On MANOVA using STATA, SAS & R

Fares Qeadan, Ph.D

Department of Internal Medicine

Division of Epidemiology, Biostatistics, & Preventive Medicine
University of New Mexico Health Sciences Center

July 13, 2015

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R
- Note on Profile Analysis

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R
- Note on Profile Analysis

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R
- Note on Profile Analysis
- Citation

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R
- Note on Profile Analysis
- Citation
- References

- Introduction
  - What is MANOVA?
  - Why is MANOVA?
  - Functional Form & Notations
  - Assumptions
  - Hypotheses
  - Remarks
- One-way MANOVA Example
  - The Problem
  - The Solution using STATA
  - The Solution using SAS
  - The Solution using R
- Note on Profile Analysis
- Citation
- References
- Links to Data and Syntax

What is MANOVA:

- Multivariate analysis of variance (MANOVA) is a statistical test for comparing multivariate means of several groups.

What is MANOVA:

- Multivariate analysis of variance (MANOVA) is a statistical test for comparing multivariate means of several groups.
- MANOVA is an extension of ANOVA such that main effects and interactions are assessed on a linear combination of a set of two or more continuous dependent variables (DVs) [1]. Think of it as ANOVA for situations when there are several continuous dependent variables.

What is MANOVA:

- Multivariate analysis of variance (MANOVA) is a statistical test for comparing multivariate means of several groups.
- MANOVA is an extension of ANOVA such that main effects and interactions are assessed on a linear combination of a set of two or more continuous dependent variables (DVs) [1]. Think of it as ANOVA for situations when there are several continuous dependent variables.
- MANOVA searches for the best linear combinations of the dependent variables, for directions in the data space, which maximizes group separation (i.e. the ratio of between-group and within-group variances) [4].

What is MANOVA:

- Multivariate analysis of variance (MANOVA) is a statistical test for comparing multivariate means of several groups.
- MANOVA is an extension of ANOVA such that main effects and interactions are assessed on a linear combination of a set of two or more continuous dependent variables (DVs) [1]. Think of it as ANOVA for situations when there are several continuous dependent variables.
- MANOVA searches for the best linear combinations of the dependent variables, for directions in the data space, which maximizes group separation (i.e. the ratio of between-group and within-group variances) [4].
- MANOVA is a two-stage test in which an overall test is first performed with subsequent tests to tease apart group differences [5].

Why is MANOVA:

- Researchers are interested in evaluating mean differences on several dependent variables simultaneously while controlling for the intercorrelations among them [2].

Why is MANOVA:

- Researchers are interested in evaluating mean differences on several dependent variables simultaneously while controlling for the intercorrelations among them [2].
- Statistically, with correlated DVs, MANOVA is a more powerful test than conducting separate ANOVAs [6] (conducting a series of ANOVAs inflates type I error rates while MANOVA helps to control for it).

Why is MANOVA:

- Researchers are interested in evaluating mean differences on several dependent variables simultaneously while controlling for the intercorrelations among them [2].
- Statistically, with correlated DVs, MANOVA is a more powerful test than conducting separate ANOVAs [6] (conducting a series of ANOVAs inflates type I error rates while MANOVA helps to control for it).
- MANOVA allows for more examinations of group differences than is the case for ANOVA (see Hypotheses section)[3].

Why is MANOVA:

- Researchers are interested in evaluating mean differences on several dependent variables simultaneously while controlling for the intercorrelations among them [2].
- Statistically, with correlated DVs, MANOVA is a more powerful test than conducting separate ANOVAs [6] (conducting a series of ANOVAs inflates type I error rates while MANOVA helps to control for it).
- MANOVA allows for more examinations of group differences than is the case for ANOVA (see Hypotheses section)[3].
- MANOVA utilizes more information from the data, using the relationship between the DVs, than does ANOVA [5].

Why is MANOVA:

- Researchers are interested in evaluating mean differences on several dependent variables simultaneously while controlling for the intercorrelations among them [2].
- Statistically, with correlated DVs, MANOVA is a more powerful test than conducting separate ANOVAs [6] (conducting a series of ANOVAs inflates type I error rates while MANOVA helps to control for it).
- MANOVA allows for more examinations of group differences than is the case for ANOVA (see Hypotheses section)[3].
- MANOVA utilizes more information from the data, using the relationship between the DVs, than does ANOVA [5].
- MANOVA may detect combined differences not found in the univariate tests [6].

Functional Form & Notations:

Using the notations of Johnson and Wichern [7], with slight modification, suppose we have $p > 1$ continuous dependent variables, then the one-way MANOVA model is:

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\epsilon}_{ij} \tag{1}$$

with $i = 1 \ldots g$ and $j = 1 \ldots n_i$ where:

- $\mathbf{y}_{ij}$ is a $p \times 1$ outcome vector for the j^{th} subject from the i^{th} treatment.
- $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_p]'$ is the overall population mean vector.
- $\boldsymbol{\tau}_i = [\tau_{i1}, \tau_{i2}, \ldots, \tau_{ip}]'$ is the i^{th} treatment effect vector for the p response variables.
- $\boldsymbol{\epsilon}_{ij}$ is the experimental error such that $\boldsymbol{\epsilon}_{ij} \sim N_p(0, \Sigma)$ with $\sum_{i=1}^{g} n_i \boldsymbol{\tau}_i = 0$.

In a matrix form, the equation in (1) could be written as

$$Y_{n\times p} = X_{n\times(g+1)}B_{(g+1)\times p} + \epsilon_{n\times p} \tag{2}$$

where $n = \sum_g n_g$,

$$Y = \begin{bmatrix} \mathbf{y}'_{11} \\ \mathbf{y}'_{12} \\ \vdots \\ \mathbf{y}'_{1n_1} \\ \mathbf{y}'_{21} \\ \vdots \\ \vdots \\ \mathbf{y}'_{gn_g} \end{bmatrix} = \begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1n_1 1} & y_{1n_1 2} & \cdots & y_{1n_1 p} \\ y_{211} & y_{212} & \cdots & y_{21p} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ y_{gn_g 1} & y_{gn_g 2} & \cdots & y_{gn_g p} \end{bmatrix}$$

and $X_{n\times(g+1)}B_{(g+1)\times p} + \epsilon_{n\times p}$ is

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_p \\ \tau_{11} & \tau_{12} & \cdots & \tau_{1p} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \tau_{g1} & \tau_{g2} & \cdots & \tau_{gp} \end{bmatrix} + \begin{bmatrix} \epsilon'_{11} \\ \epsilon'_{12} \\ \vdots \\ \epsilon'_{1n_1} \\ \epsilon'_{21} \\ \vdots \\ \vdots \\ \epsilon'_{gn_g} \end{bmatrix}$$

Assumptions:

- Normality assumption: The data (or residuals) are multivariate normally distributed for each group. So, each variable must be normal and any linear combinations of the variables must be normal (checked by Shaprio-Wilks for univariate normality (with QQplots) and Mardia's skewness and kurtosis for multivariate normality).

Assumptions:

- Normality assumption: The data (or residuals) are multivariate normally distributed for each group. So, each variable must be normal and any linear combinations of the variables must be normal (checked by Shaprio-Wilks for univariate normality (with QQplots) and Mardia's skewness and kurtosis for multivariate normality).
- Homogeneity assumption: The data from all groups have common variance-covariance matrix Σ (checked by Bartlett's test or Box's test).

Assumptions:

- Normality assumption: The data (or residuals) are multivariate normally distributed for each group. So, each variable must be normal and any linear combinations of the variables must be normal (checked by Shaprio-Wilks for univariate normality (with QQplots) and Mardia's skewness and kurtosis for multivariate normality).
- Homogeneity assumption: The data from all groups have common variance-covariance matrix Σ (checked by Bartlett's test or Box's test).
- The DVs are continuous.

Assumptions:

- Normality assumption: The data (or residuals) are multivariate normally distributed for each group. So, each variable must be normal and any linear combinations of the variables must be normal (checked by Shaprio-Wilks for univariate normality (with QQplots) and Mardia's skewness and kurtosis for multivariate normality).
- Homogeneity assumption: The data from all groups have common variance-covariance matrix Σ (checked by Bartlett's test or Box's test).
- The DVs are continuous.
- Linearity: There should be a linear relationships between the DVs (checked by conducting a scatterplot matrix between the DVs).

Assumptions:

- Normality assumption: The data (or residuals) are multivariate normally distributed for each group. So, each variable must be normal and any linear combinations of the variables must be normal (checked by Shaprio-Wilks for univariate normality (with QQplots) and Mardia's skewness and kurtosis for multivariate normality).
- Homogeneity assumption: The data from all groups have common variance-covariance matrix Σ (checked by Bartlett's test or Box's test).
- The DVs are continuous.
- Linearity: There should be a linear relationships between the DVs (checked by conducting a scatterplot matrix between the DVs).
- Absence of multivariate outliers (checked by assessing Mahalanobis Distances).

Hypotheses:

- Is there an overall treatment effect?

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \tag{3}$$

Hypotheses:

- Is there an overall treatment effect?

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \tag{3}$$

- Are the outcome means, regardless of the treatment groups, equal?

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p \tag{4}$$

Hypotheses:

- Is there an overall treatment effect?

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \tag{3}$$

- Are the outcome means, regardless of the treatment groups, equal?

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p \tag{4}$$

- Are the outcome means, for a set of the treatment groups, equal? (done by using Contrast). For example:

$$H_0 : \tau_1 = \tau_2 \tag{5}$$

Hypotheses:

- Is there an overall treatment effect?

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_g = 0 \tag{3}$$

- Are the outcome means, regardless of the treatment groups, equal?

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_p \tag{4}$$

- Are the outcome means, for a set of the treatment groups, equal? (done by using Contrast). For example:

$$H_0 : \tau_1 = \tau_2 \tag{5}$$

- If applicable, Profile Analysis *[Test of Parallelism, Coincidental (Separation) and Flatness (Level)]* and Post hoc Analysis are conducted.

Remarks:

- If the dependent variables are not correlated, separate ANOVAs are appropriate [9].

Remarks:

- If the dependent variables are not correlated, separate ANOVAs are appropriate [9].
- In most of the statistical programs used, when implementing MANOVA there are four multivariate measures: Wilks lambda, Pillai's trace, Hotelling-Lawley trace and Roys largest root. I will emphasize Wilks lambda since it demonstrates the amount of variance accounted for in the dependent variables by the independent variables and hence it can give a "Multivariate R-squared" calculated as:
Multivariate R-squared= 1 - Wilks' Lambda.

Remarks:

- If the dependent variables are not correlated, separate ANOVAs are appropriate [9].
- In most of the statistical programs used, when implementing MANOVA there are four multivariate measures: Wilks lambda, Pillai's trace, Hotelling-Lawley trace and Roys largest root. I will emphasize Wilks lambda since it demonstrates the amount of variance accounted for in the dependent variables by the independent variables and hence it can give a "Multivariate R-squared" calculated as:
  Multivariate R-squared= 1 - Wilks' Lambda.
- In this document we will give an example for one-way MANOVA only, however the analysis is similar in two-way MANOVA with the addition of having two independent factors instead of one and hence an interaction term.

The Problem (Example 1.5.1 of Christensen 2001 [8]): A study was conducted to examine the effects of two drugs on heart rates. Thirty women were randomly divided into three groups of ten. An injection was given to each person. Depending on their group, women received either a placebo, drug A, or drug B. Repeated measurements of their heart rates were taken beginning at two minutes after the injection and at five minute intervals thereafter. Four measurements were taken on each individual[1]. The data are given in Table 1.2.

[1] *The observations were taken over time on the same individual and hence correlated. Consider the heart rate measurements taken at the four times to be four DVs. This is a completely randomized design, so a one-way MANOVA is appropriate. The treatments are the two drugs and the placebo (R. Christensen).*

TABLE I.2. Heart Rate Data

	DRUG											
	Placebo				A				B			
TIME	1	2	3	4	1	2	3	4	1	2	3	4
SUBJECT												
1	80	77	73	69	81	81	82	82	76	83	85	79
2	64	66	68	71	82	83	80	81	75	81	85	73
3	75	73	73	69	81	77	80	80	75	82	80	77
4	72	70	74	73	84	86	85	85	68	73	72	69
5	74	74	71	67	88	90	88	86	78	87	86	77
6	71	71	72	70	83	82	86	85	81	85	81	74
7	76	78	74	71	85	83	87	86	67	73	75	66
8	73	68	64	64	81	85	86	85	68	73	73	66
9	76	73	74	76	87	89	87	82	68	75	79	69
10	77	78	77	73	77	75	73	77	73	78	80	70

The Solution using STATA:

- Get the Data: (Please see the last page for a link to the data and do file)

```
. use "C:\Users\Fares\Documents\Fares\manova\hrate.dta"
```

Data Editor (Browse) - [hrate]

File Edit View Data Tools

group[1] Placebo

	group	time1	time2	time3	time4
1	Placebo	80	77	73	69
2	Placebo	64	66	68	71
3	Placebo	75	73	73	69
4	Placebo	72	70	74	73
5	Placebo	74	74	71	67
6	Placebo	71	71	72	70
7	Placebo	76	78	74	71
8	Placebo	73	68	64	64
9	Placebo	76	73	74	76
10	Placebo	77	78	77	73
11	Drug_A	81	81	82	82
12	Drug_A	82	83	80	81
13	Drug_A	81	77	80	80
14	Drug_A	84	86	85	85
15	Drug_A	88	90	88	86
16	Drug_A	83	82	86	85
17	Drug_A	85	83	87	86
18	Drug_A	81	85	86	85
19	Drug_A	87	89	87	82
20	Drug_A	77	75	73	77
21	Drug_B	76	83	85	79
22	Drug_B	75	81	85	73
23	Drug_B	75	82	80	77
24	Drug_B	68	73	72	69
25	Drug_B	78	87	86	77
26	Drug_B	81	85	81	74
27	Drug_B	67	73	75	66
28	Drug_B	68	73	73	66
29	Drug_B	68	75	79	69
30	Drug_B	73	78	80	70

Variables

Filter variables here

Name | Label

group
time1
time2
time3
time4

Variables | Snapshots

Properties

Variables

Name	group
Label	
Type	str8
Format	%8s
Value label	
Notes	

Data

Filename	hrate.dta
Label	
Notes	
Variables	5
Observations	30
Size	1.17K
Memory	64M

Length: 8 Vars: 5 Order: Dataset Obs: 30 Filter: Off Mode: Browse CAP NUM

- Conduct the MANOVA test:

```
. encode group, gen(ngroup)

. manova time1 time2 time3 time4 = ngroup

                    Number of obs =         30

                    W = Wilks' lambda      L = Lawley-Hotelling trace
                    P = Pillai's trace     R = Roy's largest root
```

Source	Statistic		df	F(df1,	df2) =	F	Prob>F	
ngroup	W	0.0628	2	8.0	48.0	17.94	0.0000	e
	P	1.4371		8.0	50.0	15.96	0.0000	a
	L	6.9625		8.0	46.0	20.02	0.0000	a
	R	5.5204		4.0	25.0	34.50	0.0000	u
Residual			27					
Total			29					

e = exact, a = approximate, u = upper bound on F

This is the standard STATA output when conducting MANOVA. All four multivariate tests indicate rejection of the null hypothesis. This indicates that there are one or more differences among the four-dimensional mean vectors for the three groups. The standard output in STATA when testing MANOVA corresponds to the overall treatment effect hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$. This hypothesis is rejected ($p < 0.05$). The "Multivariate R-squared" from this model is about 93.72% which is relatively strong.

The parameters' estimates of the MANOVA model are presented in the following table:

```
. mvreg
```

Equation	Obs	Parms	RMSE	"R-sq"	F	P
time1	30	3	4.213734	0.5608	17.23592	0.0000
time2	30	3	4.755114	0.4683	11.89238	0.0002
time3	30	3	4.475447	0.5548	16.82175	0.0000
time4	30	3	4.756476	0.7105	33.1252	0.0000

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
time1						
ngroup						
Drug_B	-10	1.884439	-5.31	0.000	-13.86655	-6.13345
Placebo	-9.1	1.884439	-4.83	0.000	-12.96655	-5.23345
_cons	82.9	1.3325	62.21	0.000	80.16594	85.63406
time2						
ngroup						
Drug_B	-4.1	2.126552	-1.93	0.064	-8.463324	.2633237
Placebo	-10.3	2.126552	-4.84	0.000	-14.66332	-5.936676
_cons	83.1	1.503699	55.26	0.000	80.01466	86.18534
time3						
ngroup						
Drug_B	-3.8	2.001481	-1.90	0.068	-7.9067	.3066997
Placebo	-11.4	2.001481	-5.70	0.000	-15.5067	-7.2933
_cons	83.4	1.415261	58.93	0.000	80.49612	86.30388
time4						
ngroup						
Drug_B	-10.9	1.679947	-6.49	0.000	-14.34697	-7.453033
Placebo	-12.6	1.679947	-7.50	0.000	-16.04697	-9.153033
_cons	82.9	1.187902	69.79	0.000	80.46263	85.33737

- Test the homogeneity assumption: In this assumption, we test the null hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = 0$.

```
. quietly manova time1 = ngroup

. predict res1, residuals

. quietly manova time2 = ngroup

. predict res2, residuals

. quietly manova time3 = ngroup

. predict res3, residuals

. quietly manova time4 = ngroup

. predict res4, residuals

. mvtest covariance res1 res2 res3 res4, by(group)

Test of equality of covariance matrices across 3 samples

      Modified LR chi2 =  30.98812
      Box F(20,2616.8) =      1.21      Prob > F =  0.2362
          Box chi2(20) =     24.41   Prob > chi2 =  0.2250
```

Firstly, we get the four residuals by conducting separate ANOVAs and then use the *mvtest* function. The Box's M test suggests that the data from all groups have common variance-covariance matrix ($p = 0.225 > 0.05$) so this assumptions wasn't violated.

