Probability

Sample space & events

→ A random exp is a mechanism that produces definite outcome that cannot be predicted with certainity.

→ A sample space associated with a random exp is set of all possible outcom

→ An event is subset of sample space

eg - tossing a coin

The outcomes can be labeled h for heads and t for tail.

$S = \{h, t\}$

→ Probability of an outcome e in a sample space S is a number p between 0 & 1 that measures likelihood that e will occur on single trial of corresponding random exp.

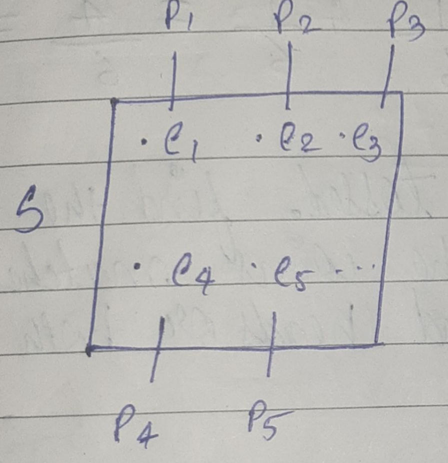
$p = 0$ impossible
$p = 1$ beingcertain

→ probability of event A is sum of probability of individual outcomes of which it is composed.

it is denoted by $P(A)$.

→ if an event $E = \{e_1, e_2, e_3, \ldots, e_k\}$

than $P(E) = P(e_1) + P(e_2) + \ldots + P(e_k)$



$P(A) = P_1 + P_2$
$P(B) = P_1 + P_2 + P_3$

$A = \{e_1, e_2\}$
$B = \{e_1, e_2, e_3\}$

→ Sum of probability of all outcomes is 1.

eg - Assign a probability to each outcome in sample space for exp that consist of tossing a single fair die. find the probability of the event E : " an even number is rolled" & T : " a no greater than two is rolled".

~~P(E)~~ - $S = \{1, 2, 3, 4, 5, 6\}$

$E = \{2, 4, 6\}$

$$P(E) = \frac{3}{6} = \frac{1}{2}$$

$T = \{3, 4, 5, 6\}$

$$P(T) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

eg- two coins are tossed. find the probability that the coins matches, i.e either both land heads or both land tails.

$E = \{TH, TT, HT, HH\}$

$$P(E) = \frac{1}{4} + \frac{1}{4} = 0.5$$

Complements, Intersections and Unions.

→ The complement of an event A is sample space S, denoted by A^c is collection of all outcome of S which are not elements of A.

eg - find complements of E & T in example of die.

$E^c = S - E$

ie all odd no.

$E^c = \{1, 3, 5\}$

$T^c = \{1, 2\}$ no. not greater than 2. ie no rolled is less than 3.

probability rule for complement

$$\boxed{P(A^c) = 1 - P(A)}$$

eg - find the probability that atleast one head will appear in 5 tosses of a fair coin.

no. of possible outcomes $= 2^5$
$= \boxed{32}$

E = atleast one head will appear in five tosses.

P(E) = sample − ~~Only~~ zero head will appear.

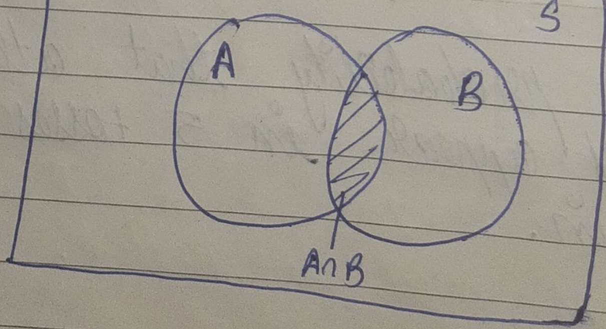
$= 1 - \frac{1}{32}$

$= \frac{31}{32} = 0.967$
$\approx 97\%$ chance.

Intersection of event:

→ intersection of event A & B, denoted by $A \cap B$, is collection of all outcomes that are elements of both of sets A & B.

→ corresponds to combining description of two events using word "and".



eg – no. should be even & greater than 2. in rolling dice.

$E \cap T = \{4, 6\}$

~~P(E~~

eg – a) find probability that no rolled is both even & greater than two.

$P(E \cap T) = \frac{2}{6} = \boxed{\frac{1}{3}}$

b) Suppose the die has been "loaded" so that $P(1) = 1/12$. $P(6) = 3/12$, & remaining four outcomes are equally likely with one another.
find probability the no. rolled is both even & greater than two.

