Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский Томский политехнический Университет»



Инженерная школа ядерных технологий

Отделение математики и математической физики

01.03.02 Прикладная математика и информатика

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Лабораторная работа № 3

**Вариант - 6**

по дисциплине:

**Математическая статистика**

|  |  |  |  |
| --- | --- | --- | --- |
| **Выполнил:** |  | | |
| студент группы | 0В21 | Дзебан А.А. |  |
|  |  |  |  |
| **Проверил:** | Шинкеев М.Л. | | |
|  |  |  |  |
|  |  |  |

Томск – 2024

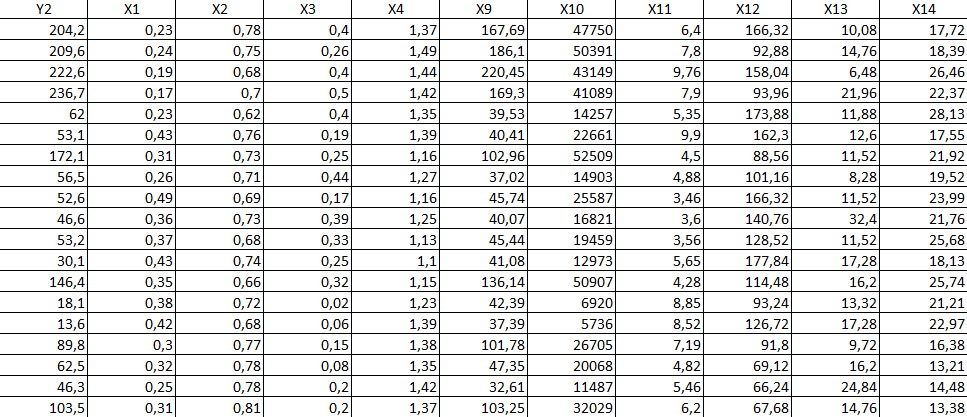
**Цели работы:**

Оценка параметров уравнения множественной линейной регрессии

**Ход работы:**

**Требуется,** для выбранных в соответствии с вариантом задания показателей (смотри таблицу 1), построить модель множественной линейной регрессии, связывающую результирующий показатель и определяющие факторы. Для устранения эффекта мультиколлинеарности, использовать метод пошагового включения факторов в уравнение регрессии (первым добавляем фактор, наиболее коррелирующий с откликом, затем последовательно добавляем факторы, наиболее коррелирующие с остатком). Указать уровни значимости коэффициентов модели и модели регрессии в целом, а также полученные оценки для коэффициента детерминации и остаточной дисперсии. Исследовать остатки модели на нормальность и однородность.

Выберем данные согласно варианту:



Действовать будем следующим образом:

1. Будем включать наиболее коррелирующий с остатком (откликом на шаге 1) фактором.
2. Временно добавляем фактор в модель.
3. Просчитываем её с добавленным коэффициентом
4. Если коэффициент не является значимым – отказываемся от его добавления
5. Если является значимым – добавляем его в финальную модель, обновляем остатки, пересчитываем модель со всеми коэффициентами.
6. Выполнять 1-5, пока либо модель перестанет быть значимой, либо коэффициентов не останется.

Так, для получения наиболее коррелирующего фактора составим матрицу корреляций с откликом

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Correlations (Spreadsheet2 in Workbook1) N=53 (Casewise deletion of missing data) | | | | | | | | | | |
| X1 | X2 | X3 | X4 | X9 | X10 | X11 | X12 | X13 | X14 |
| Y2 | -0,641424 | 0,234010 | 0,270987 | 0,361930 | 0,831650 | 0,859957 | 0,337845 | -0,301983 | -0,276249 | -0,058842 |

Наибольшим по модулю значением корреляции обладает X10

Построим линейную регрессию:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Effect |  |  |  |  |
| Y2 | Y2 std | Y2 t | Y2 p |
| Intercept | -29,1268 | 14,43639 | -2,01760 | 0,048910 |
| X10 | 0,0052 | 0,00043 | 12,03316 | 0,000000 |

Видно, что коэффициенты модели являются статистически значимыми,

|  |  |
| --- | --- |
| F | p |
| 144,7968 | 0,000000 |

Так же сама модель является статистически значимой.



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Correlations (Spreadsheet2 in Workbook1) N=53 (Casewise deletion of missing data) | | | | | | | | | |
| X1 | X2 | X3 | X4 | X9 | X10 | X11 | X12 | X13 | X14 |
| ResidsAt1 | -0,599206 | 0,309767 | 0,172198 | 0,382397 | 0,566613 | 0,000000 | 0,507875 | -0,265283 | -0,192857 | -0,152376 |

Для получения наиболее коррелирующего фактора составим матрицу корреляций с остатками регрессии.

