

## Mid-Term Exam – ST 732 – Spring 2009

Turn off cell phones and put them away as well as any other electronic devices except for your calculator. Put all your answers on separate pieces of paper (not on the test itself).

For problems 1-4. One model for longitudinal or repeated measures data is

$$Y_{hlj} = \mu + \tau_l + \gamma_j + (\tau\gamma)_{lj} + b_{hl} + e_{hlj},$$

where  $\mu$  is an overall mean,  $\tau_l$  is the deviation from the overall mean associated with the  $l$ th group,  $\gamma_j$  is the deviation associated with time  $j$ ,  $(\tau\gamma)_{lj}$  is the interaction effect for group  $l$  and time  $j$ ,  $b_{hl}$  is a random effect for the  $h$ th unit in group  $l$ , and  $e_{hlj}$  is random error. Assume that all the random quantities are independent and  $\text{Var}(b_{hl}) = \sigma_b^2$  and  $\text{Var}(e_{hlj}) = \sigma_e^2$ . Suppose that there are responses at days 10, 20, and 50,  $q = 3$  groups, and 6 units in group 1, 7 units in group 2, and 8 units in group 3.

1. For the classical split plot analysis via `proc glm`, how many lines of data are there (assuming no missing values)?

2. Why would at least one of the default tests from the code

```
proc glm;class time group unit;model y=group unit(group) time group*time;
```

be wrong? What line of code could be added to give correct default tests?

3. For the more multivariate way to analyze these data in `proc glm` with a model statement that starts with `model time10 time20 time50 = ...`, how many lines of data are there?

4. Let all the responses for the  $h$ th unit be denoted by the vector  $\mathbf{Y}_i$ .

a) What are the covariance and correlation matrices of  $\mathbf{Y}_i$ ?

b) The “regression form” for the model is  $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$ . For the second unit in group 3, write down  $\mathbf{X}_i$  and  $\boldsymbol{\beta}$ . (Note that  $\mathbf{X}_i$  is not full rank.)

c) A matrix  $\mathbf{L}$  is used to test hypotheses of the form  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$ . What  $\mathbf{L}$  and  $\mathbf{h}$  can be used to test that the means of Groups 1 and 2 are equal at day 50?

5. Using the multivariate approach in 3. above,

a) what line of code can be added to test whether there are nonlinear effects such as quadratic and cubic?

b) what does the code `manova h=group` test for? Explain in terms of the means  $E(Y_{hlj}) = \mu_{lj}$ .

The rest of the questions relate to the following experiment and data. Thirty-six golfers with average scores around 100 were randomized to three treatment groups at the beginning of June one summer. Golfers in Treatment Groups 1 and 2 were given instruction in two competing methods for swinging the golf club. Treatment Group 3 got no new instruction, but they were allowed to play lots of rounds of golf. A few golfers dropped out along the way, so there were 10 in group 1, 12 in group 2, and 8 in group 3. (For the ungolfered, the goal is to shoot “par” on 18 holes of golf, usually 70-72 on most courses. Professionals tend to score around 70 or below. The average golfer scores around 100.) The participants played at least two rounds of golf each week all summer, and we record the average of the last two weekly scores for weeks 1,2,3,6, and 12 (actually listed as Day 7,14,21,42, and 84 on the data set). The goal of the experiment was to determine if the golf instruction improved scores and which groups improved the most.

6. I was too lazy to flip the data and print out covariance and correlation matrices using `proc corr`. So I did the next best thing and used `proc mixed` with a general mean structure and a `repeated` statement with `type=un group=group` that gives essentially the same results as using `proc corr`. This is listed under “Run 1.” Make a crude plot of means. Describe briefly what you see in terms of differences over time and between treatment groups. Remembering that sample variances and correlations are very “noisy,” do you see any patterns in variances or correlations of differences between groups? Explain your intuition, but do not waste a lot of time on this.

