Appendix 2

Full statistical analysis for Monge-Nájera, Barrientos and Zúñiga. A Satellite and Ground Evaluation of Urban Vegetation and Infrastructure in the Landscape of a Tropical City: Heredia, Costa Rica

**Summary Statistics**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Count* | *Average* | *Standard deviation* | *Coeff. of variation* | *Standard error* | *Minimum* | *Maximum* |
| % Infrastructure (Satellite) | 9 | 22.8578 | 15.7527 | 68.9163% | 5.25091 | 6.68 | 48.13 |
| % Infrastructure (Ground) | 9 | 52.8778 | 12.0829 | 22.8507% | 4.02764 | 35.2 | 77.3 |
| % Vegetation (Satellite) | 9 | 77.1422 | 15.7527 | 20.4204% | 5.25091 | 51.87 | 93.32 |
| % Vegetación (Ground) | 9 | 47.0222 | 12.0215 | 25.5656% | 4.00717 | 22.7 | 64.8 |
| Total | 36 | 49.975 | 23.7303 | 47.4843% | 3.95505 | 6.68 | 93.32 |

**The StatAdvisor**

This table shows various statistics for each of the 4 columns of data. To test for significant differences amongst the column means, select Analysis of Variance from the list of Tabular Options. Select Means Plot from the list of Graphical Options to display the means graphically.

**Kruskal-Wallis Test**

|  |  |  |
| --- | --- | --- |
|  | *Sample Size* | *Average Rank* |
| % Infrastructure (Satellite) | 9 | 6.44444 |
| % Infrastructure panoramas | 9 | 19.3333 |
| % Vegetation (Satellite) | 9 | 30.5556 |
| % Vegetación panoramas | 9 | 17.6667 |

Test statistic = 23.6807 P-Value = 0.0000291233

**The StatAdvisor**

The Kruskal-Wallis test tests the null hypothesis that the medians within each of the 4 columns is the same. The data from all the columns is first combined and ranked from smallest to largest. The average rank is then computed for the data in each column. Since the P-value is less than 0.05, there is a statistically significant difference amongst the medians at the 95.0% confidence level. To determine which medians are significantly different from which others, select Box-and-Whisker Plot from the list of Graphical Options and select the median notch option.

Fig. 5

**Summary Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *% Roads* | *% Buildings* | *% Billboards* | *% Vegetation* |
| Count | 91 | 91 | 91 | 91 |
| Average | 23,396 | 28,2451 | 0,412967 | 47,946 |
| Standard deviation | 11,3793 | 17,572 | 1,21045 | 21,7507 |
| Standard error | 1,19288 | 1,84205 | 0,12689 | 2,28009 |
| Minimum | 0,0 | 0,0 | 0,0 | 3,6 |
| Maximum | 51,91 | 71,94 | 6,86 | 88,55 |
| Stnd. skewness | 0,932451 | 0,800297 | 14,842 | 0,0495287 |
| Stnd. kurtosis | -0,833598 | -1,61451 | 30,9534 | -1,74736 |

**The StatAdvisor**

This table shows summary statistics for each of the selected data variables. It includes measures of central tendency, measures of variability, and measures of shape. Of particular interest here are the standardized skewness and standardized kurtosis, which can be used to determine whether the sample comes from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant departures from normality, which would tend to invalidate many of the statistical procedures normally applied to this data. In this case, the following variables show standardized skewness values outside the expected range:

% Billboards

The following variables show standardized kurtosis values outside the expected range:

% Billboards

To make the variables more normal, you might try a transformation such as LOG(Y), SQRT(Y), or 1/Y.

**Kruskal-Wallis Test**

|  |  |  |
| --- | --- | --- |
|  | *Sample Size* | *Average Rank* |
| % Roads | 91 | 190,225 |
| % Buildings | 91 | 210,165 |
| % Billboards | 91 | 49,0165 |
| % Vegetation | 91 | 280,593 |

Test statistic = 234,81 P-Value = 0,0

**The StatAdvisor**

The Kruskal-Wallis test tests the null hypothesis that the medians within each of the 4 columns is the same. The data from all the columns is first combined and ranked from smallest to largest. The average rank is then computed for the data in each column. Since the P-value is less than 0,05, there is a statistically significant difference amongst the medians at the 95,0% confidence level. To determine which medians are significantly different from which others, select Box-and-Whisker Plot from the list of Graphical Options and select the median notch option.

