Index of Lecture 3–L

Practical Information

Today's lecture: Chapters 3–4 in Davis (2002),

- multivariate analysis for normally distributed outcomes,
  - multivariate $\sim$ multiple outcomes,
  - in the repeated measures context, the series of measures on each subject $\sim$ multivariate outcome,
- Chapter 3: one- and two-sample models,
- Chapter 4: general MANOVA model + two models/ approaches for repeated measures: profile and growth-curve analysis,
- theory and methods for multivariate analysis require matrix calculus
  $\Rightarrow$ examples focus on ideas and less on formulae.

Software for multivariate analysis:

- Stata, SAS and SPSS offer full set of classical analyses,
- R/S-Plus only limited facilities (but some non-standard add-ons).

Course schedule:

- material on multivariate normal methods expanded onto two sessions (I won't manage all of Chapter 4 today).

Analysis of Ventilation Volume Data

Multivariate model:

$$y_i = (y_{i1}, \ldots, y_{i6}) \sim \mathrm{N}_6(\mu, \Sigma), \quad i = 1, \ldots, 8,$$

where $\mu = (\mu_1, \ldots, \mu_6)$, the means at the 6 temperatures. The hypothesis of interest is

$$\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_6.$$

Approach:

- compute for each subject the 5 differences: $y^*_{ij} = y_{ij} - y_{i,j+1}$ for $j = 1, \ldots, 5$,
- test for y^*_{ij}'s the ${\mathrm{H}_0}^* : \mu^*_1 = \mu^*_2 = \ldots = \mu^*_5 = 0$, using the T^2-statistic, and

  $$F = \frac{8-6+1}{(8-1)(6-1)} T^2 = \frac{3}{35} T^2 \;\sim\; F_{6-1,8-6+1} = F_{5,3}$$

  under the hypothesis H_0 (and hence ${\mathrm{H}_0}^*$),
- calculation:

  $$T^2 = 34.155, \quad F = 2.93, \quad P = P(F_{5,3} > 2.93) = 0.20,$$

- no significant difference between temperatures,
- test has low power due to low denominator degrees of freedom for F.

Analysis of Ramus Bone Data

Multivariate model:

$$y_i = (y_{i1}, \ldots, y_{i4}) \sim \mathrm{N}_4(\mu, \Sigma), \quad i = 1, \ldots, 20,$$

where $\mu = (\mu_1, \ldots, \mu_4)$, the means at the four ages.

Analyses:

- test of age homogeneity $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_4$,
  - ∗ same approach as in previous example, leading to
    $$T^2 = 73.16,\ F = \frac{20-4+1}{(20-1)(4-1)} T^2 = \frac{17}{57} T^2 = 21.82,$$
    which in $F_{4-1,20-4+1} = F_{3,17}$-dist. $\sim P < 0.0001$,
  - ∗ H_0 is clearly significant (but not very interesting!),
- test of linearity of age effect, $\mathrm{H}_0 : \mu_j = \beta_0 + \beta_1 \,\mathrm{age}_j$,
  - ∗ due to the equi-spaced time points, $\mathrm{H}_0 \sim$ zero high-order polynomial contrasts,
  - ∗ construct derived variables:
    $$\begin{aligned}(\text{quadr.}) :\ y_{i1}^* &= (1)y_{i1} + (-1)y_{i2} + (-1)y_{i3} + (1)y_{i4},\\ (\text{cubic}) :\ y_{i2}^* &= (-1)y_{i1} + (3)y_{i2} + (-3)y_{i3} + (1)y_{i4},\end{aligned}$$
    and test hypothesis ${\mathrm{H}_0}^* : \mu_1^* = \mu_2^* = 0$,
  - ∗ calculation:
    $$T^2 = 0.038,\ F = \frac{20-2}{(20-1)2} T^2 = \frac{18}{38} T^2 = 0.018,$$
    which under ${\mathrm{H}_0}^*$ follows $F_{2,20-2} = F_{2,18}$ and is clearly non-significant,
  - ∗ conclusion: no indication at all of non-linearity.

