Data Science and Artificial Intelligence

Probability and Statistics

Introduction to Sampling Distribution

Lecture No.- 05

By- Rahul Sir

Topics to be Covered

3-4-work

10 Papers

Activities

Risk-analyst

Topic Introduction to Sampling Distribution 05

70% gain

✓ class regular + Discipline

✓ Prob + Diary → Today Targets

(Notes + Problems)

✓ Prob – DPP 1–50 questions

→ Theory → 10 Basic → Theory

→ 20 Intermediate → Logic –

→ 20 question → Telegram Post

TVF

→ friend

→ Internet

→ Teacher

15 question

→ 5 question

Teacher

✓ Don't Panic

question billions

5 ques =

X Leave

30% revise

Present Process

Do yourself

Q7. A population consists of the five numbers 2,3, 6, 8, 11. Consider all possible samples of size two which can be drawn with replacement from this population. Find (a) the mean of the population, (b) the standard deviation of the population, (c) the mean of the sampling distribution of means, (d) the standard deviation of the sampling distribution of means, i.e., the standard error of means.

(A) 6 D) 2.32

(B) 3.29

(C) 6.0

PW

Do yourself

Q8. Assume that the heights of 3000 male students at a university are normally distributed with mean 68.0 inches and standard deviation 3.0 inches. If 80 samples consisting of 25 students each are obtained, what would be the mean and standard deviation of the resulting sample of means if sampling were done (a) with replacement, (b) without replacement?

A) 0.6

B) 0.6

Introduction to Sampling Distribution

ounce (weight)

Q9. Five hundred ball bearings have a mean weight of 5.02 oz and a standard deviation of 0.30 oz. Find the probability that a random sample of 100 ball bearings chosen from this group will have a combined weight, (a) between 496 and 500 oz, (b) more than 510 oz.

Do yourself

A) 0.2164

B) 0.0015

Introduction to Sampling Distribution

Do yourself

Q10. Find the probability that in 120 tosses of a fair coin (a) between 40% and 60% will be heads, (b) 5/8 or more will be heads.

We consider the 120 tosses of the coin as a sample from the infinite population of all possible tosses of the coin. In this population the probability of heads is $p = 1/2$ and the probability of tails is $q = 1 - p = 1/2$.

(a) We require the probability that the number of heads in 120 tosses will be between 40% of 120, or 48, and 60% of 120, or 72. We proceed as in Chapter 4, using the normal approximation to the binomial distribution. Since the number of heads is a discrete variable, we ask for the probability that the number of heads lies between 47.5 and 72.5.

Introduction to Sampling Distribution

0.8197

Q13. Vehicles pass an observer in such a way that the waiting time between successive vehicles may be adequately modelled by an exponential distribution with mean 15 seconds. Particular details of each vehicle that passes are recorded on a sheet. There is room to record the details of twenty vehicles on each sheet.

What, approximately, is the probability that it takes less than 6 minutes to fill one of the sheets?

