ST102 Outline solutions to Exercise 18

1. Under the one-way ANOVA assumptions, $X_{ij} \sim_{IID} N(\mu_j, \sigma^2)$ within each $j = 1, 2, \ldots, k$. Therefore, since the X_{ij}s are independent with a common variance, σ^2, we have:

$$\bar{X}_{\cdot j} \sim N\left(\mu_j, \frac{\sigma^2}{n_j}\right) \quad \text{for } j = 1, 2, \ldots, k.$$

Hence:

$$a_j \bar{X}_{\cdot j} \sim N\left(a_j \mu_j, \frac{a_j^2 \sigma^2}{n_j}\right) \quad \text{for } j = 1, 2, \ldots, k.$$

Therefore:

$$\sum_{j=1}^{k} a_j \bar{X}_{\cdot j} \sim N\left(\sum_{j=1}^{k} a_j \mu_j, \sigma^2 \sum_{j=1}^{k} \frac{a_j^2}{n_j}\right).$$

2. We have $s_A^2 = 6.7992$, $s_B^2 = 7.2737$, $s_C^2 = 0.9247$ and $s_D^2 = 0.8419$. So we observe that although $s_A^2 \approx s_B^2$ and $s_C^2 \approx s_D^2$, the sample variances for treatments are very different across all groups, suggesting that the assumption that σ^2 is the same for all treatment levels may not be true.

3. For these $n = 30$ observations and $k = 3$ groups, we have $\bar{x}_{\cdot 1} = 12.67$, $\bar{x}_{\cdot 2} = 18.26$, $\bar{x}_{\cdot 3} = 14.96$ and $\bar{x} = 15.30$. Also:

$$\sum_{j=1}^{3} \sum_{i=1}^{10} x_{ij}^2 = 7{,}382.47.$$

Hence the total variation is:

$$\sum_{j=1}^{3} \sum_{i=1}^{10} x_{ij}^2 - n\bar{x}^2 = 7{,}382.47 - 30 \times (15.30)^2 = 359.77.$$

The between-groups variation is:

$$\begin{aligned} b = \sum_{j=1}^{3} n_j \bar{x}_{\cdot j}^2 - n\bar{x}^2 &= 10 \times ((12.67)^2 + (18.26)^2 + (14.96)^2) - 30 \times (15.30)^2 \\ &= 154.88. \end{aligned}$$

Therefore, $w = 359.77 - 154.88 = 204.89$. Hence the ANOVA table is:

Source	DF	SS	MS	F
Sector	2	154.88	77.44	10.20
Error	27	204.89	7.59	
Total	29	359.77		

To test the null hypothesis that the three types of stocks have equal price-earnings ratios, on average, we reject H_0 if:

$$f > F_{0.01,\,2,\,27} \approx 5.49.$$

Since $5.49 < 10.20$, we reject H_0 and conclude that there is strong evidence of a difference in the mean price-earnings ratios across the sectors.

4. (a) We have `A1` $= 3$, `A2` $= 284{,}400$, `A3` $= 1{,}034$ and `A4` $= 14.66$.

   (b) Since the p-value of the F test is 0.000, there exists strong evidence indicating that the mean test scores are different for children whose parents have different highest education levels.

   (c) We need to assume that we have independent observations $X_{ij} \sim N(\mu_j, \sigma^2)$ for $i = 1, 2, \ldots, n_j$ and $j = 1, 2, \ldots, k$.

5. We have $r = 5$ and $c = 3$.

   The row sample means are calculated using $\bar{X}_{i\cdot} = \sum_{j=1}^{c} X_{ij}/c$, which gives 19.77, 19.40, 19.87, 20.90 and 22.50 for $i = 1, 2, 3, 4, 5$, respectively.

   The column means are calculated using $\bar{X}_{\cdot j} = \sum_{i=1}^{r} X_{ij}/r$, which gives 22.28, 17.34 and 21.84 for $j = 1, 2, 3$, respectively.

   The overall sample mean is:

$$\bar{x} = \sum_{i=1}^{r} \frac{\bar{x}_{i\cdot}}{r} = 20.48\dot{6}.$$

   The sum of the squared observations is:

$$\sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 = 6{,}441.99.$$

   Hence:

$$\text{Total SS} = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}^2 - rc\bar{x}^2 = 6{,}441.99 - 15 \times (20.48\dot{6})^2 = 6{,}441.99 - 6{,}295.55 = 146.437.$$

$$b_{\text{row}} = c \sum_{i=1}^{r} \bar{x}_{i\cdot}^2 - rc\bar{x}^2 = 3 \times 2{,}104.83 - 6{,}295.55 = 18.924.$$

$$b_{\text{col}} = r \sum_{j=1}^{c} \bar{x}_{\cdot j}^2 - rc\bar{x}^2 = 5 \times 1{,}274.06 - 6{,}295.55 = 74.745.$$

   Residual SS $=$ Total SS $- b_{\text{row}} - b_{\text{col}} = 146.437 - 18.924 - 74.745 = 52.768$.

   To test the no row effect hypothesis $\mathrm{H}_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_5 = 0$, the test statistic value is:

$$f = \frac{(c-1)b_{\text{row}}}{\text{Residual SS}} = \frac{2 \times 18.924}{52.768} = 0.7173.$$

Under H_0, $F \sim F_{r-1,\,(r-1)(c-1)} = F_{4,\,8}$. Using Table 9 of Murdoch and Barnes' *Statistical Tables*, since $F_{0.05,\,4,\,8} = 3.84 > 0.7173$, we do not reject H_0 at the 5% significance level. We conclude that there is no evidence that the audience share depends on the city.

To test the no column effect hypothesis $\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3 = 0$, the test statistic value is:

$$f = \frac{(r-1)b_{\text{col}}}{\text{Residual SS}} = \frac{4 \times 74.745}{52.768} = 5.6660.$$

Under H_0, $F \sim F_{c-1,\,(r-1)(c-1)} = F_{2,\,8}$. Since $F_{0.05,\,2,\,8} = 4.46 < 5.6660$, we reject H_0 at the 5% significance level. Therefore, there is evidence indicating that the audience share depends on the network.

The results may be summarised in a two-way ANOVA table as follows:

Source	DF	SS	MS	F
City	4	18.924	4.731	0.7173
Network	2	74.745	37.373	5.6660
Residual	8	52.768	6.596	
Total	14	146.437		