6. ANalysis Of VAriance (ANOVA)

Setup and Notation

corresponds to intercept only model (averaging all y-values)

GMMNE

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \ \sigma^2 \boldsymbol{I})$$

sequence of model matrices

Let $\boldsymbol{X}_1 = \boldsymbol{1}$, $\boldsymbol{X}_m = \boldsymbol{X}$, and $\boldsymbol{X}_{m+1} = \underbrace{\boldsymbol{I}}_{n \times n}$.

Suppose $\boldsymbol{X}_2, \ \ldots, \boldsymbol{X}_m$ are matrices satisfying

$$\mathcal{C}(\boldsymbol{X}_1) \subset \mathcal{C}(\boldsymbol{X}_2) \subset \cdots \subset \mathcal{C}(\boldsymbol{X}_{m-1}) \subset \mathcal{C}(\boldsymbol{X}_m).$$

Let $\boldsymbol{P}_j = \boldsymbol{P}_{\boldsymbol{X}_j}$ and $r_j = \text{rank}(\boldsymbol{X}_j) \ \forall \ j = 1, \ldots, m+1.$

The Total Sum of Squares

The *total sum of squares* (also known as the *corrected total sum of squares*) is

$y^\top y$ $\qquad\qquad$ $P_1 y$

$$\sum_{i=1}^{n}(y_i - \bar{y}_.)^2 = \begin{bmatrix} \underline{y_1 - \bar{y}_.} \\ \vdots \\ \underline{y_n - \bar{y}_.} \end{bmatrix}^\top \begin{bmatrix} y_1 - \bar{y}_. \\ \vdots \\ y_n - \bar{y}_. \end{bmatrix} = [\boldsymbol{y} - \boxed{\bar{y}_.\boldsymbol{1}}]^\top[\boldsymbol{y} - \bar{y}_.\boldsymbol{1}]$$

$$= [\boldsymbol{y} - \boldsymbol{P}_1\boldsymbol{y}]^\top[\boldsymbol{y} - \boldsymbol{P}_1\boldsymbol{y}] = [\boldsymbol{I}\boldsymbol{y} - \boldsymbol{P}_1\boldsymbol{y}]^\top[\boldsymbol{I}\boldsymbol{y} - \boldsymbol{P}_1\boldsymbol{y}]$$

$$= [(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}]^\top[(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}]$$

$$= \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$$

symmetry

$$= \boldsymbol{y}^\top\boxed{(\boldsymbol{I} - \boldsymbol{P}_1)}(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$$

$(I - P_1)$ is also idempotent

$$= \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}.$$

Partitioning the Total Sum of Squares

Slide 1: $X_{m+1} = I$

$P_{m+1} = I$

$$\sum_{i=1}^{n}(y_i - \bar{y}_{\cdot})^2 = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y} = \boldsymbol{y}^\top(\boxed{\boldsymbol{P}_{m+1}} - \boldsymbol{P}_1)\boldsymbol{y}$$

$P_2 + P_3 + \ldots + P_{m+1}$

$P_1 + P_2 + P_3 + \ldots + P_m$

$$= \boldsymbol{y}^\top\left(\sum_{j=2}^{m+1}\boldsymbol{P}_j - \sum_{j=1}^{m}\boldsymbol{P}_j\right)\boldsymbol{y}$$

$$= \boldsymbol{y}^\top(\boldsymbol{P}_{m+1} - \boldsymbol{P}_m + \boldsymbol{P}_m - \boldsymbol{P}_{m-1} + \cdots + \boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$$

Partition into disjoint sums of squares

$$= \boldsymbol{y}^\top(\boldsymbol{P}_{m+1} - \boldsymbol{P}_m)\boldsymbol{y} + \ldots + \boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$$

$$= \sum_{j=1}^{m}\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}.$$

The sums of squares in the equation

$$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y} = \sum_{j=1}^{m} \boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}$$

are often arranged in an ANOVA table.

Some Additional Sum of Squares Notation

reduction in the overall/total sums of squares when using X_2 as a model matrix instead of X_1

intercept-only model

Sum of Squares	Sum of Squares
$\boxed{\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}}$	$SS(2 \mid 1)$
$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$SS(3 \mid 2)$
$\vdots$	$\vdots$
$\boldsymbol{y}^\top(\boldsymbol{P}_m - \boldsymbol{P}_{m-1})\boldsymbol{y}$	$SS(m \mid m-1)$
$\boldsymbol{y}^\top(\boldsymbol{P}_{m+1} - \boldsymbol{P}_m)\boldsymbol{y}$	$SSE = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}$

intercept only model

amount of variability that remains unexplained

Note that

$$
\begin{aligned}
SS(j+1 \mid j) &= \boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y} \\
&= \boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j \boxed{+ \boldsymbol{I} - \boldsymbol{I}})\boldsymbol{y} \\
&= \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_j - \boldsymbol{I} + \boldsymbol{P}_{j+1})\boldsymbol{y} \\
&= \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_j)\boldsymbol{y} - \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{j+1})\boldsymbol{y} \\
&= SSE_j - SSE_{j+1}
\end{aligned}
$$

reduction in sums of squares moving from model matrix X_j to X_{j+1}

Thus, $SS(j+1 \mid j)$ is the amount the error sum of square decreases when $\boldsymbol{y}$ is projected onto $\mathcal{C}(\boldsymbol{X}_{j+1})$ instead of $\mathcal{C}(\boldsymbol{X}_j)$.

