The Multivariate General Linear Model

- Extension of the univariate linear model to the multivariate case of vector observations
- The algebra is essentially the same as the univariate case
- Univariate variances are replaced by covariance matrices
- Univariate sums of squares are replaced by sums of squares and products (ssp) matrices
- Distribution theory analogous to that of the univariate case
- Test criteria are analogs of F-statistics
- There is more latitude in terms of hypotheses which can be tested

The Multivariate General Linear Model

- Let y_{ij} denote the response from subject i at time j, for $i = 1, \ldots, n$, $j = 1, \ldots, t$

- Suppose that the jth response from the ith individual was generated by the linear model

$$\begin{aligned} y_{ij} &= x_{i1}\beta_{1j} + x_{i2}\beta_{2j} + \cdots + x_{ip}\beta_{pj} + e_{ij} \\ &= \sum_{k=1}^{p} x_{ik}\beta_{kj} + e_{ij} \\ &= x_i'\beta_j + e_{ij} \end{aligned}$$

- $\beta_j = (\beta_{1j}, \ldots, \beta_{pj})'$ is a vector of p unknown parameters (specific to the jth time point)

  We assume that $p \leq n - t$

- $x_i = (x_{i1}, \ldots, x_{ip})'$ is a vector of p known coefficients (specific to the ith subject)

The Multivariate General Linear Model

- $e_i = (e_{i1}, \ldots, e_{it})'$ is a vector of t residual variates for the ith subject

- $e_i \sim N_t(0, \Sigma)$

- The $nt \times 1$ vector

$$e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \sim N_{nt}(0, I_n \otimes \Sigma),$$

  where I_n denotes the $n \times n$ identity matrix

- Thus, the $y_i = (y_{i1}, \ldots, y_{it})'$ vectors are independent $N_t(\mu_i, \Sigma)$ random vectors with

$$\mu_i = \begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{it} \end{pmatrix} = \begin{pmatrix} x_i'\beta_1 \\ \vdots \\ x_i'\beta_t \end{pmatrix}$$

Matrix Formulation

- Let Y denote the $n \times t$ data matrix:

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1t} \\ \multicolumn{3}{c}{\dotfill} \\ y_{n1} & \cdots & y_{nt} \end{pmatrix} = \begin{pmatrix} y_1' \\ \vdots \\ y_n' \end{pmatrix}$$

- Let X denote the $n \times p$ known design matrix:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \multicolumn{3}{c}{\dotfill} \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}$$

- Let B denote the $p \times t$ parameter matrix:

$$B = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1t} \\ \multicolumn{3}{c}{\dotfill} \\ \beta_{p1} & \cdots & \beta_{pt} \end{pmatrix} = (\beta_1, \cdots, \beta_t)$$

- Let E denote the $n \times t$ matrix of random errors:

$$E = \begin{pmatrix} e_{11} & \cdots & e_{1t} \\ \multicolumn{3}{c}{\dotfill} \\ e_{n1} & \cdots & e_{nt} \end{pmatrix} = \begin{pmatrix} e_1' \\ \vdots \\ e_n' \end{pmatrix}$$

Parameter Estimation

- The multivariate general linear model is

$$Y = XB + E,$$

where $\mathrm{E}(Y) = XB$ and

$$\mathrm{Var}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = I_n \otimes \Sigma$$

- The maximum likelihood estimators of B and Σ are

$$\widehat{B} = (X'X)^{-1}X'Y$$

$$\widehat{\Sigma} = \frac{1}{n}(Y - X\widehat{B})'(Y - X\widehat{B})$$

- An unbiased estimator of Σ is given by

$$S = \frac{1}{n-p}(Y - X\widehat{B})'(Y - X\widehat{B})$$

Estimation of Linear Functions of the Elements of B

- Let $\psi = a'Bc$, where $a_{(p\times 1)}$ and $c_{(t\times 1)}$ are vectors of constants

  a' operates within time points

  c operates between time points

- $\widehat{\psi} = a'\widehat{B}c$ has minimum variance among all linear unbiased estimates of ψ

  i.e., $\widehat{\psi}$ is a best linear unbiased estimate (BLUE)

- $\text{Var}(\widehat{\psi}) = (c'\Sigma c)[a'(X'X)^{-1}a]$

- This result is known as the multivariate Gauss-Markov theorem

Hypothesis Testing

- Consider the general hypothesis $H_0\colon ABC = D$
    - A is an $a \times p$ matrix of coefficients permitting "within time" hypotheses

      $\text{rank}(A) = a \leq p$
    - C is a $t \times c$ matrix of coefficients permitting "between time" hypotheses

      $\text{rank}(C) = c \leq t \leq (n - p)$
    - D is an $a \times c$ matrix of constants
- Let Q_h denote the hypothesis ssp matrix:

  $Q_h = (A\widehat{B}C - D)'[A(X'X)^{-1}A']^{-1}(A\widehat{B}C - D)$
- Let Q_e denote the residual ssp matrix:

  $Q_e = C'(Y'Y - \widehat{B}'(X'X)\widehat{B})C$

Test Statistics

- The likelihood ratio statistic is

$$\Lambda = \frac{|Q_e|}{|Q_h + Q_e|} = \prod_{i=1}^{\min(a,b)} \frac{1}{1+\lambda_i},$$

where λ_i are the solutions of the characteristic equation $|Q_h - \lambda Q_e| = 0$

  - This statistic is known as Wilks Λ
  - Λ has a multivariate beta null distribution

