**Балтийский государственный технический университет**

**«Военмех» им. Д.Ф. Устинова**

**Кафедра И7**

**«Кафедра математической статистики и прикладной математики»**

**«Математическая статистика»**

Лабораторная работа № 9

«Регрессивный анализ в пакетах STATGRAPHIC и MathCAD»

Вариант 18

**Выполнил:**

Студент Тихонов А.Ю.

Группа И383

**Преподаватель:**

Гнидин В.В.

Санкт-Петербург

2011

**Задание 1.**

Найти в пакетах STATGRAPHICS и MATHCAD оценки параметров линейной регрессии  на , доверительные интервалы для параметров и линии регрессии и проверить согласие линейной регрессии с результатами наблюдений. Принять уровень доверительной вероятности равным 0.90.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|  | 124.9 | 127.1 | 134.0 | 139.1 | 147.3 | 155.0 | 159.8 | 165.4 | 172.5 | 177.4 | 182.1 |

**Решение в пакете Statgraphics**

**Simple Regression - Prod vs. Temp**

Dependent variable: Prod

Independent variable: Temp

Linear model: Y = a + b\*X

**Coefficients**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Least Squares* | *Standard* | *T* |  |
| *Parameter* | *Estimate* | *Error* | *Statistic* | *P-Value* |
| Intercept | 116,72 | 0,943394 | 123,723 | 0,0000 |
| Slope | 6,07091 | 0,139096 | 43,6455 | 0,0000 |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 4054,15 | 1 | 4054,15 | 1904,93 | 0,0000 |
| Residual | 19,1542 | 9 | 2,12824 |  |  |
| Total (Corr.) | 4073,31 | 10 |  |  |  |

Correlation Coefficient = 0,997646

R-squared = 99,5298 percent

R-squared (adjusted for d.f.) = 99,4775 percent

Standard Error of Est. = 1,45885

Mean absolute error = 1,09587

Durbin-Watson statistic = 1,56349 (P=0,1190)

Lag 1 residual autocorrelation = 0,0509739

**The StatAdvisor**

*The output shows the results of fitting a linear model to describe the relationship between Prod and Temp. The equation of the fitted model is*

*Prod = 116,72 + 6,07091\*Temp*

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

*The R-Squared statistic indicates that the model as fitted explains 99,5298% of the variability in Prod. The correlation coefficient equals 0,997646, indicating a relatively strong relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 1,45885. 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 1,09587 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.*

**Analysis of Variance with Lack-of-Fit**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Source* | *Sum of Squares* | *Df* | *Mean Square* | *F-Ratio* | *P-Value* |
| Model | 4054,15 | 1 | 4054,15 | 1904,93 | 0,0000 |
| Residual | 19,1542 | 9 | 2,12824 |  |  |
| Lack-of-Fit | 19,1542 | 9 | 2,12824 |  |  |
| Pure Error | 0,0 | 0 |  |  |  |
| Total (Corr.) | 4073,31 | 10 |  |  |  |

**The StatAdvisor**

*The lack of fit test is designed to determine whether the selected model is adequate to describe the observed data, or whether a more complicated model should be used. The test is performed by comparing the variability of the current model residuals to the variability between observations at replicate values of the independent variable X. Unfortunately, the test can not be run in this case because there are no replicate observations at the same values of Temp.*

**Predicted Values**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *95,00%* |  | *95,00%* |  |
|  | *Predicted* | *Prediction* | *Limits* | *Confidence* | *Limits* |
| *X* | *Y* | *Lower* | *Upper* | *Lower* | *Upper* |
| 1,0 | 122,791 | 119,002 | 126,58 | 120,929 | 124,652 |
| 11,0 | 183,5 | 179,711 | 187,289 | 181,638 | 185,362 |

**The StatAdvisor**

*This table shows the predicted values for Prod using the fitted model. In addition to the best predictions, the table shows:*

*(1) 95,0% prediction intervals for new observations*

*(2) 95,0% confidence intervals for the mean of many observations*

*The prediction and confidence intervals correspond to the inner and outer bounds on the graph of the fitted model.*

**Comparison of Alternative Models**

|  |  |  |
| --- | --- | --- |
| *Model* | *Correlation* | *R-Squared* |
| Linear | 0,9976 | 99,53% |
| Squared-Y | 0,9971 | 99,42% |
| Square root-Y | 0,9970 | 99,40% |
| Exponential | 0,9958 | 99,16% |
| Reciprocal-Y | -0,9915 | 98,31% |
| Logarithmic-Y square root-X | 0,9902 | 98,06% |
| Double square root | 0,9879 | 97,59% |
| Square root-X | 0,9850 | 97,02% |
| Double squared | 0,9806 | 96,16% |
| Squared-Y square root-X | 0,9775 | 95,54% |
| Squared-X | 0,9698 | 94,05% |
| Square root-Y squared-X | 0,9635 | 92,83% |
| Logarithmic-Y squared-X | 0,9566 | 91,51% |
| Multiplicative | 0,9541 | 91,03% |
| Square root-Y logarithmic-X | 0,9481 | 89,89% |
| Logarithmic-X | 0,9416 | 88,65% |
| Reciprocal-Y squared-X | -0,9413 | 88,61% |
| Squared-Y logarithmic-X | 0,9271 | 85,96% |
| Double reciprocal | 0,8207 | 67,36% |
| S-curve model | -0,7994 | 63,90% |
| Reciprocal-X | -0,7768 | 60,35% |
| Squared-Y reciprocal-X | -0,7533 | 56,75% |
| Reciprocal-Y square root-X | <no fit> |  |
| Reciprocal-Y logarithmic-X | <no fit> |  |
| Square root-Y reciprocal-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 linear model yields the highest R-Squared value with 99,5298%. This is the currently selected model.*

**Unusual Residuals**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | *Predicted* |  | *Studentized* |
| *Row* | *X* | *Y* | *Y* | *Residual* | *Residual* |
| 1 | 1,0 | 124,9 | 122,791 | 2,10909 | 2,03 |

**The StatAdvisor**

*The table of unusual residuals lists all observations which have Studentized residuals greater than 2 in absolute value. Studentized residuals measure how many standard deviations each observed value of Prod deviates from a model fitted using all of the data except that observation. In this case, there is one Studentized residual greater than 2, but none greater than 3.*

**Influential Points**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | *Predicted* | *Studentized* |  |
| *Row* | *X* | *Y* | *Y* | *Residual* | *Leverage* |

Average leverage of single data point = 0,181818

**The StatAdvisor**

*The table of influential data points lists all observations which have leverage values greater than 3 times that of an average data point. Leverage is a statistic which measures how influential each observation is in determining the coefficients of the estimated model. In this case, an average data point would have a leverage value equal to 0,181818. There are no data points with more than 3 times the average leverage.*







**Решение в пакете MathCAD**









































































