C.W
30/10/24

Paper Source
Subject.....
Date :.....Time:.....

"Median"

- Median depends on the position

~~Step: 1~~

- Median is the middle most value in the data set.

Steps to find the median:

Step1: Sort data in ascending order:

10, 15, 18, 22, 25, 28, 33, 43, 50

Step2: Find the position:

$\therefore \frac{n+1}{2}$ (even or odd)

$\therefore \frac{9+1}{2} = 5^{th}$ order data

Step3: Find the median value.

$\therefore$ 25 = Median

For even, 10, 15, 18, 22, 25, 28, 33, 43, 50, 1000

$\frac{n+1}{2} = \frac{10+1}{2} = 5.5^{th}$ position

$\frac{25+28}{2} = 26.5$

Quiz 1 ⇒ ~~18~~ 13th November ~~(Monday)~~ Wednesday

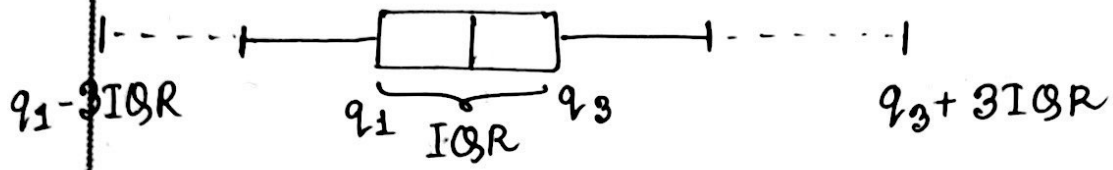
Quiz 2 ⇒ 2nd December (Monday)

Quiz 3 ⇒ 22nd January Wednesday

Mid ⇒ 11th December Wednesday

Paper Source
Subject.....
Date :.....Time:.....

Outlier : $\mu \pm 3\sigma$ | μ = mean, σ = population standard deviation



Weighted Mean:- (Fixed Value কে weight ধরতে হবে)

$\bar{X}_w$

Weighted Median:-

15 (17) (20) 22 (24)

Weighted median, $X = \frac{17+20}{2} = 18.5$

Median 20.

odd হলে normal median আর even হলে দুটো same

Mode: (Highest)

Useless

2, 2, 1, 2, 1, 2, 1

Mode 1, 2 X

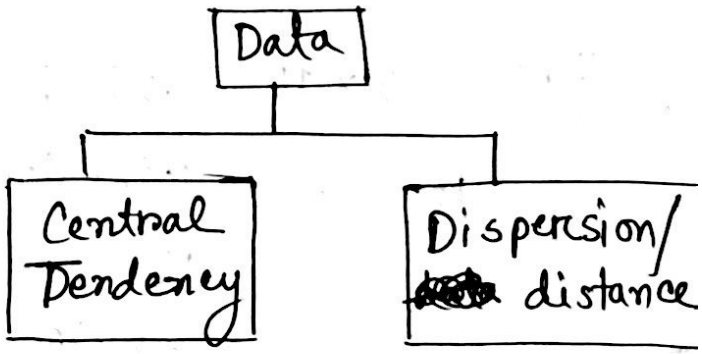
No mode ✓

2, 2, 1, 2, 1, 2

Mode 2

Dispersion: (ডেটা)

Distance can't be negative



Paper Source
Subject.....
Date :.....Time:.....

why $(n-1)$?

Variance :- $s^2 = \frac{\sum^{n}(x_i - \bar{x})^2}{n-1}$

Sample :-

$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n-1\ N}$

$s = \sqrt{s^2} = \sqrt{\text{Variance}}$

↓ standard deviation

~~Practise example:~~

Degree's of freedom :

Mean absolute deviation: $\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

median of Median absolute deviation :

∴ sort → Median, → Median Absolute deviation → sort → median

- Practise examples homework
- quantile

0 —— 25% (1st q) —— 50% (2nd q) —— 75% (3rd q)

1st quantile $= \frac{1}{4} = 25\%$.

position: $0.25 \times 15 = 3.75$ th value $= 10$

~~$0.75 \times 15 =$ th value =~~

$10 + 0.75(4\text{th} - 3\text{rd}) = 10 + 0.75(10 - 10)$

$= 10$

position: $0.75 \times 15 = 11.25$ th value

$= 15 + 0.25(12\text{th} - 11\text{th}) = 15 + 0.25(16 - 15)$

$= 15 + 0.25(1)$

~~$= 15.25$~~ $= 15.25$

Room no. 446
MW
12:15 - 1:00

Paper Source
Subject.....
Date :.....Time:.....

