Normal-Theory Methods

- Univariate methods reduce the repeated measures from each subject to a single number
  - This loss of information may not be desirable
- We now consider alternative methods for normally-distributed responses y_{ij}
  - These utilize the multivariate nature of a subject's observations
- The following methods will be studied:
  - Unstructured multivariate approach
  - Multivariate analysis of variance
    - Profile analysis
    - Growth curve analysis
  - Repeated measures ANOVA
  - Mixed linear models

The Multivariate Normal Distribution

- Let $x = (x_1, \ldots, x_p)'$ be a p-component random vector having a multivariate normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_p)'$ and $p \times p$ covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

- This can be written as $x \sim N_p(\mu, \Sigma)$

- Now consider a sample of n such vectors, $x_1 = (x_{11}, \ldots, x_{1p})'$, $\ldots$, $x_n = (x_{n1}, \ldots, x_{np})'$

- These data can be summarized in the $n \times p$ data matrix

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

Parameter Estimation

- The maximum likelihood estimate of μ is $\widehat{\mu} = \overline{x} = (\overline{x}_1, \ldots, \overline{x}_p)'$, where $\overline{x}_j = \sum_{i=1}^n x_{ij}/n$

- The maximum likelihood estimate of Σ is

$$\widehat{\Sigma} = \frac{1}{n} A,$$

where A is a $p \times p$ matrix with elements $a_{jk} = \sum_{i=1}^n (x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)$

- In matrix notation,

$$A = \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})' = \sum_{i=1}^{n} x_i x_i' - n\overline{x}\,\overline{x}'$$

- An unbiased estimator of Σ is given by

$$S = \frac{1}{n-1} A$$

The Wishart Distribution

- Let $z_1, \ldots, z_n$ be independent random vectors, with $z_i \sim N_p(0, \Sigma)$

- Let $A = \sum_{i=1}^n z_i z_i'$ (a $p \times p$ matrix)

- A has the (central) Wishart distribution with parameters n and Σ

$$A \sim W_p(n, \Sigma)$$

- The density of A is given by

$$\frac{|A|^{(n-p-1)/2} \exp\bigl(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}A)\bigr)}{2^{np/2}\, \pi^{p(p-1)/4}\, |\Sigma|^{n/2} \prod_{i=1}^p \Gamma\bigl((n+1-i)/2\bigr)}$$

for A positive definite and 0 otherwise, where $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$

- Note that A does not have a density if $n < p$

Wishart Matrices

- Let $x_1, \ldots, x_n$ be independent $N_p(\mu, \Sigma)$ random variables

- The sample covariance matrix is given by

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})'$$

- The Wishart matrix $A = (n-1)S \sim W_p(n-1, \Sigma)$

- Two important properties of Wishart matrices:

  1. The sample mean vector $\overline{x}$ and the Wishart matrix A computed from the same sample are independent

  2. If $A_1, \ldots, A_s$ are independent Wishart matrices with $A_h \sim W_p(n_h, \Sigma)$, then $\sum_{h=1}^{s} A_h \sim W_p(n, \Sigma)$, where $n = \sum_{h=1}^{s} n_h$

The Generalized T^2 Statistic

- Let $x \sim N_p(\mu, \Sigma)$
- Let nW be a $p \times p$ matrix, independent of x, such that $nW \sim W_p(n, \Sigma)$
- Then $T^2 = x'W^{-1}x$ has the generalized T^2 distribution with noncentrality parameter $\delta = \mu'\Sigma^{-1}\mu$ and degrees of freedom p and n

  $$T^2 \sim T^2_{p,n,\delta}$$

- The distribution of T^2 is related to that of the ratio of independent χ^2 random variables:

  $$F = \frac{n-p+1}{np} T^2$$

  has the $F_{p,n-p+1,\delta}$ distribution

- If $\mu = 0$, $F \sim F_{p,n-p+1}$

One-Sample Test of $H_0\colon \mu = \mu_0$

- Let $x_1, \ldots, x_n$ be a random sample from $N_p(\mu, \Sigma)$

- Suppose we wish to test $H_0\colon \mu = \mu_0$

- We will use the following results:

  1. $\sqrt{n}(\overline{x} - \mu_0) \sim N_p(\sqrt{n}(\mu - \mu_0), \Sigma)$

  2. The sample covariance matrix S is independent of $\overline{x}$

  3. $(n-1)S \sim W_p(n-1, \Sigma)$

- In this case, the generalized T^2 statistic is

$$\begin{aligned} T^2 &= \big(\sqrt{n}(\overline{x} - \mu_0)\big)' S^{-1} \big(\sqrt{n}(\overline{x} - \mu_0)\big) \\ &= n(\overline{x} - \mu_0)' S^{-1} (\overline{x} - \mu_0) \end{aligned}$$

One-Sample Test of $H_0\colon \mu = \mu_0$

- The statistic

$$F = \frac{(n-1)-p+1}{(n-1)p}\,T^2 = \frac{n-p}{(n-1)p}\,T^2$$

has the $F_{p,n-p,\delta}$ distribution, where

$$\delta = n(\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)$$

- If H_0 is true, $F \sim F_{p,n-p}$
- This test can only be used when $n > p$
- T^2 can also be derived as the likelihood ratio test of H_0
- The null distribution of T^2 is approximately valid even if the distribution of $x_1, \ldots, x_n$ is not normal (Anderson, 1984, p. 163)

