MG221:Two Sample Problems

1. After testing 100 picture tubes supplied by company A it is found that they have a mean life length of 40000 hours with a standard deviation of 1000 hours, while the test of 80 picture tubes supplied by company B yielded a mean life-length of 40500 hours and a standard deviation of 950 hours. Which supplier should be preferred and why?

2. Top speed attained by 40 solar powered light motor vehicles have a mean of 100 Km. per hour with a standard deviation of 30 Km. per hour. On the contrary the average top speed attained by 60 battery (electric) powered light motor vehicle is 105 Km. per hour with a standard deviation of 2 Km. per hour. Is there a strong evidence to conclude that on an average the battery powered light motor vehicles are capable of attaining higher top speed compared to their solar powered counter-parts?

3. Mean mileages (defined in terms of km/ltr) of two brands of four cylinder cars A and B are to be compared. 40 cars of both the brands were driven in normal city condition. Cars of brand A yielded a mean mileage of 12 km/ltr with a standard deviation of 2, while the respective statistics for cars of brand B were 14 and 2.5 respectively. Is there any significant evidence of a difference in mean mileage of the two brands of cars?

4. Features of e-mails can be studied by programs to automatically classify them as spams. Two such sets of features are under consideration for their effectiveness in detecting and blocking spams. Assume that both the sets of features are such that they do not block non-spam mails. Also assume that the sets of mails used for comparing efficacies of the two sets of features are disjoint but of the same cardinality. At least how many mails need to be tested with each set of features to compare their effectiveness in detecting spam with an error of at most $\pm 1\%$ with 95% confidence?

5. The number of hours a lap-top can run before requiring a recharge is to be compared for two brands. From earlier studies it has been estimated that the standard deviations for the two brands are 0.5 and 0.4 respectively. Assume that the same number of trials would be conducted for observing the number of hours it lasts before recharging, for both the brands. At least how many trials would be required to estimate the mean difference with a precision of ± 0.1 hours with 95% confidence?

6. The mean daily productions (in thousands of meters) in two different plants of a textile merchant are 120 and 95 respectively, with sample standard deviations 30 and 25, observed over a period of 20 days. Assume that the daily productions in the different plants are independent and each has a Normal distribution. Answer the following:

a. Is there a conclusive evidence against the assumption of homoscedasticity? Test using α=0.1.

b. Is there enough evidence to suggest that the first plant is more productive than the second one?

7. In two independent samples of 150 and 200 consumers of electricity in Bengaluru and Mumbai respectively, 120 and 50 are found to be dissatisfied with the quantity of the supply. Based on this data, would it be safe to conclude that the difference of the proprtions of

dissatisfied consumers of electricity between Bengaluru and Mumbai exceeds 50%? What can be said about the minimum value of this difference with 95% confidence?

8. If 45 of the 500 fresh recruits by Organization A leave their jobs within a year, while 25 out of 200 do so for Organization B, then is there sufficient evidence to conclude that there is a significant difference in the (one-year) attrition rates between the two orghanizations? Give a 90% confidence interval for the difference of (one-year) attrition rates between the two organizations.

9. In order to compare the performance of two financial softwares, A and B, say, 16 distinct jobs are first randomly divided into two groups of 8. The first group of jobs are then run using A while the second group of jobs are run using B. The time (in minutes) taken to finish the jobs are recorded and are as follows:

A	12	10	22	18	16	18	17	25
B	10	15	13	18	14	12	16	11

Answer the following:

a. Is there any significant evidence for the superiority of B?

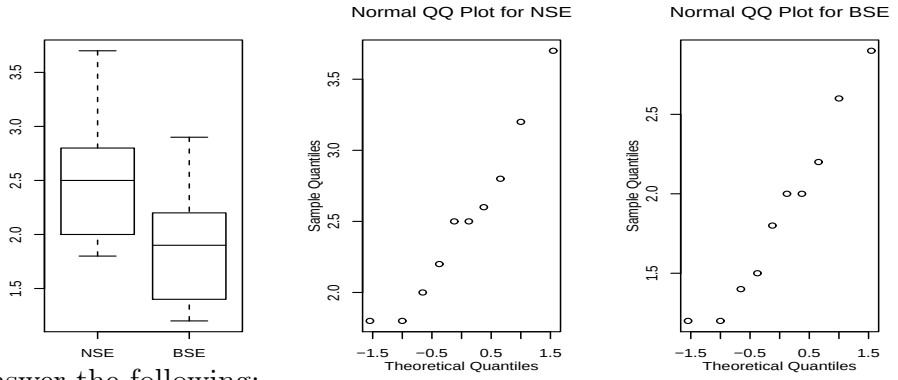
b. Give a 90% confidence interval for the difference in mean times in finishing a job using A and B.

c. Can you think of a better way to design the experiment for comparing their performances?

