PROBABILITY AND STATISTICS

UNIT -1

Trial : Conducting an experiment only one time is called trial.
Ex: Tossing of coin, Rolling of dice, we draw one card from a pack of cards, etc.,

OUTCOME: The result of a trial in a experiment is called an outcome.

EVENTS : The outcomes of an experiment are known as events.

RANDOM EXPERIMENT : An experiment whose result is not determined by us but determined by randomly and naturally is called random experiment.

SAMPLE SPACE: A set of all possible outcomes of an experiment is called sample space. Or The total no. of favourable and un favourable events is also called sample space. It is denoted by S or Ω

EXHAUSTIVE EVENTS: The total no of possible outcomes in any trial is know as exhaustive events or exhaustive cases. It is also called sample space.
Ex: i) In tossing of coin there are two exhaustive events i.e., 2^1 =2 viz., head (H) and tail (T)
ii) In throwing of a die, there are six exhaustive events i.e., 6^1 =6viz., 1,2,3,4,5,6 since any one of the six faces may come upper most
iii) In drawing two cards from a pack of cards the exhaustive number of cases is 52_{C_2} ways.
In general in tossing of n coins the exhaustive events is i.e., 2^n , in throwing of n dice the exhaustive events is i.e., 6^n

FAVOURABLE EVENTS : The no. of events are favourable for happening of an event is said to be favourable events.
Ex: In tossing of two coins at a time the no. of events favourable to getting the exactly one head is 2 i.e., { HT,TH } or at least one head is 3 i.e., { HH,HT,TH }

EQUALLY LIKELY EVENTS : Events are said to be equally likely events, if happening of one event has equal chance to happening of other event.
Ex: In tossing of a coin Head or Tail are equally likely events ii) In throwing of a die all the six faces are equally likely events.

MUTUALLY EXCLUSIVE EVENTS: Events are said to be mutually exclusive events, if happening of one event exclude the happening of other event.
Ex: In tossing of a coin Head and Tail are mutually exclusive events ii) In throwing of a die all the six faces numbered 1 to 6 are mutually exclusive events.

INDEPENDENT EVENTS: Events are said to be independent if the happening of an event is not affected by the happening of another event.
Ex: To getting a head on the coin is independent of obtaining a five on the die. Let A and B are any two independent events then $P(A \cap B) = P(A).P(B)$

DEFINITIONS OF PROBABILITY: There are three definitions of probability
I) MATHEMATICAL OR CLASSICAL OR A PRIORI DEFINITION OF PROBABILITY : If a trial result in 'n' exhaustive ,equally likely and mutually exclusive events and 'm' events are favourable to the happening of an event E, then the probability of happening of an event E is denoted by P(E) .
It is defined as $P(E) = \frac{No.of\ favourable\,events(m)}{Exhaustive\,evets(n)}$ i.e., $P(E) = \frac{m}{n}$
LIMITATIONS : i) If the various outcomes of the trial are not equally likely or equally probable
ii) If the exhaustive no. of cases in a trial is infinite.

II) STATISTICAL OR EMPIRICAL OR VON MISS DEFINITION OF PROBABILITY: If a trial is repeated a no. of times under essentially homogeneous and identical conditions then the limiting value of the ratio of the no. of times the event happens as the no. of trials, become indefinitely large, is called statistical definition of probability. Symbolically, if in 'n' trials an event E happens 'm' times then the probability of the happening of E is given by $P(E) = \lim_{n \to \infty} \frac{m}{n}$

LIMITATIONS : i) If an experiment is repeated a large number of times, the experiment conditions may not remain identical and homogeneous.

III) THE AXIOMS OF PROBABILITY: The axioms of probability are

i) The probability value always must be positive i.e., $P(E) \geq 0$

ii) The probability value always lies in between 0 to 1 i.e., $0 \leq P(E) \leq 1$

iii) The total probability equal to one i.e., $P(S) = 1$ or $P(\Omega) = 1$ or $\sum P(E) = 1$

iv) If $E_1, E_2, \ldots E_n$ are n mutually exclusive events the $P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$

CONDTIONAL PROBABILITY: If A, B are any two events then the conditional probability of occurrence of A when the event B has already happened is called conditional probability A. It is denoted by P(A/B) and this is defined as $P(A/B) = \frac{P(A \cap B)}{P(B)}$, $P(B) \neq 0$ Similarly, The conditional probability of occurrence of B when the event A has already happened is called conditional probability B. It is denoted by P(B/A) and this is defined as $P(B/A) = \frac{P(A \cap B)}{P(A)}$, when $P(A) \neq 0$.

