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**

***Цель работы:***

1. Определить коэффициенты уравнения.

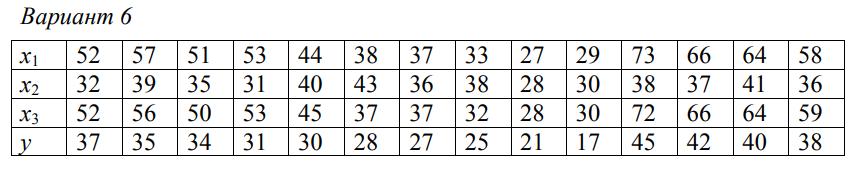
2. Выполнить статистический анализ уравнения регрессии.

**Задание на работу:**

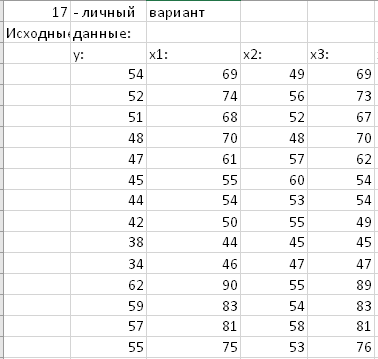
1. Рассчитать параметры линейного уравнения множественной регрессии с полным перечнем факторов.   
2. Оценить значимость уравнения в целом, используя значение множественного коэффициента корреляции и общего F-критерия Фишера.   
3. Оценить статистическую значимость параметров регрессионной модели с помощью 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.

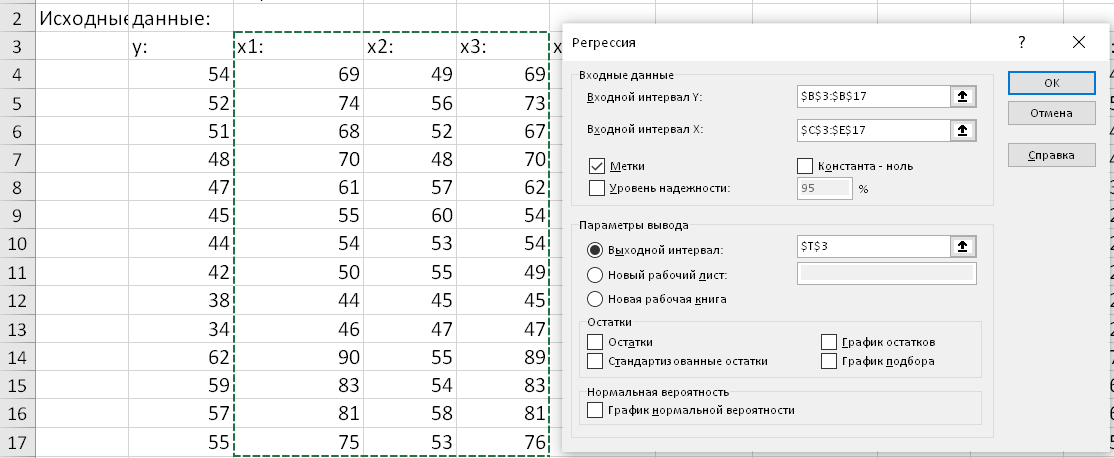
Исходные значения:  


Исходные значения для варианта 17:

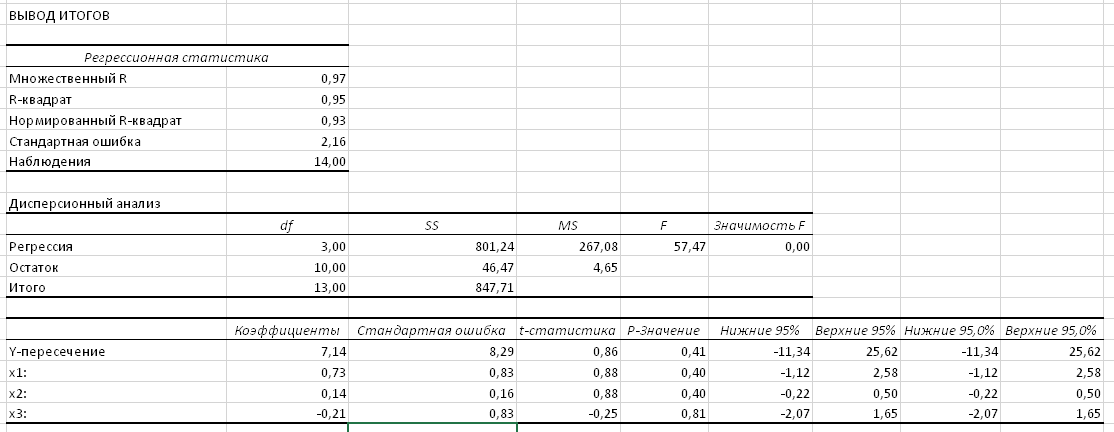


**1. Рассчитать параметры линейного уравнения множественной регрессии с полным перечнем факторов.**

Пользуясь инструментом Excel Анализ данных/Регрессия получили параметры множественной регрессии.



