L26: One-way MANOVA

1. One-way MANOVA model

   (1) An experiment
   A factor; q-levels, q-treatments; q-random response
   Univariate: $N(\mu_i, \sigma^2)$, $i = 1, ..., q$. Multivariate $N_p(\theta_i, \Sigma)$.
   Example: School district; Three specific school district; SAT scores; $\begin{pmatrix} \text{SAT} \\ \text{Household income} \end{pmatrix}$

   (2) Formulation for the populations
   $y = \theta_1 x_1 + \cdots + \theta_q x_q + \epsilon$. Here $x_i = \begin{cases} 1 & \text{level } i \\ 0 & \text{otherwise} \end{cases}$, $\epsilon \sim N(0, \Sigma)$.
   $y = (\theta_1, ..., \theta_q) \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} + \epsilon$, $\epsilon \sim N(0, \Sigma)$. Thus $y = \Theta x + \epsilon$.

   (3) Formulation for data
   The n columns of $Y \in R^{p\times n}$ are n observed response y with corresponding n values of x being the n columns of $X \in R^{q\times n}$. Then $Y \sim N_{p\times n}(\Theta X, \Sigma, I_n)$.

   Comment: One-way MANOVA fits the frame work of multivariate regression, $y = \beta x + \epsilon$ and $Y \sim N_{p\times n}(\beta X, \Sigma, I_n)$.

2. Estimating Θ and Σ

   (1) Estimator of Θ
   By the framework for regression, the LSE and MLE of Θ is $\widehat{\Theta} = YX'(XX')^{-1}$.
   Note that $Y \sim N_{p\times n}(\Theta X, \Sigma, I_n)$ represents q samples from the responses to q-treatments. Let n_i, $\overline{y}_i$ and CSSCP$_i$ be from the ith sample. Then $YX' \in R^{p\times q}$ gives the summations of y in q samples, $XX' = \text{diag}(n_1, .., n_q)$. Hence $\widehat{\Theta} = (\overline{y}_1, ..., \overline{y}_q)$.

   (2) Matrix E
   $E = Y[I - X'(XX')^{-1}X]Y' = \sum_{i=1}^{q} \text{CSSCP}_i$ gives the summation of variations within q samples and hence is often denoted by W in ANOVA.
   ANOVA specifies q different treatments. Thus the variations between treatments are the ones expected by the model. But the variations within samples are regarded as Errors.

   (3) Estimators for Σ
   By the framework from regression, $S_p = \frac{E}{n-q}$ is an UE for Σ, and $\frac{E}{n}$ is MLE for Σ.

   Comments: $\widehat{\Theta} \sim N_{p\times q}(\Theta, \Sigma, (XX')^{-1})$ and $E \sim W_{p\times p}(n - q, \Sigma)$ are independent.
   $L(\Theta, \Sigma) \le L\left(\overline{Y}, \frac{E}{n}\right) = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}$.

3. Global F-test

   (1) H_0 and $E_r = W + B$
   H_0 : The factor is not effective $\Longleftrightarrow H_0 : \theta_i = \theta_j$ for all i, j
   Under H_0, $Y \sim N_{p\times n}(\theta_1 1_n', \Sigma, I_n)$ is one sample from $N(\theta_1, \Sigma)$ with sample size n, sample mean $\overline{y}$ and CSSCP. Here $E_r = CSSCP$ is also denoted by T for total variation-covariation in samples. Write $E_r = T = E + H = W + B$. Here $H = B$ is for the variation-covariation between samples.

   (2) LRT
   By the framework from regression

$H_0 : \theta_i = \theta_j$ for all i, j versus $H_a : \theta_i = \theta_j$ for some $i \ne j$
Test statistic: $\Lambda = \frac{\lvert W\rvert}{\lvert W+B\rvert}$
Reject H_0 if $\Lambda < c$

Here c is determined by $P(\Lambda < c|H_0) \leq \alpha$.
By p-value,

$H_0: \theta_i = \theta_j$ for all i, j versus $H_a: \theta_i = \theta_j$ for some $i \neq j$ Test statistic: $\Lambda = \frac{\vert E\vert}{\vert E+H\vert}$ p-value: $P(\Lambda < \Lambda_{ob}\vert H_0)$.

