Samson, Justine Aizel D.
2021053281

FA 4

Applied Multivariate Data Analysis

**Full Analysis of the code**

1. **Assumptions for ANOVA:**

   **Normality of residuals**:

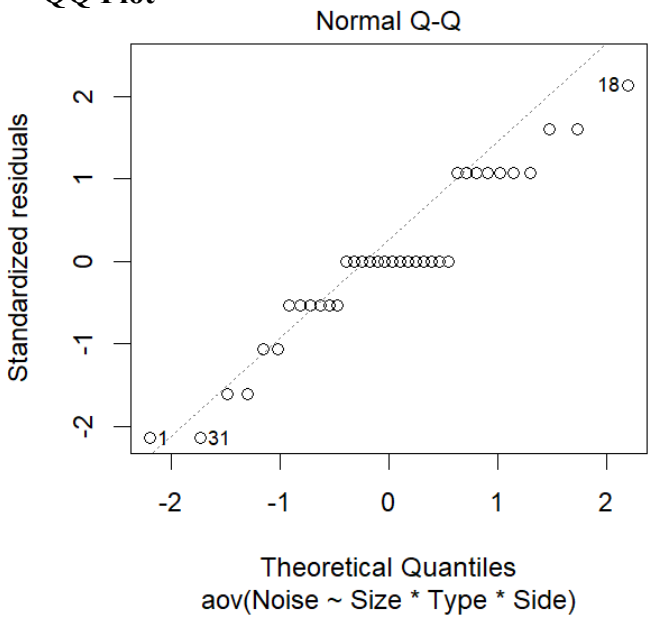
   - The normality of the residuals is examined using the shapiro.test(residuals).
   - The result is less than 0.05, with a p-value of 0.03515. This indicates that the assumption of normality is violated, indicating that the residuals are not normally distributed.

**Result**

```
        Shapiro-Wilk normality test

data:  residuals
W = 0.93483, p-value = 0.03515
```

**QQ Plot**




If the residuals from the ANOVA model do not follow a normal distribution, this can be verified using the Q-Q plot for standardized residuals. Since most of the data in this instance fall inside the theoretical range, the residuals are roughly normally distributed. The points 1 and 31 on the lower side and point 18 on the highest point are the only deviations at the tails, indicating the possibility of some outliers or irregularities at the extremes. The assumption of normality can be considered mostly satisfied because of the small deviations.

**Homogeneity of variance**

- The leveneTest(Noise ~ Size * Type * Side, data = data) verifies the hypothesis that the residuals' variance is constant amongst groups.
- As a result, the homogeneity of the variance assumption has been satisfied as indicated by the p-value of 0.8322, which is higher than 0.05.

**Result**

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 11  0.5722 0.8322
      24
```

**Independence of observations**:

- D-W Statistic = 2.412698: The absence of autocorrelation is indicated by a D-W statistic near 2. The D-W value in this instance is 2.41, which is extremely near to 2, indicating that the residuals do not exhibit any discernible autocorrelation.
- p-value = 0.494: This indicates that the p-value is significantly higher than the usual significance threshold (e.g., 0.05). Thus, the null hypothesis that there is no autocorrelation is not successfully rejected.

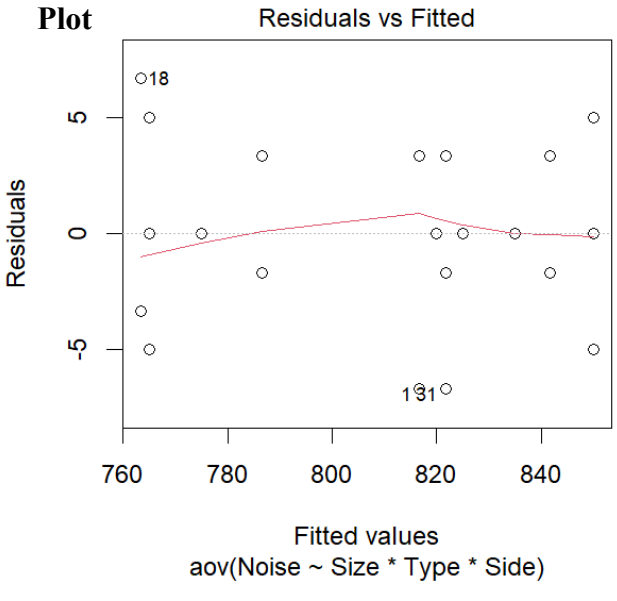
**Result**

