L23 Contrast tests in one-way ANOVA

1. Contrast tests in one-way ANOVA

   (1) Extended ANOVA table
   For one-way ANOVA $y = X\mu + \epsilon$ with $H_0 : L\mu = 0$ where $L \in R^{q\times p}$, rank$(L) = q$ and $L1_p = 0$, $H_0 \Longrightarrow E(y) \in \mathbf{R}(H)$ where $H = X(I - L^+L)$. Hence

   $$\mathcal{R}(1_n) \subset \mathcal{R}(H) \subset \mathcal{R}(X) \subset R^n.$$

   Hence there is extended ANOVA table

Source	SS	DF	MS	F	p
Model	SSM= $\|(XX^+ - 11^+)y\|^2$	$p-1$	MSM	MSM/MSE	
H	SSH= $\|(XX^+ - HH^+)y\|^2$	q	MSH	MSH/MSE	
$\mathrm{H}^\perp$	$\mathrm{SSH}^\perp = \|(HH^+ - 11^+)y\|^2$	$p-q-1$	MSE		
Error	SSE= $\|(I - XX^+)y\|^2$	$n-p$			
C.Total	SSTO= $\|(I - 11^+)y\|^2$	$n-1$			

   For testing the usefulness of the model and for the contrast test.

   (2) Filling SSTO, SSE and SSM
   Filling SSTO, SSE and SSM can be done by SAS proc anova and proc glm without specific statements. It can also be calculated with basic statistics.
   SSTO = CSS_{pooled} with DF $n-1$;
   SSE = SSW = $\mathrm{CSS}_1 + \cdots + \mathrm{CSS}_p$ with DF $n-p$;
   SSM = SSB = SSTO − SSE = $n_1(\overline{y}_1 - \overline{y})^2 + \cdots + n_p(\overline{y}_p - \overline{y})^2$ with DF $p-1$.

   (3) Filling SSH and $\mathrm{SSH}^\perp$
   $\mathrm{SSH}^\perp$ = SSM − SSH. SSH can be obtained by using SAS proc glm with contrast statements.
   We now consider how to compute SSH with basic statistics namely n_i and $\overline{y}_i$,
   Ex1: L is not unique.

   $H_0 : \mu_1 = \mu_3 = \mu_5$ and $\mu_2 = \mu_4$ is equivalent to $\begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix} \mu = 0$ and is also equivalent to $\begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -2 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix} \mu = 0$.

2. SSH

   (1) SSH= $\|[X(X'X)^{-1}L'][X(X'X)^{-1}L']^+y\|^2$
   Proof $\mathcal{R}[H,\, X(X'X)^{-1}L'] \subset \mathcal{R}(X)$ and $H'[X(X'X)^{-1}L'] = (I - L^+L)L' = 0$. With

   $$\begin{aligned} \mathrm{rank}[H, X(X'X)^{-1}L'] &= \mathrm{rank}(H) + \mathrm{rank}[X(X'X)^{-1}L'] \\ &= \mathrm{rank}(I - L^+L) + \mathrm{rank}(L') = p = \mathrm{rank}(X), \end{aligned}$$

   $\mathcal{R}(X) = \mathcal{R}[H,\, X(X'X)^{-1}L']$. Hence

   $$\begin{aligned} XX^+ &= [H,\, X(X'X)^{-1}L'][H,\, X(X'X)^{-1}L']^+ = [H,\, X(X'X)^{-1}L'] \begin{pmatrix} H^+ \\ [X(X'X)^{-1}L']^+ \end{pmatrix} \\ &= HH^+ + [X(X'X)^{-1}L'][X(X'X)^{-1}L']^+. \end{aligned}$$

   Conclusion follows.

(2) Let $\widehat{\mu} = \begin{pmatrix} \overline{y}_1 \\ \vdots \\ \overline{y}_p \end{pmatrix}$. Then SSH$= (L\widehat{\mu})'[L(X'X)^{-1}L']^{-1}(L\widehat{\mu})$.

