

Analysis of Dose-response OMIX Experiment in a two Analysis of Variance Under Ordered Constraint

Yousef El Aalamat

promotor :

Prof. dr. Ziv SHKEDY, dr. Dan LIN

**MASTER OF STATISTICS
THESIS CONTINUATION FORM**

Information on the project

Student Name: YOUSSEF SIADLANAT
Project Title: _____

Follow up of the project

ANALYSIS OF Dose-Response MIX Experiment in
a two-way analysis of Variance under ordered Constraint.

External supervisor

Name Zic Shkedy
Company _____

Date of first contact: _____

Date the project could start: _____

Intermediate contacts were:

- Weekly
- Bi-weekly
- Monthly
- Other (Define) _____

Possible relevant remarks:

Internal supervisor

Name DAN LIN

Date of first contact: _____

Date the project could start: _____

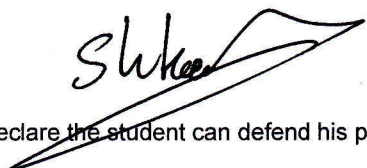
Intermediate contacts were:

- Weekly
- Bi-weekly
- Monthly
- Other (Define) _____

Possible relevant remarks:

I hereby declare the student can defend his project

Signature: _____



I hereby declare the student can defend his project

Signature: _____



Acknowledgement

I would like to express my gratitude and to acknowledge my supervisors:

Prof. Dr. ZIV SKHEDY and Dr. DAN LIN for giving me their unlimited support and very helpful explanations, remarks on this project, organization of the report, and for the time they spent with me during the preparation of this thesis. In addition I am also grateful to my lectures and colleagues for their advice and assistance when needed.

I would like to express my sincere appreciation to Prof.Dr TOMASZ BURZYKOWSKI.

Thank you for making time for me from your extremely busy schedule to guide, explain and help me during the preparation of this report. I would like to thank my parents, brother and sisters for their emotional support and encouragement.

Abstract

In data analysis, and especially when we are interested in investigating the relationship between a response variable Y (dependent variable) and a covariate X (an independent variable), one may want to investigate the trend of this relationship. This kind of relationship is of interest in a dose-response study, since in this kind of study its relationship relies on monotonicity of the true regression function. Dose-response studies are gaining a large interest in many research areas and especially in biomedical and pharmaceutical research. Testing homogeneity of means against an ordered alternative in one-dimension i.e. in one way analysis of variance, with respect to the dose levels has been discussed by several, and many test statistics have been derived in order to test homogeneity of means against ordered alternatives with respect to dose; these include: the global likelihood ratio test (Bartholomew 1961, Barlow et al. 1972, and Robertson et al. 1988), M statistic (Hu et al. 2005) and a modified M statistic (Lin et al. 2007) [11]. In this master thesis, we are more interested in a two-way analysis of variance layout, and more precisely in interaction term between the dose groups and the time levels, but in a two-way layout the alternative is dependent of the practical situation [14]. For testing the homogeneity of means (i.e. no dose effect) in a two-way layout against ordered alternatives, we use the likelihood ratio tests based on the Barlow works where the test depends on the isotonic means, so for solving this problem we use the Isotone package where we have a generalized pool-adjacent algorithm. By means of likelihood ratio test with the assumption of χ^2 distribution under the null, we compare three different models in order to find metabolites for which the saturated model i.e. model including the interaction between the two main effects of time and dose to an additive model where the main effect of the dose and time contribute additively. We use also the Hirotsu tests for testing the interaction term in a two-way layout, where these tests take into account the ordered alternatives.

KEYWORDS: dose-response study, monotonicity, two-way analysis of variance, Likelihood ratio test, Hirotsu tests.

Contents

1	Introduction and Motivation	1
2	Monotone Models	2
2.1	One Way ANOVA	2
2.2	General case	3
2.2.1	Additive Isotonic Models	3
2.2.2	General Isotonic regression	4
3	Algorithm	6
3.1	Pool-Adjacent Violators Algorithm (PAVA): One Predictor	6
3.2	Generalized Pool-Adjacent Violators Algorithm	7
4	Testing and Estimation Under Order Constrains	9
4.1	One Way ANOVA	9
4.2	Two way ANOVA	10
4.3	Two way ANOVA :Interaction	11
4.4	Likelihood Ratio Test:	14
5	Computational Complexity	15
5.1	SAS Macro	15
5.2	R code: HirotsuTests	18
6	Application	20
6.1	Model for fixed time point	20
6.1.1	Time 1	21
6.1.2	Time 2	23
6.1.3	Time 3:	25
6.2	LRT	27
6.2.1	Day 14:	27
6.2.2	Day 21:	30
6.3	Interaction:	34
7	Conclusion	40
8	Appendix	43

Chapter 1

Introduction and Motivation

Microarrays are becoming one of the essential techniques in the biological research. While in the first decade of the information explosion in biology it was centered on data that were obtained by sequencing procedures of DNA and protein, more recently microarrays and other techniques have been used to assay the levels of gene products expressed under various conditions. These techniques lead to the generation of a massive volume of such data with the hope that we can make biologically interesting inferences such as which set of genes or metabolite are co-expressed i.e. differentially expressed over the different conditions.

The regression model is an important tool for analyzing such data. Assume we have n observations of a response variable Y and a covariate X , the goal in the regression analysis is the estimation of the regression function f such that:

$$f(x) = E[Y|X = x]$$

is done by the least-squares, while in a dose-response study we make a restriction on the regression function $f(x)$ to be monotone. We study the isotonic regression, since one of its application is in dose-response study, where we expect that the response (Y) increase or decrease when we increase the doses of the compound i.e. the drug, so if the monotonicity is assumed, the regression function $f(x)$ should be monotonic. The trend of the relationship between the response and the covariate is of interest. The isotonic regression problem was studied in the begin of the sixtieth by the author *Bartholomew* [2][3]

For the additive isotonic regression we focus in this thesis on the paper by Mammen and Yu [8], where the additive factors $f_j(\cdot)$, $j = \{1, \dots, p\}$ are calculated using a backfitting algorithm. Estimation of the additive functions with the *Backfitting algorithm* has been done first by Bacchetti [1], while Morton-Jones, Diggle, Dickinson and Binks Parker [9], extend this algorithm to application in epidemiology studies. Mammen et al 2007[8] proposed an algorithm for the oracle estimator, based on the backfitting algorithm, for additive models of isotonic regression where each additive component can be estimated given that the others component are known. Hirotsu [14] proposed new test statistics for the interaction term in a two-way analysis framework i.e. for testing the null H_0 of additivity versus an ordered alternative.

Chapter 2

Monotone Models

2.1 One Way ANOVA

In this section we present the most important issues about the monotonic regression as introduced by Bartholomew[2]. It condenses only the univariate case: the case of one explanatory variables -*dose or time*- in this study; including also the tests. Note that under the term monotonic regression either an increasing(isotonic) or decreasing (antitonic) trend are included. In a simple linear regression we assume a linear relationship between a predictor $X = (x_1, x_2, \dots, x_n)$ and a given response $Y = (y_1, y_2, \dots, y_n)$ and in this regression line we are interested in minimizing the following function:

$$f(\alpha, \beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.1)$$

where $\hat{y}_i = \alpha + \beta x_i$, α and β are the regression parameters, while in the isotonic regression, the estimation of the isotonic regression line will be done in the sense of least squares under the condition that f is monotone in equation (2.1). The problem in the monotonic regression is that the researcher has no information before regarding the true regression function. Rather, the researcher can assume a specific shape that can be characterized by certain order restrictions. In case where y_i 's increase with the increasing x_i 's, we talk about the *isotonic regression*; while in the decreasing case of y_i 's, we talk about *antitonic regression*. First we are going to give some definition related to the problem in order that the reader will have a general idea about the isotonic regression. The isotonic regression method was first proposed by *Bartholomew* and *Barlow et al.* 1972 [2] [3] [4], starting with the assumption that the relationship between the response and the covariate is monotone, so formally we have an independent and identically distributed (*i.i.d*) sample $(X_i, Y_i), i = \{1, 2, \dots, n\}$; and we want to estimate the regression function:

$$f(x) = E[Y|X = x]$$

under the constraint that it is a monotonic function.

Let $X = \{x_1, x_2, \dots, x_n\}$ be an ordered set in \mathbb{R} (*real numbers*)

Definition 1:

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set with the order $x_1 \leq x_2 \leq \dots \leq x_n$.

A function f in \mathbb{R} is called isotonic function if:

$$\forall x, y \in X \quad \text{if } x \leq y \text{ then } f(x) \leq f(y).$$

◇

Definition 2:

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set with the order $x_1 \leq x_2 \leq \dots \leq x_n$, let $f(\cdot)$ be a function on the set X and $w(\cdot)$ a given positive weight function on the set X . An isotonic function \tilde{f} is an isotonic function of $f(\cdot)$ with weight $w(\cdot)$ with respect to the ordering $x_1 \leq x_2 \leq \dots \leq x_n$ if it minimizes in the class of isotonic functions $h(x)$ on X the sum:

$$\sum_{i=1}^n (f(x_i) - h(x_i))^2 w(x_i).$$

◇

2.2 General case

In the previous section we treat the isotonic regression problem for 1– dimension i.e. for one covariate, in this section we will try to introduce the isotonic regression problem when there are several explanatory variables, say $X_i = (x_{i1}, \dots, x_{ip})$ with $p > 1$. Customary/Traditionally two approaches arise when we are dealing(or when we have) more than one covariate:

- Either assuming that the covariates contribute additively to the response: Additive model.
- Either fitting a general regression model where the covariates do not contribute additively to the response.

The last approach is quite difficult in the sense for example in a 2– dimensional problems the standard techniques used to fit induce/cause problems with the estimation of the parameters and its interpretations [6].

2.2.1 Additive Isotonic Models

The additive isotonic regression model is totally different from the generalized linear models, because in the generalized regression modeling procedure the known numbers of parameters in the model affect the estimation procedures, whereas in the isotonic regression models generally the number of blocks in the estimated $f_j(\cdot)$ is not known in advance, but rather determined by the data in the additive isotonic model. Let assume that we have an independent and identically distributed vectors $(Y^1, X_1^1, X_2^1, \dots, X_d^1), \dots, (Y^n, X_1^n, X_2^n, \dots, X_d^n)$.

The additive isotonic regression models can be written as:

$$E[Y^i | X_1^i, \dots, X_d^i] = c + f_1(X_1) + f_2(X_2) + \dots + f_d(X_d), i = 1, \dots,$$

where the function $(f_j)_{j=1, \dots, d}$ are unknown with the assumption that they are monotone functions, for example isotonic functions.

In this framework we want to minimize the sum of squares of the errors, so the problem of minimization can be state as:

$$\sum_{i=1}^n (Y^i - c - \mu_1(X_1^i) - \dots - \mu_d(X_d^i))^2 \tag{2.2}$$

over the class of monotone functions satisfying the condition:

$$\int_a^b \mu_j(x_j) dx_j = 0, \quad \forall j, \quad j = 1, \dots, d. \quad (2.3)$$

for each μ_j , where $\forall j \quad x_j$ takes values in the interval $[a, b]$ and Y^i are given by:

$$Y^i = +f_1(X_1) + f_2(X_2) + \dots + f_d(X_d) + \epsilon^i, \quad i = 1, \dots, n. \quad (2.4)$$

We denote the estimators by $\hat{f}_1; \dots, \hat{f}_d, \hat{c}$.

2.2.2 General Isotonic regression

For simplicity, we will consider the case of two dimensions, i.e. two explanatory variables like in this master thesis where we have two factors: dose and time factors.

Let assume that we have two covariates ($P = 2$) X_1 and X_2 and each covariate have N_1 and N_2 ordered distinct values, with the corresponding response variable y_{ij} . In a two dimension the data can be seen as matrix with N_1 columns and N_2 rows, i.e. a D with dimension $N_1 \times N_2$.

$$\begin{bmatrix} y_{11} & y_{12} & \dots & y_{1N_1} \\ y_{21} & y_{22} & \dots & y_{2N_1} \\ \vdots & & & \vdots \\ y_{N_21} & y_{N_22} & \dots & y_{N_2N_1} \end{bmatrix}$$

each cell (i, j) in this matrix contains the response corresponding to i th observation from the first covariate and the j th observation for the second covariates. Let's give a definition about the Partial order for matrix.

Definition:(Partial order for matrix)

A matrix D is isotonic with respect to the partial order if and only if the elements y_{ij} of D accomplish the following restrictions:

$$y_{ij} \leq y_{kl} \quad \text{for every} \quad i \leq k \quad \text{and} \quad j \leq l.$$

◇

As we can see from this definition that the estimated response should be non-decreasing over columns and rows and if we projected into one dimension we will have the definition of the isotonic sequence.

It is important, when fitting an isotonic regression where the covariates do not totally contribute additively and specially in isotonic additive function that involves interactions, to realize that no additive component by itself needs to be isotonic, where f_j is assumed to be additive up to some order of interaction :

$$f(X) = \alpha_0 + \sum_{i=1}^p f_i(x_i) + \sum_{i \leq k} f_{ik}(x_i, x_k) \quad (2.5)$$

where p is the number of predictors for which the isotonic function $(f_i)_{i=1, \dots, P}$ are additive and the functions f_{ik} are the bivariate interactions (in a 2 dimension).

Tibshirani [6] state in his book: generalized additive models, that testing the lack of fit in additive model can be done by a diagnostic tool of plotting the residual with respect to the predictors $X_1 - X_2$. If this plot shows a grouping of the points it indicate lack of fit which implies the need of the interaction term between the two covariates , but this is in generalized additive models.

