Prof. Brent Coull

BIO 226: Applied Longitudinal Analysis

Homework 3 Solutions
Due Tuesday, March 10, 2015
[100 points]

Purpose:
To provide an introduction to the analysis of response profiles for longitudinal data.

Instructions:
1. For each question requiring data analysis, support your conclusions by including the relevant SAS output in your answer.
2. Include your SAS program (but not your SAS output) as an appendix to your solutions. In general, this will only be reviewed during grading to help identify a major problem affecting your answers to questions so please do not cross-reference the appendix in your answers to questions.
Late homework will not be graded unless you make arrangements with the Instructor prior to the due date/time.

Analysis of Response Profiles: Study of effects of treatment on rheumatoid arthritis:
A randomized clinical trial was completed to compare the effectiveness of 2 rheumatoid arthritis treatments. The grip strength was measured on each of the patients at 4 time points: week 0, week 1, week 2 and week 3. Grip strength is a continuous outcome. The data set is complete and balanced. Note that only a subset of patients is included in the data set for this assignment. We are most interested in determining the association between treatment and grip strength.
The data are stored in an ASCII file: compgrip.txt. Each row of the data set contains the following six variables: subject ID number, treatment indicator (1=treatment A and 2=treatment B), Y0, Y1, Y2, Y3.

Problem 1

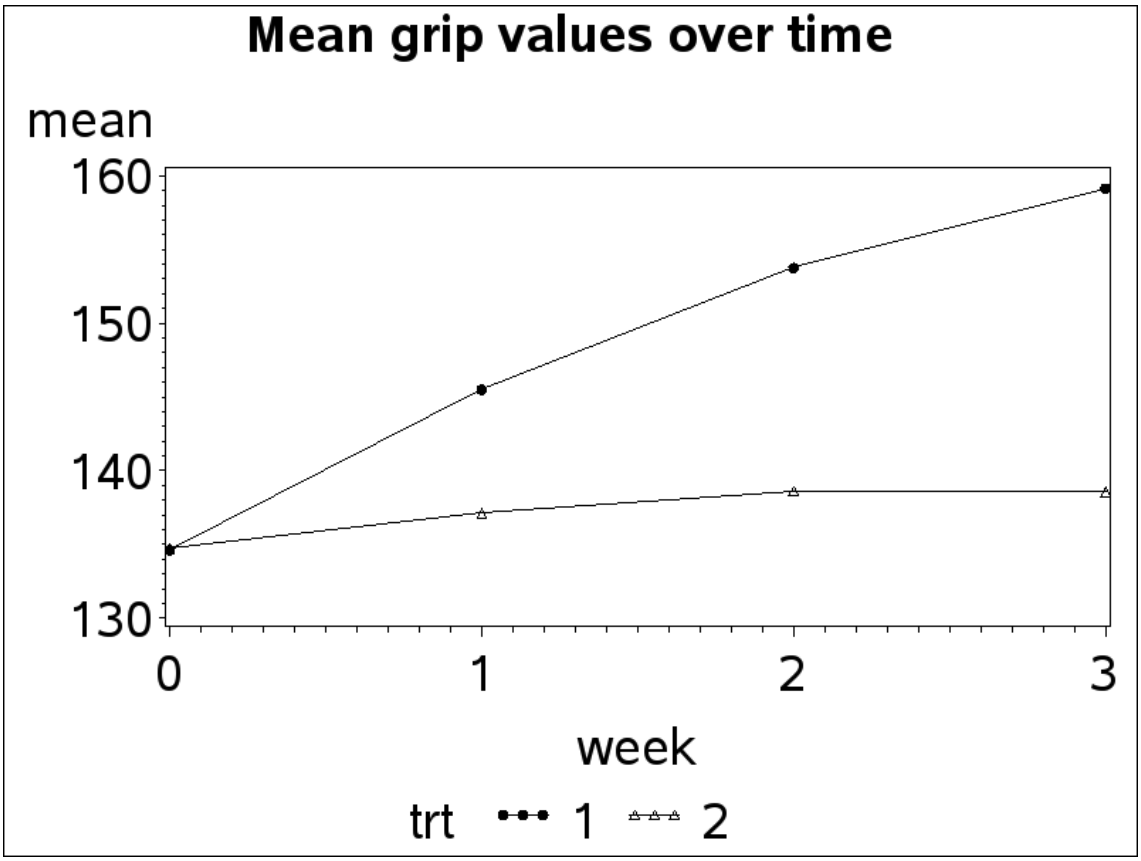
[15 points: 5 for means and std dev, 5 for plot, 5 for description] Obtain the sample size, and the sample means and standard deviations of the grip strengths at each occasion for each treatment group. On the same graph, plot the mean grip strength versus time (in weeks) for each of the two treatment groups. Describe the general characteristics of the time trends for the two groups.

```
data compgrip;
infile 'compgrip.txt';
input id trt Y0 Y1 Y2 Y3;
run;

data compgrip2;
set compgrip;
week=0; weekcat=0; Y=Y0; output;
week=1; weekcat=1; Y=Y1; output;
week=2; weekcat=2; Y=Y2; output;
```

```
week=3; weekcat=3; Y=Y3; output;
keep id trt week weekcat Y;
run;


proc sort data=compgrip2;
by trt descending week;
run;

proc means data=compgrip2 n mean std stderr;
  title 'Univariate y';
  var Y;
  by trt descending week;
  output out=meandata mean=mean std=sd;
run;

proc print data=meandata;
run;


title 'Mean grip values over time';
proc gplot data=meandata;
symbol1 color=black interpol=join value=dot;
symbol2 color=black interpol=join value=triangle;
plot mean*week=trt;
run;
```

