HOMOGENEITY OF COVARIANCE

Homogeneity of Covariance Matrix

- Assumptions of MANOVA
  - Normal Distributions
  - Independent Samples
  - Equal Covariance Matrices
- Hypothesis of Equal Covariance Matrices of m p-dimensional multi-normal distributions

$$H_o: \Sigma_1 = \ldots = \Sigma_m$$

Likelihood Function

- Definition:
  - Let $f(x_1,\ldots,x_n;\theta), \theta \in \Theta \subseteq R^k$, be the joint probability (or density) function of *n* random variables $X_1,\ldots,X_n$:
    with sample values $x_1,\ldots,x_n$

$$L(\theta, x_1,\ldots,x_n) = f(x_1,\ldots,x_n;\theta)$$

Likelihood function

- If $X_1,\ldots,X_n$ are <u>discrete</u> iid random variable with probability function $p(x,\theta)$,

then, the **likelihood function** is given by

$$L(\theta) = P(X_1 = x_1,\ldots,X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} P(x_i,\theta)$$

iid: independent and identically distributed

Likelihood function

- For continuous case

$$L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$$

- Let $X_1,\ldots,X_n$ be $N(\mu,\sigma^2)$ iid random variables.

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)=\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

The Likelihood Function

$X \sim f(x, \theta)$ pdf. parameter θ

Likelihood function

$$L(\mathcal{X}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

MLE

$$\widehat{\theta} = \arg\max_{\theta} L(\mathcal{X}; \theta)$$

log-likelihood

$$\ell(\mathcal{X}; \theta) = \log L(\mathcal{X}; \theta)$$

Example Sample $\{x_i\}_{i=1}^n$ from $N_p(\mu, \mathcal{I})$, i.e. from the pdf

$$f(x;\theta) = (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}(x-\theta)^\top (x-\theta)\right\}$$

where $\theta = \mu \in \mathbb{R}^p$ is the mean vector parameter.

The log-likelihood is

$$\ell(\mathcal{X};\theta) = \sum_{i=1}^n \log\{f(x_i;\theta)\} = \log\,(2\pi)^{-np/2} - \frac{1}{2}\sum_{i=1}^n (x_i-\theta)^\top (x_i-\theta).$$

The term $(x_i-\theta)^\top (x_i-\theta)$ equals

$$(x_i-\overline{x})^\top (x_i-\overline{x}) + (\overline{x}-\theta)^\top (\bar{x}-\theta) + 2(\overline{x}-\theta)^\top (x_i-\overline{x}).$$

If we sum up this term over $i = 1, \ldots, n$ we see that

$$\sum_{i=1}^{n} (x_i - \theta)^\top (x_i - \theta) = \sum_{i=1}^{n} (x_i - \overline{x})^\top (x_i - \overline{x}) + n(\overline{x} - \theta)^\top (\overline{x} - \theta).$$

Hence

$$\ell(\mathcal{X}; \theta) = \log(2\pi)^{-np/2} - \frac{1}{2} \sum_{i=1}^{n} (x_i - \overline{x})^\top (x_i - \overline{x}) - \frac{n}{2} (\overline{x} - \theta)^\top (\bar{x} - \theta).$$

Only the last term depends on θ and is obviously maximized for

$$\widehat{\theta} = \widehat{\mu} = \overline{x}.$$

Thus $\overline{x}$ is the MLE.

Example $\{x_i\}_{i=1}^n$ is a sample from a normal distribution $N_p(\mu, \Sigma)$

Due to the symmetry of Σ, the unknown parameter θ is in fact $\{p + \frac{1}{2}p(p+1)\}$-dimensional.

Then

$$L(\mathcal{X};\theta) = |2\pi\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^\top \Sigma^{-1}(x_i-\mu)\right\}$$

and

$$\ell(\mathcal{X};\theta) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{1}{2}\sum_{i=1}^n (x_i-\mu)^\top \Sigma^{-1}(x_i-\mu).$$

The term $(x_i-\mu)^\top \Sigma^{-1}(x_i-\mu)$ equals

$$(x_i-\overline{x})^\top \Sigma^{-1}(x_i-\overline{x}) + (\overline{x}-\mu)^\top \Sigma^{-1}(\bar{x}-\mu) + 2(\overline{x}-\mu)^\top \Sigma^{-1}(x_i-\overline{x}).$$

