**Question 1 (21 points)**

**A HomeFix has recently introduced a new hardware brand. The firm would like to estimate the proportion of people in its target market segment who are aware of the new brand. As part of a larger market research study, it was found that in a sample of 140 randomly selected individuals from the target market segment, 92 individuals were aware of the firm’s new brand.**

**The manager in charge of the new brand has stated that the brand awareness is greater than 0.6, meaning that more than 60% of the population is aware of the brand. At the 5% significance level, conduct a hypothesis test with the goal of proving his claim.**

**(4 points) Specify the null hypothesis and the alternative hypothesis. Define the parameter of your interest and write the mathematical expressions of your hypotheses in terms of your parameter (Statements in English are NOT accepted).**

**(4 points) Use StatTools to conduct the hypothesis test. Copy and paste the report generated by StatTools. You need to create the summary table of the survey data in Excel first; then you can run the analysis.**

**(3 points) Provide the results of the test in “plain English” (interpretation of "Reject"/"Don't Reject" in English) AND the meaning of the p-value.**

**(3 points) State the Type I and Type II errors of the hypothesis test in “plain English”.**

**(4 points) Construct a 95% confidence interval estimate for the population proportion of people who are aware of the new hardware brand. Copy and paste the report generated by StatTools. Interpret the confidence interval within the context of the problem in English.**

**(3 points) For future polls, the firm is interested in minimizing their marketing-research costs. The margin of error (MOE) in their pools should be no larger than B (i.e., ± 100B percentage points). To simplify the analysis, we will assume that their marketing study only has a single question that is used to estimate the proportion, p at the 5% confidence level. Based on the results in part d, estimate the sample size for future polls to ensure that MOE (=B) is no larger than 0.05 or 5%.**

1. **Specify the null hypothesis and the alternative hypothesis. Define the parameter of your interest and write the mathematical expressions of your hypotheses in terms of your parameter**

**Parameter:** Proportion of people in its target market segment who are aware of the new brand.

**Hypothesis:**

**H0:** Brand awareness is less than or equal to 0.6, meaning that less than equal to 60% of the population is aware of the brand.

**Ha:** Brand awareness is greater than 0.6, meaning that more than 60% of the population is aware of the brand

**H0:** Brand Awareness p̂ <= 0.6

**Ha:** Brand Awareness p̂ > 0.6

1. **Use StatTools to conduct the hypothesis test. Copy and paste the report generated by StatTools. You need to create the summary table of the survey data in Excel first; then you can run the analysis.**

|  |  |
| --- | --- |
| ***Hypothesis Test (Proportion)*** | **Data Set #1** |
| **Category** | **Brand Awareness** |
| **Sample Size** | 140 |
| **Sample Proportion** | 0.657 |
| **Hypothesized Proportion** | 0.6 |
| **Alternative Hypothesis** | > 0.6 |
| **Standard Error of Sample Proportion** | 0.041 |
| **z-Test Statistic** | 1.3801 |
| **p-Value** | 0.0838 |
| **Null Hypoth. at 10% Significance** | Reject |
| **Null Hypoth. at 5% Significance** | Don't Reject |
| **Null Hypoth. at 1% Significance** | Don't Reject |

1. **Provide the results of the test in “plain English” (interpretation of "Reject"/"Don't Reject" in English) AND the meaning of the p-value.**

We get the values of standard Error, z-test statistic and finally the p-value by stattools.

p-value > significant value (0.0838 > 0.05) **don’t Reject H0(Null hypothesis)**

As we can see that the p-value is greater than significant value i.e., Alpha = 0.5, we accept the Null hypothesis or we **don’t reject** the **null hypothesis**. And the population brand awareness is less than 60%.

