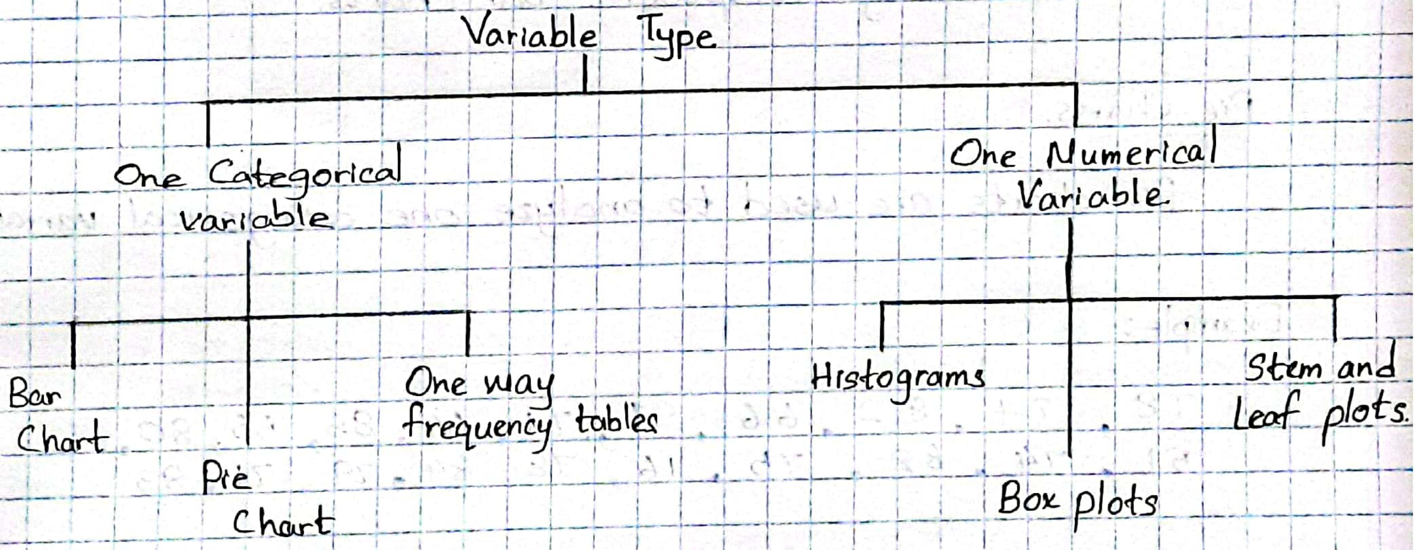
Descriptive Statistics (3)

- There are two methods that you can use under exploratory analysis. They are;

  - Graphical Methods
  - Numerical Methods.

Graphical Methods :-

- Can use to analyze both categorical and numerical variables.
- Type of graph you use depends on the type of the data available.



One wgay Frequency Tables.

Categorical Variable.

Gender	Frequency.
Male	48
Female	52

Numerical Variable

Marks	Frequency
0 - 20	12
21 - 40	8
41 - 60	42
61 - 80	56
81 - 100	10

Bar Charts.

There are several type of bar charts. For examples.

* Simple Bar Charts.
* Component bar charts / Stacked bar charts.
* Multiple bar charts / Clustered bar charts.
* Precentage component bar charts.

Pie Charts.

Pie charts are used to analyze one categorical variable

Example :-

78, 74, 82, 66, 91, 71, 64, 88, 55, 80,
51, 74, 82, 75, 16, 78, 84, 79, 71, 83

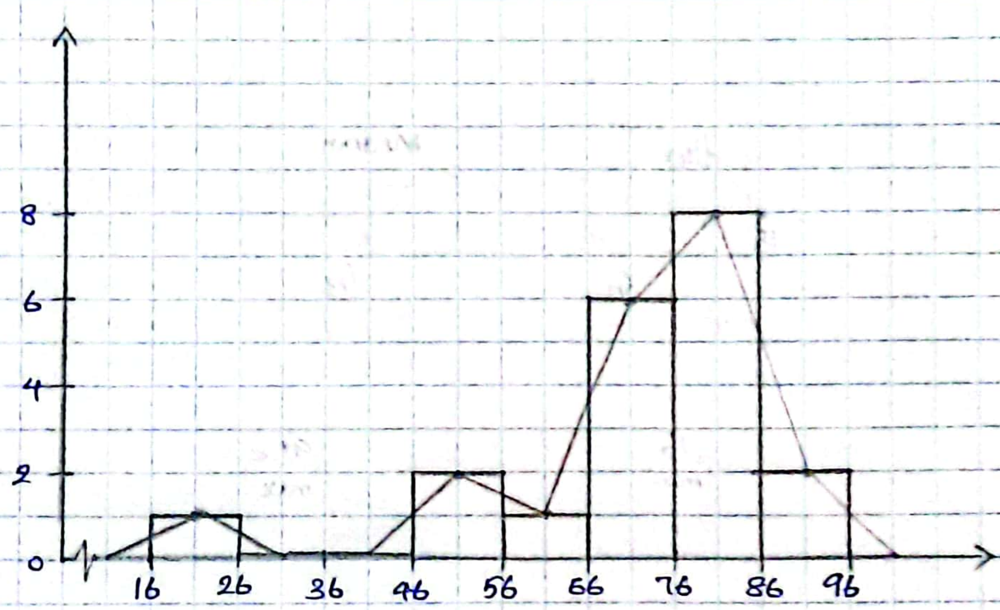
Range = Max - Min
= 91 - 16 = 75

Divide the range into required number of classes to find class width (Eg :- 8) :-

$$\text{Width of the class} = \frac{\text{Range}}{\text{No. of intervals.}}$$

$$= \frac{75}{8} = 9.375 \simeq 10$$

			frequency.
16 - 26	15.5 - 25.5	/	1
26 - 36	25.5 - 35.5		0
36 - 46	35.5 - 45.5		0
46 - 56	45.5 - 55.5	//	2
56 - 66	55.5 - 65.5	/	1
66 - 76	65.5 - 75.5	~~////~~ /	6
76 - 86	75.5 - 85.5	~~////~~ ///	8
86 - 96	85.5 - 95.5	//	2



Box Plots.

* To draw a box plot, it is need to identify the five number summary and outliers for the variables.

Five Number Summary.

* Minimum
* Maximum
* Q_1
* Q_2 (Median)
* Q_3

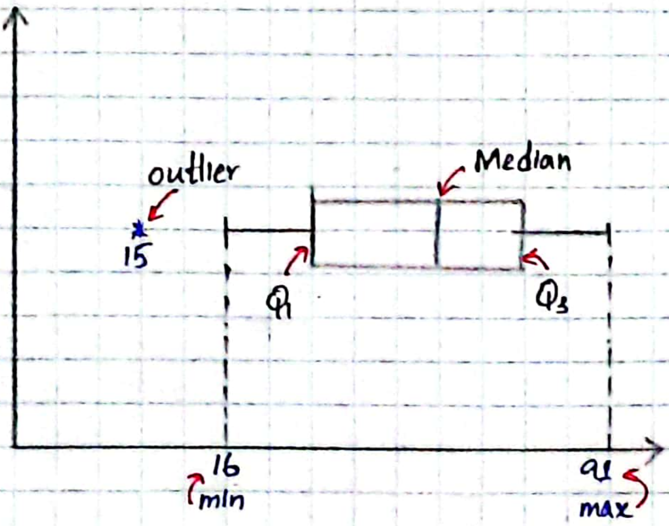
Outliers.

