L25: Testing factor effects in two-way ANOVA

1 Matrices in two-way ANOVA

(1) Samples and basic statistics in two-way ANOVA
In two-way ANOVA there are ab treatments formed by a levels of factor A and b levels of factor B. There are ab samples of responses to treatment with size n_{ij} and $\text{CSS}_{i,j}$.
There are a samples of responses to levels of A with size $n_{i.}$ and $\text{CSS}_{i.}$.
There are b samples of responses to levels of B with size $n_{.j}$ and $\text{CSS}_{.j}$.
There is a pooled sample of size $n_{..}$ and $\text{CSS}_{..}$.

(2) Matrices, spaces and projections
With respect to data vector $\mathbf{y} \in R^{n_{..}}$ let the columns of $X \in R^{n_{..}\times ab}$, $A \in R^{n_{..}\times a}$ and $B \in R^{n_{..}\times b}$ be the indicators of treatments, levels of A and levels of B respectively. Then

$$\mathcal{R}(1_{n_{..}}) \subset \left\{ \begin{array}{c} \mathcal{R}(A) \\ \mathcal{R}(B) \end{array} \right. \subset \mathcal{R}[(A,\, B)] \subset \mathcal{R}(X) \subset R^{n_{..}}.$$

Here $\mathcal{R}[(A,\, B)] = \mathcal{R}[(11^+, AA^+ - 11^+,\, BB^+ - 11^+)]$ with
$$(A,B)(A,B)^+ = 11^+ + (AA^+ - 11^+,\, BB^+ - 11^+)(AA^+ - 11^+,\, BB^+ - 11^+)^+.$$

(3) Balanced data case
When $n_{i,j} \equiv n$, $AA^+BB^+ = BB^+AA^+ = 11^+$. Hence
$$(A,\, B)(A,\, B)^+ = 11^+ + (AA^+ - 11^+) + (BB^+ - 11^+).$$

2. Testing interactions in Two-way ANOVA

(1) Extended ANOVA table
For two-way ANOVA with interactions, $E(y) \in \mathcal{R}(X)$. Under $H_0: (\alpha\beta)_{ij} = 0$ for all i, j, $E(y) \in \mathcal{R}(A,\, B)$. With $\mathcal{R}(1) \subset \mathcal{R}[(A,\, B)] \subset \mathcal{R}(X) \subset R^{n_{..}}$, there is extended ANOVA table for testing the usefulness of the model and for the interactions.

Source	SS
Model	SSM= $\|(XX^+ - 11^+)y\|^2$ = SSTO − SSE
H	SSH= $\|[XX^+ - (A,B)(A,B)^+]y\|^2$ = SSM − $\text{SSH}^\perp$
$\text{H}^\perp$	$\text{SSH}^\perp = \|(A,\, B)(A,\, B)^+ - 11^+)y\|^2$
Error	SSE= $\|(I - XX^+)y\|^2 = \sum_{i,j} \text{CSS}_{ij}$
C.Total	SSTO= $\|(I - 11^+)y\|^2 = CSS_{..}$

Here $\text{SSH}^\perp = \|[(A,B)(A,B)^+ - 11^+]y\|^2$ is the key for completing the table.

(2) Balance data case
With balanced data, by (3) in 1 and $(AA^+ - 11^+)(BB^+ - 11^+) = 0$,
$$\begin{aligned} SSH^\perp &= \|(AA^+ - 11^+)y + (BB^+ - 11^+)y\|^2 \\ &= \|(AA^+ - 11^+)y\|^2 + \|(BB^+ - 11^+)y\|^2 = SSA + SSB \end{aligned}$$
Denote SSH by SSAB. We have

Source	SS	DF
Model	SSM= $\|(XX^+ - 11^+)y\|^2$ = SSTO − SSE	$ab - 1$
A	SSA= $\|(AA^+ - 11^+)y\|^2 = \text{CSS}_{..} - \sum_i \text{CSS}_{i.}$	$a - 1$
B	SSB= $\|(BB^+ - 11^+)y\|^2 = \text{CSS}_{..} - \sum_j \text{CSS}_{.j}$	$b - 1$
AB	SSAB= $\|[XX^+ - (A,\, B)(A,\, B)^+]y\|^2$ = SSM − SSA − SSB	$(a-1)(b-1)$
Error	SSE= $\|(I - XX^+)y\|^2 = \sum_{i,j} \text{CSS}_{ij}$	$nab - ab$
C.Total	SSTO= $\|(I - 11^+)y\|^2 = CSS_{..}$	$nab - 1$

(3) SAS

The above table can be produced by

```
proc anova;
     class A B;
     model y=A B A*B;
     run;
```

For unbalanced data case replace anova by glm and use Type III SS in SAS output.