Результат регрессионного анализа имеют вид:



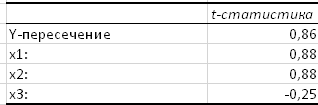
**2. Оценить значимость уравнения в целом, используя значение множественного коэффициента корреляции и общего F-критерия Фишера.**

Уравнение множественной регрессии в соответствии с результатами регрессионного анализа: **~y = 0.73\*x1 + 0.14x2 – 0.21x3 + 7.14** является статистически надежным. Действительно, Fрасч. = 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)**).

О доле вариации результативного признака y, объясненной построенным уравнением множественной регрессии лучше всего судить по значению нормированного коэффициента корреляции, в данном случае он равен 0,93. То есть построенное уравнение объясняет 93% всей вариации признака y.

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

Пользуясь инструментом Анализ данных/Регрессия получим t-статистику каждого из факторов регрессии.

****

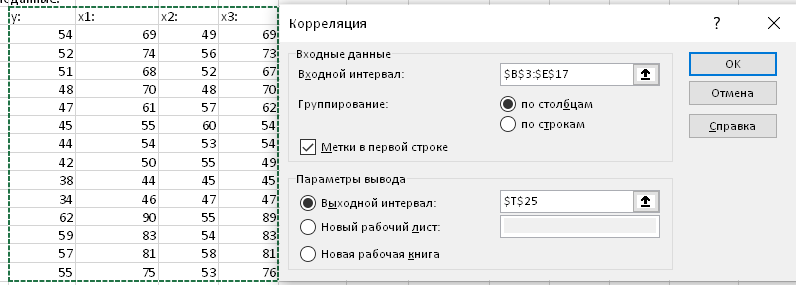
Чтобы оценить статистическую значимость параметров регрессионной модели с помощью t-критерия, найдем соответствующее нашим параметрам критическое значение с помощью функции *СТЬЮДРАСПОБР* при заданном уровне значимости 0,05 и числе степеней свободы n – m – 1. Коэффициент признается значимым, если выполняется неравенство tрасч > tкрит.

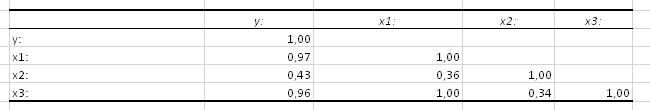
|  |  |  |  |
| --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** |
| **tрасч.** | **0.88** | **0.88** | **-0.25** |
| **tкрит.** | **2.23** | | |

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.**

Получим матрицу парных коэффициентов корреляции с помощью инструмента Анализ данных/Корреляция.



Результат корреляционного анализа имеет вид:  


Для оценки мультиколлинеарности факторов вычислим определитель матрица парных коэффициентов корреляции факторов:

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

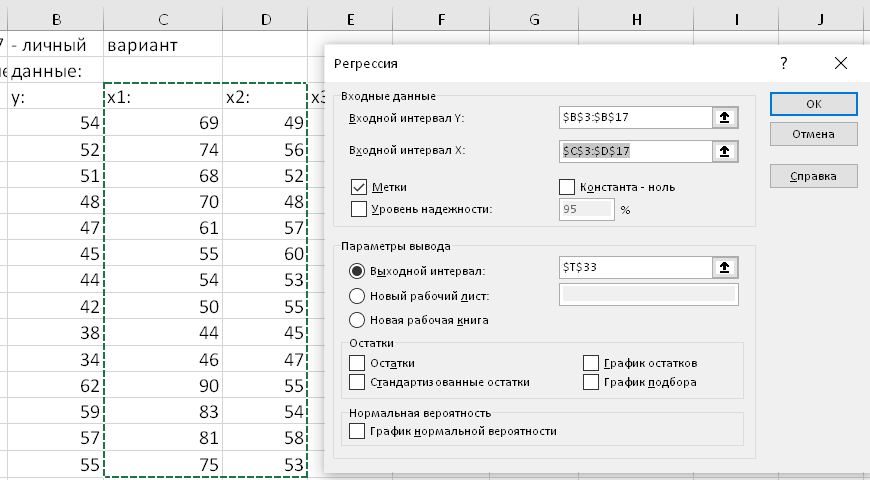
Оценка значимости мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных, т.е. Доказано, что величина имеет приближенное распределение с числом степеней свободы Если фактическое значение превосходит табличное (критическое), то гипотеза отклоняется, и мультиколлинеарность считается доказанной.