Proof Note that $(X'X)^{-1}X'y = \widehat{\mu}$ and for $A = X(X'X)^{-1}L'$, $A^+ = (A'A)^{-1}A'$. So

$$\begin{aligned}
\text{SSH} &= \|[X(X'X)^{-1}L'][X(X'X)^{-1}L']^+ y\|^2 \\
&= \|[X(X'X)^{-1}L'][L(X'X)^{-1}L']^{-1}[L(X'X)^{-1}X']y\|^2 \\
&= \|[X(X'X)^{-1}L'][L(X'X)^{-1}L']^{-1}L\widehat{\mu}\|^2 = (L\widehat{\mu})'[L(X'X)L']^{-1}(L\widehat{\mu}).
\end{aligned}$$

(3) Let $L' = (l_1, .., l_q)$. Suppose $L(X'X)^{-1}L' = \text{diag}(l_1'(X'X)^{-1}l_1, .., l_q'(X'X)^{-1}l_q)$. Write $l_i'(X'X)^{-1}l_i = \|l_i\|^2_{(X'X)^{-1}}$. Then SSH $= \sum_{i=1}^q \frac{(l_i'\widehat{\mu})^2}{\|l_i\|^2_{(X'X)^{-1}}}$.

Proof
$$\begin{aligned}
SSH &= (L\widehat{\mu})'[L(X'L']^{-1}(L\widehat{\mu}) \\
&= (l_1'\widehat{\mu}, ..., l_q'\widehat{\mu})[\text{diag}(\|l_1\|^2_{(X'X)^{-1}}, .., \|l_q\|^2_{(X'X)^{-1}})]^{-1}(l_1'\widehat{\mu}, ..., l_q'\widehat{\mu})' \\
&= \sum_{i=1}^q \frac{(l_i'\widehat{\mu})^2}{\|l_i\|^2_{(X'X)^{-1}}}.
\end{aligned}$$

3. Rewrite H_0 for contrast test for computation

(1) QR-decomposition

Rewriting H_0 equivalently as $L\mu = 0$ such that $L(X'X)^{-1}L'$ is a diagonal matrix can be done by QR-decomposition with respect to the inner product $\langle u, v\rangle_{(X'X)^{-1}}$.

Let $L\mu = 0$ be the original H_0 where $L' = (l_1, ..l_q) \in R^{p\times q}$.

Let $l_{1*} \propto l_1$; and $l_{i*} \propto l_i - \sum_{j=1}^{i-1} \frac{\langle l_i, l_{j*}\rangle_{(X'X)^{-1}}}{\|l_{j*}\|^2_{(X'X)^{-1}}} l_{j*}$, $i = 2, 3, .., q$. Then

$$L\mu = 0 \Longleftrightarrow L_*\mu = 0, \text{ Here } L_*1_p = 0 \text{ and } L_*(X'X)^{-1}L_*' \text{ is diagonal.}$$

(2) Rewrite H_0 in Ex1 such that $L(X'X)L'$ is diagonal

$$L' = (l_1, l_2, l_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \; l_{1*} \propto l_1. \text{ Let } l_{1*} = l_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

$$l_{2*} \propto l_2 - \frac{\langle l_2, l_{1*}\rangle_{(X'X)^{-1}}}{\|l_{1*}\|^2_{(X'X)^{-1}}} l_{1*} \propto \begin{pmatrix} n_1 \\ 0 \\ n_3 \\ 0 \\ -n_1 - n_3 \end{pmatrix} \overset{let}{=\!=} l_{2*}.$$

$$l_{3*} \propto l_3 - \sum_{j=1}^2 \frac{\langle l_3, l_{j*}\rangle_{(X'X)^{-1}}}{\|l_{j*}\|^2_{(X'X)^{-1}}} l_{j*} = l_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \overset{let}{=\!=} l_{3*}$$

$$\text{Thus } H_0 : \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ n_1 & 0 & n_3 & 0 & -n_1 - n_3 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix} \mu = 0.$$

