Министерство науки и высшего образования Российской Федерации

**КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ**

**Отчет   
по лабораторной работе №3  
дисциплины «Статистика»  
Множественная регрессия**

Выполнил: Соловьёв Л.А.

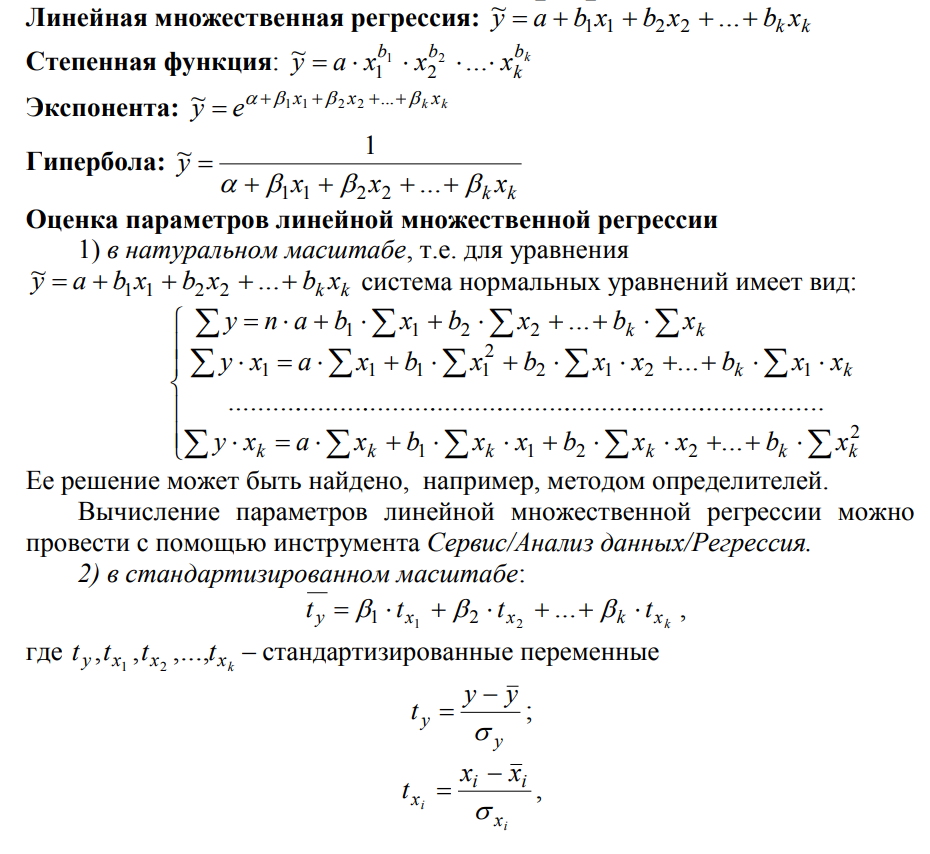
Группа: ПИ-1-22

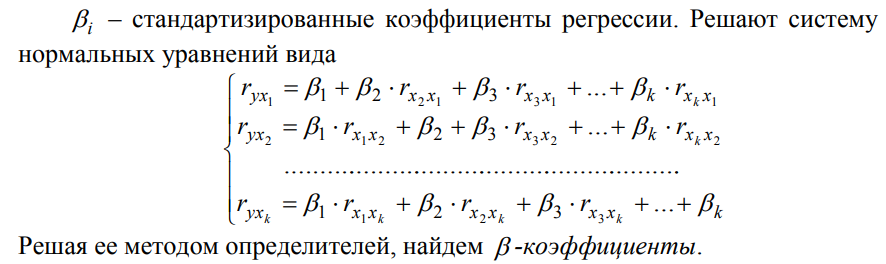
Проверил: Будникова И.К.