- The Pillai trace statistic is

$V = \text{trace}[Q_h(Q_h + Q_e)^{-1}] = \sum \theta_i$, where

θ_i are the solutions of the characteristic

equation $|Q_h - \theta(Q_h + Q_e)| = 0$

  - also known as the Barlett-Nanda-Pillai trace
  - It can be shown that $\theta_i = \lambda_i/(1 + \lambda_i)$

Test Statistics

- The Hotelling-Lawley trace statistic is

  $U = \mathrm{trace}[Q_h \, Q_e^{-1}] = \sum \lambda_i$

  Lawley (1938), Bartlett (1939), Hotelling (1947, 1951)

- Roy's (1957) maximum root statistic is

  $$\Theta = \frac{\lambda_1}{1 + \lambda_1},$$

  where λ_1 is the largest solution of the characteristic equation $|Q_h - \lambda Q_e| = 0$

  Equivalently, Θ is the largest solution of the characteristic equation $|Q_h - \theta(Q_h + Q_e)| = 0$

- In most cases, the exact null distributions of these four test criteria can not be computed and approximate tests are required

Theoretical Power Comparisons

- Λ, V, and U have been compared based on asymptotic expansions of their nonnull distributions

  Mikhail (1965, *Biometrika*)
  Pillai and Jayachandran (1967, *Biometrika*)
  Lee (1971, *Biometrika*)
  Rothenberg (1977)

- If the population characteristic roots are roughly equal, the ordering from most powerful to least powerful is $V > \Lambda > U$

- If the roots are unequal, the ordering is $U > \Lambda > V$

- These results support the use of Λ

Empirical Power Comparisons

- Ito (1962) compared the large-sample power properties of Λ and U for a simple class of alternative hypotheses

  there was little difference between these two statistics

- Pillai and Jayachandran (1967) compared all four statistics

  When the population characteristic roots were very different, U tended to have the highest power

  When the characteristic roots were equal, V was most powerful

  In the situations they considered, Θ was least powerful

Empirical Power Comparisons

- Roy, Gnanadesikan, and Srivastava (1971) compared all four statistics

  For equal population roots, V was most powerful, followed by Λ and U

  For the case of a single large population root, Θ had the highest empirical power

- Simulation studies by Schatzoff (1966) and Olson (1974)

  Θ was most powerful if the alternative was one-dimensional

  Θ was inferior if there were multiple non-zero characteristic roots

Robustness Comparisons

- All four test procedures tend to be relatively robust to departures from normality

- The limiting distributions of each criterion (suitably standardized) for non-normal Y are the same as when Y is normal

  (as long as conditions such as bounded fourth moments are satisfied)

- Olson (1974) studied the robustness under departures from covariance homogeneity and departures from normality

  Λ, U, and V were quite robust

  Θ was least robust

Profile Analysis

- Suppose that repeated measurements at t time points have been obtained from s groups of subjects

- Let n_h denote the number of subjects in group h, for $h = 1, \ldots, s$ $\quad (n = \sum_{h=1}^{s} n_h)$

- Let y_{hij} denote the response at time j from the ith subject in group h, for $h = 1, \ldots, s$, $i = 1, \ldots, n_h$, and $j = 1, \ldots, t$

- We assume that the data vectors

$$y_{hi} = (y_{hi1}, \ldots, y_{hit})'$$

  are independent and normally distributed with mean $\mu_h = (\mu_{h1}, \ldots, \mu_{ht})'$ and common covariance matrix Σ

$$y_{hi} \sim N_t(\mu_h, \Sigma)$$

Profile Analysis Model

- The model is $y_{hij} = \mu_{hj} + e_{hij}$
- In terms of the multivariate general linear model,

$$\begin{pmatrix} y'_{11} \\ \vdots \\ y'_{1n_1} \\ \hline y'_{21} \\ \vdots \\ y'_{2n_2} \\ \hline \vdots \\ \hline y'_{s1} \\ \vdots \\ y'_{sn_s} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu_{11} & \cdots & \mu_{1t} \\ \mu_{21} & \cdots & \mu_{2t} \\ \vdots & \vdots & \vdots \\ \mu_{s1} & \cdots & \mu_{st} \end{pmatrix} + \begin{pmatrix} e'_{11} \\ \vdots \\ e'_{1n_1} \\ \hline e'_{21} \\ \vdots \\ e'_{2n_2} \\ \hline \vdots \\ \hline e'_{s1} \\ \vdots \\ e'_{sn_s} \end{pmatrix}$$

or $Y = XB + E$, where Y and E are $n \times t$ matrices, X is $n \times s$, and B is $s \times t$

Profile Analysis Hypotheses

- Three general hypotheses are of interest

  H_{01}: the profiles for the s groups are parallel
  i.e., no group-by-time interaction

  H_{02}: no differences among groups

  H_{03}: no differences among time points

- H_{01} should be tested first, since the acceptance or rejection of this hypothesis affects how the other two hypotheses can be tested

- In addition, if H_{01} is rejected, we may wish to test hypotheses of the form

  H_{04}: no differences among groups within some subset of the total number of time points

  H_{05}: no differences among time points in a particular group (or subset of groups)

Test of Parallelism

- Recall that μ_{hj} is the mean response at time j in group h

- The parallelism hypothesis is

$$H_{01}\colon \begin{pmatrix} \mu_{11} - \mu_{12} \\ \mu_{12} - \mu_{13} \\ \vdots \\ \mu_{1,t-1} - \mu_{1t} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{s1} - \mu_{s2} \\ \mu_{s2} - \mu_{s3} \\ \vdots \\ \mu_{s,t-1} - \mu_{st} \end{pmatrix}$$