10. The daily maximum bid-ask spread of a stock in NSE and BSE for two consecutive weeks are as follows:

NSE	2.5	2.2	3.2	2.8	2.6	1.8	3.7	2.5	1.8	2.0
BSE	1.2	2.9	1.8	1.4	1.2	2.6	2.0	2.2	1.5	2.0

A preliminary graphical analysis of the data set is as follows:



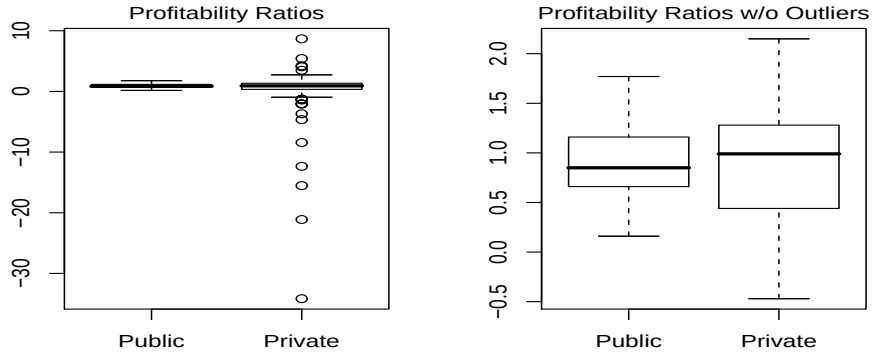
Answer the following:

a. Conduct an appropriate parametric test to see whether there is a significant difference in the mean bid-ask spread of the stock in the two exchanges.

b. Answer the same using an appropriate non-parametric test.

c. Which of the above two tests should be preferred and why?

11. Mean and standard deviations of Profitability Ratios of 29 public sector banks last financial year were 0.8876 and 0.2041 respectively, and similarly the mean and standard deviations of Profitability Ratios of 143 private sector banks last financial year were 0.2041 and 4.1476 respectively with the following distributions:



The mean and standard deviation after removing 25 outliers for the private sector banks based on 118 observations are 0.9387 and 0.5534 respectively, whose distribution along with the public sector banks (with 29 original observations) are also plotted above in the second figure. From an investor's point of view, who is interested in allocating part of her funds in the banking sector for diversification purpose and is solely interested in profitability, does it make any difference in the sector of banks (public or private) she invests in?

12. Average monthly default amounts (in thousands of Rs.) of 20 randomly selected credit card accounts of a bank, classified according to the card-holder's income level is as follows:

Monthly Income $<$ Rs.15,000:	1.5	4.8	1.4	2.6	0	2.0	3.3	0	1.4	8.1	0	0.8
Monthly Income $\geq$ Rs.15,000:	1.2	0	2.8	0	1.3	0	5.5	3.6				

Do the card-holders' income level affect the average monthly default amounts in their credit card accounts?

13. In order to assess the effect of music on productivity of workers in an embroidery sewing mill, which has only two shifts, music is played in one shift and not in the other. The number of units produced by different workers in the two different shifts are as follows:

No Music	15,	12,	14,	17,	14,	16,	12,	17		
Music	18,	16,	12,	22,	16,	18,	17,	15,	22,	16

Answer the following:

a. Assume that the productiveness of workers in the two different shifts are Normally distributed. Does music indeed improve productivity of the workers?

b. Give a 95% Confidence Interval for the expected increment in the productivity of the workers due to music.

c. What is the fault in the above design for assessing the effect of music on the productivity of workers? Suggest a better way of collecting observations for this problem.

14. Interest rates provided by two financial institutions A and B, say, for 1, 2, 3, 4 and 5 year term fixed-deposits are as follows:

Term (years) → Institution ↓	1	2	3	4	5
A	12%	15%	15%	18%	20%
B	12.5%	15%	16%	18%	18%

Is there a significant difference between the interest rates provided by A and B? Test using appropriate assumptions.

15. The number of minutes it takes to load and complete an e-transaction from 10 randomly chosen remote locations using two different web programs are as follows:

Location	1	2	3	4	5	6	7	8	9	10
Program A	4.95	3.59	3.82	4.60	2.90	3.21	2.65	1.21	2.60	1.11
Program B	4.90	3.43	3.79	4.76	3.08	3.17	2.61	1.22	2.37	1.14

After clearly stating the required assumptions, use an appropriate test to decide whether one Program is better than the other.

16. Typing speeds (in words per minute) of 12 typists in an office before and after they have gone through a month long typing training program are as follows:

Before	45	36	52	42	44	38	54	28	47	30	51	45
After	47	42	52	45	45	42	52	33	49	34	50	46

Answer the following:

a. Did the month long training program help improve the typists' speed?

b. What can be said about the minimum expected gain in typing speed as a result of the training program with 95% confidence?

17. The number of copies sold (in thousands of units) and the price (in $'s) of 10 different software products of a company in the domestic and foreign markets are as follows:

Product No.		1	2	3	4	5	6	7	8	9	10
Domestic Market	No. Sold	15	5	15	10	15	25	4	4	10	18
	Price	50	100	80	125	100	75	150	200	100	100
Foreign Markets	No. Sold	10	5	7	12	8	10	5	2	6	8
	Price	80	200	200	100	200	150	250	500	200	300

Answer the following:

a. Using an appropriate non-parametric test and stating the assumptions, show that the number of copies sold in the domestic market is significantly greater at 5% level.

b. Show that the foreign markets are however more lucrative than the domestic market, by means of a t-test, stating the assumptions.

