5-2012

A Methodology for the Analysis of Fly Activity Data.

Ruoying Wang
East Tennessee State University

A Methodology for the Analysis of Fly Activity Data

Submitted in partial fulfillment of Honors

Ruoying Wang

Honors College

Honors in Discipline in Mathematics

East Tennessee State University

March 2012

Advisor(s)

Reader 1: Dr. Edith Seier, Department of Mathematics and Statistics

Reader 2: Dr. Karl H Joplin, Department of Biology

Reader 3: Dr. Xuefeng Liu, Department of Biostatistics and Epidemiology

Abstract:

Experiments to learn about the effect of light, sex, and diet on the activity of flies generate great quantities of data that is necessary to analyze. Since different researches and students participate in the analysis of those experiments, it is convenient to have a methodology to analyze the experimental data using software so that the data can be analyzed in a uniform way. Being a double major in mathematics and biology, I am interested in:

- Deciding which statistical procedure to use to analyze the data so that the research questions of the researchers in biology are answered.
- To recommend how to implement those procedures using software in an efficient way.
- To write a prototype for the interpretation of the results.

Those are the objectives of this work. In the thesis, we first applied two-way ANOVA to analyze the effect of two selected factors, sex (female and male) and diet (liver and non-liver), on the fly activity under dark condition and under light condition, respectively. Next, we employed the repeated measures to capture how fly activity changes over time (day in this case) and to relate the changes to the selected factors, sex and diet, also under dark condition and under light condition, respectively. Finally, we did a little research on the analysis of circadian rhythms and compared the results with that obtained from honey bee activity experiments carried out before.

Keywords:

Fly activity, , ANOVA, Repeated Measures, analysis of circadian rhythm, Minitab, SAS, R.

Table of Contents

I Introduction

1.1 Experiment

1.2 Biological research interests

2 Data Production and Organization

2.1 Data Files

2.2 Software : MINITAB

2.3 From Minitab to SAS and R data files

3 Effects of sex and diet on fly activity

3.1 Two-way ANOVA

3.2 Minitab

3.3 Results

3.4 Interpretation

3.5 SAS

4 Effects of time, sex and diet on fly activity

4.1 Repeated measures

4.2 SAS

4.3 Results

4.4 Discussion

4.5 Profile Plots

5 Analysis of circadian rhythms

Acknowledge

Reference

Appendix

A. Instruction for Minitab

B. Instruction for SAS

C. Instruction for R

Chapter 1 Introduction

1.1 Experiment

Biologists believe that a variety of factors, such as sex, diet, temperature, and light, would affect the activity of flies. Following such interests, Mr. Xinguo Lu, a graduate student in Dr. Joplin's laborary at ETSU, carried out several experiements on flies focusing on fly activity as the response variable and selected factors, sex and diet (liver and non-liver), as predictor variables. They fostered a great quantity of flies in incubators. When all flies were at the age of two days, they began to determine four types of groups, female fed with liver, female fed without liver, male fed with liver, and male fed without liver. There were exactly 224 flies involved, 80 in the female-liver group, 80 in the female-nonliver group, 32 in the male-liver group, and 32 in the male-nonliver group. They put the grouped flies into test tubes on special stands which had been connected with inductors used for sensing and recording the activity for each fly per minute. The device was set up at 8:10:00 am the first day. Every minute, the device recorded the number of times each fly had activated the sensor during that minute and input the data to the computer to save. They kept the device on for the following seven days powering off it at about 8:10:00am on the eighth day, and regulated the light and dark conditions as time from 8:10:00am to 8:10:00pm the same day and from 8:10:00pm to 8:10:00am the next day, respectively. Note that during the experiment period, supplemental food and water should be supplied keeping flies alive and normally activate.

Images below were taken by Mr. Xinguo Lu, showing respectively incubators and sense device mentioned in this experiment.

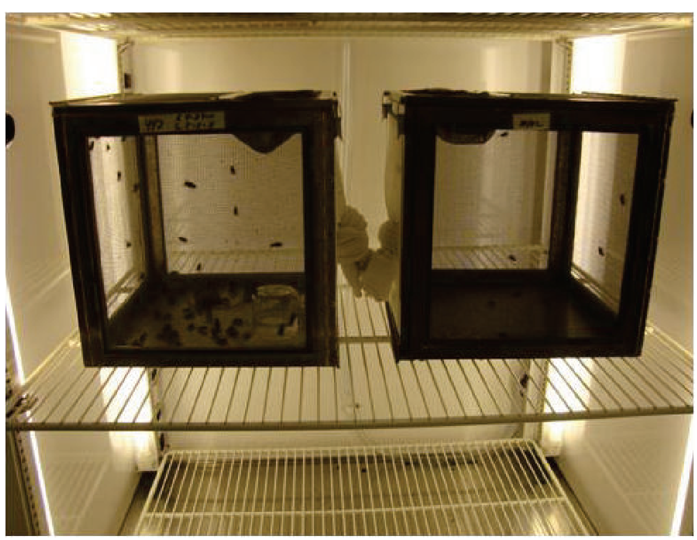


Figure 1 incubator

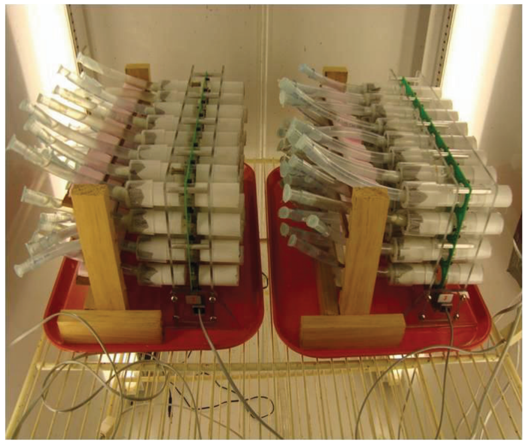


Figure 2 sense device

1.2 Biological research interests

In this work, we mainly focus on three interests.

First, we are interested in the effect of the two factors, diet and sex, on the activity of flies. At the beginning we apply two-way ANOVA on the experimental data to model the response variable, including the difference in activity during the light and dark periods, and also the activity during the light and dark periods separately related to diet and sex factors. The data are analyzed the first time in Mintab. Then, we show the two-way ANOVA on the same data using SAS.Next we add time as a third factor to model the mean response that is the activity in each day during several days acounting for correlation among repeated measures on the same fly. This interest from the biological point of view in observing several days is because flies achieve sexual maturity around the third day after emerging as flies and this fact might impact their behavior. Two problems are found out in this part: how the activity of different flies change related to the selected factors, diet and sex; how the activities of the same fly change over time (day in this case). We employ repeated measures to analyze data using SAS. Also, accounting for the two conditions, the response seperates into two, the activity during the light period and the activity of flies during the dark perior

Finally, we apply the analysis of circadian rhythms to find out whether there is a repeated pattern of activity of approximately 24 hours. Analysis on honey bee activities has indicated a sleep-wake cycle displaying a characteristic 24-hour periodicity. We are wondering if the fly activity follows such a cycle and what the period is. In this part, we use R to apply the analysis of circadian rhythms.

Chapter 2 Data Production and organization

2.1 Data Files

The original data file produced by Mr. Xinguo Lu consists of the registered activity per one-minute interval for each fly.

The original data file looks like this:

C1-D	C2	C3	C4	C5	C6	C7
time	LD	activity per hr bin	ML1-1	ML1-2	ML1-3	ML1-4
8:10:00	1	1	0	0	0	0
8:11:00	1	1	0	0	0	0
8:12:00	1	1	0	0	0	0
8:13:00	1	1	0	0	0	0
8:14:00	1	1	0	0	1	0
8:15:00	1	1	0	0	0	0
8:16:00	1	1	0	0	0	0
8:17:00	1	1	2	0	0	0
8:18:00	1	1	0	2	0	0
8:19:00	1	1	2	0	0	0
8:20:00	1	1	0	0	0	0
8:21:00	1	1	0	0	0	0

Where LD represents light-dark (1 means light and 0 means dark), ML represents male with liver (similarly, MNL means male without liver, FL means female with liver, and FNL means female without liver), ML1-1 means the first male fly fed with liver in the repeat 1, ML1-2 means the second male fly fed with liver in the repeat 2, etc.

Step by step in Minitab (details in 2.2 and appendix A.), we obtain a new data file organized to facilitate the analysis that shows the total activity per fly by hour with identifications of fly, day, age, sex, diet, dark condition or light condition the difference between the light and dark periods and the total activity during the day. Each row of the new data file corresponds to one fly.

The new data file looks like this:

C1	C2	C3-T	C4-T	C5	C6	C7	C8
Day	Age	Sex	Diet	Dark	Light	Diff_L&D	sum
4	6	M	NL	55	172	117	227
4	6	M	NL	*	*	*	*
4	6	M	NL	0	136	136	136
4	6	M	NL	0	154	154	154
4	6	M	NL	*	*	*	*
4	6	M	NL	25	115	90	140
4	6	M	NL	14	184	170	198
4	6	M	NL	0	112	112	112
4	6	M	NL	7	561	554	568
4	6	M	NL	7	168	161	175
4	6	M	NL	47	50	3	97
4	6	M	NL	*	*	*	*

Where symbol * means missing data.

2.2 Software

MINITAB

a) Create a column for the day, and a column for sex and diet. See Appendix A 1-1.
The data file will look like:

Worksheet 1 ***

↓	C1-D	C2	C3-T	C4-T	C5	C6	C7	C8	C9
	time	day	sex	diet	LD	activity per hr bin	MNL1-5	MNL1-6	MNL1-7
1	8:10:00	1	M	NL	1	1	0	5	0
2	8:11:00	1	M	NL	1	1	0	0	0
3	8:12:00	1	M	NL	1	1	0	0	0
4	8:13:00	1	M	NL	1	1	0	0	0
5	8:14:00	1	M	NL	1	1	0	0	0
6	8:15:00	1	M	NL	1	1	0	0	0
7	8:16:00	1	M	NL	1	1	0	0	0
8	8:17:00	1	M	NL	1	1	0	0	0
9	8:18:00	1	M	NL	1	1	0	0	0
10	8:19:00	1	M	NL	1	1	0	0	0
11	8:20:00	1	M	NL	1	1	0	0	0
12	8:21:00	1	M	NL	1	1	0	7	0

b) First we calculate the total activity per day for each fly. See Appendix A 1-2.
The data file will look like:

C39	C40	C41	C42	C43	C44	C45	C46	C47	C48	C49	C50	C51	C52	C53	C5
By day	By LD	Sum1	Sum2	Sum3	Sum4	Sum5	Sum6	Sum7	Sum8	Sum9	Sum10	Sum11	Sum12	Sum13	Sun
1	0	19	219	21	1	0	0	26	0	10	30	54	30	0	
1	1	47	315	6	42	55	110	14	95	21	20	57	2	160	
2	0	13	17	42	2	50	1	0	0	16	34	24	0	0	
2	1	34	914	41	155	108	19	78	152	32	30	33	3	199	
3	0	0	0	0	0	0	0	39	10	12	54	24	81	0	
3	1	134	0	17	54	450	126	78	287	133	67	85	16	371	
4	0	55	0	0	0	0	25	14	0	7	7	47	0	20	
4	1	172	0	136	154	0	115	184	112	561	168	50	13	586	
5	0	7	0	0	19	0	12	0	1	13	54	6	0	0	
5	1	94	0	101	103	0	272	185	300	179	106	130	54	336	
6	0	25	0	162	27	0	7	0	0	3	189	0	2	0	
6	1	212	0	241	88	0	227	143	284	386	115	392	92	180	

c) Then we transpose the sums and label the variables day and LD. See Appendix A1-3.
The data file will look like:

C59-T	C60	C61	C62	C63	C64	C65	C66	C67
Flies	Day1D	Day1L	Day2D	Day2L	Day3D	Day3L	Day4D	Day4L
Sum1	10	62	22	111	0	115	7	201
Sum2	8	183	5	114	54	56	32	162
Sum3	101	84	76	520	100	84	144	155
Sum4	16	32	73	46	64	723	37	219
Sum5	5	41	5	80	17	79	0	144
Sum6	32	25	102	38	0	0	0	0
Sum7	7	61	7	73	9	85	98	262
Sum8	8	33	0	138	0	87	5	76
Sum9	16	75	9	100	25	180	111	315
Sum10	20	87	0	103	0	62	62	1072
Sum11	126	13	2	47	81	73	19	109
Sum12	67	109	37	40	0	111	17	86

d) Next we merge all data into one file and make a good organization for a nice look. See Appendix A 1-4.
The data file will look like:

C1-T	C2-T	C3	C4-T	C5	C6	C7	C8	C9
Sex	Diet	Repeat	Flies	Day1D	Day1L	Day2D	Day2L	Day3D
M	NL	2	Sum43	12	28	66	120	41
M	NL	2	Sum44	43	396	30	425	118
M	NL	2	Sum45	27	610	64	908	13
M	NL	2	Sum46	277	539	0	271	0
M	NL	2	Sum47	0	88	29	134	0
M	NL	2	Sum48	0	96	0	220	0
M	L	1	Sum1	10	62	22	111	0
M	L	1	Sum2	8	183	5	114	54
M	L	1	Sum3	101	84	76	520	100
M	L	1	Sum4	16	32	73	46	64

e) Finally, we split the data day by day. See Appendix A 1-5.
For example, for day one, it will look like:

Day	Age	Sex	Diet	Dark	Light	Diff_L&D
1	3	M	NL	19	47	28
1	3	M	NL	219	315	96
1	3	M	NL	21	6	-15
1	3	M	NL	1	42	41
1	3	M	NL	0	55	55
1	3	M	NL	0	110	110
1	3	M	NL	26	14	-12
1	3	M	NL	0	95	95
1	3	M	NL	10	21	11

2.3 Going from Minitab files to SAS and R files

Once the data files are prepared in Minitab they can be exported into other formats suitable for the analysis either in SAS or R.

2.3.1 Preparing files for SAS

The easiest way is to save the Minitab file into an excel file. Once in SAS, the excel file can be imported and stored in a library. Once in SAS the command 'set nameoflibrary.name of file' will make the data available.

2.3.2 Preparing files for R

In the Minitab menu, choose: File> other files>export special text. Select the columns you want to Export, choose the option 'ANSI date file, and assign a name to the data file. From R, the data file can be read with the command 'read.table'.

Chapter 3 Effects of sex and diet on fly activity

First we will try to see the effect of diet and gender, and the possible interaction between them, on the activity of flies for a given day. For that we will use the two-way ANOVA .

For that we need to prepare the data so that we have one observation per fly for that day, as illustrated in the previous chapter.

