

PROBABILITY AND STATISTICAL DATA ANALYSIS

SECI2143-02

SEMESTER II 2019/2020

**FACTORS THAT INFLUENCES CAR FUEL CONSUMPTION**

|  |  |
| --- | --- |
| **NAME**  | SHASITHER A/L SANDRAN |
| **MATRIC NO** | A19EC0160 |
| **SECTION** | 02 |
| **LECTURER** | DR. CHAN WENG HOWE |
| **SUBMMISSION DATE** | 27 JUNE 2020 |

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| NO | TITLE | PAGE |
| 1 | Introduction | 1 |
| 2 | Methodology | 1 |
| 3 | Statistical Analysis  | 2 |
| 4 | 2- Sample Hypothesis Testing | 2 |
| 5 | Correlation | 3 - 4 |
| 6 | Regression | 5 - 6 |
| 7 | Chi-Square Test of Independence | 7 |
| 8 | Discussion | 8 |
| 9 | Conclusion | 8 |
| 10 | References | 9 |

***Introduction***

The increasing global economic recession has caused major financial problems to domestic people with increasing prices of daily goods and the cost of living. One of the most important part of people’s everyday life will be travelling. Thus, it is clear that travelling has also increased in costs day after day. The increasing prices of car fuel can be a major problem for middle-income consumer to sustain life. Therefore, this study is carried out to identify the issues involving car fuel consumption. The primary objective of this study is to analyse the fuel consumption by the car and the factors that influences the fuel consumption. This is because, this study will be able to identify which factor has the greatest influence on the fuel consumption of vehicles. There were several factors considered to conduct the statistical analysis such as distance travelled (km), speed of the vehicle (km/h) and also the fuel type (SP98 or E10 gas).

***Methodology***

The dataset used in this analysis is obtained from Kaggle which was collected by an American car driver named Andreas Wagener. The method used by him to collect the data is by writing down the data of his car’s display after each ride in regular basis. Moreover, he did also change the gas type quite often between SP98 gas and E10 gas. The data greatly helped in analysing the relationship between different factors on the fuel consumption. There were several statistical analysis was carried out including 2-sample hypothesis testing, Correlation, Regression and also Chi Square Test of Independence. A total of 351 data were used to be analysed after all the unrelated and outlier data are cleaned.

***Statistical Analysis ~ Hypothesis Testing***

1. **2-sample Hypothesis Testing (mean with unknown variance)**

Based on the data collected, there were two different type of car fuel were used by the driver which was recorded which are namely E10 gas and SP98 gas. According to the claim that different fuel type can affect the fuel consumption of a vehicle, hypothesis testing on mean with unknown variance was conducted.

Significance level : Alpha = 0.05, Assumed equal variance, σ1² = σ2²

**H0 :** There is no difference between mean distance travelled using E10 gas type and the mean distance travelled using SP98 (µ1 = µ2)

**H1 :** There is difference between mean distance travelled using E10 gas type and the mean distance travelled using SP98 (µ1 ≠ µ2)







**Conclusion :** Since t=6.123661 > 1.644854, Reject Null Hypothesis, H0. At 0.05 significance level, there is sufficient evidence that there is difference between mean distance travelled using E10 gas type and the mean distance travelled using SP98 (µ1 ≠ µ2).

1. **Correlation Analysis**

Correlation analysis is conducted to identify the relationship between distance travelled (km) and the fuel consumption (l/100km). Correlation coefficient, r will be calculated using Pearson’s Product-moment using the following formula using R Studio



Significance level : Alpha = 0.05, DF = 349

**H0 :** There is no linear correlation between distance travelled (km) and the fuel consumption (l/100km). (ρ = 0)

**H1 :** There is linear correlation exists between distance travelled (km) and the fuel consumption (l/100km). (ρ ≠ 0)

  





**Conclusion :** Since r = - 0.2543047, there is relatively weak linear correlation between distance travelled (km) and the fuel consumption (l/100km).

1. **Regression**

The linear regression test is conducted with the dependent variable, y is fuel consumption(l/100km) and the independent variable, x is speed of the vehicle(km/h).