* Before drawing the box plot we should identify the potential outliers

$$\text{Upper bound} = Q_3 + 1.5 * IQR \quad (IQR = Q_3 - Q_1)$$
$$\text{Lower band} = Q_1 - 1.5 * IQR.$$

Values outside the range are consided as outliers and marked with ~~astrick~~ asterisks (*).

Q_1, Median and Q_3 are marked as a box.
Minimum and maximum values which are not outliers, will be end point for whiskers of the box plot.



78	74	82	66	91	71	64	88	55	80
51	74	82	75	16	78	84	79	71	83

$$\text{Mean} = \frac{\text{Sum of all data values}}{\text{Number of data values}} = \frac{\sum x_i}{n}$$

$$= \frac{1442}{20}$$

$$= \underline{\underline{72.1}}$$

Median → If n is odd $= \left(\frac{n+1}{2}\right)^{th}$ term

If n is even $= \dfrac{\left(\frac{n}{2}\right)^{th} \text{ term} + \left(\frac{n}{2}+1\right)^{th} \text{ term}}{2}$

$$= \frac{10^{th} \text{ term} + 11^{th} \text{ term}}{2}$$

$$\text{Median} = \frac{80 + 51}{2}$$

$$= \underline{65.5}$$

Mode = value of the highest frequency.

= 74, 78, 71, 82 ← All 4 are modes.

- $Q_i = \left[\frac{n+1}{4} \times i\right]^{th}$ value ← Quartile

- $D_i = \left[\frac{n+1}{10} \times i\right]^{th}$ value. ← Decile

- $P_i = \left[\frac{n+1}{100} \times i\right]^{th}$ value ← Percentile

- $IQR = Q_3 - Q_1$

- Variance of population,

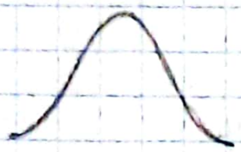
$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{N}$$

- Variance of a sample

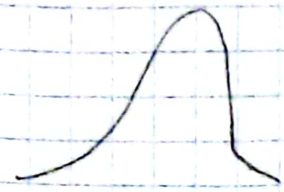
$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

- Population SD = σ (Standard deviation)
- Sample SD = s

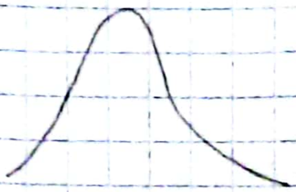
Measures of skewness.



Zero skewness



Negative skew



positive skew.

Que :- A sample of 25 plastic hinges was subjected to repealed stress cycles until failure. The number of cycles which each survived is given below.

72, 35, 63, 67, 87, 71, 64, 47, 60, 81, 39, 52, 57, 74,
43, 55, 37, 83, 48, 91, 53, 44, 94, 65, 75

35, 37, 39, 43, 44, 47, 48, 52, 53, 55, 57, 60,
63, 64, 65, 67, 71, 72, 74, 75, 81, 83, 87, 91, 9

a. Find five number summary.

Minimum = 35
Maximum = 94

$Q_1 = \left[\frac{25+1}{4} \times 1\right]^{th}$ value

$= 6.5^{th}$ value

$= 47 + \{(48-47) \times 0.5\}$ $\quad 6^{th}$ value $+ \{(7^{th}$ value $- 6^{th}$ value$) \times 0.5\}$

$Q_1 = 47 + \{1 \times 0.5\} = 47.5$

$Q_2 = \left(\frac{25+1}{4} \times 2\right)^{th}$ value

$Q_2 = \left(\frac{26}{4} \times 2\right)^{th}$ value

$Q_2 = 13^{th}$ value

$Q_2 = \underline{\underline{63}}$

$Q_3 = \left[\frac{25+1}{4} \times 3\right]^{th}$ value

$Q_3 = \left(\frac{26}{4} \times 3\right)^{th}$ value

$Q_3 = 19.5^{th}$ value $\quad 19^{th}$ value $+ \{(20^{th}$ value $- 19^{th}$ value$) \times 0.5\}$

$Q_3 = 74 + \{(75-74) \times 0.5\}$

$Q_3 = 74 + \{1 \times 0.5\}$

$Q_3 = \underline{\underline{74.5}}$

(ii) Find mode, P_{15}, D_3, mean, variance and sd.

~~mode~~ = There is no mode in the dat set.

$P_{15} = \left[\frac{25+1}{100} \times 15\right]^{th}$ value

$= \left(\frac{26}{100} \times 15\right)^{th}$ value

$= 3.9^{th}$ value.

$= 39 + \{(43-39) \times 0.9\}$

$= \underline{\underline{42.6}}$

$$D_3 = \left\{ \frac{25+1}{10} \times 3 \right\}^{th} \text{ value}$$

$$D_3 = 7.8^{th} \text{ value.}$$

$$D_3 = 48 + \{ (52-48) \times 0.8 \}$$

$$D_3 = \underline{\underline{51.2}}$$

$$\text{mean} = \frac{\text{Sum of the all data values}}{\text{Number of data values}}$$

$$= \frac{1557}{25}$$

$$= \underline{\underline{62.28}}$$

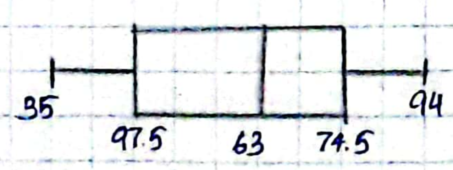
Variance, ~~s^2~~

$$s^2 = \frac{\sum_{i=1}^{25} (x_{25} - \bar{x})^2}{n-1}$$

$$s^2 = \underline{\underline{294.63}}$$

$$s_d = \sqrt{294.63}$$
$$= \underline{\underline{17.16}}$$

(iii) Draw box plot and stem and leaf pelot.



Steam	Leaves.
3	5 7 9
4	3 4 7 8
5	2 3 5 7
6	0 3 4 5 7
7	1 2 4 5
8	1 3 7
9	1 4

key · 3|7 → 37

(iv) Comment on the distribution of data.

The distribution is of a negative (left) skew. The mass of the data is concentrated on the right.

Tutorial 02

(10) Determine the mean, median and mode values for the data set:

$\{4.72, 4.71, 4.74, 4.73, 4.72, 4.71, 4.73, 4.72\}$

$4.71, 4.71, 4.72, 4.72, 4.72, 4.73, 4.73, 4.74$

$$\text{mean} = \frac{37.78}{8} = \underline{\underline{4.72}}$$

$$\text{median} = \frac{\left(\frac{8}{2}\right)^{th} \text{value} + \left(\frac{8}{2}+1\right)^{th} \text{value}}{2} = \frac{4^{th}\text{ value} + 5^{th}\text{ value}}{2}$$

$$= \frac{4.72 + 4.72}{2}$$

$$= \underline{\underline{4.72}}$$

$$\text{mode} = \underline{4.72}$$

(11) 21 bricks have a mean mass of 24.2 kg and 29 similar bricks have a mass of 23.6 kg. Determine the mean mass of the 50 bricks.

$$\text{mean mass of 50 bricks} = \frac{\frac{\Sigma x_i}{21} + \frac{\Sigma x_i}{29}}{50} \quad \frac{\Sigma x_i \times 21 + \Sigma x_i \times 29}{50}$$

mean mass of 50 bricks $= \frac{21\times(24.2\text{ kg}) + (23.6\text{ kg})\times 29}{50}$

$= \cancel{\frac{47.8\text{ kg}}{50}}$ $\quad \frac{508.2\text{kg} + 684.4\text{ kg}}{50}$

$= \frac{1192.6\text{ kg}}{50}$

$= \underline{23.852\text{kg}}$

• If not mention wether the data is a population or a sample, take it as a sample.

