

Analysis of Lead Exposure and IQ Levels

Data Analysis Report

2024-10-06

Contents

Introduction	1
Data Loading and Initial Processing	1
Initial Analysis	1
Corrected Analysis	3
Results and Discussion	5

Introduction

This report analyzes the relationship between proximity to a lead-emitting ore smelter and IQ scores in children from El Paso. The data comes from a CDC-led research team investigating the effects of low-level lead absorption on neurological function in children aged 3-15 years.

Data Loading and Initial Processing

```
# Create the dataset
data <- read_csv("lead-iq-01.csv")
# Calculate initial summary statistics

means_by_location <- data %>%
  group_by(Smelter) %>%
  summarise(
    mean_iq = mean(IQ),
    sd_iq = sd(IQ),
    n = n()
  )
```

Initial Analysis

Visualization of IQ Levels by Location

```
ggplot(data, aes(x = Smelter, y = IQ)) +
  geom_boxplot(fill = "lightblue", alpha = 0.5) +
  geom_jitter(width = 0.2, alpha = 0.5, color = "darkblue") +
  theme_minimal() +
  labs(
    title = "Distribution of IQ Scores by Proximity to Smelter",
    x = "Distance from Smelter",
    y = "IQ Score"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title = element_text(size = 10),
    axis.text = element_text(size = 9)
  )
)
```



Figure 1: Distribution of IQ Scores by Proximity to Smelter (Before Correction)

The initial boxplot reveals an unusual observation in the “Far” group with an IQ score of 999, which appears to be a data entry error.

Initial Summary Statistics

```
kable(means_by_location,
      col.names = c("Location", "Mean IQ", "SD", "N"),
      digits = 2,
      caption = "Summary Statistics of IQ Scores by Location (Before Correction)"
)
```

Table 1: Summary Statistics of IQ Scores by Location (Before Correction)

Location	Mean IQ	SD	N
Far	106.12	111.88	67
Near	89.19	12.17	57

Corrected Analysis

After consulting with the primary investigator, we learned that the IQ value of 999 should be 99. Let's correct this error and reanalyze the data.

```
# Fix the incorrect value
data$IQ[data$IQ == 999] <- 99

# Recalculate means
updated_means <- data %>%
  group_by(Smelter) %>%
  summarise(
    mean_iq = mean(IQ),
    sd_iq = sd(IQ),
    n = n()
  )
```

Updated Visualization

```
ggplot(data, aes(x = Smelter, y = IQ)) +
  geom_boxplot(fill = "lightgreen", alpha = 0.5) +
  geom_jitter(width = 0.2, alpha = 0.5, color = "darkgreen") +
  theme_minimal() +
  labs(
    title = "Distribution of IQ Scores by Proximity to Smelter (Corrected Data)",
    x = "Distance from Smelter",
    y = "IQ Score"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title = element_text(size = 10),
    axis.text = element_text(size = 9)
  )
```



Figure 2: Distribution of IQ Scores by Proximity to Smelter (After Correction)

Updated Summary Statistics

```
kable(updated_means,  
  caption = "Summary Statistics of IQ Scores by Location (After Correction)",  
  col.names = c("Location", "Mean IQ", "SD", "N"),  
  digits = 2)
```

Table 2: Summary Statistics of IQ Scores by Location (After Correction)

Location	Mean IQ	SD	N
Far	92.69	15.97	67
Near	89.19	12.17	57

Results and Discussion

After correcting the data entry error, we can see that children living far from the smelter have a mean IQ of 92.69 (SD = 15.97), while those living near the smelter have a mean IQ of 89.19 (SD = 12.17).

The corrected visualization and summary statistics show a clear difference in IQ scores between the two groups, with children living far from the smelter showing generally higher IQ scores than those living near the smelter. This suggests a possible relationship between proximity to the lead-emitting smelter and cognitive development as measured by IQ scores.