7. Next I tried a bunch of covariance structures in Runs 2-6. Pick the one you think is best and explain your choice.

8. Run 6 gives the simplest covariance structure. Using the output, write down the covariance and correlation matrix for the responses of unit one (all units have the same structure).

9. In Run 7 I have chosen a covariance structure and fit a quadratic model over time along with estimate and contrast statements.

a) Write down the estimated quadratic curves for each treatment group.

b) Give a test statistic and p-value for testing whether the linear slope of group 1 is equal to the linear slope of group 3.

c) Give a test statistic and p-value for testing whether the quadratic slope of group 1 is equal to the quadratic slope of group 2.

d) Give a test statistic and p-value for testing that the quadratic slopes are all 0.

e) Give a test statistic and p-value for testing that the quadratic slopes are all equal.

f) For a final model, what terms do you think should be dropped?

g) What estimate statement could be added to Run 7 to estimate the difference in means at day=84 for Groups 1 and 2?

-----  
Run 1

```
input unit group day score;
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=un group=group subject=unit(group) r=1,11,23 rcorr=1,11,23;
lsmeans group day group*day;
run;
```

First few lines of data (unit=golfer):

```
unit group day score
1 1 7 106.37
1 1 14 107.78
1 1 21 103.12
1 1 42 98.54
1 1 84 100.95
2 1 7 101.56
. . . .
. . . .
```

Part of the output:

Estimated R Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
1	62.2894	32.4406	17.1507	11.8377	2.8133
2	32.4406	45.3068	23.5660	12.7236	-0.1006
3	17.1507	23.5660	34.8059	14.7832	14.1995
4	11.8377	12.7236	14.7832	17.0601	17.1447
5	2.8133	-0.1006	14.1995	17.1447	29.7700

Estimated R Correlation Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.6107	0.3683	0.3631	0.06533
2	0.6107	1.0000	0.5934	0.4577	-0.00274
3	0.3683	0.5934	1.0000	0.6067	0.4411
4	0.3631	0.4577	0.6067	1.0000	0.7608
5	0.06533	-0.00274	0.4411	0.7608	1.0000

Estimated R Matrix for unit(group) 11 2

Row	Col1	Col2	Col3	Col4	Col5
1	23.6284	6.0233	1.3664	1.6213	0.1017
2	6.0233	8.6973	1.0615	0.7460	0.8350
3	1.3664	1.0615	2.5424	-0.08288	-0.6298
4	1.6213	0.7460	-0.08288	1.5626	0.5259
5	0.1017	0.8350	-0.6298	0.5259	1.5686

Estimated R Correlation Matrix for unit(group) 11 2

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.4202	0.1763	0.2668	0.01670
2	0.4202	1.0000	0.2257	0.2023	0.2261
3	0.1763	0.2257	1.0000	-0.04158	-0.3154
4	0.2668	0.2023	-0.04158	1.0000	0.3359
5	0.01670	0.2261	-0.3154	0.3359	1.0000

Estimated R Matrix for unit(group) 23 3

Row	Col1	Col2	Col3	Col4	Col5
1	47.5987	18.7588	25.5060	15.2513	27.9174
2	18.7588	27.0670	11.5750	8.3003	12.6073

3	25.5060	11.5750	22.2966	8.0971	20.4992
4	15.2513	8.3003	8.0971	20.7439	4.8854
5	27.9174	12.6073	20.4992	4.8854	30.5753

Estimated R Correlation Matrix for unit(group) 23 3

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.5226	0.7829	0.4854	0.7318
2	0.5226	1.0000	0.4712	0.3503	0.4382
3	0.7829	0.4712	1.0000	0.3765	0.7851
4	0.4854	0.3503	0.3765	1.0000	0.1940
5	0.7318	0.4382	0.7851	0.1940	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
UN(1,1)	unit(group)	group 1	62.2894
UN(2,1)	unit(group)	group 1	32.4406
.	.	.	.
UN(5,5)	unit(group)	group 3	30.5753

Fit Statistics

-2 Res Log Likelihood	720.6
AIC (smaller is better)	810.6
AICC (smaller is better)	857.1
BIC (smaller is better)	873.7

