

Design and Analysis of Single-Factor Experiments

Reference: Chapter 10 of Devore's 7th Edition

Maghsoodloo

The objective of one-factor experiments is to determine if a factor (or an input) has a significant impact on a response variable (or an output) Y . This is accomplished through an ANOVA (analysis of variance) Table. To illustrate the procedure, consider the example 10.1 on pages 370-371 of Devore's 7th edition. Here the objective is to determine if the Compression Strength (CMPS measured in lbs), the output Y , is significantly affected by the input "different types of boxes". Note that the factor "Box Type" has $\alpha = 4$ levels (1, 2, 3, 4), which are qualitative. The ultimate objective is to determine which one of the 4 types of boxes, (if any), produces the most amount of CMPS.

Further, each level of the input (or each box type) represents a normal population with population mean $E(Y_{ij}) = \mu_i$ ($i = 1, 2, 3, 4 = \alpha$) and common variance σ^2 ; note that Devore uses the uncommon notation I for total number of levels. Our null hypothesis is $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ VS the alternative H_1 : at least two μ_i 's are significantly different. Further, only under the null hypothesis do all the treatments have identical population means μ . Let y_{ij} be the actual j^{th} observation at the i^{th} level of the factor (or the i^{th} treatment), $j = 1, 2, \dots, n_i$. When $n_1 = n_2 = \dots = n_\alpha = n$, the design is called balanced. For the Table 10.1, $n_1 = n_2 = n_3 = n_4 = n = 6$ so that the design is balanced; further, $y_{34} = 671.7$, $y_{26} = 774.8$ pounds, etc.

To obtain the ANOVA table, we 1st develop the linear additive model (LAM) as follows:

$$\text{LAM: } y_{ij} \equiv y_{ij} + \mu - \mu + \mu_i - \mu_i \equiv \mu + (\mu_i - \mu) + (y_{ij} - \mu_i) \equiv \mu + \tau_i + \epsilon_{ij} \quad (44)$$

In the LAM (44), $\tau_i = \mu_i - \mu$ is called the effect of the i^{th} treatment ($i = 1, 2, \dots, \alpha$), and $\epsilon_{ij} = y_{ij} - \mu_i$, $j = 1, 2, \dots, n_i$, is called the experimental error of the j^{th} observation at the i^{th} factor level, and ϵ_{ij} 's are always assumed $\text{NID}(0, \sigma^2)$, i.e., in ANOVA one assumption is that different factor levels have the same population variance. Therefore, under model (44), our null hypothesis becomes $H_0: \mu_i = \mu$, or $H_0: \tau_i = 0$ for all i VS H_1 : at least one $\tau_i \neq 0$, i.e., under H_0 all treatment effects are hypothesized to be zero, or put differently the factor has no

impact (or effect) on the response variable Y under H_0 . Clearly, $V(\epsilon_{ij}) = \sigma_\epsilon^2 = \sigma^2$. Note that Devore uses X_{ij} for the dependent variable, but this notation is not as prevalent as using y_{ij} for the output (or the dependent variable). Recall that in calculus the dependent variable is generally denoted by y , not x .

Next, we replace the parameters in identity (44) by their corresponding point unbiased estimators in order to obtain the identity

$$y_{ij} \equiv \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}) \equiv \hat{\mu} + \hat{\tau}_i + e_{ij} = \hat{y}_{ij} + e_{ij} \quad (45)$$

For our example, $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) = \bar{y}_{i.}$, $\hat{\mu} = \bar{y}_{..} = 16380.1 / (4 \times 6) = 682.50417$, $\hat{\tau}_1 = 713.00 - 682.50417 = 30.49583, \dots, \hat{\tau}_4 = 562.01667 - 682.50417 = -120.4875$.

Note that $e_{ij} = y_{ij} - \hat{y}_{ij}$ is called the residual for the j^{th} observation at the i^{th} level of the factor, where $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\epsilon}_{ij} = \bar{y}_{i.}$, because the best predictor of ϵ_{ij} is zero. Further,

$\sum_{j=1}^{n_i} e_{ij} \equiv 0$ for all i , while this last constraint does not apply to ϵ_{ij} 's.

