Statistics 2.3.2 Exercise 6

Gear Teeth

Lucas Liona

Table of contents

Problem

A stem-and-leaf for position 7 is given, together with other plots, in Fig. 3.2.5 in the main text. The body of the plot is bimodal, and there is a group of three much smaller observations.

Data Entry and Analysis

First, let's recreate the stem-and-leaf plot from the text and analyze the distribution of gear teeth measurements at position 7.

```
# Data for position 7 based on the stem-and-leaf plot
# Each value represents the measurement of gear teeth at position 7
pos7 <- c(3.01, 3.02, 3.03, 3.08, 3.09, 3.09, 3.10, 3.10, 3.11, 3.12, 3.13,
          3.13, 3.14, 3.14, 3.14, 3.15, 3.29, 3.30, 3.30, 3.31, 3.32, 3.33)

# Display basic summary statistics
summary(pos7)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  3.010   3.092   3.130   3.156   3.255   3.330 
```

```
# Calculate standard deviation
sd(pos7)
```

```
[1] 0.1028943
```

The summary statistics show the central measures and spread of our data. Now, let's visualize this data using several methods to examine its distribution.

Stem-and-Leaf Plot

```
# Create stem-and-leaf plot
stem(pos7, scale = 2)
```

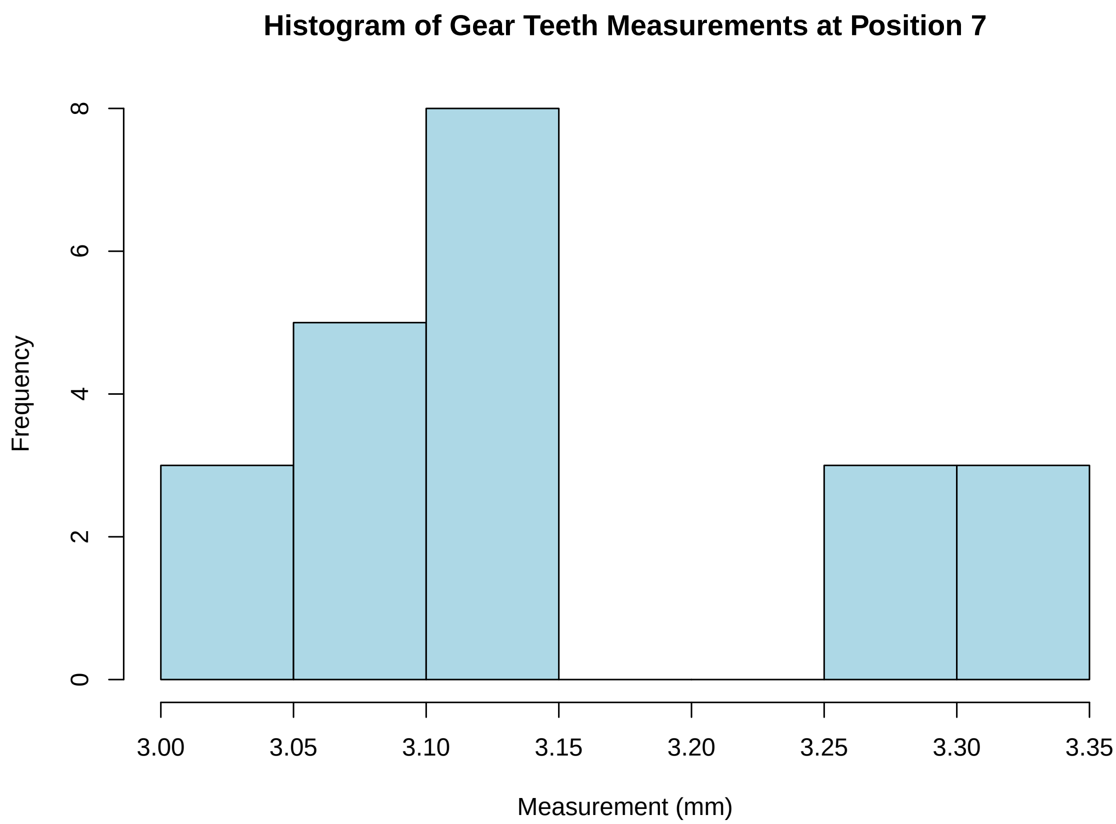
```

  The decimal point is 1 digit(s) to the left of the |

  30 | 123
  30 | 899
  31 | 001233444
  31 | 5
  32 | 
  32 | 9
  33 | 00123
```

The stem-and-leaf plot shows the distribution of the data. We can clearly see the bimodal nature of the distribution with two clusters of measurements, as well as the three smaller observations as mentioned in the problem statement.

