

## Chapter 9

# One Way MANOVA Type Models

Multivariate regression is the study of the conditional distribution  $\mathbf{y}|\mathbf{x}$  of the  $m \times 1$  vector of response variables  $\mathbf{y}$  given the  $p \times 1$  vector of nontrivial predictors  $\mathbf{x}$ . The multivariate linear model includes the following two models. i) The multivariate linear regression model of Chapter 8 has at least one quantitative predictor variable. ii) For the MANOVA model, the predictors are indicator variables. Often observations  $(Y_1, \dots, Y_m, x_1, x_2, \dots, x_p)$  are collected on the same person or thing and hence are correlated. If transformations can be found such that the  $m$  response plots and residual plots of Section 9.2 look good, and  $n \geq (m + p)^2$  (and  $n_i \geq 10m$  if there are  $p$  treatment groups and  $n = \sum_{i=1}^p n_i$ ), then the MANOVA model can often be used to efficiently analyze the data. These two plots and the DD plot of the residuals are useful for checking the model and for outlier detection.

### 9.1 Introduction

**Definition 9.1.** The **response variables** are the variables that you want to predict. The **predictor variables** are the variables used to predict the response variables.

**Notation.** A multivariate linear model has  $m \geq 2$  response variables. A *multiple linear model = univariate linear model* has  $m = 1$  response variable, but at least two nontrivial predictors, and usually a constant (so  $p \geq 3$ ). A simple linear model has  $m = 1$ , one nontrivial predictor, and usually a constant (so  $p = 2$ ). Multiple linear regression models and ANOVA models are special cases of multiple linear models.

**Definition 9.2.** The **multivariate linear model**

$$\mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \epsilon_i$$

for  $i = 1, \dots, n$  has  $m \geq 2$  response variables  $Y_1, \dots, Y_m$  and  $p$  predictor variables  $x_1, x_2, \dots, x_p$ . The  $i$ th case is  $(\mathbf{x}_i^T, \mathbf{y}_i^T) = (x_{i1}, x_{i2}, \dots, x_{ip}, Y_{i1}, \dots, Y_{im})$ . If a constant  $x_{i1} = 1$  is in the model, then  $x_{i1}$  could be omitted from the case. The model is written in matrix form as  $\mathbf{Z} = \mathbf{X}\mathbf{B} + \mathbf{E}$  where the matrices are the same as those between Definitions 8.2 and 8.3. The model has  $E(\boldsymbol{\epsilon}_k) = \mathbf{0}$  and  $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}\boldsymbol{\epsilon} = (\sigma_{ij})$  for  $k = 1, \dots, n$ . Then the  $p \times m$  coefficient matrix  $\mathbf{B} = [\boldsymbol{\beta}_1 \boldsymbol{\beta}_2 \dots \boldsymbol{\beta}_m]$  and the  $m \times m$  covariance matrix  $\boldsymbol{\Sigma}\boldsymbol{\epsilon}$  are to be estimated, and  $E(\mathbf{Z}) = \mathbf{X}\mathbf{B}$  while  $E(Y_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta}_j$ . The  $\boldsymbol{\epsilon}_i$  are assumed to be iid. The univariate linear model corresponds to  $m = 1$  response variable, and is written in matrix form as  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ . Subscripts are needed for the  $m$  univariate linear models  $\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{e}_j$  for  $j = 1, \dots, m$  where  $E(\mathbf{e}_j) = \mathbf{0}$ . For the multivariate linear model,  $\text{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \sigma_{ij} \mathbf{I}_n$  for  $i, j = 1, \dots, m$  where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix.

**Definition 9.3.** The multivariate analysis of variance (**MANOVA model**)  $\mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \boldsymbol{\epsilon}_i$  for  $i = 1, \dots, n$  has  $m \geq 2$  response variables  $Y_1, \dots, Y_m$  and  $p$  predictor variables  $X_1, X_2, \dots, X_p$ . The MANOVA model is a special case of the multivariate linear model. For the MANOVA model, the predictors are not quantitative variables, so the predictors are indicator variables. Sometimes the trivial predictor  $\mathbf{1}$  is also in the model. In matrix form, the MANOVA model is  $\mathbf{Z} = \mathbf{X}\mathbf{B} + \mathbf{E}$ . The model has  $E(\boldsymbol{\epsilon}_k) = \mathbf{0}$  and  $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}\boldsymbol{\epsilon} = (\sigma_{ij})$  for  $k = 1, \dots, n$ . Also  $E(\mathbf{e}_i) = \mathbf{0}$  while  $\text{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \sigma_{ij} \mathbf{I}_n$  for  $i, j = 1, \dots, m$ . Then  $\mathbf{B}$  and  $\boldsymbol{\Sigma}\boldsymbol{\epsilon}$  are unknown matrices of parameters to be estimated, and  $E(\mathbf{Z}) = \mathbf{X}\mathbf{B}$  while  $E(Y_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta}_j$ .

The data matrix  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}]$ . If the model contains a constant, then usually the first column of ones  $\mathbf{1}$  of  $\mathbf{X}$  is omitted from the data matrix for software such as *R* and *SAS*.

Each response variable in a MANOVA model follows an ANOVA model  $\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{e}_j$  for  $j = 1, \dots, m$  where it is assumed that  $E(\mathbf{e}_j) = \mathbf{0}$  and  $\text{Cov}(\mathbf{e}_j) = \sigma_{jj} \mathbf{I}_n$ . Hence the errors corresponding to the  $j$ th response are uncorrelated with variance  $\sigma_j^2 = \sigma_{jj}$ . Notice that the **same design matrix**  $\mathbf{X}$  of predictors is used for each of the  $m$  models, but the  $j$ th response variable vector  $\mathbf{Y}_j$ , coefficient vector  $\boldsymbol{\beta}_j$ , and error vector  $\mathbf{e}_j$  change and thus depend on  $j$ . Hence for a one way MANOVA model, each response variable follows a one way ANOVA model, while for a two way MANOVA model, each response variable follows a two way ANOVA model for  $j = 1, \dots, m$ .

Once the ANOVA model is fixed, e.g. a one way ANOVA model, the design matrix  $\mathbf{X}$  depends on the parameterization of the ANOVA model. See Chapter 3. The fitted values and residuals are the same for each parameterization, but the interpretation of the parameters depends on the parameterization.

Now consider the  $i$ th case  $(\mathbf{x}_i^T, \mathbf{y}_i^T)$  which corresponds to the  $i$ th row of  $\mathbf{X}$  and the  $i$ th row of  $\mathbf{Z}$ . Then  $\mathbf{y}_i = E(\mathbf{y}_i) + \boldsymbol{\epsilon}_i$  where

$$E(\mathbf{y}_i) = \mathbf{B}^T \mathbf{x}_i = \begin{bmatrix} \mathbf{x}_i^T \boldsymbol{\beta}_1 \\ \mathbf{x}_i^T \boldsymbol{\beta}_2 \\ \vdots \\ \mathbf{x}_i^T \boldsymbol{\beta}_m \end{bmatrix}.$$

The notation  $\mathbf{y}_i|\mathbf{x}_i$  and  $E(\mathbf{y}_i|\mathbf{x}_i)$  is more accurate, but usually the conditioning is suppressed. Taking  $E(\mathbf{y}_i|\mathbf{x}_i)$  to be a constant,  $\mathbf{y}_i$  and  $\boldsymbol{\epsilon}_i$  have the same covariance matrix. In the MANOVA model, this covariance matrix  $\boldsymbol{\Sigma}\boldsymbol{\epsilon}$  does not depend on  $i$ . Observations from different cases are uncorrelated (often independent), but the  $m$  errors for the  $m$  different response variables for the *same case* are correlated.

