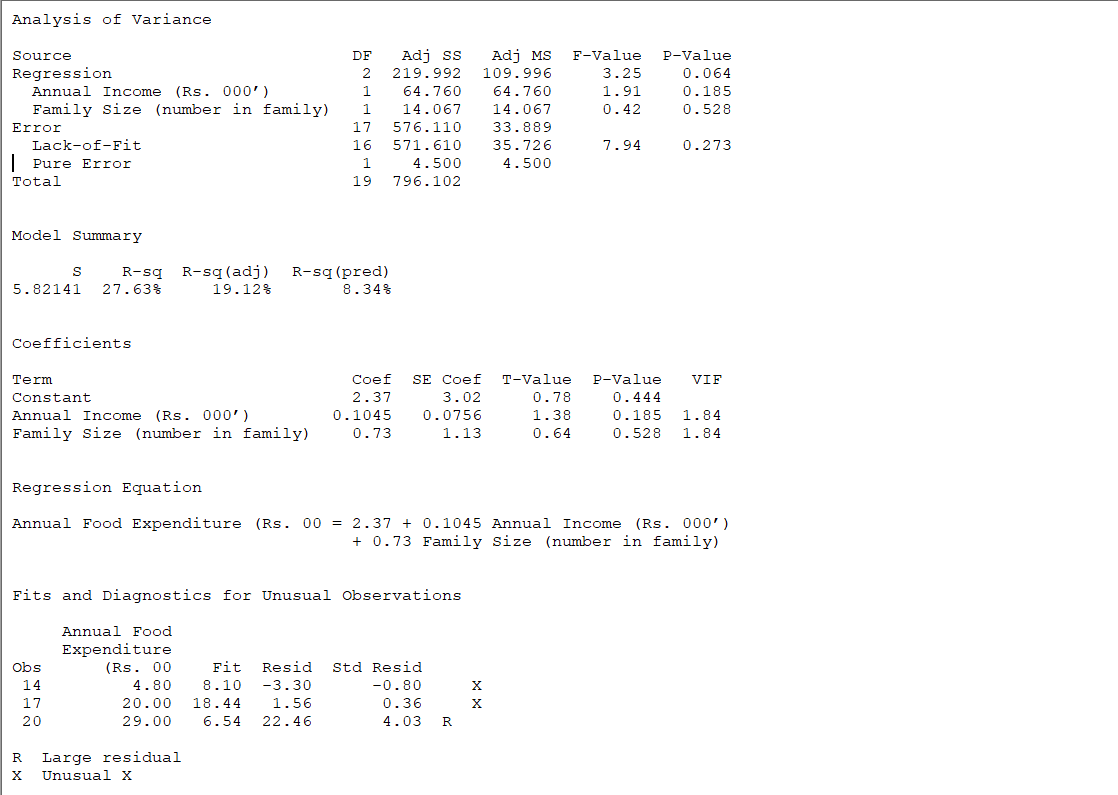
|  |  |
| --- | --- |
|  |  |
| STATISTICAL METHODS - II  Assignment 04  Multiple Regression Analysis  CST 316-2 |  |
| |  |  | | --- | --- | | UWU/CST/20/115  G.G.H.H. Prabodhana |  | |  |
|  |

1. This example examines the relationship between annual family Food Expenditure and family Income. The maintained hypothesis is that family food expenditure increases as family income increases and conversely, ceteris paribus (including family size).   
   The following set of data is obtained from 20 families in a certain area in 2023.

|  |  |  |  |
| --- | --- | --- | --- |
| Family | Annual Food Expenditure (Rs. 000’) | Annual Income (Rs. 000’) | Family Size (number in family) |
| 1 | 5.2 | 28 | 3 |
| 2 | 5.1 | 26 | 3 |
| 3 | 5.6 | 32 | 2 |
| 4 | 4.6 | 24 | 1 |
| 5 | 11.3 | 54 | 4 |
| 6 | 8.1 | 59 | 2 |
| 7 | 7.8 | 44 | 3 |
| 8 | 5.8 | 30 | 2 |
| 9 | 5.1 | 40 | 1 |
| 10 | 18.0 | 82 | 6 |
| 11 | 4.9 | 42 | 3 |
| 12 | 11.8 | 58 | 4 |
| 13 | 5.2 | 28 | 1 |
| 14 | 4.8 | 20 | 5 |
| 15 | 7.9 | 42 | 3 |
| 16 | 6.4 | 47 | 1 |
| 17 | 20.0 | 112 | 6 |
| 18 | 13.7 | 85 | 5 |
| 19 | 5.1 | 31 | 2 |
| 20 | 29 | 26 | 2 |

|  |
| --- |
|  |
|  |  |



Dependent Variable: Annual Food Expenditure

Independent Variables: Annual Income, Family Size

Coefficients:

