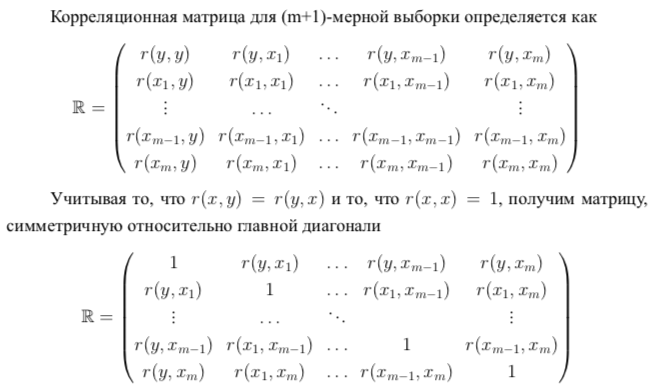
1. **Описание и постановка залачи**

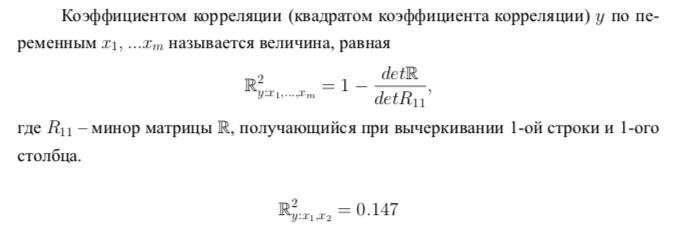
**Задача работы:** Регрессионный анализ множественной выборки

Даны отценки в 3х классах разных школ

|  |  |  |
| --- | --- | --- |
| Y | X1 | X2 |
| школа 1 | школа 2 | школа 3 |
| 2 | 4 | 5 |
| 3 | 3 | 5 |
| 4 | 4 | 3 |
| 2 | 4 | 2 |
| 3 | 4 | 4 |
| 4 | 3 | 3 |
| 5 | 5 | 2 |
| 2 | 2 | 2 |
| 2 | 2 | 3 |
| 4 | 3 | 2 |
| 5 | 5 | 4 |
| 5 | 4 | 3 |
| 5 | 3 | 4 |
| 3 | 4 | 3 |
| 2 | 2 | 5 |
| 4 | 3 | 3 |
| 4 | 4 | 4 |
| 3 | 2 | 3 |
| 5 | 5 | 4 |
| 4 | 5 | 5 |
| 5 | 3 | 5 |
| 5 | 3 | 4 |
| 3 | 4 | 5 |
| 5 | 3 | 3 |
| 5 | 4 | 5 |
| 2 | 3 | 2 |
| 3 | 4 | 3 |
| 4 | 2 | 2 |
| 2 | 4 | 3 |
| 3 | 3 | 4 |
| 4 | 4 | 5 |
| 4 | 5 | 4 |
| 4 | 4 | 2 |
| 5 | 3 | 4 |
| 5 | 5 | 3 |
| 4 | 2 | 4 |
| 3 | 2 | 5 |
| 4 | 4 | 2 |
| 5 | 2 | 5 |
| 3 | 3 | 3 |
| 2 | 5 | 4 |
| 4 | 4 | 3 |
| 4 | 5 | 4 |
| 5 | 4 | 3 |
| 3 | 3 | 4 |
| 3 | 4 | 5 |
| 2 | 2 | 4 |
| 4 | 4 | 3 |
| 5 | 3 | 4 |
| 4 | 3 | 2 |

1. **Корреляционная матрица**



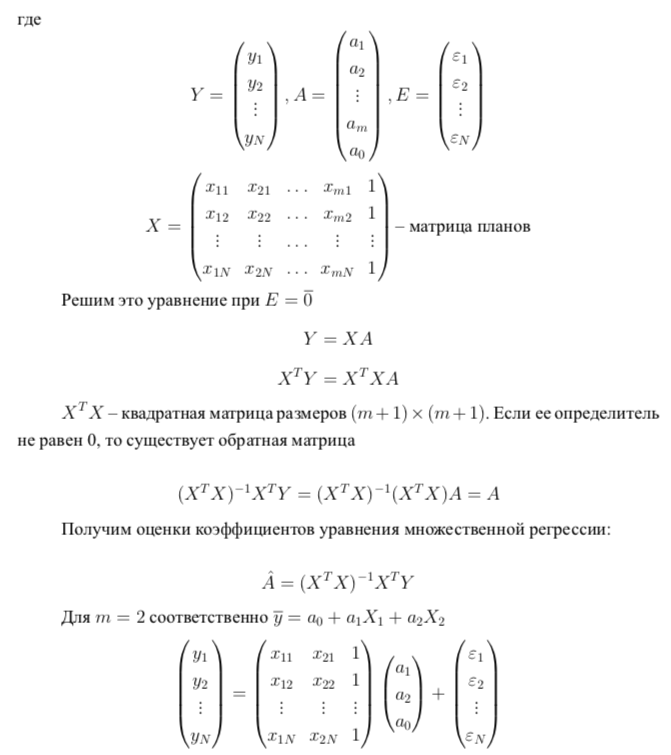




1. **Уравнение линейной множественной регрессии**

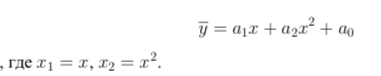
Уравнение линейной множественной регрессии записывается в виде y = a0 + a1X1 + a2X2 + ... + amXm. Оценки коэффициентов уравнения множественной регрессии получаются по матричному методу наименьших квадратов.