Let  $\hat{\mathbf{B}}$  be the MANOVA estimator of  $\mathbf{B}$ . MANOVA models are often fit by least squares. Then the **least squares estimators** are

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}_g = (\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T \mathbf{Z} = [\hat{\boldsymbol{\beta}}_1 \hat{\boldsymbol{\beta}}_2 \dots \hat{\boldsymbol{\beta}}_m]$$

where  $(\mathbf{X}^T \mathbf{X})^-$  is a generalized inverse of  $\mathbf{X}^T \mathbf{X}$ . Here  $\hat{\mathbf{B}}_g$  depends on the generalized inverse. If  $\mathbf{X}$  has full rank  $p$  then  $(\mathbf{X}^T \mathbf{X})^- = (\mathbf{X}^T \mathbf{X})^{-1}$  and  $\hat{\mathbf{B}}$  is unique.

**Definition 9.4.** The *predicted values* or *fitted values*

$$\hat{\mathbf{Z}} = \mathbf{X} \hat{\mathbf{B}} = [\hat{\mathbf{Y}}_1 \hat{\mathbf{Y}}_2 \dots \hat{\mathbf{Y}}_m] = \begin{bmatrix} \hat{Y}_{1,1} & \hat{Y}_{1,2} & \dots & \hat{Y}_{1,m} \\ \hat{Y}_{2,1} & \hat{Y}_{2,2} & \dots & \hat{Y}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{Y}_{n,1} & \hat{Y}_{n,2} & \dots & \hat{Y}_{n,m} \end{bmatrix}.$$

The *residuals*  $\hat{\mathbf{E}} = \mathbf{Z} - \hat{\mathbf{Z}} = \mathbf{Z} - \mathbf{X} \hat{\mathbf{B}} =$

$$\begin{bmatrix} \hat{\boldsymbol{\epsilon}}_1^T \\ \hat{\boldsymbol{\epsilon}}_2^T \\ \vdots \\ \hat{\boldsymbol{\epsilon}}_n^T \end{bmatrix} = [\hat{\mathbf{r}}_1 \hat{\mathbf{r}}_2 \dots \hat{\mathbf{r}}_m] = \begin{bmatrix} \hat{\epsilon}_{1,1} & \hat{\epsilon}_{1,2} & \dots & \hat{\epsilon}_{1,m} \\ \hat{\epsilon}_{2,1} & \hat{\epsilon}_{2,2} & \dots & \hat{\epsilon}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\epsilon}_{n,1} & \hat{\epsilon}_{n,2} & \dots & \hat{\epsilon}_{n,m} \end{bmatrix}.$$

These quantities can be found by fitting  $m$  ANOVA models  $\mathbf{Y}_j = \mathbf{X} \boldsymbol{\beta}_j + \mathbf{e}_j$  to get  $\hat{\boldsymbol{\beta}}_j$ ,  $\hat{\mathbf{Y}}_j = \mathbf{X} \hat{\boldsymbol{\beta}}_j$ , and  $\hat{\mathbf{r}}_j = \mathbf{Y}_j - \hat{\mathbf{Y}}_j$  for  $j = 1, \dots, m$ . Hence  $\hat{\epsilon}_{i,j} = Y_{i,j} - \hat{Y}_{i,j}$  where  $\hat{\mathbf{Y}}_j = (\hat{Y}_{1,j}, \dots, \hat{Y}_{n,j})^T$ . Finally,  $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon},d} =$

$$\frac{(\mathbf{Z} - \hat{\mathbf{Z}})^T (\mathbf{Z} - \hat{\mathbf{Z}})}{n - d} = \frac{(\mathbf{Z} - \mathbf{X} \hat{\mathbf{B}})^T (\mathbf{Z} - \mathbf{X} \hat{\mathbf{B}})}{n - d} = \frac{\hat{\mathbf{E}}^T \hat{\mathbf{E}}}{n - d} = \frac{1}{n - d} \sum_{i=1}^n \hat{\boldsymbol{\epsilon}}_i \hat{\boldsymbol{\epsilon}}_i^T.$$

The choices  $d = 0$  and  $d = p$  are common. Let  $\hat{\Sigma}_{\epsilon}$  be the usual estimator of  $\Sigma_{\epsilon}$  for the MANOVA model. If least squares is used with a full rank  $\mathbf{X}$ , then  $\hat{\Sigma}_{\epsilon} = \hat{\Sigma}_{\epsilon, d=p}$ .

## 9.2 Plots for MANOVA Models

As in Chapter 8, this section suggests using residual plots, response plots, and the DD plot to examine the multivariate linear model. The residual plots are often used to check for lack of fit of the multivariate linear model. The response plots are used to check linearity (and to detect influential cases and outliers for linearity). The response and residual plots are used exactly as in the  $m = 1$  case corresponding to multiple linear regression and experimental design models. See Olive (2010, 2017a), Olive and Hawkins (2005), and Cook and Weisberg (1999, p. 432). Chapter 8 used the response and residual plots for MLR for each response variable  $Y_j$ . The one way MANOVA model will use the response and residual plots for the one way ANOVA model for each response variable  $Y_j$ . See Chapter 3.

**Definition 9.5.** A **response plot** for the  $j$ th response variable is a plot of the fitted values  $\hat{Y}_{ij}$  versus the response  $Y_{ij}$ . The identity line with slope one and zero intercept is added to the plot as a visual aid. A **residual plot** corresponding to the  $j$ th response variable is a plot of  $\hat{Y}_{ij}$  versus  $r_{ij}$ .

**Remark 9.1.** Make the  $m$  response and residual plots for any MANOVA model. In a response plot, the vertical deviations from the identity line are the residuals  $r_{ij} = Y_{ij} - \hat{Y}_{ij}$ . Suppose the model is good, the error distribution is not highly skewed, and  $n \geq 10p$ . Then the plotted points should cluster about the identity line in each of the  $m$  response plots. If outliers are present or if the plot is not linear, then the current model or data need to be transformed or corrected. If the model is good, then each of the  $m$  residual plots should be ellipsoidal with no trend and should be centered about the  $r = 0$  line. There should not be any pattern in the residual plot: as a narrow vertical strip is moved from left to right, the behavior of the residuals within the strip should show little change. Outliers and patterns such as curvature or a fan shaped plot are bad.

For some MANOVA models that do not use replication, the response and residual plots look much like those for multivariate linear regression in Section 8.2. The response and residual plots for the one way MANOVA model need some notation, and it is useful to use three subscripts. Suppose there are independent random samples of size  $n_i$  from  $p$  different populations (treatments), or  $n_i$  cases are randomly assigned to  $p$  treatment groups with  $n = \sum_{i=1}^p n_i$ . Assume that  $m$  response variables  $\mathbf{y}_{ij} = (Y_{ij1}, \dots, Y_{ijm})^T$  are measured for the  $i$ th treatment. Hence  $i = 1, \dots, p$  and  $j = 1, \dots, n_i$ . The  $Y_{ijk}$  follow different one

way ANOVA models for  $k = 1, \dots, m$ . Assume  $E(\mathbf{y}_{ij}) = \boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{im})^T$  and  $\text{Cov}(\mathbf{y}_{ij}) = \boldsymbol{\Sigma}_\epsilon$ . Hence the  $p$  treatments have possibly different mean vectors  $\boldsymbol{\mu}_i$ , but common covariance matrix  $\boldsymbol{\Sigma}_\epsilon$ .