$P(1) = \frac{1}{12}$ $P(6) = \frac{3}{12}$

$P(2) = P(3) = P(4) = P(5) = \frac{2}{12} = P$

$P(E \cap T) = \frac{2}{12} + \frac{3}{12} = \boxed{\frac{5}{12}}$

→ Event A & B are mutually exclusive if they have no elements in common.

$P(A \cap B) = 0$

eg - find 3 choices for an event A so that event A & E "the no rolled is even" are mutually exclusive.

$P(E) = \{2, 4, 6\}$

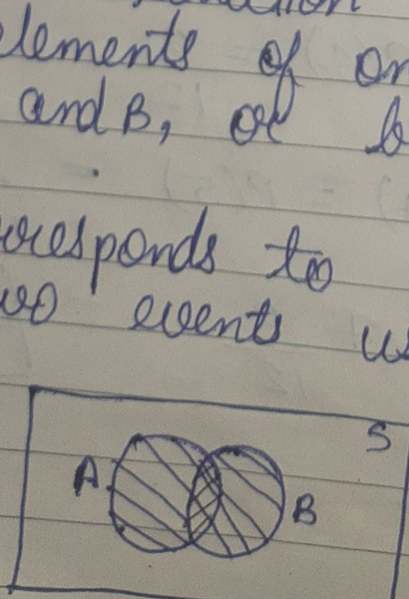
$P(A) = S - P(E)$
$= \{1, 3, 5\}, \{1, 3\}, \{1, 5\}, \{3, 5\}$
$\{1\}, \{3\}, \{5\}$.

18/10/22

<u>Union of events</u>

→ $A \cup B$ is collection of all outcomes that are elements of one or the other of the sets A and B, or both of them.

→ It corresponds to combining discriptions of two events using word "or".



Page No.
Date :

Example:-

Single die E: no rolled is even
T: no rolled is greater than two.

$E = \{2, 4, 6\}$ $S = \{1, 2, 3, 4, 5, 6\}$
$T = \{3, 4, 5, 6\}$

$E \cup T = \{2, 3, 4, 5, 6\} = S - \{1\}$

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

Example:-
two fair dice rolled.

a. both dice shows a four.
b. at least one die shows four.

$S = \{\{1,1\}\ \{1,2\}\ (1,3)\ (1,4)\ (1,5)\ (1,6)$
$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$
$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$
$(4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6)$
$(5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (6,6)$,
$(6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)\}$

E: at least one die shows 4.

$P(E) = \frac{1\cancel{6}}{36} = \frac{\cancel{1}}{\cancel{6}}\ \frac{11}{36}$

Page No.
Date :

T : both 4

$P(T) = \frac{1}{36}$

using formula :- (4 on both).

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

A : die 1 shows 4 $P(A) = 1/6$

B : die 2 shows 4 $P(B) = 1/6$

$P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$

- Conditional Probability & Independent Events :

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

→ the probability that event A has occured in a trial of a random experiment for which it is known that event B has surely occured.

eg :- die is rolled.

a) find probability that no. rolled is five, given that it is odd.

b) no. rolled is odd & given that it is 5.

Page No.
Date :

A: no. rolled is 5.

$P(A) = \frac{1}{6}$

B: no. is odd $= \{1, 3, 5\}$

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

$A \cap B = \frac{1}{6}$

a) $P(A/B) = \frac{1/6}{3/6} = \boxed{\frac{1}{3}}$

b) $P(B/A) = \frac{1/6}{1/6} = 1$.

Independent events :-

→ Accurance of event B has no effect on event A.

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

events A & B are indepent if and only if above formula is true.

die is rolled.

eg: A : {3}   B : {1, 3, 5}

are A & B independent?

$A \cap B = \{3\} = 1$ $P(A) = \frac{1}{6}$

$P(A \cap B) = 1 \times \frac{1}{3} = \frac{1}{3}$ $P(B) = \frac{\cancel{3}}{\cancel{6}\,2} = \frac{1}{2}$

$P(A \cap B) = \frac{1}{6} \times \frac{3}{6} = \frac{1}{12}$

$P(A \cap B) = \frac{1}{6}$

dependent   LHS $\neq$ RHS

Eg:- A: person has disease = 92%.

a) both test results +ve.

b) at least one of the two test results will be +ve.

a) $P(A \cap B) = P(A) \cdot P(B)$

$= (0.92)\,(0.08)$

$= 0.0736$

b) $P(A \cup B) = 0.92 + 0.08 - 0.0736$

$= 0.9264$

Page No.
Date :

eg :-
Test results +ve by lab 1 : A
lab 2 : B

given : $P(A) = P(B) = 0.92$

a) both +ve

$$P(A \cap B) = (0.92)^2 = 0.8464$$

b) at least one of two results +ve.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 2(0.92) - 0.8464$$
$$= 0.9936.$$

Random Variable.