(12) Determine the standard deviation from the mean of the following set of numbers correct to 3 significant figures.

$\{35, 22, 25, 23, 28, 33, 30\}$

$22, 23, 25, 28, 30, 33, 35$

* mean $= \frac{\sum x_i}{n}$

$= \frac{22+23+25+28+30+33+35}{7}$

$= \frac{196}{7}$

$= \underline{28}$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

$= \frac{(22-28)^2 + (23-28)^2 + (25-28)^2 + (30-28)^2 + (33-28)^2 + (35-28)^2}{7-1}$

$7\overline{)196}$ 28, 14, 56

$= \frac{36 + 25 + 9 + \cancel{25} + 25 + 49}{6}$

$$s^2 = \frac{148}{6}$$

$$s^2 = 24.67$$

$$s = \sqrt{24.67}$$
$$s = 4.966$$

Probability

Terminology

- Experiment : A process leading to a well-defined observations or outcomes that generates a set of data.
- Trial : Each repitition, if the experiment can be reapeated any number of times under identical conditions.
- Sample space (S) : The set containing all possible outcomes of an experiment.
- Finite Sample Space : Sample space that contains a finite number of outcomes.
- Continous Sample Space : Sample Space that contains an interval of values.
- Event : A subset of the sample space. Usually denoted in upper case letters.
- Simple Event : An event that corresponds to a single possible outcome.

- $\emptyset$ → impossible event
- $A \cup B$ → all outcomes that are in A or in B or in both.
- $A \cap B$ → outcomes that are both in A and B.
- A^c → all outcomes not in A, but in S.

- Mutually Exclusive Events : $A \cap B = \emptyset$ (cannot happen together)

- Collectively exhaustive Events : One of the events must occur. The set of events covers the entire sample space.

- Independent Event : Occurence of one event not affect on the occurence of other event.

- Join Events (Compound Events) : An event that corresponds to more than a single possible outc-ome.
  E.g.:- Getting an odd no. by rolling a die.

Example.

01. A balanced die is rolled. Let A be the event that an even number occurs.

Experiment : Rolling a balanced die
Sample space (S) : $S = \{1, 2, 3, 4, 5, 6\}$
Event (A) : $A = \{2, 4, 6\}$

Type of the event : compound event.

02. Consider a deck of cards. Let A - Aces, B - Black card, C - diamonds and D - Herats. Find collectively exhaustive events and mutually exclusive events.

Collectively exhausive events : B, C and D
A, B, C and D

Mutually exclusive events : B and C
B and D.

Probability

- Measure of the chance that an uncertain event will occur
- Denoted as $Pr(A)$ or $P(A)$
- The probability of S is always 1.

Example :-

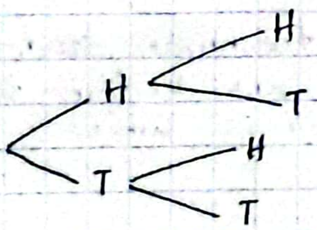
01. A balanced die is rolled. Let A be the event that an even number occurs. What is the probability of A ?

$$P(A) = \frac{\text{Number of outcomes in A}}{\text{Total number of outcomes in S}}$$

$$= \frac{3}{6}$$

$$= \underline{\underline{0.5}}$$

02. Suppose we toss two coins. Assume that all the ~~ote~~ outcomes are equally likely. Let A be the event that at least one of the coins shows up heads. Find ~~A~~ P(A)?



$\{H, H\} = \frac{1}{4}$

$\{H, T\} = \frac{1}{4}$

$\{T, H\} = \frac{1}{4}$

$$P(A) = P\{H, H\} + P\{H, T\} + P\{T, H\}$$

$$= \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$$

$$= \underline{\underline{\frac{3}{4}}}$$

Basic Properties.

- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \cap B \neq \emptyset$ then,

$$P(A \cup B) = P(A) + P(B)$$

- If $A_1, A_2, \ldots, A_k$ are mutually exclusive then,

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$$

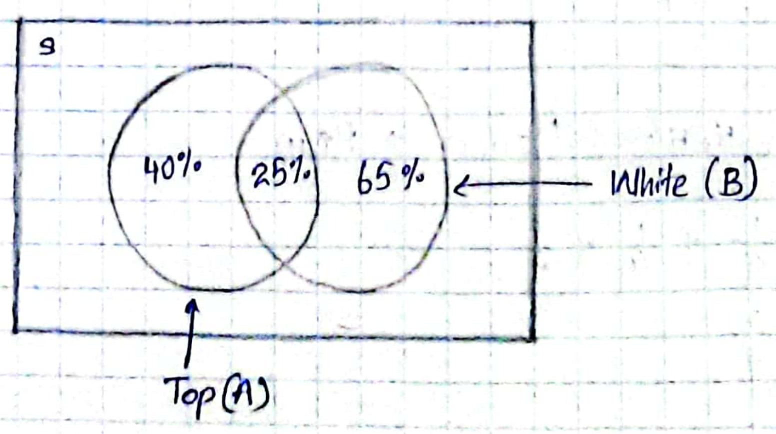
- If A and B are independent then,

$$P(A \cap B) = \cancel{A}\ P(A) * P(B)$$

Example.

01. In a large university, the ~~information~~ freshman profile for one year's fall admission says that 40% of the students were in the top 10% of their high school class, and that 65% are white, 25% of the students were white as well as were in the top 10% of their high school class

What is the probability that a freshman student selected randomly from this class either was in the top 10% of his or her high school class or is white?



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$=$$

Joint Probability.

The probability of events A and B occuring together is defined as joint probability of a joint event, A and B $[P(A \cap B)]$.

$$P(A \text{ and } B) = \frac{\text{Number of outcomes satisfying A and B}}{\text{Total number of outcomes in } S.}$$

Examples.

01. Find the probability that you will get a Black - Ace from a playing deck of cards, if a card is drawn at random.

Total no. of cards = 52
Total black cards = 26
No. of Black Aces = 2

$$\text{Probability} = \frac{2}{52}$$

$$= \frac{1}{26}$$

02. Find the probability that you will get a Red - Jack from a playing deck of cards, if a card is drawn at random.

$$\text{Probability} = \frac{2}{52}$$

$$= \frac{1}{26}$$

Marginal Probability.

The probability of a single event occurring (P(A)), without the interference of another event is known as marginal probability.

This can be thought of as an unconditional probability.

Examples.