Then for the  $k$ th response variable, the *response plot* is a plot of  $\hat{Y}_{ijk} \equiv \hat{\mu}_{ik}$  versus  $Y_{ijk}$  and the *residual plot* is a plot of  $\hat{Y}_{ijk} \equiv \hat{\mu}_{ik}$  versus  $r_{ijk}$  where  $\hat{\mu}_{ik}$  is the sample mean of the  $n_i$  responses  $Y_{ijk}$  corresponding to the  $i$ th treatment for the  $k$ th response variable. Add the identity line to the response plot and  $r = 0$  line to the residual plot as visual aids. The points in the response plot scatter about the identity line and the points in the residual plot scatter about the  $r = 0$  line, but the scatter need not be in an evenly populated band. A *dot plot* of  $Z_1, \dots, Z_n$  consists of an axis and  $n$  points each corresponding to the value of  $Z_i$ . The response plot for the  $k$ th response variable consists of  $p$  dot plots, one for each value of  $\hat{\mu}_{ik}$ . The dot plot corresponding to  $\hat{\mu}_{ik}$  is the dot plot of  $Y_{i,1,k}, \dots, Y_{i,n_i,k}$ . Similarly, the residual plot for the  $k$ th response variable consists of  $p$  dot plots, and the plot corresponding to  $\hat{\mu}_{ik}$  is the dot plot of  $r_{i,1,k}, \dots, r_{i,n_i,k}$ . Assuming the  $n_i \geq 10$ , the  $p$  dot plots for the  $k$ th response variable should have roughly the same shape and spread in both the response and residual plots. Note that  $\hat{\mu}_{ik} = \bar{Y}_{iok} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ijk}$ .

Assume that each  $n_i \geq 10$ . It is easier to check shape and spread in the residual plot. If the response plot looks like the residual plot, then a horizontal line fits the  $p$  dot plots about as well as the identity line, and there may not be much difference in the  $\mu_{ik}$ . In the response plot, if the identity line fits the plotted points better than any horizontal line, then conclude that at least some of the means  $\mu_{ik}$  differ.

**Definition 9.6.** An **outlier** corresponds to a case that is far from the bulk of the data. Look for a large vertical distance of the plotted point from the identity line or the  $r = 0$  line.

**Rule of thumb 9.1.** Mentally add 2 lines parallel to the identity line and 2 lines parallel to the  $r = 0$  line that cover most of the cases. Then a case is an outlier if it is well beyond these 2 lines.

This rule often fails for large outliers since often the identity line goes through or near a large outlier so its residual is near zero. A response that is far from the bulk of the data in the response plot is a “large outlier” (large in magnitude). Look for a large gap between the bulk of the data and the large outlier.

Suppose there is a dot plot of  $n_i$  cases corresponding to treatment  $i$  with mean  $\mu_{ik}$  that is far from the bulk of the data. This dot plot is probably not a cluster of “bad outliers” if  $n_i \geq 4$  and  $n \geq 5p$ . If  $n_i = 1$ , such a case may be a large outlier.

**Rule of thumb 9.2.** Often an outlier is very good, but more often an outlier is due to a measurement error and is very bad.

**Remark 9.2.** Rule of thumb 3.2 for the one way ANOVA  $F$  test may also be useful for the one way MANOVA model tests of hypotheses.

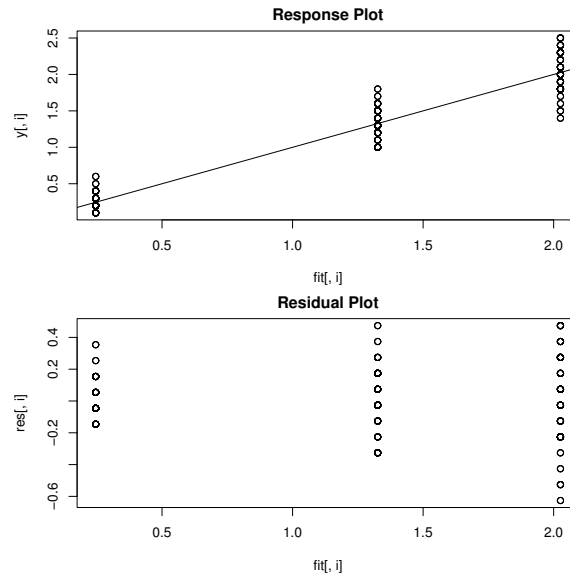
**Remark 9.3.** The above rules are mainly for linearity and tend to use marginal models. The marginal models are useful for checking linearity, but are not very useful for checking other model violations such as outliers in the error vector distribution. The RMVN DD plot of the residual vectors is a global method (takes into account the correlations of  $Y_1, \dots, Y_m$ ) for checking the error vector distribution, but is not real effective for detecting outliers since OLS is used to find the residual vectors. A DD plot of residual vectors from a robust MANOVA method might be more effective for detecting outliers. This remark also applies to the plots used in Section 8.2 for multivariate linear regression.

The RMVN DD plot of the residual vectors  $\hat{\epsilon}_i$  is used to check the error vector distribution, to detect outliers, and to display the nonparametric prediction region developed in Section 8.3. The DD plot suggests that the error vector distribution is elliptically contoured if the plotted points cluster tightly about a line through the origin as  $n \rightarrow \infty$ . The plot suggests that the error vector distribution is multivariate normal if the line is the identity line. If  $n$  is large and the plotted points do not cluster tightly about a line through the origin, then the error vector distribution may not be elliptically contoured. These applications of the DD plot for iid multivariate data are discussed in Olive (2002, 2008, 2013a) and Chapter 7. The RMVN estimator has not yet been proven to be a consistent estimator for residual vectors, but simulations suggest that the RMVN DD plot of the residual vectors is a useful diagnostic plot.

Response transformations can also be made as in Section 1.2, but also make the response plot of  $\hat{Y}_j$  versus  $Y_j$  and use the rules of Section 1.2 on  $Y_j$  to linearize the response plot for each of the  $m$  response variables  $Y_1, \dots, Y_m$ .

**Example 9.1.** Consider the one way MANOVA model on the famous iris data set with  $n = 150$  and  $p = 3$  species of iris: setosa, versicolor, and virginica. The  $m = 4$  variables are  $Y_1 = \text{sepal length}$ ,  $Y_2 = \text{sepal width}$ ,  $Y_3 = \text{petal length}$ , and  $Y_4 = \text{petal width}$ . See Becker et al. (1988). The plots for the  $m = 4$  response variables look similar, and Figure 9.1 shows the response and residual plots for  $Y_4$ . Note that the spread of the three dot plots is similar. The dot plot intersects the identity line at the sample mean of the cases in the dot plot. The setosa cases in lowest dot plot have a sample mean of 0.246 and the horizontal line  $Y_4 = 0.246$  is below the dot plots for versicolor and virginica which have means of 1.326 and 2.026. Hence the mean petal widths differ for the three species, and it is easier to see this difference in the response plot than the residual plot. The plots for the other three variables are similar. Figure 9.2 shows that the DD plot of the residual vectors suggests that the error vector distribution is elliptically contoured but not multivariate normal.

The DD plot also shows the prediction regions of Section 8.3 computed using the residual vectors  $\hat{\epsilon}_i$ . From Section 8.3, if  $\{\hat{\epsilon}|D_{\hat{\epsilon}}(\mathbf{0}, \mathbf{S}_r) \leq h\}$  is a prediction region for the residual vectors, then  $\{\mathbf{y}|D_{\mathbf{y}}(\hat{\mathbf{y}}_f, \mathbf{S}_r) \leq h\}$  is a prediction region for  $\mathbf{y}_f$ . For the one way MANOVA model, a prediction region for  $\mathbf{y}_f$  would only be valid for an  $\mathbf{x}_f$  which was observed, i.e., for  $\mathbf{x}_f = \mathbf{x}_j$ , since only observed values of the categorical predictor variables make sense. The 90% nonparametric prediction region corresponds to  $\mathbf{y}$  with distances to the left of the vertical line  $MD = 3.2$ .