If we sum up this term over $i = 1, \ldots, n$ we see that

$$\sum_{i=1}^{n} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu)$$

$$= \sum_{i=1}^{n} (x_i - \overline{x})^\top \Sigma^{-1} (x_i - \overline{x}) + n(\overline{x} - \mu)^\top \Sigma^{-1} (\overline{x} - \mu).$$

Note that

$$\begin{aligned} &(x_i - \overline{x})^\top \Sigma^{-1} (x_i - \overline{x}) \\ &= \mathrm{tr}\left\{ (x_i - \overline{x})^\top \Sigma^{-1} (x_i - \overline{x}) \right\} \\ &= \mathrm{tr}\left\{ \Sigma^{-1} (x_i - \overline{x})(x_i - \overline{x})^\top \right\}. \end{aligned}$$

We sum this up over the index i:

$$\sum_{i=1}^{n} (x_i - \mu)^\top \Sigma^{-1} (x_i - \mu)$$

$$= \mathrm{tr}\{\Sigma^{-1} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^\top\} + n(\overline{x} - \mu)^\top \Sigma^{-1} (\overline{x} - \mu)$$

$$= \mathrm{tr}\{\Sigma^{-1} n\mathcal{S}\} + n(\overline{x} - \mu)^\top \Sigma^{-1} (\overline{x} - \mu).$$

Thus the log-likelihood function for $N_p(\mu, \Sigma)$ is

$$\ell(\mathcal{X}; \theta) = -\frac{n}{2} \log|2\pi\Sigma| - \frac{n}{2} \mathrm{tr}\{\Sigma^{-1}\mathcal{S}\} - \frac{n}{2} (\overline{x} - \mu)^\top \Sigma^{-1} (\overline{x} - \mu)$$

We can easily see that the third term would be maximized by $\mu = \bar{x}$.

The MLE's are given by

$$\widehat{\mu} = \overline{x}, \quad \widehat{\Sigma} = \mathcal{S}.$$

Note that the unbiased covariance estimator $\mathcal{S}_u = \frac{n}{n-1}\mathcal{S}$ is not the MLE!

Ex: Let $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$

- If μ is unknown and $\sigma^2 = \sigma_0^2$ is known, find the MLE for μ.
- If $\mu = \mu_0$ is known and σ^2 unknown, find the MLE for σ^2.
- If μ and σ^2 are both unknown, find the MLE for $\theta(\mu,\sigma^2)$.

(a) θ=μ, $\sigma^2 = \sigma_0^2$ is known

- The likelihood function is:

$$L(\theta,\sigma_0^2) = (2\pi\sigma_0^2)^{-n/2}\exp\left(\frac{\sum_{i=1}^{n}(x_i-\theta)^2}{2\sigma_0^2}\right)$$

$$\ell(X;\theta) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S\} - \frac{n}{2}(\bar{X}-\theta)^T\Sigma^{-1}(\bar{X}-\theta)$$

- Only the last term depends on θ and is obviously maximized for

$$\hat{\theta} = \hat{\mu} = \bar{X}$$

(b) $\theta = \Sigma$, $\mu = \mu_0$ is known

- Likelihood function

$$L(\mu,\theta) = (2\pi\theta)^{-n/2}\exp\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta}\right)$$

$$\frac{\partial}{\partial\theta}(\ln L(\mu,\theta)) = -\frac{n}{2\theta} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta^2} = 0$$

$$\Rightarrow \hat{\theta} = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{n}$$

  - The unbiased estimate $\hat{\theta}_u = \frac{n}{n-1}\hat{\theta}$ is NOT MLE.

- (c) When both μ and Σ unknown, we need to differentiate both parameters, and mostly follow the same steps by part (a) and (b).

Likelihood ratio

- Define $L^*_j = \max L(X;\, \theta)$, the maxima of the likelihood for each of the hypotheses.

$$\lambda(X) = \frac{L_0^*}{L_1^*}$$

- Likelihood ratio test
- Rejection region: $R = \{x : \lambda(x) < c\}$

$$\sup_{\theta \in \Omega_0} P_\theta(x \in R) = \alpha$$

Theorem (Wilks)

- If $\Omega_1 \subset R^q$ is a q-dimensional space and if $\Omega_0 \subset \Omega_1$ is an r-dimensional subspace, then under regularity conditions for $n \to \infty$

$$\forall\ \theta \in \Omega_0 : -2 \log \lambda \xrightarrow{\mathcal{L}} \chi^2_{q-r}.$$

Test problem 1

$X_1, \ldots, X_n$, i.i.d. with $X_i \sim N_p(\mu, \Sigma)$

$H_0 : \mu = \mu_0$, $\quad \Sigma$ known, $\qquad H_1 :$ no constraints.

