Assumption: $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

21st April 2023
DS

Test of hypothesis. $H_0 : \alpha_1 = \dots = \alpha_p = 0$ (no differential effects) ag. $H_1 :$ (not H_0) ($\exists$ at least 1 diffn effect)

[If H_0 is true, then all levels of A will generate the same value of the response on an average. This means that all levels of A exert the same influence on the response variable that is there is no differential effect among the levels of A. That is why, the test for H_0 is called the test for differential effects among the levels.

The alternative hypothesis is

$H_1 : \exists$ diff effects among levels of A

i.e. $H_1 :$ all α_i's are not zero which means

$H_1 :$ at least one $\alpha_i \neq 0$

i.e $H_1 :$ (not H_0)

$H_0 : \alpha_1 = \dots = \alpha_p = 0$ ag. $H_1 :$ not H_0

The test is carried out as follows:

$$\left.\begin{array}{cccc} \underline{A_1} & \underline{A_2} & \cdots & \underline{A_p} \\ y_{11} & y_{21} & \cdots & y_{p1} \\ \vdots & \vdots & & \vdots \\ y_{1n_1} & y_{2n_2} & & y_{pn_p} \end{array}\right\} \text{Data.}$$

let us define $\bar{y}_i$ = mean of observations of the i^{th} level of A.

$$\bar{y}_1 = \text{mean of } y_{11}, y_{12}, \dots, y_{1n_1} = \frac{1}{n_1}\sum_{j=1}^{n_1} y_{1j}$$

$$\bar{y}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_{2j}$$

Hence in general $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \quad \forall\ i = 1(1)p$

and

$\bar{y} :=$ grand mean $=$ mean of all obs,

$$= \left(\sum_{i=1}^{p} n_i\right)^{-1}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}\right)$$

If H_0 is true, then the group means should be close to each other (ideally they should be equal, but since this is sample, they will be close to each other) otherwise it will be quite different.

We define SSA = sum of squares due to the levels of A

$$= \sum_{i=1}^{p} n_i (\bar{y}_i - \bar{y})^2$$

Note that $\bar{y}$, the grand mean, is the mean of the group means $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p$

The quantity can be interpreted as some sort of variation among the group means $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p$.

If H_0 is true, SSA will take values close to 0 o/w SSA will be large values.

Hence large values of SSA will indicate departure from H_0.

$\therefore$ The test statistic for H_0 will be based on the quantity SSA.

An appropriate test statistic for testing H_0 is given by

$$F_A = \frac{SSA}{SSE} \cdot \frac{(n-p)}{(p-1)} \quad \text{where } SSE = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$\therefore F_A = \frac{\sum_{i=1}^{p} n_i (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2} \cdot \frac{(n-p)}{(p-1)}$$

and we know that

$$\frac{SSE}{\sigma^2} \sim \chi^2_{p-1} \quad \text{and} \quad \frac{SSA}{\sigma^2} \sim \chi^2_{n-p}.$$

Thus, the test statistic can be rewritten as

$$F_A = \frac{SSA/(p-1)}{SSE/(n-p)} \equiv$$

Ratio of two indep χ^2 divided by their resp dof.

Hence, under H_0: $F_A \sim F_{p-1, n-p}$

High values of F_A indicate departure from H_0. Thus the critical region takes the form

$(F_A)_{obs} > k$ where k is so chosen that $P[\text{type I error}] = \alpha$

Now $P(\text{Type I error}) = \alpha$

$P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$

$P((F_A)_{obs} > k \mid F_A \sim F_{p-1; n-p}) = \alpha$

$\Rightarrow k$: upper α-point of $F_{p-1; n-p}$ dist

$\Rightarrow k = F_{p-1; n-p; \alpha}$

Test rule: Reject H_0 at α-level of significance iff $(F_A)_{obs} > F_{p-1; n-p; \alpha}$

The results of the above testing problem are summarized in an ANOVA ~~tabla~~ table as follows :

The mean square error MS is defined as $\frac{SS}{dof}$

$$\therefore MSE = \frac{SSE}{dof(SSE)} = \frac{SSE}{n-p}$$

$$\text{and} \quad MSA = \frac{SSA}{dof(SSA)} = \frac{SSA}{p-1}$$

So that $F_A = \frac{MSA}{MSE}$. The table is

ANOVA Table

Sources of variation	df	SS	MS	F - obs	Crit pt
Factor (Here drug)	$p-1$	SSA	MSA	$F_A = \frac{MSA}{MSE}$	$F_{p-1; n-p; \alpha}$
Error	$n-p$	SSE	MSE		
Total	$n-1$	TSS	—		

Conclusion : Decision taken.

Tests done

1. Test of mean of normal dist^n
   (a) σ^2 known [pop^n σ^2] — Z - test
   (b) σ^2 unknown [pop^n σ^2] — t_{n-1} test

2. Test of variance of normal dist^n
   (a) μ is known [pop^n μ] — χ^2_n test
   (b) μ is unknown [pop^n μ] — χ^2_{n-1} test

3. Test of equality of 2 means of 2 IID normal dist^n
   (a) variances are known [pop^n var's] — Z - test
   (b) " " unknown [pop^n var's] — $t_{n_1+n_2-2}$

4. Test of equality of 2 variances of 2 IID normal dist^n
   (a) when means are known [pop^n μ's] — $F_{n_1; n_2}$
   (b) when means are unknown [pop^n μ's] — $F_{n_1-1; n_2-1}$

5. ANOVA 1-way problem — $F_{p-1; n-p}$