**PROPOSAL OF PROJECT 2**

**PROBABILITY AND STATISTICAL DATA ANALYSIS**

**SECI / SCSI 2143**

|  |  |  |
| --- | --- | --- |
| Source of dataset (URL) | : | <https://www.kaggle.com/anderas/car-consume/data?select=measurements.csv> |
| Dataset description (collected by who and for what purpose?) | : | This dataset obtained from Kaggle is collected by an American car driver named Andreas Wagener. The purpose is to analyze the fuel consumption by his car on different gas types namely E10 and SP98. He wrote down data of his car’s display after each ride in regular basis and also changing the gas type quite often. |
| Variables (name of variables and type – categorical / ordinal / interval / ratio) | : | 1. Distance travelled - Interval 2. Fuel consumption(L/100km) – Interval 3. Average speed(km/h) – Interval 4. Temperature inside car - Interval 5. Temperature outside car – Interval 6. Gas type – Nominal 7. Aircond usage – Nominal 8. Climate - Nominal |
| Description of purpose of study | : | The study is conducted to find the factors that influence the fuel consumption of car. Different aspects of the car that might play a part in the fuel consumption is taken into consideration and all the relevant data are collected accordingly. |
| Specification of target population | : | The target population for this study is car owners and drivers. |
| Selection of variables (potential variables that will be selected for analysis) | : | 1. Distance travelled 6. Climate 2. Fuel consumption 3. Average speed 4. Gas type 5. Aircond usage |
| Proposed analysis (potential statistical test analysis related to the variables chosen) | : | 1. 2-sample HT : Fuel consumption and distance travelled 2. Correlation : Fuel Consumption and average speed 3. Regression : aircond usage, gas type and fuel consumption 4. Chi-Square test of independence : climate and fuel consumption |
| Expected outcome for analysis | : | We will be able to identify the factor that influences car fuel consumption the most. All the variables have some degree of influence towards fuel consumption. However, we will be able to identify the most effective way to reduce fuel consumption by identifying the most impactful factor in car fuel consumption. |

*Note: Submit this proposal within 2 weeks after briefing done by the lecturer.*

Test Analysis

1. **HT 2 Sample (mean with unknown variance ~ Equal Variance)**

🡪 E10 & SP98

\*Significance Level : Alpha = 0.05

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. 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 (distance, consumption)**

\*Significance Level : Alpha = 0.05

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**

* dependent var, y:gas consumption; independent var, x:speed

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, 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

* Does AC Usage influences fuel consumption

X2 = 254.7, DF = 1

X2.aplha = 3.841459

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