- Test the Normality assumption: In this assumption, due to the small sample size per treatment group, we test the null hypothesis $H_0 : \epsilon \sim N_4(0, \Sigma)$. *If the sample size for each drug were large, it would be appropriate to check for normality within the treatment groups [8].*

```
. mvtest norm res1* res2* res3* res4* , bivariate univariate stats(all)

Test for univariate normality
```

Variable	Pr(Skewness)	Pr(Kurtosis)	joint adj chi2(2)	joint Prob>chi2
res1	0.4686	0.8624	0.58	0.7493
res2	0.8600	0.0583	3.88	0.1436
res3	0.0770	0.8690	3.46	0.1771
res4	0.8353	0.3794	0.86	0.6500

Doornik-Hansen test for bivariate normality

Pair of variables		chi2	df	Prob>chi2
res1	res2	5.60	4	0.2307
	res3	10.82	4	0.0286
	res4	5.21	4	0.2666
res2	res3	5.77	4	0.2171
	res4	1.89	4	0.7559
res3	res4	4.35	4	0.3603

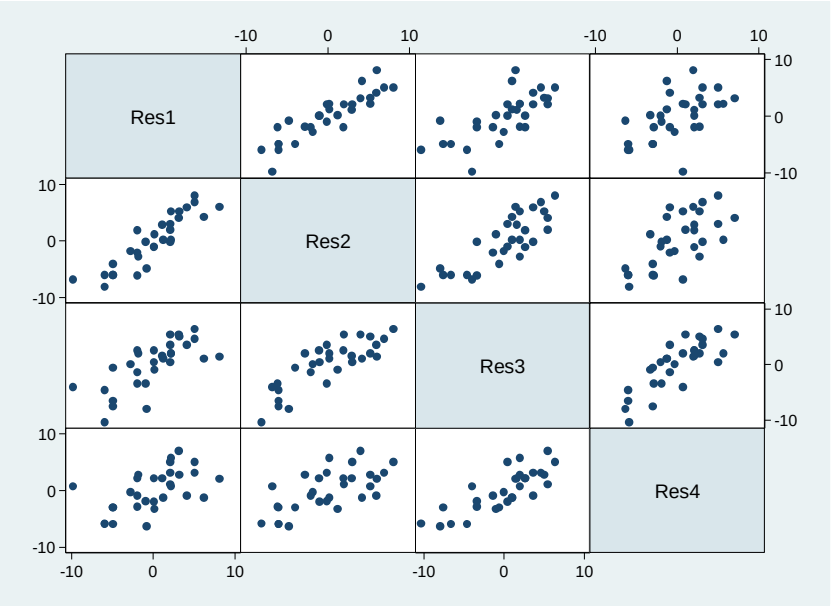
Test for multivariate normality

Mardia mSkewness =	2.215629	chi2(20) =	12.677	Prob>chi2 = 0.8908
Mardia mKurtosis =	20.61932	chi2(1) =	1.786	Prob>chi2 = 0.1814
Henze-Zirkler =	.7739534	chi2(1) =	0.280	Prob>chi2 = 0.5970
Doornik-Hansen		chi2(8) =	12.062	Prob>chi2 = 0.1485

The three formal tests above, for univariate normality, bivariate normality and multivariate normality, collectively indicate that the data are normally distributed. Only the bivariate normality of *res*1 and *res*3 was questionable since $p = 0.0286$. Nonetheless, this result shouldn't influence our inference regarding the multivariate normality assumption. This assumption is not violated and the following graphical presentations support such inference.

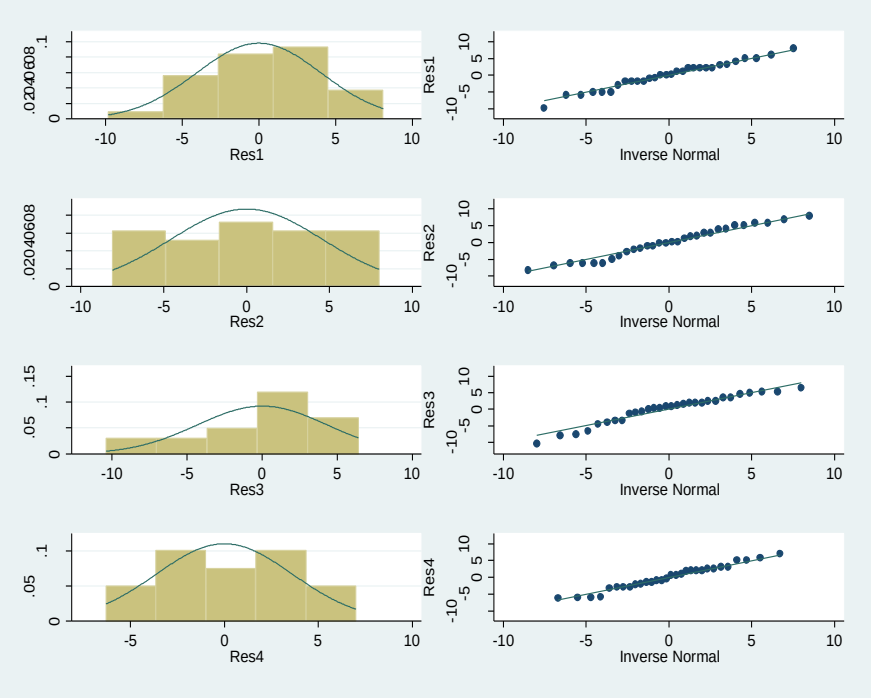
To, graphically, assess multivariate normality, we firstly examine the bivariate scatterplots for each pair of the residuals' vectors hopping to observe an elliptical shape and secondly look at the histogram of each vector of the residuals with the corresponding QQplot:

```
. gr matrix res1 res2 res3 res4
```



This graph is sufficient to establish the linearity assumption for the DVs.

```
. histogram res1, normal  name(res1h, replace) nodraw
. qnorm res1, name(res1q, replace) nodraw
. histogram res2, normal name(res2h, replace) nodraw
. qnorm res2, name(res2q, replace) nodraw
. histogram res3, normal  name(res3h, replace) nodraw
. qnorm res3, name(res3q, replace) nodraw
. histogram res4, normal name(res4h, replace) nodraw
. qnorm res4, name(res4q, replace) nodraw
. gr combine res1h res1q res2h res2q res3h res3q res4h res4q, cols(2)
```



To conduct in STATA a test for univariate normality which is similar to that in SAS or R, we use the *swilk* command which implements the Shapiro-Wilk test.

```
. swilk res1 res2 res3 res4

                    Shapiro-Wilk W test for normal data

    Variable |        Obs        W         V         z       Prob>z
-------------+-------------------------------------------------------
        res1 |         30    0.98604    0.444    -1.680    0.95355
        res2 |         30    0.96203    1.207     0.389    0.34872
        res3 |         30    0.94028    1.898     1.325    0.09254
        res4 |         30    0.97125    0.914    -0.187    0.57401
```

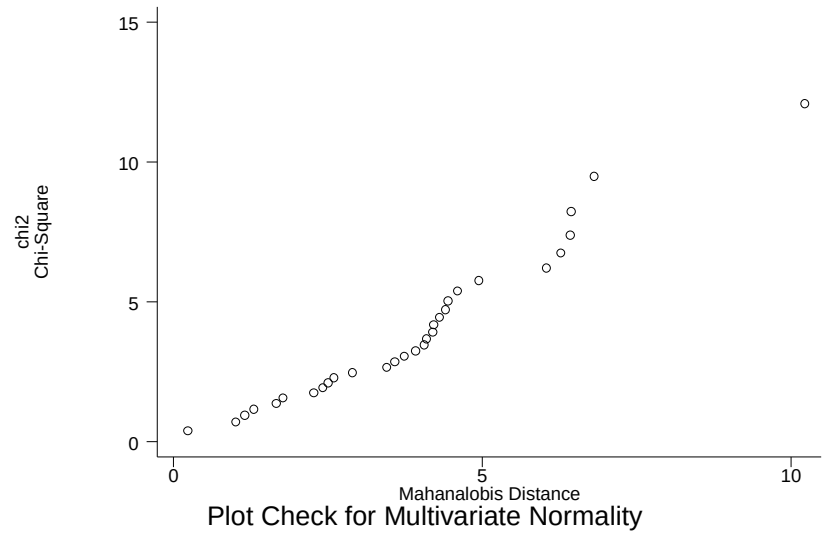
Note that:

- The normality assumption can be relaxed by appealing to the central limit theorem when the sample sizes n_i are large [10].
- Theoretically, we should examine the normality for every linear combination of the residuals. This can be time consuming so evaluating some finite number of the linear combinations is sufficient [8].
- To further examine the multivariate normality through graphical tools, one could also plot 3 dimensional scatterplots and look for elliptical shapes. This is a great tool to detect outliers.

- Test the assumption of Absence of Multivariate Outliers:

To examine multivariate outliers in the data, we use the QQPlot for the observed Mahalanobis distances (MD). We plot the ordered Mahalanobis distances versus estimated quantiles from a chi-squared distribution with p degrees of freedom and expect to see a straight-line.

```
. multnorm res1 res2 res3 res4
```




```
. display invchi2(4, 0.975)
11.143287
```

To conduct a formal test, we compute the 97.5% quantile Q of the Chi-Square distribution with p degrees of freedom using the *invchi*2 command and declare each point with MD which is greater than Q as a multivariate outlier.

The observed Mahalanobis distances of our data are presented below. Based on this data we have no multivariate outliers as none of the observations has a MD which is larger than 11.14, the 97.5% quantile of the Chi-Square distribution with 4 degrees of freedom.

	MD2	chi2
1.	.2334507	.3894305
2.	1.009522	.7107247
3.	1.15019	.9541528
4.	1.301458	1.168032
5.	1.665706	1.366477
6.	1.772314	1.556061
7.	2.272597	1.740582
8.	2.420321	1.922557
9.	2.502503	2.103842
10.	2.602005	2.285922
11.	2.896792	2.470087
12.	3.451539	2.657529
13.	3.585332	2.849415
14.	3.74001	3.046946
15.	3.915628	3.251416
16.	4.059857	3.464261
17.	4.096411	3.687134
18.	4.194732	3.921987
19.	4.209743	4.17119
20.	4.311124	4.437689
21.	4.408154	4.725257
22.	4.440107	5.038861
23.	4.597691	5.385269
24.	4.945594	5.774088
25.	6.038367	6.219663
26.	6.276779	6.744883
27.	6.42395	7.389828
28.	6.4403	8.235181
29.	6.811718	9.487729
30.	10.2261	12.09387

Note: This test is generally used to establish multivariate normality, however; we use it in here to only detect multivariate outliers.

- Test for an overall treatment effect: The null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ is rejected which indicates an existence of treatment effect. That is, at the 5% significance level, we can infer that at least one of the three treatments (Drug A, Drug B or Placebo) has a significant impact on women's heart rate.

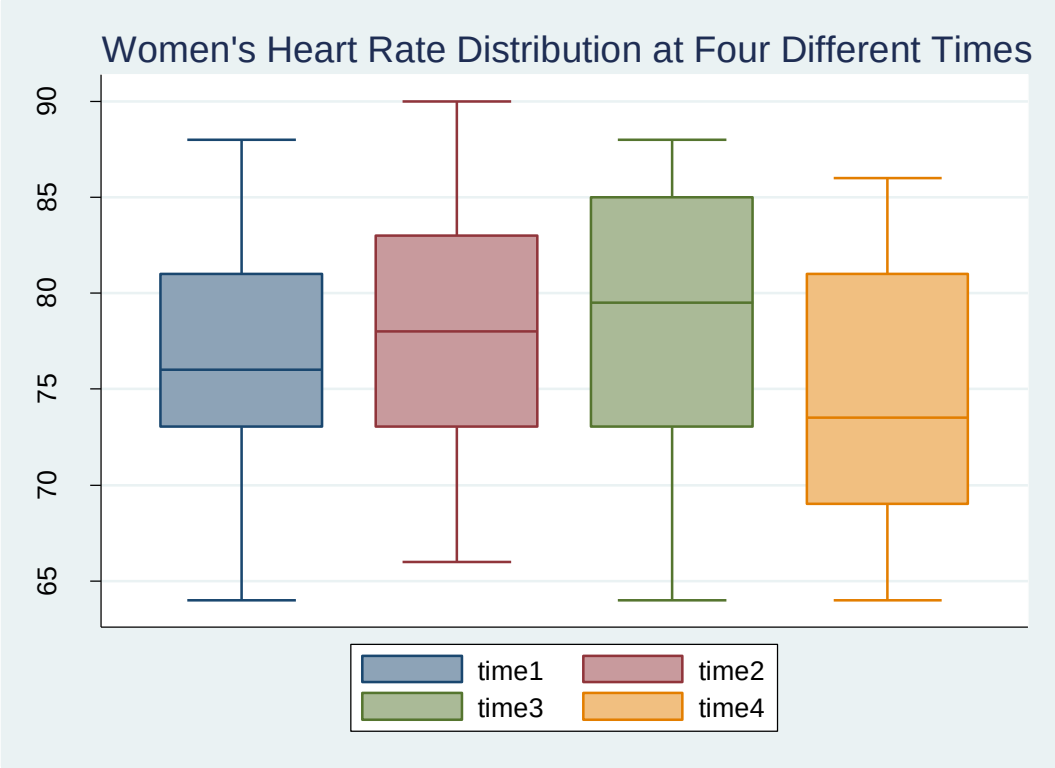
```
. encode group, gen(ngroup)

. manova time1 time2 time3 time4 = ngroup

                    Number of obs =         30

                    W = Wilks' lambda      L = Lawley-Hotelling trace
                    P = Pillai's trace     R = Roy's largest root

-----------------------------------------------------------------------------
      Source | Statistic        df    F(df1,      df2) =    F    Prob>F
-------------+---------------------------------------------------------------
      ngroup | W   0.0628        2       8.0      48.0     17.94 0.0000 e
             | P   1.4371                8.0      50.0     15.96 0.0000 a
             | L   6.9625                8.0      46.0     20.02 0.0000 a
             | R   5.5204                4.0      25.0     34.50 0.0000 u
-------------+---------------------------------------------------------------
    Residual |                  27
-------------+---------------------------------------------------------------
       Total |                  29
-----------------------------------------------------------------------------
                    e = exact, a = approximate, u = upper bound on F
```

Note that this output is the same as the default output we get from STATA when conducting a MANOVA (see page 13)].

- Test whether the four heart rate means are equal: The null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected (see STATA's output and Box-plot Figure below) which indicates, at the 5% significance level, that women's means heart rate at the four times are significantly different.

```
. quietly manova time1 time2 time3 time4 = ngroup

. matrix M = (1,-1,0,0 \ 0,1,-1,0\ 0,0,1,-1)

. matrix H = (1/3,1/3, 1/3, 1)

. manovatest , test(H) ytransform(M)

Transformations of the dependent variables
(1)     time1 - time2
(2)     time2 - time3
(3)     time3 - time4

Test constraint
(1)     .3333333*1.ngroup + .3333333*2.ngroup + .3333333*3.ngroup + _cons = 0

                    W = Wilks' lambda      L = Lawley-Hotelling trace
                    P = Pillai's trace     R = Roy's largest root

    Source | Statistic        df     F(df1,     df2) =    F   Prob>F
-----------+-----------------------------------------------------------
manovatest | W   0.3238        1       3.0      25.0    17.40 0.0000 e
           | P   0.6762                3.0      25.0    17.40 0.0000 e
           | L   2.0885                3.0      25.0    17.40 0.0000 e
           | R   2.0885                3.0      25.0    17.40 0.0000 e
-----------+-----------------------------------------------------------
  Residual |                  27
-----------------------------------------------------------------------

                    e = exact, a = approximate, u = upper bound on F
```

```
. graph box time1 time2 time3 time4, ytitle("Heart Rate") title("Women's Heart Rate Distribution at Four Different Times")
```




```
. /*95% Bonferroni C.I*/
. ci time1 time2 time3 time4, level(98.75)
```

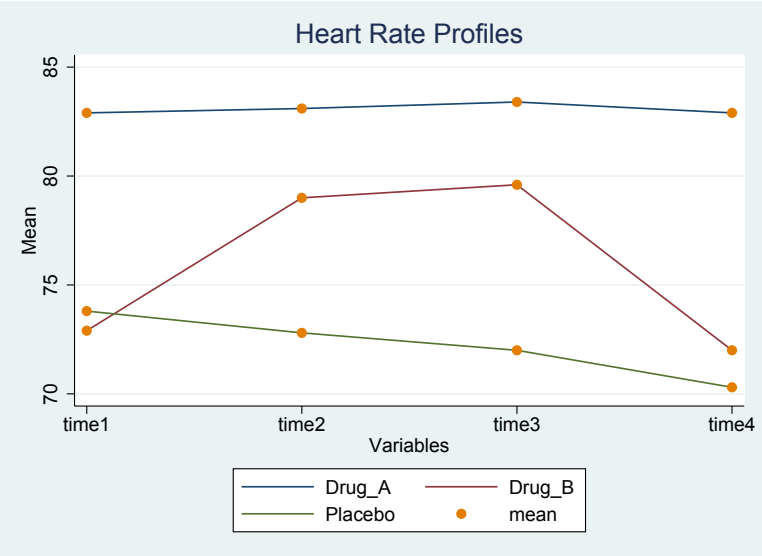
Variable	Obs	Mean	Std. Err.	[98.75% Conf.	Interval]
time1	30	76.53333	1.120071	73.55036	79.5163
time2	30	78.3	1.148863	75.24035	81.35965
time3	30	78.33333	1.181596	75.18651	81.48015
time4	30	75.06667	1.229833	71.79138	78.34195

- Test whether the four heart rate means, for Drug A and Placebo, are equal: The null hypothesis $H_0 : \tau_1 = \tau_3$ is rejected (see STATA's output below). That is, at the 5% significance level, we can infer that the impact of Drug A on women's heart rate is significantly different than that of the Placebo.

```
. quietly manova time1 time2 time3 time4 = ngroup

. matrix C = (1,0,-1,0)

. manovatest , test(C)

 Test constraint
 (1)     1.ngroup - 3.ngroup = 0

                    W = Wilks' lambda       L = Lawley-Hotelling trace
                    P = Pillai's trace      R = Roy's largest root

    Source | Statistic        df    F(df1,      df2) =    F   Prob>F
-----------+-----------------------------------------------------------
manovatest | W    0.3115       1       4.0      24.0     13.26 0.0000 e
           | P    0.6885               4.0      24.0     13.26 0.0000 e
           | L    2.2107               4.0      24.0     13.26 0.0000 e
           | R    2.2107               4.0      24.0     13.26 0.0000 e
-----------+-----------------------------------------------------------
  Residual |                  27
-----------------------------------------------------------------------
                    e = exact, a = approximate, u = upper bound on F
```

Profile Analysis: When comparing the same dependent variable between groups over several time points then profile analysis is invoked. In this analysis, one examines three diffrent hypotheses.

