

## Data Analytics (CS40003)

### Practice Set VII

#### (Topic: Correlation Analysis)

#### I. Concept Questions

1. An experiment was conducted to test the efficacy of chloromycetin in checking typhoid. In a certain hospital chloromycetin was given to 285 out of the 392 patients suffering from typhoid. The number of typhoid cases were as follows:

	Typhoid	No Typhoid	Total
Chloromycetin	35	250	285
No chloromycetin	50	57	107
Total	85	307	392

With the help of  $\chi^2$ , test the effectiveness of chloromycetin in checking typhoid.

(The  $\chi^2$  value at 5 per cent level of significance for one degree of freedom is 3.841).

2. Result of throwing die were recorded as follows:

Number falling upwards	1	2	3	4	5	6
Frequency	27	33	31	29	30	24

Is the die unbiased? Answer on the basis of Chi-square test?

- 3.

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5

No. of families	14	56	110	88	40	12
-----------------	----	----	-----	----	----	----

Is this distribution consistent with the hypothesis that male and female births are equally probable?

Apply Chi-square test.

- 4.
- 1000 babies were born during a certain week in a city of which 600 were boys and 400 girls. Use  $\chi^2$  test to examine the correctness of the hypothesis that the sex-ratio is 1 : 1 in newly born babies.
  - The percentage of smokers in a certain city was 90. A random sample of 100 persons was selected in which 85 persons were found to be smokers. Is the sample proportion significantly different from the proportion of smokers in the city? Answer on the basis of Chi-square test.

5. A group of 150 College students were asked to indicate their most liked film star from among six different well known film actors viz., A, B, C, D, E and F in order to ascertain their relative popularity. The observed frequency data were as follows:

Actors	A	B	C	D	E	F	Total
Frequencies	24	20	32	25	28	21	150

Test at 5 per cent whether all actors are equally popular.

6. The owner of a shoe store wants to know if shoe size and weight are correlated in adult males. She measures the shoe size and asks the weight of 14 consecutive customers. They are as follows:  
 size 9, 176 lbs; size 7.5, 141 lbs; size 10, 185 lbs; size 12, 202 lbs; size 9.5; 174 lbs; size 10, 150 lbs; size 10, 193 lbs; size 10.5, 237 lbs; size 13, 248 lbs; size 8, 159 lbs; size 8.5, 136 lbs; size 9.5, 174 lbs; size 9, 172 lbs; size 11, 183 lbs.

- The Pearson correlation on this data is  $r =$ 
  - .61
  - .71
  - .81
  - .91

- b. The  $r$  value can be reported significant (lowest possible) at  $p <$
- i. .05
  - ii. .05
  - iii. .01
  - iv. .001
- c. The  $df$  for this problem are
- i. 14
  - ii. 12
  - iii. 10
  - iv. 8
- d. The coefficient of determination for this problem is
- i. .65
  - ii. .81
  - iii. .809
  - iv. .99
- e. The percentage of variance in shoe size that can be accounted by weight is
- i. 65
  - ii. 81
  - iii. 267
  - iv. 90
  - v. 99

7.

Country	Alc. Consumption	Death Rate from Cirrhosis
France	24.7	46.1
Italy	15.2	23.6
Germany	12.3	23.7
Australia	10.9	7
Belgium	10.8	12.3
USA	9.9	14.2

Canada	8.3	7.4
England	7.2	3.0
Sweden	6.6	7.2
Japan	5.8	10.6
Netherland	5.7	3.7
Ireland	5.6	3.4
Norway	4.2	4.3
Finland	3.9	3.6
Isreal	3.1	5.4
total	134.2	175.5

- a. Draw a Scatter Diagram to show the association, if any, between these two variables; can you draw any conclusion/observation without doing any calculation?
  - b. Calculate the Coefficient of Correlation
8. A tobacco company statistician wishes to know whether heavy smoking is related to longevity. From a sample of recently deceased smokers, the number of cigarettes (estimated on a per day for their last five years after visits with their surviving relatives) is paired with the number of years that they lived.

Subject	Cigarettes	Years Lived
1	25	63
2	35	68
3	10	72
4	40	62
5	85	65
6	75	46
7	60	51
8	45	60
9	50	55

- a) Calculate the Pearson’s correlation coefficient of the above sample data.
- b) Test the significance of the calculation using the Pearson’s Product Moment Correlation (PPMC) table.
- c) Test the hypothesis “Cigarette smoking is injurious to health”. Hint: use t-Distribution Table.

9. Find the Pearson’s correlation coefficient for the following pairs of observations from a population.

<b>Age</b>	43	21	25	42	57	59
<b>Glucose level</b>	99	65	79	75	87	81

- a) Calculate the Pearson Product Moment Correlation Coefficient.
- b) Calculate the Coefficient Determination. What is its significance in this case?
- c) Consult the PPMC Table to test the hypothesis that “Sugar level does not increases with the age of a person”.

10. The owner of a shoe store wants to know if the shoe size and weights are correlated in adult males. He observes the 14 consecutive customers and his observation is as follows.

Observation	Size	Wight	<b>Legends</b> Size XS : Extra Small SS : Very Small S : Small M : Medium L : Large LL : Very Large XL : Extra Large  Weight VL : Very Light L : Light H : Heavy VH: Very Heavy
1	M	H	
2	XS	L	
3	L	H	
4	LL	VH	
5	M	H	
6	L	H	
7	L	VH	
8	L	VH	
9	XL	VH	
10	SS	H	
11	S	VL	
12	M	H	
13	M	H	
14	L	H	