

## RANDOM EFFECTS MODELS (Chapter 17)

So far we have studied experiments and models with only *fixed effect* factors: factors whose levels have been specifically fixed by the experimenter, and where the interest is in comparing the response for just these fixed levels.

A *random effect* factor is one that has many possible levels, and where the interest is in the variability of the response over the entire population of levels, but we only include a random sample of levels in the experiment.

*Examples:* Classify as fixed or random effect.

1. The purpose of the experiment is to compare the effects of three specific dosages of a drug on response.
2. A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To check whether this is the case, five looms are chosen at random and their output is noted at different times.
3. A manufacturer suspects that the batches of raw material furnished by his supplier differ significantly in zinc content. Five batches are randomly selected from the warehouse and the zinc content of each is measured.
4. Four different methods for mixing Portland cement are economical for a company to use. The company wishes to determine if there are any differences in tensile strength of the cement produced by the different mixing methods.

*Note:* The theory behind the techniques we discuss assumes that the population of levels of the random effect factor is infinite. However, the techniques fit well as long as the population is at least 100 times the size of the sample being observed. Situations where the population/sample size ratio is smaller than 100 require “finite population” methods which we will not cover in this class.

### *The Random-Effects One-Way Model*

For a completely randomized design, with  $v$  random selected levels of a treatment factor  $T$ , and  $t$  observations for each of these  $v$  levels, we can use the model

$$Y_{it} = \mu + T_i + \varepsilon_{it},$$

where:

Each  $\varepsilon_{it} \sim N(0, \sigma^2)$

The  $\varepsilon_{it}$ 's are independent random variables

The  $T_i$ 's are independent random variables with distribution  $N(0, \sigma_T^2)$

The  $T_i$ 's and  $\varepsilon_{it}$ 's are independent of each other.

*Caution:* The terminology can be confusing. Here is how to think of the model:

1. Each possible level of the factor T might have a different effect. “Effect of level i” is thus a random variable, hence has a certain distribution – the “distribution of effects”. One of our model assumptions is that this distribution is normal. By adjusting  $\mu$  if necessary, we may assume this distribution of effects has mean 0. We call the variance of this distribution  $\sigma_T^2$ .
2. The effect of level i of T is called (confusingly)  $T_i$ .
3. Since the levels i are randomly chosen (that is, we have a simple random sample of levels), we can say that the  $T_i$ 's are independent random variables with the same distribution  $N(0, \sigma_T^2)$ .
4. The conditions on the  $\varepsilon_{it}$ 's are the same as in previous models. They say that the observations are independently taken within each level and among levels, and all have the same normal distribution.
5. However, we also assume that the observations are made independently of the choice of levels, hence the last condition.

(Think about what these mean in, e.g., Example 2 or Example 3 above.)

*Consequences:*

$$E[Y_{it}] =$$

$$\text{Var}(Y_{it}) =$$

$$Y_{it} \sim$$

$$\text{Cov}(Y_{it}, Y_{is}) =$$

Thus: Observations within the same treatment level are correlated. Does this make sense?

*Terminology:*  $\sigma_T^2$  and  $\sigma^2$  are called *variance components*. (Why?)

*Hypothesis test:* If we wish to test whether or not the level of the factor T makes a difference in the response, what should the null and alternate hypotheses be?

$$H_0:$$

$$H_a:$$

*Least squares estimates:* Given data  $y_{it}$ ,  $i = 1, 2, \dots, v$ ,  $t = 1, 2, \dots, r_i$ , we can still use the method of least squares to obtain “fitted values”  $\hat{y}_{it} = \bar{y}_{i\cdot}$ . However, our interest here will

not be the fits, but the sums of squares obtained from them. In particular, we can still form

$$ssE = \sum_{i=1}^v \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot})^2$$

and the corresponding random variable

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

Also as with the fixed effects model, we obtain the least squares estimate  $\bar{y}_{..}$  for the submodel

$$Y_{it} = \mu + \varepsilon_{it},$$

and can form its error sum of squares  $ssE_0 = \sum_{i=1}^v \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{..})^2$  and the sum of squares for

treatment  $ssT = ssE_0 - ssE = \sum_{i=1}^v \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{..})^2 - \sum_{i=1}^v \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot})^2$ , and the corresponding

random variable

$$SST = \sum_{i=1}^v \sum_{t=1}^{r_i} (Y_{it} - \bar{Y}_{..})^2 - \sum_{i=1}^v \sum_{t=1}^{r_i} (Y_{it} - \bar{Y}_{i\cdot})^2$$

As with the fixed effects model, we can get an alternate expression for SSE (cf Equation 3.4.5):

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

We want  $E(SSE)$ , so we first need  $E(Y_{it}^2)$  and  $E(\bar{Y}_{i\cdot}^2)$

Recall:  $\text{Var}(Y) = E(Y^2) - [E(Y)]^2$ , so  
 $E(Y^2) = \text{Var}(Y) + [E(Y)]^2$

Applying this to  $Y_{it}^2$ :

$$E(Y_{it}^2) = \text{Var}(Y_{it}) + [E(Y_{it})]^2 = \underline{\hspace{10em}}$$

To find  $E(\bar{Y}_{i\cdot}^2)$ , first note that

$$\begin{aligned} \bar{Y}_{i\cdot} &= \frac{1}{r_i} \left[ \sum_{t=1}^{r_i} (\mu + T_i + \varepsilon_{it}) \right] \\ &= \mu + T_i + \frac{1}{r_i} \left[ \sum_{t=1}^{r_i} \varepsilon_{it} \right] \end{aligned}$$

Thus:  $E(\bar{Y}_{i\cdot}) =$

$$\text{Var}(\bar{Y}_{i\cdot}) =$$

$$E(\bar{Y}_{i\cdot}^2) =$$

Finally, we get

$$\begin{aligned} E(\text{SSE}) &= \sum_{i=1}^v \sum_{t=1}^{r_i} E(Y_{it}^2) - \sum_{i=1}^v r_i E(\bar{Y}_{i\cdot}^2) \\ &= \end{aligned}$$

We define  $\text{MSE} = \text{SSE}/(n-v)$  as before, and so  $E(\text{MSE}) =$  \_\_\_\_\_

Thus:

To do inference, we also need an unbiased estimator of  $\sigma_T^2$ . To this end, look at  $E(\text{SST})$ . As with the fixed effect model,

$$\text{SST} = \sum_i r_i \bar{Y}_{i\cdot}^2 - n \bar{Y}_{\cdot\cdot}^2$$

Now  $\bar{Y}_{\cdot\cdot} =$

=

So  $E(\bar{Y}_{\cdot\cdot}) =$

$$\text{Var}(\bar{Y}_{\cdot\cdot}) =$$

=

=

=

Thus  $E(\bar{Y}_{..}^2) = \text{Var}(\bar{Y}_{..}) + E(\bar{Y}_{..})^2$

=

Then

$$E(\text{SST}) = \sum_i r_i E(\bar{Y}_{i.}^2) - nE(\bar{Y}_{..}^2)$$

=

=

So

$$E(\text{MST}) = \quad =$$

where  $c =$

Thus

$$E([\text{MST} - \text{MSE}]/c) =$$

i.e.,  $[\text{MST} - \text{MSE}]/c$  is \_\_\_\_\_

Note: 1. If we have a balanced design (all  $r_i = r$ ), then  $n = vr$ , and

$c =$

2. In general, since the  $r_i$ 's sum to  $r$ ,