- Whether the curves are parallel (Parallelism)?
- Whether the curves have the same average level (Separation or Coincidental profiles)?
- Whether the average curve is horizontal (Flatness)?




```
. tabstat time1 time2 time3 time4, by(ngroup)

Summary statistics: mean
  by categories of: ngroup
```

ngroup	time1	time2	time3	time4
Drug_A	82.9	83.1	83.4	82.9
Drug_B	72.9	79	79.6	72
Placebo	73.8	72.8	72	70.3
Total	76.53333	78.3	78.33333	75.06667

We observe from the profiles plot above that Drug B is different from both Drug A and Placebo. In fact, its profile falls in between the profiles of Drug A and Placebo that both seem to be similar in their behavior over time.

Test for Parallelism: The null hypothesis tests if the two drugs and placebo have parallel profiles.

```
. /* test of parallelism */
. quietly manova time1 time2 time3 time4 = ngroup

. matrix M = (1,-1,0,0\0,1,-1,0\0,0,1,-1)

. manovatest ngroup, ytrans(M)

 Transformations of the dependent variables
 (1)     time1 - time2
 (2)     time2 - time3
 (3)     time3 - time4

                    W = Wilks' lambda      L = Lawley-Hotelling trace
                    P = Pillai's trace     R = Roy's largest root

    Source | Statistic        df     F(df1,      df2) =    F    Prob>F
-----------+-----------------------------------------------------------
    ngroup | W   0.2039          2       6.0      50.0    10.12 0.0000 e
           | P   0.9025                  6.0      52.0     7.13 0.0000 a
           | L   3.3837                  6.0      48.0    13.53 0.0000 a
           | R   3.2218                  3.0      26.0    27.92 0.0000 u
-----------+-----------------------------------------------------------
  Residual |                    27
-----------------------------------------------------------------------

                e = exact, a = approximate, u = upper bound on F
```

The previous graph of heart rate profiles clearly indicates that the parallelism hypothesis should be rejected. From the above output, we see that this hypothesis is rejected based on the four multivariate test and hence we can infer that the changes in women's heart rate are significantly NOT the same direction and pattern for the two drugs and placebo.

Test for Separation (Coincidental): The null hypothesis tests if the curves have the same average level. This hypothesis is meaningless in this situation since the parallelism hypothesis was rejected. Nonetheless, for demonstration purposes I will provide the STATA code/output.

```
. /* test of coincidental profiles (test of levels) */
. quietly manova time1 time2 time3 time4 = ngroup

. mat c2 = (1,1,1,1)

. manovatest ngroup, ytrans(c2)

Transformation of the dependent variables
(1)     time1 + time2 + time3 + time4

                    W = Wilks' lambda      L = Lawley-Hotelling trace
                    P = Pillai's trace     R = Roy's largest root

      Source | Statistic        df    F(df1,     df2) =    F   Prob>F
-------------+-------------------------------------------------------
      ngroup | W   0.4000        2       2.0      27.0    20.25 0.0000 e
             | P   0.6000                2.0      27.0    20.25 0.0000 e
             | L   1.4998                2.0      27.0    20.25 0.0000 e
             | R   1.4998                2.0      27.0    20.25 0.0000 e
-------------+-------------------------------------------------------
    Residual |                  27
-------------------------------------------------------------------------

                    e = exact, a = approximate, u = upper bound on F
```

Here is a fake example [11] in which coincidental profiles is occuring:

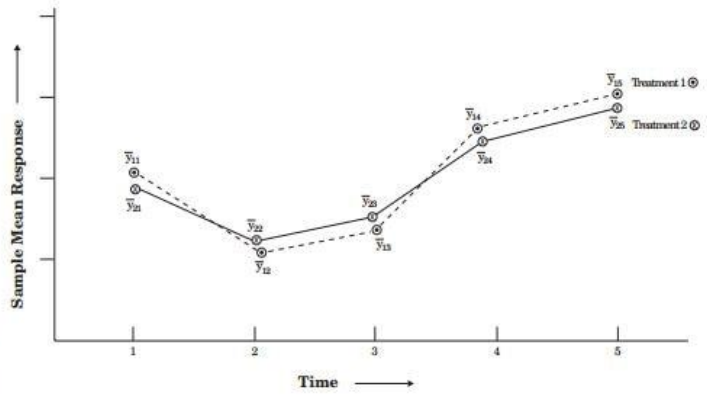


Figure 5.3
Two Sample Profiles Suggesting the Possible Coincidentalness

Test for Flatness: The null hypothesis tests if the the average curve is horizontal. This is the same as testing whether the four heart rate means are equal (see page 23). For completeness, I am providing he STATA code and output again.

```
. /* test of flatness */
. quietly manova time1 time2 time3 time4 = ngroup

. matrix M = (1,-1,0,0\0,1,-1,0\0,0,1,-1)

. matrix H = (1/3,1/3,1/3,1)

. /* Stata 10: H = (1,1/3,1/3,1/3) */
. manovatest, test(H) ytrans(M)

Transformations of the dependent variables
(1)     time1 - time2
(2)     time2 - time3
(3)     time3 - time4

Test constraint
(1)     .3333333*1.ngroup + .3333333*2.ngroup + .3333333*3.ngroup + _cons = 0

                    W = Wilks' lambda       L = Lawley-Hotelling trace
                    P = Pillai's trace      R = Roy's largest root

      Source | Statistic        df     F(df1,      df2) =    F    Prob>F
-------------+-------------------------------------------------------------
  manovatest | W    0.3238       1       3.0      25.0     17.40 0.0000 e
             | P    0.6762               3.0      25.0     17.40 0.0000 e
             | L    2.0885               3.0      25.0     17.40 0.0000 e
             | R    2.0885               3.0      25.0     17.40 0.0000 e
-------------+-------------------------------------------------------------
    Residual |                  27
-------------------------------------------------------------------------------
                    e = exact, a = approximate, u = upper bound on F
```

Note: STATA 10 or less reserves the first column of the H (test) matrix for the constant's column while STATA 11 or more reserves the last column for the same purpose. So, if you were using STATA 14 then your H matrix would be $H = (1/3, 1/3, 1/3, 1)$ and if you were using SATA 9 then it would be $H = (1, 1/3, 1/3, 1/3)$.

Post Hoc Analysis: Several methods are generally conducted after a MANOVA model including: Simultaneous confidence intervals, Multivariate contrasts, Multiple Univariate ANOVAs, Discriminant Analysis and others. For our example, I will provide the results of the Linear Discriminant Analysis (LDA) to illustrate the classification accuracy of our model.

```
. discrim lda time1 time2 time3 time4 , group(ngroup)

Linear discriminant analysis
Resubstitution classification summary
```

Key
Number Percent

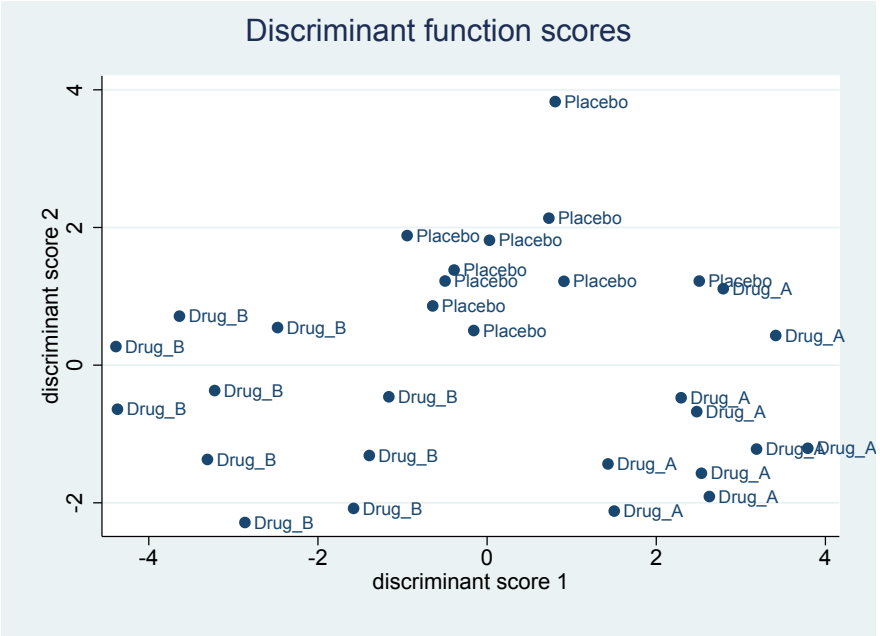
True ngroup	Classified Drug A	Drug B	Placebo	Total
Drug A	10 100.00	0 0.00	0 0.00	10 100.00
Drug B	0 0.00	10 100.00	0 0.00	10 100.00
Placebo	1 10.00	0 0.00	9 90.00	10 100.00
Total	11 36.67	10 33.33	9 30.00	30 100.00
Priors	0.3333	0.3333	0.3333	

```
. estat list, varlist misclassified
```

	Data				Classification		Probabilities		
Obs.	time1	time2	time3	time4	True	Class.	Drug_A	Drug_B	Placebo
27	76	73	74	76	Placebo	Drug_A *	0.5959	0.0000	0.4041

* indicates misclassified observations

In our model, we have only one misclassification for a Placebo into Drug A. This could be also easily seen from the following score plot.



This clear linear discrimination between the three treatments was reflected in the MANOVA analysis previously by the strong "Multivariate R-squared" of 93.72%.

The Solution using SAS:

- Get the Data: (Please see the last page for a link to the SAS syntax file)

```
data hrate;
input group $ time1 time2 time3 time4;
cards;
Placebo 80 77 73 69
Placebo 64 66 68 71
Placebo 75 73 73 69
Placebo 72 70 74 73
Placebo 74 74 71 67
Placebo 71 71 72 70
Placebo 76 78 74 71
Placebo 73 68 64 64
Placebo 76 73 74 76
Placebo 77 78 77 73
Drug_A 81 81 82 82
Drug_A 82 83 80 81
Drug_A 81 77 80 80
Drug_A 84 86 85 85
Drug_A 88 90 88 86
Drug_A 83 82 86 85
Drug_A 85 83 87 86
Drug_A 81 85 86 85
Drug_A 87 89 87 82
Drug_A 77 75 73 77
Drug_B 76 83 85 79
Drug_B 75 81 85 73
Drug_B 75 82 80 77
Drug_B 68 73 72 69
Drug_B 78 87 86 77
Drug_B 81 85 81 74
Drug_B 67 73 75 66
Drug_B 68 73 73 66
Drug_B 68 75 79 69
Drug_B 73 78 80 70
run;
```

	group	time1	time2	time3	time4
1	Placebo	80	77	73	69
2	Placebo	64	66	68	71
3	Placebo	75	73	73	69
4	Placebo	72	70	74	73
5	Placebo	74	74	71	67
6	Placebo	71	71	72	70
7	Placebo	76	78	74	71
8	Placebo	73	68	64	64
9	Placebo	76	73	74	76
10	Placebo	77	78	77	73
11	Drug_A	81	81	82	82
12	Drug_A	82	83	80	81
13	Drug_A	81	77	80	80
14	Drug_A	84	86	85	85
15	Drug_A	88	90	88	86
16	Drug_A	83	82	86	85
17	Drug_A	85	83	87	86
18	Drug_A	81	85	86	85
19	Drug_A	87	89	87	82
20	Drug_A	77	75	73	77
21	Drug_B	76	83	85	79
22	Drug_B	75	81	85	73
23	Drug_B	75	82	80	77
24	Drug_B	68	73	72	69
25	Drug_B	78	87	86	77
26	Drug_B	81	85	81	74
27	Drug_B	67	73	75	66
28	Drug_B	68	73	73	66
29	Drug_B	68	75	79	69
30	Drug_B	73	78	80	70

- Conduct the MANOVA test:

```
proc glm data=hrate order=data;
  class group;
  model time1 time2 time3 time4 = group/solution ss3;
  output out=resids r=r1 r2 r3 r4;
  manova h = group;
  run;
quit;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall group Effect
H = Type III SSCP Matrix for group
E = Error SSCP Matrix

S=2 M=0.5 N=11

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.06280100	17.94	8	48	<.0001
Pillai's Trace	1.43714881	15.96	8	50	<.0001
Hotelling-Lawley Trace	6.96245516	20.42	8	32.049	<.0001
Roy's Greatest Root	5.52036673	34.50	4	25	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

This is the standard SAS output when conducting MANOVA. All four multivariate tests indicate rejection of the null hypothesis. This indicates that there are one or more differences among the four-dimensional mean vectors for the three groups. This output corresponds to the overall treatment effect hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$. This hypothesis is rejected ($p < 0.05$). The "Multivariate R-squared" from this model is about 93.72% which is relatively strong.

The parameters' estimates of the MANOVA model are presented as follows:

Dependent Variable: time1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	612.066667	306.033333	17.24	<.0001
Error	27	479.400000	17.755556		
Corrected Total	29	1091.466667			

R-Square	Coeff Var	Root MSE	time1 Mean
0.560774	5.505750	4.213734	76.53333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	612.0666667	306.0333333	17.24	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	72.90000000	B	1.33249974	54.71	<.0001
group Placebo	0.90000000	B	1.88443920	0.48	0.6368
group Drug_A	10.00000000	B	1.88443920	5.31	<.0001
group Drug_B	0.00000000	B	.	.	.

Dependent Variable: time2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	537.800000	268.900000	11.89	0.0002
Error	27	610.500000	22.611111		
Corrected Total	29	1148.300000			

R-Square	Coeff Var	Root MSE	time2 Mean
0.468345	6.072943	4.755114	78.30000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	537.8000000	268.9000000	11.89	0.0002

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	79.00000000	B	1.50369914	52.54	<.0001
group Placebo	-6.20000000	B	2.12655172	-2.92	0.0071
group Drug_A	4.10000000	B	2.12655172	1.93	0.0644
group Drug_B	0.00000000	B	.	.	.

Dependent Variable: time3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	673.866667	336.933333	16.82	<.0001
Error	27	540.800000	20.029630		
Corrected Total	29	1214.666667			

R-Square	Coeff Var	Root MSE	time3 Mean
0.554775	5.713337	4.475447	78.33333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	673.8666667	336.9333333	16.82	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	79.60000000	B	1.41526074	56.24	<.0001
group Placebo	-7.60000000	B	2.00148093	-3.80	0.0008
group Drug_A	3.80000000	B	2.00148093	1.90	0.0684
group Drug_B	0.00000000	B	.	.	.

Dependent Variable: time4

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	934.866667	467.433333	33.13	<.0001
Error	27	381.000000	14.111111		
Corrected Total	29	1315.866667			

R-Square	Coeff Var	Root MSE	time4 Mean
0.710457	5.004186	3.756476	75.06667

Source	DF	Type III SS	Mean Square	F Value	Pr > F
group	2	934.8666667	467.4333333	33.13	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	72.00000000	B	1.18790198	60.61	<.0001
group Placebo	-1.70000000	B	1.67994709	-1.01	0.3206
group Drug_A	10.90000000	B	1.67994709	6.49	<.0001
group Drug_B	0.00000000	B	.	.	.

- Test the homogeneity assumption: In this assumption, we test the null hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = 0$.