95% percentile

position : $.95 \times 15 = 14.25$th

$= 14$th value $+ 0.25(15\text{th} - 14\text{th})$

$= 17 + 0.95(20 - 17) = 17.75$

Interquartile Range (IQR)

$Q_3 = 75$th percentile

$IQR = Q_3 - Q_1$

Range = max - min

upper limit excluded

Exclusive Interval

Inclusive Interval

140-160
160-180
180-200

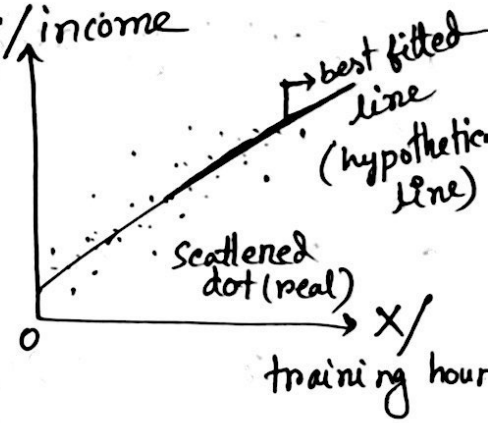
50-55
56-60
61-65

C.W
05/11/2024

Correlation

দুটো numericable variable এর ~~এ~~ মধ্যে relationship.

~~এ~~ best fitted line → যেখানে ~~সবগুলো~~ সবচেয়ে বেশি বিন্দু ঘুচ্ছিদ্ধ থাকে



$-1 \le \rho \le 1$

$-1 \le r \le 1$

-1 এর কাছাকাছি গেলে Strong ~~linear~~ negative relationship

1 " " " Strong positive relationship

~~+0.5 ++~~

Paper Source
Subject.....
Date :.....Time:.....

Rule of thumb

> 0.5 Strong positive linear correlation

< -0.5 Strong negative linear co-relation

0.0002 There is no linear correlation

0.23 positive linear correlation

$0 \leq$ probability range ≤ 1

$$r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{(N-1) S_x S_y} \qquad r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1) S_x S_y}$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \; ; \; S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

Sl no	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1	-	-	-	-	
2	-	-	-	-	

* r এর মান কখনও 0 হবে না

$r = 0.529$, There is a strong linear positive correlation.

Co Variance

$$\sigma_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{M})^2}{N} = \sigma_{xx}$$

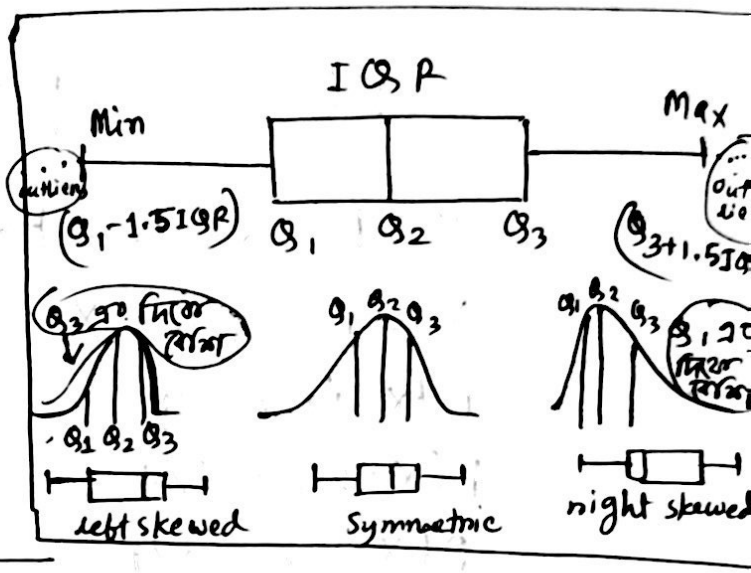
$$\sigma_{xx} = \frac{1}{N} \sum (x_i - \bar{M})(x_i - \bar{M})$$

Paper Source
Subject.....
Date :.....Time:.....

$$\therefore \sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)(Y_i-\mu)$$

$$\therefore r_{xy} = \frac{cov(xy)}{\sigma_x \sigma_y}$$

$$\therefore cov(xy) = r\,\sigma_x\,\sigma_y$$



Lecture 3

Exploratory data analysis

Attribute: Data points or samples are described by attribute.

