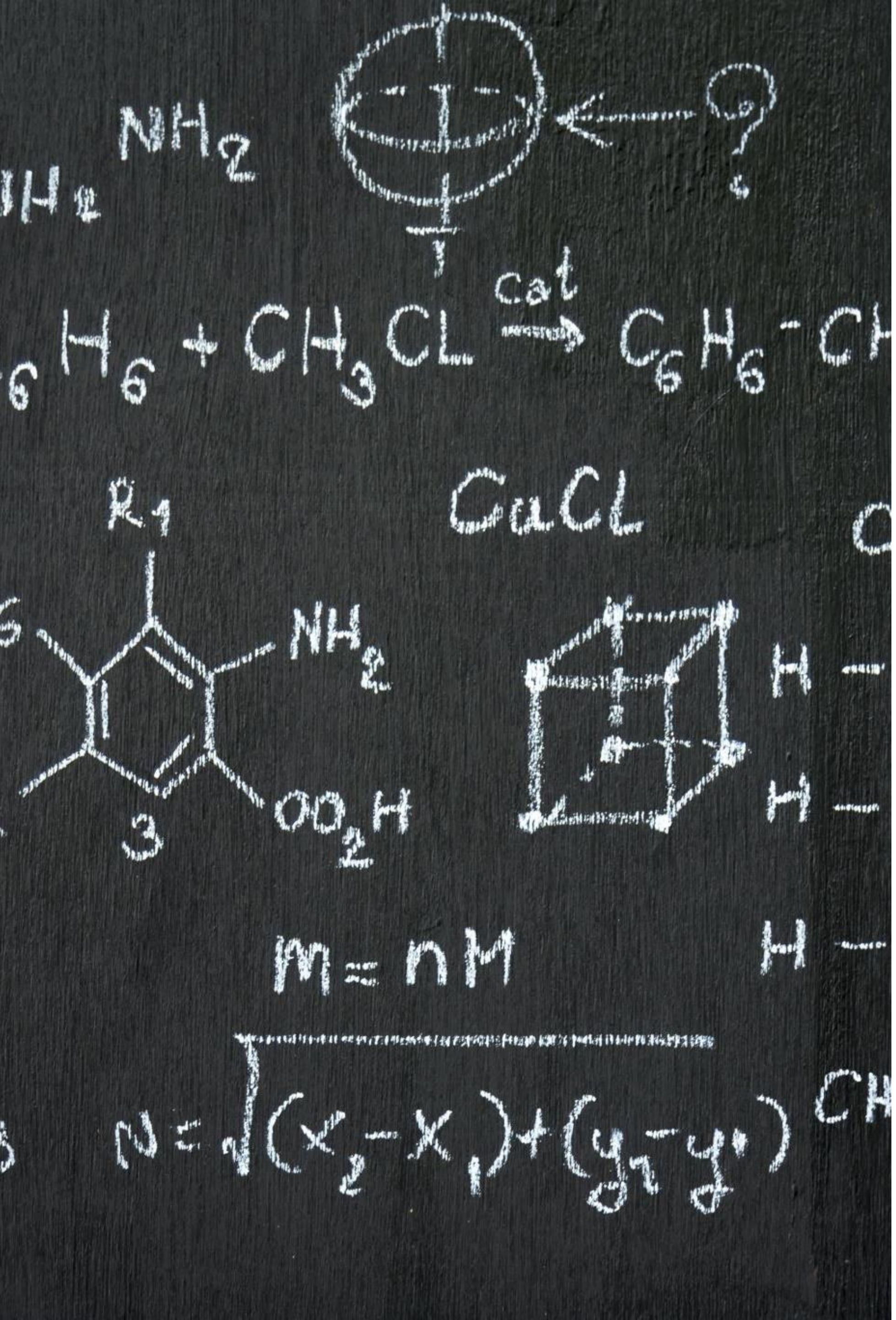
Correlation

Like covariance, but removes scale.
The *correlation coefficient* between X and Y is defined by

$$\mathrm{Cor}(X, Y) = \rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Properties:
1. ρ is the covariance of the standardized versions of X and Y.
2. ρ is dimensionless (it's a ratio).
3. $-1 \leq \rho \leq 1$. $\rho = 1$ if and only if $Y = aX + b$ with $a > 0$ and $\rho = -1$ if and only if $Y = aX + b$ with $a < 0$.