→ Numerical variable generated by random experiment.

denoted by capital letters
value → small.

→ Random variable is called discrete if it is ~~of~~ either a finite or a countable no. of possible values.

→ Random variable is called contious if

its possible values contain a whole interval of numbers.

→ The probabilities in probability distribution of random variable X must satisfy following two conditions.

(i) $0 \leq P(x) \leq 1$
(ii) $\sum P(P(x)) = 1$.

eg - A fair coin is tossed twice. let X be the number of heads that are observed.

a) Construct the probability distribution of X
b) find probability that atleast one head is observed.

$\{\{(H, T)\ (H, H)\ (T, H)\ (T, T)\} = S$

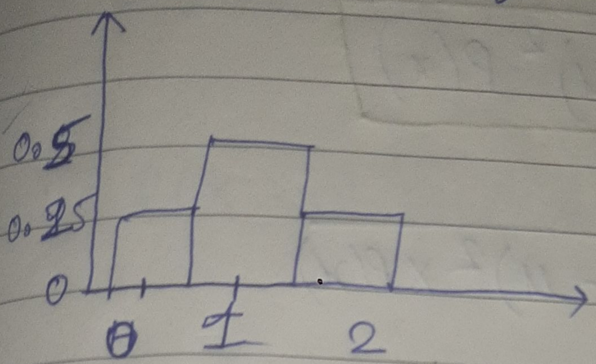
no. of heads (x)	0	1	2	~~8~~
P(x)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	~~4~~

at least one head :-

$P(x \geq 1) = \cancel{\frac{1}{4}} + \frac{2}{4} + \frac{1}{4} = \cancel{\frac{3}{4}}$

$= \frac{3}{4}$

(also draw histogram - by mam)



Mean & Standard deviation of Discrete Random Variable:-

The mean (also called expected value) of discrete random variable x is the number

$$\boxed{\mu = E(X) = \Sigma x \, P(x)}$$

eg:-

x	-2	1	2	3.5
$P(x)$	0.21	0.34	0.24	0.21

$\mu = \Sigma x \, P(x)$

$= -0.42 + 0.34 + 0.48 + 0.735$

$$\boxed{\mu = 1.135}$$

Date :

Variance

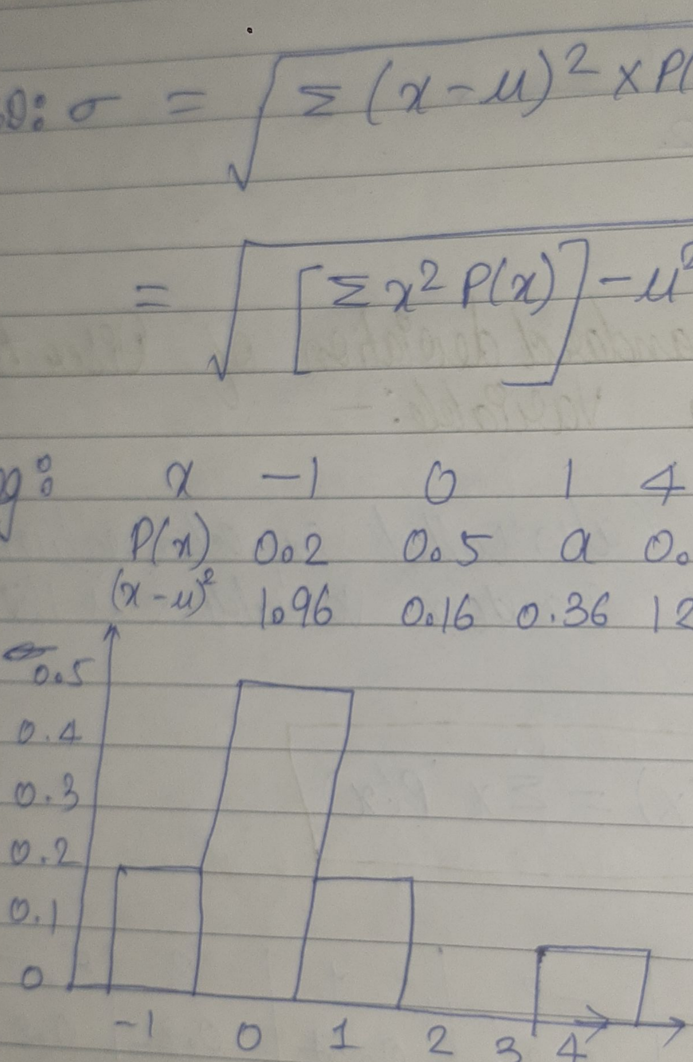
$$\boxed{\sigma^2 = \Sigma (x-\mu)^2 P(x)}$$