01. Find the probability that you will get a king from a playing deck of cards, if a card is drawn at random.

$$\text{Probability} = \frac{4}{52}$$

$$= \frac{1}{13}$$

02. Find the probability that you will get a Black card from a playing deck of cards, if a card is drawn at random.

$$\text{Probability} = \frac{26}{52}$$

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

Conditional Probability

This is the probability of one event, given that another event has already occured.

The conditional probability of an event A, given that an event B has already occured is denoted by $P(A|B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$, $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Example.

01. Of the cars on a used car lot, 70% have air conditioning and 40% have a CD player. 20% of the cars have both. What is the probability that a car has a CD player, given that it has AC?

A - Car has AC
B - Car has a CD player

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$= \frac{0.2}{0.7}$$

$$= \cancel{0.285}\ \frac{2}{7}$$

02. If two balanced dice are tossed, find the probability that the sum of the face values is 8, if the face value of the first one is 3.

Properties of Conditional Probability.

- $P(A|B) = 1 - P(A^c|B)$
- $P(B \cup C|A) = P(B|A) + P(C|A) - P(B \cap C|A)$
- Multiplication law:

$$P(A \cap B) = P(B) * P(A|B)$$
$$= P(A) * P(B|A)$$

- If A and B are independent then,

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$
$$P(A \cap B) = P(A) * P(B)$$

- For independent events $A_1, A_2, \ldots, A_k$

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1) * P(A_2) * \ldots * P(A_k)$$

Random Variables and Probability Distributions.

Random variables

A random variable x is a function defined on a sample space (S), that associates a real number, $x(\omega) = x$, with each outcome ω in S.

Random variables are denoted by using capital letters.

Random Variables.

- Discrete random variables
- Continuous random variables.

Discrete Random Variables.

- A random variable is said to be discrete, if it can assume only distinct values

or.

~~In other~~ • It can assume only countable number of values.

Ex :-

Toss a coin 5 times. Let x be the number of heads appeared Then,

$$X = 0, 1, 2, 3, 4, 5$$

Roll a die twice. Let x be the number of times 4 comes up

$$X = 0, 1, 2$$

Suppose we toss two coins. Assume that all the outcomes are equally likely (fair coins). Let Y be the number of heads appeared. Then,

$$Y = 0, 1, 2$$

Probability Distributions.

- The set of all ordered pairs $(x, Pr(X=x))$ of a discrete random varia-bles (X) is known as the probability distributions.

- This is also known as the probability mass function. (p.m.f.) and is denoted by $P_X(x)$.

Properties :-

- $P_X(x)$ refers to $P(X=x)$
- The probability distribution function is always non-negative.
- $\sum_{all\ x} P_{X}(x) = 1$
- The cumulative distribution function F of the random variable X is defined by

$$F_{X}(x) = Pr(X \leq x)$$

Expected Value and Variance.

* This is same as mean of the random variable

- Let x be a discrete random variable with p.m.f $P_{X(x)}$. Then the expected value of x, denoted by $E(x)$ is defined by

$$\rightarrow E(x) = \sum x * P_r(X = x)$$

- The variance of a random variable X is defined by

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

Let X and Y be two random variables. Then,

$[E(x)]$

- $E(c) = c$
- $E[g(x)] = \sum g(x) * Pr(X = x)$
- $E[g(x) + c] = E[g(x)] + c$
- $E[c * g(x)] = c * E[g(x)]$
- $E[X + Y] = E[X] + E[Y]$

~~Let~~ $[V(x)]$

- $V(c) = 0$
- $V[g(x) + c] = V[g(x)]$
- $V[c * g(x)] = c^2 * V[g(x)]$
- $V[X + Y] = V[X] + V[Y] + 2\,Cov(X, Y)$.
- $V[X - Y] = V[X] + V[Y] - 2\,Cov(X, Y)$
- If X and Y are independent then, $Cov(X, Y) = 0$

Discrete Probability Distribution.

Bernouli Distribution

- We can apply this for experiment which give two outcomes. (Success and Failure)
- Only one trial
- e.g.:- Tossing a coin once.

→ X - Getting the success (Getting head)

$x = 0, 1$ (tail, head)

→ $X \sim$ Bernoulli(p)

p - probability of success

$x \sim$ Bernoulli (p = 0.5)

→ $P_x(x) = p^x (1-p)^{1-x}$

x	0	1
$P_x(x)$	0.5	0.5

$P(x=0) = p^x \cdot (1-p)^{1-x}$
$= 0.5^0 \cdot (1-0.5)^{1-0}$
$= 1 \times 0.5^1$
$= 0.5$

$P(x=1) = 0.5^1 \cdot (1-0.5)^{1-1}$
$= 0.5 \times 0.5^0$
$= 0.5 \times 1 = 0.5$

→ $E(x) = p$

→ $V(x) = p(1-p)$

Binomial Distribution

- For each trial, only two possible outcome. (Success and Failure)
- No. of trails (n) are fixed.
- Probability of success (p) is constant for each and every trial.
- Trials are independent
- e.g.:- Tossing a coin 10 times

→ X - The number of success in n number of trials.

Let, x - Number of heads in 10 number of trials

$x = 0, 1, 2, 3, \ldots\ldots, 10$
The number of head appears

→ $x \overset{\text{Follows}}{\sim} Bin(n, p)$

$x \sim Bin(n = 10, p = 0.5)$

→ $P_x(x) = \binom{n}{x} p^x (1-p)^{n-x}$ ($= {}^nC_x$)

$x = 0, 1, 2, 3, \ldots, n$ [p.m.f]

x	0	1	2	....	10
$P_x(x)$					

$P(x=0) = {}^{10}C_0 \cdot 0.5^0 \cdot (1-0.5)^{10-0}$

Poission Distribution.

- For each trial, only two possible outcomes (Success and Failure)
- Trials (n) is large
- The occurences are independent of each other.
- E.g.:- Number of defects in a lot.

→ X - The number of success occurrences for a given rate of occurences. (λ)

$x \sim poisson(\lambda)$

$$P_x(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$; x = 0, 1, 2, \ldots$

$E(x) = \lambda$
$V(x) = \lambda$

$E(x) = p = \underline{0.5}$ | $\Rightarrow E(x) = np$ (ii)

$V(x) = p(1-p)$ | $\Rightarrow V(x) = np(1-p)$

$V(x) = 0.5 \times (1-0.5)$
$= 0.5 \times 0.5 = \underline{0.25}$

* An expansion of the Bernoulli distribution.
* Each trail has a Bernoulli distribution

(iii)

Example.:-

Q1. It is known that screws produced by a certain machine will be defective with probability 0.01 independently of each other. If we randomly pick 10 screws produced by this machine, what is the probability that.

(i) exactly six screws will be defective?

Let, x - No. of defective screws out of 10 screws.

$x = 0, 1, 2, \ldots, 10$

$x \sim Bin(n=10, p=0.01)$

$x = 6 \Rightarrow$

$P(x=6) = \binom{n}{x} p^{x} (1-p)^{n-x}$

$= {}^{10}C_6 \cancel{p^{10}} (0.01)^{10} (1-0.01)^{10-6}$

$= \underline{10!}$

(ii) at most 3 screw will be defective?