- In terms of the general $H_0\colon ABC = D$,

$$A_{(s-1)\times s} = (I_{s-1}, -1_{s-1})$$

$$C_{t\times(t-1)} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & -1 \end{pmatrix}$$

$$D_{(s-1)\times(t-1)} = 0$$

Tests for Differences Among Groups

- Depending on the results of the test of H_{01}, two tests of H_{02} are possible

- If the parallelism hypothesis is reasonable, the test for differences among groups can be carried out using the sum (or average) of the repeated observations from each subject

- In this case:

$$A_{(s-1)\times s} = (I_{s-1}, -1_{s-1})$$

$$C_{t\times 1} = 1_t$$

$$D_{(s-1)\times 1} = 0_{s-1}$$

- This test of H_{02} is equivalent to that from a one-way ANOVA on the totals (or means) across time from each subject

Tests for Differences Among Groups

- A multivariate test for differences among groups can also be carried out without assuming parallelism:

$$H_{02}\colon \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1t} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2t} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{s1} \\ \mu_{s2} \\ \vdots \\ \mu_{st} \end{pmatrix}$$

- In this case:

$$A_{(s-1)\times s} = (I_{s-1}, -1_{s-1})$$

$$C_{t\times t} = I_t$$

$$D_{(s-1)\times t} = 0$$

- If comparisons among groups for a subset of the t time points are of interest, the columns of C corresponding to the excluded time points can be omitted

Tests for Differences Among Time Points

- Depending on the results of the test of H_{01}, two tests of H_{03} are possible

- If the parallelism hypothesis is reasonable, the test for differences among time points can be carried out using the sum (or average) across groups of the observations at each time point

- In this case:

$$A_{1\times s} = (1, \ldots, 1) \;\text{ or }\; (1/s, \ldots, 1/s)$$

$$C_{t\times(t-1)} = \begin{pmatrix} I_{t-1} \\ -1'_{t-1} \end{pmatrix}$$

$$D_{1\times(t-1)} = 0$$

- This is equivalent to a one-sample T^2 test

Tests for Differences Among Time Points

- The preceding procedure weights each of the s groups equally and is usually appropriate

- However, if unequal group sizes result from the nature of the experimental conditions, it may be desirable to use a weighted average rather than a simple average

- In this case, $A = (n_1, \ldots, n_s)$ or

$$A = \left(\frac{n_1}{n}, \ldots, \frac{n_s}{n} \right)$$

  can be used

- Note that C and D are unchanged

Tests for Differences Among Time Points

- H_{03} can also be tested without assuming parallelism:

$$H_{03}\colon \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{s1} \end{pmatrix} = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{s2} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{st} \end{pmatrix}$$

- In this case:

$$A_{s\times s} = I_s$$

$$C_{t\times(t-1)} = \begin{pmatrix} I_{t-1} \\ -1'_{t-1} \end{pmatrix}$$

$$D_{s\times(t-1)} = 0$$

- If comparisons among time points in a particular group (or subset of groups) are of interest, the rows of A corresponding to the excluded groups can be omitted

Example

- At ages 8, 10, 12, and 14, the distance (mm) from the pituitary to the pteryomaxillary fissure was measured in 16 boys and 11 girls

- Let $\mu_b = (\mu_{b,8},\ \mu_{b,10},\ \mu_{b,12},\ \mu_{b,14})'$ and $\mu_g = (\mu_{g,8},\ \mu_{g,10},\ \mu_{g,12},\ \mu_{g,14})'$

- The profile analysis model is:

$$\begin{pmatrix} y'_{b,1} \\ \vdots \\ y'_{b,16} \\ y'_{g,1} \\ \vdots \\ y'_{g,11} \end{pmatrix} = \begin{pmatrix} 1\ 0 \\ \vdots \\ 1\ 0 \\ 0\ 1 \\ \vdots \\ 0\ 1 \end{pmatrix} \begin{pmatrix} \mu_{b,8} & \cdots & \mu_{b,14} \\ \mu_{g,8} & \cdots & \mu_{g,14} \end{pmatrix} + \begin{pmatrix} e'_{b,1} \\ \vdots \\ e'_{b,16} \\ e'_{g,1} \\ \vdots \\ e'_{g,11} \end{pmatrix}$$

or $Y = XB + E$, where Y and E are 27×4 matrices, X is 27×2, and B is 2×4

SAS Statements

```
data a;
input sex id d8 d10 d12 d14;
male=(sex=1);
female=(sex=2);
cards;
1  1 26.0 25.0 29.0 31.0
1  2 21.5 22.5 23.0 26.5
1  3 23.0 22.5 24.0 27.5
           . . .
2  9 20.0 21.0 22.0 21.5
2 10 16.5 19.0 19.0 19.5
2 11 24.5 25.0 28.0 28.0
;
```

- The derived variables `male` and `female` will be used to define the design matrix X

$$\texttt{male} = \begin{cases} 1 & \text{for boys} \\ 0 & \text{for girls} \end{cases}$$

$$\texttt{female} = \begin{cases} 1 & \text{for girls} \\ 0 & \text{for boys} \end{cases}$$

SAS Statements (Continued)

- Fit the profile analysis model:

```
proc glm;
model d8 d10 d12 d14=male female
      / noint nouni;
```