SD: $\sigma = \sqrt{\Sigma (x-\mu)^2 \times P(x)}$

$= \sqrt{\left[\Sigma x^2 P(x)\right] - \mu^2}$

eg:

x	-1	0	1	4
$P(x)$	0.2	0.5	a	0.1
$(x-\mu)^2$	1.96	0.16	0.36	12.96



a) value of a

$\Sigma P(x) = 1$

$0.2 + 0.5 + a + 0.1 = 1$

$a + 0.8 = 1$

$\boxed{a = 0.2}$

b) $P(0) = 0.5$

c) $P(X > 0) = P(1) + P(4)$
$= 0.3$

d) $P(X \geq 0) = P(0) + P(1) + P(4)$
$= 0.8$

e) $P(X \leq -2) = 0$

f) The mean μ of X

$\mu = \Sigma x\, P(x)$

$= -0.2 + 0.2 + 0.4$

$\boxed{\mu = 0.4}$

g) variance σ^2 of X.

$\sigma^2 = \Sigma (x - \mu)^2 P(x)$
$= 0.392 + 0.08 + 0.072 + 1.296$
$\boxed{\sigma^2 = 1.84}$

h) Standard deviation σ of X.

$\boxed{\sigma = 1.3565}$

ANOVA (Analysis of Variance)

→ Objective is to decide whether the means for more than two populations or treatments are identical

→ When two or more populations or treatments are compared, the characteristic that distinguishes the population or treatment from one another is "factor".

→ A single factor analysis of variance (ANOVA) problem involves comparism of k population or treatment means $\mu_1, \mu_2 \ldots \mu_K$. The objective is to test.

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_K$

against

H_a : at least two of μ's are different

ANOVA Notations :

K = no. of populations or treatment being compared.

$N = n_1 + n_2 + \ldots + n_K$ (total no. of oberservations in data set)

T = (grand total) = sum of N observations

$= n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3 + \ldots + n_K \bar{x}_K$

$\bar{\bar{x}}$ = grand mean $= \frac{}{N}$

	1	2	...	K
Population or treatment mean	μ_1	μ_2	...	μ_K
Population or treatment variance	σ_1^2	σ_2^2	...	σ_K^2
Sample size	n_1	n_2	...	n_K
Sample mean	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_K$
Sample variance	S_1^2	S_2^2	...	S_K^2

the validity of the Anova test for $H_0: \mu_1 = \mu_2 = \dots = \mu_K$ requires some assumptions.

Assumptions of Anova :-

1. Each of K response population or treatment responses distribution is normal.

2. $\sigma_1 = \sigma_2 \dots = \sigma_K$
3. Observations are independent of samples.
4. When comparing population means, random samples selected independently of one another.

When comparing treatment means, treatments are assigned at random to subjects or objects.

df → degree of freedom

→ Defenation:

→ A measure of disparity among the sample means is the treatment sum of squares, denoted by SSTr.

$$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \dots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

→ A measure of variance within the k samples, called error sum of squares and denoted by SSE.

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2$$

→ Each sum of squares has an associated df:

treatment df: $k-1$ error df $= N-k$

→ A mean square is sum of squares devided by its ~~diff~~ df.

→ Mean sq for treatment $= MSTr = \dfrac{SSTr}{k-1}$

→ Mean square for error $= MSE = \dfrac{SSE}{N-k}$.

ANOVA Table :-

Source of Variation	df	Sum of Squares	Mean Square	F
Treatment	$k-1$	SSTr	$MSTr = \frac{SSTr}{k-1}$	$F = \frac{MSTr}{MSE}$
error	$N-k$	SSE	$MSE = \frac{SSE}{N-k}$	
Total	$N-1$	SSTo		

→ Test statistic for testing the null hypothesis that k population means are equal.

$$F = \frac{MST}{MSE}$$

right tailed → Ho is rejected at level of significance α, $F \geq f_\alpha$.

