

# Business Analytics

**BSMS2002**

TA sessions

Utkarsh Sahu ([21f1001107@ds.study.iitm.ac.in](mailto:21f1001107@ds.study.iitm.ac.in))

<https://github.com/utkarsh4tech/BSMS2002>

# Tag us on discourse

- |                                   |   |                |
|-----------------------------------|---|----------------|
| 1. Prof. Dr. Malolan Sundararaman | - | @Milo          |
| 2. Prof. Swaminathan              | - | @ram158        |
| 3. TA Utkarsh Sahu                | - | @SahuUtkarsh03 |

A photograph of a crowd of people outdoors. In the foreground, a woman with dark hair tied back, wearing a dark patterned button-down shirt, looks towards the camera. To her left, a man in a white shirt is partially visible. The background is filled with other people, some looking at their phones. The scene is set against a backdrop of green trees.

**Level Sabke Niklenge**

**Lekin Niklenege Uske Jo TA sessions attend Krega :)**

# Basic Information

Quiz 1: July 7th, 2024

Quiz 2: August 4th, 2024

End term: Sep 1st 2024

$$T = 0.2Q + 0.4A + 0.4F$$

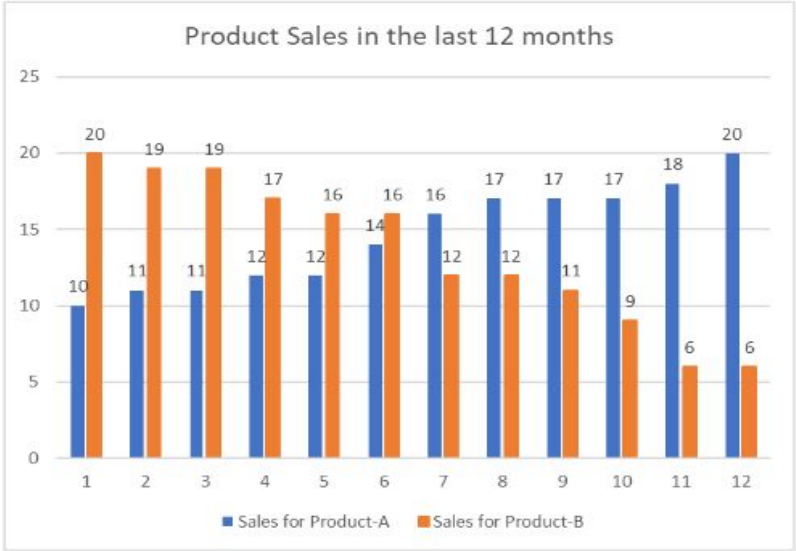
$$\text{Quiz Marks}(Q) = 0.7 * \text{Max}(Qz1, Qz2) + 0.3 * \text{Min}(Qz1, Qz2)$$

Assignment (A) = Sum of the Best 2 out of (Assignment 1, Assignment 2, Assignment 3)