Chapter 3

Algorithm

3.1 Pool-Adjacent Violators Algorithm (PAVA): One Predictor

In the univariate case i.e. when the number of predictors $p = 1$, the best known algorithm for computing the least squares solution to the isotonic regression IR, is the pool adjacent violators algorithm (Barlow *et al.*1972).[4]

Let $M_n = \{(x_i, y_i), i = 1, 2, \dots, n\}$ denote a set of n observations of one explanatory variable (x) and one outcome (y), with the assumption that the x are presorted in a increasing order i.e.the x_1, x_2, \dots, x_p form an none decreasing sequence: $x_1 \leq x_2 \leq \dots \leq x_n$ where n is the numbers of observation in the dataset. As the monotonic regression problem is an optimization problem the PAV algorithm will compute a non decreasing sequence of values $\{z_i, i = 1, 2, \dots, n\}$ such that:

$$S = \sum_{i=1}^n (z_i - y_i)^2 \rightarrow \text{minimized} \quad (3.1)$$

under the constrained that $z_i \leq z_j$ if $x_i \leq x_j \forall i, j$ such that $i \leq j$. The PAV algorithm is recursive in the sense that the optimal solution for the data set M_n is constructed by starting from the solution for M_1 , which is subsequently modified into the solution for M_2 , and so on. For example in case $n=1$, it is obvious that the optimal solution for z_1 is : $z_1 = y_1$; in case $n=2$, in this case if $y_1 < y_2$ we set $z_1 = y_1$ and $z_2 = y_2$, otherwise pool the two observed values and put $z_1 = z_2 = \frac{y_1+y_2}{2}$ and so on. In addition the algorithm has the following recursive characteristic:

1. M_k is constructed by adding to the M_{k-1} the next data point(x_k, y_k) under the condition that $\forall j$ such that $j \leq k$ we have $x_k \leq x_j$.
2. If the solution for M_{k-1} is the values z_1, z_2, \dots, z_{k-1} , then an initial solution for M_k is formed by putting $z_k = y_k$, consequently the solution for M_k is constructed by pooling adjacent values that violate the monotonicity constraints.
3. the optimal solution $\{z_i, i = 1, 2, \dots, k\}$ for M_k is composed out clusters

So the optimal solution for M_k is derived by pooling adjacent preliminary z -values that violate the monotonicity constraints, for more precision the same value $\frac{(y_k+\dots+y_{k-j})}{j+1}$ is assigned to

each following z-values: z_k, \dots, z_{k-j} , where j is the smallest integer such that z_1, \dots, z_{k-1-j} combined with the new values of z_{k-j}, \dots, z_k form a non-decreasing sequence [4].

3.2 Generalized Pool-Adjacent Violators Algorithm

In the previous section we introduced the optimization problem for the isotonic regression (IR) when we have only one covariate, this algorithm is computationally efficient, and it provides solutions that are optimal in the sense that the mean square error is minimized. While in this section I will introduce the generalized pool-adjacent-violators (GPAV) algorithm, in case when we are dealing with at least two covariates. Here for simplicity I will briefly describe the GPAV algorithm, introduced by Burdakov *et al.* 2004 [10], in case we have only two covariates. Burdakov recently generalized the PAV algorithm in order to handle typical data that include more than one covariates, but in this case different ordering of the data may leads to different solution for monotonic regression problem [10].

For the description of the generalisation of the pool adjacent violators algorithm we I use the same notation as in the previous section.

Let $M_n = \{(x_{1i}, \dots, x_{pi}, y_i), i = 1, \dots, n\}$ denote a set of n observations of p explanatory variables and one outcome i.e. response variable.

Let x_i be a vector of p -element: $x_i \in R^p$, $x_i = (x_{1i}, \dots, x_{pi})$ are the explanatory variables, we define a partial order on the set $U_n = \{x_i, i = 1, \dots, n\}$ by setting:

$$x_i \leq x_j \quad \text{if} \quad x_{ki} \leq x_{kj}, \quad \text{for } k = 1, \dots, p.$$

Sort the elements of U_n such that for each i , x_i is a minimal element of the set $V_i = \{x_j, j = i, \dots, n\}$. We define the lower cover L_i of each x_i by:

$$x_i \leq x_j$$

and the inequalities

$$x_i \leq x_k \leq x_j$$

if these two inequalities are satisfied is only if $x_i = x_k$ or $x_k = x_j$.

The generalized pool adjacent violator algorithm has the following features:

1. M_{r+1} is formed by extending M_r with a data point (x_{r+1}, y_{r+1}) , such that, $\forall i > r + 1$, either $x_{r+1} \leq x_i$, or x_{r+1} is incomparable with x_i .
2. If the clusters I_1, \dots, I_q and their associated values $z(I_1), \dots, z(I_q)$ denote a solution for M_r , then a preliminary solution for M_{r+1} is formed by introducing the cluster I_{q+1} consisting of the integer $r + 1$, and setting $z(I_{q+1}) = y_{r+1}$. Thereafter, the final solution for M_{r+1} is obtained by joining I_{q+1} with left-neighbor clusters, one by one, until the z-values violating the monotonicity constraints have been removed. A cluster I_j is called a left neighbor of I_l if there exists an $i \in I_j$ and $k \in I_l$ such that x_i belongs to the lower cover of x_k .
3. The solution $z_i, i = 1, \dots, n$ obtained for M_n is composed of clusters of identical z-values.

The solution obtained by using this algorithm are monotonic in the explanatory variables, but we should know that we are dealing with two problems when using this algorithm:

1. a cluster may have many different left neighbours.
2. the pre-sorting which ensures that, for each i , x_i is a minimal element of the set:
 $V_i = \{x_j, j = i, \dots, n\}$ may be done in different ways.

Chapter 4

Testing and Estimation Under Order Constrains

In isotonic regression models, we wish to test the homogeneity of means (H_0) against an ordered alternative (H_A) usually a non-decreasing alternative:

$$H_0 : \mu_1 \leq \mu_2 \leq \dots \leq \mu_k \text{ versus } H_1 : \mu_1 \leq \mu_2 \leq \dots \leq \mu_k,$$

where k is the numbers of the dose groups d_1, \dots, d_k with the inequality holding at least once.

4.1 One Way ANOVA

In monotonic regression, we are dealing with an ordered alternative i.e. the alternative is constrained to have a trend, so the hypotheses in the MR are:

$$H_0: \text{equality of } (\alpha_i)_{i=1,\dots,k}$$

versus

$$H_A : \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k \quad \text{OR} \quad \alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_k$$

according to the assumed trend, where k is the number of dose groups, in this study $k = 4$: three dose groups in addition to a vehicle group.

For testing the homogeneity of means against an ordered alternative in the univariate case, several testing procedures can be used, like the global likelihood ratio test (Barlow et al.1972, and Robertson et al.1988)[4] [15], the M (Hu et al. 2005) statistic and the modified M(Lin et al.2007) statistic[11].

In the microarray setting, for each gene/metabolite we fit the following ANOVA model:

$$Y_{ij} = \alpha(d_i) + \epsilon_{ij}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, n.$$

where Y_{ij} is the j th gene/metabolite expression at the i th dose level, $d_i (i = 1, 2, \dots, k)$ are the k dose levels, $\alpha(d_i)$ is the mean gene/metabolite expression at each dose level, and the error term is Gaussian with mean zero and finite variance. In the one way framework the null hypothesis of homogeneity of means which also means no dose effect is defined as follow:

$$H_0 : \alpha(d_1) = \alpha(d_2) = \dots = \alpha(d_k).$$

versus a one sided alternative hypothesis of a dose effect:

$$\begin{aligned} H_A^{Up} &: \alpha(d_1) \leq \alpha(d_2) \leq \dots \leq \alpha(d_k) \quad \text{increasing dose effect} \\ H_A^{Down} &: \alpha(d_1) \geq \alpha(d_2) \geq \dots \geq \alpha(d_k) \quad \text{decreasing dose effect} \end{aligned}$$

with at least one strict inequality.

Testing the null H_0 against H_A^{Up} or H_A^{Down} involves the estimation of the means under both the null and alternative.

Testing the homogeneity of means versus an ordered alternative using likelihood ratio test in the Gaussian case was discussed by Barlow et al.(1972)and Robertson et al. (1988). In the Gaussian framework the likelihood function is :

$$L = \frac{1}{2\pi^{N/2}} \prod_{i=1}^k \frac{1}{\sigma_i^{n_i}} \exp\left(-\frac{1}{2} \sum_{i=1}^k \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2\right) \quad (4.1)$$

where σ_i^2 is the variance for each metabolite, in this study we assume that all the metabolite have the same variance σ^2 and μ_i is the mean for the dose d_i . Since in the likelihood ratio test statistic, the variance under the null and under the alternative are compared, so this statistic, after we got the maximum likelihood estimator (ML) of the parameter σ^2 that maximize the likelihood function,became:

$$\Lambda_{01} = \left(\frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2}\right)^{N/2} \quad (4.2)$$

so we reject the null hypothesis i.e. no dose effect for small value of Λ_{01} , or equivalently we reject for large value of:

$$\bar{E}_{01}^2 = 1 - \Lambda_{01}^{2/N} = 1 - \frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2} = \frac{\hat{\sigma}_{H_0}^2 - \hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2} \quad (4.3)$$

with $\hat{\sigma}_{H_0}^2$, is the estimate of the variance under the null and $\hat{\sigma}_{H_1}^2$ is the estimate of the variance under the alternative, we should notice that the variance under the alternative depends on the isotonic means since the MLE of $\hat{\sigma}_{H_1}^2$ is:

$$\hat{\sigma}_{H_1}^2 \propto \sum_{n_i}^{j=1} (y_{ij} - \hat{\mu}_i)^2$$

where $\hat{\mu}_i$ are the isotonic means following a specified trend. Thus \bar{E}_{01}^2 became:

$$\bar{E}_{01}^2 = \frac{\sum_{ij} (y_{ij} - \hat{\mu})^2 - \sum_{ij} (y_{ij} - \hat{\mu}_j^*)^2}{\sum_{ij} (y_{ij} - \hat{\mu})^2} \quad (4.4)$$

4.2 Two way ANOVA

This section deals with the tests for monotonicity in a two-way layout for the normal means. Let assume that in a two-way analysis of variance that we have the following model:

$$Y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijl} \quad (4.5)$$

where α_i is the main dose effect, β_j the main time effect, γ_{ij} is the interaction term between the two main effects and ϵ_{ijl} is a Gaussian with mean zero and finite variance. In a two-way framework, one may be interested in testing the main effects of the dose i.e. testing

homogeneity of means (no dose effect) versus an ordered alternative, so in the same way as in the univariate case, we can use the LRT which is comparing the variance under the null and the variance under the alternative. the variance under the alternative is given by:

$$\hat{\sigma}_{H_A}^2 = \frac{cb \sum_{i=1}^a (\hat{\alpha}_i - \hat{\alpha}_i^{H_A}) + \sum_{ijl} (y_{ijl} - y_{ij})^2}{abc} \quad (4.6)$$

where a is the number of dose groups, b is the number of time points and c the number of subjects. The test statistic for testing the null hypothesis(Homogeneity of means in dose groups) against the ordered alternative is also the likelihood ratio test defined as the ratio of the variance under the alternative and the variance under the null:

$$\bar{E}_{01}^2 = 1 - \Lambda_{01}^{2/N} = 1 - \frac{\hat{\sigma}_{H_1}^2}{\hat{\sigma}_{H_0}^2} \quad (4.7)$$

So reject the null of no dose effect for small value of $\Lambda_{01}^{2/N} = \hat{\sigma}_{H_1}^2 / \hat{\sigma}_{H_0}^2$, or equivalently for large value of \bar{E}_{01}^2 in formula (4.7).

After deriving the MLE for the remains parameters in the model (4.5), it can be seen that the MLE for the main effect of time did not depend on which hypothesis for the main dose effect $(\alpha_i)_{i=1,\dots,4}$ is assumed to hold(the proof for this statement is omitted in this project). This means that for testing the time effect in the model (4.5) one can use the test statistic used in two-way analysis of variance with the F distribution under the null(: no time effect).

4.3 Two way ANOVA :Interaction

For interaction term in a two way framework, the problem of testing is quite difficult in the sense that it requires the specification of the ordered alternative appropriate to the study design, then establishing the test statistic[14].

In a two way analysis of variance, several types of ordered alternatives can be defined for the interaction term, and for each ordered alternative Hirotsu proposed appropriate methods for testing the null hypothesis of no interaction effects versus the specified ordered alternative. By comparison with the one way layout, where the problem of testing the the null hypothesis against an ordered alternative was discussed by several authors like Bartholomew [2] [3], while in the two way layout the problem was discussed by Hirotsu (1978) [14], however it is not straightforward to set up an ordered alternative appropriate to practical study.

Barlow et al *et al.1972*[4] introduced an ordered alternative, for testing the interaction term, such as:

$$\mu_{i+1,j+1} - \mu_{i+1,j} - \mu_{i,j+1} + \mu_{ij} \geq 0, \forall i, j \quad (4.8)$$

which is equivalent the the following inequalities:

$$\mu_{i+1,j+1} - \mu_{i+1,j} \geq \mu_{i,j+1} - \mu_{ij} \quad (4.9)$$

$$\mu_{i+1,j+1} - \mu_{i,j+1} \geq \mu_{i+1,j} - \mu_{ij} \quad (4.10)$$

where μ_{ij} is the expected response in the (i, j) th cell, with at least one strict inequality. Hirotsu introduced some ordered alternatives which are suitable for the most practical studies. So for testing H_0 of additivity against an ordered alternative when no natural ordering in

levels , he recommended the following ordered alternative:

for all $i \neq i' = 1, \dots, a$. and $(j = 1, \dots, b - 1)$

$$\mu_{i,j+1} - \mu_{i',j+1} \geq \mu_{i,j} - \mu_{i',j} \quad OR \quad \mu_{i,j+1} - \mu_{i',j+1} \leq \mu_{i,j} - \mu_{i',j} \quad (4.11)$$

Looking to this alternative, we can say that when the level j of the columns increase in this case ... , the differences $\mu_{ij} - \mu_{i'j}$ are increasing or decreasing respectively. For testing the H_0 of additivity versus this alternative, Hirotsu introduced three modified tests.

First before developing the Hirotsu tests, let's start with some definition of the variables. Let assume we have a dose groups, b time points and c subject divided into a dose groups, in our study a=4, b=3 and c=24. in his tests he used the arithmetic means.

