LECTURE - 7/10/22

PROBABILITY

1. Sample space $S = \{1, 2, 3, 4, 5, 6\}$

Event E: an even number is rolled

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

$P(E) = \frac{3}{6} = \frac{1}{2} = 0.5$.

Event T: number greater than two is rolled

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

$P(T) = 4/6 = 2/3 = 0.67$.

2. $S = \{HH, HT, TH, TT\}$.

$E = \{HH, TT\}$

$P(E) = 2/4 = 1/2 = 0.5$.

$\hookrightarrow P(E) = 1/4 + 1/4 = 1/2$.

3. $P(E^c) = 1 - P(E) = 1 - 0.5 = 0.5$

$P(T^c) = 1 - P(T) = 1 - 2/3 = 1/3 = 0.33$.

$E = \{2, 4, 6\}$ $\therefore E^c = \{1, 3, 5\}$

$T = \{3, 4, 5, 6\}$ $T^c = \{1, 2\}$

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

4. Atleast one head will appear in five tosses of a fair coin

Probability of Atleast one head = 1 - Probability of no head

$= 1 - \frac{1}{2^5}$

$= \frac{31}{32}$.

$= 0.9688$

$\frac{1}{12} + \frac{3}{12} + \frac{8}{12}$

Date
Page

5. Intersection: $A \cap B$.

$E = \{2, 4, 6\}$
$T = \{3, 4, 5, 6\}$
$E \cap T = \{4, 6\}$
$P(E \cap T) = 2/6 = 1/3$.

Q.

(a). Given $E: \{2, 4, 6\}$
greater than two $T: \{3, 4, 5, 6\}$
$P(E \cap T) = 2/6 = 1/3$.

(b) $1 = \frac{1}{12} + \frac{3}{12} + 4x$

$\boxed{x = P(5) = P(4) = P(6) = P(1)}$

$\Rightarrow 4x = \frac{12-4}{12} = \frac{8}{12}$

$\Rightarrow x = \frac{2}{12} = 1/6$.

$P(4) + P(6) = \frac{1}{6} + \frac{3}{12}$

$= \frac{5}{12}$

Q. $E = \{2, 4, 6\}$
any three choice, mutually exclusive
$= \{1, 3\}, \{1, 5\}, \{3, 5\}, \{1\}, \{3\}, \{5\}, \{1, 3, 5\}$.

PROBABILITY

→ Random Experiment: Definite outcome that cannot be predicted with certainity.
→ Sample space: set of all possible outcomes.
→ Event: Subset of sample space.

$P = 0 \rightarrow$ impossible event
$P = 1 \rightarrow$ Certain / Sure event.

LECTURE: 13/10/22

Union of Events :- $A \cup B$, word "or".

Q. $E = \{2, 4, 6\}$ — E: number rolled is even
$T = \{3, 4, 5, 6\}$ — T: number rolled is > than 2.
$E \cup T = \{2, 3, 4, 5, 6\}$

→ Additive rule of probability :-

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

Q. (a) $E = \{4, 4\}$.
$P(E) = \frac{1}{36}$.

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

(b) No die show 4, $5 \times 5 = 25$.
$\therefore P(E) = 1 - \frac{25}{36} = \frac{11}{36}$

$$\left.\begin{array}{l} (1, \ ) \rightarrow 5 \\ (2, \ ) \rightarrow 5 \\ (3, \ ) \rightarrow 5 \\ (5, \ ) \rightarrow 5 \\ (6, \ ) \rightarrow 5 \end{array}\right\} 25.$$

⇒ At least one die show 4 :- $P(A) = 1/6$
A: Die 1 shows 4 — $P(B) = 1/6$.
B: Die 2 shows 4. — $P(A \cap B) = \frac{1}{36}$.

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

$\Rightarrow P(A \cup B) = \frac{12-1}{36} = \frac{11}{36}$. (Ans).

→ ~~Condition~~ Conditional Probability and Independent Events:

$P(A/B)$ = Probability of occurence of A after B has occured definitely.

Rule for conditional probability:-

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

Q (a) A: number rolled is five
B: number is odd
$(A \cap B) = \{5\}$.
$B = \{1, 3, 5\}$.

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

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

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = 1/3.$$

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

(c) A: number is odd
B: number is 5
$(A \cap B) = \{5\}$
$B = \{5\}$

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

$P(A \cap B) = 1/6$.

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

$P(A/B) = P(A)$, $\Rightarrow$ occurence of B has no effect on likelihood of A.

Then $P(A \cap B) = P(A) \cdot P(B)$
$\Rightarrow$ the events are independent.

Hence, in a situation if we can compute all three probabilities $P(A)$, $P(B)$, $P(A \cap B)$

If $P(A \cap B) = P(A) \cdot P(B)$, then A and B are independent.