(I) Nominal / Categorical
(II) Ordinal
(III) Binary
(IV) Numerical

quiz 01: Slide 1 and Slide 2

13th November 2024

Book : Hayter

C.W 11/11/2024

Paper Source
Subject.....
Date :.....Time:.....

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\mu = E(X) = \sum x\, p(x) = \sum \frac{1}{n}\, x = \sum p(x)\, x$$

$2^2 = 4$, $S = \{HH, HT, TH, TT\}$

X = no. of H , $X = 2, 1, 1, 0$

$P(X=1) = 2/4$

$P(X=2) = 1/4$

$P(X=0) = 1/4$

Box plot, Histogram → Chronology

Bargraph ← Discrete Vs Continuous attribute

Attribute is the data point.

$3 \leq X \leq 10$

$X = -3, -2{\cdot}5, 10, 12, 11{\cdot}25$

$-2{\cdot}5 \leq X \leq 12$

Discrete can be negative

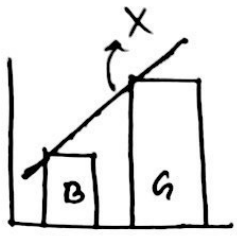
Discrete only equality matters. But continuous G doesn't matter.

Continuous $\{ P(X \geq 2) = P(X > 2)$

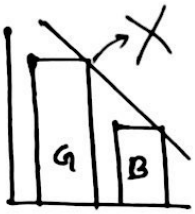
Discrete $\{ 1 - P(X < 2)$

$\Rightarrow 1 - [P(X=0) + P(X=1)]$

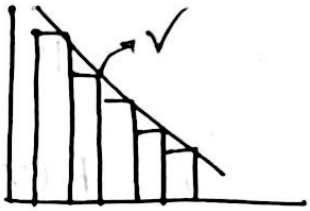
Paper Source
Subject.....
Date :.....Time:.....



Bar graph



Bar graph



Histogram

H (Effect)

S (Cause)		0 (No)	yes (1)	
	No (0)	50	25	75
	Yes (1)	20	55	75
		70	80	

C.W
13 | 11 | 2024

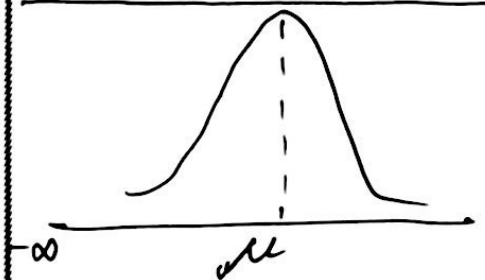
Data distribution

Normal distribution :- / Gaussian distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Discrete : pmf : probability mass function

Continuous :
pdf : prob density function



$X \sim N(\mu, \sigma^2)$
$X \sim N(2, 1)$
$\sigma = 2$

$0 < \sigma^2 < \infty$

$z = \frac{x-\mu}{\sigma}$

$z \sim N(0,1)$

Normal Distribution
* mean = meadian = mode
* symmetry

Expectation (a) = a

Paper Source
Subject.....
Date :.....Time:.....

