L21: One-way MANOVA

1. One-way MANOVA model, samples and basic statistics

   (1) Model and parameters
   Responses $\mathbf{y} \in R^p$ to q levels of a factor form q populations $N(\mu_i, \Sigma)$, $i = 1, ..., q$ with unknown parameter vectors $\mu_i \in R^p$, $i = 1, .., q$, and unknown positive definite matrix $\Sigma \in R^{p\times p}$.
   Let $\Theta' = (\mu_1, .., \mu_q) \in R^{p\times q}$ and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}$ where x_i is the indicator for the ith level of the factor, i.e., $x_i = \begin{cases} 1 & y \text{ is to the } i\text{th level of the factor} \\ 0 & \text{otherwise} \end{cases}$. Then

   $$y = \mu_1 x_1 + \cdots + \mu_q x_q + \epsilon = \Theta' x + \epsilon \text{ with } \epsilon \sim N(0, \Sigma).$$

   (2) Samples
   $y_i \in R^p$ is observed when $x = x_i \in R^q$. With $Y' = (y_1, .., y_n) \in R^{p\times n}$, $X' = (x_1, .., x_n) \in R^{q\times n}$ and $\mathcal{E} = (\epsilon_1, .., \epsilon_n) \in R^{p\times n}$,

   $$Y' = \Theta' X' + \mathcal{E} \sim N_{p\times n}(\Theta' X', \Sigma, I_n) \Longleftrightarrow Y = X\Theta + \mathcal{E}' \sim N_{n\times p}(X\Theta, I_n, \Sigma).$$

   (3) Basic statistics
   From q samples

Factor	Level 1	$\cdots$	Level q
Samples	$y_{1j} \in R^p$, $j = 1, .., n_1$	$\cdots$	$y_{qj} \in R^p$, $j = 1, .., n_q$
Statistics	n_1, $\overline{y}_1 \in R^p$, $\text{CSSCP}_1 \in R^{p\times p}$	$\cdots$	n_q, $\overline{y}_q \in R^p$, $\text{CSSCP}_q \in R^{p\times p}$

   From the pooled sample

Pooled Sample:	$y_{ij} \in R^p$, $i = 1, .., q$, $j = 1, .., n_i$
Statistics:	$n = n_1 + \cdots + n_q$, $\overline{y} = \frac{\sum_i \sum_j y_{ij}}{n} \in R^p$, $\text{CSSCP}_{pooled} \in R^{p\times p}$

   Ex1: Find CSSCP_{pooled} and CSSCP_i.

```
data a;
    infile "D:\ex.txt";
    input y1 y2  level $ @@;
proc corr CSSCP nocorr;
    var y1 y2;
    run;
proc sort;
   by level;
   run
proc corr CSSCP nocorr;
   var y1 y2 y3;
   by level;
   run;
```

2. Point estimators in one-way MANOVA

   (1) Framework of regression
   By regression data $Y = XB + \mathcal{E}' \sim N_{n\times p}(XB, I_n, \Sigma) \Longleftrightarrow Y' = B'X' + \mathcal{E} \sim N_{p\times n}(B'X', \Sigma, I_n)$, MLE for B, $\widehat{B} = (X'X)^{-1}X'Y$ is independent to error matrix $E = Y'[I - X(X'X)^{-1}X']Y$. $S = \frac{E}{n}$ is MLE for Σ and $S_u = \frac{E}{n-q}$ is an UE for Σ. With one-way MANOVA data

   $$Y = X\Theta + \mathcal{E}' \sim N_{n\times p}(XB, I_n, \Sigma) \Longleftrightarrow Y' = \Theta' X' + \mathcal{E} \sim N_{p\times n}(\Theta' X', \Sigma, I_n),$$

   $\widehat{\Theta} = (X'X)^{-1}X'Y \sim N_{q\times p}(\Theta, (X'X)^{-1}, \Sigma)$ is MLE for Θ and is an UE. $\widehat{\Theta}$ is independent to $E = Y'[I - X(X'X)^{-1}X']Y \sim W_{p\times p}(\Sigma, n - q)$. $S = \frac{E}{n}$ and $S_u = \frac{E}{n-p}$ are MLE and UE for Σ.

(2) $\widehat{\Theta}' = (\overline{y}_1, ..., \overline{y}_q)$

The columns of X are the values of q indicators, for example $X = \begin{pmatrix} 1_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1_{n_q} \end{pmatrix} \in R^{n\times q}$.