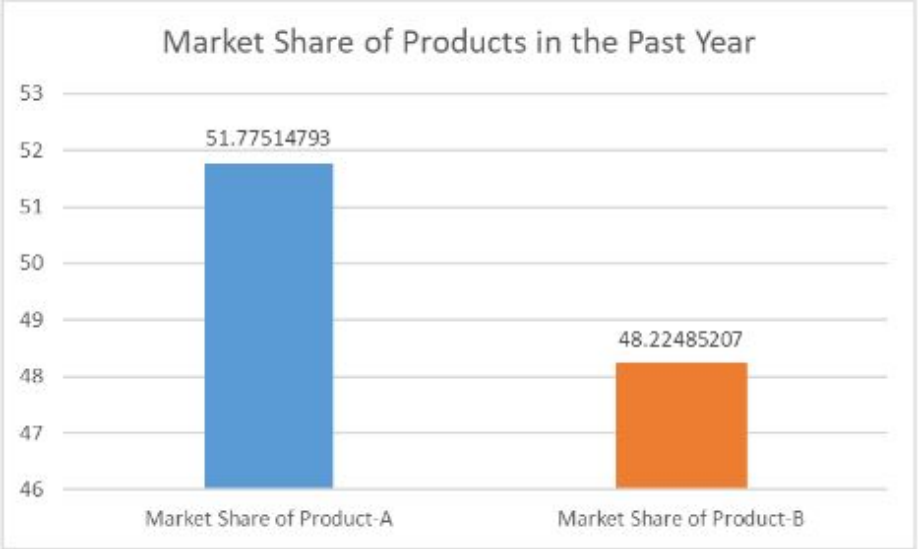
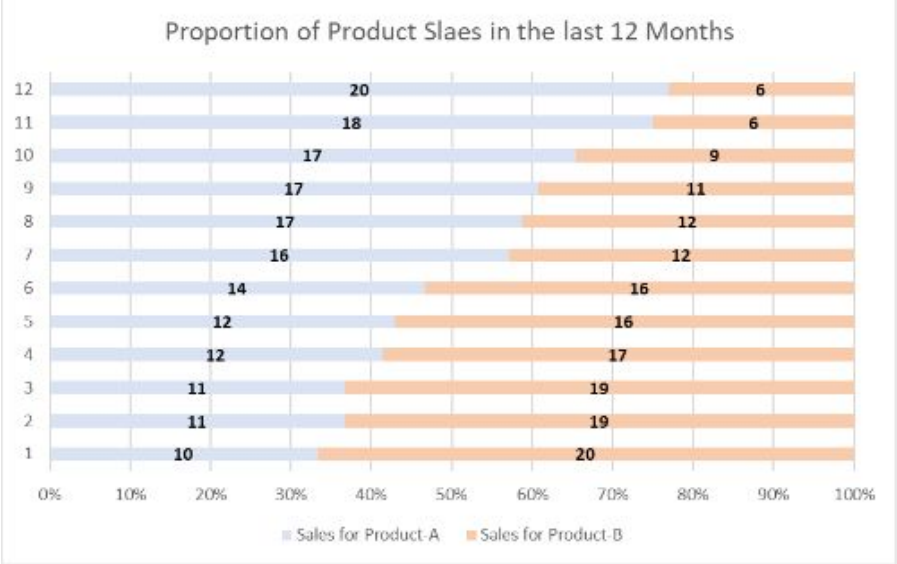
Final End Term (F) = Will be set to 45 marks and students can attempt all. Marks obtained will be capped at 40. The syllabus for the End term exam will be the contents covered in Weeks 1 to 12

Grading Doc: [Click here](#)

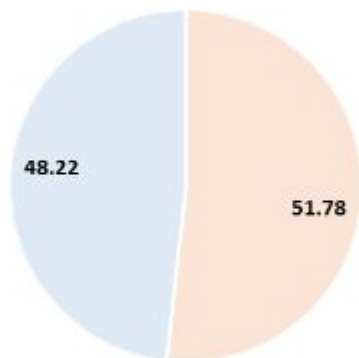
If the aim is to highlight the behaviour of the product sales in the past year, then which of the following visualizations would be best suited?



If the aim is to highlight the market share of the products in the past year, then which of the following visualizations would be best suited?

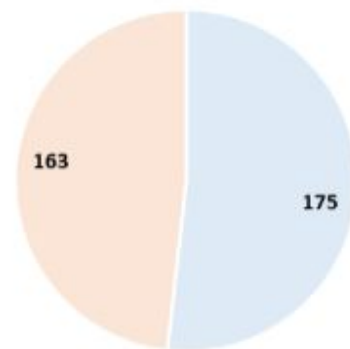


✓ Market Share of the Products in the Past Year



Market Share of Product-A Market Share of Product-B

Market Share of the Products in the Past Year



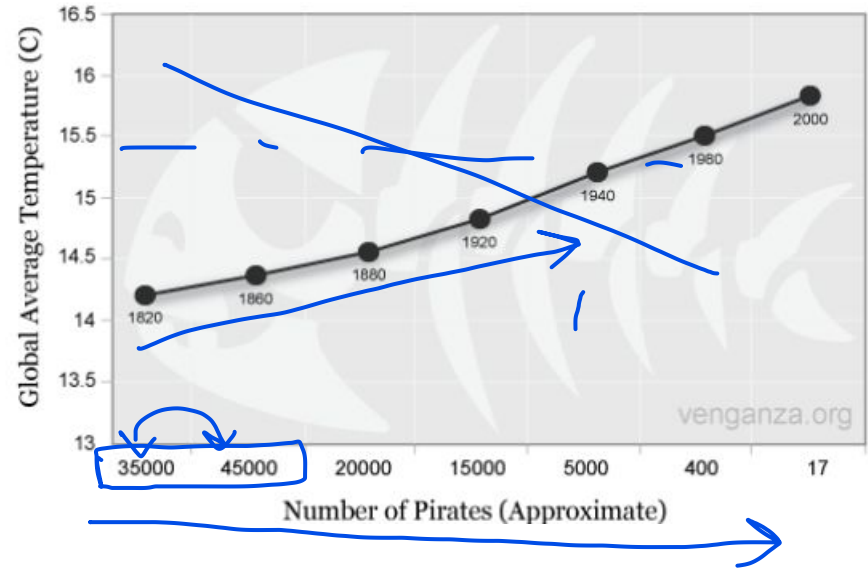
Market Share of Product-A Market Share of Product-B

From the given figure, which of the below statements are true?

