Statistics Project

* Statistics –

Statistics is the science concerned with developing and studying methods for collecting, analysing, interpreting and presenting empirical data.

1)Probability –

Type of Events :-

1. Independent Event ( Multiple Rule)
2. Mutually Exclusive Event (Addition Rule)

Probability Techniques:

1. Bayes' Theorem
2. Bionomial Distribution
3. Poissons' Distribution
4. Gaussuan Distribution

2) Descriptive Statistics -

* Fundamentals of Descriptive Statistics

Measures of Central Tendency (Mean, Median, Mode)

Measures of Variability (Range, Variance, Standard Deviation, percentile)

Measures of Distribution Shape (Skewness, Kurtosis)

3) Inferential Statistics -

* Hypothesis Testing –

1. T – test
2. Z – test

1)Probability –

Probability is simply how likely something is to happen. Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes how likely they are. The analysis of events governed by probability is called statistics.

Type of Events :-

1. Independent Event (Multiple Rule)
2. Mutually Exclusive Event (Addition Rule)

* Independent Event (Multiple Rule)

These are events where the occurrence of one event does not affect the probability of the others. These are

Formula -

P(A and B) = P(A)\*P(B)

* Event A :- It will rain in Mumbai tomorrow.
* Event B :- It will not rain in bengalore
* Mutually Exclusive Event (Addition Rule) –

These are events that cannot happen at the same time.

Formula –

P(A or B) = P(A)+P(B) – P(A/B)

* Event A :-Indian team got all out within 100 runs.
* Event B :-Virat Kohli scored double century.

Examples –

1. What is the probability of flipping a coin & getting heads, and then rolling a dice and getting 3 no it.

Ans –

Independent Event (Multiple Rule)

P(A and B) = P(A)\*P(B)

= (1/2) \* (1/6)

=0.08

1. What is the probability of rolling a 2 or a 4 in a dice.

Ans –

Mutually Exclusive Event (Addition Rule)

P(A or B) = P(A)+P(B) – P(A/B)

= ((1/6)+(1/6))-((1/6)/(1/6))

=0.33

1. Nykaa is running a marketing campaign and wants to find the probability of a customer either responding to an email offer or a social media ads.

Probability of responding to an email offer - 0.3

Probability of responding to an social media ads – 0.4

Probability of responding to both – 0.1

Ans –

Mutually Exclusive Event (Addition Rule)

P(A or B) = P(A)+P(B) – P(A/B)=0.6

Therefore , the probability of a customer responding to either email or social media ads is 0.6

1. Amazon wants to determine probability that an order will arrive on time and be in perfect condition

On time – 0.9

Perfect condition - 0.95

Ans –

Independent Event (Multiple Rule)

P(A and B) = P(A)\*P(B)

= 0.9\*0.95

=0.86

The probability of an order arriving on time & in perfect condition is 0.86

Probability Techniques:

1) Bayes' Theorem

2) Bionomial Distribution

3) Poissons' Distribution

4)Gaussuan Distribution

* Bayes' Theorem –

Bayes' Theorem It is used to find the probability when you know certain other probability.

Formula –

Probability of Event A Happening, when Event B is given

1. P(A/B) =( P(B/A) \* P(A))/P(B)

Probability of Event B happening, when Event A is given

1. P(B/A) =( P(A/B) \* P(B))/P(A)

Event A: Virat Kohli will 200 runs

Event B: Indian will get all out 350

Examples –

1. A company wants to assess the probability that a sales team member is highly skilled based on receiving a positive feedback from a customer.

Given Data:

25% of the sales team members are highly skilled. 90% of highly skilled team members received positive feedback. 70% of all sales team What is the probability that a sales team member is highly skilled given that they received positive feedback ?

Ans –

Event A: Skilled team member P(A) = 0.25

Event b: Positive feedback P(B) = 0.70

P(B|A) = 0.90

P(A/B) =( P(B/A) \* P(A))/P(B)

=(0.90\*0.25)/0.70

=32%

If somebody have received a positive feedback, that doesn't necessarily means those employees are highly skills. There is only 32% chance of being highly skilled in that case.

