

# Multivariate Statistical Analysis

## Fall 2011

C. L. Williams, Ph.D.

Lecture 21 for Applied Multivariate Analysis

# Outline

## 1 Growth Curves

# Matrices in Repeated Measures Tests

When the subject responds to a treatment or stimulus at successive time periods, the pattern of responses is often referred to as a growth curve. As in repeated measures experiments, subjects are usually human or animal.

# Comparisons via Contrasts

## Response curves as polynomials

The usual approach is to approximate the shape of the growth curve by a polynomial function of time. If the time points are equally spaced, we can use orthogonal polynomials. This approach will be described first, followed by a method suitable for unequal time intervals. Orthogonal polynomials are special contrasts that are often used in testing for linear, quadratic, cubic, and higher order trends in quantitative factors.

Suppose we administer a drug to some subjects and measure a certain reaction at 3-min intervals. Let  $\mu_1, \mu_2, \mu_3, \mu_4$ , and  $\mu_5$ , designate the average responses at 0, 3, 6, 9, and 12 min, respectively. To test the hypothesis that there are no trends in the  $\mu_1$ 's, we could test  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  or  $H_0: \mathbf{C}\boldsymbol{\mu}$  using the contrast matrix

$$\mathbf{C} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix}$$

$$T^2 = n(\mathbf{C}\bar{\mathbf{y}})' (\mathbf{CSC}')^{-1} (\mathbf{C}\bar{\mathbf{y}})' \sim T_{p-1, n-1}^2$$

The four rows of  $\mathbf{C}$  are orthogonal polynomials that test for linear, quadratic, cubic, and quartic trends in the means. As noted in Section 6.9.2, any set of orthogonal contrasts in  $\mathbf{C}$  will give the same value of  $T^2$  to test  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ . However, in this case we will be interested in using a subset of the rows of  $\mathbf{C}$  to determine the shape of the response curve.

a sample of  $n$  observation vectors  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ , for which we compute  $\bar{\mathbf{y}}$  and  $\mathbf{S}$ . The usual approach is to approximate the shape of the growth curve by a polynomial function of time. If the time points are equally spaced, we can use orthogonal polynomials. This approach will be described first, followed by a method suitable for unequal time intervals.

*Orthogonal polynomials* are special contrasts that are often used in testing for linear, quadratic, cubic, and higher order trends in quantitative factors. For a more complete description and derivation see Guttman (1982, pp. 194–207), Morrison (1983, pp. 182–188), or Rencher (2000, pp. 323–331). Here we give only a heuristic introduction to the use of these contrasts.

Suppose we administer a drug to some subjects and measure a certain reaction at 3-min intervals. Let  $\mu_1, \mu_2, \mu_3, \mu_4$ , and  $\mu_5$  designate the average responses at 0, 3, 6, 9, and 12 min, respectively. To test the hypothesis that there are no trends in the  $\mu_j$ 's, we could test  $H_0: \mu_1 = \mu_2 = \dots = \mu_5$  or  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$  using the contrast matrix

$$\mathbf{C} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & 2 & 0 & -2 & 1 \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix} \quad (6.109)$$

in  $T^2 = n(\mathbf{C}\bar{\mathbf{y}})'(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}})$ , as in (6.90). The four rows of  $\mathbf{C}$  are orthogonal polynomials that test for linear, quadratic, cubic, and quartic trends in the means. As noted in Section 6.9.2, any set of orthogonal contrasts in  $\mathbf{C}$  will give the same value of  $T^2$  to test  $H_0: \mu_1 = \mu_2 = \dots = \mu_5$ . However, in this case we will be interested in using a subset of the rows of  $\mathbf{C}$  to determine the shape of the response curve.

Table A.13 (Kleinbaum, Kupper, and Muller 1988) gives orthogonal polynomials for  $p = 3, 4, \dots, 10$ . The  $p - 1$  entries for each value of  $p$  constitute the matrix  $\mathbf{C}$ . Some software programs will generate these automatically.

As with all orthogonal contrasts, the rows of  $\mathbf{C}$  in (6.109) sum to zero and are mutually orthogonal. It is also apparent that the coefficients in each row increase and decrease in conformity with the desired pattern. Thus the entries in the first row,  $(-2, -1, 0, 1, 2)$ , increase steadily in a straight-line trend. The values in the second row dip down and back up in a quadratic-type bend. The third-row entries increase, decrease, then increase in a cubic pattern with two bends. The fourth row bends three times in a quartic curve.