Exponetial Dis $\frac{1}{\lambda} = \mu$

$E[x] = \lambda$ ← mean

$V(x) = \lambda^2$ mean

$S.D = \sqrt{\lambda}$ mean

$E[x] = \lambda = 15 + (15) + (15) + \cdots 20 \text{ times} = 20 \times 15$

$= 300$

$V(x) = \lambda^2 = (15)^2 + (15)^2 + (15)^2 + (15)^2 + 20 \text{ tims}$

$= 20 \times 225$

$= 4500$

$N(\mu, \sigma^2) \longrightarrow$ z SCORE

$\mu = 15$ SECONDS

for 20 SEC $\mu = 15 \times 20 = 300$

$$\sigma^2 = (15)^2 \times 20 = 225 \times 20 = 4500$$

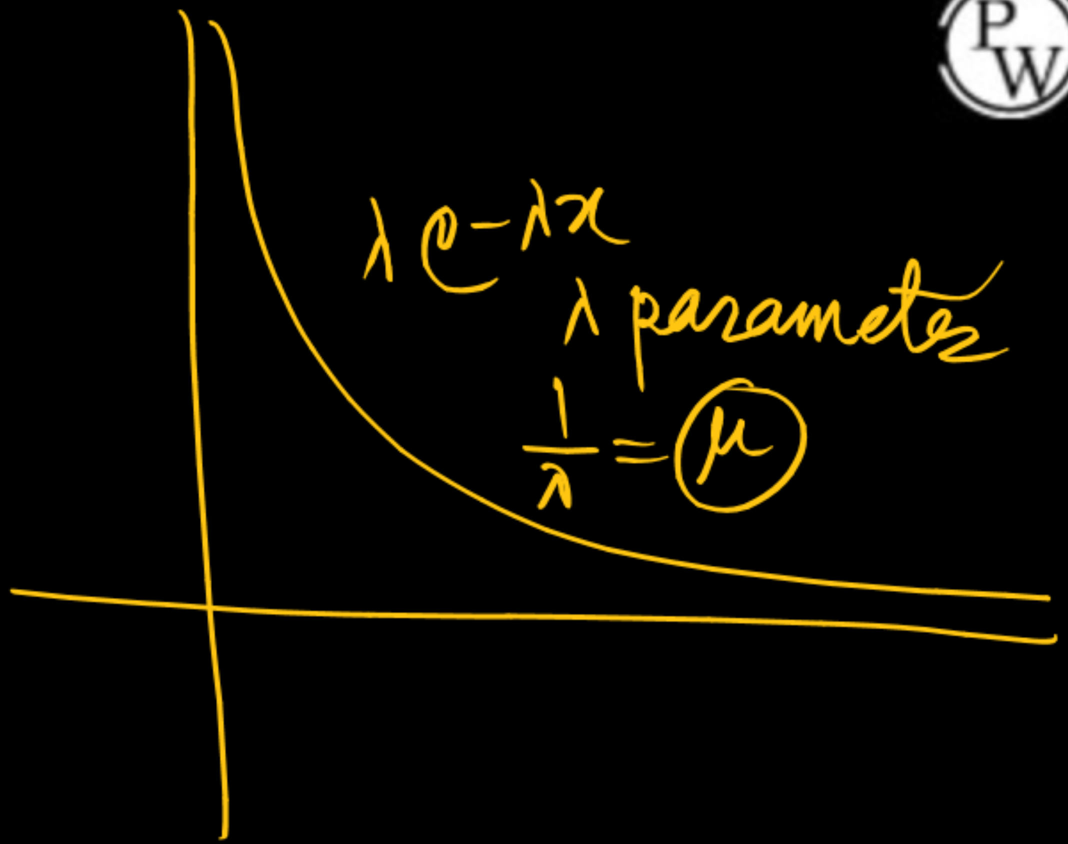
$$\underset{\mu,\ \sigma^2}{N(300, 4500)} = P\left(X < \underset{\text{SEC}}{360}\right)$$

$$= P\left(\frac{X-\mu}{\sigma} < \frac{360-\mu}{\sigma}\right)$$

$$= P\left(Z < \frac{360-300}{\sqrt{4500}}\right)$$

$$= P(Z < 0.89) = \boxed{0.81}$$

Ans



Introduction to Sampling Distribution

Q14. Suppose that the duration of a patient's visit to a dentist's surgery is a random variable X with mean 10 minutes and standard deviation 5 minutes. The dentist attends to eight patients each morning before taking a coffee break.

8 Patients

$T_8 = T_1 + T_2 + T_3 + T_4 + T_5 + T_6 + T_7 + T_8$

$T_8 = E[T_1] + E[T_2] + E[T_3] + E[T_4] + \cdots + E[T_8]$

$= 10 + 10 + 10 + 10 \cdots$ 8 times

$= 80$ minutes

$V_8 = 8 \times (5)^2 = 25 \times 8 = 200$

$\boxed{\sigma_x = (8) \times (5)^2 = 200}$

standard deviation $= \sqrt{200} = 14.11$ minutes

coffee break time $= 14.11$ minutes

Introduction to Sampling Distribution

$P[-1.77 < z < 1.77] = \boxed{0.9616}$

Q16. Suppose that the mean standing height of adult meerkats is 30 cm, and the standard deviation of their heights is 1 cm. If a sample of 50 adult meerkats is taken, what is the probability that their mean standing height, $\bar{X}_{50}$, will be within $\boxed{0.25 \text{ cm}}$ of the mean standing height of the population of adult meerkats?

