**1. Research Topic and Questions**

**Research Topic**

**The Influence of Demographics and Financial Factors on Car Purchase Price Among Indian Automobile Buyers**

**Research Questions and Hypotheses**

**RQ1: Does the type of profession (Salaried vs. Business) significantly influence the average car price?**

* **Null Hypothesis (H₀1)**: Profession does not significantly affect car price.
* **Alternative Hypothesis (Hₐ1)**: Profession significantly affects car price.

**RQ2: Is there a significant difference in car price based on marital status (Married vs. Single)?**

* **Null Hypothesis (H₀2)**: Marital status does not significantly influence car price.
* **Alternative Hypothesis (Hₐ2)**: Marital status significantly influences car price.

**RQ3: Does total salary significantly predict car price, while controlling for other factors such as profession, education level, and loan status?**

* **Null Hypothesis (H₀3)**: Total salary and other factors do not significantly predict car price.
* **Alternative Hypothesis (Hₐ3)**: Total salary and other factors significantly predict car price.

**2. Descriptive Statistics**

**Key Variables to Summarize:**

* **Age**, **Total Salary**, **Car Price** (continuous variables):
  + Mean, median, mode, standard deviation, and range.
* **Categorical Variables** (e.g., Profession, Marital Status, Education Level):
  + Frequency distributions.

**Steps in Excel:**

1. Use **Data Analysis ToolPak > Descriptive Statistics** to compute:
   * Mean, standard deviation, median, range, and more for continuous variables.
2. Use **COUNTIF** or pivot tables for categorical data frequency distribution.

**3. One-Way ANOVA**

**RQ1: Influence of Profession on Car Price**

**Objective**: Compare the average car prices of salaried vs. business professionals.

**Steps in Excel:**

1. Organize your data:
   * Create two columns: One for "Salaried" car prices, another for "Business" car prices.
2. Go to **Data > Data Analysis > ANOVA: Single Factor**:
   * Input Range: Include the two groups (Salaried and Business car prices).
   * Output: Select a range for displaying results.
3. Interpret Results:
   * Look for **p-value**: If < 0.05, reject the null hypothesis (H₀1).

**4. Independent Samples T-Test**

**RQ2: Influence of Marital Status on Car Price**

**Objective**: Determine if car prices differ significantly between married and single buyers.

**Steps in Excel:**

1. Organize data:
   * Separate car prices into two groups: "Married" and "Single".
2. Go to **Data > Data Analysis > t-Test: Two-Sample Assuming Equal Variances**:
   * Variable 1 Range: Married car prices.
   * Variable 2 Range: Single car prices.
   * Output: Select a range for displaying results.
3. Interpret Results:
   * Check the **t-statistic** and **p-value**: If p-value < 0.05, reject H₀2.

**5. Multiple Linear Regression**

**RQ3: Predicting Car Price Using Total Salary and Other Variables**

**Objective**: Assess how demographic and financial factors influence car price.

**Variables:**

* **Dependent Variable**: Car Price
* **Independent Variables**: Total Salary, Profession, Education Level, Personal Loan Status, House Loan Status.

**Steps in Excel:**

1. Ensure all independent variables are coded numerically:
   * Profession: 1 (Salaried), 0 (Business)
   * Education: 1 (Post Graduate), 0 (Graduate),
2. Go to **Data > Data Analysis > Regression**:
   * **Y Range**: Select the column for car price.
   * **X Range**: Select columns for Total Salary, Profession, Education, Personal Loan, House Loan.
   * Check **Labels** if headers are included.
   * Output: Select a range for the regression output.
3. Interpret Results:
   * **R-squared**: Indicates how well the model explains car price variability.
   * **Coefficients**: Show the influence of each variable on car price.
   * **p-values**: Identify statistically significant predictors (p < 0.05).

**6. Analysis Outputs and Interpretation**

**Descriptive Statistics:**

* Summarize key findings from the descriptive statistics (e.g., average car price, total salary).

**One-Way ANOVA:**

* Report whether the p-value shows a significant difference in car prices between salaried and business professionals.

