Statistic's

Module 3

1. AB Testing

→ A/B Testing, also known as split testing, is a research method that compares two versions of an element to determine which performs better.

- It's used in marketing, web development, and user experience (UX).
- A/B testing usually involves two varients (A and B), although the concept can be extended to multiple varients of the same variable.

2. Formulation of Null & alternate hypothesis.

- i) we have a medicine that is being manufactured and each pill is supposed to have 14 milligrams of the active ingredient. What are our null and alternate hypothesis?

→ $H_0\ \mu = 14mg$ $\quad H_a\ \mu \neq 14mg$

ii) The school principal wants to test if it is true what teacher says - that highschool juniors use the computer an average 3.2 hrs a day. what are our null and alternative hypothesis?

→ $H_0\ \mu = 3.2\ hrs$ $\quad H_a\ \mu \neq 3.2 hrs$

iii) A researcher claims that black horses are, on average, more than 30 lbs heavier than white horses, which average 1100 lbs. What is the null hypothesis, and what kind of test is this?

→ The null hypothesis would be notated $H_0: \mu \leq 1130\ lbs$. This is a right-tailed test, since the tail of the graph would be on the right. Recognize that values above 1130 would indicate that the null hypothesis be rejected.

iv) A package of gum claims that the flavor lasts more than 39 minutes. What would be the null hypothesis of a test to determine the validity of the claim? What sort of test is this?

→ The null hypothesis would be notated as $H0: \mu \leq 39$.
This is a right-tailed test, since the rejection region would consist of values greater than 39.

Q3 Permutation Test

→ A permutation test is a non-parametric test that determines the statistical significance of a model. The test involves shuffling observed data to determine how unusual an observed outcome is.

- The null hypothesis of a permutation test is that all samples come from the same distribution.
- The test statistic is calculated on the dataset and then for many random permutations of that data.

Q4 Type 1, Type 2 error's [sum or theory]

→ A Type I error (false-positive) occurs if an investigator rejects ~~a null hypothesis that is~~ investigator rejects a null a null hypothesis that is actually true in the population; a ~~hypothesis that is~~ type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population

	True	False
Rejects	Type 1 error	✓
Does not Rejects	✓	Type 2 error

Page No
Date

i) The owner of travel agency would like to determine whether or not the mean age of the agency's customers is over 24. If so, he plans to alter the destination of their special cruises and tours. If he concludes the mean age is over 24 when it is not, he makes a ______ error. If he concludes the mean age is not over 24 when it is, he makes a ______ error.

A Type II , Type II
B Type I , Type I
C Type I , Type II
✓D Type II , Type I

ii) Suppose we wish to test H: 53 vs H: 53 O μ ≤ a μ >. What will result if we conclude that the mean is greater than 53 when its true value is really 55?

A. We have made a Type I error
✓B. we have made a correct decision
C. we have made a Type II error
D. None of the above are correct

iii) A hypothesis test is used to prevent a machine from underfilling or overfilling quart bottles of beer. On the basis of sample, the machine is shut down for inspection. A through examination reveals there is nothing wrong with the filling machine. From a statistical point of view:

- Both Type I and Type II errors were made
- ✓ A Type I error was made
- A Type II error was made
- A correct decision was made

Page No.
Date

Q5 T-Test [Sum]

→ The manufacturer of certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a standard deviation of 5 months Random sample of 6 such bulbs gave the following values.

life of bulbs (in month)	$x-\bar{x}$	$(x-\bar{x})^2$
24	1	1
26	3	9
30	7	49
20	-3	9
20	-3	9
18	-5	25
$\bar{x} = 23$		$\Sigma(x-\bar{x})^2 = 102$

can you regard the producer's claim to be valid at 1% level of significance.

Sol^n. $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$= \sqrt{\frac{102}{5}}$

$= \sqrt{20.4}$

$s = 4.52$

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

$= \frac{|23-25|\sqrt{6}}{4.52}$

$= \frac{4.8989}{4.516}$

$t_{cal} = 1.084$

Calculate $t = 1.084$

$v = 5, \; t_{0.01} = 4.032$

1%

$t_{cal} < t_{table}$, then you will accept the hypothesis

Conclusion.

Calculated t-values less than the table t value ∴ Hypothesis accepted.

Therefore, the producers claim is not valid at 1% level of significance.

Page No
Date

Q6 Chi-square ~~sum~~ Test (sum)

→

$\chi^2_{0.05}$

chi-square

$v = n-1$

$v = 4-1$

$v = 3$

$\therefore \chi^2 = 9.488$

Q7 Fisher's exact test

→ Fisher's exact test is a statistical test that determines if there are nonrandom associations between two categorical variables. It's used in the analysis of contingency tables.

- It's generally used in one-tailed tests.
- Fisher's exact test calculates the ~~prob~~ probability of getting a table that is as strong due to the chance of sampling.

Q8 Multi-arm bandit algorithm

→ Multi-arm bandits offer an approach to testing, especially web testing, that allows explict optimization and more rapid decision making than the traditional statistical approach to designing experiments.

- It's commonly used in situations where you have multiple options and need to learn which one provides the best outcomes over time, such as in online advertising or recommendation systems.

Page No.
Date

9) Z - Test [sum]

→ X is a normally distributed variable with a mean of 30 and standard deviation as 4. Find probability of a dataset which is lesser than 40.

a) $P(x<40)$

b) $P(x>21)$

c) $P(30<x<35)$

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

a) $z = \frac{40-30}{4}$

$= \frac{10}{4} = 2.5$

$\Rightarrow 0.9938$ 99.38%

b) $P(x>21)$

$z = \frac{21-30}{4}$

$= \frac{-9}{4}$

$= -2.25$

0.01222 = -1.22%

$\therefore 100 - 1.22\%$

$= 98.78\%$

c) $P(30<x<35)$

$z = \frac{30-30}{4}$

$= 0$

$z = 0$

$\Rightarrow 0.5000$

$z = \frac{35-30}{4}$

$= \frac{5}{4} = 1.25$

$\Rightarrow 0.89435$

$$\begin{array}{r} 0.89435 \\ -0.5000 \\ \hline 0.3943 \end{array}$$

$\Rightarrow 39.43\%$