$\Omega_0 = \{\mu_0\}, r = 0, \Omega_1 = \mathbb{R}^p, q = p$

$$\ell(X,\theta) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S\} - \frac{n}{2}(\bar{X}-\mu)^T\Sigma^{-1}(\bar{X}-\mu)$$

$$\ell_0^* = \ell(\mu_0,\Sigma) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S\} - \frac{n}{2}(\bar{X}-\mu_0)^T\Sigma^{-1}(\bar{X}-\mu_0)$$

$$\ell_1^* = \ell(\bar{x},\Sigma) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S\}$$

$$-2\log\lambda = 2(\ell_1^* - \ell_0^*) = n(\bar{x}-\mu_0)^{\top}\Sigma^{-1}(\bar{x}-\mu_0)$$

$$-2\log\lambda \sim \chi_p^2$$

Rejection region R: $\{x \in \mathbb{R}^n \text{ such that } -2\log\lambda > \chi^2_{0.95;p}\}$

$$\ell(X,\theta) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\{\Sigma^{-1}S\} - \frac{n}{2}(\bar{X}-\mu)^T\Sigma^{-1}(\bar{X}-\mu)$$

Test problem 3

$X_i \sim N_p(\mu, \Sigma)$

$H_0 : \Sigma = \Sigma_0,\quad \mu$ unknown, $\qquad H_1 :\quad$ no constraints.

$$\ell_0^* = \ell(\overline{x}, \Sigma_0) = -\frac{1}{2} n \log|2\pi\,\Sigma_0| - \frac{1}{2} n\,\mathrm{tr}(\Sigma_0^{-1}\mathcal{S})$$

$$\ell_1^* = \ell(\overline{x}, \mathcal{S}) = -\frac{1}{2} n \log|2\pi\,\mathcal{S}| - \frac{1}{2} np$$

$$\frac{\partial}{\partial\theta}(\ln L(\mu,\theta)) = -\frac{n}{2\theta} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\theta^2} = 0$$

and thus

$$\begin{aligned} -2\log\lambda &= 2(\ell_1^* - \ell_0^*) \\ &= n\,\mathrm{tr}(\Sigma_0^{-1}\mathcal{S}) - n\log|\Sigma_0^{-1}\mathcal{S}| - np. \end{aligned}$$

Note that this statistic is a function of the eigenvalues of $\Sigma_0^{-1}\mathcal{S}$. Unfortunately, the exact finite sample distribution of $-2\log\lambda$ is very complicated. Asymptotically, we have under H_0

$$-2\log\lambda \xrightarrow{\mathcal{L}} \chi_m^2 \quad \text{as } n \to \infty$$

with $m = \frac{1}{2}\{p(p+1)\}$, since a $(p \times p)$ covariance matrix has only these m parameters as a consequence of its symmetry.

$$-2\log\lambda = n\mathrm{tr}(\Sigma_0^{-1}S) - n\log|\Sigma_0^{-1}S| - np$$

Test equality of covariance

$$H_0 : \Sigma = \Sigma_0$$

$$H_1 : \Sigma \neq \Sigma_0$$

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'$$

$$L = v(\ln|\Sigma_0| - \ln|\hat{\Sigma}| + tr\hat{\Sigma}\Sigma_0^{-1} - p)$$

- $\upsilon = n - 1$, when *n* is large enough, *L* is approximately $\chi 2$ distribution of df=*p*(*p*+1)/2.

