One-way ANOVA with random effects

Suppose we have a one-way classified data which is to be used to study the effects of a factor A on the variable of interest Y. Now, let us consider the situation that the factor A has very large number of levels, so, it is not possible to incorporate all of them. But we still want to test whether the effects of all the levels are equal or not. Therefore, we can include only a sample of these levels, and we want to test about all the levels, whether included in the experiment or not, based on the available data. The data consists only those sampled levels. So, the effects, included in the sample, will become random as they are considered as a random sample from the large population of effects.

So, in this case, the model is

$$y_{ij} = \mu + \beta_i + e_{ij}, \quad i = 1(1)k, \; j = 1(1)r \qquad \text{(2)}$$

where y_{ij} denotes the j^{th} observation corresponding to i^{th} level of the factor A,

μ denotes the general effect,

β_i denotes the random (additional) effect representing the i^{th} level of factor A,

e_{ij} are the random errors, independent to μ and β_i's.

Here, we want to test H_0: all levels in the population are equal.

against H_1: not H_0.

Let us take the following assumptions:

(i) $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$

(ii) $\beta_i \overset{iid}{\sim} N(0, \sigma_\beta^2)$

Hence, we actually test $H_0: \sigma_\beta^2 = 0$ against $H_1: \sigma_\beta^2 > 0$.

($\because$ $Var(\beta) = \sigma_\beta^2 = 0$ implies $\beta_1 = \beta_2 = \cdots\cdots$)

Now, from (2) we have

$$\bar{y}_{i0} = \mu + \beta_i + \bar{e}_{i0} \quad , \text{ where } \bar{e}_{i0} = \frac{1}{r}\sum_{j=1}^{r} e_{ij}$$

$$\bar{\bar{y}}_{00} = \mu + \beta_0 + \bar{\bar{e}}_{00} \quad , \text{ where } \beta_0 = \frac{1}{k}\sum_{i=1}^{k}\beta_i$$

$$\Rightarrow \bar{y}_{i0} - \bar{\bar{y}}_{00} = \beta_i - \beta_0 + \bar{e}_{i0} - \bar{\bar{e}}_{00} \qquad \bar{\bar{e}}_{00} = \frac{1}{k}\sum_{i=1}^{k}\bar{e}_{i0}.$$

Considering the randomness of β_i's and e_{ij}'s., we have from (2) that

$$Var(y) = Var(\beta) + Var(e)$$

$$\Rightarrow \sum_i\sum_j (y_{ij} - \bar{\bar{y}}_{00})^2 = r\sum_i (\bar{y}_{i0} - \bar{\bar{y}}_{00})^2 + \sum_i\sum_j (y_{ij} - \bar{y}_{i0})^2$$

$$\Rightarrow \quad TSS \quad = \quad SSA \quad + \quad SSE.$$

<u>Distributions of SSA and SSE</u>

Since $\beta_i \sim N(0, \sigma_\beta^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$

$\bar{y}_{i0} = \mu + \beta_i + \bar{e}_{i0} \sim$ Normal distn with

$$E(\bar{y}_{i0}) = \mu \quad , \quad Var(\bar{y}_{i0}) = \sigma_\beta^2 + \frac{\sigma_e^2}{r} = \delta^2 \text{ (say)}$$

So, $$T_i = \frac{\bar{y}_{i0} - \bar{\bar{y}}_{00}}{\delta} \sim \text{Normal distn}$$

with $$E(T_i) = \frac{1}{\delta_i}\left[E(\bar{y}_{i0}) - E(\bar{\bar{y}}_{00})\right] = 0$$

$$Var(T_i) = 1 \quad , \forall i = 1(1)k.$$

$$\therefore \frac{SSA}{r\sigma_\beta^2 + \sigma_e^2} = \sum_{i=1}^{k} T_i^2 = \frac{\sum_i (\bar{y}_{i0} - \bar{\bar{y}}_{00})^2}{\delta^2} \sim \chi^2$$

with d.f. = $k-1$.

$$\text{So, } E(SSA) = (k-1)(r\sigma_\beta^2 + \sigma_e^2)$$

$$\Rightarrow E(MSA) = r\sigma_\beta^2 + \sigma_e^2 \quad , \text{ where } MSA = \frac{SSA}{k-1}.$$

$$\frac{SSE}{\sigma_e^2} = \frac{1}{\sigma^2}\sum_i\sum_j (y_{ij} - \bar{y}_{i0})^2 = \frac{1}{\sigma_e^2}\sum_i\sum_j (e_{ij} - \bar{e}_{i0})^2$$

$$= \sum_i \left\{ \sum_j \frac{(e_{ij} - \bar{e}_{i0})^2}{\sigma_e^2} \right\}$$

$$= \sum_i u_i \text{ (say).}$$

where, $u_i \sim \chi^2_{r-1}$ using the Result 1.

