LAB TASK - 2 (21BKT0004)

💡 NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L
SLOT: L23+L24

---

1. Write down the R code to compute the coefficient of correlation between X and Y from the following data:

X :	21	23	30	54	57	58	72	78	87	90
Y :	60	71	72	83	110	84	100	92	113	135

```
> x=c(21,23,30,54,57,58,72,78,87,90)
> y=c(60,71,72,83,110,84,100,92,113,135)
> var(x)
[1] 649.5556
> var
[1] 520.8889
> var(x,y)
[1] 510.4444
```

```
> r=var(x,y)/sqrt(var(x)*var(y))
> r
[1] 0.8775417
> cor(x,y)
[1] 0.8775417
> cor.test(x,y)

  Pearson's product-moment correlation
data: x and y
t = 5.1764, df = 8, p-value = 0.0008465
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.5540299 0.9707861
sample estimates:
> cor
0.8775417
```

---

2. Write down the R code to find the rank correlation between the ranks of the variable X and Y from the following data:

$X:$	10	15	12	17	13	16	24	14	22
$Y:$	30	42	45	46	33	34	40	35	39

```
> x=c(10,15,12,17,13,16,24,14,22)
> y=c(30,42,45,46,33,34,40,35,39)
> cor.test(x,y,method="spearman")
Spearman's rank correlation rho
data: x and y
S = 72, p-value = 0.2912
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.4
```

---

3. Write down the R code to obtain the equation of the regression line of X on Y from the following data:

$X:$	4.7	8.2	12.4	15.8	20.7	24.9	31.9	35.0	39.1	38.8
$Y:$	4.0	8.0	12.5	16.0	20.0	25.0	31.0	36.0	40.0	40.0

```
> x=c(4.7,8.2,12.4,15.8,20.7,24.9,31.9,35.0,39.1,38.8)
> y=c(4.0,8.0,12.5,16.0,20.0,25.0,31.0,36.0,40.0,40.0)
> cor(x,y)
[1] 0.998992
> model=lm(y~x)
> summary.lm(model)
Call:
lm(formula = y ~ x)
Residuals:
Min 1Q Median 3Q Max
-1.3140 -0.1191 0.2321 0.3803 0.5383
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.73087 0.42971 -1.701 0.127
x 1.03589 0.01646 62.946 4.51e-12 *
---
Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.6286 on 8 degrees of freedom
Multiple R-squared: 0.998, Adjusted R-squared: 0.9977
F-statistic: 3962 on 1 and 8 DF, p-value: 4.511e-12
```

4. Write down the R code to obtain the equation of the regression plane of Y on X_1 and X_2 from the following data:

X_1 :	30	40	20	50	60	40	20	60
X_2 :	11	10	7	15	19	12	8	14
Y :	110	80	70	120	150	90	70	120

```
> x1=c(30,40,20,50,60,40,20,60)
> x2=c(11,10,7,15,19,12,8,14)
> y=c(110,80,70,120,150,90,70,120)
> model=lm(y~x1+x2)
> model
Call:
lm(formula = y ~ x1 + x2)
Coefficients:
(Intercept) x1 x2
16.8314 -0.2442 7.8488
> summary.lm(model)
Call:
lm(formula = y ~ x1 + x2)
Residuals:
1 2 3 4 5 6 7 8
14.157 -5.552 3.110 -2.355 -1.308 -11.250 -4.738 7.936
Coefficients:
```

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.8314 11.8290 1.423 0.2140
x1 -0.2442 0.5375 -0.454 0.6687
x2 7.8488 2.1945 3.577 0.0159 *
---
Signif. codes: 0 ‘*’ 0.001 ‘*’ 0.01 ‘’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 9.593 on 5 degrees of freedom
Multiple R-squared: 0.9191, Adjusted R-squared: 0.8867
F-statistic: 28.4 on 2 and 5 DF, p-value: 0.001862
```

LAB TASK - 1

NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L

1. Calculate the mean, median and quartile deviation from the following data

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	12	16	26	38	22	15	7	4

R 4.2.2 · ~/

