Probability and Statistics
Csc. 103-2067

Group A

Attempt Any Two: (2 x 10 = 20)

1. State Baye's Theorem. In a certain assembly plant, three machines B1,B2 and B3 make 30%,45%and 25% respectively, of the product. It is known from past experience that 2%, 3% and 2% of the products made by each machine, respectively, are defective. If a product were chosen randomly and found to be defective, what is the probability that it was made by machine B3.
   Ans:
   Bayes theorem:

If E1,E2,E3,……,En be the mutually exclusive events with P(Ei)≠0 (i=0,1,2…..n) then for any arbitrarily event A which is a subset of $(\sum_{i=1}^{n} Ei)$ such that P(A)>0

Then, we have,

$P(Ei/A)=\frac{P(Ei).P(A/Ei)}{\sum_{i=1}^{n} P(Ei).P(A/Ei)}$ (i=1,2,….n)

Where

- P(Ei) is called the priori probability
- P(Ei/A) is called the posteriori probability
- P(A/Ei) is the likelihood probability

Solution:

Let us define the following

E1=product from the machine B1

E2=product from machine B2

E3=product from machine B3

P(E1)=0.3 P(E2)=0.45 P(E3)=0.25

Let A denote the defective product

P(A/E1)= probability that the defective product is from machine B1=0.02

P(A/E2)= probability that the defective product is from machine B2=0.03

P(A/E3)= probability that the defective product is from machine B3=0.02

By using the bayes theorem the probability that a product found to be defective from machine B3 is given by

$P(E3/A)=\frac{P(E3).P(A/E3)}{\sum_{i=1}^{n} P(E3).P(A/E3)}$

$=\frac{0.25*0.02}{0.3*0.02+0.45*0.03}$ =ANS

2. (a) Explain point estimation and interval estimation. What are the criteria of good estimators?

Ans:

Point estimation:

a point estimate is a single number that is used to estimate the population parameter. For example, suppose the census bureau takes a sample of 10000 house holds and determines the mean housing expenditure per month $\bar{X}$. For this sample in Rs.6500. then using $\bar{X}$ as a point estimate. For μ, the bureau can state that the mean housing expenditure per month μ for all house holds is about Rs.6500. this procedure is called point estimation

Internal estimation:

An internal estimate is a range of values to estimate a population parameter.

In internal estimation, instead of a single value of a population parameter, an individual is constructed around the point estimate and it is stated that this interval is likely to contain the corresponding population parameter.

The criteria of a good estimator are:

a) unBiasness
b) consistency
c) efficiency
d) sufficiency

*explanation needed

(b) If =50 , s = 15 , n =16 and assuming that the population is normally distributed, estimate the standard error of the sample mean and estimate 99% confidence interval for the population mean μ.

Ans:

Given,

$\bar{X}$=50

S=15

n=16

standard error of sample mean

SE($\bar{X}$)=$\frac{S}{\sqrt{n}}$ = $\frac{15}{\sqrt{16}}$ = 3.75

For 99% confidence interval

1-α=0.99 α=0.01

z_α=$z_{0.01}$=2.573

CI=$\bar{X}$ ± z_α × $\frac{S}{\sqrt{n}}$

=50 ± 2.573×3.75

=50±9.65

CI is 4.35 to 59.65

3. (a) Define Karl Pearson's correlation coefficient and state its properties.

Ans:

The karl pearson correlation coefficient measures the degree of linear association between two variables. This method is popularly known as pearsonian correlation coefficient. It is unitless measure of correlation. It is the ratio of covariance between two variables to the product of the standard deviations of the two variables.

The properties of the karl pearson correlation coefficient are:

a) Correlation coefficient lies between -1 and +1. ie, $-1 \le r \le +1$
b) Correlation coefficient is geometric mean between two regression coefficient. ie, $r=\sqrt{bxy.byx}$
c) Correlation coefficient is a relative statistical measure and has no unit.

(b) The following table shows the production of coal and the number of wage earners in the coal industry over a ten year period during which the capital equipment has remained constant. Output in tons(y) 21 21 20 18 17 17 14 13 No of workers(X) 70 68 65 50 47 47 44 43 Determine the fitted regression line and predict Y for X = 55.