Jan 24 @ 1

- a. The number of pirates does not increase with the global average temperature increase
- b. There is a positive correlation between the number of pirates and the global average temperature increase
- c. There is no correlation between the number of pirates and the global average temperature increase
- d. None of the above

Global Average Temperature Vs. Number of Pirates



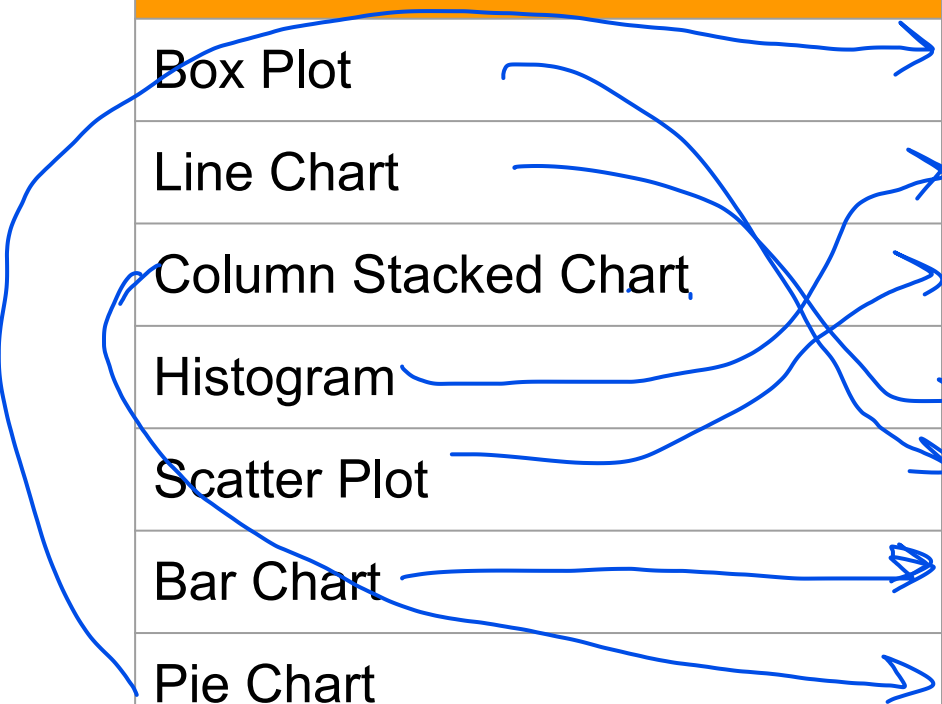
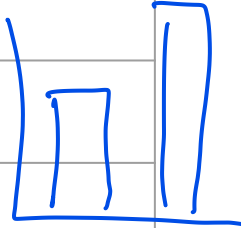


Match the charts to its most suitable purpose

Chart Type (1-7)	Purpose (A-F)
Box Plot	Proportion
Line Chart	To guess Distribution
Column Stacked Chart	Correlation Analysis
Histogram	Trend Analysis
Scatter Plot	Outlier Detection
Bar Chart	To do Comparison
Pie Chart	Compare parts of a whole

A

1,2 1,3  
2,3  
M F



Among the following graphs, which are not recommended/correct for use?