```
> x=c(5,15,25,35,45,55,65,75)
> f = c(12, 16, 26, 38, 22, 15, 7, 4)
> mean = sum(x*f)/sum(f)
> mean
[1] 34.64286
> c = cumsum(f)
> cl=cumsum(f)
> cl
[1]  12  28  54  92 114 129 136 140
> N=sum(f)
> N
[1] 140
> ml=min(which(cl>N/2))
> ml
[1] 4
> h=4
> fm=f[ml]
> fm
[1] 38
> cf=cl[ml-1]
> cd
Error: object 'cd' not found
> cf
[1] 54
> l=x[ml]-h/2
> l
[1] 33
> median=l+(((N/2)-cf)/fm)*h
> median
[1] 34.68421
> Q1=min(which(cl>N/4))
> Q1
[1] 3
> fq1=f[Q1]
> fq1
[1] 26
> cf1=cl[Q1-1]
> cf1
[1] 28
> l=x[Q1]-h/2
> l
```

```
> fq1
[1] 26
> cf1=cl[Q1-1]
> cf1
[1] 28
> l=x[Q1]-h/2
> l
[1] 23
> quartile1=l+(((N/4)-cf1)/fq1)*h
> quartile1
[1] 24.07692
> Q3=min(which(cl>3*N/4))
> Q3
[1] 5
> fq3=f[Q3]
>
> fq3
[1] 22
> cf2=cl[Q3-1]
> cf2
[1] 92
>
> l=x[Q3]-h/2
> l
[1] 43
> quartile3=l+(((3*N/4)-cf2)/fq3)*h
> quartile3
[1] 45.36364
>
```

2. Calculate the mean, median and quartile deviation from the following data

x	90 - 99	100 - 109	110 - 119	120 - 129	130 - 139	140 - 149	150 - 159	160 - 169	170 - 179	180 - 189	190 - 199	200 - 209
f	1	14	66	122	145	121	65	34	12	5	2	2

```
R 4.2.2 · ~/
> x=c(95,105,115,125,135,145,155,165,175,185,195,205)
> f=c(1,14,66,122,145,121,65,34,12,5,2,2)
> mean=sum(x*f)/sum(f)
> mean
[1] 137.5806
> c=cumsum(f)
> cl=cumsum(f)
> cl
 [1]   1  15  81 203 348 469 534 568 580 585 587 589
> N=sum(f)
> N
[1] 589
> ml=min(which(cl>N/2))
>
> ml
[1] 5
> h=5
> h
[1] 5
> fm=f[ml]
> fm
[1] 145
> cf=cl[ml-1]
>
> cf
[1] 203
> l=x[ml]-h/2
> l
[1] 132.5
> median=l+(((N/2)-cf)/fm)*h
> median
[1] 135.6552
> Q1=min(which(cl>N/4))
>
> Q1
[1] 4
> fq1=f[Q1]
> fq1
[1] 122
> cf1=cl[Q1-1]
> cf1
```

```
> fq1=f[Q1]
> fq1
[1] 122
> cf1=cl[Q1-1]
> cf1
[1] 81
> l=x[Q1]-h/2
> l
[1] 122.5
> quartile1=l+(((N/4)-cf1)/fq1)*h
> quartile1
[1] 125.2152
> Q3=min(which(cl>3*N/4))
> Q3
[1] 6
> fq3=f[Q3]
>
> fq3
[1] 121
> cf2=cl[Q3-1]
> cf2
[1] 348
> l=x[Q3]-h/2
> l
[1] 142.5
> quartile3=l+(((3*N/4)-cf2)/fq3)*h
> quartile3
[1] 146.374
>
```

LAB TASK - 3 (21BKT0004)

NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L
SLOT: L23+L24

---

1. It is known that probability of an item produced by a certain machine will be defective is 0.05. If the produced items are sent to the market in packets of 20, then write down the R code to find the number of packets containing at least, exactly and at most 2 defective items in a consignment of 1000 packets.