1. **State the Type I and Type II errors of the hypothesis test in “plain English”.**

**Type I error:** type I error occurs when Null Hypothesis is rejected but actually it is true.

**Type II error:** type II error occurs when we don’t reject the Null Hypothesis, when actually it is not true.

1. **Construct a 95% confidence interval estimate for the population proportion of people who are aware of the new hardware brand. Copy and paste the report generated by StatTools. Interpret the confidence interval within the context of the problem in English.**

|  |  |
| --- | --- |
| ***Conf. Interval (Proportion)*** | **Data Set #1** |
| **Category** | Brand Awareness |
| **Sample Size** | 140 |
| **Sample Proportion** | 0.657 |
| **Confidence Level** | 95.0% |
| **Standard Error of Proportion** | 0.040 |
| **Lower Limit** | 0.579 |
| **Upper Limit** | 0.736 |

Here we have result from stattools the resulting confidence intervals, we observe the sample proportion is 0.657 (p=0.657), which we knew, and lower limits and the upper of the interval are 0. 579 and 0.736 at confidence level 95.0%.

We can get Margin of error (MOE) from the lower and upper limit. i.e., half of the width of the interval.

MOE =

=

MOE = 0.0785

The margin of error is 0.0785 or we can say 7.85%.

1. **For future polls, the firm is interested in minimizing their marketing-research costs. The margin of error (MOE) in their pools should be no larger than B (i.e., ± 100B percentage points). To simplify the analysis, we will assume that their marketing study only has a single question that is used to estimate the proportion, p at the 5% confidence level. Based on the results in part d, estimate the sample size for future polls to ensure that MOE (=B) is no larger than 0.05 or 5%.**

.

The value of z-multiple is 1.96 at confidence interval 95%, and the estimated p̂ = 0.6.

N = ( )2 p̂ (1- p̂)

= ( )2 \*0.6 (1-0.6)

= 623.40\* 0.24

N =149

N=149 for future polls to ensure that MOE (=B) is no larger than 0.05 or 5%.

**Question 2 (29 points)**

**SmithCo sells heating oil to residential customers and would like to build a model to predict its customer’s oil consumption.**

**Oil customers are exposed to the risk of running out of fuel. Home heating oil suppliers therefore have to guarantee that the customer’s oil tank will not be allowed to run dry. Home heating oil industry try uses the concept of a degree-day, equal to the difference between the average daily temperature and 68 degree Fahrenheit. So if the average temperature on a given day is 50, the degree-days for that day will be 18. If the degree-day calculation results in a negative number, the degree-day number is recorded as 0.**

**By keeping track of the number of degree-days since the customer’s last oil fill, knowing the size of the customer’s oil tank, and estimating the customer’s oil consumptions as a function of the number of degree-days, the oil supplier can estimate when the customer is getting low on fuel and then resupply the customer. However, SmithCo has more than 2000 customers and computational burden of keeping track of all of these customers is enormous.**

**SmithCo wants to develop a consumption estimation model that is practical and reliable.**

**The file ‘Oil usage.xlsx Download Oil usage.xlsxDownload Oil usage.xlsx’ contains recent oil usage of 40 customers recent with the following variables:**

**Oil Usage: The oil consumption amounts in gallons for 40 customers.**

**DegreeDays: The number of degree-days since the last oil fill for 40 customers.**

**HomeFactor: An assessment of the home type of each of the 40 customers (levels={1,2,3,4,5}).**

**NumberPeople: The number of people residing in the home of each of the 40 customers.**

**Use StatTools to conduct the statistical analysis asked below. For questions that ask for an oil usage (or change in oil usage), use zero decimal places in your final numerical answer.**

**Solution:**

**Part A – Linear Regression (7 points)**

**Create a regression model for Oil Usage using all three variables (Degree Days, Home Factor, and Number People) as the independent variables. Let us call this Model A. Also, create the scatter plot of fit vs. Oil Usage and the residual plot when you generate the regression output.**

1. **(5 pts.) Include the Stattools regression output as Exhibit A. Copy and paste the regression output, and the plots. Write out the estimated regression equation (copy and paste the equation from Stattools report).**