Method 1: For expressing the variation among the rows, Hirotsu proposed:

$$S_{AXB}^1 = \sum_{i=1}^a \sum_{m=1}^{b-1} \frac{cbm}{b-m} \left\{ \sum_{j=1}^m \frac{y_{ij.} - y_{.j.}}{m} - (y_{i..} - y_{...}) \right\}^2 \quad (4.12)$$

however S_{AXB}^1 is not χ^2 distributed under the null H_0 of additivity, but S_{AXB}^1/σ^2 is well approximated by $d\chi_\phi^2$, where :

$$d = tr(P_b^{*'} P_b^*)^2 / (b-1), \quad and \quad \phi = (a-1)(b-1)^2 / tr(P_b^{*'} P_b^*)^2 \quad (4.13)$$

so $S_{AXB}^1 / (\phi d \sigma^2) = S_{AXB}^1 / ((a-1)(b-1)\sigma^2)$ is distributed as χ_ϕ^2 / ϕ and is independent of S_e the error sums of squares. Hence the test statistic is:

$$T_1 = \frac{S_{AXB}^1 / ((a-1)(b-1))}{S_e / (ab(c-1))} \quad (4.14)$$

and we reject this test statistic i.e. T_1 if : $T_1 > F \{ \phi, ab(c-1); \alpha \}$.

The values of $d = tr(P_b^{*'} P_b^*)^2 / (b-1)$, for $b \in \{3, \dots, 10\}$, are displayed in the following table:

Table 4.1: Values of $A = (b-1)^2 / tr(P_b^{*'} P_b^*)^2$

b	3	4	5	6	7	8	9	10
A	1.6	1.976	2.223	2.409	2.543	2.647	2.730	2.798

Method 2: Let S_{AXB}^2 be:

$$S_{AXB}^2 = c \left\{ \frac{b(b^2-1)}{12} \right\}^{-1} \sum_{i=1}^a \left\{ \sum_{j=1}^b \left(j - \frac{b+1}{2} \right) (y_{ij.} - y_{.j.}) \right\}^2 \quad (4.15)$$

we can see that S_{AXB}^2 is distributed as $\sigma^2 \chi_{a-1}^2$ under the null H_0 , and also it is a component of the S_{AXB} , thus the test statistics that he proposed is:

$$T_2 = \frac{S_{AXB}^2 / (a-1)}{S_e / (ab(c-1))} \quad (4.16)$$

so we reject the null H_0 of additivity for $T_2 > F \{ a-1, ab(c-1); \alpha \}$, for the remaining part of S_{AXB} he assumed to be an error term, but he did not pool it into the error sums of squares S_e .

Method 3: Let S_{AXB}^3 be:

$$S_{AXB}^3 = c \left(\sum_{j=1}^b l_j^2 \right)^{-1} \sum_{i=1}^a \left\{ \sum_{j=1}^b l_j (y_{ij} - y_{.j}) \right\}^2 \quad (4.17)$$

for testing the variation among a items, this sums of squares is once more a part of S_{AXB} . the l_j for $j = 1, \dots, b$; are:

$$l_j = \sqrt{j(b-j)} - \sqrt{(j-1)(b-(j-1))}$$

The test statistic is:

$$T_3 = \frac{S_{AXB}^3/(a-1)}{S_e/(ab(c-1))} \quad (4.18)$$

we reject the null of additivity for $T_3 > F\{a-1, ab(c-1); \alpha\}$

In a two way analysis of variance, testing interaction effect requires specification of the ordered alternatives appropriate to the practical situation then establishing its test [14].

In the case of the two-factor interaction framework, we are dealing with various types of alternatives. For these ordered alternatives, Hirotsu proposed appropriate methods for testing the null hypothesis. It is not straightforward to set up ordered alternatives in comparison to the one-way framework, because the contrasts expressing interaction effects are themselves Barlow et al [4].

$$S_{AXB}^{2,2} = c \left(\sum_{j=1}^b l_j^2 \right)^{-1} \sum_{i=1}^a \left\{ \sum_{j=1}^b l_j (y_{ij} - y_{.j}) \right\}^2 \quad (4.19)$$

where the l_j is :

$$l_j = \sqrt{j(b-j)} - \sqrt{(j-1)(b-(j-1))}$$

4.4 Likelihood Ratio Test:

The likelihood ratio test (LRT) is a statistical test to compare between two models. A saturated model or a comparatively complex model is compared to less complex model i.e. reduced model to see if fits the data significantly better. In case the reduced model fits significantly the data, the additional parameters in the saturated model could be ignored. The LRT is only valid if it is used to compare hierarchically nested models so the saturated model must differ from the reduced model only the addition of one or more parameters. Merely adding additional parameters will always result in a higher likelihood but the question here is the significance in fitting the model. the LRT compares the likelihood of the two models: the saturated model and the reduced model and this statistic is approximated by a χ^2_{df} distribution. To determine the statistically significance of the difference of the likelihood among the two models, we must take into account the degrees of freedom(df) which is equal to the number of the additional parameters in the saturated model.

In this master thesis we are going to do LRT sequentially starting from the saturated model and ending with the model that has only one covariates since we are dealing with two factors: dose and time factors. so for each metabolite we fit the following models:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (4.20)$$

$$y_{ijl} = \mu + \alpha_i + \beta_j + \epsilon_{ijl} \quad (4.21)$$

$$y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijl} \quad (4.22)$$

where α_i is the main dose effect, β_j the time effects, γ_{ij} the interaction term and the error terms are normally distributed with mean zero and variance: σ^2 . The use of likelihood ratio test was done for finding metabolites for which the saturated model i.e. model (4.22) was significant but also looking for metabolites where the main time effect was also significant. For the comparison between the models we assumed that the likelihood ratio test is χ^2 distributed under the null of none need of adding parameters to the model, with a degrees of freedom equal to the number of parameters added.

Chapter 5

Computational Complexity

Despite the existence of these algorithms, the practical use of monotonic regression has long been hampered by the lack of computational methods that can accommodate several explanatory variables. The implementation of these algorithms is still limited to simple case which make the analysis quite difficult in the statistical packages like SAS and R, the one used in this thesis.

When we are dealing with a microarray data one common problem is the large number of dependent variables i.e. the metabolite expression, since in our datasets we are dealing with 229 metabolite expressions.

5.1 SAS Macro

As mentioned previously, one problem arises in microarray setting is the large number of dependent variables, so for the likelihood ratio test statistics for comparing two models i.e. model (4.22) to model (4.21), we implement the following macro which is going to do for each metabolites:

1. Extracting the log-likelihood of the models: (4.22) and (4.21).
2. Then computing the likelihood ratio test.

The SAS macro is:

```
\%macro nlmodeldn(dataset=, lrt2=m=);
```

```
data templ;  
set &dataset;  
obs=_N_;  
run;
```

```
data genenamedn;  
set templ;  
\%do i=1 \%to 1;  
genename=genes;  
where obs<=&m;
```

```

output;
\%end;
keep genename;
run;

\%do i=1 \%to &m;

data temp2;
set temp1;
where obs=&i;
drop obs genes ;
run;

proc transpose data=temp2 out=temp2;
run;

data temp3;
set temp2;
keep coll;
rename coll=gene;
run;

data temp43;
merge design3 temp3;
run;
data temp42;
merge design2 temp3;
run;

ods trace on;
ods listing close;

proc nlmixed data=temp42 ;

parms    beta1=0,beta2=0,beta3=0,beta4=0,
          beta5=0, beta6=0,s2=0.5, s2u=0.5 ;

h2=exp(beta2);h3=exp(beta3);h4=exp(beta4);

pred=    beta1*d1 - h2*d2 - h3*d3 - h4*d4 + beta5*t2 + beta6*t3 + u;

model gene ~ normal(pred,s2) ;
random u~normal(0,s2u) subject=id;
ods output    FitStatistics=LR2(rename=( Value=Value2 )) ;

```

```

run;

proc nlmixed data=temp43 ;
parms    beta1=0, beta2 =0 ,  beta3 =0 ,  beta4 =0 ,
          beta5=0, beta6 =0 ,  beta7 =0 ,  beta8 =0 ,
          beta9=0, beta10=0 ,  beta11=0 ,  beta12=0 ,
          s2=0.5, s2u=0.5 ;

h1=beta1 ; h2=exp(beta2) ; h3 =exp(beta3) ; h4=exp(beta4);
h5=beta5 ; h6=exp(beta6) ; h7 =exp(beta7) ; h8=exp(beta8);
h9=beta9 ; h10=exp(beta10); h11 =exp(beta11); h12=exp(beta12);

pred =  h1*td11 - h2*td12 - h3*td13 - h4*td14 +
        h5*td21 - h6*td22 - h7*td23 - h8*td24 +
        h9*td31 - h10*td32 - h11*td33 - h12*td34 + u ;
model gene ~ normal(pred,s2) ;
random u~normal(0,s2u) subject=id;
ods output  FitStatistics=LR3(rename=( Value=Value3 )) ;
run;

ods trace off;
ods listing;
data A2dn;
merge LR2 LR3;
if (Descr ='-2 Log Likelihood') then do;
LRT2=(value2-Value3);
pvalue2=1-Probchi(LRT2,6);
output;
end;
proc print data=A2dn;
label  LRT2= ' F value '
        pValue2 = 'Prob > ChiSquare ';
format pvalue2 pvalue8.;
var LRT2 pValue2 ;
run;
data temp7;
set A2dn;
obs=_N_;
run;
data temp8;
set temp7;
keep    value2 value3 lrt2  pvalue2;
run;
proc append base=&lrt2  data=temp8  ;run;

\%end;
\%mend;

```

5.2 R code: HirotsuTests

First change the structure of the data in order to have a sub data for each dose group. the value of deno is : SSE/df where df is its degree of freedom.

- The first test statistic S_{AXB}^1 :

```

B<-Bm<-Bi<-NULL
for( i in 1:a){
  for(m in 1:(b-1)){
    B<-NULL
    for( j in 1:m){
      B[j]<-(mean_c[i , j]-mean_ac[j])/m
    }
    Bm[m]<-((c*b*m)/(b-m))*(sum(B)-(mean_bc[i]- mean(mean_bc)))^2
  }
  Bi[i]<-sum(Bm)
}
S1axb<- sum(Bi)
Saxb1<-S1axb/((a-1)*(b-1))
S1<-Saxb1/deno

```

- The second test statistic based on $S_{AXB}^{2.1}$:

```

A21<-B21<-NULL
c21<-c*(b*(b^2-1)/12)^(-1)

for( i in 1:a){
  cst<-0
  for(j in 1:b){
    B21[j]<-(j-(b+1)/2)*(mean_c[i , j]- mean_ac[j])
    cst<-cst+B21[j]
  }
  A21[i]<-cst^2
}
test21<-sum(A21)
Saxb21<-test21*c21
S21<-(Saxb21/(a-1))/deno

```

- The third test statistic based on $S_{AXB}^{2.2}$:

```

B<-l<-A<-NULL
cst<-0
for( i in 1:a){
  cst<-0
  for( j in 1:b){
    l[j]<-sqrt(j*(b-j)) - sqrt((j-1)*(b-j+1))
    B[j]<-l[j]*(mean_c[i , j]- mean_ac[j])
  }
}

```



```

                                cst<- cst+B[j]
                                }
A[i]<-cst^2
}
test<-sum(A)
l2<-NULL
for( j in 1:b){
l2[j]<- l[j]^2
}
Saxb<-c*(sum(l2))^(-1)*test
S22<-(Saxb/(a-1))/deno

```

Chapter 6

Application

The dataset used in this study is composed of the expression levels of 229 metabolites from 24 subjects, these subjects covering 4 dose groups and measured three times. So in addition to the response Y_{it} , $t = \{1, 2, 3\}$, $i \in \{1, \dots, 6\}$ i.e. the expression level of the metabolites, 2 covariates were observed, namely the *DOSE* and *TIME*, for Dose level the dosing was performed twice daily started at day1(day-14) and was continued through day21(day21) by means of three different dose of the compound JNJ-cpX and a control vehicle, the three different dosage are low dose:0.16mg/Kg, medium dose :0.63mg/Kg and high:5mg/Kg.For the Time level we have three time points: (0, 4 and 22.5h), where the 229 metabolites of each subject was measured at each time point.

In a dose response study the direction of the alternative is not known in advance, which leads to maximize twice the likelihood for both trend, then define the trend with the large maximum likelihood.

The table below displays the number of metabolites with the same trend at each time point:

Dataset	Up Trend	Down Trend
Day14	79	58
Day21	46	81

Table 6.1: number of metabolites in the two days for each trend

from this table 6.1 we can see that the majority of the metabolites in day14 with the same trend have an increasing trend while this tendency became opposite in the day21, and this with respect to the dose levels at each time point. It was also observed that the common metabolites display a different trend in both days, since 52 out of 79 metabolites in day14 with an increasing trend, their trend changed to a decreasing trend in day21 and 21 out of 58 metabolites with a decreasing trend in day14 changed to an increasing trend in day21.

6.1 Model for fixed time point

For each metabolites the following model was fitted:

$$Y_{ij} = \mu(d_i) + \epsilon_{ij}; \quad i \in \{0, \dots, 3\}, j \in \{1, 2, \dots, n_i\}; \quad (6.1)$$

where Y_{ij} is the expression of the j th metabolite at the i th dose level d_i , $\mu(d_i)$ is the mean metabolite expression at each dose level and the error terms ϵ_{ij} are independently and identically distributed with mean zero and finite variance: $\epsilon_{ij} \sim N(0, \sigma^2)$.

In this section we present the results of the model 6.1.

