3 mins

Question-1: The 2010 U.S. Census found the chance of a household being a certain size. The data is in the table ("Households by age," 2013). What is the probability that a household will have at least 5 members?

Size of household	1	2	3	4	5	6	7 or more
Probability	26.7%	33.6%	15.8%	13.7%	6.3%	2.4%	1.5%

$P(\geq 5) = P(5) + P(6) + P(7+)$

$= 6.3 + 2.4 + 1.5 = 101\%$

$= .101$

- A ● 0.33
- B ● 0.10
- C ● 0.26

3 mins

Question-2: A discrete probability distribution of scoring runs in one throw of a ball by a particular batsman in a cricket match is given in the table. Find the missing probability.

F	0	1	2	3	4	5	6
$P(F)$	0.15	0.27	0.35	0.15	0.04	0.01	

$0.15 + 0.27 + 0.35 + 0.15 + 0.04 + 0.01 + x = 1$

$\sum p_i = 1$

$x = 0.03$

- A ● 0.3
- B ● 0.02
- C ● 0.03

Activate Windows
Go to PC settings to activate Windows.

4mins

92%

Question-3: A hockey goaltender has a save percentage of [redacted]. What would be the expected number of goals scored on this goaltender in a game where she faced 35 shots?

$(1-0.92)\times 35 = 2.8$

- A 5.2
- B 2.8
- C 2
- D 2.5

Question-4: Which of the following is not a property of a Binomial Experiment?

- A All trials are identical.
- B Each trial has only two possible outcomes.
- C The probability of success may change from trial to trial.
- D The purpose of the experiment is to determine the number of successes that occurs during the n trials.

Activate Windows
Go to PC settings to activate Windows.

Amins

Expenditures of a Company (in Lakh Rupees) per Annum Over the given Years.

Year	Item of Expenditure				
	Salary	Fuel and Transport	Bonus	Interest on Loans	Taxes
1998	288	98	3.00	23.4	83
1999	342	112	2.52	32.5	108
2000	324	101	3.84	41.6	74
2001	336	133	3.68	36.4	88
2002	420	142	3.96	49.4	98

Refer to the above table and answer the following questions

Question-1: The total amount of bonus paid by the company during the given period is approximately what percent of the total amount of salary paid during this period?

- 0.1%
- 0.5%
- 1% ✓

$$\frac{\Sigma B}{\Sigma S} \times 100 = \frac{17}{1710} \times 100 \simeq 1\%$$

Activate Windows
Go to PC settings to activate Windows.