ADDITION THEOREM ON PROBABILITY : If A, B and C are any three events then

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

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

II. If A, B and C are any three mutually exclusive events then

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

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

MULTIPLICATION THEOREM ON PROBABILITY : If A, B and C are any three events then

I i) $P(A \cap B) = P(B) . P(A/B)$ or

$P(A \cap B) = P(A) . P(B/A)$

ii) $P(A \cap B \cap C) = P(A) . P(B/A) . P(C/A \cap B)$

II If A, B and C are any three independent events then

i) $P(A \cap B) = P(A) . P(B)$

ii) $P(A \cap B \cap C) = P(A) . P(B) . P(C)$

BAYES THEOREM : Suppose $E_1, E_2, \ldots, E_n$ are n mutually exclusive events of a sample space S such that $P(E_i) > 0$, i = 1,2,.., n and 'A' is any arbitrary event of S such that P(A) > 0, and A $\cup E_i = S$ then the conditional probability of E_i given A is denoted by $P(E_i/A)$ and it can be defined as

$$P(E_i/A) = \frac{P(E_i) . P(A/E_i)}{\sum P(E_i) . P(A/E_i)}$$

Where $P(E_1), P(E_2), \ldots, P(E_n)$ are termed as the a priori probabilities

$P(A/E_i)$, i = 1,2,..,n are called posterior probabilities

UNIT - 2

LARGE SAMPLE TESTS : If the sample values ($n \geq 30$) then it is called large samples. The tests belongs to small samples is called large sample tests.

Test (1): Test for significance difference between sample mean and population mean

(i) If $x_1, x_2 \dots x_n$ are n sample values are drawn from a normal population with mean μ and variance σ^2 then we calculate sample mean $\bar{x}$ and sample S.D (s) in the following way.

Sample Mean($\bar{x}$) $= \frac{\sum x}{n}$, Sample Standard Deviation(s)=$\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$, n: No. of sample values

(ii) For testing the significant difference between sample mean and population mean we formulate the null hypothesis Ho as

Ho : There is no significant difference between sample mean and population mean

i.e Ho : $\bar{x} = \mu$ Here $\bar{x}$ = sample mean, μ = Population mean (iii)

To test the above hypothesis the suitable test statistics 'Z' as

$$Z = \frac{Statistic - Parameter}{Standard\ Error} \rightarrow Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ as } n \to \infty \text{ by C LT}$$

(iv) We substitute different values in the above test we get calculated value of $|Z|$ we take table value of Z at α % level of significance (los) and compare this value calculated value of $|Z|$

(a) If calculated value of $|Z| \geq Z$ table value then we reject the null hypothesis.
i.e ., There is significant difference between sample mean and population mean.

(b) If calculated value of $|Z| < Z$ table value then we do not reject(accept) the null hypothesis.
i.e., There is no significance difference between sample mean and population mean

Conclusion can be drawn according to rejection or non-rejection of the null hypothesis.

(v) The 100(1-α)% Confidence limits for un known parameter μ is

$$\text{Statistic} \pm t_{\propto\%} \text{ (Standard Error)} = \bar{x} \pm Z_{\propto\%} \left\{\sigma/\sqrt{n}\right\}$$

Test (2): Test for significance difference between two sample means

Case I: (i)Let $x_{11}, x_{12} \dots x_{1n_1}$ be the sample values are drawn from normal population with mean μ_1 and variance σ_1^2 and $x_{21}, x_{22} \dots x_{2n_2}$ be the sample values are drawn from another normal population with mean μ_2 and variance σ_2^2. By using sample values we calculate sample means $\bar{x}_1, \bar{x}_2$, sample S.Ds S_1 , S_2 in the following way .

