Probability and Statistics
Csc. 103-2069

Group A

Q no.1) Write the algebraic computation expression for mean and standard deviation based on a given sample x1, x2, ……xn. Why they are important in statistics? Write down their properties compute the mean and standard deviation from the following scores of 10 students.

➔ Here, the given data is

x1, x2,…… xn

now, we know that,

Mean($\overline{X}$) = $\frac{\text{Total sum of items}}{Total\ no.of\ items}$

= $\frac{x1+x2+\cdots..+xn}{N}$

= $\frac{\sum xi}{N}$ [i= 1,2,…..n]

Similarly,

Standard Deviation :-

σ = $\sqrt{\frac{\sum x^2}{n}-\left(\frac{\sum x}{n}\right)^2}$

Importance of mean and Standard Deviation:

1. It is easy to calculate and simple to understand.
2. It is based on all items.
3. It is suitable for further mathematical calculations.
4. It is clearly defined.
5. It is not necessary to arrange the given data to compute the mean.
6. It is least affected by sampling fluctuations.

Now, given data is:

45, 55,35,60,55,55,65,40,45,35 and n=10

Now computing,

$\overline{X} = \frac{\sum xi}{N}$

$= \frac{45+55+35+60+55+55+65+40+45+35}{10}$

$= \frac{490}{10}$

= 49

Now, Standard Deviation

X	X^2
45	2025
55	3025
35	1225
60	3600
55	3025
55	3025
65	4225
40	1600
45	2025
35	1225
$\sum$X=490	$\sum X^2$=25000

$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

$= \sqrt{\frac{2500}{10} - \left(\frac{490}{10}\right)^2}$

$= \sqrt{2500 - (49)^2}$

$= \sqrt{2500 - 2401}$

$= \sqrt{99}$

Merits of standard deviation :-

1. It is rigidly defined.
2. It is based in all observation.
3. It is suitable for further mathematical treatment.
4. It is less effected by sampling fluctuations than other measures of dispersion.
5. It is suitable for various comparing variability.

Q no.2) Explain the terms – sample space and an event of a random experiment. State the classical and the statistical definition of probability. Which of the two definitions is most useful in statistics and why?

➔ Sample Space:

The total number of all possible outcomes in any experiment is known as exhaustive event or sample space. Sample space is denoted by S, Ω(omega), U.

For example : There are four exhaustive events for tossing coin two times. There are six exhaustive events for throwing a die.

Classical Approach:

If there are 'n' exhaustive, mutually, exclusive and equally likely events, out of which 'm' events are the occurrence of event A (say), then the probability of occurrence of events A is denotes by P(A) and is defines by

P(A) $= \frac{m}{n}$, m ≤ n

$= \frac{favourable\ no,of\ cases}{exhaustive\ or\ total\ no,of\ cases}$

Also, the probability of non-occurrence of event A is denoted by P($\overline{A}$), P(A^c) or P(A^1) and is defined as

P($\overline{A}$) = $\frac{n-m}{n}$

$= \frac{1-m}{n}$

= 1- P(A)

P($\overline{A}$) = 1- P(A)

i.e, P(A) + P($\overline{A}$)= 1

Statistical approach :-

If an experiment (trail) is repeated under similar condition for a large no. of times then the limiting value of the ratio of the number of times the occurrence of the event A(say) to the total no, of trails, provided the trails are indefinitely large and limited is unique and finite is known as the probability of occurrence of event A.

Mathematically,

P(A) =$\lim_{n \to \infty} \frac{m}{n}$

Q. no. 3) A large company wants to measure the effectiveness of newspaper advertising on sale promotion of its products. A sample of 22 cities with approximately equal populations is selected for study. The sale of the product(Y) in thousand Rs and the level of newspaper advertising expenditure(X) in thousand Rs are recorded for each or the 22 cities (n) and the recorded sum, sum of square, and the sum of cross produce X and Y are summarized below.