One-Sample Test of $H_0\colon C\mu = 0$

- Let $x_1, \ldots, x_n$ be a random sample from $N_p(\mu, \Sigma)$
- Suppose we wish to test $H_0\colon C\mu = 0$, where C is a $c \times p$ matrix of rank c ($c \leq p$)
- Let $z_i = Cx_i$, for $i = 1, \ldots, n$
- $z_1, \ldots, z_n$ are independent random vectors from the $N_c(C\mu, C\Sigma C')$ distribution
- $\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} Cx_i = C\overline{x}$
- $\overline{z} \sim N_c(C\mu, \frac{1}{n}C\Sigma C')$
- $\sqrt{n}\overline{z} \sim N_c(\sqrt{n}C\mu, C\Sigma C')$

One-Sample Test of $H_0\colon C\mu = 0$

- The sample covariance matrix of $z_1, \ldots, z_n$ is given by

$$
\begin{aligned}
S_z &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \overline{z})(z_i - \overline{z})' \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(Cx_i - C\overline{x})(Cx_i - C\overline{x})' \\
&= \frac{1}{n-1}\sum_{i=1}^{n}C(x_i - \overline{x})[C(x_i - \overline{x})]' \\
&= \frac{1}{n-1}\sum_{i=1}^{n}C(x_i - \overline{x})(x_i - \overline{x})'C' \\
&= CSC'
\end{aligned}
$$

- $S_z = CSC'$ is independent of $\overline{z}$

- $(n-1)S_z = (n-1)CSC' \sim W_c(n-1, C\Sigma C')$

One-Sample Test of $H_0\colon C\mu = 0$

- Therefore,

$$T^2 = (\sqrt{n}\overline{z})'\, S_z^{-1}\, (\sqrt{n}\overline{z})$$
$$= n(C\overline{x})'(CSC')^{-1}(C\overline{x})$$

has the $T^2_{c,n-1,\delta}$ distribution with noncentrality parameter

$$\delta = n(C\mu)'(C\Sigma C')^{-1}(C\mu)$$

- The statistic

$$F = \frac{(n-1)-c+1}{(n-1)c}\, T^2 = \frac{n-c}{(n-1)c}\, T^2$$

has the $F_{c,n-c,\delta}$ distribution

- If H_0 is true, $F \sim F_{c,n-c}$

- This test can be used if $n > c$

One-Sample Repeated Measures

- Let y_{ij} denote the response from subject i at time j, for $i = 1, \ldots, n$, $j = 1, \ldots, t$
- The $y_i = (y_{i1}, \ldots, y_{it})'$ vectors are a random sample from $N_t(\mu, \Sigma)$, where $\mu = (\mu_1, \ldots, \mu_t)'$
- Suppose that we wish to test $H_0\colon \mu_1 = \cdots = \mu_t$
- Let $y_{ij}^* = y_{ij} - y_{i,j+1}$, for $j = 1, \ldots, t-1$
- The $y_i^* = (y_{i1}^*, \ldots, y_{i,t-1}^*)'$ vectors are a random sample from $N_{t-1}(\mu^*, \Sigma^*)$, where $\mu^* = (\mu_1 - \mu_2, \mu_2 - \mu_3, \ldots, \mu_{t-1} - \mu_t)'$
- $H_0\colon \mu_1 = \cdots = \mu_t$ is then equivalent to

$$H_0^*\colon \mu^* = (0, \ldots, 0)'$$

One-Sample Repeated Measures

- The test of H_0^* can be carried out using the T^2 statistic computed from the sample mean vector and covariance matrix of the y_{ij}^* values

- $\sqrt{n}\overline{y}^* \sim N_{t-1}(\sqrt{n}\mu^*, \Sigma^*)$

- $(n-1)S^* \sim W_{t-1}(n-1, \Sigma^*)$

- $T^2 = n\overline{y}^{*\prime}\, S^{*\,-1}\, \overline{y}^* \sim T^2_{t-1,n-1,\delta^*}$, where

  $\delta^* = n\mu^{*\prime}\Sigma^{*\,-1}\mu^*$

- $$F = \frac{(n-1)-(t-1)+1}{(n-1)(t-1)}\, T^2$$

  $$= \frac{n-t+1}{(n-1)(t-1)}\, T^2$$

  has the $F_{t-1,n-t+1}$ distribution if H_0^* is true

Matrix Formulation

- $y_i^* = Cy_i$, where C is the $(t-1) \times t$ matrix

$$\begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \multicolumn{6}{c}{\dots\dots\dots\dots\dots\dots\dots\dots} \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

- $y_i^* \sim N_{t-1}(C\mu, C\Sigma C')$ and

$$T^2 = n(C\overline{y})'(CSC')^{-1}(C\overline{y})$$

- The value of T^2 is invariant with respect to the specific choice of C; another choice is

$$C = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ -1 & 0 & 1 & \dots & 0 & 0 \\ \multicolumn{6}{c}{\dots\dots\dots\dots\dots\dots\dots\dots} \\ -1 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

- Other types of hypotheses of the general form $H_0\colon C\mu = 0$ can also be tested