3. Testing the effect of one factor in two-way ANOVA without interactions

(1) Extended ANOVA table

For two-way ANOVA without interactions, $E(y) \in \mathcal{R}[(A, B)]$. Under $H_0: \alpha_i = 0$ for all i, $E(y) \in \mathcal{R}(B)$. With $\mathcal{R}(1) \subset \mathcal{R}(B) \subset \mathcal{R}(X) \subset R^{n_{\cdot\cdot}}$, there is extended ANOVA table for testing the usefulness of the model and for the effect of Factor A.

Source	SS
Model	SSM$= \|[(A,B)(A,B)^+ - 11^+]y\|^2 = \text{SSH} + \text{SSH}^\perp$
H	SSH$= \|[(A,B)(A,B)^+ - BB^+]y\|^2$
$\text{H}^\perp$	$\text{SSH}^\perp = \|(BB^+ - 11^+)y\|^2 = \text{SSB} = \text{CSS}_{\cdot\cdot} - \sum_j \text{CSS}_{\cdot j}$
Error	SSE$= \|[I - (A,B)(A,B)^+]y\|^2 = \text{SSTO} - \text{SSM}$
C.Total	SSTO$= \|(I - 11^+)y\|^2 = CSS_{\cdot\cdot}$

Here $\text{SSH} = \|[(A,B)(A,B)^+ - 11^+]y\|^2$ is the key for completing the table.

(2) Balance data case

With balanced data, by (3) in 1,

$$SSH = \|(AA^+ - 11^+)y\|^2 = \text{SSA} = \text{CSS}_{\cdot\cdot} - \sum_i \text{CSS}_{i\cdot}.$$

Denote SSH by SSA. We have

Source	SS	DF
Model	SSM$= \|[(A,B)(A,B)^+ - 11^+]y\|^2 = \text{SSA} + \text{SSB}$	$a+b-2$
A	SSA$= \|(AA^+ - 11^+)y\|^2 = \text{CSS}_{\cdot\cdot} - \sum_i \text{CSS}_{i\cdot}$	$a-1$
B	SSB$= \|(BB^+ - 11^+)y\|^2 = \text{CSS}_{\cdot\cdot} - \sum_j \text{CSS}_{\cdot j}$	$b-1$
Error	SSE$= \|[I - (A,B)(A,B)^+]y\|^2 = \text{SSTO} - \text{SSM}$	$nab-a-b+1$
C.Total	SSTO$= \|(I - 11^+)y\|^2 = CSS_{\cdot\cdot}$	$nab-1$

(3) SAS

The above table can be produced by

```
proc anova;
     class A B;
     model y=A B;
     run;
```

For unbalanced data case replace anova by glm and use Type III SS in SAS output.

Comment: Testing on factor effect should be conduct only after it is confirmed that the interactions involving the factor is not significant (no interaction).

L26: Tests in ANCOVA

1. Analysis of Covariance model (ANCOVA)

   (1) ANOVA model
   For ANOVA model $\mathbf{y} = X\mu + e$, with respect to the response vector $\mathbf{y} \in R^n$, the columns of $X \in R^{n\times p_1}$ are the values of p_1 indicators for p_1 treatments, $\mu \in R^{p_1}$ contains the mean response to the p_1 treatments and $e \sim N(0, \sigma^2 I_n)$.
   Note that $X1_{p_1} = 1_n$. So $\mathbf{R}(1_n) \subset \mathbf{R}(X)$.

   (2) Regression model
   For regression model $\mathbf{y} = Z\gamma + e$, with respect to the response vector $\mathbf{y}$, the columns of $Z \in R^{n\times p_2}$ are the values of p_2 predictors (regressors) and $e \sim N(0, \sigma^2 I_n)$.

   (3) ANCOVA model
   When dealing with the ANOVA model in (1), one may want to introduce regressors into the model. These regressors are called the covariates to the ANOVA model, the resulted model $\mathbf{y} = X\mu + Z\gamma + e$ is called an analysis of covariance model (ANCOVA).
   ANCOVA is a linear model since in $\mathbf{y} = X\mu + Z\gamma + e = (X,\, Z)\begin{pmatrix}\mu\\ \gamma\end{pmatrix} + e$ with $E(e) = 0$, $E(\mathbf{y}) = (X,\, Z)\begin{pmatrix}\mu\\ \gamma\end{pmatrix}$ is a linear function of unknown parameter vector $\begin{pmatrix}\mu\\ \gamma\end{pmatrix} \in R^{p_1+p_2}$ and lies in the linear space $\mathcal{R}[(X,\, Z)]$ where X gives the values of p_1 indicators and Z gives the values of p_2 numerical variables.