, $\sum$Y= 26953, $\sum$X= 660, $\sum Y^2$ = 35528893, $\sum X^2$= 22700, $\sum$XY= 851410

Using the above summary results:

a) Compute correlation coefficient between X and Y and coefficient of determination.
b) Fit a simpler linear regression model of Y and X using least square method and interpret the estimated slope regression coefficient.

➔

A) Here, In a large company wants to measure the effectiveness of newspaper advertising media on sale promotion of its product.

Now, given n=22, $\sum$Y= 26953, $\sum$X= 660, $\sum Y^2$ = 35528893, $\sum X^2$= 22700, $\sum$XY= 851410

Regression equation of Y to X :-

Y= a + b_{yx}X

Where, constant a and b_{yx} can be obtained by,

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

$$= \frac{22*851410-26953*660}{22*22700-435600}$$

$$= \frac{18731020-17788980}{499499-435600}$$

= 14.765

$$a = \frac{\sum X - b_{xy}\sum Y}{n}$$

$$= \frac{660-14.765*26953}{22}$$

= -18059.138

So, y= -18059.138 + 14.765X

$$\sigma_x= \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

$$= \sqrt{\frac{22700}{22} - \left(\frac{660}{22}\right)^2}$$

$$=\sqrt{1031.818 - 900}$$

= 11.481

$$\sigma_y= \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2}$$

$$= \sqrt{\frac{35528893}{22} - \left(\frac{26953}{22}\right)^2}$$

$$=\sqrt{1614949.682 - 1500959}$$

= 337.624

b_{xy} = r × $\frac{\sigma_y}{\sigma_x}$

14.765= r × $\frac{337.624}{11.481}$

r = $\frac{14.765 \times 11.481}{337.624}$

= 0.502

Co-efficient of determination (R) = 1- $\frac{SS_{error}}{SS_{total}}$

B) Here, the regression equation of line of 4 on X

Y= a +bX

Normal equation are :-

∑Y= na+b∑X

∑XY= a∑X +b$\sum X^2$

Substituting of ∑X, $\sum X^2$, ∑XY, and n=22, we get

26953= 22 × a + b × 660 ……..(i)

851410= a c 660 + b × 22700 ……(ii)

Now, solving them simultaneously, we get :

b = 14.7655

Now,

a = $\frac{26953-660 \times 14.7655}{22}$

= 782.17

Now,

Y= a +bX

Y= 782.17 + 14.76X

Group B

Q no.4) Write down the role of probability theory in a statistical with suitable example.

- Probability theory is the branch of mathematics concern with probability and the analysis of random phenomenon. The central objects of probability theory are random variables, stochastic process and events, mathematical abstraction if non-deterministic events or measured quantities that may either be single occurrence or evolve over time in an apparent random fashion.
  For e.g, If an individual coin toss or the roll of dice is considered to be a random events, then if repeated many times of sequence of random events will exhibit certain patterns, which can be studied and predicted. Two representative mathematical results describing such patterns are the law of large numbers and central limit theorem.

Q no.5) Explain discrete and continuous random variables with suitable examples. Suppose a continuous random variable X has the density function.

- A discrete random variable is one which may take only a countable number of distinct value such a 0,1,2,3,4. Discrete random variable are usually counts, If a random variable that can take only a finite numbers of distinct values, then it must be discrete. Examples of discrete random variables include the numbers if children in a family, the Friday night attendance at a cinema, the no. of patients appointment with a doctor, the number of defective light bulbs in a box of ten.

  Continuous Random Variables:
  A continuous random variable is one which takes an infinite number of possible values, continuous random variables are usually a measurement. Examples include heights, weights, the amount of sugar in an orange, the time required to run a mile.

Q no.7) In a binomial distribution with parameter n and p derive the mean and variance of the distribution.