```
proc discrim data=hrate pool=test;
class group;
var time1-time4;
run;
```

The SAS System

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
24.407926	20	0.2250

Since the Chi-Square value is not significant at the 0.1 level, a pooled covariance matrix will be used in the discriminant function.
Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

The Box's M test suggests that the data from all groups have common variance-covariance matrix ($p = 0.225 > 0.05$) so this assumptions wasn't violated.

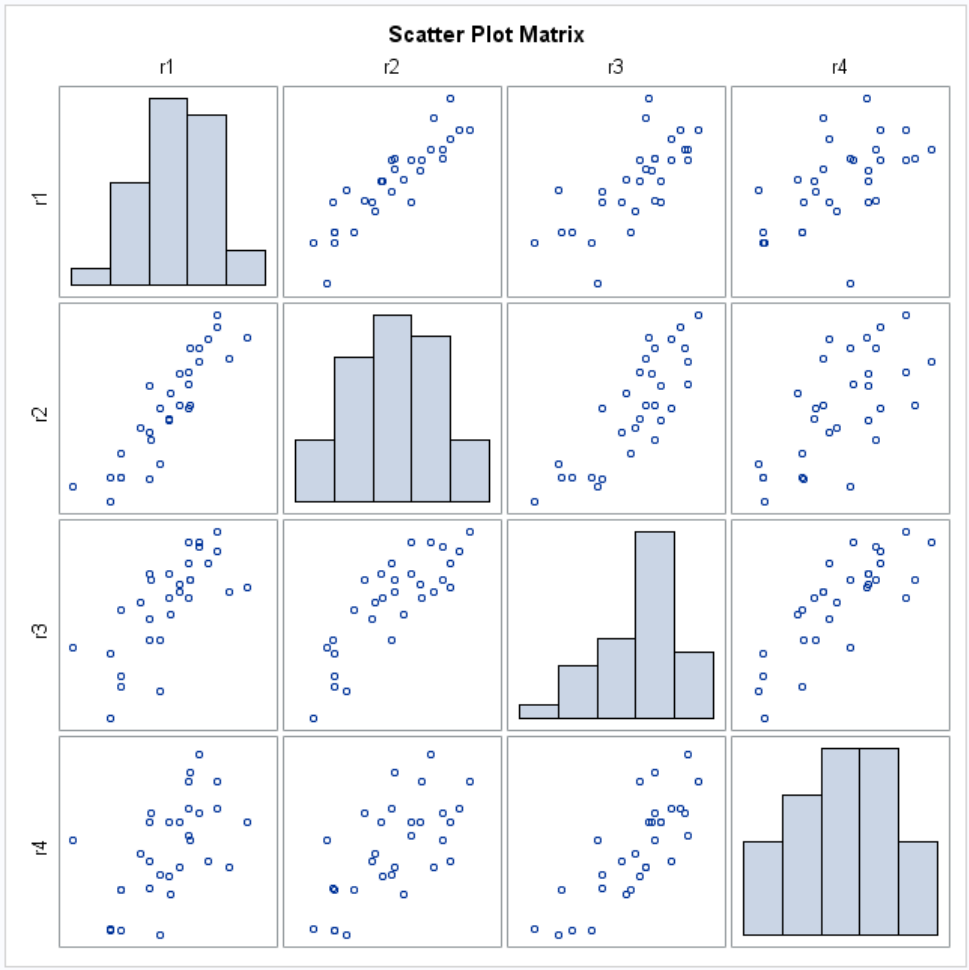
- Test the Normality assumption: To test the null hypothesis $H_0 : \epsilon \sim N_4(0, \Sigma)$, in SAS, we use the *UNIVARIATE* and *MODEL* procedures. The *UNIVARIATE* procedure provides the Shapiro-Wilk test for univariate normality and many other tests and the *MODEL* procedure provides the Mardia Skewness test for multivariate normality in addition to the the Shapiro-Wilk test for univariate normality. SAS doesn't provide the Doornik-Hansen test for bivariate normality.

```
proc model data=resids;
   r1=parm1;
   r2=parm2;
   r3=parm3;
   r4=parm4;
   fit r1 r2 r3 r4/ normal ;
run;
```

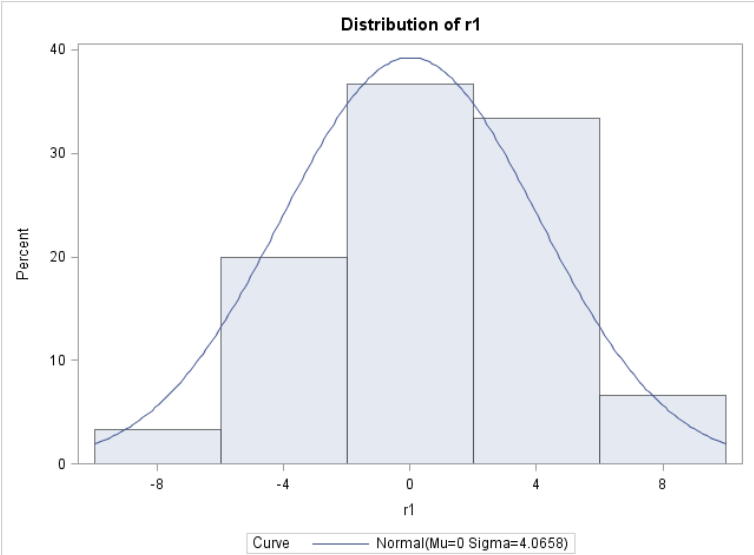
Normality Test			
Equation	Test Statistic	Value	Prob
r1	Shapiro-Wilk W	0.98	0.8877
r2	Shapiro-Wilk W	0.96	0.3230
r3	Shapiro-Wilk W	0.94	0.0946
r4	Shapiro-Wilk W	0.97	0.4637
System	Mardia Skewness	12.68	0.8908
	Mardia Kurtosis	-1.34	0.1814
	Henze-Zirkler T	0.77	0.2985

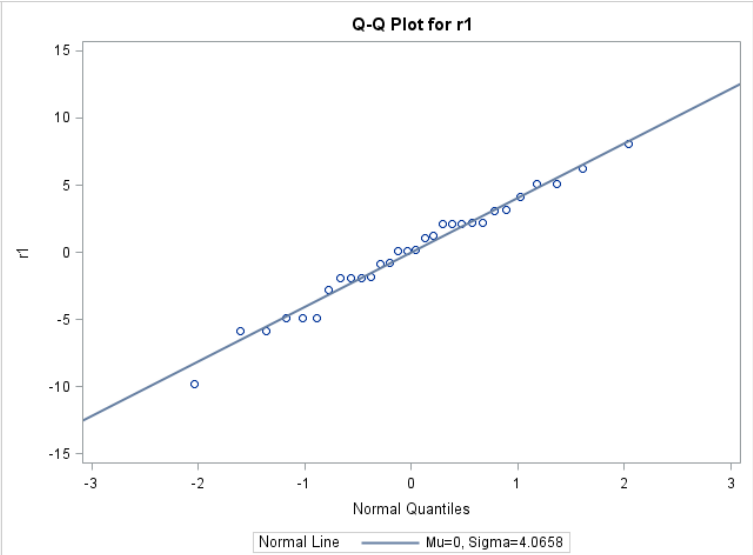
To, graphically, assess multivariate normality, we firstly examine the bivariate scatterplots for each pair of the residuals' vectors hopping to observe an elliptical shape and secondly look at the histogram of each vector of the residuals with the corresponding QQplot:

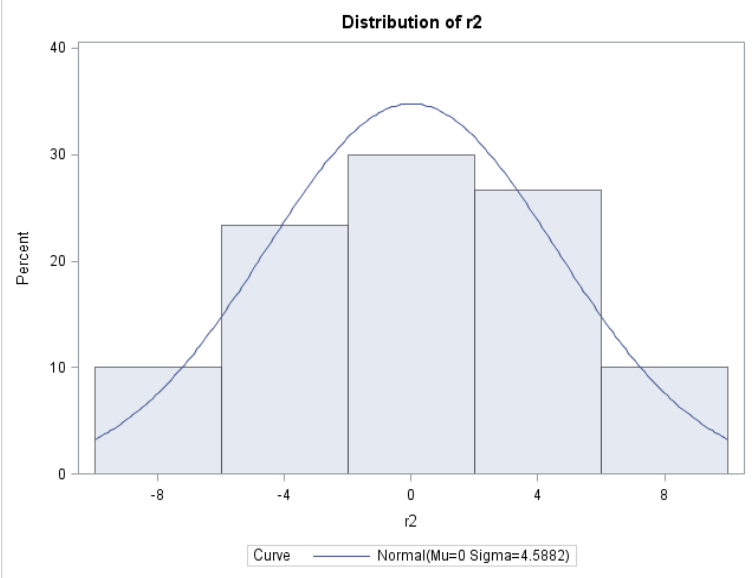
```
proc corr data=resids COV plots(maxpoints=NONE)=matrix(histogram);
   var r1 r2 r3 r4;
   ods select MatrixPlot;
run;
```

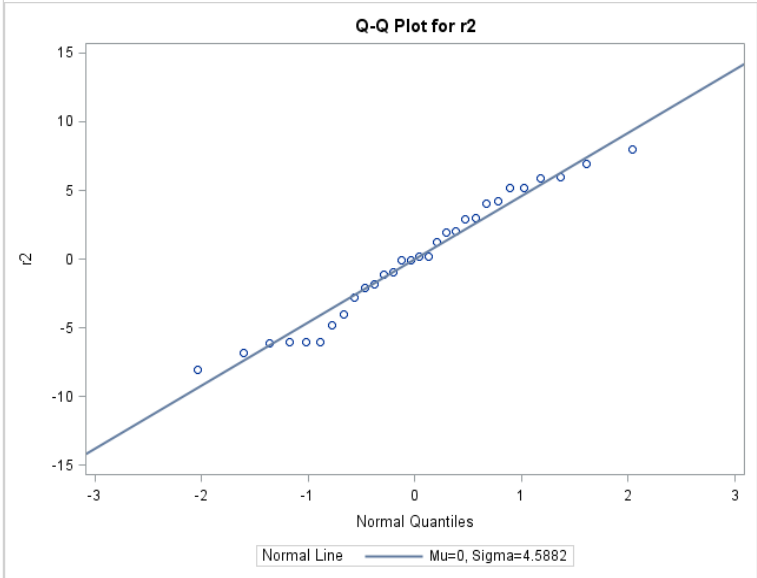


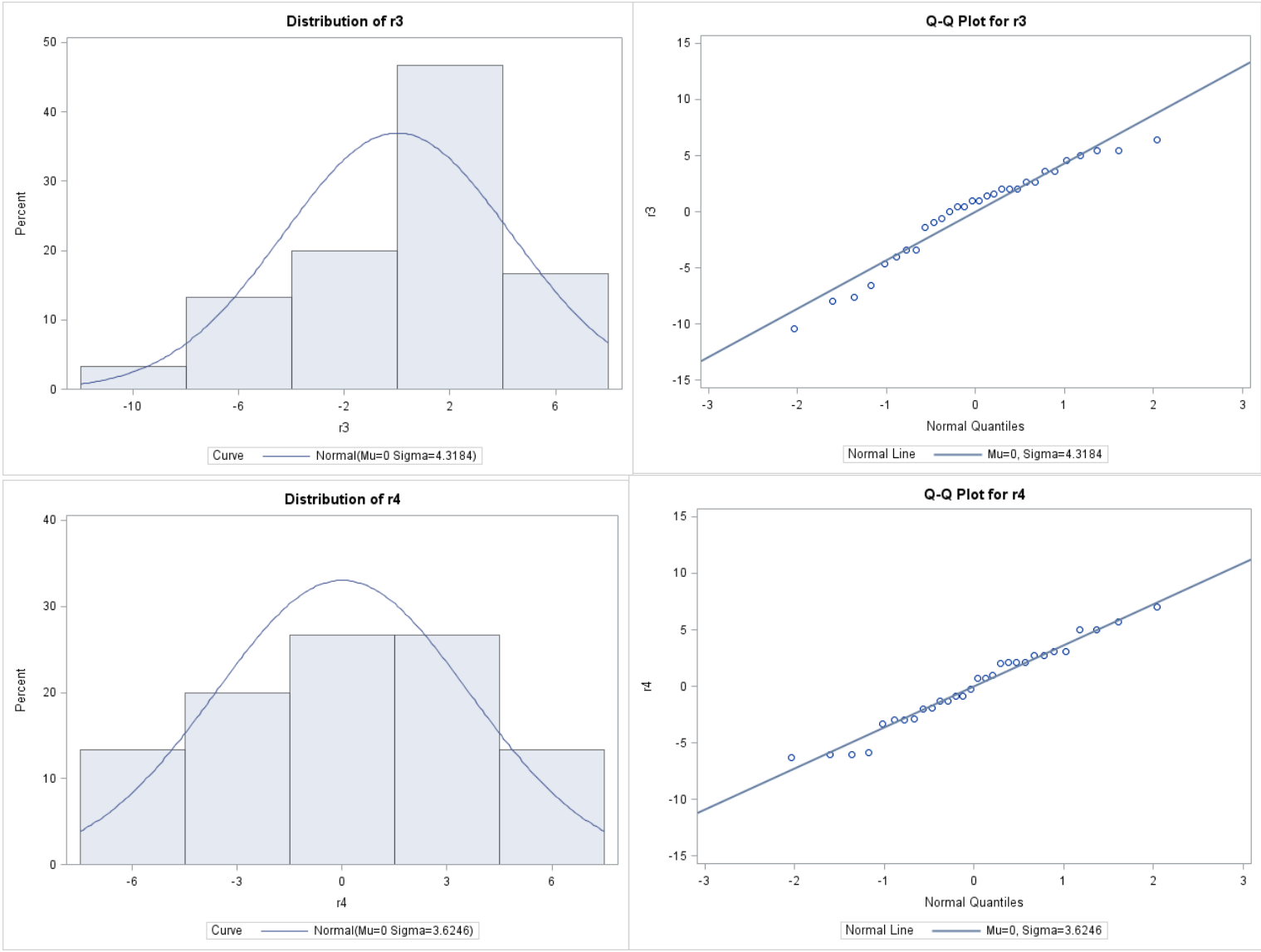
```
PROC UNIVARIATE DATA=resids NORMAL PLOT;
VAR r1 r2 r3 r4;
QQPLOT r1 r2 r3 r4 /NORMAL(MU=EST SIGMA=EST COLOR=RED L=1);
HISTOGRAM / NORMAL(COLOR=MAROON W=4) CFILL = BLUE CFRAME = LIGR;
RUN;
```










Distribution of r3
Percent
r3
Curve
Normal(Mu=0 Sigma=4.3184)
Q-Q Plot for r3
r3
Normal Quantiles
Normal Line
Mu=0, Sigma=4.3184
Distribution of r4
Percent
r4
Curve
Normal(Mu=0 Sigma=3.6246)
Q-Q Plot for r4
r4
Normal Quantiles
Normal Line
Mu=0, Sigma=3.6246

- Test the assumption of Absence of Multivariate Outliers:

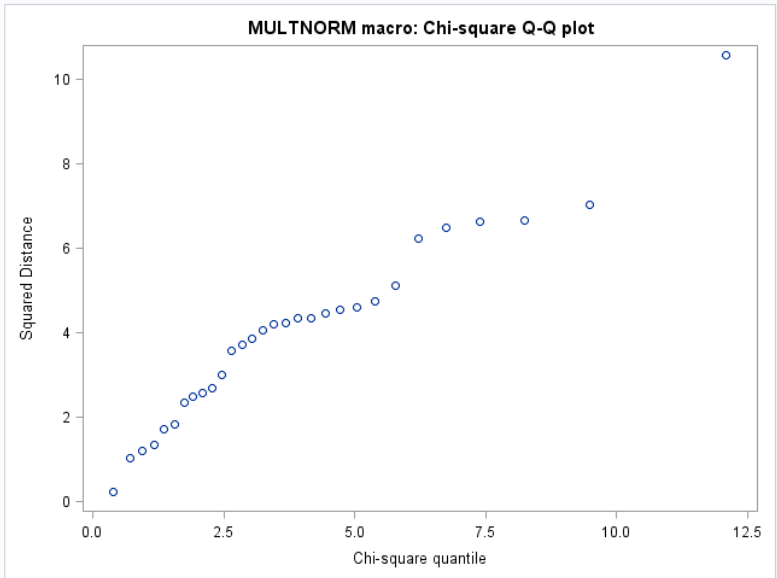
To examine multivariate outliers in the data, we use the QQPlot for the observed Mahalanobis distances (MD). This is done in SAS via either one of the macros %*multnorm* and %*cqplot* (see the last page for a link to the SAS syntax for the macros).

```
%multnorm(data=resids, var=r1 r2 r3 r4, plot=both)
```

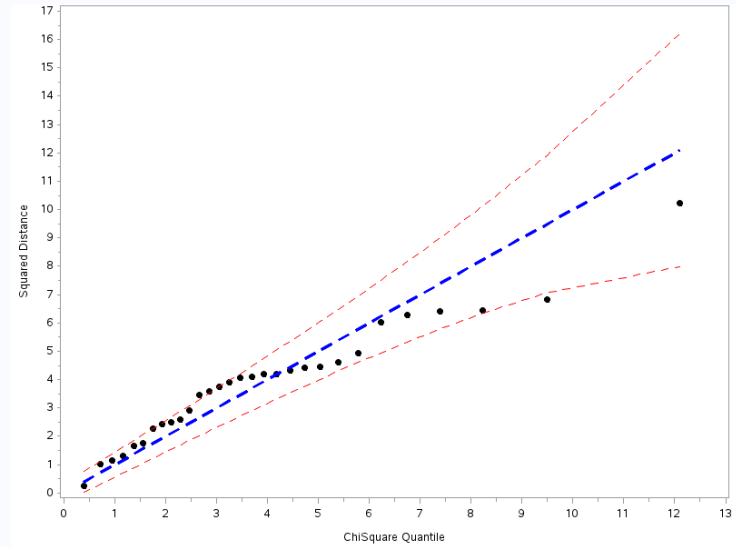
MULTNORM macro: Univariate and Multivariate Normality Tests

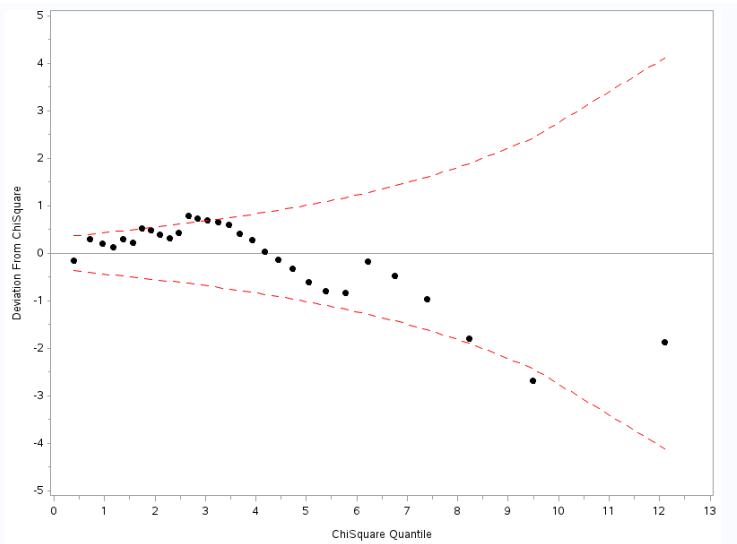
The MODEL Procedure

Normality Test			
Equation	Test Statistic	Value	Prob
r1	Shapiro-Wilk W	0.98	0.8877
r2	Shapiro-Wilk W	0.96	0.3230
r3	Shapiro-Wilk W	0.94	0.0946
r4	Shapiro-Wilk W	0.97	0.4637
System	Mardia Skewness	12.68	0.8908
	Mardia Kurtosis	-1.34	0.1814
	Henze-Zirkler T	0.77	0.2985



```
%cqplot(data=resids, var=r1-r4, nvar=4);
```





- Note that in SAS, as opposed to STATA, the Chi-square quantiles are on the x-axis instead of the y-axis.
- To get the observed Mahalanobis distances, we print the *dsq* variable from the *Cqplot* data set which was generated by the %*cqplot* macro.

The observed Mahalanobis distances of our data are presented below.

