Probability Theory

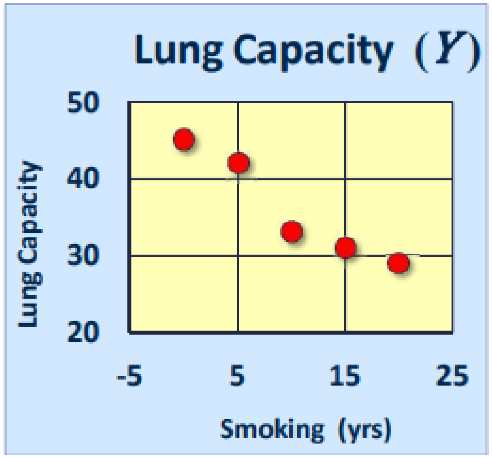
Covariance of RVs

- Random Variables may change in relation to each other. Covariance is a measure of association of two variables.
- If positive, then both variables increase or decrease together. If negative, then they vary in opposite manner.

- Covariance measures how much the movement in one variable predicts the movement in a corresponding variable
- **Example:** investigate relationship between cigarette smoking and lung capacity as shown in figure.

N	Cigarettes (X)	Lung Capacity (Y)
1	0	45
2	5	42
3	10	33
4	15	31
5	20	29

- Variables smoking and lung capacity *covary* inversely, like



Covariance of RVs

- Average product of deviation measures extent to which variables co-vary, the degree of linkage between them.

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$$

Deviation of data 1 from mean

Deviation from mean of data 2

Cigs (X)				Cap (Y)
0	-10	-90	9	45
5	-5	-30	6	42
10	0	0	-3	33
15	5	-25	-5	31
20	10	-70	-7	29
		$\sum$= -215		