**Multiple Regression - % Vegetation**

Dependent variable: % Vegetation

Independent variables:

Population Density

poblacion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 54,081 | 3,74535 | 14,4395 | 0,0000 |
| Population Density | -0,00761861 | 0,00372676 | -2,0443 | 0,0439 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 1909,67 | 1 | 1909,67 | 4,18 | 0,0439 |
| Residual | 40668,8 | 89 | 456,953 |  |  |
| Total (Corr.) | 42578,4 | 90 |  |  |  |

R-squared = 4,48506 percent

R-squared (adjusted for d.f.) = 3,41186 percent

Standard Error of Est. = 21,3764

Mean absolute error = 17,7529

Durbin-Watson statistic = 1,27697 (P=0,0001)

Lag 1 residual autocorrelation = 0,328857

Stepwise regression

Method: backward selection

F-to-enter: 4,0

F-to-remove: 4,0

Step 0:

2 variables in the model. 88 d.f. for error.

R-squared = 7,08% Adjusted R-squared = 4,97% MSE = 449,586

Step 1:

Removing variable poblacion with F-to-remove =2,45834

1 variables in the model. 89 d.f. for error.

R-squared = 4,49% Adjusted R-squared = 3,41% MSE = 456,953

Final model selected.

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between % Vegetation and 2 independent variables. The equation of the fitted model is

% Vegetation = 54,081 - 0,00761861\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between the variables at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 4,48506% of the variability in % Vegetation. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 3,41186%. The standard error of the estimate shows the standard deviation of the residuals to be 21,3764. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 17,7529 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0439, belonging to Population Density. Since the P-value is less than 0,05, that term is statistically significant at the 95,0% confidence level. Consequently, you probably don't want to remove any variables from the model.

**Simple Regression - % Vegetation vs. Population Density**

Dependent variable: % Vegetation

Independent variable: Population Density

Linear model: Y = a + b\*X

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 54,081 | 3,74535 | 14,4395 | 0,0000 |
| Slope | -0,00761861 | 0,00372676 | -2,0443 | 0,0439 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 1909,67 | 1 | 1909,67 | 4,18 | 0,0439 |
| Residual | 40668,8 | 89 | 456,953 |  |  |
| Total (Corr.) | 42578,4 | 90 |  |  |  |

Correlation Coefficient = -0,21178

R-squared = 4,48506 percent

R-squared (adjusted for d.f.) = 3,41186 percent

Standard Error of Est. = 21,3764

Mean absolute error = 17,7529

Durbin-Watson statistic = 1,27697 (P=0,0001)

Lag 1 residual autocorrelation = 0,328857

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between % Vegetation and Population Density. The equation of the fitted model is

% Vegetation = 54,081 - 0,00761861\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Vegetation and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 4,48506% of the variability in % Vegetation. The correlation coefficient equals -0,21178, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 21,3764. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 17,7529 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

Comparación con otros modelos

**Comparison of Alternative Models**

|  |  |  |
| --- | --- | --- |
| *Model* | *Correlation* | *R-Squared* |
| Multiplicative | -0,2759 | 7,61% |
| Reciprocal-Y square root-X | 0,2685 | 7,21% |
| S-curve model | 0,2671 | 7,13% |
| Reciprocal-Y logarithmic-X | 0,2664 | 7,10% |
| Logarithmic-Y square root-X | -0,2656 | 7,05% |
| Reciprocal-Y | 0,2633 | 6,93% |
| Square root-Y logarithmic-X | -0,2627 | 6,90% |
| Square root-Y reciprocal-X | 0,2580 | 6,66% |
| Double square root | -0,2496 | 6,23% |
| Logarithmic-X | -0,2486 | 6,18% |
| Exponential | -0,2471 | 6,11% |
| Reciprocal-X | 0,2464 | 6,07% |
| Reciprocal-Y squared-X | 0,2438 | 5,94% |
| Double reciprocal | -0,2426 | 5,88% |
| Square root-X | -0,2341 | 5,48% |
| Square root-Y | -0,2284 | 5,22% |
| Squared-Y reciprocal-X | 0,2270 | 5,15% |
| Squared-Y logarithmic-X | -0,2253 | 5,08% |
| Linear | -0,2118 | 4,49% |
| Squared-Y square root-X | -0,2094 | 4,38% |
| Logarithmic-Y squared-X | -0,2051 | 4,21% |
| Squared-Y | -0,1864 | 3,47% |
| Square root-Y squared-X | -0,1823 | 3,32% |
| Squared-X | -0,1645 | 2,70% |
| Double squared | -0,1395 | 1,95% |
| Logistic | <no fit> |  |
| Log probit | <no fit> |  |