Ans:

X	Y	XY	X^2	Y^2
70	21	1470	4900	441
68	21	1428	4624	441
65	20	1300	4225	400
50	18	900	2500	324
47	17	799	2209	289
47	17	799	2209	289
44	14	616	1936	196
43	13	559	1849	169
$\sum X = 434$	$\sum X = 434$	$\sum XY = 7871$	$\sum X2 = 24452$	$\sum Y2 = 2549$

Here,

n=8,

$\sum X = 434,$

$\sum X = 434,$

$\sum XY = 7871,$

$\sum X2 = 24452,$

$\sum Y2 = 2549,$

Now we know that

Regression line for Y on X is

$Y=a+b_{yx}X$

And X on Y is

$X=b_{xy}Y$

Now,

$b_{yx} = \frac{n\sum XY - \sum X \sum Y}{n\sum X2 - (\sum X)2}$

$= \frac{8\times7871 - 434\times141}{8\times24452 - (434)2}$

$= \frac{62968 - 61194}{195616 - 188356}$

$= \frac{1774}{7260} = 0.244$

Therefore, b_{yx}=0.244

a =$\frac{\sum Y - byx.\sum X}{n}$

=$\frac{141-0.244\times434}{8}$

=4.388

And the regression equation for Y on X

Y=4.388+0.244X

For X on Y

b_{xy}= $\frac{n\sum XY-\sum X\sum Y}{n\sum Y2-(\sum Y)2}$

=$\frac{1774}{511}$

=3.47

Therefore, b_{xy}=3.47

Now,

a=$\frac{\sum X - byx.\sum Y}{n}$

=$\frac{434-3.47\times141}{8}$

=-6.908

Therefore, regression equation X on Y is

X=-6.908+3.47

Again,

For X

Y=4.388+0.244×55

=4.388+13.42

=17.808

Group B

Answer any eight questions: (8 x 5 = 40)

4. The following data represent the total fat for burgers from a sample of fast-food chains. 19 31 34 35 39 43 Compute mean, median and mode then describe the shape of the distribution.
5. What is axiomatic definition of probabilities and what are its properties?

   Ans:

   The combined form of classical approach and statistical approach which gives a modern approach to define probability is called axiomatic approach. There are three different axiom. they are:

   I. 0≤P(A)≤1
   II. P(S)=1 where S=sample space
   III. If A1,A2,……,An are mutually exclusive event then P(A1V,A2V,…..,Van)
        =P(A1)+P(A2)+…..+P(An)

   And the properties are as follows:

6. If two random variables X1 and X2 have the joint probability density f(x1,x2) = for 0<x1<1, 0<x2<1 0, elsewhere Find the conditional density of X1 given X2 = x2.

Ans:[2]

$$= \int_0^1 \int_0^1 \frac{2}{3}(x1 + 2x2)dx2dx1$$

$$= \frac{2}{3}\int_0^1 \left[x1x2 + \frac{2x_2{}^2}{2}\right]\frac{1}{0}\text{dx1}$$

$$= \frac{2}{3}\int_0^1 [x1 + 1]\, dx1$$

$$=\frac{2}{3}\left[\frac{x1^2}{2} + x1\right]_0^1$$

$$=\frac{2}{3}\left[\frac{1}{2} + 1\right]$$

=1

Conditional density of X1 is given X2

f X1/X2(X1/X2)=$\frac{\mathrm{F(X1,X2)}}{F(X2)}$

$=\frac{2}{3}(X1 + 2X2) \times \frac{3x2}{4}$

$=\frac{1}{2}(x1x2 + 2x2^2)$

7. Prove that Var(X+Y) = Var(X)+Var(Y) +2Cov(X,Y).

Ans:

Var(X+Y)=E[(X+Y)(X+Y)]−E[X+Y]2

=E[X2+2XY+Y2]−(μx+μy)2

=E[X2+2XY+Y2]−μ2 x−2μxμy−μ2 y

=(E[X2]−μ2 x)+(E[Y2]−μ2 y)+2(E[XY]−μxμy)

=Var(X)+Var(Y)+2Cov(X,Y)

8. Find the first and second moments of binomial distribution and also compute variance for the binomial distribution.

Ans:

Let X~ B(n,0) be the probability mass function is given by

9. Service calls come to a maintenance center according to a Poisson process and on the average 2.7 calls come per minute. Find the probability that no more than 4 calls come in any period.

