МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

 «МАГНИТОГОРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

ИМ. Г. И. НОСОВА»

(ФГБОУ ВО «МГТУ ИМ. Г.И. НОСОВА»)

Кафедра вычислительной техники и программирования

**Отчет по лабораторной работе №6**

по дисциплине «Обработка экспериментальных данных»

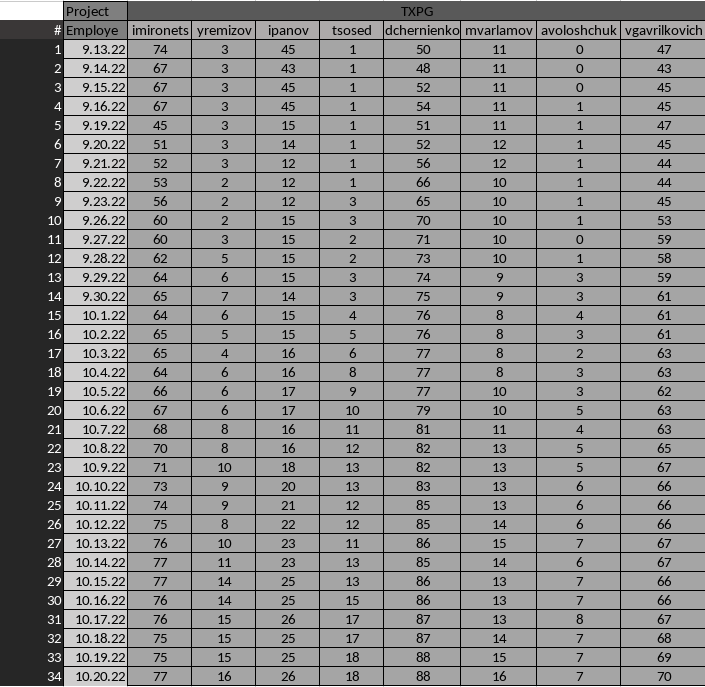
Исполнитель: Варламов М.Н., студент 4 курса, группа АВб–19–1

Руководитель: Ильина Е.А., к.п.н., доцент кафедры ВТиП.

Магнитогорск, 2022

### Построение проверки интеркорреляции и мультиколлинеарности

На рисунке 1 представлены исходные данные.



*Рисунок 1 – Исходные данные*

На основе предоставленных данных построим матрицу корреляции.

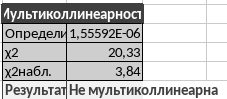


*Рисунок 2 – Матрица корреляции исходных данных*

Для расчета мультиколлинеарности матрицы необходимо вычислить определитель матрицы корреляции парных коэффициентов. В случае, если определитель матрицы равен нулю, то матрица является мультиколлинеарной, что говорит о том, что оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов будет менее надежен. Если определитель матрицы равняется единице, то матрица парных корреляций является не мультиколлинеарной.

Если же определитель матрицы отличен от ноль и единицы, то необходимо выдвинуть теорию о том, что выбранные факторы немультиколлинеарны. Для этого по критерию Пирсона вычисляется коэффициент «хи квадрат». По таблице необходимо определить критическое значение и сравнить полученные коэффициенты. В случае, если рассчитанный коэффициент больше табличного, то матрица корреляции является не мультиколлинеарной.

На рисунке 3 представлен расчёт мультиколлинеарности матрицы корреляции исходных данных.



*Рисунок 3 – Расчёт мультиколлинеарности матрицы*

Построение множества информативных и неинформативных факторов

Для построения множества информативных и неинформативных факторов необходимо в матрице межфакторной корреляции выделить максимальный по модулю коэффициент на области ниже главной диагонали. Если коэффициент является незначимым, то все факторы являются информативными и записываются в список I.

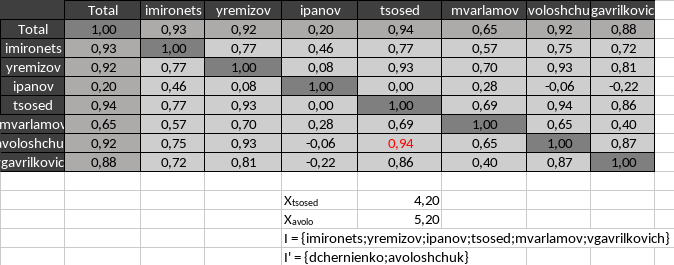
Если же коэффициент является значимым, то необходимо рассчитать сумму коэффициентов в столбце и строке, на которых находится коэффициент. Наибольшее из полученных значений является неинформативным фактором т.к. оказывает большое влияние на оставшиеся факторы. Такой фактор запишем в список . Строку и столбец с неинформативным фактором необходимо убрать из общей матрицы.