Exercise 81. Verify that equation (45) is indeed an identity, compute

$\hat{\tau}_2, \hat{\tau}_3$, and verify that $\sum_{i=1}^4 \hat{\tau}_i \equiv 0$. (b) Prove that for an unbalanced Design $\sum_{i=1}^a n_i \hat{\tau}_i \equiv 0$. In

your derivation, let $N = \sum_{i=1}^a n_i$ be the total number of all observations in the experiment. You

may also verify that the sum over j of the i^{th} residual e_{ij} is zero for all i , and that the double sum over i and j of all N residuals is always identically zero. (c) Assuming that τ_i 's are constant, use model (44) to show that the $V(y_{ij})$ is also equal to σ^2 .

We now rewrite identity (45) as $y_{ij} - \bar{y}_{..} \equiv (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}) \equiv \hat{\tau}_i + e_{ij}$. This last identity breaks down the total deviation on the LHS into 2 separate orthogonal (i.e., additive) components: one among (or Between) factor levels ($\hat{\tau}_i$) and the other Within the

same factor levels (e_{ij}). To obtain the test statistic for testing H_0 , we first square both sides of this last identity and then sum over both i and j , i.e.,

$$\sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 \equiv \sum \sum \left[(\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}) \right]^2 \quad (46)$$

Exercise 82. Show that in the above identity (46), the double sum of the cross-product terms vanishes and hence the identity reduces to

$$\begin{aligned} \sum \sum (y_{ij} - \bar{y}_{..})^2 &\equiv \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 \\ \sum \sum (y_{ij} - \bar{y}_{..})^2 &\equiv \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2. \end{aligned} \quad (47a)$$

$$SS(\text{Total}) \equiv SS(\text{Between Treatments}) + SS(\text{Within Treatments, or Subgroups})$$

The identity in (47a) is fundamental to the ANOVA of a one-factor experiment because it breaks down the total sum of squares on the LHS, denoted by $SS(\text{Total}) = SS_T$ (Corrected Total Sum of Squares), into 2 orthogonal components: the first SS on the RHS describes the variability between treatments (or levels), and 2nd SS on the RHS describes the variability within treatments, i.e.,

$$SS(\text{Total}) = SS_{\text{Total}} = SS_T = SS(\text{Treatments}) + SS(\text{Experimental Error}) \quad (47b)$$

Some authors use $SS_E = SS(\text{Error})$ for $SS(\text{Experimental Error})$, and my preference is to use $SS(\text{Model})$ for $SS(\text{Treatments})$. Further, $(\text{Total SS})/\sigma^2$ has a χ^2 distribution with $N - 1 = (\sum n_i) - 1$ degrees of freedom (df), $SS(\text{Treatments})/\sigma^2 \sim \chi_{a-1}^2$ and hence the distribution of $(SS_{\text{Error}})/\sigma^2$ follows a χ^2 with $(N - 1) - (a - 1) = (N - a)$ df .

We next prove that under linear additive model (44), $E(SS_E) = (N - a)\sigma^2$.

$$E(SS_E) = E \sum \sum (y_{ij} - \bar{y}_{i.})^2 = E \sum_{i=1}^a \sum_{j=1}^{n_i} \left[(\mu + \tau_i + \varepsilon_{ij}) - (\mu + \tau_i + \bar{\varepsilon}_{i.}) \right]^2$$

$$= \sum \mathbb{E} \sum_{j=1}^{n_i} (\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 = \sum_{i=1}^a (n_i - 1) \sigma^2 = (N - a) \sigma^2. \quad \text{QED!}$$

This last result clearly shows that an unbiased estimator of σ^2 is given by $SS_E / (N - a)$.

Before proceeding to the following Exercise, we must state that there are 2 types of the LAM (44): Fixed- and Random- Effects Models. Model (44) is fixed iff the levels of the factor are not selected at random (just like the example under consideration). When treatments are fixed, i.e., are not selected at random, then statistical conclusions pertain only to the levels under study. Since our example is one of fixed effects, then all conclusions must pertain only to box types 1, 2, 3, and 4. If the levels of a factor are selected at random from a population of treatments (like selecting 4 box types at random from a population of 40 available types), then model (44) represents the random effects model, in which case τ_i 's are no longer constants (or fixed) but are rvs assumed to be $NID(0, \sigma_\tau^2)$ and independent of ε_{ij} 's.

Exercise 83. For the fixed effects model, show that under model (44), the

$$\mathbb{E}(SS_{\text{Treatments}}) = \sum_{i=1}^a n_i \tau_i^2 + (a - 1) \sigma^2. \quad \text{Therefore, under the null hypothesis } H_0 : \tau_i = 0 \text{ for all } i, \text{ it}$$

follows that another unbiased estimator of σ^2 is given by $SS(\text{Treatments}) = SS_{\text{Tr}} / (a - 1)$. (b)

Show that for the fixed -effects model, $\sum_{i=1}^a \tau_i \equiv 0$, even if the design is unbalanced.