**The StatAdvisor**

This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the multiplicative model yields the highest R-Squared value with 7,6119%. This is 3,12683% higher than the currently selected linear model. To change models, select the Analysis Options dialog box.

Multiplicative model

**Simple Regression - % Vegetation vs. Population Density**

Dependent variable: % Vegetation

Independent variable: Population Density

Multiplicative model: Y = a\*X^b

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 5,29508 | 0,581472 | 9,10633 | 0,0000 |
| Slope | -0,242817 | 0,0896698 | -2,7079 | 0,0081 |

NOTE: intercept = ln(a)

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 2,47551 | 1 | 2,47551 | 7,33 | 0,0081 |
| Residual | 30,046 | 89 | 0,337596 |  |  |
| Total (Corr.) | 32,5215 | 90 |  |  |  |

Correlation Coefficient = -0,275897

R-squared = 7,6119 percent

R-squared (adjusted for d.f.) = 6,57383 percent

Standard Error of Est. = 0,58103

Mean absolute error = 0,438768

Durbin-Watson statistic = 1,51499 (P=0,0072)

Lag 1 residual autocorrelation = 0,177565

**The StatAdvisor**

The output shows the results of fitting a multiplicative model to describe the relationship between % Vegetation and Population Density. The equation of the fitted model is

% Vegetation = exp(5,29508 - 0,242817\*ln(Population Density))

or

ln(% Vegetation) = 5,29508 - 0,242817\*ln(Population Density)

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Vegetation and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 7,6119% of the variability in % Vegetation. The correlation coefficient equals -0,275897, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0,58103. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 0,438768 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

**Multiple Regression - % Infrastructure**

Dependent variable: % Infrastructure

Independent variables:

Population Density

poblacion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 28,2451 | 1,84205 | 15,3335 | 0,0000 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 0,0 | 0 |  |  |  |
| Residual | 27789,8 | 90 | 308,776 |  |  |
| Total (Corr.) | 27789,8 | 90 |  |  |  |

R-squared = 0,0 percent

R-squared (adjusted for d.f.) = 0,0 percent

Standard Error of Est. = 17,572

Mean absolute error = 14,7897

Durbin-Watson statistic = 1,28546 (P=0,0002)

Lag 1 residual autocorrelation = 0,318146

Stepwise regression

Method: backward selection

F-to-enter: 4,0

F-to-remove: 4,0

Step 0:

2 variables in the model. 88 d.f. for error.

R-squared = 1,52% Adjusted R-squared = -0,72% MSE = 310,998

Step 1:

Removing variable poblacion with F-to-remove =0,753178

1 variables in the model. 89 d.f. for error.

R-squared = 0,68% Adjusted R-squared = -0,44% MSE = 310,135

Step 2:

Removing variable Population Density with F-to-remove =0,605556

0 variables in the model. 90 d.f. for error.

R-squared = 0,00% Adjusted R-squared = 0,00% MSE = 308,776

Final model selected.

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between % Infrastructure and 2 independent variables. The equation of the fitted model is

% Infrastructure = 28,2451

The R-Squared statistic indicates that the model as fitted explains 0,0% of the variability in % Infrastructure. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 0,0%. The standard error of the estimate shows the standard deviation of the residuals to be 17,572. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 14,7897 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

**Simple Regression - % Infrastructure vs. Population Density**

Dependent variable: % Infrastructure

Independent variable: Population Density

Linear model: Y = a + b\*X

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 26,3211 | 3,08555 | 8,53046 | 0,0000 |
| Slope | 0,00238918 | 0,00307024 | 0,778175 | 0,4385 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 187,804 | 1 | 187,804 | 0,61 | 0,4385 |
| Residual | 27602,0 | 89 | 310,135 |  |  |
| Total (Corr.) | 27789,8 | 90 |  |  |  |

Correlation Coefficient = 0,0822072

R-squared = 0,675802 percent

R-squared (adjusted for d.f.) = -0,4402 percent

Standard Error of Est. = 17,6107

Mean absolute error = 14,7178

Durbin-Watson statistic = 1,31562 (P=0,0003)

Lag 1 residual autocorrelation = 0,303971

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between % Infrastructure and Population Density. The equation of the fitted model is

% Infrastructure = 26,3211 + 0,00238918\*Population Density

Since the P-value in the ANOVA table is greater or equal to 0,05, there is not a statistically significant relationship between % Infrastructure and Population Density at the 95,0% or higher confidence level.

