|  |  |
| --- | --- |
|  | |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ | |
| Федеральное государственное автономное образовательное учреждение высшего образования | |
| **«Дальневосточный федеральный университет» (ДВФУ)** | |
| **Институт математики и компьютерных технологий** | |
| **Департамент информационных и компьютерных систем** | |
| **ОТЧЁТ** | |
| по лабораторной работе №2  «Множественная регрессия и корреляция» | |
| по дисциплине «Эконометрика» | |
| Направление «Прикладная информатика в экономике» | |
|  | |
|  | Выполнил студент группы Б9121–09.03.03пиэ/2  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_А. А. Соломоненко |
| Проверил старший преподаватель  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Е. И. Шувалова  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  оценка |
| Г. Владивосток  2024г. | |

# **1. Условие задачи**

Цель: получить навыки построения и оценивания модели множественной регрессии.

Используя данные о стоимости автомобиля, производителе, объеме двигателя, мощности двигателя и пробеге:

1. Постройте уравнение множественной регрессии в линейной форме с полным набором факторов.
2. Установить какие факторы коллинеарны и удалить зависимые факторы
3. Постройте уравнение регрессии со статистически значимыми факторами.
4. Оцените статистическую значимость уравнений регрессии с помощью критерия Фишера.

**2. Ход работы**

Исходные данные о стоимости автомобиля, производителе, объеме двигателя, мощности двигателя и пробеге взяты с автомобильного портала Дром. Данные представлены в таблице 1.

Таблица 1 – Исходные данные

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер | Производитель | Стоимость, тыс. руб., y | Мощность, л. с., x1 | Объем двигателя, л., x2 | Пробег тыс. км., x3 |
| 1 | Nissan | 455 | 109 | 1,5 | 169 |
| 2 | Nissan | 585 | 141 | 2 | 186 |
| 3 | Honda | 770 | 88 | 1,3 | 216 |
| 4 | Toyota | 795 | 220 | 3 | 273 |
| 5 | Toyota | 835 | 69 | 1 | 132 |
| 6 | Toyota | 855 | 69 | 1 | 130 |
| 7 | Toyota | 890 | 69 | 1 | 69 |
| 8 | Toyota | 899 | 95 | 1,3 | 313 |
| 9 | Honda | 950 | 100 | 1,3 | 83 |
| 10 | Toyota | 1150 | 185 | 3,4 | 250 |
| 11 | Honda | 1211 | 118 | 1,5 | 100,5 |
| 12 | Mitsubishi | 1260 | 139 | 1,8 | 139 |
| 13 | Nissan | 1319 | 133 | 1,8 | 10 |
| 14 | Honda | 1388 | 140 | 1,8 | 79 |
| 15 | Audi | 1390 | 122 | 1,4 | 133 |
| 16 | Kia | 1480 | 150 | 2 | 83 |
| 17 | BMW | 1580 | 184 | 2 | 260 |
| 18 | Nissan | 1599 | 147 | 2 | 178 |
| 19 | Nissan | 1964 | 150 | 2 | 120 |
| 20 | Mitsubishi | 2000 | 139 | 1,8 | 28 |
| 21 | Mitsubishi | 2722 | 128 | 2,4 | 94 |
| 22 | Mercedes | 2780 | 156 | 1,6 | 50 |
| 23 | Lexus | 2800 | 150 | 2 | 77 |
| 24 | Honda | 2850 | 190 | 1,5 | 30 |
| 25 | Kia | 2850 | 202 | 2,2 | 107 |
| 26 | Kia | 2950 | 194 | 2,5 | 96 |
| 27 | Toyota | 3000 | 235 | 4,5 | 261 |
| 28 | Honda | 3050 | 192 | 1,5 | 68 |
| 29 | Kia | 3250 | 250 | 3 | 110 |
| 30 | BMW | 3350 | 258 | 2 | 93 |
| 31 | Mercedes | 3500 | 258 | 3 | 113 |
| 32 | Kia | 3649 | 202 | 2,2 | 60 |
| 33 | Audi | 3900 | 245 | 2 | 29 |
| 34 | Toyota | 3999 | 177 | 2,8 | 100 |
| 35 | BMW | 5100 | 249 | 3 | 128 |
| 36 | Mercedes | 7800 | 435 | 3 | 58 |
| 37 | Lexus | 9300 | 279 | 2,4 | 16 |

