The Logic Behind the ANOVA for the Exercise 37 on page 417 of Devore (8e) **Maghsoodloo**

In Exercise 37 on page 417 of Devore's 8th edition there are 5 treatments (or 5 independent populations), whose sample means are to be compared simultaneously in order to test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$, where μ represents the grand population average of all 5 populations (or treatments). For example, the data layout shows that y_{24} = 14.9 microns (Mu) while y_{43} = 14.4, etc; further, there are n = 6 random observations from each level (or per population). Hence, N = 5×6 = 30, and there are a total of 29 degrees of freedom in the entire experiment. In the following development, I will first break down the Total SS (sum of squares) into two orthogonal (i.e., additive) components : (1) due to differences within the five samples (or within treatments), (2) due to differences between the 5 sample means. The Total SS will be obtained from the deviations of individual observations from the grand sample means $\bar{y}_{..} = \dfrac{y_{..}}{30} = \dfrac{426.7}{30} = 14.2233\bar{3}$, while within SS will be obtained from the 6 observations in the ith treatment (or brand) from the corresponding means $\bar{y}_{i.}$, i = 1, 2, 3, 4, 5.

$$\text{SS(Total)} = \sum_{i=1}^{5\,\text{brands}} \sum_{j=1}^{n=6} (y_{ij} - 14.2233\bar{3})^2 \equiv \sum_{i=1}^{5} \sum_{j=1}^{n=6} [(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})]^2 = \ldots$$

$$\equiv \sum_{i=1}^{5} \sum_{j=1}^{n=6} [(y_{ij} - \bar{y}_{i.})^2] + \sum_{i=1}^{5} \sum_{j=1}^{n=6} [(\bar{y}_{i.} - \bar{y}_{..})^2] =$$

$$\equiv \sum_{i=1}^{5} \sum_{j=1}^{n=6} [(y_{ij} - \bar{y}_{i.})^2] + \sum_{i=1}^{5} 6[(\bar{y}_{i.} - \bar{y}_{..})^2] \equiv \sum_{i=1}^{5} \sum_{j=1}^{n=6} e_{ij}^2 + 6\sum_{i=1}^{5} \hat{\tau}_i^2 =$$

= SS(Within the 5 levels) + SS(Between Brand Means)

= SS(Residuals) + SS(Treatments = Model).

Or:

53.69366667(with 29 df) = 22.8383333 (with 5×5 df) + 30.85533333 (with 4 df)

Further, $e_{ij} = y_{ij} - \bar{y}_{i.}$ is called the ij^th^ residual while $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$ is called the effect of the i^th^ treatment (or Brand).

There are two main assumptions in Fixed-Effects ANOVA: (1) y_{ij}'s ~ NID(μ_i, σ_i^2),

(2) $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_a^2 = \sigma_\in^2 = \sigma^2$, where "a" represents the number of treatments.

The assumption 2 above implies that we may pool our sample variances

$$S_i^2 = \frac{1}{5}\sum_{j=1}^{n=6}[(y_{ij} - \bar{y}_{i.})^2] = \frac{CSS_i(\text{Within})}{5}$$ **= 1.425667, 1.363000, 0.666667, 0.882667, 0.229667 (I = 1, 2 , 3, 4, 5) from the 5 independent populations to obtain one overall estimate of error variance $\sigma_\in^2 = \sigma^2 = \sigma_y^2$. That is, from the within treatments we may estimate σ_y^2 from**

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^{5} CSS_i(W)}{5 \times 5} =$$

$$\frac{7.128333\bar{3} + 6.8150 + 3.33333\bar{3} + 4.413333\bar{3} + 1.1483333\bar{3}}{25} = \frac{22.8383333\bar{3}}{25} =$$

0.91353333 with 25 df. Another estimate of error variance may be obtained by first estimating the variance between the five population means from $\hat{\sigma}_{\bar{y}}^2$

$$= \frac{1}{4}\sum_{i=1}^{5}[(\bar{y}_{i.} - \bar{y}_{..})^2] = \frac{\sum_{i=1}^{5}\hat{\tau}_i^2}{4} = \frac{5.142555\bar{5}}{4}$$ **= 1.2856388889 with 4 df.**