➔ Mean and variance of binomial distribution

E(x) = Mean

V(x) = E(x^2) – [$E(x)^2$]

Let X~B(n,p) then probability mass of function of Binomial distribution

P(X=x) = $\begin{cases} {}^{n}_{x}c\ p^x q^{n-x} & for\ all, x = 0,1,2,\dots,n \\ 0 & otherwise \end{cases}$

Now, for mean

Mean = E(x)

=∑x.P(X=x)

$= \sum_{x=0}^{n} x\ {}^{n}_{x}C\ p^x q^{n-x}$

$= \sum_{x=0}^{n} x\ \frac{n!}{(n-x)!x!} p^x q^{n-x}$

$=\sum_{x=1}^{n} x\ \frac{n!}{(n-x)!x(x-1)!} p^x q^{n-x}$

$=\sum_{x=1}^{n} x\ \frac{n(n-1)!}{(n-x)!(x-1)!} p^x q^{n-x}$

$= n\sum_{x=1}^{n} x\ \frac{(n-1)!}{(n-x)!(x-1)!} p^x q^{n-x}$

=n $\sum_{x=1}^{n} {}^{n-1}_{x-1}C\ p^x q^{n-x}$

=np $\sum_{x=1}^{n} {}^{n-1}_{x-1}C\ p^{x-1} q^{n-x}$

=np$(p+q)^{n-1}$

Therefore, E(x) =np …………(i) [since p+q=1]

Then, For variance:

E(X^2) = $\sum x^2$p (X=x)

$= \sum_{x=0}^{n} x^2\ {}^{n}_{x}C\ p^x q^{n-x}$

$= \sum_{x=0}^{n}\{x(x-1)+x\}.\ {}^{n}_{x}C\ p^x q^{n-x}$

$=\sum_{x=0}^{n} x(x-1).\ {}^{n}_{x}C\ p^x q^{n-x} + \sum_{x=0}^{n} x.\ {}^{n}_{x}C\ p^x q^{n-x}$

$= \sum_{x=0}^{n} x(x-1).\frac{n!}{(n-x)!x!}\ p^x q^{n-x}$ + E(x)

$=\sum_{x=2}^{n} x(x-1).\frac{n!}{(n-x)!x(x-1)(x-2)!}\ p^x q^{n-x}$ + np

$= \sum_{x=2}^{n} \frac{n(n-1)(n-2!}{(n-x)!(x-2)!}\ p^x q^{n-x}$ + np

$=n(n-1)\sum_{x=2}^{n} \frac{(n-2!}{(n-x)!(x-2)!}\ p^x q^{n-x}$ + np

$= n(n-1)p^2\sum_{x=2}^{n}\ {}^{n-2}_{x-2}C\ p^{x-2} q^{n-x}$ + np

$= n(n-1)p^2\ (p+q)^{n-2}$ + np

$= n(n-p\ )^2 \times 1$ + np [since, p+q=1]

$E(X^2) = n(n-1)p^2$ + np

Now,

Variance,

V(X) = $\sum(x^2) - [E(x)]^2$

= $n(n-1)p^2$ + np – $(np)^2$

= n^2p^2 - p^2n + np -n^2p^2

= np - p^2n

= np(1-p)

Therefore,

V(X) = npq ……(ii) [since, p+q=1]

Here equation (i) and (ii) gives the expression of mean and variance.

Q no. 11) If a random variable X is normally distributed with a mean of 120 and a standard deviation of 12. Compute the following probabilities: (i) P(X<130) (ii) P(X<115) and (iii) P(110<X<130).