So $X'X = \text{diag}(n_1, ..., n_q)$ is sample size matrix and $(X'X)^{-1} = \text{diag}\left(\frac{1}{n_1}, ..., \frac{1}{n_q}\right)$.

With $Y' = (y_{11}, ..., y_{1n_1}, ..., y_{qn_q}) \in R^{p\times n}$, $Y'X = (\sum_j y_{1j}, \cdots, \sum_j y_{qj}) \in R^{p\times q}$.

Thus $\widehat{\Theta}' = Y'X(X'X)^{-1} = (\overline{y}_1, .., \overline{y}_q) \in R^{p\times q}$.

(3) Matrix $E = \text{CSSCP}_1 + \cdots + \text{CSSCP}_q$

$$\begin{aligned} Y'[I_n - X(X'X)^{-1}X'] &= Y' - (\overline{y}_1, .., \overline{y}_q)X' = (y_{11}, .., y_{1n_1}, .., y_{qn_q}) - (\overline{y}_1, .., \overline{y}_1, .., \overline{y}_q) \\ &= (y_{11} - \overline{y}_1, .., y_{1n_1} - \overline{y}_1, .., y_{qn_q} - \overline{y}_q) \in R^{p\times n} \end{aligned}.$$

$$\begin{aligned} E &= Y'[I - X(X'X)^{-1}X']Y = \{Y'[I - X(X'X)^{-1}X']\}\{Y'[I - X(X'X)^{-1}X']\}' \\ &= \text{CSSCP}_1 + \cdots + \text{CSSCP}_q. \end{aligned}$$

Comments: Matrix E measures the total variation-covariation of respose within groups and hence is also denoted as W. Both $\widehat{\Theta}'$ and matrix $E = W$ can be obtained from the SAS in Ex1.

3. Testing the usefulness of the model

(1) General LRT in one-way ANOVA

In regression suppose the reduced model by H_0 has error matrix E_r. Denote $E_r - E$ as H. Then

H_0 : _____ versus H_a : _____
Test statistic: Wilk's Lambda $\Lambda = \frac{\lvert E\rvert}{\lvert E+H\rvert}$
p-value $P(\Lambda \le \Lambda_{ob} \mid H_0)$

is a LRT scheme. Tests in one-way ANOVA fit the framework.

(2) LRT for testing the usefulness of the MANOVA model

H_0 : The model is useless $\Longleftrightarrow H_0 : \mu_1 = \cdots = \mu_q$.

Under H_0 the model becomes $Y = 1_n\mu_{\cdot} + \mathcal{E}'$ with $E_r = Y'\left(I - \frac{1_n 1_n'}{n}\right)Y = \text{CSSCP}_{pooled}$. This E_r is also denoted as T for total variation-covariation. $H = E_r - E = T - W$ measures the variation-covariation between groups and is denoted by B. Thus the MANOVA table can be obtained.

Source	Matrix	DF
Model, Between	$H = B = Y'\left[X(X'X)^{-1}X' - \frac{1_n 1_n'}{n}\right]Y$	$q-1$
Error, Within	$E = W = Y'[I_n - X(X'X)^{-1}X']Y$	$n-q$
Total	$E_r = T = Y'\left(I_n - \frac{1_n 1_n'}{n}\right)Y$	$n-1$

Comment: The matrices in MANOVA table can be quickly filled with basic statistics in (3) of 1. $E_r = T = \text{CSSCP}_{pooled}$ with DF $n-1$; $E = W = \text{CSSCP}_1 + \cdots + \text{CSSCP}_q$ with DF $n-q$. $H = B = T - E$ with DF $q-1$.

This ANOVA table reveals the relation of the matrices, but is less useful in testing since the determinants of the matrices, but not the matrices in the table are actually used. $\Lambda = \frac{\lvert E\rvert}{\lvert E+H\rvert}$.

L22 Tests in one-way MANOVA

1. Implementation of test on the usefulness of one-way MANOVA

   $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ is the response to a factor with 5 levels. To implement

   > $H_0: \mu_i = \mu_j$ for all $i, j = 1, 2, 3, 4, 5$ versus $H_a: \mu_i \neq \mu_j$ for some i, $j = 1, 2, 3, 4, 5$
   > Test statistic: $\Lambda = \frac{|E|}{|E+H|}$
   > p-value: $P(\Lambda < \Lambda_{ob}|H_0)$

   we need Λ_{ob} and p-value.

   (1) By proc reg
   Create 5 indicators for 5 levels

   ```
   proc reg;
       model y1 y2=x1 x2 x3 x4 x5/noint noprint;
       mtest x1-x2, x2-x3, x3-x4, x4-x5/print;
       run;
   ```

   displays matrices E, H, value of Λ and p-value calculated by F-distribution. This method illustrate that the essence of ANOVA is regression. But implementation is tedious.

