Probability and Statistics
Csc. 103-2069

Group A

1. Define the following three measures of dispersion – range, standard deviation and inter quartile range –by clearly state their properties. Write down a situation where range is preferred to Standard deviation. Score obtained by 10 Students in a test are given below. Compute range, and standard deviation.

42	55	35	60	55	55	65	40	45	35

Range:- Range is define as the difference between the upper limit of the largest maximum and lower limit of the smallest minimum observation

Range(R)=L-S

Here, L=largest observation

S=Small Observation

Standard deviation:- A standard deviation is defined as the positive square root of the arithmetic mean of the square of the deviation taken from the arithmetic mean. It is devoted by Greek letter σ (read as sigma). It is also defined as the rot mean square deviation from the mean.

Symbolically, $\sqrt{\frac{1}{n}\sum(X-\bar{X})^2}$

Inter quartile range:- The difference between third quartile & first quartile is called inter quartile range.

Mathematically

Quartiles (Q)=Q_3-Q_1

The difference between the upper and lower quartile is known as inter-quartile range.

Properties of standard deviation as follows

i. Standard deviation is the least possible value of rot mean square deviation.
ii. Standard deviation is independent of change of origin but not of scale.

Properties of range as follow

i. It is the simplest of all the measures of dispersion of the difference between the largest (maximum) and the small observation in the distribution

ii. Variability of two distributions having different units of measurement are need relative measure of dispersion relative measure of dispersion based on range is known as coefficient of range

Properties of inter-quartile range

i. Bless the coefficient of Q.D. implies more will be the uniformity or less will be the variability
ii. Greater the coefficient of Q.D. implies less will be the uniformity or greator will be the variability
iii. For open end classes most suitable measure of dispersion is Q.D.

Given for Range

Arranging : 35,35,40, 42, 45, 55, 55, 55, 60, 65

We know ,

Range(R) = Largest Value- Smallest Value = 65-35 = 30

Now, for standard deviation

X	Tally	F	FX	FX^2
32	II	2	70	2450
40	I	1	40	1600
42	I	1	42	1764
45	I	1	45	2025
55	III	3	165	9075
60	I	1	60	3600
65	I	1	65	4225
		N=10	$\sum FX=487$	$\sum FX^2=24739$

$\Sigma=\sqrt{\frac{\sum FX^2}{N}-\left(\frac{\sum fX}{N}\right)^2}$

$=\sqrt{\frac{24739}{10}-\left(\frac{487}{10}\right)^2}$

$=\sqrt{2473.9-2371.69}$

=10.109

2. There are three traffic lights on your way home. As you arrive at each light assume that it is either red (R) or green (G) and that it is green with probability 0.7 . Construct the sample space by listing all possible eight simple events. Assign probability to each simple event. Are the events equally likely? What is the probability that you stop no more than one time.

3. A large company wants to measure the effectiveness of radio advertising media (X) on the sale promotion (Y) of its products. A sample of 22 cities with approximately equal population is selected for study. The sales of the product in thousand Rs and the level of radio advertising expenditure in thousand Rs are recorded for each of the 22 cities (n) and sum, sum of square, and sum of cross product of X and Y are summarized below.
$\sum Y = 26953$m $\sum X = 950$, $\sum Y^2 = 35528893$, $\sum X^2 = 49250$, & $\sum XY = 1263940$
   a. Fit a simple linear regression model of Y on X using the least square method. Interpret the estimated slope coefficient.
   b. Compute R^2 and interpret it.

Solution

Given,

$\sum Y=26953$m

$\sum X = 950$

$\sum Y^2 =35528893$

$\sum X^2=49250$

$\sum XY=1263940$

n=22

The regression equation for Y on X is ,

Y= a+bX ----------(i)

Here, b is the slope coefficient

For the work , Normal equations are,

$\sum Y = na + b \sum X$ -------------------(ii)

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

Now, From(i)

26953 =22a+ b950

a = $\frac{26953 - b950}{22}$ ----------------------(iv)

Now, for (iii)

1263940 = a950 + b 49250

Dividing all by

25388 = 190 a + 9850 b

12694 =95 a + 4925 b

Now, putting value of a form (iv)

$12694 = 95\left(\frac{26953-950b}{22}\right) + 4925b$

279268=95(26953-950b) +10835b

279268 = 2560535 – 90250b + 10835b

279268 – 2560535 = -79415b

-2281267 = -79415b

28.725 = b

Putting value of b in (iv)

$a = \frac{26953 - 28.725\ x\ 950}{22} = -15.26$

therefore, regression eqn is ,

y=-15.26 +28.75X

therefore, slope is 28.75

$r = \frac{n\sum xy - \sum X \sum Y}{\sqrt{22 \times 49250-(950)^2}\sqrt{55171437}}$

$= \frac{2201330}{425.44 \times 7427.74}$

= 0.6966

Now,

R^2 =0.4852

It is coefficient of determinant. It means 48.5% of change in sales due to advertisement remaining is due to other factors.

Group B

4. Describe the scopes and limitations of statistics in empirical research.
5. Write down the properties and importance of density function of a continuous random variable. Suppose a continuous random variable X has the density function.

f(x) = $\begin{cases} k(1-x)^2 for\ 0 < x < 1 \\ \quad 0\ elsewhere \end{cases}$

Find (a) value of the constant k, and (b) E(X).