Covariance

- **Covariance formula** is a statistical formula which is used to assess the relationship between two variables.
- In simple words, **covariance** is one of the statistical measurement to know the relationship of the variance between the two variables.
- The covariance indicates how two variables are related and also helps to know whether the two variables vary together or change together.
- The covariance is denoted as Cov(X,Y)

Covariance

$$Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N}$$

$$Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N-1}$$

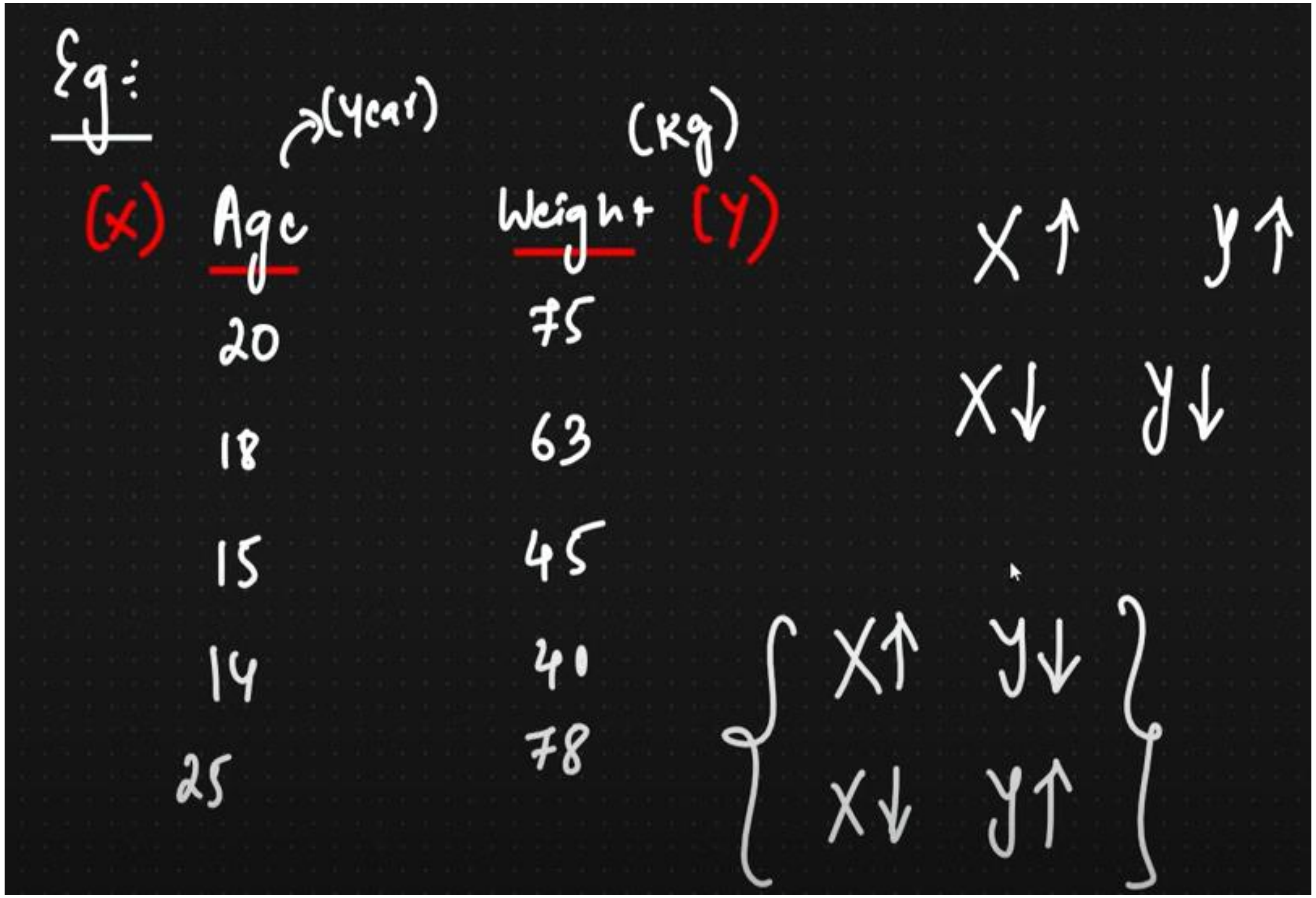
formula

Example-1

Eg:

Age (year)	Weight (Kg)
20	75
18	63
15	45
14	40
25	78

Example-1



X	Y
Economic Growth (%)	NIFTY 50 Index (%)
2.1	8
2.5	12
4.0	14
3.6	10

Example-2

x	y	$x-\bar{x}$	$y-\bar{y}$
2.1	8	-0.9	-3
2.6	12	-0.4	+1
4.0	14	1	+3
3.6	10	0.5	-1