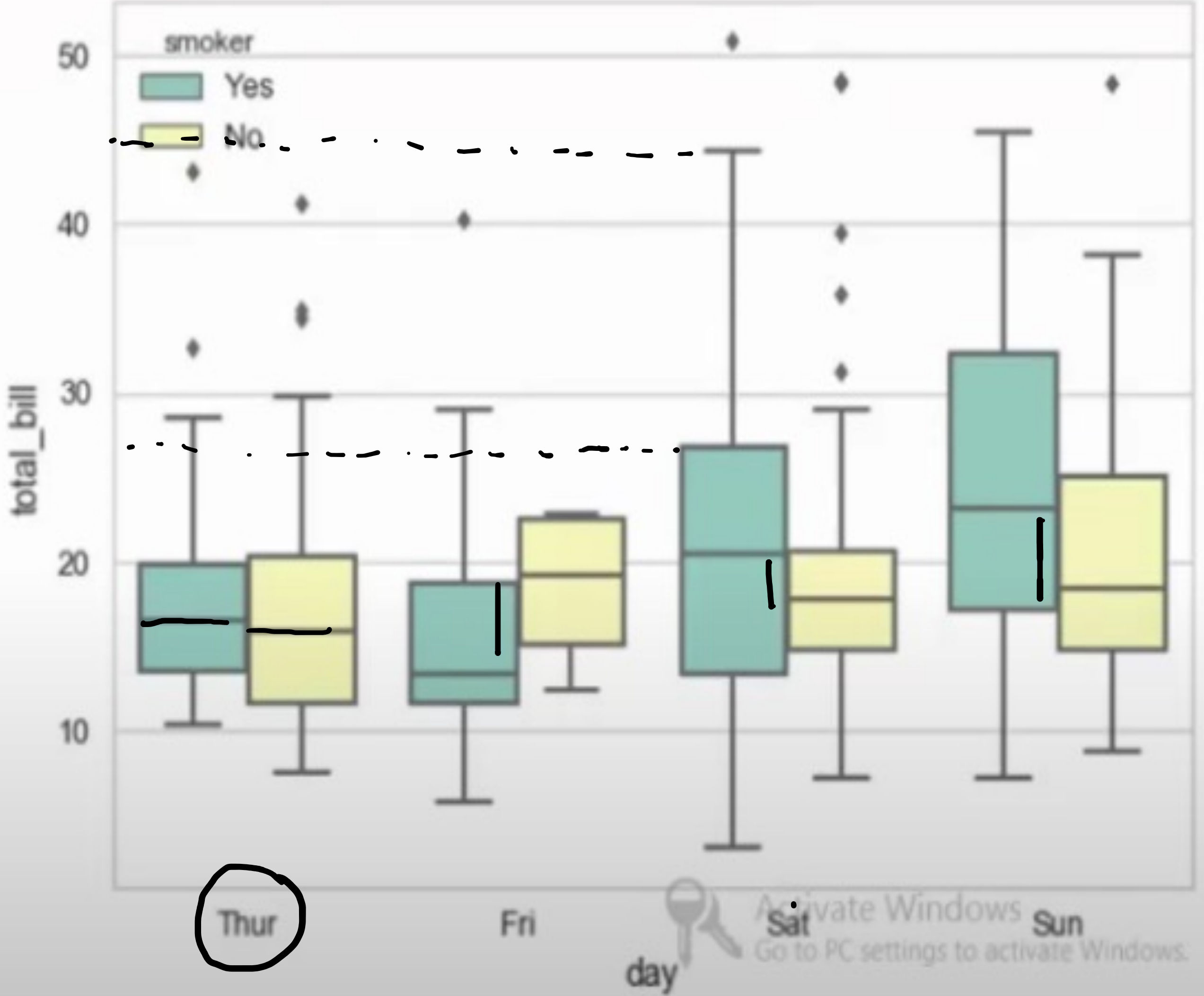
5 mins

Question-1: What is the upper bound range of the total bill for the Smokers on Saturday?

- A ● 38-45
- B ● 30-50
- C ● 38-50

Question-2: On which particular day the median total bill for both Smokers and Non-Smokers is approximately same?

- A ● Thursday
- B ● Friday
- C ● Saturday
- D ● Sunday



Activate Windows
Go to PC settings to activate Windows.

Q 7 mins (Inferential Stats)

A simple random sample of 50 adults womens is obtained, and each person's red blood cell count (in cells per microliter) is measured The sample mean is 4.63. The population standard deviation for red blood cell counts is 0.54. Construct the 95% confidence interval estimate for the mean red blood cell counts of adults.

$$CI = \mu \pm z \times \frac{\sigma}{\sqrt{n}}$$

$n = 50$

$\mu = 4.63$

$$CI = 4.63 \pm 1.96 \times \frac{0.54}{\sqrt{50}}$$

$\sigma = 0.54$

$$CI = [4.48 - 4.78]$$

$z = 1.96$

Activate Windows
Go to PC settings to activate Windows.

$=\ \longrightarrow\ \neq$

$>\ \longrightarrow\ \leq$

$<\ \longrightarrow\ \geq$

$\longrightarrow H_1 : >, <, \neq$

$\longrightarrow H_0 : =, \geq, \leq$

Formulating Null and Alternate Hypothesis

Example-1: A restaurant owner installed a new automated drink machine The machine is designed to dispense 530 mL of liquid on the medium size setting. The owner suspects that the machine may be dispensing too much in medium drinks. They decide to take a sample of 30 medium drinks to see if the average amount is significantly greater than 530 ml

all scenario

$$\begin{cases} H_1 : \mu \boxed{>} 530\,ml \\ H_0 : \mu \leq 530\,ml \end{cases}$$

$=$ $>$?

