One-way ANOVA with fixed effects:

Analysis of one-way classified data is the analysis of the data taken from an experiment which studies the effects of the factor on variable of interest Y.

The one-way ANOVA model with fixed-effects is

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1(1)k, \; j = 1(1)n_i \quad \text{—— (1)}$$

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

μ denotes the general effect,

α_i denotes the additional fixed effect corresponding to i^{th} class or level,

e_{ij} are the random errors.

Here, $\mu = \frac{1}{n}\sum_{i=1}^{k} n_i \mu_i$, where $\mu_i = \mu + \alpha_i$ i.e. $\alpha_i = \mu_i - \mu$ $\forall i = 1(1)k$.

So, $\sum_i n_i \mu_i = \mu n. \Rightarrow \sum_i n_i(\mu + \alpha_i) = \mu n$

$\Rightarrow \mu n + \sum_i n_i \alpha_i = \mu n$

$\Rightarrow \sum_i n_i \alpha_i = 0.$

Assumption: $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2) \quad \forall i, j.$

Here, we have to test $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

vs. H_1: atleast one inequality holds in H_0.

Atfirst, we have to estimate the model parameters.

Least-square estimators of μ and α_i are obtained by minimising

$\sum_i\sum_j e_{ij}^2 = \sum_i\sum_j (y_{ij} - \mu - \alpha_i)^2$ w.r.t. μ and α_i, respectively.

Solving the normal equations

$$\sum_i\sum_j y_{ij} = n\mu + \sum_i n_i \alpha_i \quad \text{—— (2)}$$

$$\sum_j y_{ij} = n_i\mu + n_i\alpha_i \quad \forall i = 1(1)k \quad \text{—— (3)}$$

we have the estimates $\hat{\mu} = \frac{1}{n}\sum_i\sum_j y_{ij} = \bar{\bar{y}}_{00}$, $\hat{\alpha}_i = \bar{y}_{i0} - \bar{\bar{y}}_{00} \; \forall i.$

Hence, model in ① can be written as

$$y_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{e}_{ij}$$
$$= \bar{\bar{y}}_{oo} + (\bar{y}_{io} - \bar{\bar{y}}_{oo}) + (y_{ij} - \bar{y}_{io})$$

So, $(y_{ij} - \bar{\bar{y}}_{oo}) = (\bar{y}_{io} - \bar{\bar{y}}_{oo}) + (y_{ij} - \bar{y}_{io})$

$$\Rightarrow (y_{ij} - \bar{\bar{y}}_{oo})^2 = (\bar{y}_{io} - \bar{\bar{y}}_{oo})^2 + (y_{ij} - \bar{y}_{io})^2 + 2(\bar{y}_{io} - \bar{\bar{y}}_{oo})(y_{ij} - \bar{y}_{io})$$

$$\Rightarrow \sum_i \sum_j (y_{ij} - \bar{\bar{y}}_{oo})^2 = \sum_i n_i (\bar{y}_{io} - \bar{\bar{y}}_{oo})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{io})^2$$

[3rd term in RHS is 0 since $\sum_j (y_{ij} - \bar{y}_{io}) = 0$]

$\Rightarrow$ Total sum of squares (TSS)

= sum of squares due to class effects (SSA)

+ sum of squares due to error (SSE)

<u>Distributions of SSA & SSE under H_0</u>

From ①, $\bar{y}_{io} = \mu + \alpha_i + \bar{e}_{io}$

$\bar{\bar{y}}_{oo} = \mu + \bar{e}_{oo}$ (since, $\sum_i^k n_i \alpha_i = 0$)

$\therefore$ $\bar{y}_{io} - \bar{\bar{y}}_{oo} = \alpha_i + \bar{e}_{io} - \bar{\bar{e}}_{oo}$

Since $e_{ij} \sim N(0, \sigma_e^2)$, $\bar{e}_{io} \sim N(0, \frac{\sigma_e^2}{n_i})$

So, $\bar{y}_{io} = \mu + \alpha_i + \bar{e}_{io} \sim$ Normal dist^n

with $E(\bar{y}_{io}) = E(\mu + \alpha_i + \bar{e}_{io}) = \mu + \alpha_i$

$Var(\bar{y}_{io}) = Var(\bar{e}_{io}) = \frac{\sigma_e^2}{n_i}$

So, $T_i = \frac{(\bar{y}_{io} - \bar{\bar{y}}_{oo})\sqrt{n_i}}{\sigma_e} \sim$ Normal dist^n

with mean $= \mu + \alpha_i - \mu = \alpha_i$ ($\because \bar{\bar{y}}_{oo}$ is an u.e. of μ)

variance $= 1$. , $i = 1(1)k$

$$\therefore \frac{SSA}{\sigma_e^2} = \sum_i T_i^2 = \sum_i \frac{n_i (\bar{y}_{io} - \bar{\bar{y}}_{oo})^2}{\sigma_e^2} \sim \chi'^2$$

with d.f. $= k-1$ and n.c.p $= \frac{1}{\sigma_e^2} \sum_i n_i \alpha_i^2$

(using the Result stated in the next page)