Example

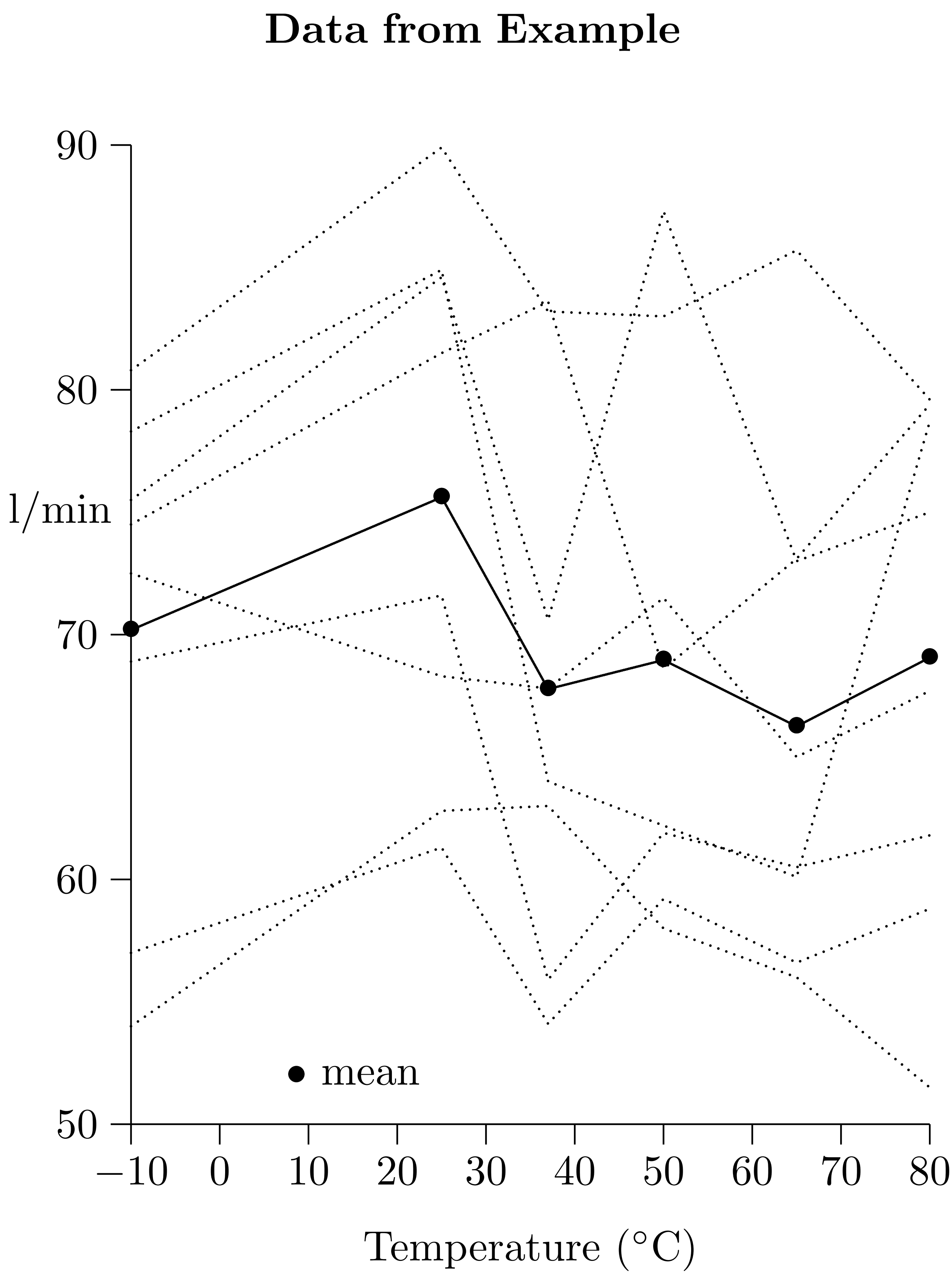
- Deal et al. (1979) measured ventilation volumes (l/min) of eight subjects under six different temperatures of inspired dry air

	Temperature (°C)					
Subject	−10	25	37	50	65	80
1	74.5	81.5	83.6	68.6	73.1	79.4
2	75.5	84.6	70.6	87.3	73.0	75.0
3	68.9	71.6	55.9	61.9	60.5	61.8
4	57.0	61.3	54.1	59.2	56.6	58.8
5	78.3	84.9	64.0	62.2	60.1	78.7
6	54.0	62.8	63.0	58.0	56.0	51.5
7	72.5	68.3	67.8	71.5	65.0	67.7
8	80.8	89.9	83.2	83.0	85.7	79.6

- Is ventilation volume affected by temperature?

Reference

Deal, E. C., McFadden, E. R., Ingram, R. H. et al. (1979). Role of respiratory heat exchange in production of exercise-induced asthma. *J Appl Physiol* **46**, 467–475.


Data from Example
90
80
l/min
70
60
50
● mean
−10
0
10
20
30
40
50
60
70
80
Temperature (°C)

SAS Statements for Example

```
 data a;
   input subject vv1-vv6;
   cards;
1 74.5 81.5 83.6 68.6 73.1 79.4
2 75.5 84.6 70.6 87.3 73.0 75.0
3 68.9 71.6 55.9 61.9 60.5 61.8
4 57.0 61.3 54.1 59.2 56.6 58.8
5 78.3 84.9 64.0 62.2 60.1 78.7
6 54.0 62.8 63.0 58.0 56.0 51.5
7 72.5 68.3 67.8 71.5 65.0 67.7
8 80.8 89.9 83.2 83.0 85.7 79.6
 ;
 proc glm;
   model vv1-vv6= / nouni;
   repeated ventvol / nou;
```

- The `nouni` option omits separate analyses for each dependent variable
- The `nou` option omits repeated measures ANOVA

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

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

    Not of great interest (and answer is obvious)

  - Is the relationship between age and bone height linear?