$P(x \leq 3) =$

$x = 0, 1, 2, 3$

$$P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

$$= {}^{10}C_0 (0.01)^0 . (1-0.01)^{10-0} + {}^{10}C_1 (0.01)^1 . (1-0.01)^9 +$$

$$ {}^{10}C_2 (0.01)^2 . (1-0.01)^8 + {}^{10}C_3 (0.01)^3 . (1-0.01)^7$$

(iii) at least 2 screw will be defective?

$P(x \geq 2)$

$x = 2, 3, 4, 5, 6, 7, 8, 9, 10$

$$P(x \geq 2) = P(x=2) + P(x=3) + P(x=4) + P(x=5) + P(x=6) +$$

$$+ P(x=7) + P(x=8) + P(x=9) + P(x=10)$$

or

$$P(x \geq 2) = 1 - P(x<2)$$

$$= 1 - \{P(x=0) + P(x=1)\}$$

$$= 1 - \{ {}^{10}C_0 (0.01)^0 . (1-0.01)^{10} + {}^{10}C_1 (0.01)^1 . (1-0.01)^9 \}$$

$=$

(iv) What is the expected number of defectives?

$$E(x) = np$$
$$= 10 \times 0.01$$
$$= \underline{0.1}$$

(v) What is the variance of defectives?

$$V(x) = np(1-p)$$
$$= 10 \times 0.01 (1-0.01)$$
$$= 0.1 \times 0.99$$
$$= \underline{0.099}$$

$SD(x) = \sqrt{V(x)} = \sqrt{0.099} = \underline{0.3146}$ ← Standard deviation.

$\boxed{\text{Tutorial 04}}$

(01) A manufacturing process produces components which are free from any faults with probability p. Find the probability that in a sample of size 50 from a large batch there are fewer than 4 faulty components when p = 0.95. Find the probability that in a sample of size 50 there are fewer than 10 faulty when p = 0.75.

Let $(x;\ )$

$x \rightarrow$ No. of faulty components out of 50 components.

$$X \sim Bin(n=50, p=0.05)$$

p is being a good component

$$p(\text{faulty}) = 1-0.95$$
$$= 0.05$$

$$P(x<4) = 1 - P(x \geq 4)$$
$$= 1 - 0.23959$$
$$= \underline{0.76041}$$

(define success as being faulty for solve this problem)

$P(\text{faulty}) = 1 - 0.75$
$= 0.25$

$P(x < 10) = 1 - P(x \geq 10)$
$= 1 - 0.83632$
$= 0.16368$

Eg.:-

Suppose that, on average, in every two pages of a book there is one typographical error. What is the probability of at least one error on a certain page of the book?

Let, x - No. of errors on a certain page

$x - 0, 1, 2, \ldots\ldots$

$x \sim \text{poisson}\ (\lambda = 0.5)$

2 pages → 1 Error

1 page → ~~0.5~~ $\frac{1}{2}$ Error

~~$P(x<1)$~~

$P(x \geq 1) = P(x=1) + P(x=2) + P(x=3) + \ldots\ldots$

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

$= 1 - \dfrac{e^{-\frac{1}{2}}\left(\frac{1}{2}\right)^0}{0!}$

$= 1 - \dfrac{e^{-\frac{1}{2}} \times 1}{1}$

$= 1 - e^{-0.5}$

$= 1 - 0.6065$

$= 0.3935$

Using table

$r = 0$
$(x \geq 0)$

$r = 1$
$(x \geq 1)$

Today is difficult Tomorrow is much more difficult But day after tomorrow is beautiful Most people die tomorrow evening

Tutorial 04

(06.) Suppose that in late summer, the Fremantle Surf Life Saving club makes an average of two surf rescues per day. Use the Poisson probability distribution to determine the probability that:

(a) More than two rescuses are made on a particular day.

Let, x - No. of recues persons made per day

a day → 2 surf rescues.

$p(x \geq 2)$

$x \sim$ poisson $(\lambda = 2)$.

$p(x \geq 2) = \cancel{1 - p(x \leq 1)}$ $p(x \geq 3)$

$\cancel{1 - e^{-2} \times 2^1}$

$= 0.32332.$ (by using table).

(b) Five surf rescues are made in a 3-day period.

y - No. of recued persons for a 3 day period.

3 day → 6 rescues

~~a day → 5/3 rescu~~

$x \sim$ poisson $(\lambda = \overset{2\times3=}{6})$

$P(y=5) = P(y \geq 5) - P(y \geq 6)$

$= 0.71494 - 0.55432$

$= \underline{0.16062}$

Poisson Approximation

* If $x \sim Bin(n,p)$, then x can be approximated with a Poisson distribution with the rate parameter (λ) being equal to np if p is quite small and n is large.

  * Usually this approximation can be used if $p < 0.1$ and $n > 50$.

Example :-

If the probability that an individual suffers an adverse reaction from a particular drug is known to be 0.001, determine the probability that out of 2000 individuals, a

(a) exactly three individuals will suffer an adverse reaction.

Let, x - No. of individuals who have adverse reaction out of 2000 individuals.

$$x \sim Bin(n = 2000, p = 0.001)$$

Since $n = 2000 > 50$ and $p = 0.001 < 0.1$ we can approximate x into a poisson distribution.

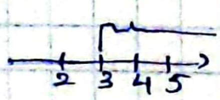
$$x \sim poisson(\lambda = np = 2000 \times 0.01 = 2)$$

$$P(x=3) = P(x \geq 3) - P(x \geq 4)$$
$$= 0.32332 - 0.14288$$
$$= \underline{0.18044}$$

(b) more than two individuals will suffer an adverse reaction.

$$x \sim poisson(\lambda = np = 2000 \times 0.001 = 2)$$

$$P(x > 2) = P(x \geq 3)$$
$$= \underline{0.32332}$$



Continuous Probability Distributions.

- A random variable is said to be continous, if it can take any value within a range.

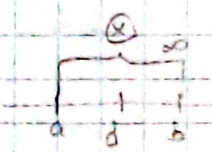
- Continous data are frequently measured in some way rather than counted.

- If x is a continuous random variable, $Pr(x=a) = 0$ for any value of a.

Example :-

28.92°c 31.33°c

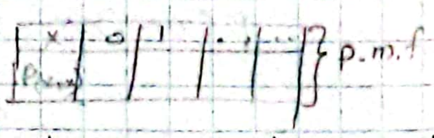
Temperature ⟹ (28°c - 32°c)
Heart beat of a patient.
Rainfall
Waiting time for a bus.



$$P(x=d) = \frac{1}{\infty} = \frac{1}{(1/0)} = \underline{\underline{0}}$$

Probability Distributions.

- For continous random variables, the probability distribution cannot be presented in a tabular form.

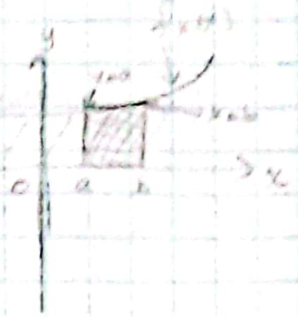


- Probability distribution function of a continous random variable is known as probability density function (pdf).

PDF - DEFINITION

The function $f_X(x)$ is a probability density function for the continuous random variable x, defined over the set of real numbers. ($\mathbb{R}$), if

- $f_X(x) \geq 0$, for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- $Pr(a < x < b) = \int_a^b f_X(x)\, dx$.