Median,

$F(x) = P(x \leq w)$

Cdf $F(x) = 0.5$

cdf up to x

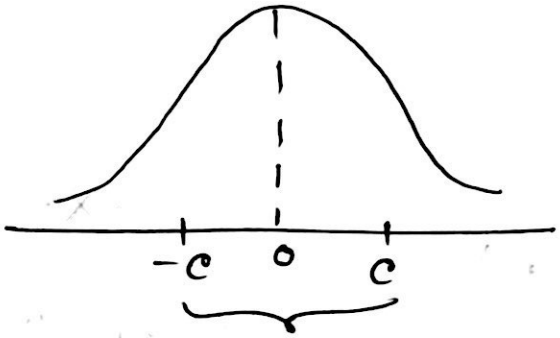
18/11/2024

Properties Normal Distribution

□ Example-1

$\mu \pm \sigma Z_{\alpha/2}$

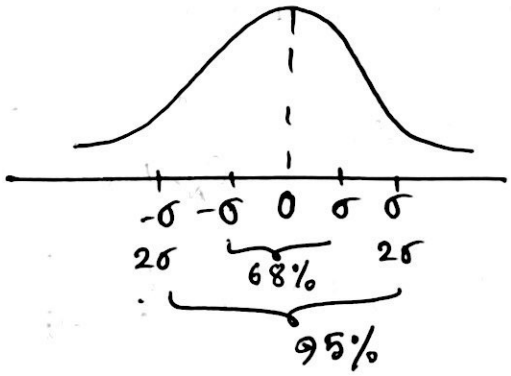
c



95% = 2σ

68% = 1σ

99.7% = 3σ



$X \sim N(\mu, \sigma^2)$

$Z \sim N(0, 1)$

σ এর value 1

$Z = \frac{X - \mu}{\sigma}$

$X \sim N(2, 4)$

~~Z এর~~ $Z \sim N(1.4, 0.15^2)$

Paper Source
Subject CSE303
Date 18/11/24 Time

$\Phi(z) = P(Z \leq z)$

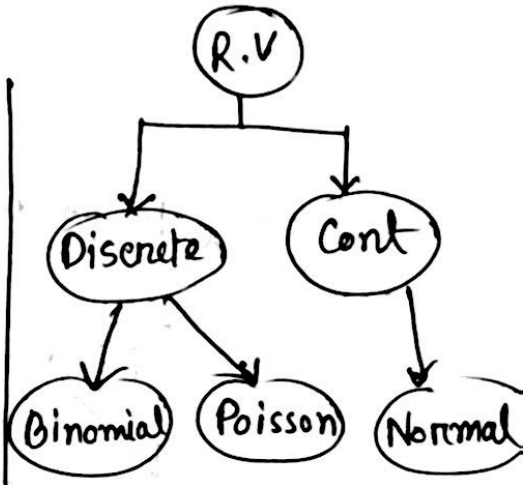
$F(x) = P(X \leq x)$

$50\% = F(x)$

$\hat{\mu} = \bar{X}$
$\hat{\sigma} = s$
} Estimated

$\bar{X} = \frac{\Sigma x_i}{n}$ $\quad s = \sqrt{\frac{\Sigma(x_i - \bar{x})}{n-1}}$

Cdf = Cumulative distribution of



Homework

~~Biono~~ Binomial Distribution

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

C.W 20/11/2024

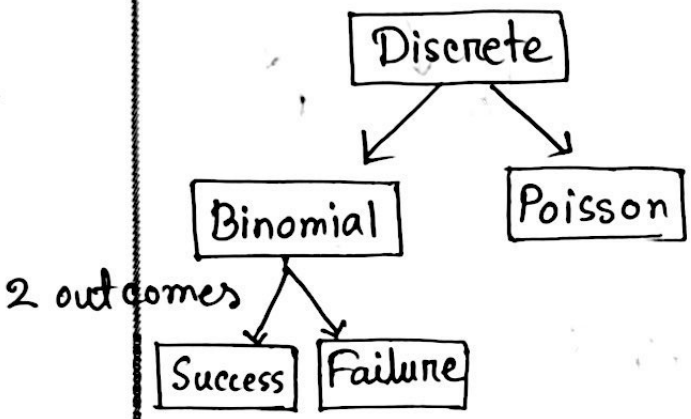
Paper Source
Subject.....
Date :.....Time:.....

Mid - 11/11/2024

Binomial Distribution

$x = 0, 1, 2, \dots, n$

$${}^{n}C_{x}\, p^{x} (1-p)^{n-x} \quad ;$$



Bino: X = number of success

p = probability " "

n = no. of trial / sample size

$p(x)$ = probability mass . (pmf)

discrete এর pmf always

$E(X) = np$

$Var(X) = npq$

$q = 1 - p$; q = probability of failure

$X \sim B(n, p)$

$\sim$ follows

Paper Source
Subject.....
Date :.....Time:.....

Poisson / Counting Distribution → Poisson

X = number of events / incident occurs during any time interval

↓ exponential

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \; ; \; x = 0, 1, 2, \ldots$$

Poisson limit 0 to ∞.