$P(A \cap B) \neq P(A) \cdot P(B)$, then A and B are dependent.

Q. $A = \{3\}$ $\quad P(A) = 1/6$ $\quad P(A \cap B) = 1/6$,
$B = \{1, 3, 5\}$ $\quad P(B) = 1/2$

$P(A \cap B) \neq P(A) \cdot P(B)$

$\therefore$ A & B are dependent.

Q. (a) both will be positive $\quad P(A) = P(B) = 0.92$

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

$= (0.92)^2 = 0.8464$

independent

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

$= 0.92 + 0.92 - 0.8464$

$= 0.9936$.

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

Mutually Exclusive Event, $P(A \cap B) = 0$.

→ A and B have no outcomes in common

→ For example $A - A^c$, $B - B^c$ are mutually exclusive

Example 18:-

$P(A) = 0.63$ $\quad P(A \cap B) = 0.27$

$P(B) = 0.34$ $\quad \therefore P(A \cup B) = 0.63 + 0.34 - 0.27$

$\Rightarrow \boxed{P(A \cup B) = 0.7}$

Example 19:-

Total no. of volunteers = 28.

(a) Speciality is medicine and speaks two or more languages

$= 2/28 = 0.0714$

(b) Speciality is medicine : A

Two or more languages : B

$P(A) = \frac{8}{28}$ $\quad \therefore P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\Rightarrow P(A \cup B) = \frac{8}{28} + \frac{6}{28} - \frac{2}{28}$

$P(B) = \frac{6}{28}$ $\quad \Rightarrow P(A \cup B) = \frac{12}{28} = \frac{3}{7} = 0.43$

(c) Speciality is something other than medicine.

$= 1 - P(\text{speciality is medicine})$

$= 1 - \frac{8}{28} = \frac{20}{28} = 0.714$

Example 21:-

(a) Teenager at first marriage.

$P(E) = \frac{125}{902} = 0.139$

(b) Teenager at first marriage, person is male.

$\cancel{P(E)} = P(E|M) = \frac{P(E \cap M)}{P(M)}$

$= \frac{\frac{43}{902}}{\frac{450}{902}}$

$\Rightarrow P(E|M) = \frac{43}{450} = 0.095$

Date
Page

Example 22:-

(a) $P(H|O) = \frac{P(H \cap O)}{P(O)} = \frac{0.09}{0.11} = \frac{9}{11} = 0.818$

(b) $P(H|O^c) = \frac{P(H \cap O^c)}{P(O^c)} = \frac{0.11}{0.89} = 0.124$

(c) Yes, overweight people tend to suffer from hypertension.

Example 24:-

$P(F) = \frac{452}{902}$ $\quad P(F \cap E) = \frac{82}{902}$

$P(E) = \frac{125}{902}$

$\therefore P(F \cap E) = 0.09$

$P(F) \cdot P(E) = 0.069$

$P(F \cap E) \neq P(F) \cdot P(E)$

$\Rightarrow$ the events are not independent.

Example 26:-

Specificity $\rightarrow$ test will be negative.

High specificity, low false +ve rate.

(a) $P(\text{test result +ve}) = 1 - 0.89 = 0.11.$

(b) Both result +ve $= 0.11 \times 0.11$ ~~= 0.121~~
$= 0.0121$

0.89 $\rightarrow$ test result is -ve.

Example 27:- $P(\text{Contraband detected}) = \cancel{0.9 \times 0.9 \times 0.2}$
$\cancel{= 0.729}$

$= 1 - P(\text{contraband not detected})$

$= 1 - P(D_1^c \cap D_2^c \cap D_3^c)$

$= 1 - P(D_1^c) \cdot P(D_2^c) \cdot P(D_3^c)$

$= 1 - (1-0.9)(1-0.9)(1-0.9)$

$\boxed{= 0.999}$ Ans.

Date
Page

or P(contraband detected)

= P(contraband detected by all three dogs)

+ P(contraband detected by one dog, undetected by two dogs)

+ P(contraband detected by two dogs, undetected by one dog)

$= P(D_1 \cap D_2 \cap D_3) + P(D_1 \cap D_2^c \cap D_3^c) \times 3 + P(D_1 \cap D_2 \cap D_3^c) \times 3$

$= (0.9)(0.9)(0.9) + (0.9)(0.1)(0.1) \times 3 + (0.9)(0.9)(0.1) \times 3$

$\boxed{= 0.999}$ (ans)

LECTURE: 14/10/22

Random variable :-

- Random variable denoted by 'X' or 'Z'.
- Actual values taken by random variable denoted by 'x' or 'z'.

	Experiment	Number X	Possible values
1.	Sum of dots when two die is rolled	Sum of number of dots on the top faces.	2, 3, 4, . . . , 11, 12

→ Probability within 0 to 1, $0 \le P(x) \le 1$.

→ $\sum P(x) = 1$.

