Chapter 10

MANOVA

10.1 Introduction

Definition 10.1. The **response variables** are the variables that you want to predict. The **predictor variables** are the variables used to predict the response variables.

Notation. The **MANOVA model** $\boldsymbol{y}_i = \boldsymbol{B}^T \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ for $i = 1, ..., n$ has $m \geq 2$ response variables $Y_1, ..., Y_m$ and d predictor variables $X_1, X_2, ..., X_d$. The ith case is $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T) = (x_{i1}, ..., x_{id}, Y_{i1}, ..., Y_{im})$. If a constant $x_{i1} = 1$ is in the model, then x_{i1} could be omitted from the case.

For the multivariate analysis of variance (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. The multivariate regression model of Chapter 12 has at least one quantitative variable.

In matrix form, the MANOVA model is $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$, and the data matrix $\boldsymbol{W} = [\boldsymbol{X} \;\; \boldsymbol{Y}]$. The $n \times m$ matrix

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

The $n \times d$ matrix $\boldsymbol{X}$ is not necessarily of full rank d, and

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,d} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_d \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}$$

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

The $d \times m$ matrix

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

The $n \times m$ matrix

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

Warning: The $\boldsymbol{e}_i$ are error vectors, not orthonormal eigenvectors.

Definition 10.2. Models in which a single response variable Y is quantitative, but all of the predictor variables are qualitative are called *analysis of variance* (ANOVA) models, *experimental design* models or *design of experiments* (DOE) models. Each combination of the levels of the predictors gives a different distribution for Y, and there are p different distributions or treatments. A predictor variable W is often called a factor and a factor level a_i is one of the categories W can take. In an ANOVA model,

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,d}\beta_d + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{10.1}$$

for $i = 1, \dots, n$. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{10.2}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of response variables, $\boldsymbol{X}$ is an $n \times d$ matrix of predictors, $\boldsymbol{\beta}$ is a $d \times 1$ vector of unknown coefficients, $\boldsymbol{e}$ is an $n \times 1$ vector

of unknown errors, and $d \geq p$. Equivalently,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,d} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}. \tag{10.3}$$

The e_i are iid with zero mean and variance σ^2, and a linear model estimator such as least squares is used to estimate the unknown parameters $\boldsymbol{\beta}$ and σ^2.

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

Once the ANOVA model is fixed, eg a one way ANOVA model, the design matrix $\boldsymbol{X}$ depends on the parameterization of the ANOVA model. The fitted values and residuals are the same for each parameterization, but the interpretation of the parameters depend on the parameterization.

Now consider the ith case $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T)$ which corresponds to the ith row of $\boldsymbol{Z}$ and the ith row of $\boldsymbol{X}$. Then

$$\begin{bmatrix} Y_{i1} = \beta_{11}x_{i1} + \cdots + \beta_{d1}x_{id} + \epsilon_{i1} = \boldsymbol{x}_i^T\boldsymbol{\beta}_1 + \epsilon_{i1} \\ Y_{i2} = \beta_{12}x_{i1} + \cdots + \beta_{d2}x_{id} + \epsilon_{i2} = \boldsymbol{x}_i^T\boldsymbol{\beta}_2 + \epsilon_{i2} \\ \vdots \\ Y_{im} = \beta_{1m}x_{i1} + \cdots + \beta_{dm}x_{id} + \epsilon_{im} = \boldsymbol{x}_i^T\boldsymbol{\beta}_m + \epsilon_{im} \end{bmatrix}$$

or $\boldsymbol{y}_i = E(\boldsymbol{y}_i) + \boldsymbol{\epsilon}_i$ where

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

The notation $\boldsymbol{y}_i|\boldsymbol{x}_i$ and $E(\boldsymbol{y}_i|\boldsymbol{x}_i)$ is more accurate, but usually the conditioning is suppressed. Taking $E(\boldsymbol{y}_i|\boldsymbol{x}_i)$ to be a constant, $\boldsymbol{y}_i$ and $\boldsymbol{\epsilon}_i$ have

the same covariance matrix. In the MANOVA model, this covariance matrix $\boldsymbol{\Sigma_\epsilon}$ 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.

Definition 10.3. The **MANOVA model** $\boldsymbol{y}_k = \boldsymbol{B}^T\boldsymbol{x}_k + \boldsymbol{\epsilon}_k$ for $k = 1, ..., n$ is written in matrix form as $\boldsymbol{Z} = \boldsymbol{XB} + \boldsymbol{E}$. The model has $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma_\epsilon} = ((\sigma_{ij}))$ for $k = 1, ..., n$. Also $E(\boldsymbol{e}_i) = \boldsymbol{0}$ while $\text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij}\boldsymbol{I}_n$ for $i, j = 1, ..., m$. Then $\boldsymbol{B}$ and $\boldsymbol{\Sigma_\epsilon}$ are unknown matrices of parameters to be estimated, and $E(\boldsymbol{Z}) = \boldsymbol{XB}$ while $E(Y_{ij}) = \boldsymbol{x}_i^T\boldsymbol{\beta}_j$. Considering the kth row of $\boldsymbol{Z}, \boldsymbol{X}$ and $\boldsymbol{E}$ shows that $\boldsymbol{y}_k^T = \boldsymbol{x}_k^T\boldsymbol{B} + \boldsymbol{\epsilon}_k^T$.

10.2 One Way ANOVA

Before describing the one way MANOVA model, it is useful to give a brief description on the one way ANOVA model.

Definition 10.4. A **lurking variable** is not one of the variables in the study, but may affect the relationships among the variables in the study. A **unit** is the experimental material assigned **treatments**, which are the conditions the investigator wants to study. The unit is *experimental* if it was randomly assigned to a treatment, and the unit is *observational* if it was not randomly assigned to a treatment.

Definition 10.5. In an **experiment**, the investigators use **randomization** to assign treatments to units. To assign p treatments to $n = n_1 + \cdots + n_p$ experimental units, draw a random permutation of $\{1, ..., n\}$. Assign the first n_1 units treatment 1, the next n_2 units treatment 2, ..., and the final n_p units treatment p.

Randomization allows one to do valid inference such as F tests of hypotheses and confidence intervals. Randomization also washes out the effects of lurking variables and makes the p treatment groups similar except for the treatment. The effects of lurking variables are present in observational studies defined in Definition 10.6.

Definition 10.6. In an **observational study**, investigators simply observe the response, and the treatment groups need to be p random samples

from p populations (the levels) for valid inference.

Example 10.1. Consider using randomization to assign the following nine people (units) to three treatment groups.

Carroll, Collin, Crawford, Halverson, Lawes,
Stach, Wayman, Wenslow, Xumong

Balanced designs have the group sizes the same: $n_i \equiv h = n/p$. Label the units alphabetically so Carroll gets 1, ..., Xumong gets 9. The *R/Splus* function `sample` can be used to draw a random permutation. Then the first 3 numbers in the permutation correspond to group 1, the next 3 to group 2 and the final 3 to group 3. Using the output shown below, gives the following 3 groups.

group 1: Stach, Wayman, Xumong
group 2: Lawes, Carroll, Halverson
group 3: Collin, Wenslow, Crawford

```
> sample(9)
[1] 6 7 9 5 1 4 2 8 3
```

Often there is a table or computer file of units and related measurements, and it is desired to add the unit's group to the end of the table. The *mpack* function `rand` reports a random permutation and the quantity groups[i] = treatment group for the ith person on the list. Since persons 6, 7 and 9 are in group 1, groups[7] = 1. Since Carroll is person 1 and is in group 2, groups[1] = 2, et cetera.

```
> rand(9,3)
$perm
[1] 6 7 9 5 1 4 2 8 3

$groups
[1] 2 3 3 2 2 1 1 3 1
```

Definition 10.7. Replication means that for each treatment, the n_i response variables $Y_{i,1}, ..., Y_{i,n_i}$ are approximately iid random variables.

Example 10.2. a) If ten students work two types of paper mazes three times each, then there are 60 measurements that are not replicates. Each

student should work the six mazes in random order since speed increases with practice. For the ith student, let Z_{i1} be the average time to complete the three mazes of type 1, let Z_{i2} be the average time for mazes of type 2 and let $D_i = Z_{i1} - Z_{i2}$. Then $D_1, ..., D_{10}$ are replicates.

b) Cobb (1998, p. 126) states that a student wanted to know if the shapes of sponge cells depends on the color (green or white). He measured hundreds of cells from one white sponge and hundreds of cells from one green sponge. There were only two units so $n_1 = 1$ and $n_2 = 1$. The student should have used a sample of n_1 green sponges and a sample of n_2 white sponges to get more replicates.

c) Replication depends on the goals of the study. Box, Hunter and Hunter (2005, p. 215-219) describes an experiment where the investigator times how long it takes him to bike up a hill. Since the investigator is only interested in his performance, each run up a hill is a replicate (the time for the ith run is a sample from all possible runs up the hill by the investigator). If the interest had been on the effect of eight treatment levels on student bicyclists, then replication would need $n = n_1 + \cdots + n_8$ student volunteers where n_i ride their bike up the hill under the conditions of treatment i.