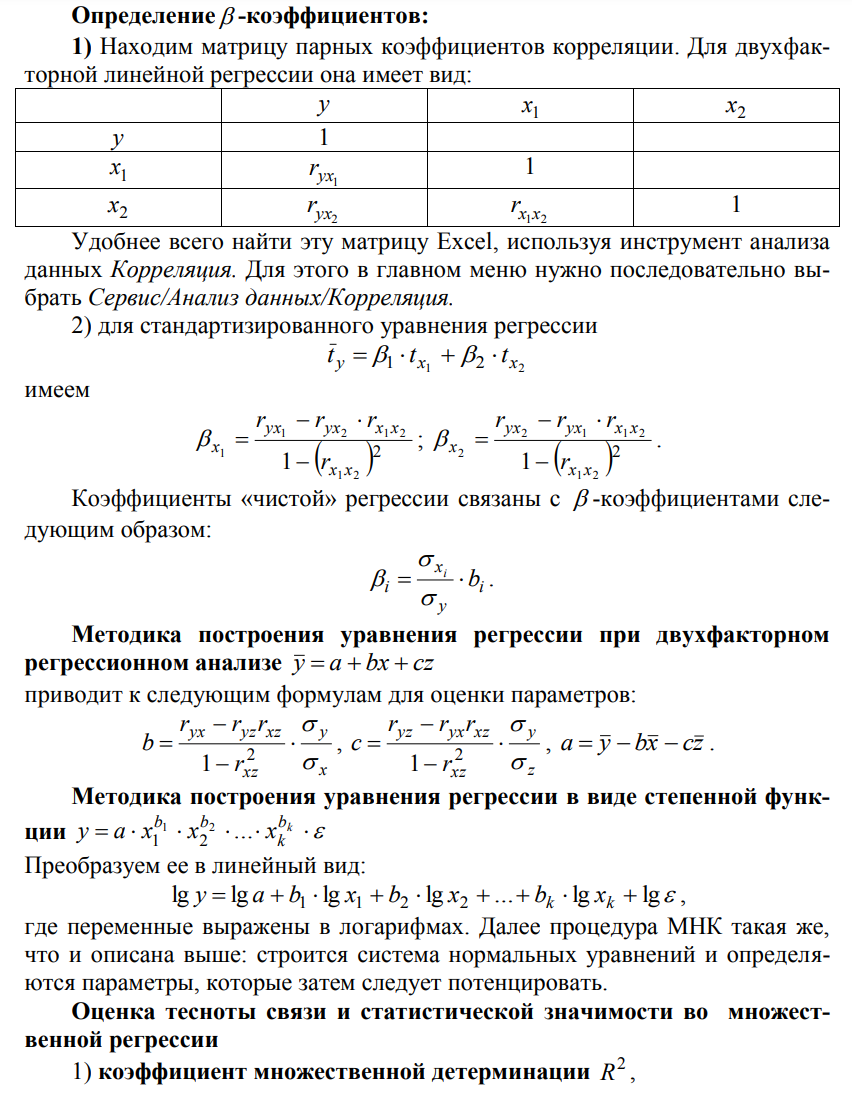
Казань 2024

**Лабораторная работа № 3**

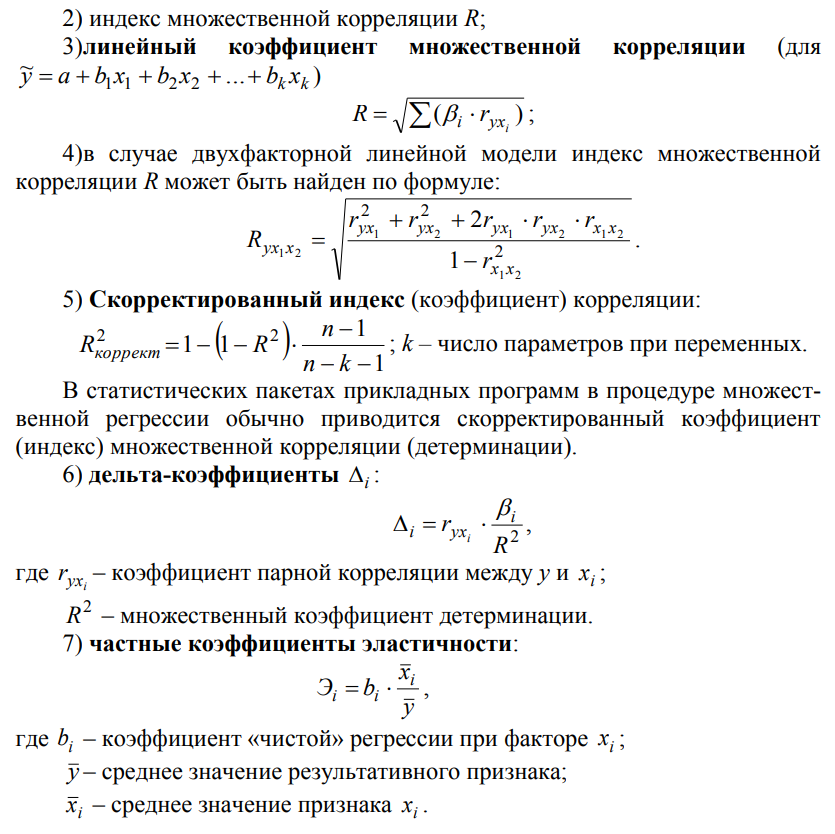
**Множественная регрессия**

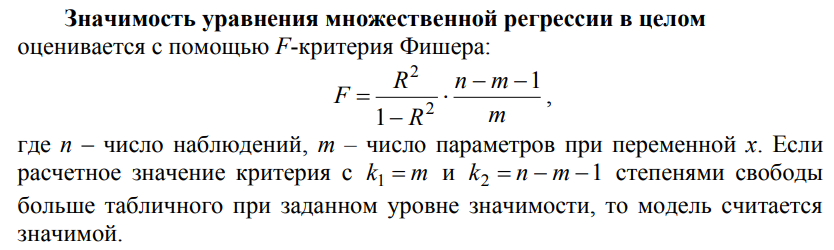
**Краткая теория  
**

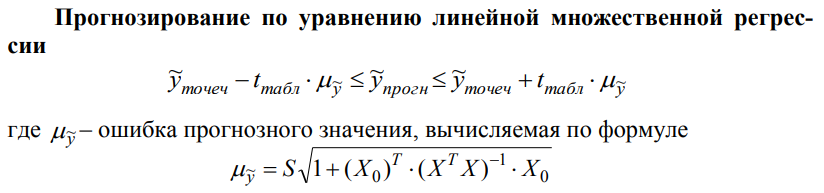
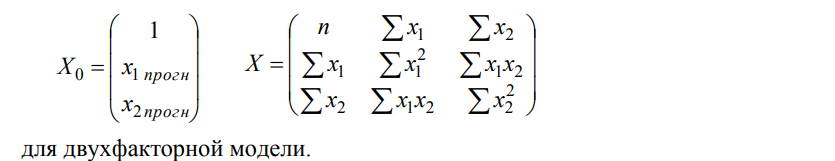
****

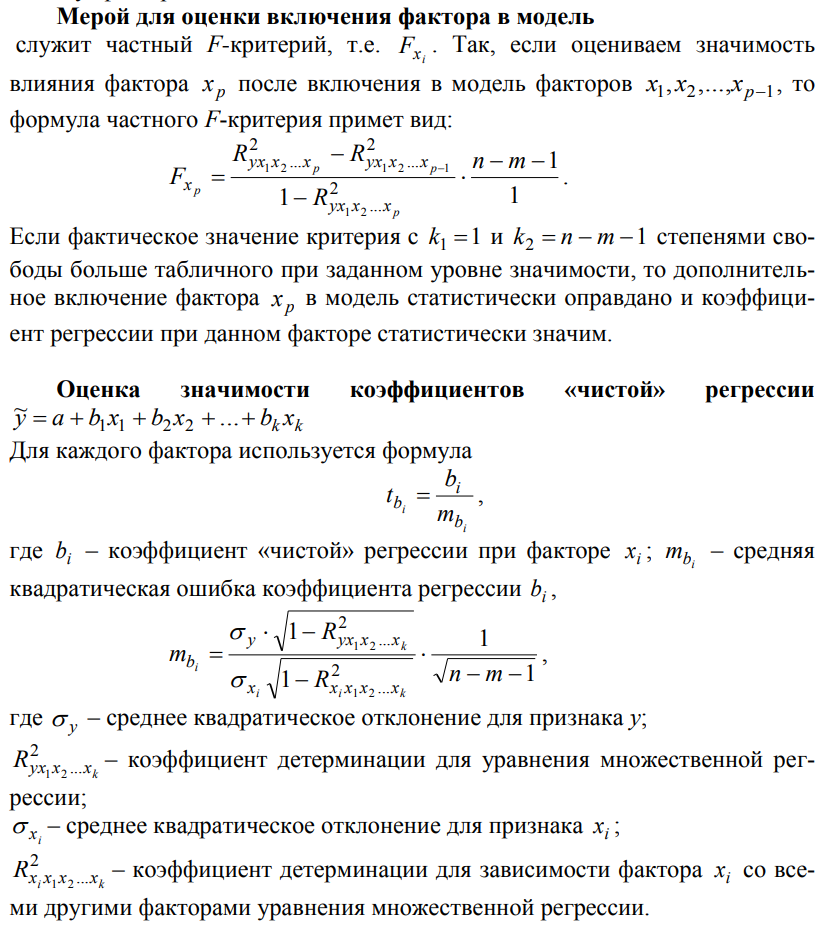
****

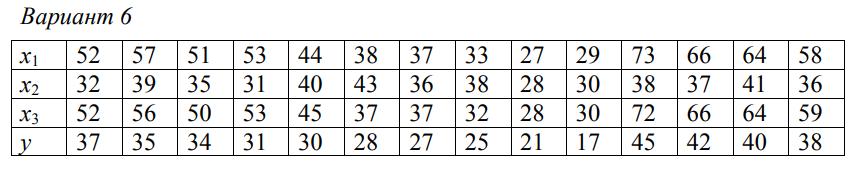
****

****

****

**  
**

****

Исследуется зависимость производительности труда y (т/ч) от уровня механизации работ X1(%), среднего возраста работников X2(лет) и энерговооруженности X3(кВт/100 работающих) по данным 14 промышленных предприятий.  
Исходные значения:  


