Ch.4 One-way classification: Random effects model (One-way ANOVA: Random factor effects model)

1 Assumptions and model

1.1 Random cell means model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, \ldots, r, \quad j = 1, \ldots, n_i$$

$\mu_1, \ldots, \mu_r$ are independent $N(\mu., \sigma_\mu^2)$;
Error terms $\varepsilon_{ij}, \quad i = 1, \ldots, r, \quad j = 1, \ldots, n_i$, are independent $N(0, \sigma^2)$;
$\{\mu_k, \quad k = 1, \ldots, r\}$ and $\{\varepsilon_{ij}, \quad i = 1, \ldots, r, \quad j = 1, \ldots, n_i\}$ are independent

Important features of model

1° The response variable Y_{ij} is normally distributed with

$$\begin{aligned} EY_{ij} &= \mu., \\ Var(Y_{ij}) &= \sigma_\mu^2 + \sigma^2. \end{aligned}$$

2° Unlike for the fixed factor levels model where all observations Y_{ij} are independent, the Y_{ij} for the random factor levels model are only independent if they pertain to different factor levels.

$$Cov(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_\mu^2 + \sigma^2, & i = i', j = j', \\ \sigma_\mu^2, & i = i', j \neq j', \\ 0, & i \neq i'. \end{cases}$$

1.2 Questions of interest

1° About σ_μ^2:

Point estimation of σ_μ^2;

Interval estimation of σ_μ^2;

Test whether $\sigma_\mu^2 = 0$; ($\sigma_\mu^2 = 0$ implies that all μ_i are equal, i.e., $\mu_i \equiv \mu.$)

Estimation of $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}$, which is the correlation coefficient between two observations Y_{ij} and $Y_{ij'}$ from the same factor level.

2° About $\mu.$:

Point estimation;

Interval estimation.

1.3 Notations

$$n_T \triangleq \sum_{i=1}^{r} n_i \qquad \text{(The total number of cases in the study)}$$

$$\bar{Y}_{i\cdot} \triangleq \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{(Sample mean of the i-th factor level)}$$

$$\bar{Y}_{\cdot\cdot} \triangleq \frac{1}{n_T}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_T}\sum_{i=1}^{r} n_i\bar{Y}_{i\cdot} \qquad \text{(The overall mean for all responses)}$$

$$SS_B \triangleq \sum_{i=1}^{r} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \qquad \text{(Sum of squares ``between groups'' (Treatment sum of squares))}$$

$$SS_W \triangleq \sum_{i=1}^{r}\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 \qquad \text{(Sum of squares ``within groups'' (Error sum of squares))}$$

$$SS_T \triangleq \sum_{i=1}^{r}\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 \qquad \text{(Total sum of squares)}$$

Breakdown of degrees of freedom

$$\underbrace{\mathrm{SS_T}}_{(n_T-1 \text{ degrees of freedom})} = \underbrace{\mathrm{SS_W}}_{(n_T-r \text{ degree of freedom})} + \underbrace{\mathrm{SS_B}}_{(r-1 \text{ degrees of freedom})}$$

2 Inference on σ^2

2.1 Test whether $\sigma_\mu^2 = 0$

$$\mathcal{H}_0 : \sigma_\mu^2 = 0 \quad \longleftrightarrow \quad \mathcal{H}_a : \sigma_\mu^2 > 0$$

If $\mathcal{H}_0$ then all μ_i are equal.