Definition 10.8. Let $f_Z(z)$ be the pdf of Z. Then the family of pdfs $f_Y(y) = f_Z(y - \mu)$ indexed by the *location parameter* μ, $-\infty < \mu < \infty$, is the *location family* for the random variable $Y = \mu + Z$ with *standard pdf* $f_Z(z)$.

Definition 10.9. A *one way fixed effects ANOVA model* has a single qualitative predictor variable W with p categories a_1, ..., a_p. There are p different distributions for Y, one for each category a_i. The distribution of

$$Y|(W = a_i) \sim f_Z(y - \mu_i)$$

where the location family has second moments. Hence all p distributions come from the same location family with different location parameter μ_i and the same variance σ^2.

Definition 10.10. The *one way fixed effects normal ANOVA model* is the special case where

$$Y|(W = a_i) \sim N(\mu_i, \sigma^2).$$

Example 10.3. The pooled 2 sample t–test is a special case of a one way ANOVA model with $p = 2$. For example, one population could be ACT

scores for men and the second population ACT scores for women. Then $W =$ *gender* and $Y =$ *score.*

Notation. It is convenient to relabel the response variable $Y_1, ..., Y_n$ as the vector $\boldsymbol{Y} = (Y_{11}, ..., Y_{1,n_1}, Y_{21}, ..., Y_{2,n_2}, ..., Y_{p1}, ..., Y_{p,n_p})^T$ where the Y_{ij} are independent and $Y_{i1}, ..., Y_{i,n_i}$ are iid. Here $j = 1, ..., n_i$ where n_i is the number of cases from the ith level where $i = 1, ..., p$. Thus $n_1 + \cdots + n_p = n$. Similarly use double subscripts on the errors. Then there will be many equivalent parameterizations of the one way fixed effects ANOVA model.

Definition 10.11. The *cell means model* is the parameterization of the one way fixed effects ANOVA model such that

$$Y_{ij} = \mu_i + e_{ij}$$

where Y_{ij} is the value of the response variable for the jth trial of the ith factor level. The μ_i are the unknown means and $E(Y_{ij}) = \mu_i$. The e_{ij} are iid from the location family with pdf $f_Z(z)$ and unknown variance $\sigma^2 = \text{VAR}(Y_{ij}) = \text{VAR}(e_{ij})$. For the normal cell means model, the e_{ij} are iid $N(0, \sigma^2)$ for $i = 1, ..., p$ and $j = 1, ..., n_i$.

The cell means model is a linear model (without intercept) of the form $\boldsymbol{Y} = \boldsymbol{X}_c \boldsymbol{\beta}_c + \boldsymbol{e} =$

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1,n_1} \\ e_{21} \\ \vdots \\ e_{2,n_2} \\ \vdots \\ e_{p,1} \\ \vdots \\ e_{p,n_p} \end{bmatrix}. \tag{10.4}$$

Notation. Let $Y_{i0} = \sum_{j=1}^{n_i} Y_{ij}$ and let

$$\hat{\mu}_i = \overline{Y}_{i0} = Y_{i0}/n_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}. \tag{10.5}$$

Hence the "dot notation" means sum over the subscript corresponding to the 0, eg j. Similarly, $Y_{00} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} Y_{ij}$ is the sum of all of the Y_{ij}.

Notice that the indicator variables used in the cell means model (10.4) are $v_{h,k} = 1$ if the hth case has $W = a_k$, and $v_{h_k} = 0$, otherwise, for $k = 1, ..., p$ and $h = 1, ..., n$. So Y_{ij} has $v_{h,k} = 1$ only if $i = k$ and $j = 1, ..., n_i$. Here $\boldsymbol{v}_k$ is the kth column of $\boldsymbol{X}_c$. The model can use p indicator variables for the factor instead of $p - 1$ indicator variables because the model does not contain an intercept. Also notice that

$$E(\boldsymbol{Y}) = \boldsymbol{X}_c \boldsymbol{\beta}_c = (\mu_1, ..., \mu_1, \mu_2, ..., \mu_2, ..., \mu_p, ..., \mu_p)^T,$$

$(\boldsymbol{X}_c^T \boldsymbol{X}_c) = \mathrm{diag}(\mathrm{n_1}, ..., \mathrm{n_p})$ and $\boldsymbol{X}_c^T \boldsymbol{Y} = (Y_{10}, ..., Y_{10}, Y_{20}, ..., Y_{20}, ..., Y_{p0}, ..., Y_{p0})^T$. Hence $(\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} = \mathrm{diag}(1/\mathrm{n_1}, ..., 1/\mathrm{n_p})$ and the OLS estimator

$$\hat{\boldsymbol{\beta}}_c = (\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} \boldsymbol{X}_c^T \boldsymbol{Y} = (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T = (\hat{\mu}_1, ..., \hat{\mu}_p)^T.$$

Thus $\hat{\boldsymbol{Y}} = \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_c = (\overline{Y}_{10}, ..., \overline{Y}_{10}, ..., \overline{Y}_{p0}, ..., \overline{Y}_{p0})^T$. Hence the ijth fitted value is

$$\hat{Y}_{ij} = \overline{Y}_{i0} = \hat{\mu}_i \tag{10.6}$$

and the ijth residual is

$$r_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i. \tag{10.7}$$

Since the cell means model is a linear model, there is an associated response plot and residual plot. However, many of the interpretations of the OLS quantities for ANOVA models differ from the interpretations for multiple linear regression (MLR) models. First, for MLR models, the conditional distribution $Y|\boldsymbol{x}$ makes sense even if $\boldsymbol{x}$ is not one of the observed $\boldsymbol{x}_i$ provided that $\boldsymbol{x}$ is not far from the $\boldsymbol{x}_i$. This fact makes MLR very powerful. For MLR, at least one of the variables in $\boldsymbol{x}$ is a continuous predictor. For the one way fixed effects ANOVA model, the p distributions $Y|\boldsymbol{x}_i$ make sense where $\boldsymbol{x}_i^T$ is a row of $\boldsymbol{X}_c$.

Also, the OLS MLR ANOVA F test for the cell means model tests $H_0 : \boldsymbol{\beta} = 0 \equiv H_0 : \mu_1 = \cdots = \mu_p = 0$, while the one way fixed effects ANOVA F test given after Definition 10.15 tests $H_0 : \mu_1 = \cdots = \mu_p$.

Definition 10.12. Consider the one way fixed effects ANOVA model. The *response plot* is a plot of $\hat{Y}_{ij} \equiv \hat{\mu}_i$ versus Y_{ij} and the *residual plot* is a

plot of $\hat{Y}_{ij} \equiv \hat{\mu}_i$ versus r_{ij}. 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 $Z_1, ..., Z_m$ consists of an axis and m points each corresponding to the value of Z_i. The response plot consists of p dot plots, one for each value of $\hat{\mu}_i$. The dot plot corresponding to $\hat{\mu}_i$ is the dot plot of $Y_{i1}, ..., Y_{i,n_i}$. The p dot plots should have roughly the same amount of spread, and each $\hat{\mu}_i$ corresponds to level a_i. If a new level a_f corresponding to $\boldsymbol{x}_f$ was of interest, hopefully the points in the response plot corresponding to a_f would form a dot plot at $\hat{\mu}_f$ similar in spread to the other dot plots, but it may not be possible to predict the value of $\hat{\mu}_f$. Similarly, the residual plot consists of p dot plots, and the plot corresponding to $\hat{\mu}_i$ is the dot plot of $r_{i1}, ..., r_{i,n_i}$.

Assume that each $n_i \geq 10$. Under the assumption that the Y_{ij} are from the same location scale family with different parameters μ_i, each of the p dot plots should have roughly the same shape and spread. This assumption is easier to judge with 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 is not much difference in the μ_i. If the identity line is clearly superior to any horizontal line, then at least some of the means differ.

Definition 10.13. 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 10.1. 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_j cases corresponding to level a_j that is far from the bulk of the data. This dot plot is probably not a cluster of "bad outliers" if $n_j \geq 4$ and $n \leq 50$. If $n_j = 1$, such a case may be a large outlier.

Rule of thumb 10.2. Often an outlier is very good, but more often an outlier is due to a measurement error and is very bad.

The assumption of the Y_{ij} coming from the same location scale family with different location parameters μ_i and the same constant variance σ^2 is a big assumption and often does not hold. Another way to check this assumption is to make a box plot of the Y_{ij} for each i. The box in the box plot corresponds to the lower, middle and upper quartiles of the Y_{ij}. The middle quartile is just the sample median of the data m_{ij}: at least half of the $Y_{ij} \geq m_{ij}$ and at least half of the $Y_{ij} \leq m_{ij}$. The p boxes should be roughly the same length and the median should occur in roughly the same position (eg in the center) of each box. The "whiskers" in each plot should also be roughly similar. Histograms for each of the p samples could also be made. All of the histograms should look similar in shape.