# **2.1 Уравнение в явном виде**

Для проведения регрессионо-корреляционного анализа все факторы должны быть количественно определенны. Фактор «Производитель» является не количественным, поэтому ему была придана количественная определенность согласно таблице 2.

Таблица 2 – Количественное определение фактора «Производитель»

|  |  |
| --- | --- |
| Производитель | Количественное определение |
| Audi | 1 |
| BMW | 2 |
| Honda | 3 |
| Kia | 4 |
| Lexus | 5 |
| Mercedes | 6 |
| Mitsubishi | 7 |
| Nissan | 8 |
| Toyota | 9 |

Для проведения дальнейшего анализа, используя пакет анализа MS Excel, вычислены параметры множественной регрессии в линейной форме с полным набором факторов. Параметры представлены на рисунке 1.

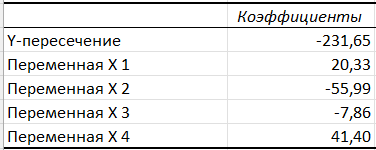


Рисунок 1 – Коэффициенты уравнения

Полученное уравнение в явном виде:

ŷ = -231,65 + 20,33 x1 - 55,86 x2 - 7,86 x3 + 41,4 x4

Характеристика поведения параметров при неизменных других:

При увеличении мощности двигателя на 1 л. с. стоимость автомобиля увеличиться на 20,33 тыс. рублей.

При увеличении объема двигателя на 1 л. стоимость автомобиля уменьшиться на 55,99 тыс. рублей.

При увеличении пробега автомобиля на 1 тыс. км. стоимость автомобиля уменьшиться на 7,86 тыс. рублей.

Критерий Фишера (F-критерий), который является показателем надежности и значимости уравнения в целом равен: 23,84, табличное значение критерия Фишера: 2,64.

Так как , уравнение надежно, значимо, показатель тесноты связи надежен и отражает устойчивую зависимость стоимости автомобиля от производителя, объема двигателя, мощности двигателя и пробеге автомобиля.

# **2.2 Анализ факторов**

Для анализа факторов была построена матрица парных корреляций, представленная на рисунке 2.

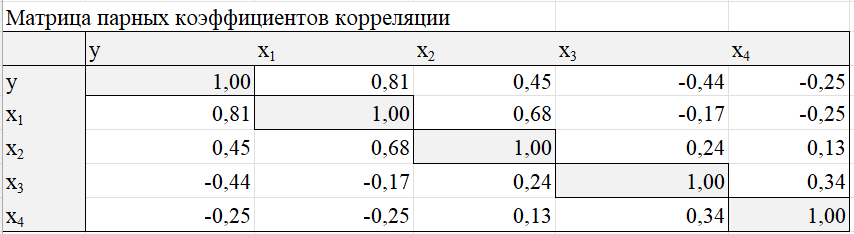


Рисунок 2 – Матрица парных корреляций

Исходя из теоретического анализа, а также коэффициента корреляции признаков x1 и x2 равного 0,68, принято решение о подробном рассмотрении.

Коэффициенты корреляции переменной x1: 0,81; 0,68; -0,17; -0,25;

Коэффициенты корреляции переменной x2: 0,45; 0,68; 0,24; 0,13;

В результате сравнения корреляции признаков x1 и x2 с другими было принято решении об исключении признака x2, так как признак x1 имеет достаточную тесноту связи с результатом, а также имеет меньшую тесноту связи с другими факторами.