**T-Test:**

* State whether marital status significantly affects car price.

**Multiple Regression:**

* Highlight which variables (e.g., total salary, profession) significantly predict car price.

Dataset used was provided by mentor/supervisor, I was never collected from the field (primary). Cleaning and coding was done on the data to ensure successful calculation of statistical methods to achieve the research objectives and test hypotheses. Sample size = 99 observations and variables are present oin the datahead below. Excel was used, (Data analysis plugin, pivot tables) for carrying out statistical methods (name them, they are below)

DataHead

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Profession | Salaried | Marrital Status | Married\_ | Education | post\_grad | No of Dependents | Personal loan | P\_loan | House Loan | H\_loan | Wife Working | Wife works | Salary | Wife Salary | Total Salary | Make | Price |
| 27 | Salaried | 1 | Single | 0 | Post Graduate | 1 | 0 | Yes | 1 | No | 0 | No | 0 | 800000 | 0 | 800000 | i20 | 800000 |
| 35 | Salaried | 1 | Married | 1 | Post Graduate | 1 | 2 | Yes | 1 | Yes | 1 | Yes | 1 | 1400000 | 600000 | 2000000 | Ciaz | 1000000 |
| 45 | Business | 0 | Married | 1 | Graduate | 0 | 4 | Yes | 1 | Yes | 1 | No | 0 | 1800000 | 0 | 1800000 | Duster | 1200000 |
| 41 | Business | 0 | Married | 1 | Post Graduate | 1 | 3 | No | 0 | No | 0 | Yes | 1 | 1600000 | 600000 | 2200000 | City | 1200000 |
| 31 | Salaried | 1 | Married | 1 | Post Graduate | 1 | 2 | Yes | 1 | No | 0 | Yes | 1 | 1800000 | 800000 | 2600000 | SUV | 1600000 |
| 28 | Salaried | 1 | Married | 1 | Graduate | 0 | 3 | Yes | 1 | Yes | 1 | No | 0 | 900000 | 0 | 900000 | Baleno | 700000 |
| 31 | Salaried | 1 | Married | 1 | Graduate | 0 | 4 | No | 0 | No | 0 | Yes | 1 | 1200000 | 600000 | 1800000 | City | 1200000 |
| 33 | Business | 0 | Married | 1 | Post Graduate | 1 | 4 | No | 0 | No | 0 | No | 0 | 1400000 | 0 | 1400000 | Baleno | 700000 |
| 34 | Business | 0 | Married | 1 | Post Graduate | 1 | 4 | No | 0 | No | 0 | No | 0 | 2000000 | 0 | 2000000 | Verna | 1100000 |
| 34 | Salaried | 1 | Married | 1 | Graduate | 0 | 3 | Yes | 1 | Yes | 1 | Yes | 1 | 1200000 | 700000 | 1900000 | i20 | 800000 |
| 35 | Salaried | 1 | Married | 1 | Post Graduate | 1 | 4 | No | 0 | No | 0 | Yes | 1 | 1300000 | 700000 | 2000000 | SUV | 1600000 |
| 35 | Salaried | 1 | Married | 1 | Graduate | 0 | 4 | Yes | 1 | Yes | 1 | m | 1 | 1400000 | 0 | 1400000 | Baleno | 700000 |

Total number of car makes;

|  |  |
| --- | --- |
| **Make** | **Car\_code** |
| i20 | 1 |
| Ciaz | 2 |
| Duster | 3 |
| City | 4 |
| SUV | 5 |
| Baleno | 6 |
| Verna | 7 |
| Luxuray | 8 |
| Creata | 9 |

