STAT/ME 424 HW 4
(due 9 AM Tue, Oct 25, 2022, in Canvas)

An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study follow (productivity improvement is measured on a scale from 0 to 100).

		i											
j		1	2	3	4	5	6	7	8	9	10	11	12
1	Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
2	Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
3	High	8.5	9.7	10.1	7.8	9.6	9.5						

1. Write down a model appropriate for the data.

2. Let μ_j denote the true mean productivity improvement for the jth level of expenditure. Obtain the values for the estimates $\hat{\mu}_j$, $j = 1, 2, 3$.

3. Obtain the residuals and plot them against the $\hat{\mu}_j$. What are your findings?

4. Make a normal quantile plot of the residuals. Does the normality assumption appear to be reasonable here?

5. The economist wishes to investigate whether location of the firm's home office is related to productivity improvement. The home office locations are as follows (U=U.S.; E=Europe):

		i											
	j	1	2	3	4	5	6	7	8	9	10	11	12
1	Low	U	E	E	E	E	U	U	U	U			
2	Moderate	E	E	E	E	U	U	U	U	U	E	E	E
3	High	E	U	E	U	U	E						

   Make side-by-side boxplots of residuals by location of home office. Does it appear that the ANOVA model could be improved by adding location of home office as a second factor? Explain.

6. Obtain the ANOVA table without using location of home office as a second factor.

7. Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Use a significance level of $\alpha = 0.05$. State your conclusion.

8. What is the significance probability (p-value) of the preceding test?

9. What appears to be the nature of the relationship between research and development expenditures and productivity improvement?

10. Estimate the mean productivity improvement for firms with high research and development expenditure levels with a 95% confidence interval.

11. Obtain a 95% confidence interval for $\mu_2 - \mu_1$. Interpret your interval estimate.

12. Obtain confidence intervals for all pairwise comparisons of the treatment means; use the Tukey procedure and a 90% simultaneous confidence level. State your findings.

13. Is the Tukey procedure employed in the preceding question the most efficient one that could be used here? Explain.

14. Obtain a 95% confidence interval for $(\mu_1 + \mu_2)/2 - \mu_3$, the difference in mean productivity improvement between firms with low or moderate research and development expenditures and firms with high expenditures. Interpret your interval estimate.

15. The sample sizes for the three treatment levels are proportional to the population sizes. The economist wishes to estimate the mean productivity gain last year *for all firms in the population*. Find a 95% confidence interval for it.

16. Using the Scheffé method, obtain 90% simultaneous confidence intervals for these 4 contrasts:

$$\mu_3 - \mu_2, \quad \mu_2 - \mu_1,$$
$$\mu_3 - \mu_1, \quad (\mu_1 + \mu_2)/2 - \mu_3.$$

What do you conclude?

17. Would the Bonferroni method be more powerful than the Scheffé method for the previous question? Explain.