Exhibit

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Multiple** | **R-Square** | **Adjusted** | **StErr of** |  |  |
| ***Summary*** | **R** | **R-Square** | **Estimate** |  |  |
|  | 0.8855 | 0.7840 | 0.7660 | 85.46750087 |  |  |
|  |  |  |  |  |  |  |
|  | **Degrees of** | **Sum of** | **Mean of** | **F-Ratio** | **p-Value** |  |
| ***ANOVA Table*** | **Freedom** | **Squares** | **Squares** |  |
| **Explained** | 3 | 954738.9266 | 318246.3089 | 43.5674 | < 0.0001 |  |
| **Unexplained** | 36 | 262968.9734 | 7304.693706 |  |  |  |
|  |  |  |  |  |  |  |
|  | **Coefficient** | **Standard** | **t-Value** | **p-Value** | **Confidence Interval 95%** | |
| ***Regression Table*** | **Error** | **Lower** | **Upper** |
| **Constant** | -218.30987 | 63.95851 | -3.4133 | 0.0016 | -348.0237 | -88.5960 |
| **DegreeDays** | 0.27508 | 0.03633 | 7.5711 | < 0.0001 | 0.2014 | 0.3488 |
| **HomeFactor** | 86.98875 | 9.63044 | 9.0327 | < 0.0001 | 67.4573 | 106.5202 |
| **NumberPeople** | 5.26724 | 10.56179 | 0.4987 | 0.6210 | -16.1531 | 26.6876 |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ***Graph Data*** | **OilUsage** | **Fit** | **Residual** |  |  |  |
| **1** | 381 | 302.7285 | 78.2715 |  |  |  |
| **2** | 171 | 301.9185 | -130.9185 |  |  |  |
| **3** | 644 | 532.8629 | 111.1371 |  |  |  |
| **4** | 19 | 11.3966 | 7.6034 |  |  |  |
| **5** | 394 | 420.3962 | -26.3962 |  |  |  |
| **6** | 153 | 250.9244 | -97.9244 |  |  |  |
| **7** | 7 | 222.5530 | -215.5530 |  |  |  |
| **8** | 319 | 311.7831 | 7.2169 |  |  |  |
| **9** | 40 | 117.7300 | -77.7300 |  |  |  |
| **10** | 121 | -2.5740 | 123.5740 |  |  |  |
| **11** | 243 | 261.4666 | -18.4666 |  |  |  |
| **12** | 200 | 297.7311 | -97.7311 |  |  |  |
| **13** | 402 | 398.0511 | 3.9489 |  |  |  |
| **14** | 118 | 206.9549 | -88.9549 |  |  |  |
| **15** | 319 | 344.0311 | -25.0311 |  |  |  |
| **16** | 185 | 98.0058 | 86.9942 |  |  |  |
| **17** | 209 | 308.3982 | -99.3982 |  |  |  |
| **18** | 467 | 451.9896 | 15.0104 |  |  |  |
| **19** | 50 | 11.9467 | 38.0533 |  |  |  |
| **20** | 153 | 84.0582 | 68.9418 |  |  |  |
| **21** | 94 | 123.2239 | -29.2239 |  |  |  |
| **22** | 574 | 489.0845 | 84.9155 |  |  |  |
| **23** | 191 | 150.1410 | 40.8590 |  |  |  |
| **24** | 679 | 529.5389 | 149.4611 |  |  |  |
| **25** | 305 | 328.1809 | -23.1809 |  |  |  |
| **26** | 85 | 57.7321 | 27.2679 |  |  |  |
| **27** | 87 | 123.9217 | -36.9217 |  |  |  |
| **28** | 170 | 174.8981 | -4.8981 |  |  |  |
| **29** | 92 | 76.2516 | 15.7484 |  |  |  |
| **30** | 35 | -13.6764 | 48.6764 |  |  |  |
| **31** | 60 | -18.6100 | 78.6100 |  |  |  |
| **32** | 507 | 392.4680 | 114.5320 |  |  |  |
| **33** | 148 | 166.0089 | -18.0089 |  |  |  |
| **34** | 83 | 255.5423 | -172.5423 |  |  |  |
| **35** | 318 | 316.5232 | 1.4768 |  |  |  |
| **36** | 85 | -5.9971 | 90.9971 |  |  |  |
| **37** | 245 | 204.5659 | 40.4341 |  |  |  |
| **38** | 56 | 69.1225 | -13.1225 |  |  |  |
| **39** | 303 | 267.2432 | 35.7568 |  |  |  |
| **40** | 10 | 103.4844 | -93.4844 |  |  |  |

