

Statistical Data Analysis Project

Alexandre De Cuyper

University of A Coruña

February 22, 2024

Introduction

- ▶ Brief overview of the data analysis project.
- ▶ Importance of the analysis: Understanding the shift in peak positions in crystals of different compositions.
- ▶ Objectives and hypotheses: Investigating if there are significant differences in peak positions among crystals.

Data Overview

- ▶ Overview of the dataset: Peaks from different crystals with compositions and varying 2 theta values.
- ▶ Key variables and their significance:
'2_theta', Condition, Value, and Cluster.

Data Preprocessing

- ▶ Cleaning and handling missing values.
- ▶ Transformation of data for analysis.

```
1 # Cleaning and transformation
2 data_clean <- na.omit(data_peaks)
3 data_long <- pivot_longer(data_clean, cols = -c('2
  _theta', Condition),
4                               names_to = "Cluster",
  values_to = "Value")
5
```

Exploratory Data Analysis

- ▶ Visualizations and summary statistics.
- ▶ Identification of patterns and trends.

```
1  # Exploratory Data Analysis (EDA)
2  ggplot(data_long, aes(x = '2_theta', y = Value,
3                        color = Condition)) +
4      geom_point() +
5      labs(title = "Exploratory Data Analysis") +
6      theme_minimal()
```

ANOVA Testing

- ▶ Analysis of Variance (ANOVA) to test hypotheses.
- ▶ Results and interpretation.

```
1 # ANOVA Testing
2 anova_result <- aov('2_theta' ~ Condition *
Cluster, data = data_long)
3 summary(anova_result)
4
```

Results and Discussion

- ▶ Interpretation of ANOVA results.
- ▶ Comparison of clusters: Assessing the significance of peak position variations.
- ▶ Implications of the findings: Understanding how composition affects peak positions.

Conclusion

- ▶ Summary of key findings.
- ▶ Limitations and areas for future research.

Any Questions?