### Relationship Between Fuel Consumption and Independent Variables

**Methods and Data Used**

1. **Variables Used:**
   * Dependent Variable: Fuel consumption (Fuel consumed per 100km)
   * Independent Variables:
   * **Total\_Idling:** Total idling time, recorded when ignition was on but no speed was detected until ignition turned off.
   * **Total duration:** Total time of truck operation per month, summing trips with speed > 0. Reflects time spent on the road, influenced by factors like traffic and weather.
   * **Offences/100km:** Total sum of app-defined offenses (speeding, braking, cornering, acceleration) per 100km.
   * **Car:** Categorical variable indicating different vehicle types (A, B, C, etc.).
2. **Distribution of Fuel Consumption**: In Annex A, by comparing three distributions - normal, lognormal, and exponential - it's evident that the normal distribution is the best choice for model fitting. It has the lowest AIC of 273.6277 compared to 278.2701 and 335.8233 from lognormal and exponential distributions, respectively.
3. **Correlation Testing**:

A diagram of a graph

Description automatically generated with medium confidence

To address multicollinearity, tests were conducted to identify and remove correlated independent variables. Two variables showed significant correlation, with coefficients larger than 0.5: total distance and total duration, total offences per 100km and car. Further analysis indicated that car and total offences' multicollinearity was not a critical issue due to high R-squared and significant effects. Total distance was excluded from the model to prevent correlation with total duration, already accounted for in total fuel consumption per 100km.

1. **Regression Modelling**: Multiple regression was used to identify statistically significant variables at a 0.01 confidence level. Significant effects were identified when rejecting the null hypothesis for independent variables with a 5% maximum possibility of error.
2. **Interaction Modelling**: Interaction effects were explored and compared against the original model's adjusted R-squared values. Interaction between car type and number of offences yielded similar adjusted R-squared values, leading to the rejection of all interaction models.
3. **Residual Analysis**: Residual analysis validated the regression analysis by confirming model assumptions.

**Regression Results**

The regression model reveals significant variables (\* at 0.01 level) impacting fuel consumption. "Total duration," "Total offences per 100km," and "Car type" exhibit significant relationships with fuel consumption.

* **Total Duration:** Positive impact on fuel consumption indicates that longer trips lead to increased consumption. Road conditions, traffic, and events like accidents contribute.
* **Offences and Fuel Efficiency:** Driving behavior, including acceleration, braking, speeding, and cornering, significantly correlates with fuel consumption. Individual driving styles affect fuel usage.
* **Car Type Distinction:** "Car type" is significant, suggesting fuel efficiency differences. Further investigation of hardware for abnormalities affecting fuel consumption is recommended.

**Recommendations**:

Based on our investigation, we propose the following recommendations:

* Reduce time spent on the road to decrease fuel consumption.
* Turn off the engine when not in use to save fuel.
* Conduct monthly review sessions with drivers, considering offence count, to encourage better driving behaviour.
* Investigate mechanical performance of vehicles to identify abnormalities affecting fuel consumption and delivery performance.

.

**Limitation of Our Analysis & Conclusion**

* **Coefficient Magnitude:** Though variables are statistically significant, coefficients don't quantify the relationship's magnitude.
* **Interdependence of Variables:** Real-world interactions among independent variables could affect results.
* **Small Dataset:** Limited data (39 rows) could impact test accuracy and robustness. Additional data would enhance reliability.