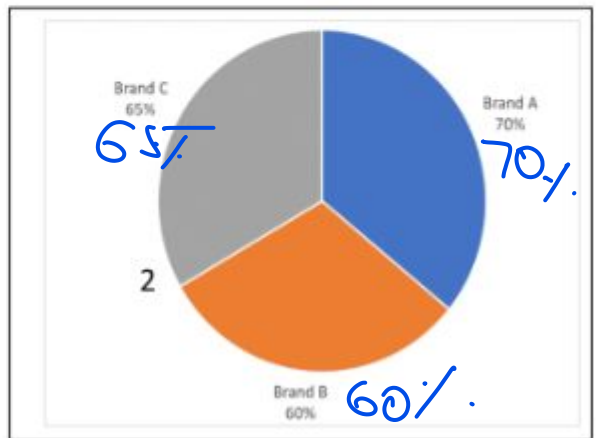
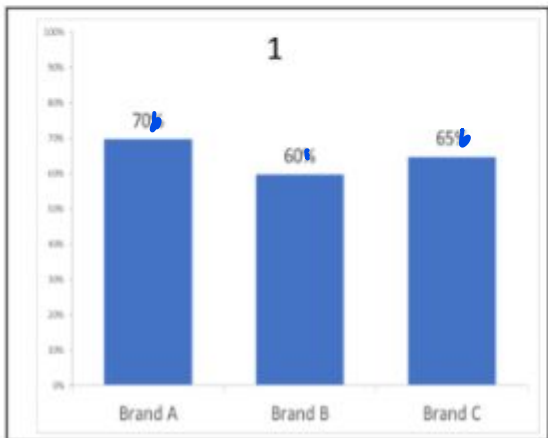
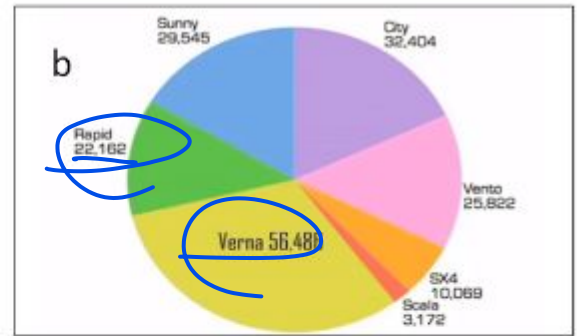
F1  
FL  
F3 2

A. (a) & (1)

B. (a) & (2)

C. (b) & (2)

D. (b) & (1)



W2

## Select the null hypothesis and alternative hypothesis for chi-square test of goodness of fit:

- ☒ a.  $H_0$ : The population follows the proposed distribution
- ☐ b.  $H_0$ : The population does not follow the proposed distribution
- ☒ c.  $H_a$ : The population does not follow the proposed distribution
- ☐ d.  $H_a$ : The population follows the proposed distribution

In the Goodness-of-Fit test, if the computed test statistic is greater than the tabulated value of the test statistic at a given significance level, then....

$\alpha = 95\%$

80%

- a. Reject the null hypothesis and conclude that the data does not come from population with proposed distribution
- b. ✓ At the specified significance level, reject the null hypothesis and conclude that the data does not come from population with proposed distribution
- c. Do not reject the null hypothesis and conclude that the data does not come from a population with proposed distribution
- d. At the specified significance level, do not reject the null hypothesis and conclude that the data does not come from a population with proposed distribution

The p-value of the chi-square goodness of fit test represents \_\_

- a. The chance of observing the sample when the null hypothesis is false.
- b. The chance of observing the sample when the alternative hypothesis is true
- c. The chance of observing the sample at the specified level of significance
- d. None of the above

If the p-value for the computed test statistic is 0.0524, then at which of the following significance levels, will you reject the null?

a. 1%

→ 0.01

b. 5%

→ 0.05

c. 10%

→ 0.1

d. 15%

→ 0.15

Reject  $H_0$

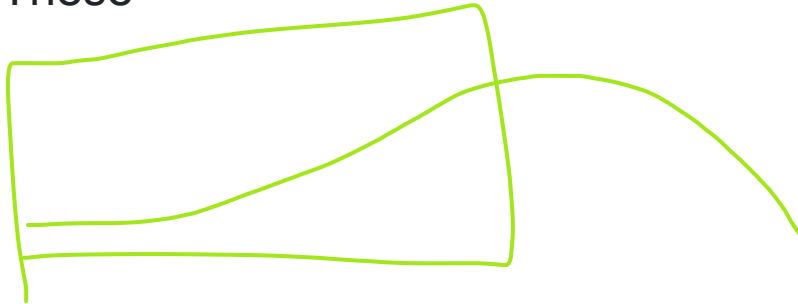
P val <  $\alpha$

0.0524

## A Distribution is Left Tailed if ...

MSQ

- a. Negative Skewness
- b. Left Tail implied mean is on Left Side
- c.  $\text{Mean} < \text{Median} < \text{Mode}$
- d. All of These



For a standard normal distribution, the Coefficient of variation (CV) is not defined properly.

