NESSDAY 8:30AM.

TEXTBOOKS

ABE 302

=> K.A. STROUD Engineering mathematics 6th edition

=> Higher Engineering mathematics S.CHand, H.K.DAS

COURSE OUTLINE.

Probability and Statistics.

Probability space theorem, Conditional probability and Independent variable, random Variables. discrete and Continous distribution, mean and Variance, binomials, Poisson and hyper geometry.

Centre Limit theorem, Statistical inference

Elementary Sample theory for normal population or mean population of Variance

Chi-square test, Simple linear regression and engineering application.

TEXTBOOKS => statistics, the exploration and analysis of Data by Roxy peck & JOY DEVORE.

~~Schum series~~ => Schaum Series,

=> Yout tube - video on statistical Inferences

Electromechanical ~~EEE~~346 (8-10) Digital elec ~~EEE324 (10-12)~~ Instrumentation EEE 372 (10 - 12)

DAY ~~ABE302 (8-10)~~, EEE352 (10-12) => power by Sanni

ABE302 (8-10), EEE 392 (Practical 10-12).

EEE324 (7-10) digital ABE302 (10-12) communic EEE 332 (12-3).

Control EEE 302 (9:00 to ----

MON 8pm Irinriaye (10 press forward from fresh). 91.5 Irinriaye (gbankogbi) (9:00 pm).

91.5 (13 press forward from fresh

Tues. 91.5 (irinriaye 9:00pm)

Wed Fresh 7:30-8:30 pm (lowujo eranko)

6 press forward after fresh is Adaba (88.9)

7 press " " " is Voice (89.9).

MON 7:30 Voice 7 press forward from fresh Oluwaloseyifunmi for gift

PROBABILITY 5-5-2021

Probability is the measure of likelihood that a particular event will occur in any one trial / experiment carried out in a prescribed condition.

NOTATION.

The probability that a certain event A will occur is denoted by $P(A) = p(A)$ and also equal to $(\bar{p}$ Success$)$.

SUCCESS / FAILURE.

When an event occur in any one trial its called success but when it fails to occur it's called failure.

In N trials there are x success, there will also be $(N-x)$ failure.

$\frac{x}{N} = P(\text{Success})$ $\frac{N-x}{N} = P(\text{failure})$

It's applied that $P(A) + P(\text{Not } A) = 1$

i.e $\boxed{\frac{x}{N} + \frac{N-x}{N} = 1}$ i.e $\boxed{P(A) + P(\bar{A}) = 1}$

Event "not A" is the complement of event A and often written as $\bar{A}$

TYPES OF PROBABILITY

→ Empirical or experimental probability.
⇒ Classical or Theoretical probability.

EMPIRICAL PROBABILITY.

This is form from what we known as experiment.

EXPECTATION ⇒ This is defined as the number of trials and probability that event A will occur.

$E = N \times P(A)$.

Where $P(A) = \frac{1}{12}$

(07066469141)

CLASSICAL PROBABILITY

The classical approach is based on theoretical number of ways ~~on~~ which its possible for an event to occur. Based on assumption Classical probability Prob(P) of an event (A) occuring is defined by;.

$P(A) = \frac{\text{Numbers of ways in which event will occur}}{\text{Total number of all possible outcomes}}$.

MUTUALLY EXCLUSIVE AND NON-MUTUALLY EXCLUSIVE EVENT

Mutually exclusive events are events which cannot occur together. Single event (note, it can occur but not d same tym)

Mutually non-exclusive events are events which occur together simultaneously when there is a pair of events e.g scoring 2's in a pair of dice.

ADDITIONAL LAWS OF PROBABILITY

For a given events (A) and (B) are mutually exclusive if A and B are mutually exclusive the probability is either event (A) or (B) but not both.

$P(A \text{ or } B) = P(A) + P(B) = \frac{x+y}{n} = \frac{x}{n} + \frac{y}{n}$

If events are non-exclusive,

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

INDEPENDENT EVENT AND DEPENDENT EVENT

Events are independent when the occurrence of event doesn't affect the probability of the occurrence of the second event (Replacement).

Events are said to be dependent when one event does affect the probability of the occurrence of the second event (non-replacement).

CONDITIONAL PROBABILITY.

Conditional probability of an event B occuring when given an event A have already taken place vice versal. In case of an event as stated above, $P(\frac{B}{A})$ is the probability

LIBARY READING

Reading. Additional law

Exp. mutually non-exclusive.

In rolling a dice',
the proba of scoring a multiple of 3.
$(3, 6) = \frac{2}{6} = P(A)$

In a rolling a multiple of ~~3~~.
$(2, 4, 6) = \frac{3}{6} = P(B)$.

The find prob of A or B occuring.

Sol.

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

To find $P(A \text{ and } B) = ?$

NOTE => prob of scoring multiple of 3 or 2 is $P(A) + P(B) = \frac{2}{6} + \frac{3}{6} = \frac{5}{6}$

Note at prob of multiple of 3 = 3, 6
" " " " 2 = 2, 4, 6.

One outcome (6) appears in both set of outcome. ∴ The outcome has been counted twice.

The $P(\text{Six}) = \frac{1}{6} = P(A \text{ and } B)$ (i.e picking 1)

$\therefore P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
$= \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{2}{3}$.

OR.

OR

multiple of 3 = 3, 6
mutiple of 2 = 2, 4, 6. → i.e 5 Elements

Since 6 ~~excist twice~~, ∴ u count 6 once and not the twice.
i.e 3, 6, 2, 4 or 3, 2, 4, 6.
(Counting the element, they are 4 elements)
Also, Since a fair die is 6 sides.

$\therefore \text{Prob of } (A \text{ or } B) = \frac{4 \text{ element}}{6}$
$= \frac{2}{3}$.

IDENTIFICATION OF EXCLUSIVE AND NON EXCLUSIVE

EXCLUSIVES.

=> A = {heart} B = {ace of spades}
=> A = {5 of clubs} B = {10 of clubs}.
=> A = {Card < 10} B = {King of diamond}.

NON-EXCLUSIVES

=> A = {ace} B = {Black card}.
=> A = {red card} B = {Jack}.

EXAMPLE.

=> A Single card is drawn from a pack of 52 playing cards.

(a) If event A = {drawing an ace} and event B = {drawing a seven}. The prob of drawing either ace or a seven i.e
$P(A \text{ or } B) = ?$

Sol.

$\boxed{\text{Note}}$ $\boxed{\text{ace} = 13}$

Sol

Bcox these are mutually exclusive events in any one trial.

$\therefore P(A \text{ or } B) = P(A) + P(B)$.

4 aces in the pack $\therefore P(A) = P(\text{ace}) = \frac{4}{52} = \frac{1}{13}$

4 sevens in the pack $\therefore P(B) = P(\text{seven}) = \frac{4}{52} = \frac{1}{13}$

$\therefore P(A \text{ or } B) = P(\text{Ace or seven}) = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$

2) If event A = {drawing a King} and event B = {drawing a red card}, then $P(A \text{ or } B) = ?$

Sol.

These are non-exclusive events since we could draw the King of hearts or the King of diamonds.

$\therefore P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

4 kings in the pack $\therefore P(A) = P(\text{King}) = \frac{4}{52} = \frac{1}{13}$

26 red cards in the pack $\therefore P(B) = P(\text{red card}) = \frac{26}{52} = \frac{1}{2}$

2 red kings in the pack $\therefore P(A \text{ and } B)$

$= P(\text{red king}) = \frac{2}{52} = \frac{1}{26}$

$\therefore P(A \text{ or } B) = \frac{1}{13} + \frac{1}{2} - \frac{1}{26} = \frac{2+13-1}{26}$

$= \frac{14}{26} = \frac{7}{13}$.

$\therefore$ for mutually exclusive $P(A \text{ or } B) = P(A) + P(B)$

for mutually non-exclusive $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

THINGS TO NOTE.

=> Mutually non-exclusive event => They are events dat happen simultaneously. e.g in rolling a die, the event of obtaining a multiple of 3 & d event of obtaining a multiple of two can occur togeda to give 6 i.e it occur together if a 6 is thrown.

THINGS TO NOTE.

=> DIE => This is a small cube. ~~Dot~~ dots mark on their faces. plural of the die is dice. On throwing a die, the outcome is d no of dots on its upper face.

=> CARDS => A pack of cards consists of four suits i.e spades, Hearts, Diamonds and clubs. Each suit consists of 13 cards, nine cards numbered 2, 3, 4 --- 10, an ACE, a KING, a QUEEN and a JACK or KNAVE. Colour of spades and clubs is black and dat of Hearts and diamonds is red. Kings, Queens and Jacks are known as face card.

=> EXHAUSTIVE EVENTS OR SAMPLE SPACE => The set of all possible outcomes of a single performance of an experiment is exhaustive events or sample space. Each outcomes is called a sample point. In case of tossing a coin once, S = (H, T) is the sample space. Two outcomes - Head and Tail - constitute an exhaustive event bcox no oda outcome is possible. (In throwing a die, there are exhaustive cases) (In throwing n dice, there are exhaustive cases of 6^n.

=> RANDOM EXPERIMENT => There are experiments in which results may be altogeda diff. even though they are performed under identical conditions. They are known as random experiments. Tossing a coin, a coin or throwing a die is random experiment. (If u throw die continuously it may not give same result) Throwing a die is trial, getting nos 1, 2 or ~ 6 is an event.

=> TRIAL AND EVENT => Performing a random experiment is called a TRIAL and outcome is termed as EVENT. Tossing of a coin is a trial and the turning up of head or tail is an event.

=> EQUALLY LIKELY EVENTS => Two events are said to be "equally likely" if one of dem can't be expected in preference to the oda. For instance, if we draw a card from well-shuffled pack, we may get any card, then the 52 diff. cases are equally likely. In a die, any no btw 1 to 6 is equally likely to come out.

=> INDEPENDENT EVENT => Two events may be independent, when the actual happening of one doesn't influence in any way d probability of the happening of the oda.

E.g The event of getting head on first coin and d event of getting tail on the second coin in a simultaneous throw of two coins are independent.

=> Mutually Exclusive Event (incompatible) => Two events are known as mutually exclusive when the occurrence of one of dem excludes the occurrence of the oda. For example, on tossing of a coin, either we get head or tail, but not both. or in rolling a die the event of throwing a six and dat of ...

Compound Event => When two or more events occur in composition with each other, the simultaneous occurrence is called a compound event. When a die is thrown, getting a 5 or 6 is a compound event.

