2/1/25
Thursday

HYPOTHESIS

M T W T F S S
Page No.:
Date:
YOUVA

Null hypothesis (H_0)
↳ hypothesis of no difference

when we rejecting the hypothesis when it is true
↳ Type I error (R T I) (Reject the true type-I error)
accepting the hypothesis when it is false
↳ Type II error

level of significance
– Maximum probability of type-I error.
ex 5%, 1%, 0.1%

degree of freedom
N – K → no. of constraint
no. of values which can be assigned at your own will.

Z-Test
Z-Test ⇒ To compare the means when sample size is large or population SD is known.
σ → ~~large~~ known
N → large – more than 30.

$$\boxed{Z = \frac{\bar{X} - \mu}{S.E.}}$$

$$\boxed{\text{Standard error (SE)} = \frac{\sigma}{\sqrt{N}}}$$

$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}}$$

ex $N = 50$ $\bar{X} = 32$ $\mu = 30$ $\sigma = 4$

$$\frac{32-30}{\frac{4}{\sqrt{50}}} \quad \frac{2}{4/\sqrt{50}} = \frac{2}{4/7} = \frac{2 \times 7}{\cancel{4}\,2}$$
$$= 4$$

~~t-test~~

7/1/25
Tuesday

– the test in which there is no assumption abt the pop^n parameters from which samples are drawn are known as non-parametric test. these are very much useful in behavioural science.

(χ^2)

the kia square test is one of the most commonly used non-parametric ~~use~~ test.

- It is used in *Eneumeration Statictics.
- It was first used by Karl Pearson in 1900.
- the quantity of kia square ~~destroys~~ describes the magnitude of discripency (difference) b/w theory & observation.
- It is computed by following formula:-

$$\boxed{\chi^2 = \sum \frac{(O-E)^2}{E}}$$

here O = observed frequency
E = expected "

- range of χ^2 value is zero to ∞.
- it always positive.
- **It is statistic not a parameter.
- if the value of χ^2 is zero it means that the observed & expected frequency completely coincidence.
- the greater the value of χ^2 there is more discripency b/w observed & expected frequency.

(observed variation)

$$\boxed{\text{Degree of freedom} = N-1 \quad (\cancel{n}(N-1))}$$

N ⇒ no. of categories/classes (not no. of samples)

no. of classes (new-1)

- the calculated value is compared with tabulated value for given degree of freedom at a certain ~~level~~ certified level of significance.
- If the calculated value is equal to or greater than tabulated value then we reject the null hypothesis & concludes that difference b/w observed & expected frequency is significant & it is

not just due to sampling of chance.
- If calculated value is less than then null hypothesis is accepted & the difference b/w expected & observed frequencies is nonsignificant & difference is just only due to chance.

~~Here~~. To test the independence of attributes the data are cross classified in a specific manner the table giving simultaneous classⁿ of the body of the data in 2 different ways is known as ~~Quantigenous table~~ contigency table

row

Gender	Graduate	Post Graduate	
Male	50 a	20 b	$R_1 = 70$
female	55 b	17 d	$R_2 = 72$
	$a+c = 105$ C_1	37 C_2	

column

$a+b+c+d = N = 150$

$$\boxed{\chi^2 = \frac{(ad-bc)^2 N}{R_1 R_2 C_1 C_2}}$$

(degree of freedom)

$$\boxed{D.F = (r-1)(c-1)}$$

(M/N) r = no. of row, c = No. of column

- If frequency of any cell is below 5 then some corrections is needed in the Yates correction

ad aur bc mein jo bada ho usme subtract karein

$$\boxed{\frac{\left(ad - bc - \frac{N}{2}\right)^2 N}{R_1 R_2 C_1 C_2}}$$

Uses of kia square

- kia square test is used to test the "goodness of fit" it enables us to acertain how appropriately the theoritical distⁿ fit with the emperical distⁿ. so, it is the "test of concordance"
- it is a test to test the independence of attributes.
- with the help of kia square we can find out whether two or more attributes are associated or not.
- It is also used to test the homogeneity.
- It helps to determine whether two or more independent sample are drawn from the same popⁿ or different popⁿ.