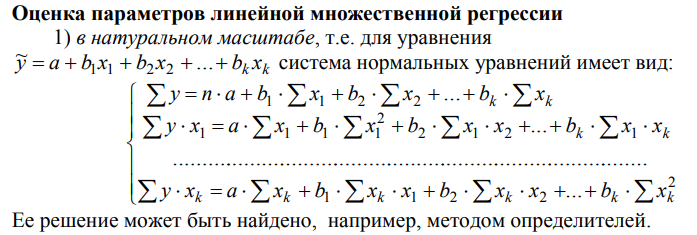
Задание на работу:

1. Рассчитать параметры линейного уравнения множественной регрессии с полным перечнем факторов.   
2. Оценить значимость уравнения в целом, используя значение множественного коэффициента корреляции и общего F-критерия Фишера.   
3. Оценить статистическую значимость параметров регрессионной модели с помощью t-критерия.   
4. Исследовать коллинеарность между факторами. При наличии мультиколлинеарности исключить какой-либо фактор из уравнения регрессии.

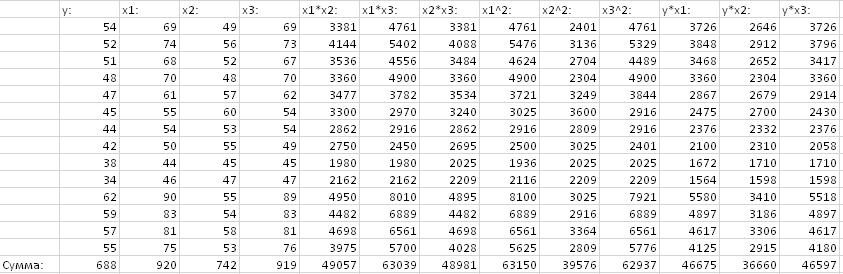
5. Построить новое уравнение множественной регрессии, провести все необходимые исследования, аналогичные проведенным выше.   
6. На основании результатов п. 5 найти

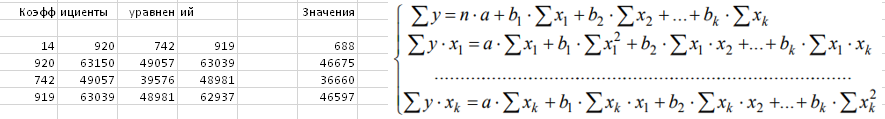
а) средние коэффициенты эластичности фактора y от независимых факторов;   
б) прогнозное значение результата при значении важнейшей объясняющей переменной, равном максимальному наблюденному значению, увеличенному на 10 %, и при значении второй объясняющей переменной, равном минимальному наблюденному значению, уменьшенному на 15%.   
в) Интервальное предсказание значения y с надежностью 0,95.

**1. Рассчитать параметры линейного уравнения множественной регрессии с полным перечнем факторов.**



Для решения данной системы составим вспомогательную таблицу вида:



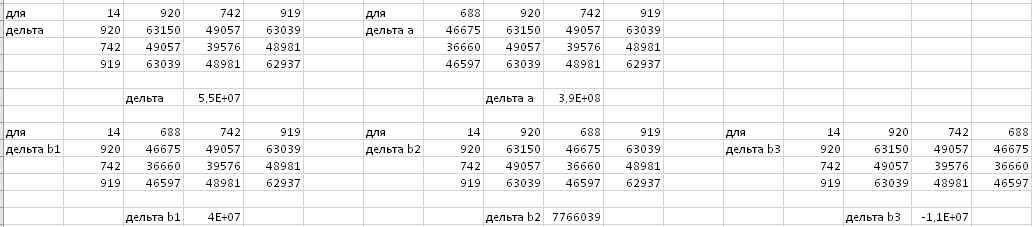
  
Составим систему уравнений:

14\*a + 920\*b1 + 742\*b2 + 919\*b3 = 688

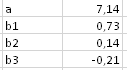
920\*a + 63150\*b1 + 49057\*b2 + 63039\*b3 = 46675

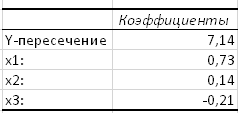
742\*a + 49057\*b1 + 39576\*b2 + 48981\*b3 = 36660

919\*a + 63039\*b1 + 48981\*b2 + 62937\*b3 = 46597

Решая систему уравнений методом определителей, составим и решим 1 основной и 4 вспомогательных определителя.  
****

Получив значения определителей вычислим значения коэффициентов уравнения регрессии по общей формуле **коэффициент = вспомогательный определитель / основной определитель**(, , ).