Критическое значение можно найти через статистическую функцию ХИ2ОБР(a; n – 1), где a – уровень значимости (по условию 0,05), а n – число степеней свободы. В нашем случае степеней свободы Получаем . Мультиколлинеарностью факторов пренебречь нельзя.

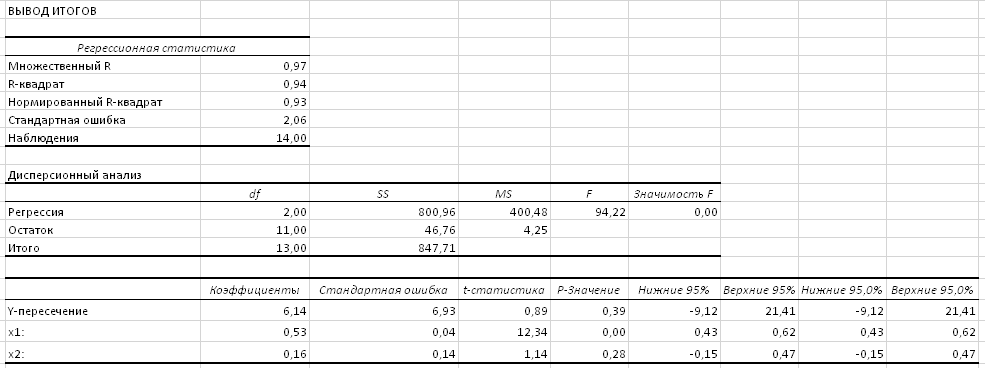
Особенно высока коллинеарность факторов x1 и x3, rx1x3 = 0.998. Один из этих факторов следует исключить из уравнения регрессии. Логично исключить тот, который имеет меньший коэффициент парной корреляции. Поскольку ryx1 = 0.97, а ryx3 = 0.96, исключаем фактор x3.

**5. Построить новое уравнение множественной регрессии, провести все необходимые исследования, аналогичные проведенным выше.**

Пользуясь инструментом Excel Анализ данных/Регрессия получили параметры множественной регрессии на факторах x1 и x2.



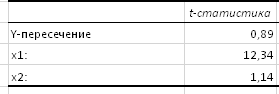
Результат регрессионного анализа имеют вид:



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

Rнорм.= 0.93, Fрасч. = 94.22, что много больше чем Fкрит.= 3.71.

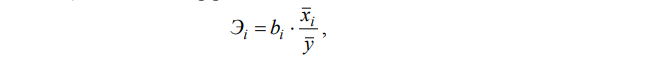
Пользуясь инструментом Анализ данных/Регрессия получим t-статистику каждого из факторов регрессии.

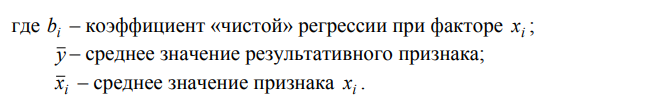
****

|  |  |  |
| --- | --- | --- |
|  | **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;**

****

****

|  |  |  |  |
| --- | --- | --- | --- |
|  | y | x1 | x2 |
| Среднее | 49.14 | 65.71 | 53 |
| Коэф. эластичности |  |  |  |

Таким образом, при изменении фактора x1 (механизации работ) на 1%, производительность труда увеличивается на 0,7%; при изменении фактора x2 (среднего возраста работников) на 1%, производительность возрастает незначительно, на 0,17%;  **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.**

в) Доверительный интервал для данного прогнозного значения y можно найти, зная предельную ошибку прогноза , где – соответствующее критическое значение критерия Стьюдента, – ошибка прогнозного значения. В нашем случае .

.

Шаг 1. Параметр S – стандартная ошибка регрессии приведен в последней регрессионной статистике S = 2,06.

Шаг 2. Матрица состоит из чисел: То есть

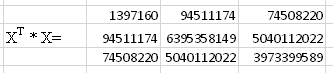
*;*

*.*

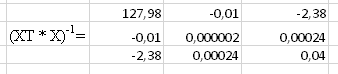
Шаг 3. Матрица X состоит из чисел .

Шаг 4.Транспонируем матрицу X. Поскольку она симметрическая, то .

Шаг 5. Найдем произведение матриц . В Exсel это можно сделать с помощью функции **МУМНОЖ**.



Шаг 6. Найдем обратную матрицу к матрице произведения . В Exсel это можно сделать с помощью функции МОБР.



Шаг 7. Найдем произведение матриц (размерность матрицы произведения 1\*3).



Шаг 8. Найдем произведение матриц (размерность матрицы произведения 1\*1, то есть только одно число).



Шаг 9.

Шаг 10.

Шаг 11. Таким образом, прогнозное значение результата будет с вероятностью 95% находиться в интервале

Вывод: в ходе выполнения лаб. работы мы научились определять коэффициенты множественной регрессии и проводить стат. анализ полученных уравнений множественной регрессии.