```
> 1000*(1-pbinom(1,20,0.05))
[1] 264.1605
> 1000*dbinom(2,20,0.05)
[1] 188.6768
> 1000*pbinom(2,20,0.05)
[1] 924.5163
```

2. A car hire firm has 2 cars which it hires out day by day. The number of demands for a car on each day follows a Poisson distribution with mean 1.5. Write down the R code to compute the proportion of days on which (i). neither car is used, (ii). at most one car is used and (iii). some demand of car is not fulfilled.

```
> #Y denotes the number of car hired a day. Y follows P(1.5)
> #p(Y=0)
> dpois(0,1.5)
[1] 0.2231302
> #p(Y<=1)
> ppois(1,1.5)
[1] 0.5578254
> #p(Y>2)
> 1-ppois(2,1.5)
[1] 0.1911532
```

3. The local corporation authorities in a certain city install 10,000 electric lamps in the streets of the city with the assumption that the life of lamps is normally distributed. If these lamps have an average life of 1,000 burning hours with a standard deviation of 200 hours, then write down the R code to calculate the number of lamps might be expected to fail in the first 800 burning hours and also the number of lamps might be expected to fail between 800 and 1,200 burning hours.

```
> #Z denotes lifetime of bulbs. Z follows N(1000,200)
> #Number of lamps with Z<800
> 10000*pnorm(800,1000,200)
[1] 1586.553
> #Number of lamps with 800<Z<1200
> 10000*(pnorm(1200,1000,200)-pnorm(800,1000,200))
[1] 6826.895
```

LAB TASK - 4 (21BKT0004)

NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L
SLOT: L23+L24

---

1. Experience has shown that 20% of a manufactured product is of top quality. In one day's production of 400 articles, only 50 are of top quality. Write down the R programming code to test whether the production of the day chosen is a representative sample at 95% confidence level.

```
> alpha<-0.05
> p0<-0.2
> p1<-50/400
> se<-sqrt(p0*(1-p0)/400)
> z<-(p1-p0)/se
> p_value<-2*(1-pnorm(abs(z)))
> if (p_value < alpha) {
+ cat("Reject H0. The production of the day chosen is not a representative
sample.")
```

```
+ } else {
+ cat("Fail to reject H0. The production of the day chosen is a representative
sample.")
+ }
Reject H0. The production of the day chosen is not a representative sample.
```

---

2. Before an increase in excise duty on tea, 800 people out of a sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Write down the R programming code to test whether the significant decrease in the consumption of tea after the increase in duty at $1\ \%$ level of significance.

```
> n1<-1000
> n2<-1200
> p1<-800/n1
> p2<-800/n2
> p<-(800+800)/(n1+n2)
> z<-(p1-p2)/sqrt(p*(1-p)*(1/n1+1/n2))
> p_value<-2*pnorm(-abs(z))
>
> alpha<-0.01
> if (p_value < alpha) {
+ cat("There was a significant decrease in tea consumption after the increase in duty at the",
alpha*100, "% level of significance.")
+ } else {
+ cat("There was no significant decrease in tea consumption after the increase in duty at the",
alpha*100, "% level of significance.")
+
+
+ }
There was a significant decrease in tea consumption after the increase in duty at the 1 % level
of significance.
```

---

3. A sample of 900 items is found to have a mean of 3.47 cm. Write down the R programming code to test whether it can be reasonably regarded as a simple sample from a population with mean 3.23 cm and SD 2.31 cm at 99% level of confidence.