Obs	trt	week	_TYPE_	_FREQ_	mean	sd
1	1	3	0	26	159.154	71.4124
2	1	2	0	26	153.769	70.9356
3	1	1	0	26	145.500	68.9401
4	1	0	0	26	134.615	71.2938
5	2	3	0	32	138.594	73.5540
6	2	2	0	32	138.625	71.4480
7	2	1	0	32	137.125	75.0315
8	2	0	0	32	134.750	72.7071

Mean grip strength increases sharply over time for individuals on treatment A while the increase is much more modest for those on treatment B. The two groups have very similar means at baseline (as would be expected in a randomized study).


Mean grip values over time
mean
160
150
140
130
0
1
2
3
week
trt
1
2

Problem 2

[10 points. It is completely ok if they do not write this out in matrix form. If they didn't write the distribution of the error terms and mention that the Σ is unstructured, 2 points off.] With baseline (week 0) and the treatment A as the reference group, write out the complete definition of the regression model for the analysis of response profiles for mean grip strength. In this model, let β denote the vector of parameters in the model for the means and assume an unstructured variance-covariance structure.

Let

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i0} \\ Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = \begin{pmatrix} \text{individual } i\text{'s grip strength at baseline} \\ \text{individual } i\text{'s grip strength at week 1} \\ \text{individual } i\text{'s grip strength at week 2} \\ \text{individual } i\text{'s grip strength at week 3} \end{pmatrix}$$

Lets consider the different covariates:

$$\begin{aligned} X_{1ij} &= 1 \text{ for all } i \text{ and } j, \\ X_{2ij} &= \begin{cases} 1 & \text{if corresponding measure at week 1} \\ 0 & \text{otherwise,} \end{cases} \\ X_{3ij} &= \begin{cases} 1 & \text{if corresponding measure at week 2} \\ 0 & \text{otherwise,} \end{cases} \\ X_{4ij} &= \begin{cases} 1 & \text{if corresponding measure at week 3} \\ 0 & \text{otherwise.} \end{cases} \\ X_{5ij} &= \begin{cases} 1 & \text{if treatment is B} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

$$Y_{ij} = \beta_1 + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \beta_5 X_{5ij} + \beta_6 X_{2ij} * X_{5ij} + \beta_7 X_{3ij} * X_{5ij} + \beta_8 X_{4ij} * X_{5ij} + e_{ij}$$

where $i = 1, \ldots, 58, \quad j = 0, \ldots, 3$. We assume $\mathbf{e}_i = \begin{bmatrix} e_{i0} \\ e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix} \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is unstructured.