a. True

$$C.V. = \sigma / \mu$$

~~No~~


b. False

c. Cannot Say, Depends on Case by Case Basis

$N \sim (0, 1)$





## Mean is zero and S.D is 1 for which Distribution

- a. Normal Distribution
- b. Poisson Distribution
- c. Uniform Distribution
- d.  Standard Normal Distribution

Let us assume the sample standard deviation to be 3.2, and the sample mean to be 10.7. What is the value coefficient of variation (CV)?

$$CV = \frac{3.2}{10.7}$$

## Which among the following is True?

-  a. QQ plot AMPLIFIES the differences in TAILS of model and sample distributions
- b. PP plot AMPLIFIES the differences in TAILS of model and sample distributions
- c. QQ plot AMPLIFIES the differences in the MIDDLE PORTION of model and sample distribution.
-  d. PP plot AMPLIFIES the differences in the MIDDLE PORTION of model and sample distribution.

A data set with 200 data points on 15 variables is provided. The chi-square goodness of fit test is conducted on a variable in the data-set, to see if it follows an exponential distribution. If the dataset is binned into 10 bins, then how many degrees of freedom does the chi-square statistic have?

a. 15

b. 12

c. 10

d. 8

Handwritten notes and calculations:

$$DOF = 10 - 1 - 1$$

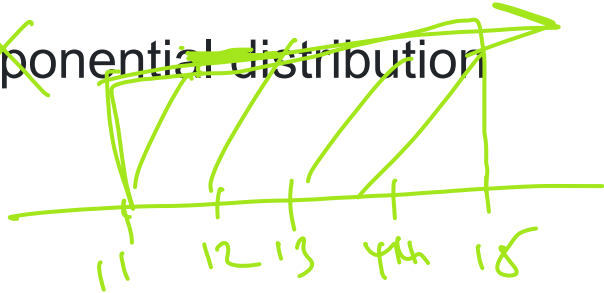
Diagram illustrating the calculation of degrees of freedom (DOF) for a chi-square goodness of fit test:

- A circle with a question mark (?) is connected by an arrow to the text  $N(\mu, \sigma^2)$ .
- Below the circle, the word "Para" is written.
- To the right of the circle, the calculation  $10 - 2 - 1$  is shown.

The table below provides the summary statistic for a random variable.  
Then what distribution could be a good fit for this random variable?  
(select all that is/are applicable)

MSD → Mo

- a. ~~Poisson distribution~~
- b. ~~Standard normal distribution~~
- c. ✓ Uniform distribution
- d. ~~Exponential distribution~~



$\mu = 12$   
 $\sigma = 2$

Summary Statistic	Value
Number of observations	300
Mean	12
Median	13
Mode	12
Std. Deviation	2
Minimum	11
Maximum	15
Skewness	0.01

The table below provides the summary statistic for a random variable. Then what distribution could be a good fit for this random variable? (select all that is/are applicable)

- a. ☒ Poisson distribution
- b. ☐ Standard normal distribution
- c. ☐ Uniform distribution
- d. ☐ Exponential distribution

Summary Statistic	Value
Number of observations	300
Mean	4.94
Median	5.00
Std. Deviation	2.22
Minimum	1
Maximum	10
Skewness	0.71

$\mu = \sigma^2$

#P=4    #C=3

Ms. X is working is with the data given in Table-1 below.

Product D has 9 obs. val.

$n = 2$

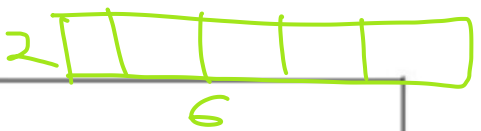
Product	Sales of a Product in a City for a given Year								
	City-1	City-2	City-3	City-1	City-2	City-3	City-1	City-2	City-3
	(1990)	(1990)	(1990)	(1991)	(1991)	(1991)	(1992)	(1992)	(1992)
A	100	90	250	120	50	120	140	20	500
B	145	300	500	175	250	250	195	230	1000
C	90	180	30	100	110	15	110	95	58
D	130	220	132	140	200	61	150	180	270

Table-1

130, 220

$\frac{2}{36} = \frac{1}{18}$

Say you want to see if the distribution of sales of **Product-D** in the Table-1 follows a **uniform distribution** within the range of 0 to 300 when split into bins as specified in Table-2.