FAVOURABLE EVENT => The events which ensure the required happening, are said to be favourable events. E.g. in throwing a die, to have the even numbers 2, 4, & 6 are favourable cases. While the favourable of even multiple are 3, 6. In a throw of two dice, the no of favourable cases of getting a sum 6 is 5 i.e (1,5) (5,1) (2,4) (4,2) (3,3)

Conditional Probability => The prob of happening an event A such that event B has already happened, is called the conditional probability of happening of A on the condition that B has already happened. It is usually denoted by $P(A/B)$.

=> Odds in favour of an events and odds against an events:

If number of favourable ways = m,

Number of not-favourable events = n

i) odds in favour of the event $= \frac{m}{n}$

ii) odds against the event $= \frac{n}{m}$

=> Classical definition of probability => If there are N equally likely, mutually exclusive, and exhaustive of events of an experiment and m of these are favourable, then the prob of the happening of the event is defined as $\frac{m}{N}$

=> Expected value => If $P_1, P_2, P_3 --- P_n$ are the prob of the events $x_1, x_2, x_3 --- x_n$ respectively then the expected values

$E(x) = P_1x_1 + P_2x_2 + P_3x_3 + --- P_nx_n$

$\therefore E(x) = \sum_{r=1}^{n} P_r x_r$

Examples.

=> A bag contains 7 white, 6 red and 5 black balls. Two balls are drawn at random. Find the prob that they will both be white?

Sol.

Total no of balls = 7+6+5 = 18.

Out of 18 balls, 2 can be drawn in $^{18}C_2$ ways.

(Single trial time)

$\therefore$ The exhaustive no of cases $= {}^{18}C_2 = \frac{18 \times 17}{2 \times 1} = 153$

Out of 7 white balls, 2 can be drawn in $^7C_2$

$^7C_2 = \frac{7 \times 6}{2 \times 1} = 21$ ways

$\therefore$ favourable no of cases = 21

Prob $= \frac{21}{153} = \frac{7}{51}$

=> Four cards are drawn from a pack of cards. Find the prob that (i) all are diamonds (ii) there is one card of each suit (iii) There are two spades and two hearts.

Sol.

4 cards can be drawn from a pack of 52 cards in;

$^{52}C_4 = \frac{52 \times 51 \times 50 \times 49}{4 \times 3 \times 2 \times 1} = 270725$

(i) There are 13 diamonds in the pack and 4 can be drawn out of them in $^{13}C_4$ ways

$\therefore$ favourable number of cases;

$^{13}C_4 = \frac{13 \times 12 \times 11 \times 10}{4 \times 3 \times 2 \times 1} = 715$.

Required Prob $= \frac{715}{270725} = \frac{11}{4165}$.

(ii) ~~Note~~ we have 4 suits.

But here we have ~~two~~ spades and two hearts = spade + spade + heart + heart = 4 suits.

Note each of dem has 13 cards.

$\therefore$ favourable no of cases $= {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1$

$= \frac{13}{1} \times \frac{13}{1} \times \frac{13}{1} \times \frac{13}{1} = 169 \times 169$

Required Prob $= \frac{169 \times 169}{270725} = \frac{2197}{20825}$

(ii) Their are 4 suits each containing 13 cards.

$\therefore$ favourable no of cases $= {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1 \times {}^{13}C_1$

Required Prob $= \frac{13 \times 13 \times 13 \times 13}{270725} = \frac{2197}{20825}$

(iii) 2 spades out of 13 can be drawn in $^{13}C_2$ ways

also 2 hearts out of 13 can be drawn in $^{13}C_2$ ways.

$\therefore$ favourable cases $= {}^{13}C_2 \times {}^{13}C_2 = \frac{13 \times 12}{2 \times 1} \times \frac{13 \times 12}{2 \times 1}$

$= 78 \times 78$

Required Prob $= \frac{78 \times 78}{270725} = \frac{468}{20825}$

⇒ A bag Contains 50 tickets numbered 1, 2, 3, ----, 50, of which Five are drawn at random and arranged in ascending order of magnitude ($x_1 < x_2 < x_3 < x_4 < x_5$), what is the prob dat $x_3 = 30$?

Sol

Exhaustive ~~Cases~~ no of Cases: $^{50}C_5$

If $x_3 = 30$, the d two tickets with no x_1 and x_2 must come out of 29 tickets numbered 1 to 29, and dis can be done in $^{29}C_2$ ways. The oda two tickets with numbers x_4 and x_5 must come out of the 20 tickets numbered 31 to 50 and dis can be done in $^{20}C_2$ ways.

∴ Favourable number of cases $= {}^{29}C_2 \times {}^{20}C_2$

Required prob $= \dfrac{{}^{29}C_2 \times {}^{20}C_2}{{}^{50}C_5}$

$= \dfrac{\frac{29\times28}{2\times1} \times \frac{20\times19}{2\times1}}{\frac{50\times49\times48\times47\times46}{5\times4\times3\times2\times1}} = \dfrac{551}{15134}$

19-5-2021 CLASS

BINOMIAL DISTRIBUTION.

Binomial formulas.

$b(x; n, p) = nC_x * p^x \times (1-p)^{n-x}$

where;

x = Total no of Successes

p = probability of a Success in an individual trial

n = number of trials

~~DEF~~

$^nC_x = \dfrac{n!}{(n-x)!}$

$P = {}^nC_r\, p^r q^{n-r}$

⇒ A Coin is tossed 10 times, what is d Prob. of getting exactly 6x

Sol

$p = 0.5$, — Success

failure $q = 1 - 0.5 = 0.5$ i.e $q = 1-p$

$n = 10$

Prob of getting $6x = 6$.

$P_6 = {}^nC_x \times p^x \times q(1-p)$

~~$= {}^{10}C_6 \times p^{0.5} \times q^{x}$~~

$= {}^{10}C_6 \times p^{6} \times (1-0.5)^{10-6}$

$= \dfrac{10!}{(10-6)!} \times 0.5^6 \times 0.5^4$

$= \dfrac{10!}{4!}$ ~~×~~ $10\times9\times9$

POISSON DISTRIBUTION

This occur ~~when~~ in real eve

Mean $= np = 10 \times 0.5 = 5$.

Variance $= npq = 10 \times 0.5 \times 0.5$

Standard deviation $= \sqrt{npq}$

POISSON DISTRIBUTION.

$$P(x = \mu) = \frac{(e^{-\mu})(\mu^x)}{x!}$$

s.t
μ = mean
$\bar{X}$ = means of sample.

$$P(x;\mu) = \frac{(e^{-\mu})(\mu^x)}{x!}$$

where

x = actual number of success dat that result from experiment.

$e = 2.71828$

μ = mean of probability.

properties.

mean $= \mu$

Variance $= \mu$

Standard deviation $= \sqrt{\mu}$

Example

=> Average no of ~~houses~~ homes sold by Solomon Company is ~~2 houses~~ 2 homes per day. What is d Prob dat exactly 3 homes will be sold tomorrow.

Sol.

$\mu = 2$ homes, since 2 homes sold per day

$x = 3$, $e = ?$

$$P(x;\mu) = \frac{(e^{-\mu})(\mu^x)}{x!}$$

$$= \frac{e^{-2} \times 2^3}{3!}$$

HYPERGEOMETRIC

When u ar taken a sample no from a sample drawn from population of size capital letter

If a population ~~comparised~~ N contains small K items of a successes and N−K failure then d prob of d hypergeometric random variable ~~is~~ n, the no of successes in a random sample of size n is;

$$P(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} =$$

$$= \frac{({}^{K}C_{k})({}^{N-K}C_{n-k})}{{}^{N}C_{n}}$$

where K = success from population

k = success from sample.

EXAMPLE.

A ~~team~~ deck of cards contain 20 cards 6 ~~red~~ red cards and 14 black cards, 5 ~~black~~ cards ar drawn random without replacing dem. What is d prob dat exact 4 red cards are drawn.

Sol.

Exactly (k) = 4, K = 6

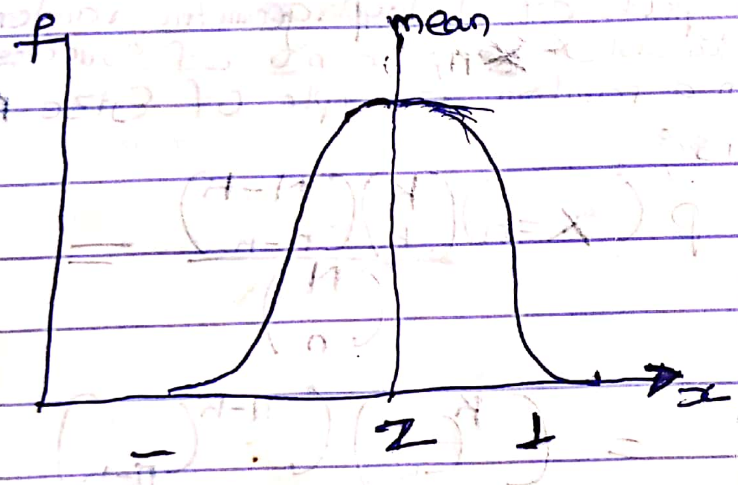
N = 20, N − K = 20 − 6 = 14

n = 5, ~~n−k = 5 − 5 = 1~~

n − k = 5 − 4 = 1

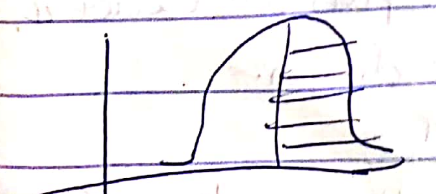
$$\frac{\binom{6}{4}\binom{14}{1}}{\binom{20}{5}} = \frac{\left({}^{6}C_{4}\right)\left({}^{14}C_{1}\right)}{{}^{20}C_{5}}$$

Normal Distribution.



Z = Size
$\bar{x}$ = mean of sample.
μ = mean of population.
σ = Standard deviation of population
n = Size of sample.

$$Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$



1 tail u subtract 0.5

2 tail u don't subtract anything.

TEXTBOOK ⇒ Advanced Engineering Mathematics by S. CHAND H.K. DASS P781

Reading

BINOMIAL DISTRIBUTION

$$P_{(r)} = {}^{n}C_{r}\, p^{r} q^{n-r}$$

n = no of trials.
r = Total no of success trials.
p = Probability of happening of an event in one trial (~~success~~)
q = Probability of not happening of an event in one trial

degree of freedom (df) $= n-1$

CLASS

CHI SQUARE (x^2)

$$x^2 = \frac{(O-E)^2}{E} \text{ OR } \Sigma \frac{(O-E)^2}{E}$$

EXAmple

class	O	E
ABC	5	5
CVE	6	5
EEE	7	5

$$x^2 = \frac{(5-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(7-5)^2}{5}$$

$$x^2 = 0 + 0.2 + 0.8 = 1.0$$

degree of freedom (df) $= n-1$

Table Value $< x^2 \rightarrow$ Non significant

" " $>$ " $\rightarrow$ Significant

~~degre~~ for rows, $n = 3$

for Column, $n = 2$.

