HW 5 Repeated Measures Analysis

Repeated measures study designs are frequently used in clinical research within Immunology. Our immune system is adaptive and learning all the time. As the body gets introduced to foreign virus or drugs, it sets off a sequence of reactions throughout the body at all levels of the immune system from things we can observe physically from a patient (like rashes) versus small molecular reactions that directly fight off and wipe out a virus.

In this particular study, researchers wanted to understand how a specific gene IFITM3 (in a healthy human) changes over time, when the people under study have volunteered to get injected with a specific Flu strain. The gene is measured on a cohort of 17 healthy individuals who agreed to participate in the study. There are a total of 8 time points (in hrs) over a period of a few days. The first time-point (Hour 0) represents measurements taken on all of the people before the flu was injected into their bodies. The remaining time points are measurements taken post flu being introduced to the body. In addition to data being collected at each time point, each patient was classified at the end of the study as being either "Symptomatic" or "Asymptomatic". Folks who are asymptomatic are not showing any flu like symptoms while folks in the symptomatic group are.

The IFIT3M gene is known to be a code for producing interferon. When interferon levels spike in the body, it is an indicator that the immune response has "kicked in" to fight the foreign intruder. When the immune response is fighting, we tend to see signs of that fighting through inflammation (flu like symptoms).

HW Question Set

1. Generate a plot of the data that allows us to view the data by time and also by symptom status. (The specifics of what plot to use like raw data, means and sds, etc is up to you).

```
proc means data=long;
class characteristics hours;
var IFITM3;
output out=geneout(drop = _type_ _freq_) mean = mean;
run;
```

