Samson, Justine Aizel
2021053281

SA #6

Applied Multivariate Data Analysis

Github Link: https://github.com/aizeljat/SA1_Samson/blob/main/SA_%236_Samson.ipynb

Introduction
The purpose of this study was to determine whether the performance of four distinct engine oil types in terms of vehicle mileage varied significantly. All four oils were used in the tests, and the mileage of each vehicle was noted. A repeated-measures ANOVA was used to examine the data. The sphericity and normality assumptions were also assessed.

Method
Four different engine oils were tested on a sample of five cars. The type of oil used was the independent variable, and the mileage was the dependent variable. In this repeated-measures design, every automobile served as its own control.

Findings
ANOVA with Repeated Measures

There were no discernible variations in mileage between the four oil types, according to the repeated-measures ANOVA ($F(3,12)=0.5679$, $p=.6466$). These findings imply that the type of engine oil has no statistically significant impact on the tested cars' mileage.

The ANOVA's p-value of 0.6466 is significantly greater than the standard significance level of 0.05. This implies that the mileage differences between the various oil types are not statistically significant. Stated differently, we are unable to draw the conclusion that the type of oil has a substantial impact on mileage based on the data presented.

Normality
The Shapiro-Wilk test was used to determine whether the mileage data for each oil was normal. The following were the outcomes:

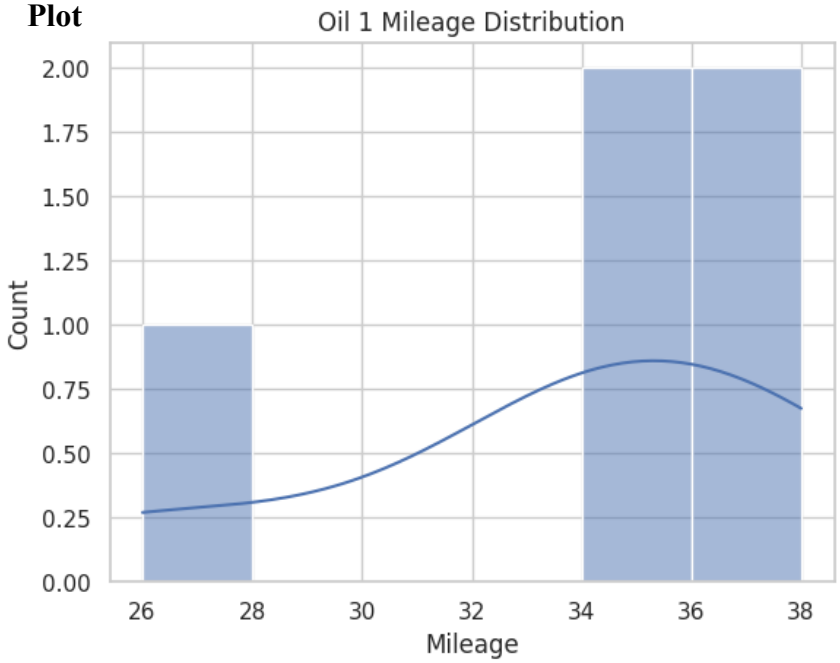
- Oil 1: $W=0.860, p=.228W = 0.860$, p = .228$W=0.860, p=.228$
- Oil 2: $W=0.814, p=.105W = 0.814$, p = .105$W=0.814, p=.105$
- Oil 3: $W=0.933, p=.616W = 0.933$, p = .616$W=0.933, p=.616$
- Oil 4: $W=0.833, p=.147W = 0.833$, p = .147$W=0.833, p=.147$

For all four oil conditions, the assumption of normalcy was satisfied because every p-value was higher than 0.05.