Example: Fair coin is tossed twice.

X, is Number of heads observed.

Atleast one head : $P(X \geq 1) = P(1) + P(2)$

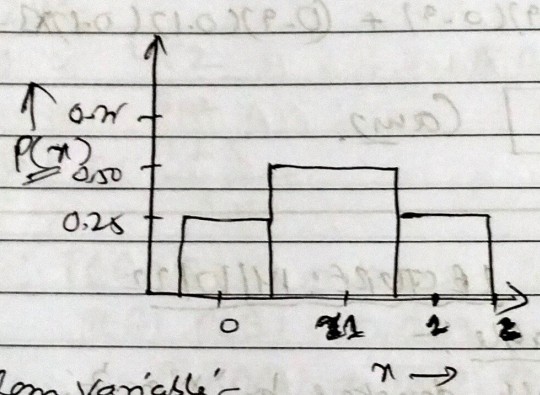
X	0	1	2
P(x)	1/4	2/4	1/4

(b) Probability atleast one head is observed $= 1 - 1/4$

$= 3/4$.

$P(X \geq 1) = P(1) + P(2)$

$= 2/4 + 1/4 = 3/4$.



Discrete Random variable

⇒ Mean of a discrete random variable $= \mu = \sum x P(x)$

Q. Mean $= -2(0.21) + 1(0.34) + 2(0.24) + (3.5)(0.2)$

$= 1.135$

→ Variance of a discrete random variable

$$\sigma^2 = \sum (x-\mu)^2 P(x)$$

$$\Rightarrow \sigma^2 = \sum x^2 P(x) - \mu^2.$$

→ Hence Standard Deviation,

$$\sigma = \sqrt{\sum x^2 P(x) - \mu^2}$$

Q.

x	-1	0	1	4
$P(x)$	0.2	0.5	a	0.1

$\because \sum P(x) = 1$

$\Rightarrow a = 0.2$

(a) $a = 0.2$

(b) $P(0) = 0.5$

(c) $P(X>0) = P(1) + P(4) = 0.2 + 0.1 = 0.3$

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

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

(f) $\mu = \sum x P(x) = 0.4$

(g) variance, $\sigma^2 = 0.2(1) + 0.2(1) + 0.1(16) - (0.4)^2$

$= 1.84$

(h) Standard Deviation $\sigma = 1.356$

Discrete Random Variable

Random variable: Numerical quantity generated by random experiment.

Random variable denoted by capital X and Z, actual values they take denoted by 'x' or 'z'.

→ A random variable is called discrete if it has either a finite or countable number of possible values.

→ A random variable is continuous if its possible value contain whole interval of numbers.

→ Example 4, examples.

LECTURE: 17/10/22

: Binomial Distribution :

→ n identical and independent trials.

Two outcomes, success and failure

Probability of success = p.

Probability of failure = q.

We say that a binomial distribution Y, with parameters 'n' & 'p'.

1. n = 125

Proportion of student female = 0.57

X: number of female student in sample.

success - student is female.

∴ p = 0.57.

∴ X is a binomial distribution, with n = 125, p = 0.57

⇒ If 'X' is a binomial random variable with parameters 'n' and 'q' then

x: actual value taken by the random variable X.

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$q = 1-p.$$

$$\therefore \boxed{P(x) = {}^nC_x \, p^x q^{n-x}} \quad (q = 1-p)$$

Example 7:

Q. It is a Binomial Distribution, since there is a success and failure

p = 0.17

q = 0.83.

(a). n = 5, p = 0.17.

$$P(x) = \frac{5!}{x!(5-x)!} (0.17)^x (0.83)^{5-x}$$

$\therefore P(0) = 0.3939$

$P(1) = 0.4033$

$P(2) = 0.1652$

$P(3) = 0.0338$

$P(4) = \cancel{0.1024}\ 0.0035$

$P(5) = 0.0001$

x	0	1	2	3	4	5
$P(x)$	0.3939	0.4034	0.1652	0.0338	0.0035	0.0001

(b) most frequent case, $X = 1$.
Most frequent number of cases, seen each day in which the victim knew the prepetrator is one.

(c) $\mu = \sum x P(x)$

$= 0 + 0.4033 + 2 \times 0.1652 + 3 \times 0.0338 + 4 \times 0.0035 + 5 \times 0.0001$

$= 0.8496$

$\rightarrow$ Mean $= np = 0.17 \times 5 = 0.85$

S.D. $= \sqrt{5 \times 0.17 \times 0.83} = 0.8399$.

$\Rightarrow$ If X is a ~~random~~ binomial random variable, with parameter 'n' and 'p' then,

$\mu = np$, $\sigma^2 = npq$, $\sigma = \sqrt{npq}$

$q = 1 - p$.