- Note that this model does not include an intercept term

- Test the parallelism hypothesis:

```
contrast 'Parallelism'
         male 1 female -1;
manova m=(1 -1  0  0,
          0  1 -1  0,
          0  0  1 -1);
```

- $A_{(1\times 2)}$ is specified using the `contrast` statement

- The transpose of $C'_{(4\times 3)}$ is specified using the `manova` statement

- D is assumed to be a matrix (or vector) of zeros

SAS Statements (Continued)

- Test for differences between boys and girls (assuming parallelism)

```
contrast 'Sex (if Parallel)'
         male 1 female -1;
manova m=(1 1 1 1);
```

- Test for differences between boys and girls (without assuming parallelism)

```
contrast 'Sex (Not Parallel)'
         male 1 female -1;
manova m=(1 0 0 0,
          0 1 0 0,
          0 0 1 0,
          0 0 0 1);
```

- (The `manova` statement is not necessary, since the default C is the identity matrix)

SAS Statements (Continued)

- Test for differences among time points (assuming parallelism and using equal weights)

```
contrast 'Time (Parallel)'
          male 0.5 female 0.5;
manova m=(1  0  0 -1,
           0  1  0 -1,
           0  0  1 -1);
```

- Test for differences among time points (assuming parallelism and using weights proportional to sample size)

```
contrast 'Time (Par., Weights)'
          male .59259 female .40741;
manova m=(1  0  0 -1,
           0  1  0 -1,
           0  0  1 -1);
```

SAS Statements (Continued)

- Test for differences among time points (without assuming parallelism)

```
contrast 'Time (Not Parallel)'
         male 1, female 1;
manova m=(1  0  0 -1,
          0  1  0 -1,
          0  0  1 -1);
```

- Test for differences among time points in boys

```
contrast 'Time (M, Not Parall)' male 1;
manova m=(1  0  0 -1,
          0  1  0 -1,
          0  0  1 -1);
```

- Test for differences among time points in girls

```
contrast 'Time (F, Not Parall)' female 1;
manova m=(1  0  0 -1,
          0  1  0 -1,
          0  0  1 -1);
```

Growth Curve Analysis

- The MANOVA approach does not require that a subject's repeated measurements are ordered
- In fact, repeated measurements obtained over time are naturally ordered
- In this case, it may be of interest to characterize trends over time using low-order polynomials
- The means at the repeated time points can then be summarized by a few coefficients, rather than by the entire vector
- When the number of responses is large, reduction to a linear or quadratic function is very useful
- Focus shifts from hypothesis testing to estimation of a substantive model for the responses

Growth Curve Analysis

- An extension of the standard MANOVA model
- Initially proposed by Potthoff and Roy (1964)
- An alternative formulation was developed by Rao (1965, 1966, 1967) and Khatri (1966)
- Grizzle and Allen (1969) unified and illustrated the methodology
- Kleinbaum (1973) generalized the model to allow missing data
- A relatively unused approach, due to:
  - unfamiliarity with the methodology
  - lack of readily available software

Potthoff-Roy Model

- $Y = XBT + E$, where
  - Y is the $n \times t$ data matrix

    y_{ij} is the response from subject i at time j
  - X is a $n \times s$ across-individual design matrix
  - B is a $s \times q$ parameter matrix
  - T is a $q \times t$ within-individual design matrix

    $\text{rank}(T) = q$, where $q \leq t$
  - E is the $n \times t$ matrix of random errors
- Each row $y_i' = (y_{i1}, \ldots, y_{it})$ of the data matrix Y has an independent multivariate normal distribution with covariance matrix Σ
- $\text{E}(Y) = XBT$ and $\text{Var}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = I_n \otimes \Sigma$

Distinction Between Profile Analysis and Growth Curve Analysis

- Suppose repeated measurements are obtained at times $j = 1, \ldots, t$ from s groups of subjects

- Let y_{hij} denote the response at time j from the ith subject in group h, for $h = 1, \ldots, s$, $i = 1, \ldots, n_h$, and $j = 1, \ldots, t$

- The profile analysis model is $y_{hij} = \mu_{hj} + e_{hij}$

- If the time trend in each group can be described by a $(q-1)$st degree polynomial $(q \leq t)$, the growth curve model is

  $$y_{hij} = \beta_{h0} + \beta_{h1}\, j + \beta_{h2}\, j^2 + \cdots + \beta_{h,q-1}\, j^{q-1} + e_{hij}$$

  (Although the functional form of the time trend is the same in each group, the parameters vary)

Distinction Between Profile Analysis and Growth Curve Analysis

- The profile analysis model is $Y = XB + E$
  - Y and E are $n \times t$ matrices
  - X is a $n \times s$ matrix of zeros and ones
  - B is a $s \times t$ matrix with (h, j)th element μ_{hj}

$$\begin{pmatrix} y'_{11} \\ \vdots \\ y'_{1n_1} \\ \hline y'_{21} \\ \vdots \\ y'_{2n_2} \\ \hline \vdots \\ \hline y'_{s1} \\ \vdots \\ y'_{sn_s} \end{pmatrix} = \begin{pmatrix} 1 \; 0 \cdots 0 \\ \vdots \; \vdots \; \cdots \; \vdots \\ 1 \; 0 \cdots 0 \\ \hline 0 \; 1 \cdots 0 \\ \vdots \; \vdots \; \cdots \; \vdots \\ 0 \; 1 \cdots 0 \\ \hline \vdots \\ \hline 0 \; 0 \cdots 1 \\ \vdots \; \vdots \; \cdots \; \vdots \\ 0 \; 0 \cdots 1 \end{pmatrix} \begin{pmatrix} \mu_{11} & \cdots & \mu_{1t} \\ \mu_{21} & \cdots & \mu_{2t} \\ \vdots & \vdots & \vdots \\ \mu_{s1} & \cdots & \mu_{st} \end{pmatrix} + \begin{pmatrix} e'_{11} \\ \vdots \\ e'_{1n_1} \\ \hline e'_{21} \\ \vdots \\ e'_{2n_2} \\ \hline \vdots \\ \hline e'_{s1} \\ \vdots \\ e'_{sn_s} \end{pmatrix}$$