8/1/25 Wednesday

ANOVA (Analysis of Variance)

- observations noted from an experiment of any character show variation. this variation is due to an ~~no~~ no. of factors known as sources of variations.
- the portion of " caused by different sources are known as components of variations.
- the ~~ano~~ ANOVA, is a simple arithematical process of sorting out the components of variations in a given data. it was developed by R.A. fischer in 1920. It is a tool by which total variation may be split up into several physical assignable components.

Role / functions of ANOVA

- It ~~Studi~~ Sorts & estimates the variance components
- It provides the test of significance.

A) Treatment ⇒ any factor in the ~~Ex~~ experiment that is controlled at different levels or values & supposed to be have an effect on the experimental units.
In other words any controlled factor or manipulations delibrately applied to the experimental units observed in exp.

B) Sum of Squares ⇒ It means the sum of squares of deviations of the variats from their mean

C) Mean Sum of Square ⇒ Synonyms of variance
It is obtained by dividing the sum of square by corresponding degree of freedom.

d) error / Residual component ⇒ component of variation which remains unexplained by different sources of variation is considered due to error.

Types

- 2 types of ANOVA
① One way ② Two way
~~is in this data are classified acc to only one criterion.~~
- There are 2 ways components of variation one is due to variation b/w the grps or treatment.
- & other is due to " among the ~~variants~~ variates within the grp.

second component is known as error variance
- if μ is the overall mean & α is the component due to grp & ϵ (abb alsilog) is component due to error then any observed value will be the sum of three parts.
A) an overall mean.
B) deviation due to the grp
C) ~~Trans~~ or chance error

Statistical model $\boxed{X_{ij} = \mu + \alpha_i + \epsilon_{ij}}$

X_{ij} = j^{th} variate of the i^{th} grp
μ = overall mean
α_i = effect of i^{th} grp
ϵ_{ij} = error component / chance.

F-test → is named in the owner of great Stastician R.A. fischer by the George A. Snedecor
- object of F-test is to find out whether the two independent estimates of population variants significantly or whether the two samples may be regarded as drawn from normal popⁿ having the same variants.

9/1/25 Thursday

- for carrying out the test of significance we calculate the ratio F & formula is.

बड़ा upper छोटा नीचे $\boxed{F = \frac{S_1^2}{S_2^2}} = \boxed{\frac{\text{larger variance}}{\text{smaller variance}}}$

S_1 is always the larger variance estimates of variants.
- here there are 2 degree of freedom
1) $\nu_1 = n_1 - 1$ 2) $\nu_2 = n_2 - 1$

Page No.:
Date:
YOUVA

The calculated value of F is compared with tabulated value for both degree of freedom at 5% or 1% level of significance.
- If the calculated value is higher than the tabulated value then we reject null hypothesis difference is significant.
& lesser then null hypothesis & difference isnt significant.

First → null hypothesis
2 → alternate))
3 → level of significance
4 → Assumption
5 → computation → i) correction factor

T_1	T_2	T_3	T_4
15	1	7	10
21	2	8	2
32	3	11	3
45	4	12	4
50	5	14	5

$$\boxed{\text{correction factor (C.F.)} = \cancel{\frac{(\sum X_{ij})^2}{N}}}$$

$$\cancel{\sum X_{ij}}$$

$\sum X_{ij}$ ⇒ sum of all observations
N → no. of samples / categories

$T_1 \neq T_2 \neq T_3 \neq T_4$

ii) Total sum of square (TSS)

$$\boxed{\sum x_{ij}^2 - C.F.}$$

iii) Between sum of square (BSS)

$$\frac{\sum T_i^2}{n_i} \quad \boxed{\sum\sum \frac{T_i^2}{n_i} - C.F.}$$

mean sum square is equal to variance

④ within sum of square (WSS)

$\boxed{WSS = TSS - BSS}$

ANOVA Table

Source of Variation	Degree of freedom	Sum of sq.	mean sum of square	F-Value
① Between treatment of grp -	$T-1$	BSS	$MSB = \frac{BSS}{T-1}$	$\boxed{F = \frac{MSB}{MSW}}$
② within grp/ treatment error component	$N-T$	WSS	$MSW = \frac{WSS}{N-T}$	
③ Total	$N-1$	TSS		