$\bar{x} = 3.0$

$\bar{y} = 11$

$r = \frac{1.6}{0.7 \times 2.7} = 0.83$

$$Cov(x,y) = \frac{2.7 - .4 + 3 - .5}{3}$$

$$= 1.6$$

$$\sigma_x = \Sigma\sqrt{\frac{(x-\bar{x})^2}{3}} = 0.74$$

$$\sigma_y = \Sigma\sqrt{\frac{(y-\bar{y})^2}{3}} = 2.5$$

Covariance

$$Cov(x,y) = \frac{\sum_{i=1}^{N}(x-\bar{x})(y-\bar{y})}{N}$$

x	y	$\bar{x}$	$\bar{y}$	$x-\bar{x}$	$y-\bar{y}$
2.1	8	3.1	11	−1	−3
2.5	12	3.1	11	−0.6	1
4.0	14	3.1	11	0.9	3
3.6	10	3.1	11	0.5	−1

Solution

Solution

$$= \frac{3 - 0.6 + 2.7 + 0.5}{3}$$

$$Cov(x,y) = \frac{4.6}{3} = 1.533 \rightarrow \text{+ve value}$$

$$Cov(x,y) = \frac{(-1)(-3) + (-0.6)\times(1) + (0.9)\times(3) + (0.5)(-1)}{4-1}$$

Solution

$$\rho_{(x,y)} = \frac{Cov(x,y)}{\sigma_x * \sigma_y} = \frac{1.533}{(0.8981)(2.58)}$$

Example-1

Day	ABC Returns	XYZ Returns
1	1.1%	3.0%
2	1.7%	4.2%
3	2.1%	4.9%
4	1.4%	4.1%
5	0.2%	2.5%

- **Daily Return for Two Stocks are given below**

Find the covariance between the two stocks (the data is the sample data)

Solution-1

Sample Covariance formula

$$Cov(x,y) = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N-1}$$

- For ABC, the mean would be (1.1 + 1.7 + 2.1 + 1.4 + 0.2) / 5 = 1.30.
- For XYZ, the mean would be (3 + 4.2 + 4.9 + 4.1 + 2.5) / 5 = 3.74.

Covariance = [(1.1 - 1.30) x (3 - 3.74)] + [(1.7 - 1.30) x (4.2 - 3.74)] + [(2.1 - 1.30) x (4.9 - 3.74)] + …/(5-1)

- = ([0.148] + [0.184] + [0.928] + [0.036] + [1.364]) / (5-1)
- = 2.66 / (5 - 1)
- = 0.665

Correlation

- Definition The correlation coefficient of X and Y, denoted by Corr(X, Y), ρX,Y, or just ρ, is defined by

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

- It represents a “scaled” covariance – correlation ranges between -1 and 1.

