Probability and Statistics
Csc. 103-2065

Group A

1. Show that the mean and variance of Poisson distribution are equal. Telephone calls enter a college switchboard on the average of two every 3 minutes. If one assumes an approximate Poisson Process What is the probability of three or more calls arriving in a 9- minute period?

Let $X \sim (\lambda)$ then proof of X is given as

$$P(X=x)=\begin{cases}\dfrac{e^{-\lambda}\lambda^x}{x!}, & \forall x=0,1,2,\dots,\infty\\ 0, & otherwise\end{cases}$$

For mean,

$$\begin{aligned}E(X) &= \textstyle\sum_x xP(x)\\ &= \textstyle\sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!}\\ &= e^{-\lambda}\textstyle\sum_{x=1}^{\infty}\frac{x\lambda^x}{x(x-1)!}\\ &= \lambda e^{-\lambda}\textstyle\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}\\ &= \lambda e^{-\lambda}\left[1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\cdots\right]\\ &= \lambda e^{-\lambda}e^{\lambda}\end{aligned}$$

$\therefore E(x)=\lambda$ ...(i)

For variance,

$$\begin{aligned}E(x^2) &= \textstyle\sum_x x^2P(x)\\ &= \textstyle\sum_{x=0}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!}\\ &= \textstyle\sum_{x=0}^{\infty}\{x(x-1)+x\}\frac{e^{-\lambda}\lambda^x}{x!}\\ &= e^{-\lambda}\left[\textstyle\sum_{x=2}^{\infty}x(x-1)\frac{\lambda^x}{x(x-1)(x-2)!}+\sum_{x=1}^{\infty}\frac{x\lambda^x}{x(x-1)!}\right]\\ &= e^{-\lambda}\left[\textstyle\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!}+\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!}\right]\\ &= e^{-\lambda}\left[\lambda^2\textstyle\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}+\lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}\right]\\ &= \lambda e^{-\lambda}\textstyle\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}\\ &= \lambda e^{-\lambda}\left[1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\cdots\right]\\ &= e^{-\lambda}\left[\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\cdots\right)+\lambda\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\cdots\right)\right]\\ &= e^{-\lambda}(\lambda^2e^{\lambda}+\lambda e^{\lambda})\end{aligned}$$

$\therefore E(x^2)=\lambda^2+\lambda$

Now, variance $V(x)=E(x^2)-[E(x)]^2=\lambda^2+\lambda-(\lambda)^2$

$\therefore V(x)=\lambda$ ...(ii)

Equation (i) and (ii) represent the expression for mean and variance of Poission distribution are equal.

Solution

Let 'X' be the random variable which denotes the number of marketing phone calls. Then,

proof of $X(PX = x) = \begin{cases} \frac{e^{-\lambda}\lambda^{x}}{x!}, & \forall x = 0,1,2,\dots,x \\ 0, & otherwise \end{cases}$

Here, we are given,

Mean $(\lambda) = 9$

Hence, the required probability= $P(x = 3) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-3}3^{9}}{9!} = 0.0027$

2. Differentiate between confidence level and level of significance. A manufacturer of flashlight batteries took a sample of 10 batteries from a day's production and used them continuously until they failed to work. The life as measured by the number of hours until failure was:

34 42 31 54 26 45 63 51 26 56

At the 0.05 level of significance, is there evidence that the mean life of the batteries is different from 10 hours?

Solution

Confidence level: In statistics, a confidence interval (C.I.) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate. It is an observed interval (i.e. it is calculated from the observation) in principle different from sample to sample, that frequently includes the parameter of interest if experiment is repeated.

Level of significance: A fixed number, most often 0.05 is referred to as significance level or level of significance. Such a number may be used either in first sense, as a cut off mark for p-values (p-value is calculated from the data), or in the second sense as a desired parameter in the test design (α depends only on the test design, and is not calculated from observed data).

Now, given,

n=10, μ=10

Calculation of $\bar{X}$ and s:

X	$d = X - 26$	d^2
34	8	64
42	16	256
31	5	25
54	28	784
26	0	0
45	19	361
63	37	1369
51	25	625
26	0	0
56	30	900
	$\sum d = 168$	$\sum d^2 = 4384$

Here, A=26, n=10

$$\bar{X} = A + \frac{\sum d}{n} = 26 + \frac{168}{10} = 42.8$$

$$s = \sqrt{\frac{1}{n-1}\left(\sum d^2\right) - \frac{(\sum d)^2}{n}}$$

$$= \sqrt{\frac{1}{10-1}\left(\sum 4384\right) - \frac{(168)^2}{10}}$$
$$= 13.17$$

Setting up the hypothesis,

Null hypothesis (H_0): $\mu_0 = \mu_1$, i.e., all the mean life of battery is the same.