$\therefore$ degree of freedom $= \underset{\text{rows}}{(n-1)}\underset{\text{column}}{(n-1)}$

$= (3-1)(2-1)$

$= (2)(1) = 2$

~~Note Suppose to be given~~

Note for a complete qn, we suppose to be given a particular percentage of level usually 5% or 0.05, den u used degree of freedom and dis 5% to locate a particular number of chi square table.

If Table Value gotten $< x^2$, it is not significance

But if " " " $> x^2$, it is significance

i.e at Significance, the theorem is accepted.

Grouped

Observe

	Small	medium	big	add
White	10	12	8	30 +
Red	5	13	10	28 +
Black	2	5	12	19
add	17 +	30 +	30	77

expected ~~as~~ = 77

(E) EXPECTED

	Small	medium	large
White	$\frac{30 \times 17}{77}$	$\frac{30 \times 30}{77}$	$\frac{30 \times 30}{77}$
Red	$\frac{17 \times 28}{77}$	$\frac{30 \times 28}{77}$	$\frac{28 \times 30}{77}$
black	$\frac{19 \times 17}{77}$	$\frac{19 \times 30}{77}$	$\frac{19 \times 30}{77}$

i.e

	Small	medium	large
white	$6.6 = 7$	$11.7 = 12$	$11.7 = 12$
Red	$6.2 = 6$	11	11
Black	~~7.11~~ = 4	7	7

$$x^2 = \Sigma \frac{(O-E)^2}{E}$$

from small

$$x^2 = \frac{(10-7)^2}{7} + \frac{(12-12)^2}{12} + \frac{8-12}{12} + \cdots$$

degree of freedom $= \underset{\text{Row}}{(3-1)}\underset{\text{Column}}{(3-1)} = 4$

Where $n = 3$ i.e 3 rows & 3 column

$n = 3$ for rows

$n = 3$ for Column

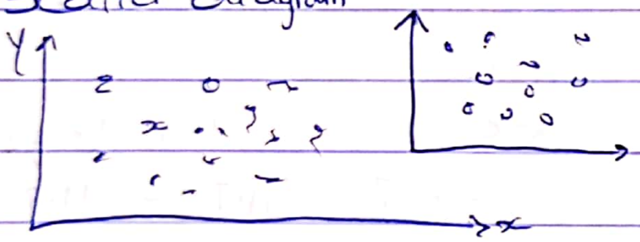
$$x^2 = \left(\frac{10-7}{7}\right) + \left(\frac{12+12}{12}\right) + \left(\frac{8+12}{12}\right) + \left(\frac{5-6}{6}\right) + \left(\frac{13-11}{11}\right) + \left(\frac{10-11}{11}\right) + \left(\frac{2-4}{4}\right) + \left(\frac{5-7}{7}\right) + \left(\frac{12-7}{7}\right).$$

MEASURE Of ASSOCIATION

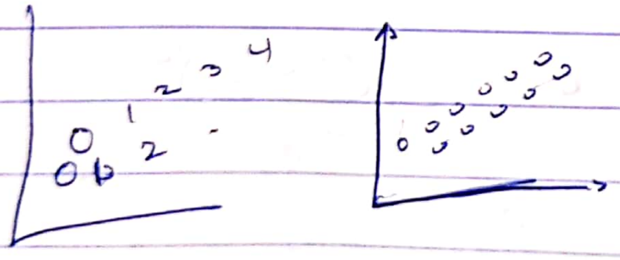
The Study of measure of association has to do with bivariance data whereby two or more variables are considered as a ~~them~~ means of viewing the relationship existing btw dem. When an increase in ~~one~~ variable leads to an increase in anoda variable we have a positive association of relationship. However when an increases in a variable leads to decrease of anoda, we ~~av~~ a -ve association of relationship.

Types of STATISTICAL RELATIONSHIP

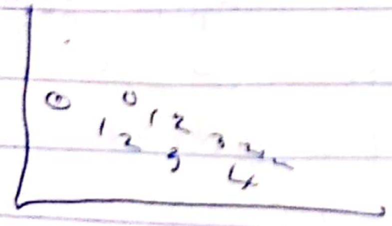
=> Scatter diagram



=> positive diagram



=> Negative diagrams



Karl Pearsons Products Moment Correction Coefficient

It the most popular mathematical method of quantifying or ~~correlat~~ measuring correlation. It is always denoted by r and is given by

$$r = \frac{\text{Covariance}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{\sigma_{xy}}{\sqrt{\sigma_x} \times \sqrt{\sigma_y}}$$

$$= \frac{\Sigma xy}{N(\sigma_x \sigma_y)}$$

$$r = \frac{\text{Covariance}(x,y)}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y}$$

$$= \frac{\Sigma xy}{N(\sigma_x \cdot \sigma_y)}$$ N = population size

where σ_x, σ_y = standard deviation of x and y respectively.

N = population size.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \cdot \Sigma y^2}}$$

$$r = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{\underbrace{(N\Sigma x^2 - (\Sigma x)^2)}_{\sigma_x}\underbrace{(N\Sigma y^2 - (\Sigma y)^2)}_{\sigma_y}}}$$

r is btw 0 and 1
When its close to 1, its relationship is HIGH but when it close to 0, its relationship is LOW.

[Reading]

CHI SQUARE

=> A die is thrown 270 times and d results of these throws are given below:

No appeared on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not.

Sol.

Null hypothesis Ho - Die is unbiased.
Under H_0, expected frequencies for each digit $= \frac{\Sigma F}{n}$

i.e $\frac{40+32+29+59+57+59}{6} = \frac{276}{6}$

$E = \frac{276}{6} = 46$.

∴ the expected value for each of them is 46 i.e

O_i	40	32	29	59	57	59
E_i	46	46	46	46	46	46
$O-E$	6	-14	-17	13	11	13
$(O_i-E_i)^2$	36	196	289	169	121	169.

$\cancel{x^2 = \frac{(O_i-E_i)^2}{E}}$ $\quad x^2 = \Sigma\left[\frac{(O_i-E_i)^2}{E_i}\right]$.

$x^2 = \frac{980}{46} = 21.30$.

degree of freedom $= n-1 = 6-1 = 5$.

from table,
At 0.05 i.e 5% with $n=5$, $x^2 = 11.07$

Since value from table < cal. x^2

i.e $11.07 < 21.30$ ∴ H_0 is rejected

i.e die is not unbiased i.e die is biased

Note, if value from table > Cal. x^2 then H_0 is accepted i.e die is unbiased

=> Records taken of d number of male and female births is 800 families having four children are as follows:

No. of males birth	0	1	2	3	4
No. of female birth	4	3	2	1	0
No. of families	32	178	290	236	94

Test whether the data ar consistent with the hypothesis dat the binomial law holds and the chance of male birth is equal to dat of female birth, namely, $p = q = 1/2$.

Sol.

H_0, the date are consistent with d hypothesis of equal probability for male and female births i.e $p=q=1/2$.

~~Using binomial distribution~~ to cal theoretical frequency.

$N(r) = N \times P(X=r)$

Where N is d total frequency.
N(r) is d no. of families with r male children

$P(X=r) = {}^nC_r\, p^r q^{n-r}$

where p and q are probability of male and female birth respectively,
n is d no of children.

$N(0)$ = No of families with 0 male children

i.e $\cancel{N(0) = 800}$ i.e $N(r) = N \times P(X=r)$

i.e $N(0) = 800 \times {}^4C_0 (\frac{1}{2})^4$

$= 800 \times 1 \times \frac{1}{2^4} = 50$

$N(1) = 800\ {}^4C_1 (\frac{1}{2})^1 (\frac{1}{2})^3 = 200$

$N(2) = 800\ {}^4C_2 (\frac{1}{2})^2 (\frac{1}{2})^2 = 300$

$N(3) = 800\ {}^4C_3 (\frac{1}{2})^1 (\frac{1}{2})^3 = 200$

$N(4) = 800\ {}^4C_4 (\frac{1}{2})^0 (\frac{1}{2})^4 = 50$

Observed freq. O_i	32	178	290	236	94
Expected freq. E_i	50	200	300	200	50
$(O_i-E_i)^2$	324	484	100	1296	1936
$\left(\frac{(O_i-E_i)^2}{E_i}\right)$	6.48	2.42	0.333	6.48	38.72

$x^2 = \Sigma \frac{(O_i-E_i)^2}{E_i} = 54.433$

~~from~~ degree of freedom $df = 5-1 = 4$
from table,
~~0.05 or 5% again~~ 4 against 0.05 i.e 5%
On chi square table,
It is 9.488 $\approx$ 9.49

Since $9.49 < 54.433$, H_0 is rejected
i.e d data are not consistent with the hypothesis dat d binomial law holds and dat the chance of a male birth is not equal to dat of a female birth.

=> Verify whether Poisson distribution can be assumed from d data given below;

No. of defect	0	1	2	3	4	5
frequency	6	13	13	8	4	3

Sol, using Poisson.

H_0; poisson fit is a good fit to d data.

Mean of d given distribution $= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{94}{47} = 2$

To fit a Poisson distribution we require m. Parameter $m = \bar{x} = 2$.

By Poisson distribution, d frequency of r success is

$N(r) = N \times e^{-m} \cdot \frac{m^r}{r!}$, N is d total frequency

$N(0) = 47 \times e^{-2} \cdot \frac{(2)^0}{0!} = 6.36 = 6$

$N(1) = 47 \times e^{-2} \cdot \left(\frac{2^1}{1!}\right) = 12.72 = 13$

$N(2) = 47 \times e^{-2} \frac{(2)^2}{2!} = 12.72 = 13$

$N(3) = 47 \times e^{-2} \frac{2^3}{3!} = 3.48 = 9$

$N(4) = 47 \times e^{-2} \left(\frac{2^4}{4!}\right) = 4.24 = 4$

$N(5) = 47 \times e^{-2} \frac{(2^5)}{5!} = 1.696 = 2$

x	0	1	2	3	4	5
O_i	6	13	13	8	4	3
E_i	6.36	12.72	12.72	8.48	4.24	1.696
$\frac{(O_i-E_i)^2}{E_i}$	0.2037	0.00616	0.00616	0.02716	0.0135	1.0026

$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E} = 1.2864$

Conclusion

The cal. value of $\chi^2 = 1.2864$.
Tabulated value of χ^2 at 5% level of significance for $r = 6-2 = 4$ df is 9.49. Since d cal. value of χ^2 is less dan dat of tabulated value. H_0 is accepted i.e Poisson distribution provides a good fit to data.