$SS(j+1 \mid j),\ j = 1, \ldots, m-1$ are called *Sequential Sums of Squares.*

SAS calls these *Type I Sums of Squares.*

Type II, Type III, Type IV

Properties of the Matrices of the Quadratic Forms

The matrices of the quadratic forms in the ANOVA table have several useful properties:

- Symmetry
- Idempotency
- $\text{rank}(P_{j+1} - P_j) = r_{j+1} - r_j$
- Zero Cross-Products

Symmetry and Idempotency

Note that $\forall\ j = 1, \ \ldots, \ m$

$$(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)^\top = \boldsymbol{P}_{j+1}^\top - \boldsymbol{P}_j^\top = \boldsymbol{P}_{j+1} - \boldsymbol{P}_j$$

and

$$\begin{aligned}(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) &= \boldsymbol{P}_{j+1}\boldsymbol{P}_{j+1} - \boldsymbol{P}_{j+1}\boldsymbol{P}_j - \boldsymbol{P}_j\boldsymbol{P}_{j+1} \\ &\quad + \boldsymbol{P}_j\boldsymbol{P}_j \\ &= \boldsymbol{P}_{j+1} - \boldsymbol{P}_j - \boldsymbol{P}_j + \boldsymbol{P}_j \\ &= \boldsymbol{P}_{j+1} - \boldsymbol{P}_j.\end{aligned}$$

By idempotency and symmetry,

$$
\begin{aligned}
\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y} &= \boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y} \\
&= \boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y} \\
&= [(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}]^\top[(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}] \\
&= ||(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}||^2 \\
&= ||\boldsymbol{P}_{j+1}\boldsymbol{y} - \boldsymbol{P}_j\boldsymbol{y}||^2 \\
&\equiv ||\widehat{\boldsymbol{y}}^{(j+1)} - \widehat{\boldsymbol{y}}^{(j)}||^2 \\
&= \sum_{i=1}^{n} \left(\widehat{y_i}^{(j+1)} - \widehat{y_i}^{(j)}\right)^2,
\end{aligned}
$$

which is why we call $\boldsymbol{y}^\top(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{y}$ a “sum of squares.”

Because rank is equal to trace for idempotent matrices, we have

$$
\begin{aligned}
\text{rank}(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) &= \text{tr}(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) = \text{tr}(\boldsymbol{P}_{j+1}) - \text{tr}(\boldsymbol{P}_j) \\
&= \text{rank}(\boldsymbol{P}_{j+1}) - \text{rank}(\boldsymbol{P}_j) \\
&= \text{rank}(\boldsymbol{X}_{j+1}) - \text{rank}(\boldsymbol{X}_j) \\
&= r_{j+1} - r_j.
\end{aligned}
$$

Zero Cross-Products

$\forall\ j < \ell$

$$
\begin{aligned}
(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)(\boldsymbol{P}_{\ell+1} - \boldsymbol{P}_\ell) &= \boldsymbol{P}_{j+1}\boldsymbol{P}_{\ell+1} - \boldsymbol{P}_{j+1}\boldsymbol{P}_\ell - \boldsymbol{P}_j\boldsymbol{P}_{\ell+1} \\
&\quad + \boldsymbol{P}_j\boldsymbol{P}_\ell \\
&= \boldsymbol{P}_{j+1} - \boldsymbol{P}_{j+1} - \boldsymbol{P}_j + \boldsymbol{P}_j \\
&= \boldsymbol{0}.
\end{aligned}
$$

Transposing both sides and using symmetry gives

$$(\boldsymbol{P}_{\ell+1} - \boldsymbol{P}_\ell)(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) = \boldsymbol{0}.$$

$A\Sigma$ is idempotent $\Longrightarrow$ $y^\top A y \sim \chi^2$

Because

$$\left(\underbrace{\boxed{\frac{\boldsymbol{P}_{j+1}-\boldsymbol{P}_j}{\sigma^2}}}_{=A}\right)\underbrace{\boxed{(\sigma^2\boldsymbol{I})}}_{=\Sigma} = \underline{\boldsymbol{P}_{j+1}-\boldsymbol{P}_j}$$

is idempotent,

$$\frac{\boldsymbol{y}^\top(\boldsymbol{P}_{j+1}-\boldsymbol{P}_j)\boldsymbol{y}}{\sigma^2} \sim \chi^2_{r_{j+1}-r_j}\left(\overbrace{\frac{\boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{P}_{j+1}-\boldsymbol{P}_j)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2}}^{ncp}\right)$$

Scaled sums of squares

for all $j = 1, \ldots, m$.

lecture 13

2-19-25