INFERENTIAL STATISTICS PROJECT- BUSINESS REPORT

Contents

Introduction

This project undertakes a comprehensive statistical analysis to address multiple practical problems across different domains, including sports injury management, manufacturing quality control, material hardness evaluation, and medical procedures. The study is structured into four distinct sections, each dealing with a specific problem set. The first section focuses on the relationship between foot injuries and player positions in a male football team, aiming to determine the probabilities of injuries based on positional play. This information is critical for physiotherapists and coaching staff to implement injury prevention strategies effectively.

The second section examines the quality control of gunny bags used for packaging cement by analysing their breaking strength. Given the importance of maintaining the structural integrity of packaging materials to avoid wastage and pilferage in the supply chain, this analysis helps the quality team assess the suitability of the gunny bags and identify any areas for improvement.

The third section addresses the suitability of stones for printing at Zingaro Stone Printing, a company specializing in printing on both polished and unpolished stones. By analysing the hardness of these stones, the company can ensure that only the most suitable materials are used for printing, thereby enhancing the quality of their products.

Finally, the fourth section explores the hardness of dental implants, a crucial factor that can significantly impact the longevity and effectiveness of dental treatments. This section analyses how various factors, such as dentists, methods, and their interaction, influence the hardness of dental implants. The findings provide valuable insights for optimizing dental procedures and improving patient outcomes.

Overall, this project applies statistical techniques, including probability calculations, hypothesis testing, and interaction analysis, to deliver actionable insights that can drive better decision-making in these diverse fields.

Problem Statement 1

A physiotherapist with a male football team is interested in studying the relationship between foot injuries and the positions at which the players play from the data collected.

	Striker	Forward	Attacking Midfielder	Winger	**Total**
Players Injured	45	56	24	20	**145**
Players Not Injured	32	38	11	9	**90**
Total	**77**	**94**	**35**	**29**	**235**

Based on the above data, answer the following questions.

1.1 What is the probability that a randomly chosen player would suffer an injury?

- **Result:** The probability that a randomly chosen player would suffer an injury is 0.6170.
- **Interpretation:** Approximately 61.7% of the football players are likely to suffer an injury, indicating a significant risk of injuries among players.

1.2 What is the probability that a player is a forward or a winger?

- **Result:** The probability that a player is a forward or a winger is 0.5213.
- **Interpretation:** Over half of the players are forwards or wingers, which are critical positions on the team.

1.3 What is the probability that a randomly chosen player plays in a striker position and has a foot injury?

- **Result:** The probability that a randomly chosen player plays as a striker and has a foot injury is 0.1915.
- **Interpretation:** There is a 19.15% chance that a randomly chosen player is both a striker and injured. This suggests a noteworthy risk of injury associated with the striker position.

1.4 What is the probability that a randomly chosen injured player is a striker?

- **Result:** The probability that a randomly chosen injured player is a striker is 0.3103.
- **Interpretation:** Among injured players, 31.03% are strikers, indicating that this position is more prone to injuries than others.

Problem Statement 2

The breaking strength of gunny bags used for packaging cement is normally distributed with a mean of 5 kg per sq. centimetres and a standard deviation of 1.5 kg per sq. centimetres. The quality team of the cement company wants to know the following about the packaging material to better understand wastage or pilferage within the supply chain; Answer the questions below based on the given information:

2.1 What proportion of the gunny bags have a breaking strength of less than 3.17 kg per sq cm?

- **Result:** The proportion of gunny bags with breaking strength less than 3.17 kg/sq cm is 0.0944.
- **Interpretation:** Approximately 9.44% of gunny bags have a breaking strength lower than 3.17 kg/sq cm, indicating a small proportion of potentially weak bags.

