ST102 Exercise 18

In this exercise you will practise one-way and two-way analysis of variance (ANOVA). In Questions 1 and 2 you are required to investigate the one-way ANOVA assumptions. Question 3 requires you to perform manual calculations to test the equality of group means. Question 4 requires you to complete and interpret computer output. Finally, Question 5 requires the manual construction of a two-way ANOVA table and tests for row and column effects.

Your answers to this problem set should be submitted as a pdf file upload to Moodle, *as directed by your class teacher*. It will be covered by your class teacher in your eighteenth class, which will take place in the week commencing Monday 13 March 2023.

1. Show that under the one-way ANOVA assumptions, for any set of constants $\{a_1, a_2, \ldots, a_k\}$, the quantity $\sum_{j=1}^{k} a_j \bar{X}_{\cdot j}$ is normally distributed with mean $\sum_{j=1}^{k} a_j \mu_j$ and variance $\sigma^2 \sum_{j=1}^{k} a_j^2 / n_j$.

2. Do the following data appear to violate the assumptions underlying one-way analysis of variance? Explain why or why not.

Treatment			
A	B	C	D
1.78	8.41	0.57	9.45
8.26	5.61	3.04	8.47
3.57	3.90	2.67	7.69
4.69	3.77	1.66	8.53
2.13	1.08	2.09	10.04

3. An indicator of the value of a stock relative to its earnings is its *price-earnings ratio*: the average of a given year's high and low selling prices divided by its annual earnings. The following table provides the price-earnings ratios for a sample of 30 stocks, ten each from the financial, industrial and utility sectors of the New York Stock Exchange. Test at the 1% significance level whether the true mean price-earnings ratios for the three market sectors are the same. Use the ANOVA table format to summarise your calculations. You may exclude the p-value.

Financial	Industrial	Utility
11.4	9.4	15.4
12.3	18.4	16.3
10.8	15.9	10.9
9.8	21.6	19.3
14.3	17.1	15.1
16.1	20.2	12.7
11.9	18.6	16.8
12.4	22.9	14.3
13.1	18.6	13.8
14.6	19.9	15.0

4. Proficiency tests are administered to a sample of 9-year-old children. The test scores are classified into four groups according to the highest education level achieved by at least one of their parents. The education categories used for the grouping are: 'less than high school', 'high school graduate', 'some college', and 'college graduate'.

   (a) Find the missing values `A1`, `A2`, `A3` and `A4` in the one-way ANOVA table below.

```
Source   DF      SS     MS      F      P
Factor   A1   45496  15165     A4  0.000
Error   275      A2     A3
Total   278  329896
```

```
                                  Individual 95% CIs For Mean Based on
                                  Pooled Standard Deviation
Level          N    Mean  StDev  -----+---------+---------+---------+----
Less than HS  41  196.83  30.23  (-----*------)
HS grad       73  207.78  29.34           (----*---)
Some college  86  223.38  34.58                     (----*---)
College grad  79  232.67  32.86                           (----*----)
                                  -----+---------+---------+---------+----
                                     195       210       225       240
Pooled Standard Deviation = 32.16
```

   (b) Test whether there are differences in mean test scores between children whose parents have different highest education levels.

   (c) State the required model conditions for the inference conducted in (b).

5. The following table shows the audience shares (in %) of three major networks' evening news broadcasts in five major cities, with one observation per cell so that there are 15 observations. Construct the two-way ANOVA table for these data (without the p-value column). Is either factor statistically significant at the 5% significance level?

		BBC	ITV	Sky
City	A	21.3	17.8	20.2
	B	20.6	17.5	20.1
	C	24.1	16.1	19.4
	D	23.6	18.3	20.8
	E	21.8	17.0	28.7