SA1 TIME SERIES Q6

Frances Aneth Rosales

2024-10-13

Contents

Import

READING THE DATA

Import

```
print(dataframe)
```

```
##     Cars Oil Mileage
## 1      1   1      36
## 2      1   2      38
## 3      1   3      30
## 4      1   4      29
## 5      2   1      34
## 6      2   2      38
## 7      2   3      30
## 8      2   4      29
## 9      3   1      34
## 10     3   2      28
## 11     3   3      38
## 12     3   4      32
## 13     4   1      38
## 14     4   2      34
## 15     4   3      20
## 16     4   4      44
## 17     5   1      26
## 18     5   2      28
## 19     5   3      34
## 20     5   4      50
```

1. Summary of ANOVA: Effect of Oil Type on Mileage

```
summary(anova_results)
```

```
##
## Error: Oil
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  3    103   34.33
##
## Error: Oil:Cars
##           Df Sum Sq Mean Sq F value Pr(>F)
## Cars       1    5.6    5.63   0.042  0.851
## Residuals  3  402.9  134.29
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 12  325.5   27.12
```

Findings:

The ANOVA results, with an **F value of 0.042** and a **p-value of 0.851** for the Cars factor, show that the type of oil utilized had no statistically significant impact on mileage. This implies that after adjusting each car's variances, the mileage differences between different oil types are **not statistically significant**. The residuals also exhibit significant change, but they provide no solid evidence that the type of oil used and mileage are **related**.

2. Summary of ANOVA: Cars as a Random Factor

```r
print(anova_summary)
```

```
## 
## Error: Cars
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  1  5.625   5.625
## 
## Error: Cars:factor(Oil)
##            Df Sum Sq Mean Sq
## factor(Oil)  3  266.6   88.85
## 
## Error: Within
##             Df Sum Sq Mean Sq F value Pr(>F)
## factor(Oil)  3  239.3   79.77   2.941 0.0762 .
## Residuals   12  325.5   27.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Findings:

The ANOVA results indicate that 5.625 is the residual sum of squares, suggesting that there is little variance in mileage that can be attributed to changes across individual cars. With an F value of 2.941 and a p-value of 0.0762, this shows that although there might be variations in mileage depending on the type of oil used, these variations are not statistically significant at the traditional alpha level of 0.05. Overall, the analysis indicates that there is a possibility that the **type of oil could affect mileage**.

```r
if ("Sphericity" %in% rownames(sphericity_test)) {
  mauchly_p_value <- sphericity_test["Sphericity", "Pr(>F)"]

  if (mauchly_p_value < 0.05) {
    cat("Mauchly's test p-value:", mauchly_p_value, "- Sphericity assumption is NOT met.\n")
  } else {
    cat("Mauchly's test p-value:", mauchly_p_value, "- Sphericity assumption is met.\n")
  }
} else {
  cat("Mauchly's test could not be performed.\n")
}
```