```
> n<-900
> xbar<-3.47
```

```
> mu0<-3.23
> sd<-2.31
> alpha<-0.01
> se<-sd/sqrt
> t<-(xbar-mu0)/se
> p_value<-pt(t,df=n-1,lower.tail=FALSE)
> if (p_value < alpha) {
+ cat("The sample can be reasonably regarded as a simple random sample from a population
with mean", mu0, "cm and SD", sd, "cm at the", (1-alpha)*100, "% level of confidence.")
+ } else {
+ cat("The sample cannot be reasonably regarded as a simple random sample from a
population with mean", mu0, "cm and SD", sd, "cm at the", (1-alpha)*100, "% level of
confidence.")
+ }
The sample can be reasonably regarded as a simple random sample from a population with
mean 3.23 cm and SD 2.31 cm at the 99 % level of confidence
```

---

4. The average mark scored by 32 boys is 72 with a standard deviation of 8, while that for 36 girls is 70 with a standard deviation of 6. Write down the R programming code to test whether the boys are performing better than girls on the basis of average mark at $5\ \%$ level of significance.

```
> n1 <- 32
> n2 <- 36
> xbar1 <- 72
> xbar2 <- 70
> sd1 <- 8
> sd2 <- 6
> alpha <- 0.05
> se <- sqrt(sd1^2/n1 + sd2^2/n2)
> t <- (xbar1 - xbar2) / se
> df <- (sd1^2/n1 + sd2^2/n2)^2 / ((sd1^2/n1)^2/(n1-1) + (sd2^2/n2)^2/(n2-1))
> p_value <- pt(t, df = df, lower.tail = FALSE)
> if (p_value < alpha) {
+ cat("The boys are performing better than girls on the basis of the average mark at the", (1-
alpha)*100, "% level of significance.")
+ } else {
+ cat("There is no significant difference in performance between boys and girls on the basis of
the average mark at the", (1-alpha)*100, "% level of significance.")
+ }
There is no significant difference in performance between boys and girls on the basis of the
average mark at the 95 % level of significance.
```

LAB TASK - 5 (21BKT0004)

NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L
SLOT: L23+L24

---

1. A random sample of 10 boys with the following IQs: 70, 120, 110, 101, 88, 83, 95, 98, 107, and 100. Write down the R programming code to test whether the data support the assumption of a population mean IQ of 100 at $5\ \%$ level of significance.

```
R 4.2.2 · ~/
> x=c(70,120,110,101,88,83,95,98,107,100)
> t.test(x,mn=100)

        One Sample t-test

data:  x
t = 21.535, df = 9, p-value = 4.726e-09
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  86.98934 107.41066
sample estimates:
mean of x
     97.2

>
```

p-value is less than 0.05 hence we reject the null hypothesis that the true mean is equal to 0.

We get, mean = 97.2

Since the p-value=0.0023 which is lesser than 0.05 ,hence the null hypothesis is rejected and it

shows that the soldiers are shorter than the sailors on the basis of average height

---

2. The mean height and the standard deviation height of 8 randomly chosen soldiers are 166.9 cm and 8.29 cm respectively. The corresponding values of 6 randomly chosen sailors are 170.3 cm and 8.50 cm respectively. Write down the R programming code to test whether the soldiers are shorter than the sailors on the basis of average height.

R 4.2.2 · ~/

```
> soldiers <- c(165.1, 167.6, 166.9, 167.6, 165.1, 168.9, 166.4, 167.6)
> sailors <- c(172.7, 168.9, 169.7, 170.3, 171.5, 167.6)
> mean_soldiers <- mean(soldiers)
> sd_soldiers <- sd(soldiers)
> mean sailors <- mean(sailors)
Error: unexpected symbol in "mean sailors"
> mean_sailors <- mean(sailors)
> sd_sailors <- sd(sailors)
> df <- length(soldiers) + length(sailors)-2
> pooled_sd <- sqrt(((length (soldiers) - 1) *sd_soldiers^2 + (length(sailors) - 1) * sd_sailors^2)/ df)
> t_stat<- (mean_soldiers mean_sailors) /(pooled_sd* sqrt (1/length (soldiers) + 1/length (sailors)))
Error: unexpected symbol in "t_stat<- (mean_soldiers mean_sailors"
> t_stat<- (mean_soldiers -mean_sailors) /(pooled_sd* sqrt (1/length (soldiers) + 1/length (sailors)))
> p_value <- 2 * pt(abs(t_stat), df, lower.tail=FALSE)
> cat("t-statistic: ", t_stat, "\n")
t-statistic:  -3.844096
> cat("p-value: ", p_value, "\n")
p-value:  0.002335042
> |
```

3. The following data relate to the marks obtained by 11 students in two sets, one held at the beginning of a year and the other at the end of the year after intensive coaching. Do the data indicate that the students have benefited by coaching at $5\,\%$ level of significance?

Test I :	19	23	16	24	17	18	20	18	21	19	20
Test II :	17	24	20	24	20	22	20	20	18	22	19

R 4.2.2 · ~/