Histogram

Let's create a histogram to further visualize the distribution:

```r
# Create histogram
hist(pos7,
     main = "Histogram of Gear Teeth Measurements at Position 7",
     xlab = "Measurement (mm)",
     col = "lightblue",
     breaks = 10)
```

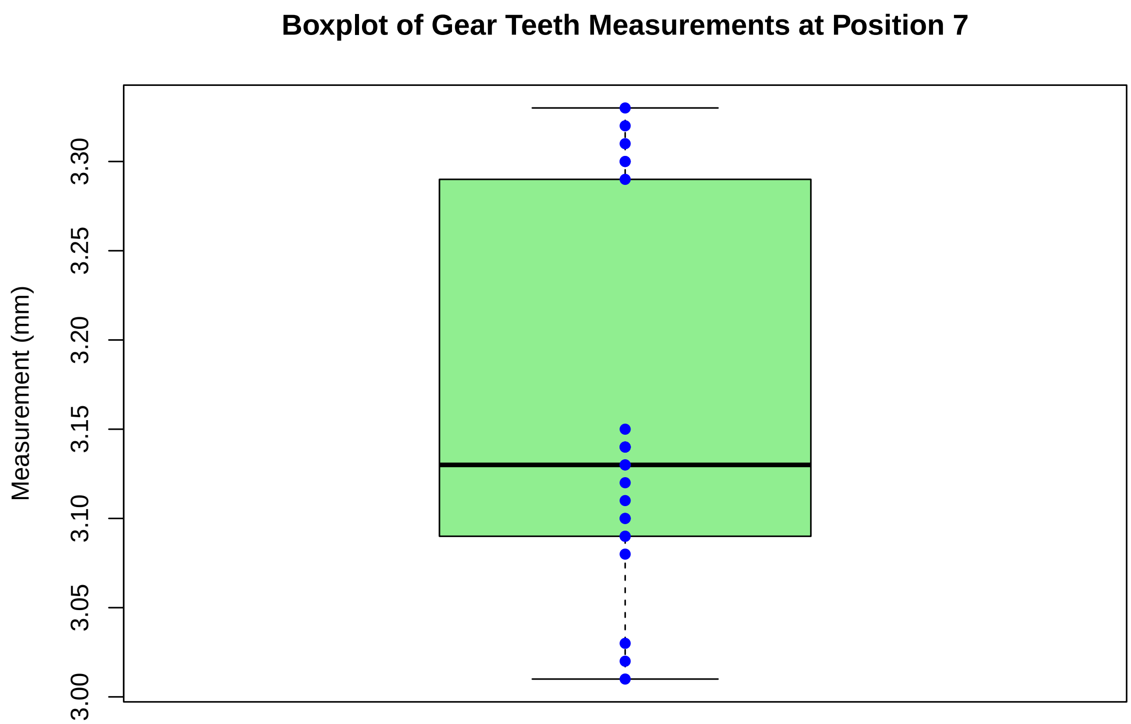


The histogram confirms the bimodal distribution we observed in the stem-and-leaf plot.

Boxplot

Now let's create a boxplot to identify potential outliers and examine the distribution's symmetry:

```
# Create boxplot
boxplot(pos7,
        main = "Boxplot of Gear Teeth Measurements at Position 7",
        ylab = "Measurement (mm)",
        col = "lightgreen")

# Add points to show actual data distribution
points(rep(1, length(pos7)), pos7, col = "blue", pch = 16)
```

