# 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
    - $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$
    - $\vdots$
    - $\mathbf{X}_{m1}, \dots, \mathbf{X}_{mn_m} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_m, \boldsymbol{\Sigma})$
  - Hypotheses $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m$ v.s. $H_1$ : otherwise
- Alternatively
  - Model: $m$ independent samples, where
    - $\mathbf{X}_{11}, \dots, \mathbf{X}_{1n_1} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu} + \boldsymbol{\tau}_1, \boldsymbol{\Sigma})$
    - $\vdots$
    - $\mathbf{X}_{m1}, \dots, \mathbf{X}_{mn_m} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu} + \boldsymbol{\tau}_m, \boldsymbol{\Sigma})$
      - 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 $i$th sample $\bar{\mathbf{X}}_i = n_i^{-1} \sum_j \mathbf{X}_{ij}$
  - Sample covariance for the $i$th 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

$$\mathbf{SSP}_\text{w} = \sum_i (n_i - 1)\mathbf{S}_i = \sum_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$$

    - Between-group SSP

$$\mathbf{SSP}_\text{b} = \sum_i n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^\top$$

    - Total (corrected) SSP

$$\mathbf{SSP}_\text{cor} = \sum_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}})(\mathbf{X}_{ij} - \bar{\mathbf{X}})^\top = \mathbf{SSP}_\text{w} + \mathbf{SSP}_\text{b}$$

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

$$SSP_{cor} = \sum_{i,j} (X_{ij} - \bar{X})(X_{ij} - \bar{X})^T$$

$$= \sum_{i,j} (X_{ij} - \bar{X}_i + \bar{X}_i - \bar{X})(X_{ij} - \bar{X}_i + \bar{X}_i - \bar{X})^T$$

$$= \sum_{i,j} \left\{ (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T + (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X})^T + (\bar{X}_i - \bar{X})(X_{ij} - \bar{X}_i)^T + (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T \right\}$$

$$= \underbrace{\sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T}_{= SSP_w} + \underbrace{\sum_{i,j} (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X})^T}_{= ①} + \underbrace{\sum_{i,j} (\bar{X}_i - \bar{X})(X_{ij} - \bar{X}_i)^T}_{= ②} + \underbrace{\sum_{i,j} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T}_{SSP_b}$$

$$① = \sum_i \underbrace{\left\{ \sum_j (X_{ij} - \bar{X}_i) \right\}}_{0_{p\times 1}} (\bar{X}_i - \bar{X})^T = 0_{p\times p}$$

$$② = ①^T = 0_{p\times p}$$

---

- MLE of $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma})$
  - Under $H_0$
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}$ for each $i$
    - $\widehat{\boldsymbol{\Sigma}} = (\sum_i n_i)^{-1} \mathbf{SSP}_{\text{cor}}$
  - Without $H_0$
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}_i = n_i^{-1} \sum_j \mathbf{X}_{ij}$
    - $\widehat{\boldsymbol{\Sigma}} = (\sum_i n_i)^{-1} \mathbf{SSP}_{\text{w}}$
- Likelihood ratio

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

$$\ln L(\mu_1, \cdots, \mu_m, \Sigma)$$
$$= \text{const} - \frac{\sum_i n_i}{2} \ln(\det \Sigma) - \frac{1}{2} \underbrace{\sum_{ij} (x_{ij} - \mu_i)^T \Sigma^{-1} (x_{ij} - \mu_i)}_{①}$$

$$① = \sum_{ij} tr\{\Sigma^{-1}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T\}$$
$$= tr[\Sigma^{-1}\{\sum_{ij}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T\}]$$

So, $\ln L(\bar{x}, \ldots, \bar{x}, (\sum_i n_i)^{-1} SSP_{cor})$
$$= \text{const} - \frac{\sum_i n_i}{2} \ln[\det\{(\sum_i n_i)^{-1} SSP_{cor}\}] - \frac{1}{2} tr\{(\sum_i n_i) SSP_{cor}^{-1} SSP_{cor}\}$$
$$= \text{const} - \frac{\sum_i n_i}{2} \cdot \ln\{(\sum_i n_i)^{-p}\} - \frac{\sum_i n_i}{2} \ln\{\det(SSP_{cor})\} - \frac{\sum_i n_i}{2} \cdot p$$

Similarly, $\ln L(\bar{x}_1, \ldots, \bar{x}_m, (\sum_i n_i)^{-1} SSP_w)$
$$= \text{const} - \frac{\sum_i n_i}{2} \cdot \ln\{(\sum_i n_i)^{-p}\} - \frac{\sum_i n_i}{2} \ln\{\det(SSP_w)\} - \frac{\sum_i n_i}{2} \cdot p$$

