→ Gaurav Chaulagain
→ 017BSCIT

Page
Date

Model Question of Statistics.

1) Solution:
let M_1 represent the event that an error occur in the first module.
let M_2 represent the event that an error occur in the second module.
let E represent the error occuring in the module.
According to the question,

$P(M_1) = 0.2$

$P(M_2) = 0.4$

$P(E/M_1) = 0.5$

$P(E/M_2) = 0.8$

$P(M_1/E) = ?$

$P(M_2/E) = ?$

$P\left(\frac{M_1}{E} \cap M_2/E\right) = ?$

We know,

$P(E) = P(M_1) \cdot P(E/M_1) + P(M_2) \cdot P(E/M_2)$

$= 0.2 \times 0.5 + 0.4 \times 0.8$

$\therefore P(E) = 0.42$

By using Baye's Theorem,

$$P(M_1/E) = \frac{P(M_1) \cdot P(E/M_1)}{P(E)}$$

$$= \frac{0.2 \times 0.5}{0.42}$$

Scanned by CamScanner

Page
Date

$\therefore P(M_1/E) = 0.238$

Again,

$$P(M_2/E) = \frac{P(M_2) \cdot P(E/M_2)}{P(E)}$$

$$= \frac{0.4 \times 0.8}{0.42}$$

$\therefore P(M_2/E) = 0.762$

since, $P(M_1/E)$ and $P(M_2/E)$ are independent,

$$P(M_1/E \cap M_2/E) = P(M_1/E) \times P(M_2/E)$$

$$= 0.238 \times 0.762$$

$\therefore P(M_1/E \cap M_2/E) = 0.181$.

Hence, the probablity of errors in both modules is 0.181.

2) solution:

Given,

X	43	37	50	51	58	105	52	45	45	10

Arranging in descending order,

X	10	37	43	45	45	50	51	52	58	105
X^2	100	1369	1849	2025	2025	2500	2601	2704	3364	11025

$\therefore \quad \Sigma X = 496$ and $\Sigma X^2 = 29562$.

Scanned by CamScanner

Page______
Date______

a) solution:

$n = 10$

Then,

Mean $(\bar{X}) = \frac{\Sigma x}{n} = \frac{496}{10} = 49.6$

for Median,

$Md = \left(\frac{n+1}{2}\right)^{th}$ item $= 5.5^{th}$ item

$= 5^{th}$ item $+ 0.5$ (6^{th} item $- 5^{th}$ item)

$= 45 + 0.5 \times (50 - 45)$

$\therefore M_d = 47.5$

for Q1,

$Q1 = \left(\frac{n+1}{4}\right)^{th}$ item $= 2.75^{th}$ item

$= 2^{nd}$ item $+ 0.75 \times$ (3^{rd} item $- 2^{nd}$ item)

$= 37 + 0.75 \times (43 - 37)$

$\therefore Q_1 = 41.5$

for Q3,

$Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item $= 8.25^{th}$ item

$= 8^{th}$ item $+ 0.25$ (9^{th} item $- 8^{th}$ item)

$= 52 + 0.25 (58 - 52)$

$\therefore Q_3 = 53.5$

Sample standard deviation $(s) = \sqrt{\frac{1}{(n-1)} \cdot (\Sigma X^2 - n\bar{X}^2)}$

$= \sqrt{\frac{1}{9} (29562 - 10 \times (49.6)^2} = 23.477$

Page
Date

b) Using 1.5 IQR rule,
for outliners,

$IQR = Q_3 - Q_1 = 53.5 - 41.5$
$= 12.$

$1.5 \times IQR = 1.5 \times 12 = 18$

Now,
upper outlier limit $= Q_3 + 1.5 \times IQR = 53.5 + 18 = 71.5$
lower outlier limit $= Q_1 - 1.5 \times IQR = 41.5 - 18 = 23.5$

The outlier range is $(23.5 - 71.5)$.

c) Since, 10 and 105 outlies from outlier range, so they are deleted.
so, new set of data obtained is

X	37	43	45	45	50	51	52	58	$\Sigma X = 381$
X^2	1369	1849	2025	2025	2500	2601	2704	3364	$\Sigma X^2 = 18437$

Here,
$n = 8$
Then,

$$\text{Mean } (\bar{X}) = \frac{\Sigma X}{n} = \frac{381}{8} = 47.625$$

for Median,

$$Md = \left(\frac{n+1}{2}\right)^{th} \text{ item} = 4.5^{th} \text{ item}$$
$$= 4^{th} \text{ item} + 0.5\,(5^{th} \text{ item} - 4^{th} \text{ item})$$
$$= 45 + 0.5 \times (50 - 45)$$
$$\therefore Md = 47.5$$

Page
Date

for Q_1,

$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item $= 2.25^{th}$ item

$= 2^{nd}$ item $+ 0.25 \times (3^{rd}$ item $- 2^{nd}$ item$)$

$= 43 + 0.25 \times (45 - 343)$

$\therefore Q_1 = 43.5$

for Q_3,

$Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item $= 6.75$ th item

$= 6^{th}$ item $+ 0.75 \times (7^{th}$ item $- 6^{th}$ item$)$

$= 51 + 0.75 \times (52 - 51)$

$= 51.75$

sample standard deviation $= \sqrt{\frac{1}{(n-1)}(\Sigma x^2 - n\bar{x}^2)}$

$= \sqrt{\frac{1}{7}(18437 - 8 \times (47.625)^2)}$

$= 6.457$

d) from ⓐ and ⓒ, we see that Median before and outliers are same i.e. 47.5. This signifies that the outliers does not affect the median.