**Fig. 9.1** Plots for  $Y_4 = \text{Petal Width}$ .

*R* commands for these two figures are shown below, and will also show the plots for  $Y_1, Y_2$ , and  $Y_3$ . The *linmodpack* function `manovalw` makes the response and residual plots while `ddplot4` makes the DD plot. The last command shows that the `pvalue = 0` for the one way MANOVA test discussed in the following section.

```
library(MASS)
y <- iris[,1:4] #m = 4 = number of response variables
group <- iris[,5]
#p = number of groups = number of dot plots
out<- manovalw(y,p=3,group=group) #right click
#Stop 8 times
ddplot4(out$res) #right click Stop
summary(out$out) #default is Pillai's test
```

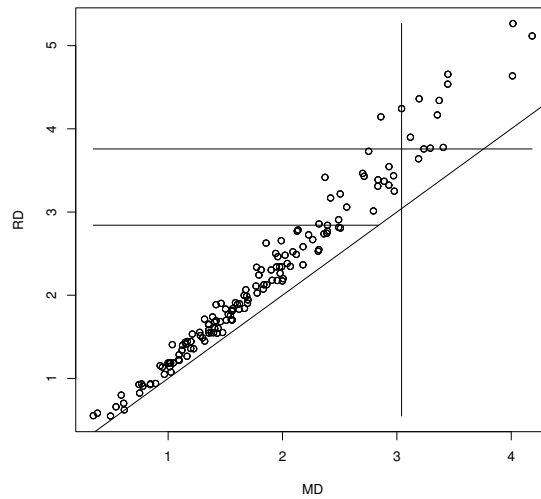


Fig. 9.2 DD Plot of the Residual Vectors for Iris Data.

### 9.3 One Way MANOVA

Using double subscripts will be useful for describing the one way MANOVA model. Suppose there are independent random samples of size  $n_i$  from  $p$  different populations (treatments), or  $n_i$  cases are randomly assigned to  $p$  treatment groups. Then  $n = \sum_{i=1}^p n_i$  and the group sample sizes are  $n_i$  for  $i = 1, \dots, p$ . Assume that  $m$  response variables  $\mathbf{y}_{ij} = (Y_{ij1}, \dots, Y_{ijm})^T$  are measured for the  $i$ th treatment group and the  $j$ th case (often an individual or thing) in the group. Hence  $i = 1, \dots, p$  and  $j = 1, \dots, n_i$ . The  $Y_{ijk}$  follow different one way ANOVA models for  $k = 1, \dots, m$ . Assume  $E(\mathbf{y}_{ij}) = \boldsymbol{\mu}_i$  and  $\text{Cov}(\mathbf{y}_{ij}) = \boldsymbol{\Sigma}_\epsilon$ . Hence the  $p$  treatments have different mean vectors  $\boldsymbol{\mu}_i$ , but common covariance matrix  $\boldsymbol{\Sigma}_\epsilon$ . (The common covariance matrix assumption can be relaxed for  $p = 2$  with the appropriate 2 sample Hotelling's  $T^2$  test.)

The one way MANOVA is used to test  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_p$ . Often  $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$ , so  $H_0$  becomes  $H_0 : \boldsymbol{\tau}_1 = \dots = \boldsymbol{\tau}_p$ . If  $m = 1$ , the one way MANOVA model is the one way ANOVA model. MANOVA is useful since it takes into account the correlations between the  $m$  response variables. Performing  $m$  ANOVA tests fails to account for these correlations, but can be a useful diagnostic. The Hotelling's  $T^2$  test that uses a common covariance matrix is a special case of the one way MANOVA model with  $p = 2$ .

Let  $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$  where  $\sum_{i=1}^p n_i \boldsymbol{\tau}_i = \mathbf{0}$ . The  $j$ th case from the  $i$ th population or treatment group is  $\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\epsilon}_{ij}$  where  $\boldsymbol{\epsilon}_{ij}$  is an error vector,  $i = 1, \dots, p$



and  $j = 1, \dots, n_i$ . Let  $\bar{\mathbf{y}} = \hat{\boldsymbol{\mu}} = \sum_{i=1}^p \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n$  be the overall mean. Let  $\bar{\mathbf{y}}_i = \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n_i$  so  $\hat{\boldsymbol{\tau}}_i = \bar{\mathbf{y}}_i - \bar{\mathbf{y}}$ . Let the residual vector  $\hat{\boldsymbol{\epsilon}}_{ij} = \mathbf{y}_{ij} - \bar{\mathbf{y}}_i = \mathbf{y}_{ij} - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\tau}}_i$ . Then  $\mathbf{y}_{ij} = \bar{\mathbf{y}} + (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}) + (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i) = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\tau}}_i + \hat{\boldsymbol{\epsilon}}_{ij}$ .

Several  $m \times m$  matrices will be useful. Let  $\mathbf{S}_i$  be the sample covariance matrix corresponding to the  $i$ th treatment group. Then the within sum of squares and cross products matrix is  $\mathbf{W} = (n_1 - 1)\mathbf{S}_1 + \dots + (n_p - 1)\mathbf{S}_p = \sum_{i=1}^p \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T$ . Then  $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \mathbf{W} / (n - p)$ . The treatment or between sum of squares and cross products matrix is

$$\mathbf{B}_T = \sum_{i=1}^p n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T.$$

The total corrected (for the mean) sum of squares and cross products matrix is  $\mathbf{T} = \mathbf{B}_T + \mathbf{W} = \sum_{i=1}^p \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}})(\mathbf{y}_{ij} - \bar{\mathbf{y}})^T$ . Note that  $\mathbf{S} = \mathbf{T} / (n - 1)$  is the usual sample covariance matrix of the  $\mathbf{y}_{ij}$  if it is assumed that all  $n$  of the  $\mathbf{y}_{ij}$  are iid so that the  $\boldsymbol{\mu}_i \equiv \boldsymbol{\mu}$  for  $i = 1, \dots, p$ .

The one way MANOVA model is  $\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\epsilon}_{ij}$  where the  $\boldsymbol{\epsilon}_{ij}$  are iid with  $E(\boldsymbol{\epsilon}_{ij}) = \mathbf{0}$  and  $\text{Cov}(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ . The MANOVA table is shown below.

Summary One Way MANOVA Table

Source	matrix	df
Treatment or Between	$\mathbf{B}_T$	$p - 1$
Residual or Error or Within	$\mathbf{W}$	$n - p$
Total (corrected)	$\mathbf{T}$	$n - 1$

If all  $n$  of the  $\mathbf{y}_{ij}$  are iid with  $E(\mathbf{y}_{ij}) = \boldsymbol{\mu}$  and  $\text{Cov}(\mathbf{y}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ , it can be shown that  $\mathbf{A} / df \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$  where  $\mathbf{A} = \mathbf{W}, \mathbf{B}_T$ , or  $\mathbf{T}$ , and  $df$  is the corresponding degrees of freedom. Let  $t_0$  be the test statistic. Often Pillai's trace statistic, the Hotelling Lawley trace statistic, or Wilks' lambda are used. Wilks' lambda

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B}_T + \mathbf{W}|} = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\sum_{i=1}^p (n_i - 1)\mathbf{S}_i|}{|(n - 1)\mathbf{S}|} = \frac{|\sum_{i=1}^p \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T|}{|\sum_{i=1}^p \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}})(\mathbf{y}_{ij} - \bar{\mathbf{y}})^T|}.$$

Then  $t_o = -[n - 0.5(m + p - 2)] \log(\Lambda)$  and  $\text{pval} = P(\chi_{m(p-1)}^2 > t_o)$ . Hence reject  $H_0$  if  $t_0 > \chi_{m(p-1)}^2(1 - \alpha)$ . See Johnson and Wichern (1988, p. 238).

The four steps of the one way MANOVA test follow.

- i) State the hypotheses  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_p$  and  $H_1 : \text{not } H_0$ .
- ii) Get  $t_0$  from output.
- iii) Get pval from output.

iv) State whether you reject  $H_0$  or fail to reject  $H_0$ . If  $pval \leq \alpha$ , reject  $H_0$  and conclude that not all of the  $p$  treatment means are equal. If  $pval > \alpha$ , fail to reject  $H_0$  and conclude that all  $p$  treatment means are equal or that there is not enough evidence to conclude that not all of the  $p$  treatment means are equal. As a textbook convention, use  $\alpha = 0.05$  if  $\alpha$  is not given.

