MG226 : Advanced Analytics
Midterm 2018

1. Let X and Y respectively denote the total amounts paid and received by a small business, in thousands of Rs., on a given work-day. It is given that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 45 \\ 50 \end{pmatrix}, \begin{bmatrix} 5 & -8 \\ 8 & 5 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 5 & 8 \\ -8 & 5 \end{bmatrix}\right)$$

Answer the following:

a. What is the probability of a work-day being profitable for that business ($Y > X$)? **[5]**

b. Give the interval with the smallest length, which has a 95% probability of capturing the amount received by the business on a given work-day. **[3]**

c. In the given graph-paper, plot the region with smallest area, which has a 95% probability of simultaneously capturing both the amounts paid and received by the business, on any given work-day. What are the minimum and maximum values of the Ys in this region? Why is this interval wider than the one obtained in **b**? **[7+2+3=12]**

d. Give the shortest interval, which has a 95% probability of capturing the amount received by the business on a given work-day, if on that day, it had made a total payment of Rs. 50,000. Why is this interval shorter than the one obtained in **b**? **[7+3=10]**

e. The (X, Y) values for the business on a certain day was (50,60) and on another day it was (37.5,42.5). Which one of these two days were more "typical" and why? **[7]**

2. A $3 \times 2 \times 2$ experiment is conducted to study the effects of Price (P), Brand (B) and Store (S) on consumers' Willingness to Purchase (Y), as follows. 60 MBA students in a class are randomly divided in 12 groups of 5 each. The 12 groups arise out of full-factorial combination of 3 Price levels of Low, Medium and High; 2 Brand levels of presence and absence of the Brand information; and 2 Store levels of presence and absence of the information on the name of the store, in which the product is available at the disclosed price level. The product experimented with, is a stereo ear-phone. The brand chosen (and is also made known to the Brand-treatment group) is Sony. The store-name given to the Store-treatment group is Walmart. Every student is shown an advertisement video of the product, and is given its Price, Brand and Store information, according to the group, the student belongs. The students respond by rating their Willingness to Purchase the product in a Lickert scale of 1-7, with 1 indicating no interest to 7 indicating extreme interest in buying the ear-phones. The (mean, standard deviation) of these Willingness to Purchase scores, in these 12 groups, are as follows:

Brand Name	Not Known			Known		
Price → Store Name ↓	Low	Medium	High	Low	Medium	High
Not Known	(2.0,0.71)	(3.6,1.82)	(4.8,0.45)	(4.6,0.55)	(5.0,1.00)	(5.0,1.22)
Known	(2.4,1.14)	(3.4,1.52)	(4.8,0.45)	(6.4,1.14)	(5.6,0.55)	(6.4,0.55)

Though the statistics reported above are *sufficient* for analyzing the data, for your convenience, SSEs of some of the fitted `aov` models in `R` notation, are further provided as follows:

Model	$Y \sim P$	$Y \sim B$	$Y \sim S$	$Y \sim P * B$	$Y \sim P * S$	$Y \sim B * S$
SSE	139.1	99.0	152.3	64.4	130.4	86.9

Answer the following:

a. Write the `aov`-model in `R` notation, which is most appropriate for this data set, with its ANOVA table. **[2+8=10]**

b. According to the model fitted in **a**, order the three factors according to their degree of effects, on a consumer's Willingness to Purchase. **[3]**

c. How many multiple comparisons are needed according to the model fitted in **a**? Give 95% Tukey-HSD confidence interval for the difference in Willingness to Purchase for the Brand name being known. Show that this effect (of Brand name being known) though is significant for Low and Medium priced ear-phones, it is not so if the Price is High. **[2+4+3×3=15]**

d. Write a brief summary of how the three factors are affecting a consumer's Willingness to Purchase. **[7]**

3. Let $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and based on a sample of size 30, it is found that $\bar{X}_1 = 1.2$, $\bar{X}_2 = 1$ and $\bar{X}_3 = 1.5$ and $\boldsymbol{S} = \begin{bmatrix} 1 & 0.5 & 1 \\ 0.5 & 0.5 & 1 \\ 1 & 1 & 3 \end{bmatrix}$. Test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$. **[13]**