If 100 positive feedbacks have been received by our employees, in that only 32 employees can be considered has highly skilled.

1. Suppose a patient goes to a doctor with Covid symptom.

The disease is rare, affecting 2% of the population. If a patient has the disease, there is 90% chance they will have this symptom. However, 10% of healthy individuals also exhibits this symptom.

The doctor wants to know the probability that the patient actually has the disease given they have the COVID symptoms.

Ans –

Event A: Disease P(A) = 0.02

Event b: symptoms P(B) = 0.10

P (B|A) = 0.90

P(A/B) =( P(B/A) \* P(A))/P(B)

=18%

Conclusion :

Even with the symptom present, the probability that the patient has COVID is about 18% .

1. Tira running an email marketing campaign and wants to understand the likelihood of a customer making a purchase after opening an email. 10% of customers makes a purchase, and 25% of all customers opens the marketing email. 40% of customers who make a purchase open the marketing email. What is the probability that a customer makes a purchase, given that they opened the marketing email ?

Ans –

Event A: Customer opens the marketing email P(A) = 0.25

Event B: Customer makes a purchase P(B) = 0.10

P(A|B) = 0.40

P(B/A) =( P(A/B) \* P(B))/P(A)

=16%

Conclusion :

If a customer has opened the marketing email , there is a 16% chance of making a purchase.

1. Boat

Event A: The customer buys a product

Event: The customer received a marketing email.

Prob that a customer buys a product from BOAT is 10% P(A) = 0.10

Prob that a customer received a marketing email, who have bought the product is 70% P(B|A) = 0.70

Prob that a customer receives a marketing email regardless of purchase is 30% P(B) =0.30

What is the probability that a customer bought a product given that they received a marketing email?

Ans – P(A/B) =( P(B/A) \* P(A))/P(B)

= 23%

Conclusion : There is a 23% percent chance that the customer bought a product, after they received marketing email.

* Bionomial Distribution -

It is a way of finding probability that helps us understand the likelihood of a certain number of success in a fixed number of trial. It applies to only those events which have only two possible outcomes: success or failure

Formula –

**P(K) = nCk\*Pk (1-P)n-k**

Example –

1. Dharmendra is a tossing a coin 5 times. Tell me the probability of getting eaxctly 3 heads.

Ans –

No. of times – (n) = 5

success : heads (P) = ½

No of success (k) = 3

**P(K) = nCk\*Pk (1-P)n-k**

=  5C3\* (0.5)3(1-0.5)5-3

=0.31%

1. Zomato send out 1000 promotional emails. Based on past campaign, we know that the probability of a customer making a purchase after receiving the email is 10%. What is the probability that exactly 120 customer will purchase.

Ans –No. of trials ( n) = 1000

p = 0.10

k = 120

**P(K) = nCk\*Pk (1-P)n-k**

=0.47%

The likelihood/ probability of exactly 120 customer making a purchase is 0.47%.

Variability –

It helps us understand the ups & downs around the average.

Formula - Variance = n \* p \* (1-p)

Example –

1. A retail chain is running a promotion, where they know based on the past data, that 20% of customers who walk inside the store will male a purchase Each day, about 100 customers visit the store.

Ans –

n=100

p=0.20

Variance = n \* p \* (1-p)

=100\*0.20\*(1-0.80)

=16%

|  |  |
| --- | --- |
| Monday | 20 |
| Tue: | 16 |
| Wed | 25 |
| Thurs: | 10 |
| Fri | 30 |

Conclusion :

The variability is 16. This tells the company that while 20 buyers per day is the average the number could vary. =+-16

1. Jio wants to understand how many of their support calls will lead to a successful resolution on the first try. From the past data, they know that 60% of calls are resolved right away on first call itself. They receive around 200 calls each day. Find the variability.