Another way to perform the one way MANOVA test is to get  $R$  output. The default test is Pillai's test, but other tests can be obtained with the  $R$  output shown below.

```
summary(out$out) #default is Pillai's test
summary(out$out, test = "Wilks")
summary(out$out, test = "Hotelling-Lawley")
summary(out$out, test = "Roy")
```

**Example 9.1, continued.** The  $R$  output for the iris data gives a Pillai's  $F$  statistic of 53.466 and  $pval = 0$ .

i)  $H_0 : \mu_1 = \dots = \mu_4$   $H_1 : \text{not } H_0$

ii)  $F = 53.466$

iii)  $pval = 0$

iv) Reject  $H_0$ . The means for the three varieties of iris do differ.

Following Mardia et al. (1979, p. 335), let  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_m$  be the eigenvalues of  $\mathbf{W}^{-1}\mathbf{B}_T$ . Then  $1 + \lambda_i$  for  $i = 1, \dots, m$  are the eigenvalues of  $\mathbf{W}^{-1}\mathbf{T}$  and  $\Lambda = \prod_{i=1}^m (1 + \lambda_i)^{-1}$ .

Following Fujikoshi (2002), let the Hotelling Lawley trace statistic  $U = tr(\mathbf{B}_T\mathbf{W}^{-1}) = tr(\mathbf{W}^{-1}\mathbf{B}_T) = \sum_{i=1}^m \lambda_i$ , and let Pillai's trace statistic  $V = tr(\mathbf{B}_T\mathbf{T}^{-1}) = tr(\mathbf{T}^{-1}\mathbf{B}_T) = \sum_{i=1}^m \frac{\lambda_i}{1 + \lambda_i}$ . If the  $\mathbf{y}_{ij} - \mu_j$  are iid with common covariance matrix  $\Sigma_\epsilon$ , and if  $H_0$  is true, then under regularity conditions  $-[n - 0.5(m + p - 2)] \log(\Lambda) \xrightarrow{D} \chi_{m(p-1)}^2$ ,  $(n - m - p - 1)U \xrightarrow{D} \chi_{m(p-1)}^2$ , and  $(n - 1)V \xrightarrow{D} \chi_{m(p-1)}^2$ . Note that the common covariance matrix assumption implies that each of the  $p$  treatment groups or populations has the same covariance matrix  $\Sigma_i = \Sigma_\epsilon$  for  $i = 1, \dots, p$ , an extremely strong assumption.

**Remark 9.4.** Another method for one way MANOVA is to use the model  $\mathbf{Z} = \mathbf{X}\mathbf{B} + \mathbf{E}$  or

$$\begin{bmatrix} Y_{111} & Y_{112} & \cdots & Y_{11m} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{1,n_1,1} & Y_{1,n_1,2} & \cdots & Y_{1,n_1,m} \\ Y_{211} & Y_{212} & \cdots & Y_{21m} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{2,n_2,1} & Y_{2,n_2,2} & \cdots & Y_{2,n_2,m} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{p,11} & Y_{p,12} & \cdots & Y_{p,1m} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{p,n_p,1} & Y_{p,n_p,2} & \cdots & Y_{p,n_p,m} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,m} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \beta_{p,2} & \cdots & \beta_{p,m} \end{bmatrix} + \mathbf{E}.$$

Then  $\mathbf{X}$  is full rank where the  $i$ th column of  $\mathbf{X}$  is an indicator for group  $i - 1$  for  $i = 2, \dots, p$ ,  $\hat{\beta}_{1k} = \bar{Y}_{pok} = \hat{\mu}_{pk}$  for  $k = 1, \dots, m$ , and

$$\hat{\beta}_{ik} = \bar{Y}_{i-1,ok} - \bar{Y}_{pok} = \hat{\mu}_{i-1,k} - \hat{\mu}_{pk}$$

for  $k = 1, \dots, m$  and  $i = 2, \dots, p$ . Thus testing  $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$  is equivalent to testing  $H_0 : \mathbf{LB} = \mathbf{0}$  where  $\mathbf{L} = [\mathbf{0} \ \mathbf{I}_{p-1}]$ . Such tests are discussed in Section 8.4. Then  $\mathbf{y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ij}$  and

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\mu}_p^T \\ \boldsymbol{\mu}_1^T - \boldsymbol{\mu}_p^T \\ \boldsymbol{\mu}_2^T - \boldsymbol{\mu}_p^T \\ \vdots \\ \boldsymbol{\mu}_{p-2}^T - \boldsymbol{\mu}_p^T \\ \boldsymbol{\mu}_{p-1}^T - \boldsymbol{\mu}_p^T \end{bmatrix}. \tag{9.1}$$

Equation (3.5) used the same  $\mathbf{X}$  for one way ANOVA model with  $m = 1$  as the  $\mathbf{X}$  used in the above one way MANOVA model. Then the MLR  $F$  test was the same as the one way ANOVA  $F$  test. Similarly, if  $\mathbf{L} = (\mathbf{0} \ \mathbf{I}_{p-1})$  then the multivariate linear regression Hotelling Lawley test statistic for testing  $H_0 : \mathbf{LB} = \mathbf{0}$  versus  $H_1 : \mathbf{LB} \neq \mathbf{0}$  is  $U = tr(\mathbf{W}^{-1}\mathbf{H})$  while the Hotelling Lawley test statistic for the one way MANOVA test with  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$  is  $U = tr(\mathbf{W}^{-1}\mathbf{B}_T)$ . Rupasinghe Arachchige Don (2018) showed that these two test statistics are the the same for the above  $\mathbf{X}$  by showing that  $\mathbf{B}_T = \mathbf{H}$ . Here  $\mathbf{H}$  is given in Section 8.4 and is not the hat matrix.

### 9.4 An Alternative Test Based on Large Sample Theory

Large sample theory can be also used to derive a competing test. Let  $\Sigma_i$  be the nonsingular population covariance matrix of the  $i$ th treatment group or population. To simplify the large sample theory, assume  $n_i = \pi_i n$  where  $0 < \pi_i < 1$  and  $\sum_{i=1}^p \pi_i = 1$ . Let  $T_i$  be a multivariate location estimator such that  $\sqrt{n_i}(T_i - \mu_i) \xrightarrow{D} N_m(\mathbf{0}, \Sigma_i)$ , and  $\sqrt{n}(T_i - \mu_i) \xrightarrow{D} N_m\left(\mathbf{0}, \frac{\Sigma_i}{\pi_i}\right)$ . Let  $\mathbf{T} = (T_1^T, T_2^T, \dots, T_p^T)^T$ ,  $\boldsymbol{\nu} = (\mu_1^T, \mu_2^T, \dots, \mu_p^T)^T$ , and  $\mathbf{A}$  be a full rank  $r \times mp$  matrix with rank  $r$ , then a large sample test of the form  $H_0 : \mathbf{A}\boldsymbol{\nu} = \boldsymbol{\theta}_0$  versus  $H_1 : \mathbf{A}\boldsymbol{\nu} \neq \boldsymbol{\theta}_0$  uses

$$\mathbf{A}\sqrt{n}(\mathbf{T} - \boldsymbol{\nu}) \xrightarrow{D} \mathbf{u} \sim N_r\left(\mathbf{0}, \mathbf{A} \operatorname{diag}\left(\frac{\Sigma_1}{\pi_1}, \frac{\Sigma_2}{\pi_2}, \dots, \frac{\Sigma_p}{\pi_p}\right) \mathbf{A}^T\right). \quad (9.2)$$

Let the Wald-type statistic

$$t_0 = [\mathbf{A}\mathbf{T} - \boldsymbol{\theta}_0]^T \left[ \mathbf{A} \operatorname{diag}\left(\frac{\hat{\Sigma}_1}{n_1}, \frac{\hat{\Sigma}_2}{n_2}, \dots, \frac{\hat{\Sigma}_p}{n_p}\right) \mathbf{A}^T \right]^{-1} [\mathbf{A}\mathbf{T} - \boldsymbol{\theta}_0]. \quad (9.3)$$

These results prove the following theorem.

