MA50259: Statistical Design of Investigations

Dr. Sandipan Roy

Lecture 4

Generalised Inverse and its properties

Let A be a square matix. A generalised inverse (G-inv) of A is any matrix A^{-} such that

$$AA^{-}A = A$$

Properties: Let $G = A^{-} = (X^T X)^{-}$

- G is also squared
- G^T is also a G-inverse of $A = X^T X$
- Let $H_X = XGX^T$ (Projection matrix or Hat matrix): Projection onto the column space of X i.e.

$$H_X X = X$$

- $H_X = XGX^T$ is invariant to the choice of G
- $H_X = XGX^T$ is symmetric
- $\text{rank}(H_X) = \text{rank}(XGX^T) = \text{rank}(X)$

Properties of the Projection Matrix H_X

- $(I - H_X)$ is invariant to the choice of G
- Both H_X and $(I - H_X)$ are idempotent
- $\text{SSE} = (Y - X\tilde{\beta})(Y - X\tilde{\beta})^T = Y^T(I - H_X)Y$ and is invariant to the choice of G
- If A is idempotent, then $\text{rank}(A) = \text{trace}(A)$
- $H_X\mathbf{1} = \mathbf{1}$, since $\mathbf{1}$ is a column of X

ANOVA for CRD

- Treatment effects model using matrix notation:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim MVN(\mathbf{0}, \sigma^2 \boldsymbol{I})$ and

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_t \end{pmatrix}$$

- We test $H_0 : \tau_1 = \tau_2 = \ldots = \tau_t$

Reduced model

$$\boldsymbol{y} = \boldsymbol{X_0}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\beta = \mu$ and

$$\boldsymbol{X_0} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Sum of Squares of Errors

- SSE under the general model :
$$\text{SSE} = \left(\mathbf{y} - \mathbf{X}\tilde{\beta}\right)^T \left(\mathbf{y} - \mathbf{X}\tilde{\beta}\right)$$
where $\tilde{\beta}$ are the solutions to normal equations.
- SSE under the null model:
$$\text{SSE}_0 = \left(\mathbf{y} - \mathbf{X_0}\hat{\mu}\right)^T \left(\mathbf{y} - \mathbf{X_0}\hat{\mu}\right) = \mathbf{y^T A y}$$
where $A = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$

Comparing SSE Under Null and Alternative

- A is idempotent and $\text{rank}(A) = \text{trace}(A) = n - 1$
- $\text{SSE}_0 > \text{SSE}$ (Fit is always worse if less parameters) and
$$\text{SSE}_0 - \text{SSE} = \mathbf{y^T B y}$$
where $\mathbf{B} = \mathbf{X G X^T} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$
- B is also idempotent and $\text{rank}(B) = \text{trace}(B) = t - 1$

Chi-square Distributions in the ANOVA Table

- If $\boldsymbol{y} \sim \text{MVN}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I})$, Then
$$\frac{\mathbf{y^T A y}}{\sigma^2} \sim \chi^2_{(\text{rank}(A), \lambda)}$$
where A is an idempotent matrix and λ is the non-centrality parameter
- under H_0, $\lambda = 0$
- In general for a balanced design with r replicates, $\lambda = \frac{r}{\sigma^2} \sum_{i=1}^{t} \left(\mu_i - \bar{\mu}\right)^2$

ANOVA Table

Sources	df	Sum of Squares	Mean Squares	F
Treatment	t-1	SST	MST $= \frac{SST}{t-1}$	
Error	n-t	SSE	MSE $= \frac{SSE}{n-t}$	$F = \frac{MST}{MSE}$
Total	n-1	SS_{Total}		

- Under H_0, $F \sim F_{t-1,n-t}$ (Central F-distribution)
- In general, $F \sim F_{t-1,n-t,\lambda}$ (Non Central F-distribution)