3.1 Two-Way ANOVA

In the simplest form, Analysis of Variance (ANOVA) is a statistical test of whether or not the means of several groups are all equal, and it is useful in comparing two, three or more means. Its fundamental technique is a partitioning of the total variation into components related to the effects used in the model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

Where y_{ijk} represents the response, μ is the general mean, α_i and β_j represent effect by the two factors, γ_{ij} represents the effect by the interaction between the two factors, and ε_{ijk} is the error part.

Basically, the total variation splits into two: the between group variation (or explained variability, also denoted with SSModel) and the within group variation (or unexplained variability).

Since there is more than factor, the explained variability or sum of squares SSModel is split in three parts coming from the factors and their possible interaction: SS(A), SS(B) and SS(AB), where A and B are the factors (in the experiments these are sex and diet) . SS(A) , SS (B) and SS(AB) are sum of squares that compare the means of each group determined by the levels of A, B and the combinations AB with the general mean. If the group means are close to each other, the sum of squares will be small.

The between group variances are denoted by MSA, MSB and MSAB for mean square between groups, the between group variation divided by its degrees of freedom.

The within group variation is due to differences within individual groups, denoted SS (W) for sum of squares within groups or SSE. The variance due to the differences within individual groups is denoted MS (W) for mean square within groups, the within group variation divided by its degrees of freedom.

The whole idea behind ANOVA is to compare the ratio of between group variance to within group variance. The F test is used for comparisons of the components of the total deviation. In one way ANOVA (Only one factor considered), the F statistic is found by dividing the between group variance by the within group variance. In the two-way ANOVA, the F statistics are calculated in a similar way (dividing each MS by MSE) when the factors are of fixed effects. A factor is said to have fixed effects when we are interested only on the specific levels that are being considered. In the case of the experiments with flies both factors are of fixed effects, diet (the levels are liver and non-liver) and sex (males and females)

The two-way ANOVA is generally an extension to the one-way ANOVA, where there are two independent categorical variables variables (factors).

The assumptions behind it are the following:

- The populations from which the samples are obtained must be normally or approximately normally distributed.
- The observations in the sample must be independent.
- The variances of the populations must be equal.

There are three sets of hypothesis with the two-way ANOVA.

- The population means of the first factor are equal.
- The population means of the second factor are equal.
- There is no interaction between the two factors.

And for each of the hypotheses, there is an F-test. The F-statistic is calculated shown below (A,B are two random factors.)

F
MS(A) / MS(W)
MS(B) / MS(W)
MS(A*B) / MS(W)

By using software, we can calculate the F-stats and corresponding P-value on which we can make conclusion whether or not to reject each null hypothes.

3.2 Analysis using MINITAB

We first use MINITAB to do two-way ANOVA on the differences between light and dark and on light and dark alone.

Since the data set is unbalanced (not the same number of observations in each group), we use the General Linear Model. See Appendix AII.

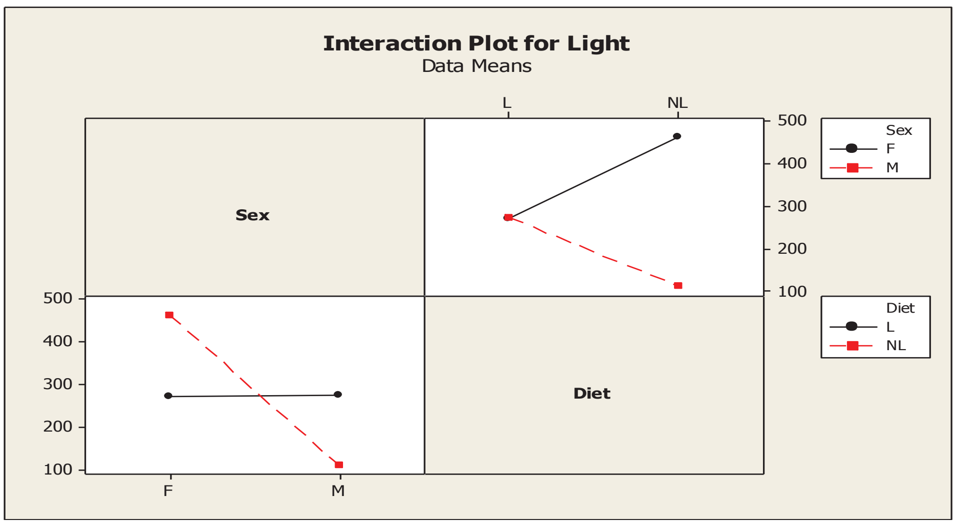
The ANOVA table in the output looks like this:

```
Analysis of Variance for Light, using Adjusted SS for Tests

Source      DF    Seq SS    Adj SS    Adj MS     F      P
Sex          1   1369978   1389252   1389252  8.94  0.003
Diet         1    451466     10157     10157  0.07  0.798
Sex*Diet     1   1439514   1439514   1439514  9.26  0.003
Error      218  33871156  33871156    155372
Total      221  37132115


S = 394.173   R-Sq = 8.78%   R-Sq(adj) = 7.53%
```

For this particular day, during the light period of the day there is interaction between diet and sex (p-value 0.003) meaning that the diet (liver/non-liver) does not equally affect females and males. The interaction plot below shows that females have a higher mean activity if they don't receive liver and the females that do not receive liver have a mean activity similar to that of males in the non-liver group.



3.3 Results.

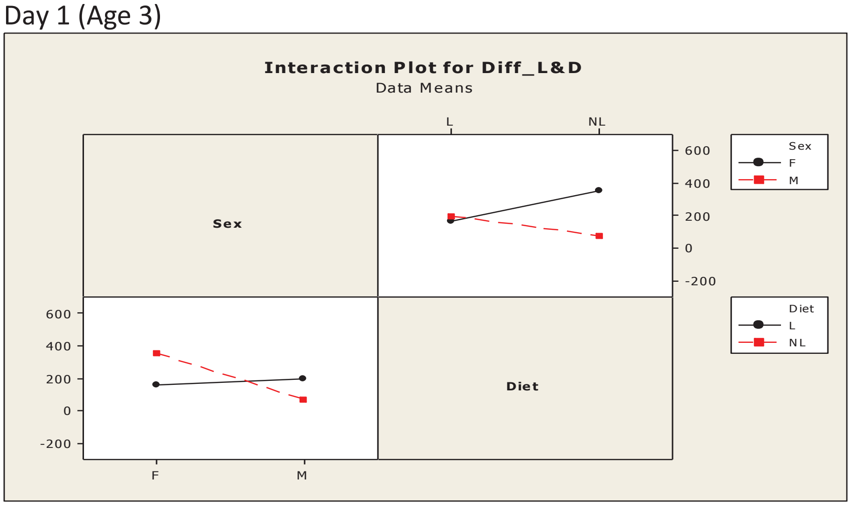
The analysis was performed for each day, first for the difference between light and dark periods, and then separately for the light and dark periods.

Part Ⅰ Difference:

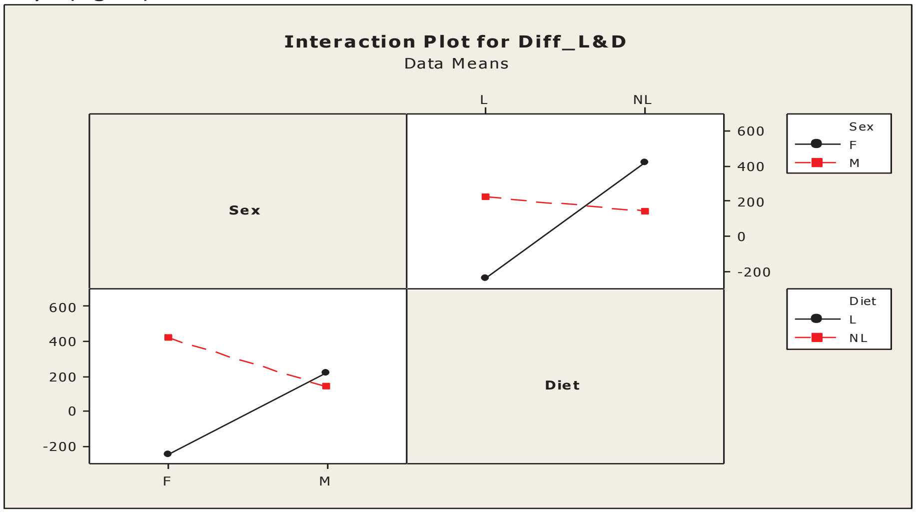
Day	Age	Factors	F Statistics	P-Value	R-squared
1	3	Sex	4.60	0.033	6.73%
		Diet	0.36	0.548	
		Sex*Diet	7.46	0.007	
2	4	Sex	0.93	0.335	16.34%
		Diet	9.29	0.003	
		Sex*Diet	14.85	0.000	
3	5	Sex	0.12	0.734	11.14%
		Diet	5.31	0.022	
		Sex*Diet	10.05	0.002	
4	6	Sex	0.01	0.91	5.36%
		Diet	0.82	0.366	
		Sex*Diet	6.36	0.012	
5	7	Sex	0.19	0.661	1.98%
		Diet	0.30	0.582	
		Sex*Diet	3.79	0.053	
6	8	Sex	0.00	0.950	0.23%
		Diet	0.18	0.674	
		Sex*Diet	0.09	0.761	
7	9	Sex	0.44	0.510	0.52%
		Diet	0.00	0.969	
		Sex*Diet	0.38	0.538	

Interaction Plots (-200-600)

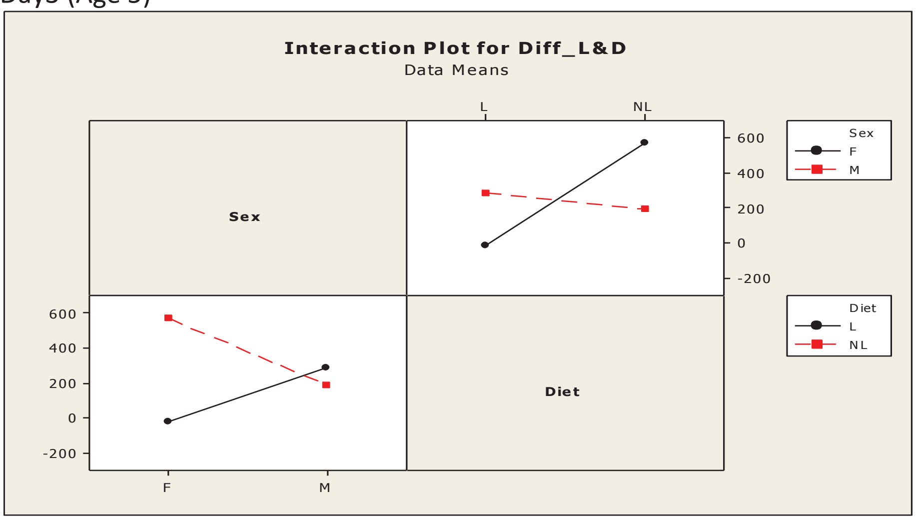
Day 1 (Age 3)



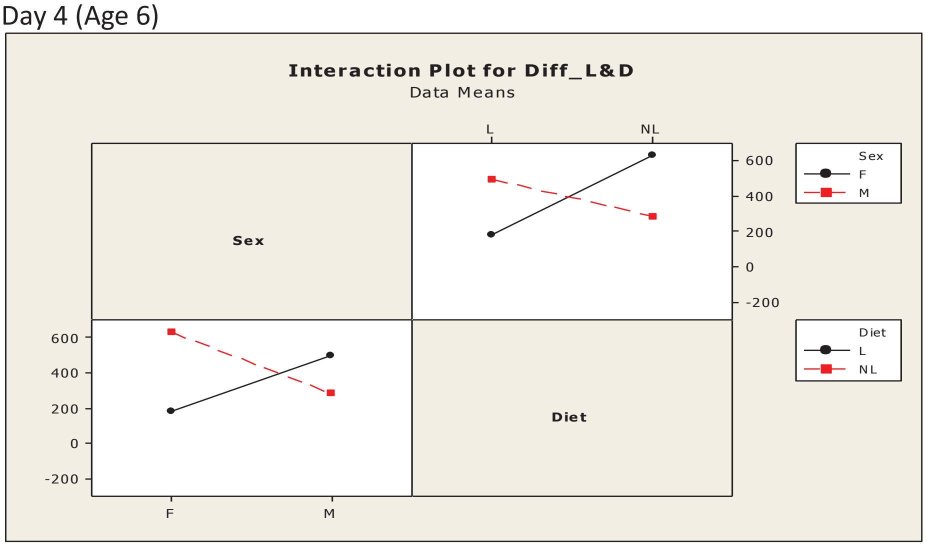
Day2 (Age 4)



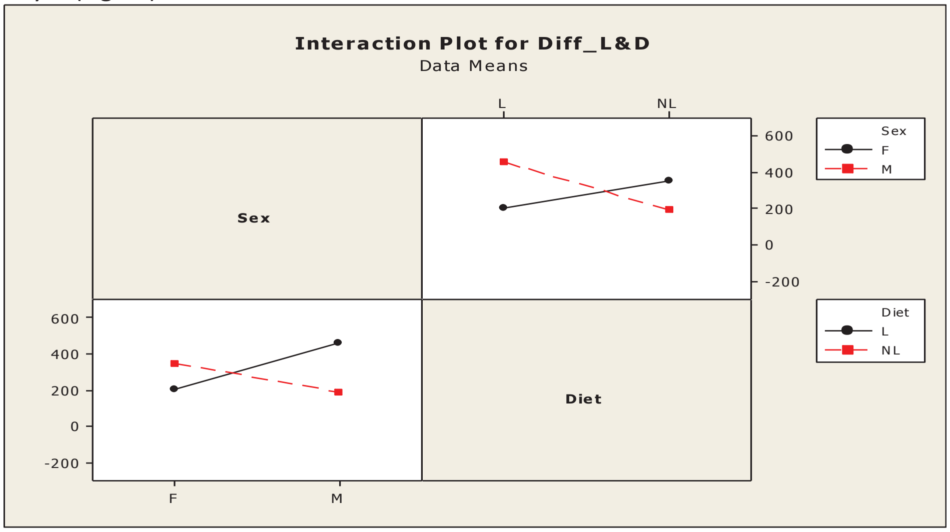
Day3 (Age 5)



