Answer to the question no 1 a

A regression model is a statistical model that estimates the relationship between one dependent variable and one or more independent variables using a line (or a plane in the case of two or more independent variables).

Definition

Multiple Regression Equation

A linear relationship between a dependent variable y and two or more independent variables $(x_1, x_2, x_3 \ldots, x_k)$

Model:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + e$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_k x_k$$

(General form of the estimated multiple regression equation)

n = sample size

k = number of independent variables

$\hat{y}$ = predicted value of the dependent variable y

$x_1, x_2, x_3 \ldots, x_k$ are the independent variables

Multiple Regression

Definition

Multiple Regression Equation

A linear relationship between a dependent variable y and two or more independent variables ($x_1, x_2, x_3 \ldots, x_k$)

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$

Notation

β_0 = the y-intercept, or the value of y when all of the predictor variables are 0

b_0 = estimate of β_0 based on the sample data

$\beta_1, \beta_2, \beta_3 \ldots, \beta_k$ are the population coefficients of the independent variables $x_1, x_2, x_3 \ldots, x_k$

$b_1, b_2, b_3 \ldots, b_k$ are the sample estimate the coefficients $\beta_1, \beta_2, \beta_3 \ldots, \beta_k$

Answer to the question no 1 b

COEFFICIENT OF DETERMINATION

The convenient way of interpreting the value of correlation coefficient is to use of square of coefficient of correlation which is called Coefficient of Determination.

The Coefficient of Determination = r^2.

Suppose: $r = 0.9$, $r^2 = 0.81$ this would mean that 81% of the variation in the dependent variable has been explained by the independent variable.

The maximum value of r^2 is 1 because it is possible to explain all of the variation in y but it is not possible to explain more than all of it.

Coefficient of Determination = Explained variation / Total variation

Interpret the values of r

$r= +1$, indicates a perfect positive relationship between x and y.

$r=-1$, indicates a perfect negative relationship between x and y.

$r=0$, means there is no linear relationship between x and y. In this case the two variables are linearly independent.

$0.5 < r < 1$, indicates a strongly positive relationship between x and y.

$0 < r < 0.5$, indicates a simple positive relationship between x and y.

$-0.5 < r < 0$ indicates a negative relationship between x and y.

$-1 < r < -0.5$ indicates a strongly negative relationship between x and y.

Answer to the question no 1 c

Same as 1b spring 22

Answer to the question no 2 a

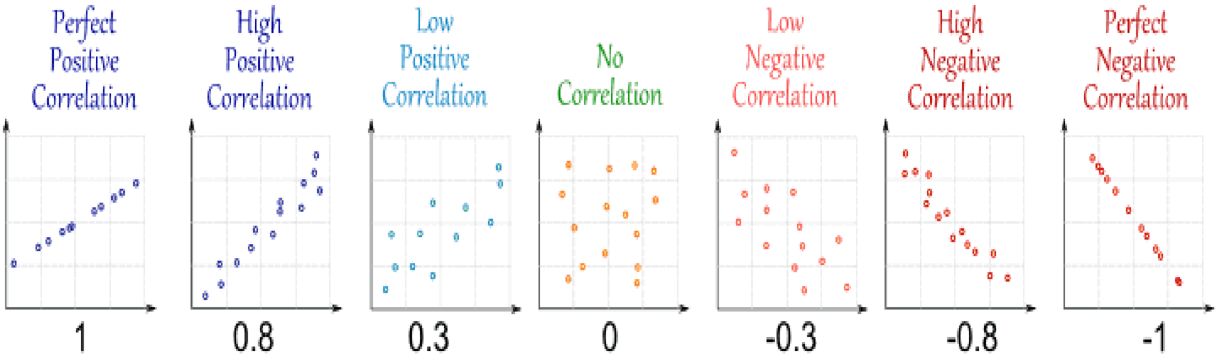
Scatter diagram method

The diagrammatic way of representing bivariate data is called scatter diagram.

Suppose, (x_1,y_1), (x_2,y_2)……….(x_n,y_n) are n pairs of observations. If the values of the variables x and y are plotted along the x-axis and y-axis respectively in the xy-plane, the diagram of dots obtained is known as a scatter diagram.