Covariance

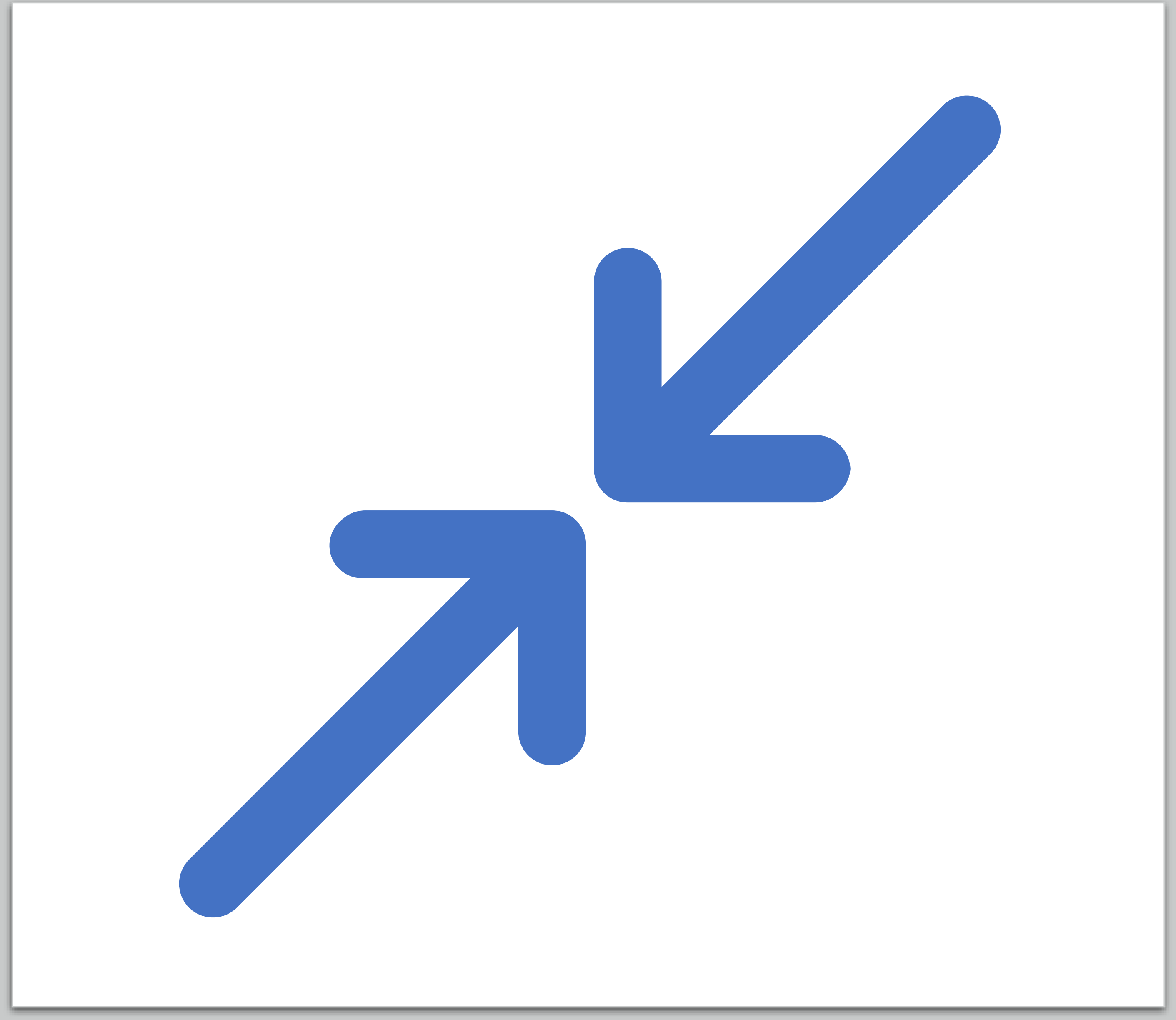
When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to one another. The covariance between two rv’ s X and Y is

$$Cov(X,Y) = E[(X - \mu(X))(Y - \mu(Y))]$$

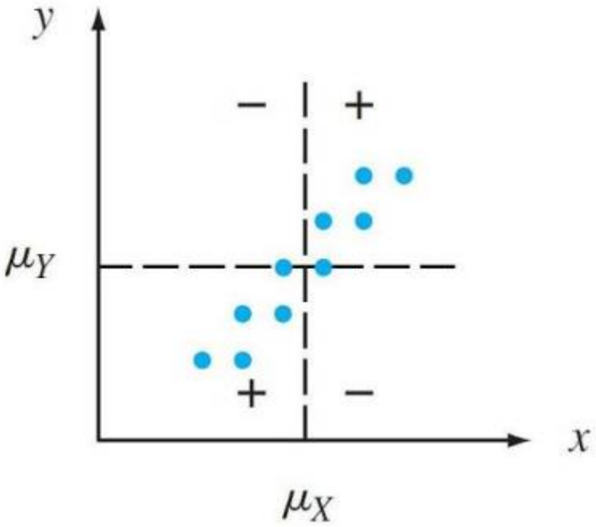
$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\, dx\, dy & X, Y \text{ continuous} \end{cases}$$

Covariance

- If both variables tend to deviate in the same direction (both go above their means or below their means at the same time), then the covariance will be positive.
- If the opposite is true, the covariance will be negative.
- X and Y are not strongly related, the covariance will be near 0

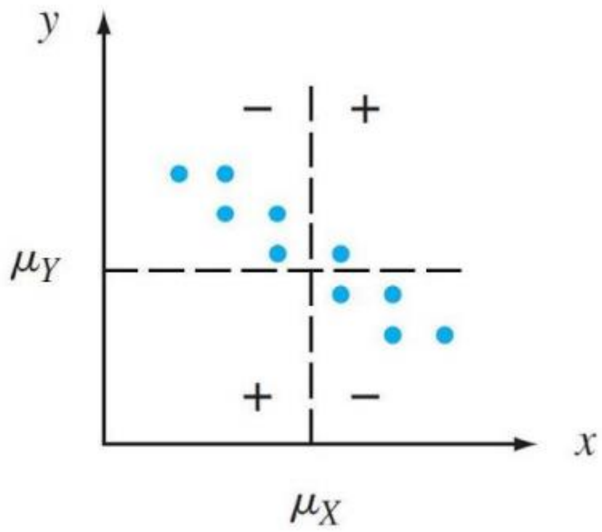


3 types of “ co-varying” :