Ans- p = 0.60

n = 200

Variance = n \* p \* (1-p)

=48%

=+-48

1. Poissons' Distribution –

It is a probability distribution that helps us find out number of times an event can occur, and the avg rate is know.

Formula -

P(k=λ) = (e-λ λx )/k!

E=2.718

K=atual no to find

λ =given average

Examples –

1. A company receives 70 emails per day on average. What is the probability that exactly 40 emails will be received on any day.

Ans -

Lamda =70

k = 40

P(k=λ) = (e-λ λk )/k!

= ((2.718)^-70\*(70)^40)/(40) !

=0.000031

1. In a factory, a machine fails 2 times a day on average. What is the probability that it will fail exactly 3 times in a day.

Ans -

Lamda=2

k=3

P(k=λ) = (e-λ λk )/k!

= ((2.718)^-2\*(2)^3)/(3) !

=0.18

1. Gaussuan Distribution –

It is a statistical method to find out how data/ things is spread out in a range.

Formula –

Z score formula

Z = (range value - μ)/ s.d

Example –

1) In a class of 100 students, the average score of an exam is 70, with a standard deviation of 10. What % of students scored between 60 & 80.

Ans –

average score (mean μ) = 70

S.d. = 10 \*Z score and percentile

For 60 Z = (60 - μ)/ s.d =(60-70)/10 =-1 15.87

For 80 Z = (80 - μ)/ s.d =(80-70)/10 =1 84.13 60 70 80

=15.87-84.13 15.87 ?

=68.26% 84.13

68.26 % of students scored between 60 to 80.

2) The avg height of Indian men is 175 cm, with a standard deviation of 8 cm. What is probability that a randomly selected man is taller than 183 cm ?

Ans –

average score (mean μ) = 175

S.d. = 8 \*Z score and percentile (google)

For 183 Z = (183 - μ)/ s.d =(183-175)/8 =1 84.13

=100-84.13 175 183

=15.87% 84.13 ?

15.87 % man are taller then 183cm 100

Inferential Statistics -

* Hypothesis Testing –

It is a statistical method used to confirm whether the sample mean can be applied to the population. Hypothesis Testing gives us the evidence to whether accept or reject the assumptions/ hypothesis.

Two Types of Hypothesis :

Null Hypothesis (H0) - The sample mean correctly represents the population mean.

Alternate Hypothesis (Ha) - The sample mean doesn't correctly represents the population mean

Null Hypothesis – Reject - Accept

Alternate Hypothesis - Accept - Reject

**Steps to be followed in Hypothesis Testing :**

1. Step 1: Decide on the business probem/ assumption/ hypothesis
2. Step 2: Define Null & Alternate Hypothesis (H0 & Ha)
3. Step 3: Perform Statistical calculation Z – test T – test
4. Step 4: Accept/ reject the H0
5. Step 5: Conclusion

Ques. Swiggy claimed that its total valuation in Dec 2023 was $14 bn.

Scenario 1 Null Hypothesis (H0): Total Valuation = $14 bn Two-tailed test

Alt Hypothesis (Ha): Total Valuation != $14 bn

The rejection is on both the sides of the normal distribution curve.

Scenario 2 Null Hypothesis (H0): Total Valuation => $14 bn One-tailed test

Alt Hypothesis (Ha): Total Valuation < $14 bn

Left-tailed test The rejection area is on the left side the ND curve

Scenario 3 Null Hypothesis (H0): Total Valuation <= $14 bn One-tailed test

Alt Hypothesis (Ha): Total Valuation > $14 bn

Right-tailed test The rejection are is on the right side of ND curve.