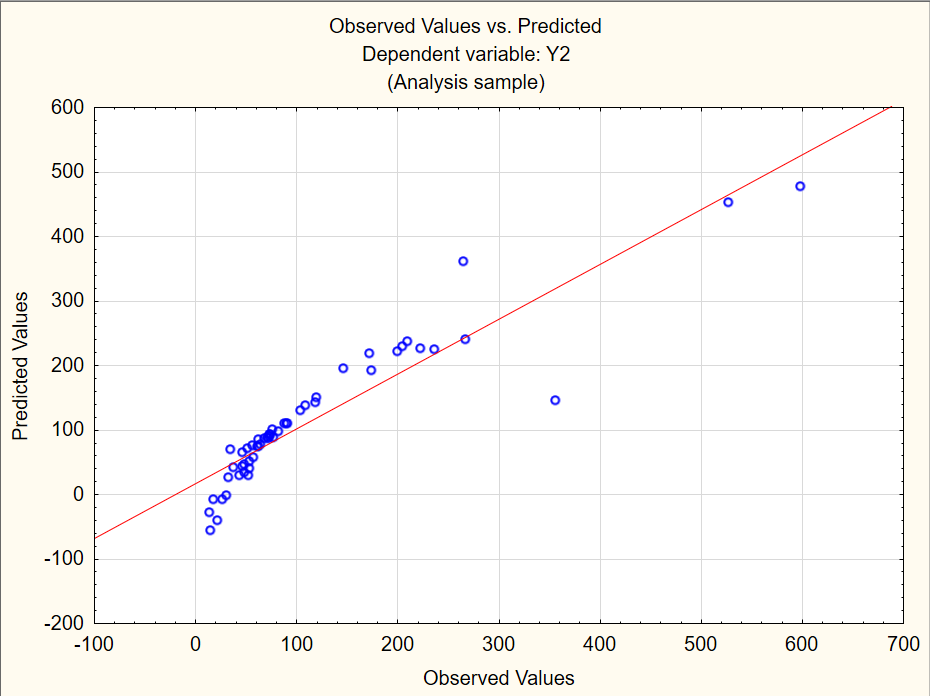
Наибольшим по модулю значением корреляции обладает параметр X1, добавим его в регрессионную модель:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N=53 | Regression Summary for Dependent Variable: Y2 (Spreadsheet2 in Workbook1) R= ,92187396 R?= ,84985160 Adjusted R?= ,84384567 F(2,50)=141,50 p<0,0000 Std.Error of estimate: 46,209 | | | | | |
| b\* | Std.Err. (of b\*) | b | Std.Err. (of b) | t(50) | p-value |
| Intercept |  |  | 115,135 | 26,24904 | 4,38627 | 0,000059 |
| X1 | -0,360760 | 0,059519 | -399,935 | 65,98221 | -6,06126 | 0,000000 |
| X10 | 0,719166 | 0,059519 | 0,004 | 0,00036 | 12,08296 | 0,000000 |

Все коэффициенты модели являются значимыми

|  |  |
| --- | --- |
| F | p |
| 141,5 | 0,00 |

Модель является статистически значимой



Составим матрицу корреляций с остатками регрессии.

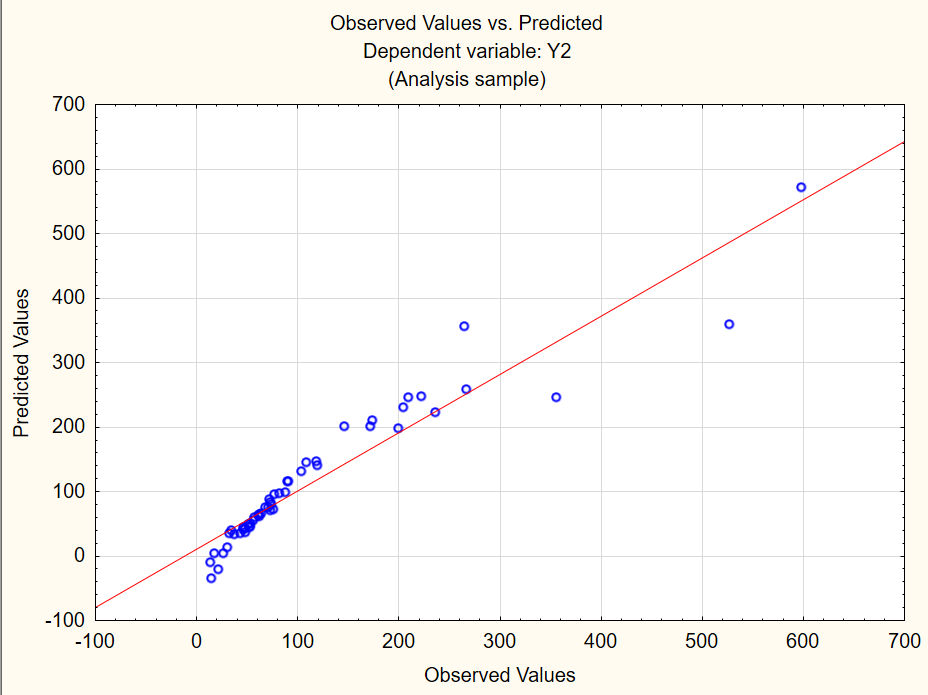
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Correlations (Spreadsheet2 in Workbook1) N=53 (Casewise deletion of missing data) | | | | | | | | | |
| X1 | X2 | X3 | X4 | X9 | X10 | X11 | X12 | X13 | X14 |
| ResidsAt2 | 0,000000 | 0,140270 | -0,156424 | 0,243745 | 0,386104 | 0,000000 | 0,240179 | 0,065357 | -0,016521 | -0,007663 |