In a one way analysis of variance framework, several testing procedures (i.e. the global likelihood ratio test, Williams, Marcus, the M and the Modified M), that take into account the order restriction of the means with respect to the increasing doses can be used, but in this project we focus only on the global likelihood ratio test discussed by Barlow et al (1972) [2] [3] [4], with the assumption that the response is normally distributed. This test statistic, in which the variance under the null (homogeneity of means in dose) and under the alternative were compared, is given by:

$$\bar{E}_{01}^2 = 1 - \Lambda_{01}^{2/N} = \frac{\sum_{ij}(y_{ij}-\hat{\mu})^2 - \sum_{ij}(y_{ij}-\hat{\mu}_j^*)^2}{\sum_{ij}(y_{ij}-\hat{\mu})^2}$$

6.1.1 Time 1

The table below table 6.2 displays the metabolites with their trend for which the dose effect was significant at the first time point in both days i.e. day14 and day21.

Day 14			Day 21	
	Metabolite	Trend	Metabolite	Trend
1	X4.Hydroxyphenylpyruvic.acid 14	Dn	Galactose.lipid.fraction 31	Dn
2	Campesterol 33	Up	Campesterol 33	Dn
3	Sphingolipid..2 56	Up	Eicosadienoic.acid.C20.2putative 54	Dn
4	Sphingolipid..3 59	Up	Sphingolipid.2 56	Dn
5	Sphingolipid..4 60	Up	Sphingolipid..3 59	Dn
6			Sphingolipid..4 60	Dn
7			Unknown.59600010. 111	Dn

Table 6.2: The significant metabolites at the first time point

From this table 6.2 we can see that in day 14 only one metabolite was decreasing while the remains metabolites Campesterol, Sphingolipid..2, Sphingolipid..3 and Sphingolipid..4 showed an increasing trend, for the day 21 all the significant metabolites have a decreasing trend. we observed also that the trend of the common metabolites was different, since their trend changed from an increasing trend in day14 up to decreasing trend in day21. The figure below fig 6.1 displays the isotonic plot for the metabolites in day14, we see that these metabolites showed an increasing trend and specially the higher dose group has an effect in comparison to the vehicle group, which can interpreted as the effect of the dose in this time for these metabolites. For the Sphingolipid 2,3 and 4 the difference was significant between the dose groups. From the figure 6.1, we see that the trend of the metabolites Campesterol and sphingolipid..2 was different in both days, and this was also observed form the metabolites sphingolipid..3 and sphingolipid..4(see fig 8.1 in appendix).

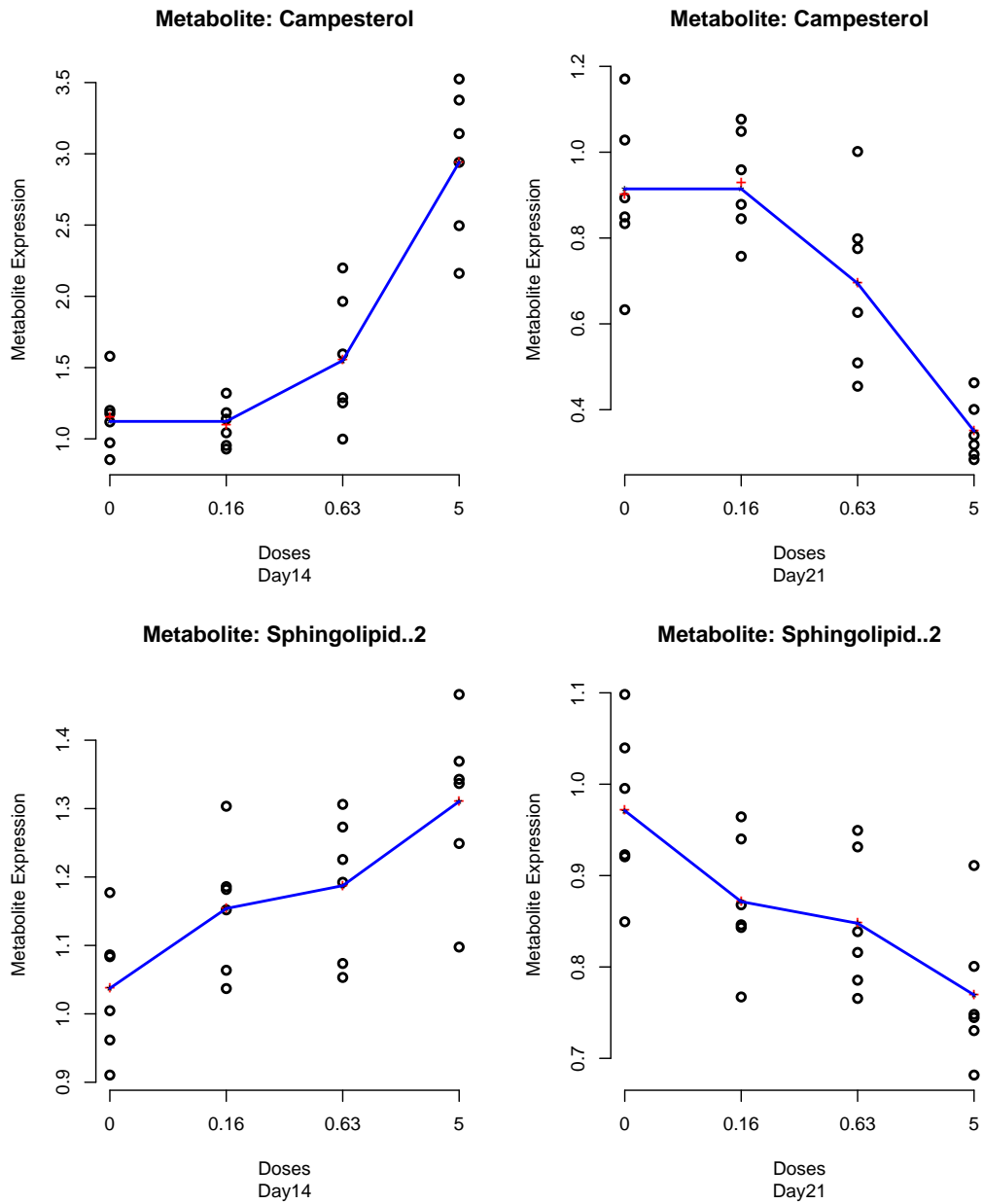


Figure 6.1: plot of the common metabolites Campesterol and Spingolipid2 in day14 and day21 respectively

6.1.2 Time 2

The table below table 6.3 displays the metabolites for which the dose effect was significant at the second time point in both days.

	Day 14		Day 21	
	Metabolite	Trend	Metabolite	Trend
1	Campesterol 33	Up	Campesterol 33	Dn
2	Sphingolipid..2 56	Up		
3	Sphingolipid..3 59	Up		
4	Sphingolipid..4 60	Up		
5	Unknown.59600010. 111	Dn		

Table 6.3: The significant metabolites at the second time point

From this table i.e. table 6.3, we see that in the second time point the number of metabolites for which the dose was significant was larger in day 14 in comparison to day 21 where we got only one metabolite, this metabolite was also common to both days with also a different trend.

From the figure 6.2 we see clearly that the trend was the same as in the previous time point i.e. the trend is increasing, from the figure 6.3 we can see that the trend was different in both days, since it was changed from an increasing trend in day14 up to a decreasing trend in dya21.

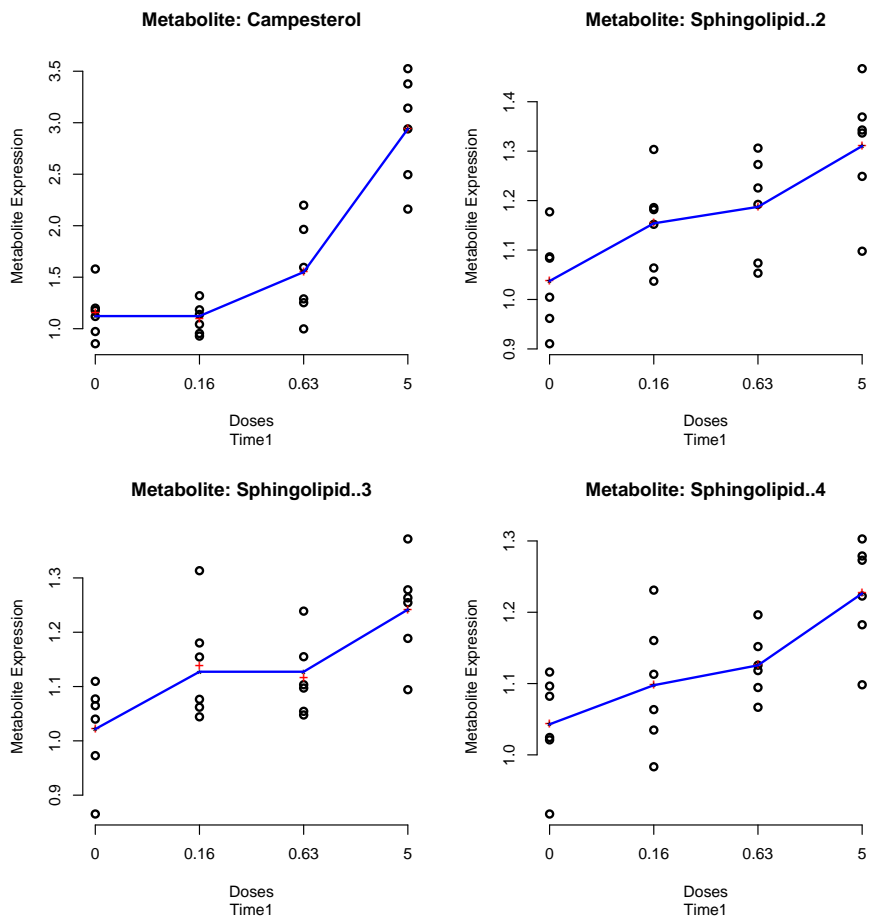


Figure 6.2: plot of the first four metabolites in day14 with up trend in the first time point

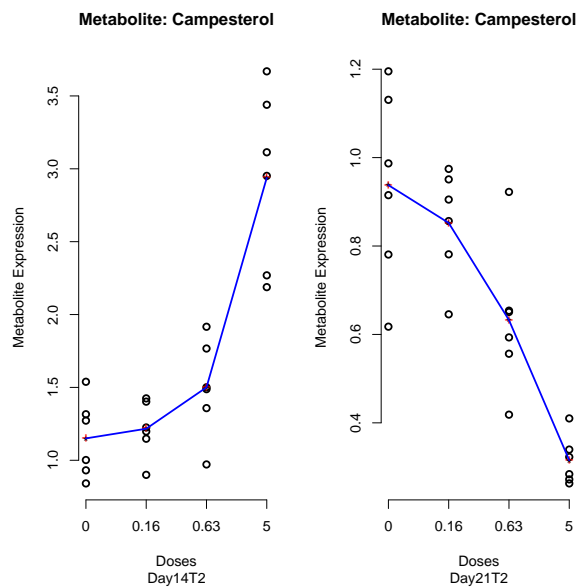


Figure 6.3: plot of the metabolite Campesterol in both days for the second time point

6.1.3 Time 3:

The table below table ?? displays the metabolites for which the dose effect was significant at the third time point in both days.

	Day 14		Day 21	
	Metabolite	Trend	Metabolite	Trend
1	Campesterol 33	Up	Campesterol 33	Dn
4	Sphingolipid..4 60	Up	Unknown..59600164 130	Dn
3	Dihydrocholesterol 67	Up		

Table 6.4: The significant metabolites at the second time point

As we can see that in day14 the metabolites *Campesterol* and *Sphingolipid.4* are common to all time points, while in day21 only *Campesterol* was the only common metabolites in this day. For the isotonic plot of the *Dihydrocholesterol* and *Unknown.59600164* metabolites see appendix figure 8.2.

the left panel in the figure 6.4 are the isotonic plot for the metabolites in day14 while the right is for the day21. We can see that the metabolite *Campesterol* has almost approximate profile at each time point for both days, while for the metabolite *Dihydrocholesterol* the isotonic means were lower for the third time point for all dose group in comparison to the first time point where it was the same.

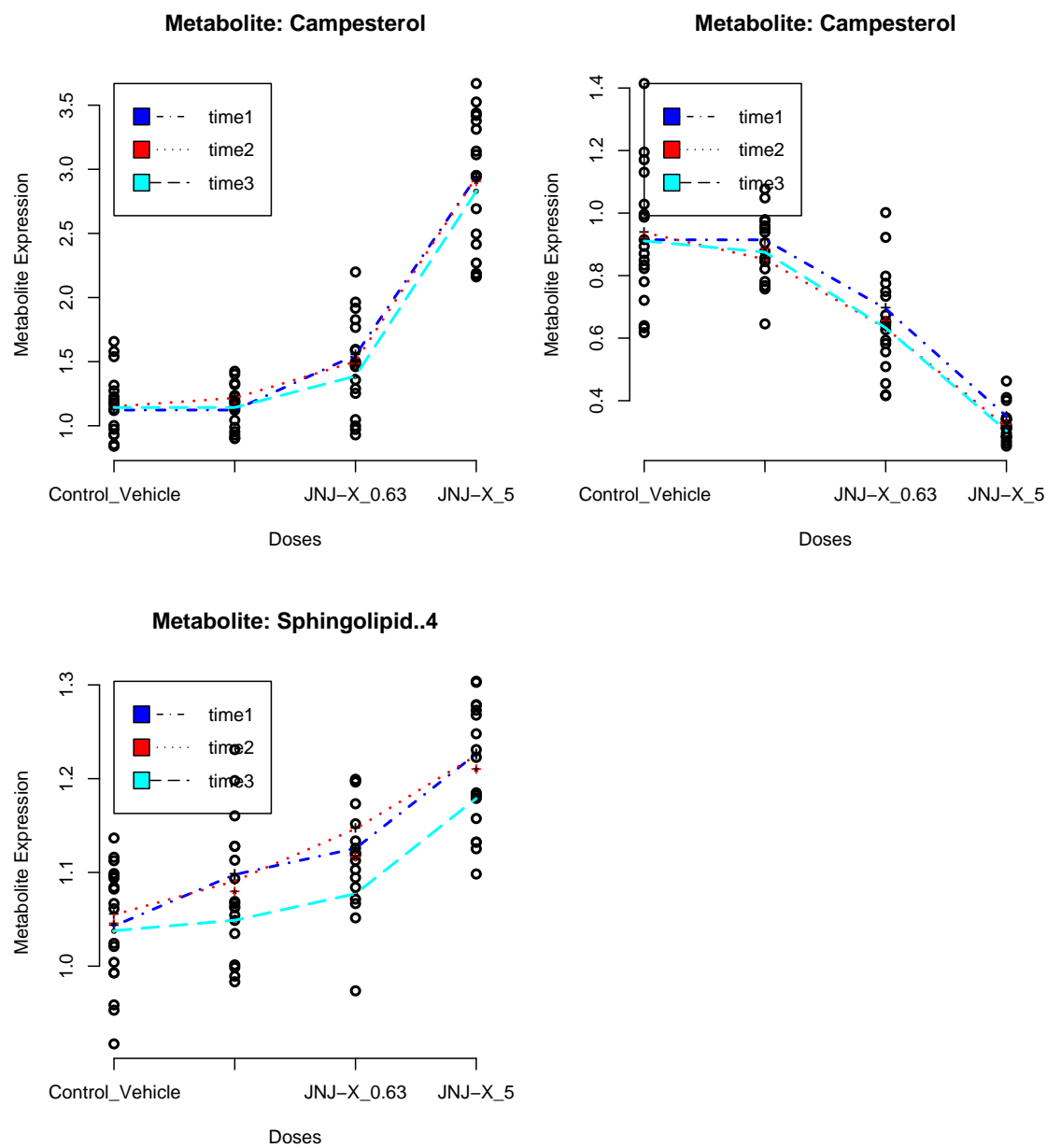


Figure 6.4: plot of the metabolites Campesterol and Sphingolipid.4 for all time points, the left panel is from day14 while the right panel is from day21.