The R-Squared statistic indicates that the model as fitted explains 0,675802% of the variability in % Infrastructure. The correlation coefficient equals 0,0822072, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 17,6107. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 14,7178 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

**Comparison of Alternative Models**

|  |  |  |
| --- | --- | --- |
| *Model* | *Correlation* | *R-Squared* |
| Squared-Y logarithmic-X | 0,1501 | 2,25% |
| Squared-Y reciprocal-X | -0,1446 | 2,09% |
| Squared-Y square root-X | 0,1430 | 2,04% |
| Squared-Y | 0,1302 | 1,70% |
| Double squared | 0,1005 | 1,01% |
| Logarithmic-X | 0,0964 | 0,93% |
| Square root-X | 0,0920 | 0,85% |
| Reciprocal-X | -0,0878 | 0,77% |
| Linear | 0,0822 | 0,68% |
| Squared-X | 0,0571 | 0,33% |
| Square root-Y logarithmic-X | 0,0443 | 0,20% |
| Double square root | 0,0418 | 0,17% |
| Square root-Y reciprocal-X | -0,0351 | 0,12% |
| Square root-Y | 0,0346 | 0,12% |
| Square root-Y squared-X | 0,0146 | 0,02% |
| Exponential | <no fit> |  |
| Reciprocal-Y | <no fit> |  |
| Logarithmic-Y square root-X | <no fit> |  |
| Reciprocal-Y square root-X | <no fit> |  |
| Multiplicative | <no fit> |  |
| Reciprocal-Y logarithmic-X | <no fit> |  |
| S-curve model | <no fit> |  |
| Double reciprocal | <no fit> |  |
| Logarithmic-Y squared-X | <no fit> |  |
| Reciprocal-Y squared-X | <no fit> |  |
| Logistic | <no fit> |  |
| Log probit | <no fit> |  |

**The StatAdvisor**

This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the squared-Y logarithmic-X model yields the highest R-Squared value with 2,25265%. This is the currently selected model.

**Simple Regression - % Infrastructure vs. Population Density**

Dependent variable: % Infrastructure

Independent variable: Population Density

Squared-Y logarithmic-X model: Y = sqrt(a + b\*ln(X))

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | -460,176 | 1097,64 | -0,419241 | 0,6761 |
| Slope | 242,419 | 169,269 | 1,43215 | 0,1556 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 2,46739E6 | 1 | 2,46739E6 | 2,05 | 0,1556 |
| Residual | 1,07066E8 | 89 | 1,20299E6 |  |  |
| Total (Corr.) | 1,09533E8 | 90 |  |  |  |

Correlation Coefficient = 0,150088

R-squared = 2,25265 percent

R-squared (adjusted for d.f.) = 1,15436 percent

Standard Error of Est. = 1096,81

Mean absolute error = 872,467

Durbin-Watson statistic = 1,46766 (P=0,0036)

Lag 1 residual autocorrelation = 0,205029

**The StatAdvisor**

The output shows the results of fitting a squared-Y logarithmic-X model to describe the relationship between % Infrastructure and Population Density. The equation of the fitted model is

% Infrastructure = sqrt(-460,176 + 242,419\*ln(Population Density))

Since the P-value in the ANOVA table is greater or equal to 0,05, there is not a statistically significant relationship between % Infrastructure and Population Density at the 95,0% or higher confidence level.