Two tailed test

Z-test

LOS : Level of Significance

|  |  |  |
| --- | --- | --- |
|  |  | Light/ Right |
| LOS (α) | Two-tailed test | One-tailed |
|  |  |  |
| 1% | 1.25 | 2.32 |
| 5% | 1.96 | 1.64 |
| 10% | 1.64 | 1.28 |

Confidence Level ('c) 99% 90%

Significance Level (α) 1% 10%

* Performing Statistical Calculation:

Z-test T-test

When to use: When to use:

1. When the sample size is large 1. When the sample is small

n >= 30 (n <30)

2. When the population S.D(σ) is known 2. When the population S.D (σ) is unknown

Example –

1. Kellogs manufacture Women Horlicks. They claim that their product is having average calories content as 200 calories, as advertised. You're a quality analyst who wants to test if this claim is correct or not. You took out a sample of 10 bottles, and found the calories in each: Kellogs claims this at 95% confidence level.

|  |  |
| --- | --- |
| 6th bottle | 200 |
| 7th bottle | 204 |
| 8th bottle | 203 |
| 9th bottle | 199 |
| 10th bottle | 201 |

Ans –

|  |  |
| --- | --- |
| 1st bottle | 197 |
| 2nd bottle | 202 |
| 3rd bottle | 205 |
| 4th bottle | 198 |
| 5th bottle | 195 |

Popualation Sample

Mean (μ) = 200 Mean (x) = 200.4

LOS = 5% n = 10

S.D. (s) = 3.20

T-test

Null Hypothesis (H0): The mean calories content is 200 calories μ = 200

Alternate Hypothesis (Ha): The mean calories content is not equal to 200 calories μ !=200

Two-tailed test

Calculate the T-test :

t = ( x - μ) / (s/ √n)

= (200.4 - 200) / (3.20/ (10)^1/2)

= 0.39

Determining the Critical Values :

Degrees of freedom (DF) = n -1

=10-1

=9 accept

LOS (α) =0.05

\*critical value calculator use: -2.26 0.39 2.26

After putting values DF=9 and LOS (α) =0.05

=2.26

Since, 0.39 < 2.26 (C.V), we will accept the Null Hypothesis (H0)

Conclusion: There is no significant difference between the sample & the advertised mean of 200 calories by Kellogs.

1. Ambuja Cement claims that the average weight of its cement bags is 1 kg, with a population standard deviation of 0.01 kg.

They claim it with 95% confidence level.

Ujjwal, the quality analyst inspects the cement bags to test the claim.

He took a sample of 50 bags & finds out the avg weight to be 0.995kg.

Ans –

Since, n > 50, Z-test

H0: The mean weight of cement bags is 1kg μ = 1 reject

Ha: The mean weight of cement bags is 1kg μ != 1

Population Sample

Mean (μ) = 1 Mean (x) = 0.995

S.D (σ) = 0.01 n = 50 -1.96 1.96

|  |  |  |
| --- | --- | --- |
|  |  | Light/ Right |
| LOS (α) | Two-tailed test | One-tailed |
|  |  |  |
| 1% | 1.25 | 2.32 |
| 5% | 1.96 | 1.64 |
| 10% | 1.64 | 1.28 |

LOS = 5%

Two - tailed test

Perform the Z-test

Z=(x - μ)/ (σ/ √n)

= (0.995 - 1) / (0.01 / (50)^1/2)

=-3.53

Since, -3.53 < CV , we reject the Null Hypothesis (H0)

Conclusion: There is significant difference between the sample mean & the Ambaji’s claim of having avg weight of 1 kg of cement bags.

Types of Error :

Type -1 Error: H0 is true, but we rejected it

Type -2 Error: H0 is false, but we accept it.

H0: The avg mileage for Kia car is 25kms/ ltr

Ha: The avg mileage for Kia car is not 25kms/ ltr

WhatsApp link:

<https://t.ly/xdmUc>

LinkedIn link:

<https://www.linkedin.com/in/yuvaraj-shivamurti-13929228b?lipi=urn%3Ali%3Apage%3Ad_flagship3_profile_view_base_contact_details%3B5AmZ0FEZRLedpoaOwux5xA%3D%3D>