Thus far, we have established that Mean square (MS) of errors, $MS(\text{Error}) = SS_E / (N - a)$, is an unbiased estimator of σ^2 even if H_0 is false, and $MS(\text{Treatments}) = SS(\text{Treatments}) / (a - 1)$ is an unbiased estimator of σ^2 only if H_0 is true. Further, we have stated that $SS(\text{Treatments}) / \sigma^2$ and $SS(\text{Error}) / \sigma^2$ have independent χ^2 distributions with $(a - 1)$ and $(N$

$- a)$ *df*, respectively. Recall from Chapter 9 that the random variable (rv), $\frac{\chi_{v_1}^2 / v_1}{\chi_{v_2}^2 / v_2}$, has

Fisher's F distribution with v_1 *df* for the numerator and v_2 *df* for the denominator.

Applying this principal to ANOVA, it follows that the statistic

$$F_0 = \frac{SS(\text{Treatments}) / [\sigma^2(a-1)]}{SS(\text{Error}) / [\sigma^2(N-a)]} = \frac{MS(\text{Treatments})}{MS(\text{Error})} \quad (48)$$

has the Fisher's F distribution with $\nu_1 = a - 1$ for the numerator, and $\nu_2 = N - a$ denominator degrees of freedom. Note that the numerator of equation (48) is expected to exceed that of

its denominator due to the fact that $E(MS_{\text{Treatments}}) = \sigma^2 + (\sum_{i=1}^a n_i \tau_i^2) / (a - 1)$. When H_0 is

true, i.e., $\tau_i = 0$ for all i , then the value of F_0 in (48) is expected to be close to 1 because $E(MS_{\text{Error}}) = \sigma^2$ whether H_0 is true or false. However, when H_0 is false, the value of F_0 is expected to far exceed 1. This constitutes a right-tailed test on $H_0 : \tau_i = 0$ for all i , i.e., H_0 is rejected at the LOS α only if the statistic F_0 in (48) exceeds $F_{\alpha}(a-1, N-a) = F_{\alpha, a-1, N-a}$.

Exercise 84. Show that the computing formulas for $SS(\text{Total})$ and

$SS(\text{Treatments})$ are given by

$$SS_T = SS(\text{Total}) = \sum \sum (y_{ij} - \bar{y}_{..})^2 = (\sum \sum y_{ij}^2) - y_{..}^2 / N = USS - CF, \quad SS(\text{treatments}) =$$

$$\sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = (\sum_{i=1}^a y_{i.}^2 / n_i) - CF, \quad \text{and hence } SS(\text{Error}) = SS_T - SS(\text{Treatments}).$$

We are now in a position to develop the ANOVA table and conduct the F-test for the Example 10.1 on pages 370-371 of your text.

Solution: $CF = 16380.1^2/24 = 11179486.50417$ (with 1 *df*), $USS = 655.5^2 + 788.3^2 + \dots + 520.0^2 = 11340700.23$ (with 24 *df*) $\rightarrow SS_T = USS - CF = 161213.72958333$ (with 23 *df*).

$SS(\text{Model}) = (4278^2 + 4541.6^2 + 4188.4^2 + 3372.1^2)/6 - CF = 127374.754583$ (with 3 *df*).

Hence $SS(\text{Error}) = SS(\text{Total}) - SS(\text{Model}) = 33838.975$ (with 20 *df*). The ANOVA Table is given in Table 2 below. Note that the value of the F statistic is highly significant ($P\text{-value} = 0.000000552545$), and thus the null hypothesis of no treatment effects (i.e., $\tau_i = 0$ for all i) must be strongly rejected. It seems that level 2 of the factor (followed by Box Type 3)

produces the highest CMPS.

