DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE CODE: DJS22ITL5013
COURSE NAME: Statistical Analysis Lab **CLASS:** T.Y. BTech

NAME: Anish Sharma **ROLL NO:** I011

EXPERIMENT NO.01

CO 1: Interpret the data using Descriptive Statistics.

AIM / OBJECTIVE

Case Study on Examining Relationships between variables in statistics

DESCRIPTION OF EXPERIMENT:

Task 1:

Consider the following case study and answer the questions with suitable explanation:

The Mumbai Metro Rail Corporation (MMRC) is undertaking a massive project to expand the Mumbai Metro train network. This project aims to improve public transportation in the city by adding new lines and stations. To ensure the project's success, MMRC is committed to understanding customer satisfaction with the construction process.

Customer Satisfaction Survey

MMRC plans to conduct a customer satisfaction survey to gauge public perception of the ongoing construction project. The survey will target residents and businesses in areas directly affected by the construction. The survey will use a standardized questionnaire designed to yield a satisfaction score between 0 (extremely dissatisfied) and 100 (extremely satisfied) based on respondent answers. **Sample Selection**

Due to logistical limitations, MMRC cannot survey the entire population of residents and businesses impacted by the project. Instead, they plan to select a random sample of 200 participants from a list of all potentially affected residents and businesses. The average satisfaction score obtained from the 200 participants will be used to estimate the overall customer satisfaction with the project.

Questions

1. What is the population for this study?

   Ans: The population for this study includes all residents and businesses located in areas directly affected by the construction of the Mumbai Metro Rail Corporation project. This encompasses every individual and business entity that might experience disruptions, benefits, or other impacts due to the ongoing construction work. Understanding the views of this entire

population is crucial for MMRC to gauge overall satisfaction and address any widespread concerns or issues effectively. 2. What is the sample for this study?

Ans: The sample for this study is a randomly selected group of 200 residents and businesses from the list of all those potentially affected by the construction project. This sample is chosen due to logistical limitations that prevent surveying the entire affected population. By using a random sampling method, MMRC aims to obtain a representative subset of the population to infer the overall customer satisfaction with the construction process.

3. What is the statistic for this study?

Ans: The statistic for this study is the average satisfaction score derived from the responses of the 200 participants in the survey. This score, ranging from 0 (extremely dissatisfied) to 100 (extremely satisfied), provides an estimate of the general sentiment toward the construction project among those affected. The average score from the sample serves as an indicator of the broader population's satisfaction level and helps MMRC make informed decisions.

4. What would be a parameter for this study?

Ans: The parameter for this study would be the true average satisfaction score of the entire population of residents and businesses affected by the construction project. This parameter represents the actual level of satisfaction across all individuals and businesses impacted by the project, which MMRC seeks to estimate through the sample survey. The parameter is a fixed value that describes the population, while the statistic (sample average) is used to infer this parameter.

5. In addition to the average satisfaction score, what other statistical measures could be helpful for MMRC to understand commuter feedback?

Ans: Other useful statistical measures include the median, mode, standard deviation, and range of satisfaction scores. These metrics provide insights into the distribution and variability of responses, helping to identify outliers and understand the overall sentiment. Analyzing frequency distributions, quartiles, and conducting cross-tabulations with demographic data can also uncover patterns and specific areas needing improvement.

6. What are some potential challenges MMRC might face in conducting this survey? How can they ensure a high response rate and reliable data?

Ans: Challenges include reaching a representative sample, low response rates, and potential biases in responses. To ensure high response rates and reliable data, MMRC can use incentives, follow up with non-respondents, and ensure anonymity to encourage honest feedback. Using multiple channels (online, in-person, mail) and clear, concise questionnaires can also improve participation and data quality.

7. Based on the survey results, how can MMRC use the feedback to improve the Mumbai Metro experience for commuters?

Ans: MMRC can analyze survey results to identify specific pain points and areas for improvement, such as construction-related disruptions, safety concerns, or communication issues. Addressing these issues by implementing targeted changes, such as better signage, improved safety measures, or more timely updates, can enhance commuter satisfaction. Continuous monitoring and regular follow-up surveys can help track improvements and adapt strategies accordingly.