H_0, H_1 — Right tail test

Activate Windows
Go to PC settings to activate Windows.

Q Hypothesis Testing :-

$\alpha = 5\%$

{Critical Value Method / p value method}

A sample of 40 new baseballs had a bounce height mean of 92.67 inches and a SD of 1.79 inches. Use a .05 sig. level to determine whether there is sufficient evidence to support the claim that the new balls have bounce heights with a mean different from 92.84 inches (a previous test figure).

① formulate null & alternate hypothesis

$$\mu \neq 92.84 \ (H_1)$$
$$\mu = 92.84 \ (H_0)$$

② which test? } 2 Tail test

③ L.CV and H.CV

④ Decision

$$L.C.V = \mu - Z \times \frac{\sigma}{\sqrt{n}} = 92.84 - 1.96 \times$$

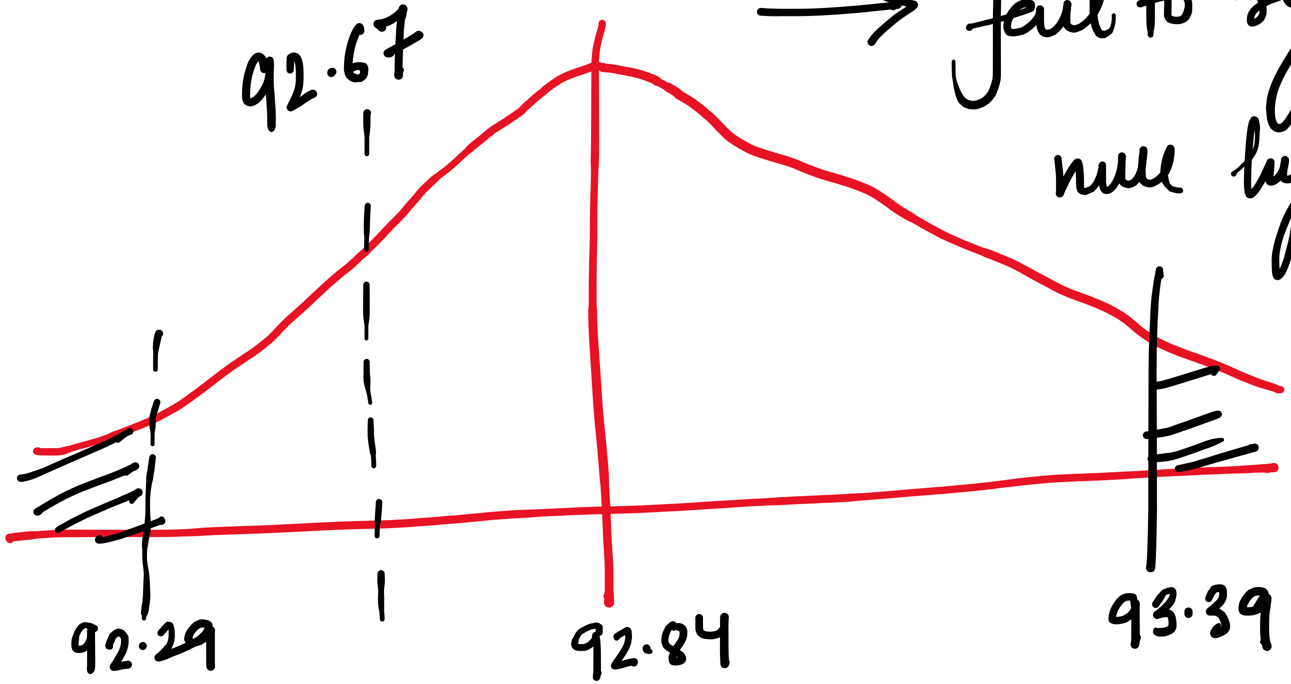
$$H.C.V = \mu + Z \times \frac{\sigma}{\sqrt{n}}$$

Activate Windows
Go to PC settings to activate Windows.

