**PROJECT 201**

**Regression Model Analysis**

**Arpitha Rao, Shreya Kulkarni, Adarsh Nair, Dhaval Bhatt.**

**San Jose State University.**

# Executive Summary:

# Introduction

# Brief description of topic

# Methodology

# Results and conclusion

1. Conduct an analysis of the response of labor supply (number of hours) to increasing hourly wages. Do labor hours increase or decrease with wage rates? What other factors affect the number of hours that people work?

* The Fitted Line Plot on Minitab has the equation of HRS=1913+80.94 WAGE, which shows an upward trend.
* This shows that as the wage increases, the number of hours worked will also increase by a factor of 80.94 (slope). And as wage decreases, the hours worked would also decrease.

Figure 1 : Fitted line plot for HRS vs Wages



2. Find the best fitting simple linear regression between HRS (Y) and WAGE (X). You may consider transformations of variables to uncover linear relationships.

Figure 2: Fitted line plot with linear regression between HRS vs Wage



Analysis of Variance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
| Regression | 1 | 51871 | 51871 | 18.47 | 0.000 |
| WAGE | 1 | 51871 | 51871 | 18.47 | 0.000 |
| Error | 37 | 103912 | 2808 |  |  |
| Lack-of-Fit | 36 | 102760 | 2854 | 2.48 | 0.471 |
| Pure Error | 1 | 1152 | 1152 |  |  |
| Total | 38 | 155783 |  |  |  |

Model Summary

|  |  |  |  |
| --- | --- | --- | --- |
| S | R-sq | R-sq(adj) | R-sq(pred) |
| 52.9948 | 33.30% | 31.49% | 27.08% |

Coefficients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Term | Coef | SE Coef | T-Value | P-Value | VIF |
| Constant | 1913.0 | 52.9 | 36.17 | 0.000 |  |
| WAGE | 80.9 | 18.8 | 4.30 | 0.000 | 1.00 |

Fits and Diagnostics for Unusual Observations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Obs | HRS | Fit | Resid | Std Resid |  |  |
| 9 | 2267.0 | 2142.7 | 124.3 | 2.38 | R |  |
| 18 | 2257.0 | 2116.6 | 140.4 | 2.69 | R |  |
| 19 | 1985.0 | 2028.2 | -43.2 | -0.94 |  | X |

*R  Large residual  
X  Unusual X*

3. Find the best multiple regression model that you think describes the relationship between HRS and the other variables in the study.

Figure : Residual plots for 95% significance level



Analysis of Variance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
| Regression | 9 | 75699.5 | 8411.06 | 36.31 | 0.000 |
| WAGE | 1 | 1293.9 | 1293.93 | 5.59 | 0.030 |
| ERSP | 1 | 5955.3 | 5955.31 | 25.71 | 0.000 |
| ERNO | 1 | 2047.6 | 2047.64 | 8.84 | 0.008 |
| NEIN | 1 | 210.0 | 209.96 | 0.91 | 0.354 |
| ASSET | 1 | 19.4 | 19.35 | 0.08 | 0.776 |
| AGE | 1 | 285.6 | 285.63 | 1.23 | 0.281 |
| DEP | 1 | 7573.8 | 7573.82 | 32.70 | 0.000 |
| RACE | 1 | 3141.4 | 3141.38 | 13.56 | 0.002 |
| SCHOOL | 1 | 168.7 | 168.68 | 0.73 | 0.405 |
| Error | 18 | 4169.5 | 231.64 |  |  |
| Total | 27 | 79869.0 |  |  |  |

Model Summary

|  |  |  |  |
| --- | --- | --- | --- |
| S | R-sq | R-sq(adj) | R-sq(pred) |
| 15.2197 | 94.78% | 92.17% | 76.35% |

Coefficients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Term | Coef | SE Coef | T-Value | P-Value | VIF |
| Constant | 1807 | 219 | 8.25 | 0.000 |  |
| WAGE | -91.3 | 38.6 | -2.36 | 0.030 | 30.08 |
| ERSP | 0.1970 | 0.0389 | 5.07 | 0.000 | 8.95 |
| ERNO | -0.1737 | 0.0584 | -2.97 | 0.008 | 3.94 |
| NEIN | 0.289 | 0.304 | 0.95 | 0.354 | 207.32 |
| ASSET | -0.0052 | 0.0181 | -0.29 | 0.776 | 314.21 |
| AGE | 3.22 | 2.90 | 1.11 | 0.281 | 23.62 |
| DEP | 73.0 | 12.8 | 5.72 | 0.000 | 8.32 |
| RACE | -2.550 | 0.692 | -3.68 | 0.002 | 21.28 |
| SCHOOL | 16.7 | 19.5 | 0.85 | 0.405 | 48.24 |

Regression Equation

|  |  |  |
| --- | --- | --- |
| HRS | = | 1807 - 91.3 WAGE + 0.1970 ERSP - 0.1737 ERNO + 0.289 NEIN - 0.0052 ASSET + 3.22 AGE + 73.0 DEP - 2.550 RACE + 16.7 SCHOOL |

Fits and Diagnostics for Unusual Observations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Obs | HRS | Fit | Resid | Std Resid |  |
| 24 | 2102.00 | 2114.28 | -12.28 | -2.02 | R |

*R  Large residual*

*Figure 4: Difference between 95% and 99% confidence level*

**

# Observations

* The co-efficient of determination r2 has good value going towards 1 which indicates there good co relation between X & Y variable i.e. hours and wages.
* Variation in P value. Variable such as wage, ERSP, ERNO, DEP, RACE all these comes under rejection region as their p-value is less that the significance level (95%)
* NEIN, Asset, Age, School comes under fail to reject as their p-value is greater than significance value. (95%)
* Considering the graph above of 95% and 99% confidence level there is not much difference in the correlation but there are few difference in the outliers.
* In 95% confidence interval there are 4 outliers and with 99% confidence interval there are 4 outliers.