6.2 LRT

In this section we present the results of the likelihood ratio tests in the case of selecting between two models. So in this section we compare between the following three models:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (6.2)$$

$$y_{ijl} = \mu + \alpha_i + \beta_j + \epsilon_{ijl} \quad (6.3)$$

$$y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijl} \quad (6.4)$$

where α is the main dose effect, β the main time effect, γ is the interaction between the dose and time effects and the error term ϵ is normally distributed with mean zero and constant variance.

6.2.1 Day 14:

The table below shows the metabolites in day 14 with their adjusted p-values for which the model (6.3) was significant in comparison to the model (6.2). From this table 6.5, we see

Day 14 Down		Day 14 Up	
ind14dn	padj.14dn	ind14up	padj.14up
16	1,45E-05	33	8,77E-15
126	0,001765	60	1,2E-05
71	0,01012	56	1,36E-05
117	0,017533	59	2,18E-05
		32	2,43E-05

Table 6.5: The top 5 significant metabolites for which the model 2 (6.3) was significant

that the model (6.3) was significant for many metabolites with an up trend while for the down trend we got only four metabolites see table 8.1 in appendix for full table.

The figure fig 6.5 displays the isotonic plot of the first two metabolites with the down trend while the figure fig 6.6 is for the increasing trend.

While for the comparison between the saturated model ,i.e. the model including interaction term between the dose groups and time points (model (6.4)), and the model without interaction term i.e. model (6.3), the smallest p value was almost equal to 1 in both trend after adjusting for their p-values to correct for multiplicity, so leading to the fact that the saturated model was not significant for any metabolites.

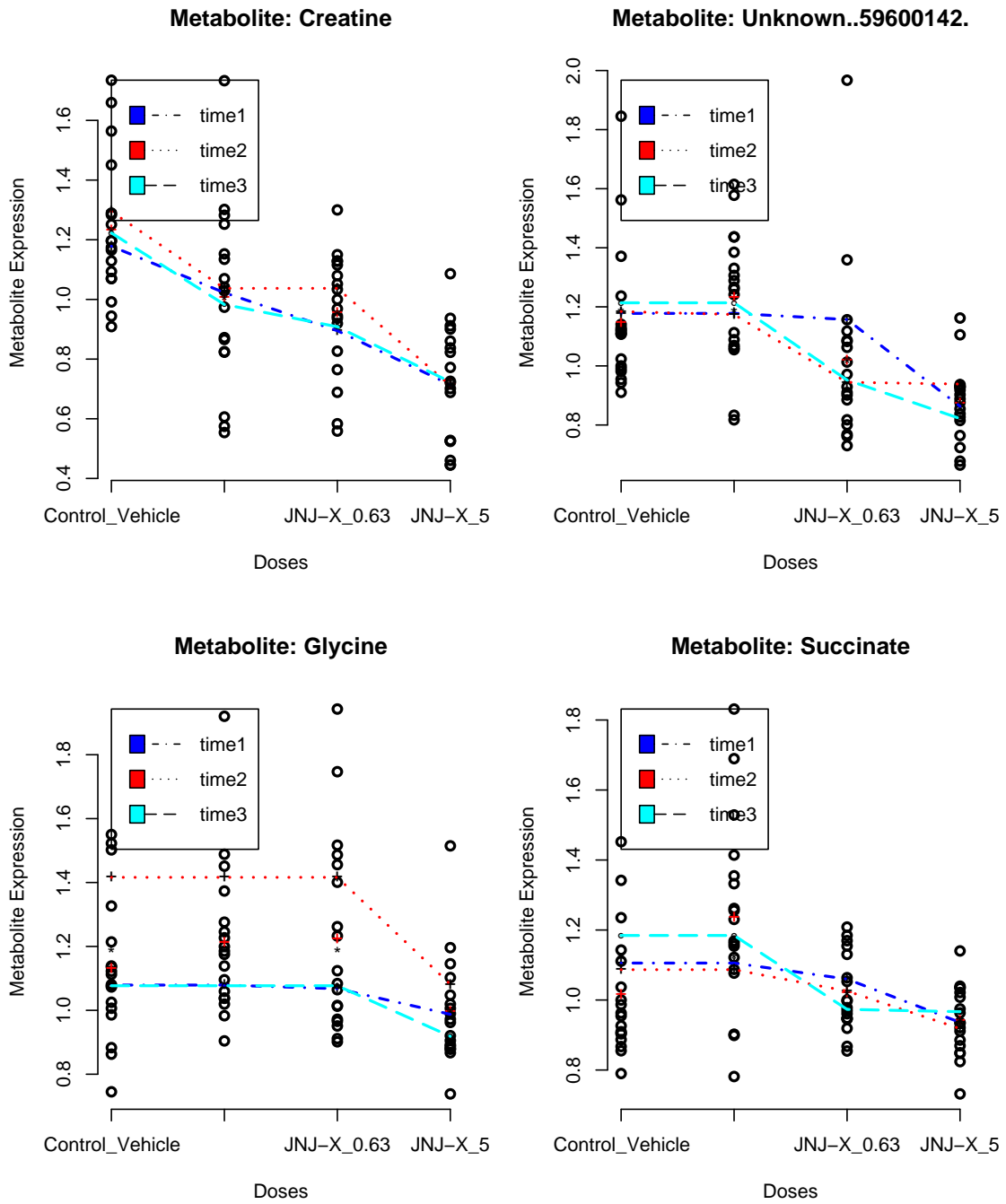


Figure 6.5: plot of the metabolite where the time effect was also significant in day14

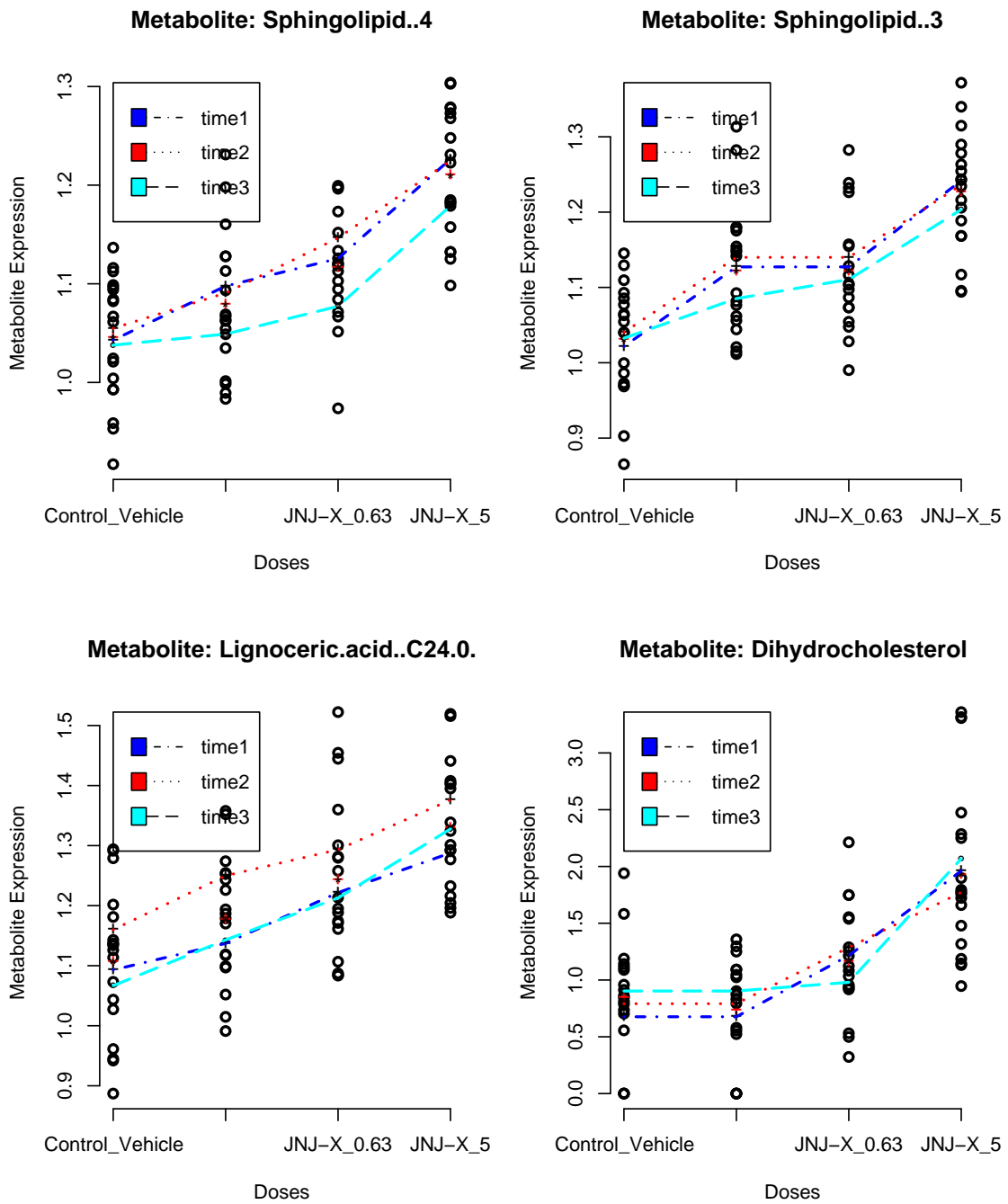


Figure 6.6: plot of the metabolite where the time effect was also significant in day14

6.2.2 Day 21:

The table below table 6.6 shows the metabolites in day21 with their adjusted p-values for which the model (6.3) was significant in comparison to the model (6.2).

Day 21 Down Trend		Day 21 Up Trend	
ind21dn	padj.21dn	ind21up	padj.21up
92	0	77	3,12E-13
73	1,67E-12	14	5,88E-09
65	2,7E-06	105	7,19E-08
209	0,00122	88	1,74E-05
208	0,001231	110	5,44E-05
45	0,001797	1	6,26E-05
50	0,002539	93	0,000978
36	0,003884	94	0,001232
		81	0,011608
		16	0,015913
		8	0,023627
		76	0,042156
		134	0,044896
		95	0,049962

Table 6.6: The significant metabolites for which the model (6.3) was significant

for the likelihood ratio test in comparison between the saturated model (6.4) and the model (6.3) it was also observed that in this case before adjusting the pvalues we got two metabolites where their pvalues were highly significant (*see table 6.7*), but after adjusting it none of them was significant.

Day 21		
Metabolite index	lrt	pvalue
Behenic.acid..C22. 45	17,8698	0,006566
Elaidic.acid..C18.trans.9.1 65	15,64099	0,015816
Unknown..59600146. 130	11,81402	0,066249
Eicosanoic.acid..C20. 36	11,43394	0,075856

Table 6.7: The significant metabolites for which the saturated model was significant before the adjustment of the pvalues

The figure below fig 6.9 displays the isotonic plot of the first two metabolites *Behenic.acid.C22.* and *Elaidic.acid.C18.trans.9.1*, we can see clearly that the interaction effects is significant since the differences in the isotonic means within each dose group for the second time point and the first time point was really significant.