➔ Given a normal distribution with mean μ = 120 and σ = 12

Here,

We are given:

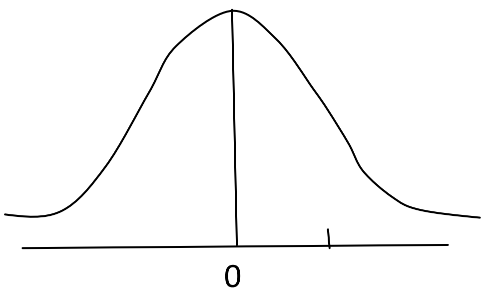
X~ N (μ, σ^2)

Where, μ = 120, σ = 12

We know,

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

For (i), (X>130)



When, X=130

0.83

So, Z= $\frac{X-\mu}{\sigma} = \frac{130-120}{12}$ = 0.83

P(Z>0.83)

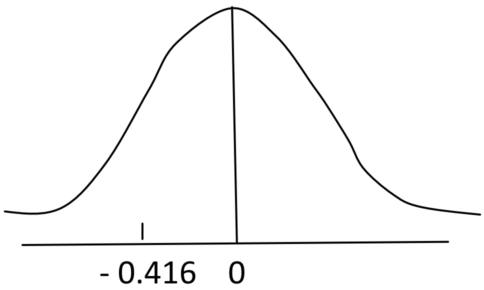
= P(0< 2<∞) – (0<2<0.83)

= 0.53 – 0.2967

= 0.2333

(ii) P(X<115)

When, X=115



Therefore, Z= $\frac{X-\mu}{\sigma}$ = $\frac{115-120}{12}$ = -0.146

So, P(X<115) = P(Z<-0.416)

= P(-∞< Z<0) – P(-0.41<Z<O)

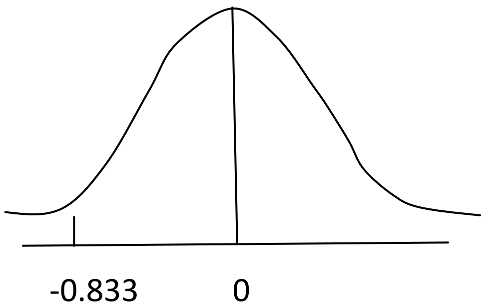
=0.5 – P(0<Z<0.41)

= 0.5- 0.1591

= 0.3409

(iii) P(110<X<130)

then, X=110



therefore, Z= $\frac{X-\mu}{\sigma}$ =$\frac{110-120}{12}$ = -0.833

= P(-0.833<X<0.833)

=P(-0.833<Z<0) + P(0<Z<0.833)

=P0<Z<0.833) + P(0<Z<0.833)

=0.2967 + 0.2967

=0.5934

Q no. 12) A manufacturer of TV sets claims that the average life of its picture tubes is at least 10 years. A sample survey of 100 of the picture tubes showed an average of 9.6 years and a standard deviation if 2.6 years. Do these results cast doubt on the claim of the manufacturer? Answer this question by setting approximation null and alterative hypothesis and testing the null hypothesis at 5% level if significance

➔ We have,

Sample mean ($\overline{X}$) = 9.6 years

Sample S.D (σ) = 2.6 years

Population mean (μ) = 10 yrs

Sample size (n) = 100>30

Level of significant

For z-test (large sample test n>30)

For test of single mean, test statistics

$$Z= \frac{\overline{X}-\mu}{\sigma/\sqrt{n}},$$

Where, $\overline{X}$ = Sample mean

μ = population mean

σ = Population standard deviation

n = sample size

If σ is not given, we use sample S.D (s) instead if σ.

To test whether there is significant difference in null and alternative hypothesis:

H_0= μ =10 , There is no significant difference in null and alternative hypothesis.

Vs, H_1= μ ≠10 (Two test tailed) there is significant difference in sample and population mean,

Test Statistics is,

$$Z= \frac{\overline{X}-\mu}{\sigma/\sqrt{n}},$$

$$= \frac{\overline{X}-\mu}{s/\sqrt{n}}, = \frac{9.6-10}{2.6/\sqrt{100}}$$

Therefore, $|Z_{cal}|$ = 1.538

Here, level of significance (α) = 5%

Now, tabulated valude of z at 5% level of significance for two tailed test,

Z_{tab} = 1.96

Since, Z_{cal} < Z_{tab} we should accept H_0 and reject H_1 with the conclusion that there is significant difference.