=> The theory predicts the proportion of beams in d four groups G_1, G_2, G_3, G_4 should be in the ratio 9:3:3:1. In an experiment with 1600 beams d nos in d four groups were 882, 313, 287 and 118. Does d experimental result support d theory

Sol

H_0 The experimental result support d theory i.e there is no significant diff btw d observed and theoretical freq. Under H_0 the theoritical frequency can be cal. as follows;

$E(G_1) = \frac{1600 \times 9}{16} = 900$

$E(G_2) = \frac{1600 \times 3}{16} = 300$

$E(G_3) = \frac{1600 \times 3}{16} = 300$

$E(G_4) = \frac{1600 \times 1}{16} = 100$

To cal d value of χ^2.

Observed freq. O_i	882	313	287	11
Expected freq. E_i	900	300	300	10
$\frac{(O_i-E_i)^2}{E_i}$	0.36	0.5633	0.5633	3

$\chi^2 = \Sigma \frac{(O_i - E_i)^3}{E_i} = 4.7266$

Conclusion. ~~To~~

Table value of χ^2 at 5% level of significance of 3df is 7.815. Since d cal. value of χ^2 is less dan dat of d tabulated value. Hence H_0 is accep i.e the experimented result support the theory.

2-6-2021

Class.

LINEAR REGRESSION

used for prediction of

This ~~known for~~ dependent variable x and tends used for ~~prediction~~ dependent variable y

$y = a + bx$ — (i)

$\Sigma y = an + b\Sigma x$ — (ii)

$\Sigma xy = a\Sigma x + b\Sigma x^2$ — (iii)

Where a ÷ b = Constants.

n = no of sample.

$y = a + bx_1 + cx_2$ — (1)

$y = ax + b\Sigma x_1 + c\Sigma x_2$ — (2)

$\Sigma x_1 y = a\Sigma x_1 + b\Sigma x_1^2 + c\Sigma x_1 x_2$ — (3)

$\Sigma x_2 y = a\Sigma x_2 + b\Sigma x_1 x_2 + c\Sigma x_2^2$ — (4)

Example

	x	y	xy	x^2
	4	5	20	16
	6	4	24	36
	7	5	35	49
	8	6	48	64
$n = 4$, Σxy	25	20	127	165

Then Sub to

$y = a + bx$

$\Sigma y = an + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$\Sigma y = an + b\Sigma x$

$20 = 4a + 25b$ — (ii)

$\Sigma xy = a\Sigma x + b\Sigma x^2$

ie $127 = 23a + 165b$ — (iii)

$\therefore y = a + bx$ — (1)

then sub a and b, dats d eqn.

ANALYSIS OF VARIANCE ONE WAY TEST.

let consider sample datas as presented below.

female = 4, 3, 4, 3, 5, 7, 6, 9, 2, 5, 3.

male = 2, 4, 4, 3, 4, 8, 7, 8, 9, 9, 2.

t-test $(t) = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

μ = mean of population.

s = standard deviatn of sample.

n = no of sample.

To find mean $\bar{x}$

$df = n - 1 = 11 - 1 = 10.$

note μ = must be given.

If the male and female is rejected u sum d two i.e f test

$f_R = \dfrac{s_1^2}{s^2}$

STATISTICS

(Engr. Dr. Mrs Folashade Fayose.)

STATISTICAL INFERENCES.

2 TYPES OF STATISTICS

=> Descriptive statistics.

=> Inferential statistics.

example of population

A d males produce in a batch in a day

3 main underlie ideas of a sample

=> A sample is likely to be a good representation of d population.

=>

Types of Sampling

=> Random Sampling => no bias, no preference. It is used to eliminate bias.

=> Stratified Sampling => The population is divided into two called Strata. i.e. dividing d populatn into groups and pick some same no of samples from each group.

=> Cluster sampling => When samples are taken from an already existed groups.

=> Systematic sampling => Dividing d group according to the range.

Ways of taking sample

=> A sample is a good representation of d population

=>

9-6-2021

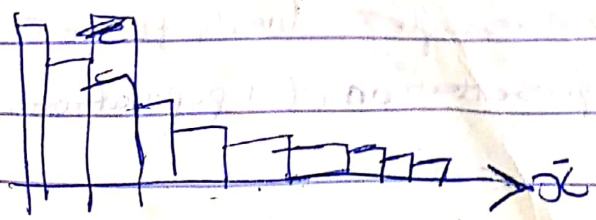
SAMPLING DISTRIBUTION.

Parameter is a population data. It is a quantity dat measure / describe a populatn e.g population mean,

i.e Parameter	Statistics
=> population mean μ	Sample mean $\bar{x}$
=> population Variance σ	" Variance
=> " Standard deviation	" standard deviation.
=> " Proportion	" proportion

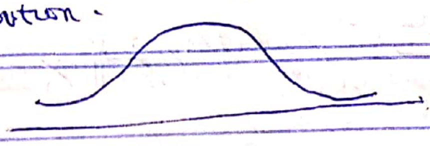
If u find d means of each samples. And took 3 out of 10 population, i.e $\binom{10}{3} = {}^{10}C_3 = 120$

If u draw their histogram



This is 10 histograms the means.

at the long run, If u do it to 120 histogram, it give a normal distribution.



bcos of disabilities in Sample mean, we have standard error.

Sample error

Mean $\bar{x} = \Sigma\left(\frac{x}{n}\right) \longrightarrow$ mean $= \Sigma \frac{x}{n}$

Variance $= \frac{\Sigma(x-\bar{x})^2}{n} \longrightarrow \frac{\sigma}{\sqrt{n}}$ => standard error

Standard deviation $= \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \longrightarrow \sqrt{\frac{\sigma}{\sqrt{n}}}$

Population $= \sqrt{\frac{\lambda(1-\lambda)}{n}}$

Properties of Sampling Distribution

For any no of samples n d centre of mean distributn is also coincid with the mean of population in ed sample and d spread of mean distribu decreases as n increases.

Larger samples lower distributn

let $\bar{x}$ = means of d observation in a random sample of size n from a population having mean μ and standard deviation σ, the mean value of the $\bar{x}$ distribu of μ and d standard deviation of $\bar{x}$ distribution by σ_x. The following rules hold.

Sample dis. for Sample mean

Rule 1: $\mu_1 = \mu$

Rule 2: $\sigma_x = \frac{\sigma}{\sqrt{n}}$

σ_x = Standard deviation of the sample

Rule 3: When populatn distribution is normal, the sampling distribution of $\bar{x}$ is also normal for any sample size

Rule 4: Central limit theory: When n is sufficiently large, the sampling distribution is well approximated by a normal curve even when the population distribution is not itself normal.

Sample distribution for sample proportion

When a random ~~sample~~

π = proportion,

parameter — Statistic.

π = Population — p–sample production

$p = \frac{\text{no of successes in the sample}}{\text{no of sample } (n)}$

Properties of sample proportion

① => The sample distribution of p is centred at π i.e $\mu_p = \pi$. ∴ p is an unbiased statistic for estimating π.

② => The standard deviation of σ_p

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

③ => As long as n is large, and $(n\pi \geq 10)$ and $n(1-\pi) \geq 10$

[Reading]

CORRELATION AND REGRESSION

~~Relations~~ STATISTICAL RELATIONSHIP is a relationship btw two series. e.g height and weight of a group of people, weight increases with increase in height. Such relationship is called STATISTICAL RELATIONSHIP.

TYPES OF DISTRIBUTION

$\bar{x} = \frac{\sum n}{n}$

=> UNIVARIATE DISTRIBUTION => This is a distribution in which there exist only one variable. e.g height of students in a class.

=> BIVARIATE distribution => This involves two variables e.g height and weight of a class.

COVARIANCE

If $(x_1,y_1), (x_2,y_2), (x_3,y_3) \cdots (x_n,y_n)$

The covariance btw X and Y is;

$$COV(X,Y) = \frac{(x_1-\bar{x}_1)(y_1-\bar{y}_1)+(x_2-\bar{x})(y_2-\bar{y})}{n}.$$

i.e, $COV(X,Y) = \frac{\sum(x_n-\bar{x})(y_n-\bar{y})}{n}.$

EXAMPLE

=> Cal the covariance of d following pairs of observations of two variates.

$(1,4), (2,2), (3,4), (4,8), (5,9), (6,12)$ $n = 6$

Sol

$\sum x_i = 1+2+3+4+5+6 = 21$

$\sum y_i = 4+2+4+8+9+12 = 39$

$\sum x_iy_i = (1\times4)+(2\times2)+(3\times4)+(4\times8)+(5\times9)+(6\times12)$

$= 4+4+12+32+45+72$

$= 169$

$n = 6$ i.e 6 brackets.

$$\boxed{COV\ x,y = \frac{\sum x_iy_i}{n} - \frac{\sum x_i}{n} \cdot \frac{\sum y_i}{n}}$$

$\therefore COV\ x,y = \frac{169}{6} - \frac{21}{6} \times \frac{39}{6}$

$= \frac{169}{6} - \frac{91}{4} = \frac{65}{12}$

$\therefore COV\ x,y = \frac{65}{12}$

CORRELATION

CORRELATION is a situation whereby wen variable x and Y are so related that increase in one is as a result of increase or decrease in the other. e.g The yield of crop varies with the amount of rainfall. i.e Variation (increase or decrease) in rainfall determines the quantity of crop that will yield.

TYPES

1 ⇒ Positive Correlation ⇒ A Correllation is said to be +ve if the increase in one variable (X) results in a corresponding increase in value of other variable (Y) on an average.

OR

If the decrease in d value of one variable X results in corresponding decrease in value of oda variable Y on an average.

2 ⇒ Negative Correlation ⇒ This is wen the increase in value of one variable X results in a corresponding decrease in the values of oda variable Y.

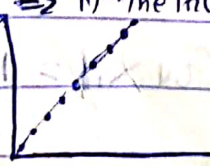
OR

If the decrease in d value of one variable X results in a corresponding increase in the values of oda variable Y.

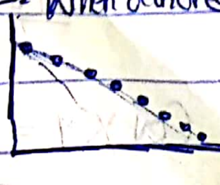
3 ⇒ LINEAR CORRELATION ⇒ When all plotted points lie approximately on a straight line.