Bin Number	Bin Range
Bin-1	Sales value less than or equal to 50
Bin-2	Sales value greater than 50 but less than or equal to 100
Bin-3	Sales value greater than 100 but <u>less than or equal to 150</u>
Bin-4	Sales value greater than 150 but less than or equal to 200
Bin-5	Sales value greater than 200 but less than or equal to 250
Bin-6	Sales value greater than 250 but less than or equal to 300

Table-2



Q

Then what is the expected frequency in any given bin?

- a. 1
- b. 1.5
- c. 2

Dist?  $\rightarrow$  Uniform

$$= \frac{\text{Total Obs. Val}}{\# \text{ Bin}}$$

$$= \frac{9}{6}$$

If a Chi-square goodness-of-fit test is run to check if the data for product D in Table-1 follows uniform distribution with bins as specified in Table-2, then what is the value of test statistics?

✓ a. 6.20 to 6.40

b. 7.50 to 7.80

$$\begin{array}{r} 0.62 \\ \hline 1.5 \end{array}$$

$$6 + 334$$

Bin	Per	Obs.f.	Exp.f.	$(E-O)/E$
1		0	1.5	1.5
2		1	1	0.167
3		4	1	0.167
4		2	1	0.167
5	1	1	1	0.167
6		1	1	0.167
6				

~~A~~ Chi-squared Goodness-Of-Fit test with the bins as specified in Table-2 is going to be carried out to check if the data on sales (whole data in Table-1) is normal or not. Then what is the degrees of freedom for the test?

$$DoF = K - P - 1$$

$$6 - 2 - 1$$

a. 5

b. 4

~~c. 3~~



A Chi-squared Goodness-Of-Fit test with the bins as specified in Table-2 and if you are told that the data on sales comes from a parent normal distribution  $\sim N(100, 100)$  then what is the degrees of freedom for the test?

a. 5

b. 4

c. 3

$\mu, \sigma^2$  has  $\sigma^2 = 100$   
 $\mu = ?$

$$K - p - 1$$
$$6 - 0 - 1$$

5

H2W

A company sells two products (A and B) in two different marketplaces. In the first marketplace, the sales are made at the Forenoon (FN). In the second marketplace, sales are made during the Afternoon (AN). Table-2 provides the data on the sales in the two marketplaces for the two products in the first 10 days on January 2023. Using this data, answer the following 6 questions

Date	Time of Day	Number of units of Product A Sold	Number of units of Product B Sold
01-01-2023	FN	31	30
02-01-2023	FN	21	41
03-01-2023	FN	16	45
04-01-2023	FN	7	56
05-01-2023	FN	13	49
06-01-2023	FN	8	40
07-01-2023	FN	8	42
08-01-2023	FN	21	44
09-01-2023	FN	13	55
10-01-2023	FN	7	42
01-01-2023	AN	31	30
02-01-2023	AN	21	41
03-01-2023	AN	16	52
04-01-2023	AN	7	57
05-01-2023	AN	13	49
06-01-2023	AN	8	55
07-01-2023	AN	8	57
08-01-2023	AN	21	44
09-01-2023	AN	13	52
10-01-2023	AN	7	62

**If the total sales of both products (A & B) on any given day are expected to be uniformly distributed across both marketplaces, then what is the expected total sales on any given day?**

**To test the hypothesis that the total sales of both products (A & B) are uniformly distributed across both marketplaces (FN & AN), a chi-square goodness-of-fit test is conducted. If the days are taken as the buckets, then what is the value of the computed test statistic?**

**In the Goodness-of-Fit test, if the computed test statistic is greater than the tabulated value of the test statistic, then (choose all that is applicable)**

- a. Reject the null hypothesis and conclude that there is strong evidence that the data does not come from a population with a specified distribution
- b. At the specified significance level, reject the null hypothesis and conclude that there is strong evidence that the data does not come from a population with a specified distribution
- c. Do not reject the null hypothesis and conclude that there is strong evidence that the data does not come from a population with a specified distribution
- d. At the specified significance level, do not reject the null hypothesis and conclude that there is strong evidence that the data does not come from a population with a specified distribution
- e. None of these



**In the Goodness-of-Fit test, if the days are taken as the buckets, what is the number of degrees of freedom if the total sales of both products (A & B) across both markets (FN & AN) is assumed to be normally distributed?**

**The question of whether sales of products A and B are independent across the different markets is being checked. Then what is the value of the computed test statistic?**

**The question of whether sales of products A and B are independent across the different markets is being checked. Then what is the number of degrees of freedom for the test?**