Least Squares Means

Effect	group	day	Estimate	Standard Error
group	1		97.4308	1.3890
group	2		91.6512	0.4514
group	3		97.9017	1.5084
day		7	99.2822	1.2538
day		14	98.2326	0.9797
day		21	94.7084	0.8485
day		42	93.6527	0.7015
day		84	92.4303	0.8775
group*day	1	7	101.24	2.4958
group*day	1	14	98.1410	2.1285
group*day	1	21	97.2070	1.8656
group*day	1	42	94.5780	1.3061
group*day	1	84	95.9900	1.7254
group*day	2	7	99.8400	1.4032
group*day	2	14	95.6292	0.8513
group*day	2	21	89.7008	0.4603
group*day	2	42	88.2525	0.3609
group*day	2	84	84.8333	0.3615
group*day	3	7	96.7687	2.4392
group*day	3	14	100.93	1.8394
group*day	3	21	97.2175	1.6695
group*day	3	42	98.1275	1.6103
group*day	3	84	96.4675	1.9550

-----  
Run 2

```
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=un subject=unit(group) r=1,11,23 rcorr=1,11,23;
```

Estimated R Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
-----	------	------	------	------	------

1	42.7299	18.1309	12.8863	8.5605	8.2170
2	18.1309	25.6630	11.2887	6.6970	3.5752
3	12.8863	11.2887	18.4184	6.9932	9.7912
4	8.5605	6.6970	6.9932	11.7014	7.1958
5	8.2170	3.5752	9.7912	7.1958	18.4893

Estimated R Correlation Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.5475	0.4593	0.3828	0.2923
2	0.5475	1.0000	0.5192	0.3865	0.1641
3	0.4593	0.5192	1.0000	0.4764	0.5306
4	0.3828	0.3865	0.4764	1.0000	0.4892
5	0.2923	0.1641	0.5306	0.4892	1.0000

Fit Statistics

-2 Res Log Likelihood	788.5
AIC (smaller is better)	818.5
AICC (smaller is better)	822.6
BIC (smaller is better)	839.5

-----  
Run 3

```
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=csh subject=unit(group) r=1,11,23 rcorr=1,11,23;
```

Estimated R Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
1	42.9074	14.2569	11.6298	9.5157	12.2878
2	14.2569	26.0955	9.0696	7.4209	9.5828
3	11.6298	9.0696	17.3644	6.0535	7.8170
4	9.5157	7.4209	6.0535	11.6251	6.3960
5	12.2878	9.5828	7.8170	6.3960	19.3849

Estimated R Correlation Matrix for unit(group) 1 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.4261	0.4261	0.4261	0.4261
2	0.4261	1.0000	0.4261	0.4261	0.4261
3	0.4261	0.4261	1.0000	0.4261	0.4261
4	0.4261	0.4261	0.4261	1.0000	0.4261
5	0.4261	0.4261	0.4261	0.4261	1.0000

Fit Statistics

-2 Res Log Likelihood	797.3
AIC (smaller is better)	809.3
AICC (smaller is better)	809.9
BIC (smaller is better)	817.7

-----  
Run 4

```
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=csh group=group subject=unit(group);
```

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Var(1)	unit(group)	group 1	66.6055
Var(2)	unit(group)	group 1	45.7956

Var(3)	unit(group)	group 1	32.7191
Var(4)	unit(group)	group 1	15.4958
Var(5)	unit(group)	group 1	32.6867
CSH	unit(group)	group 1	0.4320
Var(1)	unit(group)	group 2	22.9498
Var(2)	unit(group)	group 2	8.2790
Var(3)	unit(group)	group 2	2.7038
Var(4)	unit(group)	group 2	1.5377
Var(5)	unit(group)	group 2	1.6280
CSH	unit(group)	group 2	0.1552
Var(1)	unit(group)	group 3	42.7720
Var(2)	unit(group)	group 3	29.1288
Var(3)	unit(group)	group 3	20.4827
Var(4)	unit(group)	group 3	24.5952
Var(5)	unit(group)	group 3	29.7918
CSH	unit(group)	group 3	0.5228