6. Suppose that X and Y have joint density function

$f(x) = \begin{cases} (x+y), & if\ 0 < x, \quad y < 1 \\ 0 & \quad otherwise \end{cases}$

Find
(a) Marginal density function of X and Y, and
(b) Covariance between X and Y.

7. In a Poisson distribution with parameter λ derive the mean and variance of the distribution.

Solution,

For mean,

E(x)=$\sum_x xp(x)$

=$\sum_{x=0}^{\infty} xe^{-\lambda}.\lambda^x$

=$e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!}$

=$e^{-\lambda}\lambda\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$

=$e^{-\lambda}\lambda[1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots]$

=$e^{-\lambda}\lambda e^{\lambda}$

=E(x)=λ…………………………………………(1)

For variance,

E(x^2)= $\sum_x x^2p(x)$

=$\sum_{x=0}^{\infty} x^2\frac{e^{-\lambda}\lambda^x}{x!}$

=$e^{-\lambda}\{\sum_{x=2}^{\infty}\frac{x(x-1)\lambda^x}{x(x-1)(x-2)!}+\sum_{x=1}^{\infty}\frac{x.\lambda^x}{x(x-1)!}\}$

=$e^{-\lambda}\{\lambda^2\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}+\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}\}$

=$e^{-\lambda}\{\lambda^2\left[1+\frac{\lambda}{1!}++\frac{\lambda^2}{1!}+\cdots\right]+\lambda^2\left[1+\frac{\lambda}{1!}++\frac{\lambda^2}{2!}\right]$

=$e^{-\lambda}\{\lambda^2.e^{\lambda}+\lambda.e^{\lambda}\}$

=$\lambda^2+\lambda$

Now,

Variance V(X)=E(X^2)-E(X))^2

=$\lambda^2+\lambda-\lambda^2$

V(X)=λ

Both V(X) and E(X) i.e. variance and mean both seem to be λ as they are equal.

8. The length of life of automatic washer (X) is approximately normally distributed with mean and standard deviation equal to 3.1 and 1.2 years, respectively. Compute the probabilities (a) P(X>1), (b) P(X>2.5) and (c) P(1<X<2).

Solution

Given,

Mean(μ)=3.1

S.D(σ)=1.2

Now,

(a)

P(X>1)

Here, X=1

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

=$\frac{1-3.1}{1.2}$

=$\frac{-2.1}{1.2}$

=-1.75

Here,

P(X>1)=P(Z>-1.75)+0.50

=0.4599+0.50

=0.9599

(b)

P(X>2.5)

Here,

X=2.5

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

=$\frac{2.5-3.1}{1.2}$

=-0.5O

Here,

P(X>2.5)=P(Z>-050)

=(-0.50<Z<0)+P(0<Z<∞)

=0.1915+0.5

=0.915

(c)

P(1<X<Z)

X=1, Z=-1.75

X=2

Z=$\frac{-1.1}{1.2} = -0.916$

P(-1.75<2<-0.916) =(-1.75<Z<0)-(-0.916<Z<0)

=(0<Z<1.75)-(0<Z<0.916)

=0.4599-0.3186

=0.141300

9. If X_1, X_2, …., X_n are n independent random variables each is distributed as normal with mean μ and variance σ^2, then derive the distribution of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

10. If a continuous random variable X has exponential distribution with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & otherwise \end{cases}$$

For h>0, prove that P(X>t+h|X>t)=P(x>h), and hence prove that P(X>t + h) = P(X >t) x P(X>h).

11. If X_1, X_2, ….,X_n are n independent Bernoulli random variables with common mean p, derive the maximum likelihood estimator of p. Prove or disprove the estimator is unbiased for p ?

12. A car manufacturer claims that its car use, on average, no more than 5.5 gallons of petrol for each 100 miles. A consumer groups tests 40 of the cars and finds an average consumption of 5.65 gallons per 100 miles and a standard deviation of 1.52 gallons. Do these results cast doubt on the claim made by the manufacturer ? Answer the question by setting appropriate null and alternative hypotheses and testing the null hypothesis at 5% level of significance.

Here, we assume
H0:µ1= µ2 vs H1: µ1≠ µ2
Test statistic under Ho
Z=$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ where $\bar{X} = 5.65, \mu = 5.5$,n=40,σ=1.52
Z=$\frac{5.65-5.5}{\frac{1.52}{\sqrt{40}}}$ = 0.62
$Z_{cal,0.05}$ = 0.65
Critical region : find out the value of 2 at 0.05 level f significance
Decision: compare tab & cal. Value of 2 then take decision yourself.

13. The average length of time required to complete a certain aptitude test is claimed to be no more than 80 minutes. A sample of 25 students yielded an average of 86.5 minutes and a standard deviation of 15.4 minutes. Do these results cast doubt on the claim? Assuming that test score is normally distributed answer the query by setting appropriate null and alternative hypotheses and testing the null hypothesis at 5 % level of significance.

We assume,

H0: μ1= μ2 vs

H1: μ1≠ μ2

Test statistics

Under H0

T=$\frac{\bar{X}-\mu}{s\sqrt{n-1}}$

Where,

$\bar{X}$=86.5

μ=80

n=25

σ=15.4

T=$\frac{86.5-80}{15.4\sqrt{25-1}}$ = 0.086

Critical region:

$(Z_{0.05})_{tab}$ = find yourself

Decision: compare tab & cal value of t and take decision.