    Test of nonlinearity can be carried out using orthogonal polynomial coefficients, since the measurements are equally spaced

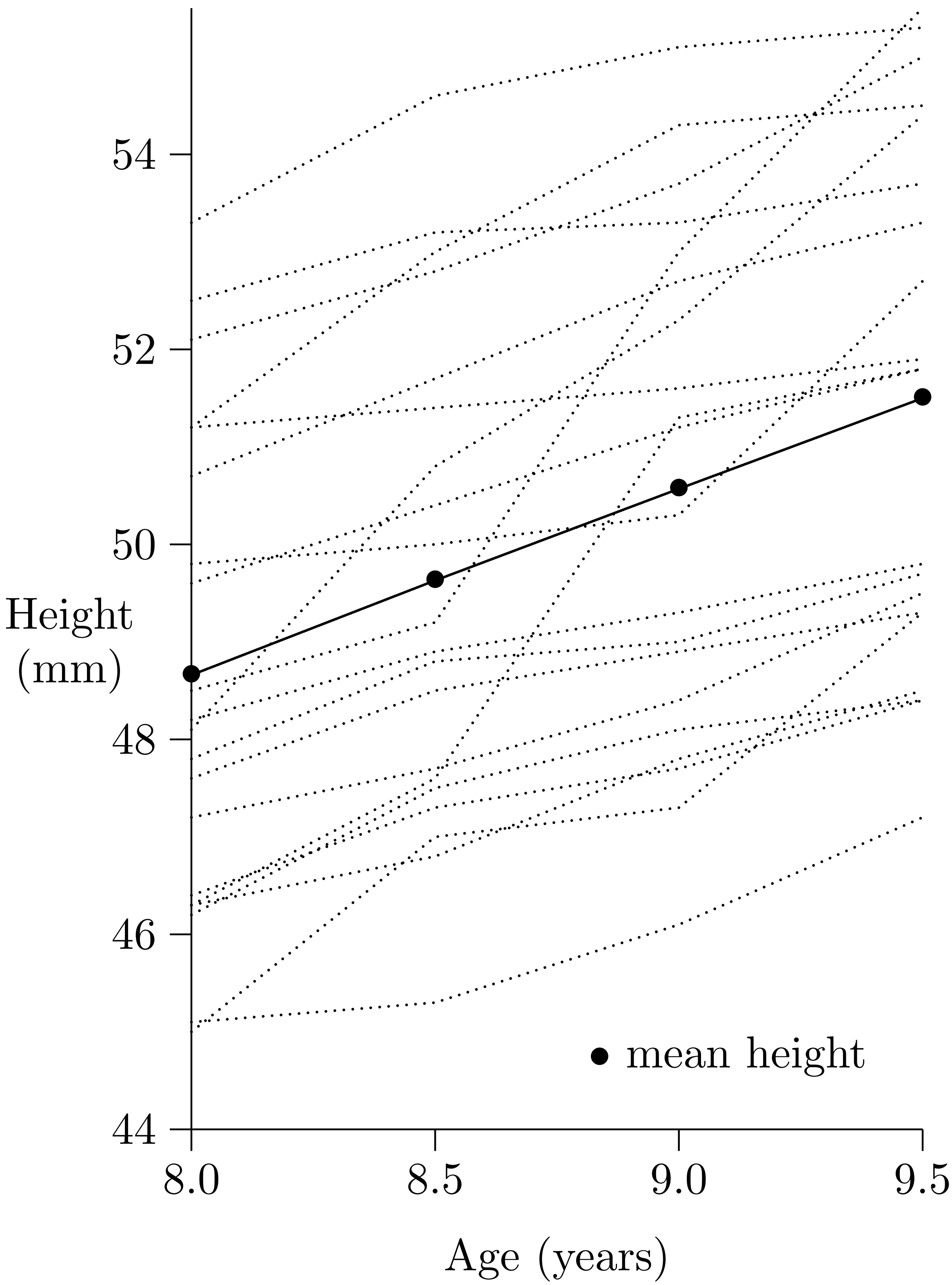
Reference

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

Data from Example



Orthogonal Polynomial Coefficients

No. of Points	Order						
3	Linear	−1	0	1			
	Quadratic	1	−2	1			
4	Linear	−3	−1	1	3		
	Quadratic	1	−1	−1	1		
	Cubic	−1	3	−3	1		
5	Linear	−2	−1	0	1	2	
	Quadratic	2	−1	−2	−1	2	
	Cubic	−1	2	0	−2	1	
	Quartic	1	−4	6	−4	1	
6	Linear	−5	−3	−1	1	3	5
	Quadratic	5	−1	−4	−4	−1	5
	Cubic	−5	7	4	−4	−7	5
	Quartic	1	−3	2	2	−3	1
	Quintic	−1	5	−10	10	−5	1

A more extensive tabulation is given in:

Pearson and Hartley, 1966, *Biometrika Tables for Statisticians, Volume I*, pp. 236–245

SAS Statements for Example

```
data a;
input subject h80 h85 h90 h95;
cards;
 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
       . . .
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
 ;
proc glm;
model h80 h85 h90 h95= / nouni;
repeated height polynomial / nou summary;
manova h=intercept m=( 1 -1 -1  1,
                      -1  3 -3  1);
```

- The `m` option of the `manova` statement tests for nonlinearity

Testing for Nonlinearity

- Orthogonal polynomial coefficients for unequally spaced time points are not tabulated (but can be generated)

- Another approach is the method of divided differences (Hills, 1968)

- Suppose that measurements are obtained at time points $x_1, \ldots, x_t$

- Let $d_j = \dfrac{1}{x_{j+1} - x_j}$, for $j = 1, \ldots, t-1$

- The test of nonlinearity is $H_0\colon C\mu = 0$, where C is the $(t-2) \times t$ matrix

$$\begin{pmatrix} -d_1 & d_1+d_2 & -d_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -d_2 & d_2+d_3 & -d_3 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -d_{t-2} & d_{t-2}+d_{t-1} & -d_{t-1} \end{pmatrix}$$

