20 October 2020

Lack of fit

Provide a test statistic for testing

H_0 : the model is fitted well or H_0 : there is no lack of fit

if there are repeated measures on y for the same set of independent variables.

Notation

y_{ij} represent the j^{th} response at the i^{th} experimental combination, $j = 1, 2, \ldots, n_i$ and $i = 1, 2, \ldots, m$;

m is the number of combinations;

n_i is the number of measures on y for the i^{th} combination, for $i = 1, \ldots, m$;

n is the total number of observations, i.e., $\sum_{j=1}^{m} n_i = n$.

Idea

There are two methods to estimate σ^2. One is from the model, i.e., $\hat{\sigma}^2 = \frac{\text{Res.S.S.}}{n-p'}$ and another one is sample variance based on the repeated measures on y, i.e., $\hat{\sigma}^2_{\text{pure error}} = \frac{\text{Pure Error S.S.}}{n-m}$.

$$
\begin{array}{ccccccc}
& \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 & = & \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2 & + & \sum_{i=1}^{m} n_i(\bar{y}_{i.}-\hat{y}_i)^2 \\
& \uparrow & & \uparrow & & \uparrow \\
& \text{Res S.S.} & & \text{Pure error S.S.} & & \text{Lack of fit S.S.} \\
\Rightarrow & \sim\ \sigma^2\chi^2(n-p',\lambda) & & \sim\ \sigma^2\chi^2(n-m) & & \sim\ \sigma^2\chi^2(m-p',\lambda)
\end{array}
$$

where $\lambda = \dfrac{\underset{\sim}{\beta}_2^T(\underset{\sim}{X}_2^T\underset{\sim}{X}_2 - \underset{\sim}{X}_2^T\underset{\sim}{X}_1(\underset{\sim}{X}_1^T\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1^T\underset{\sim}{X}_2)\underset{\sim}{\beta}_2}{\sigma^2}$ since Pure error S.S. & Lack of fit S.S. are independent.

Define the test statistic

$$
\text{F} = \frac{\sum_{i=1}^{m} n_i(\bar{y}_{i.}-\hat{y}_i)^2/(m-p')}{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2/(n-m)} \sim \text{F}(m-p', n-m)
$$

Since E(F) $\approx 1 + \dfrac{\lambda}{m-p'}$, then

$$
\text{Reject } H_0 \text{ if F} > \text{F}_\alpha(m-p', n-m)
$$

.

Remarks

1. Pure error S.S. is equal to $\sum_{i=1}^{m}(n_i-1)S_i^2$ where S_i^2 is the sample variance of repeated measures on y for the i^{th} combination of $\underset{\sim}{x}$.

2. Lack of fit S.S. is calculated as

$$
\text{Lack of fit S.S.} = \text{Res S.S.} - \text{Pure error S.S.}
$$

3. If the null hypothesis can't be rejected, $\hat{\sigma}^2$ is still equal to

$$\frac{\text{Res S.S.}}{n-p'} \ .$$

Formulaes in Sections of "Confidence Interval & Hypothesis Testing" & "Prediction" can be used without any change.

4. If the null hypothesis is rejected, $\hat{\sigma}^2$ is equal to

$$\frac{\text{Pure error S.S.}}{n-m} \ .$$

Formulaes in Sections of "Confidence Interval & Hypothesis Testing" & "Prediction" can be used by using the correct $\hat{\sigma}^2$. The degrees of freedom of $\hat{\sigma}^2$ is $n-m$.