$$L.C.V = \mu - Z \times \frac{\sigma}{\sqrt{n}} = 92.84 - 1.96 \times \frac{1.79}{\sqrt{40}}$$

$$H.C.V = \mu + Z \times \frac{\sigma}{\sqrt{n}} = 92.84 + 1.96 \times \frac{\sqrt{1.79}}{\sqrt{40}}$$

$$L.C.V = 92.29$$
$$H.C.V = 93.39$$



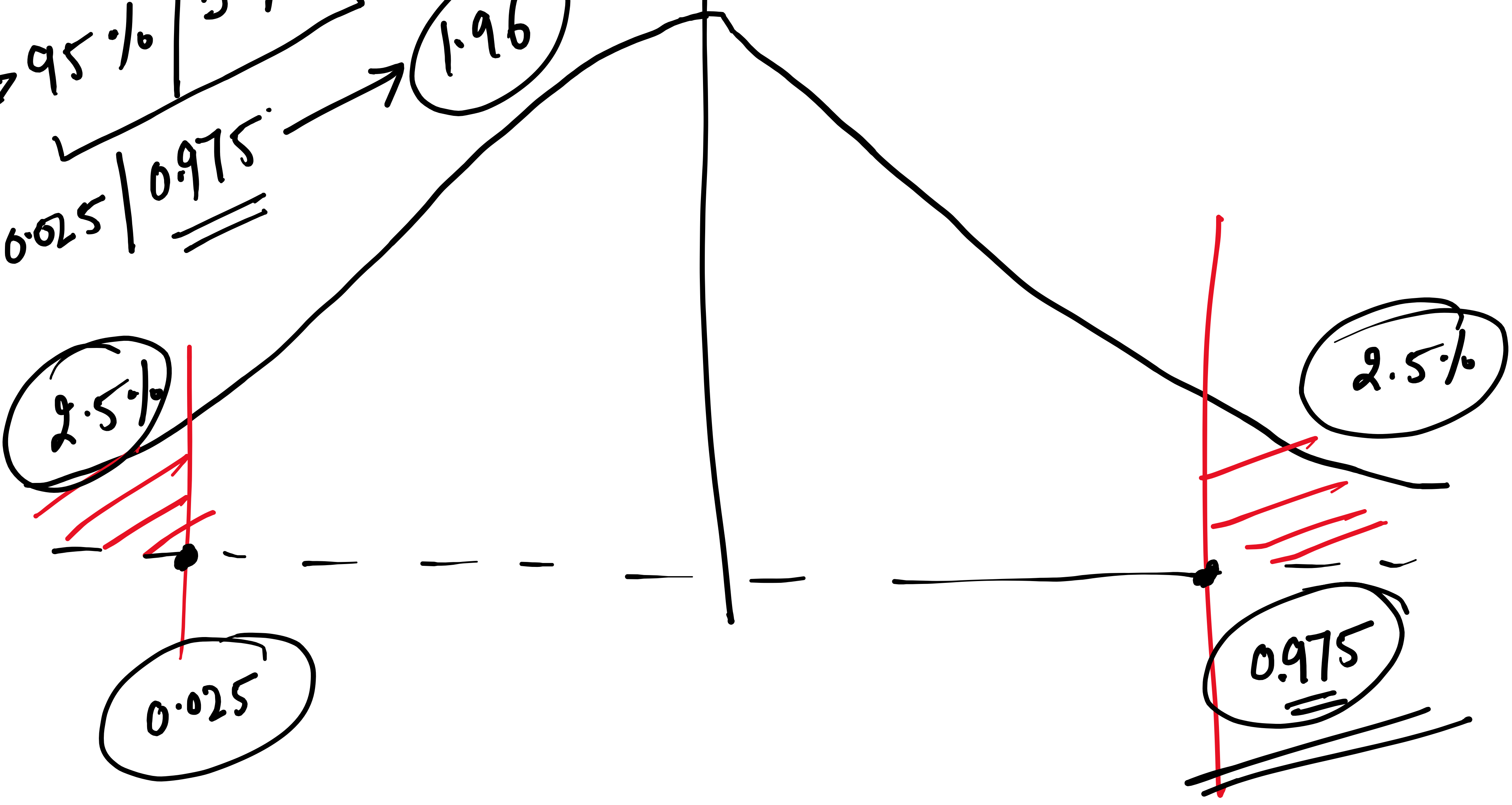