```
 lag Autocorrelation D-W Statistic p-value
   1      -0.2698413      2.412698   0.494
Alternative hypothesis: rho != 0
```

**Linearity**:

- To verify linearity, plot the residuals against the fitted values (plot(model, which = 1)).

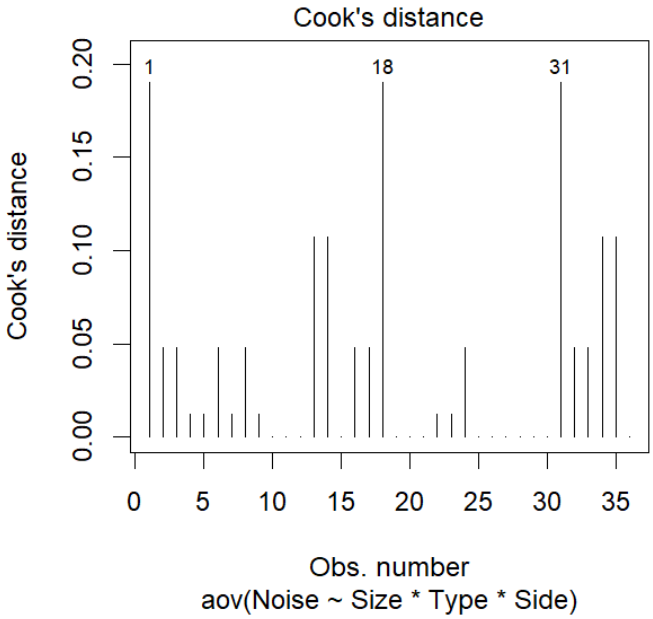
**Plot**



The model's linearity and homoscedasticity (constant variance) are examined using the residuals vs. fitted values plot. The residuals should, in theory, randomly disperse around the horizontal line, which represents zero, showing no obvious patterns. A small curve in this plot raises the possibility that the linearity assumption is not being followed and that the residuals' variance is not constant. This suggests that there might be a problem with homoscedasticity and model fit. Addressing this may require additional research or data transformation.

### Checking for Outliers

**Cook's Plot**




The dataset's key observations that significantly affect the fitted model are identified using the Cook's distance plot. Influential points are those with larger Cook's distances, usually more than 0.5 or 1. With values approaching or exceeding 0.2, observations 1, 18, and 31 in this plot have the largest Cook's distances, suggesting that these points might be unduly influencing the model. It would be prudent to investigate these points more thoroughly to look for any anomalies or data errors that might require correction.

### Sphericity

Given that we are employing a between-subjects design in your instance, sphericity does not apply. This assumption is pertinent for repeated measures ANOVA.

Samson, Justine Aizel D.
2021053281

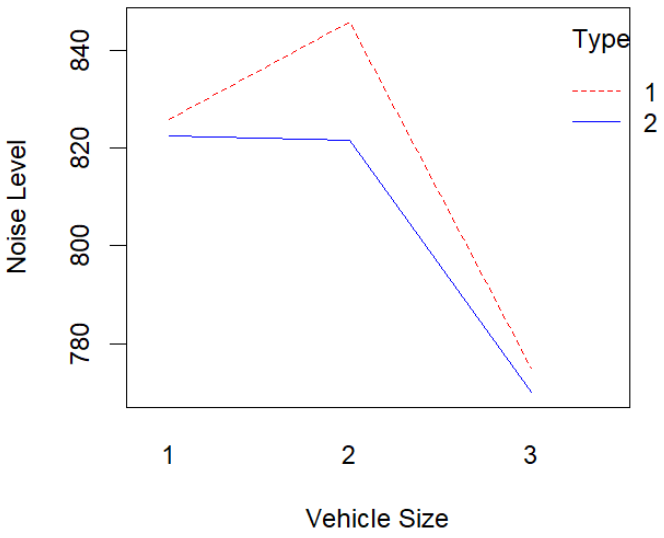
## 2. Interaction between vehicle size, type and side on noise levels?