(3) Implementation

(i) Data: Ex6.9 p304

Treatment	A	B	C
Response	$\begin{pmatrix}9\\3\end{pmatrix}, \begin{pmatrix}6\\2\end{pmatrix}, \begin{pmatrix}9\\7\end{pmatrix}$	$\begin{pmatrix}0\\4\end{pmatrix}, \begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}3\\8\end{pmatrix}, \begin{pmatrix}1\\9\end{pmatrix}, \begin{pmatrix}2\\7\end{pmatrix}$

(ii) SAS

```
data a;
  input y1 y2 id $ @@;
  datalines;
  9 3  A 6 2  A
  9 7  A 0 4  B
  2 0  B 3 8  C
  1 9  C 2 7  C
  ;
```

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

(iii) Output

$$E = \begin{pmatrix}10 & 1\\ 1 & 24\end{pmatrix} \text{ and } H = \begin{pmatrix}78 & -12\\ -12 & 48\end{pmatrix}$$

```
 Statistic                Value    F Value Num DF Den DF Pr > F

Wilks' Lambda           0.03845535  8.20    4      8     0.0062
Pillai's Trace          1.54078842  8.39    4     10     0.0031
Hotelling-Lawley Trace 9.94142259  9.94    4      4     0.0235
Roy's Greatest Root    8.07638502 20.19    2      5     0.0040
```

(iv) Test report

$H_0: \alpha_i = 0$ for all i vs $H_a: \alpha_i \neq 0$ for some i Test Statistic: $\Lambda = \frac{\vert W\vert}{\vert B+W\vert}$ p-value: $P(\Lambda \leq \Lambda_{ob}\vert H_0)$. $\Lambda = 0.03846$ p-value: $P(\Lambda < 0.03846\vert H_0) \approx P(F(4, 8) > 8.20) = 0.0062$ Reject H_0. The model is useful.

Comment: Replacing SAS "proc anova;" by "proc glm;" produces the same output.

L27 SAS for MANOVA

1. SAS for one-way MANOVA

   (1) Data in an example
   We study how to use SAS for one-way ANOVA via an example with $p = 2$, $q = 3$, $n_1 = 3$, $n_2 = 2$ and $n_3 = 3$.

Treatment	A	B	C
Response	$\begin{pmatrix}9\\3\end{pmatrix}, \begin{pmatrix}6\\2\end{pmatrix}, \begin{pmatrix}9\\7\end{pmatrix}$	$\begin{pmatrix}0\\4\end{pmatrix}, \begin{pmatrix}2\\0\end{pmatrix}$	$\begin{pmatrix}3\\8\end{pmatrix}, \begin{pmatrix}1\\9\end{pmatrix}, \begin{pmatrix}2\\7\end{pmatrix}$

   (2) SAS data set
   We create a SAS data set that contains responses y_1, y_2; indicators x_1, x_2, x_3; character variable id and $z-1 = x_1 - x_3$ and $z_2 = x_2 - x_3$ for late use.

```
data a;
  input y1 y2 x1 x2 x3 id $ @@;
  z1=x1-x3; z2=x2-x3;
  datalines;
  9 3 1 0 0 A 6 2 1 0 0 A
  9 7 1 0 0 A 0 4 0 1 0 B
  2 0 0 1 0 B 3 8 0 0 1 C
  1 9 0 0 1 C 2 7 0 0 1 C
  ;
```

   (3) Problems under the consideration
   Testing on $H_0 : \mu_x = \mu_y$ against $H_a : \mu_x \neq \mu_y$. Matrices E and H.

2. SAS

   (1) The simplest way is to use proc anova.

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

   $\Longrightarrow E = \begin{pmatrix}10 & 1\\ 1 & 24\end{pmatrix}$, $H = \begin{pmatrix}78 & -12\\ -12 & 48\end{pmatrix}$, $\Lambda = 0.038$, p-value: 0.0062.

   (2) Alternatively, it can be formulated as a regression without intercept
   $y = \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3 + \epsilon$ and we test $\mu_1 = \mu_2 = \mu_3$.