Distinction Between Profile Analysis and Growth Curve Analysis

- The growth curve model is $Y = XBT + E$
  - Y and E are $n \times t$ matrices
  - X is a $n \times s$ matrix of zeros and ones
  - B is a $s \times q$ matrix
  - T is a $q \times t$ matrix

- Thus, the expected value of Y is equal to

$$\left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline & & \vdots & \\ \hline 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{array}\right) \begin{pmatrix} \beta_{10} & \cdots & \beta_{1,q-1} \\ \beta_{20} & \cdots & \beta_{2,q-1} \\ \vdots & \vdots & \vdots \\ \beta_{s0} & \cdots & \beta_{s,q-1} \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & t \\ 1 & 4 & \cdots & t^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2^{q-1} & \ldots & t^{q-1} \end{pmatrix}$$

The Potthoff-Roy Approach

- The basic idea is to transform the growth curve model to the usual MANOVA model

- Let G be a $t \times t$ symmetric, positive-definite matrix satisfying the following conditions:
  - G must be nonstochastic *or* independent of Y
  - $TG^{-1}T'$ has rank q

- If both sides of the model $Y = XBT + E$ are post-multiplied by $G^{-1}T'(TG^{-1}T')^{-1}$,

$$\begin{aligned} YG^{-1}T'(TG^{-1}T')^{-1} &= XBTG^{-1}T'(TG^{-1}T')^{-1} \\ &\quad + EG^{-1}T'(TG^{-1}T')^{-1}, \end{aligned}$$

  or $Z = XB + E^*$, where

$$Z = YG^{-1}T'(TG^{-1}T')^{-1}$$

  is a matrix of transformed dependent variables

The Potthoff-Roy Approach

- The transformed data matrix Z has mean XB
- The rows of Z have independent $N_q(0, \Sigma^*)$ distributions, where

$$\Sigma^* = (TG^{-1}T')^{-1}TG^{-1}\Sigma G^{-1}T'(TG^{-1}T')^{-1}$$

- The growth curve model has thus been reduced to the profile analysis model
- Standard multivariate linear model theory can be used to:
  - estimate B
  - test hypotheses of the form $ABC = D$
- In particular, the linear unbiased estimator of B is

$$\widehat{B} = (X'X)^{-1}X'YG^{-1}T'(TG^{-1}T')^{-1}$$

Choice of G

- Potthoff and Roy (1964) proved that the minimum variance unbiased estimator of B is

$$\widehat{B} = (X'X)^{-1}X'Y\Sigma^{-1}T'(T\Sigma^{-1}T')^{-1}$$

- Therefore, although $\widehat{B}$ is unbiased for any G, the optimal choice is $G = \Sigma$

- In practice, Σ is usually unknown

- Potthoff and Roy (1964) suggested using an estimate of Σ obtained from an independent experiment

- They did not, however, develop the theory for allowing $G = S$, where S is the sample covariance matrix calculated from the data used to estimate B

Choice of G

- The problem is simplified when $q = t$

  i.e., when the time trend across the t points is described by a $(t-1)$st degree polynomial

- In this case,

$$\begin{aligned} Z &= YG^{-1}T'(TG^{-1}T')^{-1} \\ &= YG^{-1}T'(T')^{-1}GT^{-1} \\ &= YT^{-1}, \end{aligned}$$

  so that there is no need to choose G

- If T is an orthogonal matrix, then $Z = YT'$ and matrix inversion is not required

- Bock (1963) developed this procedure using Roy-Bargmann (1958) step-down F-tests and orthogonal polynomials

Choice of G

- When $q < t$, the simplest choice is $G = I_t$
- In this case,

$$\begin{aligned} Z &= YG^{-1}T'(TG^{-1}T')^{-1} \\ &= YT'(TT')^{-1} \end{aligned}$$

- If the time trends are parameterized using orthogonal polynomial coefficients, the transformation further simplifies to $Z = YT'$
- This simplifies the calculations and eliminates the need for matrix inversion
- However, it may not be the best choice in terms of power
- Information is lost in reducing Y to Z unless $G = \Sigma$ (or unless $\Sigma = \sigma^2 I$)

Rao-Khatri Approach

- In order to avoid the arbitrary choice of G, Khatri (1966) derived the maximum likelihood estimator of B
- Rao (1965, 1966, 1967) considered the conditional model $\mathrm{E}(Y|W) = XB + W\Gamma$ and derived a covariate-adjusted estimator of B
- If $q < t$, identical results are obtained from:
  - Khatri's maximum likelihood approach
  - Rao's covariate-adjusted approach using $t - q$ covariates
  - Potthoff and Roy's approach using $G = S$
- When $q < t$, the Potthoff-Roy approach using $G = I$ is equivalent to not using covariates in Rao's conditional model

Example

- In a dental study, the height of the ramus bone (mm) was measured in 20 boys at ages 8, $8\frac{1}{2}$, 9, and $9\frac{1}{2}$ years

- Three questions:
  - Does bone height change with age?