****

Результаты регрессионого анализа встроенного в Excel, эти значения подтвердились:  


Уравнение множественной регрессии примет вид: **~y = 0.73\*x1 + 0.14x2 – 0.21x3 + 7.14**

**2. Оценить значимость уравнения в целом, используя значение множественного коэффициента корреляции и общего F-критерия Фишера.**

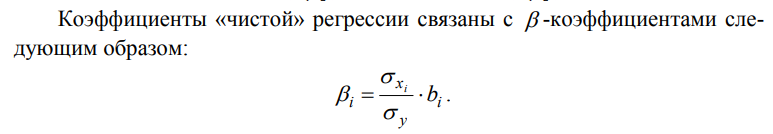
Уравнение множественной регрессии: **~y = 0.73\*x1 + 0.14x2 – 0.21x3 + 7.14**

Найдем среднеквадратические отклонения всех нужных значений:  


σx1 = sqrt(4510,71 - 65,71\* 65,71) = 13.87

σx2 = sqrt(2826,86 – 53\*53) = 4.23

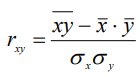
σx3 = sqrt(4496,5 – 65,64\*65,64) = 13.66

σy = sqrt(2475.57 – 49.14\*49.14) = 7.78  
****

β1 = 13.87 \* 0.73 / 7.78 = 1.3

β2 = 4.23 \* 0.14 / 7.78 = 0.08

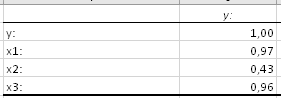
β3 = 13.66 \* -0.21 / 7.78 = -0.36

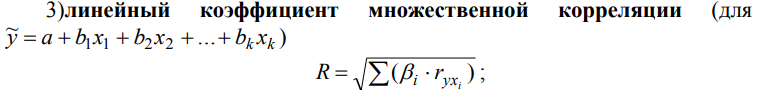
****

ryx1 = (3333.93 - 49.14 \* 65.71) / (7.78 \* 13.87) = 0.97

ryx2 = (2618.57 - 49.14 \* 53) / (7.78 \* 4.23) = 0.43

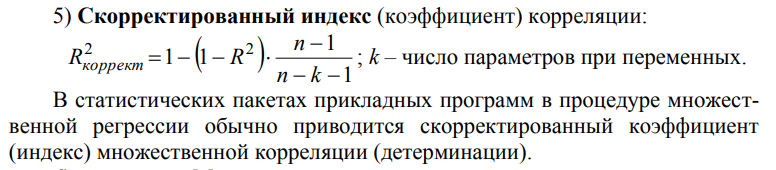
ryx3 = (3328.36 - 49.14 \* 65.64) / (7.78 \* 13.66) = 0.96

Результаты корреляционного анализа встроенного в Excel, эти значения подтвердились:  


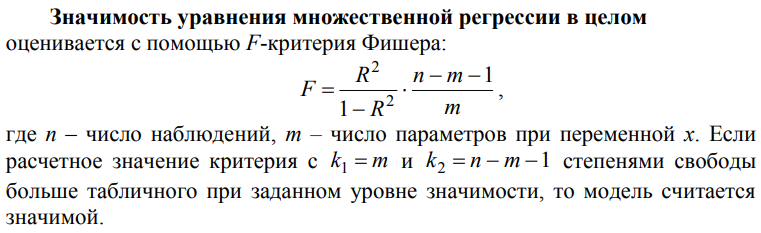
****

R = sqrt(β1\* ryx1 + β2\* ryx2 + β3\* ryx3) = sqrt(1.3 \* 0.97 + 0.08 \* 0.43 - 0.36 \* 0.96) = 0.97