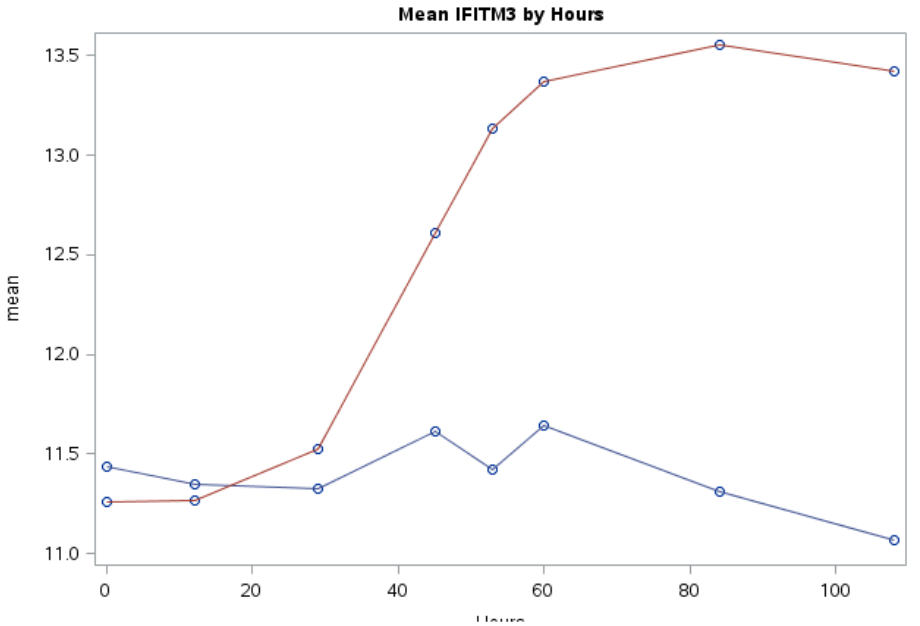
The MEANS Procedure

Analysis Variable : IFITM3

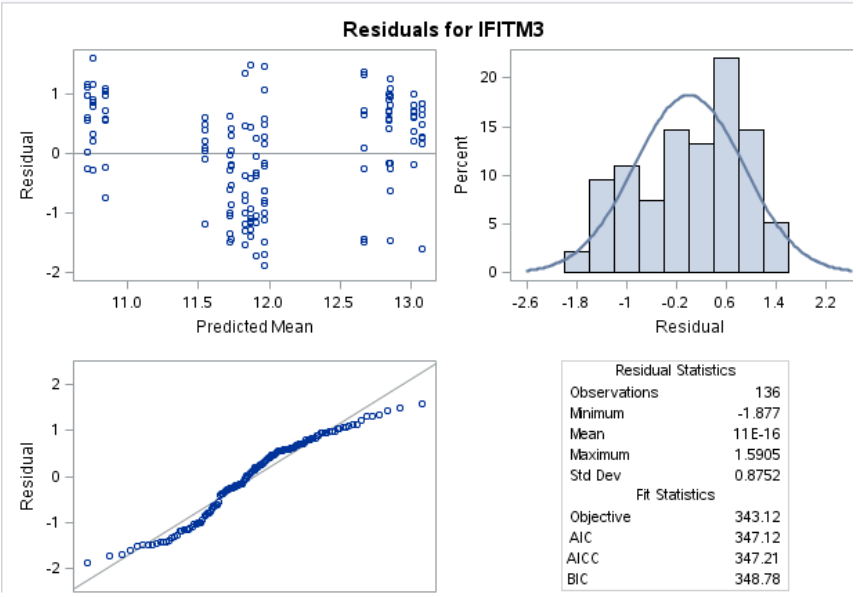
characteristics	Hours	N Obs	N	Mean	Std Dev	Minimum	Maximum
Asymp	0	8	8	11.4393168	0.5883815	10.4586925	12.3416743
	12	8	8	11.3484912	0.5167533	10.4592836	11.8512589
	29	8	8	11.3274088	0.6534070	10.0857001	11.9181688
	45	8	8	11.6122701	0.5608678	10.3552230	12.1512826
	53	8	8	11.4200548	0.6152538	10.2939075	12.1438395
	60	8	8	11.6399499	0.9570206	10.2725805	13.0158189
	84	8	8	11.3123213	0.6542124	10.1881637	12.1551457
	108	8	8	11.0705411	0.7187720	10.2250972	12.3494127
Symp	0	9	9	11.2603336	0.9680464	10.4757826	13.3575769
	12	9	9	11.2647929	0.9214736	10.3097051	13.1611850
	29	9	9	11.5285251	1.0006520	10.0845821	13.4079797
	45	9	9	12.6098627	1.1620157	11.1782123	14.0163112
	53	9	9	13.1306080	0.9379923	11.4005096	14.0866731
	60	9	9	13.3670524	0.7505357	11.4858015	13.9057538
	84	9	9	13.5544287	0.3542050	12.8418123	14.0135252
	108	9	9	13.4220260	0.3332406	12.6711858	13.8359725

```
data geneout;
set geneout;
if Hours = '.' then delete;
if characteristics = '' then delete;
run;

proc sgplot data = geneout;
scatter x = Hours y = mean;
series x = Hours y = mean / group=characteristics;
title2 'Mean IFITM3 by Hours';
run;
```



2. Using the time variable and symptom status variables, generate a repeated measures model using the Compound Symmetry structure for the correlation that exists within people. Provide a plot of the residual diagnostics and comment on the validity of the assumptions.



All the residuals look good.

3. The researchers have two specific questions. Using your knowledge of the ANOVA F-tests as well as specific contrasts to answer the following questions.

   a. Do the changes that potentially exist over time depend on which symptom status you are a part of? Provide the statistical details: test statistic, p-value, and conclusion that yields the answer to this question.

proc glm data=long plots=ALL;
class Subject Hours characteristics;
model IFITM3 = Hours characteristics Hours*characteristics / solution;
lsmeans Hours*characteristics /pdiff tdiff adjust=Tukey;
run;

The GLM Procedure

Dependent Variable: IFITM3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	108.5464403	7.2364294	12.11	<.0001
Error	120	71.6961997	0.5974683		
Corrected Total	135	180.2426400			

R-Square	Coeff Var	Root MSE	IFITM3 Mean
0.602224	6.446864	0.772961	11.98972

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Hours	7	34.25903803	4.89414829	8.19	<.0001
characteristics	1	42.57107433	42.57107433	71.25	<.0001
Hours*characteristic	7	31.71632796	4.53090399	7.58	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Hours	7	30.59331161	4.37047309	7.31	<.0001
characteristics	1	42.57107433	42.57107433	71.25	<.0001
Hours*characteristic	7	31.71632796	4.53090399	7.58	<.0001

In the GLM Type 1 & 3 p-values show that Hours, characteristics, and Hours* characteristic are significant with p-value <.0001.

   b. Are the means for the Asymptomatic group and the Symptomatic group at HR 0 statistically different? At what time point do we start seeing statistically different means between the two groups? Provide the statistical details: test statistic, p-value, and conclusion that yields the answer to this question. (Hint (Contrasts are helpful here, or look up the slicediff option of proc mixed)

proc mixed data=long plots=all;
class Subject Hours characteristics;
model IFITM3 = Hours characteristics Hours*characteristics;
repeated Hours / type=CS subject=Subject;
lsmeans Hours*characteristics / pdiff tdiff adjust=Tukey slice=hours;
run;

Tests of Effect Slices					
Effect	Hours	Num DF	Den DF	F Value	Pr > F
Hours*characteristic	0	1	105	0.23	0.6347
Hours*characteristic	12	1	105	0.05	0.8241
Hours*characteristic	29	1	105	0.29	0.5935
Hours*characteristic	45	1	105	7.05	0.0091
Hours*characteristic	53	1	105	20.74	<.0001
Hours*characteristic	60	1	105	21.14	<.0001
Hours*characteristic	84	1	105	35.64	<.0001
Hours*characteristic	108	1	105	39.20	<.0001

The means are not statistically significant to each other with a p-value of 0.6347.

We start seeing statistically different means at HR 45 with p-value of 0.0091.