DIGITAL ASSIGNMENT -2

Probability And Statistics

PMAT501L

E2-TE2 Slot

Name - Barsha Routh
Reg no - 24MCA0164

Q1. Find the equation of multiple regression plane of Z on X and Y from foll. data.

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

Solution

$$Z = a + bX + cY$$

X	Y	Z	X^2	Y^2	XY	ZX	ZY
30	11	110	900	121	330	3300	1210
40	10	80	1600	100	400	3200	800
20	7	70	400	49	140	1400	490
50	15	120	2500	225	750	6000	1800
60	19	150	3600	361	1140	9000	2850
40	12	90	1600	144	480	3600	1080
20	8	70	400	64	160	1400	560
60	14	120	3600	196	840	7200	1680
320	96	810	14600	1260	4240	35100	10470

$n = 8$

For a, b, c

$$\Sigma Z = na + b\Sigma X + c\Sigma Y$$

$$\Sigma ZX = a\Sigma X + b\Sigma X^2 + c\Sigma XY$$

$$\Sigma ZY = a\Sigma Y + b\Sigma XY + c\Sigma Y^2$$

$$810 = 8a + 320b + 96c$$

$$35100 = 320a + 14600b + 4240c$$

$$10470 = 96a + 4240b + 1260c$$

Barsha Routh, 24MCA0164

$a = \frac{2895}{172} \approx 16.83$

$b = \frac{-21}{86} \approx -0.244$

$c = \frac{675}{80} \approx 7.85$

$Z = 16.83 - 0.244X + 7.85Y$

Q2) It is known that probability of an item produced by a certain machine will be defective is 0.10. If the produced items are sent to the market in packets of 50, then find the no. of packets containing at least, exactly and atmost 5 defective items in a consignment of 1000 packets by using (i) Binomial Distribution (ii) Poisson Approximation to the Binomial Distribution.

Solution

$p = 0.10 \quad n = 50 \quad N = 100$

$P(X=k) = {}^nC_k\, p^k (1-p)^{n-k}$

Exactly 5, $P(X=5) = {}^{50}C_5 (0.10)^5 (0.90)^{45}$

$= 2118760 \times 0.00001 \times 0.04077$

≈ 0.1843

$0.1843 \times 1000 = 184.3$

Atmost 5, $P(X \leq 5) =$

$= P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= {}^5C_0 p^0 q^5 + {}^5C_1 p^1 q^4 + {}^5C_2 p^2 q^3 + {}^5C_3 p^3 q^2 + {}^5C_4 p^4 q^3 + {}^5C_5 p^5 q^0$

$= 0.0052 + 0.0286 + 0.0779 + 0.1386 + 0.1809 + 0.1849$

$P(X \leq 5) \Rightarrow 0.6161 \times 1000 = 616.1$

Barsha Routh,
24MLA0164

$P(X \geq 5) = 1 - P(X<5) = 1 - P(X \leq 4)$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)]$

$= 1 - [0.0045 + 0.0274 + 0.0822 + 0.1645 + 0.2468]$

$= 1 - 0.5254 = 0.4746$

For $N = 1000$, $0.4746 \times 1000 = 474.6$

<u>Poisson Approximation</u>

$\lambda = 5 \qquad N = 1000$

At least 5, $P(X \geq 5) = 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)]$

$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$P(X \geq 5) = 1 - [0.0067 + 0.0337 + 0.0842 + 0.1404 + 0.1755]$

$= 0.56$

$0.56 \times 1000 = 560$

<u>Exactly 5</u>, $P(X = 5)$

$P(X=5) = \frac{5^5 e^{-5}}{5!} = 0.1755$

$0.1755 \times 1000 = 175.5 \approx 176$

<u>At most 5</u>,

$= P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= 0.0067 + 0.0337 + 0.0842 + 0.1404 + 0.1755 = 0.4395$

$0.4395 \times 1000 = 439.5 \approx 440$

Barsha Routh, 24MCA0164

Q3) The finish times for marathon runners during a race are normally distributed with a mean of 200 minutes and SD of 50 mins.
(i) What is the Probability that runner will complete that marathon within 3 hrs.
(ii) What proportion of runners will complete the marathon between 3 hours and 4 hours?
(iii) Calculate the nearest min, the time by which the first 8% runners have completed the marathon.

Solution

$$Z = \frac{X-\mu}{\sigma} \quad \text{where, } X = 3, \ \mu = 200, \ \sigma = 50$$

(i)

$$Z = \frac{180-200}{50} = \frac{-20}{50} = -0.4$$

Using Z-table, the prob. at $Z = -0.4$, is 0.3446.

$$P(X \leq 180) = 0.3446$$

∴ This is the probability that runner will complete marathon within 3 hrs.