$E(X) = Var(X) = \lambda$; λ = average no. of event occur during time interval

Practise Problem:

$n = 10$

$p = 0.5$

$$P(X=6) = \binom{10}{6} p^6 (1-p)^{10-6}$$

$$= \binom{10}{6}(0.5)^6(0.5)^4 = 0.205$$

$p = 0.6$

$n = 10$

$$P(X=7) = \binom{10}{7}(0.6)^7(0.4)^3$$

$$= 0.215$$

*** for discrete, equality is everything. But for Continuous, equality is nothing.

Paper Source
Subject.....
Date :.....Time:.....

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

$$\Rightarrow 1 - P(X<7) = 1 - [P(X=0) + \dots + P(X=6)]$$

Poisson: $P(x:\mu) = (e^{-\mu} * \mu^x)/x!$

- $\lambda = 2$ per year

$$P(X=3) = \frac{e^{-2}\,2^3}{3!} = 0.180$$

*** Probability can't be negative, also can't be greater than 1.

- Hayter Book, Chapter-3

3.1 exercise (Page-159)

Archer

ⓐ $P(X=2)$

ⓑ ~~$P(X$~~ $1 - P(X<2)$

$P(X \geq 2)$

H.W: 3.1.4, 3.1.5, 3.1.6

Poisson: 3.4.6, 3.4.8

C.W
25/11/2021

Paper Source
Subject.....
Date :.....Time:.....

CORRELATION ANALYSIS (NORMAL Data)

$\chi^2 = \Sigma$
↑ Chi²

$0 \le \chi^2 \le \infty$

2 numeric value association → Correlation
2 ~~cate~~ (Cardinal / Categorical) value " → χ^2

Cause \ effect	H	$\bar{H}$	
S	250 (O_{11})	200 (O_{12})	450 R_1
$\bar{S}$	50 (O_{21})	1000 (O_{22})	1050 R_2
	300 C_1	1200 C_2	1500 GTotal

$E_{11} = \frac{450 \times 300}{1500} =$

$E_{12} = \frac{450 \times 1200}{1500} =$

$E_{21} = \frac{1050 \times 300}{1500} =$

~~Obj~~ $E_{ij} = \frac{R_i C_j}{GT = n}$

~~E11 =~~

$\chi^2 = \Sigma \left(\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right)$

$\chi^2 = \left(\frac{(250-90)^2}{90} + \frac{(200-360)^2}{360} + \frac{(150-210)^2}{210} + \frac{(1000-840)^2}{840} \right)$

$\chi^2_{cal} = 507.93.$

df (degree's of freedom) $= (r-1)(c-1) = (2-1)(2-1)$
$= 1$

Null Hypothesis

H_0: No association

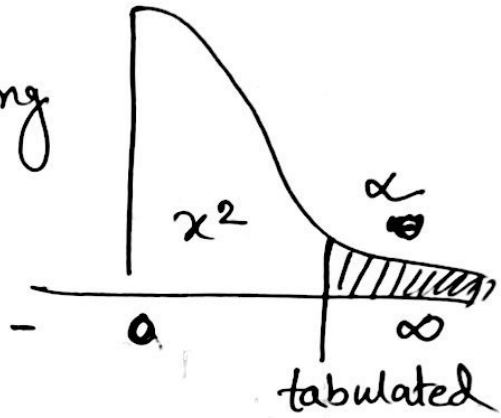
H_a / H_1: There is an association.

1%. 5%. 10%.

Paper Source
Subject.....
Date :.....Time:.....

যদি α না দেওয়া থাকে, by default $\alpha = 5\%$.

Ho : There is no association between chess playing and liking the science fiction



$\chi^2_{cal} > \chi^2_{tab} \rightarrow$ Ho rejected

We may reject Ho.

We have sufficient evidence that statistically and significantly, there is an association between the chess playing and liking the science fiction. Researcher wants to prove the alternative(H_a)

(H·W) Practise Problem

$CI + \alpha = 1$ → formula.

Hayter Book

HomeWork

Practice Problem

Paper Source
Subject
Date: 27/11/24 Time:

	Republican	Democrat	Independent	Total
Male	100	70	30	200
Female	140	60	20	220
Total	240	130	50	420

∴ Expected Values:-

$E_{11} = \frac{200 \times 240}{420} = 114.28$ | $E_{12} = \frac{130 \times 200}{420} = 61.90$ | $E_{13} = \frac{200 \times 50}{420} = 23.80$

$E_{21} = \frac{220 \times 240}{420} = 125.71$ | $E_{22} = \frac{130 \times 220}{420} = 68.09$ | $E_{23} = \frac{220 \times 50}{420} = 26.19$