$\mu + \epsilon \quad 30 + 0.25 \quad 30.25$

$\mu - \epsilon \quad 30 - 0.25 \quad 29.75$

Mean Height = 30cm

Standard deviation = 1cm

$n = 50$ samples

In sampling Distribution $N\left(\mu, \frac{\sigma^2}{n}\right)$

$\boxed{\mu = 30}$

$\sigma^2 = \frac{\sigma^2}{n} = \frac{(1)^2}{50}$

$\boxed{\sigma = \sqrt{0.02}}$

30cm mean

$P[29.75 \leq \bar{X}_{50} \leq 30.25]$

Slide 18

$$P\left(29.75 \le \overline{X}_{50} \le 30.25\right) = P\left(\frac{29.75-\mu}{\sigma} \le \frac{\overline{X}_{50}-\mu}{\sigma} \le \frac{30.25-\mu}{\sigma}\right)$$

$$= P\left[\frac{29.75-30.00}{\sqrt{0.02}} \le Z \le \frac{30.25-30}{\sqrt{0.02}}\right]$$

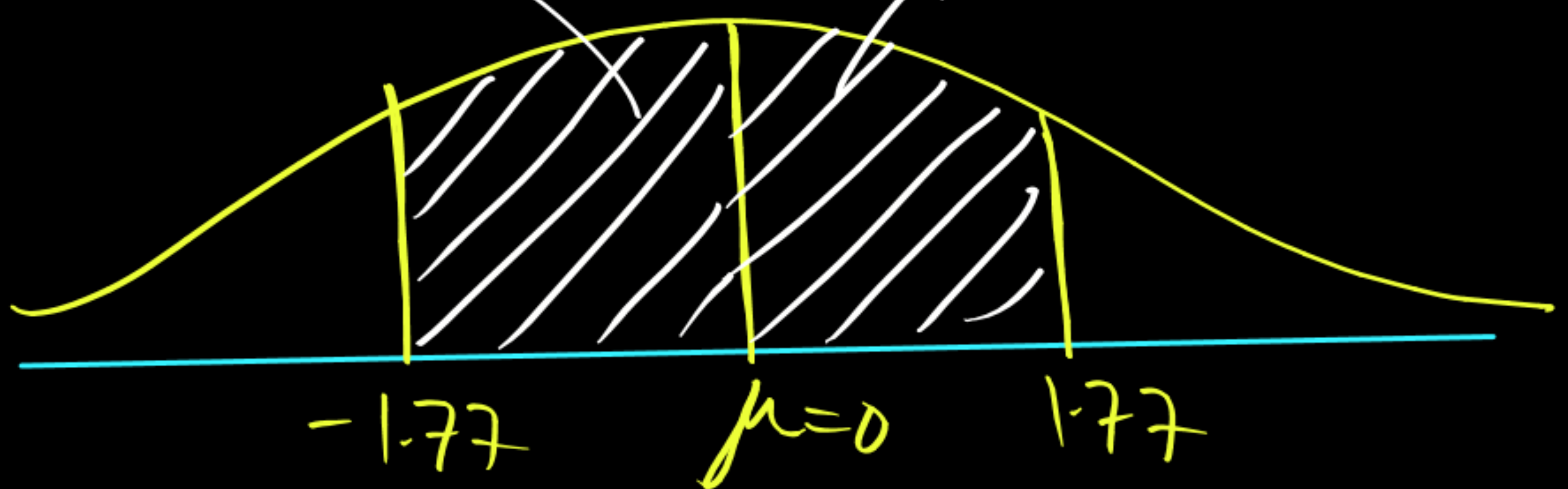
$$= P\left[-1.776 \le Z \le 1.776\right]$$