Properties.

Let X be a continous random variable with a p.d.f. ($f_X(x)$), defined over the set of real numbers ($\mathbb{R}$).

- The c.d.f. $F_X(x) = Pr(X \leq x) = \int_{-\infty}^{x} f_X(x)\, dx$.

  $f_X(x) : -\infty < x < \infty$

- $E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$ $\quad E(x) = \sum (x_i\, P(X = x_i))$

  E.g.:- $E(x) = \int_{-\infty}^{\infty} (x \cdot f_X(x))\, dx$

  $E(x^2) = \int_{-\infty}^{\infty} (x^2 \cdot f_X(x))\, dx$

- $V[g(x)] = E[g(x)^2] - \{E[g(x)]\}^2$

  E.g.:- $V(x) = E(x^2) - \{E(x)\}^2$

Ex.:-

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continous random variable X having the probability density function,

$$f_X(x) = \begin{cases} Cx^2 & -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(1) Find the value of c.

Let, x - Error in reaction temperature in °C

$$f_X(x) = cx^2 \; ; \; -1 < x < 2$$

Since $\int_{-1}^{2} f_X(x)\, dx = 1$

$\int a f(x) = a \int f(x)$

$\int x^n dx = \frac{x^{n+1}}{n+1}$

$$\int_{-1}^{2} (cx^2)\, dx = 1$$

$$c \int_{-1}^{2} x^2\, dx = 1$$

$$c \cdot \left[\frac{x^3}{3}\right]_{-1}^{2} = 1$$

$$c \cdot \left[\frac{8}{3} - \frac{(-1)}{3}\right] = 1$$

$$c \cdot \frac{9}{3} = 1$$

$$c = \frac{1}{3}$$

(2) Find $Pr(0 < x \leq 1)$

$$P(0 < x \leq 1) = \int_0^1 f_x(x)\, dx = \int_0^1 (cx^2)\, dx$$

$$= c \int_0^1 x^2\, dx$$

$$= c \cdot \left[\frac{x^3}{3}\right]_0^1$$

$$= c \cdot \left[\frac{1}{3} - \frac{0}{3}\right]$$

$$= c \cdot \frac{1}{3}$$

$$= \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

(3) Find the expected value and the variance.

$$E(x) = \int_{-1}^{2} x \cdot f_x(x)\, dx$$

$$\cancel{E(x)} = \int_{-1}^{2} x \cdot c x^2\, dx$$

$$= c \cdot \int_{-1}^{2} x^3\, dx$$

$$= c \cdot \left[\frac{x^4}{4}\right]_{-1}^{2}$$

$$= \frac{1}{3} \cdot \left[\frac{16}{4} - \frac{1}{4}\right]$$

$$= \cancel{\frac{1}{3} \cdot \frac{15}{3}} \qquad \frac{1}{3} \times \frac{15}{4}$$

$$= \cancel{\frac{5}{4}} \qquad \frac{5}{4} = 1.25$$

$$V(x) = E(x^2) - \{E(x)\}^2$$

$$\cancel{V(x)}$$

$$V(x) = \frac{11}{5} - \left(\frac{5}{4}\right)^2$$

$$V(x) = \frac{11}{5} - \frac{25}{16}$$

$$V(x) = \frac{176 - 125}{80} = \frac{51}{80}$$

$$E(x^2) = \int_{-1}^{2} (x^2 \cdot f_x(x))\, dx$$

$$E(x^2) = \int_{-1}^{2} x^2 \cdot c x^2\, dx$$

$$= c \cdot \int_{-1}^{2} x^4\, dx$$

$$= \frac{1}{3} \cdot \left[\frac{x^5}{5}\right]_{-1}^{2}$$

$$= \frac{1}{3} \cdot \left(\frac{32}{5} - \frac{(-1)}{5}\right)$$

$$= \frac{1}{3} \times \frac{33}{5}$$

$$= \frac{11}{5}$$

$\frac{11}{16}$ 176

(4) Find the c.d.f.

$$f_X(x) = P(X \le r) \; ; \quad (-1 < r < x < 2)$$

r is in with in the range

$$F_X(r) = \int_{-1}^{r} f_X(x)\,dx$$

$$= \int_{-1}^{r} (cx^2)\,dx = c\int_{-1}^{r} x^2\,dx = c.\left[\frac{x^3}{3}\right]_{-1}^{r}$$

$$= c.\left[\frac{r^3}{3} - \frac{(-1)^3}{3}\right]$$

$$= \frac{1}{3}.\left(\frac{r^3+1}{3}\right)$$

$$F_X(r) = \frac{r^3+1}{9}.$$

$$F_X(x) = \frac{x^3+1}{9} \; ; \; (-1 < x < 2) \quad \Leftarrow \quad P(X \le x)$$

They ask To calculate $p(x \le 1)$ how do it from c.d.f

$$\therefore p(x \le 1) = F_X(1) = \frac{1^3+1}{9} = \frac{2}{9}$$

Tutorial 05

Q1. An officer is always late to the ~~officest~~ office and arrives within the grace period of ten minutes after the start. Let x be the time that elapses between the start and the time the officer signs in with a probability density function.

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

where $k > 0$ is a constant.

(a) Compute the value of k.

Let x :- the time that elapses between the start and the time the officer signs.

$$f_x(x) = kx^2 \; ; \; 0 \le x \le 10$$

Since, $\int_0^{10} f_x(x)\,dx = 1$

$$\int_0^{10} (kx^2)\,dx = 1$$

$$k \int_0^{10} x^2\,dx = 1$$

$$k \cdot \left[\frac{x^3}{3}\right]_0^{10} = 1$$

$$k \cdot \left[\frac{1000}{3} - \frac{0}{3}\right] = 1$$

$$k \cdot \frac{1000}{3} = 1$$

$$k = \frac{3}{1000} = 0.003$$

(b) Find the cumulative distribution function of x.

$$F_x(r) = P(x \le r) \; ; \; (0 \le r \le x \le 10)$$

$$F_x(r) = \int_0^r f_x(x)\,dx$$

$$= \int_0^r (kx^2)\,dx = k\int_0^r x^2\,dx$$

$$= k\left[\frac{x^3}{3}\right]_0^r = \frac{3}{1000}\left[\frac{r^3}{3} - \frac{0^3}{3}\right]$$

CamScanner

$$F_x(r) = \frac{3}{1000}\left(\frac{r^3}{3}\right)$$

$$F_x(r) = \frac{\cancel{3} r^3}{\cancel{3} \times 1000}$$

$$F_x(r) = \frac{r^3}{1000} \quad \Big\} \quad F_x(x) = \frac{x^3}{1000} \; ; \; 0 \leq x \leq 10$$

c) ~~P(x<3)~~ Find the probability that the he arrives less than 3 minutes after the start of the office.

$$p(x<3) = F_x(3) = \frac{3^3}{1000} = \frac{27}{1000}$$

(d) Calculate the mean and variance of x.

$$E(x) = \int_0^{10} x \cdot f(x)\, dx$$

$$= \int_0^{10} x \cdot k x^2 \, dx$$

$$= k \int_0^{10} x^3 \, dx$$

$$= \frac{3}{1000}\left[\frac{x^4}{4}\right]_0^{10}$$

$$= \frac{3}{1000} \times \left[\left\{\frac{10\,000}{4}\right\} - 0\right]$$

$$= \frac{3}{\cancel{1000}} \times \frac{10\,\cancel{000}}{4}$$

$$= \frac{30}{4} = 7.5$$

$$V(x) = E(x^2) - \{E(x)\}^2$$
$$= 60 - (7.5)^2$$
$$= 60 - 56.25$$
$$= 3.75$$

$$E(x^2) = \int_0^{10} x^2 f_x(x)\, dx$$
$$= \int_0^{10} (x^2 \cdot kx^2)\, dx$$
$$= k \cdot \int_0^{10} x^4\, dx$$
$$= \frac{3}{1000} \left[\frac{x^5}{5} \right]_0^{10}$$
$$= \frac{3}{1000} \times \frac{100000}{5}$$
$$= \frac{300}{5} = 60$$

$$SD(x) = \sqrt{V(x)} = \sqrt{3.75}$$
$$= 1.9365$$

Continuous Probability Distribution.