- One-way ANOVA Table (for the balanced case)

Source of Variation	df	Sum of Squares (SS)	Mean Square (MS)	F^*-value	Pr($> F^*$) (p-value)
Between treatments	$r-1$	SS_B	$MS_B = \frac{SS_B}{r-1}$	$F^* = \frac{MS_B}{MS_W}$	
Error or Residuals (within treatments)	$r(n-1)$	SS_W	$MS_W = \frac{SS_W}{r(n-1)}$		
Total	$rn-1$	SS_T			

- Derivation

$$\begin{aligned}
SS_W &\triangleq \sum_{i=1}^r \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i\cdot})^2 = \sum_{i=1}^r \sum_{j=1}^n \left\{\mu_i + \varepsilon_{ij} - \frac{1}{n}\sum_{j=1}^n (\mu_i + \varepsilon_{ij})\right\}^2 = \sum_{i=1}^r \sum_{j=1}^n \{\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}\}^2 \\
SS_B &\triangleq \sum_{i=1}^r n(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum_{i=1}^r n\Big\{\frac{1}{n}\sum_{j=1}^n (\mu_i + \varepsilon_{ij}) - \frac{1}{rn}\sum_{i=1}^r \sum_{j=1}^n (\mu_i + \varepsilon_{ij})\Big\}^2 \\
&= \sum_{i=1}^r n\Big\{(\mu_i - \bar{\mu}_\cdot) + (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})\Big\}^2
\end{aligned}$$

$$SS_W \text{ and } SS_B \text{ are independent}, \quad \frac{SS_W}{\sigma^2} \sim \chi^2_{r(n-1)}, \qquad \frac{SS_B}{\sigma^2 + n\sigma_\mu^2} \sim \chi^2_{r-1}.$$

Under $\mathcal{H}_0$,

$$F^* = \frac{MS_B}{MS_W} \sim F_{r-1, r(n-1)}$$

Apex Enterprises Example, page 1036.

Set up the data frame, as in the fixed factor levels case.

```
> y = read.table("CH25TA01.DAT")
> y
    V1 V2 V3
1   76  1  1
     ......
20  79  5  4

>  data = y[,1]
>  officer = factor(rep(LETTERS[1:5],c(4,4,4,4,4)))
>  candidate.df = data.frame(officer,data)
```

Look at the data set graphically.

```
> plot(officer,data)
> plot(candidate.df)
```

The ANOVA table is obtained by

```
> summary(aov(data~officer,candidate.df))
               Df   Sum Sq   Mean Sq  F value   Pr(>F)
    officer     4   1579.70  394.93   5.389     0.006803
     Residuals  15  1099.25   73.28
```

What's your conclusion?

2.2 Estimation of σ_μ^2

- An unbiased estimator of σ_μ^2:

$$\hat{\sigma}_\mu^2 \triangleq \frac{MS_W - MS_B}{n}$$

Occasionally, $\hat{\sigma}_\mu^2 < 0$.

- It is not possible to construct an exact CI for σ_μ^2.

2.3 Confidence interval for $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}$

A pivotal quantity

$$\frac{MS_W/(\sigma_\mu^2+\sigma^2)}{MS_B/\sigma^2} \sim F_{r-1,r(n-1)}$$

The probability statement for $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}$:

$$P\Big(\frac{L}{1+L} \le \frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2} \le \frac{U}{1+U}\Big) = 1-\alpha,$$

where

$$L = \frac{1}{n}\Big[\frac{MS_W}{MS_B}\Big(\frac{1}{qf(1-\alpha/2;r-1,r(n-1))}\Big)-1\Big],$$
$$U = \frac{1}{n}\Big[\frac{MS_W}{MS_B}\Big(\frac{1}{qf(\alpha/2;r-1,r(n-1))}\Big)-1\Big].$$

Question: Is it possible to obtain $L < 0$? What are you going to do in such a case?

3 Inference on $\mu.$

- An unbiased estimator of $\mu.$:

$$\hat{\mu}. \triangleq \bar{Y}_{..} \sim N\Big(\mu., \frac{n\sigma_\mu^2+\sigma^2}{rn}\Big).$$

- A pivotal quantity

$$\frac{\bar{Y}_{..}-\mu.}{MS_W/rn} \sim t_{r-1}$$

The $1-\alpha$ CI for $\mu.$

$$\bar{Y}_{..} \pm qt(1-\alpha/2;r-1)\frac{MS_W}{rn}$$