Tutorial 4

CHEN Xiao

Department of Mathematics

March 12, 2020

Review

There are n random samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from the $N_p(\boldsymbol{\mu}, \Sigma)$. Let $Z_j = \mathbf{a}'\boldsymbol{X}_j, j = 1, 2, \ldots, n$ Then the sample mean and variance of the observed values $z_1, z_2, \ldots, z_n$ are

$$\bar{z} = \mathbf{a}'\overline{\mathbf{x}} \quad \text{and} \quad s_z^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean vector and covariance matrix of the $\mathbf{X}_j$'s respectively.

Given a particular $\mathbf{a}$, the confidence interval is that set of $\mathbf{a}'\mu$ values for which

$$|t| = \left| \frac{\sqrt{n}\,(\mathbf{a}'\overline{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})}{\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}}} \right| \leq t_{n-1}(\alpha/2)$$

Simultaneously for all a, the interval

$$\left(\mathbf{a}'\bar{X} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}}, \quad \mathbf{a}'\bar{X} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}} \right)$$

will contain a' μ with probability $1 - \alpha$

Example 1

Perspiration from 20 healthy females was analyzed. Three components, X_1 = sweat rate, X_2 = sodium content, and X_3 = potassium content, were measured. For the data the computer calculation provides the sample mean vector, the sample covariance matrix and its inverse as follows:

$$\bar{\mathbf{x}} = \begin{pmatrix} 4.640 \\ 45.400 \\ 9.965 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} 2.879 & 10.010 & -1.810 \\ 10.010 & 199.788 & -5.640 \\ -1.810 & -5.640 & 3.628 \end{pmatrix}$$

and

$$\mathbf{S}^{-1} = \begin{pmatrix} .586 & -.022 & -.258 \\ -.022 & .006 & -.002 \\ .258 & -.002 & .402 \end{pmatrix}$$

Solve the following two problems under the significant level $\alpha = 0.1$

(a) Construct the simultaneous confidence interval for $2\mu_1 - \mu_3$ and μ_2.
(b) For three variables, testing $H_0 : \boldsymbol{\mu} = (5, 45, 10)'$ against $H_1 : \boldsymbol{\mu} \neq (5, 45, 10)'$.
(Useful formula: the T^2 statistic for a test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ is given by

$$T^2 = n\,(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\,\mathbf{S}^{-1}\,(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$

its corresponding F-statistic is $F = \frac{n-p}{(n-1)p}T^2$. The latter follows $F(p, n-p)$ distribution when the null hypothesis is true.
$F_{3,17}(.10) = 2.44, F_{1,19}(0.10) = 2.9899, t_{19}(0.05) = 1.7921.$)

Solutions

(a) Suppose the 20 random samples $X_1, X_2, \ldots, X_{20}$ came from the $N_3(\mu, \Sigma)$. $\bar{x}$ and S are the sample mean vector and covariance matrix of the X_j 's respectively. Let $\boldsymbol{a_1}' = (2, 0, -1)$, $\boldsymbol{a_2}' = (0, 1, 0)$
The simultaneous confidence interval for $2\mu_1 - \mu_3$ and μ_2.

$$\left(\mathbf{a}_1'\bar{\boldsymbol{x}} - \sqrt{\frac{57}{340}F_{3,17}(0.1)\mathbf{a}_1'\mathbf{S}\mathbf{a}_1}, \quad \mathbf{a}_1'\bar{\boldsymbol{x}} + \sqrt{\frac{57}{340}F_{3,17}(0.1)\mathbf{a}_1'\mathbf{S}\mathbf{a}_1}\right)$$

and

$$\left(\mathbf{a}_2'\bar{\boldsymbol{x}} - \sqrt{\frac{57}{340}F_{3,17}(0.1)\mathbf{a}_2'\mathbf{S}\mathbf{a}_2}, \quad \mathbf{a}_1'\bar{\boldsymbol{x}} + \sqrt{\frac{57}{340}F_{3,17}(0.1)\mathbf{a}_2'\mathbf{S}\mathbf{a}_2}\right)$$

(b)$\boldsymbol{\mu}_0 = (5, 45, 10)'$

$$T^2 = 20\,(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\,\mathbf{S}^{-1}\,(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) = 1.6758$$

$$F = \frac{n-p}{(n-1)p}\,T^2 = \frac{17}{19 \times 3}\,T^2 = 0.4998 < F_{3,17}(.10) = 2.44$$

Thus H_0 can't be rejected.