R2 = 0.97 \* 0.97 = 0.95



Норм. R = 1 – (1 – 0.95) \* ((14 - 1) / (14 - 3 - 1)) = 0.93

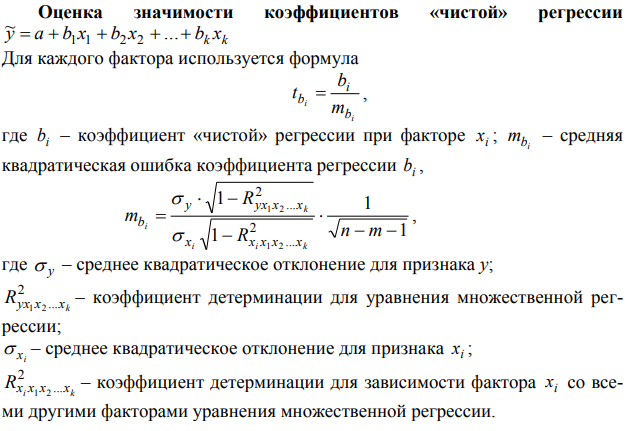


Fрасч. = 0.95 / (1 – 0.95) \* ((14 – 3 - 1) / 3) = 57.47

Расчётное значение критерия Фишера равно 57.47, оно превышает соответствующее табличное значение 3.71(FРАСПОБР(0.05, 3**(3 фактора)**, 10**(14 – 1 - 1)**). Найденное уравнение множественной регрессии статистически надежно.

Нормированный коэффициент корреляции равен 0.93, т.е. построенное уравнение объясняет 93% всей вариации признака y.

**3. Оценить статистическую значимость параметров регрессионной модели с помощью t-критерия.**

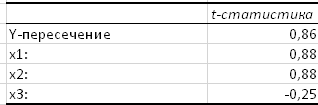


σx1 = 13.87

σx2 = 4.23

σx3 =13.66

σy = 7.78  
R2 = 0.97 \* 0.97 = 0.95

****

|  |  |  |  |
| --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** |
| **tрасч.** | **0.88** | **0.88** | **-0.25** |
| **tкрит.** | **2.23** | | |

Таким образом, ни один из факторов не имеет статистически значимого коэффициента регрессии, и построенное уравнение для прогнозирования непригодно.

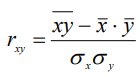
**4. Исследовать коллинеарность между факторами. При наличии мультиколлинеарности исключить какой-либо фактор из уравнения регрессии.**

σx1 = 13.87

σx2 = 4.23

σx3 =13.66

σy = 7.78

****

rx1x1 = 1

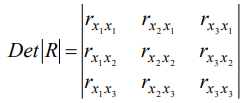
rx1x2 =rx2x1 = 3504.07 – 65.71 \*53 / (13.87 \* 4.23) = 0.36

rx1x3 =rx3x1 = 4502.79 – 65.71 \*65.64 / (13.87 \* 13.66) = 0.998

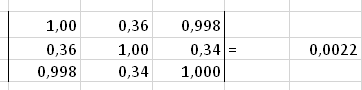
rx2x2 = 1

rx2x3 =rx3x2  = 3498.64 – 53 \* 65.64 / (4.23 \* 13.66) = 0.34

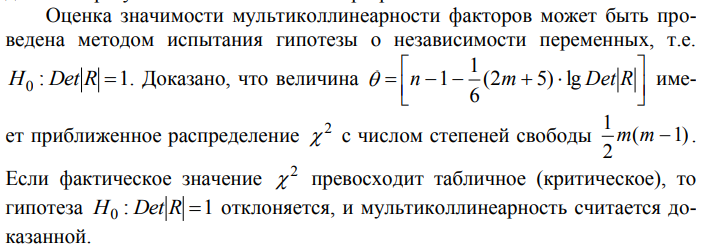
rx3x3 = 1



Составим и вычислим определитель матрицы парных коэффициентов корреляции факторов:

****

Поскольку определитель матрицы межфакторной корреляции близок к нулю, имеем мультиколлинеарность факторов и вытекающую отсюда ненадежность результатов множественной регрессии.



θ = 14 – 1 – (1/6)\*(2\*3 + 5)\* lg(0.002) = 17.86

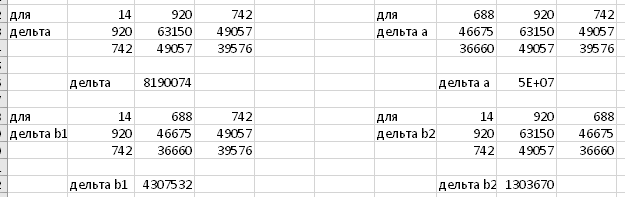
**χ**2крит. = ХИ2ОБР(0.05, (1/2) \* 3 \* (3 - 1)) = 7.81  
Т.к. **χ**2крит.< θ, то мультиколлинеарностьтю пренебречь нельзя.  
Особенно высока коллинеарность факторов x1 и x3, rx1x3 = 0.998. Один из этих факторов следует исключить из уравнения регрессии. Логично исключить тот, который имеет меньший коэффициент парной корреляции. Поскольку ryx1 = 0.97, а ryx3 = 0.96, исключаем фактор x3.  
**5. Построить новое уравнение множественной регрессии, провести все необходимые исследования, аналогичные проведенным выше.**

Составим новое уравнение множественной регрессии на основании x1 и x2 .

