Hand in solutions for ALL questions by February 24, 2022

1. The table below gives a 28×4 data matrix related to weights of bark deposits of 28 trees in the four directions north (N), east (E), south (S), and west (W). The data set is generated by the SAS program CORKDATA.SAS.

N	E	S	W		N	E	S	W
72	66	76	77		91	79	100	75
60	53	66	63		56	68	47	50
56	57	64	58		79	65	70	61
41	29	36	38		81	80	68	58
32	32	35	36		78	55	67	60
30	35	34	26		46	38	37	38
39	39	31	27		39	35	34	37
42	43	31	25		32	30	30	32
37	40	31	25		60	50	67	54
33	29	27	36		35	37	48	39
32	30	34	28		39	36	39	31
63	45	74	63		50	34	37	40
54	46	60	52		43	37	39	50
47	51	52	45		48	54	57	43

   (a) Check for univariate normality for each of the four variables. Check for multivariate normality for all four variables using the Q-Q plot of the squared Mahalanobis distances.

   (b) Assume multivariate normality, test the null hypothesis that the mean thickness of cork is the same in all directions at the 5% significance level.

   (c) The subject matter knowledge suggested that the cork borings are likely to be uniform in the N and S directions and also uniform in E and W directions. Assuming multivariate normality, formulate a hypothesis for this suggestion and test it at the 5% level. Find a 95% confidence interval for the contrast of means: $N - E - W + S$. Draw a relevant conclusion about this suggestion.

2. Mosteller and Tukey (1977) considered data from The Coleman Report on the relationships between several variables and mean verbal test scores for sixth graders at 20 schools in the New England and Mid-Atlantic regions of the United States. Two variables, $x_1 =$ the percentage of sixth-graders' fathers employed in white collar jobs and $x_2 =$ the sixth-graders' mothers' mean number of years of schooling, are given in the following table. The data are stored in the file SIXTHGRADERS.SAS

   (a) Check for univariate normality for x_1, $\arcsin(\sqrt{x_1/100})$ and $\sqrt{x_1}$. Which one is more closer to normality?

   (b) Adopt the transformations you chose in (a), test the null hypothesis that the percentage of white collar fathers is 50% and that the mothers have on average 12 years of schooling. Use 0.05 significance level.

x_1	x_2	x_1	x_2
28.87	12.38	12.20	11.33
20.10	10.34	22.55	10.51
69.05	14.10	14.30	11.72
65.40	14.20	31.79	12.33
29.59	12.00	11.60	11.38
44.82	12.82	68.47	14.02
77.37	13.70	42.64	12.95
24.67	11.53	16.70	12.05
65.01	13.11	86.27	15.37
9.99	11.23	76.73	14.74

3. Five male and five female patients suffering from a form of asthma were given a treatment to relieve the condition. The severity of the condition was measured by forced expiratory volume (FEV) and in this trial the FEV was measured at the commencement of treatment and at fixed times during the following two hours. The following table reports the increase in FEV after commencement for each patient, a positive increase denoting an improvement in condition. The data are stored in the file FEV.DAT

Sex	Patient	Time after commencement (minutes)				
		5	10	30	60	120
M	1	13	20	16	20	15
	2	22	27	31	21	17
	3	20	28	27	23	19
	4	18	26	18	18	9
	5	22	23	22	16	10
	Mean	19.0	24.8	22.8	19.6	14.0
F	1	18	17	20	18	18
	2	31	24	21	26	20
	3	14	16	17	20	17
	4	25	24	31	26	18
	5	36	28	24	26	29
	Mean	24.8	21.8	22.6	23.2	20.4

You may assume multinormality and common population covariance matrix.

(a) Construct a plot of the mean response for males and females over different times.

(b) Denote $\boldsymbol{\mu}_M$ and $\boldsymbol{\mu}_F$ respectively the mean responses for males and females over 5 different time points.

   (i) Assuming the same covariance matrices for male and female patients, test $H_0 : \boldsymbol{\mu}_M = \boldsymbol{\mu}_F$ at the 5% significance level.

   (ii) Describe in words what $H_0 : \boldsymbol{C}(\boldsymbol{\mu}_M - \boldsymbol{\mu}_F) = \boldsymbol{0}$ means, where

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

   Hence, perform a test for H_0 at the 5% significance level.