**1. Оценка уравнения регрессии**.

Определим вектор оценок коэффициентов регрессии. Согласно методу наименьших квадратов, вектор *s* получается из выражения: s = (XTX)-1XTY

К матрице с переменными Xj добавляем единичный столбец:

|  |  |  |
| --- | --- | --- |
| 1 | 9.3 | 13.195 |
| 1 | 10.2 | 14.996 |
| 1 | 9.6 | 15.511 |
| 1 | 8.8 | 20.18 |
| 1 | 8.1 | 23.472 |
| 1 | 7.2 | 25.231 |
| 1 | 9.6 | 19.868 |
| 1 | 12.7 | 28.203 |
| 1 | 17.9 | 33.948 |
| 1 | 24.4 | 35.452 |
| 1 | 27.5 | 36.236 |
| 1 | 26.5 | 36.007 |

Матрица Y

|  |
| --- |
| 171.4 |
| 185.9 |
| 297.6 |
| 213.2 |
| 228.2 |
| 242.9 |
| 237.4 |
| 226.2 |
| 207.8 |
| 194.2 |
| 182.4 |
| 179.1 |

Матрица XT

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9.3 | 10.2 | 9.6 | 8.8 | 8.1 | 7.2 | 9.6 | 12.7 | 17.9 | 24.4 | 27.5 | 26.5 |
| 13.195 | 14.996 | 15.511 | 20.18 | 23.472 | 25.231 | 19.868 | 28.203 | 33.948 | 35.452 | 36.236 | 36.007 |

Умножаем матрицы, (XTX)

В матрице, (XTX) число 12, лежащее на пересечении 1-й строки и 1-го столбца, получено как сумма произведений элементов 1-й строки матрицы XT и 1-го столбца матрицы X

Умножаем матрицы, (XT∙Y)

Находим обратную матрицу (XTX)-1

Вектор оценок коэффициентов регрессии равен

Уравнение регрессии (оценка уравнения регрессии)

Y = 229.6968-3.9397X1 + 1.6103X2

Интерпретация коэффициентов регрессии. Константа оценивает агрегированное влияние прочих (кроме учтенных в модели хi) факторов на результат Y и означает, что Y при отсутствии xi составила бы 229.6968. Коэффициент b1 указывает, что с увеличением x1 на 1, Y снижается на 3.9397. Коэффициент b2 указывает, что с увеличением x2 на 1, Y увеличивается на 1.6103.

**2. Матрица парных коэффициентов корреляции R**.

Матрица A, составленная из Y и X.

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 171.4 | 9.3 | 13.195 |
| 1 | 185.9 | 10.2 | 14.996 |
| 1 | 297.6 | 9.6 | 15.511 |
| 1 | 213.2 | 8.8 | 20.18 |
| 1 | 228.2 | 8.1 | 23.472 |
| 1 | 242.9 | 7.2 | 25.231 |
| 1 | 237.4 | 9.6 | 19.868 |
| 1 | 226.2 | 12.7 | 28.203 |
| 1 | 207.8 | 17.9 | 33.948 |
| 1 | 194.2 | 24.4 | 35.452 |
| 1 | 182.4 | 27.5 | 36.236 |
| 1 | 179.1 | 26.5 | 36.007 |

Транспонированная матрица.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 171.4 | 185.9 | 297.6 | 213.2 | 228.2 | 242.9 | 237.4 | 226.2 | 207.8 | 194.2 | 182.4 | 179.1 |
| 9.3 | 10.2 | 9.6 | 8.8 | 8.1 | 7.2 | 9.6 | 12.7 | 17.9 | 24.4 | 27.5 | 26.5 |
| 13.195 | 14.996 | 15.511 | 20.18 | 23.472 | 25.231 | 19.868 | 28.203 | 33.948 | 35.452 | 36.236 | 36.007 |

Матрица XT∙X.

|  |  |  |  |
| --- | --- | --- | --- |
| 12 | 2566.3 | 171.8 | 302.299 |
| 2566.3 | 562798.67 | 35192.65 | 63546.404 |
| 171.8 | 35192.65 | 3105.3 | 4946.233 |
| 302.299 | 63546.404 | 4946.233 | 8443.359 |

Полученная матрица имеет следующее соответствие:

|  |  |  |  |
| --- | --- | --- | --- |
| ∑n | ∑y | ∑x1  | ∑x2  |
| ∑y | ∑y2  | ∑x1 y | ∑x2 y |
| ∑x1  | ∑yx1  | ∑x1 2  | ∑x2 x1  |
| ∑x2  | ∑yx2  | ∑x1 x2  | ∑x2 2  |

Найдем парные коэффициенты корреляции.

Значения парного коэффициента корреляции свидетельствует о умеренной линейной связи между x1 и y.

Значения парного коэффициента корреляции свидетельствует о не сильной линейной связи между x2 и y.

