc** **sort** data=idcs;by id;**run**;

**proc** **sort** data=sr;by id;**run**;

\*Combine Random Effect Estimates with Coping Styles;

**data** re;

 merge idcs(keep=id cs) sr(keep=id estimate);

 label estimate='Random Effect Estimate';

**run**;

\*Histograms of Random Effects by Coping Style;

**proc** **univariate** data=re;

 class cs;

 histogram estimate/kernel midpoints=(-**3** to **3** by **0.5**);;

 title 'Histograms of Random Effects Estimates by Coping Style';

 label cs='Coping Style';

**run**;

\*Or overlapping...;

goptions colors=(black white);

**proc** **gchart** data = re;

 vbar estimate/subgroup=cs midpoints=(-**1.5** to **3** by **.25**);

 title 'Histograms of Random Effects Estimates by Coping Style';

**run**; **quit**;

\*Set up axes;

goptions reset = all;

axis1 minor=none label = ("Trial");

axis2 minor=none label=(angle=**90** "Log2(Pain Tolerance(secs))") order=(-**2** to **3** by **1**);

\*Profile plot of residual;

**proc** **gplot** data=outp;

 plot resid\*trial=id/nolegend haxis=axis1 vaxis=axis2;

 title 'Profile Plot of Pain Data Residuals';

 symbol i=join v=none c=black r=**64**;

**run**;

\*Plot separate lines for each coping style and treatment;

/\* Since the repeat option 'r=' in the symbol statements uses the order of subject IDs (regardless how the data is sorted), and ID is not coded in the order of coping style, we need to create a column of new IDs that are in the order of coping style for the symbol statements\*/

**proc** **sort** data=Idcs; by CS; **run**;

**data** IDcs; set IDcs;

IDnew=\_n\_;

**run**;

/\* copy the new IDs to the residuals data by merging datasets \*/

**proc** **sort** data=outp; by id; **run**;

**proc** **sort** data=idcs; by id; **run**;

**data** outp (drop=count percent); merge idcs outp; by id; **run**;

\*Red and black lines for Coping Style on one plot;

**proc** **gplot** data=outp;

 title 'Profile Plot of Residuals by Coping Style';

 title2 color=black "attenders ------ distracters " color=red "--------";

 symbol i=join v=none c=black r=**32** /\*32 attenders black\*/;

 symbol2 i=join v=none c=red r=**32** /\*32 distraters red\*/;

 plot resid\*trial = IDnew/ haxis=axis1 vaxis=axis2 nolegend;

**run**;

\*ON SEPARATE PLOTS;

\*Coping style on separate plots;

**proc** **sort** data=outp; by cs treatment;**run**;

/\*Figure 10.3, page 337\*/

**proc** **gplot** data = outp;

 by cs;

 symbol1 value=none color=black interpol = join r = **32**;

 plot resid\*trial=id/ nolegend haxis=axis1 vaxis=axis2 name='f10\_3\_';

 title 'Profile Plot of Residuals by Coping Style';

**run**; **quit**;

\*Print both plots on one page;

**proc** **greplay** igout = work.gseg tc=sashelp.templt template=h2 nofs;

 \*igout= catalogs where graphs will be stored;

 \*nofs= don't want the GREPLAY window to appear;

 \*tc is template catalog, sashelp.templt has some already created templates;

 \*template = h2, we want the template h2 which has 2 plots horizontally;

 treplay **1**:f10\_3\_ **2**:f10\_3\_1 ;

**run**; **quit**;

\*By Coping Style and Treatment;

**proc** **gplot** data=outp;

 title 'Profile Plot of Residuals by Coping Style and Treament';

 by cs treatment;

 plot resid\*trial = id/ haxis=axis1 vaxis=axis2 nolegend name='f10\_3n\_';

**run**;

\*Create a new template to get 2x3 since it does not exist already...;

\*Create a temporary catalog called tempcat;

 **proc** **greplay** tc=tempcat nofs;

 tdef newtemp des='Six panel template 2x3'

 **1**/llx=**0** lly=**50**

 ulx=**0** uly=**100**

 urx=**33.3** ury=**100**

 lrx=**33.3** lry=**50**

 color=black

 **2**/llx=**33.3** lly=**50**

 ulx=**33.3** uly=**100**

 urx=**66.6** ury=**100**

 lrx=**66.6** lry=**50**

 color=black

 **3**/llx=**66.6** lly=**50**

 ulx=**66.6** uly=**100**

 urx=**100** ury=**100**

 lrx=**100** lry=**50**

 color=black

 **4**/llx=**0** lly=**0**

 ulx=**0** uly=**50**

 urx=**33.3** ury=**50**

 lrx=**33.3** lry=**0**

 color=black

 **5**/llx=**33.3** lly=**0**

 ulx=**33.3** uly=**50**

 urx=**66.6** ury=**50**

 lrx=**66.6** lry=**0**

 color=black

 **6**/llx=**66.6** lly=**0**

 ulx=**66.6** uly=**50**

 urx=**100** ury=**50**

 lrx=**100** lry=**0**

 color=black

 ;

 template newtemp;

 list template;

**quit**;

\*Print all 6 plots on one page;

**proc** **greplay** igout = work.gseg tc=tempcat template=newtemp nofs;

 treplay **1**:f10\_3n\_ **2**:f10\_3n\_1 **3**:f10\_3n\_2 **4**:f10\_3n\_3 **5**:f10\_3n\_4 **6**:f10\_3n\_5;