Testing for Nonlinearity

- If the measurements in the previous example had been obtained at ages 8, 8.5, 9, and 10,

$$d_1 = \frac{1}{8.5 - 8} = 2, \quad d_2 = \frac{1}{9 - 8.5} = 2,$$
$$d_3 = \frac{1}{10 - 9} = 1$$

  and

$$C = \begin{pmatrix} -2 & 4 & -2 & 0 \\ 0 & -2 & 3 & -1 \end{pmatrix}$$

- If the time points $x_1, \ldots, x_t$ are equally spaced, then $d_1 = \cdots = d_{t-1} = 1$ and

$$C = \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \end{pmatrix}$$

- This approach can be generalized to test the adequacy of higher-order polynomials

Comments on the One-Sample Unstructured Multivariate Approach

Positive:

- Assumes only multivariate normality
- Covariance structure is not specified
- Analogous to the univariate paired-t test

Negative:

- Uses up df in estimating covariance parameters
- Has low power when denominator df is small
- Can only be used when the number of linearly independent components of the hypothesis is less than the number of subjects

  e.g., in order to test homogeneity, $n \geq t$
- Can not be easily adapted for situations in which there are missing data

Unstructured Multivariate Approach

The Two-Sample Problem

- Repeated measurements at t time points are obtained from two independent groups of subjects

		Time Point				
Group	Subject	1	$\ldots$	j	$\ldots$	t
1	1	y_{111}	$\cdots$	y_{11j}	$\cdots$	y_{11t}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	i	y_{1i1}	$\cdots$	y_{1ij}	$\cdots$	y_{1it}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	n_1	y_{1n_11}	$\cdots$	y_{1n_1j}	$\cdots$	y_{1n_1t}
2	1	y_{211}	$\cdots$	y_{21j}	$\cdots$	y_{21t}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	i	y_{2i1}	$\cdots$	y_{2ij}	$\cdots$	y_{2it}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	n_2	y_{2n_21}	$\cdots$	y_{2n_2j}	$\cdots$	y_{2n_2t}

Unstructured Multivariate Approach

The Two-Sample Problem

- $y_{1i} \sim N_t(\mu_1, \Sigma)$, for $i = 1, \ldots, n_1$, where

  $y_{1i} = (y_{1i1}, \ldots, y_{1it})'$ and $\mu_1 = (\mu_{11}, \ldots, \mu_{1t})'$

- $y_{2i} \sim N_t(\mu_2, \Sigma)$, for $i = 1, \ldots, n_2$, where

  $y_{2i} = (y_{2i1}, \ldots, y_{2it})'$ and $\mu_2 = (\mu_{21}, \ldots, \mu_{2t})'$

- We wish to test $H_0\colon \mu_1 = \mu_2$

- $\overline{y}_h \sim N_t\Big(\mu_h, \dfrac{1}{n_h}\Sigma\Big)$, for $h = 1, 2$, and

$$\overline{y}_1 - \overline{y}_2 \sim N_t\left(\mu_1 - \mu_2, \Big(\frac{1}{n_1} + \frac{1}{n_2}\Big)\Sigma\right)$$

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\Big(\overline{y}_1 - \overline{y}_2\Big) \sim N_t\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\Big(\mu_1 - \mu_2\Big), \Sigma\right)$$

Unstructured Multivariate Approach
The Two-Sample Problem

- The pooled estimator of the covariance matrix Σ is given by

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2},$$

where

$$S_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \overline{y}_h)(y_{hi} - \overline{y}_h)'$$

is the sample covariance matrix in group h

- $(n_h - 1)S_h \sim W_t(n_h - 1, \Sigma)$

- $(n_1 - 1)S_1 + (n_2 - 1)S_2 \sim W_t(n_1 + n_2 - 2, \Sigma)$

- Therefore, $(n_1 + n_2 - 2)S \sim W_t(n_1 + n_2 - 2, \Sigma)$

Unstructured Multivariate Approach

The Two-Sample Problem

- $T^2 = \dfrac{n_1 n_2}{n_1 + n_2} \left(\overline{y}_1 - \overline{y}_2\right)' S^{-1} \left(\overline{y}_1 - \overline{y}_2\right)$ has the $T^2_{t, n_1+n_2-2, \delta}$ distribution with noncentrality parameter

$$\delta = \frac{n_1 n_2}{n_1 + n_2} \left(\mu_1 - \mu_2\right)' \Sigma^{-1} \left(\mu_1 - \mu_2\right)$$

- $F = \dfrac{(n_1 + n_2 - 2) - t + 1}{(n_1 + n_2 - 2)t} T^2$

$$= \frac{n_1 + n_2 - t - 1}{(n_1 + n_2 - 2)t} T^2$$

  has the $F_{t, n_1+n_2-t-1, \delta}$ distribution

- If $H_0\colon \mu_1 = \mu_2$ is true, $F \sim F_{t, n_1+n_2-t-1}$

- Note that this test assumes $\Sigma_1 = \Sigma_2$

Tests of Other Hypotheses

- Suppose we wish to test $H_0\colon C(\mu_1 - \mu_2) = 0$, where C is a $c \times t$ matrix of rank c ($c \leq t$)