Example 10.4. Kuehl (1994, p. 128) gives data for counts of hermit crabs on 25 different transects in each of six different coastline habitats. Let Z be the count. Then the response variable $Y = \log_{10}(Z+1/6)$. Although the counts Z varied greatly, each habitat had several counts of 0 and often there were several counts of 1, 2 or 3. Hence Y is not a continuous variable. The cell means model was fit with $n_i = 25$ for $i = 1, ..., 6$. Each of the six habitats was a level. Figure 10.1a and b shows the response plot and residual plot. There are 6 dot plots in each plot. Because several of the smallest values in each plot are identical, it does not always look like the identity line is passing through the six sample means $\overline{Y}_{i0}$ for $i = 1, ..., 6$. In particular, examine the dot plot for the smallest mean (look at the 25 dots furthest to the left that fall on the vertical line FIT ≈ 0.36). Random noise (jitter) has been added to the response and residuals in Figure 10.1c and d. Now it is easier to compare the six dot plots. They seem to have roughly the same spread.

The plots contain a great deal of information. The response plot can be used to explain the model, check that the sample from each population (treatment) has roughly the same shape and spread, and to see which populations have similar means. Since the response plot closely resembles the residual plot in Figure 10.1, there may not be much difference in the six populations. Linearity seems reasonable since the samples scatter about the identity line. The residual plot makes the comparison of "similar shape" and "spread" easier.

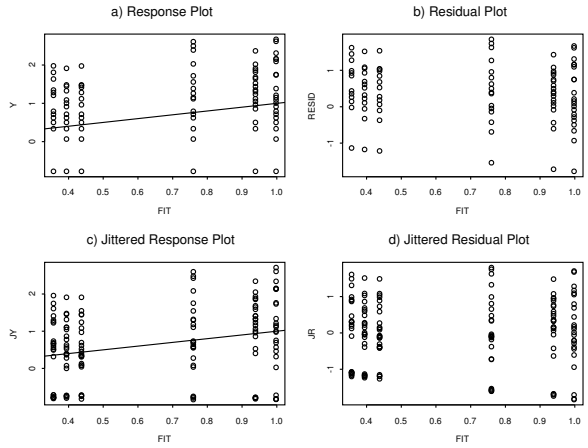

Figure 10.1: Plots for One Way ANOVA Model for Crab Data

Definition 10.14. a) The *total sum of squares*

$$SSTO = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{00})^2.$$

b) The *treatment sum of squares*

$$SSTR = \sum_{i=1}^{p} n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2.$$

c) The residual sum of squares or *error sum of squares*

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{io})^2.$$

Definition 10.15. Associated with each SS in Definition 10.14 is a degrees of freedom (df) and a mean square $= SS/df$. For SSTO, df $= n-1$ and $MSTO = SSTO/(n-1)$. For SSTR, df $= p-1$ and $MSTR = SSTR/(p-1)$. For SSE, $df = n-p$ and $MSE = SSE/(n-p)$.

Let $S_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2/(n_i - 1)$ be the sample variance of the ith group. Then the MSE is a weighted sum of the S_i^2:

$$\hat{\sigma}^2 = MSE = \frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i} r_{ij}^2 = \frac{1}{n-p}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2 =$$

$$\frac{1}{n-p}\sum_{i=1}^{p}(n_i-1)S_i^2 = S_{pool}^2$$

where S_{pool}^2 is known as the pooled variance estimator.

The ANOVA table is the same as that for MLR, except that SSTR replaces the regression sum of squares. The MSE is again an estimator of σ^2. The ANOVA F test tests whether all p means μ_i are equal. Shown below is an ANOVA table given in symbols. Sometimes "Treatment" is replaced by "Between treatments," "Between Groups," "Model," "Factor" or "Groups." Sometimes "Error" is replaced by "Residual," or "Within Groups." Sometimes "p-value" is replaced by "P", "$Pr(> F)$" or "PR > F."

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Treatment	p-1	SSTR	MSTR	Fo=MSTR/MSE	for Ho:
Error	n-p	SSE	MSE		$\mu_1 = \cdots = \mu_p$

Note that the software output uses pvalue for pval, an estimate of the pvalue.

Be able to perform the 4 step fixed effects one way ANOVA F test of hypotheses:
i) State the hypotheses Ho: $\mu_1 = \mu_2 = \cdots = \mu_p$ and Ha: not Ho.
ii) Find the test statistic $F_o = MSTR/MSE$ or obtain it from output.
iii) Find the pval from output or use the F–table: pval =

$$P(F_{p-1,n-p} > F_o).$$

iv) State whether you reject Ho or fail to reject Ho. If the pval $< \delta$, reject Ho and conclude that the mean response depends on the level of the factor. Otherwise fail to reject Ho and conclude that the mean response does not depend on the level of the factor. Give a nontechnical sentence.

Rule of thumb 10.3. If

$$\max(S_1, ..., S_p) \leq 2\min(S_1, ..., S_p),$$

then the one way ANOVA F test results will be approximately correct if the response and residual plots suggest that the remaining one way ANOVA model assumptions are reasonable. See Moore (2000, p. 512). If all of the

$n_i \geq 5$, replace the standard deviations by the ranges of the dot plots when examining the response and residual plots.

Remark 10.1. If the units are a representative sample of some population of interest, then randomization of units into groups makes the assumption that $Y_{i1}, ..., Y_{i,n_i}$ are iid hold to a useful approximation for large sample theory. Random sampling from populations also induces the iid assumption. Linearity can be checked with the response plot, and similar shape and spread of the location families can be checked with both the response and residual plots. Also check that outliers are not present. If the p dot plots in the response plot are approximately symmetric, then the sample sizes n_i can be smaller than if the dot plots are skewed.

Remark 10.2. When the assumption that the p groups come from the same location family with finite variance σ^2 is violated, the one way ANOVA F test may not make much sense because unequal means may not imply the superiority of one category over another. Suppose Y is the time in minutes until relief from a headache and that $Y_{1j} \sim N(60, 1)$ while $Y_{2j} \sim N(65, \sigma^2)$. If $\sigma^2 = 1$, then the type 1 medicine gives headache relief 5 minutes faster, on average, and is superior, all other things being equal. But if $\sigma^2 = 100$, then many patients taking medicine 2 experience much faster pain relief than those taking medicine 1, and many experience much longer time until pain relief. In this situation, predictor variables that would identify which medicine is faster for a given patient would be very useful.

```
fat1  fat2  fat3  fat4 One way Anova for Fat1 Fat2 Fat3 Fat4
64    78    75    55   Source     DF SS     MS     F     P
72    91    93    66   treatment  3  1636.5 545.5  5.41  0.0069
68    97    78    49   error      20 2018.0 100.9
77    82    71    64
56    85    63    70
95    77    76    68
```

Example 10.5. The output above represents grams of fat (minus 100 grams) absorbed by doughnuts using 4 types of fat. See Snedecor and Cochran (1967, p. 259). Let μ_i denote the mean amount of fat$i$ absorbed by doughnuts, $i = 1, 2, 3$ and 4. a) Find $\hat{\mu}_1$. b) Perform a 4 step ANOVA F test.

Solution: a) $\hat{\boldsymbol{\beta}}_{1c} = \hat{\mu}_1 = \overline{Y}_{10} = Y_{10}/n_1 = \sum_{j=1}^{n_1} Y_{1j}/n_1 =$
$(64 + 72 + 68 + 77 + 56 + 95)/6 = 432/6 = 72.$

b) i) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ H_a: not H_0

ii) F = 5.41

iii) pval = 0.0069

iv) Reject H_0, the mean amount of fat absorbed by doughnuts depends on the type of fat.

Definition 10.16. A **contrast** $C = \sum_{i=1}^{p} k_i \mu_i$ where $\sum_{i=1}^{p} k_i = 0$. The estimated contrast is $\hat{C} = \sum_{i=1}^{p} k_i \overline{Y}_{i0}$.

If the null hypothesis of the fixed effects one way ANOVA test is not true, then not all of the means μ_i are equal. Researchers will often have hypotheses, before examining the data, that they desire to test. Often such a hypothesis can be put in the form of a contrast. For example, the contrast $C = \mu_i - \mu_j$ is used to compare the means of the ith and jth groups while the contrast $\mu_1 - (\mu_2 + \cdots + \mu_p)/(p-1)$ is used to compare the last $p-1$ groups with the 1st group. This contrast is useful when the 1st group corresponds to a standard or control treatment while the remaining groups correspond to new treatments.

Assume that the normal cell means model is a useful approximation to the data. Then the $\overline{Y}_{i0} \sim N(\mu_i, \sigma^2/n_i)$ are independent, and

$$\hat{C} = \sum_{i=1}^{p} k_i \overline{Y}_{i0} \sim N\left(C, \sigma^2 \sum_{i=1}^{p} \frac{k_i^2}{n_i}\right).$$

Hence the standard error

$$SE(\hat{C}) = \sqrt{MSE \sum_{i=1}^{p} \frac{k_i^2}{n_i}}.$$

The degrees of freedom is equal to the MSE degrees of freedom $= n - p$.