The estimated regression formula used is as followed



**H0 :** There is no linear relationship between speed of the vehicle(km/h) and the fuel consumption(l/100km) (β1 = 0)

**H1 :** There exists linear relationship between speed of the vehicle(km/h) and the fuel consumption(l/100km) (β1 ≠ 0)

**ŷ = 5.106855 - 0.008087x**

****

****

Coefficient of Determination, R2  = 0.015 🡪 Weaker Linear Relationship between speed of the vehicle(km/h) and the fuel consumption(l/100km). Some but not all of the variation in fuel consumption(l/100km) is explained by variation in speed of the vehicle(km/h).

**Conclusion :** Since t= - 0.01191543 > -1.967, fail to reject null hypothesis, H0. There is not sufficient evidence that speed of the vehicle(km/h) affects the fuel consumption(l/100km).

1. **Chi Square Test of Independence**

Chi-Square test is conducted on the relationship between the variables aircond usage and fuel consumption.

**H0 :** Fuel Consumption is independent of aircond usage

**H1 :** Fuel Consumption is dependent of aircond usage



 



The X-Squared value obtained from Chi-Square Test is 254.7 with the degree of freedom equal to 1.



**Conclusion :** Since 254.7 > 3.841459, reject null hypothesis, HO. There is sufficient evidence that the AC Usage influences fuel consumption (l/100km).

***Discussion***

In analysing the factors that influences car fuel consumption based on the dataset collected, we were able to understand the contributing factors towards car fuel consumption. There were several aspects analysed in this case study in order to identify the greatest contributor in car fuel consumption. Based on the results, we were able to make several assumptions on the case study. From the 2-sample hypothesis testing which involves the gas type and fuel consumption, it can be found that different gas type can possibly influence car fuel consumption. Therefore, it is important for car drivers to properly identify the right fuel type for their vehicle with efficient fuel consumption in order to save costs on car fuel.

Next, by conducting correlation analysis on distance travelled with car fuel consumption, we can assume that distance travelled by the vehicle does not have much influence on the car fuel consumption based on the data collected. This might not be true in general case, but the analysis says quite the opposite. This could be because of several reasons from collecting the data by the dataset owner who is also a car driver. Apart from that, Regression test which was conducted to identify the relationship between speed of car (km/h) and fuel consumption, the analysis output says that the speed of the car does not influences car fuel consumption. This could be true in some cases but car drivers should also keep in mind that other aspects of the car specifications such as the engine capacity could possibly affect the fuel consumption from speed travelled by the car.

In the Chi-Square test of Independence, we can confirm that the aircond (AC) usage in the car affects car fuel consumption. Therefore, it is important to have a good practise in using the AC in the car. In the case of running out of fuel, turning off the AC usage can effectively reduce the car fuel consumption and help move the vehicle a few extra miles. From this we can see that different factors have different level of influence towards the car fuel consumption.

***Conclusion***

As a conclusion, it is important to understand the factors that could possibly affect the fuel consumption in our car. The statistical analysis conducted helped me to understand the different factors that has a major role in car fuel consumption. By conducting the analysis using several different statistical analysis techniques makes the interpretation of the results easy to understand and reflect on the point the test output trying to hypothesize.

***References***

1. M Redsell (1993, Jan 1). Factors affecting Car Fuel Consumption. Retrieved from [https://doi.org/10.1243%2FPIME\_PROC\_1993\_207\_155\_02](https://doi.org/10.1243/PIME_PROC_1993_207_155_02)
2. David L.Greene (2007, May 17). Estimated Speed/Fuel Consumption relationships for a large sample of cars. Retrieved from [https://doi.org/10.1016/0360-5442(81)90006-2](https://doi.org/10.1016/0360-5442%2881%2990006-2)
3. Sylvain Weber (2018, Oct 16). Travel distance, fuel efficiency, and vehicle weight: An estimation of the rebound effect. Retrieved from [file:///C:/Users/ASUS/Downloads/WP14-03.pdf](file:///C%3A%5CUsers%5CASUS%5CDownloads%5CWP14-03.pdf)