- Let $z_{hi} = Cy_{hi}$, for $h = 1, 2$

- Since $\overline{y}_1 - \overline{y}_2 \sim N_t\left(\mu_1 - \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right)$,

$$\overline{z}_1 - \overline{z}_2 \sim N_c\left(C\left(\mu_1 - \mu_2\right), \left(\frac{n_1 + n_2}{n_1 n_2}\right)C\Sigma C'\right)$$

- Let $S_{zh} = CS_hC'$ denote the sample covariance matrix of the z_i vectors from group h and let

$$S_z = \frac{(n_1 - 1)S_{z1} + (n_2 - 1)S_{z2}}{n_1 + n_2 - 2}$$

denote the pooled covariance matrix

- $(n_1 + n_2 - 2)S_z \sim W_c(n_1 + n_2 - 2, C\Sigma C')$

Tests of Other Hypotheses

- Therefore,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{y}_1 - \overline{y}_2)' C' \left(CSC'\right)^{-1} C(\overline{y}_1 - \overline{y}_2)$$

has the $T^2_{c,n_1+n_2-2,\delta}$ distribution with noncentrality parameter

$$\delta = \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)' C' (C\Sigma C')^{-1} C(\mu_1 - \mu_2)$$

- $F = \dfrac{(n_1 + n_2 - 2) - c + 1}{(n_1 + n_2 - 2)c} T^2$

$$= \frac{n_1 + n_2 - c - 1}{(n_1 + n_2 - 2)c} T^2$$

has the $F_{c,n_1+n_2-c-1,\delta}$ distribution

- If $H_0\colon C\mu_1 = C\mu_2$ is true, $F \sim F_{c,n_1+n_2-c-1}$

Hypothesis of Parallelism

- A weaker, and often more realistic, hypothesis is that the μ-profiles in the two groups are parallel

  i.e., the μ_1 and μ_2 profiles differ only by a constant vertical shift

- This hypothesis can be expressed as:

$$\begin{aligned} H_0\colon \mu_{12} - \mu_{11} &= \mu_{22} - \mu_{21}, \\ \mu_{13} - \mu_{12} &= \mu_{23} - \mu_{22}, \\ &\vdots \\ \mu_{1t} - \mu_{1,t-1} &= \mu_{2t} - \mu_{2,t-1} \end{aligned}$$

  or as $H_0\colon C(\mu_1 - \mu_2) = 0$, where C is the $(t-1) \times t$ matrix

$$\begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

Example

- A study conducted in two groups of children (16 boys and 11 girls)

- At ages 8, 10, 12, and 14, the distance (mm) from the center of the pituitary to the pteryomaxillary fissure was measured

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

- Are the profiles for boys and girls the same?

$$H_0\colon\ \mu_b = \mu_g$$

- Are the profiles for boys and girls parallel?

$$\begin{aligned} H_0\colon\ & \mu_{b,10} - \mu_{b,8} = \mu_{g,10} - \mu_{g,8}, \\ & \mu_{b,12} - \mu_{b,10} = \mu_{g,12} - \mu_{g,10}, \\ & \mu_{b,14} - \mu_{b,12} = \mu_{g,14} - \mu_{g,12} \end{aligned}$$

