**Solutions:**

1. The longitudinal data from an insulin study contain 36 rabbits where 12 rabbits were randomly assigned to each of 3 groups: group 1 rabbits received the standard insulin mixture, group 2 rabbits received a mixture containing 1% less protamine than the standard, and group 3 rabbits received a mixture containing 5% less pro- tamine. Rabbits were injected with the assigned mixture at time 0, and blood sugar measurements taken on each rabbit at the time of injection (time 0) and 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 hours post-injection.

   The data file *"insulin"* is on the course website. The variables appearing in columns are: (1) rabbit id, (2) insulin group, and (3-9) response (blood sugar level) at 7 time points.

   (a) **(10 points)** Create a spaghetti plot of the data with separate panels for each group. Comment on the heterogeneity in the data.

```
Here we'll do separate plots (panels) for each group;
Proc SGpanel data = insulin;
PanelBy group / columns=3;
series x=hour y=Ins / group =id LineAttrs= (pattern=1);
run;
```
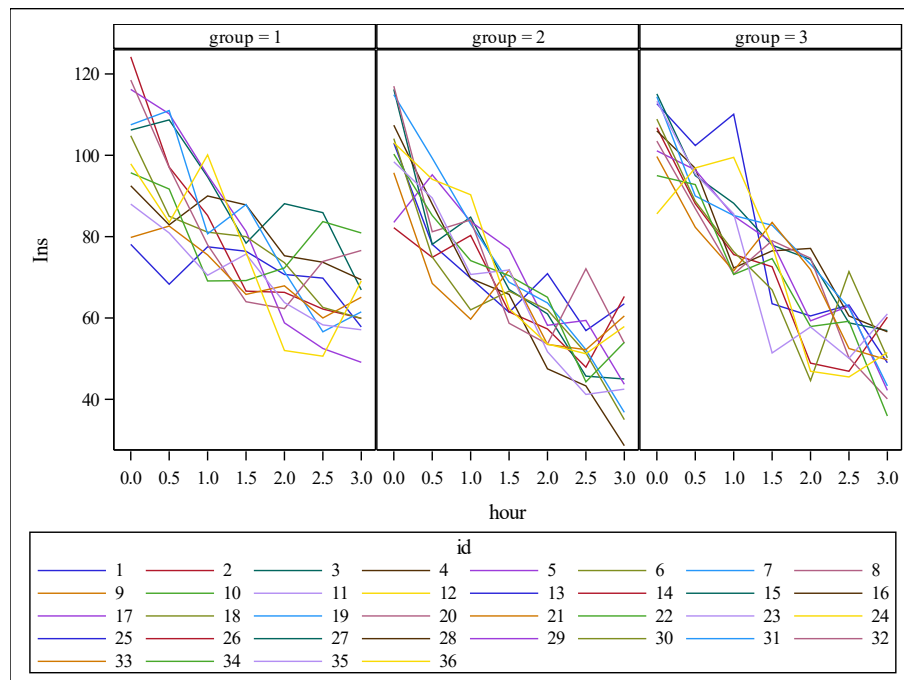

The amount within subject variability in the data appears to be relatively high. The rankings of the insulin values bounce around a lot over the study period. There doesn't appear to be much heterogeneity in the data.