(a) positive covariance;

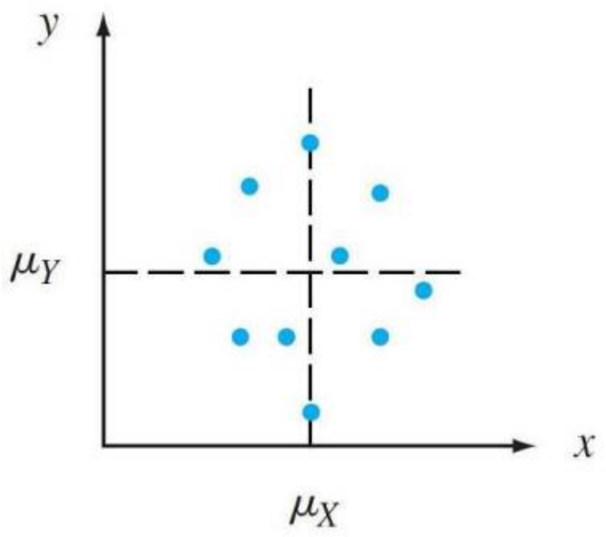
3 types of " co-varying" :



(b) negative covariance;

3 types of “ co-varying” :

-



(c) covariance near zero

Example-1

- An insurance agency services customers who have both a homeowner's policy and an automobile policy. For each type of policy, a deductible amount must be specified.
- For an automobile policy, the choices are \$100 and \$250, whereas for a homeowner's policy, the choices are \$0, \$100, and \$200. Suppose an individual "Rohit" is selected at random from the agency's files.
- Let X = his deductible amount on the auto policy and Y = his deductible amount on the homeowner's policy.
- Suppose the joint pmf is given by the insurance company the accompanying joint probability table:

			y	
$p(x, y)$		0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

- What is the covariance between X and Y?

Correlation

Definition The correlation coefficient of X and Y, denoted by Corr(X, Y), ρX,Y, or just ρ, is defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

It represents a “scaled” covariance – correlation ranges between -1 and 1.

Discrete Random Variable

Binomial Distribution:

- All the trails are independent
- Number (n) of trails is finite
- The Probability (p) of the success is same of each trials

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

Discrete Random Variable

Binomial Distribution:

- All the trails are independent
- Number (n) of trails is finite
- The Probability (p) of the success is same of each trials

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

Example:

a. A coin toss 3-times, find the probability of 2-Heads.

b. A coin toss 10-times, find the probability of 5-Heads.

Binomial Distribution

Q1. The probability that man aged 60 will live up to 70 is 0.65 out of 10 men. Now aged 60, find the probability:

1. At least 7 will live up to 70

2. Exactly 9 will live up to 70

3. At most 9 will live up to 70

Binomial Distribution

Q2. Out of 800 families with 5 children each, how many families would be expected to have

- 3 boys
- 5 girls
- Either 2 or 3 boys
- At least 2 girls

Binomial Distribution

Q3. The Probability that a pen manufactured by a company will be defective is 1/10. If 12 such pen are manufactured. Find the probability that:

1. Exactly 2 will be defective
2. None will be defective
3. At least 2 will be defective

Binomial Distribution

Q4. Medical professionals use the binomial distribution to model the probability that a certain number of patients will experience side effects as a result of taking new medications.

E.g., suppose it is known that 5% of adults who take a certain medication experience negative side effects. We can use a Binomial Distribution Calculator to find the probability that more than a certain number of patients in a random sample of 100 will experience negative side effects.

- P(X > 5 patients experience side effects) = **0.38400**
- P(X > 10 patients experience side effects) = **0.01147**
- P(X > 15 patients experience side effects) = **0.0004**

Binomial Distribution

Q5. Banks use the binomial distribution to model the probability that a certain number of credit card transactions are fraudulent.

For example, suppose it is known that 2% of all credit card transactions in a certain region are fraudulent. If there are 50 transactions per day in a certain region, we can use a Binomial Distribution Calculator to find the probability that more than a certain number of fraudulent transactions occur in a given day:

- P(X > 1 fraudulent transaction) = **0.26423**
- P(X > 2 fraudulent transactions) = **0.07843**
- P(X > 3 fraudulent transactions) = **0.01776**