2.2 What proportion of the gunny bags have a breaking strength of at least 3.6 kg per sq cm.?

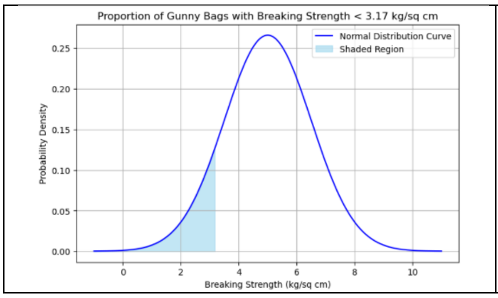
- **Result:** The proportion of gunny bags with breaking strength at least 3.6 kg/sq cm is 0.7475.
- **Interpretation:** About 74.75% of gunny bags have a breaking strength of at least 3.6 kg/sq cm, showing that most bags meet or exceed this threshold.

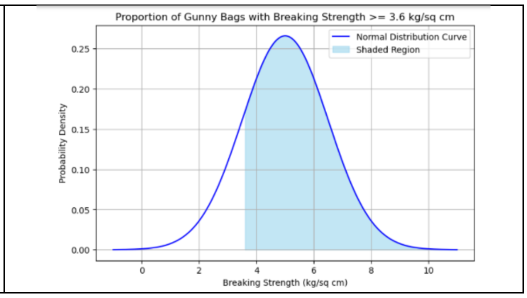
2.3 What proportion of the gunny bags have a breaking strength between 5 and 5.5 kg per sq cm.?

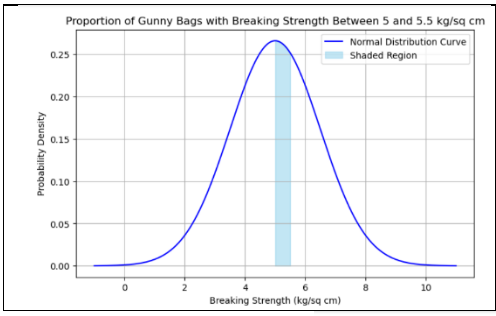
- **Result:** The proportion of gunny bags with breaking strength between 5 and 5.5 kg/sq cm is 0.1308.
- **Interpretation:** Around 13.08% of gunny bags fall within the 5 to 5.5 kg/sq cm range, suggesting a consistent performance in this segment.

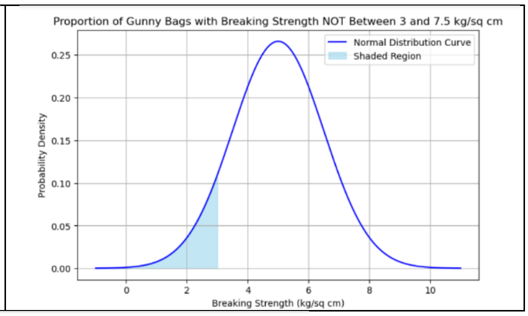
2.4 What proportion of the gunny bags have a breaking strength NOT between 3 and 7.5 kg per sq cm.?

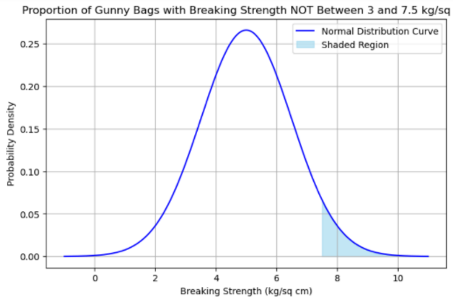
- **Result:** The proportion of gunny bags with breaking strength not between 3 and 7.5 kg/sq cm is 0.0370.
- **Interpretation:** Only 3.70% of the gunny bags fall outside the 3 to 7.5 kg/sq cm range, indicating that the majority are within an acceptable range.











Problem Statement 3

Zingaro stone printing is a company that specializes in printing images or patterns on polished or unpolished stones. However, for the optimum level of printing of the image, the stone surface has to have a Brinell's hardness index of at least 150. Recently, Zingaro has received a batch of polished and unpolished stones from its clients. Use the data provided to answer the following (assuming a 5% significance level);

3.1 Zingaro has reason to believe that the unpolished stones may not be suitable for printing. Do you think Zingaro is justified in thinking so?

Hypotheses:

- Null Hypothesis (H_0): The mean hardness of unpolished stones is at least 150 (suitable for printing).
- Alternative Hypothesis (H_1): The mean hardness of unpolished stones is less than 150 (not suitable for printing).