$$\bar{x}_1 = \frac{\sum x_{1i}}{n_1}, \ \bar{x}_2 = \frac{\sum x_{2i}}{n_2}, S_1 = \sqrt{\frac{\sum x_{1i}^2}{n_1} - \left(\frac{\sum x_{1i}}{n_1}\right)^2} \quad S_2 = \sqrt{\frac{\sum x_{2i}^2}{n_2} - \left(\frac{\sum x_{2i}}{n_2}\right)^2}$$

(ii) For testing the significant difference between two sample means we formulate the null hypothesis as

Ho: There is no significant difference between two sample means

i.e Ho : $\bar{x}_1 = \bar{x}_2$, Here $\bar{x}_1$ = First sample mean, $\bar{x}_2$ = Second sample mean

(iii) To test the above hypothesis the suitable test statistics ' Z ' as

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if known population S.D(**σ**)

Or

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if un known population S.D (**σ**)

Rest of the procedure as usual

(v) The 100(1-α)% Confidence limits for un known parameter μ_1-μ_2 is

$$\text{Statistic} \pm t_{\propto\%} \text{ (Standard Error)} = |\bar{x}_1 - \bar{x}_2| \pm t_{\propto\%} \left\{ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \; or \; \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$$

Case II : If two sample values are drawn from Single population then for testing the significant difference between two sample means we use the following test statistic as

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if known population S.D (**σ**)

Or

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if un known population S.D(**σ**)

Where $$\hat{\sigma}^2 = \frac{n_1 . s_1^2 + n_2 . s_2^2}{n_1 + n_2}$$

Test (3): Test for significance difference between sample proportion and population proportion.

(i)

If $x_1, x_2 \ldots x_n$ be n sample values are drawn from a binomial population with mean $\boldsymbol{nP}$ and variance $\boldsymbol{nPQ}$**.** Let we denote the special no of characters in the sample is denoted with x. with the help of sample values we find sample proportion by using the following formula

Sample Proportion $(\boldsymbol{p}) = \frac{\text{Special no.of characters in the sample } (x)}{\text{Sample size } (n)}$, i.e., $p = \frac{x}{n}$

(ii) For testing significant difference between sample proportion and population proportion we formulate the null hypothesis Ho as

Ho: There is no significant difference between sample proportion and population proportion

i.e Ho : $\boldsymbol{p} = P$ Here p = sample proportion, P = Population Proportion

(iii) To test the above hypothesis the suitable test statistics ' Z ' as

$$Z = \frac{\text{Statistic} - \text{Parameter}}{\text{Standard Error}} \rightarrow Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1) \text{ as } n \to \infty \text{ by C LT}$$

(v) The 100(1-α)% Confidence limits for un known parameter $\boldsymbol{P}$ is $p \pm z_{\propto\%} \left\{ \sqrt{\frac{PQ}{n}} \right\}$

Test(4): Test for significance difference between two sample proportions

Case I: (i) Let $x_{11}, x_{12} \dots x_{1n_1}$ be the n_1 sample values are drawn from binomial population with mean $n_1 P_1$ and variance $n_1 P_1 Q_1$ and $x_{21}, x_{22} \dots x_{2n_2}$ be the n_2 sample values are drawn from another binomial population with mean $n_2 P_2$ and variance $n_2 P_2 Q_2$.
By using sample values we calculate sample proportions in the following way .

$$p_1 = \frac{\sum x_{1i}}{n_1} \qquad p_2 = \frac{\sum x_{2i}}{n_2}$$

(ii)For testing the significant difference between two sample proportions. We formulate the null hypothesis as **Ho:** There is no significant difference between two sample proportions

i.e Ho : $p_1 = p_2$, Here p_1 = First sample proportion , p_2 = Second sample proportion

(iii) To test the above hypothesis the suitable test statistics ' Z ' as

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if known population proportion (P)

Or

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if un known population proportion(***P***)

Rest of the procedure as a usual

(v) The 100(1-α)% Confidence limits for un known parameter $\boldsymbol{P_1}$-$\boldsymbol{P_2}$ is