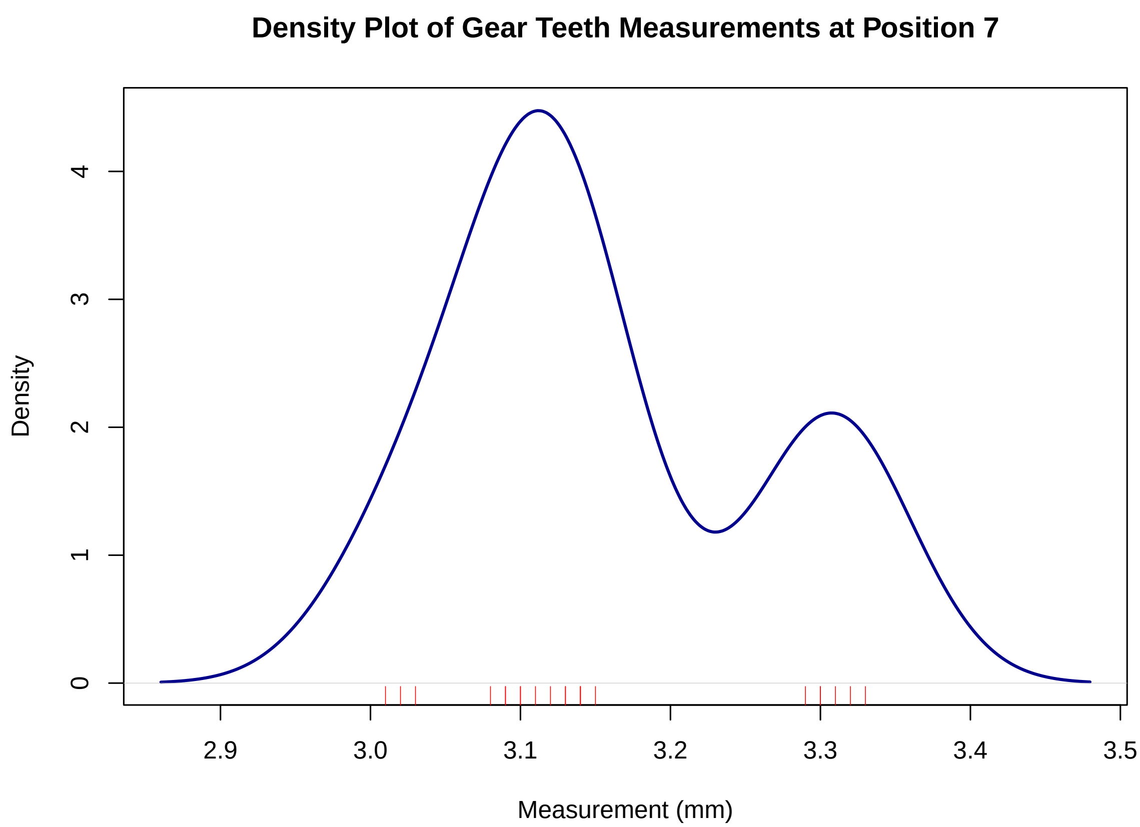


The boxplot shows the five-number summary (minimum, first quartile, median, third quartile, and maximum) visually. The points overlaid on the boxplot show the actual data distribution.

Density Plot

A density plot can help visualize the shape of the distribution:

```r
# Create density plot
plot(density(pos7),
     main = "Density Plot of Gear Teeth Measurements at Position 7",
     xlab = "Measurement (mm)",
     col = "darkblue",
     lwd = 2)

# Add rug plot to show actual data points
rug(pos7, col = "red")
```



The density plot clearly shows the bimodal nature of the distribution, with two peaks corresponding to the two clusters of measurements.

Identification of Groups

Based on our analysis, we can identify three distinct groups in the data:

1. A group of three smaller measurements around 3.01-3.03 mm
2. A cluster of measurements around 3.08-3.15 mm
3. Another cluster of measurements around 3.29-3.33 mm

Let's separate these groups for further analysis:

```
# Define the groups
group1 <- pos7[pos7 < 3.05]  # Smaller observations
group2 <- pos7[pos7 >= 3.05 & pos7 < 3.20]  # First mode
group3 <- pos7[pos7 >= 3.20]  # Second mode

# Display group sizes and summary statistics
cat("Group 1 (smaller observations):", length(group1), "measurements\n")
```

```
Group 1 (smaller observations): 3 measurements
```

```
cat("Range:", min(group1), "-", max(group1), "\n\n")
```

```
Range: 3.01 - 3.03
```

```
cat("Group 2 (first mode):", length(group2), "measurements\n")
```

```
Group 2 (first mode): 13 measurements
```

```
cat("Range:", min(group2), "-", max(group2), "\n\n")
```

```
Range: 3.08 - 3.15
```

```
cat("Group 3 (second mode):", length(group3), "measurements\n")
```

```
Group 3 (second mode): 6 measurements
```