$\Rightarrow$ Cumulative Probability Distribution

$P(X \leq x) = P(0) + P(1) + P(2) + \dots + P(x)$

Example 9:-

Q. Question might $\rightarrow$ when it is 'true'.

p: Probability that student guessed 'true'

$\therefore p = 1/2$

$q = 1/2$

$n = 10,\ p = 1/2,\ q = 1/2.$

(a) $\therefore P(6) = \cancel{{}^{10}C_6 (p)^{1/2} (q)^{1/2}} \cdot {}^{10}C_6 \left(\frac{1}{2}\right)^{10-6} \left(\frac{1}{2}\right)^{6}$

$\Rightarrow P(6) = {}^{10}C_6 \left(\frac{1}{2}\right)^{10}$

$\therefore \boxed{P(6) = 0.2051}$

(b) $P(X \geq 6) = P(6) + P(7) + P(8) + P(9) + P(10)$

$= \left({}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}\right) \times \left(\frac{1}{2}\right)^{10}$

$= 0.3769$

Example 10:

Q. $p = 1/3,\ n = 5.$

(a) $P(X \leq 1) = P(0) + P(1)$

$= {}^5C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^5 + {}^5C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^4$

$= \boxed{0.4609}$

(b) $P(X \leq x) = P(0) + P(1) + P(2) + \dots + P(x)$

$P(X \leq x) \geq 0.95.$

For $x = 2,\ P(X \leq 2) = 0.7901$

$x = 3,\ P(X \leq 3) = 0.9547$

$\therefore \boxed{\text{minimum } 3}$

→ he has enough of the day's supplies i.e almost condition

Variance:- $\sigma^2 = x^2 P(x) - \mu^2$

Standard Deviation:- $\sigma = \sqrt{x^2 P(x) - \mu^2}$

In case of a binomial random variable i.e. discrete random variable,

mean $= \mu = np$

variance $= \sigma^2 = npq$

Standard Deviation $= \sigma = \sqrt{npq}$. $(q = 1-p)$.

<u>Atleast means</u> $P(X \geq x)$

<u>Atmost</u> $P(X \leq x)$

LECTURE :- 3/11/22. <u>ANOVA (Analysis of Variance)</u>

→ <u>Single factor ANOVA</u>: involves comparison of K population or treatment means μ_1, μ_2, μ_3, The objective is to test whether all the K population means are the same.

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

against, H_a : Atleast two of the μ's are different.

⇒ <u>ANOVA Notation</u> :-

K = number of population or treatments being compared

(i) Population or treatment	1	2	.....	K
(ii) Population or treatment mean	μ_1	μ_2	.....	μ_K
(iii) Population or treatment variance	σ_1^2	σ_2^2	.....	σ_K^2
(iv) Sample size	n_1	n_2	.....	n_K
(v) Sample mean	$\bar{x}_1$	$\bar{x}_2$	.....	$\bar{x}_K$
(vi) Sample variance	s_1^2	s_2^2	.....	s_K^2.

→ Combined sample size.

$N = n_1 + n_2 + \cdots + n_k$ (the total number of observation in the data set).

T = grand total = sum of all N observations
$= n_1 \bar{x}_1 + n_2 \bar{x}_2 + \cdots + n_k \bar{x}_k$.

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

⇒ Assumptions for ANOVA:

1. Each of the population or treatment response distribution is normal.
2. $\sigma_1 = \sigma_2 = \cdots = \sigma_k$ (The 'k' normal distribution are identical standard deviation)
3. The observations in the sample from any particular one of the k population or treatments are independent of one another.
4. When comparing population mean, 'k' random samples are selected independently of one another.

⇒ Test procedure of ANOVA:

(a) SSTr (Treatment sum of squares): measure of disparity among sample means.

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

(b) SSE (Error sum of squares): measure of variation within the 'k' samples.

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

$$\boxed{SSTo = SSTr + SSE}$$

one way ≡ taking one random ~~variable~~ sample from each population under consideration.

Each sum of squares has an associated df:

df_1 = Treatment df $= K-1$.
df_2 = Error df $= N-K$

} df: degree of freedom.

A mean square is sum of square divided by its df.

mean square for treatments $= MSTr = \frac{SSTr}{K-1}$

mean square for error $= MSE = \frac{SSE}{N-K}$

Number of error degree of freedom comes from adding the number of degree of freedom associated with each of the sample variance:

$$(n_1-1)+(n_2-1)+(n_3-1)+\dots+(n_K-1)$$
$$=(n_1+n_2+n_3+\dots+n_K)-K$$
$$=N-K.$$

⇒ Simple ANOVA Factor Table :-

Source of Variation	df	Sum of Squares	Means Square	F
Treatment	$K-1$	$SSTr$	$MSTr = \frac{SSTr}{K-1}$	$F = \frac{MSTr}{MSE}$
Error	$N-K$	SSE	$MSE = SSE/(N-K)$	
Total	$N-1$	$SSTo$		

Test statistic, for testing null hypothesis that K population means are equal.