F is always right tailed.

Que $\alpha = 5\%$

no. of student = 11

" subjects = 4

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: not all four population means are equal.

	Mathematics	english	education	Biology
$\bar{x}$	2.9027	3.3354	3.35909	3.0154

$N = n_1 + n_2 + n_3 + n_4$ $\quad n_1 = n_2 = n_3 = n_4$

$= 4n = 44$

$T = n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3 + n_4\bar{x}_4$

$= 11[2.9027 + 3.3357 + 3.35909 + 3.0154]$

$= 138.84849$

$$\bar{\bar{x}} = \frac{T}{N} = \frac{138.84849}{44}$$

$$\boxed{\bar{\bar{x}} = 3.1556}$$

$SSTr = 11[0.06395 + 0.083232 + 0.0414 + 0.01965]$

$= \boxed{1.73052}$

$$MSTr = \frac{SSTr}{k-1} = \frac{1.73052}{4-1}$$

$$\boxed{MSTr = 0.57684}$$

Page No.
Date :

$$S^2 = \frac{\Sigma (x-\bar{x})^2}{n-1}$$

$$= [0.0977 + 0.05166 + 0.0045 + 0.1621 + 0.1389 + 0.15 + 0.1389 + 0.07144 + 0.041 + \cancel{0.9883}\ 0.955 + 0.069] / 10$$

$\boxed{S_1^2 = 0.188}$

$S_2^2 = 0.147$

$S_3^2 = 0.20984$

$S_4^3 = 0.157$

$SSE = n \cdot 10 [0.70184]$

$\boxed{SSE = 7.0184}$

$$MSE = \frac{SSE}{N-K} = \frac{7.0184}{44-4}$$

$\boxed{MSE = 0.17546}$

$$F = \frac{MST_r}{MSE} = \frac{0.57684}{0.17546}$$

$\boxed{F = 3.2875}$

Page No.
Date :

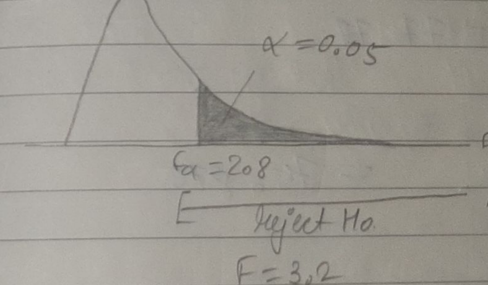
$df_1 = 3$
$df_2 = 40$

from table, $\alpha = 0.05$

$F\alpha = 2.8387$

$F_0 > F\alpha$

$\therefore$ null hypothesis rejected.



7/11
Que $\alpha = 1\% = 0.01$

→ whether these data provide sufficient evidence to confirm the belief that at least one of two treatments affects the average survival time of mice with thymic leukemia.

$H_0: \mu_1 = \mu_2 = \mu_3$

H_a: not all hypothesis are equal at $\alpha = 0.01$

Treatment
1 2 3

$\bar{x}$

$\bar{x}_1 = \frac{1116}{16} = 69.75$

$\bar{x}_2 = \frac{700}{9} = 77.77$

$\bar{x}_3 = \frac{607}{8} = 75.875$

$N = 16 + 9 + 8 = 33$

$T = n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3$
$= 1116 + 700 + 607$
$= 2423$

$\bar{\bar{x}} = \frac{T}{N} = 73.42$

$SSTr = 16(13.46) + 9(18.92) + 8(2.45)$
$= \cancel{405.24}\ 433.83$

$s_1^2 = \frac{349\cancel{5}.7}{16-1} = 34.6$

$s_2^2 = \frac{421}{9-1} = 52.7$

$s_3^2 = \frac{214.9}{8-1} = 30.7$

$SSE = 15(34.6) + 8(52.7) + 7(30.7)$
$= 1155.5$

$df_1 = k - 1 = 2$

$df_2 = N - k$
$= 33 - 3 = 30$

$MSTr = \frac{SSTr}{k-1} = \frac{433.83}{2} = 216.9$

$MSE = \frac{SSE}{N-k} = \frac{1155.5}{30} = 38.51$

$F = \frac{MSTr}{MSE} = \frac{216.9}{38.51} = \boxed{5.632}$

$F_\alpha = \cancel{2.48872}\ 5.390$

$\cancel{F > \alpha}$ $F > F_\alpha$

$\therefore H_0$ is rejected.