(b) **(10 points)** Create a plot that has the groups means over time on the same plot with different colors (if you'll print/submit in color) or line-types (if you print/submit in black and white). What trends might be appropriate (e.g., profile, linear, quadratic, etc.)?

```
proc sort data=insulin;
by group hour;
run;

*Calculate the mean by group and hour;
proc means mean data=insulin;
by group hour;
var Ins;
output out = MN_GRP_dat mean = mn_GRP_Ins;

proc print data = MN_GRP_dat;
run;


*Plot the mean insulin by hour and group (with color);
Proc SGplot data = MN_GRP_dat;
series x=hour y=mn_GRP_Ins / group =group LineAttrs=(pattern=1 thickness=3);
run;
```
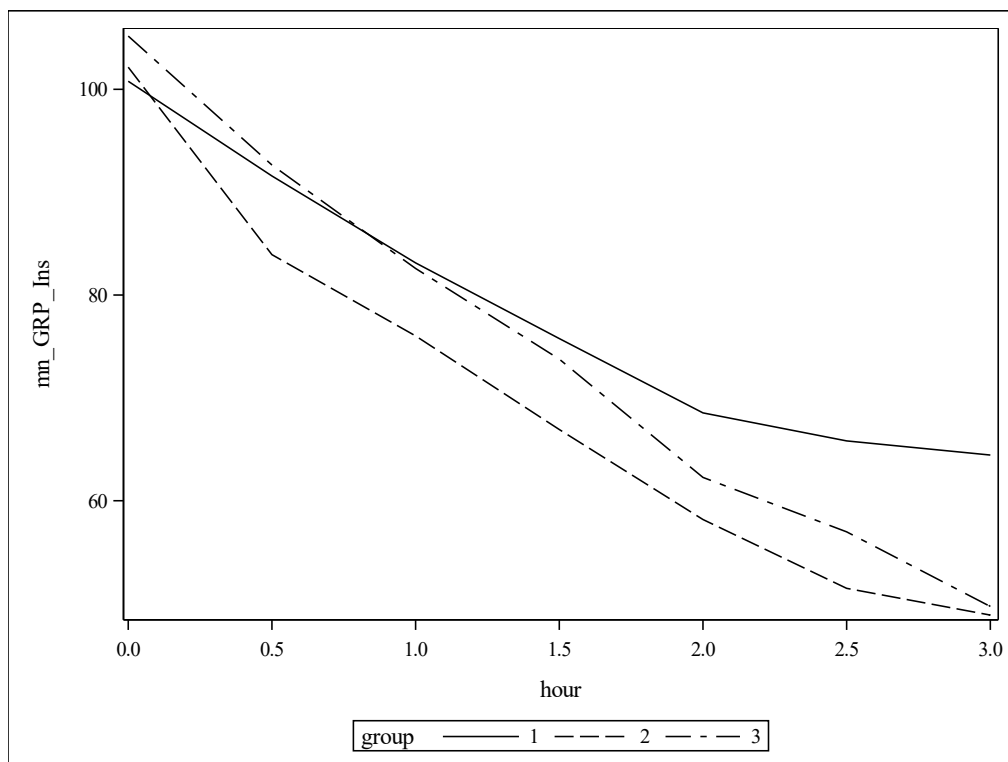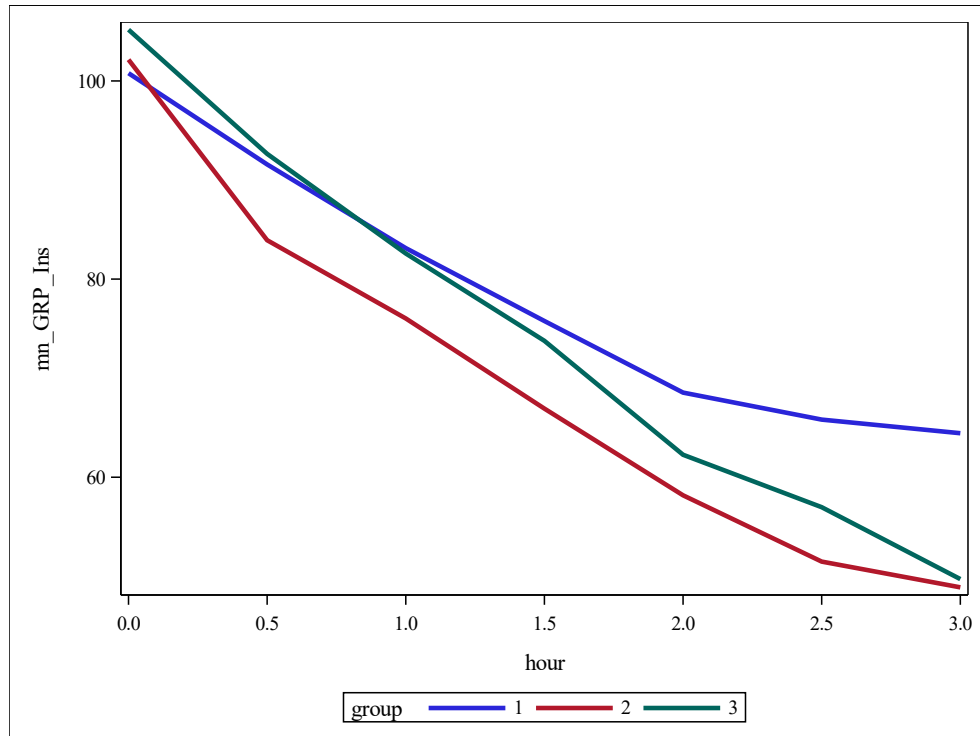


```
*Plot the mean insulin by hour and group (with different line types);
Proc SGplot data = MN_GRP_dat;
series x=hour y=mn_GRP_Ins / group =group LineAttrs=(color=1 thickness=3);
```

```
run;
```


The trends here appear to be linear. Considering the small sample size (n=36) this is likely the best way to model time.