The R-Squared statistic indicates that the model as fitted explains 2,25265% of the variability in % Infrastructure. The correlation coefficient equals 0,150088, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 1096,81. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 872,467 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

**Multiple Regression - % Roads**

Dependent variable: % carreteras

Independent variables:

Population Density

poblacion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 19,5801 | 1,94034 | 10,091 | 0,0000 |
| Population Density | 0,0047388 | 0,00193071 | 2,45443 | 0,0161 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 738,83 | 1 | 738,83 | 6,02 | 0,0161 |
| Residual | 10915,2 | 89 | 122,643 |  |  |
| Total (Corr.) | 11654,1 | 90 |  |  |  |

R-squared = 6,33968 percent

R-squared (adjusted for d.f.) = 5,28732 percent

Standard Error of Est. = 11,0744

Mean absolute error = 8,85134

Durbin-Watson statistic = 1,46589 (P=0,0035)

Lag 1 residual autocorrelation = 0,264989

Stepwise regression

Method: backward selection

F-to-enter: 4,0

F-to-remove: 4,0

Step 0:

2 variables in the model. 88 d.f. for error.

R-squared = 9,35% Adjusted R-squared = 7,29% MSE = 120,052

Step 1:

Removing variable poblacion with F-to-remove =2,92109

1 variables in the model. 89 d.f. for error.

R-squared = 6,34% Adjusted R-squared = 5,29% MSE = 122,643

Final model selected.

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between % Roads and 2 independent variables. The equation of the fitted model is

% Roads = 19,5801 + 0,0047388\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between the variables at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 6,33968% of the variability in % carreteras. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 5,28732%. The standard error of the estimate shows the standard deviation of the residuals to be 11,0744. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 8,85134 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0161, belonging to Population Density. Since the P-value is less than 0,05, that term is statistically significant at the 95,0% confidence level. Consequently, you probably don't want to remove any variables from the model.

**Simple Regression - % Roads vs. Population Density**

Dependent variable: % carreteras

Independent variable: Population Density

Linear model: Y = a + b\*X

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 19,5801 | 1,94034 | 10,091 | 0,0000 |
| Slope | 0,0047388 | 0,00193071 | 2,45443 | 0,0161 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 738,83 | 1 | 738,83 | 6,02 | 0,0161 |
| Residual | 10915,2 | 89 | 122,643 |  |  |
| Total (Corr.) | 11654,1 | 90 |  |  |  |

Correlation Coefficient = 0,251787

R-squared = 6,33968 percent

R-squared (adjusted for d.f.) = 5,28732 percent

Standard Error of Est. = 11,0744

Mean absolute error = 8,85134

Durbin-Watson statistic = 1,46589 (P=0,0035)

Lag 1 residual autocorrelation = 0,264989

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between % Roads and Population Density. The equation of the fitted model is

% Roads = 19,5801 + 0,0047388\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Roads and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 6,33968% of the variability in % carreteras. The correlation coefficient equals 0,251787, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 11,0744. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 8,85134 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

**Comparison of Alternative Models**

|  |  |  |
| --- | --- | --- |
| *Model* | *Correlation* | *R-Squared* |
| Square root-Y reciprocal-X | -0,3470 | 12,04% |
| Square root-Y logarithmic-X | 0,3273 | 10,71% |
| Reciprocal-X | -0,3222 | 10,38% |
| Logarithmic-X | 0,3061 | 9,37% |
| Double square root | 0,3006 | 9,03% |
| Square root-X | 0,2819 | 7,94% |
| Square root-Y | 0,2680 | 7,18% |
| Squared-Y reciprocal-X | -0,2628 | 6,90% |
| Linear | 0,2518 | 6,34% |
| Squared-Y logarithmic-X | 0,2495 | 6,22% |
| Squared-Y square root-X | 0,2292 | 5,25% |
| Square root-Y squared-X | 0,2102 | 4,42% |
| Squared-Y | 0,2041 | 4,16% |
| Squared-X | 0,1977 | 3,91% |
| Double squared | 0,1591 | 2,53% |
| Exponential | <no fit> |  |
| Reciprocal-Y | <no fit> |  |
| Logarithmic-Y square root-X | <no fit> |  |
| Reciprocal-Y square root-X | <no fit> |  |
| Multiplicative | <no fit> |  |
| Reciprocal-Y logarithmic-X | <no fit> |  |
| S-curve model | <no fit> |  |
| Double reciprocal | <no fit> |  |
| Logarithmic-Y squared-X | <no fit> |  |
| Reciprocal-Y squared-X | <no fit> |  |
| Logistic | <no fit> |  |
| Log probit | <no fit> |  |

**The StatAdvisor**

This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the square root-Y reciprocal-X model yields the highest R-Squared value with 12,0413%. This is 5,70163% higher than the currently selected linear model. To change models, select the Analysis Options dialog box.