To further illustrate how the orthogonal polynomials pinpoint trends in the means when testing  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , consider the three different patterns for  $\boldsymbol{\mu}$  depicted in Figure 6.4, where  $\boldsymbol{\mu}'_a = (8, 8, 8, 8, 8)$ ,  $\boldsymbol{\mu}'_b = (20, 16, 12, 8, 4)$ , and  $\boldsymbol{\mu}'_c = (5, 12, 15, 12, 5)$ . Let us denote the rows of  $\mathbf{C}$  in (6.109) as  $\mathbf{c}'_1, \mathbf{c}'_2, \mathbf{c}'_3$ , and  $\mathbf{c}'_4$ . It is clear that  $\mathbf{c}'_i \boldsymbol{\mu}_a = 0$  for  $i = 1, 2, 3, 4$ ; that is, when  $H_0: \mu_1 = \dots = \mu_5$  is true, all four comparisons confirm it. If  $\boldsymbol{\mu}$  has the pattern  $\boldsymbol{\mu}_b$ , only  $\mathbf{c}'_1 \boldsymbol{\mu}_b$  is nonzero. The other rows are not sensitive to a linear pattern. We illustrate this for  $\mathbf{c}'_1$  and  $\mathbf{c}'_2$ :

$$\mathbf{c}'_1 \boldsymbol{\mu}_b = (-2)(20) + (-1)(16) + (0)(12) + (1)(8) + (2)(4) = -44,$$

$$\mathbf{c}'_2 \boldsymbol{\mu}_b = 2(20) - 16 - 2(12) - 8 + 2(4) = 0.$$



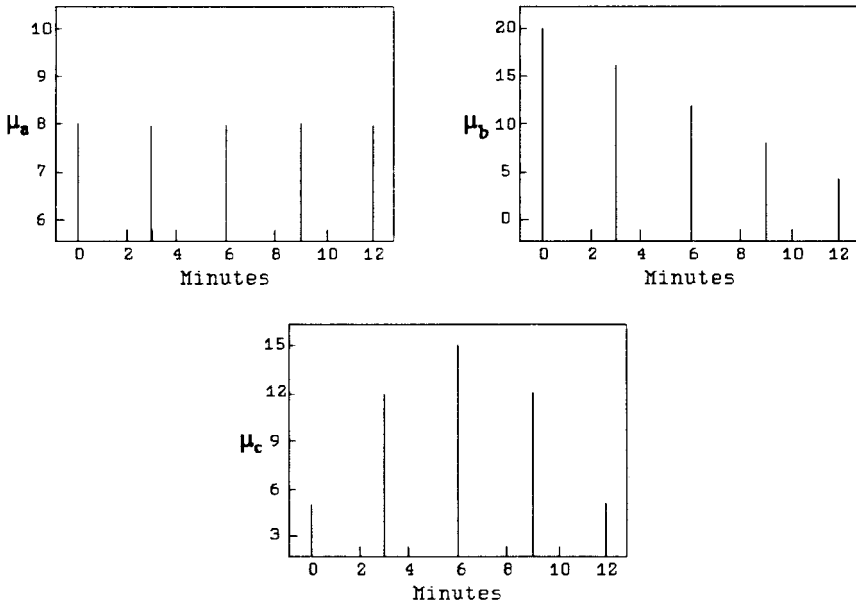


Figure 6.4. Three different patterns for  $\mu$ .

For  $\mu_c$ , only  $c'_2\mu_c$  is nonzero. For example,

$$c'_1\mu_c = -2(5) - 12 + 12 + 2(5) = 0,$$

$$c'_2\mu_c = 2(5) - 12 - 2(15) - 12 + 2(5) = -19.$$

Thus each orthogonal polynomial independently detects the type of curvature it is designed for and ignores other types. Of course real curves generally exhibit a mixture of more than one type of curvature, and in practice more than one orthogonal polynomial contrast may be significant.

To test hypotheses about the shape of the curve, we therefore use the appropriate rows of  $C$  in (6.109). Suppose we suspected a priori that there would be a combined linear and quadratic trend. Then we would partition  $C$  as follows:

$$C_1 = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \end{pmatrix},$$

$$C_2 = \begin{pmatrix} -1 & 2 & 0 & -2 & 1 \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix}.$$

We would test  $H_0: C_1\mu = \mathbf{0}$  by

$$T^2 = n(C_1\bar{y})'(C_1SC_1)^{-1}(C_1\bar{y}),$$

which is distributed as  $T^2_{2,n-1}$ , where 2 is the number of rows of  $C_1$ ,  $n$  is the number of subjects in the sample, and  $\bar{y}$  and  $S$  are the mean vector and covariance matrix for

the sample. Similarly,  $H_0: \mathbf{C}_2\boldsymbol{\mu} = \mathbf{0}$  is tested by

$$T^2 = n(\mathbf{C}_2\bar{\mathbf{y}})'(\mathbf{C}_2\mathbf{S}\mathbf{C}_2')^{-1}(\mathbf{C}_2\bar{\mathbf{y}}),$$

which is  $T_{2,n-1}^2$ . In this case we might expect the first to reject  $H_0$  and the second to accept  $H_0$ .