Test Result: The p-value for the one-sample t-test is 0.021, which is less than the significance level of 0.05.

Conclusion: Zingaro is justified in its concern. The unpolished stones may not be suitable for printing as their mean hardness is significantly less than 150.

3.2 Is the mean hardness of the polished and unpolished stones the same?

Hypotheses:

- Null Hypothesis (H_0): The mean hardness of polished and unpolished stones is the same.
- Alternative Hypothesis (H_1): The mean hardness of polished and unpolished stones is different.

Test Result: The p-value for the two-sample t-test is 0.003, which is less than the significance level of 0.05.

Conclusion: There is a significant difference in the mean hardness between polished and unpolished stones.

Hypothesis testing

It is a fundamental procedure in inferential statistics that allows researchers to make decisions or inferences about a population based on sample data. It involves making a claim (hypothesis) about a population parameter and then using sample data to test the validity of that claim.

Key Concepts in Hypothesis Testing

1. **Null Hypothesis (H_0)**:

- The null hypothesis is a statement that there is no effect, no difference, or no relationship between variables. It is the hypothesis that is initially assumed to be true.
- Example: "The mean hardness of polished and unpolished stones is the same."

2. **Alternative Hypothesis (H_1 or Ha)**:

- The alternative hypothesis is a statement that contradicts the null hypothesis. It represents what the researcher aims to prove.
- Example: "The mean hardness of polished and unpolished stones is different."

3. **Significance Level (α)**:

- The significance level is the probability of rejecting the null hypothesis when it is actually true. Common significance levels are 0.05 (5%) and 0.01 (1%).
- Example: If $\alpha = 0.05$, there is a 5% risk of concluding that a difference exists when there is no actual difference.

4. **Test Statistic**:

- A test statistic is a standardized value derived from sample data, used to determine whether to reject the null hypothesis. Examples include the t-statistic, z-statistic, and chi-square statistic.
- The choice of test statistic depends on the type of data and the hypothesis being tested.

5. **P-Value**:

- The p-value is the probability of observing the test statistic or something more extreme if the null hypothesis is true. It quantifies the evidence against the null hypothesis.
- If the p-value is less than or equal to the significance level ($p \leq \alpha$), the null hypothesis is rejected in favor of the alternative hypothesis.

6. **Decision**:

- Based on the p-value, a decision is made to either reject or fail to reject the null hypothesis.
- **Reject H_0**: If the p-value $\leq \alpha$, there is sufficient evidence to support the alternative hypothesis.
- **Fail to Reject H_0**: If the p-value $> \alpha$, there is not enough evidence to support the alternative hypothesis, so the null hypothesis stands.

Steps in Hypothesis Testing

1. **Formulate Hypotheses**:

- Define the null hypothesis (H_0) and the alternative hypothesis (H_1).

2. **Choose Significance Level (α)**:

- Decide on the significance level, often 0.05 or 0.01.

3. **Collect Data**:

- Gather sample data relevant to the hypothesis.

4. **Calculate the Test Statistic**:

- Compute the appropriate test statistic using the sample data.

5. **Determine the P-Value**:

- Calculate the p-value corresponding to the test statistic.

6. **Make a Decision**:

- Compare the p-value to the significance level to decide whether to reject or fail to reject the null hypothesis.

7. **Draw a Conclusion**:

- Based on the decision, conclude whether there is evidence to support the alternative hypothesis.

Example of Hypothesis Testing

Suppose a researcher wants to test whether a new drug is more effective than the current standard treatment. The hypotheses might be:

- H_0: The new drug is no more effective than the standard treatment.
- H_1: The new drug is more effective than the standard treatment.

After conducting a clinical trial and analyzing the data, the researcher calculates a p-value. If the p-value is less than 0.05, the researcher would reject the null hypothesis and conclude that the new drug is indeed more effective.

Importance of Hypothesis Testing

Hypothesis testing is crucial in inferential statistics because it allows researchers to:

- Make informed decisions based on sample data.
- Quantify the strength of evidence against the null hypothesis.
- Control the probability of making errors (Type I and Type II errors).