Exponential Distribution

The distribution is usually used to model life times. (There is a link to the Poisson distribution)

$x \sim Exp(\lambda)$

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

$$\boxed{\begin{aligned} E(x) &= 1/\lambda \\ V(x) &= 1/\lambda^2 \end{aligned}}$$

E.g. :- life time of batteries
life time of bulbs

- Widely used in waiting line (or queuing) theory to model the length of time between arrivals in process.

Ex. :- duration between two customers at Bank ATMs, To model patients entering to an accident ward.

Example :-

Q1 The time, in hours, during which an electrical generator is operational is a random variable that follows an exponential distribution with a mean of 160. What is the probability that a generator of this type will be operational for,

a) Less than 40 hours ?

Let x - The time, in hours, which an electrical generator is operational.

Given that $x \sim Exp\left(\lambda = \frac{1}{160}\right)$

$$\left\{ \begin{array}{l} E(x) = 1/\lambda \\ \lambda = \frac{1}{160} \end{array} \right.$$

$f(x) = \lambda . e^{-\lambda x} \; : x \geq 0$

$$P(x<40) = \cancel{P(x \geq 40)} \; P \int_0^{40} f(x) \, dx$$

$$= \int_0^{40} (\lambda . e^{-\lambda x}) \, dx$$

$$= \lambda . \int_0^{40} e^{-\lambda x} \, dx$$

$$= \cancel{\lambda} . \left[\frac{e^{-\lambda x}}{-\cancel{\lambda}} \right]_0^{40}$$

$$= \left[- e^{-\lambda x} \right]_0^{40}$$

$$= -e^{-1/160 \times 40} - \left(- e^{-\frac{1}{160} \times 0} \right)$$

$$= -e^{-1/4} + e^0$$

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

$$\int_a^b e^{ax} dx = \left[\frac{e^{ax}}{a} \right]_a^b$$

divide by ~~coeffi~~ dirivative of x

(b) Between 60 and 160 hours?

$$P(60 < x < 160) = \int_{60}^{160} f(x)\,dx$$

$$= \int_{60}^{160} (\lambda \cdot e^{-\lambda x})\,dx$$

$$= \lambda \int_{60}^{160} e^{-\lambda x}\,dx$$

$$= \lambda \cdot \left[\frac{e^{-\lambda x}}{-\lambda}\right]_{60}^{160}$$

$$= \left(-e^{-\lambda x}\right)_{60}^{160}$$

$$= -e^{-1/160 \times 160} - \left(-e^{-1/160 \times 60}\right)$$

$$= -e^{-1} + e^{-3/8}$$

$$= \cancel{1.0562}\ 0.32$$

(c) More than 200 hours.

$$P\,\cancel{10}(x > 200) = \int_{200}^{\infty} f(x)\,dx$$

$$= \int_{200}^{\infty} (\lambda \cdot e^{-\lambda x})\,dx$$

$$= \lambda \int_{200}^{\infty} e^{-\lambda x}\,dx$$

$$= \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_{200}^{\infty}$$

$$= \left[-e^{-\lambda x}\right]_{200}^{\infty}$$

$$P(x > 200) = -e^{-1/160 \times 200} - \left(-e^{-1/160 \times \infty}\right)$$

$$= -e^{-5/4} + e^{-\infty}$$

$$= e^{-\infty} - e^{-5/4}$$

$$=$$

$$P(x>200) = \int_{200}^{\infty} f(x)\,dx = 1 - P(x \leq 200)$$

$$= 1 - \int_{0}^{200} f(x)\,dx$$

$$= 1 - \int_{0}^{200} (\lambda \cdot e^{-\lambda x})\,dx$$

$$= 1 - \lambda \int_{0}^{200} e^{-\lambda x}\,dx$$

$$= 1 - \lambda \left[\frac{\cancel{\lambda}\, e^{-\lambda x}}{-\lambda} \right]_{0}^{200}$$

$$= 1 - \cancel{\lambda} \left(-e^{-\lambda x} \right)_{0}^{200}$$

$$= 1 - \cancel{\lambda} \left[-e^{-1/160 \times 200} - \left(-e^{-1/160 \times 0} \right) \right]$$

$$= 1 - \left[-e^{-5/4} + e^{0} \right]$$

$$= 1 - e^{0} + e^{-5/4}$$

$$= \cancel{1} - \cancel{1} + e^{-5/4}$$

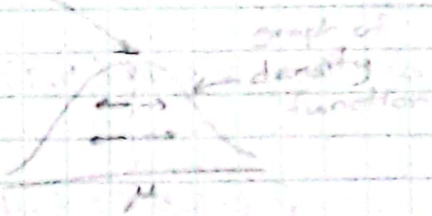
$$=$$

Normal Distribution / Gaussian Distribution

* This is most commonly used distribution.
* This is bell shaped distribution and perfectly symmetric around μ.

$$x \sim N(\mu, \sigma^2)$$

$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]} \qquad -\infty < x < \infty$$



$$\boxed{\begin{aligned} E(x) &= \mu \\ V(x) &= \sigma^2 \end{aligned}}$$

Standard Normal Distribution

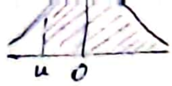
- Normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is known as the Standard Normal Distribution.
- Evaluating probabilities with Normal requires complex integratio
- To simplify the procedure, statistical tables are defined.
- But, tables for each combination of μ and σ^2 cannot be crea -ted.
- So, tables are only for the standard normal distribution

Normal ⟶ Standard Normal :

If $x \sim N(\mu, \sigma^2)$, Then

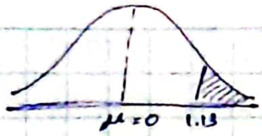
$$z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

-7.25
2
3.



P(z > u)

Normal Distribution - Examples :-



Q1. For $Z \sim N(0,1)$, calculate $Pr(Z \geq 1.13)$.

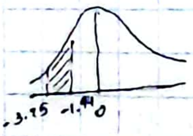
u → 0.03
↓
1.1   0.12924 } Stat table.