Non-Parametric test

→ Non-Parametric test is a hypothesis test that does not require any specific conditions concerning the shape of population or the value of any population parameters.

→ Applied to categorical data.

→ less efficient than parametric tests.

→ Sign test is non-parametric test that is used to test whether or not two groups are equally sized.

Assumptions:-

1. Data distribution - The sign test is a non-parametric test, so we do not assume that the data is normally distributed.

2. Two sample - Data should be from two samples. The population may differ for two samples.

3. dependent Sample - should be paired sample or matched. Also known as 'before-after' sample.

Types of Sign test:-

1. One Sample: We set up the hypothesis so that + and - signs are the values of random variables having equal size.

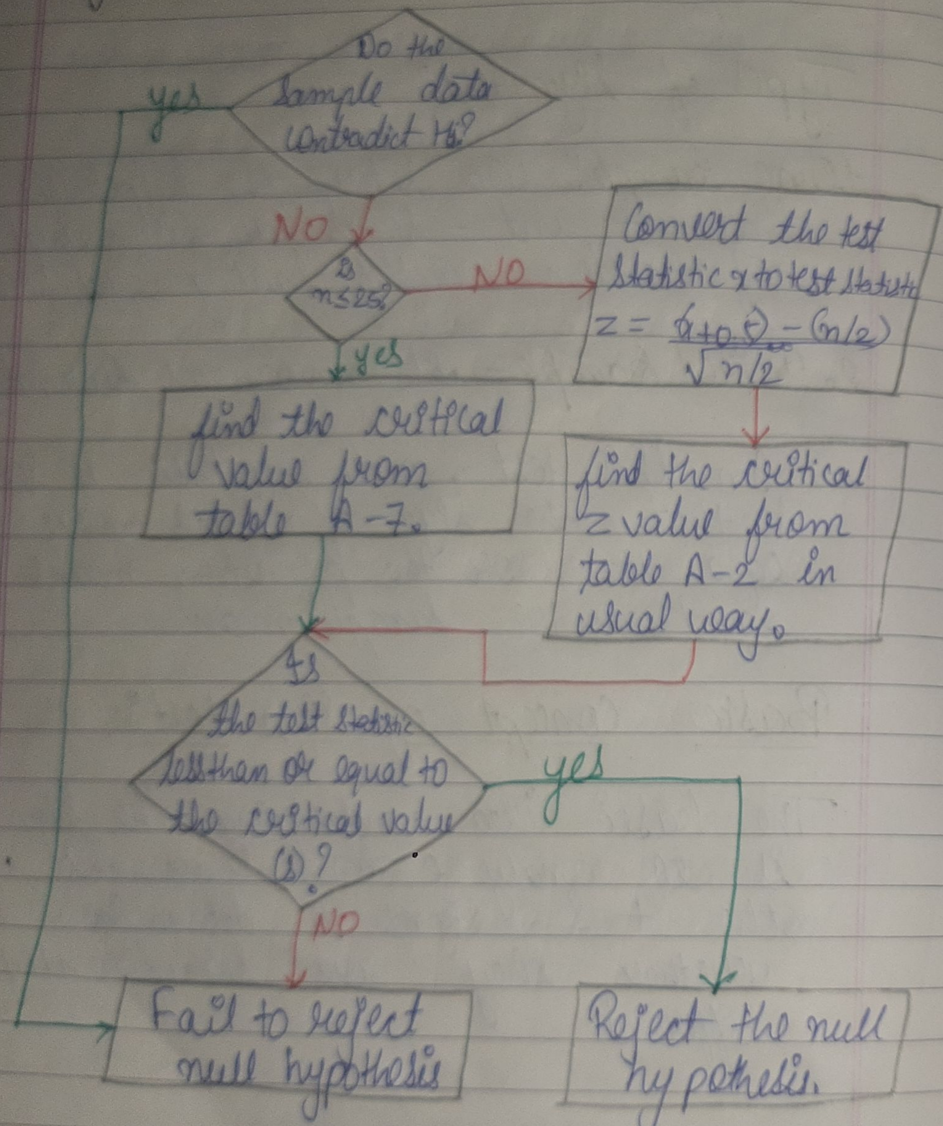
2. Paired Sample - Also known as Alternative

→ In this test, null hypothesis is set up so that the sign of + or - are of equal size, or the population means are equal to the sample mean.