∴ Calculation of χ^2:-

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(100-114.28)^2}{114.28} + \frac{(70-61.90)^2}{61.90} + \frac{(30-23.80)^2}{23.80} + \frac{(140-125.71)^2}{125.71}$$

$$+ \frac{(60-68.09)^2}{68.09} + \frac{(20-26.19)^2}{26.19}$$

$\chi^2_{\text{calculated}} = 4.459 + 4.0486 = 8.5076$

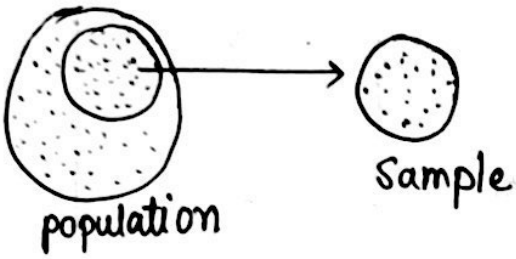
degrees of freedom $= (2-1)(3-1) = 2$

$\chi^2_{\text{tabular}} = 7.824$

Significance level $\alpha = 0.02$

∴ We may reject H_0
There is significant relation between gender and political party. Researcher wants to prove H_a

Paper Source
Subject.....
Date :.....Time:.....



$\mu \quad \sigma \quad \sigma^2 \quad \rho \quad n$
$\bar{x}, s, s^2, r, N$

BBHS 2022
HIES 2022
MICS 2022

$X \sim N(\mu, \sigma^2)$
$X \sim B(n, p)$
$X \sim P(d)$

population list → Sampling f

4 pillars of Sampling
- Simple random sampling (SRS)
- Straightified random = (condition থাকবে) (within Homogenity)
- Systematic = = (within Homogenity)
- Cluster (গুচ্ছ) = = (Within Heterogenity)

SRS সমগুলোতে থাকে

SRS has no conditions.

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

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

$E(\bar{x}) - \mu = 0$ (Unbiased)

$E(\bar{x}) - \mu \neq 0$ (Biased) 0.0001 হলেও biased

same $E(\hat{\mu}) - \mu = \underbrace{E(\hat{\theta}) - \theta}_{\text{বইয়ে দেওয়া}} = \text{Bias} \neq 0$

Exhausted = Fullfilled

Mutually Exclusive = No common thing ($\cap$) $A \cap B = \phi$

Collectively Exhausted = Fullfilled ($\cup$) $A \cup B = S$

Book : Cocran (Sampling Test)

$$n = \left(\frac{z\sigma}{E}\right)^2$$

E = ~~margi~~ margine of error

σ = standard deviation

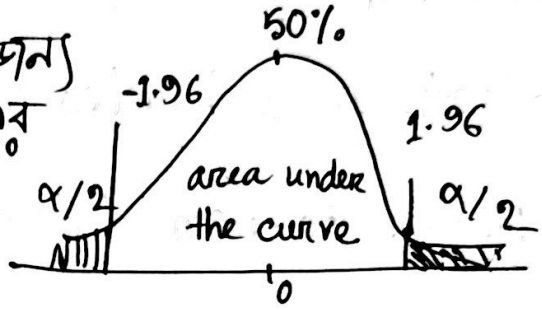
Quiz – Slide 3 & 4. (Monday)

C.W
02/12/2024

Paper Source
Subject.....
Date :.....Time:.....

Mid - 11 Dec 2024

0.5 এর নিচের value এর জন্য negative z table. 0.5 এর উপর হলে (+) z table.
Normally closest lower value নিবো



$\alpha = 5\% = 0.05$

$\frac{\alpha}{2} = 0.025$

$\alpha = 10\% = 0.1$

$p = -1.65$

$$n = \left(\frac{z\sigma}{E}\right)^2$$

{ E = margin of Error
σ = std.

$CI + \alpha = 1$

$CI = 90\%$ ∴ $\alpha = 10\%$

$$n = \frac{z^2 pq}{E^2}$$

{ E = margin of Error
z = standard normal
q = 1 - p

Sampling Distribution of a statistics :-

If you increase the sample size, central

$\bar{X}, S, S^2$

Sampling distribution এ Variance কে যদি Square root করে তাহলে

$Var(X) = SD(X)$
$Var(\bar{X}) = SE(\bar{X})$

2-12-21
Lab work

LAB

Paper Source
Subject.....
Date :.....Time:.....