If we have no a priori expectations as to the shape of the curve, we could proceed as follows. Test the overall hypothesis  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , and if  $H_0$  is rejected, use each of the rows of  $\mathbf{C}$  separately to test  $H_0: \mathbf{c}'_i\boldsymbol{\mu} = 0, i = 1, 2, 3, 4$ . The respective test statistics are

$$T^2 = n(\mathbf{C}\bar{\mathbf{y}})'(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}}),$$

which is  $T_{4,n-1}^2$ , and

$$t_i = \frac{\mathbf{c}'_i\bar{\mathbf{y}}}{\sqrt{\mathbf{c}'_i\mathbf{S}\mathbf{c}_i/n}}, \quad i = 1, 2, 3, 4,$$

each of which is distributed as  $t_{n-1}$  (see Example 6.9.2).

In a case where  $p$  is large so that  $\boldsymbol{\mu}$  has a large number of levels, say 10 or more, we would likely want to stop testing after the first four or five rows of  $\mathbf{C}$  and test the remaining rows in one group. However, for larger values of  $p$ , most tables of orthogonal polynomials give only the first few rows and omit those corresponding to higher degrees of curvature. We can find a matrix whose rows are orthogonal to the rows of a given matrix as follows. Suppose  $p = 11$  so that  $\mathbf{C}$  is  $10 \times 11$  and  $\mathbf{C}_1$  contains the first five orthogonal polynomials. Then a matrix  $\mathbf{C}_2$ , with rows orthogonal to those of  $\mathbf{C}_1$ , can be obtained by selecting five linearly independent rows of

$$\mathbf{B} = \mathbf{I} - \mathbf{C}'_1(\mathbf{C}_1\mathbf{C}'_1)^{-1}\mathbf{C}_1, \quad (6.110)$$

whose rows can easily be shown to be orthogonal to those of  $\mathbf{C}_1$ . The matrix  $\mathbf{B}$  is not full rank, and some care must be exercised in choosing linearly independent rows. However, if an incorrect choice of  $\mathbf{C}_2$  is made, the computer algorithm should indicate this as it attempts to invert  $\mathbf{C}_2\mathbf{S}\mathbf{C}'_2$  in  $T^2 = n(\mathbf{C}_2\bar{\mathbf{y}})'(\mathbf{C}_2\mathbf{S}\mathbf{C}'_2)^{-1}(\mathbf{C}_2\bar{\mathbf{y}})$ .

Alternatively, to check for significant curvature beyond the rows of  $\mathbf{C}_1$  without finding  $\mathbf{C}_2$ , we can use the test for additional information in a subset of variables in Section 5.8. We need not find  $\mathbf{C}_2$  in order to find the overall  $T^2$ , since, as noted in Section 6.9.2, any full rank  $(p-1) \times p$  matrix  $\mathbf{C}$  such that  $\mathbf{C}\mathbf{j} = \mathbf{0}$  will give the same value in the overall  $T^2$ -test of  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ . We can conveniently use a simple contrast matrix such as

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

in

$$T^2 = n(\mathbf{C}\bar{\mathbf{y}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{y}}), \tag{6.111}$$

which is  $T_{p-1, n-1}^2$ . Let  $p_1$  be the number of orthogonal polynomials in  $\mathbf{C}_1$  and  $p_2$  be the number of rows of  $\mathbf{C}_2$  if it were available; that is  $p_1 + p_2 = p - 1$ . Then the test statistic for the  $p_1$  orthogonal polynomials in  $\mathbf{C}_1$  is

$$T_1^2 = n(\mathbf{C}_1\bar{\mathbf{y}})'(\mathbf{C}_1\mathbf{S}\mathbf{C}_1')^{-1}(\mathbf{C}_1\bar{\mathbf{y}}), \tag{6.112}$$

which is  $T_{p_1, n-1}^2$ . We wish to compare  $T_1^2$  in (6.112) to  $T^2$  in (6.111) to check for significant curvature beyond the rows of  $\mathbf{C}_1$ . However, the test for additional information in a subset of variables in Section 5.8 was for the two-sample case. We can adapt (5.29) for use with the one-sample case, as follows. The test for significance of any curvature remaining after that accounted for in  $\mathbf{C}_1$  is made by comparing

$$(n - p_1 - 1) \frac{T^2 - T_1^2}{n - 1 + T_1^2}$$

with the critical value  $T_{\alpha, p_2, n-p_1-1}^2$ .

We now describe an approach that can be used when the time points are not equally spaced. It may also be of interest in the equal-time-increment case because it provides an estimate of the response function.

Suppose we observe the response of the subject at  $p$  time points  $t_1, t_2, \dots, t_p$  and that the average response  $\mu$  at any time point  $t$  is a polynomial in  $t$  of degree  $k < p$ :

$$\mu = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^k.$$

This holds for each point  $t_r$  and the corresponding average response  $\mu_r$ . Thus our hypothesis becomes

$$H_0: \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 t_1 + \beta_2 t_1^2 + \dots + \beta_k t_1^k \\ \beta_0 + \beta_1 t_2 + \beta_2 t_2^2 + \dots + \beta_k t_2^k \\ \vdots \\ \beta_0 + \beta_1 t_p + \beta_2 t_p^2 + \dots + \beta_k t_p^k \end{pmatrix}, \tag{6.113}$$

which can be expressed in matrix notation as

$$H_0: \boldsymbol{\mu} = \mathbf{A}\boldsymbol{\beta}, \tag{6.114}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^k \\ 1 & t_2 & t_2^2 & \cdots & t_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_p & t_p^2 & \cdots & t_p^k \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}.$$

In practice, it may be useful to transform the  $t_r$ 's by subtracting the mean or the smallest value in order to reduce their size for computational purposes.

