27 October 2020

One categorical variable

Single-degree-of-freedom Comparison

Contrast: Any linear function of μ_i in the form

$$\omega = \sum_{i=1}^{m} c_i \mu_i, \quad \text{where} \sum_{i=1}^{m} c_i = 0,$$

For testing

$$H_0 : \sum_{i=1}^{m} c_i \mu_i = 0,$$

use the test statistic

$$F = \frac{\left(\sum_{i=1}^{m} c_i \bar{y}_{i.} \right)^2}{\hat{\sigma}^2 \sum_{i=1}^{m} (c_i^2 / n_i)} \sim F_{1, \sum_{i=1}^{m} n_i - m}$$

Define $SSW = \dfrac{\left(\sum_{i=1}^{m} c_i \bar{y}_{i.} \right)^2}{\sum_{i=1}^{m} (c_i^2 / n_i)}$.

Orthogonal contrasts: The two contrasts

$$\omega_1 = \sum_{i=1}^{m} b_i \mu_i \text{ and } \omega_2 = \sum_{i=1}^{m} c_i \mu_i$$

are said to be orthogonal if $\sum_{i=1}^{m} b_i c_i / n_i = 0$ or when the n_i's are all equal to n if $\sum_{i=1}^{m} b_i c_i = 0$.

Then,

$$Reg S.S. = SSW_1 + SSW_2 + \ldots + SSW_{m-1}$$

if these single-degree-of-freedom contrasts $(SSW_1, SSW_2, \ldots, SSW_{m-1})$ are orthogonal to each other.

We normally find one or two highly significant contrasts and all of the remaining contrasts are not significant. Then, we can explain the reason why *Reg.S.S.* is significant through the significant contrasts.

Two categorical variables

Model I (Regression Model)

Categorical variable I (Factor A) a levels $\Rightarrow (a-1)$ dummy variables

Categorical variable II (Factor B) b levels $\Rightarrow (b-1)$ dummy variables

$$y_k \;=\; \beta_0 + \sum_{i=1}^{a-1} \alpha_i * g_{i,k} + \sum_{j=1}^{b-1} \beta_j * c_{j,k} + \sum_{i=1}^{a-1}\sum_{j=1}^{b-1} \gamma_{ij} * g_{i,k} * c_{j,k} + e_k$$

for $k = 1, \ldots, N$, where $g_{i,k} = 1$ if k^{th} observation is in i^{th} level of Factor A and $g_{i,k} = 0$ otherwise; $c_{j,k} = 1$ if k^{th} observation is in j^{th} level of Factor B and $c_{j,k} = 0$ otherwise.

Model II (ANOVA)

$$y_{ijk} \;=\; \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

for $i = 1, 2, \ldots, a,\;\; j = 1, \ldots, b,\;\; k = 1, \ldots, n_{ij},\; \alpha_a = 0,\;\; \beta_b = 0,\;\; \gamma_{aj} = 0$ for each j, $\;\gamma_{ib} = 0$ for each i.

Re-parameterization of Model II

$$\Rightarrow y_{ijk} \;=\; \mu_{ij} + e_{ijk}$$

for $i = 1, 2, \ldots, a,\;\; j = 1, \ldots, b,\;\; k = 1, \ldots, n_{ij}$

Based on Model II,

$$\begin{aligned}
\hat{\mu}_{ij} &= \bar{y}_{ij.} \\
Var(\hat{\mu}_{ij}) &= \frac{\sigma^2}{n_{ij}} \\
Cov(\hat{\mu}_{ij}, \hat{\mu}_{kl}) &= 0 \\
\text{Res.S.S.} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2 \\
\Rightarrow \hat{\sigma}^2 &= \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2}{\sum_{i=1}^{a}\sum_{j=1}^{b} (n_{ij} - 1)}
\end{aligned}$$

$\Rightarrow$

$(1-\alpha)\%$ C.I. for μ_{ij}:

$$\bar{y}_{ij.} \pm t_{\alpha/2, N-ab)} \hat{\sigma} \sqrt{\frac{1}{n_{ij}}}$$

For testing $H_0 : \mu_{ij} = \mu_{ij0}$,

$$t = \frac{\bar{y}_{ij.} - \mu_{ij0}}{\hat{\sigma}\sqrt{\dfrac{1}{n_{ij}}}}$$

Reject H_0 if $|t_{obs}| > t_{\alpha/2, N-ab}$.

<u>Test all population means are equal</u>

$H_o :\ \mu_{ij} = \mu$ (in Model IIA) is equivalent to $H_0 :\ \alpha_i = 0, \beta_j = 0, \gamma_{ij} = 0$ (in Model I)

$$\begin{aligned}
\text{Res. S. S.} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{ij.})^2 \quad \text{with d.f. } = \sum_{i=1}^{a}\sum_{j=1}^{b}(n_{ij} - 1) \\
\text{Total S. S.} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}}(y_{ijk} - \bar{y}_{...})^2 \quad \text{with d.f. } = \sum_{i=1}^{a}\sum_{j=1}^{b}n_{ij} - 1
\end{aligned}$$

$\Rightarrow$ Reg.S.S. = Total S.S. - Res.S.S. with d.f. $= ab - 1$.

From Section 4 of Chapter 1,

$$F = \frac{\text{Reg.S.S.}/(ab-1)}{\text{Res.S.S.}/(N-ab)}$$

Reject H_0 if $F_{obs} > F_\alpha(ab-1, N-ab)$