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

The test is right tailed. H_0 is rejected at level of significance 'α' if $F \geq F_\alpha$.

LECTURE :- 4/11/22

$$F = \frac{MSTr}{MSE} \qquad df_1 = K-1$$
$$df_2 = N-K$$

F distribution always right tailed

$$\boxed{SSTo = SSTr + SSE}$$

Example's 5% level of significance.

Mathematics	English	Education	Biology
2.59	3.64	4.00	2.78
3.13	3.19	3.59	3.51
2.97	3.15	2.80	2.65
2.50	3.78	2.39	3.16
2.53	3.03	3.47	2.94
3.29	2.61	3.59	2.32
2.83	3.20	3.74	2.58
3.17	3.30	3.77	3.21
2.70	3.54	3.13	3.23
3.88	3.25	3.00	3.57
2.64	4.00	3.47	3.22
$\bar{x}_1 = 2.903$	$\bar{x}_2 = 3.335$	$\bar{x}_3 = 3.359$	$\bar{x}_4 = 3.015$

$n_1 = n_2 = n_3 = n_4 = 11$, $K = 4$

$n = 11 + 11 + 11 + 11 = 44$

$\therefore df_1 = K-1 = 4-1 = 3$

$df_2 = 44 - 4 = 40$

$df_1 = 3$
$df_2 = 40$

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

$$\therefore s_1^2 = \frac{1.8806}{10} = 0.1881$$

$$s_2^2 = 0.1475$$

$S_3^2 = 0.228$ $\quad$ $T = 136.732$; $N = 44$

$S_4^2 = 0.1564$ $\quad$ $\bar{\bar{x}} = 3.152$

$\therefore$ ~~$MSA_x = 10x$~~

$SSA_x = 11(0.157) = 1.728 \Rightarrow MSA_x = 0.576$

$SSE = 7.202 \Rightarrow MSE = 0.1801$

$$\therefore f = \frac{0.576}{0.1801} = 3.198.$$

$df_1 = 3$, $df_2 = 40$, $\alpha = 0.05$, $f_\alpha = \cancel{2.8303}$ 2.8387.

$\therefore f \geq f_\alpha \Rightarrow$ $\boxed{H_0, \text{ null hypothesis is rejected.}}$

<u>Example:</u>

$k = 3$ $\quad$ $\alpha = 0.01$

$N = 33$

Treatment 1	Treatment 2	Control group
71	77	81
72	67	79
75	79	73
80	75	71
60	81	84
65	72	77
63	71	67
78	84	75
75	91	
73		
72		
65		
61		

69		
64		
71		
$n_1 = 16$	$n_2 = 9$	$n_3 = 8$

$df_2 = N - K = 30$

$df_1 = K - 1 = 3 - 1 = 2$

$\bar{x}_1 = 69.75$

$\bar{x}_2 = 77.78$

$\bar{x}_3 = 75.88$

$T = 2423.06$

$N = 33$

$\therefore \bar{x} = 73.4261$

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

$\therefore S_1^2 = 517/15 = 34.46$

$S_2^2 = 421.56/8 = 52.69$

$S_3^2 = \frac{214.872}{7} = 30.696$

$SSTr = 16(69.75 - 73.4261)^2 + 9(77.78 - 73.4261)^2 + 8(75.88 - 73.4261)^2$

$\Rightarrow SSTr = 435$

$\therefore MSTr = \frac{435}{2} = 217.5$

$SSE = 15 \times 34.46 + 8 \times 52.69 + 7 \times 30.696$

$\Rightarrow SSE = 1153.43$

$MSE = \frac{1153.43}{30} = 38.448$

$$\therefore F = \frac{217.5}{38.448} = 5.656$$

$$\therefore F = 5.656$$
$$F_\alpha = 5.350 \qquad \alpha = 0.01,\ df_1 = 2,\ df_2 = 20$$

$\therefore F > F_\alpha \Rightarrow$ Null hypothesis, H_0 is rejected.

LECTURE 1 - 11/11/22

Non parametric test :- ⇒ **Sign Test** is a non parametric test that is used to test whether or not two groups ~~are having same size~~ are equally sized. Also called **binomial sign test**.

Assumptions:

(1) **Data distribution** :- sign test is non parametric test, don't assume data is normally distributed

(2) **Two sample** :- Data should be from two samples.

(3) ~~Dep~~ ~~Dep~~ **Dependent sample** :- ~~Dep~~ Dependent sample should be a paired sample or matched. It is '**before - after**' sample.

→ One of easiest non parametric test is sign test.