**Estimate Regression Equation:**

Oil usage = -218.3099 + **DegreeDays**\*(0.2751) + **HomeFactor**\*(86.9887) + **NumberPeople**\*(5.2672)

This is the Estimated Regression Equation to predict Oil Usage. The value of the Oil usage (dependent variable) is predicted to me -218.3099 (constant) when all the independent variables (Degree Days, Home Factor, Number People) are zero.

1. **(2 pts.) Provide an economic interpretation of the coefficient of NumberPeople.**

Coefficients describe the size of the effect on the dependent variable (Oil usage) by the independent variables, and constant is the value **oil usage** to have when all the independent variables are equal to zero.

**Coefficient of NumberPeople**

The oil usage is predicted to be increased by 5.2672, when NumberPeople variable goes up by one. Means the size of effect of NumberPeople of the Oil Usage is 5.2672.

**Model B – Adding Categorical Variables (13 points)**

**Model A treats the HomeFactor variable as a numerical variable. Build a model, which treats the HomeFactor variable as a categorical variable. Let us refer to this model as Model B. Create the scatter plot of fit vs. Oil Usage and the residual plot.**

1. **(5 pts.) Include the Stattools regression output as Exhibit B. Copy and paste the regression output, and the plots. Write out the estimated regression equation (copy and paste the equation from Stattools report).**