```
proc print data=Cqplot;
 var dsq _z_;
 run;
```

Obs	dsq	_z_
1	0.2335	0.3894
2	1.0095	0.7107
3	1.1502	0.9542
4	1.3015	1.1680
5	1.6657	1.3665
6	1.7723	1.5561
7	2.2726	1.7406
8	2.4203	1.9226
9	2.5025	2.1038
10	2.6020	2.2859
11	2.8968	2.4701
12	3.4515	2.6575
13	3.5853	2.8494
14	3.7400	3.0469
15	3.9156	3.2514
16	4.0599	3.4643
17	4.0964	3.6871
18	4.1947	3.9220
19	4.2097	4.1712
20	4.3111	4.4377
21	4.4082	4.7253
22	4.4401	5.0389
23	4.5977	5.3853
24	4.9456	5.7741
25	6.0384	6.2197
26	6.2768	6.7449
27	6.4240	7.3898
28	6.4403	8.2352
29	6.8117	9.4877
30	10.2261	12.0939

- Test for an overall treatment effect: The null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ is rejected which indicates an existence of treatment effect. That is, at the 5% significance level, we can infer that at least one of the three treatments (Drug A, Drug B or Placebo) has a significant impact on women's heart rate.

```
proc glm data=hrate order=data;
  class group;
  model time1 time2 time3 time4 = group/solution ss3;
  output out=resids r=r1 r2 r3 r4;
  manova h = group;
  run;
quit;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall group Effect
H = Type III SSCP Matrix for group
E = Error SSCP Matrix

S=2 M=0.5 N=11

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.06280100	17.94	8	48	<.0001
Pillai's Trace	1.43714881	15.96	8	50	<.0001
Hotelling-Lawley Trace	6.96245516	20.42	8	32.049	<.0001
Roy's Greatest Root	5.52036673	34.50	4	25	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

Note that this output is the same as the default output we get from SAS when conducting a MANOVA (see page 33)].

- Test whether the four heart rate means are equal: The null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected (see SAS output and Box-plot Figure below) which indicates, at the 5% significance level, that women's means heart rate at the four times are significantly different.

Method I:

```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
Manova M=TIME1-TIME2, TIME2-TIME3, TIME3-TIME4 H=INTERCEPT/SUMMARY;
  run;
quit;
```

Method II:

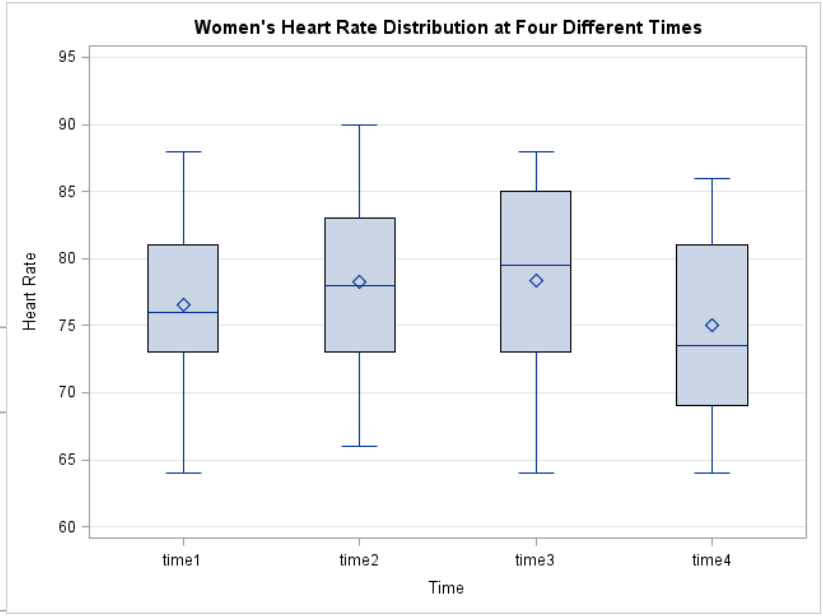
```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
CONTRAST "Horizontal"  intercept 1;
MANOVA H=GROUP M=(1 -1 0 0,
                  1 0 -1 0,
                  1 0 0 -1)/PRINTE PRINTH;
  run;
quit;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Horizontal Effect on the Variables Defined by the M Matrix Transformation
H = Contrast SSCP Matrix for Horizontal
E = Error SSCP Matrix

S=1 M=0.5 N=11.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.32378036	17.40	3	25	<.0001
Pillai's Trace	0.67621964	17.40	3	25	<.0001
Hotelling-Lawley Trace	2.08851346	17.40	3	25	<.0001
Roy's Greatest Root	2.08851346	17.40	3	25	<.0001

```
data box;
set hrate;
id=_N_;
RUN;
PROC TRANSPOSE DATA=BOX OUT=BOX2;
BY ID;
RUN;
PROC SGPLOT DATA =BOX2;
vbox COL1/category=_NAME_;
XAXIS TYPE = DISCRETE GRID;
YAXIS LABEL = 'Heart Rate' GRID VALUES = (60 TO 95 BY 5);
XAXIS LABEL = 'Time';
TITLE "Women's Heart Rate Distribution at Four Different Times";
RUN;
```




```
/***The 95% Bonferroni C.I****/
data null;
call symput('bonalpha',0.05/4);
run;
proc means data=hrate alpha=&bonalpha n mean lclm uclm;
var time1-time4;
run;
```

The MEANS Procedure

Variable	N	Mean	Lower 98.75% CL for Mean	Upper 98.75% CL for Mean
time1	30	76.5333333	73.5503646	79.5163020
time2	30	78.3000000	75.2403543	81.3596457
time3	30	78.3333333	75.1865125	81.4801542
time4	30	75.0666667	71.7913794	78.3419540

- Test whether the four heart rate means, for Drug A and Placebo, are equal: The null hypothesis $H_0 : \tau_1 = \tau_3$ is tested via using the **contrast** statement. In here H_0 is rejected (see SAS output below). That is, at the 5% significance level, we can infer that the impact of Drug A on women's heart rate is significantly different than that of the Placebo.

```
PROC GLM data = hrate;
CLASS group;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
CONTRAST "Drug A vs. Placebo" GROUP 1 0 -1 ;
MANOVA h = GROUP;
quit;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Drug A vs. Placebo Effect
H = Contrast SSCP Matrix for Drug A vs. Placebo
E = Error SSCP Matrix

S=1 M=1 N=11

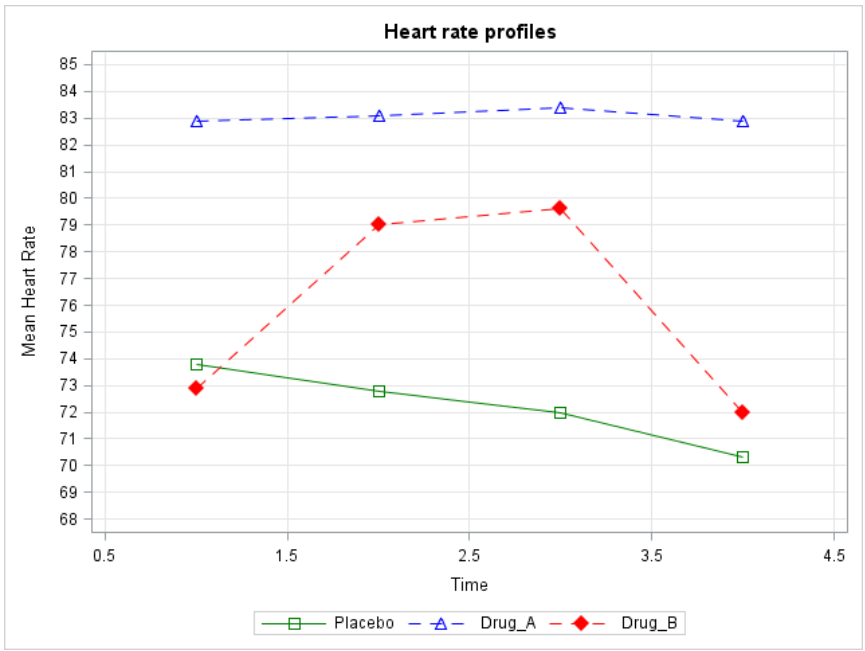
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.31146233	13.26	4	24	<.0001
Pillai's Trace	0.68853767	13.26	4	24	<.0001
Hotelling-Lawley Trace	2.21066118	13.26	4	24	<.0001
Roy's Greatest Root	2.21066118	13.26	4	24	<.0001

Profile Analysis: The profiles plot and table are presented below using SAS.

```
proc means mean data = hrate;
   class group;
   var time1 time2 time3 time4;
   OUTPUT OUT=hrate2;
 run;
data hrate2;
 set hrate2;
 where _stat_="MEAN" and group ne "";
 run;
proc transpose data=hrate2 out=hrate3;
 by group;
 run;
data hrate3;
 set hrate3;
 if _name_ in ("_TYPE_", "_FREQ_") then delete;
 run;
data hrate3;
 set hrate3;
 if _name_="time1" then Time=1;
 if _name_="time2" then Time=2;
 if _name_="time3" then Time=3;
 if _name_="time4" then Time=4;
 run;
data hrate3;
 set hrate3;
 IF GROUP="Placebo" then h_rate1=col1;
 IF GROUP="Drug_A" then h_rate2=col1;
 IF GROUP="Drug_B" then h_rate3=col1;
 RUN;
PROC SGPLOT DATA =HRATE3;
  SERIES X = Time Y = h_rate1 / LEGENDLABEL = 'Placebo'
  MARKERS markerattrs=(symbol=square size=10 color=green)
 lineattrs=(color=green) ;
  SERIES X = Time Y = h_rate2 / LEGENDLABEL = 'Drug_A'
  MARKERS markerattrs=(symbol=triangle size=10 color=blue)
 lineattrs=(color=blue pattern=dash) ;
  SERIES X = Time Y = h_rate3 / LEGENDLABEL = 'Drug_B'
  MARKERS markerattrs=(symbol=diamondfilled size=10 color=red)
 lineattrs=(color=red pattern=dash);
 XAXIS LABEL = 'Time' GRID VALUES = (0.5 TO 4.5 );
 YAXIS LABEL = 'Mean Heart Rate' GRID VALUES = (68 TO 85 );
  TITLE 'Heart rate profiles';
 RUN;
```

The MEANS Procedure

group	N Obs	Variable	Mean
Drug_A	10	time1	82.9000000
		time2	83.1000000
		time3	83.4000000
		time4	82.9000000
Drug_B	10	time1	72.9000000
		time2	79.0000000
		time3	79.6000000
		time4	72.0000000
Placebo	10	time1	73.8000000
		time2	72.8000000
		time3	72.0000000
		time4	70.3000000



Test for Parallelism: The null hypothesis tests if the two drugs and placebo have parallel profiles.

Method I:

```
proc glm data=hrate;
   class group;
   model time1 time2 time3 time4 = group/nouni ;
manova M= TIME1-TIME2, TIME2-TIME3, TIME3-TIME4 H=GROUP/SUMMARY ;
RUN;
QUIT;
```

Method II:

```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
CONTRAST "PARALLEL" GROUP 1 0 -1
                    GROUP 0 1 -1;
MANOVA H=GROUP M=(1 -1 0 0, 1 0 -1 0, 1 0 0 -1)/PRINTE PRINTH;
RUN;
QUIT;
```

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall group Effect on the Variables Defined by the M Matrix Transformation
H = Type III SSCP Matrix for group
E = Error SSCP Matrix

S=2 M=0 N=11.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.20386608	10.12	6	50	<.0001
Pillai's Trace	0.90245442	7.13	6	52	<.0001
Hotelling-Lawley Trace	3.38365968	13.85	6	31.616	<.0001
Roy's Greatest Root	3.22178634	27.92	3	26	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

Test for Separation: The null hypothesis tests if the curves have the same average level. This hypothesis is meaningless in this situation since the parallelism hypothesis was rejected. Nonetheless, for demonstration purposes I will provide the SAS code and output.

Method I:

```
proc glm data=hrate;
  class group;
  model time1 time2 time3 time4 = group/nouni ;
MANOVA M=time1+time2+time3+time4 H=group/summary;
run;
QUIT;
```

Method II:

```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
MANOVA H=GROUP M=(1 1 1 1)/PRINTE PRINTH;
RUN;
QUIT;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall group Effect on the Variables Defined by the M Matrix Transformation
H = Type III SSCP Matrix for group
E = Error SSCP Matrix

S=1 M=0 N=12.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.40003404	20.25	2	27	<.0001
Pillai's Trace	0.59996596	20.25	2	27	<.0001
Hotelling-Lawley Trace	1.49978729	20.25	2	27	<.0001
Roy's Greatest Root	1.49978729	20.25	2	27	<.0001

Test for Flatness: The null hypothesis tests if the the average curve is horizontal. This is the same as testing whether the four heart rate means are equal (see page 45). For completeness, I am providing he SAS code and output again.

Method I:

```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
Manova M=TIME1-TIME2, TIME2-TIME3, TIME3-TIME4 H=INTERCEPT/SUMMARY;
  run;
quit;
```

Method II:

```
PROC GLM DATA=HRATE;
CLASS GROUP;
MODEL TIME1 TIME2 TIME3 TIME4=GROUP/NOUNI;
CONTRAST "Horizontal"  intercept 1;
MANOVA H=GROUP M=(1 -1 0 0,
                  1 0 -1 0,
                  1 0 0 -1)/PRINTE PRINTH;
  run;
quit;
```

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Horizontal Effect on the Variables Defined by the M Matrix Transformation
H = Contrast SSCP Matrix for Horizontal
E = Error SSCP Matrix

S=1 M=0.5 N=11.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.32378036	17.40	3	25	<.0001
Pillai's Trace	0.67621964	17.40	3	25	<.0001
Hotelling-Lawley Trace	2.08851346	17.40	3	25	<.0001
Roy's Greatest Root	2.08851346	17.40	3	25	<.0001

Post Hoc Analysis: Several methods are generally conducted after a MANOVA model including: Simultaneous confidence intervals, Multivariate contrasts, Multiple Univariate ANOVAs, Discriminant Analysis and others. For our example, I will provide the results of the Linear Discriminant Analysis (LDA) to illustrate the classification accuracy of our model.