**Simple Regression - % Roads vs. Population Density**

Dependent variable: % carreteras

Independent variable: Population Density

Square root-Y reciprocal-X model: Y = (a + b/X)^2

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 5,44095 | 0,258159 | 21,0759 | 0,0000 |
| Slope | -402,822 | 115,404 | -3,49054 | 0,0008 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 18,4052 | 1 | 18,4052 | 12,18 | 0,0008 |
| Residual | 134,445 | 89 | 1,51062 |  |  |
| Total (Corr.) | 152,85 | 90 |  |  |  |

Correlation Coefficient = -0,347006

R-squared = 12,0413 percent

R-squared (adjusted for d.f.) = 11,053 percent

Standard Error of Est. = 1,22907

Mean absolute error = 0,921527

Durbin-Watson statistic = 1,55639 (P=0,0126)

Lag 1 residual autocorrelation = 0,2206

**The StatAdvisor**

The output shows the results of fitting a square root-Y reciprocal-X model to describe the relationship between % Roads and Population Density. The equation of the fitted model is

% Roads = (5,44095 - 402,822/Population Density)^2

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Roads and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 12,0413% of the variability in % carreteras. The correlation coefficient equals -0,347006, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 1,22907. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 0,921527 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is less than 0,05, there is an indication of possible serial correlation at the 95,0% confidence level. Plot the residuals versus row order to see if there is any pattern that can be seen.

****

**Multiple Regression - % Billboards**

Dependent variable: % Billboards

Independent variables:

Population Density

poblacion

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| CONSTANT | 0,0191938 | 0,20681 | 0,0928086 | 0,9263 |
| Population Density | 0,000488999 | 0,000205784 | 2,37627 | 0,0196 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 7,86726 | 1 | 7,86726 | 5,65 | 0,0196 |
| Residual | 124,0 | 89 | 1,39326 |  |  |
| Total (Corr.) | 131,867 | 90 |  |  |  |

R-squared = 5,96604 percent

R-squared (adjusted for d.f.) = 4,90948 percent

Standard Error of Est. = 1,18036

Mean absolute error = 0,626436

Durbin-Watson statistic = 1,68981 (P=0,0564)

Lag 1 residual autocorrelation = 0,15262

Stepwise regression

Method: backward selection

F-to-enter: 4,0

F-to-remove: 4,0

Step 0:

2 variables in the model. 88 d.f. for error.

R-squared = 6,44% Adjusted R-squared = 4,31% MSE = 1,402

Step 1:

Removing variable poblacion with F-to-remove =0,444949

1 variables in the model. 89 d.f. for error.

R-squared = 5,97% Adjusted R-squared = 4,91% MSE = 1,39326

Final model selected.

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between % Billboards and 2 independent variables. The equation of the fitted model is

% Billboards = 0,0191938 + 0,000488999\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between the variables at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 5,96604% of the variability in % Billboards. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 4,90948%. The standard error of the estimate shows the standard deviation of the residuals to be 1,18036. This value can be used to construct prediction limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 0,626436 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0,05, there is no indication of serial autocorrelation in the residuals at the 95,0% confidence level.

In determining whether the model can be simplified, notice that the highest P-value on the independent variables is 0,0196, belonging to Population Density. Since the P-value is less than 0,05, that term is statistically significant at the 95,0% confidence level. Consequently, you probably don't want to remove any variables from the model.

**Simple Regression - % Billboards vs. Population Density**

Dependent variable: % Billboards

Independent variable: Population Density

Linear model: Y = a + b\*X

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 0,0191938 | 0,20681 | 0,0928086 | 0,9263 |
| Slope | 0,000488999 | 0,000205784 | 2,37627 | 0,0196 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 7,86726 | 1 | 7,86726 | 5,65 | 0,0196 |
| Residual | 124,0 | 89 | 1,39326 |  |  |
| Total (Corr.) | 131,867 | 90 |  |  |  |

Correlation Coefficient = 0,244255

R-squared = 5,96604 percent

R-squared (adjusted for d.f.) = 4,90948 percent

Standard Error of Est. = 1,18036

Mean absolute error = 0,626436

Durbin-Watson statistic = 1,68981 (P=0,0564)

Lag 1 residual autocorrelation = 0,15262

**The StatAdvisor**

The output shows the results of fitting a linear model to describe the relationship between % Billboards and Population Density. The equation of the fitted model is

% Billboards = 0,0191938 + 0,000488999\*Population Density

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Billboards and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 5,96604% of the variability in % Billboards. The correlation coefficient equals 0,244255, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 1,18036. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 0,626436 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0,05, there is no indication of serial autocorrelation in the residuals at the 95,0% confidence level.