(c) **(10 points)** Fit a full interaction model using a **profile analysis** with the following covariance matrices. Hand in the estimates of the covariance matrices and model fit statistics (AIC and BIC) for each. Nothing else.

    i. Unstructured (heterogeneous and symmetric)

```
proc mixed data=insulin;
class ID group hour;
model Ins = group hour group*hour;
repeated hour/type=UN subject=ID;
run;
```

Estimated R Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	142.10	55.3734	27.3998	-20.4943	8.1449	11.6053	-9.8090
2	55.3734	100.88	49.9615	4.0966	1.8378	3.5719	-16.8321
3	27.3998	49.9615	116.31	-4.6389	-15.3237	-0.1450	3.7044
4	-20.4943	4.0966	-4.6389	59.1381	21.6112	-10.9575	-27.5573
5	8.1449	1.8378	-15.3237	21.6112	88.2290	29.7329	6.4097

Estimated R Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
6	11.6053	3.5719	-0.1450	-10.9575	29.7329	89.1641	23.7737
7	-9.8090	-16.8321	3.7044	-27.5573	6.4097	23.7737	95.7134

Fit Statistics

-2 Res Log Likelihood	1721.6
AIC (Smaller is Better)	1777.6
AICC (Smaller is Better)	1785.7
BIC (Smaller is Better)	1822.0

ii. Compound Symmetry

```
proc mixed data=insulin;
class ID group hour;
model Ins = group hour group*hour;
repeated hour/type=CS subject=ID;
run;
```

Estimated R Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	98.7919	6.7365	6.7365	6.7365	6.7365	6.7365	6.7365
2	6.7365	98.7919	6.7365	6.7365	6.7365	6.7365	6.7365
3	6.7365	6.7365	98.7919	6.7365	6.7365	6.7365	6.7365
4	6.7365	6.7365	6.7365	98.7919	6.7365	6.7365	6.7365
5	6.7365	6.7365	6.7365	6.7365	98.7919	6.7365	6.7365
6	6.7365	6.7365	6.7365	6.7365	6.7365	98.7919	6.7365
7	6.7365	6.7365	6.7365	6.7365	6.7365	6.7365	98.7919

Fit Statistics

-2 Res Log Likelihood	1766.1
AIC (Smaller is Better)	1770.1
AICC (Smaller is Better)	1770.1
BIC (Smaller is Better)	1773.2

iii. Heterogeneous Compound Symmetry.

```
proc mixed data=insulin;
class ID group hour;
model Ins = group hour group*hour;
repeated hour/type=CSH subject=ID;
run;
```

Estimated R Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	140.80	6.9994	7.5786	5.5232	6.5803	6.6238	6.9966
2	6.9994	98.5251	6.3397	4.6203	5.5045	5.5409	5.8528
3	7.5786	6.3397	115.51	5.0026	5.9600	5.9995	6.3371
4	5.5232	4.6203	5.0026	61.3493	4.3436	4.3723	4.6184
5	6.5803	5.5045	5.9600	4.3436	87.0789	5.2091	5.5023
6	6.6238	5.5409	5.9995	4.3723	5.2091	88.2348	5.5387
7	6.9966	5.8528	6.3371	4.6184	5.5023	5.5387	98.4468

Fit Statistics

-2 Res Log Likelihood	1759.7
AIC (Smaller is Better)	1775.7
AICC (Smaller is Better)	1776.3
BIC (Smaller is Better)	1788.4

(d) **(10 points)** Based on the estimated covariance matrices what do you think is best and why?

The diagonal term of the unstructured and CSH covariance matrices do appear to bounce around a good bit. There's no consistent pattern though and the sample size is small, so it may be that the homogeneous is the best. The off-diagonal values of the unstructured have estimates that are close to zero and even negative. This is likely just noise as the sample size is small. To me, this says that the compound symmetric is likely the best.

(e) **(10 points)** For the models that were fit in (c), which model has the best fit according to AIC and BIC?