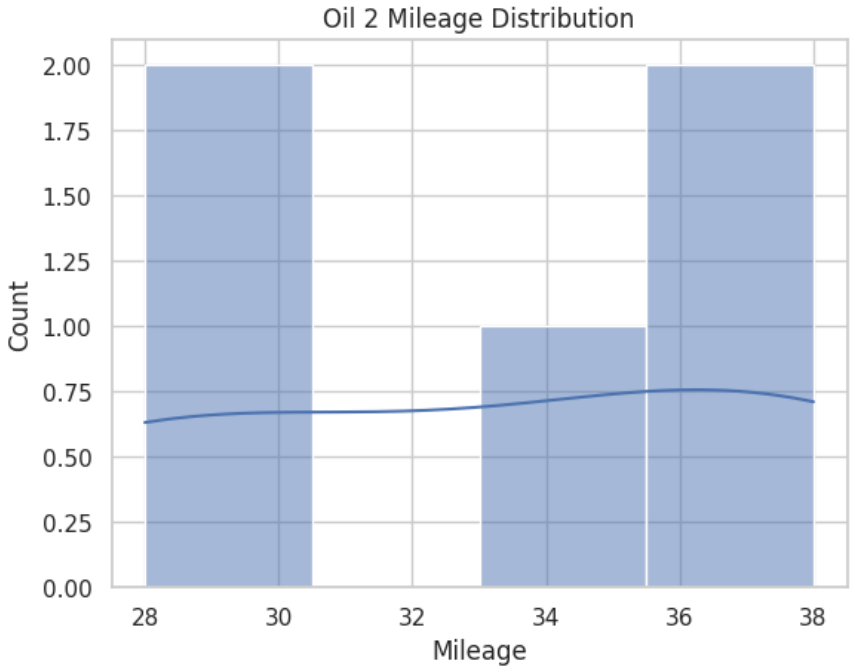
Sphericity
The assumption of sphericity was not broken, according to Mauchly's test for sphericity ($W=0.287$, $p=.659$). Consequently, there was no need to adjust for sphericity.

Samson, Justine Aizel
2021053281

Plot

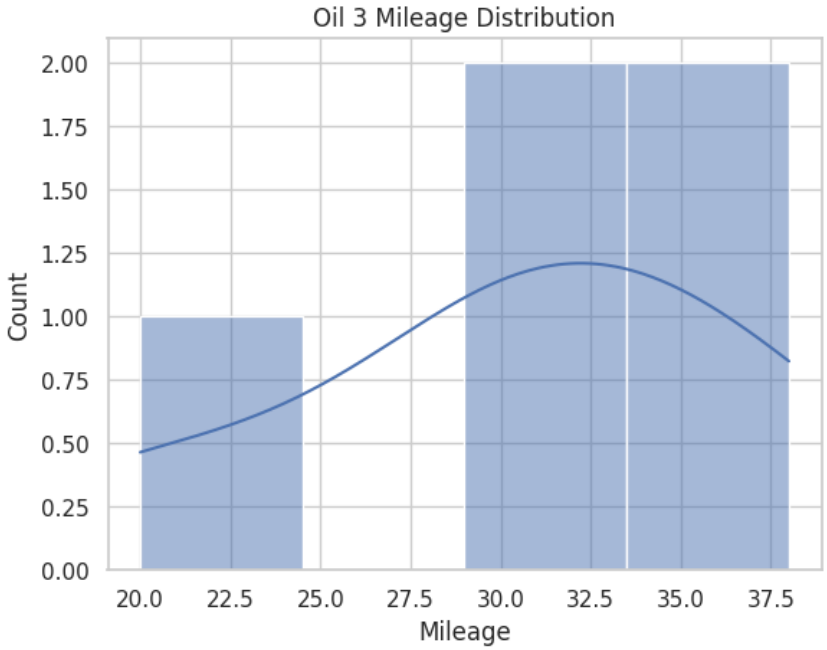


With the y-axis representing the number of observations and the x-axis displaying mileage values ranging from 26 to 38, the plot illustrates the distribution of mileage for Oil 1. The highest counts (roughly 2) occur in the 34–36-mile range, which is visible as a peak. With only a slight count seen at the lower end, between 26 and 28, the data points to a bias towards higher mileage values. As mileage increases, the distribution exhibits a unimodal trend, peaking and then slightly declining.