Two-Sample Problem

Statistical model:

- $y_{1i} \sim \mathrm{N}_t(\mu_1, \Sigma)$ for $i = 1, \ldots, n_1$, where $y_{1i} = (y_{1i1}, \ldots, y_{1it})'$ and $\mu_1 = (\mu_{11}, \ldots, \mu_{1t})'$,
- $y_{2i} \sim \mathrm{N}_t(\mu_2, \Sigma)$ for $i = 1, \ldots, n_2$, where $y_{2i} = (y_{2i1}, \ldots, y_{2it})'$ and $\mu_2 = (\mu_{21}, \ldots, \mu_{2t})'$,
- note the assumption of equal variances in the two groups.

Hypothesis of interest: $\mathrm{H}_0 : \mu_1 = \mu_2$.

Statistical procedure (generalization of univariate t-test):

- pooled covariance matrix: $S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$,
- T^2 test statistic:

$$T^2 = [(1/n_1) + (1/n_2)]^{-1}(\bar{y}_1 - \bar{y}_2)' S^{-1} (\bar{y}_1 - \bar{y}_2),$$

- under H_0: $F = \frac{n_1+n_2-t-1}{(n_1+n_2-2)t} T^2 \sim F_{t, n_1+n_2-t-1}$.

More general hypothesis of interest: $\mathrm{H}_0 : C(\mu_1 - \mu_2) = 0$, for a $c \times t$ matrix C,

- compute derived variables $z_{1i} = C y_{1i}$ and $z_{2i} = C y_{2i}$,
- test hypothesis $\mathrm{H}_0 : \mu_{z1} = \mu_{z2}$ using T^2 (for z's), and

$$F = \frac{n_1+n_2-c-1}{(n_1+n_2-2)c} T^2 \;\sim\; F_{c, n_1+n_2-c-1} \text{ under } \mathrm{H}_0.$$

First Analysis of Dental Data

Multivariate model:

$$\begin{aligned} y_{bi} &= (y_{bi1}, \ldots, y_{bi4}) \sim \mathrm{N}_4(\mu_b, \Sigma), \quad i = 1, \ldots, 16, \\ y_{gi} &= (y_{gi1}, \ldots, y_{gi4}) \sim \mathrm{N}_4(\mu_g, \Sigma), \quad i = 1, \ldots, 11, \end{aligned}$$

where $\mu_b = (\mu_{b1}, \ldots, \mu_{b4})$ and $\mu_g = (\mu_{g1}, \ldots, \mu_{g4})$, the means at the four ages.

Analyses:

- test of equal boy and girl means $\mathrm{H}_0 : \mu_b = \mu_g$,
  - ∗ calculation:

    $$T^2 = 16.51, \; F = \tfrac{16+11-4-1}{(16+11-2)4} T^2 = \tfrac{22}{100} T^2 = 3.63,$$

    which in $F_{4,16+11-4-1} = F_{4,22}$-dist. $\sim P = 0.020$,
  - ∗ test is (weakly) significant — evidence of gender differences,

- test of parallel curves for boys and girls,
  - ∗ compute differences: $y^*_{bi1} = y_{bi2} - y_{bi1}, \ldots,$ $y^*_{bi3} = y_{bi4} - y_{bi3}$ (and similarly for the girls),
  - ∗ test equality between boys and girls for y^*-variables,

    $$T^2 = 8.79, \; F = \tfrac{16+11-3-1}{(16+11-2)3} T^2 = \tfrac{23}{75} T^2 = 2.695,$$

    which in $F_{3,16+11-3-1} = F_{3,23}$-dist. $\sim P = 0.070$,
  - ∗ test is close to significant — some indication of non-parallel curves.

More Analyses of Dental Data

- test of parallelism: on previous slide, $P = 0.070$,
- test of no group differences — two versions:
  - without assuming parallelism: on previous slide, $P = 0.020$,
  - assuming parallelism: univariate t-test between two groups (assuming same variances; similar without assumption),

$$t = 3.05, \quad \text{df} = 25, \quad P = 0.005,$$

- test of no time differences — two versions:
  - without assuming parallelism: test no reviewed yet:

$$F = 11.46, \quad \text{df} = (6, 46), \quad P < 0.0001,$$

  - assuming parallism: T^2-test similar to previous ones,

$$F = 36.4, \quad \text{df} = (3, 23), \quad P < 0.0001,$$

Conclusion:

- some indication of non-parallel curves,
- clear group and time differences.