**Theorem 9.1.** Under the above conditions,  $t_0 \xrightarrow{D} \chi_r^2$  if  $H_0$  is true.

This test is due to Rupasinghe Arachchige Don and Olive (2019), and a special case was used by Zhang and Liu (2013) and Konietzschke et al. (2015) with  $T_i = \bar{\mathbf{y}}_i$  and  $\hat{\Sigma}_i = \mathbf{S}_i$ . The  $p = 2$  case gives analogs to the two sample Hotelling's  $T^2$  test. See Rupasinghe Arachchige Don and Pelawa Watagoda (2018). The  $m = 1$  case gives analogs of the one way ANOVA test. If  $m = 1$ , see competing tests in Brown and Forsythe (1974a,b), Olive (2017a, pp. 200-202), and Welch (1947, 1951).

For the one way MANOVA type test, let  $\mathbf{A}$  be the block matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & -\mathbf{I} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & -\mathbf{I} \end{bmatrix}.$$

Let  $\mu_i \equiv \boldsymbol{\mu}$ , let  $H_0 : \mu_1 = \dots = \mu_p$  or, equivalently,  $H_0 : \mathbf{A}\boldsymbol{\nu} = \mathbf{0}$ , and let

$$\mathbf{w} = \mathbf{AT} = \begin{bmatrix} T_1 - T_p \\ T_2 - T_p \\ \vdots \\ T_{p-2} - T_p \\ T_{p-1} - T_p \end{bmatrix}. \tag{9.4}$$

Then  $\sqrt{n}\mathbf{w} \xrightarrow{D} N_{m(p-1)}(\mathbf{0}, \boldsymbol{\Sigma}\mathbf{w})$  if  $H_0$  is true with  $\boldsymbol{\Sigma}\mathbf{w} = (\boldsymbol{\Sigma}_{ij})$  where  $\boldsymbol{\Sigma}_{ij} = \frac{\boldsymbol{\Sigma}_p}{\pi_p}$  for  $i \neq j$ , and  $\boldsymbol{\Sigma}_{ii} = \frac{\boldsymbol{\Sigma}_i}{\pi_i} + \frac{\boldsymbol{\Sigma}_p}{\pi_p}$  for  $i = j$ . Hence

$$t_0 = n\mathbf{w}^T \hat{\boldsymbol{\Sigma}}\mathbf{w} = \mathbf{w}^T \left( \frac{\hat{\boldsymbol{\Sigma}}\mathbf{w}}{n} \right)^{-1} \mathbf{w} \xrightarrow{D} \chi_{m(p-1)}^2$$

as the  $n_i \rightarrow \infty$  if  $H_0$  is true. Here

$$\frac{\hat{\boldsymbol{\Sigma}}\mathbf{w}}{n} = \begin{bmatrix} \frac{\hat{\boldsymbol{\Sigma}}_1}{n_1} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \dots & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \\ \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_2}{n_2} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \dots & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} & \dots & \frac{\hat{\boldsymbol{\Sigma}}_{p-1}}{n_{p-1}} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p} \end{bmatrix} \tag{9.5}$$

is a block matrix where the off diagonal block entries equal  $\hat{\boldsymbol{\Sigma}}_p/n_p$  and the  $i$ th diagonal block entry is  $\frac{\hat{\boldsymbol{\Sigma}}_i}{n_i} + \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p}$  for  $i = 1, \dots, (p-1)$ .

Reject  $H_0$  if

$$t_0 > m(p-1)F_{m(p-1), d_n}(1-\delta) \tag{9.6}$$

where  $d_n = \min(n_1, \dots, n_p)$ . See Theorem 2.25. It may make sense to relabel the groups so that  $n_p$  is the largest  $n_i$  or  $\hat{\boldsymbol{\Sigma}}_p/n_p$  has the smallest generalized variance of the  $\hat{\boldsymbol{\Sigma}}_i/n_i$ . This test may start to outperform the one way MANOVA test if  $n \geq (m+p)^2$  and  $n_i \geq 40m$  for  $i = 1, \dots, p$ .

If  $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$  and  $\hat{\boldsymbol{\Sigma}}_i$  is replaced by  $\hat{\boldsymbol{\Sigma}}$ , we will show that for the one way MANOVA test that  $t_0 = (n-p)U$  where  $U$  is the Hotelling Lawley statistic. For the proof, some results on the vec and Kronecker product will be useful. Following Henderson and Searle (1979),  $vec(\mathbf{G})$  and  $vec(\mathbf{G}^T)$  contain the same elements in different sequences. Define the permutation matrix  $\mathbf{P}_{r,m}$  such that

$$vec(\mathbf{G}) = \mathbf{P}_{r,m}vec(\mathbf{G}^T) \tag{9.7}$$

where  $\mathbf{G}$  is  $r \times m$ . Then  $\mathbf{P}_{r,m}^T = \mathbf{P}_{m,r}$ , and  $\mathbf{P}_{r,m}\mathbf{P}_{m,r} = \mathbf{P}_{m,r}\mathbf{P}_{r,m} = \mathbf{I}_{rm}$ . If  $\mathbf{C}$  is  $s \times m$  and  $\mathbf{D}$  is  $p \times r$ , then

$$\mathbf{C} \otimes \mathbf{D} = \mathbf{P}_{p,s}(\mathbf{D} \otimes \mathbf{C})\mathbf{P}_{m,q}. \tag{9.8}$$

Also

$$(\mathbf{C} \otimes \mathbf{D})\text{vec}(\mathbf{G}) = \text{vec}(\mathbf{D}\mathbf{G}\mathbf{C}^T) = \mathbf{P}_{p,s}(\mathbf{D} \otimes \mathbf{C})\text{vec}(\mathbf{G}^T). \quad (9.9)$$

If  $\mathbf{C}$  is  $m \times m$  and  $\mathbf{D}$  is  $r \times r$ , then  $\mathbf{C} \otimes \mathbf{D} = \mathbf{P}_{r,m}(\mathbf{D} \otimes \mathbf{C})\mathbf{P}_{m,r}$ , and

$$[\text{vec}(\mathbf{G})]^T(\mathbf{C} \otimes \mathbf{D})\text{vec}(\mathbf{G}) = [\text{vec}(\mathbf{G}^T)]^T(\mathbf{D} \otimes \mathbf{C})\text{vec}(\mathbf{G}^T). \quad (9.10)$$

**Theorem 9.2.** For the one way MANOVA test using  $\mathbf{A}$  as defined below Theorem 9.1, let the Hotelling Lawley trace statistic  $U = \text{tr}(\mathbf{W}^{-1}\mathbf{B}_T)$ . Then

$$(n-p)U = t_0 = [\mathbf{A}\mathbf{T} - \boldsymbol{\theta}_0]^T \left[ \mathbf{A} \text{diag} \left( \frac{\hat{\boldsymbol{\Sigma}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}}{n_2}, \dots, \frac{\hat{\boldsymbol{\Sigma}}}{n_p} \right) \mathbf{A}^T \right]^{-1} [\mathbf{A}\mathbf{T} - \boldsymbol{\theta}_0].$$

Hence if the  $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$  and  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_p$  is true, then  $(n-p)U = t_0 \xrightarrow{D} \chi_{m(p-1)}^2$ .