Evaluation yields,

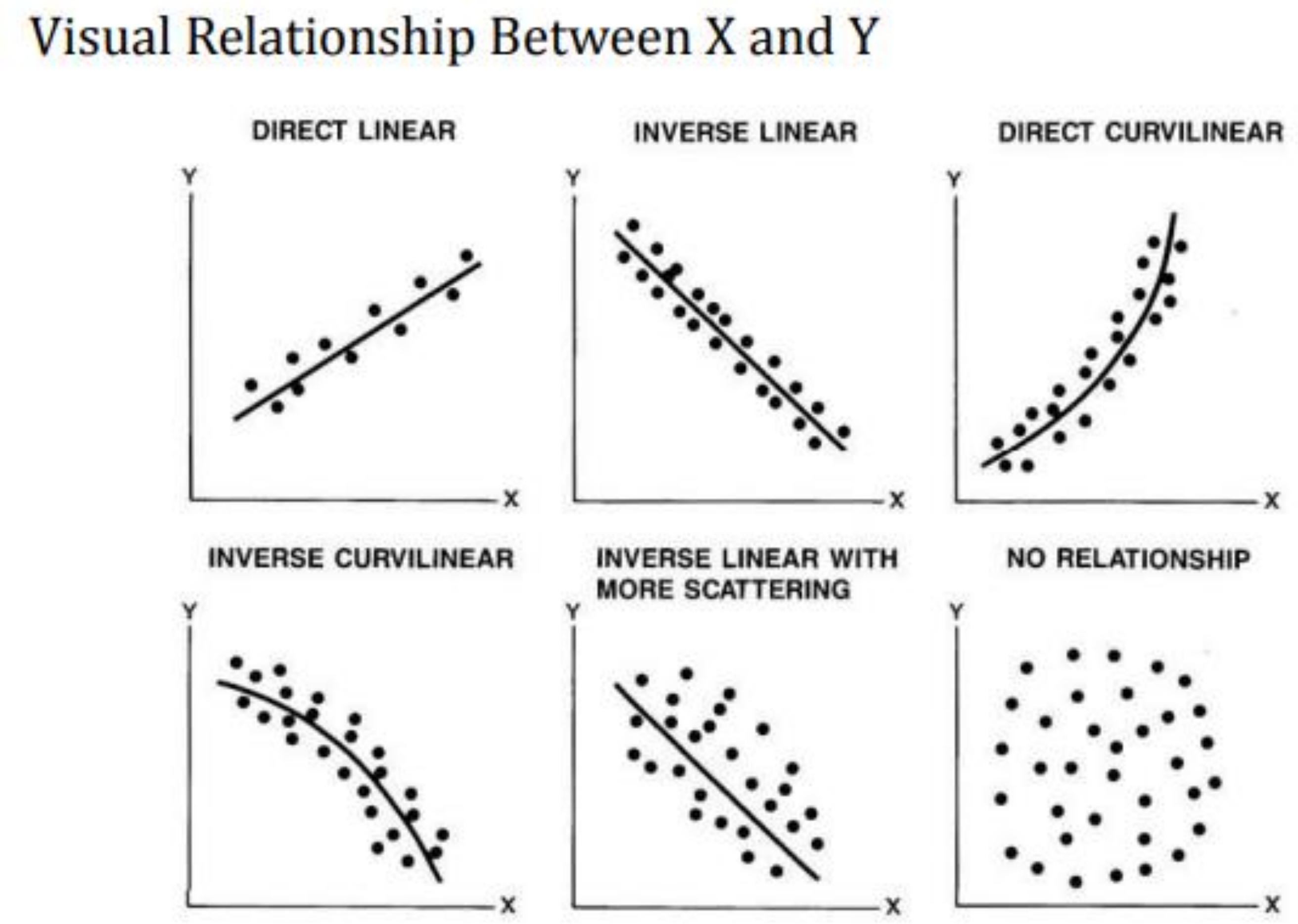
$$S_{xy} = \frac{1}{4}(-215) = -53.75$$

Correlation of RVs

- A measure which determines the standard change in one variable due to change in the other variable.
- Correlation is of two types, i.e. positive correlation or negative correlation.
- Correlation can take any value between -1 to +1, where in values close to +1 represents strong positive correlation and values close to -1 is an indicator of strong negative correlation.
- Measures of correlation:
  - Scatter diagram
  - Rank correlation coefficient

Correlation of RVs

- Correlation using scatter plot



Correlation of RVs

- Correlation coefficient

$$Corr(x,y) = \frac{\sum_{i=1}^{n}(x_i - x')\,(y_i - y')}{\sqrt{\sum_{i=1}^{n}(x_i - x')^2 \sum_{i=1}^{n}(y_i - y')^2}}$$

- Covariance, Cov(X,Y) is dependent upon the units of X & Y.
- Correlation, Corr(X,Y), scales covariance by the standard deviations of X & Y so that it lies between 1 & –1

$$Corr(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

Where σ is the Standard deviation

Common Distributions of RVs

- Uniform distribution
- Poisson distribution
- Normal distribution
- Standard normal distribution

Common Distributions of RVs

The Uniform Distribution

A random variable X is said to be *uniformly distributed* in $a \le x \le b$ if its density function is

$$f(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

and the distribution is called a *uniform distribution*.

The distribution function is given by

$$F(x) = P(X \le x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & x \ge b \end{cases}$$

The mean and variance are, respectively,

$$\mu = \frac{1}{2}(a+b), \qquad \sigma^2 = \frac{1}{12}(b-a)^2$$

Common Distributions of RVs

The Poisson Distribution

Let X be a discrete random variable that can take on the values 0, 1, 2, . . . such that the probability function of X is given by

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots \tag{13}$$

where λ is a given positive constant. This distribution is called the *Poisson distribution* (after S. D. Poisson, who discovered it in the early part of the nineteenth century), and a random variable having this distribution is said to be *Poisson distributed*.