(ii)

$$Z = \frac{240-200}{50} = \frac{40}{50} = 0.8$$

From z table prob. at $z = 0.8$ is 0.7881.
And at, $Z = -0.4$ is 0.3446 (from (i))

So, $P(180 \leq X \leq 240) = P(Z=0.8) - P(Z=-0.4)$

$P(180 \leq X \leq 240) = 0.7881 - 0.3446 = 0.4435$

So, 44.35% of runners will complete the marathon between 3 and 4 hours.

(iii). From (ii) we know, probability that $X \leq 240$ mins is 0.7881.

So, $P(X \geq 240) = 1 - P(X \leq 240)$

$P(X \geq 240) = 1 - 0.7881 = 0.2119$

So, Prob. that runner will complete the marathon after 4 hrs is 0.2119 or 21.19%

(iv) First we will find $P(X \leq X_{8\%}) = 0.08$

From Z-table, Z score corresponding to a cumulative prob of 0.08 is -1.405 approx.

Now,

Converting z-score back to original,

$$X = \mu + Z \times \sigma$$

$$X = 200 + (-1.405) \times 50$$

$$X = 200 - 70.25 = 129.75 \text{ mins.}$$

So, first 8% runners will complete in 130 mins approx.

Q4) The student welfare office of a certain university polled random sample of 1000 male students and found that 720 were in favor of new grading system. At same time, 695 out of random sample of 900 female students were in favor of the new system. Do the results indicate a significant diff. in proportion of male & female students who favor the new grading system at 95% level of confidence?

Solution Male students

$n_1 = 1000$, $x_1 = 720$, $p_1 = \frac{720}{1000} = 0.72$

Female students

$n_2 = 900$, $x_2 = 695$, $p_2 = \frac{695}{900} = 0.7722$

Null Hypothesis; $H_0: p_1 = p_2$ (No difference)

Alternative Hypothesis; $H_a: p_1 \neq p_2$ (diff. in proportion)

$$z = \frac{(p_1 - p_2)}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

P or pooled proportion: $P = \frac{x_1 + x_2}{n_1 + n_2}$

$$= \frac{1415}{1900} \approx 0.7447$$

Barsha Routy
24MCA0164

Now, $z = \dfrac{0.72 - 0.7722}{\sqrt{0.7447 \times (1 - 0.7447) \times \left(\frac{1}{1000} + \frac{1}{900}\right)}}$

$z = \dfrac{0.72 - 0.7722}{0.02003} \approx -2.606$

for two tailed test at 95% confidence level, the critical z value is ± 1.96

So, -2.606, that falls outside the range $-1.96 \leq z \leq 1.96$

$\because$ z is less than -1.96,
we reject null hypothesis.

Q5. Write the detailed Report on Applications of Correlation & Regression Analysis or PD or Sampling Technique in Science, Engineering or Technology Oriented Problems for minimum 1 page.

Solution

CORRELATION AND REGRESSION

INTRODUCTION

Correlation and regression analysis are essential statistical tools widely used in science, engineering & technology. These technique help identify relationship between variables, enabling researchers and professionals to make informed decisions based on data.

- Correlation Analysis

It assesses the strength and direction of the relationship between two or more variably. The correlation coefficient, ranging from -1 to 1, quantifies this relationship, with values close to 1 indicating a strong positive correlation and values close to -1 indicating a strong negative correlation.

Barsha Routh, 24MCA 0164

Applications in Science

① Environmental Science :- Eg- to analyse relation between temperature and pollution level.

② Healthcare :- eg- to study relationship between lifestyle factors and health outcomes.

③ Agriculture :- to study relationship between crop yield and factors like rainfall, soil fertilizers, etc.

Regression Analysis

Regression analysis extends correlation analysis by modelling the relation between a dependent variable & one or more independent variables. It allows for prediction and the quantification of the impact of changes in independent variables on the dependent variable.

Application in Engineering:

(i) Structural Engineering → Vital in predicting the load balancing capacity of structures.

(ii) Quality Control → helps identify factors influencing product quality.

(iii) System Engineering → Aids in optimizing complex systems by modelling relationship between system variables.

Application in Technology :-

① ML And AI : Regression analysis forms foundation for ML algo. Techniques such as linear regression, logistic regression are commonly used for predictive modelling and classification tasks.

② Finance and Economics → Used to forecast stock prices and assess the impact of economic indicators on financial markets.

Barsha Routh, 24MCA0164

3. Information Technology → helps identify factors that influence software performance.

CONCLUSION

Correlation and regression analysis are powerful statistical tools with diverse applications across science, engineering and technology. Their ability to reveal relationships between variables and predict outcomes makes them invaluable for decision making and problem solving. Through effective application of correlation and regression analysis, researchers and professionals can harness data to inform practice, improve efficiencies and foster innovation.