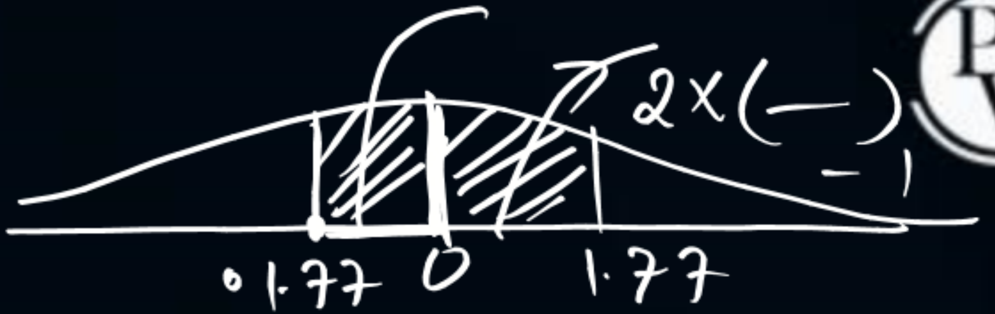
$$= 2 \times 0.9616 - 1$$

$$= \underline{\underline{0.92}}$$

$$\left\{\begin{array}{l} P[-1.76 \le Z \le 0] \\ + P[0 \le Z \le 1.76] \end{array}\right.$$

$(\quad) + (\quad)$





Q17. A van has a weight carrying capacity, or 'payload', of 1600 kg. The company that owns the van wishes to transport 30 small but heavy items which it knows from long experience have a mean weight of 50 kg and a standard deviation of 10 kg. What is the approximate probability that the total weight of the 30 items will be less than the payload of the van?

$P(\bar{X} < 30)$

$= P\left[\dfrac{\bar{X}-\mu}{\sigma} \le \dfrac{30-\mu}{\sigma/\sqrt{n}}\right]$

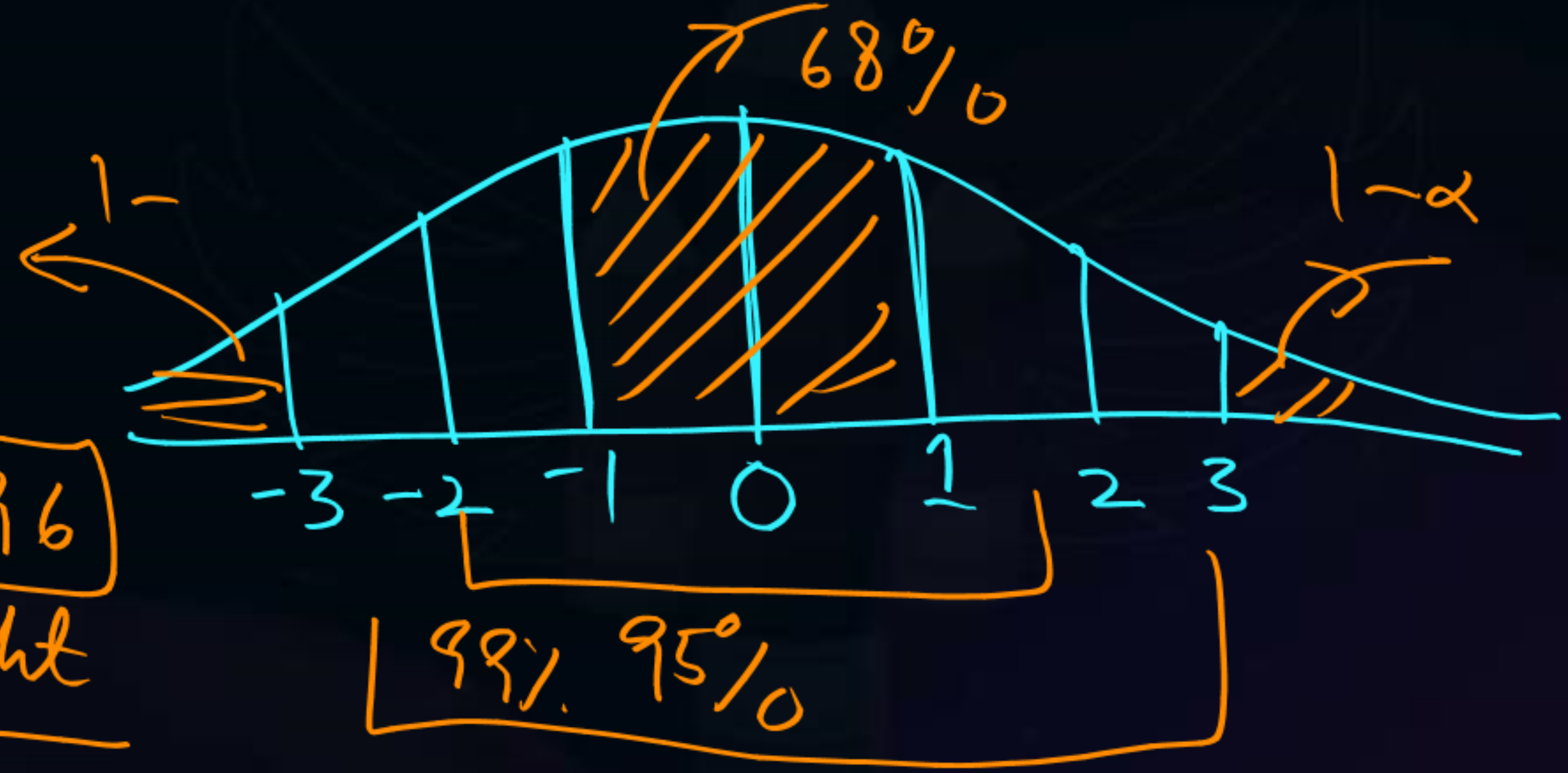
$= P\left(Z \le \dfrac{30-\mu}{\sigma/\sqrt{n}}\right)$

