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: '$2_t heta$', *Condition*, *Value*, *andCluster*.

Data Preprocessing

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

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

```

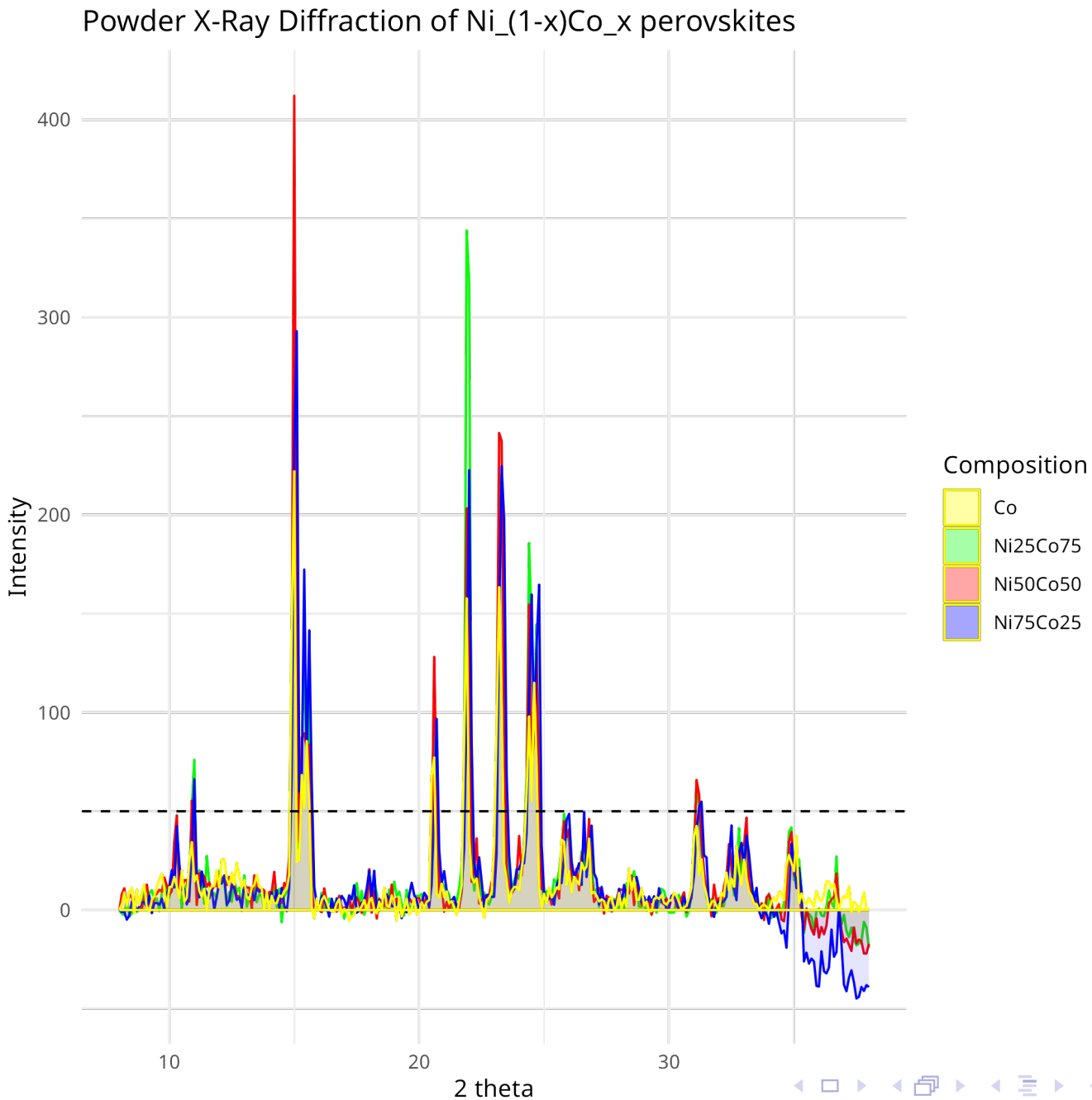
Exploratory Data Analysis

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

```
  # Exploratory Data Analysis (EDA)
  ggplot(data, aes(x = `2_theta`)) +
geom_ribbon(aes(ymin = 0, ymax = `Ni25Co75`, fill = "Ni25Co75"),
  alpha = 0.1, color = "green"
) +
geom_ribbon(aes(ymin = 0, ymax = `Ni50Co50`, fill = "Ni50Co50"),
  alpha = 0.1, color = "red"
) +
geom_ribbon(aes(ymin = 0, ymax = `Ni75Co25`, fill = "Ni75Co25"),
  alpha = 0.1, color = "blue"
) +
geom_ribbon(aes(ymin = 0, ymax = `Co`, fill = "Co"),
  alpha = 0.1, color = "yellow"
) +
labs(
  title = "Powder X-ray Diffraction of Ni_(1-x)Co_x perovskites",
  x = "2 theta",
  y = "Intensity",
  fill = "Composition"
) +
scale_fill_manual(values = c(
  "Ni75Co25" = "blue", "Ni50Co50" = "red",
  "Ni25Co75" = "green", "Co" = "yellow"
)) +
theme_minimal()

```

Visualization of X-ray Diffraction



ANOVA Testing

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

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

```

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?