Or if you prefer to replace the X's with something more descriptive:

$$Y_{ij} = \beta_1 + \beta_2 \text{week1}_{ij} + \beta_3 \text{week2}_{ij} + \beta_4 \text{week3}_{ij} + \beta_5 \text{trtB}_{ij} + \beta_6 \text{week1}_{ij} * \text{trtB}_{ij} + \beta_7 \text{week2}_{ij} * \text{trtB}_{ij} + \beta_8 \text{week3}_{ij} * \text{trtB}_{ij} + e_{ij}$$

where $i = 1, \ldots, 58, \quad j = 0, \ldots, 3$.

We could alternatively write the model in matrix form, by looking at each time point and plugging in the appropriate values for the time variables, for example, for baseline $j = 0$, $\text{week1}_{i0} = 0$, $\text{week2}_{i0} = 0$, $\text{week3}_{i0} = 0$. Then $Y_{i0} = \beta_1 + \beta_5 \text{trtB}_{i0}$. So we have

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i0} \\ Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix} = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \qquad i = 1, \ldots, 58$$

where

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & X_{5i0} & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & X_{5i1} & X_{5i1} & 0 & 0 \\ 1 & 0 & 1 & 0 & X_{5i2} & 0 & X_{5i2} & 0 \\ 1 & 0 & 0 & 1 & X_{5i3} & 0 & 0 & X_{5i3} \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \end{pmatrix}$$

In the alternative notation,

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & \text{trtB}_{i0} & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \text{trtB}_{i1} & \text{trtB}_{i1} & 0 & 0 \\ 1 & 0 & 1 & 0 & \text{trtB}_{i2} & 0 & \text{trtB}_{i2} & 0 \\ 1 & 0 & 0 & 1 & \text{trtB}_{i3} & 0 & 0 & \text{trtB}_{i3} \end{pmatrix}$$

We assume $\mathbf{e}_i \sim \text{N}(\mathbf{0}, \boldsymbol{\Sigma})$, so $\mathbf{Y}_i \sim \text{N}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is unstructured.

Problem 3

[15 points: 10 points for the proc mixed results, 5 points for concluding that the groups are not parallel] Using PROC MIXED, fit the model described in question 2. Include the SOLUTION (or S) option on the MODEL statement to obtain estimates and standard errors for the components of β. Include the CHISQ option on the MODEL statement and hence evaluate whether there is evidence of a difference in the pattern of change over time in mean grip strength between the two treatment groups. Justify your answer.

There is evidence that the pattern of change over time in mean grip strength is different between the two treatment groups (i.e. not parallel), because the trt*week interactions are significant (p-value 0.0014 or 0.0031 depending on whether you prefer chisq or F tests) using a Multivariate Wald test.

```
proc sort data=compgrip2; by descending trt; run;

proc sort data=compgrip2; by descending week; run;

proc mixed data=compgrip2 order=data;
class id trt week;
model y=trt week trt*week / s chisq;
repeated week / type=un subject=id r rcorr;
run;
```

Solution for Fixed Effects

Effect	trt	week	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			134.62	14.1360	56	9.52	<.0001
trt	2		0.1346	19.0311	56	0.01	0.9944

The Mixed Procedure

Solution for Fixed Effects

Effect	trt	week	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1		0	.	.	.	.
week		3	24.5385	3.9255	56	6.25	<.0001
week		2	19.1538	3.4572	56	5.54	<.0001
week		1	10.8846	3.0531	56	3.57	0.0008
week		0	0	.	.	.	.
trt*week	2	3	-20.6947	5.2849	56	-3.92	0.0002
trt*week	2	2	-15.2788	4.6545	56	-3.28	0.0018
trt*week	2	1	-8.5096	4.1104	56	-2.07	0.0430
trt*week	2	0	0	.	.	.	.
trt*week	1	3	0	.	.	.	.
trt*week	1	2	0	.	.	.	.
trt*week	1	1	0	.	.	.	.
trt*week	1	0	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
trt	1	56	0.34	0.34	0.5596	0.5620
week	3	56	30.76	10.25	<.0001	<.0001
trt*week	3	56	15.58	5.19	0.0014	0.0031

Problem 4

[15 points] Provide an interpretation for the estimate of each component of β.

β_1 - average grip strength at baseline for those on treatment A.

β_2 - average change in grip strength between week 1 and baseline for those on treatment A.

β_3 - average change in grip strength between week 2 and baseline for those on treatment A.