Coding and data cleaning;

|  |  |
| --- | --- |
| Personal loan | P\_loan |
| Yes | 1 |
| No | 0 |

|  |  |
| --- | --- |
| Marrital Status | Married\_ |
| Single | 0 |
| Married | 1 |

|  |  |
| --- | --- |
| Profession | Salaried |
| Salaried | 1 |
| Business | 0 |

|  |  |
| --- | --- |
| Education | post\_grad |
| Post Graduate | 1 |
| Graduate | 0 |

|  |  |
| --- | --- |
| House Loan | H\_loan |
| No | 0 |
| Yes | 1 |

|  |  |
| --- | --- |
| Wife Working | Wife works |
| No | 0 |
| Yes | 1 |

Sample size = 99 observations

Tasks

* Use reverse engineering to come up with research topic
* Develop research topic and questions (relevant followed by Null and alternative hypotheses for each question)
* Use Descriptive statistics
* One way ANOVA
* Independent sample t-test
* Multiple linear regression involving one key variable and all other relevant variables

Results

1. Descriptive statistics for Categorical variables

|  |  |
| --- | --- |
| **Row Labels** | **Count of Profession** |
| Business | 35 |
| Salaried | 64 |
| **Grand Total** | **99** |

|  |  |
| --- | --- |
| **Row Labels** | **Count of Marrital Status** |
| Married | 84 |
| Single | 15 |
| **Grand Total** | **99** |
|  |  |
|  |  |
| **Row Labels** | **Count of Education** |
| Graduate | 43 |
| Post Graduate | 56 |
| **Grand Total** | **99** |

|  |  |
| --- | --- |
| **Row Labels** | **Count of Make** |
| Baleno | 19 |
| Ciaz | 12 |
| City | 10 |
| Creata | 14 |
| Duster | 7 |
| i20 | 12 |
| Luxuray | 2 |
| SUV | 19 |
| Verna | 4 |
| **Grand Total** | **99** |

|  |  |
| --- | --- |
| **Row Labels** | **Count of Personal loan** |
| No | 67 |
| Yes | 32 |
| **Grand Total** | **99** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Personal loan** | **Column Labels** |  |  |
| **Row Labels** | **Business** | **Salaried** | **Grand Total** |
| Baleno | 8 | 11 | 19 |
| Ciaz | 4 | 8 | 12 |
| City | 3 | 7 | 10 |
| Creata | 3 | 11 | 14 |
| Duster | 4 | 3 | 7 |
| i20 | 5 | 7 | 12 |
| Luxuray |  | 2 | 2 |
| SUV | 6 | 13 | 19 |
| Verna | 2 | 2 | 4 |
| **Grand Total** | **35** | **64** | **99** |

|  |  |
| --- | --- |
| **Row Labels** | **Count of Wife Working** |
| Baleno | 19 |
| Ciaz | 12 |
| City | 10 |
| Creata | 14 |
| Duster | 7 |
| i20 | 12 |
| Luxuray | 2 |
| SUV | 19 |
| Verna | 4 |
| **Grand Total** | **99** |

|  |  |
| --- | --- |
| **Row Labels** | **Count of Wife Working** |
| Baleno | 19 |
| Ciaz | 12 |
| City | 10 |
| Creata | 14 |
| Duster | 7 |
| i20 | 12 |
| Luxuray | 2 |
| SUV | 19 |
| Verna | 4 |
| **Grand Total** | **99** |

1. Descriptive Statistics for Continuous variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Age* | *Price* | *Salary* | *Wife Salary* | *Total Salary* |
|  |  |  |  |  |  |
| Mean | 36.31313131 | 1194040.404 | 1736363.636 | 534343.4343 | 2270707.07 |
| Standard Error | 0.627752067 | 43990.05697 | 67701.53108 | 60849.50762 | 105607.103 |
| Median | 36 | 1200000 | 1600000 | 500000 | 2100000 |
| Mode | 36 | 1600000 | 1400000 | 0 | 1400000 |
| Standard Deviation | 6.246054207 | 437695.5404 | 673621.729 | 605444.9563 | 1050777.41 |
| Sample Variance | 39.01319316 | 1.91577E+11 | 4.53766E+11 | 3.66564E+11 | 1.1041E+12 |
| Kurtosis | -0.263890781 | 4.198933895 | -0.010220746 | -0.498407118 | 0.47977443 |
| Skewness | 0.534932463 | 1.141602503 | 0.584393751 | 0.765588432 | 0.88007404 |
| Range | 25 | 2890000 | 3600000 | 2100000 | 5000000 |
| Minimum | 26 | 110000 | 200000 | 0 | 200000 |
| Maximum | 51 | 3000000 | 3800000 | 2100000 | 5200000 |
| Sum | 3595 | 118210000 | 171900000 | 52900000 | 224800000 |
| Count | 99 | 99 | 99 | 99 | 99 |