```
proc discrim data=hrate pool=test listerr out=misclassified;
  class group;
  var time1-time4;
  run;
```

The SAS System

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.HRATE
Resubstitution Summary using Linear Discriminant Function

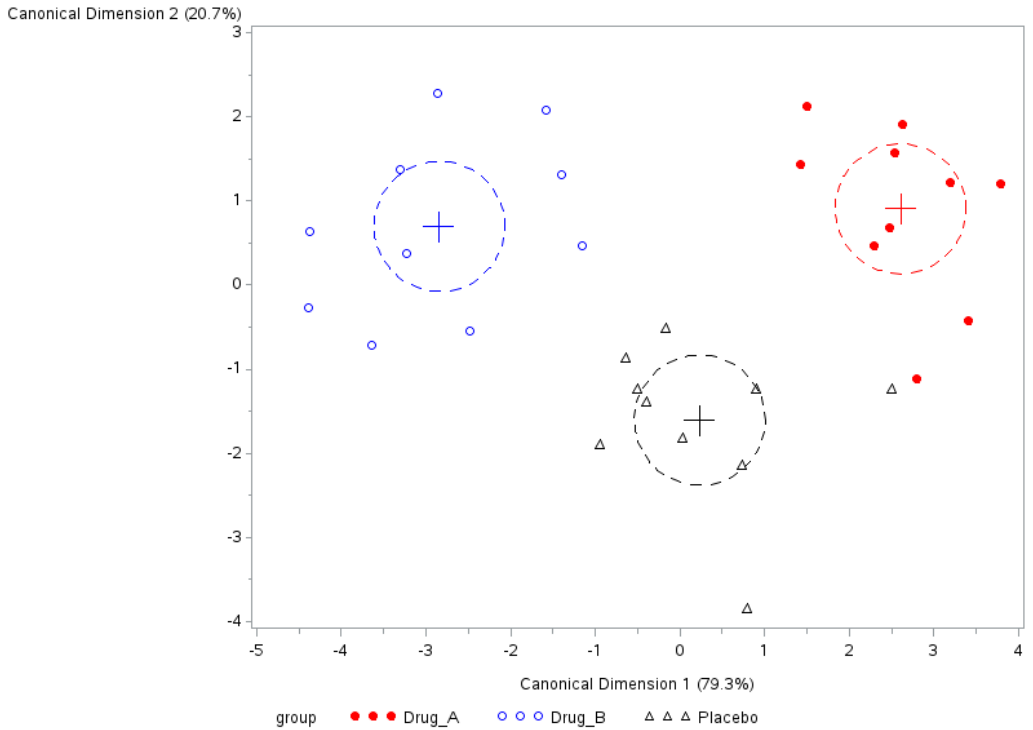
Number of Observations and Percent Classified into group				
From group	Drug_A	Drug_B	Placebo	Total
Drug_A	10 100.00	0 0.00	0 0.00	10 100.00
Drug_B	0 0.00	10 100.00	0 0.00	10 100.00
Placebo	1 10.00	0 0.00	9 90.00	10 100.00
Total	11 36.67	10 33.33	9 30.00	30 100.00
Priors	0.33333	0.33333	0.33333	

The DISCRIM Procedure
Classification Results for Calibration Data: WORK.HRATE
Resubstitution Results using Linear Discriminant Function

Posterior Probability of Membership in group						
Obs	From group	Classified into group		Drug_A	Drug_B	Placebo
9	Placebo	Drug_A	*	0.5959	0.0000	0.4041

In our model, we have only one misclassification for a Placebo into Drug A. This could be also easily seen from the following score plot generated by the SAS %*canplot* Macro (see link in last page).

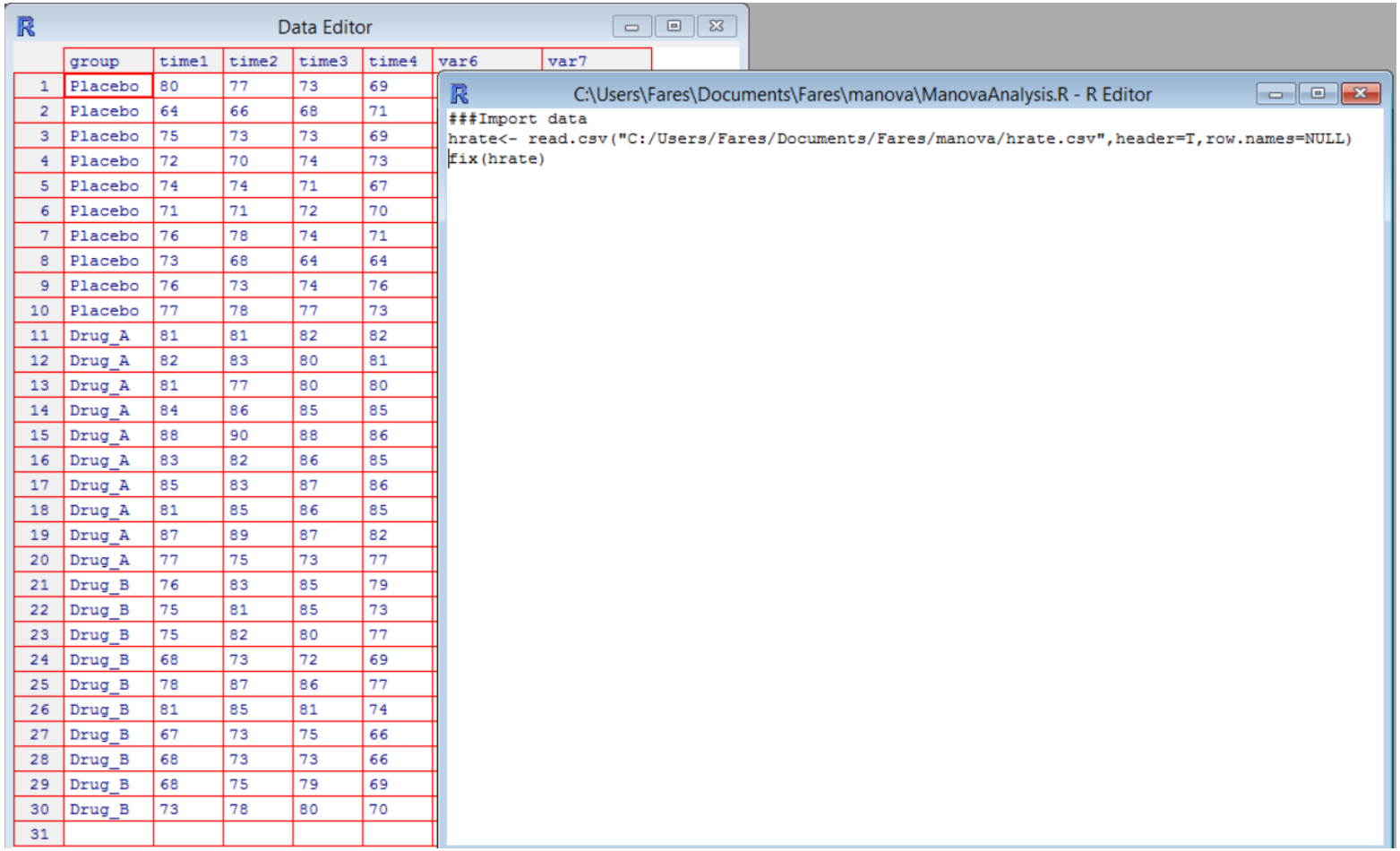
```
%canplot(data=hrate, var=time1-time4, class=group,colors=red blue black);
```



This clear linear discrimination between the three treatments was reflected in the MANOVA analysis previously by the strong "Multivariate R-squared" of 93.72%.

The Solution using R:

- Get the Data: (Please see the last page for a link to the data and R file)



	group	time1	time2	time3	time4	var6	var7
1	Placebo	80	77	73	69		
2	Placebo	64	66	68	71		
3	Placebo	75	73	73	69		
4	Placebo	72	70	74	73		
5	Placebo	74	74	71	67		
6	Placebo	71	71	72	70		
7	Placebo	76	78	74	71		
8	Placebo	73	68	64	64		
9	Placebo	76	73	74	76		
10	Placebo	77	78	77	73		
11	Drug_A	81	81	82	82		
12	Drug_A	82	83	80	81		
13	Drug_A	81	77	80	80		
14	Drug_A	84	86	85	85		
15	Drug_A	88	90	88	86		
16	Drug_A	83	82	86	85		
17	Drug_A	85	83	87	86		
18	Drug_A	81	85	86	85		
19	Drug_A	87	89	87	82		
20	Drug_A	77	75	73	77		
21	Drug_B	76	83	85	79		
22	Drug_B	75	81	85	73		
23	Drug_B	75	82	80	77		
24	Drug_B	68	73	72	69		
25	Drug_B	78	87	86	77		
26	Drug_B	81	85	81	74		
27	Drug_B	67	73	75	66		
28	Drug_B	68	73	73	66		
29	Drug_B	68	75	79	69		
30	Drug_B	73	78	80	70		
31							

- Conduct the MANOVA test:

```
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> summary(fit, test="Wilks")
          Df    Wilks approx F num Df den Df    Pr(>F)
group      2 0.062801   17.942      8     48 4.824e-12 ***
Residuals 27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(fit, test="Pillai")
          Df Pillai approx F num Df den Df    Pr(>F)
group      2 1.4371   15.958      8     50 2.181e-11 ***
Residuals 27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(fit, test="Hotelling-Lawley")
          Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
group      2           6.9625   20.017      8     46 1.317e-12 ***
Residuals 27
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(fit, test="Roy")
          Df    Roy approx F num Df den Df    Pr(>F)
group      2 5.5204   34.502      4     25 7.681e-10 ***
Residuals 27
---
```

Note: The summary of the *manova* function in R doesn't output the results of the four tests ("Pillai", "Wilks", "Hotelling-Lawley" and "Roy") at once. It provides the results of one test at a time. To get the results of one of the four tests, one needs to specify the name of the test within the *summary* command by using the option *test* = "....". Alternatively, one could use the *lm* and *Manova* functions to have all four tests printed together as follows.

```
> library(car)
> fit<-lm(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> table<- Manova(fit)
> summary(table,multivariate=TRUE)

Type II MANOVA Tests:

Sum of squares and products for error:
      time1 time2 time3 time4
time1 479.4 483.7 363.0 237.5
time2 483.7 610.5 475.6 319.7
time3 363.0 475.6 540.8 366.4
time4 237.5 319.7 366.4 381.0

------------------------------------------

Term: group

Sum of squares and products for the hypothesis:
         time1 time2    time3    time4
time1 612.0667 430.5 449.6667 740.4333
time2 430.5000 537.8 600.4000 616.7000
time3 449.6667 600.4 673.8667 659.9333
time4 740.4333 616.7 659.9333 934.8667

Multivariate Tests: group
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            2  1.437149 15.95836      8     50 2.1807e-11 ***
Wilks             2  0.062801 17.94242      8     48 4.8238e-12 ***
Hotelling-Lawley  2  6.962455 20.01706      8     46 1.3168e-12 ***
Roy               2  5.520367 34.50229      4     25 7.6810e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

The parameters' estimates of the MANOVA model are presented as follows:

```
> fit<-lm(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> summary(fit)
Response time1 :

Call:
lm(formula = time1 ~ group, data = hrate)

Residuals:
   Min     1Q Median     3Q    Max
 -9.80  -1.90   0.15   2.20   8.10

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    82.900      1.332  62.214  < 2e-16 ***
groupDrug_B   -10.000      1.884  -5.307 1.34e-05 ***
groupPlacebo   -9.100      1.884  -4.829 4.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.214 on 27 degrees of freedom
Multiple R-squared:  0.5608,    Adjusted R-squared:  0.5282
F-statistic: 17.24 on 2 and 27 DF,  p-value: 1.501e-05


Response time2 :

Call:
lm(formula = time2 ~ group, data = hrate)

Residuals:
   Min     1Q Median     3Q    Max
 -8.10  -3.70   0.05   3.75   8.00

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    83.100      1.504  55.264  < 2e-16 ***
groupDrug_B    -4.100      2.127  -1.928   0.0644 .
groupPlacebo  -10.300      2.127  -4.844 4.64e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.755 on 27 degrees of freedom
Multiple R-squared:  0.4683,    Adjusted R-squared:  0.429
F-statistic: 11.89 on 2 and 27 DF,  p-value: 0.0001977
```

```
Response time3 :

Call:
lm(formula = time3 ~ group, data = hrate)

Residuals:
   Min     1Q Median     3Q    Max
 -10.4   -2.9    1.0    2.6    6.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    83.400      1.415  58.929  < 2e-16 ***
groupDrug_B    -3.800      2.001  -1.899   0.0684 .
groupPlacebo  -11.400      2.001  -5.696 4.73e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.475 on 27 degrees of freedom
Multiple R-squared:  0.5548,    Adjusted R-squared:  0.5218
F-statistic: 16.82 on 2 and 27 DF,  p-value: 1.802e-05


Response time4 :

Call:
lm(formula = time4 ~ group, data = hrate)

Residuals:
   Min     1Q Median     3Q    Max
-6.300 -2.675  0.200  2.550  7.000

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    82.900      1.188  69.787  < 2e-16 ***
groupDrug_B   -10.900      1.680  -6.488 5.91e-07 ***
groupPlacebo  -12.600      1.680  -7.500 4.55e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.756 on 27 degrees of freedom
Multiple R-squared:  0.7105,    Adjusted R-squared:  0.689
F-statistic: 33.13 on 2 and 27 DF,  p-value: 5.409e-08


> |
```

- Test the homogeneity assumption: In this assumption, we test the null hypothesis $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = 0$.

```
> library(biotools)
> boxM(hrate[,2:5], hrate[,1])

        Box's M-test for Homogeneity of Covariance Matrices

data:  hrate[, 2:5]
Chi-Sq (approx.) = 24.4079, df = 20, p-value = 0.225

> |
```

The Box's M test suggests that the data from all groups have common variance-covariance matrix ($p = 0.225 > 0.05$) so this assumptions wasn't violated.

- Note that $hrate[, 2:5]$ contains the dependent variables time1, time2, time3 and time4 while $hrate[, 1]$ contains the independent variable group (i.e. an indicator variable for Drug A, Drug B and Placebo).

- Test the Normality assumption: To test the null hypothesis $H_0 : \epsilon \sim N_4(0, \Sigma)$, in R, we firstly use the Shapiro-Wilk test for univariate normality. Secondly, to be consistent with STATA, we use the Doornik-Hansen test for bivariate normality. Thirdly, we use the Mardia Skewness test for multivariate normality to be consistent with both STATA and SAS.

Shapiro-Wilk test for univariate normality:

```
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> resid<-data.frame(residuals(fit))
> shapiro.test(resid$time1)

        Shapiro-Wilk normality test

data:  resid$time1
W = 0.9825, p-value = 0.8877

> shapiro.test(resid$time2)

        Shapiro-Wilk normality test

data:  resid$time2
W = 0.9607, p-value = 0.323

> shapiro.test(resid$time3)

        Shapiro-Wilk normality test

data:  resid$time3
W = 0.9406, p-value = 0.09455

> shapiro.test(resid$time4)

        Shapiro-Wilk normality test

data:  resid$time4
W = 0.9671, p-value = 0.4637
```

Doornik-Hansen test for bivariate normality:

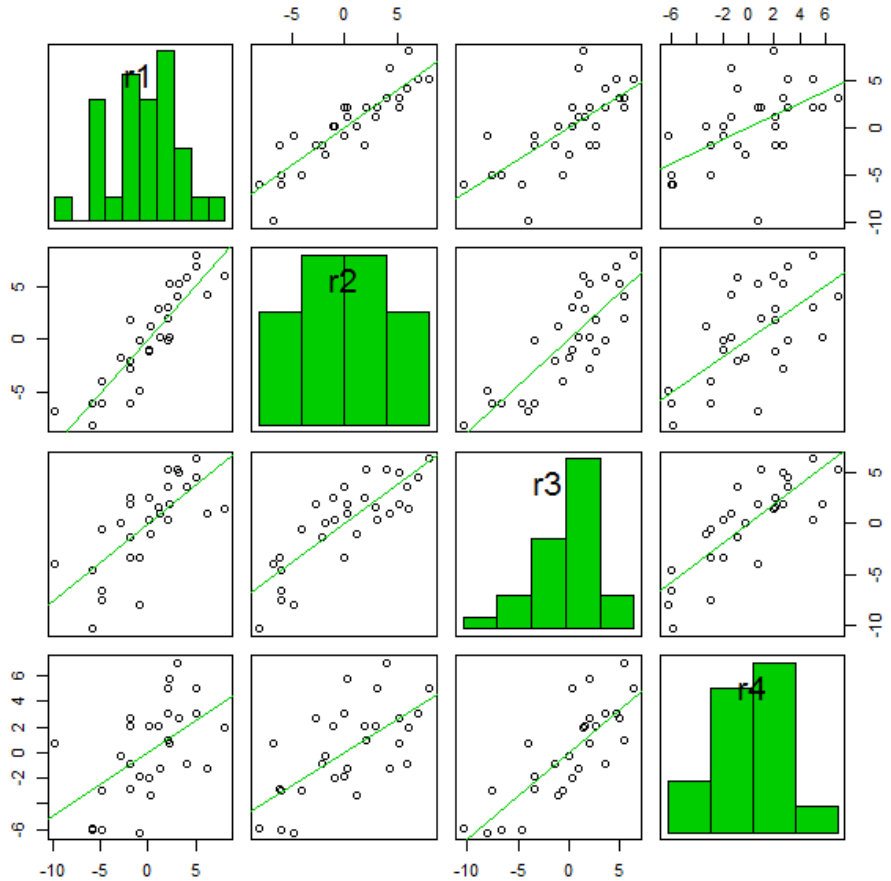
```
> library(asbio)
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> resid<-data.frame(residuals(fit))
> DH.test(resid[,c(1,2)])$multi
         E df P(Chi > E)
1 5.604932  4  0.2306587
> DH.test(resid[,c(1,3)])$multi
         E df P(Chi > E)
1 10.82174  4 0.02864212
> DH.test(resid[,c(1,4)])$multi
       E df P(Chi > E)
1 5.2086  4  0.2665556
> DH.test(resid[,c(2,3)])$multi
         E df P(Chi > E)
1 5.768796  4  0.2170929
> DH.test(resid[,c(2,4)])$multi
         E df P(Chi > E)
1 1.890408  4  0.7559069
> DH.test(resid[,c(3,4)])$multi
       E df P(Chi > E)
1 4.3531  4  0.3603231
>
```

Mardia Skewness test for multivariate normality:

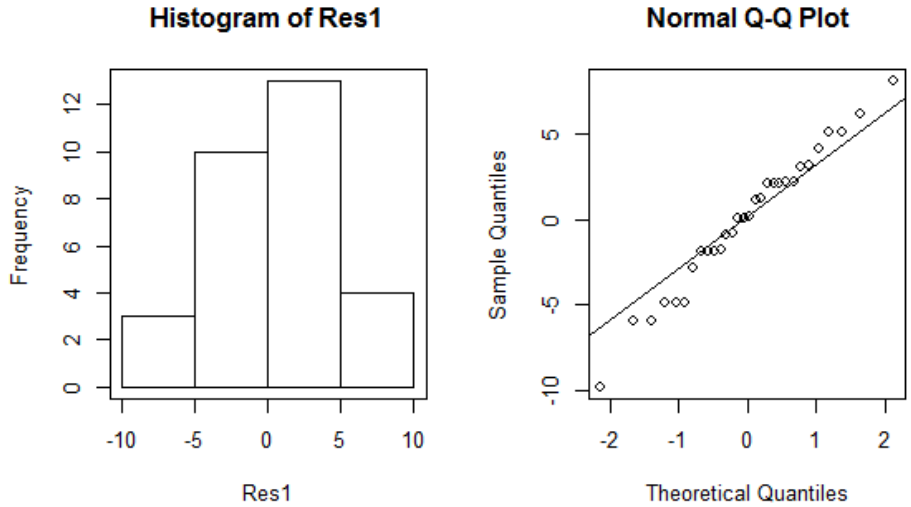
```
> library(MVN)
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> resid<-data.frame(residuals(fit))
> result<-mardiaTest(resid[,1:4], qqplot = TRUE)
> result
   Mardia's Multivariate Normality Test
---------------------------------------
   data : resid[, 1:4]

   g1p             : 2.215629
   chi.skew        : 11.07814
   p.value.skew    : 0.9441708

   g2p             : 20.61932
   z.kurtosis      : -1.336331
   p.value.kurt    : 0.1814412

   chi.small.skew  : 12.67667
   p.value.small   : 0.8908219

   Result          : Data are multivariate normal.
---------------------------------------
```