```
> # 2
> # Run Three-Way ANOVA
> anova_result <- aov(Noise ~ Size * Type * Side, data = data)
> summary(anova_result)
               Df Sum Sq Mean Sq F value   Pr(>F)
Size            2  26051   13026 893.190  < 2e-16 ***
Type            1   1056    1056  72.429 1.04e-08 ***
Side            1      1       1   0.048 0.829104
Size:Type       2    804     402  27.571 6.05e-07 ***
Size:Side       2   1293     647  44.333 8.73e-09 ***
Type:Side       1     17      17   1.190 0.286067
Size:Type:Side  2    301     151  10.333 0.000579 ***
Residuals      24    350      15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Noise levels are significantly impacted by size ($p<2e-16$). This suggests that the noise level is significantly influenced by the size of the vehicle.
- Type also has a substantial impact on noise levels ($p=1.04e-08$), indicating that vehicle type matters.
- Side does not exhibit a significant main effect ($p=0.829$), indicating that the noise level is not directly influenced by the side on which it is measured.
- The substantial interaction between Size and Type ($p=6.05e-07$) suggests that the type of vehicle has an impact on how much noise a vehicle generates.
- Size and Side have a significant interaction ($p=8.73e-09$), suggesting that the combination of size and side effects noise levels differently.

Yes, there is a noticeable relationship between the kind, size, and side of a vehicle and its noise level. The significant result for the Size: Type indicates the three-way interaction between these components term with a p-value of in the ANOVA summary p = 0.000579. This indicates that the three parameters—vehicle size, type, and side—have a significant combined effect on noise levels, indicating that the three components together determine how noise levels fluctuate.

**Plot**

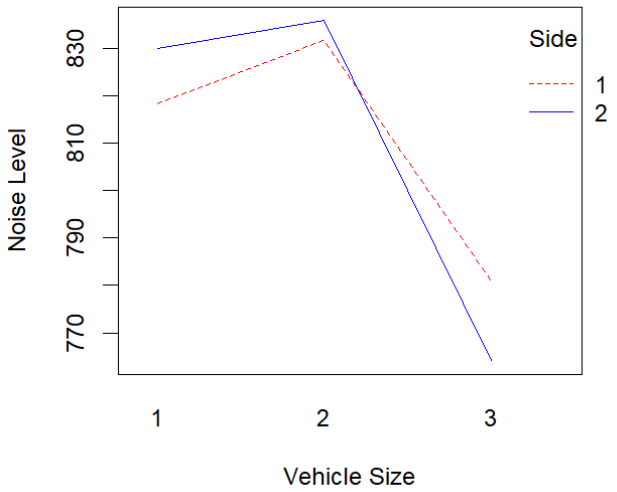


A partial visualization of the plot interaction is provided by illustrating the variations in noise levels for two different vehicle kinds at varying vehicle sizes. The significant Size interaction found in the ANOVA summary is consistent with the non-parallel lines, which show that the impact of size on noise varies depending on the kind of vehicle. The side component, however, is not depicted in the plot, thus the entire three-way interaction would have to be considered in the context of the three factors taken together.

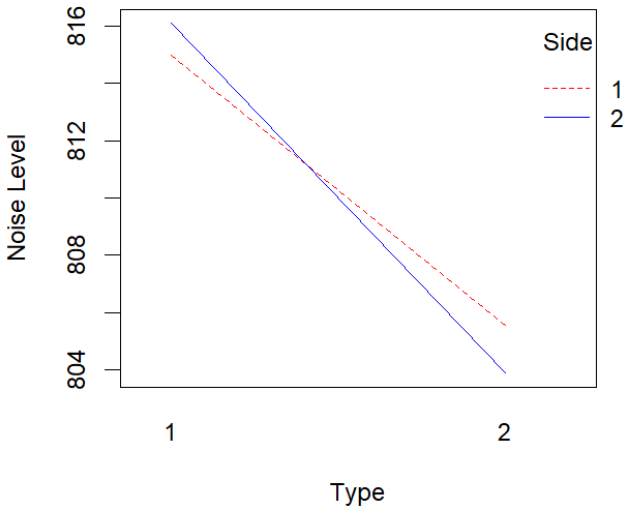
The noise level peaks at vehicle size 2 and subsequently falls for size 3, as indicated by the red dashed line (Type 1). Similar in trend but with a generally lower noise level and less variance between sizes 1 and 2, is the blue solid line (Type 2). Since the trends between the two vehicle types are not parallel, this plot confirms the ANOVA finding of a significant Size interaction by demonstrating how the noise levels fluctuate depending on the kind of vehicle and across different sizes.