    Not of great interest, since answer is obvious

  - Is there a linear relationship between age and bone height?

  - What is the model for predicting bone height from age?

Reference

Elston, R. C. and Grizzle, J. E. (1962). Estimation of time-response curves and their confidence bands. *Biometrics* **18**, 148–159.

Data from Example

	Age (years)			
Subject	8	$8\frac{1}{2}$	9	$9\frac{1}{2}$
1	47.8	48.8	49.0	49.7
2	46.4	47.3	47.7	48.4
3	46.3	46.8	47.8	48.5
4	45.1	45.3	46.1	47.2
5	47.6	48.5	48.9	49.3
6	52.5	53.2	53.3	53.7
7	51.2	53.0	54.3	54.5
8	49.8	50.0	50.3	52.7
9	48.1	50.8	52.3	54.4
10	45.0	47.0	47.3	49.3
11	51.2	51.4	51.6	51.9
12	48.5	49.2	53.0	55.5
13	52.1	52.8	53.7	55.0
14	48.2	48.9	49.3	49.8
15	49.6	50.4	51.2	51.8
16	50.7	51.7	52.7	53.3
17	47.2	47.7	48.4	49.5
18	53.3	54.6	55.1	55.3
19	46.2	47.5	48.1	48.4
20	46.3	47.6	51.3	51.8

Application to Example

- In this example, $n = 20$, $t = 4$, $s = 1$

- Since there is a single group of subjects, X is the 20×1 matrix $(1, \ldots, 1)'$

- We will first choose $q = t = 4$

- If T is the 4×4 matrix of orthogonal polynomial coefficients,

$$Z = YG^{-1}T'(TG^{-1}T')^{-1} = YT^{-1} = YT'$$

- Thus, it is not necessary to choose G and matrix inversion is not required

- We will use this model to test if the nonlinear components of the time effect are statistically significant

SAS Statements

(q=4, Standardized Orth. Poly. Coefficients)

```
data a;
input subject h80 h85 h90 h95;
* standardized orth. poly. coefficients;
sop0=(   h80 + h85 + h90 + h95)/2;
sop1=(-3*h80-  h85+  h90+3*h95)/sqrt(20);
sop2=(   h80-  h85-  h90+  h95)/2;
sop3=(  -h80+3*h85-3*h90+  h95)/sqrt(20);
cards;
 1 47.8 48.8 49.0 49.7
 2 46.4 47.3 47.7 48.4
         . . .
19 46.2 47.5 48.1 48.4
20 46.3 47.6 51.3 51.8
;
proc glm;
model sop0-sop3= / nouni;
manova h=intercept m=(1 0 0 0);
manova h=intercept m=(0 1 0 0);
manova h=intercept m=(0 0 1 0);
manova h=intercept m=(0 0 0 1);
manova h=intercept m=(0 0 1 0,
                      0 0 0 1);
```

Comments

- The constant and linear age effects are highly significant

- The quadratic and cubic effects of age are nonsignificant, both individually and jointly

- We will now model the effects of age on ramus height using a linear growth curve model ($q = 2$)

- Computations are simpler using orthogonal polynomial coefficients

- Interpretation is simpler using the matrix

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8.0 & 8.5 & 9.0 & 9.5 \end{pmatrix}$$

- We will first use $G = I_4$ and then consider $G = S$ (the sample covariance matrix)

Linear Model, $G = I_4$

- $Z = YG^{-1}T'(TG^{-1}T')^{-1} = YT'(TT')^{-1}$

- The transformation is computed as follows:

$$TT' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8.0 & 8.5 & 9.0 & 9.5 \end{pmatrix} \begin{pmatrix} 1 & 8.0 \\ 1 & 8.5 \\ 1 & 9.0 \\ 1 & 9.5 \end{pmatrix} = \begin{pmatrix} 4 & 35 \\ 35 & 307.5 \end{pmatrix}$$

$$(TT')^{-1} = \begin{pmatrix} 61.5 & -7 \\ -7 & 0.8 \end{pmatrix}$$

$$T'(TT')^{-1} = \begin{pmatrix} 5.5 & -0.6 \\ 2.0 & -0.2 \\ -1.5 & 0.2 \\ -5.0 & 0.6 \end{pmatrix}$$

- The SAS statements are:

```
data b; set a;
pi0=5.5*h80+2.0*h85-1.5*h90-5.0*h95;
pi1=-.6*h80-0.2*h85+0.2*h90+0.6*h95;
proc glm;
model pi0 pi1=;
```

Linear Model, $G = S$

- In a one-sample problem, the sample covariance matrix S can be computed using PROC CORR:

```
proc corr nosimple cov; var h80 h85 h90 h95;
```

- In this example,

$$S = \begin{pmatrix} 6.32997 & 6.18908 & 5.77700 & 5.35579 \\ 6.18908 & 6.44934 & 6.15342 & 5.78526 \\ 5.77700 & 6.15342 & 6.91800 & 6.77421 \\ 5.35579 & 5.78526 & 6.77421 & 7.18316 \end{pmatrix}$$

- For the case in which $G = S$, the transformation $Z = YG^{-1}T'(TG^{-1}T')^{-1}$ is computed as follows:

$$G^{-1} = \begin{pmatrix} 2.6933 & -2.8416 & 0.0498 & 0.2334 \\ -2.8416 & 4.1461 & -1.5651 & 0.2555 \\ 0.0498 & -1.5651 & 3.8824 & -2.4379 \\ 0.2334 & 0.2555 & -2.4379 & 2.0585 \end{pmatrix}$$