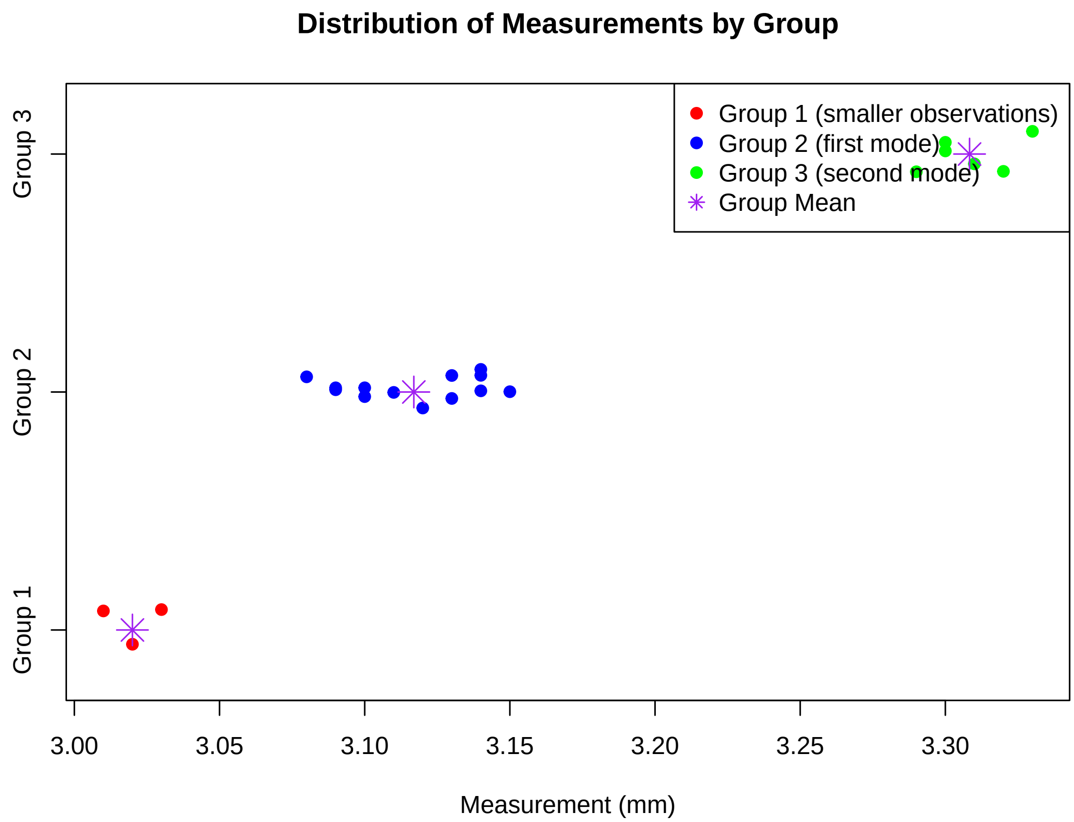
```
cat("Range:", min(group3), "-", max(group3), "\n")
```

```
Range: 3.29 - 3.33
```

Visual Comparison of Groups

Let's create a more visual representation of the three groups:

```r
# Create jittered stripchart for visualization
stripchart(list("Group 1" = group1, "Group 2" = group2, "Group 3" = group3),
           method = "jitter",
           pch = 19,
           col = c("red", "blue", "green"),
           main = "Distribution of Measurements by Group",
           xlab = "Measurement (mm)")

# Add group means
points(c(mean(group1), mean(group2), mean(group3)), c(1, 2, 3),
       pch = 8, cex = 2, col = "purple")

# Add legend
legend("topright",
       legend = c("Group 1 (smaller observations)",
                  "Group 2 (first mode)",
                  "Group 3 (second mode)",
                  "Group Mean"),
       pch = c(19, 19, 19, 8),
       col = c("red", "blue", "green", "purple"))
```



Conclusions

From our analysis, we can conclude:

1. The gear teeth measurements at position 7 show a clear bimodal distribution, with two distinct clusters of values.
2. There are three much smaller measurements (around 3.01-3.03 mm) that are separated from the main body of the data.
3. The first mode consists of measurements around 3.08-3.15 mm, while the second mode consists of measurements around 3.29-3.33 mm.
4. This pattern suggests that there might be two different manufacturing processes or settings being used, resulting in two distinct groups of measurements. The three smaller measurements could represent defective parts or measurements taken under different conditions.

5. For quality control purposes, it would be important to investigate why there are two distinct clusters of measurements and whether this bimodal distribution is acceptable for the intended use of the gears.

The bimodal distribution and the presence of the three much smaller observations align with what was described in the problem statement regarding Fig. 3.2.5 in the main text.