STT 843: Multivariate Analysis

9. Multivariate Analysis of Variance

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

1. MANOVA (one-way)
2. MANOVA Table (one-way)
3. Wilks' Λ
4. Simultaneous Confidence Intervals for Treatment Effects
5. Repeated measure designs and growth curves

MANOVA (one-way)

Comparing means from g groups

- Sample from population 1: $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$
- Sample from population 2: $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$
- $\vdots$
- Sample from population g: $\mathbf{x}_{g1}, \mathbf{x}_{g2}, \ldots, \mathbf{x}_{gn_g}$

independent random samples $\mathbf{x}_{\ell j} \sim N(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma})$ $\leftarrow$ $\boldsymbol{\Sigma}$ is the common covariance matrix

- Instead of testing

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g \text{ vs. } H_1 : \text{ at least two } \boldsymbol{\mu}\text{'s are unequal}$$

we usually reparameterize

$$\boldsymbol{\mu}_\ell = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell \quad \leftarrow \text{ treatment effect}$$

- Thus $\mathbf{x}_{\ell j} \sim N(\boldsymbol{\mu} + \boldsymbol{\tau}_\ell, \boldsymbol{\Sigma})$ and

$$H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g$$

- Our model:

$$\mathbf{x}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \mathbf{e}_{\ell j}, \quad \ell = 1, \ldots, g, \quad j = 1, \ldots, n_\ell$$

- For uniqueness (identifiability), we impose the constraint

$$\sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \mathbf{0}$$

- Decomposition of sample:

$$\underset{\text{observed}}{\mathbf{x}_{\ell j}} = \underset{\substack{\uparrow \\ \text{overall} \\ \text{sample} \\ \text{mean} \\ \hat{\boldsymbol{\mu}}}}{\bar{\mathbf{x}}} + \underset{\substack{\text{estimated} \\ \text{treatment} \\ \text{effect} \\ \hat{\boldsymbol{\tau}}_\ell}}{(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})} + \underset{\substack{\text{residual} \\ \hat{\mathbf{e}}_{\ell j}}}{(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)}$$

- Multivariate analog of total (corrected) sum of squares is

$$\underbrace{\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}})(\shortparallel)'}_{\substack{\text{total corrected} \\ \text{sum of squares} \\ \text{and cross} \\ \text{products matrix}}} = \underbrace{\sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})(\shortparallel)'}_{\substack{=\mathbf{H} \\ \text{“Between”} \\ \text{groups} \\ \text{matrix}}} + \underbrace{\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\shortparallel)'}_{\substack{=\mathbf{E} \\ \text{“Within” groups} \\ \text{matrix} \\ = \sum_{\ell=1}^{g}(n_\ell - 1)\mathbf{S}_\ell}}$$

Notes:

- Assuming no linear dependencies, $\text{rank}\{\mathbf{H}\} = \min(p, \nu_H)$
- $\mathbf{S}_\ell$ is the covariance matrix for the ℓ^{th} sample. So,

$$E\left\{\frac{1}{\left(\sum_{\ell=1}^{g} n_\ell\right) - g}\mathbf{E}\right\} = \mathbf{\Sigma}$$

where $\text{rank}\{\mathbf{E}\} = \min(p, \nu_E)$

MANOVA TABLE (one-way)

Source	SS Matrix	d.f.
Treatment	$\mathbf{H}$	$\nu_H = g - 1$
Error	$\mathbf{E}$	$\nu_E = (\sum_{\ell=1}^{g} n_\ell) - g$
Total (corrected)	$\mathbf{H} + \mathbf{E}$	$(\sum_{\ell=1}^{g} n_\ell) - 1$

The likelihood ratio test of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g$ rejects H_0 when

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \leq \Lambda_{\alpha, p, \nu_H, \nu_E}$$

- $$\mathbf{E} + \mathbf{H} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}})(\mathbf{x}_{lj} - \bar{\mathbf{x}})^T .$$
- We have that $\mathbf{E} \sim W_p(n-g, \mathbf{\Sigma})$ and under H_0, $\mathbf{H} \sim W_p(g-1, \mathbf{\Sigma})$ independently of $\mathbf{E}$ (how?), hence $\mathbf{H} + \mathbf{E} \sim W_p(n-1, \mathbf{\Sigma})$.
- The exact (null) distributions of (transformations of) Λ^* are known in some special cases: $p = 1, 2$ and $g = 2, 3$ - see Table 6.3.

