Analyses of Repeated Measure Data

INTRODUCTION

Repeated measures data are encountered in a wide variety of disciplines such as medical research, behavioral science, agriculture, or ecology. A repeated measures design is one in which at least one of the factors consists of repeated measurements on the same subjects or experimental units, under different conditions. Therefore, repeated measure experiments can be considered as a type of factorial experiment, with group and time as the two factors. For example, blood pressure was measured on a subject at 1 month, 3 month, 6 month, and 12 month to monitor the benefit of an anti-hypertension medication. The "time" factor is the 1, 3, 6, and 12 months. The "group" is the medication. One of the unique features for the analyses of repeated measure model is that the "within subject" and "between subject" variability should be accounted for in order to correctly interpret the results.

The objectives of repeated measures data analyses are to examine and compare response trends over time. This can involve comparisons of groups at specific time or average over time. It also can involve comparisons of times within a group, which is the "within subject" factor. These are objectives common to any factorial experiment. The important feature of repeated measures experiment that requires special attention in data analyses is the correlation among the responses on the same subject/experiment unit over time.

STATISTICAL METHODS FOR ANALYZING REPEATED MEASURES DATA

Responses measured on the same experiment unit are correlated as they contain a common contribution from the experiment unit. Measures on the same experiment unit closed in time tend to be more highly correlated than measures taken far apart in time, that is, the structure of variance for repeated measures often change with time. Therefore, the covariance matrix for repeated measures is usually complicated as it reflects the correlation over time within an experiment unit.

In repeated measures data analyses, the effects of interest are "between-subject" effect, "within-subject" effect, and the interaction between "within" and "between" subject effect. For the example described above, it is of interest to assess the "between-subject" effect, i.e. the "GROUP"effect, of the anti-hypertension medication among subjects who are on it verse subjects who are not on it. It is also of interest to assess the "within-subject" or "TIME" effect of the anti-hypertension medication for each subject who is on it over time (at 1,3,6, & 12 months). It is also important to assess whether the benefit of medication overtime is consistent, that is, if there is a "GROUP" by "TIME" interaction.

There are several statistical methods for analyzing repeated measures data. These include:

1. Univariate analyses: separate analyses at each time point time;
2. Univariate Analyses of Variance;
3. Mixed model analysis;

The following section demonstrates each of the analysis methods and also to discuss the advantages and shortcomings. We will use the example above and focus on the factors of "time" (month) and "medication" (Yes, No). The analyses will use SAS Enterprise Guide; the PROC GLM and MIXED procedures will be used to carry out the calculation.

The following data points represent the systolic measurements; these data are hypothetical and are used for demonstration only.

Subject	Medication	Month			
		M1	M3	M6	M12
1	Y	101	114	102	102
2	Y	110	102	93	96
3	Y	117	94	110	92
4	Y	100	102	100	101
5	Y	110	111	110	102
6	Y	99	99	100	104
7	Y	103	107	110	102
8	Y	119	103	95	92
9	Y	107	102	100	105
10	Y	101	99	100	102
11	N	100	119	102	119
12	N	112	139	160	149
13	N	110	121	110	119
14	N	115	124	122	126
15	N	140	149	110	139
16	N	112	109	112	111
17	N	132	120	140	119
18	N	103	132	139	121
19	N	111	129	112	111
20	N	111	114	147	144
21	Y	101	114	102	102

Figure 1 depicts the trend for blood pressure over time for subjects who are on medication verse subjects who are not on it.

Figure 1. Blood Pressure over Time.
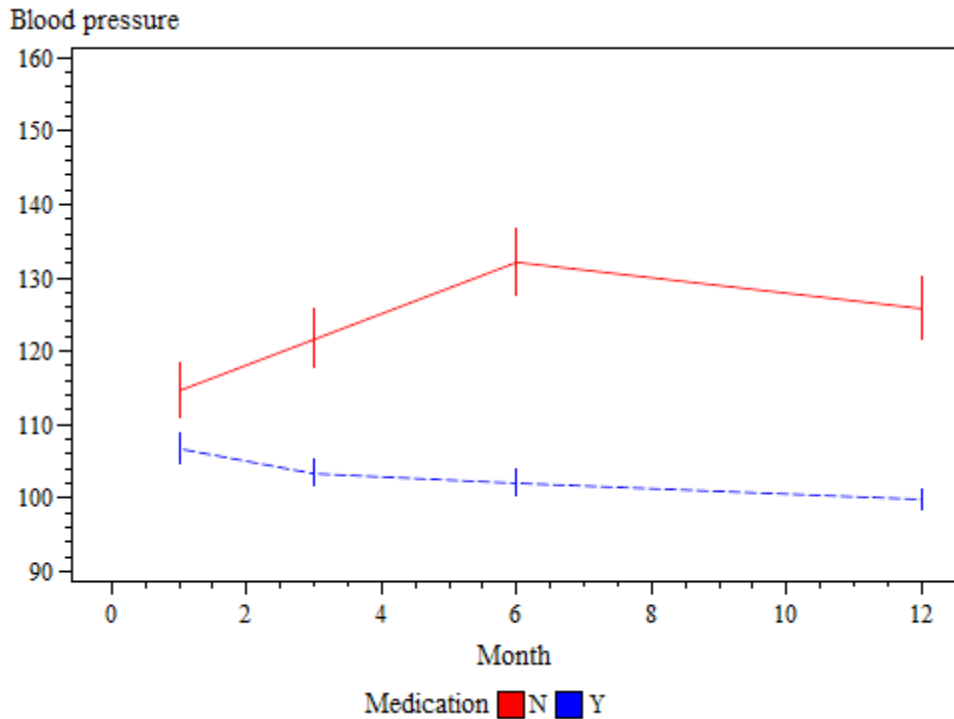
1. Univariate Analyses

Univariate analyses examine the effect of medication at each time point separately. It does not make statistical comparisons among time. The information about the blood pressure over time between groups and the change within subject is not being account for.