Fit Statistics	
-2 Res Log Likelihood	746.8
AIC (smaller is better)	782.8
AICC (smaller is better)	788.7
BIC (smaller is better)	808.0

-----  
Run 5

```
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=cs group=group subject=unit(group);
```

Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
Variance	unit(group)	group 1	23.1906
CS	unit(group)	group 1	14.6559
Variance	unit(group)	group 2	6.4430
CS	unit(group)	group 2	1.1568
Variance	unit(group)	group 3	14.3165
CS	unit(group)	group 3	15.3398

Fit Statistics	
-2 Res Log Likelihood	787.4
AIC (smaller is better)	799.4
AICC (smaller is better)	800.0
BIC (smaller is better)	807.8

-----  
Run 6

```
proc mixed data=one;class unit group day;
model score=group day group*day /solution ddfm=kr;
repeated day / type=cs subject=unit(group);
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	unit(group)	9.3336
Residual		14.0668

Fit Statistics	
-2 Res Log Likelihood	813.9
AIC (smaller is better)	817.9
AICC (smaller is better)	817.9
BIC (smaller is better)	820.7

-----

Run 7

```
input unit group day score;
day2=day**2;
proc mixed data=one;class unit group;
model score=group group*day group*day2 / noint solution ddfm=kr;
```

```
estimate 'estimate 1' group 1 -1 0;
estimate 'estimate 2' group 1 0 -1;
estimate 'estimate 3' group 0 1 -1;
estimate 'estimate 4' group*day 1 -1 0;
estimate 'estimate 5' group*day 1 0 -1;
estimate 'estimate 6' group*day 0 1 -1;
estimate 'estimate 7' group*day2 1 -1 0;
estimate 'estimate 8' group*day2 1 0 -1;
estimate 'estimate 9' group*day2 0 1 -1;
contrast 'contrast 1' group 1 -1 0, group 1 0 -1;
contrast 'contrast 2' group*day 1 -1 0, group*day 1 0 -1;
contrast 'contrast 3' group*day2 1 -1 0, group*day2 1 0 -1;
```

Solution for Fixed Effects

Effect	group	Estimate	Standard Error	DF	t Value	Pr >  t
group	1	102.47	3.0259	13.9	33.86	<.0001
group	2	100.64	1.3278	23.1	75.80	<.0001
group	3	97.2799	2.9054	14.3	33.48	<.0001
day*group	1	-0.3039	0.1144	34.3	-2.66	0.0119
day*group	2	-0.4095	0.05200	35.3	-7.87	<.0001
day*group	3	0.03456	0.1187	21.6	0.29	0.7738
day2*group	1	0.002703	0.001133	36	2.39	0.0225
day2*group	2	0.002643	0.000460	36.9	5.74	<.0001
day2*group	3	-0.00058	0.001200	23.2	-0.49	0.6305

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	3	25.5	2540.01	<.0001
day*group	3	46.7	22.41	<.0001
day2*group	3	49.6	12.64	<.0001

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr >  t
estimate 1	1.8237	3.3044	19.4	0.55	0.5873
estimate 2	5.1866	4.1949	27.8	1.24	0.2266
estimate 3	3.3629	3.1944	21	1.05	0.3044
estimate 4	0.1056	0.1256	48.1	0.84	0.4049
estimate 5	-0.3384	0.1649	52.6	-2.05	0.0451
estimate 6	-0.4440	0.1296	30.5	-3.43	0.0018
estimate 7	0.000060	0.001223	47.7	0.05	0.9610
estimate 8	0.003288	0.001651	55.6	1.99	0.0513
estimate 9	0.003228	0.001285	30.5	2.51	0.0175

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
contrast 1	2	24.3	0.80	0.4588
contrast 2	2	45.6	5.82	0.0056
contrast 3	2	46.6	3.18	0.0505