# **2.3 Уравнение без коррелирующих признаков**

Вычисленные коэффициенты множественной регрессии в линейной форме с новым набором факторов представлены на рисунке 3.

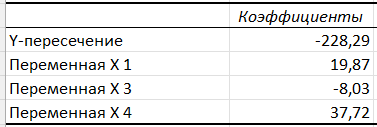


Рисунок 3 – Коэффициенты уравнения

Уравнение без коррелирующих признаков в явном виде:

ŷ = -228,29 + 19,87 x1 – 8,03 x3 + 37,72x4

Критерий Фишера (F-критерий), который является показателем надежности и значимости уравнения в целом равен 32,76, табличное значение критерия Фишера - 2,85.

Так как , уравнение надежно, значимо, показатель тесноты связи надежен и отражает устойчивую зависимость стоимости автомобиля от производителя, мощности двигателя и пробеге автомобиля.

Характеристика поведения параметров при неизменных других:

При увеличении мощности двигателя на 1 л. с. стоимость автомобиля увеличиться на 19,87 тыс. рублей.

При увеличении пробега автомобиля на 1 тыс. км. стоимость автомобиля уменьшиться на 8,03 тыс. рублей.

# **2.4 Стандартизация уравнения**

Рассчитанные стандартизированные параметры β представлены на рисунке 4.

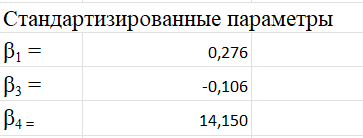


Рисунок 4 – Стандартизированные параметры

Уравнение в стандартизированном виде:

Влияние факторов:

1. наибольшее влияние имеет пробег автомобиля;
2. среднее влияние оказывает мощности двигателя;
3. наименьшее влияние имеет пробег автомобиля.

# **2.6 Теорема Гаусса-Маркова**

Исследование случайных отклонений.

На рисунке 5 представлена зависимость остатков от теоретических значений y. Точки распределены равномерно, что показывает, что остатки являются случайными величинами, и использование метода наименьших квадратов (МНК) оправдано. Теоретические значения ŷx хорошо соответствуют фактическим значениям y. Дисперсия случайной ошибки не меняется в зависимости от номера наблюдения, поэтому условие гомоскедастичности выполняется.

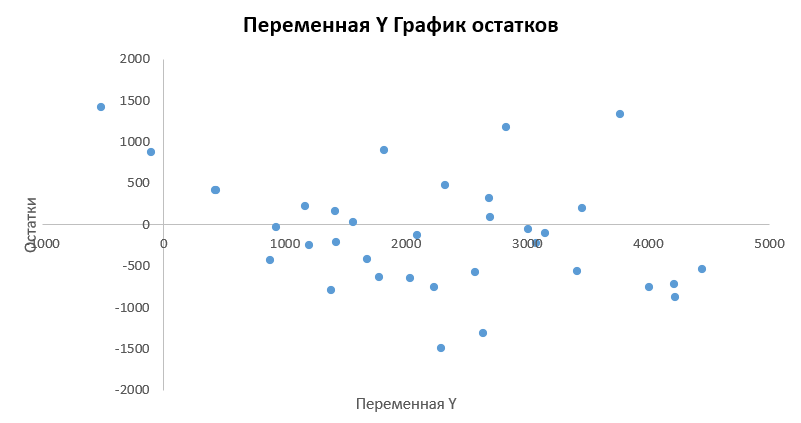




Рисунок 5 - График зависимости остатков от значений y

Построен график зависимости случайных остатков от фактора x1. Остатки на графике представлены в виде случайных величин, они не зависят от значений x1, что свидетельствует о несмещённости оценок коэффициентов регрессии. Данный график представлен на рисунке 6.





Рисунок 6 - График зависимости остатков от переменной x1

Построен график зависимости случайных остатков от фактора x3. Остатки на графике представлены в виде случайных величин, они не зависят от значений x3, что свидетельствует о несмещённости оценок коэффициентов регрессии. Данный график представлен на рисунке 7.