   (2) By proc anova
   Suppose variable level has 5 values identify 5 levels of the factor.

   ```
   proc anova;
       class level;
       model y1 y2=level/nouni;
       manova h=level/printe printh;
       run;
   ```

   displays matrices E, H, value of Λ and p-value calculated by F-distribution.

   (3 By proc glm
   Both regression and anova are linear models. proc glm (general linear model) can do some of the things proc reg and proc anova can do.

   ```
   proc glm;
       class level;
       model y1 y2=level/nouni;
       manova h=level/printe printh;
       run;
   ```

   displays matrices E, H, value of Λ and p-value calculated by F-distribution.

2. Contrast tests

   (1) Contrast test
   After $H_0: \mu_1 = \cdots = \mu_5$ is rejected one may want to check the equivalent groups. For example $H_0: \mu_1 = \mu_3 = \mu_5$ and $\mu_2 = \mu_4$. This H_0 can be written as $H_0: \mu_1 - \mu_3 = 0$, $\mu_3 - \mu_5 = 0$ and $\mu_2 - \mu_4 = 0$ or equivalently $H_0: \mu_1 - \mu_3 = 0$, $\mu_1 + \mu_3 - 2\mu_5 = 0$ and $\mu_2 - \mu_4 = 0$. The test is called a contrast test since $\mu_1 - \mu_3$, $\mu_1 + \mu_3 - 2\mu - 5$ and $\mu_2 - \mu_4$ are all contrasts of $\mu_1, .., \mu_5$, linear combinations of $\mu_1, .., \mu_5$ with coefficients summed to 0.

H_0 : $\mu_1 - \mu_3 = 0$, $\mu_3 - \mu_5 = 0$ and $\mu_2 = \mu_4$ versus
H_a : at least one equation in H_0 is false
Test statistic: Wilk's Lambda $\Lambda = \frac{|E|}{|E+H|}$
p-value: $P(\Lambda \leq \Lambda_{ob}|H_0)$

We need Λ_{ob} and p-value.

(2) Implementation by contrast statement in proc glm

```
proc glm;
     class level;
     model y1 y2=level/nouni;
     contrast "two groups" level 1 0 -1 0 0,
           level 0 0 1 0 -1, level 0 1 0 -1 0;
     manova h=level/printh printe;
     run;
```

displays E, H, Λ and p-values

3. Two-sample test

(1) T^2-test and Wilk's Lambda test
For $H_0 : \mu_x = \mu_y$ versus $H_a : \mu_x \neq \mu_y$, there is a scheme using

$$T^2 = (\overline{x} - \overline{y})' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_u \right]^{-1} (\overline{x} - \overline{y}) \text{ with p-value } P(T^2(p,\, n-2) > T^2_{ob});$$

and a LRT using Wilk's Lambda

$$\Lambda = \frac{|E|}{|E+H|} \text{ with p-value } P\left(F(p,\, n-p-1) > \frac{n-p-1}{(n-2)p} T^2_{ob} \right).$$

(2) $\Lambda = \left(1 + \frac{T^2}{n-2}\right)^{-1}$

Proof: Sketch. With $q = 2$, $E_r = E + \frac{n_1 n_2}{n}(\overline{x} - \overline{y})(\overline{x} - \overline{y})'$. So

$$\begin{aligned} 1 \cdot \left| E + \tfrac{n_1 n_2}{n}(\overline{x} - \overline{y})(\overline{x} - \overline{y})' \right| &= \begin{vmatrix} 1 & \sqrt{\frac{n_1 n_2}{n}}(\overline{x} - \overline{y})' \\ -\sqrt{\frac{n_1 n_2}{n}}(\overline{x} - \overline{y}) & E \end{vmatrix} \\ &= |E| \cdot \left[1 + \tfrac{n_1 n_2}{n}(\overline{x} - \overline{y})' E^{-1} (\overline{x} - \overline{y}) \right]. \end{aligned}$$

Thus $|E_r| = |E| \cdot \left(1 + \frac{T^2}{n-2}\right)$. Hence $\Lambda = \frac{|E|}{|E+H|} = \left(1 + \frac{T^2}{n-2}\right)^{-1} \Longleftrightarrow T^2 = \left(\frac{1}{\Lambda} - 1\right)(n-2)$.

Comment: While T^2-test is textbook version of the test, the implementation can be through MANOVA by using proc anova or proc glm.