Exhibit

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Multiple** | **R-Square** | **Adjusted** | **StErr of** |  |  |
| ***Summary*** | **R** | **R-Square** | **Estimate** |  |  |
|  | 0.8967 | 0.8040 | 0.7684 | 85.03688466 |  |  |
|  |  |  |  |  |  |  |
|  | **Degrees of** | **Sum of** | **Mean of** | **F-Ratio** | **p-Value** |  |
| ***ANOVA Table*** | **Freedom** | **Squares** | **Squares** |  |
| **Explained** | 6 | 979075.9321 | 163179.322 | 22.5658 | < 0.0001 |  |
| **Unexplained** | 33 | 238631.9679 | 7231.271753 |  |  |  |
|  |  |  |  |  |  |  |
|  | **Coefficient** | **Standard** | **t-Value** | **p-Value** | **Confidence Interval 95%** | |
| ***Regression Table*** | **Error** | **Lower** | **Upper** |
| **Constant** | -190.7429 | 79.7308 | -2.3923 | 0.0226 | -352.9565 | -28.5294 |
| **DegreeDays** | 0.2906 | 0.0417 | 6.9633 | < 0.0001 | 0.2057 | 0.3756 |
| **NumberPeople** | 9.8934 | 11.6392 | 0.8500 | 0.4014 | -13.7867 | 33.5736 |
| **HomeFactor = 2** | 144.3963 | 46.3486 | 3.1154 | 0.0038 | 50.0994 | 238.6933 |
| **HomeFactor = 3** | 217.7831 | 41.3350 | 5.2687 | < 0.0001 | 133.6863 | 301.8798 |
| **HomeFactor = 4** | 314.8318 | 46.5896 | 6.7575 | < 0.0001 | 220.0444 | 409.6191 |
| **HomeFactor = 5** | 347.6091 | 44.2600 | 7.8538 | < 0.0001 | 257.5615 | 437.6566 |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ***Graph Data*** | **OilUsage** | **Fit** | **Residual** |  |  |  |
| **1** | 381 | 314.8038 | 66.1962 |  |  |  |
| **2** | 171 | 277.2719 | -106.2719 |  |  |  |
| **3** | 644 | 508.2906 | 135.7094 |  |  |  |
| **4** | 19 | 29.8471 | -10.8471 |  |  |  |
| **5** | 394 | 393.7925 | 0.2075 |  |  |  |
| **6** | 153 | 278.1963 | -125.1963 |  |  |  |
| **7** | 7 | 196.1270 | -189.1270 |  |  |  |
| **8** | 319 | 347.2198 | -28.2198 |  |  |  |
| **9** | 40 | 129.2084 | -89.2084 |  |  |  |
| **10** | 121 | -24.4169 | 145.4169 |  |  |  |
| **11** | 243 | 271.2086 | -28.2086 |  |  |  |
| **12** | 200 | 284.2130 | -84.2130 |  |  |  |
| **13** | 402 | 429.3143 | -27.3143 |  |  |  |
| **14** | 118 | 188.3036 | -70.3036 |  |  |  |
| **15** | 319 | 362.7703 | -43.7703 |  |  |  |
| **16** | 185 | 117.0255 | 67.9745 |  |  |  |
| **17** | 209 | 271.1329 | -62.1329 |  |  |  |
| **18** | 467 | 422.8441 | 44.1559 |  |  |  |
| **19** | 50 | 30.4284 | 19.5716 |  |  |  |
| **20** | 153 | 110.9460 | 42.0540 |  |  |  |
| **21** | 94 | 138.1337 | -44.1337 |  |  |  |
| **22** | 574 | 457.7083 | 116.2917 |  |  |  |
| **23** | 191 | 162.2445 | 28.7555 |  |  |  |
| **24** | 679 | 568.2375 | 110.7625 |  |  |  |
| **25** | 305 | 355.4932 | -50.4932 |  |  |  |
| **26** | 85 | 91.7879 | -6.7879 |  |  |  |
| **27** | 87 | 104.9035 | -17.9035 |  |  |  |
| **28** | 170 | 188.4016 | -18.4016 |  |  |  |
| **29** | 92 | 45.8809 | 46.1191 |  |  |  |
| **30** | 35 | -0.9720 | 35.9720 |  |  |  |
| **31** | 60 | -50.0165 | 110.0165 |  |  |  |
| **32** | 507 | 414.7588 | 92.2412 |  |  |  |
| **33** | 148 | 149.3705 | -1.3705 |  |  |  |
| **34** | 83 | 210.9598 | -127.9598 |  |  |  |
| **35** | 318 | 333.7069 | -15.7069 |  |  |  |
| **36** | 85 | -41.0187 | 126.0187 |  |  |  |
| **37** | 245 | 215.4187 | 29.5813 |  |  |  |
| **38** | 56 | 86.5089 | -30.5089 |  |  |  |
| **39** | 303 | 277.3119 | 25.6881 |  |  |  |
| **40** | 10 | 74.6537 | -64.6537 |  |  |  |

**Estimated regression equation:**

**Oil Usage**=-190.7429 + **DegreeDays** (0.2906) + **NumberPeople** (9.8934) + **HomeFactor=2**(144.3963) + **HomeFactor=3**(217.7831) + **HomeFactor=4**(314.8318) + **HomeFactor=5**(347.6091)

This is the Estimation Regression Equation for the prediction of Oil Usage.

The homefactor=1 is excluded because 2 or more explanatory variables have exact multicollinearity.