**Comparison of Alternative Models**

|  |  |  |
| --- | --- | --- |
| *Model* | *Correlation* | *R-Squared* |
| Square root-Y squared-X | 0,3126 | 9,77% |
| Square root-Y | 0,2886 | 8,33% |
| Squared-X | 0,2678 | 7,17% |
| Double square root | 0,2625 | 6,89% |
| Linear | 0,2443 | 5,97% |
| Square root-Y logarithmic-X | 0,2263 | 5,12% |
| Square root-X | 0,2209 | 4,88% |
| Logarithmic-X | 0,1896 | 3,59% |
| Double squared | 0,1888 | 3,56% |
| Squared-Y | 0,1658 | 2,75% |
| Square root-Y reciprocal-X | -0,1461 | 2,14% |
| Squared-Y square root-X | 0,1458 | 2,12% |
| Reciprocal-X | -0,1223 | 1,50% |
| Squared-Y logarithmic-X | 0,1205 | 1,45% |
| Squared-Y reciprocal-X | -0,0699 | 0,49% |
| Exponential | <no fit> |  |
| Reciprocal-Y | <no fit> |  |
| Logarithmic-Y square root-X | <no fit> |  |
| Reciprocal-Y square root-X | <no fit> |  |
| Multiplicative | <no fit> |  |
| Reciprocal-Y logarithmic-X | <no fit> |  |
| S-curve model | <no fit> |  |
| Double reciprocal | <no fit> |  |
| Logarithmic-Y squared-X | <no fit> |  |
| Reciprocal-Y squared-X | <no fit> |  |
| Logistic | <no fit> |  |
| Log probit | <no fit> |  |

**The StatAdvisor**

This table shows the results of fitting several curvilinear models to the data. Of the models fitted, the square root-Y squared-X model yields the highest R-Squared value with 9,77032%. This is 3,80427% higher than the currently selected linear model. To change models, select the Analysis Options dialog box.

**Simple Regression - % Billboards vs. Population Density**

Dependent variable: % Billboards

Independent variable: Population Density

Square root-Y squared-X model: Y = (a + b\*X^2)^2

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 0,115295 | 0,0729433 | 1,58061 | 0,1175 |
| Slope | 1,28279E-7 | 4,13221E-8 | 3,10438 | 0,0026 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 3,13863 | 1 | 3,13863 | 9,64 | 0,0026 |
| Residual | 28,9855 | 89 | 0,325679 |  |  |
| Total (Corr.) | 32,1241 | 90 |  |  |  |

Correlation Coefficient = 0,312575

R-squared = 9,77032 percent

R-squared (adjusted for d.f.) = 8,7565 percent

Standard Error of Est. = 0,570683

Mean absolute error = 0,36577

Durbin-Watson statistic = 1,73401 (P=0,0853)

Lag 1 residual autocorrelation = 0,130209

**The StatAdvisor**

The output shows the results of fitting a square root-Y squared-X model to describe the relationship between % Billboards and Population Density. The equation of the fitted model is

% Billboards = (0,115295 + 1,28279E-7\*Population Density^2)^2

Since the P-value in the ANOVA table is less than 0,05, there is a statistically significant relationship between % Billboards and Population Density at the 95,0% confidence level.

The R-Squared statistic indicates that the model as fitted explains 9,77032% of the variability in % Billboards. The correlation coefficient equals 0,312575, indicating a relatively weak relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0,570683. This value can be used to construct prediction limits for new observations by selecting the Forecasts option from the text menu.

The mean absolute error (MAE) of 0,36577 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in your data file. Since the P-value is greater than 0,05, there is no indication of serial autocorrelation in the residuals at the 95,0% confidence level.

