voters in city A. Use 5% level. (CO3-K4)

9. (a) The following data relate to the marks obtained by 11 students before and after intensive coaching . Do the data indicate that the students have benefited by the intensive coaching (CO3-K4)

Marks before coaching	19	23	16	24	17	18	20	18	21	19	20
Marks after coaching	17	24	20	24	20	22	20	20	18	22	19

OR

(b) In one sample of 8 observations the sum of squares of deviations of the sample values from the sample mean was 84.4 and in another sample of 10 observations it was 102.6 . Test the significant difference between the sample variances at 1% level of significance (CO3-K4)

10. (a) A sample of 48 tools produced by a machine shows the following sequence of good (G) and defective (D) tools :

G G G G G G D D G G G G G G G G G G D D D D G G G G G G D G G G G G G G G D D G G G G G D G G.

Test the randomness of the sequence at the 0.05 significance level (CO4-K4)

OR

(b) Find out whether the distributions of the two diets is same are not using median test. (CO4-K4)

Diet A: 16.3,10.1,10.7,13.5,14.9,11.8,14.3.10.2,12,14.7,23.6,15.1, 14.5,18.4,13.2,14,24.2,27.2,15.9,16.4,22.3

Diet B: 21.3,23.8,15.4,25.1,19.6,12.1,13.9,18.8,19.2,15.3,20.1, 14.8,18.9,20.7,21.1,26.4,18.6,15.8,16.2,17.8,23.6,24.3

***

Course Code				**23STMAP234**			
Title of the Course				**Inferential Statistics Lab**			
Offered to: (Program)				**B.Sc.(Hons) - Statistics**			
L	**0**	**T**	**0**	**P**	**2**	**C**	**1**
Year of Introduction:	**2024-25**			**Semester:**			**3**
Course Category:	**MAJOR**			**Course Relates to:**		**Local, Regional, National, Global**	
Year of Revision	**NA**			**Percentage:**		**NA**	
Type of the Course:				**SKILL DEVELOMENT**			
Crosscutting Issues of the Course:				**NA**			
Pre-requisites, if any				**23STMAL121, 23STMAL122**			

Course Description:

An inferential statistics lab course is designed to provide hands-on experience in applying statistical methods to real-world data. While the theoretical concepts are typically covered in a lecture-based course, the lab focuses on the practical implementation of these concepts using statistical software.

Course Aims and Objectives:

S. No	COURSE OBJECTIVES
1	**Develop proficiency in statistical software:** Students will become proficient in using statistical software packages (like SPSS, R, Python, or Excel) to perform complex statistical analyses
2	**Apply statistical concepts to real data:** Students will analyze real-world datasets to draw meaningful conclusions and make informed decisions
3	**Understand the assumptions and limitations of statistical tests:** Students will learn to critically evaluate the appropriateness of different statistical tests for given data and research questions
4	**Develop critical thinking and problem-solving skills:** Students will learn to interpret statistical results, identify potential issues, and communicate findings effectively.

Course Outcomes

At the end of the course, the student will be able to…

NO	COURSE OUTCOME	BTL	PO	PSO
CO1	apply statistical analysis that can test hypotheses under parametric approaches	**K3**	**2**	**2**
CO2	apply statistical analysis that can test hypotheses under non-parametric approaches.	**K3**	**2**	**2**
CO3	inferences for various non – parametric methods for Two samples using excel	**K4**	**2**	**2**
CO4	inferences of various large samples using Excel.	**K4**	**2**	**2**
CO5	inferences of various small samples	**K4**	**2**	**2**

For BTL: K1: Remember; K2: Understand; K3: Apply; K4: Analyze; K5: Evaluate; K6: Create

CO-PO-PSO MATRIX									
CO NO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**
CO1	3								3
CO2	3								3
CO3	3								3
CO4	3								3
CO5	3								3

Use the codes 3, 2, 1 for High, Moderate and Low correlation Between CO-PO-PSO respectively

Course Structure

Practical 1:Large sample test for single mean & difference of means

Applications

I. **Manufacturing:** Checking if the mean production rate of a machine is as expected
II. **Healthcare:** Evaluating if a new treatment improves the average patient outcome compared to a known standard
III. **Education:** Assessing if there's a significant difference in test scores between two teaching methods