$P(Z < 1.82)$

$= \underline{0.96}$

$= \boxed{0.96}$

Right



$n = 30$ Items

mean $= 30 \times 50 = 1500$

variance $= 30 \times (10)^2 = 3000$

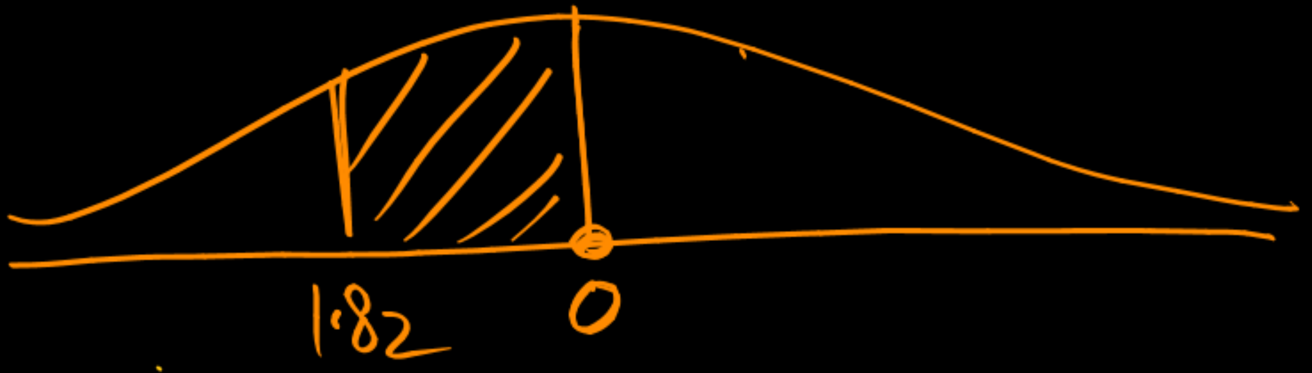
$N(1500, 3000)$

$P(T_{30} < 1600) = P(X < 1600)$

$P\left(\frac{X-\mu}{\sigma} < \frac{1600-\mu}{\sigma}\right)$

$= P\left(Z < \frac{1600-1500}{\sqrt{3000}}\right)$

$= P(Z < 1.82 \text{ or } 1.83)$



$= \underline{\underline{0.96}}$

$\sigma^2 = 3 \times (5)^2 = 75$

$\mu = 1004 \times 3 = 3012$

Q18. Suppose that the volume of orange juice in bottles, each of which is labelled as containing 1 litre, is normally distributed with mean 1004 ml and standard deviation 5 ml.

(a) Find the probability that the total contents of three randomly selected bottles will be less than 3 litres.

(b) Find the probability that the mean contents of a random sample of four bottles will be less than 1 litre.