4 ⇒ PERFECT CORRELATION ⇒

A ⇒ Positive perfect correlation ⇒ If the increase in one variable X is proportional to increase in d oda variable Y. The graph will be straight line.



B ⇒ Negative perfect correlation ⇒ When d increase in one variable is proportional to d decrease in the oda variable. The graph is also a straight line.



KARL PEARSON'S COEFFICIENT OF CORRELATION

$$r = \frac{\sum XY}{\sqrt{(\sum X^2)(\sum Y^2)}} = \frac{p}{\sigma_x \sigma_y}$$

$$= \frac{\text{Covariance}(x, y)}{\sqrt{\text{Variance}\,x}\ \sqrt{\text{Variance}\,y}}$$

where $X = x - \bar{x}$, $Y = y - \bar{y}$

$$\text{Covariance}, (p) = \frac{\sum XY}{n}$$

and σ_x, σ_y, being d standard deviation of these series

EXAMPLES.

⇒ Cal. the coefficient of correlation btw x and y series from the following data;

$\sum(x-\bar{x})^2 = 136$ $\quad \sum(y-\bar{y})^2 = 138$

$\sum((x-\bar{x})(y-\bar{y}))^2 = 122.$

Sol

Since $p = \frac{(x-\bar{x})(y-\bar{y})}{n} = \frac{\sum XY}{n}$

Also note; $(x-\bar{x})$ = mean deviation

$(x-\bar{x})^2$ = Variance

$\sqrt{V} = \sqrt{(x-\bar{x})^2}$ = Standard deviation

formular to use;

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

$$\text{i.e } r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(\text{Standard deviation of } x)(\text{Standard deviation of } y)}$$

$$r = \frac{122}{(\sqrt{136})(\sqrt{138})}$$

$$r = \frac{122}{11.66 \times 11.75} = \frac{122}{[illegible]} = 0.89$$

REGRESSION
(Coefficient of correlation of grouped data)

x	y	$\bar{x}=65$, $X = x-\bar{x}$	$\bar{y}=66$, $Y = y-\bar{y}$	$X^2=(x-\bar{x})^2$	$Y^2=(y-\bar{y})^2$	$XY=(x-\bar{x})(y-\bar{y})$
78	84	13	18	169	324	234
36	51	-29	-15	841	225	435
98	91	33	25	1089	625	825
25	60	-40	-6	1600	36	240
75	68	10	2	100	4	20
82	62	17	-4	289	16	-68
90	86	25	20	625	400	500
62	58	-3	-8	9	64	24
65	53	0	-13	0	169	0
39	47	-26	-19	676	361	494
$\Sigma x=650$	$\Sigma y=660$	$\Sigma X=0$	$\Sigma Y=0$	$\Sigma X^2=5398$	$\Sigma Y^2=2224$	$\Sigma XY=2704$

$$r = \frac{122}{11.66 \times 11.25} = \frac{122}{137.005} = 0.89$$

$\therefore r = 0.89$

=> Tenstud students got the following % of marks in Economics and Statistics.

Roll no	1	2	3	4	5	6	7	8	9	10
Marks in Economics	78	36	98	25	75	82	90	62	65	39
Marks in Statistics	84	51	91	60	68	62	86	58	53	47

Cal. the coefficient of Correlation.

Sol.

Let x = marks in Economics

y = marks in Statistics

$$\bar{x} = \frac{78+36+98+25+75+82+90+62+65+39}{10}$$

$$\bar{x} = \frac{650}{10} = 65$$

$$\bar{y} = \frac{84+51+91+60+68+62+86+58+53+47}{10}$$

$$\bar{y} = \frac{660}{10} = 66.$$

$\therefore \Sigma X^2 = 5398$, $\Sigma Y^2 = 2224$

$\Sigma XY = 2704.$

$$r = \frac{\Sigma XY}{\sqrt{(\Sigma X^2)(\Sigma Y^2)}} = \frac{2704}{(\sqrt{5398})(\sqrt{2224})}$$

$$r = \frac{2704}{73.4 \times 47.1} = \frac{2704}{3457} = 0.78$$

$\therefore r = 0.78$

REGRESSION

LINE OF REGRESSION

A line of regression is the straight line which gives the best fit in the least square sense to the given frequency. The the Curve is a straight line.

EQNS TO THE LINES OF REGRESSION

NOTE.

Regression of y on x is;

$$\boxed{y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})}$$

Regression of x on y is;

$$\boxed{x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})}$$

The slope of the line of regression;

$$b = r\frac{\sigma_y}{\sigma_x}.$$

$$\therefore b_{yx} = r\frac{\sigma_y}{\sigma_x}, \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

i.e b_{yx} and b_{xy} are called coefficient of regression.

$$b_{yx} \cdot b_{xy} = \left(r\frac{\sigma_y}{\sigma_x}\right)\left(r\frac{\sigma_x}{\sigma_y}\right) = r^2$$

EXAMPLES.

=> Find d correlation coefficient btw x & y, when the lines of regression are;

$2x - 9y + 6 = 0$, $x - 2y + 1 = 0$

Sol

Let line of regression of x on y be $2x - 9y + 6 = 0$

" " " " y " x " $x - 2y + 1 = 0$

$2x - 9y + 6 = 0 \Rightarrow x = \frac{9}{2}y - 3$

i.e $b_{xy} = 9/2.$

$x - 2y + 1 = 0 \Rightarrow y = \frac{1}{2}x + \frac{1}{2}$

Since $r^2 = b_{xy} \cdot b_{yx}$

$r = \sqrt{\frac{9}{2} \times \frac{1}{2}} = \sqrt{\frac{9}{4}} = \frac{3}{2} > 1$

Since $\frac{3}{2} > 1$, this is incorrect bcoz the answer must be lesser than 1.
~~Switch Changing~~ ∴ The chosen regression line is incorrect

~~Switch~~ Interchanging d two eqns;

Let the line of regression of y on x be $2x - 9y + 6 = 0$

$y = \frac{2x}{9} + \frac{6}{9}$

i.e $b_{yx} = \frac{2}{9}$

let the line of regression of x on y be $x - 2y + 1 = 0$

$x = 2y - 1$

i.e $b_{xy} = 2$.

$r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{2 \times \frac{2}{9}} = \frac{2}{3}$

∴ The Correlation Coefficient btw x and y is $2/3$.

⇒ The regression eqns calculated from a given set of observations for two random variables are;
$x = -0.4y + 6.4$, $y = -0.6x + 4.6$.
Cal (a) Mean of x's $(\bar{x})$.
(b) mean of y's $(\bar{y})$
(c) The correlation Coefficient btw x and y (r)

Sol.

$x = -0.4y + 6.4$ ——— (1)
$y = -0.6x + 4.6$ ——— (2)

Since eqn (1) and (2) passes through the point $(\bar{x}, \bar{y})$ (a) and (b)

∴ eqn becomes

$\bar{x} = -0.4\bar{y} + 6.4$ ——— (3)
$\bar{y} = -0.6\bar{x} + 4.6$ ——— (4)

Solving eqn 3 and 4 simultaneously

$\bar{x} = 6, \quad \bar{y} = 1$.

(c)

from eqn (1) The coefficient of x on y $= b_{xy} = r\frac{\sigma_x}{\sigma_y} = -0.4$ ——— (5)

from eqn (2)

The Coefficient of y on x $b_{yx} = r\frac{\sigma_y}{\sigma_x} = -0.6$

The correlation coefficient r;

$r^2 = b_{xy} \cdot b_{yx}$

$r^2 = r\frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x}$

$r^2 = (-0.4)(-0.6) = 0.24$.

$r = \sqrt{0.24} = \cancel{\pm\sqrt{0.49}}$

$r = \pm 0.49$.

from eqn (5) and (6), σ_x and σ_y are (always) +ve, so r is negative.

i.e $r = -0.49$.

16-6-2021 CLASS

CONFIDENCE INTERVAL

⇒ Introduction to Confidence interval for one mean (σ known) (youtube)
Sigma i.e standard deviate.

Channel ⇒ jb statistics.

⇒ Confidence Interval and marging error Ap Statistics.

Channel; Khan Academy.

⇒ Solved problems and Submit.

To be submitted on wed 23, 2021

Scanned with CamScanner

Reading

SAMPLING OF VARIABLES

POPULATION => This is a group of individuals under study. It is also known as UNIVERSE.

SAMPLING.

A part selected from the population is called SAMPLE. The process of selection of a Sample is SAMPLING.

RANDOM SAMPLE => This is one in which each member of population has an equal chance of being included in it.

$^{N}C_{n}$ means there are diff samples of size n that can be picked up from a population of size N

PARAMETERS AND STATISTICS

The statistical Constants of the population such as mean (μ), Standard deviation (σ) are called Parameters. Parameters are denoted by Greek words

The mean ($\bar{x}$), standard deviation (s) of a sample are known as STATISTICS. Statistics are denoted by Roman Letters.

Symbols for population and samples

Population

Samples

Population	Sample
Parameters	Statistics
=> Population size = N	Sample size = n
=> Population mean = μ	Sample mean = $\bar{x}$
=> Population standard deviation = σ	sample standard deviation = s
=> Population proportion = p	sample proportion = $\bar{p}$

AIMS OF SAMPLES

STATISTICAL INFERENCES => The logic of the sample theory is the Logic of Induction. In induction, we pass from a particular (sample) to general (POPULATION). This types of generalisation here is known as STATISTICAL INFERENCE.

TYPES OF SAMPLING

=> Purposive sampling

=> Random sampling

=> Stratified Sampling

=> Systematic sampling.

SAMPLING DISTRIBUTION

From a population N, a sample are drawn of equal size n. Find out the mean of each sample. The means of samples are not equal. The means with their respective frequencies are grouped. The frequency distribution so formed is known as SAMPLING DISTRIBUTION of the mean. Similarly, sampling distribution of standard deviation we can have.

STANDARD ERROR (S.E)

This is the standard deviation of the sampling distribution. For assessing a diff btw expected value and observed value, standard error is used. Reciprocal of standard error is PRECISION.

SAMPLING DISTRIBUTION OF MEANS FROM INFINITE POPULATION.

Let x = random variable denoting the measurement of the xteristics.

let μ = population mean.

Variance of x, $Var(x) = \sigma^2$

Expected value of x, $E(x) = \mu$

Then the Sample mean ($\bar{x}$) is d sum of n random Variables $x_1, x_2, ---, x_n$ each being divided by n.

i.e $x_1, x_2, ---, x_n$ are independent random Variable fromd infinitely large population.