Statistic ± $t_{\propto\%}$ (Standard Error) = $|p_1 - p_2| \pm t_{\propto\%} \left\{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}} \; or \; \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right\}$

Case II : If two sample values are drawn from Single population then for testing the significant difference between two sample proportions we use the following test statistic as

$$t = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if known population proportion(P)

Or

$$t = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad N(0,1) \text{ as } n_1, n_2 \to \infty \text{ by CLT}$$, if un known population proportion(P)

Where $$\hat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

UNIT - 3

Small Sample Tests : If the sample values (n < 30) then it is called small samples. The tests belongs to small samples is called small sample tests.

Small sample tests for Means

Test (1): Test for significance difference between sample mean and population mean

If $x_1, x_2 \ldots x_n$ are n sample values drawn from a normal population with mean μ and variance $\boldsymbol{\sigma}^2$ then we calculate sample mean $\bar{x}$ and sample S.D (s)in the following way Sample

Mean$(\bar{x}) = \frac{\sum x}{n}$, Sample Standard Deviation(s)=$\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$, n: No. of sample values

For testing significant difference between sample mean and population mean we formulate the null hypothesis Ho as

Ho : There is no significant difference between sample mean and population mean

i.e Ho : $\bar{x} = \mu$ Here $\bar{x}$ = sample mean, μ = Population mean

To test the above hypothesis the suitable test statistics ' t ' as

$$t = \frac{Statistic - Parameter}{Standard\ Error} \rightarrow t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \sim t_{(n-1)d.f}$$

We substitute different values in the above test we get calculated value of $|\boldsymbol{t}|$ and compare this value with table value of $\boldsymbol{t}$ at α % level of significance (los) with $(n_1 - 1)d.f$

Rest of the procedure as a usual

Test (2): Test for significant difference between two sample means for small samples

If two independent sample values $x_1, x_2 \ldots x_n$ and $y_1, y_2 \ldots y_n$ of sizes n_1 and n_2 have been drawn from two normal populations with means μ_x and μ_y. By using sample values we calculate sample means $\bar{x}, \bar{y}$, sample variances S_x , S_y in the following way .

$$\bar{x} = \frac{\sum x}{n}, \quad \bar{y} = \frac{\sum y}{n} \quad , \quad s_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}, \qquad s_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

For testing the significant difference between two sample means we formulate the null hypothesis as Ho: There is no significant difference between two sample means

i.e Ho : $\bar{x} = \bar{y}$, Here $\bar{x}$ = First sample mean, $\bar{y}$ = Second sample mean

To test the above hypothesis the suitable test statistics ' t ' as

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)d.f}$$

Where $$\hat{\sigma}^2 = \frac{n_1.s_1^2 + n_2.s_2^2}{n_1+n_2-2}$$

We substitute different values in the above test we get calculated value of $|\boldsymbol{t}|$ and compare this value with table value of $\boldsymbol{t}$ at α % level of significance (los) with $(n_1 + n_2 - 1)d.f$

Rest of the procedure as a usual

Paired t - test for difference of means (or) Correlated variables

Let $(x_i\, y_i), i = 1,2,\ldots,n$ be the paired sample values are drawn from a normal population. we consider when the sample size are equal i.e $n = \mathrm{n}_1 = \mathrm{n}_2$. The problem is to test the sample means differ significantly or not.

For example we want to test the performance of the students before giving the coaching and after giving the coaching , the performance of the patients before giving the drug and after giving the drug. Let x_i and $y_i (i = 1,2,\ldots,n)$ be the readings in hours of sleep, on the i[th] individual, before and after the drug is given respectively. For testing the significant difference between the means of paired sample values or correlated samples , we formulate the null hypothesis as follows.

Ho : The drug is not significant.

To test the above hypothesis we use the following test statistic t as

$$t = \frac{\bar{d}}{Sd/\sqrt{n-1}} \sim t_{(n-1)d.f}$$

Where $d = x_i - y_i$, $\bar{d} = \frac{\sum d}{n}$, $Sd = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$

We substitute different values in the above test we get calculated value of $|\boldsymbol{t}|$ and compare this value with table value of $\boldsymbol{t}$ at α % level of significance (los) with $(n-1)d.f$

Rest of the procedure as a usual

F-TEST Or Testing the significant difference between the two sample variances.