Consider a family of null hypotheses for contrasts $\{Ho : \sum_{i=1}^{p} k_i \mu_i = 0$ where $\sum_{i=1}^{p} k_i = 0$ and the k_i may satisfy other constraints$\}$. Let δ_S denote the probability of a type I error for a single test from the family where a type I error is a false rejection. The **family level** δ_F is an upper bound on the (usually unknown) size δ_T. Know how to interpret $\delta_F \approx \delta_T =$
P(of making at least one type I error among the family of contrasts).

Two important families of contrasts are the family of all possible contrasts and the family of pairwise differences $C_{ij} = \mu_i - \mu_j$ where $i \neq j$. The

Scheffé multiple comparisons procedure has a δ_F for the family of all possible contrasts while the Tukey multiple comparisons procedure has a δ_F for the family of all $\binom{p}{2}$ pairwise contrasts.

To interpret output for multiple comparisons procedures, the underlined means or blocks of letters besides groups of means indicate that the group of means are not significantly different.

Example 10.6. The output below uses data from SAS Institute (1985, p. 126-129). The mean nitrogen content of clover depends on the strain of clover (3dok1, 3dok5, 3dok7, compos, 3dok4, 3dok13). Recall that means μ_1 and μ_2 are significantly different if you can conclude that $\mu_1 \neq \mu_2$ while μ_1 and μ_2 are not significantly different if there is not enough evidence to conclude that $\mu_1 \neq \mu_2$ (perhaps because the means are approximately equal or perhaps because the sample sizes are not large enough).

Notice that the strain of clover 3dok1 appears to have the highest mean nitrogen content. There are 4 pairs of means that are not significantly different. The letter B suggests 3dok5 and 3dok7, the letter C suggests 3dok7 and compos, the letter D suggests compos and 3dok4, while the letter E suggests 3dok4 and 3dok13 are not significantly different.

```
Means with the same letter are not significantly different.
     Waller Grouping          Mean      N    strain
                   A        28.820      5    3dok1

                   B        23.980      5    3dok5
                   B
              C    B        19.920      5    3dok7
              C
              C    D        18.700      5    compos
                   D
              E    D        14.640      5    3dok4
              E
              E             13.260      5    3dok13
```

Definition 10.17. Graphical Anova for the one way model uses the residuals as a reference set instead of a t, F or normal distribution. The scaled treatment deviations or scaled effect $c(\overline{Y}_{i0} - \overline{Y}_{00}) = c(\hat{\mu}_i - \overline{Y}_{00})$ are scaled to have the same variability as the residuals. A dot plot of the scaled deviations is placed above the dot plot of the residuals. Assume that

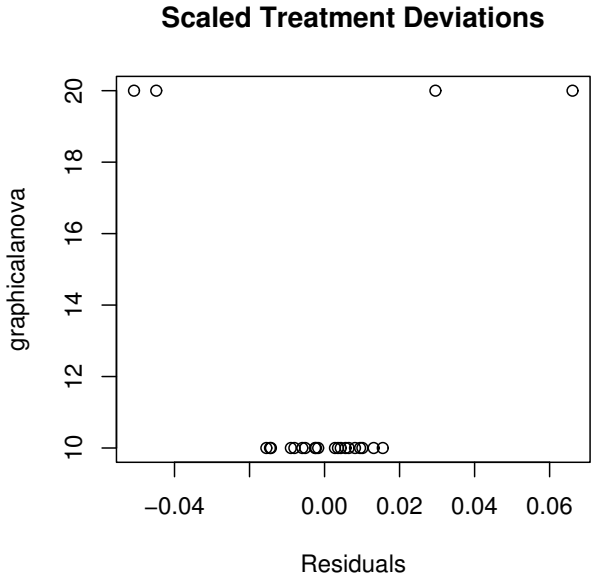

Figure 10.2: Graphical Anova

$n_i \equiv h = n/p$ for $i = 1, ..., p$. For small $n \leq 40$, suppose the distance between two scaled deviations (A and B, say) is greater than the range of the residuals $= \max(r_{ij}) - \min(r_{ij})$. Then declare μ_A and μ_B to be significantly different. If the distance is less than the range, do not declare μ_A and μ_B to be significantly different. Scaled deviations that lie outside the range of the residuals are significant (so significantly different from the overall mean).

For $n \geq 100$, let $r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n)}$ be the order statistics of the residuals. Then instead of the range, use $r_{(\lceil 0.975n \rceil)} - r_{(\lceil 0.025n \rceil)}$ as the distance where $\lceil x \rceil$ is the smallest integer $\geq x$, eg $\lceil 7.7 \rceil = 8$. So effects outside of the interval $(r_{(\lceil 0.025n \rceil)}, r_{(\lceil 0.975n \rceil)})$ are significant. See Box, Hunter and Hunter (2005, p. 136, 166). A derivation of the scaling constant $c = \sqrt{(n-p)/(p-1)}$ is given in Section 10.5.

```
ganova(x,y)
sdev    0.02955502  0.06611268 -0.05080048 -0.04486722
Treatments "A"      "B"        "C"          "D"
```

Example 10.7. Cobb (1998, p. 160) describes a one way ANOVA design used to study the amount of calcium in the blood. For many animals, the

body's ability to use calcium depends on the level of certain hormones in the blood. The response was 1/(level of plasma calcium). The four groups were A: Female controls, B: Male controls, C: Females given hormone and D: Males given hormone. There were 10 birds of each gender, and five from each gender were given the hormone. The output above uses the `mpack` function *ganova* to produce Figure 10.2.

In Figure 10.2, the top dot plot has the scaled treatment deviations. From left to right, these correspond to C, D, A and B since the output shows that the deviation corresponding to C is the smallest with value -0.050. Since the deviations corresponding to C and D are much closer than the range of the residuals, the C and D effects yielded similar mean response values. A and B appear to be significantly different from C and D. The distance between the scaled A and B treatment deviations is about the same as the distance between the smallest and largest residuals, so there is only marginal evidence that the A and B effects are significantly different.

Since all 4 scaled deviations lie outside of the range of the residuals, all effects A, B, C and D appear to be significant.

10.2.1 Response Transformations for ANOVA Models

A model for an experimental design is $Y_i = E(Y_i) + e_i$ for $i = 1, ..., n$ where the error $e_i = Y_i - E(Y_i)$ and $E(Y_i) \equiv E(Y_i|\boldsymbol{x}_i)$ is the expected value of the response Y_i for a given vector of predictors $\boldsymbol{x}_i$. Many models can be fit with least squares (OLS or LS) and are linear models of the form

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$$

for $i = 1, \ldots, n$. Often $x_{i,1} \equiv 1$ for all i. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ design matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. If the fitted values are $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$, then $Y_i = \hat{Y}_i + r_i$ where the residuals $r_i = Y_i - \hat{Y}_i$.

The applicability of an experimental design model can be expanded by allowing response transformations. An important class of *response transformation models* adds an additional unknown transformation parameter λ_o,

such that

$$Y_i = t_{\lambda_o}(Z_i) \equiv Z_i^{(\lambda_o)} = E(Y_i) + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i.$$

If λ_o was known, then $Y_i = t_{\lambda_o}(Z_i)$ would follow the linear model for the experimental design.

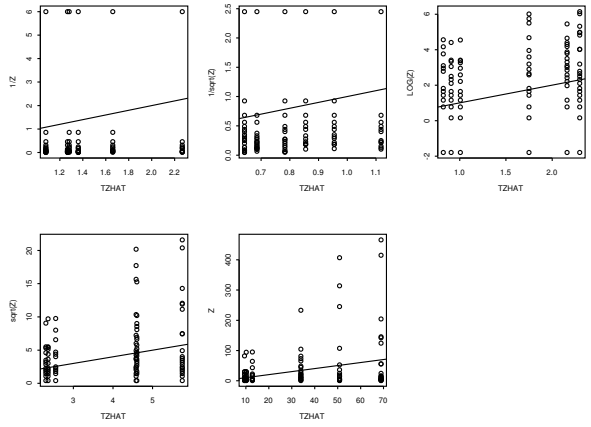

Figure 10.3: Transformation Plots for Crab Data

Definition 10.20. Assume that **all** of the values of the "response" Z_i are **positive**. A *power transformation* has the form $Y = t_\lambda(Z) = Z^\lambda$ for $\lambda \neq 0$ and $Y = t_0(Z) = \log(Z)$ for $\lambda = 0$ where $\lambda \in \Lambda_L = \{-1, -1/2, 0, 1/2, 1\}$.

A graphical method for response transformations computes the fitted values $\hat{W}_i$ from the experimental design model using $W_i = t_\lambda(Z_i)$ as the "response." Then a plot of the $\hat{W}$ versus W is made for each of the five values of $\lambda \in \Lambda_L$. For many experimental design models, the plotted points follow the identity line in a (roughly) evenly populated band if the experimental design model is reasonable for $(\hat{W}, W)$. An exception is the one way ANOVA model where there will be p dot plots of roughly the same shape and spread that scatter about the identity line. If more than one value of $\lambda \in \Lambda_L$ gives a linear plot, consult subject matter experts and use the simplest or most reasonable transformation. Note that Λ_L has 5 models, and the graphical method selects the model with the best response plot. After selecting the transformation, the usual checks should be made. In particular, the transformation plot is also the response plot, and a residual plot should be made.