```
> test1=c(19,23,16,24,17,18,20,18,21,19,20)
> test2=c(17,24,20,24,20,22,20,20,18,22,19)
> t.test(test1, test2, paired=TRUE)

	Paired t-test

data:  test1 and test2
t = -1.3772, df = 10, p-value = 0.1985
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -2.6179307  0.6179307
sample estimates:
mean difference 
             -1 

> 
```

p-value=0.19 is greater than 0.05(5% significance value) hence we accept the null hypothesis. Hence the students have no benefited from the coaching.

---

4. Two random samples drawn from two normal populations with the following observations.

Sample I :	21	24	25	26	27	
Sample II :	22	27	28	30	31	36

Write down the R programming code to test whether the two populations have the same variance at $5\ \%$ level of significance.

```
R 4.2.2 · ~/
> sample1 <- c(21, 24, 25, 26, 27)
> sample2 <- c(22, 27, 28, 30, 31, 36)
> var1 <- var (sample1)
> var2 <- var (sample2)
> F_stat<- var1/var2
> df1 <- length (sample1) - 1
> df2 <- length (sample2) - 1
> p_value <- pf (F_stat, df1, df2, lower.tail = FALSE)
> cat ("F-statistic: ", F_stat, "\n")
F-statistic:  0.2453704
> cat ("p-value: ", p_value, "\n")
p-value:  0.9009257
>
```

Since the p-value=0.9 is greater than the significance level(0.05) we will accept the null hypothesis.

Hence the two populations have the same variance at 5 % level of significance

LAB TASK - 6 (21BKT0004)

💡 NAME: SUGAM B KUBER
ROll : 21BKT0004
SUB: Probability and Statistics
CODE: BMAT202L
SLOT: L23+L24

---

1. The following table gives the number of fatal road accidents that occurred during the 7 days of a week. Write down the R programming code to test whether the accidents are uniformly distributed over the week at $95\ \%$ level of confidence.

Day :	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Number :	8	14	16	12	11	14	9

```
> observed <- c(8,14,16,12,11,14,9)
> expected <- rep(sum(observed)/7,7)
>
chisq.test(observed,p=expected,rescale.p=TRU
E)
Chi-squared test for given probabilities
```

```
data: observed
X-squared = 4.1667, df = 6, p-value = 0.6541
```

---

2. A total number of 3759 individuals were interviewed according to gender and decision in a public opinion survey on a political proposal with the results as in the following table. Write down the R programming code to test the hypothesis that there is no association between gender and attitude $5\ \%$ level of significance.

	Decision		
	Favoured	**Opposed**	**Undecided**
Male	1154	475	243
Female	1103	442	342

```
> observed <-
matrix(c(1154,475,243,1103,442,342),nrow=2,byrow=TRUE)
> rownames(observed) <-c("Male","Female")
> colnames(observed) <-c("Flavored","Opposed","Undecided")
> chisq.test(observed)
Pearson's Chi-squared test
data: observed
X-squared = 19.034, df = 2, p-value = 7.358e-05
```

---

3. A random sample is selected from each of 3 makes of ropes (Type 1, Type 2 and Type 3) and their breaking strength (in certain units) are measured with the results in the following table.

Type 1 :	70	72	75	80	83		
Type 2 :	60	65	57	84	87	73	
Type 3 :	100	110	108	112	113	120	107

Write down the R programming code to test whether the breaking strengths of the ropes differ significantly at $5\ \%$ level of significance.

