Analysis of Variance

Kenneth Chu Study Notes August 19, 2012

1 One-way Analysis of Variance

Proposition 1.1

$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{k=1}^{n_g}(X_{gk}-X_{\bullet\bullet})^2 &= \frac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{j=1}^{n_g}(X_{gk}-X_{g\bullet})^2 \;+\; \frac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{k=1}^{n_g}(X_{g\bullet}-X_{\bullet\bullet})^2 \\
&= \frac{1}{\sigma^2}\sum_{g=1}^{G} n_g\left(\frac{1}{n_g}\sum_{j=1}^{n_g}(X_{gk}-X_{g\bullet})^2\right) \;+\; \frac{1}{\sigma^2}\sum_{g=1}^{G} n_g\,(X_{g\bullet}-X_{\bullet\bullet})^2
\end{aligned}$$

Remark 1.2

- *Single-mean model (Null Model or Reduced Model):*

$$X_{gk} = \mu + \epsilon_{gk}, \quad X_{gk} \sim N(\mu, \sigma^2)$$

- *Multiple-mean model (Full Model):*

$$X_{gk} = \mu_g + \epsilon_{gk}, \quad X_{gk} \sim N(\mu_g, \sigma^2)$$

- *The dimension of* ***observation space*** *is* $N := \sum_{g=1}^{G} n_g$*, which is simply the number of observed values we have collected. The dimension of the single-mean model space is of course just* 1*. Hence, the dimension of the error subspace for the single-mean model is* $N-1$*.*

- *Similarly, the dimension of the multiple-mean model space is* G*, which is (rather obviously) the number of parameters* $(\mu_1, \mu_2, \ldots, \mu_G)$ *in this model. Consequently, the dimension of the error subspace for the multiple-mean model is* $N-G$*.*

- *Hence, the* ***degree of freedom*** *of the "error reduction" term*

$$\frac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{k=1}^{n_g}(X_{g\bullet}-X_{\bullet\bullet})^2$$

  when going from the single-mean (reduced) model to the multiple-mean (full) model is

$$N-G = (N-1)-(N-G)$$

- *The* F*-statistic is:*

$$F \;=\; \frac{SSBG/\mathrm{df}(SSBG)}{SSWG/\mathrm{df}(SSWG)} \;=\; \frac{\text{null-to-full model sum-of-squared-error reduction}}{\text{square of Euclidean norm of full-model error}} \;\sim\; F^{\,G-1}_{\,N-G}$$

	degree of freedom (dimension of error subspace)	*Euclidean-geometric interpretation*	*common ANOVA nomenclature*
$\dfrac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(X_{gk}-X_{\bullet\bullet}\right)^2$	$N-1$	*square of Euclidean norm of single-mean model error*	*total sum of squares*
$\dfrac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{j=1}^{n_g}\left(X_{gk}-X_{g\bullet}\right)^2$	$N-G$	*square of Euclidean norm of multiple-mean model error*	*within-group sum of squares,* $\dfrac{1}{\sigma^2}SSWG$
$\dfrac{1}{\sigma^2}\sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(X_{g\bullet}-X_{\bullet\bullet}\right)^2$	$G-1$	*error reduction from single- to multiple-mean model*	*between-group sum of squares,* $\dfrac{1}{\sigma^2}SSBG$

References