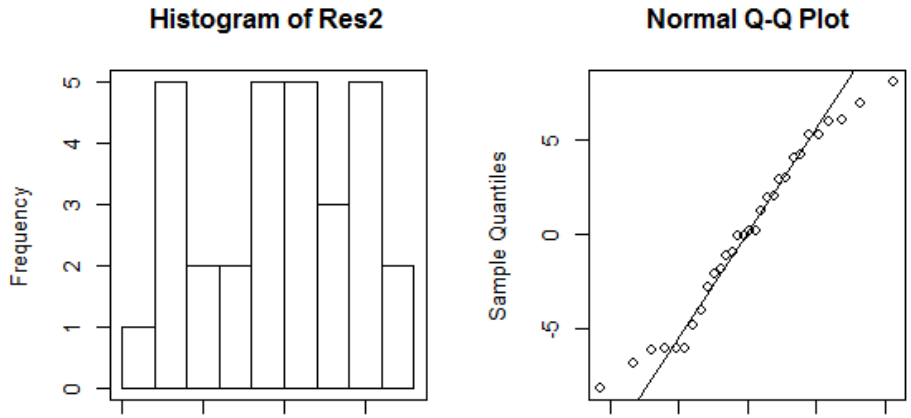
To, graphically, assess multivariate normality, we firstly examine the bivariate scatterplots for each pair of the residuals' vectors hopping to observe an elliptical shape and secondly look at the histogram of each vector of the residuals with the corresponding QQplot:

```
> library(car)
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> resid<-data.frame(residuals(fit))
> names(resid)<-c("r1","r2","r3","r4")
> scatterplotMatrix(resid[,1:4],diagonal="histogram",smooth=FALSE)
>
```

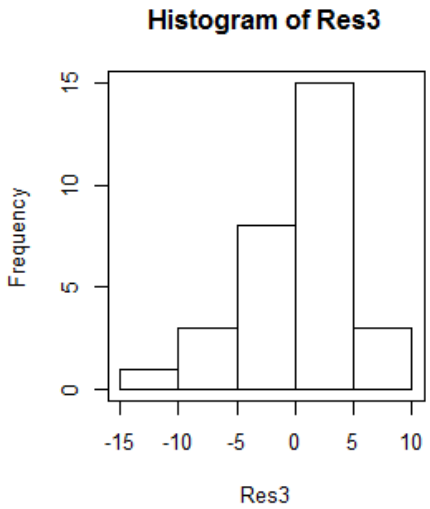


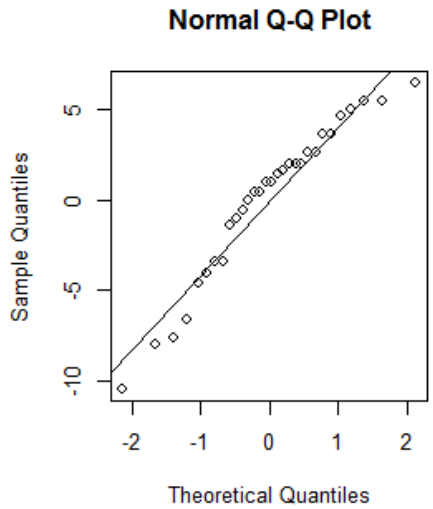
```
> par(mfrow=c(2,2))
> hist(resid[,1], main='Histogram of Res1', xlab='Res1')
> box()
> qqnorm(resid[,1])
> qqline(resid[,1])
> hist(resid[,2], main='Histogram of Res2', xlab='Res2')
> box()
> qqnorm(resid[,2])
> qqline(resid[,2])
```

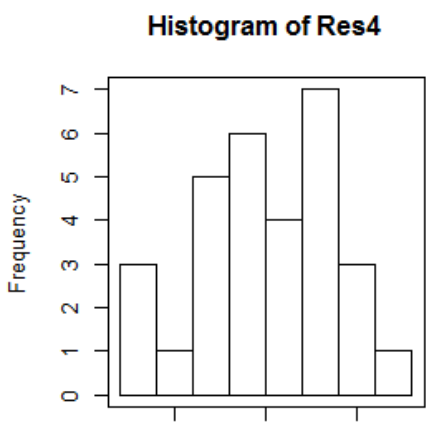


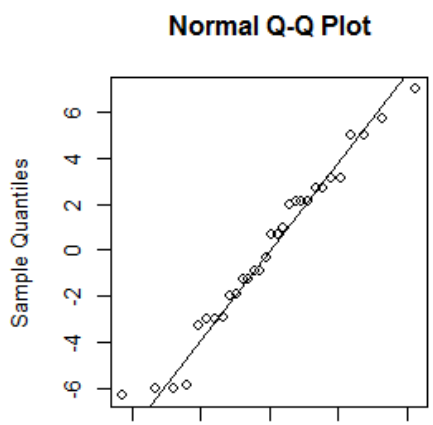


```
> par(mfrow=c(2,2))
> hist(resid[,3], main='Histogram of Res3', xlab='Res3')
> box()
> qqnorm(resid[,3])
> qqline(resid[,3])
> hist(resid[,4], main='Histogram of Res4', xlab='Res4')
> box()
> qqnorm(resid[,4])
> qqline(resid[,4])
```





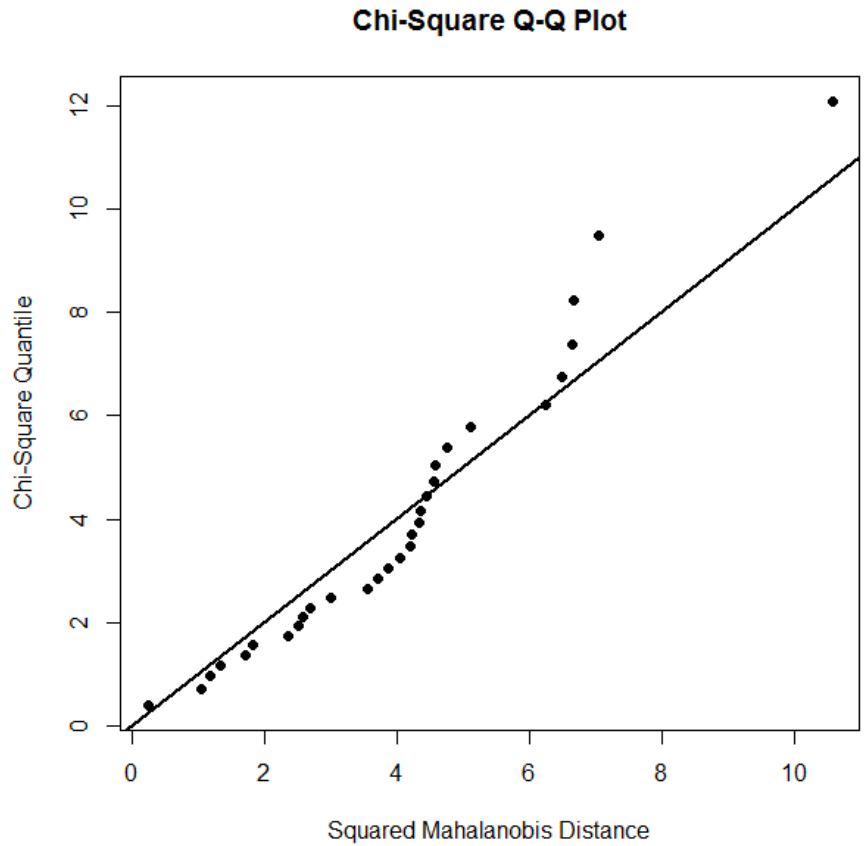




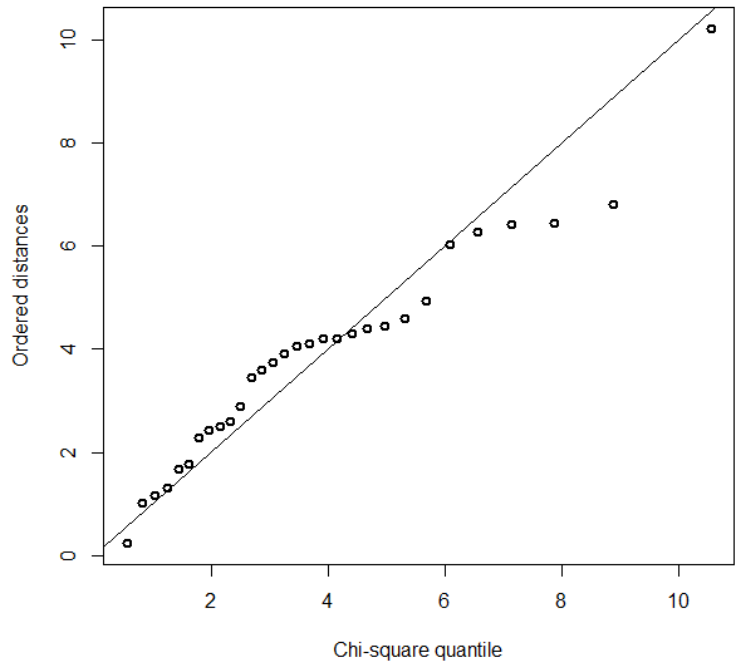
- Test the assumption of Absence of Multivariate Outliers:

To examine multivariate outliers in the data, we use the QQPlot for the observed Mahalanobis distances (MD). This is done in R via either the *mardiaTest* function or the *chisplott* function provided by Everitt [12].

```
> library(MVN)
> fit <- manova(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> resid<-data.frame(residuals(fit))
> result<-mardiaTest(resid[,1:4], qqplot = TRUE)
```



```
> chisplot <- function(x) {
+     if (!is.matrix(x)) stop("x is not a matrix")
+     ### determine dimensions
+     n <- nrow(x)
+     p <- ncol(x)
+     xbar <- apply(x, 2, mean)
+     S <- var(x)
+     S <- solve(S)
+     index <- (1:n)/(n+1)
+     xcent <- t(t(x) - xbar)
+     di <- apply(xcent, 1, function(x,S) x %*% S %*% x,S)
+     quant <- qchisq(index,p)
+     plot(quant, sort(di), ylab = "Ordered distances",
+          xlab = "Chi-square quantile", lwd=2,pch=1)
+     abline(c(0,1))
+     cbind(sort(di),quant)
+ }
> chisplot(residuals(fit))
```



The observed Mahalanobis distances of our data are presented below.

```
> chisplot(residuals(fit))
                    quant
11   0.2334508  0.5567919
6    1.0095219  0.8219317
14   1.1501895  1.0428797
3    1.3014583  1.2430104
30   1.6657063  1.4314567
10   1.7723141  1.6130119
12   2.2725970  1.7906445
15   2.4203204  1.9664189
5    2.5025029  2.1419085
16   2.6020053  2.3184095
28   2.8967921  2.4970658
7    3.4515393  2.6789492
25   3.5853323  2.8651188
18   3.7400106  3.0566695
4    3.9156285  3.2547782
27   4.0598569  3.4607530
21   4.0964108  3.6760910
29   4.1947323  3.9025504
24   4.2097424  4.1422485
17   4.3111238  4.3977969
19   4.4081532  4.6724987
22   4.4401072  4.9706470
1    4.5976909  5.2979967
13   4.9455949  5.6625492
23   6.0383670  6.0759395
20   6.2767796  6.5560911
9    6.4239505  7.1328461
8    6.4402999  7.8617911
26   6.8117169  8.8668125
2   10.2261046 10.5394374
```

- Note that the quantiles in R are computed slightly different than that in SAS or STATA.

- Test for an overall treatment effect: The null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ is rejected which indicates an existence of treatment effect. That is, at the 5% significance level, we can infer that at least one of the three treatments (Drug A, Drug B or Placebo) has a significant impact on women's heart rate.

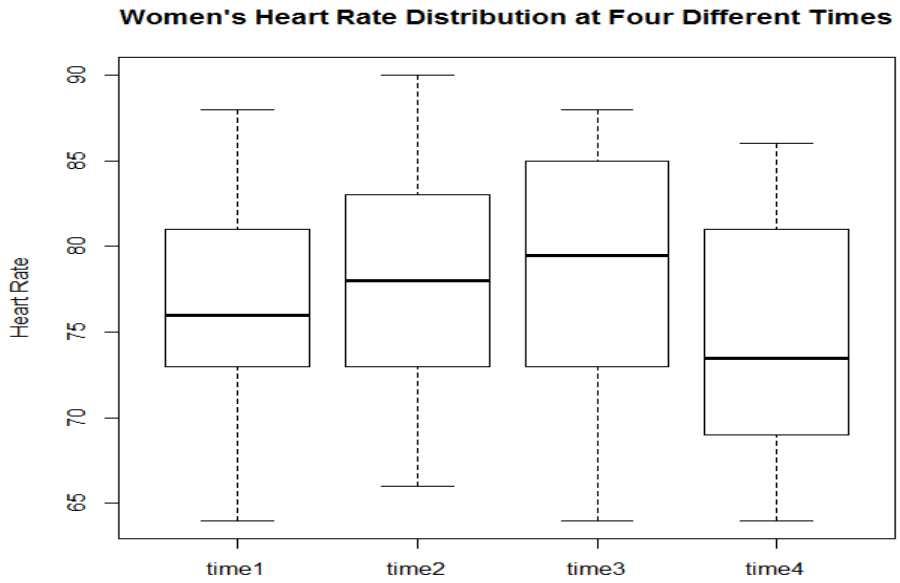
```
> library(car)
> fit<-lm(cbind(time1,time2,time3,time4) ~ group, data=hrate)
> table<- Manova(fit)
> summary(table,multivariate=TRUE)

Type II MANOVA Tests:

Sum of squares and products for error:
      time1 time2 time3 time4
time1 479.4 483.7 363.0 237.5
time2 483.7 610.5 475.6 319.7
time3 363.0 475.6 540.8 366.4
time4 237.5 319.7 366.4 381.0

----------------------------------------

Term: group

Sum of squares and products for the hypothesis:
         time1 time2    time3    time4
time1 612.0667 430.5 449.6667 740.4333
time2 430.5000 537.8 600.4000 616.7000
time3 449.6667 600.4 673.8667 659.9333
time4 740.4333 616.7 659.9333 934.8667

Multivariate Tests: group
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            2  1.437149 15.95836      8     50 2.1807e-11 ***
Wilks             2  0.062801 17.94242      8     48 4.8238e-12 ***
Hotelling-Lawley  2  6.962455 20.01706      8     46 1.3168e-12 ***
Roy               2  5.520367 34.50229      4     25 7.6810e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Note that this output is the same as the default output we get from R when conducting a MANOVA (see page 55)].

- Test whether the four heart rate means are equal: The null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected (see R output and Box-plot Figure below) which indicates, at the 5% significance level, that women's means heart rate at the four times are significantly different.

```
> fit2<-lm(cbind(time1,time2,time3,time4) ~1, data=hrate)
> C<- matrix(c(1))
> M <- matrix(c(1, 0, 0, -1, 1, 0, 0, -1, 1, 0, 0, -1), nrow = 4, by = TRUE)
> linearHypothesis(model = fit2, hypothesis.matrix = C, P = M)

 Response transformation matrix:
      [,1] [,2] [,3]
time1    1    0    0
time2   -1    1    0
time3    0   -1    1
time4    0    0   -1

Sum of squares and products for the hypothesis:
           [,1]        [,2]         [,3]
[1,]   93.633333  1.76666667 -173.133333
[2,]    1.766667  0.03333333   -3.266667
[3,] -173.133333 -3.26666667  320.133333

Sum of squares and products for error:
          [,1]      [,2]       [,3]
[1,]  411.36667  29.23333 -304.86667
[2,]   29.23333 210.96667  -48.73333
[3,] -304.86667 -48.73333  477.86667

Multivariate Tests:
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1 0.4107893 6.274671      3     27 0.0022645 **
Wilks             1 0.5892107 6.274671      3     27 0.0022645 **
Hotelling-Lawley  1 0.6971857 6.274671      3     27 0.0022645 **
Roy               1 0.6971857 6.274671      3     27 0.0022645 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

> ## Alternatively(see [13])
> ############################
> X = as.matrix(hrate[,2:5])
> C = rbind(c(-1, 1, 0, 0), c(0, -1, 1, 0), c(0, 0, -1, 1))
> Y = X%*%t(C)
> dbar = colMeans(Y)
> df1 = length(dbar)
> S = cov(Y)
> cov.est = S/nrow(Y)
> dfS = nrow(Y) - 1
> Tsqd = t(dbar)%*%solve(cov.est, dbar)
> df2 = dfS - df1 + 1
> Fcalc = df2*Tsqd/(df1*dfS)
> pval = pf(Fcalc, df1, df2, lower.tail = 0)
> Fcalc
         [,1]
[1,] 6.274671
> pval
             [,1]
[1,] 0.002264542
> |
```

```
> boxplot(hrate[,2:5],ylab="Heart Rate",
+ main="Women's Heart Rate Distribution at Four Different Times")
> |
```



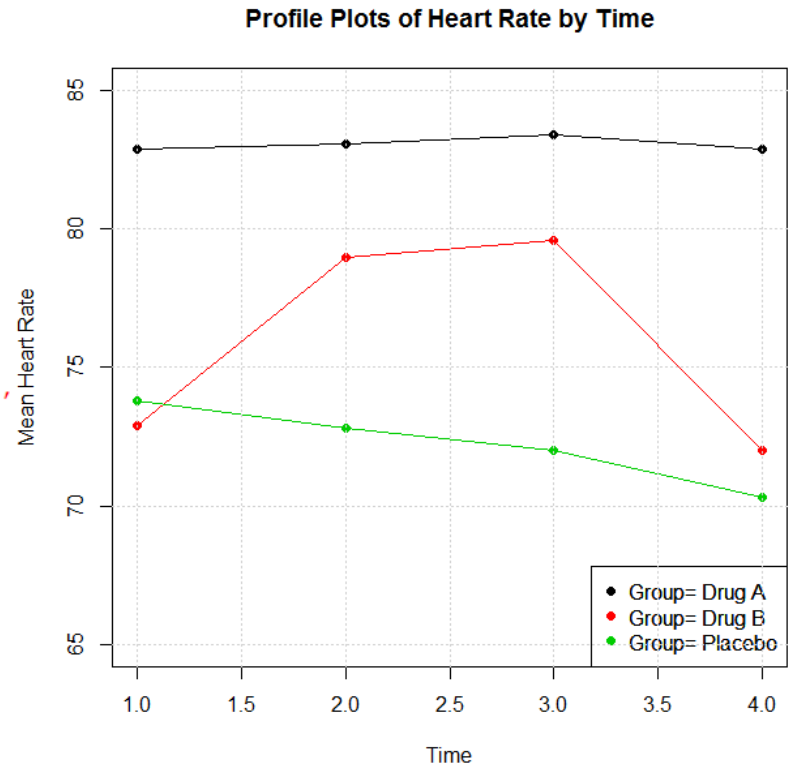

```
> p <- ncol(hrate[,2:5])
> n <- nrow(hrate[,2:5])
> a <- 1 - 0.025/p
> tf <- qt(a, n - 1)
> xbar <- colMeans(hrate[,2:5])
> S <- cov(hrate[,2:5])
> x1L <- xbar[1] - tf * sqrt(S[1, 1]/n)
> x1U <- xbar[1] + tf * sqrt(S[1, 1]/n)
> x2L <- xbar[2] - tf * sqrt(S[2, 2]/n)
> x2U <- xbar[2] + tf * sqrt(S[2, 2]/n)
> x3L <- xbar[3] - tf * sqrt(S[3, 3]/n)
> x3U <- xbar[3] + tf * sqrt(S[3, 3]/n)
> x4L <- xbar[4] - tf * sqrt(S[4, 4]/n)
> x4U <- xbar[4] + tf * sqrt(S[4, 4]/n)
>  xbar
   time1    time2    time3    time4
76.53333 78.30000 78.33333 75.06667
> cat("time1 95% Bonferroni C.I:(", x1L, "-", x1U, ")", "\n", "time2 95% Bonferroni C.I:(", x2L, "-", x2U, ")",
+       "\n", "time3 95% Bonferroni C.I:(", x3L, "-", x3U, ")", "\n","time4 95% Bonferroni C.I:(", x4L, "-", x4U, ")")
time1 95% Bonferroni C.I:( 73.55036 - 79.5163 )
 time2 95% Bonferroni C.I:( 75.24035 - 81.35965 )
 time3 95% Bonferroni C.I:( 75.18651 - 81.48015 )
 time4 95% Bonferroni C.I:( 71.79138 - 78.34195 )> |
```

- Test whether the four heart rate means, for Drug A and Placebo, are equal: The null hypothesis $H_0 : \tau_1 = \tau_3$ is tested via using **c(0, 0, 1)** within the **linearHypothesis** function. In here H_0 is rejected (see R output below). That is, at the 5% significance level, we can infer that the impact of Drug A on women's heart rate is significantly different than that of the Placebo.