1. **(2 pts.) Provide an economic interpretation of the coefficient of (HomeFactor level = 5).**

Coefficients describe the size of the effect on the dependent variable (Oil usage) by the independent variables. constant (-190.7429) is the value oil usage have when all the independent variables are equal to zero.

Coefficient of HomeFactor level = 5:

The oil usage is predicted to be increased by 347.6091, when HomeFactor level = 5. Means the size of effect of HomeFactor level = 5 on the Oil Usage is 347.6091.

1. **(2 pts.) According to Model B estimated above, by how much higher/lower is the average oil consumption of customers in HomeFactor level 2 compared to the average oil consumption of customers in HomeFactor level 4, when DegreeDays and NumberPeople remain the same?**

Coefficients of HomeFactor = 2 is **144.3963** and HomeFactor = 4 is **314.8318.**

When the homefactor level=2 the value of the Oil Usage increased by 144.3963 and on the other hand if the homefactor level=4 the value of the Oil Usage increases by 314.8318. The effect on Oil Usage by level=4 is more than the Level=2.

1. **(2 pts.) Provide an economic interpretation of the coefficient of DegreeDays.**

Coefficient of DegreeDays:

The oil usage is predicted to be increased by 0.2906, when DegreeDays is increased by one. Means the size of effect of DegreeDays on the Oil Usage is 0.2906. The effect of the DegreeDays on oil Usage is very low compared to other.

1. **(2 pts.) Compare the performance of two models (Model A and Model B). Explain why use dummies for HomeFactor instead of the variable itself?**

**Modal A**: R-square =0.7840

**Oil Usage** = -218.3099+DegreeDays\*(0.2751) +HomeFactor\*(86.9887) +NumberPeople\*(5.2672)

**Modal B**: R-square =0.8040

**Oil Usage**= -190.7429 + DegreeDays (0.2906) + NumberPeople (9.8934) + HomeFactor=2(144.3963) + HomeFactor=3(217.7831) + HomeFactor=4(314.8318) + HomeFactor=5(347.6091)

**Why use dummies:**

We are using dummy variables instead of variable itself because it represents multiple groups in a single regression equation. And the effect of every single category is been observed by using dummy variable. Every factor has own different effect of the observed value (oil usage).

R-square is the how well the regression model fitting the observer value (oil Usage).

Also we can see from the comparison of the Modal A and Modal B, the value of R-square for the Model A (0.7840) is less the R square of Model B (0.8040).

**Part C – Adding Interactions (9 points)**

**Next, suppose it is conjectured that the DegreeDays varies by HomeFactor. To account for this conjecture, we augment Model B with interaction terms between DegreeDays and HomeFactor. Let us call this model Model C. Create the scatter plot of fit vs. Oil Usage and the residual plot.**

**a. (5 pts) Include the StatTools regression output as Exhibit C. Copy and paste the regression output, and the plots. Write out the estimated regression equation (copy and paste the equation from StatTools report).**

**b. (2 pts) According to Model C estimated above, by how much higher/lower is the average oil consumption of customers in HomeFactor level 2 compared to the average oil consumption of customers in HomeFactor level 4 when DegreeDays = 1000 and NumberPeople is the same?**

**c. (2 pts) Estimate the oil consumption of a customer with DegreeDays =380, NumberPeople =4, HomeFactor = 1.**

1. **(5 pts) Include the StatTools regression output as Exhibit C. Copy and paste the regression output, and the plots. Write out the estimated regression equation (copy and paste the equation from StatTools report).**