However, forming the F statistic as the ratio of $\dfrac{\hat{\sigma}_{\bar{y}}^2}{\hat{\sigma}_y^2}$ would be exactly like comparing Apples and Oranges because the numerator estimates the variance between means while the denominator estimates the variance amongst the individual measurements within the same treatment! Therefore, we convert the variance of means in the numerator by

multiplying it by the size of each sample n = 6 in order to convert $\hat{V}(\bar{y})$ to $\hat{V}(y)$. This implies that another estimate of σ_y^2 may be obtained from $n\times\hat{\sigma}_{\bar{y}}^2$ = 6×1.2856388889 = 7.713833333. Hence, the correct value of the F statistic is $F_0 = \dfrac{7.71383333\overline{3}}{0.91353333\overline{3}} = \dfrac{\text{MS(Between Samples)}}{\text{MS(Within Samples)}}$ **= 8.443954, and the corresponding Pr level of testing $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$ is given by *P-value* = $P(F_{4,\,25} \geq 8.443954)$ = 0.00018715. Since this *P-value* is much smaller than 5%, we can strongly reject $H_0 : \tau_i = 0$ and conclude that at least two of the 5 population means differ significantly. This implies that 5 treatments (or Brands) have a statistically significant impact on the dependent variable "Motor Vibration". Then, Tukey's Post-ANOVA must be used to ascertain where the exact differences are.**

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

Step 1: Arrange the a = 5 brand means in ascending order

$\bar{y}_{5.}$ = 13.08 $\bar{y}_{3.}$ =13.67 $\bar{y}_{1.}$ =13.68 $\bar{y}_{4.}$ =14.73 $\bar{y}_{2.}$ = 15.95

Step 2: Obtain the critical value of Tukey's (Studentized Range) from Table A10 (appendix p. A-20 of the 8th edition), $Q_{0.05,5,25}$, and use it to compute the 95% half confidence interval band as shown below

$$w_{ik} = Q_{0.05,5,25} \times \sqrt{\frac{MS_{Error}(1/n_i + 1/n_k)}{2}}, \quad (i \neq k,\ i,\ k = 1, 2, , \ldots a)$$

$$= 4.15\times\sqrt{\frac{0.9135333(1/6+1/6)}{2}} = 4.15\times\sqrt{\frac{0.9135333}{6}} = 4.15\times se(\bar{y}) = 1.6193.$$

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

$\underline{\bar{y}_{5.} = 13.08 \quad \bar{y}_{3.} = 13.67 \quad \bar{y}_{1.} = 13.68}$ $\bar{y}_{4.}$ =14.73 $\bar{y}_{2.}$ = 15.95

<u>Step 4:</u> **Use the underlined means to draw the conclusion that they come statistically come from the same population from the standpoint of population means. Thus, Brands 5, 3 and 1 means are statistically the same; Brand 4 mean is statistically different from 5, and brand 2 mean differs significantly from those of 5, 3, and 1. Note that i and k are 2 distinct levels of treatments.**

Additional Information: **USS = 6122.789917; CF = 6069.0962465; CSS = USS − CF = 53.693666667; USS_1= 1130.53000786 ; USS_2 =1533.23 ; USS_3 =1123.99997368; USS_4 = 1306.83996159; USS_5 =1028.18997437; $y_{i.}$ = 82.10, 95.7, 82.0, 88.4, 78.50 ⟶ $y_{..}$ = 426.700.**

SS(Model) = $\dfrac{82.10^2 + 95.7^2 + 82.0^2 + 88.4^2 + 78.50^2}{6}$ **− CF = 30.8553333**

SS(Error or Residuals) = SS(Total) − SS(Model or Brands) = 22.8383333.

The ANOVA Table for the Exercise 37 on p. 417 of Devore)

SOURCE OF VARIATION	DF	SS	MS = SS/DF	F_0 = MS_{Tr}/MS_E	PR LEVEL = *P-value*
Total	29	53.693666667			
Model (OR Brands)	4	30.8553333	7.713833333	8.443954	R-Sq=57.47%
ERROR	25	22.8383333	0.91353333	$F_{0.05,4,25}$ =	0.00018715

In order to test $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots = \sigma_a^2 = \sigma_\epsilon^2 = \sigma^2$, **use the Bartlett statistic** $\chi_0^2 = q/c$, **where** $q = (N - a)\ln(S_p^2) - \sum_{i=1}^{a}(n_i - 1)\ln(S_i^2)$, $c = 1 + [\sum_{i=1}^{a}(n_i - 1)^{-1} - (N - a)^{-1}]/[3(a-1)]$, **and reject** H_0 **only if** $\chi_0^2 > \chi_{0.05,a-1}^2$.