1 litre = 1000 ml

$\mu = 1004$ ml

$\sigma = 5$ ml

$P(X < 3000\text{ml}) = P\left(\frac{x-\mu}{\sigma} < \frac{3000-\mu}{\sigma}\right) = P\left(z < \frac{3000-\mu}{\sigma}\right)$

$N(3012, 75) = P\left[z < \frac{3000-3012}{\sqrt{75}}\right] = P[z < -1.39]$

$= \underline{0.0823}$ Ans

$$P(X<1000) = P\left(\frac{X-\mu}{\sigma} < \frac{1000-\mu}{\sigma}\right)$$

$$= P\left[Z < \frac{1000-\mu}{\sigma}\right]$$

$$N\left(1004, \frac{25}{4}\right) = P\left(Z < \frac{1000-1004}{\sqrt{6.25}}\right]$$

$$= P[Z < -1.6]$$

$$= \underline{0.054} \text{ Ans}$$

$\mu = 1004$

$$V(X) = \frac{1}{4} \times (5)^2$$

$$= \frac{25}{4}$$

Introduction to Sampling Distribution

✓ BEST – 2 min

Q19. Rather than keep an accurate record of individual transactions, the holder of a bank account records only individual deposits into and withdrawals from his account to the nearest pound. Suppose that the error in individual records may be adequately modelled by a continuous uniform distribution $U\left(-\frac{1}{2}, \frac{1}{2}\right)$ and that he makes 400 transactions in a particular year.

0.91

(a) What, approximately, is the distribution of the error in his estimate of his bank balance at the end of the year? Hint: it will be useful to remember that for $X \sim U(a, b)$, $\boxed{E(X) = (a + b)/2 \text{ and } V(X) = (b-a)^2/12.}$

(b) Find the probability that the error in his estimate is less than £10.

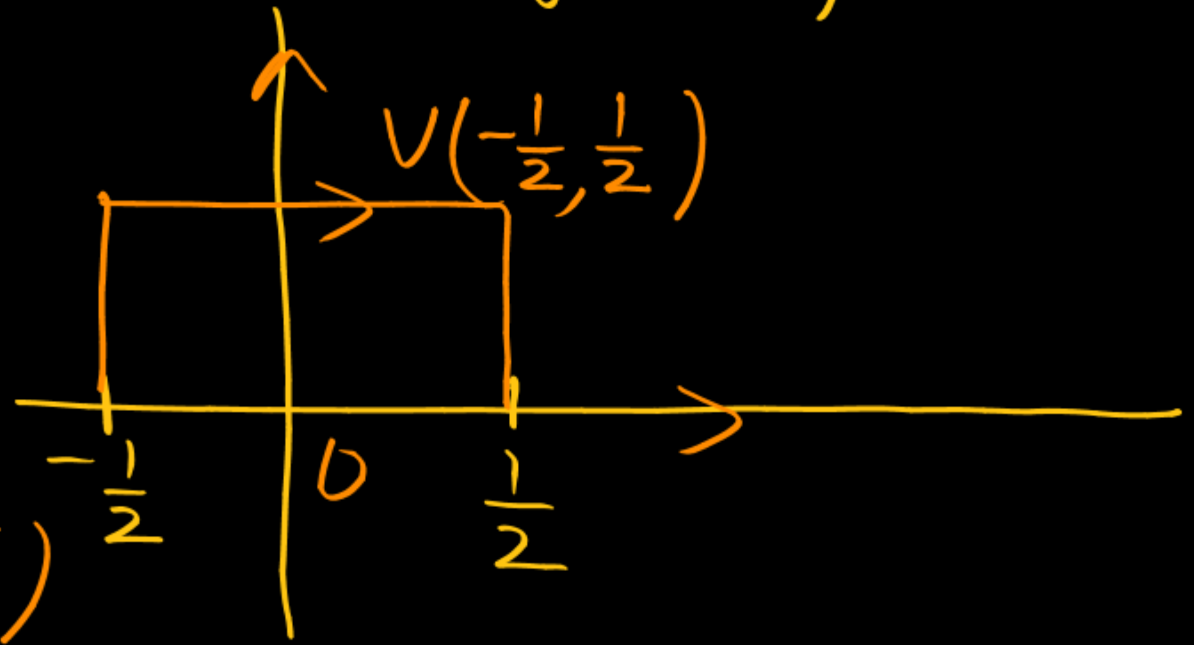
Probability

Z score

$$E[X] = \frac{a+b}{2} = \frac{-\frac{1}{2}+\frac{1}{2}}{2} = 0$$

$$V(X) = \frac{(b-a)^2}{12} = \frac{\left(\frac{1}{2}+\frac{1}{2}\right)^2}{12} = \frac{1}{12}$$