Wilks' Λ

- Note: Reject for small values of Λ. As in univariate anova F-test, we "accept" when total SS $(\mathbf{E} + \mathbf{H})$ is dominated by error $(\mathbf{E})$.
- Note: We sometimes refer to the subscripts of the Λ_{p,ν_H,ν_E} distribution as "dimension," "numerator df," and "denominator df" (e.g., $\Lambda_{\text{dim}\ ,\mathrm{df}_{\text{num}}\ ,\mathrm{df}_{\text{den}}}$)

Properties of Wilk's Λ

1. For statistic to be obtained, we need $\nu_E \geq p$.
2. Degrees of freedom ν_H and ν_E are the same as in analogous univariate case; e.g., one-way model: $\nu_H = g - 1$ and $\nu_E = \sum_{\ell=1}^{g} n_\ell - g$
3. Let $\lambda_1, \ldots, \lambda_s$ be the s non-zero eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $s = \min(p, \nu_H)$. Then $\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$.
4. Critical value $\Lambda_{\alpha,p,\nu_H,\nu_E}$ decreases as p increases. Thus, adding variables decreases power unless variables contribute to separation.
5. When $\nu_H = 1$ or $\nu_H = 2$ or $p = 1$ or $p = 2$, Λ can be transformed to follow an F distribution.

- If $\nu_H = 1$

$$\frac{\nu_E - p + 1}{p}\frac{1 - \Lambda}{\Lambda} \sim F_{p,\nu_E - p + 1}$$

- If $\nu_H = 2$

$$\frac{\nu_E - p + 1}{p}\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2p,2(\nu_E - p + 1)}$$

- If $p = 1$

$$\frac{\nu_E}{\nu_H}\frac{1-\Lambda}{\Lambda} \sim F_{\nu_H,\nu_E}$$

- If $p = 2$

$$\frac{(\nu_E-1)}{\nu_H}\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2\nu_H,2(\nu_E-1)}$$

6. Approximate tests

For $p > 2$ or $\nu_H > 2$ and n large

$$\chi^2 = -\left[\nu_E - \frac{1}{2}(p - \nu_H + 1)\right] \ln \Lambda \overset{\text{approx}}{\sim} \chi^2_{p\nu_H}$$

Approximately valid when $p^2 + \nu_H^2 \leq \frac{1}{3}\left[\nu_E - \frac{1}{2}(p - \nu_H + 1)\right]$

- More correct approximate distribution for Λ (exact when ν_H or p is 1 or 2):

$$F = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \frac{df_2}{df_1} \overset{\text{approx}}{\sim} F_{df_1, df_2}$$

$$\begin{aligned}
df_1 &= p\nu_H \\
df_2 &= wt - \frac{1}{2}(p\nu_H - 2) \\
w &= \nu_E + \nu_H - \frac{1}{2}(p + \nu_H + 1) \\
t &= \begin{cases} \sqrt{\frac{p^2\nu_H^2 - 4}{p^2 + \nu_H^2 - 5}} & \text{for } p^2 + \nu_H^2 - 5 > 0 \quad (\text{ or } \quad p + \nu_H > 3) \\ 1 & \text{for } p^2 + \nu_H^2 - 5 \le 0 \quad (\text{ or } \quad p + \nu_H \le 3) \end{cases}
\end{aligned}$$

Other MANOVA Tests

Let $(\lambda_1, \dots, \lambda_s)$ be the ordered eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $s = \min(p, \nu_H) =$ rank of $\mathbf{H}$

- Roy's Largest Root:

$$\theta = \lambda_1$$

- Note: SAS and most authors denote Roy's Largest Root as λ_1 (the largest root of $\mathbf{E}^{-1}\mathbf{H}$). If Roy's Largest Root is $\xi_1 = \frac{\lambda_1}{1+\lambda_1}$, it is the largest root of $(\mathbf{E} + \mathbf{H})^{-1}\mathbf{H}$.

- Approximate F-statistic (used by SAS):

$$F_\theta = \frac{(\nu_E - d + \nu_H)}{d} \lambda_1$$

is an upper bound for "true F " which is distributed

$$F_{d}, \nu_E - d + \nu_H$$

where $(d = \max(p, \nu_H))$
F_θ-test is anti-conservative (yields lower bound on p-value)

Pillai's Trace

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$$
$$= \mathrm{tr}\left\{(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}\right\} = \sum_{i=1}^{s} \xi_i$$

where $\xi_1, \ldots, \xi_s$ are the s ordered eigen-values of $(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}$

- Note 1:

$$\left.\begin{array}{l} \mathbf{E}^{-1}\mathbf{H} \text{ is analagous to } \frac{\text{between SS}}{\text{within SS}} \\ (\mathbf{E}+\mathbf{H})^{-1}\mathbf{H} \text{ is analagous to } \frac{\text{between SS}}{\text{total SS}} \end{array}\right\} \begin{array}{c} \text{"Large Ratio"} \\ \text{Reject } H_0 \end{array}$$