Y =XA+E,



A = ( )

полиномиальная регрессия имеет вид:



теоретическое линейное уравнение множественной регрессии имеет вид:

Y = β0 + β1X1 + β2X2 + ... + βmXm + ε

β0 - свободный член, определяющий значение Y, в случае, когда все объясняющие переменные Xj равны 0.

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

……………………………………….

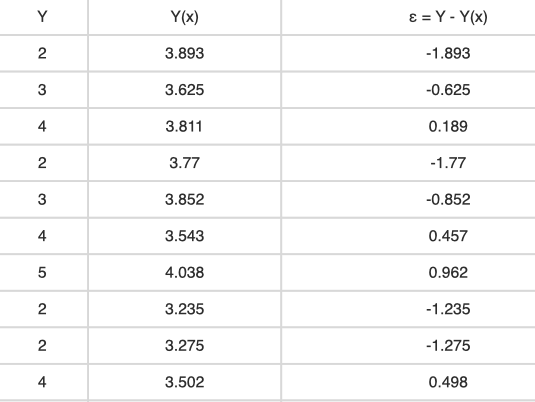
rxmy=rxmx1•β1 + rxmx2•β2 + ... + βm

Для наших данных (берем из матрицы парных коэффициентов корреляции):   
0.244 = β1 + 0.0203β2  β1 = 0.243;   
0.0441 = 0.0203β1 + β2  β2 = 0.0392;

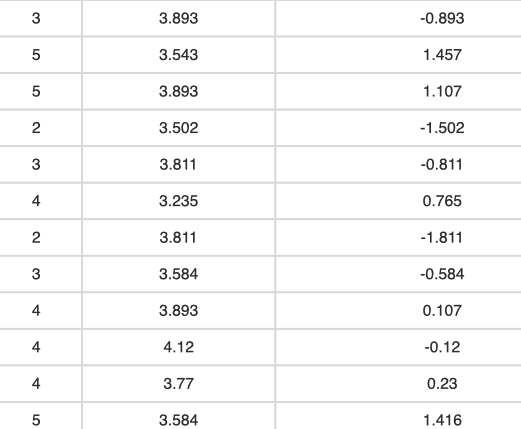
**Анализ параметров уравнения регрессии**

Для несмещенной оценки дисперсии проделаем следующие вычисления:

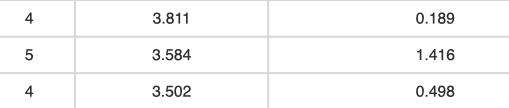
Несмещенная ошибка ε = Y - Y(x) = Y - X\*s (абсолютная ошибка аппроксимации)











∑E^2 = 53,053

1. **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:

Значимость уровня регрессии:

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.

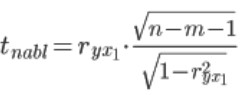
Табличное значение при степенях свободы :

( 2 ; 47 ) = 3.15

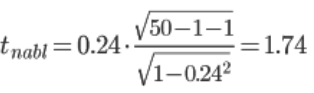
Поскольку фактическое значение F < Fkp, то коэффициент детерминации статистически не значим и уравнение регрессии статистически ненадежно (совместная незначимость коэффициентов при факторах xi подтверждается).

4) **Проверка значимости частных коэффициентов уравнения регрессий**

Рассчитаем наблюдаемые значения t-статистики для ryx1 по формуле:



где m = 1 - количество факторов в уравнении регрессии



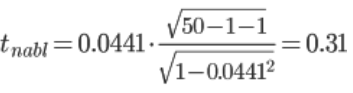
По таблице Стьюдента находим Tтабл

tкрит(n-m-1;α/2) = (48;0.025) = 2.009

Поскольку tнабл < tкрит, то принимаем гипотезу о равенстве 0 коэффициента корреляции.

Коэффициент корреляции статистически - не значим

Рассчитаем наблюдаемые значения t-статистики для ryx2 по формуле:



Поскольку tнабл < tкрит, то принимаем гипотезу о равенстве 0 коэффициента корреляции.

Коэффициент корреляции статистически - не значим

1. **Вывод**

В результате расчетов было получено уравнение множественной регрессии:

Y = 2.6171 + 0.2679X1 + 0.04075X2

Возможна экономическая интерпретация параметров модели:

увеличение X1 на 1 ед.изм. приводит к увеличению Y в среднем на 0.268 ед.изм.;

увеличение X2 на 1 ед.изм. приводит к увеличению Y в среднем на 0.0407 ед.изм.

По максимальному коэффициенту β1=0.243 делаем вывод, что наибольшее влияние на результат Y оказывает фактор X1. Установлено также, что параметры модели статистически не значимы.