Alternative hypothesis (H_1): $\mu_0 \neq \mu_1$, i.e., the mean life of battery is different.

The test statistic is,

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$$

Now,

$$t_{cal} = \frac{42.8-10}{13.17/\sqrt{10}} = 7.88$$

The level of significance $(\alpha) = 5\% = 0.05$

Now, tabulated value of t at $\alpha = 0.05$ and $n - 1 = 10 - 1 = 9$ d.f. is,

$$t_{tab} = t_{0.05,9} = 2.262$$

Since, $t_{cal} > t_{tab}$, then we reject H_0 and conclude that the mean life of battery is different.

3. A chemical company, wishing to study the effect of extraction time on the efficiency of an extraction operation, obtained the data shown in the following table:

Extraction Time in	27	45	41	19	35	39	19
Extraction efficiency	57	64	80	46	62	72	52

(a) Fit a straight line to the given data by the method of least squares and use it to predict the extraction efficiency one can expect when the extraction time is 35 minutes.

(b) Determine the coefficient of determination and interpret its meaning.

<u>Solution</u>

Here,

The regression equation of the line of Y on X is,

$Y = a + bx$ …(i)

Normal equation are,

$\sum Y = na + b\sum X$ …(ii)

$\sum XY = a\sum X + b\sum X^2$ …(iii)

and, the regression equation of line of X on Y is,

$\sum X = n = a' + b'y$ …(iv)

Normal equations are,

$\sum X = na' + b'\sum Y'$ …(v)

$\sum XY = a'\sum Y + b'\sum Y^2$ …(vi)

Now, calculation table,

X	Y	X^2	Y^2	XY
27	57	729	3249	1539
45	64	2025	4096	2880
41	80	1681	6400	3280
19	46	361	2116	874
35	62	1225	3844	2170
39	72	1521	5184	2808
19	52	361	2704	988

$\sum X = 225$	$\sum Y = 433$	$\sum X^2 = 7903$	$\sum Y^2 = 27593$	$\sum XY = 14539$

n=7

Now, putting the value in (ii)

$$433 = 7a + 225b$$

$$a = \frac{433-225b}{7} \text{ ...(1)}$$

Again, putting the value in equation (iii)

$$14539 = 225a + 7903b \text{ ...(2)}$$

Now, putting value of 'a' from (1),

$$14539 = \left(\frac{433-225b}{7}\right)225 + 7903b$$

$$\therefore b = 1.006$$

Now, putting the value of 'b' in equation (i),

$$a = \frac{433-1.006\times 2}{7}$$

$$\therefore a = 29.52$$

Now, on the line, Y on X,

$$Y = a + bx$$

$\therefore$ The required equation is $Y = 29.52 + 1.006x$.

Now,

(a) From above question, if extraction time is 35 minutes,

$$y = 29.52 + 1.006 \times 35 = 64.73$$

(b) Here, correlation coefficient is

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{7\times 14539 - 225\times 433}{\sqrt{7\times 7903 - 225^2}\sqrt{7\times 37593 - 433^2}} = 0.23$$

Now, for coefficient of determination,

$$r^2 = (0.23)^2 = 0.0529$$

which means that 5.29% of efficiency change is due to time and remaining of the efficiency change is due to other factors.

Group B

4. The following are the numbers of minutes that a person had to wait for the bus to work on 15 working days: 10,1,13,9,5,9,2,10,3,8,6,17,2,10 and 15. Find mean, median, mode and describe the shape of the distribution.

Solution

Arranging them in ascending order,

1, 2, 2, 3, 5, 6, 8, 9, 9, 10, 10, 10, 13, 15, 17.

Calculation table:

x	f	cf	fx	fx^2
1	1	1	1	1
2	2	3	4	8
3	1	4	3	9
5	1	5	5	25

6	1	6	6	36
8	1	7	8	64
9	2	9	18	162
10	3	12	30	300
13	1	13	13	169
15	1	14	15	225
17	1	15	17	285
	$N = 15$		$\sum fx = 115$	$\sum fx^2 = 1284$

Now,

Mean $(\bar{X}) = \frac{\sum fx}{N} = \frac{115}{15} = 7.67$

Median $(M_d) = \left(\frac{N+1}{2}\right)^{th}$ $item = \left(\frac{16}{2}\right)^{th}$ $item = 8^{th}\ item$

$\therefore\ M_d = 9$

Now, the highest frequency is 10

$\therefore$ Mode $(M_o) = 10$

Again, shape of distribution

$\bar{X} < M_o$ so the distribution is negatively skewed.

$$\sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2} = \sqrt{\frac{1284}{15} - \left(\frac{115}{15}\right)^2} = \sqrt{85.6 - 58.8289} = 5.17$$

Coefficient of skewness,

$$S_{kp} = \frac{\bar{X} - M_o}{\sigma} = \frac{7.67 - 10}{5.17} = -0.4507$$

5. The probability that an integrated circuit chip will have defective etching is 0.12 ,the probability that it will have a crack defect is 0.29, and the probability that it has both defects is 0.07. What is the probability that a newly manufactured chip will have either an etching or a crack defects.