Comparing Mean Vectors from Two Populations

Result 4.7 If $X_{11}, X_{12}, \ldots, X_{1n_1}$ is a random sample of size n_1 from $N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \ldots, \boldsymbol{X}_{2n_2}$ is an independent random sample size n_2 from $N_p(\boldsymbol{\mu}_2, \Sigma)$, then
$T^2 = \left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]' \left[\frac{1}{n_1} + \frac{1}{n_2} S_{pooled}\right]^{-1} \left[\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)\right]$ is distributed as

$$\frac{(n_1 + n_2 - 2)\, p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1}$$

Consequently,

$$P\left(T^2 \leq c^2\right) = 1 - \alpha$$

where

$$c^2 = \frac{(n_1 + n_2 - 2)\, p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1}(\alpha)$$

Testing for Equality of Covariance Matrics

With g populations, the null hypothesis is

$$H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$$

where Σ_l is the covariance matrix for the l th population, $l = 1, 2, \ldots, g$, and Σ is the presumed common covariance matrix. The alternative hypothesis is that at least two of the covariance matrices are not equal. Here n_l is the sample size for the l th group, $\mathbf{S}_l$ is the l th group sample covariance matrix and $\mathbf{S}_{\text{pool}}$ is the pooled sample covariance matrix given by

$$\mathbf{S}_{pooled} = \frac{1}{\sum_l (n_l - 1)} \left\{ (n_1 - 1)\,\mathbf{S}_1 + (n_2 - 1)\,\mathbf{S}_2 + \cdots + (n_g - 1)\,\mathbf{S}_g \right\}$$

Set

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right]$$

$$M = \left\{\left[\sum_\ell (n_\ell - 1)\right] \ln |\mathbf{S}_{\text{pooled}}| - \sum_\ell \left[(n_\ell - 1) \ln |\mathbf{S}_\ell|\right]\right\}$$

where p is the number of variables and g is the number of groups. Then Box's test statistic

$$C = (1-u)M = (1-u)\left\{\left[\sum_l (n_l - 1)\right] \ln |\mathbf{S}_{\text{pooled}}| - \sum_l \left[(n_l - 1) \ln |\mathbf{S}_l|\right]\right\}$$

has an approximate χ^2 distribution with

$$\nu = g\frac{1}{2}p(p+1) - \frac{1}{2}p(p+1) = \frac{1}{2}p(p+1)(g-1)$$

degrees of freedom. At significance level α, reject H_0 if $C > \chi^2_{p(p+1)(g-1)/2}(\alpha)$

Example 2

Fifty bars of soap are manufactured in each of two ways. Two characteristics, $X_1 =$ lather and $X_2 =$ mildness, are measured. The summary statistics for the bars produced by methods 1 and 2 are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

(a) Testing $H_0 : \Sigma_1 = \Sigma_2$ against $\Sigma_1 \neq \Sigma_2$ at the significant level $\alpha = 0.10$.
(b) Testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ at the significant level $\alpha = 0.10$. where $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are mean vectors for the bars produced by method 1 and 2.Σ_1 and $\boldsymbol{\Sigma}_2$ are their covariance matrices. (Box's test statistic $C = (1 - u)M$ where

$$u = \left[\sum_{\ell} \frac{1}{(n_\ell - 1)} - \frac{1}{\sum_\ell (n_\ell - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right]$$

$$M = \left\{\left[\sum_{\ell} (n_\ell - 1)\right] \ln |\mathbf{S}_{\text{pooled}}| - \sum_{\ell} [(n_\ell - 1) \ln |\mathbf{S}_\ell]\right\}$$

$F_{2,97}(0.10) = 2.3581, \chi^2_3(0.10) = 6.2514\big)$

Soluitons

(a)$n_1 = n_2 = 50, p = 2, g = 2,$

$S_{pooled} = \frac{1}{n_1+n_2-2}[(n_1-1)S_1 + (n_2-1)S_2] = \left[\begin{array}{cc} 2 & 1 \\ 1 & 5 \end{array}\right]$

$u = \left[\frac{1}{n_1-1} + \frac{1}{n_2-1} - \frac{1}{n_1+n_2-2}\right]\left[\frac{2p^2+3p-1}{6(p+1)(g-1)}\right] = 0.0221$

$M = (n_1+n_2-2)\ln|\mathbf{S}_{\text{pooled}}| - (n_1-1)\ln|\mathbf{S}_1| - (n_2-1)\ln|\mathbf{S}_2| = 2.4815$