Составим систему линейных уравнений:

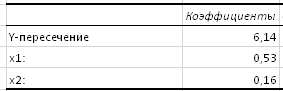
14\*a + 920\*b1 + 742\*b2 + = 688

920\*a + 63150\*b1 + 49057\*b2 = 46675

742\*a + 49057\*b1 + 39576\*b2 = 36660  
Решая систему методом определителей, составим и решим 1 основной и 3 вспомогательных определителей:  
****

Получив значения определителей вычислим значения коэффициентов уравнения регрессии по общей формуле **коэффициент = вспомогательный определитель / основной определитель**(, , ).

****

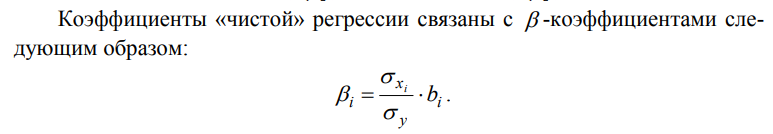
Результаты регрессионого анализа встроенного в Excel, эти значения подтвердились:  


Уравнение множественной регрессии примет вид: **~y = 0.53\*x1 + 0.16x2 + 6.14**

Найдем среднеквадратические отклонения всех нужных значений:  

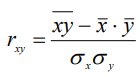

σx1 = sqrt(4510,71 - 65,71\* 65,71) = 13.87

σx2 = sqrt(2826,86 – 53\*53) = 4.23

σy = sqrt(2475.57 – 49.14\*49.14) = 7.78  
****

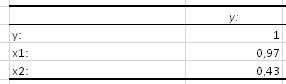
β1 = 13.87 \* 0.73 / 7.78 = 1.3

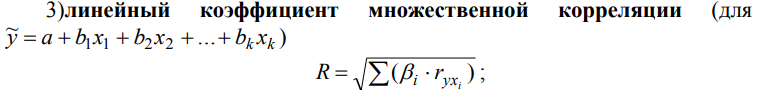
β2 = 4.23 \* 0.14 / 7.78 = 0.08

****

ryx1 = (3333.93 - 49.14 \* 65.71) / (7.78 \* 13.87) = 0.97

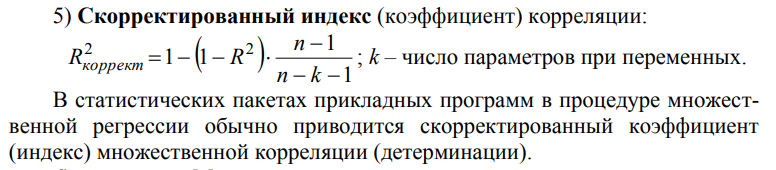
ryx2 = (2618.57 - 49.14 \* 53) / (7.78 \* 4.23) = 0.43

Результаты корреляционного анализа встроенного в Excel, эти значения подтвердились:  


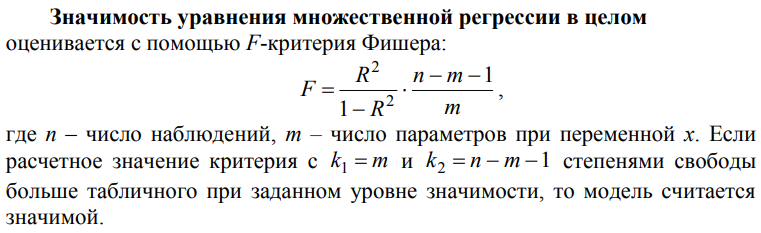
****

R = sqrt(β1\* ryx1 + β2\* ryx2) = sqrt(1.3 \* 0.97 + 0.08 \* 0.43) = 0.97

R2 = 0.97 \* 0.97 = 0.94



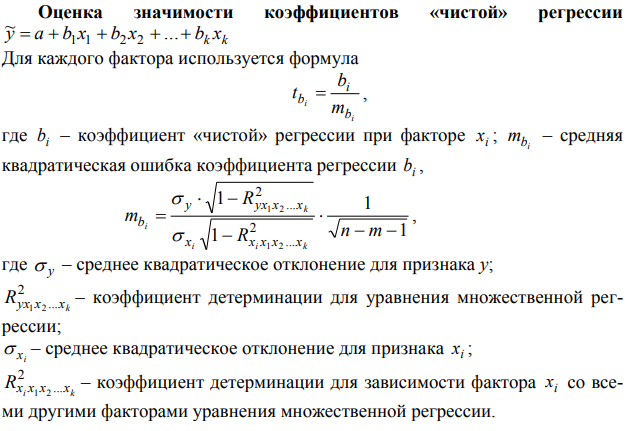
Норм. R = 1 – (1 – 0.94) \* ((14 - 1) / (14 - 2 - 1)) = 0.93