Table 2 (The ANOVA Table for the Example 10.1 of Devore)

SOURCE OF VARIATION	<i>df</i>	SS	MS = SS/ <i>df</i>	$F_0 =$ MS_{BT}/MS_E	PR LEVEL = <i>P-value</i> = $\hat{\alpha}$
Total	23	161213.72953			
Model (OR Box Types)	3	127374.75453	42458.25153	25.0943	0.0 ⁶ 552545
ERROR	20	33838.975	1691.94875	$F_{0.05,3,20} =$	3.0984

Exercise 85. Show that for a balanced design $E(e_{ij}) = 0$, and $V(e_{ij}) = (n-1)\sigma^2/n$, and thus the $se(e_{ij}) = [(n-1)MS_E/n]^{1/2}$, where e_{ij} is the ij^{th} residual. (b) Use the results of part (a) to compute the values of the 3 Studentized residuals r_{21} , r_{15} and r_{34} . (c) Redo the ANOVA table for the Example 10.1 coding the data by subtracting 650 from all y_{ij}' s.

Exercise 86. Work Exercises 10.2, 10.6 and 10.9 on pages 378-379 of Devore.

Obtain a 95% CI for μ_1 of Exercise 10.2, and a 95% CI for $\mu_1 - \mu_4$ of Exercise 10.9 on p. 379 of Devore.

Exercise 87. Carefully study section 10.2 (pp. 379-383) of your text and work Exercise 10.11, 10.17 and 10.18 on page 384-5 of your text.

Tukey's Post-ANOVA For Fixed-Effects Only if H_0 is Rejected

Step 1: Arrange the $a = 4$ level means in ascending order

$$\bar{y}_4 = 562.02$$

$$\bar{y}_3 = 698.07$$

$$\bar{y}_1 = 713.00$$

$$\bar{y}_2 = 756.93$$

Step 2: Obtain the critical value of Tukey's (Studentized Range) from Table A.10 (p. 682), $Q_{0.05,4,20}$, and use it to compute the 95% half confidence interval band as shown below

$$w_{ij} = Q_{0.05,4,20} \times \sqrt{\frac{MS_{\text{Error}}(1/n_i + 1/n_j)}{2}}$$

$$= 3.96 \times \sqrt{\frac{1691.94875(1/6 + 1/6)}{2}} = 3.96 \times \sqrt{\frac{1691.94875}{6}} = 3.96 \times se(\bar{y}) = 66.4987$$

Step 3: Underline all the means in step 1 which do not differ by as much as w_{ij} .

$$\bar{y}_4 = 562.02 \quad \bar{y}_3 = 698.07 \quad \bar{y}_1 = 713.00 \quad \bar{y}_2 = 756.93$$

Step 4: Use the underlined means to draw the conclusion that they come statistically from the same population from the standpoint of population means.

Thus, Box types 3, 1, and 2 means are statistically the same; all the other pairs of means are significantly different at the 5% level.

THE RANDOM- EFFECTS MODEL (SECTION 10-3, pp. 391-393)

Recall that this model is valid only when the levels of a factor are randomly selected from a population of levels (e.g., selecting 4 box types at random from a population of 40 available types). Box types would then form a random qualitative factor but the conclusions drawn from the ANOVA Table would pertain to the entire population of 40 types. Note that there will not be any interest in determining which 2 of the actually selected treatments (or box types) are significantly different (i.e., Tukey's procedure and orthogonal contrasts will not be applicable) but rather the objective is to determine if there is significant variation in the entire population of treatments (or box types). Therefore, our null hypothesis is $H_0 : \sigma_{\tau}^2$

$= 0$ VS $H_1 : \sigma_\tau^2 > 0$, which again is clearly a right-tailed test. The ANOVA Table is obtained exactly as in the case of fixed-effects, but the post-ANOVA procedure is quite different. In the case of a random-effects model, if the F-test rejects the null hypothesis, then it is essential to estimate the components of variance in the LAM : $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, where $\mu = E(y_{ij})$ for all $i = 1, 2, \dots, a$, and $j = 1, 2, \dots, n$.

Note that we are considering only the simpler balanced case (i.e., $n_i = n$ for all i) for the random-effects model. Further, τ_i 's are rvs assumed $NID(0, \sigma_\tau^2)$ so that the constraint

$\sum_{i=1}^a \tau_i = 0$ no longer applies to the random-effects model. The experimental errors, ϵ_{ij} 's, are as before $NID(0, \sigma_\epsilon^2)$ and independent of τ_i 's. This leads to $V(y_{ij}) = \sigma_\tau^2 + \sigma_\epsilon^2 = \sigma_\tau^2 + \sigma^2$, i.e., there are two variance components.

