3.5 Reduced Models and the Extra Sum-of-Squares F-test in Single-Factor ANOVA

RETURN to the Research Problem on pages 9-12:
Do exercise level and sodium intake level (explanatory variables) have an effect on blood pressure (response variable)?
Table 1: $df_{Error} = n - k = 48 - 6 = 42$
Table 2: $df_E = n - k = 48 - 2 = 46$
Table 3: $df_E = n - k = 48 - 3 = 45$

(d) At the 5% significance level, perform the most appropriate test to determine whether there is a difference in mean SBP between the Typical Na groups and the Reduced Na groups after accounting for the effect of exercise (No, Moderate, and vigorous).

> If there is no mean difference in mean SBP between typical Na and reduced Na in the no exercise group, then $\mu_{NT} = \mu_{NR}$.
> If there is no mean difference in mean SBP between typical Na and reduced Na in the moderate exercise group, then $\mu_{MT} = \mu_{MR}$.
> If there is no mean difference in mean SBP between typical Na and reduced Na in the vigoruous exercise group, then $\mu_{VT} = \mu_{VR}$.

$H_0 : \mu_{NT} = \mu_{NR} \text{ and } \mu_{MT} = \mu_{MR} \text{ and } \mu_{VT} = \mu_{VR}$ - 3-mean model: reduced model.
$H_a : \mu_{NT},\ \mu_{NR},\ \mu_{MT},\ \mu_{MR},\ \mu_{VT},\ \mu_{VR}$ - 6-mean model: full model.

Using the ANOVA table for comparison of all six means (full model) (Table 1).
And the ANOVA table for comparison of exercise groups (reduced model) (Table 3), we get:
$df_E\ (full) = n - k = 42$
$df_E\ (reduced) = n - k = 45$

$$F = \frac{[SS_{E(Reduced)} - SS_{E(Full)}]\ /\ [df_{E(Reduced)} - df_{E(Full)}]}{SS_{E(Full)}/df_{E(Full)}} = \frac{(4415.938 - 3343.875)/(45-42)}{3343.875/42} = 4.488$$

$$df = [df_{E(Reduced)} - df_{E(Full)},\ df_{E(Full)}] = [45 - 42, 42] = [3, 42] \approx [3, 40]$$

$\text{P-value: } 0.005 < P < 0.01$. This is very strong evidence against H_0.
$P < \alpha\ (0.05),\ \text{therefore reject } H_0$.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in mean SBP between the Typical Na groups and the Reduced NA groups, after accounting for the effect of exercise.

(e) At the 5% significance level, perform the most appropriate test to determine whether there is a difference in mean SBP between exercise groups after accounting for the effect of the level of sodium (Typical or Reduced Na).

$H_0 : \mu_{NT} = \mu_{MT} = \mu_{VT}$ and $\mu_{NR} = \mu_{MR} = \mu_{VR}$ - reduced model: two-mean model.
$H_a : \mu_{NT},\ \mu_{NR},\ \mu_{MT},\ \mu_{MR},\ \mu_{VT},\ \mu_{VR}$ - 6-mean model: full model.

Using the ANOVA table for comparison of all six means (full model) (Table 1).
And the ANOVA table for comparison of exercise groups (reduced model) (Table 2), we get:
$df_E\ (full) = n - k = 48 - 6 = 42$
$df_E\ (reduced) = n - k = 48 - 2 = 46$

$$F = \frac{[SS_E(Reduced) - SS_E(Full)]\ /\ [df_E(Reduced) - df_E(Full)]}{SS_E(Full)\ /\ df_E(Full)} = \frac{(4946.292 - 3343.875)/(46-42)}{3343.875/42} = \frac{400.60425}{79.616071} = 5.032$$

$df = [4, 42] \approx [4, 40]$
P-value: 0.001 < P < 0.005. There is very strong evidence against H_0.
$P < \alpha\ (0.05)$, therefore reject H_0.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in mean SBP between exercise groups after accounting for the effect of the level of sodium (Typical or Reduced).

(f) Suppose a pooled t-test was performed to test for the difference in mean SBP between the Typical Na groups and the Reduced Na groups. What would be the df, the t-statistic and the exact P- value?

This t-test would be testing the same thing as the F-test shown in the SPss output in Table 2 (part d), which also tests for the difference in mean SBP between the Typical NA groups and the Reduced NA groups. Therefore, the required values can be obtained from that table.
The df of the t-test would be the same as the denominator df of that F-test, that is, df = n - k = 48 - 2 = 16. The t-statistic would be $t = \sqrt{F} = \sqrt{9.80848} = 3.132$, with an exact P value of P = -0.003018.

3.6 The Kruskal-Wallis test (Nonparametric Equivalent of One-Way ANOVA)

Research Problem: Pitfall traps are inserted into the soil at ground level in three grasslands (A, B and C) in order to determine whether there is a difference in the abundance of ants in the three grasslands. Test this hypothesis at the 10% significance level.

Step 1: The purpose is to compare k populations

- 3 independent random samples
  However:
  - The data are neither normal nor lognormal
  - Sample size is < 30, therefore the Central Limit Theorem does not apply
  - Therefore, the Kruskal-Wallis Test must be performed
- Same shape distributions, as indicated in the NPPs above
- Sample size of all groups ≥ 5.

Step 2: H_0: There is no difference in the abundance of ants in the three grasslands.
H_a: There is a difference in the abundance of ants in the three grasslands (at least two are different).

Step 3: Calculate the test statistic H
Rank the data from lowest to highest, assigning average ranks where there are tied observations.
>>>>>>>>>>

	Grassland A		Grassland B		Grassland C	
	No. of ants	Rank	No. of ants	Rank	No. of ants	Rank
	168	28	0	3	144	26
	0	3	13	14	1	7.5
	62	21	28	16	3	10.5
	0	3	32	17.5	135	24.5
	1	7.5	18	15	45	20
	135	24.5	4	12.5	3	10.5
	0	3	41	19	122	23
	0	3	1	7.5	4	12.5
	155	27			110	22
	32	17.5				
	1	7.5				
Sum of ranks (R_j)		**R_1 = 145**		**R_2 =104.5**		**R_3=156.5**
Sample size (n_j)		**11**		**8**		**9**

$$n = 11 + 8 + 9 = 28$$

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

$$H = \frac{12}{28(28+1)} \left[\frac{145^2}{11} + \frac{104.5^2}{8} + \frac{156.5^2}{9} \right] - 3(28+1)$$

= (0.01478)(5997.75600) – 87 = 1.647

df = k – 1 = 3 – 1 = 2 Examining the Chi-square table, P > 0.20
There is weak evidence against H_0. P > α (0.10), therefore do not reject H_0.

Conclusion: At the 10% significance level, the data do not provide sufficient evidence to conclude that there is a difference in the abundance of ants in the three grasslands.
>>>>>>>>>>