2. Tests in ANCOVA

   (1) Test the usefulness of the model
   Because H_0 : The model is useless $\Longleftrightarrow H_0 :\ \mu = \mu_. 1_{p_1}$ and $\gamma = 0 \Longrightarrow E(\mathbf{y}) \in \mathcal{R}(1_n)$ and $\mathcal{R}(1) \subset \mathcal{R}[(X,\, Z)] \subset R^n$, there is an ANOVA able for testing the usefulness of the model.

Source	SS	DF	MS	F	p
Model	SSM= $\|[(X,Z)(X,Z)^+ - 11^+]y\|^2$	$\text{rank}[(X,Z)] - 1$	MSM	–	–
Error	SSE= $\|[I - (X,Z)(X,Z)^2]y\|^2$	$n - \text{rank}[(X,Z)]$	MSE		
C.Total	SSTO= $\|(I - 11^+)y\|^2$	$n - 1$			

   (2) Test the ANOVA part
   Because H_0 : The ANOVA part is useless $\Longleftrightarrow H_0 :\ \mu = \mu_. 1_{p_1} \Longrightarrow E(\mathbf{y}) \in \mathcal{R}[(1_n,\, Z)]$ and $\mathcal{R}(1) \subset \mathcal{R}[(1_n,\, Z)] \subset \mathcal{R}[(X,\, Z)] \subset R^n$, there is decomposition of SSM in (1) for testing the usefulness of the ANOVA part.

Source	SS	DF	MS	F	p
Model	SSM= $\|[(X,Z)(X,Z)^+ - 11^+]y\|^2$	$\text{rank}[(X,Z)] - 1$	MSM	–	–
H	SSH= $\|[(X,Z)(X,Z)^+ - (1,Z)(1,Z)^+]y\|^2$	$\text{rank}[(X,Z)] - \text{rank}[(1,Z)]$	MSH	–	–
$\text{H}^\perp$	$\text{SSH}^\perp$ = $\|[(1,Z)(1,Z)^+ - 11^+]y\|^2$	$\text{rank}[(1,Z)] - 1$			
Error	SSE= $\|[I - (X,Z)(X,Z)^2]y\|^2$	$n - \text{rank}[(X,Z)]$	MSE		
C.Total	SSTO= $\|(I - 11^+)y\|^2$	$n - 1$			

   (3) Test the regression part
   Because H_0 : The regression part is useless $\Longleftrightarrow H_0 :\ \gamma = 0 \Longrightarrow E(\mathbf{y}) \in \mathcal{R}(X)$ and $\mathcal{R}(1) \subset \mathcal{R}(X) \subset \mathcal{R}[(X,\, Z)] \subset R^n$, there is decomposition of SSM in (1) for testing the usefulness of the regression part.

Source	SS	DF	MS	F	p
Model	SSM= $\|[(X,Z)(X,Z)^+ - 11^+]y\|^2$	$\text{rank}[(X,Z)] - 1$	MSM	–	–
H	SSH= $\|[(X,Z)(X,Z)^+ - XX^+]y\|^2$	$\text{rank}[(X,Z)] - \text{rank}(X)$	MSH	–	–
$\text{H}^\perp$	$\text{SSH}^\perp = \|(XX^+ - 11^+)y\|^2$	$\text{rank}(X) - 1$			
Error	SSE= $\|[I - (X,Z)(X,Z)^2]y\|^2$	$n - \text{rank}[(X,Z)]$	MSE		
C.Total	SSTO= $\|(I - 11^+)y\|^2$	$n - 1$			

3. SAS

(1) SAS code

```
data a;
   infile "D:\ex.txt";
   input y x c $ @@;
proc glm;
   class c;
   model y=c x;
   run;
```

(2) SAS output
SAS put three tests in two tables

Source	DF	SS	MS	F	Pr>F
Model	3	1847	616	4.44	0.126
Error	3	416	139		
C.Total	6	2263			

Source	DF	Type III SS	MS	F	Pr>F
x	1	1430	1430	10.31	0.049
c	2	418	209	1.51	0.353

The first table shows that the model is not useful. The second table reveals that the regression part is useful but the ANOVA part is not. So next we try regression only.

(3) Try regression only

```
proc reg;
   model y=x;
   run;
```

Source	DF	SS	MS	F	Pr>F
Model	1	1430	1430	8.57	0.0327
Error	5	833	167		
C.Total	6	2263			

The output shows that the regression model is useful.