Further, likelihood ratio $\lambda$
$$= \frac{L(\bar{x}, \ldots, \bar{x}, (\sum_i n_i)^{-1} SSP_{cor})}{L(\bar{x}_1, \ldots, \bar{x}_m, (\sum_i n_i)^{-1} SSP_w)}$$
$$= \exp\{\ln L(\bar{x}, \cdots, \bar{x}, (\sum_i n_i)^{-1} SSP_{cor}) - \ln L(\bar{x}_1, \cdots, \bar{x}_m, (\sum_i n_i)^{-1} SSP_w)\}$$
$$= \exp[\frac{\sum_i n_i}{2}\{\ln \det(SSP_w) - \ln \det(SSP_{cor})\}]$$
$$= \{\det(SSP_w) / \det(SSP_{cor})\}^{\sum_i n_i/2}$$

- Wilk's lambda test statistic
$$\Lambda = \lambda^{2/\sum_i n_i} = \frac{\det(\mathbf{SSP}_{\mathrm{w}})}{\det(\mathbf{SSP}_{\mathrm{cor}})}$$
  - Under $H_0$: $\Lambda \sim$ Wilk's lambda distribution $\Lambda(\mathbf{\Sigma}, \sum_i n_i - m, m-1)$
    - Since $\mathbf{SSP}_{\mathrm{w}} \sim W_p(\mathbf{\Sigma}, \sum_i n_i - m)$ and $\mathbf{SSP}_{\mathrm{b}} \sim W_p(\mathbf{\Sigma}, m-1)$
    - 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 $\alpha$
$$\Big\{x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{mn_m} : \{(p+m)/2 - \sum_i n_i + 1\} \ln \Lambda \geq \chi^2_{1-\alpha, p(m-1)}\Big\}$$
$$= \left\{x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{mn_m} : \Lambda \leq \exp\Big\{\frac{\chi^2_{1-\alpha,p(m-1)}}{(p+m)/2 - \sum_i n_i + 1}\Big\}\right\}$$

- $p$-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

```
options(digits = 4)
install.packages('car')
tear <- c(
  6.5, 6.2, 5.8, 6.5, 6.5, 6.9, 7.2, 6.9, 6.1, 6.3,
  6.7, 6.6, 7.2, 7.1, 6.8, 7.1, 7.0, 7.2, 7.5, 7.6
)
gloss <- c(
  9.5, 9.9, 9.6, 9.6, 9.2, 9.1, 10.0, 9.9, 9.5, 9.4,
  9.1, 9.3, 8.3, 8.4, 8.5, 9.2, 8.8, 9.7, 10.1, 9.2
)
opacity <- c(
  4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0, 3.9, 1.9, 5.7,
  2.8, 4.1, 3.8, 1.6, 3.4, 8.4, 5.2, 6.9, 2.7, 1.9
)
(X <- cbind(tear, gloss, opacity))
(rate <- factor(gl(2,10,length=nrow(X)), labels=c("Low", "High")))

# Bartlett's approximation to Wilks lambda distribution
X_low <- X[rate == 'Low',]
X_high <- X[rate == 'High',]
n <- nrow(X); p <- ncol(X); m <- 2
SSPcor = (n-1)*cov(X)
SSPw <- (nrow(X_low) - 1)*cov(X_low) + (nrow(X_high) - 1)*cov(X_high)
(Lambda <- det(SSPw)/det(SSPcor))
(cri.point = exp(qchisq(0.95, p*(m-1))/((p+m)/2-n+1)))
Lambda <= cri.point
(p.val = 1-pchisq(((p+m)/2-n+1)*log(Lambda), p*(m-1)))

# Rao's approximation to Wilks lambda distribution
summary(manova(X ~ rate), test = 'Wilks')
summary(car::Manova(lm(X ~ rate)), test.statistic='Wilks')
```

- Report: Testing hypotheses $H_0$: no `RATE` effect on film characteristics v.s. $H_1$: otherwise, we carried on the Wilk's lambda test and obtained 0.4136 as the value of test statistic. The corresponding $p$-value (resp. rejection region) was 0.002227 (resp. $(-\infty, 0.6227]$). So, at the .05 level, there was statistical evidence against $H_0$, i.e., we rejected $H_0$ and believed that there was an effect from `RATE` on film characteristics.