Basic Concept of Sign test:-

The basic idea underlying the sign test is to analyze the frequencies of the + & - signs to determine whether they are significantly different.

Sign test procedure :-



Requirements :-

1. The sample data have been randomly selected

2. There is no requirement that the sample data come from a population with a particular distribution, such as normal distribution.

Notation for sign test :-

x = the no. of times the less frequent sign occurs

n = the total no. of +ve & -ve sign combined.

Test Statistic :-

for $n \le 25$: x (the no of times the less frequent sign occurs)

for $n > 25$: $z = \frac{(x + 0.5) - \frac{n}{2}}{\frac{\sqrt{n}}{2}}$

Critical values:

for $n \le 25$, critical values are in table A-7

$n > 25$, table A-2.

Key Concept:

If two sets of data have equal medians, no of positive signs should be approximately equal to no of -ve signs.

Examples:

yield of corn from diff seeds.

$\alpha = 0.05$

there is no diff. betn the yields from the regular & kiln-dried seed.

∴ two tailed

Regular	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Kilndried	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Sign	-	+	-	+	-	+	-	-	+	-	-

4 +ve   7 -ve.

$n \le \cancel{25}$ $n = 11$

$x \neq 4$ $\alpha = 0.05$

$x = 4$

from the table, critical value = 1

fail to reject H_0.

Que Out of 325 types of corps, 295 are one type. Use the sign test at 0.05 significance level to test the claim that method used to identify the corps ~~are~~ has no effect on the selection of type of corp.

Null hypothesis:

Alternative hypothesis:

$n = 325$

$\alpha = 0.05$ two tailed ∴ $\alpha = 0.25$

$x = 325 - 295 = \boxed{30}$

$n > 25$

$$z = \frac{(30 + 0.5) - \frac{325}{2}}{\frac{\sqrt{325}}{2}}$$

$$= \frac{-132}{9.013} = \boxed{-14.64}$$

critical value = -1.96.

Critical values:

for $n \le 25$, critical values are in table A7

$n > 25$, table A-2.

Key Concept:

If two sets of data have equal medians, no of positive signs should be approximately equal to no of -ve signs.

Examples:

yield of corn from diff seeds.

$\alpha = 0.05$

there is no diff. betn the yields from the regular & kiln-dried seed.

∴ two tailed

Regular	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Kilndried	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Sign	-	+	-	+	-	+	-	-	+	-	-

4 +ve 7 -ve.

$n \le 25$

$x = 4$

$n = 11$

$\alpha = 0.05$

$x = 4$

from the table, critical value = 1

fail to reject Ho.

Que Out of 325 types of corps, 295 are one type. Use the sign test at 0.05 significance level to test the claim that method used to identify the corps ~~are~~ has no effect on the selection of type of corp.

Null hypothesis:

Alternative hypothesis:

$n = 325$

$\alpha = 0.05$ two tailed ∴ $\alpha = 0.025$

$x = 325 - 295 = \boxed{30}$

$n > 25$

$$z = \frac{(30 + 0.5) - \frac{325}{2}}{\frac{\sqrt{325}}{2}}$$

$$= \frac{-132}{9.013} = \boxed{-14.64}$$

critical value = -1.96.

Paired t- test :-

A paired t test is used to compare two population means where you have two sample in which observation is one sample can be paired with other sample.

→ eg:- before - after observations on same subjects.

Procedure :-

let x = test score before the module

y = test score after the module

To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

1. Calculate the difference ($d_i = y_i - x_i$) betn two observation on each pair, making sure you distinguish betn +ve & -ve differences.

2. Calculate the mean difference, $\bar{d}$

3. Calculate the SD of the differences, S_d, as use to calculate the standard error of mean difference,

$$SE(\bar{d}) = \frac{S_d}{\sqrt{n}}$$

4. Calculate the t-statistics, which is given by $T = \frac{\bar{d}}{SE(\bar{d})}$. under the null hypothesis, this statistic follows a t-distribution with $n-1$ degree of freedom.

5. Use tables of the t-distribution to compare ~~wi th~~ your value for T to the t_{n-1} distribution this will give p-value for the paired test.

NOTE: for this test to be valid the differences only need to be approximately normally distributed.

for any extreme outliers, it is not advised to use paired t-test.

→ P value is known as level of marginal significance within the hypothesis testing that represents the probability of occurance of given event.