Activity - 1

Real-world data collection

a) Collect data on a variable (e.g., height, weight, income) from a large
b) sample ($n > 30$).
c) Calculate the sample mean and standard deviation.
d) Formulate a hypothesis about the population mean.
e) Conduct a Z-test to test the hypothesis.
f) Interpret the results in context

Activity – 2

Comparative data collection

a) Collect data on two groups (e.g., males vs. females, treatment group
b) vs. control group) for a variable of interest.
c) Calculate sample means and standard deviations for both groups.
d) Formulate a hypothesis about the difference in population means.
e) Conduct a Z-test to test the hypothesis.

f) Interpret the results in context

Practical 2: Large sample test for single proportion & difference of proportions

Applications

I. **Market research:** Testing if the proportion of customers who prefer a new product exceeds a certain threshold

II. **Medical research:** Evaluating the effectiveness of a vaccine by comparing the proportion of vaccinated individuals who contract a disease to a historical control group

Activity 1

Real-world data collection:

a) Collect data on a categorical variable (e.g., success/failure, yes/no) from a large sample ($n > 30$).
b) Calculate the sample proportion.
c) Formulate a hypothesis about the population proportion.
d) Conduct a Z-test to test the hypothesis.
e) Interpret the results in context

Activity 2

Comparative data collection:

a) Collect data on a categorical variable for two groups (e.g., males vs. females, treatment group vs. control group).
b) Calculate sample proportions for both groups.
c) Formulate a hypothesis about the difference in population proportions.
d) Conduct a Z-test to test the hypothesis.
e) Interpret the results in context

Practical 3: Large sample test for difference of standard deviations

Applications

I. **Process Variability:** Comparing the variability of two production processes can help identify which process is more consistent.

II. **Data Reliability:** In experimental studies, comparing the variability of measurements can assess the reliability of the data collection method

Activity

a) **Data Collection and Visualization:**
b) Collect data on two different groups (e.g., heights of male and female students).
c) Calculate the mean and standard deviation for each group.
d) Visualize the data using box plots or histograms to compare the spread.
e) Discuss the implications of a larger or smaller standard deviation in this context.

Practical 4: Large sample test for correlation coefficient

Applications

I. **Education:** Investigating the relationship between student test scores and class size.

II. **Psychology:** Studying the correlation between IQ and income level.

Activity

Real-world Data Analysis

a) **Data collection:** Collect data on two variables from a large sample (e.g., height and weight of students, hours studied and exam scores).
b) **Data cleaning:** Ensure data is clean and free from outliers.
c) **Calculate correlation coefficient:** Use statistical software to calculate the correlation

coefficient (r).

d) **Hypothesis testing:**
Set up null and alternative hypotheses
Calculate the test statistic (t-value).
Determine the p-value.
Make a decision based on the p-value (reject or fail to reject H0).
Interpretation: Explain the meaning of the results in context

Practical 5: Small sample test for single mean & difference of means

Applications

I. **Quality control:** Testing the mean weight of a small batch of products
II. **Psychology:** Studying the cognitive abilities of a small sample of individuals

Activity 1

Simulation-based activities

a) Use statistical software to generate random samples from a normal distribution with a known mean.
b) Calculate sample means and conduct t-tests for different sample sizes.
c) Investigate the impact of sample size on the power of the test.
d) Explore Type I and Type II errors through simulations

Activity 2

a) Comparative data collection
b) Collect data on a variable for two small groups (e.g., treatment group vs. control group, male vs. female).
c) Calculate sample means and standard deviations for both groups.
d) Formulate a hypothesis about the difference in population means.
e) Conduct a t-test for independent samples to test the hypothesis.
f) Interpret the results in context.