The multiple regression analysis results for predicting annual family food expenditure indicate the following: The intercept (constant) is 2.3695, with a standard error of 3.023, a t-value of 0.784, and a p-value of 0.444. This suggests that the intercept is not statistically significant, as indicated by the p-value being greater than 0.05. The 95% confidence interval for the intercept ranges from -4.009 to 8.748, which includes zero, indicating high uncertainty in this estimate. For annual income, the coefficient is 0.1045, meaning that for every additional Rs. 000' in annual income, the annual food expenditure is expected to increase by 0.1045 Rs. 000'. However, with a standard error of 0.076, a t-value of 1.382, and a p-value of 0.185, this relationship is not statistically significant. The 95% confidence interval for this coefficient spans from -0.055 to 0.264, including zero, further confirming the lack of statistical significance. Similarly, the coefficient for family size is 0.7270, suggesting that each additional family member is associated with an increase in annual food expenditure by 0.7270 Rs. 000'. The standard error is 1.128, the t-value is 0.644, and the p-value is 0.528, all indicating that this predictor is not statistically significant. The 95% confidence interval ranges from -1.654 to 3.108, which includes zero. In summary, the analysis shows that neither annual income nor family size significantly predicts annual family food expenditure in the given data set.

**Model Statistics:**

R-squared: 0.276

Adjusted R-squared: 0.191

F-statistic: 3.246

Prob (F-statistic): 0.0640

Number of Observations: 20

**Interpretation:**

* R-squared: 27.6% of the variability in annual food expenditure can be explained by the model (annual income and family size).
* Adjusted R-squared: 19.1% (adjusted for the number of predictors in the model).

**Coefficient Interpretation:**

* The Intercept (constant) is 2.3695, which means that when both annual income and family size are zero, the annual food expenditure would be approximately Rs. 2.3695 (though this scenario is unrealistic in practical terms).
* For every additional Rs. 1000 of annual income, the annual food expenditure increases by approximately Rs. 0.1045.
* For each additional member in the family, the annual food expenditure increases by approximately Rs. 0.7270.