Scatter diagrams for different values of r are as follows:

- **Scatter diagrams for different values of r are as follows:**



Answer to the question no 2 b

Application Problem-1: Obtain the rank correlation co-efficient for the following data:

A:	80	75	90	70	65	60
B:	65	70	60	75	85	80

Solution: Calculating Table

A	B	Rank of A (x)	Rank of B (y)	d = x-y	d^2
80	65	2	5	-3	9
75	70	3	4	-1	1
90	60	1	6	-5	25
70	75	4	3	1	1
65	85	5	1	4	16
60	80	6	2	4	16
Total					

$$R = \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 4}{6(6^2-1)}$$

$$= -0.94$$

Conclusion: There exist strongly negative relationship between A and B.

Answer to the question no 3b

Application problem: The distribution of number of stores according to size in 3 areas is given in following table:

Area	Store size		
	Large(L)	Medium(M)	Small(S)
A	30	45	75
B	150	125	275
C	20	130	150

Find the probabilities (i) P(M); (ii) P[BM] (iii) P(B U L) and (iv) P(A | M) (v) Are the events A and L independent?

Solution:

Area	Store size			
	Large(L)	Medium(M)	Small(S)	**Total**
A	30	45	75	**150**
B	150	125	275	**550**
C	20	130	150	**300**
Total	**200**	**300**	**500**	**1000**

Here total number of stores is 1000.That is N(S)=1000.

(i) Here M is the event of medium size store. Then N(M)= 300

Therefore $P[M] = \frac{N(M)}{N(S)} = \frac{300}{1000} = 0.300$

(ii) $B \cap M = BM$.

$P[BM] = \frac{N(BM)}{N(S)} = \frac{125}{1000} = 0.125$

(iii) P(B U L)=P[B]+P[L]-P[BL]

$$= \frac{N(B)}{N(S)} + \frac{N(L)}{N(S)} \quad \frac{N(BL)}{N(S)}$$

$$= \frac{550}{1000} + \frac{200}{1000} \quad \frac{150}{1000}$$

=0.60

(iv) $P(A \mid M) = \frac{p(AM)}{P(M)} = \frac{\frac{N(AM)}{N(S)}}{\frac{N(M)}{N(S)}} = \frac{N(AM)}{N(S)} = \frac{45}{300} = 0.15$

(v) Are the events A and L independent?

Solution: The events A and L are independent if P [AL] = P [A] P [L].

Here, P [AL] = $\frac{N(AL)}{N(S)} = \frac{30}{1000}$

And P [A] P [L]= $\frac{N(A)}{N(S)} \times \frac{N(L)}{N(S)} = \frac{150}{1000} \times \frac{200}{1000} = \frac{30}{1000} = P[AL]$

Hence the events A and L are independent.

Answer to the question no 4 a

Random variable with example.
A random variable is a real valued function whose values are determined with the outcomes of a random experiment. It is usually denoted by X,Y,Z etc and the value of the random variable denoted by x,y,z.

MATHEMATICAL EXPECTATION

If X is a discrete or continuous random variable with probability function or probability density function f(x).then the mathematical expectation of X is usually denoted by E[X] or μ and defined by

$$\mu = E[X] = \sum x f(x) \text{ ; If X is a discrete random variable.}$$

$$= \int_{\infty}^{\infty} x f(x) dx \text{ ; If X is a continuous random variable.}$$

Probability function and probability density function.

Probability function: A function f(x) of a discrete random variable X is called a probability function if it satisfies the following two conditions:

(i) $f(x) \geq 0$
(ii) $\sum f(x) = 1$

Probability density function: A function f(x) of a continuous random variable X is called a probability function if it satisfies the following two conditions:

(i) $f(x) \geq 0$

(ii) $\int_{\infty}^{\infty} f(x) dx = 1$

Answer to the question no 4 b

Same as 3(b) spring 22

Answer to the question no 5 a

standard normal variate

Standard normal variate: A variate is called standard normal variate if its mean 0 and variance 1 respectively. It is denoted by Z where,

$$Z = \frac{X - \mu}{\sigma}$$

Properties of normal distribution:

1. Normal distributions are symmetric around their mean.
2. The mean, median, and mode of a normal distribution are equal.
3. The area under the normal curve is equal to 1.0.
4. Normal distributions are denser in the center and less dense in the tails.
5. Normal distributions are defined by two parameters, the mean (μ) and the standard deviation (σ).
6. 68% of the area of a normal distribution is within one standard deviation of the mean.
7. Approximately 95% of the area of a normal distribution is within two standard deviations of the mean.