As an example of obtaining the ANOVA Table for a random-effects model see Example 10.10, on pp. 392-3 of your text. In order to estimate the 2 components of variance, σ_τ^2 and σ_ϵ^2 , we must derive $E(MS_E)$ and $E(MS_{Treatments})$. As in the case of fixed-effects, it can easily be verified that $E(MS_E) = \sigma^2$. We now prove that, for the balanced case, under the LAM (44), $E(MS_{Treatments}) = n\sigma_\tau^2 + \sigma^2$. **Proof:**

$$\begin{aligned} E(MS_{Treatments}) &= E \sum_{i=1}^a \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 / (a-1) = n E \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 / (a-1) \\ &= \frac{n}{a-1} E \sum [(\mu + \tau_i + \bar{\epsilon}_{i.}) - (\mu + \bar{\tau} + \bar{\epsilon}_{..})]^2 = \frac{n}{a-1} E \sum [(\tau_i - \bar{\tau}) + (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})]^2 \\ &= \frac{n}{a-1} \left[E \sum (\tau_i - \bar{\tau})^2 + E \sum (\bar{\epsilon}_{i.} - \bar{\epsilon}_{..})^2 \right] = \frac{n}{a-1} \left[(a-1)\sigma_\tau^2 + (a-1)\sigma_\epsilon^2 \right] = \sigma^2 + n\sigma_\tau^2. \end{aligned}$$

If the design were unbalanced, then $E(MS_{Treatments}) = \sigma^2 + \sigma_\tau^2 \left[N - \sum_{i=1}^a (n_i^2 / N) \right] / (a-1)$.

As before, an unbiased estimator of σ^2 is clearly MS_E , and the above proof shows that $E(MS_{\text{Treatments}}) - \sigma^2 = n \sigma_\tau^2$. Inserting $E(MS_E)$ for σ^2 in this last equation results in $E(MS_{\text{Treatments}}) - E(MS_E) = n \sigma_\tau^2$, or $E[(MS_{\text{Treatments}} - MS_E)/n] = \sigma_\tau^2$. This last equality clearly shows that an unbiased estimator of σ_τ^2 is indeed $(MS_{\text{Treatments}} - MS_E)/n$. We are now in a position to estimate the components of variance for the Example 10.10 on p. 392. From the ANOVA Table on page 393, the unbiased estimate of σ_ϵ^2 is $MS(\text{Error}) = 16.17$ and the unbiased estimate of σ_τ^2 is $\hat{\sigma}_\tau^2 = (1862.1 - 16.17)/3 = 615.31$. This implies that 97.44% of variation in travel times is attributed to the rails.

Exercise 88. Work Exercise 41 and on p. 395 of your text.

THE RANDOMIZED COMPLETE BLOCK DESIGN (RCBD)

Consider an experiment where it is desired to determine if there are significant differences among 3 computer languages (Java, VB, and C++) in coding a complex problem. Obviously we need human subjects for the DOE (or DOX) so that the 3 computer coding methods can be analyzed for efficiency. Suppose we select 10 programmers at random who are competent at all 3 languages to design 3 programs each, and as the response variable we measure the length of time that each programmer takes to complete a set of workable codes. Note that each programmer forms a Block and acts as his/her own control, and the experiment would be an Incomplete Block Design only if some of the programmers could not code in at least one of the 3 languages. The data are provided in Table 3 below. Further, randomization takes place only within each block. Before developing the LAM and the corresponding $E(SS)$ for the model, we obtain the ANOVA Table and draw appropriate conclusions first. The ANOVA calculations are: the $CF = 93.9^2 / 30 = 293.907$ (with 1 *df*), the $USS = 2.5^2 + 3.1^2 + \dots + 4.1^2 = 319.030$ (with 30 *df*) $\rightarrow SS_T = SS(\text{Total}) = 25.123$ (with 29 *df*). The treatment SS is computed using the subtotals pertaining to the 3 languages, i.e., $SS(\text{Treatments}) = (31^2 + 34.8^2 + 28.1^2) / 10 - CF = 2.2580$ (with 2 *df*); $SS(\text{Blocks}) = (7.8^2 + 9.4^2 + \dots + 14.0^2) / 3 - CF = 18.749667$ (with 9 *df*), and thus $SS(\text{Residuals}) = 4.115333$

Table 3. (Experimental Layout)

	programmer 1	2	3	4	5	6	7	8	9	BLK 10	y_i
Java	2.5 Hrs	3.2	3.1	3.7	2.4	2.1	3.3	3.5	1.9	5.3	31.0
VB	3.1	2.5	2.8	4.7	2.9	2.9	4.4	4.4	2.5	4.6	34.8
C++	2.2	3.7	2.7	3.9	2.4	1.7	2.6	3.0	1.8	4.1	28.1
y_j	7.8 Hrs.	9.4	8.6	12.3	7.7	6.7	10.3	10.9	6.2	14.0	$y_{..} = 93.9$

(with 18 df). The ANOVA is given in Table 4.