$\therefore \frac{SSE}{\sigma_e^2} \sim \chi^2_{\sum_{i=1}^{k}(r-1)} \equiv \chi^2_{n-k}$ since $n = rk$.

So, $E(SSE) = \sigma_e^2 (n-k)$.

$\Rightarrow E(MSE) = \sigma_e^2$, where $MSE = \frac{SSE}{n-k}$

$\Rightarrow$ MSE is an unbiased estimator of σ_e^2.

We already have that $E(MSA) = r\sigma_\beta^2 + \sigma_e^2$

$\Rightarrow E\left(\frac{MSA - MSE}{r}\right) = \sigma_\beta^2$

$\Rightarrow \frac{1}{r}(MSA - MSE)$ is an unbiased estimator of σ_β^2.

Note that when H_0 is true, $E(MSA) = E(MSE) = \sigma_e^2$.

Therefore we define the following test-statistic for testing H_0:

$$F_0 = \frac{MSA}{MSE} = \frac{SSA/k-1}{SSE/n-k} \sim F_{k-1, n-k}$$

H_0 will be rejected at given α if the observed $F_0 > F_{\alpha;\overline{k-1},\overline{n-k}}$, where $F_{\alpha;\overline{k-1},\overline{n-k}}$ represents the upper α-point of the F-distn with d.f. $(k-1)$ and $(n-k)$.

ONE-WAY ANOVA Table with random effects

Sources of Variations	d.f.	SS	MS	F_0	F at level 1% or 5%
Between Classes	$k-1$	$SSA = r\sum_i (\bar{y}_{i0} - \bar{\bar{y}}_{00})^2$	$MSA = \frac{SSA}{k-1}$	$F_0 = \frac{MSA}{MSE}$	At 1%, $F_{0.01;\overline{k-1},\overline{n-k}}$
Error	$n-k$	$SSE = \sum_i\sum_j (y_{ij} - \bar{y}_{i0})^2$	$MSE = \frac{SSE}{n-k}$		At 5%, $F_{0.05;\overline{k-1},\overline{n-k}}$
Total	$n-1$	$TSS = \sum_i\sum_j (y_{ij} - \bar{\bar{y}}_{00})^2$	—	—	—

Remark 1. In the one-way random effects model, the observations have the same expectation μ and same variance, $\sigma_y^2 = \sigma_\beta^2 + \sigma_e^2$. However, the observations are not statistically independent. This dependence can be expressed in terms of the intra-class correlation coefficient, which is nothing but the ordinary correlation coefficient between any two observations, y_{ij} and $y_{ij'}$ $(j \neq j')$, of the same class:

$$\rho = \frac{E(y_{ij}-\mu)(y_{ij'}-\mu)}{E(y_{ij}-\mu)^2} = \frac{E[(b_i+e_{ij})(b_i+e_{ij'})]}{\sigma_y^2}$$

$$= \frac{E(b_i^2) + E[b_i(e_{ij}+e_{ij'})] + E(e_{ij}.e_{ij'})}{\sigma_Y^2}$$

$$= \frac{E(b_i^2)+0}{\sigma_Y^2} = \frac{Var(b_i)}{\sigma_Y^2} \quad (\because E(b_i)=0 \ \forall i)$$

$$= \frac{\sigma_\beta^2}{\sigma_Y^2} = \frac{\sigma_\beta^2}{\sigma_\beta^2+\sigma_e^2}$$

Result: (Fundamental, Vol. 2, Page-59, Problem No. 1.14)
For the random effects model ②, the following is a consistent estimator of the intra-class correlation coefficient

$$\rho = \frac{\sigma_\beta^2}{\sigma_y^2}:$$

$$\hat{\rho} = \frac{MSA - MSE}{r\left[MSE + \frac{1}{r}(MSA-MSE)\right]} = \frac{MSA - MSE}{MSA + (r-1)\,MSE}.$$

Proof. Hint. Show that $\hat{\sigma}_\beta^2$ and $\hat{\sigma}_Y^2$ is consistent for σ_β^2 and σ_Y^2, respectively. Then using an appropriate result [i.e. if $\hat{\theta}$ is consistent for θ, then for any one-to-one function $g(\cdot)$, $g(\hat{\theta})$ is consistent for $g(\theta)$], complete the proof.