5 November 2020

One categorical variable + one continuous variable

Model I - Regression Model

$$y_i \;=\; \beta_0 + \sum_{j=1}^{m-1} \beta_{g_j} * g_{i,j} + \beta_1 * x_{i1} + \sum_{j=1}^{m-1} \beta_{1g_j} * g_{i,j} * x_{i1} + e_i$$

for $i = 1, \ldots, n$, where $g_{ij} = 1$ if i^{th} observation is in j^{th} level and $g_{ij} = 0$ otherwise. The terms of $g_{i,j} * x_{i1}$, for $j = 1, \ldots, m-1$, are called interaction terms.

Model II - ANCOVA Model

$$y_{ij} = \quad \gamma_{0i} + \gamma_{1i} x_{ij,1} + e_{ij}, \qquad\qquad i = 1, \ldots, m,\; j = 1, \ldots, n_i$$

Write

$$\underset{\sim}{\beta} = \left(\begin{array}{c} \gamma_{01} \\ \gamma_{11} \\ \hline \gamma_{02} \\ \gamma_{12} \\ \hline \vdots \\ \vdots \\ \hline \gamma_{0m} \\ \gamma_{1m} \end{array}\right)$$

then,

$$\underset{\sim}{X}^T \underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1^T \underset{\sim}{X}_1 & \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{X}_2^T \underset{\sim}{X}_2 & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{X}_m^T \underset{\sim}{X}_m \end{pmatrix} \qquad \underset{\sim}{X}^T \underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{X}_1^T \underset{\sim}{Y}_1 \\ \underset{\sim}{X}_2^T \underset{\sim}{Y}_2 \\ \vdots \\ \underset{\sim}{X}_m^T \underset{\sim}{Y}_m \end{pmatrix}$$

$\Rightarrow$

$$\hat{\underset{\sim}{\beta}} \;=\; \left(\begin{array}{c} \hat{\gamma}_{01} \\ \hat{\gamma}_{11} \\ \hline \hat{\gamma}_{02} \\ \hat{\gamma}_{12} \\ \hline \vdots \\ \vdots \\ \hline \hat{\gamma}_{0m} \\ \hat{\gamma}_{1m} \end{array}\right) \;=\; \begin{pmatrix} \left(\underset{\sim}{X}_1^T \underset{\sim}{X}_1\right)^{-1} \underset{\sim}{X}_1^T \underset{\sim}{Y}_1 \\ \left(\underset{\sim}{X}_2^T \underset{\sim}{X}_2\right)^{-1} \underset{\sim}{X}_2^T \underset{\sim}{Y}_2 \\ \vdots \\ \vdots \\ \left(\underset{\sim}{X}_m^T \underset{\sim}{X}_m\right)^{-1} \underset{\sim}{X}_m^T \underset{\sim}{Y}_m \end{pmatrix}$$

and

$$\hat{\sigma}^2 \;=\; \frac{\sum_{i=1}^{m} Res.S.S.|_i}{\sum_{i=1}^{m} (n_i - 2)}$$

As

$$\begin{aligned}
\beta_1 + \beta_{1g_1} &= \gamma_{11} \\
\vdots \quad & \quad \vdots \\
\beta_1 + \beta_{1g_{m-1}} &= \gamma_{1,m-1} \\
\beta_1 &= \gamma_{1m}
\end{aligned}$$

$H_o:\ \beta_{1g_1} = \beta_{1g_2} = \cdots = \beta_{1g_{m-1}} = 0$ (in Model I) is equivalent to $H_0:\ \gamma_{11} = \gamma_{12} = \ldots = \gamma_{1m}$ (in Model II)

For example, when $m = 3$, write $H_0:\ \gamma_{11} = \gamma_{12} = \gamma_{13}$ as

$$H_0: \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma_{01} \\ \gamma_{11} \\ \gamma_{02} \\ \gamma_{12} \\ \gamma_{03} \\ \gamma_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Then, use the test statistic in Section 4.2.3. of Chapter 1 to handle.

Can't reject H_0, then re-write

Model I

$$y_i = \beta_0 + \sum_{j=1}^{m-1} \beta_{g_j} * g_{i,j} + \beta_1 * x_{i1} + e_i$$

for $i = 1, \ldots, n$, where $g_{ij} = 1$ if i^{th} observation is in j^{th} level and $g_{ij} = 0$ otherwise.

Model II

$$y_{ij} = \gamma_{0i} + \beta_1 x_{1,ij} + e_{ij}, \qquad i = 1, \ldots, m,\ j = 1, \ldots, n_i$$

It is suggested to perform estimation or hypothesis testing based on Model I (transformed to centered model).