**Proof.** Let  $\mathbf{B}$  and  $\mathbf{X}$  be as in Remark 9.4. Let  $\mathbf{L} = [\mathbf{0} \ \mathbf{I}_{p-1}]$  be an  $s \times p$  matrix with  $s = p-1$ . For this choice of  $\mathbf{X}$ ,  $U = \text{tr}(\mathbf{W}^{-1}\mathbf{B}_T) = \text{tr}(\mathbf{W}^{-1}\mathbf{H})$  by Remark 9.4. Hence by Theorem 8.6,

$$(n-p)U = [\text{vec}(\mathbf{L}\hat{\mathbf{B}})]^T [\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} \otimes (\mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T)^{-1}] [\text{vec}(\mathbf{L}\hat{\mathbf{B}})]. \quad (9.11)$$

Now  $\text{vec}([\mathbf{L}\hat{\mathbf{B}}]^T) = \mathbf{w} = \mathbf{A}\mathbf{T}$  of Equation (9.4) with  $T_i = \bar{\mathbf{y}}_i$ . Then

$$t_0 = \mathbf{w}^T \left( \frac{\hat{\boldsymbol{\Sigma}}\mathbf{w}}{n} \right)^{-1} \mathbf{w}$$

where

$$\frac{\hat{\boldsymbol{\Sigma}}\mathbf{w}}{n} = \mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T \otimes \hat{\boldsymbol{\Sigma}}$$

is given by Equation (9.5) with each  $\hat{\boldsymbol{\Sigma}}_i$  replaced by  $\hat{\boldsymbol{\Sigma}}$ . Thus  $t_0 =$

$$[\text{vec}([\mathbf{L}\hat{\mathbf{B}}]^T)]^T [(\mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T)^{-1} \otimes \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1}] [\text{vec}([\mathbf{L}\hat{\mathbf{B}}]^T)]. \quad (9.12)$$

Then  $t_0 = (n-p)U$  by Equation (9.10) with  $\mathbf{G} = \mathbf{L}\hat{\mathbf{B}}$ .  $\square$

Hence the one way MANOVA test is a special case of Equation (9.3) where  $\boldsymbol{\theta}_0 = \mathbf{0}$  and  $\hat{\boldsymbol{\Sigma}}_i \equiv \hat{\boldsymbol{\Sigma}}$ , but then Theorem 9.1 only holds if  $H_0$  is true and  $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$ . Note that the large sample theory of Theorem 9.1 is trivial compared to the large sample theory of  $(n-p)U$  given in Theorem 9.2. Fujikoshi (2002) showed  $(n-m-p-1)U \xrightarrow{D} \chi_{m(p-1)}^2$  while  $(n-p)U \xrightarrow{D} \chi_{m(p-1)}^2$  by Theorem 9.2 if  $H_0$  is true under the common covariance matrix assumption. There is no contradiction since  $(m+1)U \xrightarrow{P} 0$  as the  $n_i \rightarrow \infty$ . Note the  $\mathbf{A}$  is  $m(p-1) \times mp$ .

For tests corresponding to Theorem 9.1, we will use bootstrap with the prediction region method of Chapter 4 to test  $H_0$  when  $\hat{\Sigma}\mathbf{w}$  or the  $\hat{\Sigma}_i$  are unknown or difficult to estimate. To bootstrap the test  $H_0 : \mathbf{A}\boldsymbol{\nu} = \boldsymbol{\theta}_0$  versus  $H_1 : \mathbf{A}\boldsymbol{\nu} \neq \boldsymbol{\theta}_0$ , use  $Z_n = \mathbf{AT}$ . Take a sample of size  $n_j$  with replacement from the  $n_j$  cases for each group for  $j = 1, 2, \dots, p$  to obtain  $T_j^*$  and  $\mathbf{T}_1^*$ . Repeat  $B$  times to obtain  $\mathbf{T}_1^*, \dots, \mathbf{T}_B^*$ . Then  $Z_i^* = \mathbf{AT}_i^*$  for  $i = 1, \dots, B$ . We will illustrate this method with the analog for the one way MANOVA test for  $H_0 : \mathbf{A}\boldsymbol{\theta} = \mathbf{0}$  which is equivalent to  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_p$ , where  $\mathbf{0}$  is an  $r \times 1$  vector of zeroes with  $r = m(p-1)$ . Then  $Z_n = \mathbf{AT} = \mathbf{w}$  given by Equation (9.4). Hence the  $m(p-1) \times 1$  vector  $Z_i^* = \mathbf{AT}_i^* = ((T_1^* - T_p^*)^T, \dots, (T_{p-1}^* - T_p^*)^T)^T$  where  $T_j$  is a multivariate location estimator (such as the sample mean, coordinatewise median, or trimmed mean), applied to the cases in the  $j$ th treatment group. The prediction region method fails to reject  $H_0$  if  $\mathbf{0}$  is in the resulting confidence region.

We may need  $B \geq 50m(p-1)$ ,  $n \geq (m+p)^2$ , and  $n_i \geq 40m$ . If the  $n_i$  are not large, the one way MANOVA test can be regarded as a regularized estimator, and can perform better than the tests that do not assume equal population covariance matrices. See the simulations in Rupasinghe Arachchige Don and Olive (2019).

If  $H_0 : \mathbf{A}\boldsymbol{\nu} = \boldsymbol{\theta}_0$  is true and if the  $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$  for  $i = 1, \dots, p$ , then

$$t_0 = [\mathbf{AT} - \boldsymbol{\theta}_0]^T \left[ \mathbf{A} \operatorname{diag} \left( \frac{\hat{\Sigma}}{n_1}, \frac{\hat{\Sigma}}{n_2}, \dots, \frac{\hat{\Sigma}}{n_p} \right) \mathbf{A}^T \right]^{-1} [\mathbf{AT} - \boldsymbol{\theta}_0] \stackrel{D}{\rightarrow} \chi_r^2.$$

If  $H_0$  is true but the  $\boldsymbol{\Sigma}_i$  are not equal, we may be able to get a bootstrap cutoff by using

$$t_{0i}^* = [\mathbf{AT}_i^* - \mathbf{AT}]^T \left[ \mathbf{A} \operatorname{diag} \left( \frac{\hat{\Sigma}}{n_1}, \frac{\hat{\Sigma}}{n_2}, \dots, \frac{\hat{\Sigma}}{n_p} \right) \mathbf{A}^T \right]^{-1} [\mathbf{AT}_i^* - \mathbf{AT}] = \\ D^2_{\mathbf{AT}_i^*} \left( \mathbf{AT}, \mathbf{A} \operatorname{diag} \left( \frac{\hat{\Sigma}}{n_1}, \frac{\hat{\Sigma}}{n_2}, \dots, \frac{\hat{\Sigma}}{n_p} \right) \mathbf{A}^T \right).$$

## 9.5 Summary

1) The **multivariate linear model**  $\mathbf{y}_i = \mathbf{B}^T \mathbf{x}_i + \boldsymbol{\epsilon}_i$  for  $i = 1, \dots, n$  has  $m \geq 2$  response variables  $Y_1, \dots, Y_m$  and  $p$  predictor variables  $x_1, x_2, \dots, x_p$ . The  $i$ th case is  $(\mathbf{x}_i^T, \mathbf{y}_i^T) = (x_{i1}, x_{i2}, \dots, x_{ip}, Y_{i1}, \dots, Y_{im})$ . If a constant  $x_{i1} = 1$  is in the model, then  $x_{i1}$  could be omitted from the case. The model is written in matrix form as  $\mathbf{Z} = \mathbf{XB} + \mathbf{E}$ . The model has  $E(\boldsymbol{\epsilon}_k) = \mathbf{0}$  and  $\operatorname{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}\boldsymbol{\epsilon} = (\sigma_{ij})$  for  $k = 1, \dots, n$ . Also  $E(\mathbf{e}_i) = \mathbf{0}$  while  $\operatorname{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \sigma_{ij}\mathbf{I}_n$  for

$i, j = 1, \dots, m$ . Then  $\mathbf{B}$  and  $\Sigma_{\epsilon}$  are unknown matrices of parameters to be estimated, and  $E(\mathbf{Z}) = \mathbf{XB}$  while  $E(Y_{ij}) = \mathbf{x}_i^T \boldsymbol{\beta}_j$ .

