

### Reparametrization of the one-way model.

Suppose  $n_i$  are all equal, and equal to  $J$ . Also, let the number of groups be  $k = I$ . Then  $\sum_{i=1}^I n_i = IJ$ , and  $\bar{y}_i = \sum_{j=1}^J y_{ij}/J$ , for  $i = 1, \dots, I$ .

$\bar{y}_{..} = \sum_{i=1}^I \sum_{j=1}^J y_{ij}/(IJ)$ . Further,

$SS_W = \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_i)^2$  has d.f.  $(IJ - I)$ ;

$SS_B = \sum_{i=1}^I n_i (y_i - \bar{y}_{..})^2 = J \sum_{i=1}^I (y_i - \bar{y}_{..})^2$  has d.f.  $I - 1$ .

We can rewrite the model,  $y_{ij} = \mu_i + \epsilon_{ij}$ ,  $\epsilon_{ij} \sim N(0, \sigma^2)$  i.i.d. as follows.

$\mu_i = \bar{\mu}_{..} + (\mu_i - \bar{\mu}_{..}) = \mu + \alpha_i$ , where  $\bar{\mu}_{..} = \sum_{i=1}^I \mu_i/I$  and  $\alpha_i = \mu_i - \bar{\mu}_{..}$ . Then,  $\sum_{i=1}^I \alpha_i = \alpha_{..} = \sum_{i=1}^I (\mu_i - \bar{\mu}_{..}) = 0$ . Further,  $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$  is the same as  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{I-1} = 0$  ( $\alpha_{..} = 0$  implies that  $\alpha_I = -\sum_{i=1}^{I-1} \alpha_i = 0$  also.)

Similarly write

$\bar{\epsilon}_i = \bar{\epsilon}_{..} + \bar{\epsilon}_i - \bar{\epsilon}_{..}$ , so that

$\epsilon_{ij} = \bar{\epsilon}_{..} + (\bar{\epsilon}_i - \bar{\epsilon}_{..}) + (\epsilon_{ij} - \bar{\epsilon}_i)$ . Therefore

$$\sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij}^2 = \sum_{i=1}^I \sum_{j=1}^J \bar{\epsilon}_{..}^2 + \sum_{i=1}^I \sum_{j=1}^J (\bar{\epsilon}_i - \bar{\epsilon}_{..})^2 + \sum_{i=1}^I \sum_{j=1}^J (\epsilon_{ij} - \bar{\epsilon}_i)^2,$$

since  $\bar{\epsilon}_{..} \sum_{i=1}^I (\bar{\epsilon}_i - \bar{\epsilon}_{..}) = 0$ ,  $\bar{\epsilon}_{..} \sum_{i=1}^I \sum_{j=1}^J (\epsilon_{ij} - \bar{\epsilon}_i) = 0$  and

$\sum_{i=1}^I \sum_{j=1}^J (\bar{\epsilon}_i - \bar{\epsilon}_{..})(\epsilon_{ij} - \bar{\epsilon}_i) = \sum_{i=1}^I (\bar{\epsilon}_i - \bar{\epsilon}_{..}) \sum_{j=1}^J (\epsilon_{ij} - \bar{\epsilon}_i) = 0$ .

Now, since  $\epsilon_{ij} = y_{ij} - \mu - \alpha_i$ , we get  $\bar{\epsilon}_i = \bar{y}_i - \mu - \alpha_i$ ,  $\bar{\epsilon}_{..} = \bar{y}_{..} - \mu$ , and further, from above,

$$\begin{aligned} & \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu - \alpha_i)^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{..} - \mu)^2 + \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_i - \bar{y}_{..} - \alpha_i)^2 + \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_i)^2. \end{aligned}$$

Least squares estimates subject to  $\sum_{i=1}^I \alpha_i = 0$  may be obtained simply by examination of the above, and they are:

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\alpha}_i = \bar{y}_i - \bar{y}_{..},$$

and hence  $RSS = \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_i)^2$ .

Under  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{I-1} = 0$ , we have

$$\begin{aligned} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu - \alpha_i)^2 &\equiv \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu)^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{i..} - \mu)^2 + \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{i.} - \bar{y}_{i..})^2 + \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{i.})^2, \end{aligned}$$

so that, then,  $\sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu)^2$  is minimized when  $\hat{\mu} = \bar{y}_{i..}$  (with  $\alpha_i = 0$ ). We then get

$$\begin{aligned} \text{RSS}_{H_0} &= \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{i.} - \bar{y}_{i..})^2 + \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{i.})^2 \\ &= J \sum_{i=1}^I (\bar{y}_{i.} - \bar{y}_{i..})^2 + \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{i.})^2. \end{aligned}$$

Therefore,

$$\text{RSS}_{H_0} - \text{RSS} = J \sum_{i=1}^I (\bar{y}_{i.} - \bar{y}_{i..})^2.$$

Note that all these can be done by just inspection, even though we have derived these previously using other methods. The simplicity of this approach, however, is very useful for higher-way classification models.

### One-way ANOVA with equal number of observations per group.

source	d.f	SS	MS	F
Treatments	$I - 1$	$J \sum (\bar{y}_{i.} - \bar{y}_{i..})^2$	$\text{SS}_B / (I - 1)$	$\frac{J \sum (\bar{y}_{i.} - \bar{y}_{i..})^2 / (I-1)}{\sum \sum (y_{ij} - \bar{y}_{i.})^2 / (IJ-I)}$
Error	$IJ - I$	$\sum \sum (y_{ij} - \bar{y}_{i.})^2$	$\text{SS}_W / (IJ - I)$	
Total (C)	$IJ - 1$	$\sum \sum (y_{ij} - \bar{y}_{i..})^2$		

This approach of reparametrization and decomposition generalizes to higher-way classification where there are substantial simplifications.

## 2-factor Analysis or 2-way ANOVA

**Example.** An engineer is designing a battery for use in a device that will be subjected to some extreme temperature variations. The only design parameter that he can select at this time is the plate material for the battery, and he has three possible choices. When the device is manufactured and shipped

to the field, the engineer has no control over the temperature extremes that the device will encounter, and he knows from past experience that temperature may impact the effective battery life. However, temperature can be controlled in the product development laboratory for the purposes of testing.

The engineer decides to test all three plate materials at three different temperature levels, 15°F, 70°F and 125°F (-10, 21 and 51 degree C), as these temperature levels are consistent with the product end-use environment. Four batteries are tested at each combination of plate material and temperature, and the 36 tests are run in random order.

Question 1. What effects do material type and temperature have on the life of the battery?

Question 2. Is there a choice of material that would give uniformly long life regardless of temperature? (Robust product design?)