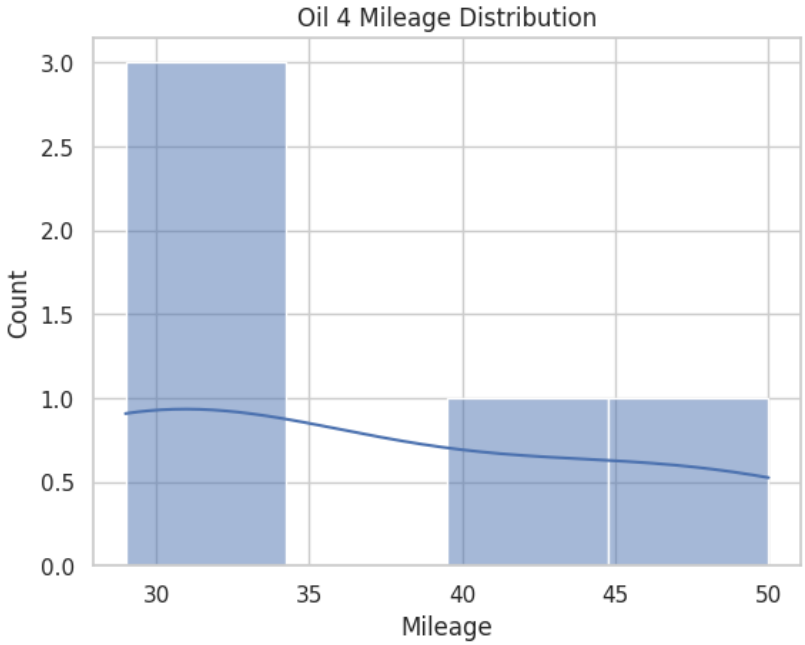


The histogram for Oil 2's mileage distribution displays two noticeable peaks in a reasonably symmetrical pattern. This implies that although the data does not exhibit extreme skewness or kurtosis, it may not have a perfectly normal distribution. With a p-value of 0.105 from the Shapiro-Wilk test, Oil 2 did not significantly deviate from normalcy at the 0.05 level.

Consequently, the Oil 2 mileage data logically satisfies the repeated-measures ANOVA's normality assumption.



The Oil 3 mileage information spans a range of roughly 20 to 37.5 miles. Between thirty and thirty-five miles, the highest frequency of data points is observed. The distribution clearly peaks at 32 miles, suggesting that most cars using Oil 3 get about this distance. With fewer data points as mileage rises above 35 miles, the curve points to a slightly skewed right distribution.



Oil 4's mileage data shows a notable peak between 30 and 35 miles, with the majority of the data falling between 30 and 50 miles. Beyond 40 miles, the distribution stays comparatively low after a steep decline after 35 miles. Oil 4 does not typically achieve higher mileage, as

evidenced by the small number of data points in the 40–50-mile range. The distribution is more left-skewed, with most vehicles achieving lower mileage.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
1	2	-0.4	0.9997	-12.6557	11.8557	False
1	3	-3.2	0.8766	-15.4557	9.0557	False
1	4	3.2	0.8766	-9.0557	15.4557	False
2	3	-2.8	0.9127	-15.0557	9.4557	False
2	4	3.6	0.8345	-8.6557	15.8557	False
3	4	6.4	0.4636	-5.8557	18.6557	False

The mileage differences between the engine oil pairs were compared using the Tukey HSD post-hoc test. No statistically significant differences were found in any of the pairwise comparisons because all p-values were higher than 0.05 and all comparisons' confidence intervals included zero. This implies that none of the different types of oil have appreciable variations in mileage. Because of this, we are unable to rule out the null hypothesis for any pairwise comparisons, which lends more credence to the idea that engine oil type has little bearing on vehicle mileage.