If two independent sample values $x_1, x_2 \ldots x_n$ and $y_1, y_2 \ldots y_n$ of sizes n_1 and n_2 have been drawn from two normal populations with same variances σ^2 .

With the help of sample values we calculate sample variances by using the following formulae.

First sample variance $S_1^2 = \frac{\sum x^2}{n_1} - \left(\frac{\sum x}{n_1}\right)^2$, Second sample variance $S_2^2 = \frac{\sum y^2}{n_2} - \left(\frac{\sum y}{n_2}\right)^2$,

For testing the significant difference between two sample variances, we formulate the null hypothesis as
Ho : There is no significant difference between two sample variances or population variances
i.e Ho : $S_1^2 = S_2^2$

To test the above hypothesis we use the following test statistic F as

$$F = \frac{n_1 S_1^2/(n_1-1)}{n_2 S_2^2/(n_2-1)} \sim F_{\{(n_1-1),(n_2-1)\}}\ d.f \quad \text{if } S_1^2 > S_2^2$$

Or

$$F = \frac{n_2 S_2^2/(n_2-1)}{n_1 S_1^2/(n_1-1)} \sim F_{\{(n_2-1),(n_1-1)\}}\ d.f \quad \text{if } S_2^2 > S_1^2$$

We substitute different values in the above test we get calculated value of F and compare this value with table value of F at α % level of significance (los) with $(n_1-1),(n_2-1)$ or $(n_2-1),(n_1-1) d.f$

Rest of the procedure as usual

χ^2 TEST (Or) Testing the significant difference between the sample variance and population variance :

If $x_1, x_2 \dots x_n$ are n sample values drawn from a normal population with a specified variance σ^2. With the help of sample values we calculate sample variance by using the following formula.

Sample Variance $$S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

For testing significant difference between sample variance s^2 and population variance σ^2we formulate the null hypothesis Ho as

Ho : There is no significant difference between sample variance(S^2) and population variance(σ^2)

To test the hypothesis the test statistic is $\chi^2 = \frac{n\, s^2}{\sigma^2} \sim \chi^2_{(n-1)d.f}$

We substitute different values in the above test we get calculated value of χ^2 and compare this value with table value of χ^2 at α % level of significance (los) with $(n-1)d.f$

Rest of the procedure as usual

χ^2 Test for goodness of fit

Let O_i (i=1,2,….,n) be the observed(given) frequencies and E_i(i=1,2,….,n) be the corresponding expected frequencies then to test the significant difference between observed and expected frequencies, we formulate the null hypothesis Ho as follows

Ho: There no significant difference between observed and expected frequencies

To test the above hypothesis the test statistic χ^2 as , $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(n-1)d.f}$

We substitute different values in the above test we get calculated value of χ^2 we take table value of χ^2 at α % level of significance (los) with $(n-1)\ d.f$

Rest of the procedure as usual

χ^2 Test for independence of attributes.

In the given table cell frequency are known as observed frequency i.e O_{ij}. From this observed frequencies we calculate expected frequencies i.e (E_{ij}) by using the following formula

Expected Frequency $E_{ij} = \frac{i^{th}\ Row\ total\ (A_i) \times j^{th}\ Column\ total(B_j)}{Grand\ total\ (N)}$ **i.e** $E_{ij} = \frac{(A_i) \times (B_j)}{(N)}$

For testing independence of the attributes or Testing the significant difference between observed and expected frequencies , we formulate the null hypothesis as follows.

Ho: The two attributes are independent or

Ho: There is no significant difference between observed and expected frequencies.

To test the above hypothesis to be true the suitable test statistics χ^2 **as**

$$\chi^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)d.f}$$ **, r = No. of rows, c = No. of columns**

If we substitute different values in the above formula then we get calculated value of χ^2 and compare this value with table value of χ^2 at α % level of significance (los) with $(r-1)(c-1)d.f$