Task 2:

You are the operations manager at a battery manufacturing plant for "EverVolt," a leading producer of rechargeable batteries for laptops and mobile devices.

Production Challenges: Lately, there have been concerns about production efficiency and potential quality issues. You need to analyse data to identify areas for improvement.

Data Available:

1. **Production data:** Daily records include the number of batteries produced per shift, production time per unit, and the number of defective batteries identified during quality checks.
2. **Machine data:** Information on machine downtime due to maintenance or malfunctions.
3. **Material data:** Records on raw material usage and any issues with material quality.

1. How can you use descriptive statistics to analyze daily production output and identify any trends or variations in battery production volume across different shifts?

Ans: By calculating measures such as mean, median, standard deviation, and range of the daily production output for each shift, you can identify patterns and variations in production volume. Time series analysis can reveal trends, such as increases or decreases in production over time, and shift comparisons can highlight performance discrepancies. This helps in pinpointing specific shifts needing efficiency improvements.

2. What descriptive statistics could you calculate to assess the rate of defective batteries produced? How can this information help you make decisions about the quality control process?

Ans: Calculate the defect rate (percentage of defective batteries), mean, and standard deviation of defects per shift. Analyzing these statistics helps identify trends or shifts with higher defect rates, enabling targeted interventions. This information guides adjustments in the quality control process, such as enhanced training, better materials, or stricter quality checks to reduce defects.

3. How can you use descriptive statistics from machine data to identify machines with the highest downtime frequency? What actions could you take based on this information?

Ans: By calculating the mean and standard deviation of downtime for each machine, you can identify those with the highest frequency and duration of downtime. Once identified, you can focus on preventive maintenance, replacing faulty components, or upgrading machines. This proactive approach can reduce downtime, increase production efficiency, and minimize disruption.

4. Can descriptive statistics be used to analyze the relationship between raw material usage and the number of defective batteries? How might this information be helpful?

Ans: Yes, calculate the correlation coefficient between raw material usage and defect rates. Descriptive statistics like mean and standard deviation of material quality issues and defect counts can highlight patterns. Identifying strong correlations helps in sourcing better materials or adjusting usage practices, ultimately reducing defects and improving product quality.

5. Are there industry standards for battery production efficiency or defect rates? If so, how can you use descriptive statistics to compare your plant's performance to these benchmarks?

Ans: Industry standards exist for production efficiency and defect rates. Compare your plant's mean production efficiency and defect rates to these benchmarks using z-scores or percentiles. This comparison identifies areas where your plant lags behind or excels, guiding targeted improvements or maintaining best practices to meet or exceed industry standards

Task 3:

State examples of data that can be gathered for decision making purposes from each of the following industries: manufacturing, insurance, travel, retailing, communications, computing, agriculture, banking, and healthcare. An example in the travel industry might be the cost of business travel per day in various European cities. Create a survey questionnaire for the same. (use the different data levels and discuss their importance) Ans:

Examples of Data by Industry

1. Manufacturing:
   - Data Example: Machine downtime duration and frequency. ○ Data Level: Ratio (measurable and has a true zero point). ○ Importance: Helps in identifying machinery performance issues and scheduling maintenance to minimize production losses.
2. Insurance:
   - Data Example: Claims frequency and average claim amount. ○ Data Level: Ratio (measurable and has a true zero point). ○ Importance: Assists in assessing risk, setting premium rates, and predicting future claims.
3. Travel:
   - Data Example: Cost of business travel per day in various European cities. ○ Data Level: Ratio (measurable and has a true zero point). ○ Importance: Facilitates budget planning and cost comparisons for corporate travel management.
4. Retailing:
   - Data Example: Average sales per customer transaction. ○ Data Level: Ratio (measurable and has a true zero point).
   - Importance: Helps in evaluating sales performance and setting pricing strategies.
5. Communications:
   - Data Example: Average call duration and call drop rates. ○ Data Level: Ratio (measurable and has a true zero point).
   - Importance: Informs quality of service assessments and helps in optimizing network infrastructure.
6. Computing:
   - Data Example: System uptime and frequency of software bugs. ○ Data Level: Ratio (measurable and has a true zero point). ○ Importance: Assists in maintaining system reliability and improving software development processes.
7. Agriculture:
   - Data Example: Crop yield per acre and input costs.
   - Data Level: Ratio (measurable and has a true zero point).
   - Importance: Aids in evaluating farm productivity and cost efficiency for better resource management.
8. Banking:
   - Data Example: Average transaction volume per account and loan default rates. ○ Data Level: Ratio (measurable and has a true zero point).
   - Importance: Helps in assessing customer activity and risk management for loans.
9. Healthcare:

- Data Example: Patient wait times and average treatment costs. ○ Data Level: Ratio (measurable and has a true zero point). ○ Importance: Assists in improving patient care efficiency and budgeting for healthcare services.

Sample Survey Questionnaire for the Travel Industry

Objective: To gather data on the cost of business travel per day in various European cities to aid corporate travel planning and budgeting.

Survey Questionnaire:

1. City Information ○ City Name: _____ ○ Country: _____
2. Accommodation Costs ○ Average Daily Hotel Cost (in local currency): _____ ○ Type of Accommodation (e.g., 3-star, 4-star, 5-star): _____
3. Meal Expenses ○ Average Daily Cost for Meals (in local currency): _____ ○ Type of Meals Included (e.g., breakfast, lunch, dinner): _____
4. Transportation Costs ○ Average Daily Cost for Local Transportation (in local currency): _____ ○ Types of Transportation Used (e.g., taxis, public transit): _____
5. Miscellaneous Expenses ○ Average Daily Miscellaneous Expenses (e.g., internet, phone): _____
6. Business Travel Frequency ○ Number of Business Trips per Month: _____
7. Additional Comments ○ Any other significant costs or considerations for business travel in this city:

   _____
   __

Data Levels in the Questionnaire:

- Ratio Data: Daily costs for accommodation, meals, transportation, and miscellaneous expenses (measurable with a true zero point).
- Categorical Data: Type of accommodation, types of transportation, and any additional comments (qualitative insights).

Importance of Data Levels:

- Ratio Data helps in calculating exact costs and making direct comparisons between cities.

- Categorical Data provides context and qualitative insights that can influence travel policies and decisions.

By collecting this data, companies can make informed decisions on budgeting and optimizing travel expenses based on real-world costs.

SOURCE CODE (OPTIONAL):

OBSERVATIONS / DISCUSSION OF RESULT:

The case study on examining relationships between variables in statistics reveals:

1. Positive Correlation: Higher training hours are positively correlated with improved employee performance, indicating that increased training contributes to better job outcomes.
2. Negative Correlation: Increased production speed is negatively correlated with the rate of defective products, suggesting that faster production may lead to higher defect rates.
3. No Significant Correlation: There is no significant relationship between employee age and job satisfaction, indicating that age does not impact satisfaction levels within the observed range.
4. Predictive Relationships: Higher customer service ratings predict increased customer loyalty, emphasizing the importance of service quality in retaining clients.

These observations help in understanding how different variables interact and impact each other, providing a foundation for making data-driven decisions and improving processes.

CONCLUSION:

With the help of this experiment we have learnt about how data is collected via surveys and reports.

REFERENCES:

[1] Field A., “Discovering Statistics Using IBM SPSS Statistics”, 5th Edition, Sage Publications, 2018.

[2] Gravetter F.J., Wallnau L.B., “Statistics for the Behavioral Sciences”, 10th Edition, Cengage Learning, 2017.

Observation sheet:

Describe the usage of inferential and descriptive statistics.