Значения парного коэффициента корреляции свидетельствует о сильной линейной связи между x2 и x1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Признаки x и y | ∑xi |  | ∑yi |  | ∑xi∙yi |  |
| Для y и x1 | 171.8 | 14.317 | 2566.3 | 213.858 | 35192.65 | 2932.721 |
| Для y и x2 | 302.299 | 25.192 | 2566.3 | 213.858 | 63546.404 | 5295.534 |
| Для x1 и x2 | 302.299 | 25.192 | 171.8 | 14.317 | 4946.233 | 412.186 |

Дисперсии и среднеквадратические отклонения.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Признаки x и y |  |  |  |  |
| Для y и x1 | 53.808 | 1164.502 | 7.335 | 34.125 |
| Для y и x2 | 68.997 | 1164.502 | 8.306 | 34.125 |
| Для x1 и x2 | 68.997 | 53.808 | 8.306 | 7.335 |

Матрица парных коэффициентов корреляции R:

|  |  |  |  |
| --- | --- | --- | --- |
| - | y | x1 | x2 |
| y | 1 | -0.5154 | -0.3242 |
| x1 | -0.5154 | 1 | 0.8457 |
| x2 | -0.3242 | 0.8457 | 1 |

**Частные коэффициенты корреляции**.

Частные коэффициенты корреляции вычисляются по формуле:

где Rij - алгебраическое дополнение элемента rij матрицы R.

Теснота связи не сильная.

Теснота связи низкая.

Теснота связи сильная.

При сравнении коэффициентов парной и частной корреляции видно, что из-за влияния межфакторной зависимости между xi происходит завышение оценки тесноты связи между переменными.

**Анализ мультиколлинеарности**.

1. Анализ мультиколлинеарности на основе матрицы коэффициентов корреляции.

В нашем случае r(x1x2) имеют |r|>0.7, что говорит о мультиколлинеарности факторов и о необходимости исключения одного из них из дальнейшего анализа.

2. Ридж-регрессия.

Наиболее детальным показателем наличия проблем, связанных с мультиколлинеарностью, является коэффициент увеличения дисперсии, определяемый для каждой переменной как:

где Rj2 коэффициент множественной детерминации в регрессии Xj на прочие X.

О мультиколлинеарности будет свидетельствовать VIF от 4 и выше хотя бы для одного j.

По данному критерию мультиколлинеарность отсутствует.

3. Критерием плохой обсуловленности является высокая величина отношения λmax/λmin максимального и минимального собственных чисел матрицы XTX — называемого показателем обусловленности. Это соотношение также позволяет судить о степени серьезности проблем мультиколлинеарности: показатель обусловленности в пределах от 10 до 100 свидетельствует об умеренной коллинеарности, свыше 1000 — об очень серьезной коллинеарности.

Для оценки β-коэффициентов применим МНК. При этом система нормальных уравнений будет иметь вид:

rx1y=β1+rx1x2•β2 + ... + rx1xm•βm

rx2y=rx2x1•β1 + β2 + ... + rx2xm•βm

...

rxmy=rxmx1•β1 + rxmx2•β2 + ... + βm

Для наших данных (берем из матрицы парных коэффициентов корреляции):

-0.515 = β1 + 0.846β2

-0.324 = 0.846β1 + β2

Данную систему линейных уравнений решаем методом Гаусса: β1 = -0.847; β2 = 0.392;

Искомое уравнение в стандартизованном масштабе: ty=β1tx1+β2tx2

Расчет β-коэффициентов можно выполнить и по формулам:

Стандартизированная форма уравнения регрессии имеет вид:

ty = -0.847x1 + 0.392x2

Найденные из данной системы β–коэффициенты позволяют определить значения коэффициентов в регрессии в естественном масштабе по формулам:

**3. Анализ параметров уравнения регрессии**.

Перейдем к статистическому анализу полученного уравнения регрессии: проверке значимости уравнения и его коэффициентов, исследованию абсолютных и относительных ошибок аппроксимации

Для несмещенной оценки дисперсии проделаем следующие вычисления:

Несмещенная ошибка ε = Y - Y(x) = Y - X∙s (абсолютная ошибка аппроксимации)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Y | Y(x) | ε = Y - Y(x) | ε2 |  | |ε : Y| |
| 171.4 | 214.305 | -42.905 | 1840.825 | 1802.71 | 0.25 |
| 185.9 | 213.659 | -27.759 | 770.573 | 781.668 | 0.149 |
| 297.6 | 216.852 | 80.748 | 6520.186 | 7012.667 | 0.271 |
| 213.2 | 227.523 | -14.323 | 205.135 | 0.433 | 0.0672 |
| 228.2 | 235.581 | -7.381 | 54.485 | 205.683 | 0.0323 |
| 242.9 | 241.96 | 0.94 | 0.884 | 843.418 | 0.00387 |
| 237.4 | 223.868 | 13.532 | 183.106 | 554.21 | 0.057 |
| 226.2 | 225.077 | 1.123 | 1.262 | 152.317 | 0.00497 |
| 207.8 | 213.841 | -6.041 | 36.496 | 36.703 | 0.0291 |
| 194.2 | 190.655 | 3.545 | 12.569 | 386.45 | 0.0183 |
| 182.4 | 179.704 | 2.696 | 7.268 | 989.627 | 0.0148 |
| 179.1 | 183.275 | -4.175 | 17.431 | 1208.142 | 0.0233 |
|   |   |   | 9650.22 | 13974.029 | 0.922 |