(I) Vs code , Conda

(II) How to integrate conda with Vs code

(III) What is python (or conda) environment?

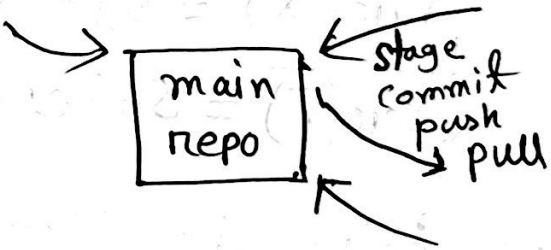
(IV) How to maintain multiple environment?

(V) ~~In~~ Create an account on GitHub.

(VI) Install git in your device.

(VII) How to connect with a remote repo using ~~Single Vs~~ VS code?

- ☐ Git basics : repo, branch, stage, commit, push, pull, merge
- ☐ clone ← download or import
- ☐ Install necessary extension on VS code for version control (github)
- ☐ Create your ~~dist~~ first project and push it on github
- ☐ remote device
  local device → (origin)



- ☐ Data analysis from Team Data set.

Paper Source
Subject.....
Date :.....Time:.....

No classes on Monday.

Sampling distribution of a statistics

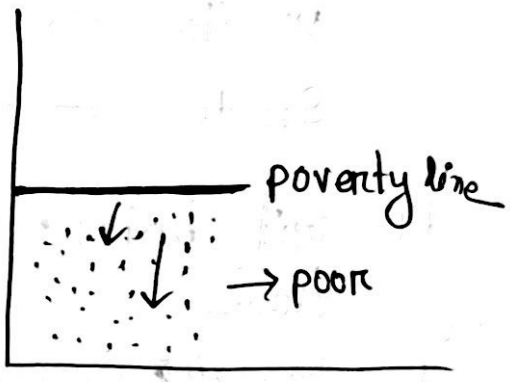
Sample size

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \text{(Mean based)}$$

$$n = \frac{pqz^2}{E^2} \quad \text{(proportion based)}$$

prob. distribution → x / y / z → over the population

Normal, poisson, Binomial



Sampling Distribution (Sampling distribution of statistics)

$\sqrt{Var(X)} = Std(X)$

$\sqrt{Var(\bar{X})} = Std\,error(\bar{X})$

$\bar{X} = \hat{\mu}$

$s = \hat{\sigma}$

$\wedge$ = estimation

Central limit Theorem

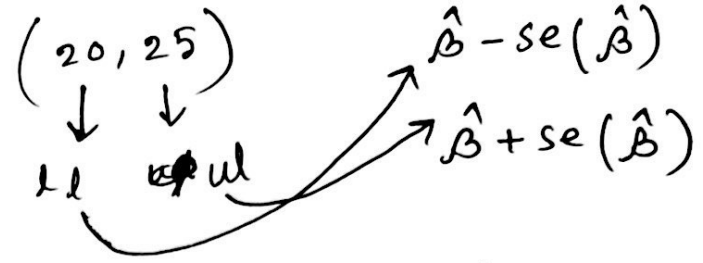
$X \sim (\mu, \sigma^2)$

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Standard error = $SE = \frac{s}{\sqrt{n}} = \frac{\hat{\sigma}}{\sqrt{n}}$

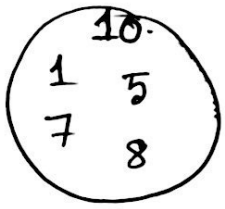
Paper Source
Subject.....
Date :.....Time:.....

- $\hat{\beta} \pm se(\hat{\beta})$



- 95% CI means, if I draw 100 times sample we will get 100 $\bar{X}$'s. Among the hundred $\bar{X}$'s 95% $\bar{X}$ will be very close / represent the μ.

- $N = 35$
  $n = 5$



- Bootstrap & Jackknife → syllabus এ নাই
  ↓ ↓
  with replacement without replacement

Mid — 11-12-2024

Slide 1-5
Definition থাকবে না, Math থাকবে
Slide-5 → Theoretical / Descriptive

$E = E(\hat{\theta} - \theta) = \mu$

কখন সবচেয়ে safe side যখন p না থাকে (50%. 50%)
তখন highest sample size পাওয়া যায়,
Standard error বের করা,

Z table, chi square,	formula sheet