Mean	$\mu = \lambda$
Variance	$\sigma^2 = \lambda$
Standard deviation	$\sigma = \sqrt{\lambda}$

When p is small and n is fixed, Mean = λ = np, where

- **n** is the Number of Trails
- **p** is Probability of Success

Common Distributions of RVs

The Normal Distribution

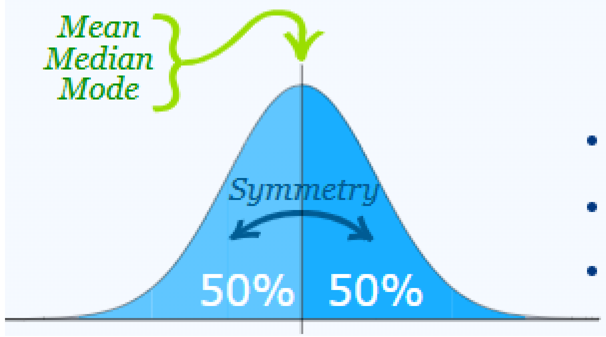
One of the most important examples of a continuous probability distribution is the *normal distribution*, sometimes called the *Gaussian distribution*. The density function for this distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty \tag{4}$$

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv \tag{5}$$

If X has the distribution function given by (5), we say that the random variable X is *normally distributed* with mean μ and variance σ^2.



The **Normal Distribution** has:

- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater than the mean

Common Distributions of RVs

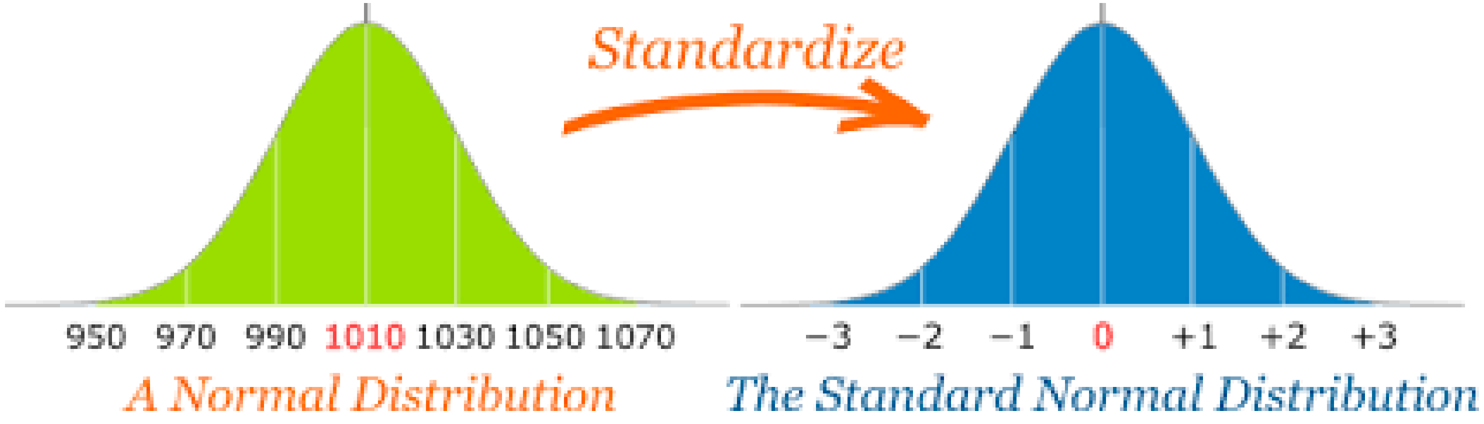
Standard normal distribution, also known as the z-distribution

- In this distribution, the **mean (average) is 0** and the **standard deviation (a measure of spread) is 1**.
- This creates a **bell-shaped curve** that is symmetrical around the mean ie. 0.
- The random variable of a standard normal distribution is known as the standard score or a z-score.

$$z = (X - \mu) / \sigma$$

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Where
$-\infty < z < \infty$



Central Limit Theorem

When large samples usually greater than thirty are taken into consideration then the distribution of sample arithmetic mean approaches the normal distribution irrespective of the fact that random variables were originally distributed normally or not.

Let us assume we have a random variable X.

Let σ be its standard deviation and μ is the mean of the random variable.

Now as per the Central Limit Theorem, the sample mean $\bar{X}$ will approximate to the normal distribution which is given as $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$.