```
## Mauchly's test could not be performed.
```

3. Q-Q Plot

```
ggplot(dataframe, aes(sample = Mileage)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "Q-Q Plot of Mileage", x = "Theoretical Quantiles", y = "Sample Quantiles") +
  theme_minimal()
```
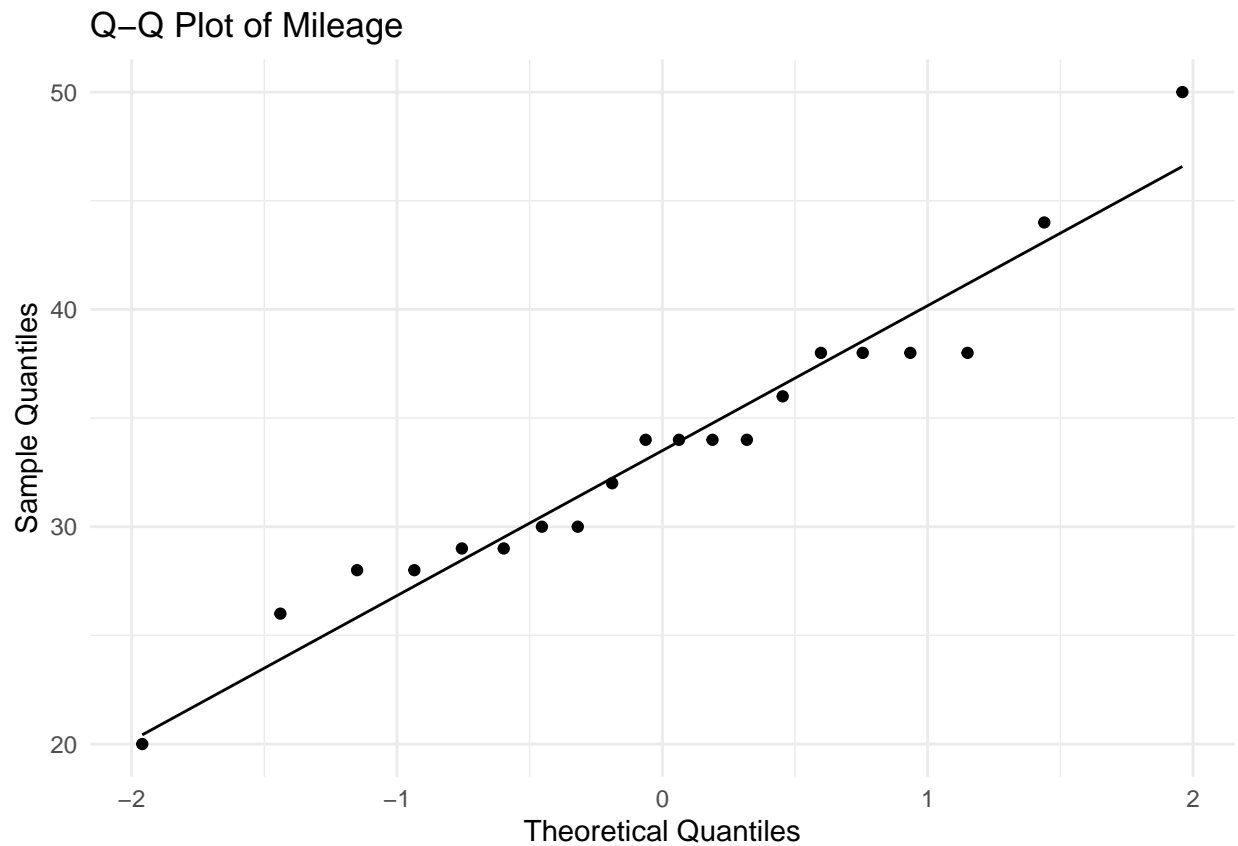
Q–Q Plot of Mileage

Findings:

A normal distribution is indicated by the Q-Q plot of mileage, which shows that the data follows a **45-degree reference line**. However, there are also variants, particularly in the **tails**, which suggest outliers or unusual patterns in both **high and low mileage values**. Thus, implicating a **deviating from normality**.

4. Boxplot: Mileage vs Oil