NOW;

Expected of x_1 ($E(x_1)$) $= \mu$, Variance of x_1, $Var(x_1) = \sigma^2$

Expected of x_2, $E(x_2) = \mu$, Variance of x_2, $Var(x_2) = \sigma^2$

$\therefore$

Expected Sample mean, $E(\bar{x}) = E\left[\frac{x_1 + x_2 + --- x_n}{n}\right]$.

$$E(\bar{x}) = \frac{1}{n}Ex_1 + \frac{1}{n}E(x_2) + ---- + \frac{1}{n}E(x_n)$$

Since $E(x_n) = \mu$

$$\therefore E(\bar{x}) = \frac{1}{n}\mu + \frac{1}{n}\mu + --- + \frac{1}{n}\mu$$

$$= \mu$$

$$\text{i.e } E(\bar{x}) = \mu$$

$\therefore$ The expected value of sample mean $E(\bar{x})$ is desame as population mean μ.

For Variance.

Variance of Sample mean, $Var(\bar{x}) = Var\left[\frac{x_1 + x_2 + -- + x_n}{n}\right]$

$$Var(\bar{x}) = Var\left(\frac{x_1}{n}\right) + Var\left(\frac{x_2}{n}\right) + ---- + Var\left(\frac{x_n}{n}\right)$$

$$Var(\bar{x}) = \frac{1}{n}Var(x_1) + \frac{1}{n}Var(x_2) + -- + \frac{1}{n}Var(x_n).$$

Since $Var(x_n) = \sigma^2$.

$$\therefore Var(\bar{x}) = \frac{1}{n}\sigma^2 + \frac{1}{n}\sigma^2 + -- + \frac{1}{n}\sigma^2$$

$$\therefore Var(\bar{x}) = \frac{\sigma^2}{n}$$

$\therefore$ The Variance of Sample mean ($Var(\bar{x})$) is the Variance of population mean (σ^2) divided by the Sample size (n)

STANDARD ERROR OF THE MEAN

This is also known as STANDARD DEVIATION ($\frac{\sigma}{\sqrt{n}}$) i.e $\sqrt{\text{Variance}}$

$$\therefore \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

SAMPLING WITH REPLACEMENT

When sampling is done with replacement, so dat the population is back to dsame form b4 the next Sample member is picked up, we have;

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}$$

Standard deviation, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

SAMPLING WITHOUT REPLACEMENT FROM FINITE POPULATION.

$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Also $\sigma_{\bar{x}} = \sqrt{Var(\bar{x})}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

If $\frac{n}{N}$ is very small,

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ approximately.

SAMPLING FROM NORMAL POPULATION

If $x \sim N(\mu, \sigma^2)$ then it follows that

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}).$$

Ass

⇒ Introduction to Confidence interval for one mean (σ known) youtube Channel ⇒ jbstatistics.

Sol.

Note $\bar{x}$ is a point estimator of μ
sample mean — population mean

The confidence interval ;

$\bar{x} \pm$ margin of Error.

for confidence interval method to be reasonable we require ;

⇒ A simple random sample from the population of interest (It is very important regardless of sample size).

of normal distributed population (This is not important for large sample size). Also, population standard deviation σ must be known.

Confidence Interval formula; .

Confidence Interval for $\mu = A(1-\alpha)100\%$

$$\boxed{\bar{x} \pm \underbrace{Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}_{\text{margin of error of confidence interval}}}$$

$$\sigma_{\bar{x}} = \frac{\text{Standard deviation of sampling distribution}}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$(1-\alpha)100\%$ is a confidence interval level which is commonly 95%.

Note α is gotten from the confidence interval level.

e.g if confidence interval level $= 95\%$

i.e $95\% = 0.95$.

$\alpha = \frac{100\% - 95\%}{2} = \frac{5\%}{2} = 2.5\%$

$\therefore \alpha = 0.025$.

To Z table, $Z(\frac{\alpha}{2}) = 1.96$

i.e $0.95 + 0.025 = 0.975$.

On z-table, 0.9750 is under 1.9 under 0.06 i.e $1.9 + 0.06 = 1.96$,

Example

The ratio of the length of the second digit (index finger) to that of the fourth digit (ring finger), known as the 2D:4D ratio is often studied by researchers.

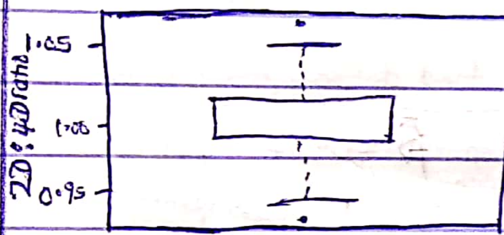
Sol

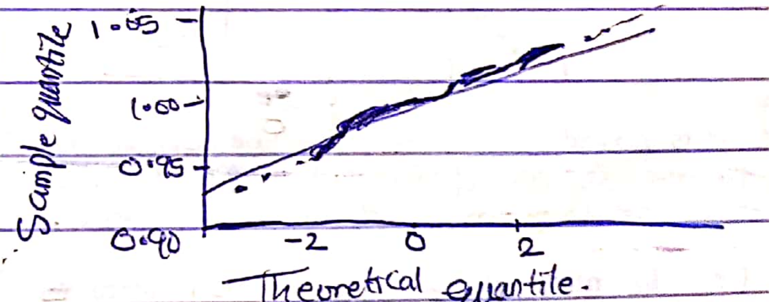
A study investigated the 2D:4D ratio in women of European descent at a large college. A sample of 135 women yielded an average 2D:4D ratio of 0.988.
i.e (sample mean) $(\bar{x}) = 0.988$.

What is the 95% confidence interval for the population mean?

Suppose it is known that $\sigma = 0.028$.

The box plot and quantile quantile plot





$\bar{x} = 0.988 \quad \sigma = 0.028, \quad n = 135$

What is a 95% confidence interval for the population mean?

$\bar{x} \pm Z(\frac{\alpha}{2}) \times \frac{\sigma}{\sqrt{n}}$ Note $Z(\alpha/2) = 1.96$

Substituting d value

$0.988 \pm 1.96 \times \frac{0.028}{\sqrt{135}}$

$= 0.988 \pm 0.0047$

$= (0.988 + 0.0047, 0.988 - 0.0047)$

$= (0.983, 0.993)$

∴ We can be 95% Confident that the population mean lies in this interval (0.983, 0.993)
i.e We can be 95% Confident dat the true mean 2D:4D ratio for women of European descent at this college lies btw 0.983 and 0.993.

Q2

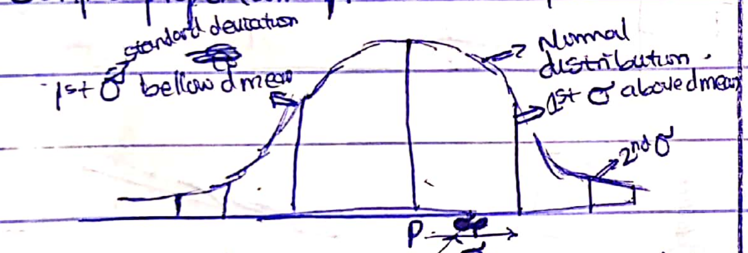
Confidence intervals and margin of error AP statistics Khan Academy.
Channen => Khan academy.

SOL

In an election betw A and B Candidate, in a population size of n=100, people who support Candidate A is 54% and people who support " B is 58% • Sampling distributions of sample proportions (for n=100).

To find d sampling distribution.

Sample proportion $\hat{p} = 0.54$, $n = 100$



It is normal distribution becx we assume dat the true proportion (proportion dat support A) isn't too close to zero or not too close to one.

i.e The mean of this sampling distribution is ~~going~~ to be d actual population proportion (p)
i.e $\bar{x} = P$
standard deviation for
$\sigma_{\hat{p}}$ = sample proportion for this sampling distribution
$$\sigma_p = \sqrt{\frac{P(1-P)}{n}}$$
For $n=100$, $\hat{p}$ => sample proportion $= 0.54$

i.e to find the position where 0.54 lies on the Normal distribution above.
Note Actual population P is unknown.

QST What is d probability dat $\hat{p} = 0.54$ is within $2\sigma_{\hat{p}}$ of P?

It will be approximately 95%

i.e ther is a 95% prob. dat population proportion P is [illegible] of $\hat{p} = 0.54$.

Since $P = ?$ = true proportion
Using anoda formular for estimating for population proportion.
Since Sample proportion $\hat{p} = 0.54$

Using Standard Error (SE) formular
$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$SE_{\hat{p}} = \sqrt{\frac{0.54(1-0.54)}{100}} = 0.05.$$
∴ with 95% Confidence, btw [(0.54 − (2×0.05) i.e 0.44) and (0.54 + (2×0.05) = 0.64)]
0.44 and 0.64 voters support A.

The Confidence Interval = 0.44 to 0.64.

The related Idea to the confidence interval is d notion of margin of error.

Margin of error = $2SE_{\hat{p}}$
i.e Margin of Error = 2×0.05 = 0.1
Margin of Error below sample proportion is 0.44 − 0.1 = 0.34
Margin of Error above Sample proportion is 0.64 + 0.1 = 0.74

~~To tighten up the intervals on average~~
To tighten up the Intervals on average?
U lower the margin of error by increasing the no of size of population (sample size).

23-6-2021 CLASS

NOTE. Sampled distribution problem

⇒ The average male drinks 2L of water when active outdoors with (with a standard deviation of 0.7L) you are planning a full day nature trip for 50 men and will bring 110L of water. What is the probability that you will run out.

SOL

Population mean $\mu = 2L$

$\sigma = 0.7L$ $n = 50$

$P > 110L$

Sample mean $\bar{x} = \frac{110}{50} = 2.2L$

(Variance) $= \frac{\sigma^2}{n} = \frac{(0.7)^2}{50}$

$\sigma_{\bar{x}}$ (S.D = $\sqrt{\text{Variance}}$ $= \frac{\sigma}{\sqrt{n}} = \frac{0.7}{\sqrt{50}}$

$\sigma_{\bar{x}} = 0.099$.

note if $n > 30$, it means the diagram is tending to normal distribution

To find the no of standard deviation at which it's far from the centre

$\frac{\bar{x} - \mu}{\sigma} = \frac{2.2-2}{0.099} = 2.02$

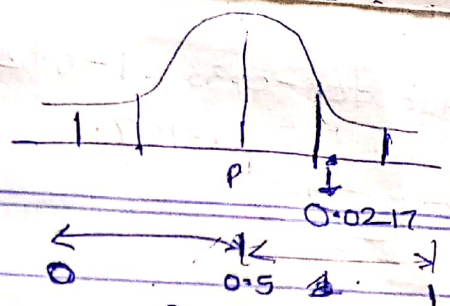
i.e after two standard deviation.

from Z-table, $2.02 = 0.9783$.