Q.No.5) Solution:

According to the question

=)

Page
Date

Page
Date

X	Y	X^2	Y^2	XY
6	40	36	1600	240
7	55	49	3025	385
7	50	49	2500	350
8	41	64	1681	328
10	17	100	289	170
10	26	100	676	260
15	16	225	256	240
$\Sigma X = 63$	$\Sigma Y = 245$	$\Sigma X^2 = 623$	$\Sigma Y^2 = 10027$	$\Sigma XY = 1973$

a) Response variable are those variables that we are trying to measure and explain. In our case, response variable is processed requests.

fitting simple regression line,

To fit $Y = a + bx$ —— (*)

At first, we have,

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

$$245 = 7a + 63b$$

or, $7a + 63b = 245$ —— (i)

secondly, we have

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

or, $1973 = 63a + 623b$ —— (ii)

Solving (i) & (ii),

$a = 72.28$ and $b = -4.14$

Scanned by CamScanner

Page
Date

plugging in the value of 'a' and 'b' in (*),

$Y = 72.28 - 4.14x$ ——— (A)

b) The regression coefficient of the given problem is -4.14. This means that Y decreases 4.14 per unit change in x.

c) for correlation coefficient,

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \cdot \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{7 \times 1973 - 63 \times 245}{\sqrt{7 \times 623 - (63)^2} \cdot \sqrt{7 \times 10027 - (245)^2}}$$

$$= \frac{-1624}{14\sqrt{2} \times 22\sqrt{21}}$$

$\therefore r = -0.813$

for coefficient of determination,

$R^2 = r^2$

$= (-0.813)^2$

$= 0.661$.

d) Given,

when, $x = 30$

Then, from (A),

$Y = 72.28 - 4.14 \times 30 = -51.92$

It is not possible to predict efficiency for data size of 30 gigabytes because Y is negative.

Scanned by CamScanner

Page
Date

when $x = 12$,
$$Y = 72.28 - 4.14 \times 12 = 22.6.$$

5) solution:

a) for dataset A,
$$\bar{X} = \frac{\Sigma X}{n} = \frac{439}{30} = 14.633$$

$$s = \sqrt{\frac{1}{n-1} \Sigma (X-\bar{X})^2} = \sqrt{\frac{1}{29} \times 1348.97} = 6.82$$

$$C.V = \frac{s}{\bar{X}} \times 100\% = \frac{6.82}{14.633} \times 100\% = 46.6\%.$$

Mean is less than median, so the data is negatively skewed.

for dataset B,
$$\bar{X} = \frac{\Sigma X}{n} = \frac{625}{30} = 20.833$$
$$s = \sqrt{\frac{1}{n-1} \Sigma (X-\bar{X})^2} = \sqrt{\frac{1}{29} \times 540.17} = 4.31$$

Page
Date

$$C.V = \frac{s}{\bar{x}} \times 100\% = \frac{4.31}{20.833} \times 100\% = 20.68\%$$

Mean is less than median, so the data is negatively skewed.

For dataset C,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1239}{30} = 41.3$$

$$s = \sqrt{\frac{1}{29} \times 12836.3} = 21.038$$

$$\therefore C.V = \frac{s}{\bar{x}} \times 100\% = \frac{21.038}{41.3} \times 100\% = 50.9\%$$

Mean is greater than median, so the data is positively skewed.

On comparing three data set the variability is more in dataset C and less in dataset B.

b) On the basis of box-plots, since the left whisker of dataset A and B are larger than right whisker, these two dataset are negatively skewed whereas data set C has longer right whisker, so it is positively skewed. The shape of distribution on the basis of box-plot is found to be same as calculation done in (a).

6) Solution:

(a) Let D be the event that ~~defect~~ represents the defective rate of the device.

Then, According to question,

$P(D) = 0.03 = p$.

no. of item $(n) = 20$.

$P(X \geq 1) = ?$

Now,

$p + q = 1$.

or $q = 1 - p = 1 - 0.03 = 0.97$

Then, $x \sim B(20, 0.03)$ and $n = 20$.

The pmf of Binomial Distribution is given by

$$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}.$$

Now,

$$P(X \geq 1) = 1 - P(X=0)$$

$$= 1 - {}^{20}C_{0}\,(0.03)^{0} \cdot (0.97)^{20}$$

$\therefore P(X \geq 1) = 0.46$

(b)

Page
Date

7) solution:

let x = Number of message.

