Binomial Probability Distribution

- only 2 possible outcomes
- "SUCCESS" or "FAILURE"

Not according to real world success it may be according to condition
=> Failure in question can also be success in this condition

- We denotes SUCCESS with p and FAILURE with q

$$p+q=1 \Rightarrow q=1-p$$

- Every trial is independent
- Because there are only 2 outcome that doesn't depend on each other.
- Replacement

$$P(X=x) = \binom{n}{x} p^x q^{n-x}$$

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

- where

n = total number of trials
p = probability of SUCCESS
q = probability of FAILURE
x = number of successes in n trials

$$\boxed{\begin{array}{c} 0 < p.d < 1 \\ \text{Always} \end{array}}$$

Average Result:

$$\mu = np$$

Variance:

$$\sigma = \sqrt{np(1-p)}$$

Hypergeometric Probability Distribution

- Independant Success isn't met
- Good for finite and small number of population where the population falls into separat catagories.
- Without Replacement ✓
- Every trial is dependant.

Eg

⇒ Selecting 5 members from a commitee of 3 chemist and 6 physicists. If we select 1 chemist then there will be 2 chemist to be select and same as for the physicist. So, the sample space doesn't remains constant.

- Selection is from different finite catagoory.

$$P(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

Where

y = number of successes in trials n
N = total number of elements
r = number of successes in population
$N-r$ = number of failure in population
n = number of trials (sample space)
$n-y$ = number of failures in trials

Describing the equ.

$$p(y) = \frac{(\text{\# of way to get } y \text{ out of } r \text{ success})(\text{\# of ways to get } n-y \text{ out of } N-r \text{ failure})}{(\text{Number of ways to choose } n \text{ out of } N)}$$

$$p(y) = \frac{\text{\# of ways to get a group of } n \text{ with } y \text{ successes}}{\text{\# of ways to pick a group of } n \text{ out of } N}$$

$$p(y) = \frac{(y \text{ successes} \cap n-y \text{ failures})}{(\text{choices of } n \text{ out of } N)}$$

- always $0 < \text{probability} < 1$
- sum of all probabilities must be 1.

Expected value: (Averege)

$$\mu = \frac{nr}{N} \Rightarrow \mu = np$$

Variance:

$$\sigma^2 = \frac{nr}{N}$$

- The probability of SUCCESS changes on each trial
- Fixed number of trials

POISSON Distribution

- Rate of Successes

Rate = Occurance per some unit

e.g

The hourly arrivals to a machine in a production process

→ Random variable (X) counts the number of Occurances

→ Every occurance is independent

- The random variable must be discrete.
- The rate does 't have to be discrete, because its average

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ or } \frac{\mu^x e^{-\mu}}{x!}$$

x = rate of occurance
e = 2.78

Variance is $\sigma^2 = \lambda^2$. Therefore, Standard deviation is λ.

- It doesn't have an upper bound
- Lowerbound is 0. cant have negative acount of occurances.

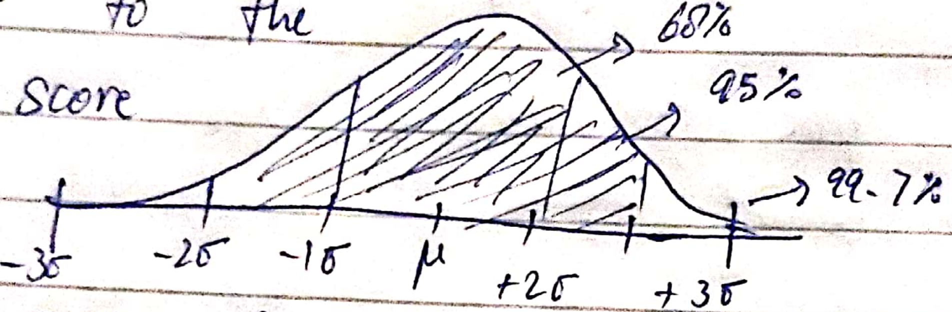
Normal Deviation

→ Probability ⇒ From given Charts

+ve and −ve

$$\boxed{Z = \frac{x-\mu}{\sigma}}$$

→ Chart always gives you the value left to the Z - score



→ Bell looking curve

→ Falls between −3σ standard deviation to +3σ

→ All data should be in same unit

For Standardized Data

$$Z = \frac{x-\mu}{\sigma}$$

μ → Mean

σ → Standard Deviation

σ^2 → Variance

Scanned with CamScanner

Regression

Linear Regression / Relation b/w variables

The dependance of one variable, called the dependant variable. others are called independant variables.

→ The relation between the expected value of the dependant variable and the independant variable is called regression relation.

→ Linear regression is generally written as:

$$D.V \rightarrow Y = a + b\,X\ (+\,\varepsilon) \rightarrow \text{error term}$$

$\uparrow$ I.V — //can be neglected

Where

$a \Rightarrow$ the value of Y when $X = 0$ and called Y- intercept

$b \Rightarrow$ indicates the change in Y when one-unit change in X and called slop of the line

$$a = \frac{\Sigma X^2 \Sigma Y - \Sigma X \Sigma XY}{n\Sigma X^2 - (\Sigma X)^2}$$

$$b = \frac{\Sigma (X-\bar{X})(Y-\bar{Y})}{\Sigma (X-\bar{X})^2}$$

-OR-

$$b = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2}$$

Correlation

→ Measure of the degree to which any two variables vary together

→ No distinction between dependent and Independent Variable.

→ If (the) one variable increase as the other one decreases, the correlation is said to be negative or inverse.

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}}$$