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

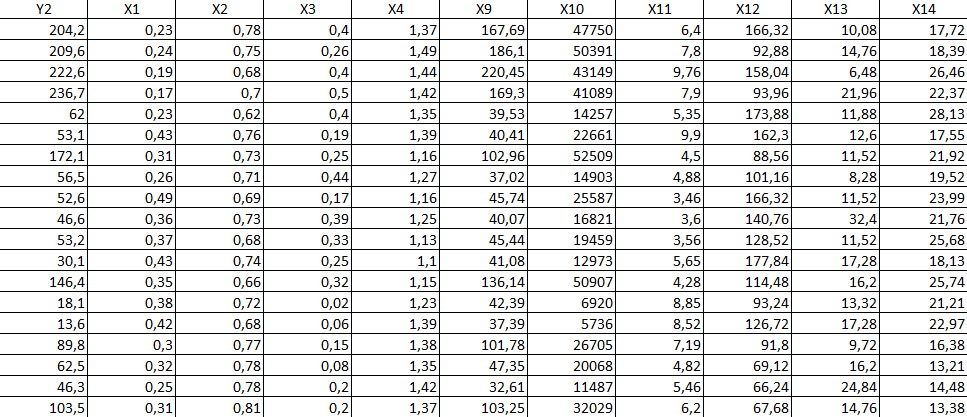
**Цели работы:**

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.

Выходные данные и важные метрики для моделей:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Добавл Фактор** | **Коэффициент** | **P-значение** | **R^2** | **Скорр R^2** | **F-стат** | **P-значение (F-стат)** | **Статистика ДУ** |
| X10 | 0,005170025 | 1,62444E-16 | 0,739525928 | 0,734418593 | 144,7968394 | 1,62444E-16 | 2,016737942 |
| X9 | 0,608048239 | 8,0761E-10 | 0,878420291 | 0,873557103 | 180,6264183 | 1,3229E-23 | 2,557630234 |
| X1 | -222,0450777 | 0,001080035 | 0,902455113 | 0,896482977 | 151,1109455 | 9,28405E-25 | 2,892406961 |

Итоговая модель имеет следующий вид:

Уровни значимости для всех коэффициентов итоговой модели:

|  |  |
| --- | --- |
| Const | 0.069 |
| X10 | 0.00 |
| X9 | 0.00 |
| X1 | 0.001 |

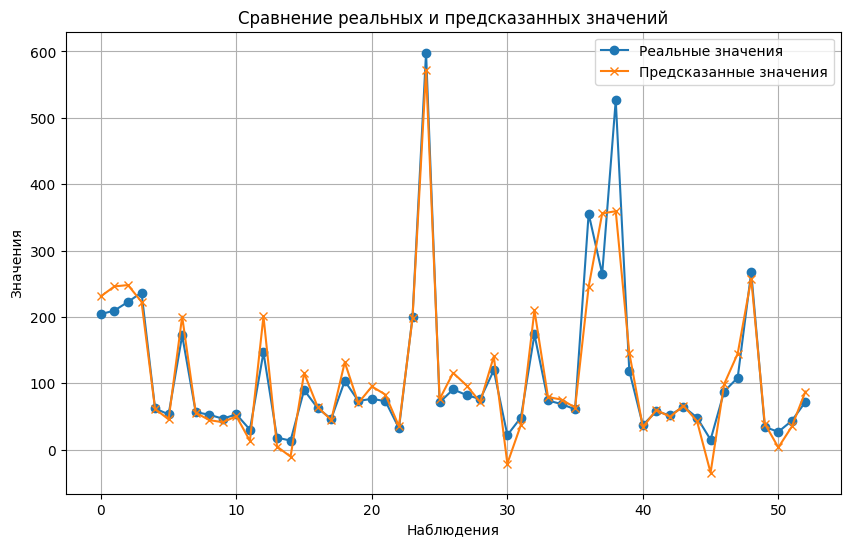
Уровень статистик:

F-statistic: 151.1

Prob (F-statistic): 9.28e-25

R-squared: 0.902

График предсказанных и реальных значений



Для проверки остатков на однородность воспользуемся тестом Шапиро-Уилка:

Тест Шапиро-Уилка: статистика = 0.7925, p-value = 0.0000

Остатки не распределены нормально (p-value ≤ 0.05).

Для проверки остатков на однородность воспользуемся тестом Бреуша-Пагана:

Тест Бреуша-Пагана: статистика = 15.9015, p-value = 0.0012

Остатки гетероскедастичны (p-value ≤ 0.05).

**Вывод:** хотя модель обладает высоким значением R^2 ~ 0.902, остатки модели не являются однородными или нормальными, что вообще говоря будет отрицательно влиять на поведение предсказанных далее значений.