Linear Model, $G = S$

$$G^{-1}T' = \begin{pmatrix} 0.13501104 & 0.05931779 \\ -0.00512515 & 0.85011176 \\ -0.07088109 & -1.12416465 \\ 0.10952328 & 1.65380106 \end{pmatrix}$$

$$TG^{-1}T' = \begin{pmatrix} 0.16852809 & 1.43906596 \\ 1.43906596 & 13.29412054 \end{pmatrix}$$

$$(TG^{-1}T')^{-1} = \begin{pmatrix} 78.42126986 & -8.48896924 \\ -8.48896924 & 0.99413772 \end{pmatrix}$$

$$G^{-1}T'(TG^{-1}T')^{-1} = \begin{pmatrix} 10.084191 & -1.087135 \\ -7.618493 & 0.888635 \\ 3.984414 & -0.515867 \\ -5.450112 & 0.714366 \end{pmatrix}$$

- The SAS statements are:

```
data b; set a;
ps0=10.08419058*h80-7.61849306*h85
    +3.98441438*h90-5.45011190*h95;
ps1=-1.08713454*h80+0.88863538*h85
    -0.51586713*h90+0.71436629*h95;
proc glm;
model ps0 ps1=;
```

Example

- A study conducted in 16 boys and 11 girls
- At ages 8, 10, 12, and 14, the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure was measured
- The change in the pituitary-pteryomaxillary distance during growth is important in orthodontal therapy
- The goals are to:
  - Describe the distance in boys and girls as simple functions of age
  - Compare the functions for boys and girls

Reference

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

Dental Measurements

Group	ID	Age 8	Age 10	Age 12	Age 14
Boys	1	26.0	25.0	29.0	31.0
	2	21.5	22.5	23.0	26.5
	3	23.0	22.5	24.0	27.5
	4	25.5	27.5	26.5	27.0
	5	20.0	23.5	22.5	26.0
	6	24.5	25.5	27.0	28.5
	7	22.0	22.0	24.5	26.5
	8	24.0	21.5	24.5	25.5
	9	23.0	20.5	31.0	26.0
	10	27.5	28.0	31.0	31.5
	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
	Mean	22.9	23.8	25.7	27.5
Girls	1	21.0	20.0	21.5	23.0
	2	21.0	21.5	24.0	25.5
	3	20.5	24.0	24.5	26.0
	4	23.5	24.5	25.0	26.5
	5	21.5	23.0	22.5	23.5
	6	20.0	21.0	21.0	22.5
	7	21.5	22.5	23.0	25.0
	8	23.0	23.0	23.5	24.0
	9	20.0	21.0	22.0	21.5
	10	16.5	19.0	19.0	19.5
	11	24.5	25.0	28.0	28.0
	Mean	21.2	22.2	23.1	24.1

Outline of Analyses

1. Fit growth curve model with $q = t = 4$ using standardized orthogonal polynomial coefficients
   - matrix inversion and/or computation of the pooled covariance matrix S not required
   - test joint significance of constant, linear, quadratic, and cubic terms to determine degree of polynomial

2. Fit reduced covariate-adjusted model using standardized orthogonal polynomial coefficients
   - test equality of parameters for boys and girls
   - compare with Potthoff-Roy estimated parameters when $G = S$

3. Fit Potthoff-Roy reduced polynomial model with T defined on the natural time scale

SAS Statements

(q=4, Standardized Orth. Poly. Coefficients)

```
data a;
input sex id d8 d10 d12 d14;
male=(sex=1);
female=(sex=2);
* standardized orth. poly. coefficients;
sop0=(   d8 + d10 + d12 + d14)/2;
sop1=(-3*d8-  d10+  d12+3*d14)/sqrt(20);
sop2=(   d8-  d10-  d12+  d14)/2;
sop3=(  -d8+3*d10-3*d12+  d14)/sqrt(20);
cards;
1  1 26.0 25.0 29.0 31.0
         . . .
2 11 24.5 25.0 28.0 28.0
 ;
proc glm;
model sop0-sop3=male female / noint nouni;
contrast 'Both Sexes' male 1, female 1;
manova m=(1 0 0 0);
manova m=(0 1 0 0);
manova m=(0 0 1 0);
manova m=(0 0 0 1);
manova m=(0 0 1 0,
          0 0 0 1);
```

Covariate-Adjusted Linear Model

- Since nonlinear effects are nonsignificant, quadratic and cubic effects will be used as covariates

- The SAS statements are:

```
proc glm;
model sop0 sop1=male female sop2 sop3
      / noint;
contrast 'Both Sexes' male 1, female 1;
manova m=(1 0,
          0 1);
proc glm;
model sop0 sop1=male female sop2 sop3
      / noint;
contrast 'Sex' male 1 female -1;
manova m=(1 0,
          0 1);
```

- The first model tests joint effects in boys and girls, while the second model tests equality of effects for boys and girls

Potthoff-Roy Linear Model with $G = S$

- In a multi-sample problem, PROC DISCRIM can be used to compute the pooled sample covariance matrix S

```
proc discrim pcov; class sex;
var d8 d10 d12 d14;
```

- In this example,

$$G = S = \begin{pmatrix} 5.41545 & 2.71682 & 3.91023 & 2.71023 \\ 2.71682 & 4.18477 & 2.92716 & 3.31716 \\ 3.91023 & 2.92716 & 6.45574 & 4.13074 \\ 2.71023 & 3.31716 & 4.13074 & 4.98574 \end{pmatrix}$$