18. An HR consultant group claims to be able to improve soft-skills of techies by more than 1.5 times through its training method. An organization suffering from lack of soft-skills of its junior engineers, first internally assessed the soft-skills of 8 of its junior engineers by an expert committee giving a score between 0 to 100, and then sent them to the above mentioned training programme. The same expert committee then again assessed their soft skills (by giving a score between 0 to 100 using the same criteria as before) upon the completion of the programme with the following observations:

Eng. No.	1	2	3	4	5	6	7	8
Before	27	42	32	19	39	25	21	31
After	61	70	58	35	42	39	35	41

Answer the following:

a. After writing down the assumptions in precise terms with respect to this problem, perform an appropriate t-test for verifying the claim of the HR consultant group.

b. Answer the same as above (both assumptions and hypothesis test) using a rank-based distribution free test.

c. As the proponent of the idea of sending the junior engineers to the programme, what kind of feed back should you give to the HR consultants?

19. Let X and Y respectively denote the investment amounts and returns on investments. The summary statistics of yearly returns and investment amounts (in crores of Rs.) of 10 ventures in Agriculture and 12 ventures in Real Estate are as follows:

Variable	$R = Y/X$		$L = \log(R)$	
Statistics	Mean	Variance	Mean	Variance
Agriculture	0.2576	0.0050	-1.3893	0.0844
Real Estate	0.1043	0.0006	-2.2812	0.0478

It is hypothesized that $Y = \rho \times X \times \varepsilon$, where ρ is the yearly rate of return of an investment, which depends on the type of an investment, and the error $\epsilon = \log \varepsilon \sim N(0, \sigma^2)$. Answer the following:

a. Why the observed rate of return R cannot be directly compared for the two types of investments by means of a two-sample t-test under the hypothesized model?

b. Show that such a direct comparison of R also lacks empirical validation.

c. Show that under the hypothesized model, L can however be compared for the two types of investments by means of a two-sample t-test, which also enjoys empirical support.

d. Which investment is expected to bring higher returns?

20. In a typical auditing exercise usually a preliminary report is first prepared by a "senior" which is subsequently reviewed by an "audit manager". Audit managers frequently work with a familiar regular group of seniors. This familiarity with the work of sub-ordinate(s) has both positive and negative implications on effectiveness of the audit review process in terms of its objectivity, and as well as on the performance appraisal of the subordinates. Thus in order to assess the audit managers' ability to objectively evaluate the quality of seniors' work, and to test whether "good" managers are less susceptible to prior impressions of their subordinates, an experiment is conducted as follows.

40 audit seniors and 20 audit managers are chosen from a well-known audit firm and each audit manager is matched with two audit seniors with whom the audit manager is familiar. The firm's performance appraisal system rates one of the seniors as good, and the other as average. The firm has also classified the audit managers as either good or average with 10 of them selected from each of these two categories. Only one auditing case is identified for this experiment. Each senior reads the description of this auditing case and then writes a

memo to his or her paired manager on the key issues of which the manager should be aware. Each senior's identity appears on the corresponding memo, from which the managers know whether the memo is written by a good senior or not, and then the managers evaluate the two memos separately on a ll-point scale from 0 (very poor memo) to 10 (extremely good memo). This is Stage 1 of the experiment. Stage 2 is conducted about three months later so that the managers' memories of the case and the grades they had assigned to the seniors are hopefully cleared. Also in Stage 1 the managers are unaware that they will be asked to evaluate the memos for a second time. In Stage 2 each manager again evaluates the quality of the same memos written by the same two seniors they had evaluated in Stage 1. This time each senior's identification was removed from the memos and the managers did not even know whether they are evaluating memos written by seniors familiar to them. Following table gives the summary statistics of the ratings provided by the managers to their corresponding seniors:

↓ Managers Seniors→		Stage 1			Stage 2			Difference in Stages 1 & 2
		Good	Average	Difference	Good	Average	Difference	
(0)		(1)	(2)	(3)=(1)-(2)	(4)	(5)	(6)=(4)-(5)	(7)=(3)-(6)
Good	Mean:	6.5	4.3	2.2	5.7	3.9	1.8	0.4
	SD:	2.07	1.57	1.14	2.06	1.60	1.23	0.70
Average	Mean:	7.0	4.9	2.1	5.8 2.62	5.7	0.1	2.0
	SD:	2.45	2.28	1.10		1.77	1.37	0.84
All	Mean:	6.75	4.6	2.15	5.75	4.8	0.95	1.2
	SD:	2.22	1.93	1.09	2.29	1.88	1.53	0.80

Answer the following assuming normality whenever required:

a. Do the managers' tend to give higher ratings to seniors when they know they are good? Explain why this difference cannot be attributed to the familiarity of the managers to their seniors.

b. Are the managers good at discerning between good and average seniors?

c. Show that even though good managers as such may not be very good at discerning a good senior from an average one, nevertheless they are better at it than an average manager.

d. Show that in general familiarity might affect a manager's rating of his or her seniors, however good managers tend to be more objective in their ratings.

e. Comment on the design of this experiment.