⇒ **Types of sign test** :-

→ **One sample** : '+' and '-' signs are values of random variable having equal size

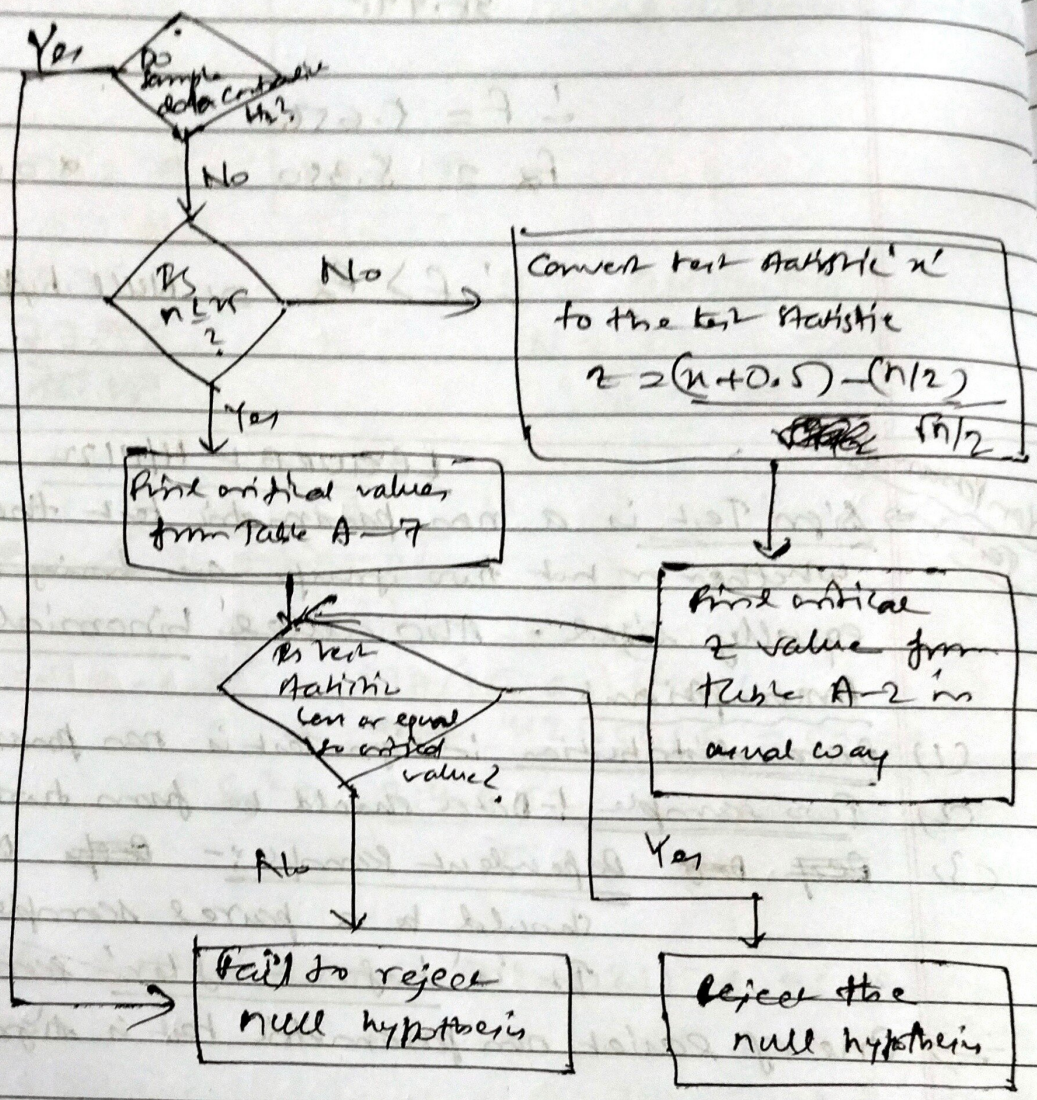
→ **Paired sample** : Alternative to paired 't'-test.

⇒ **Basic concept of the sign test** :-

The basic idea is to analyze the frequencies of plus and minus sign to determine whether they are significantly different.

$\boxed{\text{Sign test procedure}}$

Sign Test Procedure:



⇒ **Requirements:**

(i) Sample data have been randomly selected.

(ii) No requirement that the sample data come from a population with a particular distribution, such as normal distribution.

⇒ **Notation of sign test:**

- x = number of times the less frequent sign occurs.
- n = total number of positive and negative signs combined.

→ Test statistic:

$n \leq 25$ x: No. of times the less frequent sign occurs.

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

→ Critical value:

$n \leq 25$ critical x value in table A-7

$n > 25$ critical z value in table A-2.

⇒ Key concept underlying the use of sign test:

→ Two sets of data have equal medians, number of positive sign approximately equal to number of negative sign.

Examples:

1. $n = 11 \leq 25$, significance level 0.05, Two tailed test. $H_0: \mu = \mu_0$

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

7 -ve
4 +ve

⇒ 7 -ve
4 +ve

Null Hypothesis:- There is difference between yeild from regular and kiln dried seed.