15/1/25 Wednesday

Two Way ANOVA

- in 2 way ANOVA the data are classified on the basis of 2 criteria -

In this there are 2 null hypothesis
no difference b/w 1 criteria
" " " 2 "

alternate hypothesis → difference b/w 1 criteria
" " 2 "

It is maximum probability of type-1 error →

	G_1 x2	G_2 x2	G_3 x2	Sum of each grp
T_1	7 49	8 64	6	21
T_2	10 100	11 121	3	24
T_3	4 16	7	9	20
	8 64	6	9	23
sum of each	29	32	27	88

$$\boxed{\text{Correction factor} = \frac{(\Sigma X)^2}{N} = \frac{(88)^2}{12}}$$

Total sum of square (TSS

$$\boxed{TSS = \Sigma x^2 - C.F.}$$

Between sum of square / Treatment

$$BSS\,T = \Sigma x^2 - C.F.$$

$$\cancel{BSSG} \left[\frac{(\Sigma \cdot T_1)^2}{N_1} + \frac{(\Sigma \cdot T_2)^2}{N_2} + \frac{(\Sigma \cdot T_3)^2}{N_3} + \frac{(\Sigma \cdot T_4)^2}{N_4}\right] - C.F.$$

$$BSSG = \left\{\frac{(\Sigma G_1)^2}{K_1} + \frac{(\Sigma G_2)^2}{K_2} + \frac{(\Sigma G_3)^2}{K_3}\right\} - C.F.$$

Within sum of square

$$WSS = TSS - BSST - BSSG$$
$$WSS = TSS - (BSST + BSSG)$$

ANOVA Table

Source of Variation	Degree of freedom	Sum of sq.	mean sum of sq.	F-value
① B/w treatment	$T-1$	BSST	$\frac{BSST}{T-1}$	$\frac{MSBT}{MSW}$
② B/w grp (block)	$G-1$	BSSG	$\frac{BSSG}{G-1}$	$\frac{MSBG}{MSW}$
★★★★ ③ Within sum of square/component	$(T-1)(G-1)$	WSS	$\frac{WSS}{(T-1)(G-1)}$	
④ Total	$N-1$	TSS		

Design of Experiment

- concept ~~wa~~ of design of experiment given by R.A. fischer.
- experimental design concerned the arranging of treatments or variables in a such a manner that the inferences & conclusions regarding the effects of these treatments can be easily done & their reliability is measured.

** Statistical inference ⇒ systematic drawing of a conclusion or conclusions from given data

16/1/25 Thursday - Principles of design of Exp

- there are 3 " " " " " ".

1. replication ⇒
2. Randomization
3. local control

- the repitation of treatments under investigation is known as replication.
- it is essential that treatment should atleast be duplicated or preferably replicated so that an estimate of error may be obtained & the significance of the observed treatment difference is be trusted.
- rep^n provides means for estimating error for different test of significance.
- ~~so there~~ 2 ~~different~~ advantages of replication :-

1) It ↑ses the precision by ↓ing the error.
2) It provide the estimate of error.

- the no. of replication should be such as to provide atleast **12** degree of freedom for error.

② Randomization

- the allocation of ~~interval~~ treatments to the different experimental units by a random process.
- It minimizes the chances of biasness.
- It facilitates equal environmental conditions to different treatments.
- It should be accompanied by sufficient no. of repl^n.
- by repl^n & randomization valid estimate of error is obtained.

③ Local control ⇒ the principal of making use of greater homogenity in a grp of experimental units for reducing experimental error.

- this is done by dividing the experimental material into homogenous grp of units.
- local control is meaningful when the treatments are replicated several times.

Advantages

① when experimental material is heterogenous w.r.t certain characters by local control. we are in pos^n to study all the treatments for their relative merits under different sets of homogenous conditions.

② the chance error is also diminished & precision of estimates of treatment effects is ↑sed.