The following method of testing  $H_0$  is due to Rao (1959, 1973). The model  $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\beta}$  is similar to a regression model  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  (see Section 10.2.1). However, in this case, we have  $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$  rather than  $\sigma^2\mathbf{I}$ , as in the standard regression assumption. In place of the usual regression approach of seeking  $\boldsymbol{\beta}$  to minimize  $\text{SSE} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$  [see (10.4) and (10.6)], we use a standardized distance as in (3.80),  $(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}})' \mathbf{S}^{-1}(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}})$ . The value of  $\boldsymbol{\beta}$  that minimizes  $(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}})' \mathbf{S}^{-1}(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}})$  is

$$\hat{\boldsymbol{\beta}} = (\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\bar{\mathbf{y}} \quad (6.115)$$

[see Rencher (2000, Section 7.8.1)], and  $H_0: \boldsymbol{\mu} = \mathbf{A}\boldsymbol{\beta}$  can be tested by

$$T^2 = n(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}})' \mathbf{S}^{-1}(\bar{\mathbf{y}} - \mathbf{A}\hat{\boldsymbol{\beta}}), \quad (6.116)$$

which is distributed as  $T_{p-k-1, n-1}^2$ . The dimension of  $T^2$  is reduced from  $p$  to  $p - k - 1$  because  $k + 1$  parameters have been estimated in  $\hat{\boldsymbol{\beta}}$ . The  $T^2$ -statistic in (6.116) is usually given in the equivalent form

$$T^2 = n(\bar{\mathbf{y}}'\mathbf{S}^{-1}\bar{\mathbf{y}} - \bar{\mathbf{y}}'\mathbf{S}^{-1}\mathbf{A}\hat{\boldsymbol{\beta}}). \quad (6.117)$$

The mean response at the  $r$ th time point,

$$\begin{aligned} \mu_r &= \beta_0 + \beta_1 t_r + \beta_2 t_r^2 + \cdots + \beta_k t_r^k \\ &= (1, t_r, t_r^2, \dots, t_r^k)\boldsymbol{\beta} = \mathbf{a}'_r \boldsymbol{\beta}, \end{aligned}$$

can be estimated by

$$\hat{\mu}_r = \mathbf{a}'_r \hat{\boldsymbol{\beta}}. \quad (6.118)$$

Simultaneous confidence intervals for all possible  $\mathbf{a}'\boldsymbol{\beta}$  are given by

$$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm \frac{T_\alpha}{\sqrt{n}} \sqrt{\mathbf{a}'(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{a} \left(1 + \frac{T^2}{n-1}\right)}, \quad (6.119)$$

where  $T_\alpha = \sqrt{T_{\alpha, k+1, n-1}^2}$  is from Table A.7 and  $T^2$  is given by (6.116) or (6.117).

The intervals in (6.119) for  $\mathbf{a}'\boldsymbol{\beta}$  include, of course,  $\mathbf{a}'_r\boldsymbol{\beta}$  for the  $p$  rows of  $\mathbf{A}$ , that is, confidence intervals for the  $p$  time points. If  $\mathbf{a}'_r\boldsymbol{\beta}$ ,  $r = 1, 2, \dots, p$ , are the only values of interest, we can shorten the intervals in (6.119) by using a Bonferroni coefficient  $t_{\alpha/2p}$  in place of  $T_\alpha$ :

$$\mathbf{a}'_r\hat{\boldsymbol{\beta}} \pm \frac{t_{\alpha/2p}}{\sqrt{n}} \sqrt{\mathbf{a}'_r(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{a}_r \left(1 + \frac{T^2}{n-1}\right)}, \tag{6.120}$$

where  $t_{\alpha/2p} = t_{\alpha/2p, n-1}$ . Bonferroni critical values  $t_{\alpha/2p, v}$  are given in Table A.8. See procedures 2 and 3 in Section 5.5 for additional comments on the use of  $t_{\alpha/2p}$  and  $T_\alpha$ .

**Example 6.10.1.** Potthoff and Roy (1964) reported measurements in a dental study on boys and girls from ages 8 to 14. The data are given in Table 6.16.