Date
Page

A:- ∵ Since the sample size is (25),
x = no. of times less frequent sign
$\Rightarrow x = 4$. [∵ +'ve nos are 4 only]

$\therefore \boxed{x = 4}$

Critical value of x, when level of significance = 0.05, is
$x_{0.025}$ ~~$x_{0.05}$~~ = 1 $(\alpha/2 = 0.025)$

$\left\{ \begin{array}{l} n = 11, \text{ two tailed} \\ \alpha/2 = 0.05/2 = 0.025 \end{array} \right.$ [∵ $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$]

~~∵ $x < x_{0.05}$, Null hypothesis is rejected.~~

∵ $x > x_{0.05}$ ⇒ fail to reject null hypothesis.

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

A:- $n = 325$
$x = 325 - 295 = 30$.
H_0: Method to identify the corps has effect on selection of type of corps.
H_a: Method doesn't have effect on selection

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

$$\Rightarrow z = \frac{-132}{9.0124} = -14.64$$

$$\boxed{z = -14.64}$$

$\alpha/2 = 0.025$

$z_{\alpha/2} = z_{0.025} = -1.96$

$\therefore$ ~~$z > z_\alpha$~~ $z < z_\alpha$.

$\Rightarrow$ Null hypothesis rejected.

$\Rightarrow$ Doesn't have effect.

<u>Non parametric Test</u>:- A non parametric test is a hypothesis test that doesn't require any specific conditions concerning the shape of population or value of any population parameters.

- Non parametric test don't require normally distributed populations.
- Applied to categorical data.
- Less efficient than parametric test. Stronger evidence required to reject null hypothesis.

<u>Ex</u>:- Sign Test | Binomial sign test.

$\Rightarrow$ Sign test is a weak test.

Paired T-Test :-

LECTURE : 14/11/22

=> **Paired T-test** : Compare two population means where we have two samples, in which observation in one sample can be paired with observation in the other sample.

(i) x = before the module

y = after the module

Difference, $d_i = y_i - x_i$

d = before + after = post − pre

(ii) mean difference, $\bar{d}$.

(iii) S.D. of the difference = s_d

Standard error of the mean difference, $SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$

(iv) Calculate the value of T using,

$$T = \frac{\bar{d}}{SE(\bar{d})}$$

$(n-1)$ degree of freedom.

(v) Compare T and t_{n-1}, this will give p value for paired t-test.

=> Approximately normally distributed, for test to be valid.

p-value : probability value.

Probability of occurrence of the event.

p-value small → Null hypothesis is rejected

Alternative hypothesis accepted.

→ Similarity between paired and unpaired t-test :-

Data in normal distribution.

Unpaired T-test	Paired T-test
1. Two groups independent	1. Dependent
2. Sample size not be equal	2. Data taken from subjects that have been measured twice.
3. Compares mean data of the two groups	

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

Student	Pre module (x_i)	Post module (y_i)	$d_i (= y_i - x_i)$
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1
			$\Sigma d = +41$

$$\therefore \bar{d} = \frac{+41}{20} = +2.05 \qquad \Sigma d^2 = 237$$

$$S_d = \sqrt{\frac{237 - 20(-2.05)^2}{19}}$$

$$= \sqrt{8.05} = 2.837$$

$$\therefore T = \frac{+2.05}{\frac{2.837}{\sqrt{20}}} = +3.232$$

~~$\therefore T = +3.232$~~

~~$\therefore \frac{3.23 - 3.20}{3.30 - 3.20} = p = 0.002$~~

~~$\therefore T = +3.32$~~

$\therefore T = +3.232$

Degree of freedom $= n-1 = 20-1 = 19$

for d.f. $= 19$

t_{n-1}	p
3.20	0.002
3.30	0.002

$\therefore$ when $T = +3.232$, $p = 0.002$

p value Table :-

p- value	Decision
(i) p-value > 0.05	result not ~~significantly~~ statistically significant, ~~don't~~ don't reject null hypothesis.
(ii) pvalue < 0.05	result statistically significant, reject null hypothesis
(iii) p value ~~< 0.05~~ < 0.01	highly statistically significant, reject null hypothesis

one way ANOVA

LECTURE 1 - 17/11/22

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

$K = 3$

$N = 7 \times 3 = 21$

$\alpha = 0.05$

$df_e = N - K = 18$

$df_t = K - 1 = 2$.

→ $H_0: \mu_0 = \mu_{50} = \mu_{100}$ → H_a: all are not equal.