**run**; **quit**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*WEIGHTLOSS DATA\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*SET GRAPH OPTIONS FOR WEIGHT DATA;

 goptions reset = all;

 symbol i=join v=none c=black r=**38**;

 axis1 label= (angle=**90** "Weight(lbs)") order=(**130** to **270** by **10**);

 axis2 order =(**1** to **8** by **1**) label=(a=**0** 'Week') minor=none;

 axis3 label= (angle=**90** "Weight(lbs)") order=(-**15** to **15** by **5**);

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Model 1: Random Intercept

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **mixed** data=b236.weight1 method=ml covtest noitprint;

 class id;

 model weight=/s outp=pred;

 random intercept/subject=id type=un g s;

 title 'Model 1: Random Intercept';

**run**;

\*Profile Plot of Fitted Values;

**proc** **gplot** data=pred;

 title 'Predicted Weight - RI';

 plot pred\*week=id/ vaxis=axis1 haxis=axis2 nolegend name='preda';

**run**;

\*The Profile plot of fitted values isn't that interesting in this simple model but it can be quite interesting in more complex models, see the book cover for an example. ;

\*Profile Plot of Residuals;

**proc** **gplot** data=pred;

 title 'Residual Weight - RI';

 plot resid\*week=id/ vaxis=axis3 haxis=axis2 nolegend name='resida';

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Model 2: Fixed Intercept and Slope Random Intercept

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **mixed** data=b236.weight1 method=ml covtest noitprint;

 class id;

 model weight=week/s outp=pred2;

 random intercept/subject=id type=un g s;

 title 'Model 2: Fixed Intercept and Slope Random Intercept';

**run**;

\*Profile Plot of Predicted;

**proc** **gplot** data=pred2;

 title 'Predicted Weight – Fixed Slope, RI';

 plot pred\*week=id/ vaxis=axis1 haxis=axis2 nolegend name='predb';

**run**;

\*Profile Plot of Residuals;

**proc** **gplot** data=pred2;

 title 'Residual Weight – Fixed Slope, RI';

 plot resid\*week=id/ vaxis=axis3 haxis=axis2 nolegend name='residb';

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Model 3: Unstructured Mean, Random Intercept

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**proc** **mixed** data=b236.weight1 method=ml covtest noitprint;

 class id week;

 model weight=week/s outp=pred3;

 random intercept/subject=id type=un g s;

 title 'Model 3: Unstructured Mean, Random Intercept';

 ods output solutionr=sr3;

**run**;

\*Profile Plot of Predicted;

**proc** **gplot** data=pred3;

 title 'Predicted Weight - Unstructured Mean, Random Intercept';

 plot pred\*week=id/ vaxis=axis1 haxis=axis2 nolegend name='predc';

**run**;

\*Profile Plot of Residuals;

**proc** **gplot** data=pred3;

 title 'Residual Weight - Unstructured Mean, Random Intercept';

 plot resid\*week=id/ vaxis=axis3 haxis=axis2 nolegend name='residc';

**run**;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Model 4: Unstructured Mean, RI and Slope

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*need both a categorical and continuous version of the week variable;

**data** tweight1;

 set b236.weight1;

 cont\_week=week;

**run**;

**proc** **mixed** data=tweight1 method=ml covtest noitprint;

 class id week;

 model weight=week/s outp=pred4;

 random intercept cont\_week/subject=id type=un g s;

 title 'Model 4: Unstructured Mean, RI and Slope';

 ods output solutionr=sr4;

**run**;

\*Profile Plot of Predicted;

**proc** **gplot** data=pred4;

 title 'Predicted Weight - Unstructured Mean, RI and Slope';

 plot pred\*week=id/ vaxis=axis1 haxis=axis2 nolegend name='predd';

**run**;

\*Profile Plot of Residuals;

**proc** **gplot** data=pred4;

 title 'Residual Weight - Unstructured Mean, RI and Slope';

 plot resid\*week=id/ vaxis=axis3 haxis=axis2 nolegend name='residd';

**run**;

\*Display multiple plots per page, and incidentally induce a better shape for the plots.

\*All 4 predicted plots on one page 2x2;

**proc** **greplay** igout = work.gseg tc=sashelp.templt template=l2r2 nofs;

 treplay **1**:preda **2**:predc **3**:predb **4**:predd;

**run**; **quit**;

\*All 4 predicted plots on one page - 1x4;

**proc** **greplay** igout = work.gseg tc=sashelp.templt template=h4 nofs;

 treplay **1**:preda **2**:predb **3**:predc **4**:predd;

**run**; **quit**;

\*All 4 residual plots on one page - 2x2;

/\*Figure 10.5, page 339\*/

**proc** **greplay** igout = work.gseg tc=sashelp.templt template=l2r2 nofs;

 treplay **1**:resida **2**:residc **3**:residb **4**:residd;

**run**; **quit**;

\*All 4 residual plots on one page - 1x4;

**proc** **greplay** igout = work.gseg tc=sashelp.templt template=h4 nofs;

 treplay **1**:resida **2**:residb **3**:residc **4**:residd;

**run**; **quit**;

\*TODO: Which shape do you like best? Why?

\*SCATTER PLOT OF INTERCEPT VS. SLOPE FOR MODEL 4;

\*Extra step needed to modify data set since we have both RI and Slope....;

**data** sr4mod;

 set sr4;

 by id;

 retain intercept;

 if first.id then intercept=estimate;

 else slope=estimate;

 if last.id then output;

 keep id intercept slope;

**run**;

\*Scatter Plot of Intercept vs. Slope;

**proc** **gplot** data=sr4mod;

 title 'Scatterplot of Intercept vs. Slope, Weight Model 4';

 symbol i=none v=star r=**38**;**run**;

 axis1 label= (angle=**90** "Slope");

 axis2 label=(angle=**0** 'Intercept');

 plot slope\*intercept/ vaxis=axis1 haxis=axis2 nolegend;

**run**;

ods rtf close;