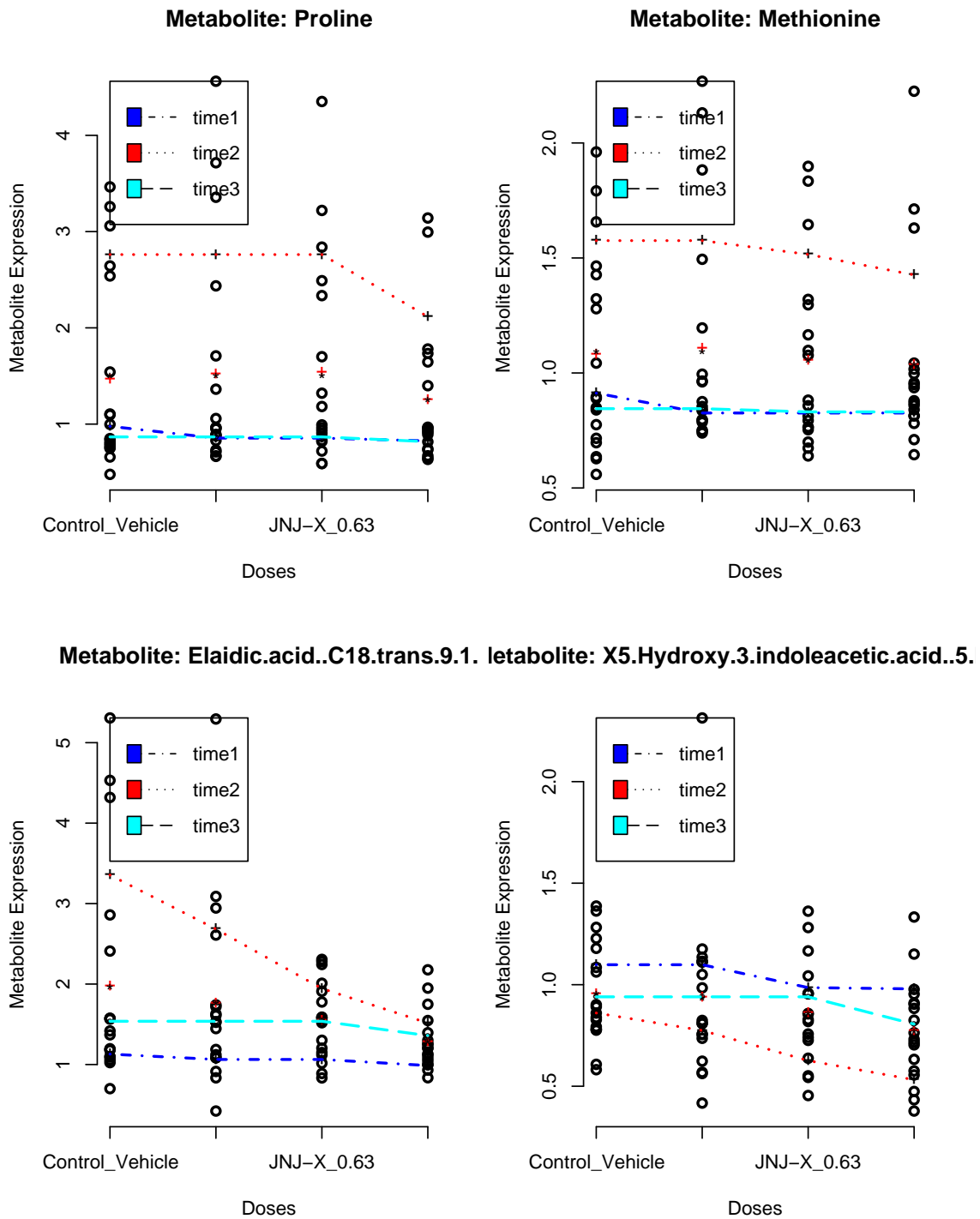


Figure 6.7: plot of the first four metabolites where the time effect was also significant in day21 down

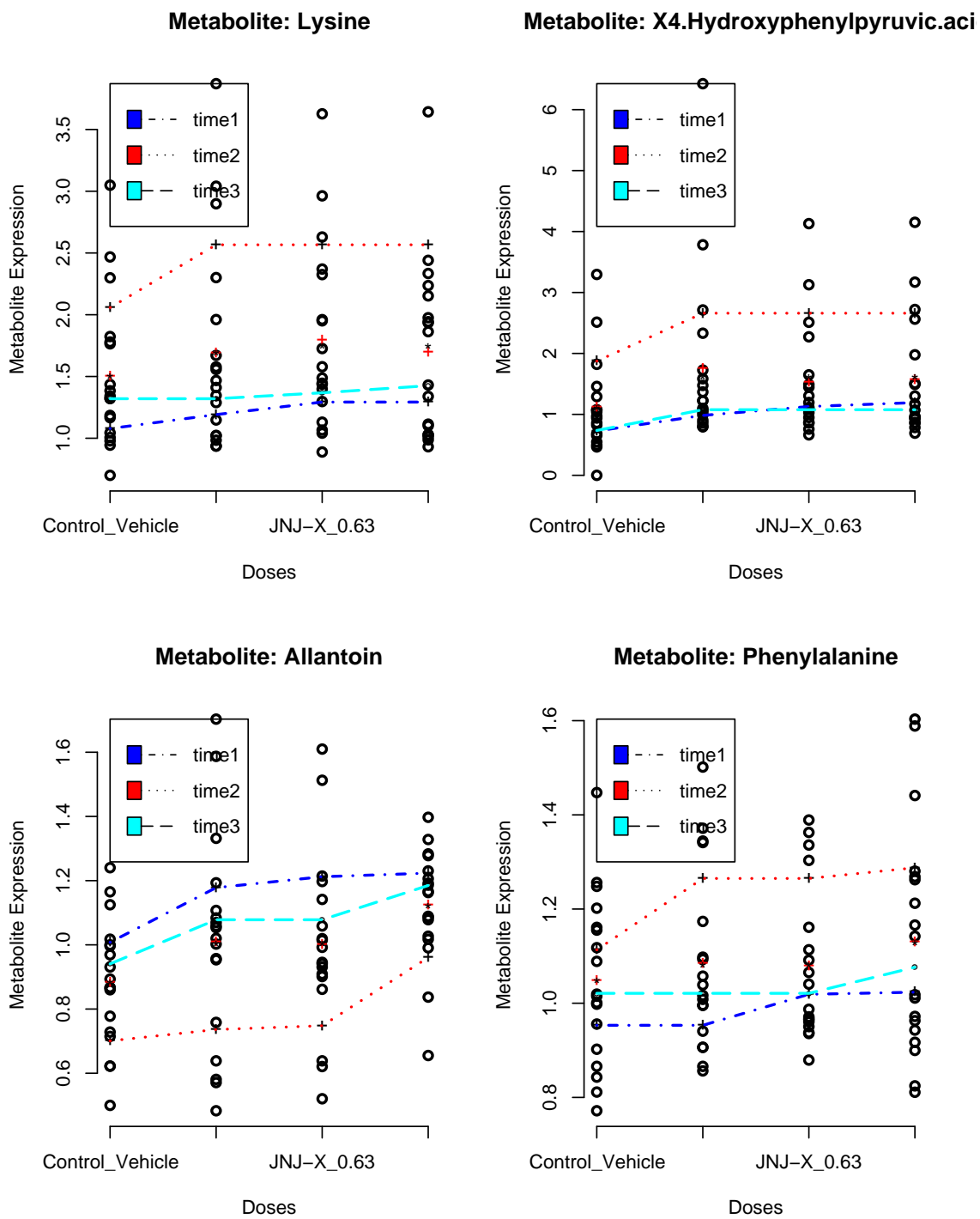


Figure 6.8: plot of the first four metabolites where the time effect was also significant in day21 up

Metabolite: Behenic.acid..C22.0.

Metabolite: Elaidic.acid..C18.trans.9

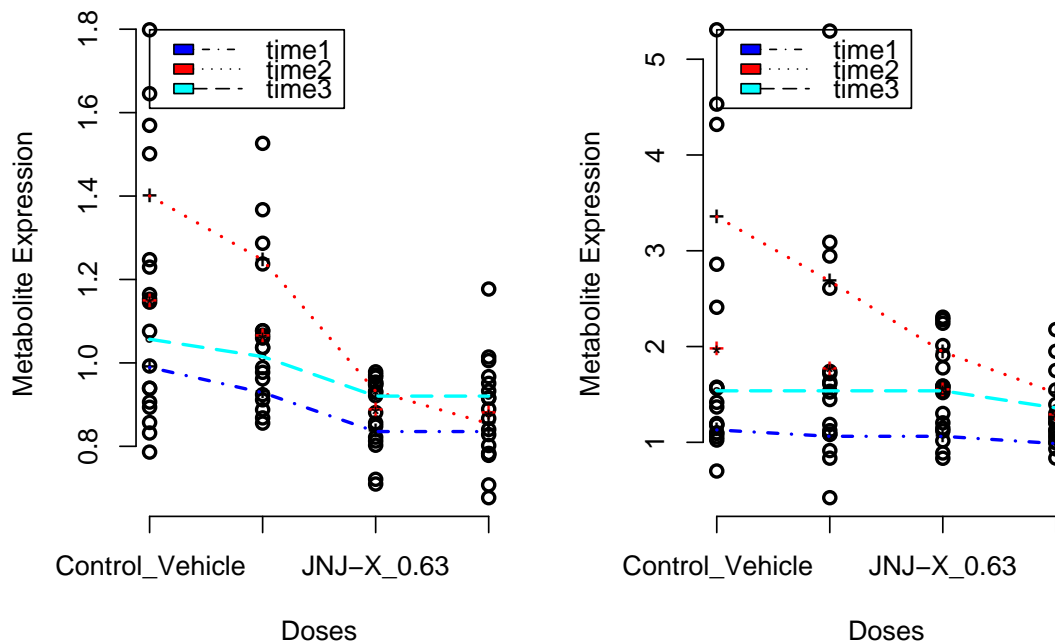


Figure 6.9: plot of the first two metabolites where the interaction effect was also significant in day21 down but without correcting for multiplicity

from this figure fig 6.9 we can see that the differences in the isotonic means between the second and third time point decrease with a decreasing dose levels, but the fact that the LRT did not find any metabolites for which the saturated model was significant can be also explained by the fact that only 2 out of 81 metabolites with a decreasing trend in all time point was significant before correcting for multiplicity.

6.3 Interaction:

In this section we will present the results from the work of Hirotsu, he present three different tests in order to test the interaction effect in case when we have an ordered alternative. The table below table 6.8 displays the metabolites for which the interaction term was significant based on at least one test of the Hirotsu tests for day14, although in both datasets we observed that the first test statistics was more sensitive in the sense that we were able to get a larger number of metabolites.

Hirotsu Tests for Day 14			
Metabolite index	S_{AxB}^1	S_{AxB}^{21}	S_{AxB}^{22}
175	0,010514	0,018473	0,018473
38	0,013069	0,039842	0,039842
133	0,018955	0,052668	0,052668
131	0,029222		
224	0,029395		
199	0,048749		
137	0,060341		

Table 6.8: The significant metabolites for which the interaction term was significant based on the Hirotsu tests

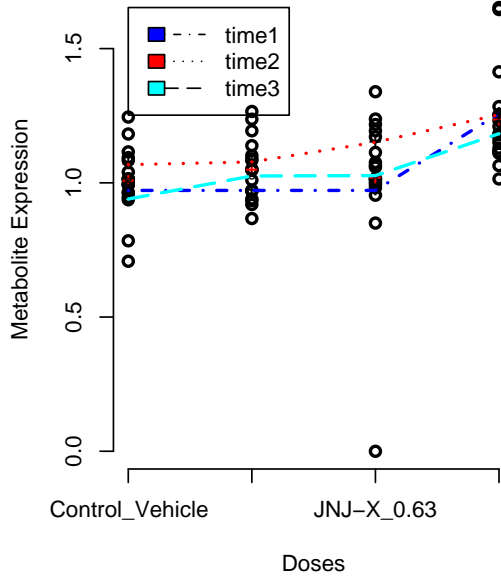
The following table displays the top 10 of the metabolites for which the Hirotsu tests was significant in day21. It was observed that in day21 many metabolites was significant for the interaction term, which can be interpreted as : the effect of the drug was also sensitive to the time, for the full table see table 6.9.

Hirotsu Tests Day 21			
Metabolite index	S_{AxB}^1	S_{AxB}^{21}	S_{AxB}^{22}
130	9,78E-05	0,000307	0,000307
174	0,004073	0,014511	0,014511
16	0,004436	0,012932	0,012932
40	0,00464	0,02017	0,02017
46	0,00464	0,015796	0,015796
140	0,008851	0,040982	0,040982
128	0,016416	0,061942	0,061942
112	0,020257	0,03198	0,03198
27	0,020631	0,049873	0,049873
187	0,025504	0,036099	0,036099

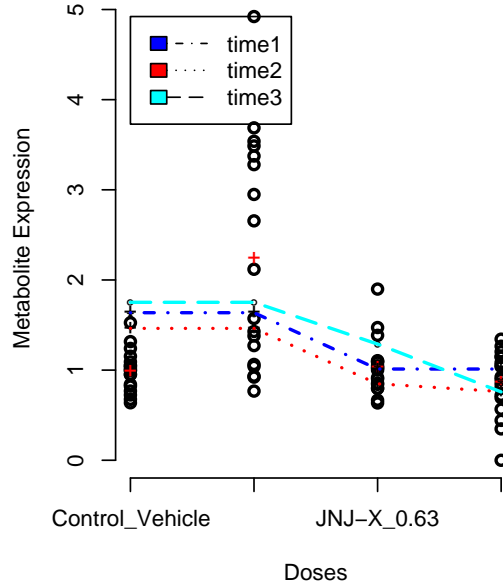
Table 6.9: The Top 10 significant metabolites for which the interaction term was significant based on the Hirotsu tests

We observed that in day21 we got more metabolites for which the Hirotsu test was significant, leading to the significance of the interaction term while in day 14 we got less metabolites. we observed also that the Hirotsu tests were able to find the two metabolites in day 21 for which their p-values was significant before the correcting for multiplicity.

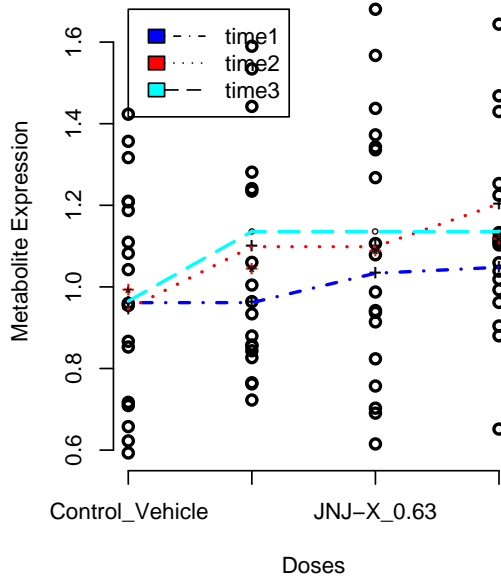
Metabolite: Phosphate..lipid..Phospholipid



Metabolite: Unknown..59600147.



