# STAT 3690 Lecture 15

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 04, 2022

## Testing for equality of covariance matrices (J&W Sec. 6.6)

- Model: $m$ independent samples, where
  - $\mathbf{X}_{11}, \ldots, \mathbf{X}_{1n_1} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
  - $\vdots$
  - $\mathbf{X}_{m1}, \ldots, \mathbf{X}_{mn_m} \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$
- Hypotheses $H_0 : \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_m$ v.s. $H_1$ : otherwise
- MLE of $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m)$
  - Under $H_0$
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}_i = n_i^{-1} \sum_j \mathbf{X}_{ij}$
    - $\widehat{\boldsymbol{\Sigma}}_i = (\sum_i n_i)^{-1} \mathbf{SSP}_{\text{w}} = (\sum_i n_i)^{-1} \sum_{ij} (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$ for all $i$
  - No restriction on $\boldsymbol{\Sigma}_i$
    - $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{X}}_i = n_i^{-1} \sum_j \mathbf{X}_{ij}$
    - $\widehat{\boldsymbol{\Sigma}}_i = n_i^{-1}(n_i - 1)\mathbf{S}_i = n_i^{-1} \sum_j (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)^\top$
- Likelihood ratio
$$\lambda = \prod_i \left[ \frac{\det\{n_i^{-1}(n_i - 1)\mathbf{S}_i\}}{\det\{(\sum_i n_i)^{-1}(\sum_i n_i - m)\mathbf{S}_{\text{pool}}\}} \right]^{n_i/2}$$
  - $\mathbf{S}_{\text{pool}} = (\sum_i n_i - m)^{-1} \mathbf{SSP}_{\text{w}}$

$\sum_i n_i = n$

Let $\theta = (\mu_1, \cdots, \mu_m, \Sigma_1, \cdots, \Sigma_m)$

$\Theta_0 = \{\theta : \mu_1 = \cdots = \mu_m, \Sigma_1 > 0, \cdots, \Sigma_m > 0\}$

$\Theta = \{\theta : \mu_1 \in \mathbb{R}, \cdots, \mu_m \in \mathbb{R}, \Sigma_1 > 0, \cdots, \Sigma_m > 0\}$

$$\ln L(\theta) = \text{const} - (n/2) \ln \det \Sigma_i - \underbrace{\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (X_{ij} - \mu_i)^T \Sigma_i^{-1} (X_{ij} - \mu_i)}_{(1)}$$

$$(1) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} tr\{\Sigma_i^{-1} (X_{ij} - \mu_i)(X_{ij} - \mu_i)^T\}$$

$$= tr\left[\sum_{i=1}^{m} \Sigma_i^{-1} \left\{\sum_{j=1}^{n_i} (X_{ij} - \mu_i)(X_{ij} - \mu_i)^T\right\}\right]$$

If we plug $\hat{\mu}_i = \bar{X}_i$ and $\hat{\Sigma}_i = n^{-1} SSP_w$ into (1), then

$$(1) = tr\left[\sum_{i=1} \left(\sum_{i=1}^{m} n_i\right) SSP_w^{-1} \{(n_i - 1) S_i\}\right]$$

$$= tr\left[\left(\sum_{i=1}^{m} n_i\right) \cancel{SSP_w^{-1} \left\{\sum_{i=1} (n_i - 1) S_i\right\}}\right]$$

$$= tr\left\{\left(\sum_{i=1}^{m} n_i\right) I_p\right\} = p \sum_{i=1}^{m} n_i$$

If we plug $\hat{\mu}_i = \bar{X}_i$ and $\hat{\Sigma}_i = n_i^{-1} (n_i - 1) S_i$ into (1), then

$$(1) = tr\left[\sum_{i=1}^{m} n_i \cancel{(n_i - 1)^{-1} S_i^{-1} \{(n_i - 1) S_i\}}\right]$$

$$= tr\left\{\left(\sum_{i=1}^{m} n_i\right) I_p\right\} = p \sum_{i=1}^{m} n_i$$

So, $\sup_{\theta \in \Theta_0} L(\theta) = \exp\left\{\text{const} - \sum_{i=1}^{m} (n_i/2) \ln \det (n^{-1} SSP_w)\right\}$

$$= \exp\left[\text{const} - \sum_{i=1}^{m} (n_i/2) \ln \det \{n^{-1} (n-m) S_{pool}\}\right]$$

$$\sup_{\theta \in \Theta} L(\theta) = \exp\left[\text{const} - \sum_{i=1}^{m} (n_i/2) \ln \det \{n_i^{-1} (n_i - 1) S_i\}\right]$$

$$\Rightarrow \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \exp \sum_{i=1}^{m} (n_i/2) \ln \frac{\det\{n_i^{-1} (n_i - 1) S_i\}}{\det\{n^{-1} (n-m) S_{pool}\}}$$

$$= \prod_{i=1}^{m} \left[\frac{\det\{n_i^{-1} (n_i - 1) S_i\}}{\det\{n^{-1} (n-m) S_{pool}\}}\right]^{n_i/2}$$

---

- Box's M test statistic (a modification of LRT)

$$M = -2 \ln \prod_i \left(\frac{\det \mathbf{S}_i}{\det \mathbf{S}_{\text{pool}}}\right)^{(n_i - 1)/2}$$

- Under $H_0$

$$(1-u)M \approx \chi^2(p(p+1)(m-1)/2)$$

  * $u = \{\sum_i (n_i - 1)^{-1} - (\sum_i n_i - m)^{-1}\}\{6(p+1)(m-1)\}^{-1}(2p^2+3p-1)$

- Rejection region at level $\alpha$

$$\left\{x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{mn_m} : (1-u)M \geq \chi^2_{1-\alpha, p(p+1)(m-1)/2}\right\}$$

- $p$-value

$$1 - F_{\chi^2_{1-\alpha,p(p+1)(m-1)/2}}\{(1-u)M\}$$

---

- Exercise: factors in producing plastic film (continued)
  - Check the equality of covariance matrices for `RATE="Low"` and `RATE="High"`

```r
install.packages('heplots')
options(digits = 4)
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")))
(additive <- factor(gl(2,5,length=nrow(X)), labels=c("Low", "High")))

result = heplots::boxM(lm(X~rate))
result$statistic
result$p.value
```

- Report: Testing hypotheses $H_0$: the covariance matrix does not vary with the level of `RATE` v.s. $H_1$: otherwise, we carried on the Box's M test and obtained 4.017 as the value of test statistic. The corresponding $p$-value was 0.6743. So, at the .05 level, there was no strong statistical evidence against $H_0$, i.e., we did not reject $H_0$ and believed that the covariance matrix does not vary with the level of `RATE`.