So, $E(SSA) = \sigma_e^2\left(\overline{k-1} + \frac{1}{\sigma_e^2}\sum_i n_i\alpha_i^2\right) = \sigma_e^2(k-1) + \sum_i n_i\alpha_i^2$

Now, define $MSA = \dfrac{SSA}{\text{d.f. of } SSA} = \dfrac{SSA}{k-1}$

$\therefore\ E(MSA) = \sigma_e^2 + \frac{1}{k-1}\sum_i n_i\alpha_i^2 \geq \sigma_e^2$

Under H_0, $E(MSA) = \sigma_e^2$ $\quad (\because H_0: \alpha_i = 0 \ \forall\ i = 1(1)k)$

equivalently, MSA is an unbiased estimator of σ_e^2.

Again, $y_{ij} - \bar{y}_{i0} = \mu + \alpha_i + e_{ij} - (\mu + \alpha_i + \bar{e}_{i0}) = e_{ij} - \bar{e}_{i0}$

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

~~$\Rightarrow e_{ij} - \bar{e}_{i0} \sim$ Normal dist^n with~~

~~mean~~ $= E(e_{ij} - \bar{e}_{i0}) = 0$

~~variance~~ $\cancel{= Var(e_{ij} - \bar{e}_{i0})}$

$\cancel{= V(e_{ij}) + V(\bar{e}_{i0}) - 2\,cov(e_{ij}, \bar{e}_{i0})}$

$\cancel{= \sigma_e^2 + \frac{\sigma_e^2}{n_i} - \frac{2}{n_i} cov\left(e_{ij}, \sum_j e_{ij}\right)}$

$\cancel{= \sigma_e^2\left(1 - \frac{1}{n_i}\right)}$

Result: If $x_1, x_2, \dots, x_n \overset{iid}{\sim} N(\theta, \sigma^2)$, where θ and σ^2 both are unknown. Then,

$$T = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \chi^2_{n-1}, \quad \text{where } E(\bar{x}) = \theta$$

If $E(x_i - \bar{x}) = \delta_i \neq 0$, then $T \sim \chi'^2$ with d.f $= n-1$, n.c.p $= \frac{1}{\sigma^2}\sum_i \delta_i^2$.

Using the above Result,

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

$$= \sum_i \left\{ \frac{\sum_{j=1}^{n_i}(e_{ij} - \bar{e}_{i0})^2}{\sigma_e^2} \right\}$$

$$= \sum_i u_i, \quad \text{where } u_i \sim \chi^2_{n_i - 1}$$

$$\sim \chi^2_{\sum_{i=1}^{k}(n_i - 1)} \equiv \chi^2_{n-k}$$

$\therefore\ \dfrac{SSE}{\sigma_e^2} \sim \chi^2_{n-k} \Rightarrow E(SSE) = \sigma_e^2(n-k)$.

So, $MSE = \dfrac{SSE}{d.f.\text{ of } SSE} = \dfrac{SSE}{n-k}$

$\therefore E(MSE) = \sigma_e^2$

So, MSE is always an unbiased estimator of σ_e^2.

If H_0 is true, $E(MSA) = E(MSE)$.

Since, under H_0, $\dfrac{SSA}{\cancel{\sigma_e^2}} \sim \chi^2_{k-1}$, independent of SSE

Then, we define $F_0 = \dfrac{MSA}{MSE} = \dfrac{(SSA/k-1)}{(SSE/n-k)} \sim F_{k-1,\, n-k}$

Thus, H_0 is rejected at specified level α if for the given values

$$F_0 = \frac{MSA}{MSE} > F_{\alpha;\, \overline{k-1},\, \overline{n-k}},$$

where $F_{\alpha;\, \overline{k-1},\, \overline{n-k}}$ is the upper α-point of the F-distn with d.f. $= (k-1, n-k)$

ONE-WAY ANOVA Table

Source of variation	df	SS	MS	F_0	F at level 1% or 5%
Between Classes	$k-1$	$SSA = \sum_i n_i (\bar{y}_{i0} - \bar{\bar{y}}_{00})^2$	$MSA = \dfrac{SSA}{k-1}$	$\dfrac{MSA}{MSE}$	At 1%, $F_{0.01}: (k-1), (n-k)$
Error	$n-1$	$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i0})^2$	$MSE = \dfrac{SSE}{n-k}$		At 5%, For $F_{0.05}: (k-1), (n-k)$,