Project I

Requirement:

- Each group works on the assigned topic.

- In each report, there must student names and IDs on the cover page, the table of content, and the questions.

- R-Studio must be used to analyze the data set and the codes must be inside framed environments. Detailed explanations must be provided to receive full credit.

Project I - Topic 1

1. A study was conducted in one city to identify how people use motorcycles, bicycles, and buses. The records are given in the following table

	Motorcycles	Bicycles	Buses
Female	25	125	100
Male	75	205	120

   At the significance level of 5%, can we conclude that there is a difference in the use of means of transport to work in the two groups of men and women?

2. A sociological survey was conducted in 5 cities A, B, C, D, E asking the respondents about their level of satisfaction with the city where they live. The results are as follows:

City	Levels of satisfaction		
	Very Satisfied	Relatively Satisfied	Very Satisfied
A	220	121	65
B	130	207	75
C	84	54	24
D	156	95	43
E	122	164	73

   At the significance level $\alpha = 3\%$, test whether the level of life satisfaction is equally distributed in the 5 cities above.

3. The following table shows the number of daily newspapers sold in 5 urban districts:

Day	Districts				
	District 1	District 2	District 3	District 4	District 5
Monday	22	18	22	18	18
Tuesday	21	18	22	18	19
Wednesday	25	25	25	19	20
Thursday	24	24	18	20	22
Friday	28	19	15	22	25
Saturday	30	22	28	25	25

   Is there any significant difference in the amount of newspapers sold in the 5 districts at $\alpha = 2\%$? Is the amount of newspapers sold affected by days of week?

4. The following table shows the rental rates in 5 cities

City A	900	1200	850	1320	1400	1150	975	
City B	625	640	775	1000	690	550	840	750
City C	415	400	420	560	780	620	800	390
City D	410	310	320	280	500	385	440	
Cities E	340	425	275	210	575	360		

   At the significance level 5%, is there any significant difference in rental rates in the five cities mentioned above?

Project I - Topic 2

1. Health examination of workers in five building of a battery factory shows the following records for their blood lead levels

Observation	Factor level				
	F1	F2	F3	F4	F5
1	0.25	0.22	0.25	0.31	0.22
2	0.28	0.25	0.26	0.33	0.28
3	0.32	0.24	0.28	0.30	0.28
4	0.22	0.28	0.25	0.29	0.25
5	0.22	0.31	0.22	0.25	0.30
6		0.21	0.28		
7		0.22	0.31		

Compare the blood lead levels among workers in the above factory at the significance level $\alpha = 3\%$.

2. The following table shows the income of the skilled workers in Sweden in 1930

Age group	Income levels					
	0-1	1-2	2-3	3-4	4-6	> 6
40-50	71	430	1072	1609	1178	158
50-60	54	324	894	1202	903	112

Is there any difference in the income of skilled workers between these two age groups at the significance level $\alpha = 5\%$?

3. The number of students arriving late at five high schools on different days of week are given the following table:

Days of week	High School			
	A	B	C	D
Monday	5	4	5	7
Tuesday	4	5	3	2
Wednesday	4	3	4	5
Thursday	4	4	3	2

Is there any significant difference in the number of late arrivals among different days of the week at the significance level $\alpha = 1\%$?

4. In a scientific experiment, the thickness of the nickel coating obtained from various types of plating tanks is measured.

Thickness of nickel coating	Plating tank		
	A	B	C
4-8	32	51	68
8-12	123	108	80
12-16	10	26	26
16-20	41	24	28
20-24	19	20	28

At the significance level of $\alpha = 0.05$, test the hypothesis: the coating thickness does not depend on the type of plating tank used.

Project I - Topic 3

1. Trachoma has 4 stages T1, T2, T3 and T4. The results of trachoma examination in 3 provinces A, B, C are given in the following table:

	Stages			
	T1	T2	T3	T4
A	47	189	807	1768
B	53	746	1387	946
C	16	228	438	115

   At the significance level of 1%, determine whether the distribution of the stages of trachoma is different in the three provinces.

2. The following table shows data on the number of cancer deaths in the United States, Japan, and the UK. Cancers are classified by the type of tissue in which the cancer originates.