Средняя ошибка аппроксимации

Оценка дисперсии равна:

se2=(Y-Y(X))T(Y-Y(X))=9650.22

Несмещенная оценка дисперсии равна:

Оценка среднеквадратичного отклонения (*стандартная ошибка для оценки Y*):

Найдем оценку ковариационной матрицы вектора k = S2 • (XTX)-1

Дисперсии параметров модели определяются соотношением S2i = Kii, т.е. это элементы, лежащие на главной диагонали

**Множественный коэффициент корреляции (Индекс множественной корреляции).**

R = = = 0.5563

Коэффициент множественной корреляции можно определить через матрицу парных коэффициентов корреляции:

где Δr - определитель матрицы парных коэффициентов корреляции; Δr11 - определитель матрицы межфакторной корреляции.

Коэффициент множественной корреляции

Аналогичный результат получим при использовании других формул:

Связь между признаком Y и факторами Xi умеренная.

**Коэффициент детерминации**.

R2= 0.55632 = 0.3094

Более объективной оценкой является скорректированный коэффициент детерминации:

Чем ближе этот коэффициент к единице, тем больше уравнение регрессии объясняет поведение Y.

Добавление в модель новых объясняющих переменных осуществляется до тех пор, пока растет скорректированный коэффициент детерминации.

**5. Проверка гипотез относительно коэффициентов уравнения регрессии (проверка значимости параметров множественного уравнения регрессии)**.

1) t-статистика

Tтабл (n-m-1;α/2) = (9;0.025) = 2.685

Статистическая значимость коэффициента регрессии b0 подтверждается.

Статистическая значимость коэффициента регрессии b1 не подтверждается.

Статистическая значимость коэффициента регрессии b2 не подтверждается.

*Доверительный интервал для коэффициентов уравнения регрессии*.

Определим доверительные интервалы коэффициентов регрессии, которые с надежность 95% будут следующими:

(bi - tтабл∙Sbi; bi + tтабл∙Sbi)

b0: (229.697 - 2.685∙32.079 ; 229.697 + 2.685∙32.079) = (143.565;315.828)

b1: (-3.94 - 2.685∙2.414 ; -3.94 + 2.685∙2.414) = (-10.422;2.543)

Поскольку найденный интервал включает 0, то коэффициент b1 не значим.

b2: (1.61 - 2.685∙2.132 ; 1.61 + 2.685∙2.132) = (-4.114;7.335)

Поскольку найденный интервал включает 0, то коэффициент b2 не значим.

**6. Проверка общего качества уравнения множественной регрессии**.

F-статистика. Критерий Фишера.

Проверим гипотезу об общей значимости - гипотезу об одновременном равенстве нулю всех коэффициентов регрессии при объясняющих переменных:

H0: R2 = 0; β1 = β2 = ... = βm = 0.

H1: R2 ≠ 0.

Проверка этой гипотезы осуществляется с помощью F-статистики распределения Фишера (правосторонняя проверка).

Если F < Fkp = Fα ; n-m-1, то нет оснований для отклонения гипотезы H0.

Табличное значение при степенях свободы k1 = 2 и k2 = n-m-1 = 12 - 2 - 1 = 9, Fkp(2;9) = 4.2565

Поскольку фактическое значение F < Fkp, то коэффициент детерминации статистически не значим и уравнение регрессии статистически ненадежно (совместная незначимость коэффициентов при факторах xi подтверждается).

**Выводы**.

В результате расчетов было получено уравнение множественной регрессии: Y = 229.6968-3.9397X1 + 1.6103X2. Возможна экономическая интерпретация параметров модели: увеличение X1 на 1 ед.изм. приводит к уменьшению Y в среднем на 3.94 ед.изм.; увеличение X2 на 1 ед.изм. приводит к увеличению Y в среднем на 1.61 ед.изм. Статистическая значимость уравнения проверена с помощью коэффициента детерминации и критерия Фишера. Установлено, что в исследуемой ситуации 30.94% общей вариабельности Y объясняется изменением факторов Xj. Установлено также, что один или несколько параметров модели статистически не значимы.

Согласно матрице парных коэффициентов корреляции r(x1x2) имеют |r|>0.7, что говорит о мультиколлинеарности факторов и о необходимости исключения одного из них из дальнейшего анализа.

Решение было получено и оформлено с помощью сервиса:

[Уравнение множественной регрессии](https://math.semestr.ru/regress/corel.php)

Вместе с этой задачей решают также:

[Уравнение парной линейной регрессии](https://math.semestr.ru/corel/corel.php)

[Выявление тренда методом аналитического выравнивания](https://math.semestr.ru/trend/analis.php)

[Уравнение нелинейной регрессии](https://math.semestr.ru/corel/noncorel.php)

[Проверка на автокорреляцию](https://math.semestr.ru/corel/autocorrelation.php)

[Системы эконометрических уравнений](https://math.semestr.ru/regress/systems.php)

[Метод статистических уравнений зависимостей](https://math.semestr.ru/regress/system.php)