Definition 10.21. A *transformation plot* is a plot of $(\hat{W}, W)$ with the identity line added as a visual aid.

In the following example, the plots show $t_\lambda(Z)$ on the vertical axis. The label "TZHAT" of the horizontal axis are the fitted values that result from using $t_\lambda(Z)$ as the "response" in the software.

For one way ANOVA models with $n_i \equiv m \geq 5$, look for a transformation plot that satisfies the following conditions. i) The p dot plots scatter about the identity line with similar shape and spread. ii) Dot plots with more skew are worse than dot plots with less skew or dot plots that are approximately symmetric. iii) Spread that increases or decreases with TZHAT is bad.

Example 10.4, continued. Following Kuehl (1994, p. 128), let C be the count of crabs and let the "response" $Z = C + 1/6$. Figure 10.3 shows the five *transformation plots*. The transformation $\log(Z)$ results in dot plots that have roughly the same shape and spread. The transformations $1/Z$ and $1/\sqrt{Z}$ do not handle the 0 counts well, and the dot plots fail to cover the identity line. The transformations $\sqrt{Z}$ and Z have variance that increases with the mean.

Remark 10.4. The graphical method for response transformations can be used for design models that are linear models, not just one way ANOVA models. The method is nearly identical to that of Chapter 12, but Λ_L only has 5 values. The **log rule** states that if all of the $Z_i > 0$ and if $\frac{\max(Z_i)}{\min(Z_i)} \geq 10$, then the response transformation $Y = \log(Z)$ will often work.

10.3 One Way MANOVA

Using double subscripts will be useful for describing the one way MANOVA model. Suppose there independent random samples from p different populations (treatments), or $n = \sum_{i=1}^{p} n_i$ and n_i cases are randomly assigned to p treatment groups. Then the group sample sizes are n_i for $i = 1, ..., p$. Assume that m response variables $\boldsymbol{y}_{ij} = (Y_{ij1}, ..., Y_{ijm})^T$ are measured for the ith treatment. Hence $i = 1, ..., p$ and $j = 1, ..., n_i$. The Y_{ijk} follow different one way ANOVA models for $k = 1, ..., m$. Assume $E(\boldsymbol{y}_{ij}) = \boldsymbol{\mu}_i$ and $\text{Cov}(\boldsymbol{y}_{ij}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Hence the p treatments have different mean vectors $\boldsymbol{\mu}_i$, but common covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. (This 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 $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$. Often $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$, so H_0 becomes $H_0 : \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_p$. 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 $m = 2$.

Let $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\tau}_i$ where $\sum_{i=1}^{p} n_i \boldsymbol{\tau}_i = 0$. The jth case from the ith population or treatment group is $\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_j + \boldsymbol{e}_{ij}$ where $\boldsymbol{e}_{ij}$ is an error vector, $i = 1, ..., p$ and $j = 1, ..., n_i$. Let $\overline{\boldsymbol{y}} = \hat{\boldsymbol{\mu}} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}/n$ be the overall mean. Let $\overline{\boldsymbol{y}}_i = \sum_{j=1}^{n_i} \boldsymbol{y}_{ij}/n_i$ so $\hat{\boldsymbol{\tau}}_i = \overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}}$. Let the residual $\hat{\boldsymbol{\epsilon}}_{ij} = \boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i = \boldsymbol{y}_{ij} - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\tau}}_i$. Then $\boldsymbol{y}_{ij} = \overline{\boldsymbol{y}} + (\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}}) + (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i) = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\tau}}_i + \hat{\boldsymbol{\epsilon}}_{ij}$.

Let $\boldsymbol{S}_i$ be the sample covariance matrix corresponding to the ith treatment group. Then the within sum of squares and cross products matrix is $\boldsymbol{W} = (n_1 - 1)\boldsymbol{S}_1 + \cdots + (n_p - 1)\boldsymbol{S}_p = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)^T$. Then $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \boldsymbol{W}/(n - p)$. The treatment or between sum of squares and cross products matrix is $\boldsymbol{B} = \sum_{i=1}^{p} n_i (\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}})(\overline{\boldsymbol{y}}_i - \overline{\boldsymbol{y}})^T$. The total corrected (for the mean) sum of squares and cross products matrix is $\boldsymbol{T} = \boldsymbol{B} + \boldsymbol{W} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}})(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}})^T$. Note that $\boldsymbol{T}/(n - 1)$ is the usual sample covariance matrix if it is assumed that all n of the $\boldsymbol{y}_{ij}$ are iid so that the $\boldsymbol{\mu}_i \equiv \boldsymbol{\mu}$ for $i = 1, ..., p$.

The one way MANOVA model is $\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\epsilon}_{ij}$ where the $\boldsymbol{\epsilon}_{ij}$ are iid with $E(\boldsymbol{\epsilon}_{ij}) = \boldsymbol{0}$ 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	$\boldsymbol{B}$	$p - 1$
Residual or Error or Within	$\boldsymbol{W}$	$n - p$
Total (corrected)	$\boldsymbol{T}$	$n - 1$

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

$$\Lambda = \frac{|\boldsymbol{W}|}{|\boldsymbol{B} + \boldsymbol{W}|} = \frac{|\boldsymbol{W}|}{|\boldsymbol{T}|}$$

is good if the iid $\boldsymbol{\epsilon}_{ij} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$. Then $t_o = -[n-1-(m+p)/2]\log(\Lambda)$ and pval $= P(\chi^2_{m(p-1)} > t_0)$. Hence reject H_0 if $t_0 > \chi^2_{m(p-1)}(1-\alpha)$. See Johnson and Wichern (1988, p. 238).

The four steps of the one way MANOVA test follow.
i) State the hypotheses $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ and H_1 : 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 α is not given.

Rule of thumb 10.4. In the one way MANOVA model, $\boldsymbol{Y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{e}_j$ is a one way ANOVA model for $j = 1, ..., m$. To check the one way MANOVA model, make the m response and residual plots corresponding to the m one way ANOVA models. Make a DD plot of the n residual vectors. Response transformations can be done as in Section 10.2.1. If the n_i are large, make p DD plots of the $\boldsymbol{y}_{ij}$ for $i = 1, ..., p$. Also if the n_i are large, make p plots of $D_{ij}(\overline{\boldsymbol{y}}_i, \boldsymbol{S}_i)$ versus $D_{ij}(\overline{\boldsymbol{y}}_i, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}})$ to check that the common covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is an adequate assumption. The plotted points in these p plots should cluster tightly about the identity line if n_i is large and the covariance matrix of the ith treatment group is approximately $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$.

10.4 Summary

1) The **fixed effects one way ANOVA** model has one qualitative explanatory variable called a **factor** and a quantitative response variable Y_{ij}. The factor variable has p levels, $E(Y_{ij}) = \mu_i$ and $V(Y_{ij}) = \sigma^2$ for $i = 1, ..., p$ and $j = 1, ..., n_i$. **Experimental units** are randomly assigned to the treatment levels.

2) Let $n = n_1 + \cdots + n_p$. In an **experiment**, the investigators use randomization to randomly assign n units to treatments. Draw a random permutation of $\{1, ..., n\}$. Assign the first n_1 units to treatment 1, the next n_2 units to treatment 2, ..., and the final n_p units to treatment p. Use $n_i \equiv h = n/p$ if possible. Randomization washes out the effect of lurking variables.

3) The 4 step fixed effects one way ANOVA F test has steps
i) Ho: $\mu_1 = \mu_2 = \cdots = \mu_p$ and Ha: not Ho.
ii) Fo = MSTR/MSE is usually given by output.
iii) The pval = P($F_{p-1,n-p} > Fo$) is usually given by output.
iv) If the pval $< \delta$, reject Ho and conclude that the mean response depends on the level of the factor. Otherwise fail to reject Ho and conclude that the mean response does not depend on the level of the factor. Give a nontechnical sentence.

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Treatment	p-1	SSTR	MSTR	Fo=MSTR/MSE	for Ho:
Error	n-p	SSE	MSE		$\mu_1 = \cdots = \mu_p$

4) Shown is an ANOVA table given in symbols. Sometimes "Treatment" is replaced by "Between treatments," "Between Groups," "Model," "Factor" or "Groups." Sometimes "Error" is replaced by "Residual," or "Within Groups." Sometimes "p-value" is replaced by "P", "$Pr(> F)$" or "PR > F."