(3) For Ex1, find SSH

Using H_0 in (2). SSH $= \frac{(\overline{y}_1 - \overline{y}_3)^2}{\frac{1}{n_1} + \frac{1}{n_3}} + \frac{(n_1\overline{y}_1 + n_3\overline{y}_3 - (n_1+n_3)\overline{y}_5)^2}{n_1 + n_3 + \frac{(n_1+n_3)^2}{n_5}} + \frac{(\overline{y}_2 - \overline{y}_4)^2}{\frac{1}{n_2} + \frac{1}{n_4}}$.

L24: Testing the interactions in Two-way ANOVA

1. Two-way ANOVA with interaction

   (1) Model
   Factor A has a levels; Factor B has b levels; There are ab treatments with mean μ_{ij}.

   $$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

   where $\sum_i \alpha_i = \sum_i (\alpha\beta)_{ij} = 0$ and $\sum_j \beta_j = \sum_j (\alpha\beta)_{ij} = 0$.

   (2) Samples
   There are ab cell samples with sizes n_{ij}, mean $\overline{y}_{ij}$ and CSS$_{ij}$, $i = 1, .., a$; $j = 1, .., b$;
   a row samples with sizes $n_{i.}$, mean $\overline{y}_{i.}$ and CSS$_{i.}$, $i = 1, .., a$;
   b column samples with sizes $n_{.j}$, mean $\overline{y}_{.j}$ and CSS$_{.j}$, $j = 1, .., b$;
   and one pooled sample has size $n_{..}$, mean $\overline{y}_{..}$ and CSS$_{pooled}$.
   Let $\mathbf{y} \in R^{n_{..}}$ be data vector, the columns of $X \in R^{n_{..} \times ab}$ be indicators for ab cells, and $\mu \in R^{ab}$ be cell mean vector. Then $\mathbf{y} = X\mu + e$, $e \sim N(0, \sigma^2 I_n)$ and $X 1_{ab} = 1_{n_{..}}$. So $\mathcal{R}(1_{n_{..}}) \subset \mathcal{R}(X) \subset R^{n_{..}}$.

   (3) The usefulness of the model
   For testing $H_0 : \mu_{ij} = \mu_{st}$ for all (i, j) and (s, t), we have

Source	SS	DF	MS	F	p
Model	SSM= $\|(XX^+ - 11^+)y\|^2$	ab-1	MSM	MSM/MSE	
Error	SSE= $\|(I - XX^+)y\|^2$	n-ab	MSE		
C.Total	SSTO= $\|(I - 11+)y\|^2$	n-1			

   Here $SSTO = CSS_{pooled}$ and SSE= $\sum_i \sum_j CSS_{ij}$.

2. Testing the interaction

   (1) Testing the interactions
   For two-way ANOVA we consider testing $H_0 : (\alpha\beta)_{ij} = 0$ for all i, j.
   The columns of $A \in R^{n_{..} \times a}$ and $B \in R^{n_{..} \times b}$ are indicators for a levels of A and b levels of B. With $\alpha \in R^a$, $\beta \in R^b$ with $1_a'\alpha = 0$ and $1_b'\beta = 0$, under H_0,

   $$\begin{aligned} E(y) &= 1_{n_{..}}\mu_{..} + A\alpha + B\beta = 1_{n_{..}}\mu_{..} + A(I_a - 11^+)\gamma + B(I_b - 11^+)\gamma_* \\ &= 1_{n_{..}}\mu_{..} + (A - 1_{n_{..}}1_a^+)\gamma + (B - 1_{n_{..}}1_b^+)\gamma_* \in \mathcal{R}(1_{n_{..}},\ A - 11_a^+,\ B - 11_b^+) \\ &\overset{claim}{=\!=} \mathcal{R}(1_{n_{..}}1_{n_{..}}^+,\ AA^+ - 11^+,\ BB^+ - 11^+). \end{aligned}$$

   Proof $\supset$ holds since $(11^+,\ AA^+ - 11^+,\ BB^+ - 11^+) = (1,\ A - 11_a^+,\ B - 11_b^+) \begin{pmatrix} 1^+ & 1_a^+ & 1_b^+ \\ 0 & I_a & 0 \\ 0 & 0 & I_b \end{pmatrix}$.