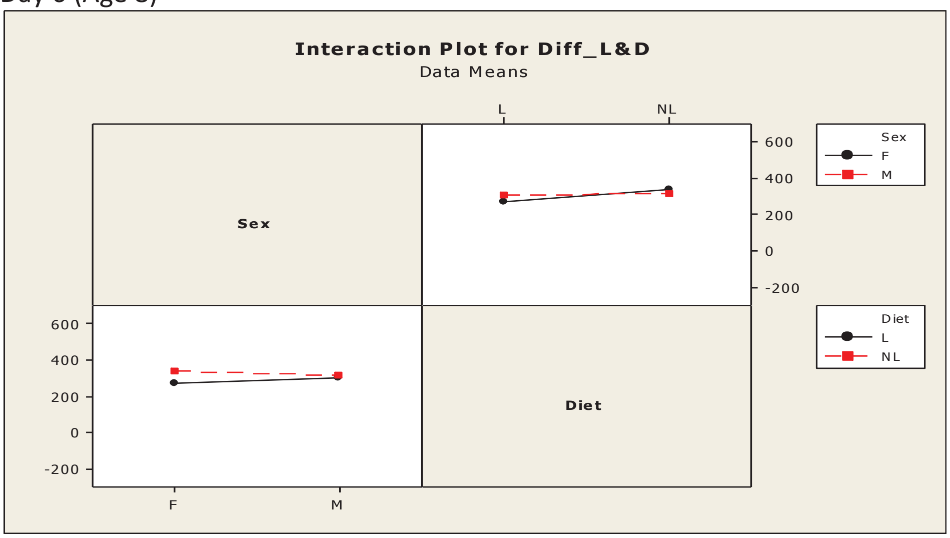
Day 4 (Age 6)



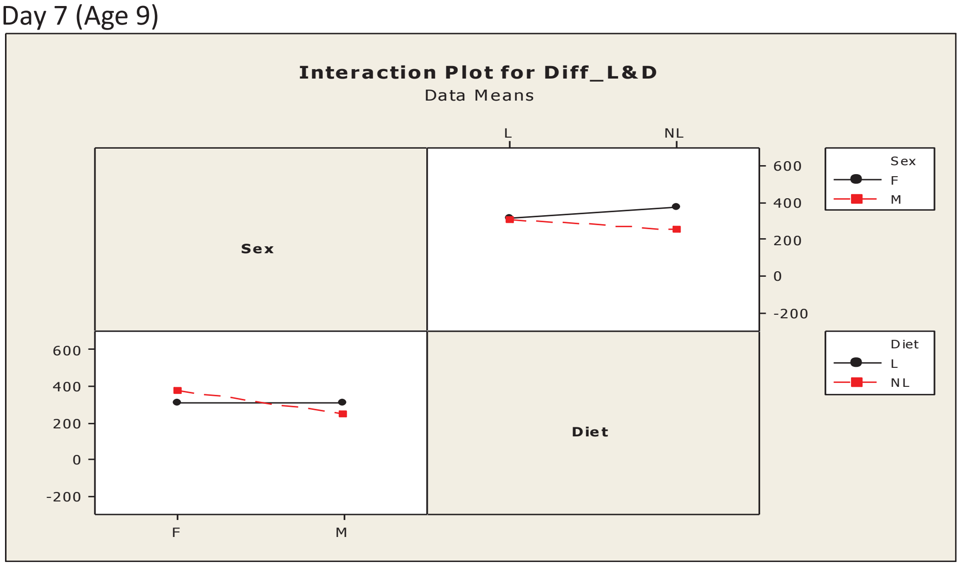
Day 5 (Age 7)



Day 6 (Age 8)



Day 7 (Age 9)

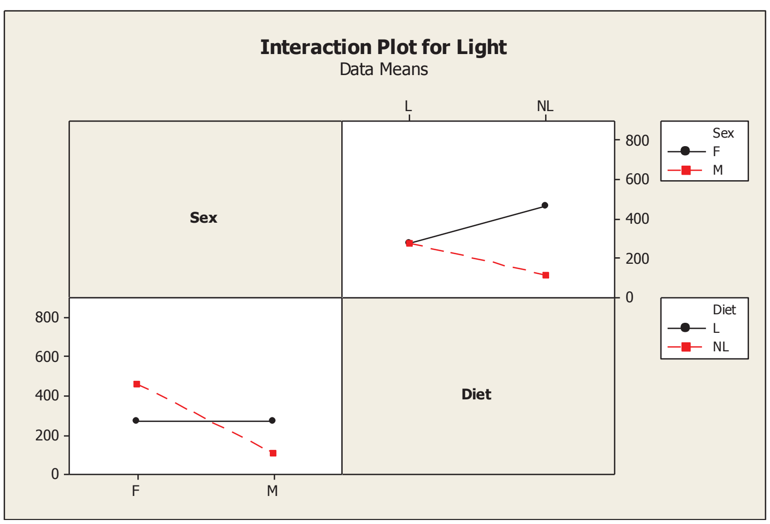


Part II Light:

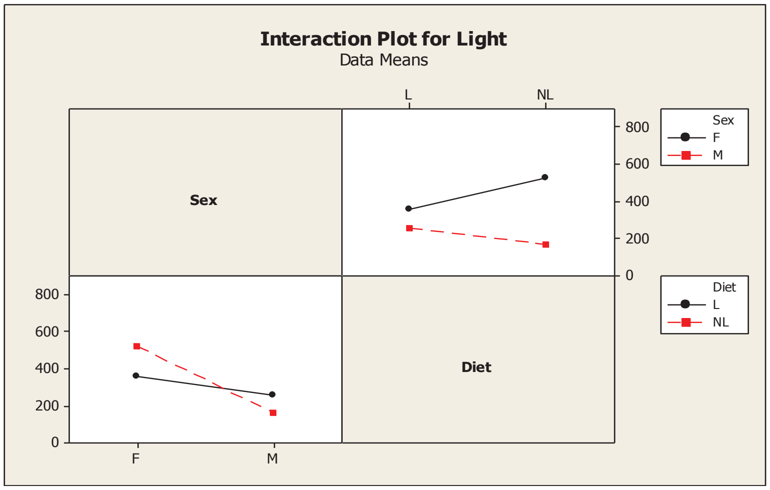
Day	Age	Factors	F Statistics	P-Value	R-squared
1	3	Sex	8.94	0.003	8.78%
		Diet	0.07	0.798	
		Sex*Diet	9.26	0.003	
2	4	Sex	10.50	0.001	6.79%
		Diet	0.29	0.593	
		Sex*Diet	3.29	0.071	
3	5	Sex	16.94	0.000	7.99%
		Diet	0.51	0.476	
		Sex*Diet	0.35	0.554	
4	6	Sex	5.77	0.017	4.19%
		Diet	0.09	0.762	
		Sex*Diet	2.43	0.121	
5	7	Sex	3.45	0.065	3.72%
		Diet	0.92	0.337	
		Sex*Diet	4.03	0.046	
6	8	Sex	5.23	0.023	4.66%
		Diet	1.22	0.270	
		Sex*Diet	0.85	0.358	
7	9	Sex	9.50	0.002	10.27%
		Diet	1.97	0.163	
		Sex*Diet	4.48	0.036	

Interaction Plots (0-900)

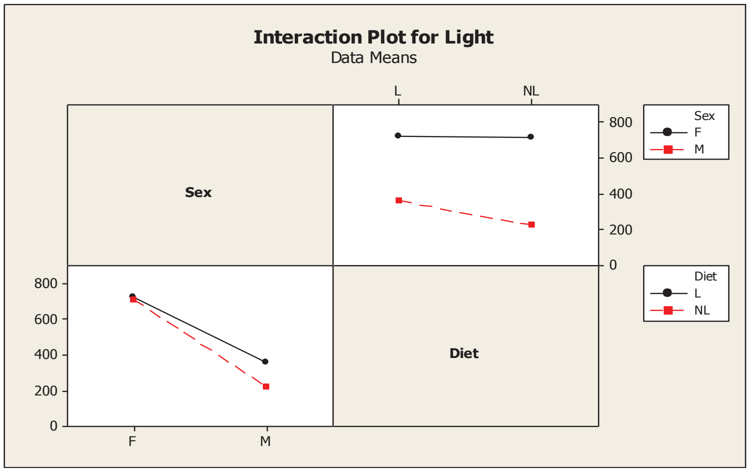
Day1 (Age 3)



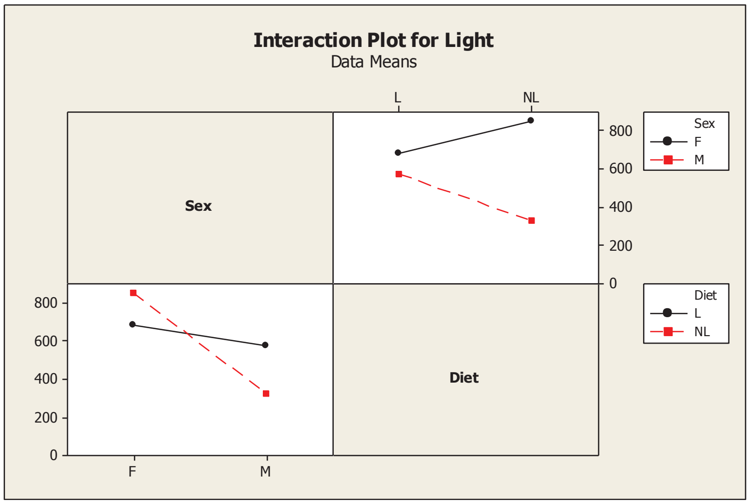
Day2 (Age 4)



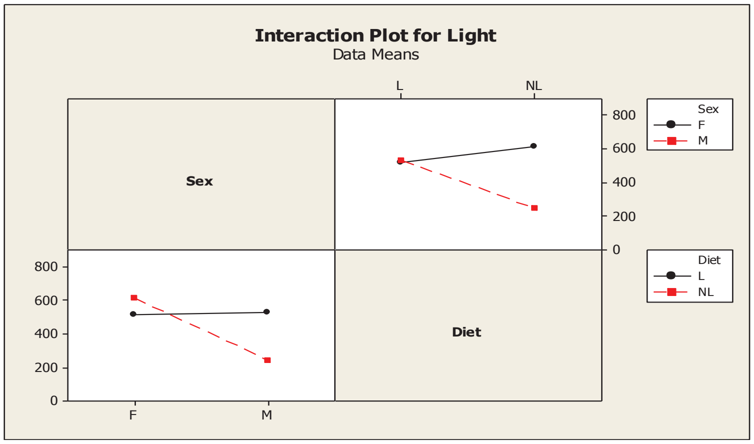
Day3 (Age 5)



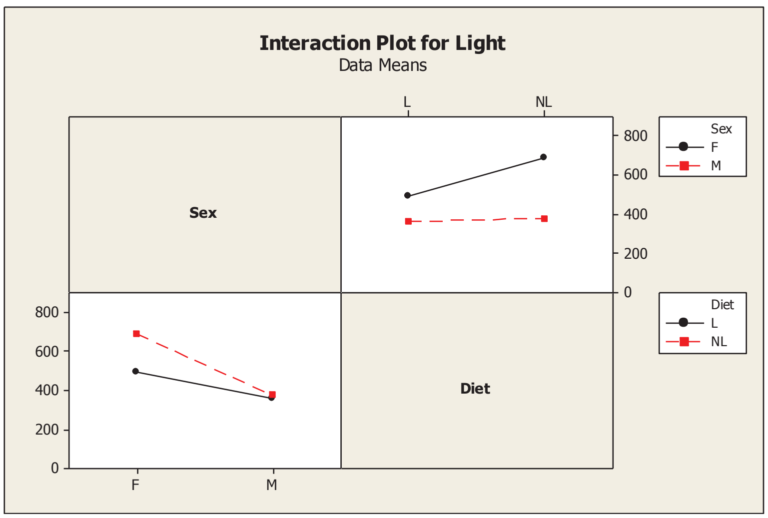
Day4 (Age 6)



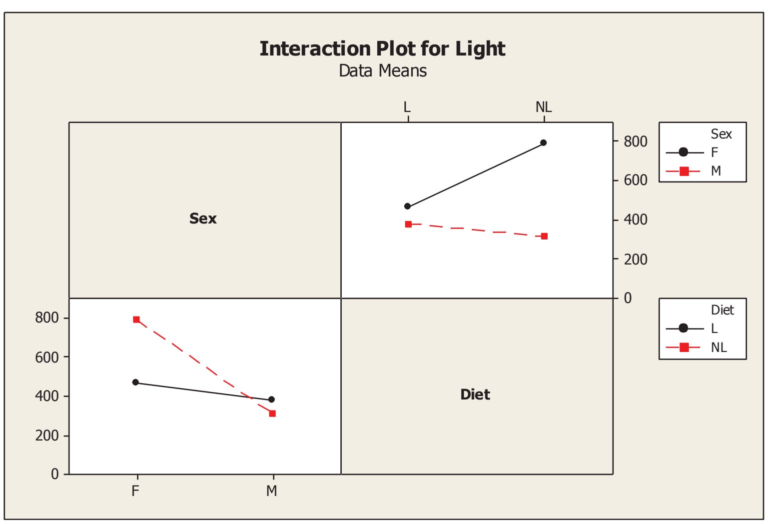
Day5 (Age 7)



Day6 (Age 8)



Day7 (Age 9)

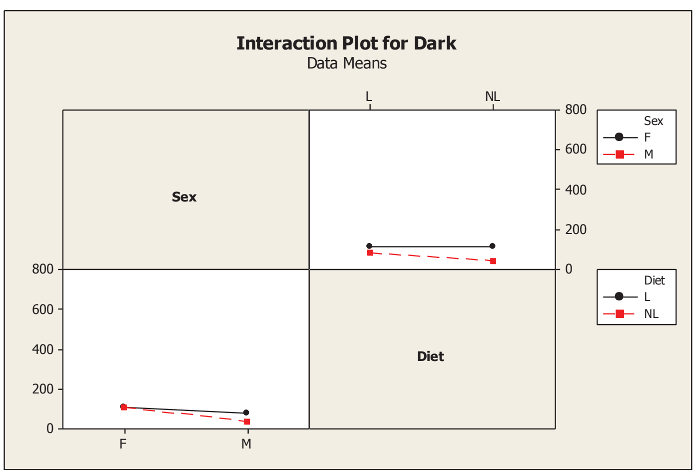


Part III Dark:

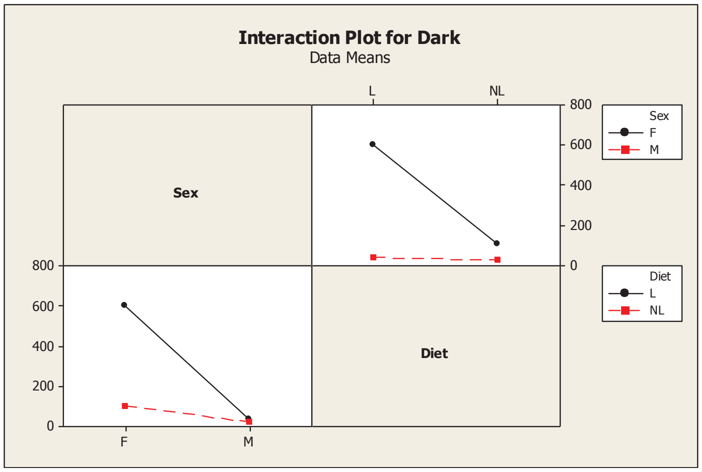
Day	Age	Factors	F Statistics	P-Value	R-squared
1	3	Sex	2.73	0.100	1.50%
		Diet	0.43	0.513	
		Sex*Diet	0.41	0.524	
2	4	Sex	14.44	0.000	17.04%
		Diet	9.07	0.003	
		Sex*Diet	8.18	0.005	
3	5	Sex	18.07	0.000	20.62%
		Diet	12.25	0001	
		Sex*Diet	9.23	0003	
4	6	Sex	10.15	0.002	8.12%
		Diet	2.79	0.096	
		Sex*Diet	1.70	0.193	
5	7	Sex	9.02	0.003	4.44%
		Diet	0.20	0.657	
		Sex*Diet	0.04	0.839	
6	8	Sex	9.48	0.002	6.27%
		Diet	0.80	0.372	
		Sex*Diet	0.65	0.421	
7	9	Sex	6.99	0.009	8.00%
		Diet	2.29	0.132	
		Sex*Diet	2.63	0.107	

Interaction Plots (0-800)

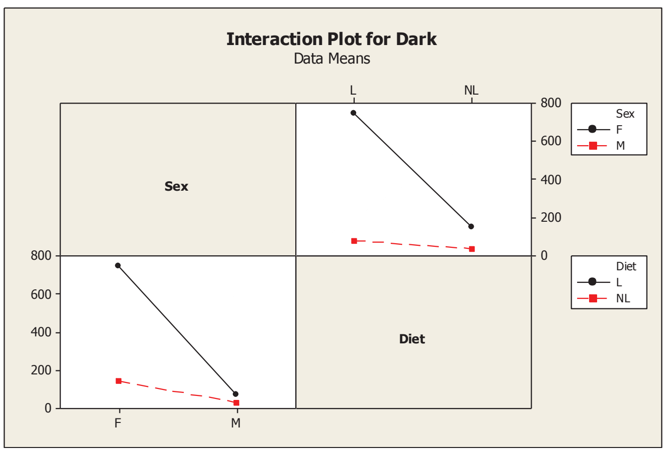
Day1 (Age 3)



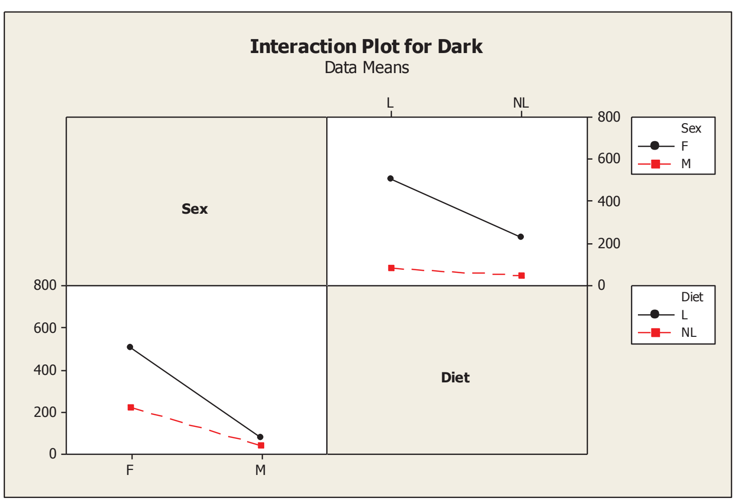
Day2 (Age4)



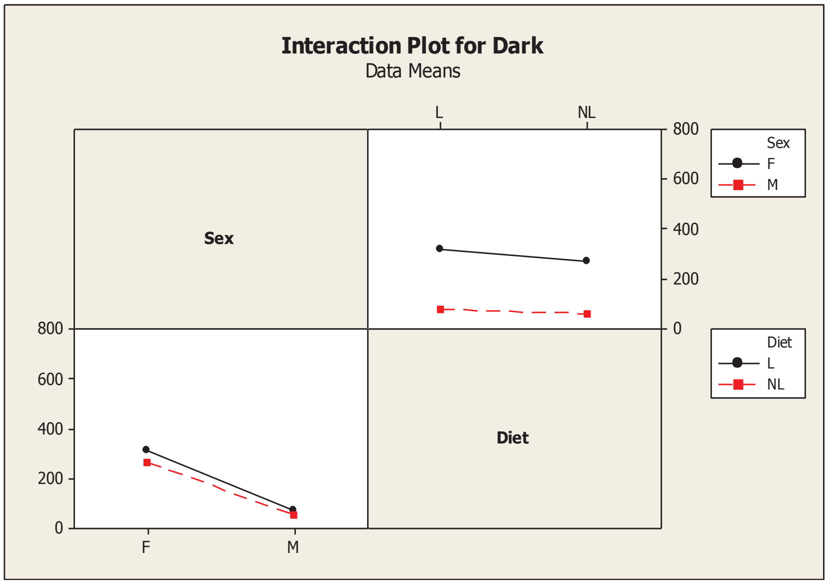
Day3 (Age5)



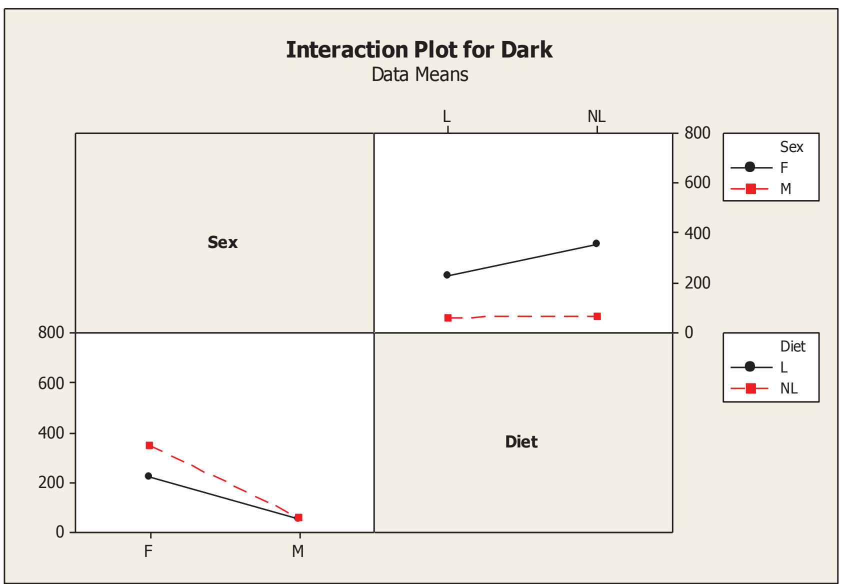
Day4 (Age 6)



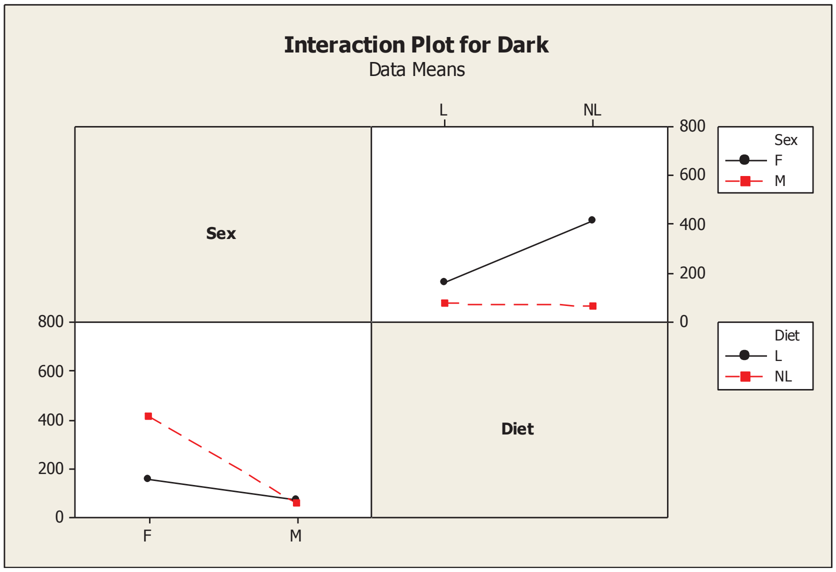
Day5 (Age 7)



Day6 (Age 8)



Day7 (Age 9)



3.4 Interpretation of the results.

Part I Difference

By checking the table as well as the interaction plots, interaction between sex and diet is significant for activity difference between light and dark for each day. The patterns are about similar for the first five days (Age 3 to Age 7). In the first five days, the interaction is very significant for activity difference. Clearly the female with non-liver gets more activity difference than the female with liver, while the male with liver behaves more activity difference. In the last two days, the interaction is not so significant. Also, after the first day, the scales for female get smaller and smaller, while the scales for male get larger first, and then get smaller in the last two days.

Part II Light

By checking the table as well as the interaction plots, interaction between sex and diet is not very significant for light activities on each day except the first day. Generally, the activity of female flies is larger than that of male flies under the light. Patters of plots are about similar for days after the first day. For females, those with non- liver are more active than those fed with liver. And the difference between the activities of female without liver and of female with liver gets larger and larger. This indicates that more liver needed as female flies grow. For male flies, those with liver are more active than those without liver. The reason might be that male flies with liver are more powerful than those without liver.

Part III Dark

By checking the table as well as the interaction plots, interaction between sex and diet is not significant for activities of flies under dark condition. Generally, the sum of activities of female flies is larger than that of male flies. For male flies, basically, there is no difference of difference between male without liver and male with liver because males don't need liver for growth and they don't move during night. For females, those with liver are more active than those without liver in the first several days, however, especially in the last two days, those without liver are obviously more active. This performance verifies that female flies need liver to grow so that they try to find liver and increase the activity even during night.

3.5 SAS

We do both ANOVA and General Linear Model using SAS, and compare the results. See Appendix B I.

The outputs are given below:

A.

The ANOVA Procedure

Dependent Variable: Light Light

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2916722.89	972240.96	2.55	0.0568
Error	198	75427755.13	380948.26		
Corrected Total	201	78344478.02			

R-Square	Coeff Var	Root MSE	Light Mean
0.037229	119.1298	617.2101	518.0990

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Sex	1	1372308.136	1372308.136	3.60	0.0592
Diet	1	429.385	429.385	0.00	0.9733
Sex*Diet	1	1543985.367	1543985.367	4.05	0.046

B.

The GLM Procedure

Dependent Variable: Light Light

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2916722.89	972240.96	2.55	0.0568
Error	198	75427755.13	380948.26		
Corrected Total	201	78344478.02			

R-Square	Coeff Var	Root MSE	Light Mean
0.037229	119.1298	617.2101	518.0990

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sex	1	1313398.598	1313398.598	3.45	0.0648
Diet	1	352142.359	352142.359	0.92	0.3375
Sex*Diet	1	1535675.553	1535675.553	4.03	0.046

Note the two outputs are different between ANOVA and GLM, for unbalanced data. Why? Which one should we use in this case?

In SAS, proc anova is only used when there are an equal number of observations in each of the ANOV cells (balanced design). proc glm is a much more general procedure that will work with any balanced or unbalanced design (unbalanced meaning an unequal number of observations in each cell). Therefore, in this case, proc anova is not appropriate and it would give inaccurate outputs. We should use glm.

Chapter 4 Effects of time, sex and diet on fly activity

In this chapter, we added time as one of selected factors. The repeated measures method is a good way to capture how activity of flies changes over time and to relate the changes to selected factors (diet, and sex in this case).

4.1 Repeated Measures

A repeated measures study refers to data sets with multiple measurements of a response variable on the same experimental units. Basically, there are two factors, treatment(s) and time. Treatment is called the between-individual factor because levels of treatment can change only between individuals. Time is called within-individual factor because different measurements on the same individual are at different times. In repeated measures experiments, interest concerns on 1) between-individual changes in the response and 2) within individual changes in the response over time. Actually, the assessment of within-individual changes in the response over time can only be achieved within a repeated measures study design. For example, in a cross-sectional study, where the response is measured at a single time, one can only obtain estimates of between-individual differences in the response while not obtain measures of how individuals change over time. On the other hand, a repeated measures study that measures an individual at different times could capture how individuals change over time. As mentioned, there are two problems of interest in this experiment on flies: how the activities of different flies change related to the factors, diet and sex; how the activities of the same fly change over time. Therefore, in this case, the repeated measures method is a good choice to solve the problems.

In the experiment described in chapter 1, day is the within-individual factor, both sex and diet is the between-individual factors, and activities of flies are the response variable. Note that in this case, we cannot simply treat activities of one fly each day as the response because each day the same fly activity is measured twice actually, under the light condition and under dark condition. Thus, the response separates into two responses, the light activity and the dark activity, due to the independent assumption behind the repeated measures.

The model is in the ANOVA context:

$$y_{ijmk} = \mu + \alpha_i + \beta_j + \rho_m + \theta_{ij} + \pi_{im} + \sigma_{jm} + \varepsilon_{ijmk}$$

, Where y_{ijmk} is the response, μ is the general mean, $\alpha_i, \beta_j and \rho_m$ respectively represent the effects of three factors, $\theta_{ij}, \pi_{im} and\ \sigma_{jm}$ respectively represent the effects of interactions between two factors, and ε_{ijmk} is the error part.

4.2 SAS

We use SAS to apply the repeated measures on the experimental data. In SAS, there are three general types of repeated measures analysis. One method is called univariate analysis of variance, using PROC GLM with the RANDOM statement. Another method applies multivariate and univariate analysis methods to linear transformation of the repeated measures. It uses PROC GLM with the REPEATED statement. The third one applies methods based on the mixed model with special parametric structure on the mixed model with special parametric structure on the covariance matrices. It is applied in PROC MIXED, typically using the REPEATED statement. Note that both PROC GLM and PROC MIXED test within individual variability for repeated measures analysis of variance. PROC GLM is basically a fixed-effects procedure that can handle class and continuous variables.

