fixed effect v.s. random effect

Random Effects Models

Hao Zhang

Introduction

Example. Solutions of alcohol are used for calibrating Breathalyzers. The following data show the alcohol concentrations of samples of alcohol solutions taken from six bottles of alcohol solution randomly selected from a large batch. The objective is to determine if all bottles in the batch have the same alcohol concentrations.

Bottle	Concentration			
1	1.4357	1.4348	1.4336	1.4309
2	1.4244	1.42321	1.42131	1.4256
3	1.4153	1.4137	1.4176	1.4164
4	1.4331	1.4325	1.4312	1.4297
5	1.4252	1.4261	1.4293	1.4272
6	1.4179	1.4217	1.4191	1.4204

Conclusion will be applied to the population of our bottles.

We are not interested in any difference between the six bottles used in the experiment. We therefore treat the six bottles as a random sample from the population and use a random effects model. If we were interested in the six bottles, we would use a fixed effects model.

One Random Effect

The Random-Effects One-Way Model

For a completely randomized design, with v randomly selected levels of a treatment factor T, the random-effects one-way model is

$$\begin{gathered} Y_{it} = \mu + T_i + \epsilon_{it}, \\ \epsilon_{it} \sim N\left(0, \sigma^2\right), \quad T_i \sim N\left(0, \sigma_T^2\right), \end{gathered}$$

ϵ_{it} 's and T_i 's are all mutually independent ,

$$t = 1, \ldots, r_i, \quad i = 1, \ldots, v.$$

Note the model parameters are μ, σ^2 and σ_T^2.

$$E\left[Y_{it}\right] = E[\mu] + E\left[T_i\right] + E\left[\epsilon_{it}\right] = \mu.$$

The variance of Y_{it} is

Sum of the 2 variances

$$\operatorname{Var}\left(Y_{it}\right) = \operatorname{Var}\left(\mu + T_i + \epsilon_{it}\right) = \operatorname{Var}\left(T_i\right) + \operatorname{Var}\left(\epsilon_{it}\right) + 2 \operatorname{Cov}\left(T_i, \epsilon_{it}\right) = \sigma_T^2 + \sigma^2,$$

since T_i and ϵ_{it} are mutually independent and so have zero covariance. Therefore, the distribution of Y_{it} is

$$Y_{it} \sim N\left(\mu, \sigma_T^2 + \sigma^2\right).$$

The two components σ_T^2 and σ^2 of the variance of Y_{it} are known as variance components. Observations on the same treatment are correlated, with

$$\operatorname{Cov}(Y_{it}, Y_{is}) = \operatorname{Cov}(\mu + T_i + \epsilon_{it}, \mu + T_i + \epsilon_{is}) = \operatorname{Var}(T_i) = \sigma_T^2.$$

Estimation of σ^2

$$\text{SSE} = \sum_{i=1}^{v} \sum_{t=1}^{r_i} Y_{it}^2 - \sum_{i=1}^{v} r_i \bar{Y}_{i.}^2.$$

Remember that the variance of a random variable X is calculated as $\operatorname{Var}(X) = E\left[X^2\right] - (E[X])^2$. So, we have

$$E\left[Y_{it}^2\right] = \operatorname{Var}(Y_{it}) + \left(E\left[Y_{it}\right]\right)^2 = \left(\sigma_T^2 + \sigma^2\right) + \mu^2.$$

Now,

$$\bar{Y}_{i.} = \mu + T_i + \frac{1}{r_i} \sum_{t=1}^{r_i} \epsilon_{it},$$

so

$$\operatorname{Var}\left(\bar{Y}_{i.}\right) = \sigma_T^2 + \frac{\sigma^2}{r_i} \text{ and } E\left[\bar{Y}_{i.}\right] = \mu.$$

Consequently,

$$E\left[\bar{Y}_{i.}^2\right] = \left(\sigma_T^2 + \frac{\sigma^2}{r_i}\right) + \mu^2.$$

Thus,

$$\begin{aligned} E[SSE] &= \sum_{i=1}^{v} \sum_{t=1}^{r_i} \left(\sigma_T^2 + \sigma^2 + \mu^2\right) - \sum_{i=1}^{v} r_i \left(\sigma_T^2 + \frac{\sigma^2}{r_i} + \mu^2\right) \\ &= n\sigma^2 - v\sigma^2 \quad \left(\text{where } n = \sum_{i=1}^{v} r_i\right) \\ &= (n-v)\sigma^2, \end{aligned}$$

giving

$$E[MSE] = E[SSE/(n-v)] = \sigma^2.$$

Therefore the MSE is an unbiased estimator of σ^2.