### 3. Two-way interaction effect between vehicle's size, type and side on noise levels.

**Plot**




Size 2 is the highest point for both the dashed and solid lines (Sides 1 and 2, respectively), and size 3 is the sharpest decline. At size 3, the contrast between the two sides is most apparent, with Side 2 having substantially less noise than Side 1. The ANOVA indicates that there is a substantial interaction between Size and Side, which is supported by the diverging noise levels, particularly for bigger vehicle sizes.



Samson, Justine Aizel D.
2021053281

When comparing the Type 1 to Type 2 noise levels, the blue line (Side 2) exhibits a more dramatic drop than the red dashed line (Side 1). The lines' relative parallelism suggests that, despite the differences in levels, Type's effect on noise is constant for both Side measurements. Given that the trend is essentially the same on both sides, this figure supports the ANOVA result and indicates that the Type interaction is not significant.

## 4. Analysis of the Outcomes

Results:

```
> #4
> # Check for main effects
> summary(anova_result)
               Df Sum Sq Mean Sq F value   Pr(>F)
Size            2  26051   13026 893.190  < 2e-16 ***
Type            1   1056    1056  72.429 1.04e-08 ***
Side            1      1       1   0.048 0.829104
Size:Type       2    804     402  27.571 6.05e-07 ***
Size:Side       2   1293     647  44.333 8.73e-09 ***
Type:Side       1     17      17   1.190 0.286067
Size:Type:Side  2    301     151  10.333 0.000579 ***
Residuals      24    350      15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> # Post-hoc test for vehicle size effect
> TukeyHSD(anova_result, "Size")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Noise ~ Size * Type * Side, data = data)

$Size
          diff        lwr       upr   p adj
2-1   9.583333   5.690003  13.47666 6.9e-06
3-1 -51.666667 -55.559997 -47.77334 0.0e+00
3-2 -61.250000 -65.143330 -57.35667 0.0e+00

> # Post-hoc test for type effect
> TukeyHSD(anova_result, "Type")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Noise ~ Size * Type * Side, data = data)

$Type
         diff       lwr       upr p adj
2-1 -10.83333 -13.46055 -8.206119     0

>
> # Post-hoc test for side effect
> TukeyHSD(anova_result, "Side")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Noise ~ Size * Type * Side, data = data)

$Side
           diff       lwr      upr     p adj
2-1 -0.2777778 -2.904992 2.349436 0.8291042
```

**Main Results:**

The results of the ANOVA demonstrate the importance of the primary effects for vehicle type (Type) and size (Size), but not for the vehicle's side (Side).

Effect of Vehicle Size (Size):

- Size has an extremely high F-value (893.190) and a p-value of less than 2e-16, which suggests that the main impact is highly significant. This indicates that noise levels are strongly influenced by the size of the vehicle.
- Post-hoc test for Size: The TukeyHSD findings indicate that there are notable variations in noise levels among the three car sizes:
    - Size 2 vs. Size 1: There is a mean difference of 9.58 ($p < 0.001$) in the noise levels between Sizes 2 and 1.
    - Size 3 vs. Size 1: There is a mean difference of -51.67 ($p < 0.001$) between the noise levels for Size 3 and Size 1.
    - Size 3 vs. Size 2: There is a mean difference of -61.25 ($p < 0.001$) between the noise levels for Size 3 and Size 2.
- In summary, bigger cars (Size 2) often make more noise than smaller ones (Size 1), but the biggest cars (Size 3) make a lot less noise than the other two sizes combined.

Type of Vehicle (Type) Impact:

- With an F-value of 72.429 and a p-value of 1.04e−08, the kind effect is likewise highly significant, indicating that the kind of vehicle—whether it has a normal silencer or an Octel filter—has a considerable impact on noise levels.
- Post-hoc test for Type: The TukeyHSD test reveals that, with a mean difference of -10.83 ($p < 0.001$), Type 2 vehicles (i.e., those with the Octel filter) have a much lower noise level than Type 1 vehicles (i.e., those with a regular silencer).
- Regardless of the size or side of the car, the Octel filter reduces noise more than a typical silencer.