- Note 2:

$$\xi_i = \frac{\lambda_i}{1+\lambda_i} \text{ and } \lambda_i = \frac{\xi_i}{1-\xi_i}$$

Approximate F-statistic (used in SAS):

$$F_V = \frac{(2N+s+1)}{(2m+s+1)} \left(\frac{V}{s-V} \right) \sim F_{s(2m+s+1),s(2N+s+1)}$$

where

$$s = \min\left(\nu_H, p\right)$$
$$m = \frac{1}{2}\left(|\nu_H - p| - 1\right)$$
$$N = \frac{1}{2}\left(\nu_E - p - 1\right)$$

Lawley-Hotelling Trace

$$
\begin{aligned}
U &= \sum_{i=1}^{s} \lambda_i \\
&= \operatorname{tr}\left\{\mathbf{E}^{-1}\mathbf{H}\right\}
\end{aligned}
$$

Approximate F-statistic (used in SAS):

$$
F_u = \frac{2(sN+1)}{s^2(2m+s+1)} U \sim F_{s(2m+s+1),2(sN+1)}
$$

$\rightarrow$ Also known as "Hotelling's generalized T^2"

Why four test statistics?

- All 4 are exact tests (i.e., have size α), but when H_0 not true they have different power
- For $p = 1, \mu_1, \dots, \mu_k$ can be ordered along 1 dimension (line) and F-test is U.M.P.
- For $p > 1, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k$ are points in $s = \min(p, \nu_H)$ dimensions. But means may in fact occupy only a subspace of the s dimensions; e.g., they may lie close to a line (1-D) or a plane (2-D).

Comparison of Multivariate Test Statistics Performance

Condition	**"Best" Statistic**
Concentrated Effect (one eigenvalue is large)	Roy's Largest Root
Diffuse Effect (spread across many dimensions)	Wilks' Lambda or Hotelling-Lawley
Small Samples or Violations of Assumptions	Pillai's Trace

Diffuse Effect

- Wilks' Lambda ($\Lambda = \prod \frac{1}{1+\lambda_i}$): Because it is a product-based statistic, it incorporates the "contribution" of every single eigenvalue. If multiple λ_i are greater than zero, their product compounds, making it easier to reject the null hypothesis.
- Hotelling-Lawley Trace ($T = \sum \lambda_i$): This is a direct sum of all eigenvalues. It captures the total volume of the treatment effect across all dimensions.

Small Samples or Violations of Assumptions

- Wilks' Lambda is a product. If a single eigenvalue is very small (potentially due to noise or a small sample size causing a near-singular matrix), the entire product is severely affected.
- Pillai's Trace is a sum, which acts as an averaging mechanism. This "averaging" prevents any single outlier eigenvalue from dominating the test statistic.
- When group sizes are unequal and covariance matrices differ, Wilks' Lambda, Hotelling-Lawley, and Roy's Root become highly susceptible to Type I errors (falsely claiming a significant result).
- In small sample sizes (n is close to p):The estimation of the Error matrix $\mathbf{E}$ is unstable.

Simultaneous Confidence Intervals for Treatment Effects

- Let τ_{ki} be the i-th component of τ_k. Since τ_k is estimated by $\hat{\tau}_k = \bar{\mathbf{x}}_k - \bar{\mathbf{x}}$ and $\hat{\tau}_{ki} - \hat{\tau}_{li} = \bar{x}_{ki} - \bar{x}_{li}$ is the difference between two independent sample means.
- Note that $Var(\hat{\tau}_{ki} - \hat{\tau}_{li}) = Var(\overline{X}_{ki} - \overline{X}_{li}) = \left(\frac{1}{n_k} + \frac{1}{n_l}\right)\frac{E_{ii}}{n-g}$, where E_{ii} is the ith diagonal element of $\mathbf{E}$ and $n = \sum_k n_k$

- There are p variables and $g(g-1)/2$ pairwise differences, critical value is $t_{n-g}(\alpha/2m)$, where $m = pg(g-1)/2$
- With confidence at least $(1-\alpha)$, $\tau_{ki} - \tau_{li}$ belongs to $(\overline{X}_{ki} - \overline{X}_{li}) \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{E_{ii}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}$ for all components $i = 1, \ldots, p$.