In this work, we focus on mixed measurements, one of repeated measures, which use PROC MIXED procedure in SAS. Although we typically use PROC MIXED with the repeated statement, we use the

RANDOM statement sometimes. In this statement, it assumes that there is no relation between observations over time, and its structure is called compound symmetric structure. However, this structure is probably not realistic because two observations close in time are likely to be more highly correlated than two observations far apart in time. The REPEATED statement not only models variation between subjects as the RANDOM statement does, but also is used to model the covariance structure within subjects. Basically, PROC MIXED with the REPEATED statement provides three covariance types:

- Compound symmetric structure (CS in SAS), is also named covariance pattern models in which we assume the between-individual variance and the within-individual variance are confounded.
- Autoregressive order one structure (AR (1) in SAS), is also named random effect covariance structures in which the between-individual variance and the within-individual variance are not confounded.
- Unstructured model (UN in SAS), is also called unstructured covariance in which no mathematical pattern is imposed on the covariance matrix.

Note that the choice of the three covariance structures would affect the result for which reason we need to decide which to assume in the model for final inference. The decision process can be assisted by using two model-fit criteria computed by PROC MIXED, Akaike's Information Criterion (AIC in SAS) and Schwarz' Bayesian Criterion (SBC in SAS). It appears that the covariance structure with values of the criteria closest to zero is considered most desirable.

4.3 Results of the repeated measures analysis

We use the three types of structures in mixed measures (compound symmetric, autoregressive order one, and unstructured), run in SAS and make the comparison. Recall that in this experiment, we determine time, diet, and sex as predictor variable, and that the response separates into two: light activity and dark activity.

The procedures are shown in Appendix B-2

4.3.1 Compound symmetric structure

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	20.54	20.54	<.0001	<.0001
diet	1	236	0.94	0.94	0.3328	0.3338
sex*diet	1	236	13.34	13.34	0.0003	0.0003
day	6	1416	35.81	5.97	<.0001	<.0001
sex*day	6	1416	6.44	1.07	0.3756	0.3762
diet*day	6	1416	6.89	1.15	0.3315	0.3322
sex*diet*day	6	1416	3.99	0.66	0.6781	0.67

(a) Tests for light–fly-activity

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	41.61	41.61	<.0001	<.0001
diet	1	236	3.66	3.66	0.0557	0.0569
sex*diet	1	236	2.13	2.13	0.1442	0.1456
day	6	1416	11.34	1.89	0.0786	0.0794
sex*day	6	1416	14.94	2.49	0.0207	0.0212
diet*day	6	1416	38.17	6.36	<.0001	<.0001
sex*diet*day	6	1416	32.89	5.48	<.0001	<.0001

(b) Tests for dark-fly-activity

4.3.2 Autoregressive order one structure

Type 3 Tests of Fixed Effects for light

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	26.08	26.08	<.0001	<.0001
diet	1	236	1.19	1.19	0.2751	0.2762
sex*diet	1	236	16.94	16.94	<.0001	<.0001
day	6	1416	26.54	4.42	0.0002	0.0002
sex*day	6	1416	5.32	0.89	0.5040	0.5043
diet*day	6	1416	5.39	0.90	0.4945	0.4948
sex*diet*day	6	1416	4.07	0.68	0.6677	0.667

(a) [Tests for light–fly-activity

Type 3 Tests of Fixed Effects for DARK

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	40.90	40.90	<.0001	<.0001
diet	1	236	3.60	3.60	0.0578	0.0591
sex*diet	1	236	2.10	2.10	0.1478	0.1491
day	6	1416	11.25	1.87	0.0810	0.0819
sex*day	6	1416	15.41	2.57	0.0173	0.0177
diet*day	6	1416	27.96	4.66	<.0001	0.0001
sex*diet*day	6	1416	25.61	4.27	0.0003	0.0003

(b) Tests for dark-fly-activity

4.3.3 Unstructure structure

Type 3 Tests of Fixed Effects FOR LIGHT

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	20.54	20.54	<.0001	<.0001
diet	1	236	0.94	0.94	0.3328	0.3338
sex*diet	1	236	13.34	13.34	0.0003	0.0003
day	6	236	31.77	5.29	<.0001	<.0001
sex*day	6	236	4.76	0.79	0.5744	0.5754
diet*day	6	236	7.06	1.18	0.3157	0.3198
sex*diet*day	6	236	4.38	0.73	0.6249	0.6253

(a) Tests for light–fly-activity

Type 3 Tests of Fixed Effects FOR DARK

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
sex	1	236	41.61	41.61	<.0001	<.0001
diet	1	236	3.66	3.66	0.0557	0.0569
sex*diet	1	236	2.13	2.13	0.1442	0.1456
day	6	236	17.32	2.89	0.0082	0.0098
sex*day	6	236	21.17	3.53	0.0017	0.0023
diet*day	6	236	20.29	3.38	0.0025	0.0032
sex*diet*day	6	236	19.04	3.17	0.0041	0.0052

(b) Tests for dark-fly-activity

4.3.4 Covariance Structure Comparison

dark

Fit Statistics

-2 Res Log Likelihood	25131.3
AIC (smaller is better)	25135.3
AICC (smaller is better)	25135.3
BIC (smaller is better)	25142.3

Light

Fit Statistics

-2 Res Log Likelihood	25759.7
AIC (smaller is better)	25763.7
AICC (smaller is better)	25763.7
BIC (smaller is better)	25770.7

(a) Statistics for compound symmetric structure

dark

Fit Statistics

-2 Res Log Likelihood	24996.6
AIC (smaller is better)	25000.6
AICC (smaller is better)	25000.6
BIC (smaller is better)	25007.5

Light

Fit Statistics

-2 Res Log Likelihood	25721.9
AIC (smaller is better)	25725.9
AICC (smaller is better)	25725.9
BIC (smaller is better)	25732.9

(b) Statistics for autoregressive order one structure

dark

Fit Statistics

-2 Res Log Likelihood	24661.0
AIC (smaller is better)	24717.0
AICC (smaller is better)	24718.0
BIC (smaller is better)	24814.4

light

Fit Statistics

-2 Res Log Likelihood	25523.9
AIC (smaller is better)	25579.9
AICC (smaller is better)	25580.9
BIC (smaller is better)	25677.4

(c) Statistics for unstructured structure

4.4 Discussion of the repeated measures analysis

4.4.1 Compound symmetric structure

First, we check the test table for light-fly-activity. Since Sex-P-value is .0001 as shown in the test table, the factor 'sex' has significant effect on the between-individual changes in the response. Note that Diet-P-value is .3838; the effect of diet is not significant on the between-individual changes in the response. And since the Day-P-value is less than .0001, we know that the within-individual changes are significant over time. At last, the P-value of sex*diet*day is .67, which means that the effect of interaction between sex and diet on the response has no difference on different days. Similarly, we check the test table for dark-fly-activity, which comes out a result. The factor 'sex' still has significant effect on the response, and factor 'diet' makes no difference on the response. However, the interaction between sex and diet has significant effect on the response over day in these days.

4.4.2 Autoregressive order one structure

Also, we check the test for light-fly-activity first and then the test for dark-fly-activity. Note that all statistics involved are very close to those obtained in compound symmetric structure. Thus, the conclusions are the same with those made in compound symmetric structure.

4.4.3 Unstructure structure

Again, the statistics are almost the same with those obtained in the two structures above, and also the conclusion.

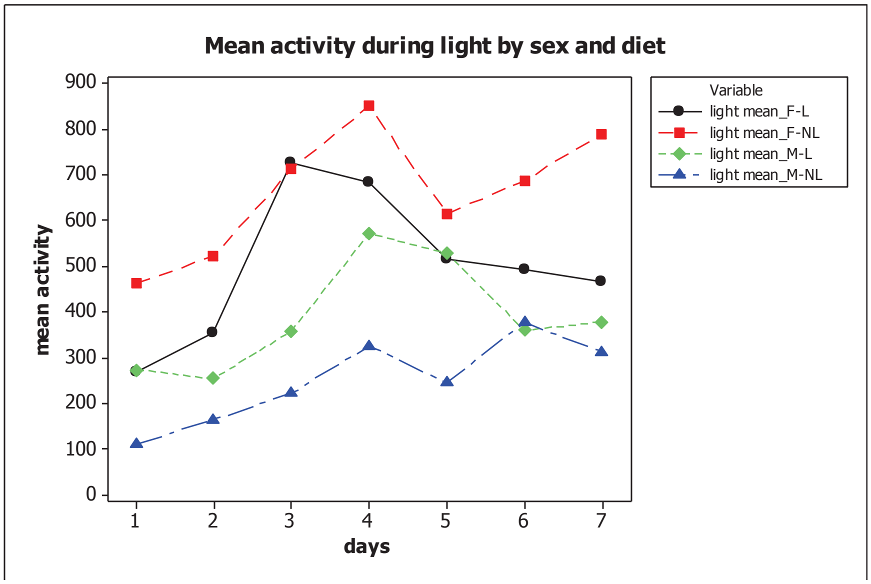
4.4.4 Comparison of structures

Although the statistics obtained in the three structures are similar, the unstructured covariance is better than the others for the smaller values (closer to zero) in terms of the light-fly-activity response, while the autoregressive order one structure is the better one referring to the dark-fly-activity response.

4.5 Profile Plots

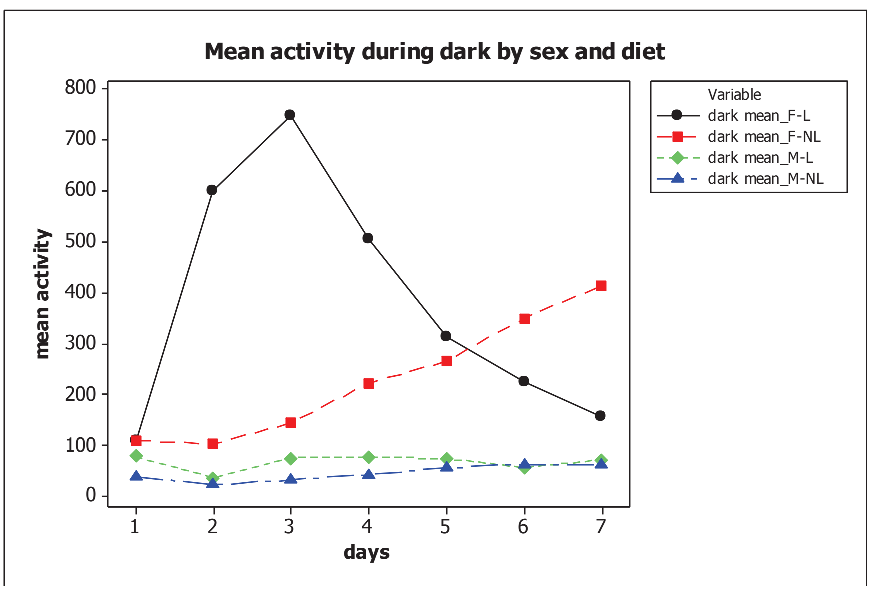
The following plots display the mean activity for all flies in each group throughout all the days that the experiment lasted. The profile plots are prepared with Minitab by calculating the means by groups and using scatterplots with connecting lines. And it is a good summary for the whole story.

4.5.1 Mean activity during light



From the plot, the activity trend of each group goes up at the starting and then down. In average, the most active day is the fourth day for each group and female fly is more active than male fly. And during the light, Female fly fed without liver is most active, which might because they need move to find food.

4.5.2 Mean activity during dark



From the plot, during dark, similarly the female fly is more active than the male fly in general. The male flies act as the same way over the seven days no matter if they are fed with liver or not. For female group with liver, the activity trend is to go up at first and then down. Also, obviously, female group with liver is the most active, which might because they have more energy for their movement in dark.

Chapter 5 Analysis of circadian rhythms

The sleep-week cycle, with its characteristic intervals of activity alternating with restfulness that recur with a periodicity approximating the 24-hour day-night cycle, is the prototypical example of a behavior that demonstrates a circadian rhythm. Based on analysis of circadian rhythms applied on records of behaviors of honey bees, they follow 24 hours circadian rhythms. In this case, we are interested in the most important frequency for flies and then find out how important the circadian rhythm is for this fly. Circadian rythms are studied from experiments that keep the organisms in constant darkness or free-run experiements, external cues of light and dark are absent.

We use R to do the analysis of circadian rhythms.

a. Create "objects" for particular flies and read it.
   1. First, open up 'FNL Activity' in origindata file. It look like this:

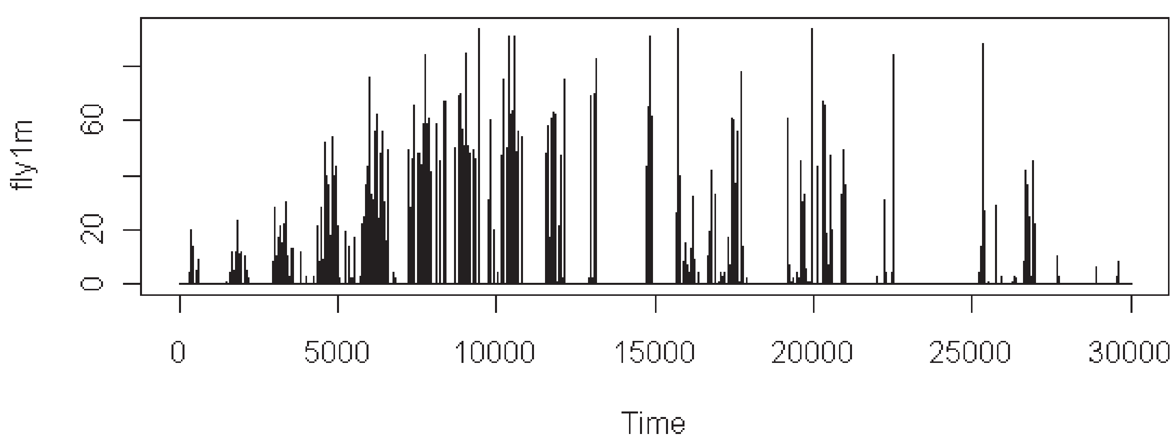
time	monitor	LD	activity per bin	FNL1-1	FNL1-2	FNL1-3	FI
8:09:00	1	1	1	0	0	0	
8:10:00	1	1	1	1	0	0	
8:11:00	1	1	1	3	8	0	
8:12:00	1	1	1	0	0	0	
8:13:00	1	1	1	1	3	0	
8:14:00	1	1	1	0	0	0	
8:15:00	1	1	1	0	0	0	
8:16:00	1	1	1	4	0	0	
8:17:00	1	1	1	0	0	0	
8:18:00	1	1	1	0	0	0	
8:19:00	1	1	1	0	0	0	
8:20:00	1	1	1	0	0	0	

   2. Select FNL1-1 to FNL1-8 and create "objects" in a new file which could be read in R. See Appendix C.
      The data file looks like this:

0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	3
3	8	0	0	0	3	0	0
0	0	0	0	0	2	0	0
1	3	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	6
0	0	0	0	0	0	0	0
0	0	0	0	0	0	4	0
0	0	0	0	0	0	8	0
0	0	0	0	0	0	1	0
1	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

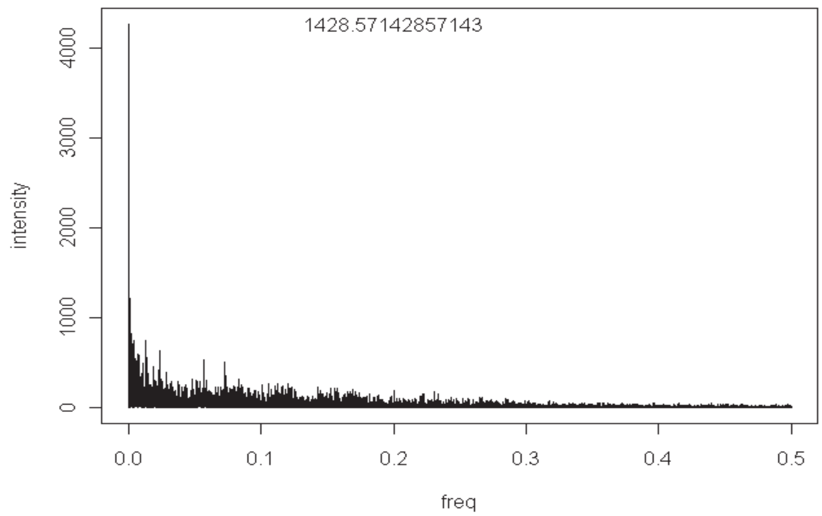
Each column represents one fly.