```
ggplot(dataframe, aes(x = factor(Oil), y = Mileage)) +
  geom_boxplot() +
  labs(title = "Mileage by Engine Oil Type", x = "Engine Oil", y = "Mileage") +
  theme_minimal()
```


Mileage by Engine Oil Type
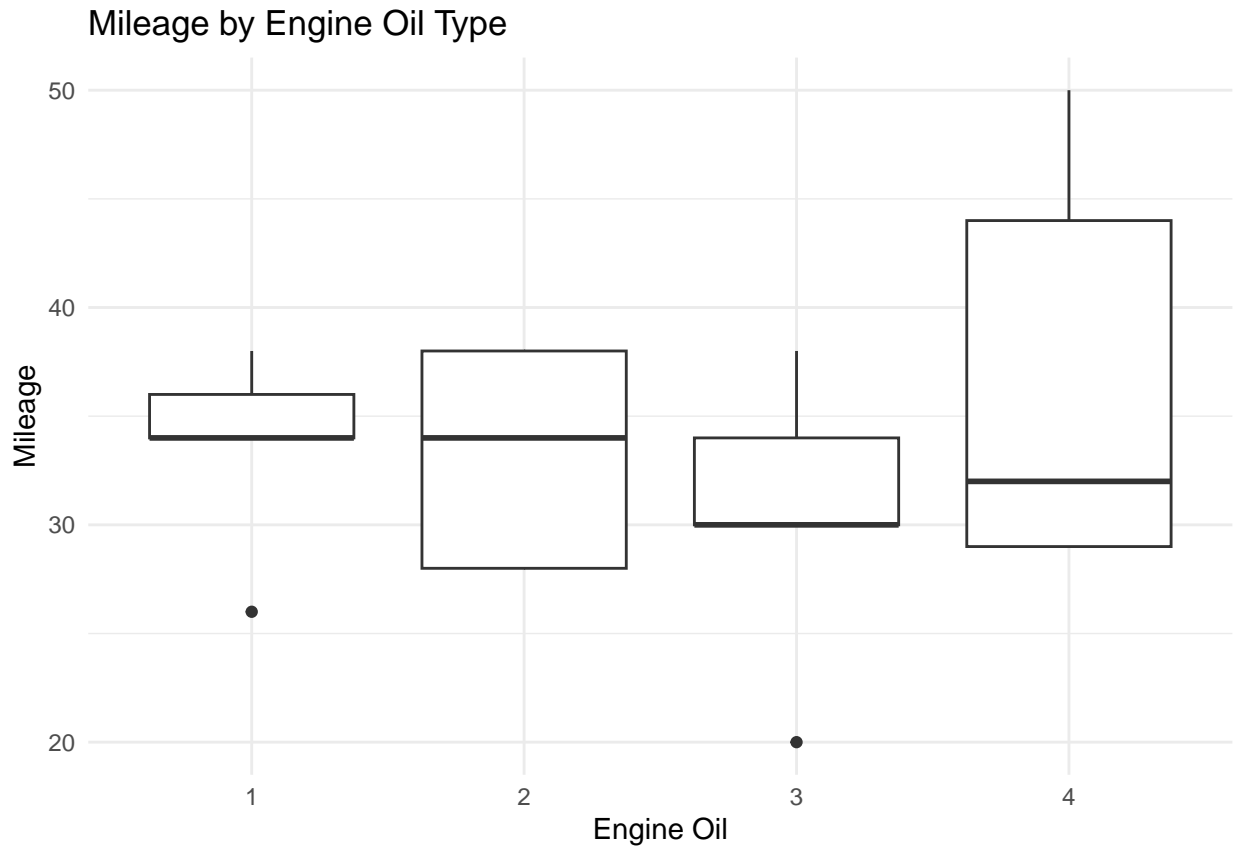
Findings:

As shown in our **boxplot**, Oil type 1 has a **fairly limited range** with one lower outlier, whereas oil type 4 has the **largest variance** in distance and the **broadest range** within a single quarter. When compared to the other oils, oil type 4 generally has a **greater median mileage**, while oil type 3 has the **lowest**.

5. POST-HOC

```
posthoc_results <- emmeans(anova_results, ~ Oil)
```

```
## Note: re-fitting model with sum-to-zero contrasts
```

```
## NOTE: Results are based on intra-block estimates and are biased.
```

```
tukey_results <- pairs(posthoc_results, adjust = "tukey")
print(tukey_results)
```

```
##  contrast    estimate   SE df t.ratio p.value
##  Oil1 - Oil2     -2.0 7.72 12  -0.259  0.9936
##  Oil1 - Oil3      7.4 7.72 12   0.958  0.7750
##  Oil1 - Oil4     18.7 7.72 12   2.421  0.1257
##  Oil2 - Oil3      9.4 7.72 12   1.217  0.6285
##  Oil2 - Oil4     20.7 7.72 12   2.680  0.0819
##  Oil3 - Oil4     11.3 7.72 12   1.463  0.4875
##
## P value adjustment: tukey method for comparing a family of 4 estimates
```

Findings:

There were no **significant differences** in mileage between the various oil types according to the **Tukey post-hoc analysis**, with the biggest difference being **18.7** between Oil 1 and Oil 4. The **p-value** of **0.1257**, however, was not statistically significant.

ANALYSIS AND CONCLUSION

The results of the ANOVA analysis indicate that the type of oil used does not have a statistically significant effect on mileage, as shown by the p-value of 0.851. The boxplot demonstrates that oil type 4 had the highest median mileage, but the variation between oil types was not significant. The Q-Q plot suggests that while the data follows a normal distribution, there are outliers in both the high and low mileage values. Furthermore, the Tukey post-hoc test reveals no significant differences between the oil types, with the largest difference being 18.7 between Oil 1 and Oil 4, but this was not statistically significant (p-value = 0.1257). Overall, the analysis shows no strong evidence that the type of oil used affects the mileage significantly.