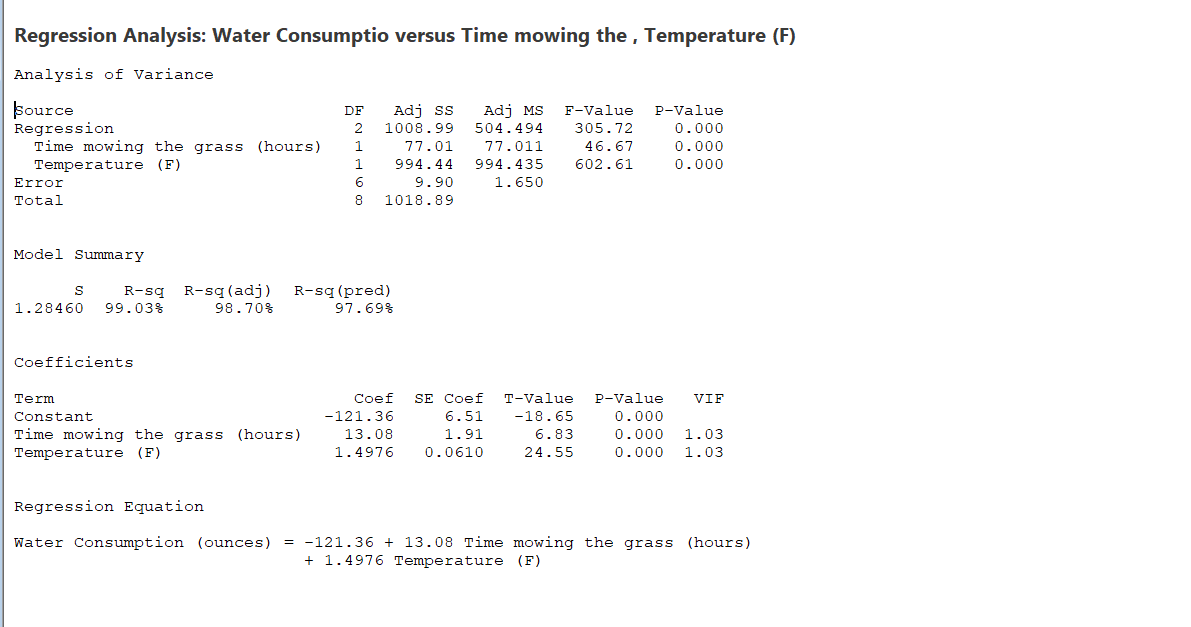
**Statistical Significance:**

* The p-values for the coefficients of Annual Income (0.185) and Family Size (0.528) are greater than 0.05, indicating that these variables are not statistically significant predictors of annual food expenditure at the 5% significance level.
* The overall model's F-statistic has a p-value of 0.0640, suggesting that the model is not statistically significant at the 5% level but is close to being significant.

In summary, relationship between annual food expenditure and the predictors (annual income and family size) is positive, the predictors are not statistically significant at the 5% level. The model explains a moderate portion (27.6%) of the variance in food expenditure, but further investigation with a larger dataset or additional predictors might be needed for more conclusive results.

1. Assume that during a three-hour period spent outside, a person recorded the outside temperature, the time spent mowing the grass, and their water consumption. Use Minitab to find the Multiple Correlation Coefficient, RRR, to determine if the amount of water consumed is dependent on the temperature and the time spent mowing the grass. Use the data given below to build the relationship among these variables. Write a summary report on your findings, which should include all relevant and sufficient hypotheses.

|  |  |  |
| --- | --- | --- |
| Water Consumption (ounces) | Time mowing the grass (hours) | Temperature (F) |
| 16 | 1.85 | 75 |
| 20 | 1.25 | 83 |
| 25 | 1.50 | 85 |
| 27 | 1.75 | 85 |
| 32 | 1.15 | 92 |
| 48 | 1.75 | 97 |
| 48 | 1.60 | 99 |
| 24 | 1.45 | 84 |
| 31 | 1.36 | 91 |



This analysis investigates the relationship between water consumption (dependent variable) and two independent variables. The time spent mowing the grass and the temperature (in Fahrenheit). The goal is to determine if these factors significantly impact water consumption during a three-hour period.

Hypotheses:

* Null Hypothesis (H0): Water consumption is not dependent on temperature and time spent mowing the grass.
* Alternative Hypothesis (H1): Water consumption is dependent on temperature and time spent mowing the grass.

A multiple regression analysis was performed using Minitab, considering water consumption as the response variable and both time spent mowing the grass and temperature as predictor variables. The data includes the following measurements:

* Water Consumption (ounces)
* Time mowing the grass (hours)
* Temperature (F)

**Regression Model:**

The regression equation shows that water consumption increases with both the time spent mowing the grass and the temperature. Specifically:

* For every additional hour spent mowing the grass, water consumption increases by approximately 13.08 ounces.
* For every additional degree Fahrenheit in temperature, water consumption increases by approximately 1.4976 ounces.

**Statistical Significance:**

The p-values for both predictors (time mowing the grass and temperature) are 0.000, indicating that both variables are statistically significant at the 0.05 level.

The overall model is also statistically significant (p-value = 0.000).

**Model Fit:**

The R-squared value of 99.03% indicates that 99.03% of the variability in water consumption can be explained by the time spent mowing the grass and the temperature.

The adjusted R-squared value of 98.70% accounts for the number of predictors in the model, providing a more accurate measure of model fit.

The predicted R-squared value of 97.69% suggests that the model will perform well on new data.

**Variance Inflation Factor (VIF):**

The VIF values for both predictors are 1.03, indicating that there is no multicollinearity problem among the predictors.

The analysis provides strong evidence that both the time spent mowing the grass and the temperature significantly affect water consumption. The model explains a very high proportion of the variability in water consumption, making it a reliable tool for predicting water consumption based on these two factors.