The data matrix  $\mathbf{W} = [\mathbf{X} \ \mathbf{Z}]$  except usually the first column  $\mathbf{1}$  of  $\mathbf{X}$  is omitted if  $x_{i,1} \equiv 1$ . The  $n \times m$  matrix

$$\mathbf{Z} = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,m} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n,1} & Y_{n,2} & \dots & Y_{n,m} \end{bmatrix} = [\mathbf{Y}_1 \ \mathbf{Y}_2 \ \dots \ \mathbf{Y}_m] = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix}.$$

The  $n \times p$  matrix

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & \dots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,p} \end{bmatrix} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_p] = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

where often  $\mathbf{v}_1 = \mathbf{1}$ .

The  $p \times m$  matrix

$$\mathbf{B} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,m} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \beta_{p,2} & \dots & \beta_{p,m} \end{bmatrix} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2 \ \dots \ \boldsymbol{\beta}_m].$$

The  $n \times m$  matrix

$$\mathbf{E} = \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \dots & \epsilon_{1,m} \\ \epsilon_{2,1} & \epsilon_{2,2} & \dots & \epsilon_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \epsilon_{n,2} & \dots & \epsilon_{n,m} \end{bmatrix} = [\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_m] = \begin{bmatrix} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{bmatrix}.$$

2) The univariate linear model is  $Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \dots + x_{i,p}\beta_p + e_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i = \boldsymbol{\beta}^T \mathbf{x}_i + e_i$  for  $i = 1, \dots, n$ . In matrix notation, these  $n$  equations become  $\mathbf{Y} = \mathbf{XB} + \mathbf{e}$ , where  $\mathbf{Y}$  is an  $n \times 1$  vector of response variables,  $\mathbf{X}$  is an  $n \times p$  matrix of predictors,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of unknown coefficients, and  $\mathbf{e}$  is an  $n \times 1$  vector of unknown errors.

3) Each response variable in a multivariate linear model follows a univariate linear model  $\mathbf{Y}_j = \mathbf{XB}_j + \mathbf{e}_j$  for  $j = 1, \dots, m$  where it is assumed that  $E(\mathbf{e}_j) = \mathbf{0}$  and  $\text{Cov}(\mathbf{e}_j) = \sigma_{jj} \mathbf{I}_n$ .

4) In a MANOVA model,  $\mathbf{y}_k = \mathbf{B}^T \mathbf{x}_k + \boldsymbol{\epsilon}_k$  for  $k = 1, \dots, n$  is written in matrix form as  $\mathbf{Z} = \mathbf{XB} + \mathbf{E}$ . The model has  $E(\boldsymbol{\epsilon}_k) = \mathbf{0}$  and  $\text{Cov}(\boldsymbol{\epsilon}_k) = \Sigma_{\epsilon} = (\sigma_{ij})$  for  $k = 1, \dots, n$ . Each response variable in a MANOVA model follows



an ANOVA model  $\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{e}_j$  for  $j = 1, \dots, m$  where it is assumed that  $E(\mathbf{e}_j) = \mathbf{0}$  and  $\text{Cov}(\mathbf{e}_j) = \sigma_{jj}\mathbf{I}_n$ .

5) The **one way MANOVA** model is as above where  $\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{e}_j$  is a one way ANOVA model for  $j = 1, \dots, m$ . Check the model by making  $m$  response and residual plots and a DD plot of the residual vectors  $\hat{\boldsymbol{\epsilon}}_i$ .

6) The one way MANOVA model is a generalization of the Hotelling's  $T^2$  test from 2 groups to  $p \geq 2$  groups, assumed to have different means but a common covariance matrix  $\boldsymbol{\Sigma}_\boldsymbol{\epsilon}$ . Want to test  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_p$ . This model is a multivariate linear model so there are  $m$  response variables  $Y_1, \dots, Y_m$  measured for each group. Each  $Y_i$  follows a one way ANOVA model for  $i = 1, \dots, m$ .

7) For the one way MANOVA model, make a DD plot of the residual vectors  $\hat{\boldsymbol{\epsilon}}_i$  where  $i = 1, \dots, n$ . Use the plot to check whether the  $\boldsymbol{\epsilon}_i$  follow a multivariate normal distribution or some other elliptically contoured distribution. We want  $n \geq (m + p)^2$  and  $n_i \geq 10m$ .

8) For the one way MANOVA model, write the data as  $Y_{ijk}$  where  $i = 1, \dots, p$  and  $j = 1, \dots, n_i$ . So  $k$  corresponds to the  $k$ th variable  $Y_k$  for  $k = 1, \dots, m$ . Then  $\hat{Y}_{ijk} = \hat{\mu}_{ik} = \bar{Y}_{iok}$  for  $i = 1, \dots, p$ . So for the  $k$ th variable, the means  $\mu_{1k}, \dots, \mu_{pk}$  are of interest. The residuals are  $r_{ijk} = Y_{ijk} - \hat{Y}_{ijk}$ . For each variable  $Y_k$  make a response plot of  $\bar{Y}_{iok}$  versus  $Y_{ijk}$  and a residual plot of  $\bar{Y}_{iok}$  versus  $r_{ijk}$ . Both plots will consist of  $p$  dot plots of  $n_i$  cases located at the  $\bar{Y}_{iok}$ . The dot plots should follow the identity line in the response plot and the horizontal  $r = 0$  line in the residual plot for each of the  $m$  response variables  $Y_1, \dots, Y_m$ . For each variable  $Y_k$ , let  $R_{ik}$  be the range of the  $i$ th dot plot. If each  $n_i \geq 5$ , we want  $\max(R_{1k}, \dots, R_{pk}) \leq 2 \min(R_{1k}, \dots, R_{pk})$ . The one way MANOVA model may be reasonable for the test in point 9) if the  $m$  response and residual plots satisfy the above graphical checks.

9) The four steps of the one way MANOVA test follow.

i) State the hypotheses  $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_p$  and  $H_1 : \text{not } H_0$ .

ii) Get  $t_0$  from output.

iii) Get pval from output.

iv) State whether you reject  $H_0$  or fail to reject  $H_0$ . If  $\text{pval} \leq \alpha$ , reject  $H_0$  and conclude that not all of the  $p$  treatment means are equal. If  $\text{pval} > \alpha$ , fail to reject  $H_0$  and conclude that all  $p$  treatment means are equal or that there is not enough evidence to conclude that not all of the  $p$  treatment means are equal. Give a nontechnical sentence as the conclusion, if possible. As a textbook convention, use  $\alpha = 0.05$  if  $\alpha$  is not given.

10) The one way MANOVA test assumes that the  $p$  treatment groups or populations have the same covariance matrix:  $\boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_p$ , but the test has some resistance to this assumption. See points 6) and 8).

## 9.6 Complements

The `linmodpack` function `manbtsim2` simulates the bootstrap tests corresponding to Theorem 9.1 using the sample mean, coordinatewise median, and coordinatewise 25% trimmed mean. The function `manbtsim4` adds the test corresponding to Equation (9.6). The function `manbtsim` is like `manbtsim2`, but adds  $T_{RMVN}$  from Definition 7.17 to the simulation, making the simulation very slow. The prediction region method was proven to work for the sample mean, coordinatewise median, and coordinatewise trimmed means in Rupasinghe Arachchige Don and Olive (2019). We only conjecture that the prediction region method works for  $T_{RMVN}$ .

## 9.7 Problems

**10.1\***. If  $\mathbf{X}$  is of full rank and least squares is used to fit the MANOVA model, then  $\hat{\boldsymbol{\beta}}_i = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_i$ , and  $\mathbf{Y}_i = \mathbf{X} \boldsymbol{\beta}_i + \mathbf{e}_i$ . Treating  $\mathbf{X} \boldsymbol{\beta}_i$  as a constant,  $\text{Cov}(\mathbf{Y}_i, \mathbf{Y}_j) = \text{Cov}(\mathbf{e}_i, \mathbf{e}_j) = \sigma_{ij} \mathbf{I}_n$ . Using this information, show  $\text{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j) = \sigma_{ij} (\mathbf{X}^T \mathbf{X})^{-1}$ .