5) A *dot plot* of $Z_1, ..., Z_h$ consists of an axis and h points each corresponding to the value of Z_i. The *response plot* is a plot of $\hat{Y}$ versus Y. For the one way ANOVA model, the response plot is a plot of $\hat{Y}_{ij} = \hat{\mu}_i$ versus Y_{ij}. Often the identity line with unit slope and zero intercept is added as a visual aid. Vertical deviations from the identity line are the residuals $r_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i$. The plot will consist of p dot plots that scatter about the identity line with similar shape and spread if the fixed effects one way ANOVA model is appropriate. The ith dot plot is a dot plot of $Y_{i,1}, ..., Y_{i,n_i}$. Assume that each $n_i \geq 10$. 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 is not much difference in the μ_i. If the identity line is clearly superior to any horizontal line, then at least some of the means differ.

6) The *residual plot* is a plot of $\hat{Y}$ versus residual $r = Y - \hat{Y}$. The plot will consist of p dot plots that scatter about the $r = 0$ line with similar shape and spread if the fixed effects one way ANOVA model is appropriate. The ith dot plot is a dot plot of $r_{i,1}, ..., r_{i,n_i}$. Assume that each $n_i \geq 10$. Under the assumption that the Y_{ij} are from the same location scale family with

different parameters μ_i, each of the p dot plots should have roughly the same shape and spread. This assumption is easier to judge with the residual plot than with the response plot.

7) Rule of thumb: If $\max(S_1, ..., S_p) \leq 2\min(S_1, ..., S_p)$, then the one way ANOVA F test results will be approximately correct if the response and residual plots suggest that the remaining one way ANOVA model assumptions are reasonable.

8) The **cell means model** for the fixed effects one way ANOVA is $Y_{ij} = \mu_i + e_{ij}$ where Y_{ij} is the value of the response variable for the jth trial of the ith factor level for $i = 1, ..., p$ and $j = 1, ..., n_i$. The μ_i are the unknown means and $E(Y_{ij}) = \mu_i$. The e_{ij} are iid from the location family with pdf $f_Z(z)$, zero mean and unknown variance $\sigma^2 = V(Y_{ij}) = V(e_{ij})$. For the normal cell means model, the e_{ij} are iid $N(0, \sigma^2)$. The estimator $\hat{\mu}_i = \overline{Y}_{i0} = \sum_{j=1}^{n_i} Y_{ij}/n_i = \hat{Y}_{ij}$. The ith residual is $r_{ij} = Y_{ij} - \overline{Y}_{i0}$, and $\overline{Y}_{00}$ is the sample mean of all of the Y_{ij} and $n = \sum_{i=1}^{p} n_i$. The total sum of squares SSTO $= \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{00})^2$, the treatment sum of squares SSTR $= \sum_{i=1}^{p} n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2$, and the error sum of squares SSE $= \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i0})^2$. The MSE is an estimator of σ^2. In the ANOVA table, SSTO, SSTR and SSE have $n-1$, $p-1$ and $n-p$ degrees of freedom.

9) Let $Y_{i0} = \sum_{j=1}^{n_i} Y_{ij}$ and let

$$\hat{\mu}_i = \overline{Y}_{i0} = Y_{i0}/n_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}.$$

Hence the "dot notation" means sum over the subscript corresponding to the 0, eg j. Similarly, $Y_{00} = \sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}$ is the sum of all of the Y_{ij}. Be able to find $\hat{\mu}_i$ from data.

10) The applicability of a DOE (design of experiments) model can be expanded by allowing response transformations. An important class of *response transformation models* is

$$Y = t_{\lambda_o}(Z) = E(Y) + e = \boldsymbol{x}^T\boldsymbol{\beta} + e$$

where the subscripts (eg Y_{ij}) have been suppressed. If λ_o was known, then $Y = t_{\lambda_o}(Z)$ would follow the DOE model. Assume that **all** of the values of the "response" Z are **positive**. A **power transformation** has the form $Y = t_\lambda(Z) = Z^\lambda$ for $\lambda \neq 0$ and $Y = t_0(Z) = \log(Z)$ for $\lambda = 0$ where $\lambda \in \Lambda_L = \{-1, -1/2, 0, 1/2, 1\}$.

11) A graphical method for response transformations computes the fitted values $\hat{W}$ from the DOE model using $W = t_\lambda(Z)$ as the "response" for each of the five values of $\lambda \in \Lambda_L$. Let $\hat{T} = \hat{W}$ = TZHAT and plot TZHAT vs $t_\lambda(Z)$ for $\lambda \in \{-1, -1/2, 0, 1/2, 1\}$. These plots are called **transformation plots**. The residual or error degrees of freedom used to compute the MSE should not be too small. Choose the transformation $Y = t_{\lambda^*}(Z)$ that has the best plot. Consider the one way ANOVA model with $n_i > 4$ for $i = 1, ..., p$. i) The dot plots should spread about the identity line with similar shape and spread. ii) Dot plots that are approximately symmetric are better than skewed dot plots. iii) Spread that increases or decreases with TZHAT (the shape of the plotted points is similar to a right or left opening megaphone) is bad.

12) The transformation plot for the selected transformation is also the response plot for that model (eg for the model that uses $Y = \log(Z)$ as the response). Make all of the usual checks on the DOE model (residual and response plots) after selecting the response transformation.

13) The **log rule** says try $Y = \log(Z)$ if $\max(Z)/\min(Z) > 10$ where $Z > 0$ and the subscripts have been suppressed (so $Z \equiv Z_{ij}$ for the one way ANOVA model).

14) **Graphical Anova** for the **one way ANOVA** model makes a dot plot of scaled treatment deviations (effects) above a dot plot of the residuals. For small $n \leq 40$, suppose the distance between two scaled deviations (A and B, say) is greater than the range of the residuals $= \max(r_{ij}) - \min(r_{ij})$. Then declare μ_A and μ_B to be significantly different. If the distance is less than the range, do not declare μ_A and μ_B to be significantly different. Assume the $n_i \equiv m$ for $i = 1, ..., p$. Then the ith scaled deviation is $c(\overline{Y}_{i0} - \overline{Y}_{00}) = c\hat{\alpha}_i = \tilde{\alpha}_i$ where $c = \sqrt{df_e/df_{treat}} = \sqrt{\frac{n-p}{p-1}}$.

15) Assume that the residual degrees of freedom are large enough for testing. Then the response and residual plots contain much information. Linearity and constant variance may be reasonable if the p dot plots have roughly the same shape and spread, and the dot plots scatter about the identity line. The p dot plots of the residuals should have similar shape and spread, and the dot plots scatter about the $r = 0$ line. It is easier to check linearity with the response plot and constant variance with the residual plot. Curvature is often easier to see in a residual plot, but the response plot can be used to check whether the curvature is monotone or not. The response

plot is more effective for determining whether the signal to noise ratio is strong or weak, and for detecting outliers or influential cases.

16) In a MANOVA model, $\boldsymbol{y}_k = \boldsymbol{B}^T\boldsymbol{x}_k + \boldsymbol{\epsilon}_k$ for $k = 1, ..., n$ is written in matrix form as $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$. The model has $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = ((\sigma_{ij}))$ for $k = 1, ..., n$. Each response variable in a MANOVA model follows an ANOVA model $\boldsymbol{Y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{e}_j$ for $j = 1, ..., m$ where it is assumed that $E(\boldsymbol{e}_j) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}_j) = \sigma_{jj}\boldsymbol{I}_n$.

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

18) The four steps of the one way MANOVA test follow.
i) State the hypotheses $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ and H_1 : 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 means are equal. If pval $> \alpha$, fail to reject H_0 and conclude that all p means are equal or that there is not enough evidence to conclude that not all of the p means are equal. As a textbook convention, use $\alpha = 0.05$ if α is not given.

10.5 Summary

1) The **multivariate linear model** $\boldsymbol{y}_i = \boldsymbol{B}^T\boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ for $i = 1, ..., n$ has $m \geq 2$ response variables $Y_1, ..., Y_m$ and p predictor variables $X_1, X_2, ..., X_p$. The ith case is $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T) = (x_{i1}, x_{i2}, ..., x_{ip}, Y_{i1}, ..., Y_{im})$. 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 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$. The model has $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = ((\sigma_{ij}))$ for $k = 1, ..., n$. Also $E(\boldsymbol{e}_i) = \boldsymbol{0}$ while $\text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij}\boldsymbol{I}_n$ for $i, j = 1, ..., m$. Then $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ are unknown matrices of parameters to be estimated, and $E(\boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{B}$ while $E(Y_{ij}) = \boldsymbol{x}_i^T\boldsymbol{\beta}_j$.

The data matrix $\boldsymbol{W} = [\boldsymbol{X} \;\; \boldsymbol{Y}]$ except usually the first column $\boldsymbol{1}$ of $\boldsymbol{X}$ is omitted if $X_1 = 1$. The $n \times m$ matrix

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

The $n \times p$ matrix

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

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

The $p \times m$ matrix

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

The $n \times m$ matrix

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

Warning: The $\boldsymbol{e}_i$ are error vectors, not orthonormal eigenvectors.

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

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

4) 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 $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$. This model is a multivariate linear model so there are m response variables $Y_1, ..., Y_m$ measured for each group. Each Y_i follows a one way ANOVA model for $i = 1, ..., m$.