→ fail to reject null hypothesis

2 tail test

→ 95% | 5%

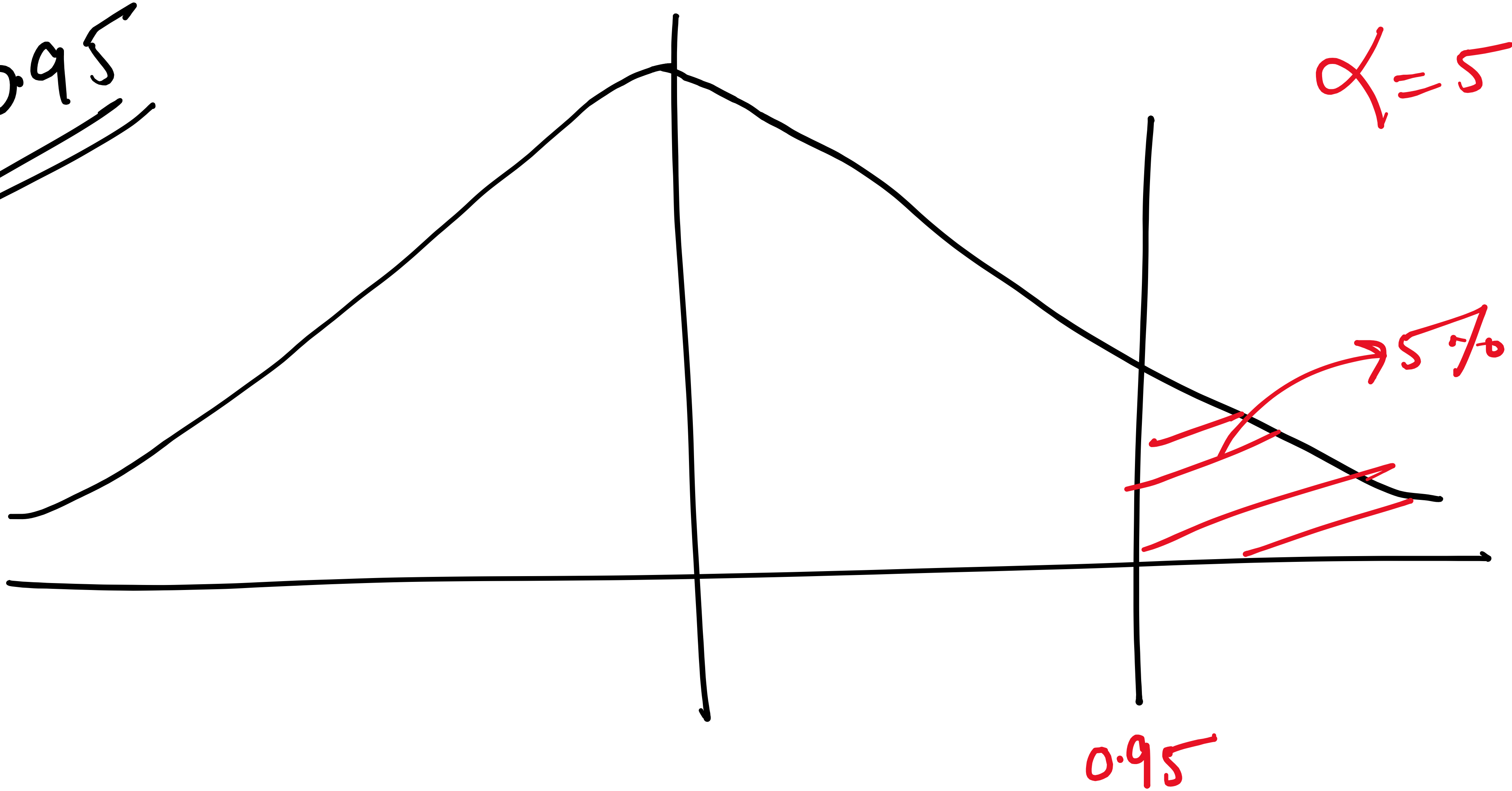
$Z = 0.025 \mid 0.975 \rightarrow 1.96$

5%



$Z = 095$

$\alpha = 5\% \,/\, 95\%$



→ LR

Let's create a numerical problem where you can calculate the R-squared value for a given line of best fit.