10. In a photographic process, the developing time of prints may be looked upon as a random variable having the normal distribution with a mean of 16.28 seconds and a standard deviation of 0.12 second. Find the probability that it will take (i) anywhere from 16.00 to 16.50 seconds to develop one of the prints, (ii) at least 16.20 seconds to develop one of the prints.

Ans:

Here we are given,

X~N(μ-σ^2)

Where,

μ=16.28

$\sigma = 0.12$

we know,

$\frac{\bar{X}-\mu}{\sigma}=Z \sim (0,1)$

for (1)

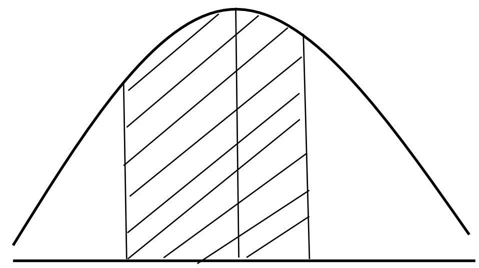
P(16<x<16.50)

When X=16, $Z=\frac{16.00-16.28}{0.12}$

$=-2.33$

When X=16.50, $Z=\frac{16.50-16.28}{0.12}$

$=1.83$



Now, P(16.00<x<16.50)

$=P(z<-2.33)+P(z<1.83)$

$=P(-\infty<z<0)-P(-2.33<z<\infty)+0.5-P(0<z<1.83)$

$=0.5-P(-2.33<z<\infty)+0.5-P(0<z<1.83)$

$=0.5-0.099+0.5-0.336$

$=0.565$

For(2)

At least 16.20 sec to develop of the prints

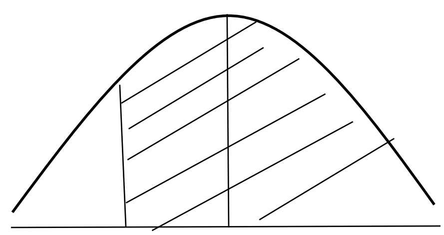
P(X>16.20)

When,

X=16.20

$Z=\frac{16.20-16.28}{0.12}$

$=-0.66$



Now,

$P(X>16.20)=P(z>-0.66)$

$=P(-0.66<z<0)+P(0<z<\infty)$

$=P(0<z<0.66)+0.5$

$=0.2546+0.5$

$=0.7546$

11. Obtain the maximum likelihood estimate for mean (μ) and variance (σ^2) of the normal distribution.
12. Define canonical definition of t-distribution. Discuss some of its properties.

 Ans:

 Let 'X' be the continuous random variable, then 'X' is said to following t-distribution if its probability density function is given as.

$$P(X=x)=\frac{1}{\sqrt{n}\beta\left(\frac{1}{2},\frac{n}{2}\right)}\cdot\frac{1}{\left(1+\frac{x^2}{n}\right)}-\frac{n+1}{n} \qquad -\infty\le t\le\infty$$

We write,

$X\sim t_n$ to denote X follows t-distribution with n d.f

Some of its properties are:

*for the test of significance of single mean

Here,

Test statistic

$t=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ if population variance is given.

$t=\frac{\bar{X}-\mu}{s/\sqrt{n-1}}$ if sample variance is given

*for the test of significance of two sample mean

$t=\frac{\overline{X1}-\overline{X2}}{\sqrt{s^2\left(\frac{1}{n1}+\frac{1}{n2}\right)}}$ where $S^2=\frac{n1s1^2+n2s2^2}{n1+n2}$

13. It is claimed that an automobile is driven on the average more than 20,000 kilometers per year. To test this claim, a random sample of 100 automobiles owners are asked to keep a record of the kilometers they travel. Would you agree with this claim if the random sample showed an average of 23,500 kilometers with a standard deviation of 3900 kilometers?

Ans:

Here we assume

$H_o=\mu1=\mu2$ vs

$H_1=\mu1\neq\mu2$

Test statistics under H_o

n=100, $\bar{X}=20000$, μ=12500, SD=3900

we are using Z test

$Z=\frac{\bar{X}-\mu}{sd/\sqrt{\sigma}}$

$=\frac{20000-12500}{3900/\sqrt{100}}$

$|Z|_{cal}=8.97$

Critical region: compare yourself through Z table using 0.05% level of significance