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

   $\Longrightarrow E = \begin{pmatrix}10 & 1\\ 1 & 24\end{pmatrix}$, $H = \begin{pmatrix}78 & -12\\ -12 & 48\end{pmatrix}$, $\Lambda = 0.038$, p-value: 0.0062.
   This method needs q indicator variable.

   (3) It can also be formulated as a regression with intercept
   $y = \mu_. + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \epsilon$ where $\alpha_1 + \alpha_2 + \alpha_3 = 0$, i.e.,
   $y = \mu_. + \alpha_1(x_1 - x_3) + \alpha_2(x_2 - x_3) + \epsilon$ and we test $\alpha_1 = \alpha_2 = 0$
   Thus new variables $z_1 = x_1 - x_3$ and $z_2 = x_2 - x_3$ are needed.

```
proc reg;
   model y1 y2=z1 z2/noprint;
   mtest /print;
   run;
```

$\Longrightarrow E = \begin{pmatrix} 10 & 1 \\ 1 & 24 \end{pmatrix}$, $H = \begin{pmatrix} 78 & -12 \\ -12 & 48 \end{pmatrix}$, $\Lambda = 0.038$, p-value: 0.0062.

(4) proc glm can do the job of proc anova

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

$\Longrightarrow E = \begin{pmatrix} 10 & 1 \\ 1 & 24 \end{pmatrix}$, $H = \begin{pmatrix} 78 & -12 \\ -12 & 48 \end{pmatrix}$, $\Lambda = 0.038$, p-value: 0.0062.

3. Two-sample problem

(1) Two sample test

$H_0: \mu_x = \mu_y$ vs $H_a: \mu_x \neq \mu_y$
Test statistic: $T^2 = (\overline{x} - \overline{y})' \left(\frac{n}{n_1 n_2} S_p \right)^{-1} (\overline{x} - \overline{y})$
p-value: $P(T^2(p, n-2) > T^2_{ob})$

(2) The implementation can be carried out by proc anova

```
proc anova;
   class id;
   model y1 y2=id/nouni;
   manova h=id/print;
   run;
```

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

Proof. For two-sample problem $E = Y(I - JJ^+)Y'$ where $J = \begin{pmatrix} 1_{n_1} & 0 \\ 0 & 1_{n_2} \end{pmatrix}$ and $S_p = \frac{E}{n-2}$.
Under H_0, $E_r = Y(I - 1_n 1_n^+)Y'$. So $H = E_r - E = Y(JJ^+ - 11^+)Y'$. Here

$$\begin{aligned} Y(JJ^+ - 11^+) &= \left[\left(\overline{x} - \tfrac{n_1 \overline{x} + n_2 \overline{y}}{n} \right) 1'_{n_1},\ \left(\overline{y} - \tfrac{n_1 \overline{x} + n_2 \overline{y}}{n} \right) 1'_{n_2} \right] \\ &= \left[\tfrac{n_2(\overline{x} - \overline{y})}{n} 1'_{n_1},\ \tfrac{-n_1(\overline{x} - \overline{y})}{n} 1'_{n_2} \right] = (\overline{x} - \overline{y}) \left(\tfrac{n_2}{n} 1'_{n_1},\ \tfrac{-n_1}{n} 1'_{n_2} \right). \end{aligned}$$

It follows that $H = [Y(JJ^+ - 11^+)][Y(JJ^+ - 11^+)]' = \frac{n_1 n_2}{n} (\overline{x} - \overline{y})(\overline{x} - \overline{y})'$.
Consider $\begin{vmatrix} 1 & -\frac{n_1 n_2}{n} (\overline{x} - \overline{y})' \\ \overline{x} - \overline{y} & E \end{vmatrix}$. We have

$$\left| E + \frac{n_1 n_2}{n} (\overline{x} - \overline{y})(\overline{x} - \overline{y}) \right| = |E| \cdot \left[1 + \frac{n_1 n_2}{n} (\overline{x} - \overline{y})' E^{-1} (\overline{x} - \overline{y}) \right].$$

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

(3) Comment
For testing on $H_0: \mu_x - \mu_y = \delta_0$ the second sample is modified to be the one from a population with mean $\mu_y + \delta_0$ in implementation.