$U\left(-\frac{1}{2}, \frac{1}{2}\right)$



400 Transcation $\quad U(a,b) \longrightarrow N(\mu, \sigma^2)$

$$U(a,b) \longrightarrow N\left(0, \frac{1}{12}\right)$$

$$\text{400 Transcation} \longrightarrow N\left(0, 400 \times \frac{1}{12}\right]$$

$$N\left(0, \frac{100}{3}\right) \text{ Ans}$$

$$\left.\begin{array}{l} \mu = 0 \\ \sigma^2 = \frac{100}{3} \end{array}\right]$$

$$P(-10 < X < 10) \Rightarrow P\left(\frac{-10-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{10-\mu}{\sigma}\right)$$

$$\underline{\sigma = \frac{\sigma}{\sqrt{n}}}$$

$$\left[10 \to \text{Error Estimate}\right]$$

$$\Rightarrow P\left(\frac{-10-0}{\sqrt{100/3}} < Z < \frac{10-0}{\sqrt{100/3}}\right)$$

$$\Rightarrow P[-1.73 < Z < 1.73]$$

$$\Rightarrow P[-1.73 < Z < 0] + P[0 < Z < 1.73]$$

$$\Rightarrow \underline{\underline{0.916}}$$

Non Parameter

DataSET $\longrightarrow$ degree of Freedom 1

$\rightarrow$ Expectation is given

$E[X] = \mu$ ✓

degree of Freedom = 2

$E[X]$, $E[X^2]$, $V(X)$, $\sqrt{V(X)}$

degree of freedom = 15

70 students
Average = 50

$\rightarrow$ one student = 50

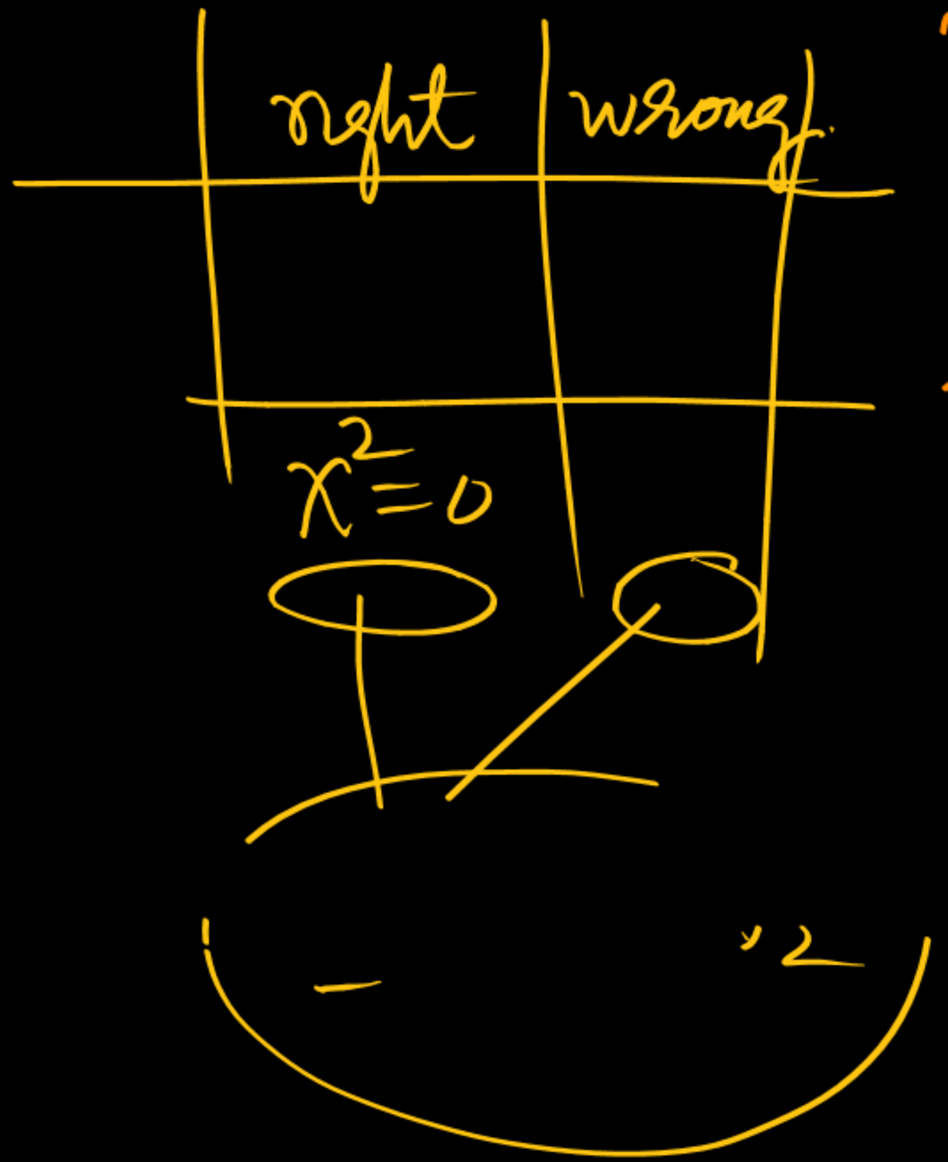
degree of freedom 1

Degree of Freedom

$X_1 + X_2 = 2$

Two variable

constraint

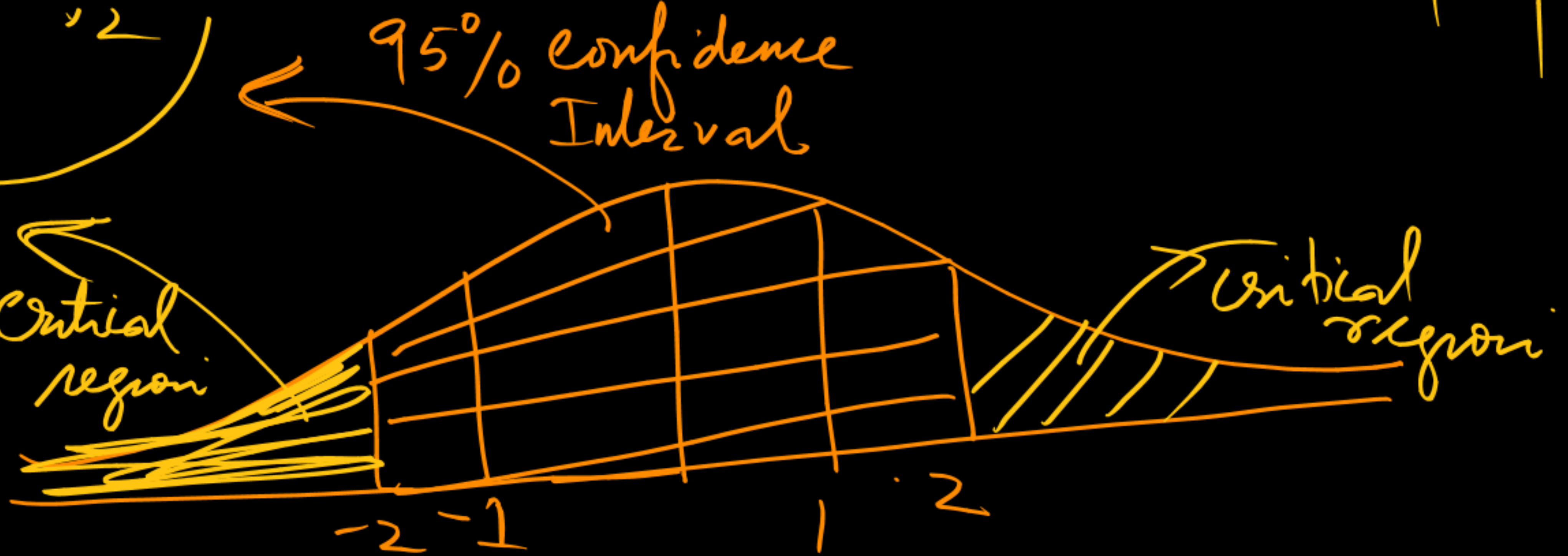


$$\chi^2_{(1)} = \sum_{i=1}^{n} \frac{O_i^2}{E_i} \Rightarrow \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$\chi^2 = 0$ (chi square TEST — Hypothesis Accept)

$\chi^2 > 0$ (chi-square)

Significance



THANK - YOU

95%

✓ LEAST method ✓

✓ t-Test F-Test (goodness of fit)

✓ correlation / regression