Given,

For etching or a crack defects,

$P(E) = 0.12$

$P(C) = 0.29$

$P(E \cap C) = 0.07$

Probability of having etching or crack defects

$$P(E \cup C) = P(E) + P(C) - P(E \cap C) = 0.12 + 0.29 - 0.07 = 0.34$$

Now, we know that having neither of the defects is given by $P(\overline{E \cup C})$

We know,

$P(E \cup C) + P(\overline{E \cup C}) = 1$

$$\therefore P(\overline{E \cup C}) = 1 - 0.34 = 0.66$$

6. An importer is offered a shipment of machine tools for Rs. 140,000 and the probabilities that he will be able to sell them for Rs. 180,000 Rs, 170,000 or Rs.150,000 are 0.32, 0.55 and 0.13 respectively. What is the importers expected gross profit?

Solution

Let X_1, X_2, X_3 be the profit from selling in 180000, 170000, 150000.
Then,

$$X_1 = 180000 - 140000 = 40000$$
$$X_2 = 170000 - 140000 = 30000$$
$$X_3 = 150000 - 140000 = 100000$$

Now, expected gross profit,

$$E(X) = E(X_1) \times P_{x_1} + E(X_2) \times P_{x_2} + E(X_3) \times P_{x_3}$$
$$= 40000 \times 0.32 + 30000 \times 0.55 + 10000 \times 0.13$$
$$= 30600$$

7. If Two random variables have the joint density

f(x,y) = $\frac{6}{5}(x + y^2)$ for 0<x<1, 0<y<1

0 elsewhere

Find the probability that 0.2<X<0.5 and 0.4<Y<0.6

8. It is believed that 80% of Nepalese do not have any health insurance. Suppose this is true and let X equal the number with no health insurance in a random sample of n = 12 Nepalese.

i) Write the probability model of X?
ii) Give the mean and variance of X.
iii) Find P(X>2).

9. Given a random variable having the normal distribution with μ and σ^2= 1.5625, find the probabilities that it will take on a value (i) greater than 16.8, (ii) less than 14.9 (iii) between 13.6 and 18.8.

10. Define canonical definition of Chi square distribution and write its density function and its some properties.

Chi-square distribution is a non-parametric test developed in statistical theory and is specially used for experiments incorporated with enumerative (count) type data such as classification of the given responses into a number of categories.

Some of its properties are:

a. All the sample items should be independent.
b. For d.f.=1 i.e. k=2, no e_i's should be less than 5.
   For k>2 i.e. d.f.>1, χ^2-test for one sample case is not feasible if more than 20% of the e_i's are less than 5 or any e_i<1.
c. The overall frequency size must be reasonably large. Normally it should not be less than 50 irrespective to the number of class.

d. The constraints if any must be linear i.e.,

$$\sum(o_i - e_i) = 0 \Rightarrow \sum o_i - \sum e_i$$

It's density function: do yourself

11. Obtain the maximum likelihood estimate for the parameter π(proportion of success) of binomial distribution.

Solution

The likelihood function is given by

$$f(x_1, x_2, \ldots, x_n|\lambda) = \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$
$$= \frac{e^{-\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x! \ldots x_n!}$$

$\log f(x_1, x_2, \ldots, x_n|\lambda) = n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log c$

where, $c = \sum_{i=1}^{n} x_i!$ does not depend on f

$\frac{d \log f(x_1,x_2,\ldots,x_n|\lambda)}{d\lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$

By equation to zero, we obtain that the maximum likelihood estimate $\hat{\lambda}$ equal

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and so the maximum likelihood estimator is given by

$$d(x_1, \ldots, x_n) = \sum_{i=1}^{n} X_i$$

12. The average zinc concentration recovered from a sample of zinc measurements in 36 different locations is found to be 2.6 grams per milliliter. Find the 95% confidence interval for the mean zinc concentration. Assume that the population standard deviation is 0.3.

Here, we assume

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

The test statistic is,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} \ldots \text{(i)}$$

where,

$\mu = 2.6$

$n = 3.6$

$s = 0.3$

.Find t_{cal} by substituting these values in equation (i). Then, tally it with the tabulated value of t, i.e., t_{tab} and hence derive the conclusion. ☺

13. Explain the purpose of regression analysis. State model and assumptions of simple linear regression.