Compound Symmetric structure works the best among all information criteria.

(f) **(10 points)** Complete a likelihood ratio test between the following structures.

To each of the results coincide with the results from AIC and BIC?

i. Unstructured and Heterogeneous Compound Symmetry

For the Unstructured vs CSH:
D =1759.7 -1721.6 = 38.1
df = 28-8 = 20
A chi-squared with 20 df and alpha of 0.01 = 37.57, which is less than D.
More specifically the p-value = 0.0086

This shows that the unstructured fits significantly better than the CSH. This does not agree with AIC or BIC, which both prefer the CSH model.

(g) **(10 points)** Using the model that fit best from (c), test whether the time profiles of means are different in the groups.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	2	33	12.10	0.0001
time	6	198	121.67	<.0001
group*time	12	198	1.75	0.0587

The interaction type III test is insignificant (though close). As a result, we cannot say that the mean profiles of the groups are different.

(h) **(10 points)** Does time have a significant impact on the response (this *may* require you to fit another model)?

In the model with no interaction and the model with an interaction the "hour" variable is found to be significant. This is an indication that there is a significant time effect.

(i) **(10 points)** Using the model that fit best from (c), test and give an estimate of the difference in the mean response level at 0.5 hour from baseline (0 hour) in group 1.

There are many ways to do this test. Below I ran a model that had time 0 and group 1 as the referent group. The estimate of the difference is -9.21. In that model the significance test for hour=0.5 is a test that the means at 0 and 0.5 hours are the same for group 1.

Another way to complete this would be to use the 'pdiff' option.

Solution for Fixed Effects

Effect	group	time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			100.78	2.8693	33	35.13	<.0001
group	2		1.3667	4.0577	33	0.34	0.7384
group	3		4.3917	4.0577	33	1.08	0.2870
group	1		0	.	.	.	.
time		0.5	-9.2083	3.9170	198	-2.35	0.0197
time		1	-17.6583	3.9170	198	-4.51	<.0001
time		1.5	-25.0167	3.9170	198	-6.39	<.0001
time		2	-32.2417	3.9170	198	-8.23	<.0001
time		2.5	-34.9667	3.9170	198	-8.93	<.0001
time		3	-36.3417	3.9170	198	-9.28	<.0001
time		0	0	.	.	.	.
group*time	2	0.5	-9.0167	5.5394	198	-1.63	0.1052
group*time	2	1	-8.4667	5.5394	198	-1.53	0.1280
group*time	2	1.5	-10.2083	5.5394	198	-1.84	0.0668
group*time	2	2	-11.7333	5.5394	198	-2.12	0.0354
group*time	2	2.5	-15.7000	5.5394	198	-2.83	0.0051
group*time	2	3	-16.9250	5.5394	198	-3.06	0.0026
group*time	2	0	0	.	.	.	.
group*time	3	0.5	-3.3250	5.5394	198	-0.60	0.5490
group*time	3	1	-4.9333	5.5394	198	-0.89	0.3742
group*time	3	1.5	-6.3917	5.5394	198	-1.15	0.2500
group*time	3	2	-10.6667	5.5394	198	-1.93	0.0556
group*time	3	2.5	-13.2333	5.5394	198	-2.39	0.0178
group*time	3	3	-19.1167	5.5394	198	-3.45	0.0007
group*time	3	0	0	.	.	.	.
group*time	1	0.5	0	.	.	.	.
group*time	1	1	0	.	.	.	.

Solution for Fixed Effects

Effect	group	time	Estimate	Standard Error	DF	t Value	Pr > |t|
group*time	1	1.5	0	.	.	.	.
group*time	1	2	0	.	.	.	.
group*time	1	2.5	0	.	.	.	.
group*time	1	3	0	.	.	.	.
group*time	1	0	0	.	.	.	.

(j) **(10 points)** Using the model that fit best from (c), interpret at least two of the parameters in context of the problem. Have one of the parameters you interpret be from an interaction.

Various interpretations accepted.

Among rabbits in group 1, the average blood sugar level at 2 hours was 32.24 ng/mL less than the average blood sugar at baseline (95% CI: -39.85, -24.63).

The difference in average blood sugar between baseline and 3 hours among rabbits in group 2 is 16.92 ng/mL less than the difference in average blood sugar between baseline and 3 hours among rabbits in group 1 (95% CI: 6.06, 27.8).