Types of cancer	Countries		
	US	Japan	UK
Colorectal cancer	11	5	5
Breast cancer	15	3	7
Stomach cancer	3	22	3
Other types of	41	30	15

   Compare the cancer death rate in the three countries at the significance level of $\alpha = 1\%$

3. The sales of 4 stores of a company (million/month) are give in the following table:

Month	Stores			
	1	2	3	4
1	12.3	14.2	15.6	17.2
2	12.6	12.4	17.2	15.8
3	11.6	11.5	18.2	12.2
4	15.2	11.6	12.5	
5	18.6		11.8	
6	17.1			

   At the significance level of $\alpha = 5\%$, compare the sales of these stores.

4. At the level of significance $\alpha = 5\%$, compare the business performance of some industries in the four urban districts on the basis of the sales of some stores given in the following table.

Industries	Districts			
	1	2	3	4
Refrigeration	2.5, 2.7, 2.0, 3.0	13.1, 3.5, 2.7	2.0, 2.4	5.0, 5.4
Construction Materials	0.6, 10.4	15.0	9.5, 9.3, 9.1	19.5, 17.5
Computer services	1.2, 1.0, 9.8, 1.8	2.0, 2.2, 1.8	1.2, 1.3, 1.2	5.0, 4.8, 5.2

Project I - Topic 4

1. In order to compare advertising costs in four different newspapers (with the same advertising conditions), a sample of 7 advertising articles was collected from each newspaper and the following results (in thousand VND).

Newspaper A	57	65	50	45	70	62	48
Newspaper B	72	81	64	55	90	38	75
Newspaper C	35	42	58	59	46	60	61
Newspaper D	73	85	92	68	82	94	66

   Find the p-value to determine if there is any significant difference in advertising costs among these newspapers.

2. The following table gives the counts of hair color of a random sample of people.

Hair color	Male	Female
Black	56	32
Red	37	66
Brown	84	90
Yellow	19	38

   At the significance level of 3%, determine whether hair color and gender are correlated.

3. In oder to check whether the primary occupations and secondary occupations affect the average income of households, the following record of average incomes is collected.

Primary occupations	Secondary occupations			
	(1)	(2)	(3)	(4)
Rice cultivation	3.5	7.4	8.0	3.5
Fruit farming	5.6	4.1	6.1	9.6
Breeding	4.1	2.5	1.8	2.1
Services	7.2	3.2	2.2	1.5

   Draw a conclusion at the significance level of 3%.

4. The saponin content (mg) of medicinal plant of the same type that were harvested at different times of seasons in three regions is given the following table.

Seasons	Time	Regions		
		South	Central	North
Dry season	Early	2.4	2.1	3.2
	Mid	2.4	2.2	3.2
	Late	2.5	2.2	3.4
Rainy	Early	2.5	2.2	3.4
	Mid	2.5	2.3	3.5
	Late	2.6	2.3	3.5

   Is there any significant difference in the saponin content among the seasons and the regions? Do the two factors season and region interact? Use the significance level of 5%.

Project I - Topic 5

1. A dairy farm raised three breeds of dairy cows A, B, C. The amount of milk of the cows is given in the following table:

Type of cows	The amount of milk		
	Little	Medium	Much
A	92	37	46
B	53	15	19
C	75	19	12

At the significance level of $\alpha = 0.05$, can we conclude that the amount of milk produced by these types of cows is different?

2. At the significance of $\alpha = 5\%$, compare the business performance of some industries in the four urban districts on the basis of the sales of some stores given in the following table.

Industries	Districts			
	1	2	3	4
Refrigeration	2.5, 2.7, 2.0, 3.0	13.1, 3.5, 2.7	2.0, 2.4	5.0, 5.4
Construction Materials	0.6, 10.4	15.0	9.5, 9.3, 9.1	19.5, 17.5
Computer services	1.2, 1.0, 9.8, 1.8	2.0, 2.2, 1.8	1.2, 1.3, 1.2	5.0, 4.8, 5.2

3. A group of US businessmen is categorized by their annual incomes and their ages. The results are as follows.

Age range	Income levels		
	Under $100000	$100000 - $400000	Over $400000
Under 40	6	9	5
40-54	18	19	8
Over 54	11	12	17

At the 1% significance level, are the age and the income level related or not?

4. The following table shows the number of daily newspapers sold in 5 urban districts:

Day	Districts				
	District 1	District 2	District 3	District 4	District 5
Monday	22	18	22	18	18
Tuesday	21	18	22	18	19
Wednesday	25	25	25	19	20
Thursday	24	24	18	20	22
Friday	28	19	15	22	25
Saturday	30	22	28	25	25

Is there any significant difference in the amount of newspapers sold in the 5 districts at $\alpha = 1\%$? Is the amount of newspapers sold affected by the days of week?