Fрасч. = 0.94 / (1 – 0.94) \* ((14 – 2 - 1) / 2) = 94.22

Расчётное значение критерия Фишера равно 94.22, оно превышает соответствующее табличное значение 3.71(FРАСПОБР(0.05, 2**(2 фактора)**, 11**(14 – 2 - 1)**). Найденное уравнение множественной регрессии статистически надежно.

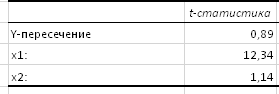
Нормированный коэффициент корреляции равен 0.93, т.е. построенное уравнение объясняет 93% всей вариации признака y.



σx1 = 13.87

σx2 = 4.23

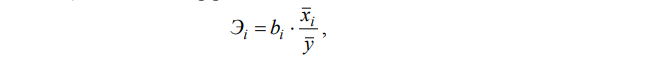
σy = 7.78  
R2 = 0.97 \* 0.97 = 0.94

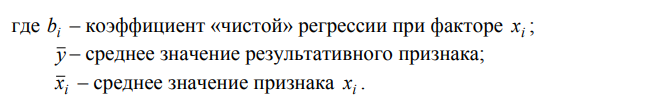
****

|  |  |  |
| --- | --- | --- |
|  | **x1** | **x2** |
| **tрасч.** | **12.34** | **1.14** |
| **tкрит.** | **2.20** | |

Таким образом, при весьма удовлетворительной значимости уравнения регрессии в целом, мы добились значимости коэффициента регрессии при переменной x1. **6. На основании результатов п. 5 найти**

**а) средние коэффициенты эластичности фактора y от независимых факторов;**

****

****

Э1 = 0.53 \* 65.71 / 49.14 = 0.7

Э2 = 0.16 \* 53 / 49.14 = 0.17

Таким образом, при изменении фактора x2 (среднего возраста работников) на 1%, производительность возрастает незначительно, на 0,17%; при изменении фактора x1 (механизации работ) на 1%, производительность труда увеличивается на 0,7%. **б) прогнозное значение результата при значении важнейшей объясняющей переменной, равном максимальному наблюденному значению, увеличенному на 10 %, и при значении второй объясняющей переменной, равном минимальному наблюденному значению, уменьшенному на 15%.**

Максимальное наблюденное значение фактора x1 = 90  
Минимальное наблюденное значение фактора x2 = 45

Прогнозные значения факторов:

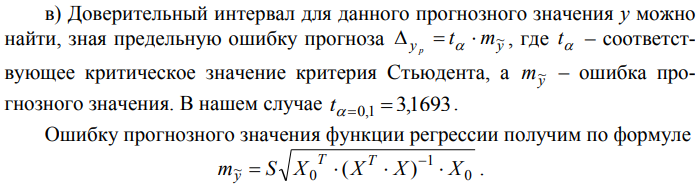
x1прогн. = 90 \* 1.1 = 99

x2прогн. = 45 \* 0.85 = 38.25

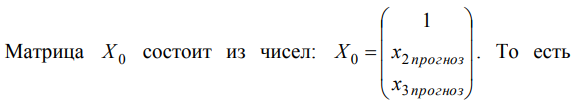
Тогда ~yпрогн. = 0.53\*99 + 0.16\*38.25 + 6.14 = 52.37 + 6.12 + 6.14 = 64.3

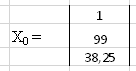


**в) Интервальное предсказание значения y с надежностью 0,95.**

****Шаг 1.

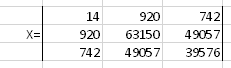
  
S = 2.06

Шаг 2.  
****





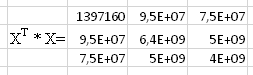
Шаг 3.



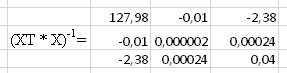
Шаг 4.

XT = X

Шаг 5.



Шаг 6.



Шаг 7.



Шаг 8.



Шаг 9.

m~y = 2,06 \* sqrt(9,86) = 6.47

Шаг 10.

Deltayпрогн. = 20,52

Шаг 11.

Таким образом, прогнозное значение результата будет с вероятностью 95% находиться в интервале 64,3± 20,52.