To illustrate the methods of this section, we use the data for the boys alone. In Example 6.10.2 we will compare the growth curves of the boys with those of the girls. We first test the overall hypothesis  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , where  $\mathbf{C}$  contains orthogonal polynomials for linear, quadratic, and cubic effects:

$$\mathbf{C} = \begin{pmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}. \tag{6.121}$$

**Table 6.16. Dental Measurements**

Subject	Girls' Ages in Years				Subject	Boys' Ages in Years			
	8	10	12	14		8	10	12	14
1	21.0	20.0	21.5	23.0	1	26.0	25.0	29.0	31.0
2	21.0	21.5	24.0	25.5	2	21.5	22.5	23.0	26.5
3	20.5	24.0	24.5	26.0	3	23.0	22.5	24.0	27.5
4	23.5	24.5	25.0	26.5	4	25.5	27.5	26.5	27.0
5	21.5	23.0	22.5	23.5	5	20.0	23.5	22.5	26.0
6	20.0	21.0	21.0	22.5	6	24.5	25.5	27.0	28.5
7	21.5	22.5	23.0	25.0	7	22.0	22.0	24.5	26.5
8	23.0	23.0	23.5	24.0	8	24.0	21.5	24.5	25.5
9	20.0	21.0	22.0	21.5	9	23.0	20.5	31.0	26.0
10	16.5	19.0	19.0	19.5	10	27.5	28.0	31.0	31.5
11	24.5	25.0	28.0	28.0	11	23.0	23.0	23.5	25.0
					12	21.5	23.5	24.0	28.0
					13	17.0	24.5	26.0	29.5
					14	22.5	25.5	25.5	26.0
					15	23.0	24.5	26.0	30.0
					16	22.0	21.5	23.5	25.0

From the 16 observation vectors we obtain

$$\bar{\mathbf{y}} = \begin{pmatrix} 22.88 \\ 23.81 \\ 25.72 \\ 27.47 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 6.02 & 2.29 & 3.63 & 1.61 \\ 2.29 & 4.56 & 2.19 & 2.81 \\ 3.63 & 2.19 & 7.03 & 3.24 \\ 1.61 & 2.81 & 3.24 & 4.35 \end{pmatrix}.$$

To test  $H_0: \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ , we calculate

$$T^2 = n(\mathbf{C}\bar{\mathbf{y}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{y}}) = 77.957,$$

which exceeds  $T_{.01,3,15}^2 = 19.867$ . We now test  $H_0: \mathbf{c}'_i\boldsymbol{\mu} = 0$  for each row of  $\mathbf{C}$  to determine the shape of the growth curve. For the linear effect, using the first row,  $\mathbf{c}'_1$ , we obtain

$$t_1 = \frac{\mathbf{c}'_1\bar{\mathbf{y}}}{\sqrt{\mathbf{c}'_1\mathbf{S}\mathbf{c}_1/n}} = 7.722 > t_{.005,15} = 2.947.$$

The test of significance of the quadratic component using the second row yields

$$t_2 = \frac{\mathbf{c}'_2\bar{\mathbf{y}}}{\sqrt{\mathbf{c}'_2\mathbf{S}\mathbf{c}_2/n}} = 1.370 < t_{.025,15} = 2.131.$$

To test for a cubic trend, we use the third row of  $\mathbf{C}$ :

$$t_3 = \frac{\mathbf{c}'_3\bar{\mathbf{y}}}{\sqrt{\mathbf{c}'_3\mathbf{S}\mathbf{c}_3/n}} = -.511 > -t_{.025,15} = -2.131.$$

Thus only the linear trend is needed to describe the growth curve.

To model the curve for each variable, we use (6.113),

$$\mu_r = \beta_0 + \beta_1 t_r, \quad r = 1, 2, 3, 4, \quad \text{or} \\ \boldsymbol{\mu} = \mathbf{A}\boldsymbol{\beta},$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & -3 \\ 1 & -1 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

The values in the second column of  $\mathbf{A}$  are obtained as  $t = \text{age} - 11$ . By (6.115), we obtain

$$\hat{\boldsymbol{\beta}} = (\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\bar{\mathbf{y}} = \begin{pmatrix} 25.002 \\ .834 \end{pmatrix},$$

and our prediction equation is

$$\begin{aligned} \hat{\mu} &= 25.002 + .834t = 25.002 + .834(\text{age} - 11) \\ &= 15.828 + .834(\text{age}). \end{aligned} \quad \square$$

**6.10.2 Growth Curves for Several Samples**

For the case of several samples or groups, the data layout would be similar to that in Table 6.9, where the  $p$  levels of factor  $A$  represent time points. Assuming the time points are equally spaced, we can use orthogonal polynomials in the  $(p - 1) \times p$  contrast matrix  $\mathbf{C}$  and express the basic hypothesis in the form  $H_0: \mathbf{C}\bar{\boldsymbol{\mu}} = \mathbf{0}$ , where  $\bar{\boldsymbol{\mu}} = \sum_{i=1}^k \boldsymbol{\mu}_i/k$ . This is equivalent to  $H_0: \bar{\mu}_{.1} = \bar{\mu}_{.2} = \dots = \bar{\mu}_{.p}$ , which compares the means of the  $p$  time points averaged across groups. As in Section 6.9.3, let us denote the sample mean vectors for the  $k$  groups as  $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_k$ , with grand mean  $\bar{\mathbf{y}}_{..}$  and pooled covariance matrix  $\mathbf{S}_{pl} = \mathbf{E}/\nu_E$ . For the overall test of  $H_0: \mathbf{C}\bar{\boldsymbol{\mu}} = \mathbf{0}$  we use the test statistic