Table 4. (The ANOVA Table for the Data of Table 3)

Source of Variation	df	SS	MS	F_0	$F_{0.05}(2, 18)$	P -value
Total	29	25.123				
Computer Languages	2	2.258	1.129	4.9381	3.55	0.01951
Programmers (or Blocks)	9	18.74967				
Residuals (or Left-Over's)	18	4.115333	0.22863			

The ANOVA Table 4 clearly indicates that there are significant differences among the 3 programming languages (i.e., the Treatments) at the Pr level $\hat{\alpha} = 0.01951$, i.e., we must reject the null hypothesis $H_0: \tau_1 = \tau_2 = \tau_3 = 0$ at levels of significance as small as 0.01951.

Exercise 89. Analyze the result of the above experiment as a completely randomized design (CRD), i.e., the order of the 30 observations were completely randomized. ANS: $F_0 = 1.3332$.

We now develop the LAM for a RCBD. As before we start with an identity!

$$\begin{aligned} y_{ij} &\equiv y_{ij} + \mu - \mu + \mu_i - \mu_i + \mu_j - \mu_j \equiv \mu + (\mu_i - \mu) + (\mu_j - \mu) + (y_{ij} - \mu_i - \mu_j + \mu) \equiv \\ &\equiv \mu + \tau_i + B_j + \epsilon_{ij} \end{aligned} \quad (49)$$

Model (49) shows that $\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{B}_j + 0 = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$. Thus, the ij^{th} residual is given by $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$. The term $B_j = \mu_j - \mu$ in (49) is called the effect of the j^{th} block, and the definitions of the other 2 terms on the RHS of (49) are self-explanatory! Again, we replace the model parameters in equation Eq. (49) by their corresponding point unbiased estimators and obtain:

$$\begin{aligned} y_{ij} &\equiv \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \\ y_{ij} - \bar{y}_{..} &\equiv (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \end{aligned} \quad (50)$$

We 1st square both sides of (50) and then sum from $i=1$ to " a " (the number of treatments) and also sum from $j=1$ to " b " (the no. of blocks), i.e.,

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 &\equiv b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum \sum (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \\ SS_{\text{Total}} &\equiv b \sum_{i=1}^a \hat{\tau}_i^2 + a \sum_{j=1}^b \hat{B}_j^2 + \sum \sum e_{ij}^2 \end{aligned} \quad (51)$$

Note that the 1st term on the RHS of (51) gives $SS_{\text{Treatments}} = SS_{\text{Tr}}$, the 2nd term gives $SS(\text{BLKS})$, and the last term gives $SS(\text{Residuals})$.

Exercise 90. Show that equation (51) is indeed an identity because the double sum of the 3 cross-product terms on the RHS vanish.

Exercise 91. For the mixed-effects (i.e., fixed treatments but random blocks), show that

$$E(SS_{\text{Treatments}}) = \sigma^2(a-1) + b \sum_{i=1}^a \tau_i^2, \text{ and } E(SS_{\text{Blocks}}) = (b-1)(\sigma^2 + a\sigma_\beta^2), \text{ and } E(SS_E) =$$

$(a-1)(b-1)\sigma^2$, where σ_{β}^2 represents the variance of the random blocks. Further, show that for a RCBD $V(e_{ij}) = [(a-1)(b-1)/ab]\sigma_{\epsilon}^2$.

It should be quite clear that identity (51) is fundamental to the analysis of RCBD experiments because it breaks down the Total SS on the LHS into 3 orthogonal (i.e., additive) components: the 1st SS on the RHS is due to Treatments, the 2nd is due to Blocks, and the 3rd SS is due to Residuals. The *df* of Treatments is $(a-1)$, that of Blocks is $(b-1)$, and hence the *df* of Residuals is $(ab-1) - (a-1) - (b-1) = (a-1)(b-1)$, where $ab = N$.

Exercise 92. Work Exercise 36 on page 350 of Devore using ANOVA procedure and compare your result against that of the paired t-test. (b) Then work Exercise 38 on page 351 of Devore in 2 different ways: (i) using the paired t-test, (ii) using ANOVA for a RCBD, and show that your results are exactly identical.