STAT 3690 Lecture 13

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Feb 28, 2022

Testing for equality of population means (one-way multivariate analysis of variance (1-way MANOVA), J&W Sec. 6.4)

- Generalization of two-sample problem
  - Model: m independent samples, where
    - $\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n_1} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$
    - $\vdots$
    - $\mathbf{X}_{m1}, \ldots, \mathbf{X}_{mn_m} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_m, \boldsymbol{\Sigma})$
  - Hypotheses $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m$ v.s. H_1 : otherwise
- Alternatively
  - Model: m independent samples, where
    - $\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n_1} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu} + \boldsymbol{\tau}_1, \boldsymbol{\Sigma})$
    - $\vdots$
    - $\mathbf{X}_{m1}, \ldots, \mathbf{X}_{mn_m} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu} + \boldsymbol{\tau}_m, \boldsymbol{\Sigma})$
      - Identifiability: $\sum_i \boldsymbol{\tau}_i = 0$ otherwise there are infinitely many models that lead to the same data-generating mechanism.
  - Hypotheses $H_0 : \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_m = 0$ v.s. H_1 : otherwise
- Alternatively
  - Model: $\mathbf{X}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \mathbf{E}_{ij}$ with $\mathbf{E}_{ij} \overset{\text{iid}}{\sim} MVN_p(\mathbf{0}, \boldsymbol{\Sigma})$
    - Identifiability: $\sum_i \boldsymbol{\tau}_i = 0$
  - Hypotheses $H_0 : \boldsymbol{\tau}_1 = \cdots = \boldsymbol{\tau}_m = 0$ v.s. H_1 : otherwise

---

- Sample means and sample covariances
  - Sample mean for the ith sample $\bar{\mathbf{X}}_i = n_i^{-1} \sum_j \mathbf{X}_{ij}$
  - Sample covariance for the ith sample $\mathbf{S}_i = (n_i - 1)^{-1} \sum_j (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$
  - Grand mean $\bar{\mathbf{X}} = \sum_i n_i \bar{\mathbf{X}}_i / \sum_i n_i = \sum_{ij} \mathbf{X}_{ij} / \sum_i n_i$
  - Sum of squares and cross products matrix (SSP)
    - Within-group SSP
$$\mathbf{SSP}_{\mathrm{w}} = \sum_i (n_i - 1)\mathbf{S}_i = \sum_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$$
    - Between-group SSP
$$\mathbf{SSP}_{\mathrm{b}} = \sum_i n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^\top$$
    - Total (corrected) SSP
$$\mathbf{SSP}_{\mathrm{cor}} = \sum_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}})(\mathbf{X}_{ij} - \bar{\mathbf{X}})^\top = \mathbf{SSP}_{\mathrm{w}} + \mathbf{SSP}_{\mathrm{b}}$$

- Exercise: verify the decomposition $\mathbf{SSP}_{\text{cor}} = \mathbf{SSP}_{\text{w}} + \mathbf{SSP}_{\text{b}}$.

- MLE of $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma})$
  - Under H_0
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}$ for each i
    - $\widehat{\boldsymbol{\Sigma}} = (\sum_i n_i)^{-1}\mathbf{SSP}_{\text{cor}}$
  - Otherwise
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}_i = n_i^{-1}\sum_j \mathbf{X}_{ij}$
    - $\widehat{\boldsymbol{\Sigma}} = (\sum_i n_i)^{-1}\mathbf{SSP}_{\text{w}}$
- Likelihood ratio

$$\lambda = \left\{\frac{\det(\mathbf{SSP}_{\text{w}})}{\det(\mathbf{SSP}_{\text{cor}})}\right\}^{\sum_i n_i/2}$$

- Wilk's lambda test statistic

$$\Lambda = \lambda^{2/\sum_i n_i} = \frac{\det(\mathbf{SSP}_{\text{w}})}{\det(\mathbf{SSP}_{\text{cor}})}$$

  - Under H_0: $\Lambda \sim$ Wilk's lambda distribution $\Lambda(\boldsymbol{\Sigma}, \sum_i n_i - m, m-1)$
    - Since $\mathbf{SSP}_{\text{w}} \sim W_p(\boldsymbol{\Sigma}, \sum_i n_i - m)$ and $\mathbf{SSP}_{\text{b}} \sim W_p(\boldsymbol{\Sigma}, m-1)$
    - When $\sum_i n_i - m$ is large (i.e., $(p+m)/2 - \sum_i n_i + 1 \ll 0$), Bartlett's approximation

$$\{(p+m)/2 - \sum_i n_i + 1\} \ln \Lambda \approx \chi^2(p(m-1))$$

- Rejection region at level α

$$\begin{aligned}&\left\{x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{mn_m} : \{(p+m)/2 - \sum_i n_i + 1\} \ln \Lambda \ge \chi^2_{1-\alpha, p(m-1)}\right\} \\ = &\left\{x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{mn_m} : \Lambda \le \exp\left\{\frac{\chi^2_{1-\alpha,p(m-1)}}{(p+m)/2 - \sum_i n_i + 1}\right\}\right\}\end{aligned}$$

- p-value

$$1 - F_{\chi^2(p(m-1))}\left[\{(p+m)/2 - \sum_i n_i + 1\} \ln \Lambda\right]$$

- Exercise: factors in producing plastic film
  - W. J. Krzanowski (1988) *Principles of Multivariate Analysis.* A User's Perspective. Oxford UP, pp. 381.
  - Three response variables (tear, gloss and opacity) describing measured characteristics of the resultant film
  - A total of 20 runs
  - One factor `RATE` (rate of extrusion, 2-level, low or high) in the production test