Since the whole diagram is 1.

∴ to find the position is $1 - 0.9783$

$= 0.0217$.



Prob dat $P > 110L = 2\%$.

⇒ Suppose samples of 100 items were drawn from a population with a mean of 200 and a standard deviation 30. What proportion of sample will have mean,

(a) greater dan 202

(b) less dan 199.

(c) between 198.5 and 203

SOL.

$n = 100$, $\mu = 200$, $\sigma = 30$

(i) $Z = \frac{x - \mu}{\sigma_{\bar{x}}}$

Since $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$Z = \frac{202 - 200}{30/\sqrt{100}} = \frac{2}{3} = 0.67$

under Z table, $0.67 = 0.7486$

i.e 0.6 under 0.07

position $= 1 - 0.7486 = 0.2514$

Prob. $= 25\%$.

(ii) $Z = \frac{199 - 200}{\frac{\sigma}{\sqrt{n}}} = \frac{-1}{3} = 0.33$

~~because d ans is negative,~~

under -ve table, $-0.33 = 0.3707$

but if +ve table was given,

Check 0.3 under 0.03 in the table

$= 0.6293$

ACtual z value for $-0.33 = 1 - 0.6293$

$= 0.3707.$

Prob $- = 37\%$

(ii)

$$Z = \frac{195.5-200}{3} = \frac{-1.5}{3} = -0.50$$

under negative table, 0.5 and 0.00

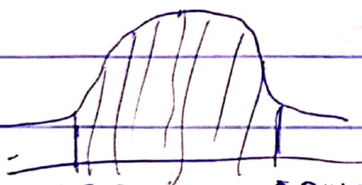
$Z_{-0.5} = 0.3085$

Also for 203.

$$Z = \frac{203-200}{3} = 1.00$$

under +ve table, 1.0 under 0.00

$Z_{1.0} = 0.8413.$



0.3085 0.8413

Since 0 0.5 1.

To get d actual point,

$0.8413 - 0.3085 = 0.5328.$

$= 53\%$

NOTE there are two types of z tables.

1. The one of 1 like above to identify first row and first column under zero will be ~~0.000~~ 0.5
2. The one with 0.5 to identify, first row first column, u will see 0.0000

Confidence interval

>> The ratio of the length of the second digit (index finger) to that of the fourth digit (ring finger) known as the 2D:4D is often studied by researchers. A study investigated the 2D:4D at a large College.

A sample of 135 women yielded an average of 2D:4D ratio 0.988. What is d Confidence interval for the population mean? Suppose it is known that $\sigma = 0.028$.

Sol.

$\bar{x} = 0.988$, $\sigma = 0.028$, $n = 135$

Since $(1-\alpha) = 95\%$

i.e $\alpha = 5\% = 0.05.$

Recall that

$$\bar{x} \pm Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$$= 0.988 \pm \frac{Z(0.028)}{2} \times \frac{0.028}{\sqrt{135}}$$

Assumptions in Confidence intervals are 95%, 90% and 99.9%

Ass

A sample poll of 100 voters chosen at random from all voters in a given district indicated dat 55% of dem were in favour of a particular Candidate, find (a) 95% (b) 99% (c) 99.73% Confidence limits for the proportion of all the voters in favour of this Candidates.

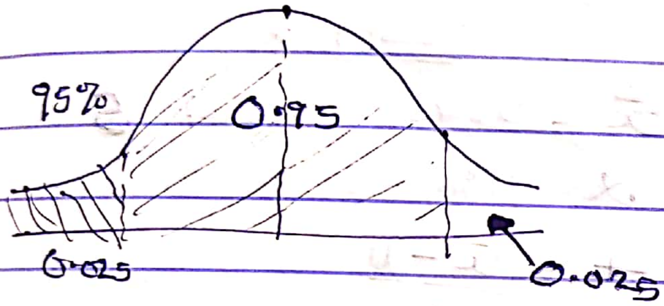
30-6-2021 CLASS

CONFIDENCE INTERVAL.

Let π be the population proportion of the defectives nails -

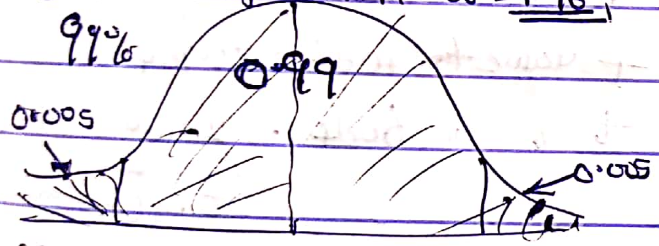
p = sample proportion = $\frac{\text{no of defective in the sample}}{n}$

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$



0.95+0.025   0.975 = 1.96.

to check where 0.975 lies on z table that is 1.9 under 0.06 which gives $1.9+0.06 = \underline{1.96}$,



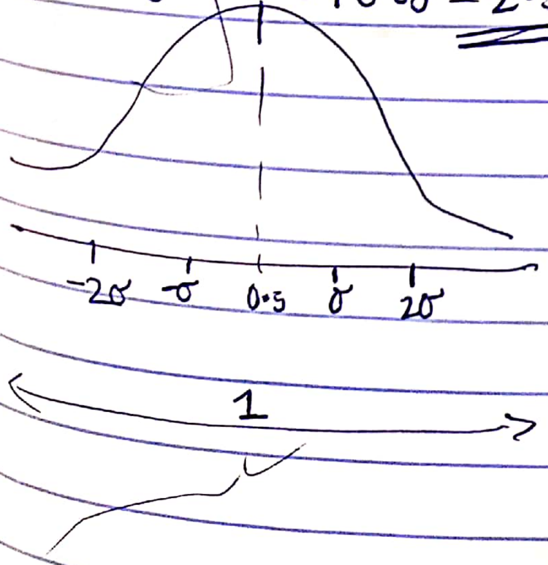
0.99+0.005 = 0.995

Z value of 0.995 = 2.58.

~~2.58H = 2.58 can be found~~

0.9951 = 0.995 Can be found on z table

2.5 under 0.08 = 2.5+0.08 = 2.58,



EXAMPLE

The variance of bolt heigh σ^2 produce in a industry is 100 while the mean height of 16 poles randomly selected from the industry is 185. Find 95% confidence intervals for the population mean and make confidence statement.

Sol,

$\sigma^2 = 100$ i.e $\sigma = 10$.

$\bar{x} = 185$, $n = 16$

$\bar{x} \pm Z_c \frac{\sigma}{\sqrt{n}}$

$185 \pm 1.96 \left(\frac{10}{\sqrt{16}}\right)$

$= 185 \pm 4.9$

Upper limit = 185+4.9 = 189.9

lower limit = 185-4.9 = 180.1

Confidence Statement => At 95% confidence level (probability) the interval 180.10 and 189.9 contain the mean height of the bolt population.

EXAMPLE

To illustrate sample estimate

$p = \frac{\text{number of success in the sample}}{n}$

$p = \frac{537}{1013} = 0.530$

Now we want to generate confidence interval for the point estimate.

Firstly assume a confidence level say 95%

$\pi = p \pm Z^* \sqrt{\frac{p(1-p)}{n}}$

$(0.499, 0.561)$ $np \geq 10$

Scanned with CamScanner

CONFIDENCE INTERVAL FOR μ

When σ is unknown and usefull when the sample size is ~~large~~.

$\bar{x} = \mu$, $\sigma_x = \frac{\sigma}{\sqrt{n}}$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

σ is unknown and the sample size is small, use t-table.

$$-1.96 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

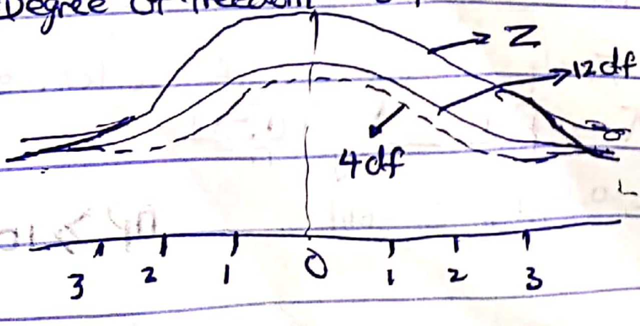
$$\bar{x} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{x} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

T DISTRIBUTION.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

INTRODUCTION TO T-DISTRIBUTION.

- The spread of t-distribution is more than that of normal distribution.
- As the no of samples increases, your t-distribution tends towards normal distribution.
- t-distribution are meant for small sample size only
- Degree of freedom $df = n - 1$



Here

EXAMPLE.

=> Assuming df @ 99.5%
t-value → 4.32

=> from a sample of 49 students the mean mathematics score in a school was found to be 55% with a SD of 15. find the 99% Confidence interval for the real mean score.

Sol

$\bar{x} = 55\%$, $n = 49$ $s = 15$

$\alpha = 99\%$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

If ur table is not having the d.f

t value for 40df = 2.704

t " " 60df = 2.660

5.364

Through Interpolation and extrapolation

$\frac{5.364}{2} = 2.682$

$\simeq 2.7$

1 pdf. First pdf. Reading

PG 7

Inferential Statistics helps us to draw conclusion beyond the data we have from the population.

Statistical Inferences is the process of drawing conclusion about population parameters based on a sample taken from the population.

SAMPLE => This is the data collected from population.

TYPES OF STATISTICAL ANALYSIS

Descriptive Statistics.
Inferential statistics.

STATISTICAL INFERENCE 2nd pdf

Sampling distribution of $\bar{x}$ provides information about the long run behaviour of $\bar{x}$ when sample after sample is selected.

Sample proportion is the fraction of individual or objects in a sample that have some characteristics of interest OR the population of individual in a sample that possesses a particular property.

SAMPLING DISTRIBUTION of MEAN.

Rule 1 => $\mu_{\bar{x}} = \mu$

Rule 2 => $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

For detail PG 2

Go to class on 9-6-2021

UNBIASED ESTIMATE => This is a statistics whose mean values equal to the value of the population xteristics being estimated.

The Greek letter π denotes the population of success in a population.

$p = \frac{\text{no of success in the sample}}{n}$

Statistics Population

=> $\mu_p = \pi$

$\therefore$ p is an unbiased statistic for estimating π.

=> Standard deviation of p is

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ As long as n is large, $n\pi \geq 10$ and $n(1-\pi) \geq 10$. Also sample size is less than 10% of the population size.