Параметр X9 наиболее всего коррелирует с остатками, добавим его в модель:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Effect |  |  |  |  |
| Y2 (Param.) | Y2 (Std.Err) | Y2 (t) | Y2 (p) |
| Intercept | 46,782 | 25,17084 | 1,85857 | 0,069101 |
| X1 | -222,045 | 63,90354 | -3,47469 | 0,001080 |
| X9 | 0,445 | 0,08659 | 5,14047 | 0,000005 |
| X10 | 0,003 | 0,00035 | 9,73774 | 0,000000 |

Свободный член модели перестал быть статистически значимым

|  |  |
| --- | --- |
| F | p |
| 151,1 | 0,00 |



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | Correl |  |  |  |  |  |  |  |  |  |
| X1 | X2 | X3 | X4 | X9 | X10 | X11 | X12 | X13 | X14 |
| ResidsAt3 | 0,000000 | 0,115188 | -0,074626 | 0,156218 | 0,000000 | 0,000000 | -0,117573 | 0,044802 | 0,043594 | -0,039347 |

Остановимся, поскольку коэффициенты корреляции модели перестали быть значимыми.

Итоговая модель:

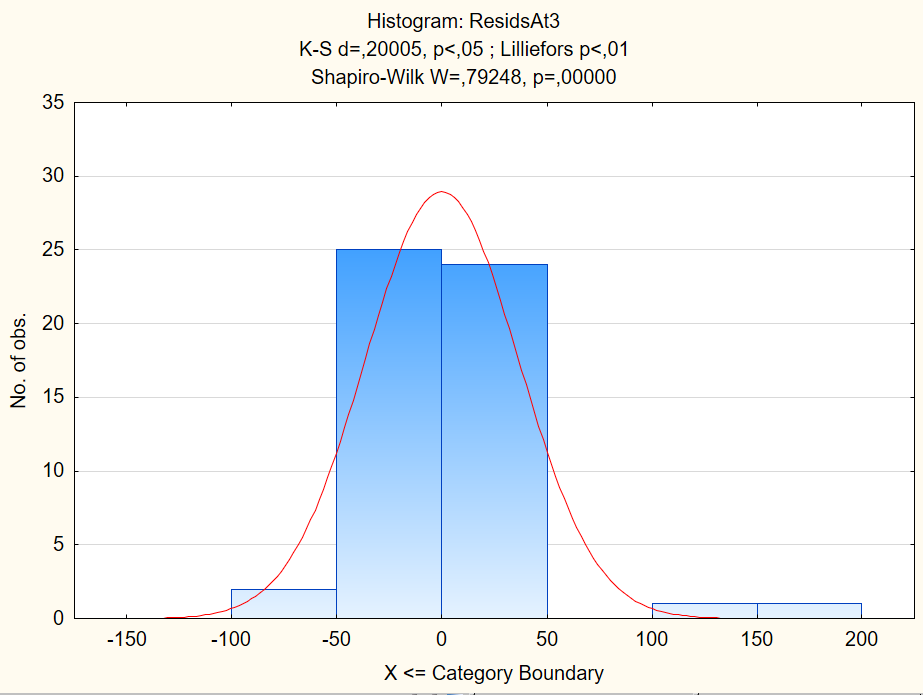
Параметры модели:

|  |  |  |
| --- | --- | --- |
| MS Residual | R2 | Adj, R2 |
| 1415,51779 | 0,902455 | 0,896483 |

С помощью теста Шапиро-Уилка проверим нормальность остатков:

|  |  |
| --- | --- |
| SW | P-Value |
| 0,79248 | 0,000 |

То есть остатки не распределены нормально.



|  |  |
| --- | --- |
| LM | P-Value |
| 10,48 | 0,02 |

Таким образом остатки гетероскедастичны.