$\bar{x}_1 = 57/7 = 8.14$

$\bar{x}_2 = 47/7 = 6.71$

$\bar{x}_3 = 21/7 = 3$

$T = 125$ ∴ $\bar{\bar{x}} = \frac{125}{21} = 5.952$

$N = 21$

∴ SST = 98.8335

$s_1^2 = \frac{\Sigma(x_i - \bar{x}_1)^2}{n-1} = 0.4762$

$s_2^2 = 0.8715$

$s_3^2 = 4/6 = 0.666$

SSB = 10.2818

→ MST = ~~22.8444~~ 49.2667

MSB = 0.5714

$\boxed{F = 86.2241}$

$F_\alpha = 3.8546$ ($\alpha = 0.05$)

$\therefore$ $F > F_\alpha$ → Null hypothesis is rejected.

Minitab

⇒ One way ANOVA:

→ Response data in different columns ⇒ 0mg, 50mg, 100mg

→ Confidence interval 95%

→ Graphs: Interval plots.
Individual value plot.
Box plot
Histogram plot.

P value = 0 ⇒ Null hypothesis rejected

⇒ Wilcoxon Signed Rank Test:

→ Popular non parametric test, for matched or pair data i.e. dependent data.

→ Like the sign test it is based on different scores, but in addition to analyzing the signs of differences, it also takes into account the magnitude of the observed difference.

$$Z = \frac{|R - \text{mean}|}{\text{Standard Deviation}}$$

R = Sum of ranks (smaller, larger, or any one).

n = Number of samples (After excluding 0 difference)

$$\text{Mean} = \frac{n(n+1)}{4}$$

$$S.D. = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Q. A drug is given to 11 patients and difference in their BP were recorded to be :-

Before Drug A	After drug A
112	116
113	120
~~118~~ 118	117
~~115~~ 120	125
~~116~~ 119	126
~~117~~ 113	111
~~118~~ 110	111
~~119~~ 122	117
126	126
115	112
119	129

Use the Wilcoxon Signed rank test to test the hypothesis that ~~has~~ the drug has no effect on change on BP?

Ans H_0: Drug A doesn't effect BP.

H_a: Drug A affects BP.

Before	After	d	|d|	R
112	116	-4	4	5
113	120	-7	7	8.5
118	117	1	1	1.5
120	125	-5	5	6.5
119	126	-7	7	8.5
113	111	2	2	3
110	111	-1	1	1.5
122	117	+5	5	6.5
126	126	0		

115	112	3	3	4
119	129	-10	10	10

$R^+ = 1.5 + 3 + 4 + 6.5 = 15$

R^- = rank of -ve sign number

$= 8 + 8.5 + 6.5 + 8.5 + 1.5 + 10$

$= 40.$

$n = 10$ → n is value of top rank.

$$\text{mean} = \frac{10 \times (10+1)}{4} = \frac{10 \times 11}{4} = 27.5$$

$$S.D. = \sqrt{\frac{10 \times 11 \times 21}{24}} = 9.811$$

$$Z = \frac{|40 - 27.5|}{9.811} = \frac{12.5}{9.811} = 1.274$$

$\therefore \boxed{Z_{cal} = 1.274}$

$\alpha = 0.05$, $Z_{crit} = 1.96.$

$Z_{0.05} = 1.96.$

~~$Z_{tab} > Z_{cal}$~~

LECTURE: 18/11/22

Q.

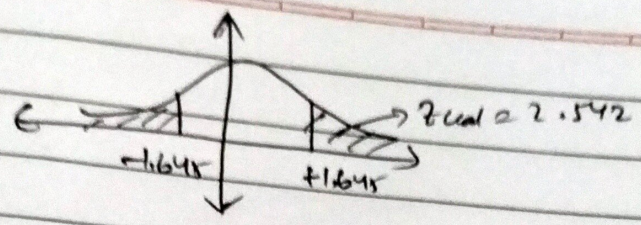
a: $H_0: \mu = 48,750$ $\alpha = 0.1$

$H_a: \mu \neq 48,750$ $\frac{\alpha}{2} = 0.05$

(i) $$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{57,505 - 48,750}{\frac{6892}{\sqrt{40}}}$$

$$\Rightarrow Z_{cal} = \frac{2755 \times \sqrt{40}}{6892} = 2.542$$

$Z_{0.05} = -1.645.$



$\Rightarrow$ Null hypothesis is rejected.

$\Rightarrow \mu \neq 48750$.

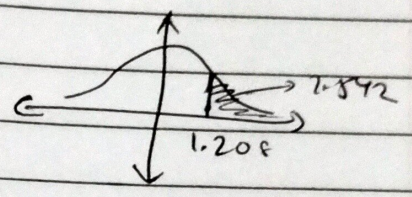
(ii) $H_0: \mu = 48,750$

$H_1: \mu > 48,750$.

$\alpha = 0.1$

$z_{cal} = 2.542$

$z_{0.1} = -1.2088$



$\Rightarrow$ Null hypothesis is rejected

$\Rightarrow \mu > 48750$.