Project I - Topic 6

1. Using linear regression, model the following data

X	50	130	170	270	90	210	50	130	270	240	170	210	90	210	90	240	50	240
Y	15	115	215	335	95	295	55	155	295	315	175	275	75	255	115	35	275	315

2. One researcher wanted to examine the response times of men and women to different types of signals. The subjects were asked to press the ENTER button on the computer keyboard as soon as they recognized the signal. The duration (measured in seconds) between the time the signal was emitted and the time the object hitting the button was recorded. Here are the results for 15 men and 15 women.

Gender	Sound	Light	Pulse
	10.0	6.0	9.1
	7.2	3.7	5.8
Male	6.8	5.1	6.0
	6.0	4.0	4.0
	5.0	3.2	5.1
	10.5	6.6	7.3
	8.8	4.9	6.1
Female	9.2	2.5	5.2
	8.1	4.2	2.5
	13.4	1.8	3.9

Draw a conclusion at the significance level of $\alpha = 5\%$. Do the factors gender and signal interact?

3. To study whether the size of a company affects the advertising effectiveness, a survey of 356 customers's opinion was collected and the following table was obtained.

Company size category	Advertising effectiveness		
	High	Moderate	Low
Small	20	52	32
Medium	53	47	28
Large	67	32	25

At the significance of $\alpha = 0.1$, is there enough evidence to conclude that that the company size affects the advertising effectiveness?

4. The number of students arriving late at five high schools on different days of week are given the following table:

Days of week	High School			
	A	B	C	D
Monday	5	4	5	7
Tuesday	4	5	3	2
Wednesday	4	3	4	5
Thursday	4	4	3	2

Is there any significant difference in the number of late arrivals among different days of the week at the significance level $\alpha = 5\%$?

Project I - Topic 7

1. An agronomist conducts an experiment to test the effects of three types of fertilizer on tomato plants. The following table shows the number of tomatoes produced by the plants.

Types of fertilizer		
A	B	C
24	21	16
18	26	22
27	32	19
28	25	17

At the significance level of $\alpha = 5\%$, compare the average number of tomatoes produced when these types of fertilizer are used.

2. The following table gives the counts of hair color of a random sample of people.

Hair color	Male	Female
Black	56	32
Red	37	66
Brown	84	90
Yellow	19	38

At the significance level of 3%, determine whether hair color and gender are correlated.

3. A survey was conducted for various social classes to determine how their satisfaction changes over last year. The data follow.

Opinion	Class		
	Workers	Farmers	Intellectuals
Increased	100	300	20
Unchanged	200	400	30
Decreasing	50	80	5
No comment	30	70	5

At the significance level of $\alpha = 2\%$, is there a significant difference in the levels of satisfaction among the social classes?

4. In oder to check whether the primary occupations and secondary occupations affect the average income of households, the following record of average incomes is collected.

Primary occupations	Secondary occupations			
	(1)	(2)	(3)	(4)
Rice cultivation	3.5	7.4	8.0	3.5
Fruit farming	5.6	4.1	6.1	9.6
Breeding	4.1	2.5	1.8	2.1
Services	7.2	3.2	2.2	1.5

Draw a conclusion at the significance level of 2%.

Project I - Topic 8

1. The following table shows the the diameter X and the height Y of 20 trees respectively.

X	2.3	2.5	2.6	3.1	3.4	3.7	7.3	3.9	4	4.1	4.1	4.2	4.4	4.7	5.1	5.5	5.8	6.2	6.9	6.9
Y	7	8	4	4	6	6	14	12	8	5	7	8	7	9	10	13	7	11	11	16

Fit the linear regression model and check all assumptions.

2. A company wants target 3 overseas markets. To compare the potential market share of the company and that of some competitors in the markets, their products were sold to potential customers. The following results were obtained.

	Markets		
	A	B	C
The company	55	38	24
Competitor 1	28	30	21
Competitor 2	20	18	31
Other competitors	47	64	74

Draw a conclusion at the significance level of 5%.

3. The river water level was measured at several locations in a province on the same day and the data follow.

Measurement time	Measurement location			
	F1	F2	F3	F4
1	5.5	4.9	4.6	4.5
2	4.6	5.1	4.8	6.2
3	5.8	6.5	5.8	4.8
4	5.9	5.4	5.1	4.8
5	6.0	6.1	6.2	6.5
6	6.7		7.1	6.8
7	7.2			

At the significance level $\alpha = 2\%$, are the average river water levels each day at these locations significantly different?