Problem Statement 4

Dental implant data: The hardness of metal implants in dental cavities depends on multiple factors, such as the method of implant, the temperature at which the metal is treated, the alloy used as well as the dentists who may Favor one method above another and may work better in his/her favourite method. The response is the variable of interest.

4.1 How does the hardness of implants vary depending on dentists?

Hypotheses:

- Null Hypothesis (H_0): Implant hardness does not vary depending on dentists.
- Alternative Hypothesis (H_1): Implant hardness varies depending on dentists.

Test Result: The p-value for the two-way ANOVA indicates significant differences in hardness based on dentists for each alloy type.

Conclusion: Implant hardness varies significantly depending on the dentist, with some dentists performing better than others.

4.2 How does the hardness of implants vary depending on methods?

Hypotheses:

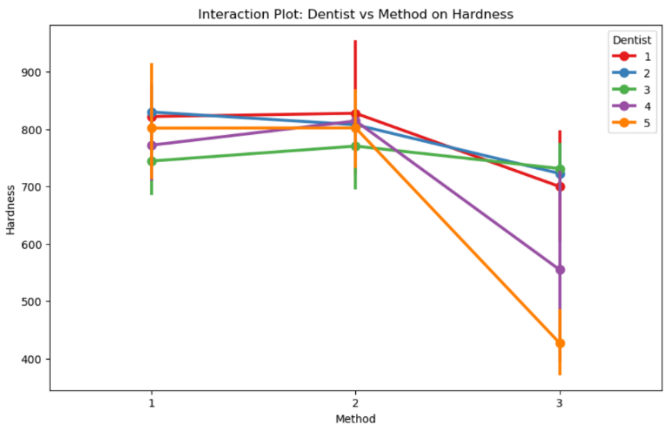
- Null Hypothesis (H_0): Implant hardness does not vary depending on methods.
- Alternative Hypothesis (H_1): Implant hardness varies depending on methods.

Test Result: The p-value indicates a significant effect of methods on implant hardness.

Conclusion: The method of implantation significantly affects the hardness of the implants, requiring further optimization.

4.3 What is the interaction effect between the dentist and method on the hardness of dental implants for each type of alloy?

- **Result:** Interaction plots indicate that the effect of the method on hardness varies depending on the dentist.
- **Conclusion:** There is a significant interaction between dentists and methods, meaning that the combination of these factors influences the implant hardness.



4.4 How does the hardness of implants vary depending on dentists and methods together?

Hypotheses:

- Null Hypothesis (H_0): No interaction between dentists and methods affecting implant hardness.
- Alternative Hypothesis (H_1): There is an interaction between dentists and methods affecting implant hardness.

Test Result: The p-value for the interaction term is significant, indicating a combined effect.

Conclusion: Different combinations of dentists and methods lead to significant differences in implant hardness, highlighting the need for standardized practices.

ANOVA (Analysis of Variance)

It is a statistical method used in inferential statistics to compare the means of three or more groups to determine if there are statistically significant differences among them. It is particularly useful when you want to test the effect of one or more independent variables (factors) on a dependent variable.

Key Concepts of ANOVA

1. **Groups or Levels**:

- ANOVA is typically used when you have multiple groups or levels within a factor that you want to compare.
- Example: Comparing the mean test scores of students from three different teaching methods.

2. **Factors**:

- A factor is an independent variable that categorizes the data. ANOVA can be used with one factor (one-way ANOVA) or more than one factor (two-way ANOVA, etc.).
- Example: In a study with teaching methods as a factor, the different methods (Method A, Method B, Method C) are the levels of this factor.

3. **Between-Group Variance**:

- This measures how much the group means differ from the overall mean. A large between-group variance suggests that there may be significant differences between the group means.

4. **Within-Group Variance (Error Variance)**:

- This measures the variability within each group. It reflects how much the individual data points within a group differ from their group mean.