CONFIDENCE INTERVAL 3rd pdf

It is explained betw lower and upper endpoint of the interval.

$p = \frac{\text{no of the sample that possess the property of interest}}{n}$

when n is large,

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

CONFIDENCE INTERVAL FOR POPULATION MEAN (μ)

~~Prop of sampling distribution~~

=> $\mu_{\bar{x}} = \mu$

=> $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

=> As long as n is large ($n \geq 30$), the sampling distribution of $\bar{x}$ is approximately normal even when d population distribution itself is not normal.

Formular for Confidence interval for a population mean μ

=> $\bar{x}$ is sample mean from random sample
=> The sample size n is large ($n \geq 30$)
=> The sample standard deviation is
=> Population standard deviation is known i.e when σ is known

$\therefore$ $\bar{x} \pm (Z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$

CONDITION NECCESSARY TO CONSIDER B4 USING CONFERENCE INTERVAL (CI)

$$\bar{x} \pm Z_{(\text{critical value})}\left(\frac{\sigma}{\sqrt{n}}\right).$$

1) σ must be known (population standard deviation must be known)

2) Sample size n must be large $n \geq 30$

Then use this formular for C.I

$$\bar{x} \pm \left(Z_{\text{critical value}}\right) \times \left(\frac{\sigma}{\sqrt{n}}\right).$$

NOTE the following;

Recall dat for a normal population distribution the sampling distribution ~~is normal~~ of $\bar{x}$ is normal even for small sample sizes.

$\therefore$ If n is small but population distribution is normal, d same confidence interval formula just introduced can still be used

i.e if n is small $(n < 30)$ and it is reasonable to believe dat d distribution of values in the population is normal, a confidence interval for μ (when σ is known) is i.e σ must also be known.

$$\bar{x} \pm \left(Z_{\text{critical value}}\right)\left(\frac{\sigma}{\sqrt{n}}\right).$$

CONFIDENCE INTERVAL for μ WHEN σ IS UNKNOWN

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

The above formular is used if ~~σ is~~ σ is unknown and n is large $(n \geq 30)$

T - DISTRIBUTION

This t-distribution is used only when σ is unknown and n is small $(n < 3$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

POINT ESTIMATION

CONFIDENCE INTERVAL FOR π

population proportion.

$$p = \frac{\text{no of successes in the sample}}{n}$$

PG 8

$\Rightarrow \mu_p = \pi$ — This is used when n is large

$\Rightarrow \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

$\Rightarrow$ As long as n $(n\pi \geq 10)$ and $n(1-\pi) \geq 10$ and sample size is less dan 10% of the population size

$$p \pm \left(Z_{\text{critical value}}\right) \times \left(\sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

dsame

If π is unknown, as long as sample size is large i.e n is large we can use $\sqrt{\frac{p(1-p)}{n}}$ in place of $\sqrt{\frac{\pi(1-\pi)}{n}}$

dsame

i.e $p \pm \left(Z_{\text{critical value}}\right)\left(\sqrt{\frac{p(1-p)}{n}}\right).$

i.e the following are to be note;

$\Rightarrow np \geq 10$ and $n(1-p) \geq 10$

$\Rightarrow$ Sample size n is less dan 10% of the population size if sampling is without replacement

$\Rightarrow$ The sample can be regarded as a random sample from the population of interest.

$\therefore$ If $np \geq 10$ and $n(1-p) \geq 10$

Then use $p \pm Z_{\text{critical value}}\left(\sqrt{\frac{p(1-p)}{n}}\right)$

Continue

Books minitab

+ (i) $np \geq 5$ (ii) $n(1-p) \geq 5$.

. (i) $np \geq 10$ (ii) $n(1-p) \geq 10$ 1990–2000s ~~Current~~

. (i) $np \geq 15$ (ii) $n(1-p) \geq 15$ Current

If the above Condition is met, the sampling distribution of a sample proportion is (i) approximately normal (ii) mean $\mu = p$.

(iii) $\sigma = \sqrt{\frac{p(1-p)}{n}}$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

where $\hat{p} = \frac{x}{n}$

p = success

NOTE GENERALLY.

SAMPLING DISTRIBUTION (SD)

S.D for mean $= \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

Note $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

∴ The above formula is used the mention Confidence Interval, Confidence limit, or probability or if they mention mean) in the qn.

Note for Confidence interval = (lower, upper)

" Confidence limit = (Value ± Value)

SAMPLING DISTRIBUTION FOR STANDART DEVIATION

$$S \pm Z_{\frac{\alpha}{2}} \left(\frac{S}{\sqrt{n}}\right).$$

where S = Standard deviation of Sample

Example is Ass with electric bulb.

The above formulars is used when $n \geq 30$.

Sampling distribution for variance

" " median

the above is not consider for now.

NORMAL DISTRIBUTION

Under normal distribution, to find the final answer,

=> If u are using 2-tail Z-table, u will Subtract that value from 1 which will give u the exact final answer.

=> If u are using 1-tail Z-table, u Subtract that Z value being gotten from 0.5 which will give u the exact final value.

IDENTIFYING 1 tail or 2 tail Z-table.

For 2-tail Z-table, first row first column u will see 0.5000.

For 1 tail Z-table, first row first column, u will see 0.0000.

If $n < 30$ and σ is known, u can also use the above formular.

NOTE under population proportion using;

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$np \geq 10$ and also. $n\pi \geq 10$

$n(1-p)$ or $n(1-\pi) \geq 10$

15 Curren

met, the sample proportion mean $\mu = p$.

Also it is used if $n \geq 30$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

any where u see dat Cal any percentage (99%, 90% e.tc) lesser, greater or ~~btw~~ something, know dat is normal distribution.

$\bar{x}$ = Sample mean.

μ = population mean.

LLY.

LTION (SD)

$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

Example is Ass with 22 L of water last 6 leaves, ie ~~6 pa~~ last 6 pages first column and Second column.

Here, u must know the type of ~~tab~~ Z table u use either 2 tail (subtract what u get from Z table from 1) and if it is 1 tail (u subtract the Z value from 0.5)

NOTE At 2 tail Z table, 1st row, first column, u see 0.5000
At 1 tail " " " " " " 0.0000

ed the mention

limit, a

mean) in the

CONFIDENCE INTERVAL (CI)

Generally for CI, we use

(lower, upper) $\mu = \bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ as the formular.

Value)

It also applies when $n \geq 30$.

If $n < 30$ and σ is known, u can also use the above formular.

STANDART

NOTE under population proportion using;

$$p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$np \geq 10$ and also. $n\pi \geq 10$

... of sample

But in d qn, if u see proportion in SD, CI or Confidence limit, use.

$$\pi = p \pm Z^* \sqrt{\frac{p(1-p)}{n}} \rightarrow \text{success}$$

where p = Sample proportion:

given by $p = \frac{\text{no of success in the sample}}{n}$

p = ~~succe~~ no of success

Note $Z^* = Z_{\frac{\alpha}{2}}$

also $q = 1-p \Rightarrow$ failure.

~~we can rewrite the formular that~~

$$\pi = p \pm Z^* \sqrt{\frac{pq}{n}}$$

or

$$\pi = p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}.$$

Here also, $n \geq 30$.

U can also use $\sqrt{\frac{\pi(1-\pi)}{n}}$ in place of $\sqrt{\frac{p(1-q)}{n}}$, which gives;

$$\pi = p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}}$$

NOTE; if $n < 30$ and σ is unknown, U use T-distribution table which leads us to T-distribution.

T-DISTRIBUTION.

$n < 30$, $\sigma = \text{unknown}$. $\Rightarrow$ population standard deviation

$\mu = \bar{x} \pm T^* \frac{S}{\sqrt{n}}$ for CI ($S \Rightarrow$ sample standard deviation)

i.e $\bar{x} \pm T_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

But on normal t-distribution formular is $T = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$

i.e $T = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$ Which can be used in normal distribution but only used when $n < 30$ and σ = unknown.

Note;
Standard deviation is also known or called standard error.

i.e Standard deviation = Standard Error.

Youtube video.

28-7-2021 class

NOTE

$\Rightarrow$ All examples and assignment

$\Rightarrow$ Note when to use lower T, upper T and middle T-table.

$\Rightarrow$ know when to use T-table and Z-table

HYPOTHESIS

NULL hypothesis is d one that have dat u wanted to test.

Alternative is when it is opposite i.e d statement is not true or less or greater than the sample.

And they are all population mean μ.

Population mean

If Population Standard deviation is unknown u use T-distribution $T = \frac{x - \mu}{\frac{S}{\sqrt{n}}}$

but if σ is known, use.

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

$\Rightarrow$ The breaking strength of cable has the mean $\mu = 1800$ lb and standard deviation $\sigma = 100$ lb by a new technique of producing cable and a sample of 50 cables is $\bar{x} = 1850$ lb. Can we support this distribution in sampling distr-?

Sol

$\mu = 1800$ lb, $\sigma = 100$ lb

$n = 50$ $\bar{x} = 1850$

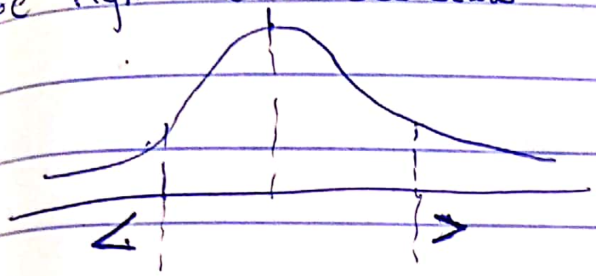
Now ~~producing~~ of new lowering strength

$H_0 : \mu = 1800$, Null hypothesis

$H_a : \mu > 1800$

If Ha is greater, u use 1 tail

the right tail. See below.



Note Significant level = 1 − confidence level.

Test Statistics ⇒ Z table

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1850 - 1800}{\frac{100}{\sqrt{50}}} = 3.54$$

from d question, 99% is $\alpha = 0.01$

= 0.9998

= ~~1 − 0.9998~~ = 0.0002

Standards are 95%, 99% 99.73%

find while p-value of 99% in 1 tail test is 2.33 instead of 2.58 in 2 tail.

Using One tail 99%

= 2.33 = Ha

Since Ha < 3.54 (z score), so the test is wrong.

∴ H_0 is not Correct

3.54 − 3.5 = 0.04

1.4 + 0.06 = 1.46 0.945 = 975

If ~~P value~~ Ha

Note;

Since Pvalue (2.33) ≮ Z score 3.54

∴ Ha is Correct and H_0 is not correct.

But if Pvalue > Z score,

∴ Ha is not Correct and H_0 is Correct.