4. The following table show the average cost, measured as 1000 VND expended for each service at each city.

Cities	Type of service		
	I	II	III
I	61	52	69
II	58	51	61
III	68	64	79

At the significance level $\alpha = 5\%$, compare the costs for three types of services in three cities.

Project I - Topic 9

1. A sociological survey was conducted in 5 cities A, B, C, D, E asking the respondents about their level of satisfaction with the city where they live. The results are as follows:

City	Levels of satisfaction		
	Very Satisfied	Relatively Satisfied	Very Satisfied
A	220	121	63
B	130	207	75
C	84	54	24
D	156	95	43
E	122	164	73

At the significance level $\alpha = 2\%$, test whether the level of life satisfaction is equally distributed in the 5 cities above.

2. The following table gives the counts of hair color of a random sample of people.

Gender	Hair color			
	Black	Red	Yellow	Brown
Male	56	37	84	19
Female	32	66	90	38

At the significance level of 3%, determine whether hair color and gender are correlated.

3. In oder to check whether the primary occupations and secondary occupations affect the average income of households, the following record of average incomes is collected.

Primary occupations	Secondary occupations			
	(1)	(2)	(3)	(4)
Rice cultivation	3.5	7.4	8.0	3.5
Fruit farming	5.6	4.1	6.1	9.6
Breeding	4.1	2.5	1.8	2.1
Services	7.2	3.2	2.2	1.5

Draw a conclusion at the significance level of 1%.

4. The following table shows the world records for running 1 mile in the 20th century:

X	1911	1913	1915	1923	1931	1933	1934	1937	1942	1942
Y	4 : 15.4	4 : 14.6	4 : 12.6	4 : 10.4	4 : 09.2	4 : 07.6	4 : 06.8	4 : 06.4	4 : 06.2	4 : 06.2
X	1942	1943	1944	1945	1954	1954	1957	1958	1962	1964
Y	4 : 04.6	4 : 02.6	4 : 01.6	4 : 01.4	3 : 59.4	3 : 58	3 : 57.2	3 : 54.5	3 : 54.4	3 : 54.1
X	1965	1966	1967	1975	1975	1979	1980	1981	1981	1985
Y	3 : 53.6	3 : 51.3	3 : 51.1	3 : 51	3 : 49.4	3 : 49	3 : 48.9	3 : 48.8	3 : 48.7	3 : 46.5

(a) Calculate the correlation coefficient between X (in years) and Y (in seconds).

(b) Determine the regression line.

(c) Check the assumptions of linear regression models.

(Note: 3:48.7 = 3 minutes 48.7 seconds).

Project I - Topic 10

1. The following table records a sample of 11 observations (x_i, y_i).

X	0.9	1.22	1.32	0.77	1.3	1.2	1.32	0.95	1.45	1.3
Y	−0.3	0.1	0.7	−0.28	−0.25	0.02	0.37	−0.70	0.55	0.35

   Fit the linear regression model and check all assumptions.

2. The following table shows the income of the skilled workers in Sweden in 1930

Age group	Income levels					
	0-1	1-2	2-3	3-4	4-6	> 6
40-50	71	430	1072	1609	1178	158
50-60	54	324	894	1202	903	112

   Is there any difference in the income of skilled workers between these two age groups at the significance level $\alpha = 2\%$?

3. A group of US businessmen is categorized by their annual incomes and their ages. The results are as follows.

Age range	Income levels		
	Under $100000	$100000 - $400000	Over $400000
Under 40	6	9	5
40-54	18	19	8
Over 54	11	12	17

   At the 5% significance level, are the age and the income level related or not?

4. At the significance of $\alpha = 5\%$, compare the business performance of some industries in the four urban districts on the basis of the sales of some stores given in the following table.

Industries	Districts			
	1	2	3	4
Refrigeration	2.5, 2.7, 2.0, 3.0	13.1, 3.5, 2.7	2.0, 2.4	5.0, 5.4
Construction Materials	0.6, 10.4	15.0	9.5, 9.3, 9.1	19.5, 17.5
Computer services	1.2, 1.0, 9.8, 1.8	2.0, 2.2, 1.8	1.2, 1.3, 1.2	5.0, 4.8, 5.2