Repeated measure designs and growth curves

- A variation on repeated measures - treatment $l(l = 1, \ldots, g)$ is applied to each of n_l subjects and then a certain characteristic is monitored over time.
- For instance each of $n = \sum n_l$ plants is fertilized (using one of the g fertilizers) and their weights are measured at times $t_1, \ldots, t_p$.
- Any one plant is viewed as an observation from a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with $\boldsymbol{\mu}_i =$ mean size at time t_i.

Potthoff-Roy model for quadratic growth

The 'Potthoff-Roy model for quadratic growth' (there are others) takes

$$\boldsymbol{\mu} = \begin{pmatrix} \beta_0 + \beta_1 t_1 + \beta_2 t_1^2 \\ \vdots \\ \beta_0 + \beta_1 t_p + \beta_2 t_p^2 \end{pmatrix} = \mathbf{B}\boldsymbol{\beta}$$

in an obvious notation. The coefficient vectors $\boldsymbol{\beta}$ vary from one group to another.

- Thus the sample data are

$$\{\mathbf{x}_{lj} \mid j = 1, \ldots, n_l, l = 1, \ldots, g\}$$

with $\mathbf{x}_{lj} \sim N_p(\boldsymbol{\mu}_l = \mathbf{B}\boldsymbol{\beta}_l, \boldsymbol{\Sigma})$.

- We wish to compare the curves in varying groups. In the unrestricted model, with no assumed structure on μ_l, the MLEs are $\hat{\boldsymbol{\mu}}_l = \bar{\mathbf{x}}_l$ and $\hat{\boldsymbol{\Sigma}} = \mathbf{E}/n$, where $\mathbf{E} = (n-g)\mathbf{S}_{\text{pooled}}$, exactly as in one-way MANOVA.
- To test the adequacy of a particular growth curve model $\boldsymbol{\mu}_l = \mathbf{B}\beta_l$, where $\mathbf{B}$ is $p \times (q+1)$ (for instance when a q^{th} -order polynomial is fitted), we must find the MLE's of the $\boldsymbol{\beta}_l$.

These minimize the trace in the exponent of the likelihood, which is:

$$
\begin{aligned}
&\Sigma^{-1}\left\{\sum_l \sum_j (\mathbf{x}_{lj} - \mathbf{B}\beta_l)(\cdots)^T\right\} \\
=&\Sigma^{-1}\left\{\sum_l \sum_j ((\mathbf{x}_{lj} - \overline{\mathbf{x}}_l) + (\overline{\mathbf{x}}_l - \mathbf{B}\beta_l))(\cdots)^T\right\} \\
=&\Sigma^{-1}\mathbf{W} + \sum_l n_l (\overline{\mathbf{x}}_l - \mathbf{B}\beta_l)^T \Sigma^{-1} (\overline{\mathbf{x}}_l - \mathbf{B}\beta_l) \\
=&\Sigma^{-1}\mathbf{W} + \sum_l n_l \left\|\Sigma^{-1/2}\overline{\mathbf{x}}_l - \Sigma^{-1/2}\mathbf{B}\beta_l\right\|^2 .
\end{aligned}
$$

Thus $\hat{\beta}_l$ minimizes $\left\|\Sigma^{-1/2}\overline{\mathbf{x}}_l - \Sigma^{-1/2}\mathbf{B}\beta_l\right\|^2$; by standard least squares theory this is

$$\hat{\beta}_l = \left(\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}_l.$$

- It is usual (because it is much simpler) to take $\hat{\Sigma} = \mathbf{S}_{\text{pooled}}$, where $\mathbf{S}_{\text{pooled}} = \frac{1}{n-g}\mathbf{E}$
- To test a particular growth model, i.e. $H_0 : \boldsymbol{\mu} = \mathbf{B}\boldsymbol{\beta}$ (for any fixed $\mathbf{B}$ and $\boldsymbol{\beta}_{(q+1)\times 1}$, not necessarily representing quadratic effects) we fit without restrictions and then under the hypothesis, obtaining

$$\Lambda^* = \frac{|\mathbf{E}|}{|\mathbf{E}_q|}$$

where

$$\mathbf{E}_q = \sum_{l=1}^{g}\sum_{j=1}^{n_l}\left(\mathbf{x}_{lj} - \mathbf{B}\hat{\beta}_l\right)\left(\mathbf{x}_{lj} - \mathbf{B}\hat{\beta}_l\right)^T$$

$$= \mathbf{E} + \sum_{l} n_l\left(\bar{\mathbf{x}}_l - \mathbf{B}\hat{\beta}_l\right)\left(\bar{\mathbf{x}}_l - \mathbf{B}\hat{\beta}_l\right)^T .$$

Bartlett's approximation:

$$-\left(n - \frac{p-q+g}{2}\right) \log \Lambda^* \sim \chi^2_{(p-q-1)g}$$