Балтийский государственный технический университет «Военмех» им. Д.Ф.Устинова

Кафедра прикладной математики и информатики (И7).

Математическая статистика.

Лабораторная работа №9.

Регрессионный анализ в пакетах Statgraphics и MathCad.

Вариант 12.

**Выполнил**

Романов Н.

группа И383

**Проверил**

Гнидин В.В.

Санкт-Петербург

2011

**Задание 1:**

Найти оценки параметров линейной регрессии *y* на *x*, доверительные интервалы для параметров и линии регрессии и проверить согласие линейной регрессии результатам наблюдений. Принять уровень доверительной вероятности равным 0.90.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Y | 6.32 | 6.52 | 6.65 | 7.26 | 7.49 | 7.83 | 8.13 | 8.40 | 8.58 | 9.01 | 9.05 |

**Решение в пакете StatGraphics:**

**Regression Analysis** - Linear model: Y = a + b\*X

-----------------------------------------------------------------------------

Dependent variable: y

Independent variable: x

-----------------------------------------------------------------------------

Standard T

Parameter Estimate Error Statistic P-Value

-----------------------------------------------------------------------------

Intercept 5,98618 0,0743865 80,4741 0,0000

Slope 0,293818 0,0109677 26,7894 0,0000

-----------------------------------------------------------------------------

Analysis of Variance

-----------------------------------------------------------------------------

Source Sum of Squares Df Mean Square F-Ratio P-Value

-----------------------------------------------------------------------------

Model 9,4962 1 9,4962 717,67 0,0000

Residual 0,119087 9 0,0132319

-----------------------------------------------------------------------------

Total (Corr.) 9,61529 10

Correlation Coefficient = 0,993788

R-squared = 98,7615 percent

R-squared (adjusted for d.f.) = 98,6239 percent

Standard Error of Est. = 0,11503

Mean absolute error = 0,089124

Durbin-Watson statistic = 2,00142 (P=0,3451)

Lag 1 residual autocorrelation = -0,126187

**The StatAdvisor**

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

*y = 5,98618 + 0,293818\*x*

*Since the P-value in the ANOVA table is less than 0.01, there is a statistically significant relationship between y and x at the 99% confidence level.*

*The R-Squared statistic indicates that the model as fitted explains 98,7615% of the variability in y. The correlation coefficient equals 0,993788, indicating a relatively strong relationship between the variables. The standard error of the estimate shows the standard deviation of the residuals to be 0,11503. 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 0,089124 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.*

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

-----------------------------------------------------------------------------

Source Sum of Squares Df Mean Square F-Ratio P-Value

-----------------------------------------------------------------------------

Model 9,4962 1 9,4962 717,67 0,0000

Residual 0,119087 9 0,0132319

-----------------------------------------------------------------------------

Lack-of-Fit 0,119087 9 0,0132319

Pure Error 0,0 0

-----------------------------------------------------------------------------

Total (Corr.) 9,61529 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 x.*

**Predicted Values**

------------------------------------------------------------------------------

90,00% 90,00%

Predicted Prediction Limits Confidence Limits

X Y Lower Upper Lower Upper

------------------------------------------------------------------------------

10,0 8,92436 8,6899 9,15883 8,82185 9,02688

110,0 38,3062 36,2037 40,4087 36,2143 40,3981

------------------------------------------------------------------------------

**The StatAdvisor**

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

*(1) 90,0% prediction intervals for new observations*

*(2) 90,0% 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 0,9938 98,76%

Square root-Y 0,9923 98,46%

Exponential 0,9902 98,05%

Square root-X 0,9867 97,36%

Reciprocal-Y -0,9844 96,91%

Multiplicative 0,9595 92,06%

Logarithmic-X 0,9490 90,06%

Double reciprocal 0,8328 69,35%

S-curve -0,8133 66,15%

Reciprocal-X -0,7925 62,81%

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 98,7615%. This is the currently selected*

*model.*

*Unusual Residuals*

----------------------------------------------------------------------------

Predicted Studentized

Row X Y Y Residual Residual

----------------------------------------------------------------------------

3 3,0 6,65 6,86764 -0,217636 -2,72

11 11,0 9,05 9,21818 -0,168182 -2,07

----------------------------------------------------------------------------

**The StatAdvisor**

*The table of unusual residuals lists all observations which have Studentized residuals greater than 2.0 in absolute value. Studentized residuals measure how many standard deviations each observed value of y deviates from a model fitted using all of the data except that observation. In this case, there are 2 Studentized residuals greater than 2.0, but none greater than 3.0.*

**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 MathCad:**



































































