The univariate data analyses can be assess using SAS PROC GLM by the following statements:

```
PROC SORT;
CLASS MEDICATION;
PROC GLM;
MODEL M1 M3 M6 M12 =MEDICATION;
LSMEANS MEDICATION /PDIFF;
Run;
```

The analysis results from comparing at each time point are displayed below:

Medication	M1 LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
N	114.600000	0.0958
Y	106.700000	
	M3 LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
N	121.600000	0.0008
Y	103.300000	
Medication	M6 LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
N	132.100000	<.0001
Y	102.000000	
Medication	M12 LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
N	125.800000	<.0001
Y	99.800000	

The p-value comparing the medication effect at each time point indicated that the blood pressure among those subjects that are on medication was not statistically different from those subjects that are not at Month 1; at the other time points medication did help to lower blood pressure.

One caution about this analysis is that the p-value is not adjusted for the multiple comparisons (at each time point) and the type I error is inflated. One way to correct it is to apply multiple comparison correction, for example, Bonferroni correction method, to correct the p-value for testing.

2. Univariate Analyses of Variance

Univariate analysis of variance (ANOVA) treated the data as if they were from a split-plot design with subject as the whole-plot experiment units and time points as the sub-plot units. If the measurements have equal variance over the time, and if the correlation within each subject is the same between paired time points, then ANOVA is a valid method as it assumes a symmetric covariance matrix. The data can be analyzed by SAS PROC GLM procedure:

PROC GLM ;

CLASS SUBJECT MONTH MEDICATION;

MODEL BP1=MEDICATION SUBJECT(MEDICATION) MONTH MONTH*MEDICATION;

RANDOM SUBJECT(MEDICATION);

LSMEANS MEDICATION /STDERR E=SUBJECT(MEDICATION);

LSMEANS MEDICATION*MONTH/PDIFF ;

RUN;

The results from the analyses are displayed below:

Month	Medication	BP1 LSMEAN	LSMEAN Number
1	N	114.600000	1
1	Y	106.700000	2
3	N	121.600000	3
3	Y	103.300000	4
6	N	132.100000	5
6	Y	102.000000	6
12	N	125.800000	7
12	Y	99.800000	8

Least Squares Means for effect Month*Medication
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: BP1

i/j	1	2	3	4	5	6	7	8
1		0.0238	0.0442	0.0016	<.0001	0.0005	0.0017	<.0001
2	0.0238		<.0001	0.3214	<.0001	0.1722	<.0001	0.0472
3	0.0442	<.0001		<.0001	0.0032	<.0001	0.2217	<.0001
4	0.0016	0.3214	<.0001		<.0001	0.7035	<.0001	0.3075
5	<.0001	<.0001	0.0032	<.0001		<.0001	0.0691	<.0001
6	0.0005	0.1722	<.0001	0.7035	<.0001		<.0001	0.5200
7	0.0017	<.0001	0.2217	<.0001	0.0691	<.0001		<.0001
8	<.0001	0.0472	<.0001	0.3075	<.0001	0.5200	<.0001	

The results from this analysis are different at Month 1 compared to the univariate analyses; at Month 3 the medication effect was also more prominent.

3. Mixed model analyses

As noted above，analyses for repeated measures data require special attention to the covariance structure due to the correlation among the measurements within each subject. Methods 1 discussed above ignored the correlation between measurements and Method 2 imposed strong assumption regarding the covariance structure. Ignoring the covariance issues may result in incorrect conclusions and imposing not-sensible assumptions to the covariance matrix is not an efficient analysis. The general mixed model discussed here allows the capability to address regarding the covariance matrix.

There are several covariance matrix choices to fit the data in SAS PROC MIXED procedure. One criterion used to assess the appropriateness of the model is to select a covariance structure with a smallest **Akaike information criterion** (**AIC**).

The data can be analyzed using SAS PROC MIXED procedure:

PROC MIXED ;
CLASS SUBJECT MEDICATION MONTH;
MODEL BP1=MEDICATION MONTH MEDICATION*MONTH;
REPEATED /SUBJECT=SUBJECT TYPE=CSH ;
LSMEANS MEDICATION*MONTH/PDIFF;
RUN;

The heterogeneous compound symmetry covariance structure has a smallest AIC among others that were used and was selected as the final model. The comparisons of the blood pressure at each time point were:

Differences of Least Squares Means									
Effect	Medication	Month	Medication	Month	Estimate	Standard Error	DF	t Value	Pr > |t|
Medication*Month	N	1	Y	1	7.9000	4.8252	54	1.64	0.1074
Medication*Month	N	3	Y	3	18.3000	4.3620	54	4.20	0.0001
Medication*Month	N	6	Y	6	30.1000	4.9125	54	6.13	<.0001
Medication*Month	N	12	Y	12	26.0000	4.5761	54	5.68	<.0001

The blood pressure at Month 1 was not statistically significant, while Months 3, 6, and 12 the medication did lower blood pressure among subjects who are on it.

CONCLUSIONS

For repeated measures data, the analyses that count for the correlation among measurements within a subject are better use the information and provide a more precise estimate. Particularly, for a dataset where sample size is small and several measurements are taken within a subject, the analyses utilizing mixed model will provide valid and efficient analyses.