Practical 6: Small sample test for correlation coefficient

Applications

- **Medical Research:** Investigating the correlation between a new drug's efficacy and patient age in a small clinical trial.
- **Psychology:** Studying the relationship between intelligence quotient (IQ) and creativity in a small group of individuals

Activity

a) Simulation Activity
b) **Generate random data:** Use statistical software to generate pairs of random data with different correlation coefficients (e.g., strong positive, weak negative, no correlation).
c) **Calculate correlation coefficients:** Calculate the correlation coefficient for each dataset.
d) **Hypothesis testing:** Conduct hypothesis tests for each correlation coefficient.
e) **Discussion:** Discuss the impact of sample size and the true population correlation on the results.

Practical 7:Paired t-test (paired samples).

Applications

Before-and-after designs:

- Evaluating the effectiveness of a treatment or intervention.

Examples: Weight loss programs, smoking cessation, educational programs.

Repeated measures designs:

- Analysing data collected over time from the same individuals.
  Examples: Tracking blood pressure changes over time, monitoring patient symptoms before, during, and after treatment

Activity 1

Simulation-based Learning

a) **Generate paired data:** Use statistical software to generate paired data with a specified mean difference.
b) **Conduct paired t-tests:** Perform paired t-tests on the generated data for different sample sizes.
c) **Explore power:** Investigate the impact of sample size and effect size on the power of the paired t-test.
d) **Type I and Type II errors:** Discuss the concepts of Type I and Type II errors in the context of paired t-tests

Activity 2

Real-world Data Collection and Analysis

a) **Collect paired data:** Students can collect paired data on themselves (e.g., heart rate before and after exercise, weight before and after a diet).
b) **Data analysis:** Calculate the difference scores, mean difference, and standard deviation of the differences.
c) **Hypothesis testing:** Formulate null and alternative hypotheses, calculate the t-statistic, and determine the p-value.
d) **Interpretation:** Draw conclusions about the effectiveness of the intervention or change.

Practical 8: Small sample test for single variance (χ^2 test)

Applications

a) **Quality control:** Testing if the variance of a manufacturing process meets specified quality standards.
b) **Finance:** Assessing the volatility (variance) of a financial asset compared to a benchmark.

Activity

Real-world Data Collection and Analysis

a) **Collect data:** Students can collect data on a variable from a small sample (e.g., reaction times, measurement errors).
b) **Calculate variance:** Calculate the sample variance.
c) **Hypothesis testing:** Formulate null and alternative hypotheses about the population variance.
d) **Chi-square test:** Conduct a chi-square test for a single variance.
e) **Interpretation:** Draw conclusions about the population variance based on the test results.

Practical 9:Small sample test for difference of variances (F test)

Applications

a) Assessing the consistency of two different measurement methods.
b) Comparing the variability of two manufacturing processes.

c) Evaluating the reliability of two different test instruments

Activity 1

a) Real-world Data Collection and Analysis

b) **Collect data:** Students can collect two sets of data from different groups (e.g., reaction times for two groups of people, measurement errors from two instruments).

c) **Calculate variances:** Calculate the sample variances for both groups.

d) **Hypothesis testing:** Formulate null and alternative hypotheses about the population variances.

e) **F-test:** Conduct an F-test to compare the variances.

f) **Interpretation:** Draw conclusions about the equality of population variances based on the test results.

Activity 2

Case Studies

a) **Provide real-world case studies:** Present case studies involving F-tests for comparing variances (e.g., manufacturing processes, measurement instruments).

b) **Data analysis:** Provide students with sample data and ask them to conduct an F-test.

c) **Interpretation:** Discuss the implications of the results in the context of the case study.

d) **Practical 10:** χ^2 test for goodness of fit and independence of attributes

e) Applications

f) **Market Research:** To determine if product preferences match historical data.

g) **Quality Control:** To check if production output conforms to expected distribution.

Practical 10: Chi- Square test for goodness of fit and independence of attributes

Applications

a) Market Research: To determine if product preferences match historical data.

b) Quality Control: To check if production output conforms to expected distribution.

Activity 1

Product Preference Survey

a) Conduct a survey asking people about their preferred brand of a product.

b) Compare the observed frequencies with the expected frequencies based on market share data.

c) Conduct a chi-square goodness of fit test to determine if the preferences match the market share.