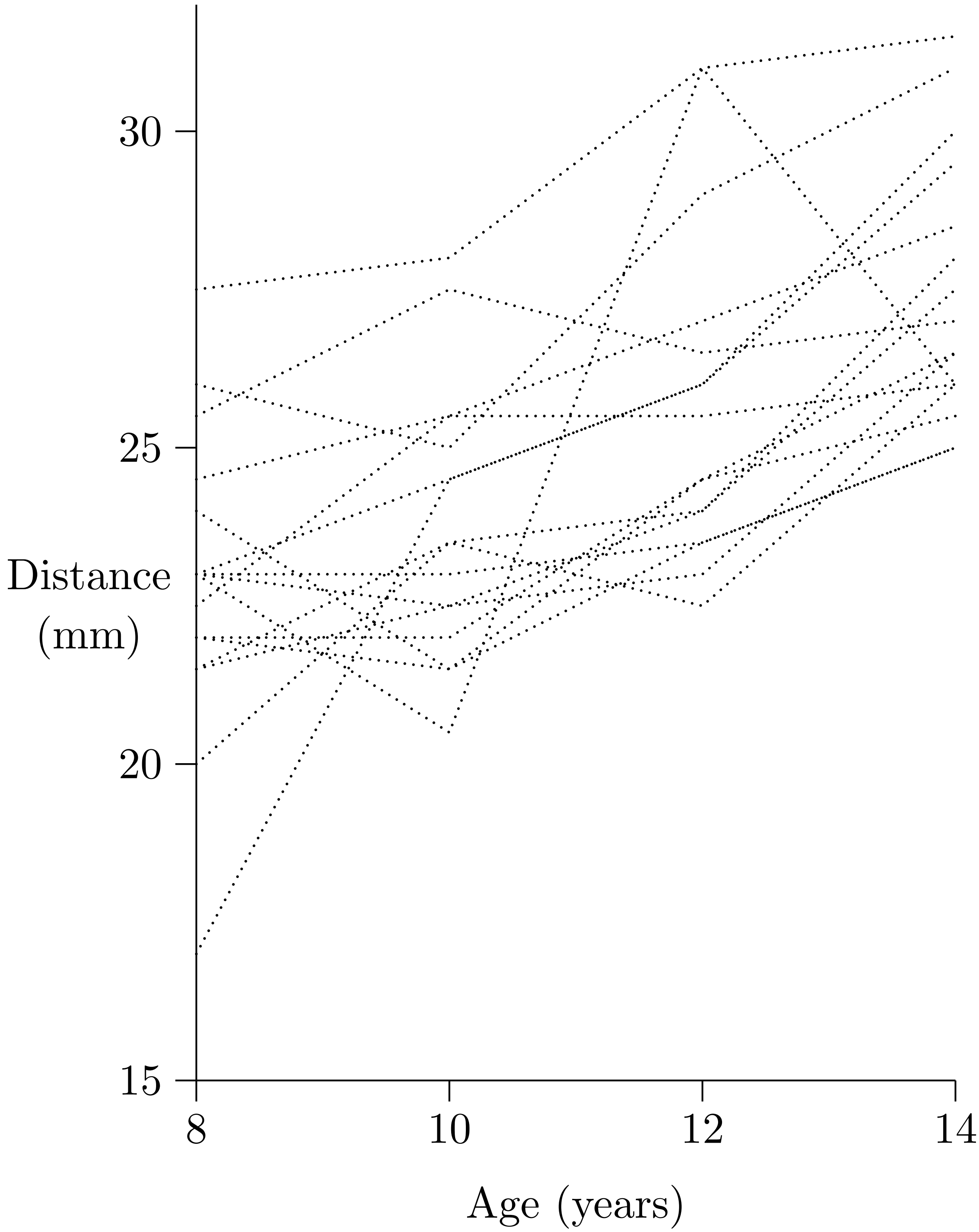
Reference

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.

