Mean Vectors and Variance-Covariance Matrices

Inference for 1 and 2 Samples

Univariate Statistics

Let $Y_1, Y_2, \ldots, Y_N \sim$ *iid* $N(\mu, \sigma^2)$

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$$

Univariate Test for σ²

$$\mathrm{H}_0 : \sigma^2 = \sigma_0^2$$

$$\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$$

Test Statistic: $$\mathrm{X}^2 = \frac{(N-1)s^2}{\sigma_0^2} \sim \chi^2_{N-1}$$

This test is highly sensitive to violations of the assumption of normality.

Multivariate Statistics

Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N \sim$ *iid* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\bar{\mathbf{Y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_i$$

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right)\left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right)'$$

Multivariate Test for Σ

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$$
$$H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0$$

LRT Test given in Johnson (1998), p. 398

This test is highly sensitive to violations of the assumption of multivariate normality.

Test for Sphericity

$$H_0 : \mathbf{\Sigma} = \sigma^2 \mathbf{I}$$
$$H_1 : \mathbf{\Sigma} \neq \sigma^2 \mathbf{I}$$

LRT Test produced by PROC FACTOR using METHOD=ML option

This test is highly sensitive to violations of the assumption of multivariate normality.

Test for Compound Symmetry

$$H_0 : \boldsymbol{\Sigma} = \sigma^2 \left[(1-\rho)\mathbf{I} + \rho\mathbf{J}\right] = \sigma^2 \begin{pmatrix} 1 & \dots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}$$

$$H_1 : \sim H_0$$

LRT Test given in Johnson (1998), p. 403

This test is highly sensitive to violations of the assumption of multivariate normality.

Huynh-Feldt Conditions

$$H_0 : \boldsymbol{\Sigma} = \eta\mathbf{I} + \boldsymbol{\gamma}\mathbf{1}' + \mathbf{1}\boldsymbol{\gamma}'$$

$$H_1 : \boldsymbol{\Sigma} \neq \eta\mathbf{I} + \boldsymbol{\gamma}\mathbf{1}' + \mathbf{1}\boldsymbol{\gamma}'$$

Let $\mathbf{C}$ be a p×(p-1) matrix of orthogonal constrasts such that

$$\mathbf{C}'\mathbf{C} = \mathbf{I}_{p-1}$$

Huynh-Feldt Conditions

Then Σ satisfies the Huynh-Feldt conditions if and only if

$$\boldsymbol{\Sigma}^* = \mathbf{C}'\boldsymbol{\Sigma}\mathbf{C} = \eta\mathbf{I}$$

i.e., Σ* meets the sphericity condition

Huynh-Feldt Conditions

- A test for the Huynh-Feldt conditions can be obtained from PROC GLM with the PRINTE option in the REPEATED statement. The appropriate test is the Sphericity Test associated with the Orthogonal Components.
- This test is highly sensitive to violations of the assumption of multivariate normality.

Implications

Note that:

Sphericity $\Rightarrow$ Compound Symmetry
$\Rightarrow$ Huynh-Feldt Conditions

Test for Independence

$$H_0 : \boldsymbol{\Sigma} = diag\left(\sigma_1^2 \quad \cdots \quad \sigma_p^2\right) = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2 \end{pmatrix}$$

$$H_1 : \sim H_0$$

LRT Test produced by PROC FACTOR using METHOD=ML option

This test is highly sensitive to violations of the assumption of multivariate normality.

Equality of Several Variance-Covariance Matrices

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_m$$
$$H_1 : \sim H_0$$

LRT Test produced by PROC DISCRIM using POOL=TEST option

This test is highly sensitive to violations of the assumptions of multivariate normality of the observations in each group.

Univariate Inference about μ

Assume $X_1, X_2, \cdots, X_N \sim iid\ N\left(\mu, \sigma^2\right)$

Want to test: $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

Test statistic: $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{N}} \sim t_{N-1}$

An Alternative Test Statistic

$$t^2 = \frac{\left(\bar{X} - \mu_0\right)^2}{s^2 / N}$$

$$= N\left(\bar{X} - \mu_0\right)\left(s^2\right)^{-1}\left(\bar{X} - \mu_0\right)$$

$$\sim F_{1,N-1}$$

100(1-α)% Confidence Interval for μ

$$\bar{X} \pm t_{N-1,\alpha/2} \frac{s}{\sqrt{N}}$$

Multivariate Inference about μ

Assume $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N \sim iid\ \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Want to test:

$$\mathrm{H}_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

$$\mathrm{H}_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

Test statistic:

$$\mathrm{T}^2 = N\left(\bar{\mathbf{X}} - \boldsymbol{\mu}_0\right)'\left(\mathbf{S}\right)^{-1}\left(\bar{\mathbf{X}} - \boldsymbol{\mu}_0\right)$$

$$\sim \text{Hotelling's } \mathrm{T}^2_{p,N-1}$$

$$\sim \frac{(N-1)p}{N-p}\mathrm{F}_{p,N-p}$$

Simultaneous Confidence Statements

We wish to construct a 100(1-α)% Confidence Interval for $\mathbf{a}'\boldsymbol{\mu}$, true for <u>all</u> $\mathbf{a}$:

Method 1:

$$\mathbf{a}'\bar{\mathbf{X}} \pm \sqrt{\left[\frac{p(N-1)}{N-p}\mathrm{F}_{p,N-p;\alpha}\right]\left[\frac{1}{N}\mathbf{a}'\mathbf{S}\mathbf{a}\right]}$$

Simultaneous Confidence Statements

Method 2: Use Bonferroni's Inequality

For a fixed (small) number of $\mathbf{a}$'s,

$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_r$

$$\mathbf{a}_i'\bar{\mathbf{X}} \pm t_{N-1;\alpha/2r}\sqrt{\frac{1}{N}\mathbf{a}'\mathbf{S}\mathbf{a}}$$

for $i = 1, \cdots, r$

The Two-Sample Problem

$$\mathbf{X}_{11}, \mathbf{X}_{12}, \cdots, \mathbf{X}_{1N_1} \sim N_p\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right)$$

$$\mathbf{X}_{21}, \mathbf{X}_{22}, \cdots, \mathbf{X}_{2N_2} \sim N_p\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\right)$$

$$\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

$$\mathrm{H}_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

Case 1: Equal Σ's

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$

then a pooled estimate of $\boldsymbol{\Sigma}$ is

$$\mathbf{S}_p = \frac{(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2}{N_1 + N_2 - 2}$$

Case 1: Equal Σ's

Test Statistic:

$$\mathrm{T}^2 = \frac{N_1 N_2}{N_1 + N_2}\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\right)' \mathbf{S}_p^{-1}\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\right)$$

where $\dfrac{N_1 + N_2 - p - 1}{p\left(N_1 + N_2 - 2\right)}\mathrm{T}^2 \sim \mathrm{F}_{p,\, N_1+N_2-p-1}$

Case 2: Unequal Σ's

A test statistic valid for large samples is:

$$U = \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right)' \left[\frac{1}{N_1}\mathbf{S}_1 + \frac{1}{N_2}\mathbf{S}_2\right]^{-1} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right)$$

$$\sim \chi^2_p$$

Simultaneous Confidence Statements

We wish to construct a 100(1-α)% Confidence Interval for $\mathbf{a}'(\mu_1 - \mu_2)$,

true for <u>all</u> $\mathbf{a}$:

Method 1: using Hotelling's T^2

$$\mathbf{a}'\left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right) \pm c_0 \sqrt{\left[\frac{N_1 + N_2}{N_1 N_2}\mathbf{a}'\mathbf{S}_p\mathbf{a}\right]}$$

where $c_0 = \left[\frac{p(N_1 + N_2 - 2)}{N_1 + N_2 - p - 1} F_{p, N_1+N_2-p-1;\alpha}\right]$

Simultaneous Confidence Statements

Method 2: Using Bonferroni's Inequality

For a fixed (small) number of $\mathbf{a}$'s,

$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$

$$\mathbf{a}_j'\left(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2\right) \pm c_0 \sqrt{\left[\frac{N_1 + N_2}{N_1 N_2}\mathbf{a}'\mathbf{S}_p\mathbf{a}\right]}$$

where $c_0 = t_{N_1+N_2-2;\alpha/2m}$

Paired Samples

Observation	Time 1	Time 2	Difference
1	$\mathbf{y}_{11}$	$\mathbf{y}_{12}$	$\mathbf{d}_1$
2	$\mathbf{y}_{21}$	$\mathbf{y}_{22}$	$\mathbf{d}_2$
⋮	⋮	⋮	⋮
N	$\mathbf{y}_{N1}$	$\mathbf{y}_{N2}$	$\mathbf{d}_N$

Paired Samples - Hypothesis

The hypothesis

$H_0:\mu_1=\mu_2$

is equivalent to

$H_0:\delta=0$

where

$\delta=\mu_1-\mu_2$

Paired Samples – Test Statistic

- Let $\mathbf{d}_i=\mathbf{y}_{i1}-\mathbf{y}_{i2}$
- This is now just a one sample problem with the **d**'*s*