- Bartlett 1954 modified the equation as

$$L' = \left[1 - \frac{1}{6v-1}\left(2p + 1 - \frac{2}{p+1}\right)\right] L \sim \chi^2_{p(p+1)/2}$$

$$L' = \left[1 - \frac{1}{6v-1}\left(2p+1-\frac{2}{p+1}\right)\right] L \sim \chi^2_{p(p+1)/2}$$

Example

設一樣品變積矩陣為

$$\hat{\Sigma} = \begin{pmatrix} 3.42 & 2.60 & 1.89 \\ & 8.00 & 6.51 \\ \text{對稱} & & 9.62 \end{pmatrix}$$

其樣品大小 $n = 20$，$p = 3$，其假設檢定為

$$H_0 : \Sigma = \Sigma_0 = \begin{pmatrix} 4 & 3 & 2 \\ & 6 & 5 \\ & & 10 \end{pmatrix}$$

利用 (7.1) 及 (7.2) 式

$$|\Sigma_0| = 86,\ \left|\hat{\Sigma}\right| = 88.6355,\ v = 19,\ tr\hat{\Sigma}\Sigma_0^{-1} = 3.2217$$

$$L = 19(\ln 86 - \ln 88.6355 + 3.2217 - 3) = 3.6388$$

$$L' = \left[1 - \frac{1}{6\times 19-1}\left(2\times 3+1-\frac{2}{3+1}\right)\right] \times 3.6388$$

$$= 0.9425 \times 3.6388 = 3.4296 < \chi^2_{0.05,\,6} = 12.592$$

Homogeneity of Covariance Matrix

- Hypothesis of Equal Covariance Matrices of m p-dimensional multivariate normal distributions

$$H_o: \Sigma_1 = \ldots = \Sigma_m$$

Homogeneity of Covariance Matrix

$\mathbf{S_j}$ be the pxp sample covaraince matrix of sample j

with sample size n_j and degrees of freedom n_j-1

The pooled pxp covariance matrix is given as

$$\mathbf{S} = \frac{1}{\sum_{j=1}^{m}(n_j-1)} \sum_{j=1}^{m}(n_j-1)\mathbf{S_j}$$

Homogeneity of Covariance Matrix

$$\mathrm{M}=\sum_{\mathrm{j}=1}^{\mathrm{m}}(\mathrm{n}_{\mathrm{j}}\text{-}1)\ln\left|\mathbf{S}\right|-\sum_{\mathrm{j}=1}^{\mathrm{m}}(\mathrm{n}_{\mathrm{j}}\text{-}1)\ln\left|\mathbf{S}_{\mathrm{j}}\right|$$

Scale factor

$$C=1-\frac{2p^2+3p\text{-}1}{6(p+1)(m\text{-}1)}\left[\sum_{j=1}^{m}\frac{1}{(n_j\text{-}1)}-\sum_{j=1}^{m}\frac{1}{\sum_{j=1}^{m}(n_j\text{-}1)}\right]$$

利用Box(1953)值公式得為

$$\mathrm{X}=\mathrm{MC}\sim\chi^2_{(\mathrm{m}\text{-}1)\mathrm{p}(\mathrm{p}+1)/2}$$

Homogeneity of Covariance Matrix

- Reaction Times

  32 male and 32 female normal subjects reacted to visual stimuli preceded by warning intervals of different lengths in 0.5 and 0.15 seconds

Homogeneity of Covariance Matrix

The sample covariance matrices for male and female

$$\mathbf{S}_M = \begin{pmatrix} 4.32 & 1.88 \\ & 9.18 \end{pmatrix} \text{ and } \mathbf{S}_F = \begin{pmatrix} 2.52 & 1.90 \\ & 10.06 \end{pmatrix}$$

The pooled covariance matrix

$$\mathbf{S} = \begin{pmatrix} 3.42 & 1.89 \\ & 9.62 \end{pmatrix}$$

$$\begin{aligned} M &= 2(31\text{-}1)\ln|\mathbf{S}| \text{-} (31\text{-}1)[\ln|\mathbf{S}_\mathbf{M}| + \ln|\mathbf{S}_\mathbf{M}|] \\ &= 62\ln(29.328) \text{-} 30[\ln(36.123) + \ln(21.741)] \\ &= 2.82 \end{aligned}$$

$C=0.965$ and $X=2.82\text{x}0.965=2.72 < \chi^2_{0.05,3} = 7.81$ qchisq, 都是用右尾檢定

Fail to reject the null hypothesis of equal covariance matrix

R package

- ## Homogeneity of covariances test
- install.packages("covTestR")
- library(covTestR)
- homogeneityCovariances(x, ..., covTest = BoxesM)