$$T^2 = N(\mathbf{C}\bar{\mathbf{y}}_{..})'(\mathbf{C}\mathbf{S}_{pl}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}}_{..}), \tag{6.122}$$

which is  $T_{p-1, \nu_E}^2$  as in (6.93), where  $N = \sum_{i=1}^k n_i$  for unbalanced data or  $N = kn$  for balanced data. The corresponding degrees of freedom for error is given by  $\nu_E = N - k$  or  $\nu_E = k(n - 1)$ . A test that the average growth curve (averaged over groups) has a particular form can be tested with  $\mathbf{C}_1$ , containing a subset of the rows of  $\mathbf{C}$ :

$$T^2 = N(\mathbf{C}_1\bar{\mathbf{y}}_{..})'(\mathbf{C}_1\mathbf{S}_{pl}\mathbf{C}'_1)^{-1}(\mathbf{C}_1\bar{\mathbf{y}}_{..}), \tag{6.123}$$

which is distributed as  $T_{p_1, \nu_E}^2$ , where  $p_1$  is the number of rows in  $\mathbf{C}_1$ .

The growth curves for the  $k$  groups can be compared by the interaction or parallelism test of Section 6.9.3 using either  $\mathbf{C}$  or  $\mathbf{C}_1$ . We do a one-way MANOVA on  $\mathbf{C}\mathbf{y}_{ij}$  or  $\mathbf{C}_1\mathbf{y}_{ij}$ , or equivalently calculate by (6.96),

$$\Lambda = \frac{|\mathbf{C}\mathbf{E}\mathbf{C}'|}{|\mathbf{C}(\mathbf{E} + \mathbf{H})\mathbf{C}'|} \quad \text{or} \quad \Lambda = \frac{|\mathbf{C}_1\mathbf{E}\mathbf{C}'_1|}{|\mathbf{C}_1(\mathbf{E} + \mathbf{H})\mathbf{C}'_1|}, \tag{6.124}$$

which are distributed as  $\Lambda_{p-1, k-1, \nu_E}$  and  $\Lambda_{p_1, k-1, \nu_E}$ , respectively.

**Example 6.10.2.** In Example 6.10.1, we found a linear trend for the growth curve for dental measurements of boys in Table 6.16. We now consider the growth curve for the combined group and also compare the girls' group with the boys' group.

The two sample sizes are unequal and we use (6.33) to calculate the  $\mathbf{E}$  matrix for the two groups,

$$\mathbf{E} = \begin{pmatrix} 135.39 & 67.88 & 97.76 & 67.76 \\ 67.88 & 103.76 & 72.86 & 82.71 \\ 97.76 & 72.86 & 161.39 & 103.27 \\ 67.76 & 82.71 & 103.27 & 124.64 \end{pmatrix},$$

from which we obtain  $\mathbf{S}_{pl} = \mathbf{E}/v_E$ . Using the  $\mathbf{C}$  matrix in (6.121), we can test the basic hypothesis of equal means for the combined samples,  $H_0: \mathbf{C}\bar{\boldsymbol{\mu}} = \mathbf{0}$ , using (6.122):

$$\begin{aligned} T^2 &= N(\mathbf{C}\bar{\mathbf{y}}_{..})'(\mathbf{C}\mathbf{S}_{pl}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}}_{..}) \\ &= 118.322 > T_{.01,3,25}^2 = 15.538. \end{aligned}$$

To test for a linear trend, we use the first row of  $\mathbf{C}$  in (6.123):

$$\begin{aligned} T^2 &= N(\mathbf{c}'_1\bar{\mathbf{y}}_{..})'(\mathbf{c}'_1\mathbf{S}_{pl}\mathbf{c}_1)^{-1}(\mathbf{c}'_1\bar{\mathbf{y}}_{..}) \\ &= 99.445 > T_{.01,1,25}^2 = 7.770. \end{aligned}$$

This is, of course, the square of a  $t$ -statistic, but in the  $T^2$  form it can readily be compared with the preceding  $T^2$  using all three rows of  $\mathbf{C}$ . The linear trend is seen to dominate the relationship among the means.

We now compare the growth curves of the two groups using (6.124). For  $\mathbf{C}$ , we obtain

$$\begin{aligned} \Lambda &= \frac{|\mathbf{C}\mathbf{E}\mathbf{C}'|}{|\mathbf{C}(\mathbf{E} + \mathbf{H})\mathbf{C}'|} = \frac{1.3996 \times 10^8}{1.9025 \times 10^8} \\ &= .736 > \Lambda_{.05,3,1,25} = .717. \end{aligned}$$

For the linear trend, we have

$$\begin{aligned} \Lambda &= \frac{|\mathbf{c}'_1\mathbf{E}\mathbf{c}_1|}{|\mathbf{c}'_1(\mathbf{E} + \mathbf{H})\mathbf{c}_1|} = \frac{1184.2}{1427.9} \\ &= .829 < \Lambda_{.05,1,1,25} = .855. \end{aligned}$$