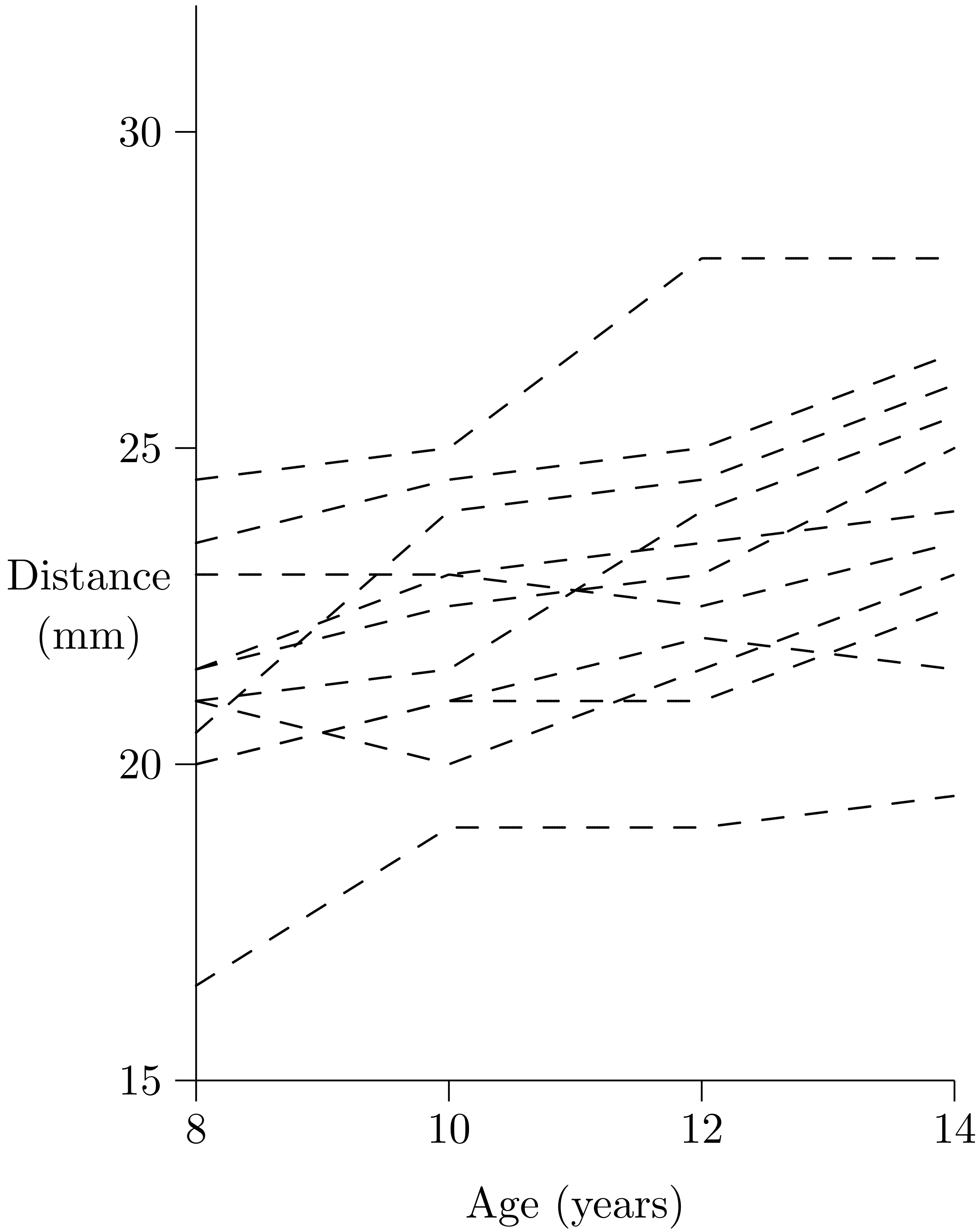
Dental Measurements

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

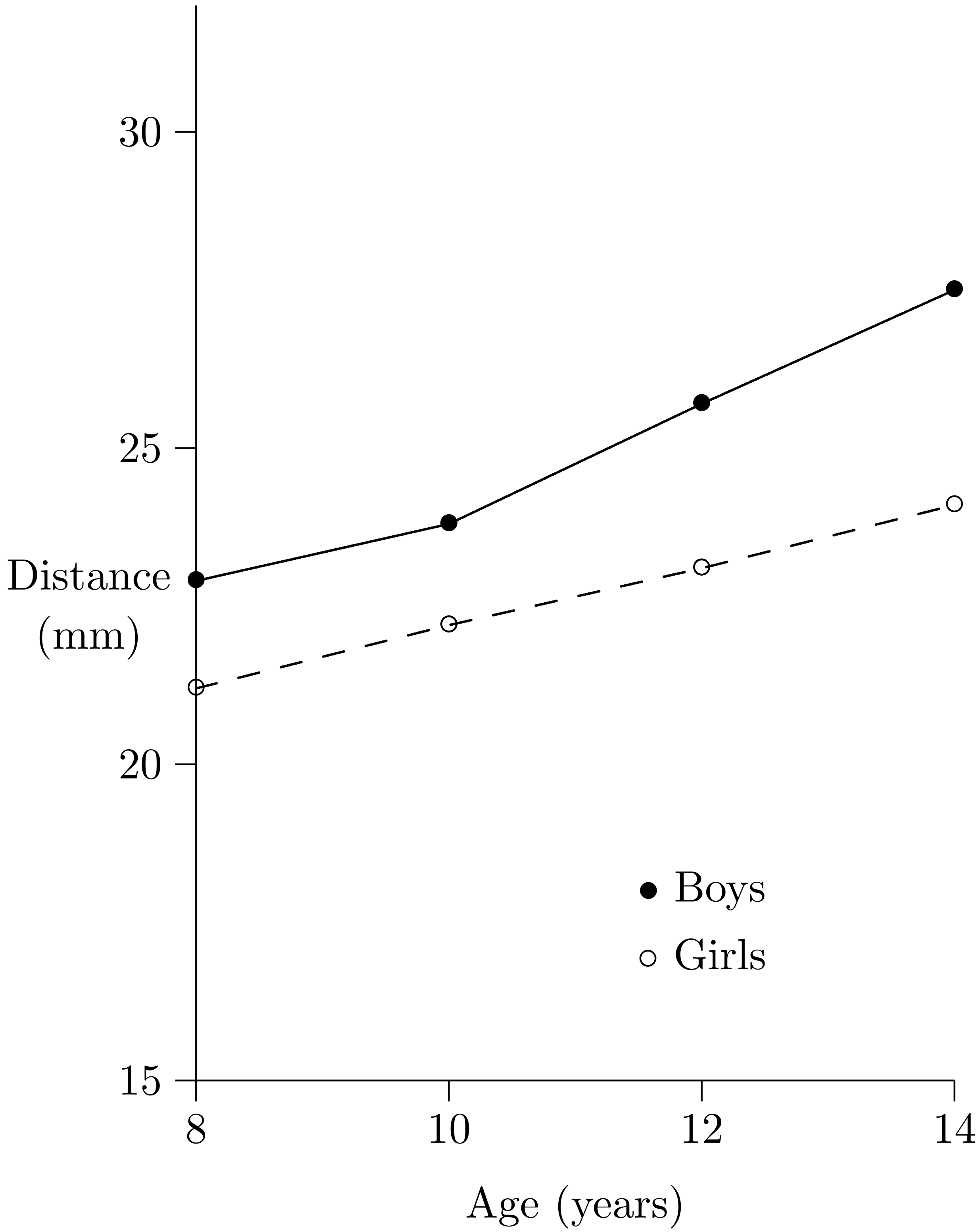
Dental Measurements in Boys



Dental Measurements in Girls



Mean Dental Measurements



SAS Statements

- The variable `sex` is coded 1 for boys, 2 for girls

```
data a;
input sex id d8 d10 d12 d14;
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
;
proc glm;
model d8 d10 d12 d14=sex / nouni;
manova h=sex;
manova h=sex m=(-1  1  0  0,
                 0 -1  1  0,
                 0  0 -1  1);
```

- The first `manova` statement tests $H_0\colon \mu_b = \mu_g$, while the second tests parallelism

Test of Equality of Covariance Matrices

- Bartlett's modification of the likelihood ratio test can be used to test $H_0\colon \Sigma_1 = \Sigma_2 = \cdots = \Sigma_s$
- Implemented in the SAS DISCRIM procedure

```
proc discrim pool=test:
class class-variable;
var list-of-variables;
```

  - *class-variable* defines the s groups
  - *list-of-variables* defines the t components of the multivariate normal distribution
- The asymptotic distribution of the test criteria used in PROC DISCRIM is $\chi^2_{(s-1)t(t+1)/2}$
- Parhizgari and Prakash (1989) implement an improved approximation
- Although this test is unbiased, it is not robust to non-normality

What if Covariance Matrices are Unequal?

- If $\Sigma_1 \neq \Sigma_2$, the significance level of the T^2 test of $H_0\colon \mu_1 = \mu_2$ depends on Σ_1 and Σ_2

- If the difference between Σ_1 and Σ_2 is small, or if the sample sizes n_1 and n_2 are large, there is no practical effect

- Otherwise, the nominal significance level may be distorted

- If $n_1 = n_2 = n/2$, the null hypothesis can be tested using a $T^2_{t,(n-2)/2}$ statistic

  (In comparison, $T^2_{t,n-2}$ assuming $\Sigma_1 = \Sigma_2$)

- If $n_1 < n_2$, $H_0\colon \mu_1 = \mu_2$ can be tested using a T^2_{t,n_1-1} statistic