5) For the one way MANOVA model, make a DD plot of the residuals $\hat{\boldsymbol{\epsilon}}_i$ where $i = 1, ..., n$. Use the plot to check whether the $\boldsymbol{\epsilon}_i$ follow a multivariate

normal distribution or some other elliptically contoured distribution. Want $n > 10p$.

6) For the one way MANOVA model, write the data as Y_{ijk} where $i = 1, ..., p$ and $j = 1, ..., n_i$. So k corresponds to the kth variable Y_k for $k = 1, ..., m$. Then $\hat{Y}_{ijk} = \hat{\mu}_{ik} = \overline{Y}_{i0k}$ for $i = 1, ..., p$. So for the kth variable, mean $\mu_{1k}, ..., \mu_{pk}$ are of interest. The residuals are $r_{ijk} = Y_{ijk} - \hat{Y}_{ijk}$. For each variable Y_k make a response plot of $\overline{Y}_{i0k}$ versus Y_{ijk} and a residual plot of $\overline{Y}_{i0k}$ versus r_{ijk}. Both plots will consist of p dot plots of n_k cases located at the $\overline{Y}_{i0k}$. 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 $Y_1, ..., Y_m$. For each variable Y_k, let R_{ik} be the range of the ith dot plot. If each $n_i \geq 5$, want $\max(R_{1k}, ..., R_{pk}) \leq 2 \min(R_{1k}, ..., R_{pk})$. The one way MANOVA model may be reasonable if the m response and residual plots satisfy the above graphical checks.

7) The four steps of the one way MANOVA test follow.
i) State the hypotheses $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ and H_1 : 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. Give a nontechnical sentence as the conclusion, if possible.

8) The one way MANOVA test assumes that $\boldsymbol{\Sigma}_{\boldsymbol{x}_1} = \cdots = \boldsymbol{\Sigma}_{\boldsymbol{x}_p}$, but has some resistance to this assumption. See point 6).

9) Know how to use randomization to assign units to treatment groups with the *R/Splus* function `sample` that is used to draw a random permutation of $\{1, 2, ..., n\}$. If the units are $a_1, ..., a_9$ and the `sample(9)` command gives 6 7 9 5 1 4 2 8 3, then a_6, a_7 and a_9 are assigned treatment 1, a_5, a_1 and a_4 are assigned treatment 2, and a_2, a_8 and a_3 are assigned treatment 3.

10.6 Complements

Four good tests on the design and analysis of experiments (ANOVA) are Box, Hunter and Hunter (2005), Cobb (1998), Kuehl (1994) and Ledolter and Swersey (2007). Also see Olive (2010, ch. 5-9). Section 10.2 followed Olive (2010, ch. 5) closely.

All of the parameterizations of the one way fixed effects ANOVA model yield the same predicted values, residuals and ANOVA F test, but the interpretations of the parameters differ. The cell means model is a linear model (without intercept) of the form $\boldsymbol{Y} = \boldsymbol{X}_c\boldsymbol{\beta}_c + \boldsymbol{e} =$ that can be fit using OLS. The OLS MLR output gives the correct fitted values and residuals but an incorrect ANOVA table. An equivalent linear model (with intercept) with correct OLS MLR ANOVA table as well as residuals and fitted values can be formed by replacing any column of the cell means model by a column of ones $\mathbf{1}$. Removing the last column of the cell means model and making the first column $\mathbf{1}$ gives the model $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1}x_{p-1} + e$ given in matrix form by (10.8).

It can be shown that the OLS estimators corresponding to (10.8) are $\hat{\beta}_0 = \overline{Y}_{p0} = \hat{\mu}_p$, and $\hat{\beta}_i = \overline{Y}_{i0} - \overline{Y}_{p0} = \hat{\mu}_i - \hat{\mu}_p$ for $i = 1, ..., p-1$. The cell means model has $\hat{\beta}_i = \hat{\mu}_i = \overline{Y}_{i0}$.

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1,n_1} \\ e_{21} \\ \vdots \\ e_{2,n_2} \\ \vdots \\ e_{p,1} \\ \vdots \\ e_{p,n_p} \end{bmatrix}. \quad (10.8)$$

Graphical Anova uses scaled treatment effects = scaled treatment deviations $\tilde{d}_i = cd_i = c(\overline{Y}_{i0} - \overline{Y}_{00})$ for $i = 1, ..., p$. Following Box, Hunter and Hunter (2005, p. 166), suppose $n_i \equiv m = n/p$ for $i = 1, ..., n$. If Ho $\mu_1 = \cdots = \mu_p$ is true, want the sample variance of the scaled deviations to be approximately equal to the sample variance of the residuals. So want $1 \approx \frac{\frac{1}{p}\sum_{i=1}^{p} c^2 d_i^2}{\frac{1}{n}\sum_{i=1}^{n} r_i^2} = F_0 = \frac{MSTR}{MSE} = \frac{SSTR/(p-1)}{SSE/(n-p)} = \frac{\sum_{i=1}^{p} m d_i^2/(p-1)}{\sum_{i=1}^{n} r_i^2/(n-p)}$

since $SSTR = \sum_{i=1}^{p} m(\overline{Y}_{i0} - \overline{Y}_{00})^2 = \sum_{i=1}^{p} md_i^2$. So

$$F_0 = \frac{\sum_{i=1}^{p} c^2 \frac{n}{p} d_i^2}{\sum_{i=1}^{n} r_i^2} = \frac{\sum_{i=1}^{p} \frac{m(n-p)}{p-1} d_i^2}{\sum_{i=1}^{n} r_i^2}.$$

Equating numerators gives

$$c^2 = \frac{mp}{n} \quad \frac{(n-p)}{(p-1)} = \frac{(n-p)}{(p-1)}$$

since $mp/n = 1$. Thus $c = \sqrt{(n-p)/(p-1)}$.

For Graphical Anova, see Box, Hunter and Hunter (2005, p. 136, 150, 164, 166) and Hoaglin, Mosteller, and Tukey (1991). The R package `granova`, available from (http://streaming.stat.iastate.edu/CRAN/) and authored by R.M. Pruzek and J.E. Helmreich, may be useful.

The *modified power transformation family*

$$Y_i = t_\lambda(Z_i) \equiv Z_i^{(\lambda)} = \frac{Z_i^\lambda - 1}{\lambda}$$

for $\lambda \neq 0$ and $t_0(Z_i) = \log(Z_i)$ for $\lambda = 0$ where $\lambda \in \Lambda_L$.

Box and Cox (1964) give a numerical method for selecting the response transformation for the modified power transformations. Although the method gives a point estimator $\hat{\lambda}_o$, often an interval of "reasonable values" is generated (either graphically or using a profile likelihood to make a confidence interval), and $\hat{\lambda} \in \Lambda_L$ is used if it is also in the interval.

There are several reasons to use a coarse grid Λ_L of powers. First, several of the powers correspond to simple transformations such as the log, square root, and reciprocal. These powers are easier to interpret than $\lambda = .28$, for example. Secondly, if the estimator $\hat{\lambda}_n$ can only take values in Λ_L, then sometimes $\hat{\lambda}_n$ will converge in probability to $\lambda^* \in \Lambda_L$. Thirdly, Tukey (1957) showed that neighboring modified power transformations are often very similar, so restricting the possible powers to a coarse grid is reasonable.

The graphical method for response transformations is due to Olive (2004, 2010: ch. 5). A variant of the method would plot the residual plot or both the response and the residual plot for each of the five values of λ. Residual plots are also useful, but they do not distinguish between nonlinear monotone relationships and nonmonotone relationships. See Fox (1991, p. 55). Alternative methods are given by Cook and Olive (2001) and Box, Hunter and Hunter (2005, p. 321).

A **randomization test** for the one way ANOVA model has H_0: *the different treatments have no effect*. This null hypothesis is also true if all p pdfs $Y|(W = a_i) \sim f_Z(y - \mu)$ are the same. An impractical randomization test uses all $M = \frac{n!}{n_1! \cdots n_p!}$ ways of assigning n_i of the Y_{ij} to treatment i for $i = 1, ..., p$. Let F_0 be the usual F statistic. The F statistic is computed for each of the M permutations and H_0 is rejected if the proportion of the M F statistics that are larger than F_0 is less than δ. The distribution of the M F statistics is approximately $F_{p-1,n-p}$ for large n when H_0 is true. The power of the randomization test is also similar to that of the usual F test. See Hoeffding (1952). These results suggest that the usual F test is semiparametric: the pvalue is approximately correct if n is large and if all p pdfs $Y|(W = a_i) \sim f_Z(y - \mu)$ are the same.

Let $[x]$ be the integer part of x, eg $[7.7] = 7$. Olive (2011b) shows that practical randomization tests that use a random sample of $\max(1000, [n \log(n)])$ permutations have level and power similar to the tests that use all M possible permutations. See Ernst (2009) and the *mpack* function `rand1way` for R code.