Exhibit

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Multiple** | **R-Square** | **Adjusted** | **StErr of** |  |  |
| ***Summary*** | **R** | **R-Square** | **Estimate** |  |  |
|  | 0.9736 | 0.9479 | 0.9365 | 44.52250698 |  |  |
|  |  |  |  |  |  |  |
|  | **Degrees of** | **Sum of** | **Mean of** | **F-Ratio** | **p-Value** |  |
| ***ANOVA Table*** | **Freedom** | **Squares** | **Squares** |  |
| **Explained** | 7 | 1154275.784 | 164896.5406 | 83.1864 | < 0.0001 |  |
| **Unexplained** | 32 | 63432.11609 | 1982.253628 |  |  |  |
|  |  |  |  |  |  |  |
|  | **Coefficient** | **Standard** | **t-Value** | **p-Value** | **Confidence Interval 95%** | |
| ***Regression Table*** | **Error** | **Lower** | **Upper** |
| **Constant** | -35.4342 | 41.8091 | -0.8475 | 0.4030 | -120.5965 | 49.7281 |
| **DegreeDays** | 0.3464 | 0.0300 | 11.5489 | < 0.0001 | 0.2853 | 0.4074 |
| **HomeFactor** | 2.2153 | 10.1726 | 0.2178 | 0.8290 | -18.5057 | 22.9363 |
| **NumberPeople** | 13.8213 | 5.9427 | 2.3258 | 0.0265 | 1.7165 | 25.9261 |
| **Interaction(DegreeDays,HomeFactor = 1)** | -0.2777 | 0.0356 | -7.7955 | < 0.0001 | -0.3503 | -0.2052 |
| **Interaction(DegreeDays,HomeFactor = 2)** | -0.1193 | 0.0390 | -3.0641 | 0.0044 | -0.1987 | -0.0400 |
| **Interaction(DegreeDays,HomeFactor = 4)** | 0.1171 | 0.0335 | 3.4955 | 0.0014 | 0.0489 | 0.1854 |
| **Interaction(DegreeDays,HomeFactor = 5)** | 0.2288 | 0.0416 | 5.4988 | < 0.0001 | 0.1440 | 0.3136 |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ***Graph Data*** | **OilUsage** | **Fit** | **Residual** |  |  |  |
| **1** | 381 | 320.2417 | 60.7583 |  |  |  |
| **2** | 171 | 173.6188 | -2.6188 |  |  |  |
| **3** | 644 | 648.0699 | -4.0699 |  |  |  |
| **4** | 19 | 52.8849 | -33.8849 |  |  |  |
| **5** | 394 | 415.7244 | -21.7244 |  |  |  |
| **6** | 153 | 207.4566 | -54.4566 |  |  |  |
| **7** | 7 | 92.6054 | -85.6054 |  |  |  |
| **8** | 319 | 300.7802 | 18.2198 |  |  |  |
| **9** | 40 | 112.2136 | -72.2136 |  |  |  |
| **10** | 121 | 86.4565 | 34.5435 |  |  |  |
| **11** | 243 | 268.2880 | -25.2880 |  |  |  |
| **12** | 200 | 136.3788 | 63.6212 |  |  |  |
| **13** | 402 | 450.4158 | -48.4158 |  |  |  |
| **14** | 118 | 113.7271 | 4.2729 |  |  |  |
| **15** | 319 | 379.4359 | -60.4359 |  |  |  |
| **16** | 185 | 114.8851 | 70.1149 |  |  |  |
| **17** | 209 | 178.7425 | 30.2575 |  |  |  |
| **18** | 467 | 478.9740 | -11.9740 |  |  |  |
| **19** | 50 | 53.3389 | -3.3389 |  |  |  |
| **20** | 153 | 122.3238 | 30.6762 |  |  |  |
| **21** | 94 | 115.7913 | -21.7913 |  |  |  |
| **22** | 574 | 553.7267 | 20.2733 |  |  |  |
| **23** | 191 | 142.4939 | 48.5061 |  |  |  |
| **24** | 679 | 671.9701 | 7.0299 |  |  |  |
| **25** | 305 | 332.6861 | -27.6861 |  |  |  |
| **26** | 85 | 119.5470 | -34.5470 |  |  |  |
| **27** | 87 | 105.5145 | -18.5145 |  |  |  |
| **28** | 170 | 173.6661 | -3.6661 |  |  |  |
| **29** | 92 | 68.6053 | 23.3947 |  |  |  |
| **30** | 35 | 22.7189 | 12.2811 |  |  |  |
| **31** | 60 | 57.4411 | 2.5589 |  |  |  |
| **32** | 507 | 431.1162 | 75.8838 |  |  |  |
| **33** | 148 | 116.0166 | 31.9834 |  |  |  |
| **34** | 83 | 65.4192 | 17.5808 |  |  |  |
| **35** | 318 | 344.8001 | -26.8001 |  |  |  |
| **36** | 85 | 48.0815 | 36.9185 |  |  |  |
| **37** | 245 | 203.8323 | 41.1677 |  |  |  |
| **38** | 56 | 91.0490 | -35.0490 |  |  |  |
| **39** | 303 | 275.5615 | 27.4385 |  |  |  |
| **40** | 10 | 75.4008 | -65.4008 |  |  |  |