- The transformation $Z = YG^{-1}T'(TG^{-1}T')^{-1}$ is computed as follows:

$$T = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{20} & -1/\sqrt{20} & 1/\sqrt{20} & 3/\sqrt{20} \end{pmatrix}$$

Potthoff-Roy Linear Model with $G = S$

$$G^{-1} = \begin{pmatrix} 0.37168 & -0.15407 & -0.19490 & 0.06194 \\ -0.15407 & 0.57220 & 0.05082 & -0.33905 \\ -0.19490 & 0.05082 & 0.43363 & -0.28713 \\ 0.06194 & -0.33905 & -0.28713 & 0.63038 \end{pmatrix}$$

$$G^{-1}T' = \begin{pmatrix} 0.04232484 & -0.21690806 \\ 0.06494694 & -0.24067261 \\ 0.00120890 & 0.02372727 \\ 0.03306573 & 0.39292648 \end{pmatrix}$$

$$TG^{-1}T' = \begin{pmatrix} 0.07077320 & -0.02046346 \\ -0.02046346 & 0.46821105 \end{pmatrix}$$

$$(TG^{-1}T')^{-1} = \begin{pmatrix} 14.31048448 & 0.62544880 \\ 0.62544880 & 2.16312460 \end{pmatrix}$$

$$G^{-1}T'(TG^{-1}T')^{-1} = \begin{pmatrix} 0.4700241 & -0.4427271 \\ 0.7788937 & -0.4799838 \\ 0.0321401 & 0.0520811 \\ 0.7189420 & 0.8706298 \end{pmatrix}$$

Potthoff-Roy Linear Model with $G = S$

- The SAS statements are:

```
data b; set a;
sops0= 0.47002414*d8+0.77889373*d10
      +0.03214013*d12+0.71894200*d14;
sops1=-0.44272714*d8-0.47998385*d10
      +0.05208114*d12+0.87062985*d14;
proc glm;
model sops0 sops1=male female / noint;
```

- The estimated constant and linear age parameters for boys and girls are identical to those from the covariate-adjusted model

- Differences between the two models:
  - Standard errors of estimated parameters
  - Test statistics and degrees of freedom for hypothesis tests

Potthoff-Roy Linear Model

(Natural Time Scale, $G = S$)

- For ease of interpretation, the linear model will now be fit using the matrix

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix}$$

$$G^{-1}T' = \begin{pmatrix} 0.08464969 & -0.03889577 \\ 0.12989387 & 0.35251199 \\ 0.00241780 & 0.13270736 \\ 0.06613146 & 2.48466667 \end{pmatrix}$$

$$TG^{-1}T' = \begin{pmatrix} 0.28309282 & 2.93099025 \\ 2.93099025 & 39.59177541 \end{pmatrix}$$

$$(TG^{-1}T')^{-1} = \begin{pmatrix} 15.12612431 & -1.11979123 \\ -1.11979123 & 0.10815623 \end{pmatrix}$$

$$G^{-1}T'(TG^{-1}T')^{-1} = \begin{pmatrix} 1.323977 & -0.098997 \\ 1.570051 & -0.107328 \\ -0.112033 & 0.011646 \\ -1.781995 & 0.194679 \end{pmatrix}$$

Potthoff-Roy Linear Model

(Natural Time Scale, $G = S$)

- The SAS statements are:

```
data b; set a;
ps0=1.32397685*d8+1.57005103*d10
    -.11203261*d12-1.78199527*d14;
ps1=-.09899680*d8-0.10732765*d10
    +.01164570*d12+0.19467875*d14;
proc glm;
model ps0 ps1=male female / noint;
contrast 'Both Sexes' male 1, female 1;
manova m=(1 0,
          0 1);
proc glm;
model ps0 ps1=male female / noint nouni;
contrast 'Sex' male 1 female -1;
manova m=(1 0);
manova m=(0 1);
manova m=(1 0,
          0 1);
```

Potthoff-Roy Linear Model (Natural Time Scale, $G = S$)

- The resulting model is:

	Boys		Girls	
	Estimate	S.E.	Estimate	S.E.
Constant	15.842	0.972	17.425	1.173
Linear Age	0.827	0.082	0.476	0.099

- The slopes for boys and girls are significantly different ($p = 0.01$)

- The intercepts for boys and girls are not significantly different ($p = 0.3$)

- All hypothesis tests involving slopes, as well as the joint tests of intercepts and slopes, are identical to those from the orthogonal polynomial parameterization

Potthoff-Roy Linear Model

(Natural Time Scale, $G = I$)

- $Z = YG^{-1}T'(TG^{-1}T')^{-1} = YT'(TT')^{-1}$, where

$$TT' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix} \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} = \begin{pmatrix} 4 & 44 \\ 44 & 504 \end{pmatrix}$$

$$(TT')^{-1} = \begin{pmatrix} 6.30 & -0.55 \\ -0.55 & 0.05 \end{pmatrix}$$

$$T'(TT')^{-1} = \begin{pmatrix} 1.9 & -0.15 \\ 0.8 & -0.05 \\ -0.3 & 0.05 \\ -1.4 & 0.15 \end{pmatrix}$$

- The resulting model is:

	Boys		Girls	
	Estimate	S.E.	Estimate	S.E.
Constant	16.341	1.019	17.373	1.228
Linear Age	0.784	0.086	0.480	0.104