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; (

SAT

(2) Formulation for the populations

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

(3) Formulation for data

The n columns of $Y \in \mathbb{R}^{p \times n}$ are n observed response y with corresponding n values of x being the *n* columns of $X \in \mathbb{R}^{q \times n}$. 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 *i*th sample. Then $YX' \in \mathbb{R}^{p \times q}$ 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 \overline{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. $L(\Theta, \Sigma) \leq L\left(\overline{Y}, \frac{E}{n}\right) = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}.$

- 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 $H_0, Y \sim N_{p \times n}(\theta_1 \mathbf{1}'_n, \Sigma, I_n)$ 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

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

- (3) Implementation
 - (i) Data: Ex6.9 p304



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

Statistic	Value	F	Value	Num	DF	Den	DF	Pr >	·F	
Wilks' Lambda	0.03845535		8.20	4		8		0.00	62	
Pillai's Trace	1.54078842		8.39	4		10		0.00	31	
Hotelling-Lawley Trace	9.94142259		9.94	4		4		0.02	35	
Roy's Greatest Root	8.07638502	2	0.19	2		5		0.00	40	

(iv) Test report

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

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	А	В	С
Response	$\begin{pmatrix} 9\\ 3 \end{pmatrix}, \begin{pmatrix} 6\\ 2 \end{pmatrix}, \begin{pmatrix} 9\\ 7 \end{pmatrix}$	$\begin{pmatrix} 0\\4 \end{pmatrix}, \begin{pmatrix} 2\\0 \end{pmatrix}$	$\binom{3}{8}, \binom{1}{9}, \binom{2}{7}$

(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.



(3) Problems under the consideration Testing on H_0 : $\mu_A = \mu_B$ and $\mu_B = \mu_C$ against H_a : at least one equation in H_0 is false. Matrices E and H.

2. SAS

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

 $\implies E = \begin{pmatrix} 10 & 1 \\ 1 & 24 \end{pmatrix}, H = \begin{pmatrix} 78 & -12 \\ -12 & 48 \end{pmatrix}, \Lambda = 0.038, p$ -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 $\mu_1 = \mu_2 = \mu_3$.

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

(3) It can also be formulated as a regression with intercept $y = \mu_{\cdot} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \epsilon$ where $\alpha_1 + \alpha_2 + \alpha_3 = 0$, i.e., $y = \mu_{\cdot} + \alpha_1 (x_1 - x_3) + \alpha_2 (x_2 - x_3) + \epsilon$ 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;

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

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

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

- 3. Two-sample problem
 - (1) Two sample test

$$H_0: \mu_x = \mu_y \text{ vs } H_a: \mu_x \neq \mu_y$$

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: $P(T^2(p, n-2) > T_{ob}^2)$

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

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

$$Y(JJ^{+} - 11^{+}) = \left[\left(\overline{x} - \frac{n_{1}\overline{x} + n_{2}\overline{y}}{n} \right) \mathbf{1}'_{n_{1}}, \left(\overline{y} - \frac{n_{1}\overline{x} + n_{2}\overline{y}}{n} \right) \mathbf{1}'_{n_{2}} \right] \\ = \left[\frac{n_{2}(\overline{x} - \overline{y})}{n} \mathbf{1}'_{n_{1}}, \frac{-n_{1}(\overline{x} - \overline{y})}{n} \mathbf{1}'_{n_{2}} \right] = (\overline{x} - \overline{y}) \left(\frac{n_{2}}{n} \mathbf{1}'_{n_{1}}, \frac{-n_{1}}{n} \mathbf{1}'_{n_{2}} \right)$$

It follows that $H = [Y(JJ^+ - 11^+)][Y(JJ^+ - 11^+)]' = \frac{n_1n_2}{n}(\overline{x} - \overline{y})(\overline{x} - \overline{y})'.$ Consider $\begin{vmatrix} 1 & -\frac{n_1n_2}{n}(\overline{x} - \overline{y})' \\ \overline{x} - \overline{y} & E \end{vmatrix}$. 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.