Thus the overall comparison does not reach significance, but the more specific comparison of linear trends does give a significant result.  $\square$

### 6.10.3 Additional Topics

Jackson and Bryce (1981) presented methods of analyzing growth curves based on univariate linear models. Snee (1972) and Snee Acuff, and Gibson (1979) proposed the use of eigenvalues and eigenvectors of a matrix derived from residuals after fitting the model. If one of the eigenvalues is dominant, certain simplifications result. Bryce (1980) discussed a similar simplification for the two-group case. Geisser (1980) and Fearn (1975, 1977) gave the Bayesian approach to growth curves, including estimation and prediction. Zerbe (1979a, b) provided a randomization test requiring fewer assumptions than normal-based tests.



### 6.11 TESTS ON A SUBVECTOR

#### 6.11.1 Test for Additional Information

In Section 5.8, we considered tests of significance of the additional information in a subvector when comparing two groups. We now extend these concepts to several groups and use similar notation.

Let  $\mathbf{y}$  be a  $p \times 1$  vector of measurements and  $\mathbf{x}$  be a  $q \times 1$  vector measured in addition to  $\mathbf{y}$ . We are interested in determining whether  $\mathbf{x}$  makes a significant contribution to the test of  $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k$  above and beyond  $\mathbf{y}$ . Another way to phrase the question is, Can the separation of groups achieved by  $\mathbf{x}$  be predicted from the separation achieved by  $\mathbf{y}$ ? It is not necessary, of course, that  $\mathbf{x}$  represent new variables. It may be that  $\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$  is a partitioning of the present variables, and we wish to know if the variables in  $\mathbf{x}$  can be deleted because they do not contribute to rejecting  $H_0$ .

We consider here only the one-way MANOVA, but the results could be extended to higher order designs, where various possibilities arise. In a two-way context, for example, it may happen that  $\mathbf{x}$  contributes nothing to the  $A$  main effect but does contribute significantly to the  $B$  main effect.

It is assumed that we have  $k$  samples,

$$\begin{pmatrix} \mathbf{y}_{ij} \\ \mathbf{x}_{ij} \end{pmatrix}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, n,$$

from which we calculate

$$\mathbf{E} = \begin{pmatrix} \mathbf{E}_{yy} & \mathbf{E}_{yx} \\ \mathbf{E}_{xy} & \mathbf{E}_{xx} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_{yy} & \mathbf{H}_{yx} \\ \mathbf{H}_{xy} & \mathbf{H}_{xx} \end{pmatrix},$$

where  $\mathbf{E}$  and  $\mathbf{H}$  are  $(p + q) \times (p + q)$  and  $\mathbf{E}_{yy}$  and  $\mathbf{H}_{yy}$  are  $p \times p$ .

Then

$$\Lambda(\mathbf{y}, \mathbf{x}) = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \tag{6.125}$$

is distributed as  $\Lambda_{p+q, \nu_H, \nu_E}$  and tests the significance of group separation using the full vector  $\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$ . In the balanced one-way model, the degrees of freedom are  $\nu_H = k - 1$  and  $\nu_E = k(n - 1)$ . To test group separation using the reduced vector  $\mathbf{y}$ , we can compute

$$\Lambda(\mathbf{y}) = \frac{|\mathbf{E}_{yy}|}{|\mathbf{E}_{yy} + \mathbf{H}_{yy}|}, \tag{6.126}$$

which is distributed as  $\Lambda_{p, \nu_H, \nu_E}$ .

To test the hypothesis that the extra variables in  $\mathbf{x}$  do not contribute anything significant to separating the groups beyond the information already available in  $\mathbf{y}$ ,

we calculate

$$\Lambda(\mathbf{x}|\mathbf{y}) = \frac{\Lambda(\mathbf{y}, \mathbf{x})}{\Lambda(\mathbf{y})}, \quad (6.127)$$

which is distributed as  $\Lambda_{q, v_H, v_E - p}$ . Note that the dimension of  $\Lambda(\mathbf{x}|\mathbf{y})$  is  $q$ , the number of  $x$ 's. The error degrees of freedom,  $v_E - p$ , has been adjusted for the  $p$   $y$ 's. Thus to test for the contribution of additional variables to separation of groups, we take the ratio of Wilks'  $\Lambda$  for the full set of variables in (6.125) to Wilks'  $\Lambda$  for the reduced set in (6.126). If the addition of  $\mathbf{x}$  makes  $\Lambda(\mathbf{y}, \mathbf{x})$  sufficiently smaller than  $\Lambda(\mathbf{y})$ , then  $\Lambda(\mathbf{x}|\mathbf{y})$  in (6.127) will be small enough to reject the hypothesis.