3. Read in R. See Appendix C.
4. Create an 'object' for each fly, and we do the time series plot and if the fly died before the end. (you can cut the data accordingly) See Appendix C.
   We choose Fly1-1 for example.
   Its time series plot is shown below.



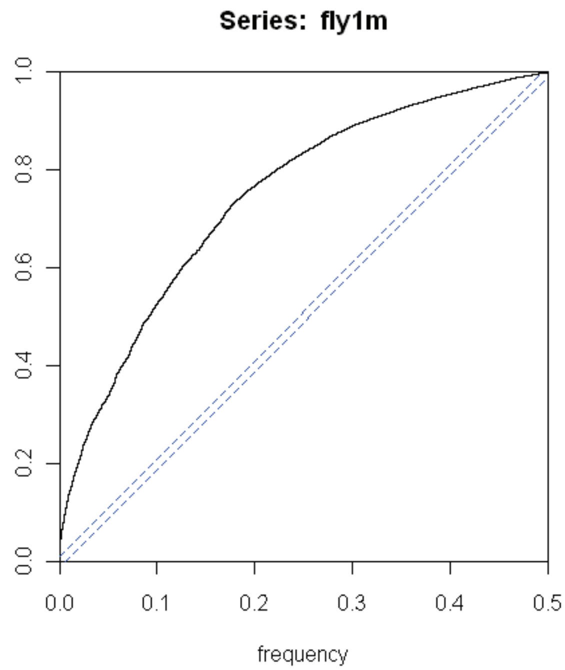
Note that the fly dies after 30,000.

5. Next do the Periodogram to find out the most frequency for the particular fly. Since the perioplot function which produces the plot of the periodogram, we need put the program in R. See Appendix C.
   The periodogram plot is shown below.



The most important frequency is a very low one, the period is 1438. Since the data are produced by minute, we can divide 1438 by 60 to have the period in hours 1438/60= 23.96667 ~ 24 hrs.

6. Now let's see how important the circadian rhythm is for this fly. Program see Appendix C. The significant plot is shown below.



Almost only 4% of the total variability is attributed to the circadian rhythm and cycles longer than one day, including any possible trend.

7. At last, calculate the periodogram in numerical version. Also, the perionum function for calculation is not a plug in function in R, we need put it in. See Appendix in R. The output is shown below.

	frequency	intensity	period	cumulative%
[11,]	3.333333e-04	3.539956e+02	3000.000	1.452763e+00
[12,]	3.666667e-04	4.189092e+02	2727.273	1.543684e+00
[13,]	4.000000e-04	1.073459e+03	2500.000	1.776669e+00
[14,]	4.333333e-04	6.012089e+02	2307.692	1.907157e+00
[15,]	4.666667e-04	1.829375e+01	2142.857	1.911128e+00
[16,]	5.000000e-04	5.817232e+01	2000.000	1.923753e+00
[17,]	5.333333e-04	9.250819e+01	1875.000	1.943832e+00
[18,]	5.666667e-04	1.509914e+02	1764.706	1.976603e+00
[19,]	6.000000e-04	6.140529e+01	1666.667	1.989931e+00
[20,]	6.333333e-04	3.670101e+02	1578.947	2.069587e+00
[21,]	6.666667e-04	2.269770e+03	1500.000	2.562223e+00
[22,]	7.000000e-04	4.266218e+03	1428.571	3.488171e+00
[23,]	7.333333e-04	2.304749e+03	1363.636	3.988398e+00
[24,]	7.666667e-04	2.586659e+01	1304.348	3.994013e+00
[25,]	8.000000e-04	4.188875e+02	1250.000	4.084929e+00

Acknowledge

I especially appreciate Dr. Seier, who is thesis supervisor, and Dr. Xuefeng Liu and Dr. Karl H Joplin, who are the other readers of the thesis.

I am also thankful to Mr. Xinguo Lu, a graduate student in the department of biology. He conducted the experiment, provided the original experimental data, and gave some interpretations to the results at a biological angle.

References

Crawley, Michael J., (2005), *Statistics: an introduction using R*, England: John Wiley & Sons Ltd.

Fitzmaurice, G.M., Laird, N.M., Ware, J.H. (2011), *Applied Longitudinal Analysis*, New York: Wiley.

Little, R.C. , Milliken, G.A., Stroup, W.W. and Wolfinger, R.D. (1997), *SAS System for Mixed Models*, Cary NC: The SAS Institute Inc.

Marasinghe, Mervyn G., Kennedy, Willian J. (2008), *SAS for Data Analysis,* Ames, IA: Springer Science and Business Media.

Wei, William W.S., (2004), *Time Seies Analysis: Univariate and Multivariate Methods*, NY: Pearson Addison Wesley.

Wonnacott, Thomas H., Wonnacott, Ronald J. (2003), *Introductory Statistics*, NY: John Wiley and Sons.

Yarandi, H. N. (2003), "Comparison of PROC MIXED and PROC GLM for Analysis of Repeated Measures Data", *SESUG Conference proceedings*

http://analytics.ncsu.edu/sesug/2003/SD06-Yarandi.pdf

Appendix

Appendix A--- MINITAB PART

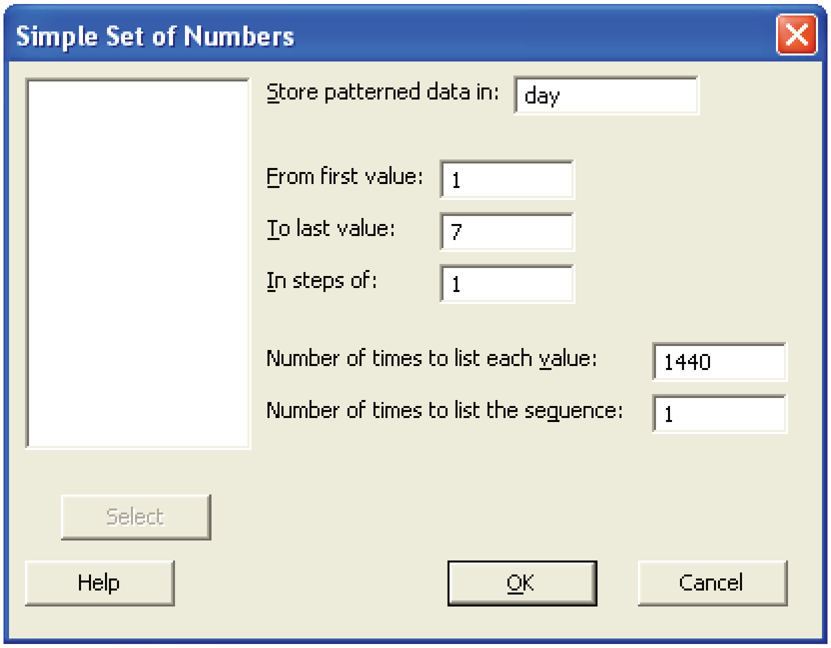
I Data preparation and organization

The original Data file (check origindata folder in thesis data folder):

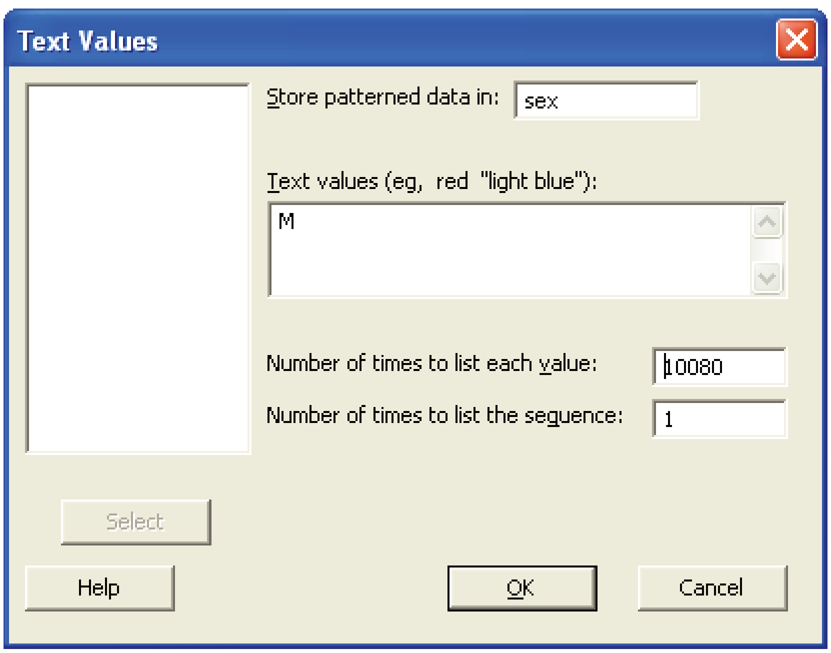
C1-D	C2	C3	C4	C5	C6	C7
time	LD	activity per hr bin	ML1-1	ML1-2	ML1-3	ML1-4
8:10:00	1	1	0	0	0	0
8:11:00	1	1	0	0	0	0
8:12:00	1	1	0	0	0	0
8:13:00	1	1	0	0	0	0
8:14:00	1	1	0	0	1	0
8:15:00	1	1	0	0	0	0
8:16:00	1	1	0	0	0	0
8:17:00	1	1	2	0	0	0
8:18:00	1	1	0	2	0	0
8:19:00	1	1	2	0	0	0
8:20:00	1	1	0	0	0	0
8:21:00	1	1	0	0	0	0

1. Preparation of the data
    f) Create a column for the day, and a column for sex and diet. In Minitab, we use EDITOR> insert columns, and create 3 columns.
    g) Use CALC>Make Patterned Data to create the values for each one of these variables

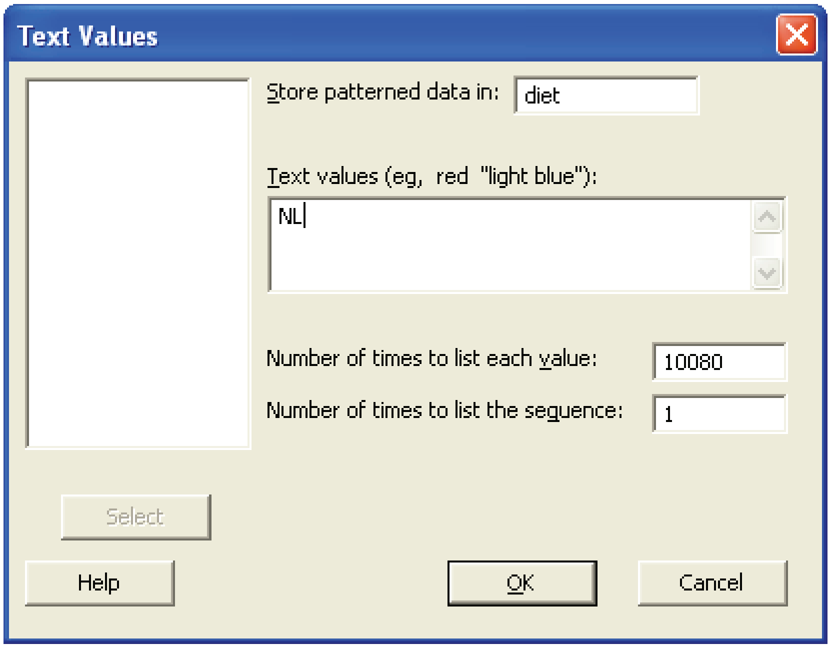
For day, there are 1440 observations per day, divide the total number of observations per 1440 to know how many days you have information for. In the example there were 7 days.