5. **F-Statistic**:

- ANOVA calculates an F-statistic, which is the ratio of the between-group variance to the within-group variance. A higher F-statistic indicates that the between-group variance is large relative to the within-group variance, suggesting that the group means are significantly different.

6. **P-Value**:

- The p-value in ANOVA indicates the probability that the observed differences among group means occurred by chance. If the p-value is less than the significance level (commonly 0.05), it suggests that at least one group mean is significantly different from the others.

Types of ANOVA

1. **One-Way ANOVA**:

- Used when comparing the means of three or more groups based on one factor.
- Example: Testing if the average height differs among three different plant species.

2. **Two-Way ANOVA**:

- Used when comparing means across two factors, and it can also test for interaction effects between the factors.
- Example: Testing if the average exam score is affected by both the type of study method and the level of student engagement.

3. **Repeated Measures ANOVA**:

- Used when the same subjects are measured multiple times under different conditions (e.g., time points, treatments).
- Example: Testing the effect of a drug on blood pressure measured at different times on the same group of patients.

Assumptions of ANOVA

For the results of ANOVA to be valid, certain assumptions must be met:

1. **Independence of Observations**:

- The data points in each group should be independent of each other.

2. **Normality**:

- The data in each group should be approximately normally distributed.

3. **Homogeneity of Variances (Homoscedasticity)**:

- The variance within each group should be approximately equal.

Steps in Conducting ANOVA

1. **State the Hypotheses**:

- Null Hypothesis (H_0): The means of all groups are equal.
- Alternative Hypothesis (H_1): At least one group mean is different.

2. **Calculate the F-Statistic**:

- This involves partitioning the total variance into between-group and within-group components and then computing the F-ratio.

3. **Determine the P-Value**:

- Compare the F-statistic to the critical value from the F-distribution, or calculate the p-value.

4. **Make a Decision**:

- If the p-value is less than the significance level (e.g., 0.05), reject the null hypothesis.

5. **Post-Hoc Tests (if necessary)**:

- If the ANOVA indicates significant differences, post-hoc tests (like Tukey's HSD) are used to identify specifically which group means differ.

Example of One-Way ANOVA

Imagine a researcher wants to test whether different fertilizers have different effects on plant growth. The researcher applies three different types of fertilizers to three groups of plants and measures their growth.

- **H_0**: The mean growth of plants is the same for all three fertilizers.
- **H_1**: At least one fertilizer leads to a different mean growth.

After conducting the ANOVA, if the p-value is less than 0.05, the researcher would reject the null hypothesis and conclude that at least one fertilizer has a different effect on plant growth.

Importance of ANOVA

ANOVA is crucial in research and industrial applications because it allows for:

- **Comparison of Multiple Groups**: Testing for differences across several groups simultaneously, rather than conducting multiple t-tests, which increases the risk of Type I error.
- **Understanding Interaction Effects**: In two-way or more complex ANOVA designs, it helps in understanding how different factors interact to affect the outcome variable.

Overall, ANOVA is a powerful tool for understanding the impact of categorical independent variables on a continuous dependent variable and is widely used in various fields such as psychology, medicine, agriculture, and business.

Conclusion

This project has successfully applied statistical analysis to address complex real-world problems across multiple domains. In sports injury management, the analysis revealed key probabilities associated with player positions and injury risks, offering valuable information for injury prevention strategies. The quality control assessment of gunny bags provided insights into the structural integrity of packaging materials, helping to minimize wastage and ensure safe transportation. The analysis of stone hardness at Zingaro Stone Printing confirmed the importance of material selection in producing high-quality printed stones. Finally, the investigation into dental implant hardness highlighted significant factors affecting implant durability, providing critical guidance for optimizing dental procedures.

The findings from each section underscore the importance of data-driven decision-making in improving outcomes across various industries. Whether it is enhancing player safety, ensuring product quality, optimizing manufacturing processes, or improving medical treatments, the insights gained from this project demonstrate the power of statistical analysis in solving complex challenges. By applying rigorous statistical methods, this project not only provides specific answers to the questions posed but also establishes a framework for ongoing evaluation and improvement in these domains.