```
> fit2<-lm(cbind(time1,time2,time3,time4) ~group, data=hrate)
> linearHypothesis(model = fit2,  c(0, 0, 1))

Sum of squares and products for the hypothesis:
       time1  time2 time3 time4
time1 414.05 468.65 518.7 573.3
time2 468.65 530.45 587.1 648.9
time3 518.70 587.10 649.8 718.2
time4 573.30 648.90 718.2 793.8

Sum of squares and products for error:
      time1 time2 time3 time4
time1 479.4 483.7 363.0 237.5
time2 483.7 610.5 475.6 319.7
time3 363.0 475.6 540.8 366.4
time4 237.5 319.7 366.4 381.0

Multivariate Tests:
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1 0.6885377 13.26397      4     24 7.7196e-06 ***
Wilks             1 0.3114623 13.26397      4     24 7.7196e-06 ***
Hotelling-Lawley  1 2.2106612 13.26397      4     24 7.7196e-06 ***
Roy               1 2.2106612 13.26397      4     24 7.7196e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Profile Analysis: The profiles plot and table are presented below using R.

```
> Profile.Means <- t(aggregate(hrate[, 2:5], by = list(hrate[,1]),FUN = mean))
> Profile.Means
        [,1]      [,2]      [,3]
Group.1 "Drug_A" "Drug_B" "Placebo"
time1   "82.9"   "72.9"   "73.8"
time2   "83.1"   "79.0"   "72.8"
time3   "83.4"   "79.6"   "72.0"
time4   "82.9"   "72.0"   "70.3"
> for (i in 1:3)
+   {
+     if (i == 1)
+       {
+          plot(Profile.Means[-c(1), i], type = "l", col = i, ylim = c(65, 85),
+                main = "Profile Plots of Heart Rate by Time",
+                ylab = "Mean Heart Rate",xlab = "Time")
+          points(Profile.Means[-c(1), i], type = "p", pch = 16, col = i)
+       } else {
+      points(Profile.Means[-c(1), i], type = "l", col = i)
+ points(Profile.Means[-c(1), i], type = "p", pch = 16, col = i)
+    }
+   }
>
> legend("bottomright", pch = 16,legend = mylegend, col = (1:3))
>
> labels1<-c("Drug A", "Drug B", "Placebo")
> mylegend <- paste(paste("Group=", labels1))
> legend("bottomright", pch = 16, legend = mylegend, col = (1:3))
> grid()
>
```



We observe from the profiles plot above that Drug B is different from both Drug A and Placebo. In fact, its profile falls in between the profiles of Drug A and Placebo that both seem to be similar in their behavior over time.

Test for Parallelism: The null hypothesis tests if the two drugs and placebo have parallel profiles.

```
> fit2<-lm(cbind(time1,time2,time3,time4) ~group, data=hrate)
> C <- cbind(rep(0, 3-1), diag(1, 3-1))
> M <- matrix(c(1, 0, 0, -1, 1, 0, 0, -1, 1, 0, 0, -1), nrow = 4, by = TRUE)
> linearHypothesis(model = fit2, hypothesis.matrix = C, P = M)

 Response transformation matrix:
      [,1] [,2] [,3]
time1    1    0    0
time2   -1    1    0
time3    0   -1    1
time4    0    0   -1

Sum of squares and products for the hypothesis:
           [,1]       [,2]       [,3]
[1,]  288.86667  43.43333 -274.46667
[2,]   43.43333  10.86667  -30.23333
[3,] -274.46667 -30.23333  288.86667

Sum of squares and products for error:
      [,1]  [,2]  [,3]
[1,] 122.5 -14.2 -30.4
[2,] -14.2 200.1 -18.5
[3,] -30.4 -18.5 189.0

Multivariate Tests:
                 Df test stat  approx F num Df den Df     Pr(>F)
Pillai            2  0.902454  7.126148      6     52 1.4109e-05 ***
Wilks             2  0.203866 10.123036      6     50 2.6176e-07 ***
Hotelling-Lawley  2  3.383660 13.534639      6     48 6.4250e-09 ***
Roy               2  3.221786 27.922148      3     26 2.7304e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Test for Separation: The null hypothesis tests if the curves have the same average level. This hypothesis is meaningless in this situation since the parallelism hypothesis was rejected. Nonetheless, for demonstration purposes I will provide the R code and output.

```
> ####Test for Separation (coincidental profiles)
> ##############################################
> fit2<-lm(cbind(time1,time2,time3,time4) ~0+group, data=hrate)
> C<- matrix(c(1, 0, -1, 0, 1, -1),ncol = 3, by =T)
> M <- matrix(c(1, 1, 1, 1), nrow = 4, by = TRUE)
> linearHypothesis(model = fit2, hypothesis.matrix = C, P = M)

 Response transformation matrix:
      [,1]
time1    1
time2    1
time3    1
time4    1

Sum of squares and products for the hypothesis:
         [,1]
[1,] 9753.867

Sum of squares and products for error:
       [,1]
[1,] 6503.5

Multivariate Tests:
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            2  0.599966 20.24713      2     27 4.2492e-06 ***
Wilks             2  0.400034 20.24713      2     27 4.2492e-06 ***
Hotelling-Lawley  2  1.499787 20.24713      2     27 4.2492e-06 ***
Roy               2  1.499787 20.24713      2     27 4.2492e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Test for Flatness: The null hypothesis tests if the the average curve is horizontal. This is the same as testing whether the four heart rate means are equal (see page 69). For completeness, I am providing the R code and output again.

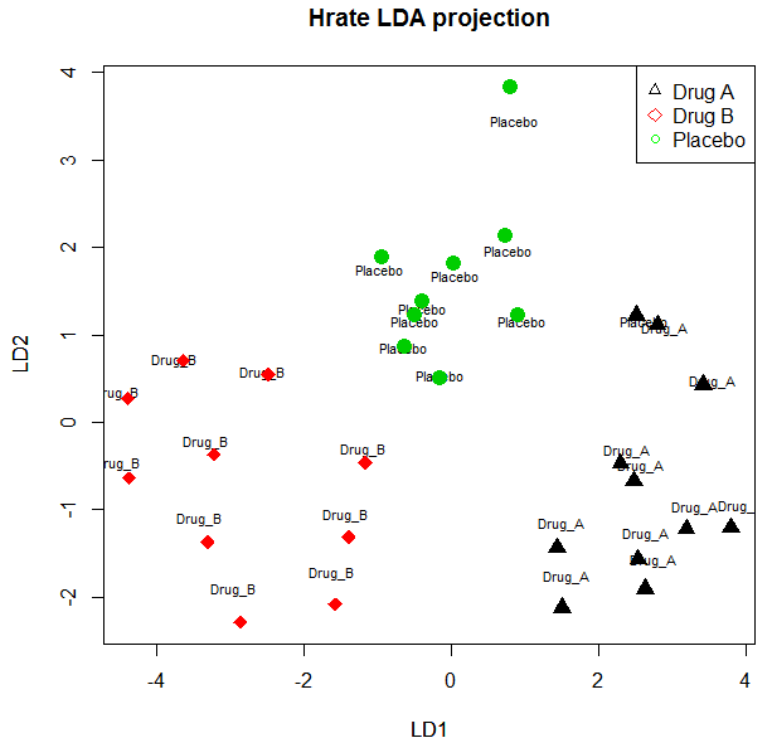
```
> fit2<-lm(cbind(time1,time2,time3,time4) ~1, data=hrate)
> C<- matrix(c(1))
> M <- matrix(c(1, 0, 0, -1, 1, 0, 0, -1, 1, 0, 0, -1), nrow = 4, by = TRUE)
> linearHypothesis(model = fit2, hypothesis.matrix = C, P = M)

 Response transformation matrix:
      [,1] [,2] [,3]
time1    1    0    0
time2   -1    1    0
time3    0   -1    1
time4    0    0   -1

Sum of squares and products for the hypothesis:
            [,1]        [,2]        [,3]
[1,]   93.633333  1.76666667 -173.133333
[2,]    1.766667  0.03333333   -3.266667
[3,] -173.133333 -3.26666667  320.133333

Sum of squares and products for error:
          [,1]      [,2]       [,3]
[1,]  411.36667  29.23333 -304.86667
[2,]   29.23333 210.96667  -48.73333
[3,] -304.86667 -48.73333  477.86667

Multivariate Tests:
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1 0.4107893 6.274671      3     27 0.0022645 **
Wilks             1 0.5892107 6.274671      3     27 0.0022645 **
Hotelling-Lawley  1 0.6971857 6.274671      3     27 0.0022645 **
Roy               1 0.6971857 6.274671      3     27 0.0022645 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>

> ## Alternatively(see [13])
> ############################
> X = as.matrix(hrate[,2:5])
> C = rbind(c(-1, 1, 0, 0), c(0, -1, 1, 0), c(0, 0, -1, 1))
> Y = X%*%t(C)
> dbar = colMeans(Y)
> df1 = length(dbar)
> S = cov(Y)
> cov.est = S/nrow(Y)
> dfS = nrow(Y) - 1
> Tsqd = t(dbar)%*%solve(cov.est, dbar)
> df2 = dfS - df1 + 1
> Fcalc = df2*Tsqd/(df1*dfS)
> pval = pf(Fcalc, df1, df2, lower.tail = 0)
> Fcalc
         [,1]
[1,] 6.274671
> pval
             [,1]
[1,] 0.002264542
>
```

Post Hoc Analysis: Several methods are generally conducted after a MANOVA model including: Simultaneous confidence intervals, Multivariate contrasts, Multiple Univariate ANOVAs, Discriminant Analysis and others. For our example, I will provide the results of the Linear Discriminant Analysis (LDA) to illustrate the classification accuracy of our model.

```
> library(MASS)
> dis = lda(group~time1+time2+time3+time4,data=hrate)
> dis
Call:
lda(group ~ time1 + time2 + time3 + time4, data = hrate)

Prior probabilities of groups:
   Drug_A    Drug_B   Placebo
0.3333333 0.3333333 0.3333333

Group means:
         time1 time2 time3 time4
Drug_A    82.9  83.1  83.4  82.9
Drug_B    72.9  79.0  79.6  72.0
Placebo   73.8  72.8  72.0  70.3

Coefficients of linear discriminants:
              LD1         LD2
time1  0.3841577  0.1985048
time2 -0.3046955 -0.1678573
time3 -0.1833077 -0.1215921
time4  0.3255901 -0.0956523

Proportion of trace:
   LD1    LD2
0.7929 0.2071
> ###Assess the accuracy of the prediction
> clasify <- predict(dis)$class
> table(clasify,group)
         group
clasify   Drug_A Drug_B Placebo
  Drug_A      11      0       0
  Drug_B       0     10       0
  Placebo      0      0       9
>
```

In our model, we have only one misclassification for a Placebo into Drug A. This could be also easily seen from the following score plot generated by R.

```
> library(MASS)
> dis = lda(group~time1+time2+time3+time4,data=hrate)
> plot(dis,xaxt='n',yaxt='n')
> coef(dis)
              LD1         LD2
time1  0.3841577  0.1985048
time2 -0.3046955 -0.1678573
time3 -0.1833077 -0.1215921
time4  0.3255901 -0.0956523
> par(new=TRUE)
> lda.hrate <- lda(group ~ ., hrate)
> lda.pred <- predict(lda.hrate, hrate)
>
> plot(lda.pred$x,
+      pch=as.numeric(lda.pred$class)+16,
+      cex=1.6,
+      col=lda.pred$class)
>
> legend("topright",
+         col=c("black", "red", "green"),
+           pch=c(2,5,1),legend=c("Drug A", "Drug B", "Placebo"))
>
> title(main="Hrate LDA projection")
>
```



This clear linear discrimination between the three treatments was reflected in the MANOVA analysis previously by the strong "Multivariate R-squared" of 93.72%.

Note on Profile Analysis:

A *profile* is a broken line that graphically joins interdependent observations that are measured, generally over time, on the same experimental unit. Profile analysis is a sequential procedure which addresses the following three questions:

- Are the profiles parallel? (looks for Group by Time interaction)
- If so, are the profiles coincidental? (looks for the between groups difference)
- If so, are the profiles horizontal (flat)? (looks for the difference between the DVs means)

Note that:

1. Profile analysis is used only when the DVs are measured on the same scale. If the DVs are measured on different scales, profile analysis could be conducted on the standardized z-scores of the DVs instead.
2. Profile analysis is considered as the multivariate equivalent of repeated measures or mixed ANOVA.
3. As a multivariate method, profile analysis doesn't allow subjects with missing responses.

How to cite this work:

This work was funded by the NIH grants (1U54GM104944-01A1) through the National Institute of General Medical Sciences (NIGMS) under the Institutional Development Award (IDeA) program and the UNM Clinical & Translational Science Center (CTSC) grant (UL1TR001449). Thus, to cite this work please use:

Fares Qeadan (2015). On MANOVA using STATA, SAS & R. A short course in biostatistics for the Mountain West Clinical Translational Research Infrastructure Network (grant 1U54GM104944) and UNM Clinical & Translational Science Center (CTSC) (grant UL1TR001449). University of New Mexico Health Sciences Center. Albuquerque, New Mexico.

References


[1]. Stevens, J. P. (2002). Applied multivariate statistics for the social sciences. Mahwah, NJ: Lawrence Erlbaum Associates.


[2]. James H. Bray, Scott E. Maxwel (1985). Multivariate Analysis of Variance, Issue 54. SAGE Publications, Inc.


[3]. Lisa L. Harlow (2005). The Essence of Multivariate Thinking: Basic Themes and Methods. Psychology Press.


[4]. Gerry P. Quinn, Michael J. Keough (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press.


[5]. Andy Field, Jeremy Miles, Zo Fiel (2012). Discovering Statistics Using R. Sage Publications Ltd.


[6]. Joseph F. Hair Jr, William C. Black, Barry J. Babin, Rolph E. Anderson (2009). Multivariate Data Analysis. Prentice Hall.


[7]. Richard A. Johnson, Dean W. Wichern (2001). Applied Multivariate Statistical Analysis. Prentice Hall.


[8]. Ronald Christense (2001). Advanced Linear Modeling. Springer.


[9]. Sverre Grimnes, Orjan G. Martinse (2008). Bioimpedance and Bioelectricity Basics. Academic Press.


[10] Samuel Kotz, N. Balakrishnan (2006). Encyclopedia of Statistical Sciences (Volume 8). Wiley-Interscience.


[11] Ravindra Khattree, Dayanand N. Naik (2000). Applied Multivariate Statistics With SAS Software. SAS Press/ John Wiley and Sons (Copublished).


[12] Everitt, B. S. (2006). An R and S-PLUS companion to multivariate analysis. Springer Science and Business Media.


[13] Douglas Wiens. `http://www.mathstat.ualberta.ca/~wiens/stat575/stat575.html`. Accessed on July 11, 2015.

Thank you.

For questions, Email: FQeadan@salud.unm.edu

For STATA:

Data: http://www.mathalpha.com/MANOVA/hrate.dta

Do file: http://www.mathalpha.com/MANOVA/stataManova.do

For SAS:

Syntax: http://www.mathalpha.com/MANOVA/ManovaAnalysis.sas

Macro: http://www.mathalpha.com/MANOVA/multnorm.sas

Macro: http://www.mathalpha.com/MANOVA/cqplot.sas

Macro: http://www.mathalpha.com/MANOVA/canplot.sas

For R:

Data: http://www.mathalpha.com/MANOVA/hrate.csv

Script: http://www.mathalpha.com/MANOVA/ManovaAnalysis.R