Another alternative to one way ANOVA is to use feasible weighted least squares (FWLS) on the cell means model with $\sigma^2 \boldsymbol{V} = diag(\sigma_1^2, ..., \sigma_p^2)$ where σ_i^2 is the variance of the ith group for $i = 1, ..., p$. Then $\hat{\boldsymbol{V}} = diag(S_1^2, ..., S_p^2)$ where $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i0})^2$ is the sample variance of the Y_{ij}. Hence the estimated weights for FWLS are $\hat{w}_{ij} \equiv \hat{w}_i = 1/S_i^2$. Then the FWLS cell means model has $Y = \boldsymbol{X}_c \boldsymbol{\beta}_c + \boldsymbol{e}$ as in (10.4) except $\mathrm{Cov}(\boldsymbol{e}) = diag(\sigma_1^2, ..., \sigma_p^2)$.

Hence $\boldsymbol{Z} = \boldsymbol{U}_c \boldsymbol{\beta}_c + \boldsymbol{\epsilon}$. Then $\boldsymbol{U}_c^T \boldsymbol{U}_c = diag(n_1 \hat{w}_1, ..., n_p \hat{w}_p)$, $(\boldsymbol{U}_c^T \boldsymbol{U}_c)^{-1} = diag(S_1^2/n_1, ..., S_p^2/n_p) = (\boldsymbol{X} \hat{\boldsymbol{V}}^{-1} \boldsymbol{X}^T)^{-1}$, and $\boldsymbol{U}_c^T \boldsymbol{Z} = (\hat{w}_1 Y_{10}, ..., \hat{w}_p Y_{p0})^T$. Thus

$$\hat{\boldsymbol{\beta}}_{FWLS} = (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T = \hat{\boldsymbol{\beta}}_c.$$

That is, the FWLS estimator equals the one way ANOVA estimator of $\boldsymbol{\beta}$ based on OLS applied to the cell means model. The ANOVA F test generalizes the pooled t test in that the two tests are equivalent for $p = 2$. The FWLS procedure is also known as the Welch one way ANOVA and generalizes the Welch t test. The Welch t test is thought to be much better than the pooled t test. See Brown and Forsythe (1974ab), Kirk (1982, p. 100, 101, 121, 122) and Welch (1947, 1951).

In matrix form $\boldsymbol{Z} = \boldsymbol{U}_c\boldsymbol{\beta}_c + \boldsymbol{\epsilon}$ becomes

$$\begin{bmatrix} \sqrt{\hat{w}_1}Y_{1,1} \\ \vdots \\ \sqrt{\hat{w}_1}Y_{1,n_1} \\ \sqrt{\hat{w}_2}Y_{21} \\ \vdots \\ \sqrt{\hat{w}_2}Y_{2,n_2} \\ \vdots \\ \sqrt{\hat{w}_p}Y_{p,1} \\ \vdots \\ \sqrt{\hat{w}_p}Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} \sqrt{\hat{w}_1} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sqrt{\hat{w}_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\hat{w}_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \sqrt{\hat{w}_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{\hat{w}_p} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{\hat{w}_p} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2,n_2} \\ \vdots \\ \epsilon_{p,1} \\ \vdots \\ \epsilon_{p,n_p} \end{bmatrix}. \tag{10.9}$$

Four tests for $Ho : \mu_1 = \cdots = \mu_p$ can be used if Rule of Thumb 10.3: $\max(S_1, ..., S_p) \leq 2\min(S_1, ..., S_p)$ fails. Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$, and let $Y_{(1)} \leq Y_{(2)} \cdots \leq Y_{(n)}$ be the order statistics. Then the rank transformation of the response is $\boldsymbol{Z} = rank(\boldsymbol{Y})$ where $Z_i = j$ if $Y_i = Y_{(j)}$ is the jth order statistic. For example, if $\boldsymbol{Y} = (7.7, 4.9, 33.3, 6.6)^T$, then $\boldsymbol{Z} = (3, 1, 4, 2)^T$. The first test performs the one way ANOVA F test with $\boldsymbol{Z}$ replacing $\boldsymbol{Y}$. See Montgomery (1984, p. 117-118). Two of the next three tests are described in Brown and Forsythe (1974b). Let $\lceil x \rceil$ be the smallest integer $\geq x$, eg $\lceil 7.7 \rceil = 8$. Then the Welch (1951) ANOVA F test uses test statistic

$$F_W = \frac{\sum_{i=1}^p w_i(\overline{Y}_{i0} - \tilde{Y}_{00})^2/(p-1)}{1 + \frac{2(p-2)}{p^2-1}\sum_{i=1}^p (1 - \frac{w_i}{u})^2/(n_i - 1)}$$

where $w_i = n_i/S_i^2, u = \sum_{i=1}^p w_i$ and $\tilde{Y}_{00} = \sum_{i=1}^p w_i\overline{Y}_{i0}/u$. Then the test statistic is compared to an F_{p-1,d_W} distribution where $d_W = \lceil f \rceil$ and

$$1/f = \frac{3}{p^2-1}\sum_{i=1}^p (1 - \frac{w_i}{u})^2/(n_i - 1).$$

For the modified Welch (1947) test, the test statistic is compared to an $F_{p-1,d_{MW}}$ distribution where $d_{MW} = \lceil f \rceil$ and

$$f = \frac{\sum_{i=1}^p (S_i^2/n_i)^2}{\sum_{i=1}^p \frac{1}{n_i-1}(S_i^2/n_i)^2} = \frac{\sum_{i=1}^p (1/w_i)^2}{\sum_{i=1}^p \frac{1}{n_i-1}(1/w_i)^2}.$$

Some software uses f instead of d_W or d_{MW}, and variants on the denominator degrees of freedom d_W or d_{MW} are common.

The modified ANOVA F test uses test statistic

$$F_M = \frac{\sum_{i=1}^p n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2}{\sum_{i=1}^p (1 - \frac{n_i}{n})S_i^2}$$

The test statistic is compared to an F_{p-1,d_M} distribution where $d_M = \lceil f \rceil$ and

$$1/f = \sum_{i=1}^p c_i^2/(n_i - 1)$$

where

$$c_i = (1 - \frac{n_i}{n})S_i^2/[\sum_{i=1}^p (1 - \frac{n_i}{n})S_i^2].$$

The `mpack` function *anovasim* can be used to compare the five tests.

Huberty and Olejnik (2006) and Khattree and Naik (1999, ch. 4) are useful reference for MANOVA. Mardia (1971) notes that the one way MANOVA test based on Pillai's trace V is robust to nonnormality, especially when all of the treatment sample sizes are the same: $n_i \equiv h$. Permutation tests offer an alternative. See, for example, Anderson (2001).

10.7 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

10.1*. In the MANOVA model, $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}_i$, and $\boldsymbol{Y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{e}_i$. Treating $\boldsymbol{X}\boldsymbol{\beta}_i$ as a constant, $\text{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij}\boldsymbol{I}_n$. Using this information, show $\text{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_j) = \sigma_{ij}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

10.2. SAS Institute (1985, p. 498 - 501) describes a one way MANOVA model. There are two groups for gender: female and male. There were $p = 4$ (skull measurements) variables $X_1 = length$ $X_2 = basilar$, $X_3 = zygomat$ and $X_4 = postorb$. There were $n_1 = 18$ females and $n_2 = 22$ males measured. Suppose $t_0 = 0.9567$ and pvalue $= 0.6566$. Here t_o was Wilk's lambda, but the other three test statistics gave the same pvalue. Do a 4 step one way MANOVA test.

10.3. Suppose the 15 units are 1 Adatorwovor, 2 Adhikari, 3 Alanzi, 4 Alsibiani, 5 AlTalib, 6 Fan, 7 Kuo, 8 Lamsal, 9 Liu, 10 Meyer, 11 Peiris, 12 Rathnayake, 13 Rupasinghe, 14 Schroeppel and 15 Watagoda. Use the following output to allocate the 15 units to three groups of 5. Show the three groups.

```
> sample(15)
 [1]  6  3  4  2  1 10  7  5 12 15 13  8 14 11  9
```

R/Splus Problems

Warning: Use the command *source("G:/mpack.txt")* **to download the programs. See Preface or Section 15.2.** Typing the name of the `mpack` function, eg *ddplot*, will display the code for the function. Use the `args` command, eg *args(ddplot)*, to display the needed arguments for the function.

10.4. The Johnson and Wichern (1988, p. 262) turtle data gives the length, width and height of painted turtle shells. There is a sample of 24 female and a sample of 24 male turtles.

a) The *R* command for this part make the response and residual plots for each of the three variables. Click the rightmost mouse button and highlight *Stop* to advance the plot. When you have the response and residual plots for one variable on the screen, copy and paste the two plots into *Word*. Do this three times, once for each variable. The male turtles are smaller than the female turtles.

b) The *R* command for this plot makes a DD plot of the residuals and adds the lines corresponding to the three prediction regions of Section 5.2. The robust cutoff is larger than the semiparametric cutoff. Place the plot in *Word*. Do the residuals appear to follow a multivariate normal distribution?