Answer to the question no 6 a

Procedure of test of hypothesis

i. Set of hypothesis

ii. Set up of suitable significance level

iii. Determination of suitable test statistic

iv. Determine the critical region

v. Doing computation

vi. Making decision

3.Chi-square ( )test

a. To test the significance of a specified population variance

b. To test the goodness of fit of a distribution

c. To test the independence of attributes

d. To test the equality of several correlation coefficients

e. To test the homogeneity of several tests

f. To test the homogeneity of several population variance

Answer to the question no 6 b

Problem, Chapter-8: Two hundred Engineers were interviewed and classified according to their results and job satisfaction. The distribution of graduates by results and job satisfaction are given in the following contingency table:

Results	Job satisfaction	
	Yes	No
Excellent	20	70
Good	45	65

Compute the value of Chi-square for the above data.

Solution: Computation table

Results	Job satisfaction		Total
	Yes	No	
Excellent	20 (O_{11})	70(O_{12})	90
Good	45(O_{21})	65(O_{22})	110
Total	65	135	200

We know, $$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Here,

$$E_{11} = \frac{90 \times 65}{200} = 29.25,$$

$$E_{12} = \frac{90 \times 135}{200} = 60.75,$$

$$E_{21} = \frac{110 \times 65}{200} = 35.75,$$

$$E_{22} = \frac{110 \times 135}{200} = 74.25$$

Therefore, $$\chi^2 = \frac{(20-29.25)^2}{29.25} + \frac{(70-60.75)^2}{60.75} + \frac{(45-33.75)^2}{33.75} + \frac{(65-74.25)^2}{74.25}$$

$$= 2.93+1.41+3.75+1.15 = 9.24$$

Answer to the question no 7 a

Mutually exclusive events: Two events are said to be mutually exclusive if they have no common points. If A and B are two mutually exclusive events, then AB $= \emptyset$.

Application Problem: Suppose A and B are two mutually exclusive events with P[A]=.35 and P[B]=.15. Find (i) P[A$\cup$B] (ii) P[$\overline{A}$] (iii) P[A$\cap$B] (iv) P[$\overline{A} \cup \overline{B}$]

Solution:

(i) According to axiom 3.

P[AUB]= P[A]+P[B]= 0.35+0.15=0.50

(ii) P[$\overline{A}$]=1-0.35=0.65

(iii) P[A$\cap$B]=0; Since A and B are mutually exclusive.

(iv) According to De Morgan's law, $\overline{A} \cup \overline{B} = \left(\overline{A \cap B}\right)$

$P\,(\overline{A} \cup \overline{B}) = P\left(\overline{A \cap B}\right)$=1- P[A$\cap$B]=1-0=1

Answer to the question no 7 b

Application problem: A discrete random variable X has the following probability function:

Values of X:x	0	1	2	3	4
f (x)	0.12	0.18	k	0.30	0.16

(i) Find the value of k, Compute (ii) P[X>3] ; (iii) P[1< X < 4] ; and (iv) P[X<1].

(i) Since, $\sum f(x) = 1$

Or, (0.76+k) = 1

Or, k =1-0.76

Or, k = 0.24

(ii) P[X>3] = P[X=4] = 0.16

(iii) P[1< X < 4]= P[X=2]+ P[X=3]

= k+0.30

= 0.24+0.30 = 0.54

(iv) P[X<1] = P[X=0]

= 0.12

MATHEMATICAL EXPECTATION

If X is a discrete or continuous random variable with probability function or probability density function f(x).then the mathematical expectation of X is usually denoted by E[X] or μ and defined by

μ = E[X] $= \sum x f(x)$; If X is a discrete random variable.

$= \int_{-\infty}^{\infty} x f(x) dx$; If X is a continuous random variable.

Properties of variance of a random variable:

(i) If b is a constant then V[b] = 0

(ii) If X is a random variable, then V [aX+b] = a^2 V[X] , Where a and b constant.

(iii) If X is a random variable with expectation E[X], then
$V(X) = E[X-E(X)]^2 = E(X^2) - [E(X)]^2$

(iv) If X and Y are random variables then V[X+Y] = V[X] + V[Y]

(v) If X and Y are random variables then V[X-Y] = V[X] + V[Y]