```
> type1 <- c(70,72,75,80,83)
> type2 <- c(60,65,57,84,87,73)
```

```
> type3 <- c(100,110,108,112,113,120,107)
> rope_data <- data.frame(Type=c(rep("Type 1", length(type1)),
rep("Type 2", length(type2)), rep("Type 3",length(type3))),Strength =
c(type1, type2, type3))
> rope_aov <- aov(Strength ~ Type, data = rope_data)
> summary(rope_aov)
Df Sum Sq Mean Sq F value Pr(>F)
Type 2 5838 2919.2 38.89 1.16e-06 ***
Residuals 15 1126 75.1
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

---

4. A company appoints 4 salesman $(A, B, C \ \& \ D)$ and observes their sales in 3 seasons (Summer, Winter $\&$ Monsoon). The figures (Rs. in Lakhs) are given in the following table.

Seasons	Treatments			
	A	**B**	**C**	**D**
Summer	36	36	21	35
Winter	28	29	31	32
Mansoon	26	28	29	29

Write down the R programming code to perform an analysis of variance at $5\ \%$ level of significance.

```
> seasons = rep(c("summer", "winter", "monsoon"), 4)
> treatments = c(rep('a', 3), rep('b', 3), rep('c', 3), rep('d', 3))
> seasons
[1] "summer" "winter" "monsoon" "summer" "winter" "monsoon"
"summer" "winter" "monsoon" "summer"
[11] "winter" "monsoon"
> treatments
[1] "a" "a" "a" "b" "b" "b" "c" "c" "c" "d" "d" "d"
> figures = c(36, 28, 26, 36, 29, 28, 21, 31, 29, 35, 32, 29)
> figures
[1] 36 28 26 36 29 28 21 31 29 35 32 29
> aov(figures ~ seasons + treatments)
Call:
aov(formula = figures ~ seasons + treatments)
Terms:
seasons treatments Residuals
Sum of Squares 32 42 136
Deg. of Freedom 2 3 6
```

```
Residual standard error: 4.760952
Estimated effects may be unbalanced
```

---

5. The following data resulted from an experiment to compare three burners $(B_1, B_2 \,\&\, B_3)$. A Latin square design was used as the tests were made on 3 engines and were spread over 3 days.

Days	**Engines**		
	Engine 1	**Engine 2**	**Engine 3**
Day 1	B_1 – 16	B_2 – 17	B_3 – 20
Day 2	B_2 – 16	B_3 – 21	B_1 – 15
Day 3	B_3 – 15	B_1 – 12	B_2 – 13

Write down the R programming code to test the hypothesis that there is no difference between (i). days, (ii). engines and (iii). burners at $5\ \%$ level of significance.

```
> days = c("day1", "day2", "day3")
> engines = c(rep("eng1", 3), rep("eng2", 3), rep("eng3", 3))
> burners = c('b1', 'b2', 'b3', 'b2', 'b3', 'b1', 'b3', 'b1', 'b2')
> data = c(16, 16, 15, 17, 21, 12, 20, 15, 13)
> df = data.frame(engines, days, burners, data)
> df
engines days burners data
1 eng1 day1 b1 16
2 eng1 day2 b2 16
3 eng1 day3 b3 15
4 eng2 day1 b2 17
5 eng2 day2 b3 21
6 eng2 day3 b1 12
7 eng3 day1 b3 20
8 eng3 day2 b1 15
9 eng3 day3 b2 13
> mylm = lm(data ~ days + engines + burners, df)
> anova(mylm)
Analysis of Variance Table
Response: data
Df Sum Sq Mean Sq F value Pr(>F)
days 2 34.889 17.4444 22.429 0.04268 *
engines 2 1.556 0.7778 1.000 0.50000
burners 2 30.889 15.4444 19.857 0.04795 *
Residuals 2 1.556 0.7778
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```