Problem Statement

Given the following dataset of X (independent variable) and Y (dependent variable):

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{\sum (y_a - y_p)^2}{\sum (y_a - \bar{y})^2}$$

X	Y
1	2
2	3
3	5
4	4
5	6

$y = 0.8x + 1.6$

→ Calculate R^2

$$y_p = 0.8x + 1.6$$

X	y_a	y_p	$y_a - y_p$	$(y_a - y_p)^2$	$(y_a - \bar{y})^2$
1	2	2.4	-0.4	0.16	4
2	3	3.2	-0.2	0.04	1
3	5	4.0	1	1	1
4	4	4.8	-0.8	0.64	0
5	6	5.6	.4	0.16	4
	$\Sigma y = 20$			Σa	Σb

$\bar{y} = 4$

$$\bar{y} = \frac{\Sigma y}{N} = \frac{20}{5} = 4$$

$$R^2 = 1 - \frac{a}{b}$$

$a = 2$

$b = 10$

$$R^2 = 1 - \frac{2}{10}$$

$= 80\%$

Que

2.71

Problem Statement

Given the following dataset of X (independent variable) and Y (dependent variable), where Y represents a binary outcome (0 or 1). Given a logistic regression model with the equation:

(Logistic Reg)

$$\left\{ P = \frac{1}{1+e^{-(ax+b)}} \right\}$$

Tasks

1.Calculate the probability P(Y=1|X) for X=2.0 using the ~~logistic~~ function.

2.Determine the predicted class for X=2.0 based on a threshold of 0.5.

$> 0.5 \rightarrow 1$

$\leq 0.5 \rightarrow 0$

$a = 1$

$b = -3$

X	Y
0.5	0
1.0	0
1.5	0
2	1
2.5	1
3	1
3.5	1
4	1

$$\left\{ P = \frac{1}{1+e^{-(2+(-3))}} = \frac{1}{1+e^{1}} \right.$$

$$\simeq \underline{0.2689} \} < 0.5$$

Predicted Class = 0

Confusion Matrix:-

$$
\begin{array}{cc|c|c|}
 & & \text{Predicted} & \\
 & & + & - \\
\hline
\text{Actual} & + & 100 & 30 \\
\hline
 & - & 20 & 150 \\
\hline
\end{array}
$$

* Accuracy
* Precision
* Recall
* F1-Score

TP = 100

TN = 150

FP = 20

FN = 30

$\rightarrow$ Accuracy $= \dfrac{TP + TN}{TP + TN + FP + FN} = \dfrac{100 + 150}{100 + 150 + 20 + 30} = \dfrac{250}{300}$

$= 83.33\%$

$\rightarrow$ Recall $= \dfrac{TP}{TP + FN} = \dfrac{100}{100 + 30} = \dfrac{100}{130}$

$\approx 76.92\%$

→ Precision $= \dfrac{TP}{TP+FP} = \dfrac{100}{100+20} = \dfrac{100}{120}$

$\simeq 83.33\%$

→ F1-Score $= \dfrac{2 \times P \times R}{(P+R)} = \dfrac{2 \times 0.83 \times .77}{.83+0.77}$

$= 79.8\%$

→ Calculate the VIF value for these features X1 & X2 having a R^2 of 92%. Also advise to keep these features or not?

* $VIF = \frac{1}{1-R^2} \rightarrow VIF = \frac{1}{1-0.92} = \frac{10}{0.08} = \frac{100}{8}$

$VIF = 12.5$ { Not to keep as $\underline{\underline{VIF \geq 5}}$ }