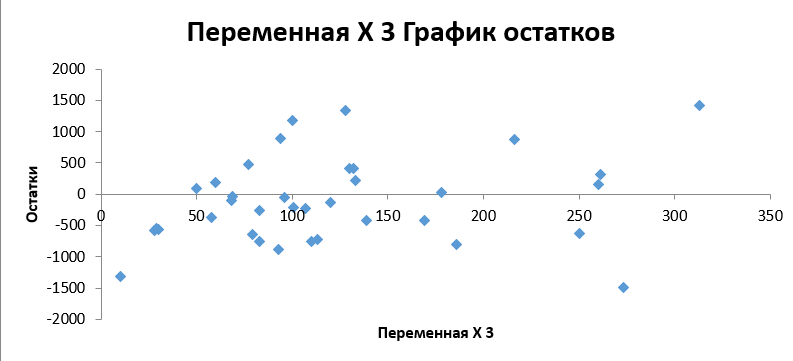




Рисунок 7 - График зависимости остатков от переменной x3

Построен график зависимости случайных остатков от фактора x4. Остатки на графике представлены в виде случайных величин, они не зависят от значений x4, что свидетельствует о несмещённости оценок коэффициентов регрессии. Данный график представлен на рисунке 8.

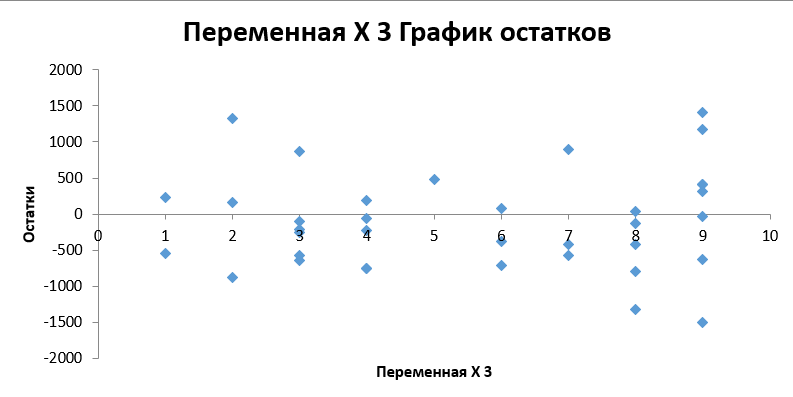




Рисунок 8 - График зависимости остатков от переменной x4

График нормального распределения отображает стандартный вид и соответствует ожидаемому распределению. Случайные ошибки на нем распределены нормально. График представлен на рисунке 9.

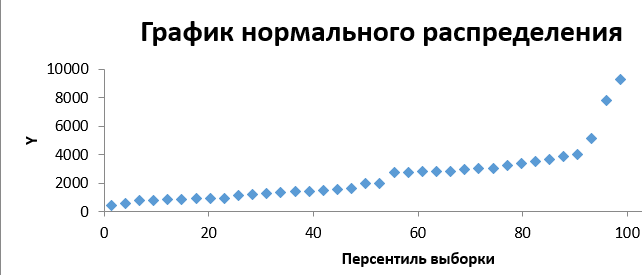


Рисунок 9 – График нормального распределения

Условия Гаусса-Маркова выполняются, оценки параметров регрессии являются несмещёнными, состоятельными и эффективными.

# **3. Полученные результаты**

В результате выполнения данной лабораторной работы, используя данные о стоимости автомобиля, производителе, объеме двигателя, мощности двигателя и пробеге:

1. Построено уравнение множественной регрессии в линейной форме с полным набором факторов.
2. Установлено какие факторы коллинеарны и удален зависимый фактор.
3. Построено уравнение регрессии со статистически значимыми факторами.
4. Произведена оценка статистической значимости уравнений регрессии с помощью критерия Фишера.
5. Произведено исследование случайных отклонений используя теорему Гаусса-Маркова.