β_4 - average change in grip strength between week 3 and baseline for those on treatment A.

β_5 - average difference in grip strength at baseline between those on treatment B versus A.

β_6 - average difference in the change in grip strength between week 1 and baseline for those on treatment B vs A.

β_7 - average difference in the change in grip strength between week 2 and baseline for those on treatment B vs A.

β_8 - average difference in the change in grip strength between week 3 and baseline for those on treatment B vs A.

Problem 5

[15 points: 10 for the numbers, 5 for saying identical] Show how the estimates for the components of β can be used to construct the sample means at each measurement time in each treatment group. Compare these estimated means with the sample means that you obtained in question 1. Do they differ? If so, can you suggest why?

Treatment A, week 0 $= \beta_1 = 134.62$

Treatment A, week 1 $= \beta_1 + \beta_2 = 134.62 + 10.8846 = 145.5046$

Treatment A, week 2 $= \beta_1 + \beta_3 = 134.62 + 19.1538 = 153.7738$

Treatment A, week 3 $= \beta_1 + \beta_4 = 134.62 + 24.5385 = 159.1585$

Treatment B, week 0 $= \beta_1 + \beta_5 = 134.62 + 0.1346 = 134.7546$

Treatment B, week 1 $= \beta_1 + \beta_2 + \beta_5 + \beta_6 = 134.62 + 10.8846 + 0.1346 - 8.5096 = 137.1296$

Treatment B, week 2 $= \beta_1 + \beta_3 + \beta_5 + \beta_7 = 134.62 + 19.1538 + 0.1346 - 15.2788 = 138.6296$

Treatment B, week 3 $= \beta_1 + \beta_4 + \beta_5 + \beta_8 = 134.62 + 24.5385 + 0.1346 - 20.6947 = 138.5984$

These are almost identical to the means from question 1, but slightly different from rounding error. This is the case because we are fitting the saturated model.

Problem 6

[15 points: 10 for the code, 5 for the same conclusion as question 3. If people did not do a LRT test, but just looked at the significance of the interaction term trt*week, 7 points off.] Now conduct a profile analysis (week, treatment, week*treatment) on the difference vector $D = (Y1 - Y0, Y2 - Y0, Y3 - Y0)$. Show that this analysis yields the same conclusions as those obtained from the full profile analysis fit in question 3.

By Lecture 7 slide 37, we know that the original test of "parallelism of profiles" now becomes a joint test of main effect of trt and the trt*week interaction. As in question 3, there is evidence that the pattern of change over time in mean grip strength is different between the two treatment groups, because the interaction trt*week and the main effect trt are jointly significant (p-value 0.0026).

```
data compgrip_diff;
set compgrip;
D1 = Y1 - Y0;
D2 = Y2 - Y0;
D3 = Y3 - Y0;
run;

data compgrip_diff2;
set compgrip_diff;
week=1; weekcat=1; D=D1; output;
week=2; weekcat=2; D=D2; output;
week=3; weekcat=3; D=D3; output;
keep id trt week weekcat D;
run;

proc sort data=compgrip_diff2; by descending trt; run;

proc sort data=compgrip_diff2; by descending week; run;
```

```
title 'FULL MODEL';
proc mixed data=compgrip_diff2 order=data method=ml;
class id trt week;
model D=trt week trt*week / s chisq;
repeated week / type=un subject=id r rcorr;
run;
title;

title 'REDUCED MODEL';
proc mixed data=compgrip_diff2 order=data method=ml;
class id trt week;
model D=week / s chisq;
repeated week / type=un subject=id r rcorr;
run;
title;
```

FULL MODEL

Fit Statistics

-2 Log Likelihood	1413.6

REDUCED MODEL

Fit Statistics

-2 Log Likelihood	1427.8

```
/* 1427.8 - 1413.6 = 14.2 */

title 'pval';
DATA pvalues;
chsq = SDF('chisquare',14.2,3);
RUN;
PROC PRINT DATA=pvalues;
RUN;
title;
```

Obs	chsq
1	.00264518

Problem 7

[15 points: 5 points for writing down the model, 5 points for interpretations, 5 points for fitting the model]
Write down a precise definition of the ANCOVA model, just as you did in question 2 for the profile analysis. Then run this ANCOVA analysis, and interpret the parameters of the ANCOVA model, as you did in question 4 for the profile analysis. There is no need to explicitly compare the parameters from the ANCOVA model to the parameters from the profile analysis model.