Activity 2

Gender and Major

a) Collect data on the gender and major of a group of students.

b) Create a contingency table to display the observed frequencies.

c) Calculate the expected frequencies under the assumption of independence.

d) Conduct a chi-square test of independence to determine if there is a relationship between gender and major.

Practical 11: Nonparametric tests for single sample(run test, sign test and Wilcoxon signed rank test)

Applications

Run Test

a) Detecting randomness:

b) **Quality control:** To check if a production process

c) is in control.
d) **Financial analysis:** To analyse the randomness of stock price movements.
e) Sign Test
f) **Comparing median to a value:** Used to test if the median of a sample is different from a specified value.
g) **Before-after comparisons:** To analyse changes in paired data when the magnitude of change is not important.

Quality control: To assess if a process has shifted from a target value

a) Wilcoxon Signed-Rank Test
b) **Comparing paired data:** Used to compare paired observations when the data is not normally distributed.
c) **Before-after designs:** To analyze changes in paired data when the magnitude of change is important.
d) **Clinical trials:** To compare treatment effects between paired samples

Activity 1 Run Test

Coin Flipping Experiment: Toss a coin multiple times and record the outcomes as heads (H) or tails (T). Count the number of runs (consecutive sequences of the same outcome). Test if the sequence is
Random

Activity 2 Sign Test

Medication Effectiveness: Collect data on patients' pain levels before and after taking a new medication. Convert the differences to plus or minus signs based on improvement or worsening. Test if the median difference is different from zero.

Activity 3 Wilcoxon Signed-Rank Test

Weight Loss Program: Measure participants' weights before and after a weight loss program. Calculate the differences and rank their absolute values. Test if the median weight loss is different from zero

Practical 12: Nonparametric tests for two independent samples (Median test, wilcoxon – Mann- Whitney - U test, Wald - Wolfowitz' s runs test)

Applications

a) Median Test
b) **Comparing groups based on median:** Determines if two or more groups differ significantly in terms of their medians.
c) **Example:** Comparing the median income of different educational levels
d) Wilcoxon-Mann-Whitney U Test
e) **Comparing two independent groups:** Determines if there's a difference in distribution between two independent groups.
f) **Example:** Comparing the pain levels of two treatment groups
g) Wald-Wolfowitz Runs Test
h) **Testing for randomness:** Determines if a sequence of data is random.
i) **Example:** Analyzing the pattern of stock price increases and decreases

Activity 1 Median Test

Plant Growth Experiment: Measure the heights of plants under two different growing conditions. Calculate the median height for each group. Conduct a median test to compare plant growth

Activity 2 Wilcoxon-Mann-Whitney U Test

Drug Effectiveness: Collect data on pain levels for two groups of patients, one receiving a placebo and the other a new drug. Conduct a Wilcoxon-Mann-Whitney U test to compare pain levels between the groups.

Activity 3 Wald-Wolfowitz Runs Test

Quality Control: Inspect a sequence of products for defects.

Convert the data to a sequence of defective (D) and non-defective (N) items. Conduct a Wald-Wolfowitz runs test to check for randomness in the defect pattern.

Lab Manual:

1. Inferential Statistics using Excel – Prepared by Department of Statistics, PBSCAS

Reference Book:

Gips, M. A. (2015). *Mastering Excel: A Problem-Solving Approach* (2nd Edition). Wiley.

SEE (LAB) Model Question Paper

23STMAP234: Inferential Statistics Lab

Offered to B.Sc. Hons STATISTICS

Max. Marks: 50 **Max. Time: 3Hrs**

Pass. Min: 20

(A)	**Evaluation Procedure**	**35 Marks**
I	Experiments (Exam & Execution)	**30 Marks**
II	Viva	**3 Marks**
III	Record	**2 Marks**
(B)	**CONTINUOUS ASSESMENT(Internal)**	**15 MARKS**

15 marks for the continuous assessment (Day to day work in the laboratory shall be evaluated for 15 marks by the concerned laboratory teacher based on the regularity/ record/viva). Laboratory teachers are mandated to ensure that every student completes 80%-90% of the lab assessments.

TOTAL: (A)+(B) = 50 MARKS