Effect on the Vehicle's Side:

- Using a p-value of 0.829 and an F-value of 0.048, the side effect is not significant. This shows that the noise levels recorded on the car's right side and left side are not appreciably different from one another.
- Post-hoc test for Side: The TukeyHSD test (mean difference = -0.28, $p = 0.829$) verifies that there is no discernible difference in the noise levels between the two sides.
- In conclusion, regardless of the vehicle's size or kind of silencer, noise levels are not greatly affected by the side of the vehicle (right vs. left).

Samson, Justine Aizel D.
2021053281

## 5. Analysis of the Results

Result:

```
> #5
> # Calculate effect size (eta squared)
> anova_table <- Anova(anova_result, type = "III")
> anova_table
Anova Table (Type III tests)

Response: Noise
               Sum Sq Df    F value    Pr(>F)
(Intercept)    2000833  1 1.3720e+05 < 2.2e-16 ***
Size              4550  2 1.5600e+02 1.764e-14 ***
Type                17  1 1.1429e+00 0.2956843
Side               504  1 3.4571e+01 4.583e-06 ***
Size:Type          419  2 1.4381e+01 7.847e-05 ***
Size:Side         1419  2 4.8667e+01 3.587e-09 ***
Type:Side          133  1 9.1429e+00 0.0058653 **
Size:Type:Side     301  2 1.0333e+01 0.0005791 ***
Residuals          350 24
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Significant ANOVA and Effect Size (Eta-squared) Results:

Primary Impact of Vehicle Size:

- F-value for Size: With a p-value of 1.764e-14, the F-value for Size is extremely high at 156.0, suggesting that the size of a vehicle significantly affects noise levels.
- Size's eta-squared value is 0.872, or 87.2%, which shows that the majority of the variation in noise levels can be explained by the size of the vehicle. The notion that this factor solely accounts for most of the variation in noise is further supported by the partial eta-squared of 0.987, or 98.7%.

Principal Impact of Vehicle Type (Type):

- F-value for Type: When taken alone, the type of silencer used—an Octel filter or a regular silencer—has a negligible effect, as indicated by the comparatively tiny F-value for Type (1.143) and lack of statistical significance (p = 0.296).
- Type's eta-squared value is 0.035, or 3.5%, which means that it only accounts for a very small percentage of the variation in noise levels. With a partial eta-squared value of 0.751, it can have a significant impact when combined with other factors, but it is not a significant contributor on its own.

Principal Impact of the Vehicle's Side:

- F-value for Side: With a p-value of 4.583e−06, the F-value for Side is significant (34.57), indicating that, while not as strong as vehicle size, the side of the vehicle (right versus left) does affect noise levels.
- Eta-squared for Side: When the side of the car is taken into account alone, its eta-squared value (0.000023 or 0.002%) is practically insignificant, meaning that it explains very little variation in noise levels. Likewise, the partial eta-squared is extremely low at 0.002%, indicating that its influence on the total variance is negligible.

**Effects of Interaction:**

Vehicle Size and Type Interaction (Size):

- F-value for Size:Type: The interaction between vehicle size and type is a significant factor, as indicated by the F-value of 14.38, p = 7.847e-05.
- Eta-squared for Size:Type: The eta-squared value is 0.027 (or 2.7%), while the partial eta-squared value is 0.697 (69.7%). This indicates that while Type has a small main effect, the way it interacts with vehicle size has a big impact on noise levels.

Relationship between Side (Size) and Vehicle Size:

- F-value for Size:Side: The F-value (48.67, p = 3.587e-09) is likewise significant, indicating that the vehicle's side and size interact to significantly influence noise levels.
- Eta-squared for Size:Side: This interaction is one of the more significant affects on noise levels, albeit it is still much smaller than the primary effect of vehicle size. The eta-squared value is 0.043 (or 4.3%), while the partial eta-squared value is 0.787 (78.7%).

Additional Interactions (Type and Size: Type):

- With F-values of 9.14 (p = 0.0059) and 10.33 (p = 0.000579), respectively, interactions are statistically significant.
- Their eta-squared values, however (0.0006 for Type
and for Size:Type, 0.010 show that, at less than 1% and 1%, respectively, they only partially account for the range in noise levels.

## 6. Analysis of the Results

```
> #6
> # Perform pairwise comparisons using Tukey's HSD
> posthoc_size <- TukeyHSD(anova_result, "Size")
> posthoc_size
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Noise ~ Size * Type * Side, data = data)