**Estimated regression equation:**

**Model C**

**Oil Usage** = -35.4342 + **DegreeDays**(0.3464) + **HomeFactor**(2.2153) + **NumberPeople**(13.8213) + **Interaction(DegreeDays,HomeFactor = 1)**(-0.2777) + **Interaction(DegreeDays,HomeFactor = 2**)(-0.1193) + **Interaction(DegreeDays,HomeFactor = 4)(**0.1171) + **Interaction(DegreeDays,HomeFactor = 5)(**0.2288)

1. **(2 pts) According to Model C estimated above, by how much higher/lower is the average oil consumption of customers in HomeFactor level 2 compared to the average oil consumption of customers in HomeFactor level 4 when DegreeDays = 1000 and NumberPeople is the same?**

**Oil usage on Homefactor level =2and DegreeDays = 1000 :**

**Oil Usage** = -35.4342 + DegreeDays(0.3464) + HomeFactor(2.2153) + NumberPeople(13.8213) + Interaction(DegreeDays,HomeFactor = 2)(-0.1193)

**Oil Usage** = -35.4342 + **DegreeDays**(0.2271)+ **HomeFactor**(2.2153) + **NumberPeople**(13.8213)

**Oil Usage** = -35.4342 + 1000(0.2271) + 2 (2.2153) + Numberpeople(13.8213)

= 196.0964 + Numberpeople(13.8213)

**Oil usage on Homefactor level =4 and DegreeDays = 1000:**

**Oil Usage** = -35.4342 + DegreeDays(0.3464)+ Interaction(DegreeDays,HomeFactor = 4)(0.1171) + HomeFactor(2.2153) + NumberPeople(13.8213)

**Oil Usage** = -35.4342 + **DegreeDays**(0.4635) + **HomeFactor**(2.2153) + **NumberPeople**(13.8213)

**Oil Usage** = -35.4342 + 1000(0.4635) + 4(2.2153) + NumberPeople (13.8213)

= 426.927 + Numberpeople(13.8213)

The Oil consumption by the homefactor level=2 is less/lower than the homefactor level=4 customer with the DegreeDays=1000.

1. **(2 pts) Estimate the oil consumption of a customer with DegreeDays =380, NumberPeople =4, HomeFactor = 1.**

**Oil Usage** = -35.4342 + DegreeDays(0.3464) + HomeFactor(2.2153) + NumberPeople(13.8213) + Interaction(DegreeDays,HomeFactor = 1)(-0.2777)

**Oil Usage** = -35.4342 + DegreeDays(0.3464-0.2777) + HomeFactor(2.2153) + NumberPeople(13.8213)

**Oil Usage** = -35.4342 + 380 (0.3464-0.2777) + 1(2.2153) + 4(13.8213)

**Oil Usage** = 48.1723

From the Regressoin Equation, the oil usage is estimated 48.1723 of a customer with DegreeDays =380, NumberPeople =4, HomeFactor = 1.