If we are interested in the effect of adding a single  $x$ , then  $q = 1$ , and (6.127) becomes

$$\Lambda(x|y_1, \dots, y_p) = \frac{\Lambda(y_1, \dots, y_p, x)}{\Lambda(y_1, \dots, y_p)}, \quad (6.128)$$

which is distributed as  $\Lambda_{1, v_H, v_E - p}$ . In this test we are inquiring whether  $x$  reduces the overall  $\Lambda$  by a significant amount. With a dimension of 1, the  $\Lambda$ -statistic in (6.128) has an exact  $F$ -transformation from Table 6.1,

$$F = \frac{1 - \Lambda}{\Lambda} \frac{v_E - p}{v_H}, \quad (6.129)$$

which is distributed as  $F_{v_H, v_E - p}$ . The statistic (6.128) is often referred to as a *partial  $\Lambda$ -statistic*; correspondingly, (6.129) is called a *partial  $F$ -statistic*.

In (6.128) and (6.129), we have a test of the significance of a variable in the presence of the other variables. For a breakdown of precisely how the contribution of a variable depends on the other variables, see Rencher (1993; 1998, Section 4.1.6).

We can rewrite (6.128) as

$$\Lambda(y_1, \dots, y_p, x) = \Lambda(x|y_1, \dots, y_p)\Lambda(y_1, \dots, y_p) \leq \Lambda(y_1, \dots, y_p), \quad (6.130)$$

which shows that Wilks'  $\Lambda$  can only decrease with an additional variable.

**Example 6.11.1.** We use the rootstock data of Table 6.2 to illustrate tests on subvectors. From Example 6.1.7, we have, for all four variables,  $\Lambda(y_1, y_2, y_3, y_4) = .1540$ . For the first two variables, we obtain  $\Lambda(y_1, y_2) = .6990$ . Then to test the significance of  $y_3$  and  $y_4$  adjusted for  $y_1$  and  $y_2$ , we have by (6.127),

$$\Lambda(y_3, y_4|y_1, y_2) = \frac{\Lambda(y_1, y_2, y_3, y_4)}{\Lambda(y_1, y_2)} = \frac{.1540}{.6990} = .2203,$$

which is less than the critical value  $\Lambda_{.05, 2, 5, 40} = .639$ .

Similarly, the test for  $y_4$  adjusted for  $y_1, y_2,$  and  $y_3$  is given by (6.128) as

$$\begin{aligned}\Lambda(y_4|y_1, y_2, y_3) &= \frac{\Lambda(y_1, y_2, y_3, y_4)}{\Lambda(y_1, y_2, y_3)} = \frac{.1540}{.2460} \\ &= .6261 < \Lambda_{.05,1,5,39} = .759.\end{aligned}$$

For each of the other variables, we have a similar test:

$$\begin{aligned}y_3: \quad \Lambda(y_3|y_1, y_2, y_4) &= \frac{.1540}{.2741} = .5618 < \Lambda_{.05,1,5,39} = .759, \\ y_2: \quad \Lambda(y_2|y_1, y_3, y_4) &= \frac{.1540}{.1922} = .8014 > \Lambda_{.05,1,5,39} = .759, \\ y_1: \quad \Lambda(y_1|y_2, y_3, y_4) &= \frac{.1540}{.1599} = .9630 > \Lambda_{.05,1,5,39} = .759.\end{aligned}$$

Thus the two variables  $y_3$  and  $y_4$ , either individually or together, contribute a significant amount to separation of the six groups.  $\square$

### 6.11.2 Stepwise Selection of Variables

If there are no variables for which we have a priori interest in testing for significance, we can do a data-directed search for the variables that best separate the groups. Such a strategy is often called *stepwise discriminant analysis*, although it could more aptly be called stepwise MANOVA. The procedure appears in many software packages.

We first describe an approach that is usually called *forward selection*. At the first step calculate  $\Lambda(y_i)$  for each individual variable and choose the one with minimum  $\Lambda(y_i)$  (or maximum associated  $F$ ). At the second step calculate  $\Lambda(y_i|y_1)$  for each of the  $p - 1$  variables not entered at the first step, where  $y_1$  indicates the first variable entered. For the second variable we choose the one with minimum  $\Lambda(y_i|y_1)$  (or maximum associated partial  $F$ ), that is, the variable that adds the maximum separation to the one entered at step 1. Denote the variable entered at step 2 by  $y_2$ . At the third step calculate  $\Lambda(y_i|y_1, y_2)$  for each of the  $p - 2$  remaining variables and choose the one that minimizes  $\Lambda(y_i|y_1, y_2)$  (or maximizes the associated partial  $F$ ). Continue this process until the  $F$  falls below some predetermined threshold value, say,  $F_{in}$ .

A *stepwise* procedure follows a similar sequence, except that after a variable has entered, the variables previously selected are reexamined to see if each still contributes a significant amount. The variable with smallest partial  $F$  will be removed if the partial  $F$  is less than a second threshold value,  $F_{out}$ . If  $F_{out}$  is the same as  $F_{in}$ , there is a very small possibility that the procedure will cycle continuously without stopping. This possibility can be eliminated by using a value of  $F_{out}$  slightly less than  $F_{in}$ . For an illustration of the stepwise procedure, see Example 8.9.