$Size
          diff        lwr       upr   p adj
2-1   9.583333   5.690003  13.47666 6.9e-06
3-1 -51.666667 -55.559997 -47.77334 0.0e+00
3-2 -61.250000 -65.143330 -57.35667 0.0e+00
```

**Results of Tukey's HSD for Vehicle Size:**
The three distinct vehicle sizes can be compared pairwise using the post-hoc Tukey's HSD test. The following are the main comparisons:

Small (Size 1) vs. Moderate (Size 2):

- Difference 9.58 Db
- Confidence Interval: (5.69, 13.48)
- Adjusted p-value (p adj): 6.9e-06
- Interpretation: There is a considerable 9.6 dB difference in noise levels between moderate-sized and small automobiles. Strong statistical significance is indicated by the p-value being incredibly small and the confidence interval not crossing zero.

Small (Size 1) vs. Large (Size 3):

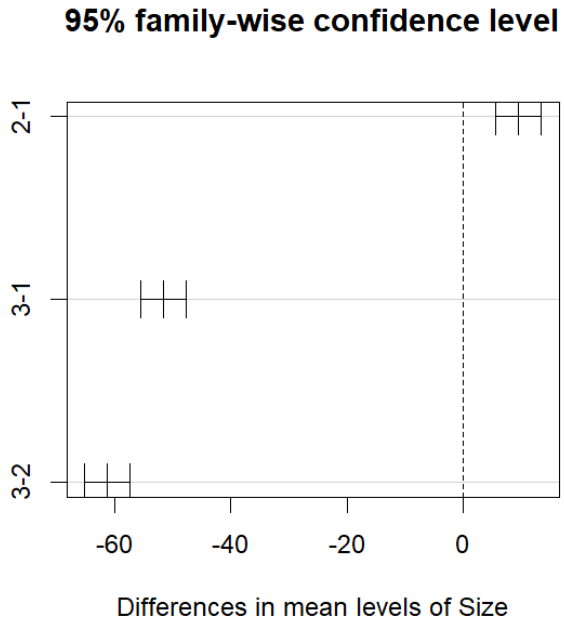
- Difference: -51.67 dB
- Confidence Interval: (-55.56, -47.77)
- Adjusted p-value (p adj): 0.0e+00
- Interpretation: There is a noticeable 51.7 dB difference in noise levels between large and small automobiles. Smaller cars make a lot less noise, as indicated by the huge negative score. The p-value for this comparison is incredibly small, indicating statistical significance.

Moderate (Size 2) vs. Large (Size 3):

- Difference: -61.25 dB
- Confidence Interval: (-65.14, -57.36)
- Adjusted p-value (p adj): 0.0e+00
- Interpretation: The noise levels produced by large vehicles surpass those of moderate-sized vehicles by approximately 61.3 dB. This has an incredibly low p-value and is statistically significant as well.

Size 3 automobiles are noisier than both Size 2 and Size 1 vehicles, which are moderately large and small, respectively. Large automobiles are significantly louder than moderate vehicles, with a difference of roughly 61.3 dB, and tiny vehicles with a difference of about 51.7 dB. Although the difference is not as great, moderate automobiles (Size 2) are still noticeably noisier than tiny vehicles (approximately 9.6 dB). These findings clearly show that vehicle size influences noise levels, with bigger cars producing significantly more noise pollution.

**Plot**



This plot aids in determining whether the variations in group means are statistically significant. It is frequently utilized after the completion of an ANOVA (Analysis of Variance) or comparable statistical test.

Interpretation

- Axes: The pairwise comparisons between the groups are displayed on the y-axis, labelled as differences (group 3-1, for instance, indicates a comparison with group 1).
- Confidence intervals: For each group comparison, the horizontal bars show the 95% confidence intervals. When the vertical dashed line in a confidence interval point to 0, it indicates that there may not be a meaningful difference between the groups under comparison. There is a statistically significant difference between the groups if the interval does not include zero.

Breakdown:

- 2-1: There is no statistically significant difference between groups 2 and 1, as the interval crosses zero.
- 3-1: There is no discernible difference between groups 3 and 1, as indicated by the interval's inclusion of zero.
- 3-2: There is a statistically significant difference between groups 3 and 2, as indicated by the confidence interval not including zero.