   $\subset$ is left for your exercise.

   (2) Extended ANOVA table
   Denote $(11^+,\ AA^+ - 11^+,\ BB^+ - 11^+)$ as H. Then

   $$\mathcal{R}(1) \subset \mathcal{R}(H) \subset \mathcal{R}(X) \subset R^{n_{..}}$$

   So the test can be carried out with the extended ANOVA table

Source	SS	DF	MS	F	p
Model	SSM= $\|(XX^+ - 11^+)y\|^2$	ab-1	MSM	MSM/MSE	
AB	SSAB= $\|XX^+ - HH^+)y\|^2$	(a-1)(b-1)	MSAB	MSAB/MSE	
AB$^\perp$	SSAB$^\perp$ = $\|(HH^+ - 11^+)y\|^2$	a+b-2			
Error	SSE= $\|(I - XX^+)y\|^2$	n-ab	MSE		
C.Total	SSTO= $\|(I - 11+)y\|^2$	n-1			

3. Computations

(1) Under balanced data
In h, $11^+(AA^+ - 11^+) = 0$ and $11^+(BB^+ - 11^+) = 0$. Under balanced data with $n_{ij} \equiv n$,

$$\begin{aligned} AA^+BB^+ &= A(A'A)^{-1}A'B(B'B)^{-1}B' = A\left(\frac{1}{nb}\right)(n1_a1_b'\left(\frac{1}{na}\right)B \\ &= (A1_a)\frac{1}{nab}(B1_b)' = 1_{n..}\frac{1}{n..}1_{n..}' = 11^+ \end{aligned}$$

Hence $(AA^+ - 11^+)(BB^+ - 11^+) = 0$ and $HH^+ = 11^+ + (AA^+ - 11^+) + (BB^+ - 11^+)$

(2) Extended ANOVA table

$$\begin{aligned} \text{SSAB}^\perp &= \|(HH^+ - 11^+)y\|^2 = \|(AA^+ - 11^+)y + (BB^+ - 11^+)y\|^2 \\ &\overset{denoted}{=\!=} \text{SSA+SSB} = [\text{CSS}_{pooled} - \textstyle\sum_i \text{CSS}_{i.}] + [\text{CSS}_{pooled} - \textstyle\sum_j \text{CSS}_{.j}]. \end{aligned}$$

SSAB=SSM-SSA-SSB. We have the extended ANOVA table for the test

Source	SS	DF	MS	F	p
Model	SSM= $\|(XX^+ - 11^+)y\|^2$	ab-1	MSM	MSM/MSE	
A	SSA= $\|(AA^+ - 11^+)y\|^2$	a-1	MSA	MSA/MSE	
B	SSB= $\|(BB^+ - 11^+)y\|^2$	b-1	MSB	MSB/MSE	
A*B	SSAB= $\|XX^+ - HH^+)y\|^2$	(a-1)(b-1)	MSAB	MSAB/MSE	
Error	SSE= $\|(I - XX^+)y\|^2$	n-ab	MSE		
C.Total	SSTO= $\|(I - 11+)y\|^2$	n-1			

(3) SAS for the above table

```
data a;
     infile "D:\ex.txt";
     input y A $ B $ @@;
proc anova;
     class A B;
     model y=A B A*B;
     run;
```

Caution: For unbalanced data proc anova is not valid. Use proc glm instead.