If you chose not to center Y_{i0}, this is your model:

$$Y_{ij} = \beta_1 + \beta_2 \text{week2}_{ij} + \beta_3 \text{week3}_{ij} + \beta_4 \text{trtB}_{ij} + \beta_5 \text{week2}_{ij} * \text{trtB}_{ij} + \beta_6 \text{week3}_{ij} * \text{trtB}_{ij} + \beta_7 Y_{i0} + e_{ij}$$

where $i = 1, \ldots, 58, \quad j = 1, \ldots, 3$. We assume $\mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix} \sim \text{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is unstructured.

β_1 - mean grip strength in week 1 among people in treatment A who have zero grip strength at baseline

β_2 - average change in grip strength between week 2 and week 1, among people in treatment A who have the same baseline measurement.

β_3 - average change in grip strength between week 3 and week 1, among people in treatment A who have the same baseline measurement

β_4 - average difference in grip strength between treatments A and B at week 1, among people who have the same baseline measurement

β_5 - average difference in the change in grip strength between week 2 and week 1 for those on treatment A versus B, comparing people with the same baseline measurement

β_6 - average difference in the change in grip strength between week 3 and week 1 for those on treatment A versus B, comparing people with the same baseline measurement

β_7 - for every unit increase in baseline Y_{i0}, the grip strength is predicted to be β_7 higher, among measurements at the same time and same treatment group.

If you chose to center Y_{i0}, this is your model (changes from above are in blue):

$$Y_{ij} = \beta_1 + \beta_2 \text{week2}_{ij} + \beta_3 \text{week3}_{ij} + \beta_4 \text{trtB}_{ij} + \beta_5 \text{week2}_{ij} * \text{trtB}_{ij} + \beta_6 \text{week3}_{ij} * \text{trtB}_{ij} + \beta_7 (Y_{i0} - \text{mean}(Y_{i0})) + e_{ij}$$

where $i = 1, \ldots, 58, \quad j = 1, \ldots, 3$. We assume $\mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix} \sim \text{N}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is unstructured.

β_1 - mean grip strength in week 1 among people in treatment A who have the average grip strength at baseline

β_2 - average change in grip strength between week 2 and week 1, among people in treatment A who have the same baseline measurement.

β_3 - average change in grip strength between week 3 and week 1, among people in treatment A who have the same baseline measurement

β_4 - average difference in grip strength between treatments A and B at week 1, among people who have the same baseline measurement

β_5 - average difference in the change in grip strength between week 2 and week 1 for those on treatment A versus B, comparing people with the same baseline measurement
β_6 - average difference in the change in grip strength between week 3 and week 1 for those on treatment A versus B, comparing people with the same baseline measurement
β_7 - for every unit increase in baseline Y_{i0}, the grip strength is predicted to be β_7 higher, among measurements at the same time and same treatment group.

```
data compgrip_ancova;
set compgrip;
week=1; weekcat=1; Y=Y1; output;
week=2; weekcat=2; Y=Y2; output;
week=3; weekcat=3; Y=Y3; output;
keep id trt week weekcat Y Y0;
run;

/* Without centering */

proc sort data=compgrip_ancova; by descending trt; run;
proc sort data=compgrip_ancova; by descending week; run;

proc mixed data=compgrip_ancova order=data;
class id trt week;
model Y = trt week trt*week Y0 / s chisq;
repeated week / type = un subject = id r corr;
run;

/* With centering */

proc means data=compgrip n mean std stderr;
  title 'Baseline Y';
  var Y0;
run;

/* Now take the mean(Y0) from the output and do something like
data compgrip_ancova;
set compgrip_ancova;
Y0c = Y0 - mean(Y0);
run;
*/

proc sort data=compgrip_ancova; by descending trt; run;
proc sort data=compgrip_ancova; by descending week; run;

proc mixed data=compgrip_ancova order=data;
class id trt week;
model Y = trt week trt*week Y0c / s chisq;
repeated week / type = un subject = id r corr;
```

```
run;
```