$p(p+1)(g-1)/2 = 3$

$$C = (1-u)M = 2.4267 < \chi_3^2(0.10) = 6.2514$$

(b)$\Sigma_1 = \Sigma_2$ If $\mu_1 = \mu_2$, then
$T^2 = [\bar{x}_1 - \bar{x}_2]' \left[(\frac{1}{n_1} + \frac{1}{n_2}) S_{\text{pooled}} \right]^{-1} [\bar{x}_1 - \bar{x}_2]$ is

$$\text{distributed as} \quad \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1}$$

$$T^2 = [\bar{x}_1 - \bar{x}_2]' \left[(\frac{1}{n_1} + \frac{1}{n_2}) S_{\text{pooled}} \right]^{-1} [\bar{x}_1 - \bar{x}_2] = 52.4722$$

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)\, p} T^2 = 25.9684 > F_{2,97}(0.10) = 2.3581$$

Then H_0 can be rejected.

Exmaple 3

Using Moody's bond ratings, samples of 20 Aa (middle-high quality) corporate bonds and 20 Baa(top-medium quality) corporate were selected. For each of the corresponding companies, the ratio $X_1 =$ current ratio (a measure of short-term liquidity), $X_2 =$ debt to equity ratio (a measure of financial risk or leverage) were recorded. The summary statistics are as follows:

Aa bound companies:

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.287 \\ .347 \end{bmatrix}, \quad \mathbf{S}_1 = \begin{bmatrix} .459 & -.026 \\ -.026 & .030 \end{bmatrix}, \quad n_1 = 20$$

Baa bond companies:

$$\bar{\mathbf{x}}_2 = \begin{bmatrix} 2.404 \\ .524 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} .944 & .002 \\ .002 & .024 \end{bmatrix}, \quad n_2 = 20$$

(a) Test $H_0 : \Sigma_1 = \Sigma_2$ vs $H_1 : \Sigma_1 \neq \Sigma_2$ at the significant level $\alpha = 0.05$.
(b) Test $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$ vs $H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq 0$ at the significant level $\alpha = 0.05$
(Useful formula: the T^2 statistics for two populations is given by

$$T^2 = [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\text{pooled}} \right]^{-1} [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]$$

when $\Sigma_1 = \Sigma_2$, and when $\Sigma_1 \neq \Sigma_2$

$$T^2 = [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\tfrac{1}{n_1} \mathbf{S}_1 + \tfrac{1}{n_2} \mathbf{S}_2 \right]^{-1} [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]$$
$$\mathbf{S}_{pooled} = \tfrac{n_1 - 1}{n_1 + n_2 - 2} \mathbf{S}_1 + \tfrac{n_2 - 1}{n_1 + n_2 - 2} \mathbf{S}_2$$

$F_{2,37}(0.05) = 3.25, \chi^2_2(0.05) = 5.99, \chi^2_3(0.05) = 7.8147$)

(a) $S_{pooled} = \frac{1}{n_1+n_2-2}[(n_1-1)S_1 + (n_2-1)S_2] = \begin{bmatrix} 0.7015 & -0.012 \\ -0.0120 & 0.027 \end{bmatrix}$

$u = \left[\frac{1}{n_1-1} + \frac{1}{n_2-1} - \frac{1}{n_1+n_2-2}\right]\left[\frac{2p^2+3p-1}{6(p+1)(g-1)}\right] = 0.0570$

$M = (n_1+n_2-2)\ln|\mathbf{S}_{\text{pooled}}| - (n_1-1)\ln|\mathbf{S}_1| - (n_2-1)\ln|\mathbf{S}_2| = 3.3238$

$p(p+1)(g-1)/2 = 3$

$$C = (1-u)M = 3.1344 < \chi^2_3(0.05) = 7.8147$$

(b)$\Sigma_1 = \Sigma_2$ If $\mu_1 - \mu_2 = 0$, then
$T^2 = [\bar{x}_1 - \bar{x}_2]' \left[(\frac{1}{n_1} + \frac{1}{n_2}) S_{\text{pooled}} \right]^{-1} [\bar{x}_1 - \bar{x}_2]$ is

$$\text{distributed as} \quad \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1}$$

$$T^2 = [\bar{x}_1 - \bar{x}_2]' \left[(\frac{1}{n_1} + \frac{1}{n_2}) S_{\text{pooled}} \right]^{-1} [\bar{x}_1 - \bar{x}_2] = 12.15328$$

$F = \frac{n_1+n_2-p-1}{(n_1+n_2-2)p} T^2 = 5.9167 > F_{2,37}(0.05) = 3.25$ Then H_0 can be rejected.