Metabolite: Unknown..59600150.



Metabolite: Unknown..59600154.

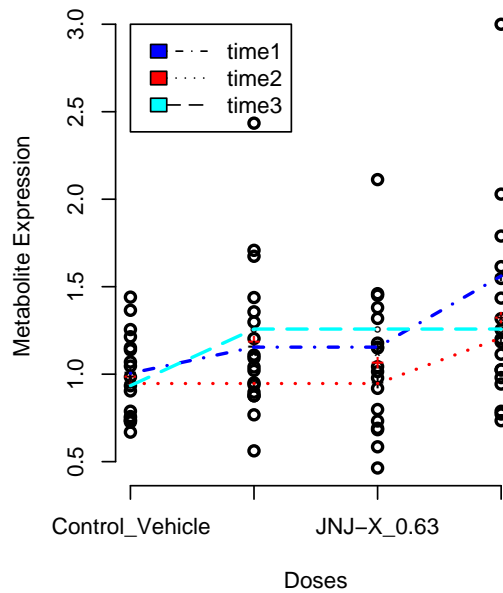
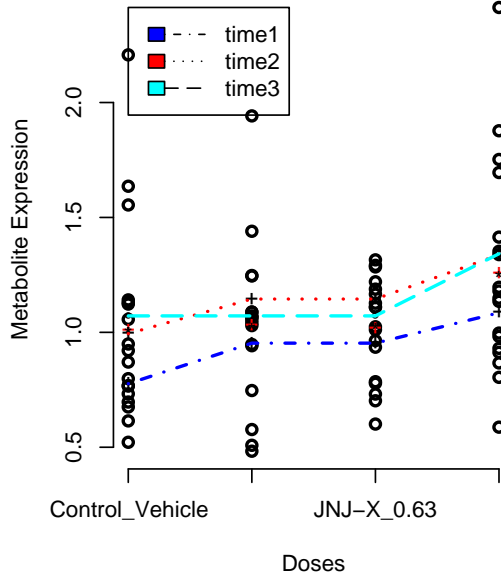
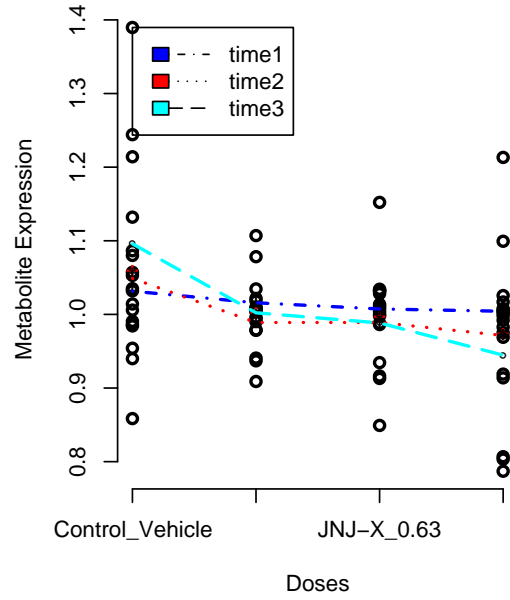


Figure 6.10: plot of the metabolites where the interaction effect was also significant in day14

Metabolite: Ceramide..d18.1.C24.1..d18.2.C2



Metabolite: Unknown..69600052.



Metabolite: Progesterone

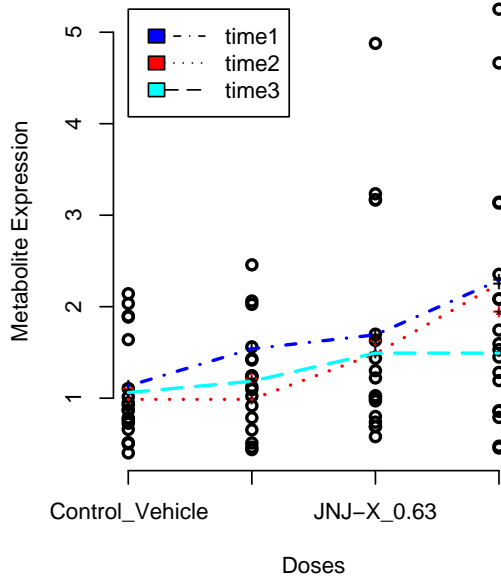


Figure 6.11: plot of the first four metabolites where the time effect was also significant in day14

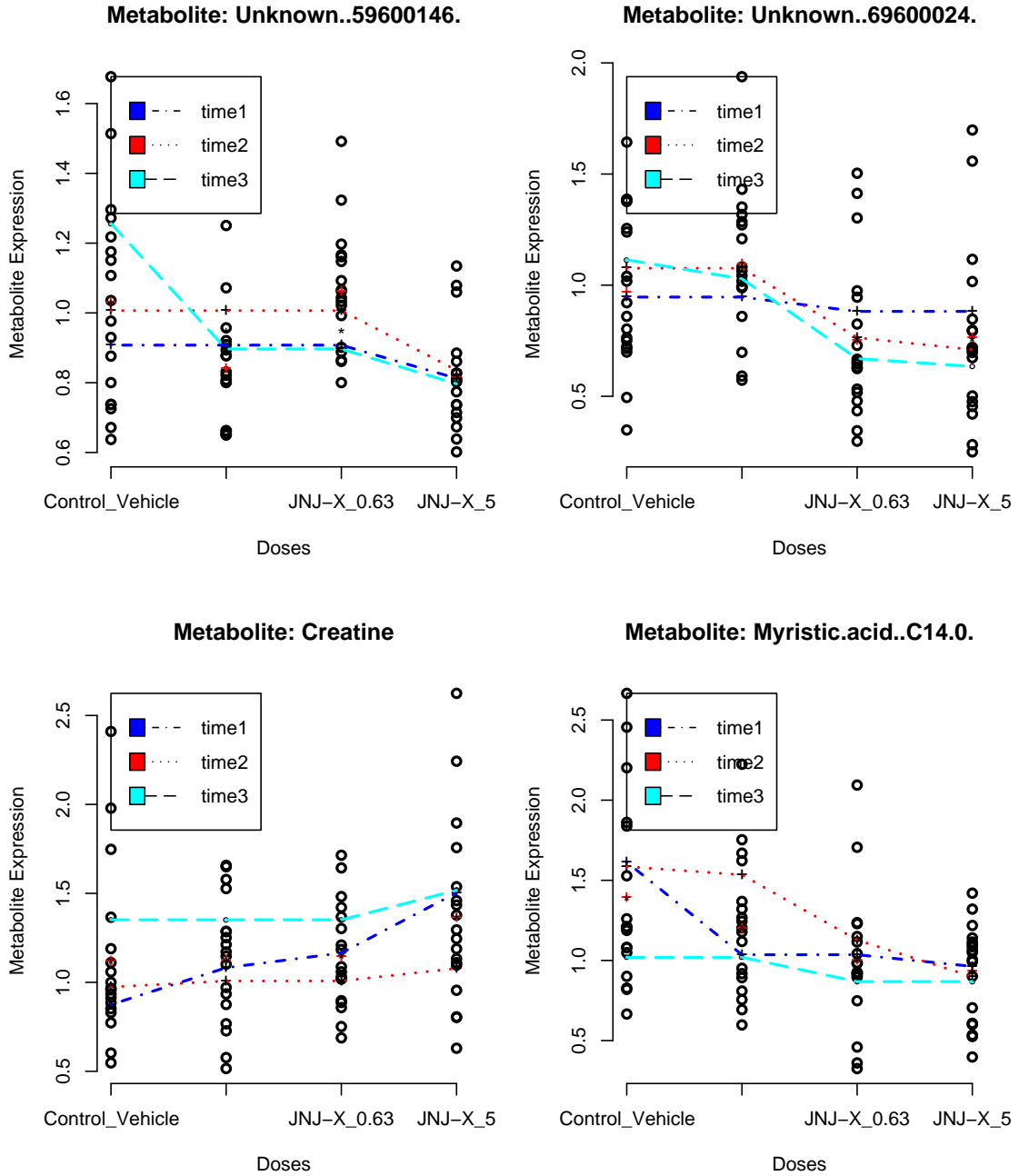


Figure 6.12: plot of the first four metabolites where the time effect was also significant in day21

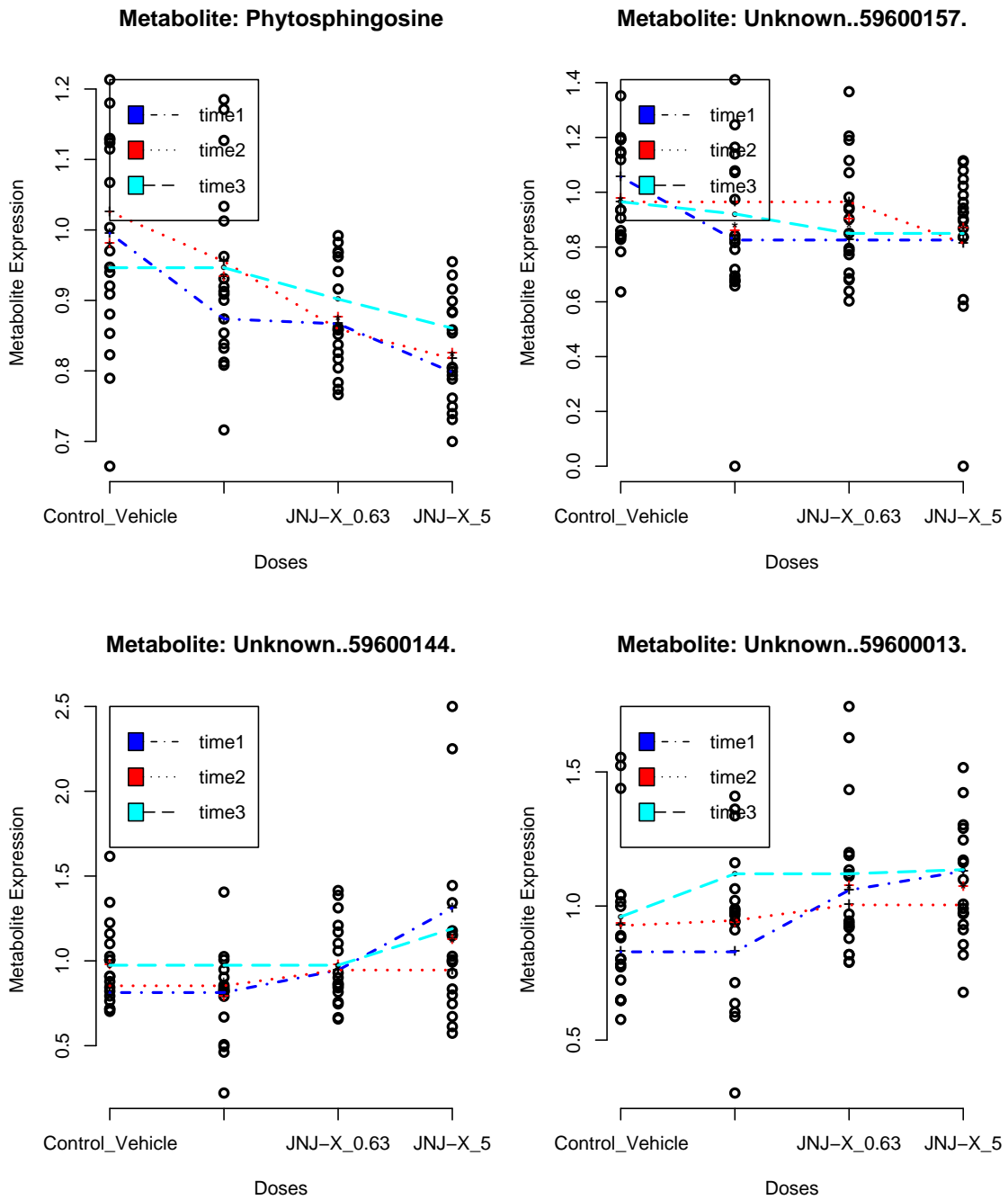


Figure 6.13: plot of the first four metabolites where the time effect was also significant in day21

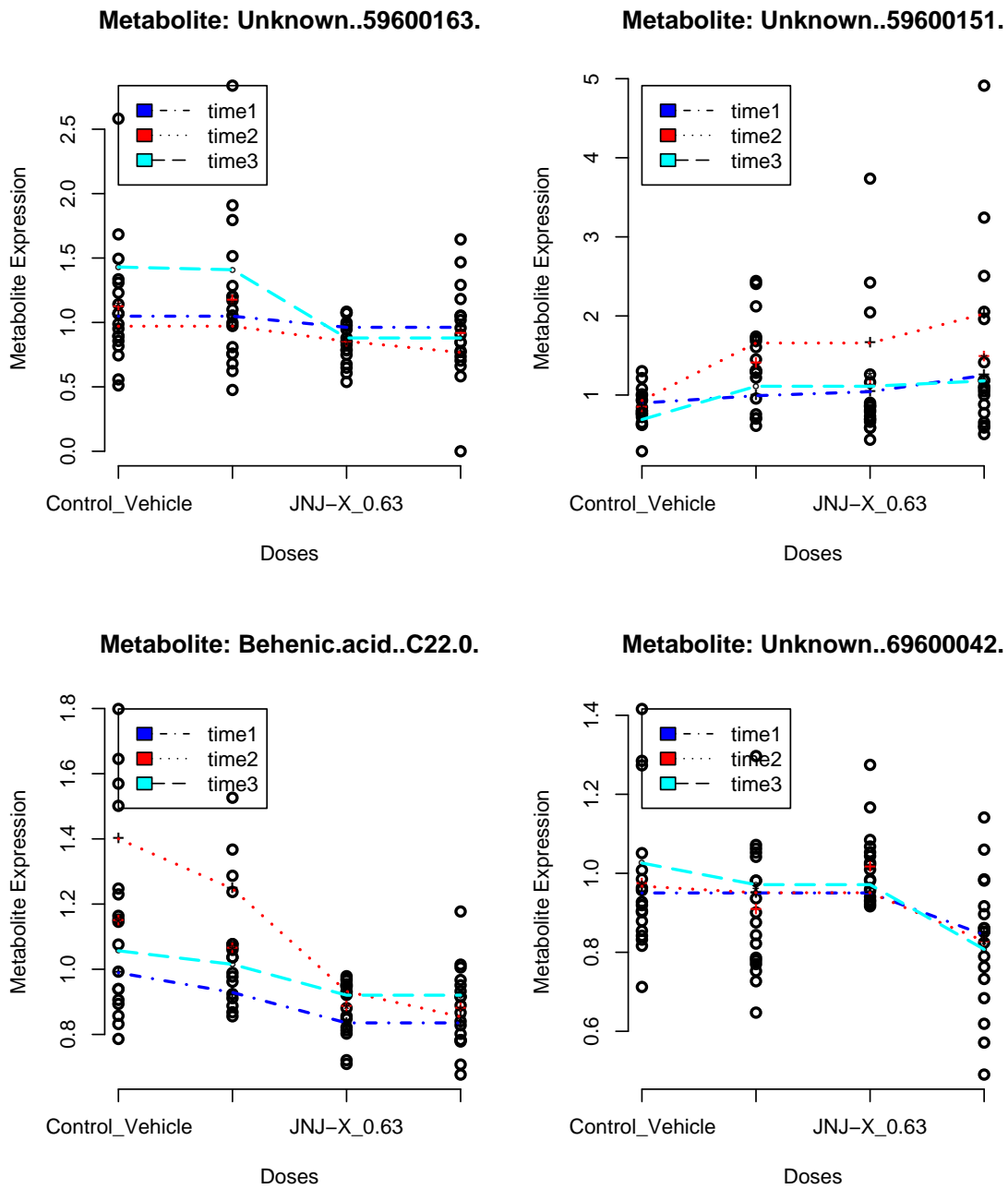


Figure 6.14: plot of the first four metabolites where the time effect was also significant in day21