$Pr(Z \geq 1.13) = 0.12924$

Q2. For $x \sim N(5,4)$, calculate $Pr(-2.5 < x < 1.13)$

$$Pr(-2.5 < x < 1.13) = Pr\left(\frac{-2.5-5}{2} < \frac{x-\mu}{\sigma} < \frac{1.13-5}{2}\right)$$

$$= Pr\left(-3.75 < Z < -1.935\right)$$



$$= Pr(-3.75 < Z < -1.94)$$

$$= Pr(Z > -3.75) - Pr(Z > -1.94)$$

$$= 0.99991 - 0.97381$$

$$= 0.02610$$

Q3. The actual marks for FCS of Metro students revealed that they were normally distributed with a mean mark of 45 and a standard deviation of 22. What is the probability that a randomly chosen student will pass? (Assume that pass mark is 45).

$\mu = 45$
$\sigma^2 = 22^2$

Let x - Actual marks for FCS of a metro student.

Given that $x \sim N(\mu = 45, \sigma^2 = 22^2)$

$$P(x \geq 45) = P\left(\frac{x-\mu}{\sigma} \geq \frac{45-45}{22}\right)$$

$$= P(Z \geq 0)$$

$$= 0.5$$

Tutorial 05

① Suppose that we are told that the heights of adult males in a particular region of the world are normally distributed with a mean of 70 inches and standard deviation of 2 inches.

(a) Approximately what proportion of adult males are taller than 73 inches?

$$x \sim N(\mu = 70, \sigma^2 = 2^2)$$

$$P(x \geq 73) = P\left(\frac{x-\mu}{\sigma} \geq \frac{73-70}{2}\right)$$

$$= P(z \geq 1.5)$$

$$= \cancel{0.22663} \quad \underline{0.06681}$$

(b) What proportion of adult males are between 72 and 73 inches?

$$P(72 < x < 73) = P\left(\frac{72-70}{2} < \frac{x-\mu}{\sigma} < \frac{73-70}{2}\right)$$

$$= P(\cancel{0.5}\ 1 < z < \cancel{0.75}\ 1.5)$$

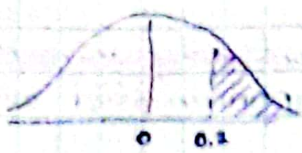
$$= P(z > \cancel{0.5}\ 1) - P(z > \cancel{0.75}\ 1.5)$$

$$= \cancel{0.30854 - 0.22663} \quad 0.15866 - 0.06681$$

$$= \cancel{0.08191} \quad \underline{0.09185}$$

(c) What height corresponds to the point where 20% of all adult males are greater than this height?

Let Height = a



$$P(x > a) = 20\% = 0.2$$

$$P\left(\frac{x-\mu}{\sigma} > \frac{a-70}{2}\right) = 0.2$$

$$P\left(z > \underbrace{\frac{a-70}{2}}_{b}\right) = 0.2$$

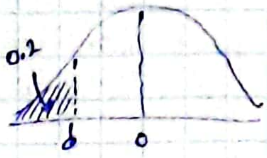
$$0.84 = \frac{a-70}{2}$$

$$1.68 + 70 = a$$

$$a = \underline{71.68} \text{ inches}$$

(d) What height corresponds to the point where 20% of all adult males are less than this height?

Let Height - b



$P(x<b) = 0.2$

$P\left(\frac{x-\mu}{\sigma} < \frac{b-70}{2}\right) = 0.2$

$P\left(z < \underbrace{\frac{b-70}{2}}_{d}\right) = 0.2$

$1 - P(z \geq d) = 0.2$

$P(z \geq d) = 1 - 0.2 = 0.8$

$d = -0.84$

$\frac{b-70}{2} = -0.84$

$b = -1.68 + 70$

$b = 68.32$ inches

Approximating Binomial Probabilities.

- For $x \sim Bin(n,p)$ this approximation can be used if n is large and p is moderate.
- A general rule can be defined as, np and n(1-p) is greater than 5.
- Can be approximated with a r.v. with a distribution $N(np, np(1-p))$.
- A continuity correction is needed because a discrete distribution is approximated with a continuous distribution.

$X \sim Bin(n,p) \xrightarrow{np>5 \ \& \ n(1-p)>5} Y \sim N(\mu, \sigma^2)$

if $np>5$ and $n(1-p)>5$
can use this approximation

$\mu = np$
$\sigma^2 = np(1-p)$

Continuity Correction.

If $X \sim Bin(n,p)$ is approximated with a r.v. $Y \sim N(np, np(1-p))$

$$P_r(X \le a) = P(Y < a+0.5)$$
$$P_r(X \ge a) = P(Y > a-0.5)$$
$$P_r(X < a) = P(Y < a-0.5)$$
$$P_r(X > a) = P(Y > a+0.5)$$
$$P_r(X = a) = P(a-0.5 < Y < a+0.5)$$

Approximating Poisson Probabilities :

- If $x \sim$ Poisson (λ) then if λ is greater than 20, the approximation can be used.
- Can be approximated with a r.v. with a distribution $N(\lambda, \lambda)$.
- A continuity correction is ~~used~~ needed because a discrete distributio is approximoated with a continuous distribution (just as in the case of the Binomial to Normal approximation.)

$\lambda > 20$ then you can approximate this to Normal

if both $np > 5$ and $n(1-p) > 5$ can use this approximation

$Y \sim N(\lambda, \lambda)$ $\mu = \lambda$

$\sigma^2 = \lambda$

e.g.:- Suppose that a sample of $n = 1600$ tires of the same type are obtained at random from an ongoing production process in which 8% of all such tires produced are defective. What is the probability that in ~~which~~ such a sample 150 or fewer tires will be defective?

Let x = No. of defective tires out of 1600 tires.

$$x \sim (n = 1600, p = 0.08)$$

$np = 1600 \times 0.08 = 128 > 5$

$np(1-p) = 1600 \cancel{\times 0.08} (1-0.08) = \cancel{320}\ 1472 > 5.$

Since, $np > 5$ and $np(1-p) > 5$, we can approximate x into Normal distribution.

$$Y \sim N\left(\mu = np = 128,\ \sigma^2 = np(1-p) = 117.76\right)$$

$$P(x \le 150) = P(Y < 150.5)$$
$$= P(Y \le 150.5)$$
$$= P\left(\frac{Y-\mu}{\sigma} < \frac{150.5 - 128}{\sqrt{117.76}}\right)$$

$$= P(z < 2.07)$$
$$= 1 - P(z \ge 2.07)$$
$$= 1 - 0.01923$$
$$\approx \cancel{683}$$
$$= \underline{\underline{0.98077}}$$

E.g.:- The annual number of earthquakes registering at least 2.5 on the Richter Scale and having an epicenter within 40 miles of down town Memphis follows a Poisson distribution with mean 22.5. What is the probability that at least 25 such earthquakes will strike next year?

Let x - No. of earthquakes occured during any year.

Given that $x \sim \text{Poisson}(\lambda = 22.5)$

Since, $\lambda = 22.5 > 20$, we can approximate x into Normal distribution.

$$Y \sim N\left(\mu = 22.5,\ \sigma^2 = 22.5\right)$$

$$P(x \ge 25) = P(Y > 24.5)$$
$$= P\left(\frac{Y-\mu}{\sigma} > \frac{24.5 - 22.5}{\sqrt{22.5}}\right)$$
$$= P(z > 0.42)$$
$$= \underline{\underline{0.33724}}$$

4.74