We can show that SSE/σ^2 has a χ_{n-v}^2 distribution. Hence the confidence bound for σ^2 can be computed as under fixed-effects models, that is,

$$\boxed{\sigma^2 \leq \frac{ssE}{\chi_{n-v,1-\alpha}^2},}$$

where $\chi_{n-v,1-\alpha}^2$ is the percentile of the chi-squared distribution with $n-v$ degrees of freedom and with probability of $1-\alpha$ in the right-hand tail.

Estimation of σ_T^2

$$SST = \sum_{i=1}^{v} r_i \bar{Y}_{i.}^2 - n\bar{Y}_{..}^2$$

Using the same type of calculation as in Sect. 17.3.2 above, we have

$$\bar{Y}_{..} = \mu + \frac{1}{n}\sum_i r_i T_i + \frac{1}{n}\sum_{i=1}^{v}\sum_{t=1}^{r_i}\epsilon_{it}.$$

So

$$E\left[\bar{Y}_{..}\right] = \mu \text{ and } \operatorname{Var}\left(\bar{Y}_{..}\right) = \frac{\sum r_i^2}{n^2}\sigma_T^2 + \frac{n}{n^2}\sigma^2.$$

Also, from (17.3.3),

$$E\left[\bar{Y}_{i.}\right] = \mu \text{ and } \operatorname{Var}\left(\bar{Y}_{i.}\right) = \sigma_T^2 + \frac{\sigma^2}{r_i}.$$

Therefore,

$$\begin{aligned} E[SST] &= \sum_{i=1}^{v} r_i\left(\sigma_T^2 + \frac{\sigma^2}{r_i} + \mu^2\right) - n\left(\frac{\sum r_i^2}{n^2}\sigma_T^2 + \frac{\sigma^2}{n} + \mu^2\right) \\ &= \left(n - \frac{\sum r_i^2}{n}\right)\sigma_T^2 + (v-1)\sigma^2 \end{aligned}$$

Therefore

$$E\left[\frac{MST - MSE}{c}\right] = \sigma_T^2.$$

where $c = \left(n - \Sigma r_i^2/n\right)/(v-1)$.

Note this unbiased estimator for σ_T^2 is not always positive.

Testing Equality of Treatment Effects

Consider testing

$$H_0 : \sigma_T^2 = 0 \text{ agains } H_1 : \sigma_T^2 > 0.$$

It can be shown that

$$SST/\left(c\sigma_T^2 + \sigma^2\right) \sim \chi_{v-1}^2$$

and

$$\mathrm{SSE}/\sigma^2 \sim \chi_{n-v}^2$$

and that SST and SSE are independent. Consequently, we have

$$\frac{MST/\left(c\sigma_T^2 + \sigma^2\right)}{MSE/\sigma^2} \sim \frac{\chi_{v-1}^2/(v-1)}{\chi_{n-v}^2/(n-v)} \sim F_{v-1,n-v}$$

Therefore under H_0, we have

$$\boxed{\frac{MST}{MSE} \sim F_{v-1,n-v}.}$$

Hence,

$$\boxed{\text{reject } H_0^T \text{ if } \frac{msT}{msE} > F_{v-1,n-v,\alpha}}$$

where α is the level of significance.

ANOVA Table

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Ratio	Expected mean square
Treatments	$v-1$	ssT	$\frac{ssT}{v-1}$	$\frac{msT}{msE}$	$c\sigma_T^2 + \sigma^2$
Error	$n-v$	ssE	$\frac{ssE}{n-v}$	σ^2	
Total	$n-1$	sstot			

Computational formulae

$$ssss = \sum_i r_i \bar{y}_{i.}^2 - n\bar{y}_{..}^2 \qquad \text{ssE} = \sum_i \sum_t y_{it}^2 - \sum_i r_i \bar{y}_{i.}^2$$
$$sstot = \sum_i \sum_t y_{it}^2 - n\bar{y}_{..}^2 \quad c = \frac{n^2 - \sum_i^2}{n(v-1)}$$