Chapter 7

Conclusion

Dose-response studies are gaining a large interest in many research areas, but in such studies we are more interested in the trend of the relationship between the response variable \hat{Y} and the covariates X . Modeling the interaction term is not straightforward in a two-way layout since the ordered alternative depends on the practical study in the sense of having the ordering in both factors dose and times or only in one factor [14]. For the homogeneity of means i.e. no dose effect, in a two-way layout, it can be done by means of the likelihood ratio test like in one-way Anova expect that the variance under the alternative in this case is equal to (4.6). For the interaction term, it is quite complicated since the ordered alternative depends on the practical study, so for the interaction term we first use the likelihood ratio test to compare between a saturated model, i.e. model including interaction between the dose and time effects, but the results of the likelihood ratio after correcting for multiple testing showed a none interaction effects for all metabolites even without correcting for multiplicity the saturated model (6.4) on page 27 was significant for only two metabolites in dataset day21(see table 6.7 on page 30).

The tests proposed by Hirotsu for testing the interaction term under an ordered alternative, were able for finding metabolites for which the interaction was significant, so using these tests we were able to find metabolites for which increasing the level of the time, the differences of the means: $\mu_{ij} - \mu_{i'j}$ increase or decrease for $i' \leq i$, although without using the isotonic means and partitioning the sum of squares of the interaction. It was also observed that the first test statistic of Hirotsu S_{AXB}^1 was able to detect a larger number of metabolites for which the interaction term was significant, in comparison to the remain two tests $S_{AXB}^{2.1}$ and $S_{AXB}^{2.2}$ see table 6.8 on page 34 and 8.2 in appendix. It was also observed that in our datasets the two test statistics $S_{AXB}^{2.1}$ and $S_{AXB}^{2.2}$ find the same metabolites with respectively the same p-values. For further research, one can try to use the resampling based method to get the distribution of the likelihood ratio test statistic in case of comparing between two models, For the Hirotsu tests one can try to replace the arithmetic means for the dose groups by its isotonic means by means of GPAV algorithm [10], and see if the Hirotsu tests became more sensitive to the interaction effect or not.

Bibliography

- [1] Bacchetti, P. (1989). Additive isotonic models. *Journal. Amer. Statist. Assoc* 84 289-294.
- [2] D.J. Bartholomew (1959a). A test of homogeneity for ordered alternatives. *Biometrika* 46, 36-48.
- [3] D.J. Bartholomew (1959b). A test of homogeneity for ordered alternativesII. *Biometrika* 46.
- [4] Barlow R.E, Bartholomew D.J, Bremner J.M, and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*. Wiley, New York, 1972.
- [5] Dykstra R. and Robertson T. (1982). An algorithm for isotonic regression for two or more independent variables. *The Annals of Statistics* 10, 708 -716.
- [6] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall: London, 1990.
- [7] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *L^AT_EX: The Elements of Statistical Learning*. Springer-Verlag, New York.
- [8] E. Mammen and K. Yu,2007. *Additive Isotonic Regression*.
- [9] T. Morton-Jones, P. Diggle, L. Parker, HO. Dickinson, K. Binks(2000). Additive Isotonic Regression Models in Epidemiology.*Statistics in Medicine*,19(6), Pages 849 - 859.
- [10] O. Burdakov, A. Grimvall, and M. Hussian. A generalised PAV algorithm for monotonic regression in several variables. Proceedings of COMPSTAT, 23-27 August, 2004, Prague, 2004a.
- [11] D. Lin, Z. Shkedy, D. Yekutieli, T. Burzykowski, H. Göhlmann, A. De Bondt, T. Perera, T. Geerts and L. Bijmens(2007). Testing for Trends in Dose-Response Microarray Experiments: A Comparison of Several Testing Procedures, Multiplicity and Resampling-Based Inference. *Statistical Applications in Genetics and Molecular Biology* , 6(1), Article 26.
- [12] S. Pramana, D. Lin,P. Haldermans,Z. Shkedy and T. Verbeke. IsoGene:An R Package for Analyzing Dose-response Studies in Microarray Experiments. *Journal of Statistical Software*
- [13] J de Leeuw, K Hornik and P Mair. Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software*.

- [14] Hirotsu, C. (1978). Ordered Alternatives for Interaction effects. *Biometrika*, 65(3),561-570.
- [15] Robertson, T., Wright, F.T. and R.L. Dykstra: Order restricted statistical inference, December 1989, Vol. 30 Issue: Number 1 p316-316.

Chapter 8

Appendix

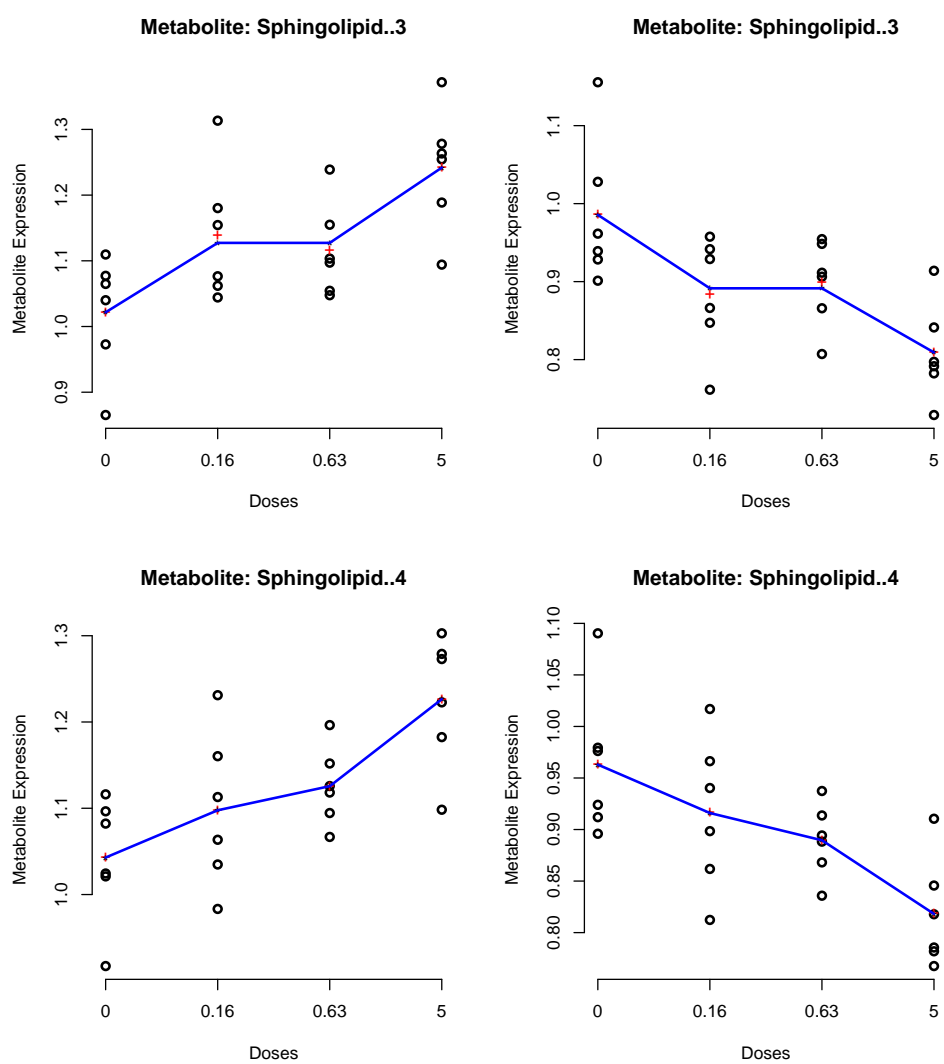


Figure 8.1: plot of the first four metabolites where the time effect was also significant in day21

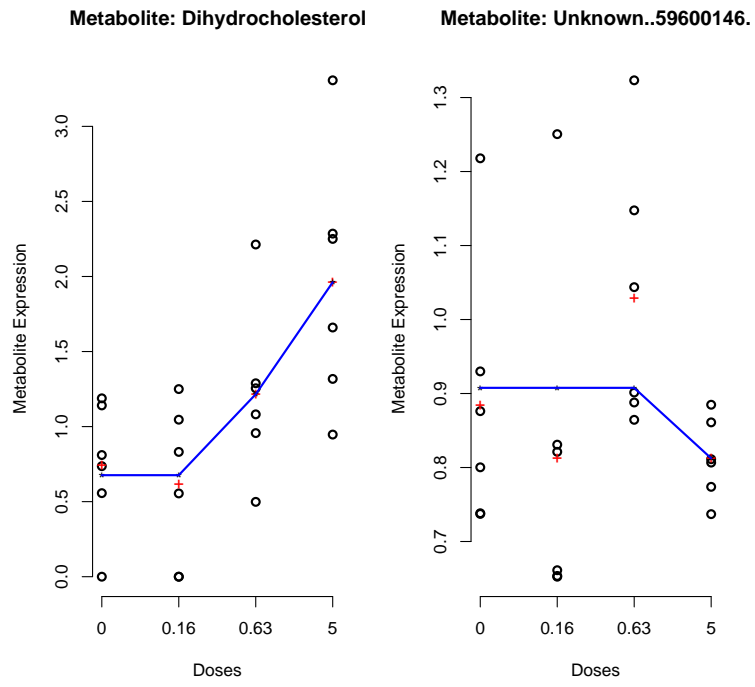


Figure 8.2: plot of the first four metabolites where the time effect was also significant in day21

Day 14 Down		Day 14 Up	
ind14dn	padj.14dn	ind14up	padj.14up
16	1,45E-05	33	8,77E-15
126	0,001765	60	1,2E-05
71	0,01012	56	1,36E-05
117	0,017533	59	2,18E-05
		32	2,43E-05
		67	2,44E-05
		53	0,000488
		58	0,000488
		187	0,000636
		28	0,000828
		46	0,000828
		20	0,001041
		36	0,001041
		63	0,001218
		177	0,001377
		65	0,003374
		41	0,004054
		45	0,004054
		54	0,004054
		35	0,004456
		153	0,00919
		39	0,010729
		22	0,011447
		49	0,011533
		38	0,013247
		210	0,01337
		31	0,016814
		23	0,017651
		40	0,01815
		51	0,023808
		19	0,034457
		208	0,036593
		18	0,039965
		218	0,039965
		74	0,040979
		62	0,045503
		79	0,049124

Table 8.1: The significant metabolites for which the model 2 eq3 was significant

ind14	padj.1	padj.21	padj.22
130	9,78E-05	0,000307	0,000307
131	9,78E-05	0,000307	0,000307
212	9,78E-05	0,000307	0,000307
69	0,000135	0,000307	0,000307
177	0,001404		
123	0,004063	0,01637	0,01637
174	0,004073	0,014511	0,014511
16	0,004436	0,012932	0,012932
40	0,00464	0,02017	0,02017
42	0,00464	0,014511	0,014511
46	0,00464	0,015796	0,015796
147	0,00464	0,02017	0,02017
117	0,007757	0,025933	0,025933
140	0,008851	0,040982	0,040982
129	0,011986	0,020894	0,020894
128	0,016416		
17	0,01771	0,026373	0,026373
112	0,020257	0,03198	0,03198
18	0,020631		
27	0,020631	0,049873	0,049873
99	0,020631	0,026907	0,026907
114	0,020631	0,03198	0,03198
199	0,022866	0,046603	0,046603
172	0,024097		
80	0,025504	0,045422	0,045422
187	0,025504	0,036099	0,036099
152	0,028819	0,046603	0,046603
3	0,029038		
202	0,029038		
62	0,029122	0,046603	0,046603
144	0,029122	0,046603	0,046603
134	0,033444		
45	0,036939		
66	0,036939	0,049428	0,049428
189	0,037329	0,046603	0,046603
173	0,044707		

Table 8.2: The significant metabolites based on the Hirotzu tests in day21

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:

Analysis of Dose-response OMICS Experiment in a Two-Way Analysis of Variance Under Ordered Constraint

Richting: **Master of Statistics-Bioinformatics**

Jaar: **2010**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

El Aalamat, Yousef

Datum: **1/02/2010**