In the column 'sex' write the appropriate letter for your data file



Same for the diet



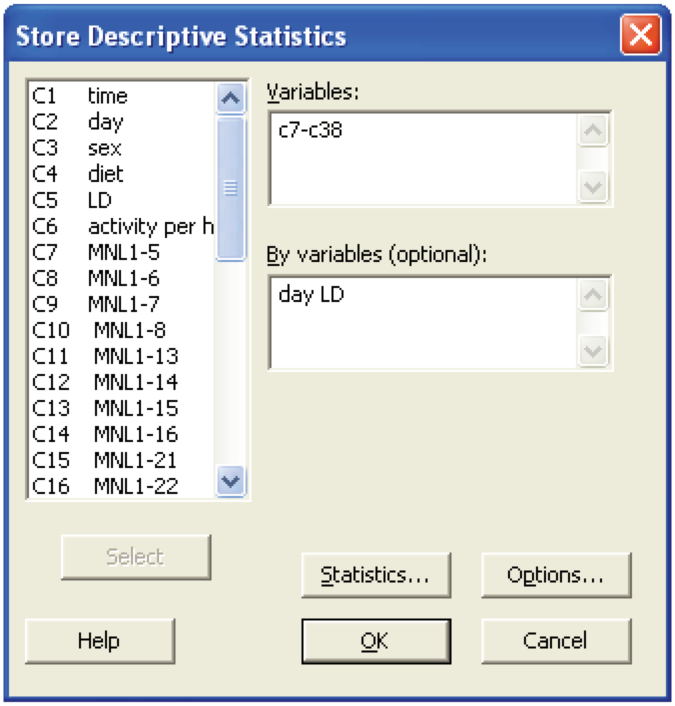
Now the data file should look like this:

Worksheet 1 ***

↓	C1-D	C2	C3-T	C4-T	C5	C6	C7	C8	C9
	time	day	sex	diet	LD	activity per hr bin	MNL1-5	MNL1-6	MNL1-7
1	8:10:00	1	M	NL	1	1	0	5	0
2	8:11:00	1	M	NL	1	1	0	0	0
3	8:12:00	1	M	NL	1	1	0	0	0
4	8:13:00	1	M	NL	1	1	0	0	0
5	8:14:00	1	M	NL	1	1	0	0	0
6	8:15:00	1	M	NL	1	1	0	0	0
7	8:16:00	1	M	NL	1	1	0	0	0
8	8:17:00	1	M	NL	1	1	0	0	0
9	8:18:00	1	M	NL	1	1	0	0	0
10	8:19:00	1	M	NL	1	1	0	0	0
11	8:20:00	1	M	NL	1	1	0	0	0
12	8:21:00	1	M	NL	1	1	0	7	0

2. Calculating the total activity per fly per day and L or D during that day

Stat>Basic statistics>Store Descriptive Statistics



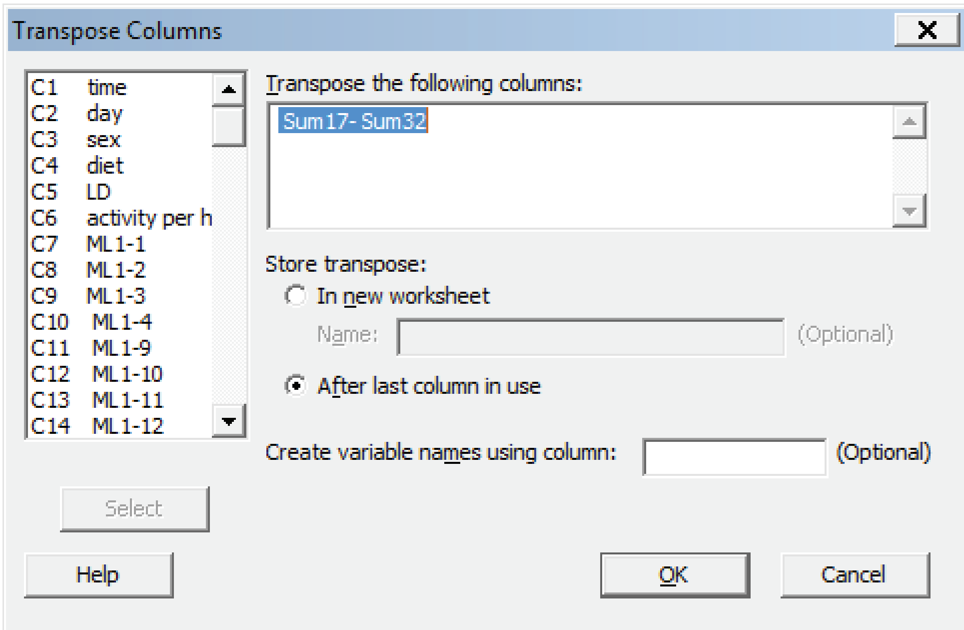
Now the totals per light period (L/D) for each day and for each fly have been calculated

The data file should look like this, where the columns 'sum' corresponding to the total activity per day per fly by LD:

C39	C40	C41	C42	C43	C44	C45	C46	C47	C48	C49	C50	C51	C52	C53	C5
By day	By LD	Sum1	Sum2	Sum3	Sum4	Sum5	Sum6	Sum7	Sum8	Sum9	Sum10	Sum11	Sum12	Sum13	Sun
1	0	19	219	21	1	0	0	26	0	10	30	54	30	0	
1	1	47	315	6	42	55	110	14	95	21	20	57	2	160	
2	0	13	17	42	2	50	1	0	0	16	34	24	0	0	
2	1	34	914	41	155	108	19	78	152	32	30	33	3	199	
3	0	0	0	0	0	0	0	39	10	12	54	24	81	0	
3	1	134	0	17	54	450	126	78	287	133	67	85	16	371	
4	0	55	0	0	0	0	25	14	0	7	7	47	0	20	
4	1	172	0	136	154	0	115	184	112	561	168	50	13	586	
5	0	7	0	0	19	0	12	0	1	13	54	6	0	0	
5	1	94	0	101	103	0	272	185	300	179	106	130	54	336	
6	0	25	0	162	27	0	7	0	0	3	189	0	2	0	
6	1	212	0	241	88	0	227	143	284	386	115	392	92	180	

3. Transpose the sums and label the variables day and LD:

Data> Transpose Columns



The data file should like this:

C4-T	C5	C6	C7	C8	C9	C10	C11	C12	C13
Flies	Day1D	Day1L	Day2D	Day2L	Day3D	Day3L	Day4D	Day4L	Day5I
Sum1	19	47	13	34	0	134	55	172	
Sum2	219	315	17	914	0	0	0	0	
Sum3	21	6	42	41	0	17	0	136	
Sum4	1	42	2	155	0	54	0	154	
Sum5	0	55	50	108	0	450	0	0	
Sum6	0	110	1	19	0	126	25	115	
Sum7	26	14	0	78	39	78	14	184	
Sum8	0	95	0	152	10	287	0	112	
Sum9	10	21	16	32	12	133	7	561	

4. We merge all data into a file which looks like this:

C1-T	C2-T	C3	C4-T	C5	C6	C7	C8	C9
Sex	Diet	Repeat	Flies	Day1D	Day1L	Day2D	Day2L	Day3D
M	NL	2	Sum43	12	28	66	120	41
M	NL	2	Sum44	43	396	30	425	118
M	NL	2	Sum45	27	610	64	908	13
M	NL	2	Sum46	277	539	0	271	0
M	NL	2	Sum47	0	88	29	134	0
M	NL	2	Sum48	0	96	0	220	0
M	L	1	Sum1	10	62	22	111	0
M	L	1	Sum2	8	183	5	114	54
M	L	1	Sum3	101	84	76	520	100
M	L	1	Sum4	16	32	73	46	64

5. We split the data day by day and calculate the difference (check in byday folder in thesis data folder):

For day 1:

Day	Age	Sex	Diet	Dark	Light	Diff_L&D
1	3	M	NL	19	47	28
1	3	M	NL	219	315	96
1	3	M	NL	21	6	-15
1	3	M	NL	1	42	41
1	3	M	NL	0	55	55
1	3	M	NL	0	110	110
1	3	M	NL	26	14	-12
1	3	M	NL	0	95	95
1	3	M	NL	10	21	11
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

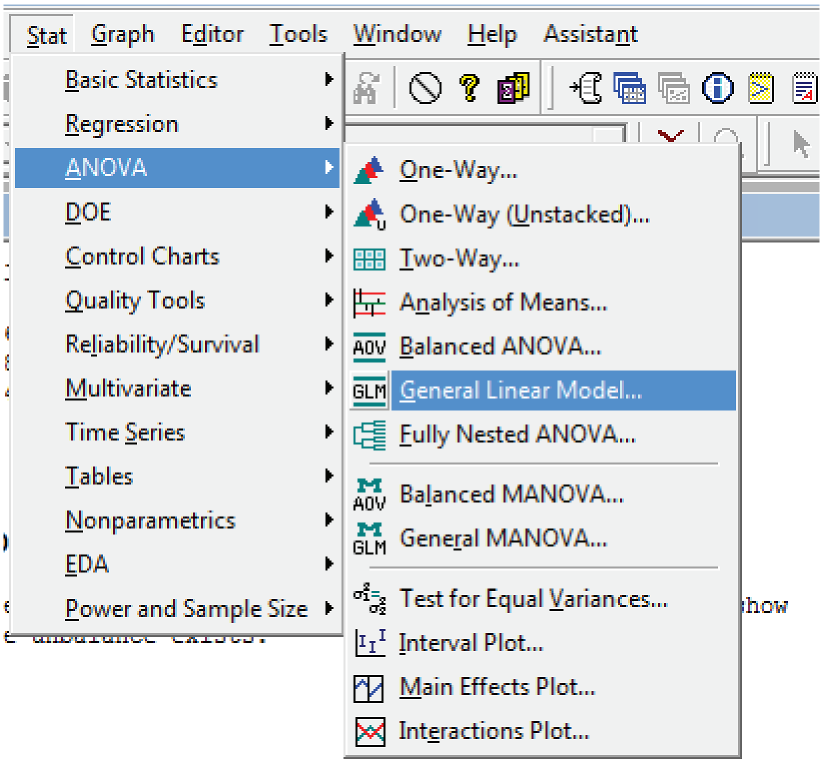
Then we get a file folder look like this:

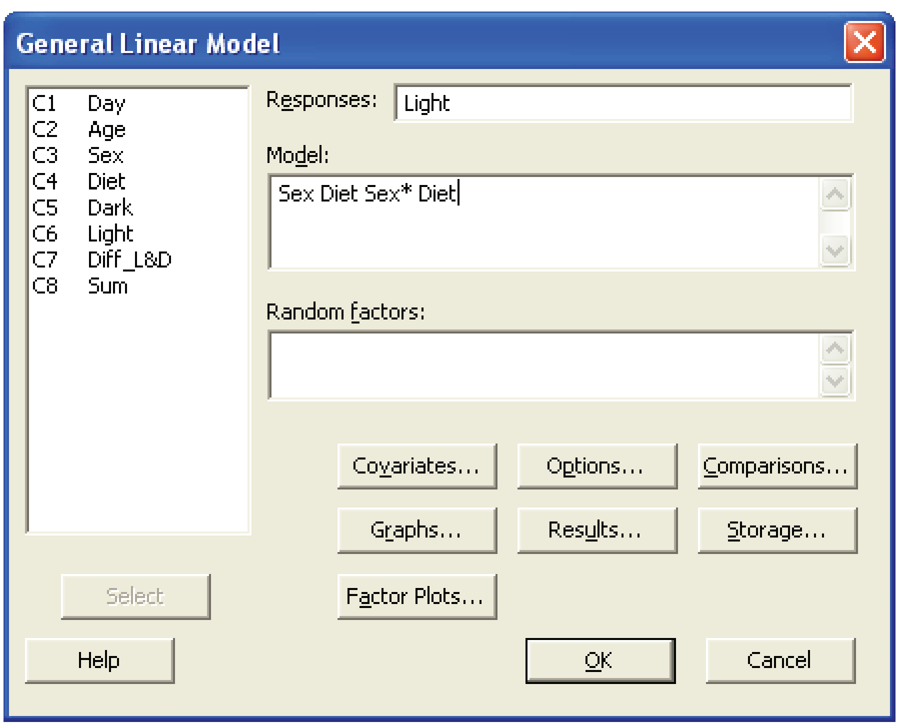
Day1.MTW
Day2.MTW
Day3.MTW
Day4.MTW
Day5.MTW
Day6.MTW
Day7.MTW

Appendix A MINITAB PART

II Two-way ANOVA

Stat>ANOVA>General Linear Model



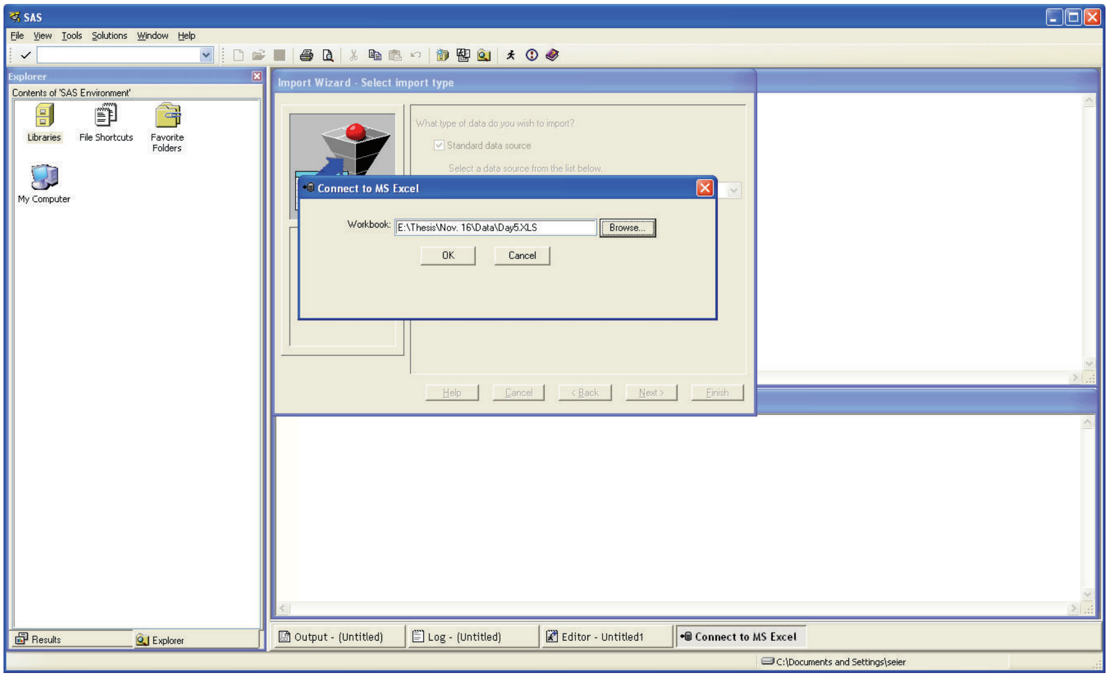


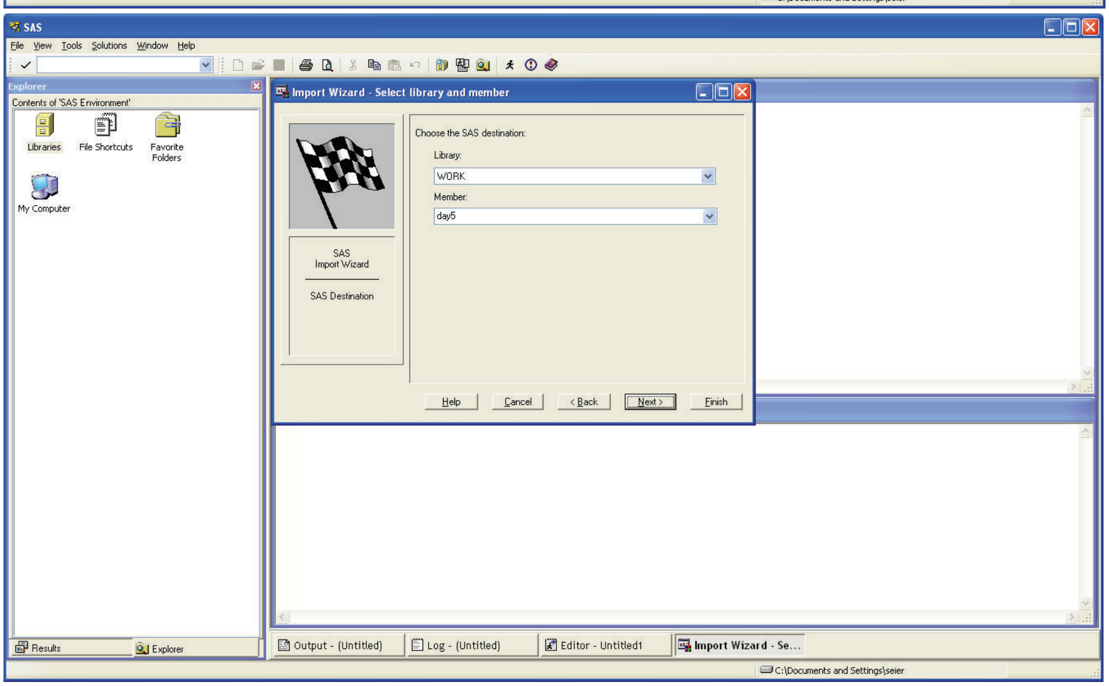
APPENDIX B SAS

I Two-way ANOVA and General Linear Model in SAS

a. Two-way ANOVA

1. To export the data file from Minitab into Excel and read it in SAS.





2. Program to do the analysis of day 5

```
options  ps=60 ls=80;
data dayfive;
set work.day5;
run;
proc print data=dayfive (obs=10);
run;
proc anova;
class Sex Diet ;
model  Light=Sex Diet Sex*Diet;
proc anova;
class Sex Diet ;
model  Dark=Sex Diet Sex*Diet;
proc anova;
class Sex Diet ;
model Diff_L_D=Sex Diet Sex*Diet;
run;
```

