Lead-IQ Data Analysis Report

Yifan Zhao

2022-10-07

Project Setup

In this study, we are interested in the lead-IQ dataset, which is described in detail in the `Background/DataInfo.md` document.

Let's read in the data first, and take a look at the variables in the dataset.

```
# Read in lead-iq-01.csv
lead_IQ <- read.csv("../DataRaw/lead-iq-01.csv", stringsAsFactors = T)
# Note: Since the report will be generated in the ./Reports folder,
# relative path is used here.

# Print head and tail of the dataset
kable(head(lead_IQ), row.names = T)
```

	Smelter	IQ
1	Far	70
2	Far	85
3	Far	86
4	Far	76
5	Far	96
6	Far	94

```
kable(tail(lead_IQ), row.names = T)
```

	Smelter	IQ
119	Near	95
120	Near	77
121	Near	74
122	Near	96
123	Near	91
124	Near	78

```
# Check dimensions
dim(lead_IQ)
```

```
## [1] 124   2
```

```
# Check dataset summary
summary(lead_IQ)
```

```
##  Smelter         IQ
##  Far :67   Min.   : 46.00
##  Near:57   1st Qu.: 81.50
##            Median : 91.00
##            Mean   : 98.34
##            3rd Qu.: 99.25
##            Max.   :999.00
```

Data Correction

After a brief conversation with the primary investigator, it was identified that one individual who had an IQ value of "999" was in fact supposed to have a value of "99". We need to correct that and save the processed data in the `DataProcessed` folder.

```
# Change "999" to "99"
lead_IQ$IQ[which(lead_IQ$IQ == 999)] <- 99

# Check the summary of the current dataset
summary(lead_IQ)
```

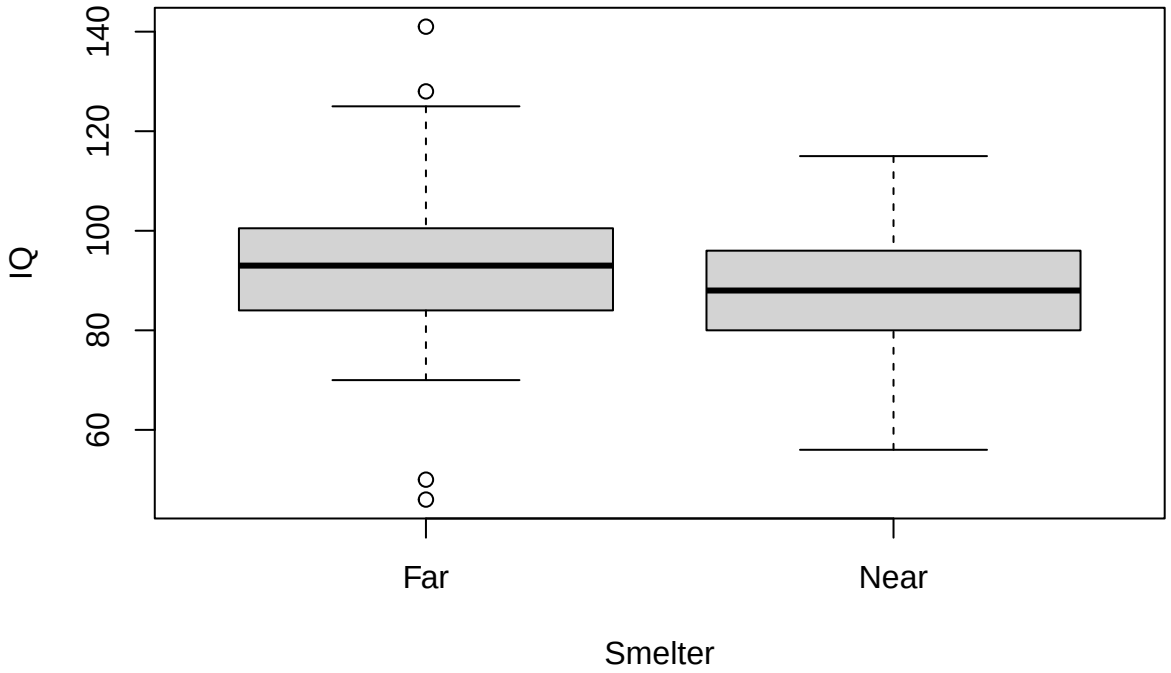
```
##  Smelter         IQ
##  Far :67   Min.   : 46.00
##  Near:57   1st Qu.: 81.50
##            Median : 91.00
##            Mean   : 91.08
##            3rd Qu.: 99.00
##            Max.   :141.00
```

```
# Save the processed data in csv format in the DataProcessed folder
write.csv(lead_IQ, "../DataProcessed/lead-iq-corrected.csv")
```

Boxplot

It will be great to show the IQ levels by location status in a box plot.

```
boxplot(IQ ~ Smelter, data = lead_IQ)
```



As we can see from the figure above, the IQ levels in Group Far are generally higher than those in Group Near, but the IQ levels in Group Far have a wider spread than those in Group Near.

Statistics Summary

- The mean of the IQ is currently 91.08.
- The standard deviation of the IQ is 14.4.
- In Group Near:
  - the mean of the IQ is 89.19;
  - the standard deviation of the IQ is 12.17.
- In Group Far:
  - the mean of the IQ is 92.69;
  - the standard deviation of the IQ is 15.97.