One Way ANOVA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Anova: Single Factor** | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Business | 35 | 39200000 | 1120000 | 1.15E+11 |  |  |
| Salaried | 35 | 44110000 | 1260286 | 3.33E+11 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 3.44E+11 | 1 | 3.44E+11 | 1.539032 | 0.219025 | 3.981896256 |
| Within Groups | 1.52E+13 | 68 | 2.24E+11 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 1.56E+13 | 69 |  |  |  |  |

**Independent Samples T-Test -**

|  |  |  |
| --- | --- | --- |
| **t-Test: Two-Sample Assuming Equal Variances** | |  |
|  |  |  |
|  | *Married* | *Single* |
| Mean | 1221547.619 | 1040000 |
| Variance | 1.98784E+11 | 1.32571E+11 |
| Observations | 84 | 15 |
| Pooled Variance | 1.89228E+11 |  |
| Hypothesized Mean Difference | 0 |  |
| df | 97 |  |
| t Stat | 1.488900709 |  |
| P(T<=t) one-tail | 0.069878826 |  |
| t Critical one-tail | 1.66071461 |  |
| P(T<=t) two-tail | 0.139757653 |  |
| t Critical two-tail | 1.984723186 |  |

**Multiple Linear Regression**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.788095 |  |  |  |  |  |  |  |
| R Square | 0.621093 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.566672 |  |  |  |  |  |  |  |
| Standard Error | 284321.7 |  |  |  |  |  |  |  |
| Observations | 99 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 11 | 1.16608E+13 | 1.06007E+12 | 14.42471827 | 2.14839E-15 |  |  |  |
| Residual | 88 | 7.11381E+12 | 80838801234 |  |  |  |  |  |
| Total | 99 | 1.87746E+13 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 721247 | 120274.2767 | 5.996685645 | 4.35406E-08 | 482227.1774 | 960266.8795 | 482227.2 | 960266.9 |
| Wife Salary | 0.11478 | 0.154661494 | 0.742134899 | 0.45998187 | -0.192577528 | 0.422136913 | -0.19258 | 0.422137 |
| Total Salary | 0.270907 | 0.060440063 | 4.48223928 | 2.21506E-05 | 0.1507949 | 0.391018749 | 0.150795 | 0.391019 |
| Salary | 0 | 0 | 65535 | #NUM! | 0 | 0 | 0 | 0 |
| No of Dependents | 34482.36 | 28709.11136 | 1.201094605 | #NUM! | -22570.96726 | 91535.68481 | -22571 | 91535.68 |
| Salaried | 19090.87 | 64416.60399 | 0.296365657 | 0.767649156 | -108923.595 | 147105.3334 | -108924 | 147105.3 |
| Married\_ | -221349 | 120255.8173 | -1.840649849 | 0.069043076 | -460332.0188 | 17634.31496 | -460332 | 17634.31 |
| post\_grad | -36323.3 | 59425.20689 | -0.611244559 | 0.542613447 | -154418.4457 | 81771.77699 | -154418 | 81771.78 |
| Personal\_loan | -130546 | 77532.14485 | -1.683760898 | 0.095772148 | -284624.4395 | 23533.25186 | -284624 | 23533.25 |
| House\_loan | -142526 | 77881.68985 | -1.830029782 | 0.07063118 | -297299.3048 | 12247.681 | -297299 | 12247.68 |
| Wife works | -36738.7 | 131046.5447 | -0.280348685 | 0.779868 | -297166.1965 | 223688.7436 | -297166 | 223688.7 |
| Car\_make | 6702.345 | 12656.68366 | 0.52954982 | 0.59775753 | -18450.1546 | 31854.84372 | -18450.2 | 31854.84 |