b. General Linear Model

Program to do the analysis of day 5

```
options  ps=60 ls=80;
data dayfive;
set work.day5;
run;
proc print data=dayfive (obs=10);
run;
proc anova;
class Sex Diet ;
model  Light=Sex Diet Sex*Diet;
proc glm;
class Sex Diet ;
model  Dark=Sex Diet Sex*Diet;
proc anova;
class Sex Diet ;
model Diff_L_D=Sex Diet Sex*Diet;
run;
```

II Repeated measures in SAS

a. Program for compound symmetric structure

1. light-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model light= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=cs subject=fly R RCORR;
run;
```

2. dark-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model dark= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=cs subject=fly R RCORR;
run;
```

b. Program for autoregressive order one structure

1. light-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model light= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=ar(1) subject=fly R RCORR;
run;
```

2. dark-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model dark= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=ar(1) subject=fly R RCORR;
run;
```

c. Program for unstructured structure

1. Light-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model light= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=UN subject=fly R RCORR;
run;
```

2. dark-fly-activity response

```
option ps=1000 ls=80;
data flies;
infile 'e:alldaysflies.dat';
input  fly sex$ diet$ day dark light;
run;
proc mixed;
class fly sex diet day;
model dark= sex diet sex*diet day sex*day diet*day day*sex*diet/S CHISQ;
REPEATED day/ type=UN subject=fly R RCORR;
run;
```

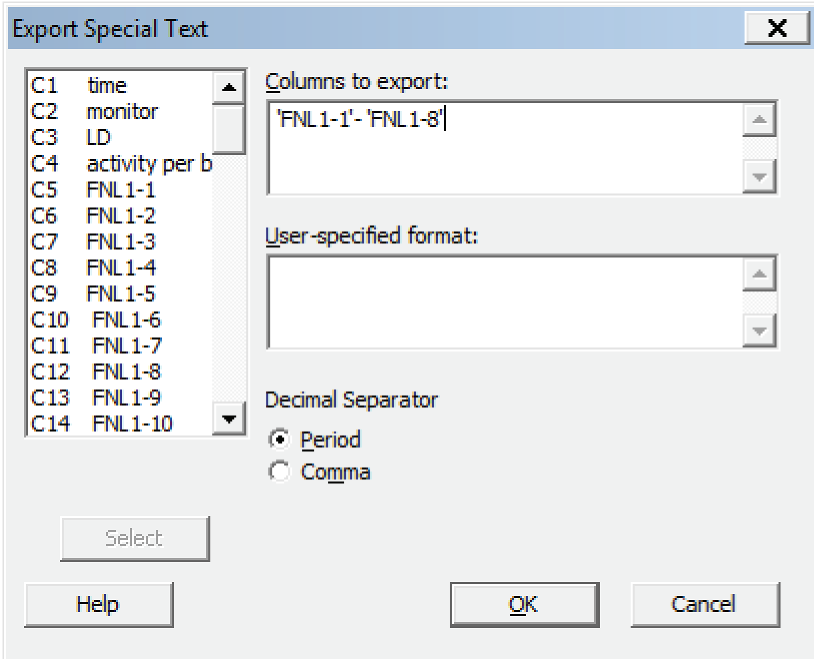
Appendix C R PART (Complete on March 12)

Analysis of Circadian Rhythms in R

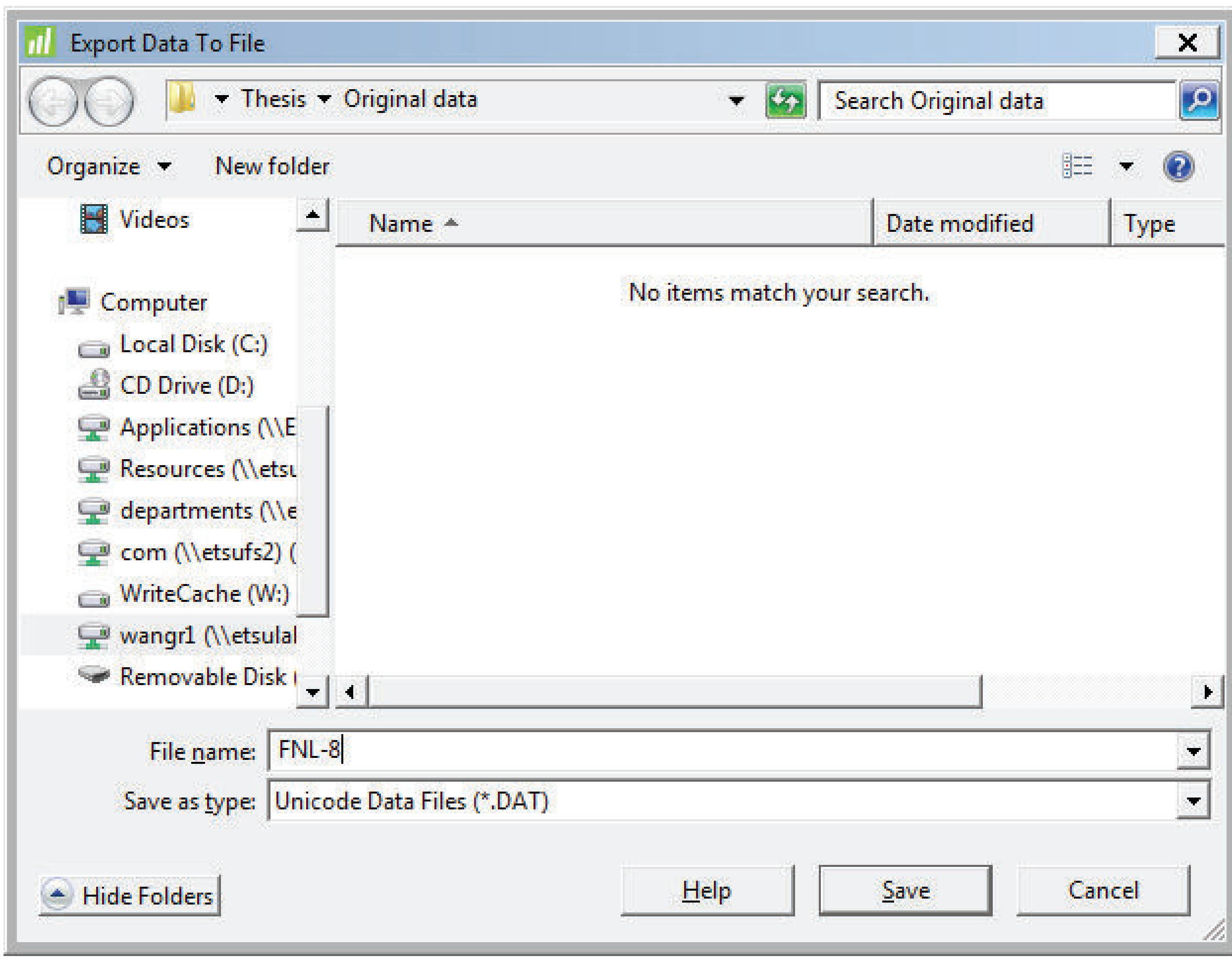
1. Create "objects" for particular flies and read it.
   a. First, open up 'FNL Activity' in origindata file. It looks like this:

time	monitor	LD	activity per bin	FNL1-1	FNL1-2	FNL1-3	Fl
8:09:00	1	1	1	0	0	0	
8:10:00	1	1	1	1	0	0	
8:11:00	1	1	1	3	8	0	
8:12:00	1	1	1	0	0	0	
8:13:00	1	1	1	1	3	0	
8:14:00	1	1	1	0	0	0	
8:15:00	1	1	1	0	0	0	
8:16:00	1	1	1	4	0	0	
8:17:00	1	1	1	0	0	0	
8:18:00	1	1	1	0	0	0	
8:19:00	1	1	1	0	0	0	
8:20:00	1	1	1	0	0	0	

   b. Select FNL1-1 to FNL1-8 and create "objects" in a new file which could be read in R.
      File>Other Files...>Export Special Text...



c. Click Ok to save data in a new file of .dat form and name it FNL-8.



2. Read in R.

Note: In R when the data file has only one column you use 'scan'; if it has several columns we use 'read.table'.

In this case, since there are eight columns, we use 'read.table'.

eightflies<-read.table('e:FNL-8.dat')

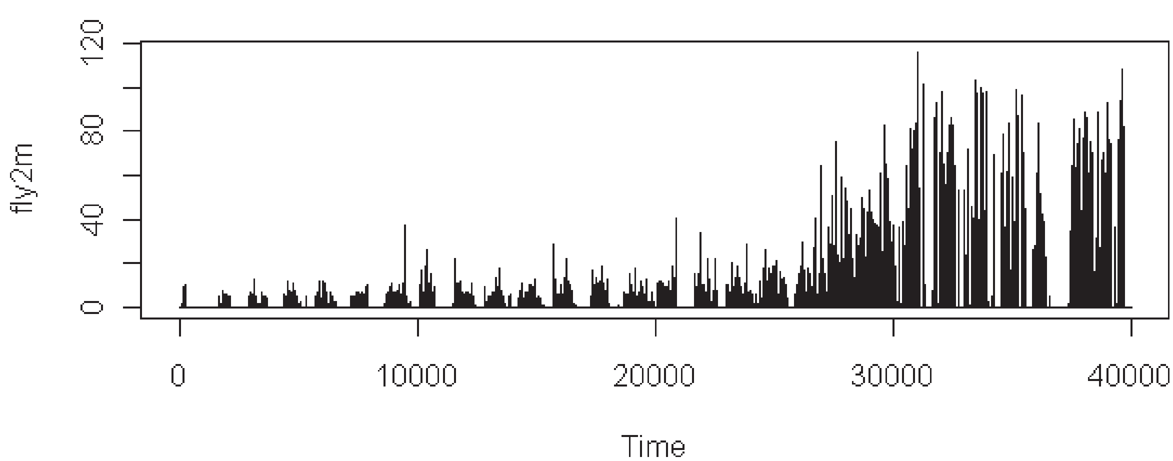
3. Now create an 'object' for each fly, and we do the time series plot and if the fly died before the end. (you can cut the data accordingly)

   a. Fly1-1

fly1<-eightflies[,1]
ts.plot(fly1)
fly1m<-fly1[1:30000]
ts.plot(fly1m)x<-scan('e:onebeesu.dat')

   b. Fly1-2

fly2<-eightflies[,2]
ts.plot(fly2)
fly2m<-fly2[1:40000]
ts.plot(fly2m)



For a fly like this one the trend will dominate the periodogram we won't be able to perceive the circadian rhythm even if it existed.

4. This is the perioplot function that will produce the plot of the periodogram.

```
perioplot<-function(x){
adjx=x-mean(x);

tf=fft(adjx);
nf=length(tf); n2=nf/2+1;
 pritf<-tf[c(1:n2)];
intensity<-(abs(pritf^2))/nf;
nyquist=1/2; pfreq<-seq(0,nf/2,by=1);
 freq<-pfreq/(length(pfreq)-1)*nyquist;
intmax<-max(intensity)
posmax<-max.col(t(intensity))
 freqmax<-(freq[posmax])
 maxper<-1/freqmax
 plot(freq,intensity,type="l")
text(0.2,intmax, label= maxper)}

perioplot(fly1m)
```

5. Now let's see how important the circadian rhythm is for this fly.

```
cpgram(fly1m)
```

6. This is the perionum function that will calculate the periodogram (numerical version).

```
## this function calculates the periodogram and gives numerical output
## output is: frequency, intensity, period and % cumulative intensity
perionum<-function(x){
adjx=x-mean(x);                   # substracts the mean of the series
tf=fft(adjx);                     # calculates finite Fourier transform
nf=length(tf); n2=nf/2+1;          # decides the number of frequencies
 pritf<-tf[c(1:n2)];         # takes the elements of the Fourier transform
intensity<-(abs(pritf^2))/nf;      # calculates the ordinates of periodogram
nyquist=1/2; pfreq<-seq(0,nf/2,by=1); # preparation for frequencies
 freq<-pfreq/(length(pfreq)-1)*nyquist; # calculates frequencies
periods<-1/freq;                    # calculates periods
cumint<-cumsum(intensity)*100/(max(cumsum(intensity))); # accumulates
## now the output is organized in columns
perfrint<-array(c(freq,intensity,periods,cumint),dim=c(length(freq), 4));
colnames(perfrint)<-c('frequency',  'intensity',  'period',  'cumulative%');
perfrint[0:25,]}


perionum(fly1m)
```