Average number of message per hour $(\lambda) = 9$

Its pmf is $p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

a) Probablity of recieving at least five messages during the next hour

$P(X \geq 5) = 1 - \{P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)\}$

$= 1 - \{\frac{e^{-9}9^{0}}{0!} + \frac{e^{-9}9^{1}}{1!} + \frac{e^{-9}9^{2}}{2!} + \frac{e^{-9}9^{3}}{3!} + \frac{e^{-9}9^{4}}{4!}\}$

$= 1 - e^{-9}\{\frac{9^{0}}{0!} + \frac{9^{1}}{1!} + \frac{9^{2}}{2!} + \frac{9^{3}}{3!} + \frac{9^{4}}{4!}\}$

$\therefore P(X \geq 5) = 0.945$

b) Probablity of recieving exactly seven message during the next hour:

$P(X=7) = \frac{e^{-9}9^{7}}{7!}$

$\therefore P(X=7) = 0.117$

8) solution:

Given,

$$f(x) = \begin{cases} C(10-x)^2, & 0 < x < 10 \\ 0, & \text{Otherwise} \end{cases}$$

Scanned by CamScanner

we know,

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

or, $$\int_{0}^{10} c(10-x^2)\, dx = 1.$$

or, $$c\left[10x - \frac{x^3}{3}\right]_0^{10} = 1$$

or, $$c\left[\left(10\times10 - \frac{10^3}{3}\right) - \left(10\times0 - \frac{0^3}{3}\right)\right] = 1.$$

or, $c = ..$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

$$c\int_{0}^{10} (10-x)^2\, dx = 1.$$

or, $$c\int_{0}^{10} (100 - 10x + x^2)\, dx = 1.$$

or, $$c\left[100x - \frac{10x^2}{2} + \frac{x^3}{3}\right]_0^{10} = 1.$$

or, $$c = \frac{3}{1000}$$

Page
Date

Now,

$$P(1 \le x \le 2) = \int_1^2 c(10-x)^2\,dx$$

$$= \frac{3}{1000}\int_1^2 (100 - 20x + x^2)\,dx$$

$$= \frac{3}{1000}\left[100x - \frac{20x^2}{2} + \frac{x^3}{3}\right]_1^2$$

$\therefore P(1 \le x \le 2) = 0.217$

q) solution:

According to the question:

R_1	R_2	$d = R_1 - R_2$	d^2
5	10	-5	25
2	5	-3	9
9	1	8	64
8	3	5	25
1	8	-7	49
10	6	4	16
3	2	1	1
4	7	-3	9
6	9	-3	9
7	4	3	9
			$\Sigma d^2 = 216$

Then,

$$R = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 216}{10 \times 99}$$

$\therefore R = -0.309$

since, the correlation is negative between Dell and HP. so, if one of them increases, another decreases.

10) A continuous random variable X assuming non negative values is said to follow an exponential distribution with parameter $\lambda > 0$, if its probablity density function is given by

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

solution:

let X = time of job.

$\lambda = 3/hr$

i) we have,

expected value $(E(X)) = \frac{1}{\theta} = \frac{1}{3}$ hours = 20 minutes betwe^n job.

ii)

5 min $= \frac{1}{12}$ hrs.

Page
Date

now,

$$P\left(X < \frac{1}{12}\text{ hrs}\right) = \int_0^{1/12} 3e^{-3x}\,dx = \left[\frac{3e^{-3x}}{-3}\right]_0^{1/12}$$

$$= -e^{-3/12} + 1$$

$$= 1 - e^{-1/4}$$

$$\therefore P\left(X < \frac{1}{12}\text{ hrs}\right) = 0.22$$

11) Solution:

let X be the lifetime of the electronic component. The ~~exponent~~ expectation of 5000 hours and a standard deviation of 1000 hours follows normal distribution i.e.

$$N \sim (5000,\ 100^2)$$

so,

Expectation $(E(x))$ = Mean $(\mu) = 5000$

standard deviation $(\sigma) = 100$

let z, be the standard normal variate, such that

$$z = \frac{X-\mu}{\sigma} = \frac{X-5000}{100} \quad \text{——— (i)}$$

=)

a) soln.

$P(X<5012) = P\left(\frac{X-5000}{100} < \frac{5012-5000}{100}\right)$

$= P(Z < 0.12)$

$= P(-\infty < Z < 0) + P(0 < Z < 0.12)$

$\therefore P(X<5012) = 0.548.$

b) $P(4000 < X < 6000) = P\left(\frac{4000-5000}{100} < Z < \frac{6000-5000}{100}\right)$

$= P(-10 < Z < 10)$

$$= \int_{-10}^{10} \frac{1}{6\sqrt{2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{6}\right)^2} dz$$

$$= \int_{-10}^{10} \frac{1}{6\sqrt{2}} e^{-\frac{1}{2}z^2} dz$$

$= \frac{1}{\cancel{}2}$

c) $P(X>7000) = P\left(\frac{X-5000}{100} > \frac{7000-5000}{100}\right)$

$= P(Z > 20)$

$$= \int_{20}^{\infty} \frac{1}{6\sqrt{2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{6}\right)^2} dz = 0.$$