Review

The "distance " of the point $[x_1, x_2, \ldots, x_p]'$ to origin

$$\begin{aligned} (\text{ distance })^2 = & a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{pp}^2 x_p^2 \\ & + 2\left(a_{12}x_1x_2 + a_{13}x_1x_3 + \ldots + a_{p-1,p}x_{p-1}x_p\right) \end{aligned}$$

A geometric interpretation based on the eigenvalues and eigenvectors of the matrix A.

For example, suppose $p = 2$, Then the points $\mathbf{x}' = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}^2 + 2a_{12}x_1x_2 = c^2$$

By the spectral decomposition,

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2'$$

so

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1 \left(\mathbf{x}'\mathbf{e}_1\right)^2 + \lambda_2 \left(\mathbf{x}'\mathbf{e}_2\right)^2$$

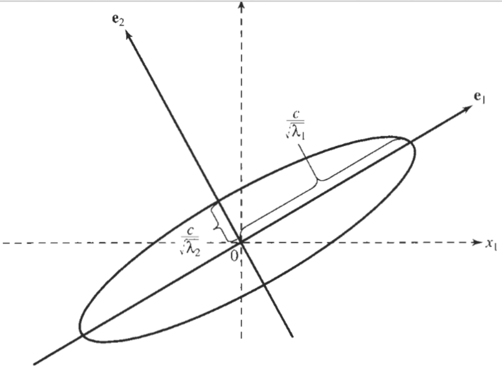

Figure 2.6 Points a constant distance c from the origin ($p = 2, 1 \leq \lambda_1 < \lambda_2$).

Assignment 1 5

Consider the sets of points (x_1, x_2) whose "distance" from the origin are given by

$$c^2 = 4x_1^2 + 3x_2^2 + 2\sqrt{2}x_1x_2$$

for $c^2 = 1$ and for $c^2 = 4$. Determine the major and minor axes of the ellipse of constant distances and their associated length. Sketch the ellipse of constant distances and comment on their positions.

Solutions

$$c^2 = 4x_1^2 + 3x_2^2 + 2\sqrt{2}x_1x_2 = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 4 & \sqrt{2} \\ \sqrt{2} & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Let $A = \begin{pmatrix} 4 & \sqrt{2} \\ \sqrt{2} & 3 \end{pmatrix}$, then $|A - \lambda I| = (\lambda - 2)(\lambda - 5), \quad \lambda_1 = 2, \lambda_2 = 5$

When $\lambda = 2, \quad e_1 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$

When $\lambda = 5, \quad e_2 = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$

For $\mathrm{c}^2 = 1$, major axe is $\dfrac{1}{\sqrt{2}} \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$, minor axe is $\dfrac{1}{\sqrt{5}} \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$

For $\mathrm{c}^2 = 4$, major axe is $\sqrt{2}\begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$, minor axe is $\frac{2}{\sqrt{5}}\begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$ Let $\boldsymbol{x} = (x_1, x_2)'$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1 \left(\mathbf{x}'\mathbf{e}_1\right)^2 + \lambda_2 \left(\mathbf{x}'\mathbf{e}_2\right)^2 = c^2$$

Assume that the original coordinate axes are rotated through an angle of θ

$$\begin{aligned} \mathrm{x}_\theta &= x_1 \cos\theta + x_2 \sin\theta \\ \mathrm{y}_\theta &= -x_1 \sin\theta + x_2 \cos\theta \end{aligned}$$

Let $\mathrm{x}_\theta = x_1 \cos\theta + x_2 \sin\theta = \mathbf{x}'\mathbf{e}_1 = x_1 \cos\frac{1}{\sqrt{3}} + x_2 \sin\frac{\sqrt{2}}{\sqrt{3}}$

$$\theta = \arccos(-\frac{1}{\sqrt{3}}) = 0.6959$$

Thus, the ellipse rotates an angle between 90 and 180 from the original axes.

Are the following distance functions valid for the distance from the origin ? Explain. (a) $x_1^2 + 4x_2^2 + x_1x_2 = (\text{ distance })^2$ (b) $x_1^2 - 2x_2^2 = (\text{ distance })^2$

Soluitons

The points $\mathbf{x}' = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}^2 + 2a_{12}x_1x_2 = c^2$$

If A is a positive define matrix, the distant function is valid.

(a) Because $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 4 \end{pmatrix}$ is a positive define matrix, and the origin is $(0, 0)$, so the distant function is valid.

(b) Because $\begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$ is not a positive define matrix, so the distant function is not valid.