- by repl^n & local control error is diminished.

I
Replication

II
Randomization

III
Local control

Valid estimate
the error

~~to~~ diminish
the error

Accuracy

the accuracy of a measurement signifies the closeness ~~which~~ with which a measurement approaches the true value / closeness to true value.

- the lesser the deviation from true value the greater is the accuracy of any measurement.

Precision

- denotes closeness with which measurement approaches the average value
- the lesser the variability of measurement from their average value the greater the precision of these measurements & their estimates.
- accuracy implies the lack of bias while precision involves the control of variability.
- by bias we mean the systematic error.
- by variability means the random error
- random error may be defined as one whose expectations is zero that is such that in a series of observations difference b/w the true value

Two way Anova ← (RBD → Randomized block design)

Page No:
Date:

(population parameter) & the sample estimates ↑ses on the average as the no. of observation ↑ses.

- a systematic error or bias on other hand that will never be diminished nearly by ↑ing the no. of observations bcz that is due to such a source of error which can't be controlled by ↑ing the no. of observations or repeating the experiments unless the source of error is removed.

28/1/25 Tuesday

CRD ≠ IMP.

[Completely randomized Design]

It is equivalent ~~two~~ to one way Anova.

- this is the simplest type of design in which whole experimental material is divided into a no. of experimental units depending upon the no. of treatment & no. of replications for each.

After that treatments are alloted by units entirely by chance (Randomly)

* random allocation of treatment *

Str of ANOVA

no. of treatment $= n/t$

" " replication $= r$

total no. of experimental units $= n \times r / t \times r = N$

If treatment have vary no. of replication r_1, r_2, r_3 --- so on then total no. of experimental units $= r_1 + r_2 + r_3 - - - -$

there are two sources of variation -

1. within treatment → component provides a basis for
2. B/W "

estimation of error.
degree of freedom total $= n-1$
b/w treatment $= t-1 / n-1$
within treatment $\rightarrow N-T / N-n$ $\quad N =$ total no. of unit
$n =$ no. of treatment.
after this write the whole procedure of ANOVA
(puri table & procedure likna hai)

** # Merits

- In this design any no. of treatments & replicates may be used.
- the no. of replicates may be vary from treatment to treatment.
- the stastical analysis of data is very easy.
- the method of analysis remain simple when result from some units are rejected or missing.
- the relative loss of information due to missing data is smaller than that with any other design.
- this design is specially useful in small experiment where the experimental material is scared / limited & homogenous.
- this design provides max. no. of degree of freedom for estimation of error as compared to other designs for a given no. of treatment & a given no. of experimental units.

demerits

- It is not much accurate as no restriction on the randomization of treatment.

** in CRD principle of local control is not followed. **

Uses → It should be used when experimental material is limited & homogenous.

when it is expected that some of units will destroyed or will fail to respond.

RBD (randomized block design).

- In this design whole exp. material is divided into homogenous grp each of which constitutes a single replication. each of these groups further divided into a no. of experimental units which are equal in all aspects.
- treatments are applied to these units by any random process.
- In this design the no. of blocks must be equal to no. of replication fixed for each treatment whereas no. of plots/individual in each block should be equal to no. of treatment.
- exp. error within each block are kept to be as small as practically possible & variation from block to block as great as possible.
- randomization of treatment in each block afreshed.

Merits

when exp. material is heterogenous we go for RBD & any no. of treatment & any no. of replication.

It is ~~repli~~ equivalent to two way Anova.

1. Grouping

Degree of freedom is more in CRD then RBD.

2. no. of individuals in a block / group = no. of treatments

no. of individual for a treatment = no. of block / group.

for treatment = 4
group / block = 3

	T_1	T_2	T_3	T_4
B_1	x_{11}	x_{12}	x_{13}	x_{14}
B_2	x_{21}	x_{22}	x_{23}	x_{24}
B_3	x_{31}			

~~x_{41}~~

Two way ANOVA

** $\boxed{(B-1)(T-1)}$

$B-1=2$

$T-1=3$

$(B-1)(T-1)=6$

$N-1=11$