→ If P value is small than reject null hypothesis.

P value	decision
$P > 0.05$	The result is not statistically significant & hence don't reject null hypothesis
$P < 0.05$	reject H_0
$P < 0.01$	reject H_0

Paired Vs Unpaired T-test.

Similarity: both assume data from normal distribution

Characteristics of Unpaired T-test:-

- Two groups taken should be independent
- sample size may differ
- compares mean of data
- 95% confidence interval for mean difference is calculated.

Characteristics of Paired T-test:-

- Data is taken from subjects who have been measured twice.
- 95% confidence interval is obtained from the difference betn the two sets of joined of observations.

Que

Student	Premodule Score	Post module score
1	18	22
2	21	25
3	16	17
4	22	24
5	19	16
6	24	29
7	14	20

S	~~Post~~ Pre-M	~~Pre~~ Post-M	S	Pre	Post
8	21	23	15	18	18
9	23	19	16	20	24
10	18	20	17	12	18
11	14	15	18	22	25
12	16	15	19	15	19
13	16	18	20.	17	16.
14	19	26			

	d	E.
1	4	-1.95
2	4	-1.95
3	1	1.05
4	2	0.5
5	-3	5.05
6	5	-2.95
7	3	-0.95
8	2	0.5
9	-4	6.05
10	2	0.5
11	1	1.05
12	-1	3.05
13	2	0.5
14	7	-4.95
15	0	2.05
16	4	-1.95
17	6	-3.95
18	3	-0.95
19	4	-1.95
20	-1	3.05

$\Sigma d = 41$

$\bar{d} = \frac{41}{20} = 2.05$

$Sd = \sqrt{\frac{146.94}{19}} = 2.78.$

$SE = \frac{Sd}{\sqrt{n}} = \frac{2.78}{\sqrt{20}} = 0.62$

Page No.
Date :

One Way Anova

Que One factors with at least two levels, levels are independent

Researchers want to test a ~~mini~~ new anxiety medication. They split participants into 3 conditions (0mg, 5mg and 100mg), then ask them to rate their anxiety level on a scale of 1-10. Are there any differences betn the 3 conditions using $\alpha = 0.05$?

0mg	50mg	100mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	2

$\sigma_1^2 = 0.4761$
$\sigma_2^2 = 0.3313$
$\sigma_3^2 = 0.666$

$N = n_1 + n_2 + n_3 = 7(3) = 21$

$T = n_1\overline{x_1} + n_2\overline{x_2} + n_3\overline{x_3}$

$= 7(8.142 + 6.714 + 3)$

$= 127.992$

$\overline{\overline{x}} = \frac{T}{N} = \boxed{5.952}$

Page No.
Date :

$\overline{x_1} = 8.142$
$\overline{x_2} = 6.714$
$\overline{x_3} = 3$

$SSTr = n_1(\overline{x_1} - \overline{\overline{x}})^2 + n_2(\overline{x_2} - \overline{\overline{x}})^2 + n_3(\overline{x_3} - \overline{\overline{x}})^2$

$= 7[4.7961 + 0.5806 + 8.7143]$

$\boxed{SSTr = 98.637}$

~~SSE~~ $MSTr = \frac{SSTr}{K-1} = \frac{98.637}{3-1} = \boxed{49.31}$

$SSE = (n_1-1)s_1^2 + (n_2-2)s_2^2 + (n_3-1)s_3^2$
$= 6[1.474]$
$= \cancel{8.84}\ 10.2859$

~~F =~~ $MSE = \frac{SSE}{N-K} = \cancel{0.491}\ 0.5711$

$F = \frac{MSTr}{MSE} = 86.22$

$df_1 = 2 \quad df_2 = 18$
from table, $\alpha = 0.05$
$F_\alpha = 3.55$

$F_\alpha < F$
$\therefore$ Rejected H_0

$P = 0.00$
Pvalue < 0.01
Reject H_0

One - way Anova

1. Define Null & Alternative hypothesis
2. State alpha
3. Calculate degree of freedom
4. State decision rule
5. Calculate test statistic
6. State result
7. State conclusions

Mini- tab

Stats → anova → One way anova
↓
response data are in seperate coloum for each factor
↓
factor (Select 0mg, 50mg, 100mg) → confidence level 0.95 ($\alpha = 0.05$)
↓
graph (→ interval plot) → results (select every checkbox)
↓
OK

H_0 : all the means are equal
H_a : atleast one mean is different