Данный алгоритм необходимо повторять до тех пор, пока не сформируются списки информативных и неинформативных факторов.

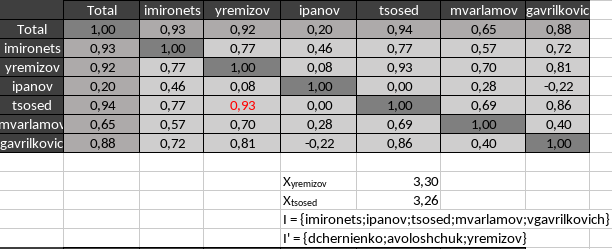
На рисунках 4-6 изображен процесс формирования списка информативных и неинформативных факторов.



*Рисунок 4 – Формирование списка информативных и неинформативных факторов (1 шаг)*



*Рисунок 5 – Формирование списка информативных и неинформативных факторов (2 шаг)*



*Рисунок 6 – Формирование списка информативных и неинформативных факторов (3 шаг)*

В результате мы получили матрицу, состоящую только из значимых факторов (рис 7).



*Рисунок 7 – Матрица информативных факторов*

Список значимых и незначимых факторов представлен на рисунке 8.

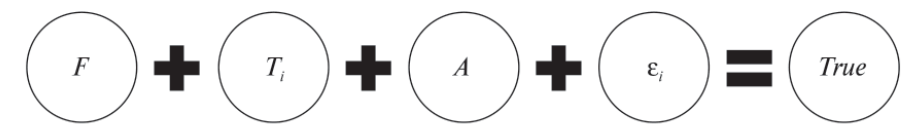
о



*Рисунок 8 – Список информативных и неинформативных факторов*

Выполнить построение линейной и мультипликативной моделей

Для оценки применимости построенной эмпирической модели для последующего прогнозирования и управления используют три статистических показателя и результаты анализа остатков. В зависимости от вида уравнения набор показателей изменяется согласно схеме, приведенной на рисунке 5. Положительное решение о применимости эмпирической модели – при истинности всех требований к этим показателям (рисунок 9).

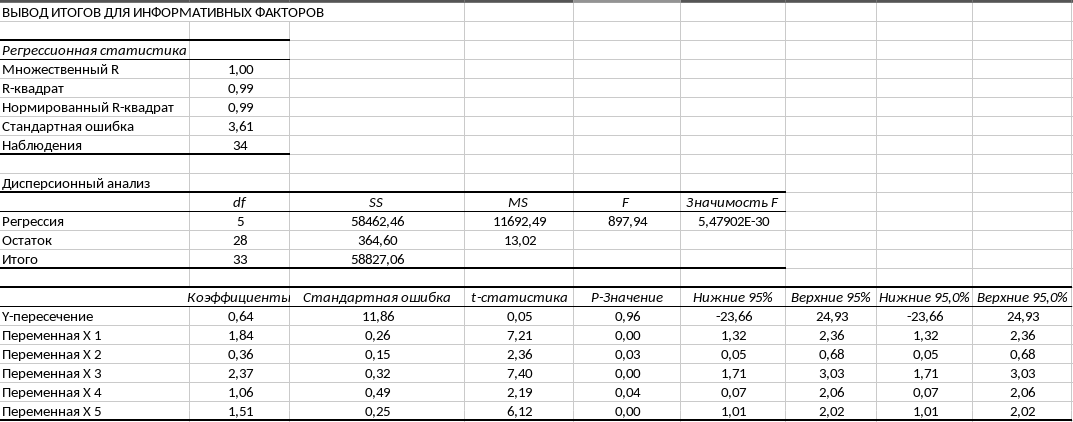


*Рисунок 9 – Схема принятия положительного решения о применимости эмпирической модели*

На рисунке 9 приняты некоторые обозначения:

* F – Статистика Фишера
* Ti – Статистика Стьюдента
* A – Относительная ошибка апроксимации
* Ei – Остатки

Для расчета необходимых статистик воспользуемся регрессионным анализом. С помощью статистического пакета excel выполним данный анализ для информативных факторов. На рисунке 10 представлен результат регрессионного анализа.



*Рисунок 9 – Регрессионный анализ информативных факторов*

С помощью рассчитанного анализа составим линейное уравнение со всеми значимыми факторами.

Критерий Фишера равен 0,65. Уравнение регрессии является ненадёжным, так как уровень значимости больше 5%.

Значение статистики Стьюдента для коэффициента равно 7,21 на уровне значимости большем, чем 5%, следовательно, не значимо отличается от нуля.

Значение статистики Стьюдента для коэффициента равно 2,36 на уровне значимости большем, чем 5%, следовательно, не значимо отличается от нуля.

Значение статистики Стьюдента для коэффициента равно 7,40 на уровне значимости большем, чем 5%, следовательно, не значимо отличается от нуля.

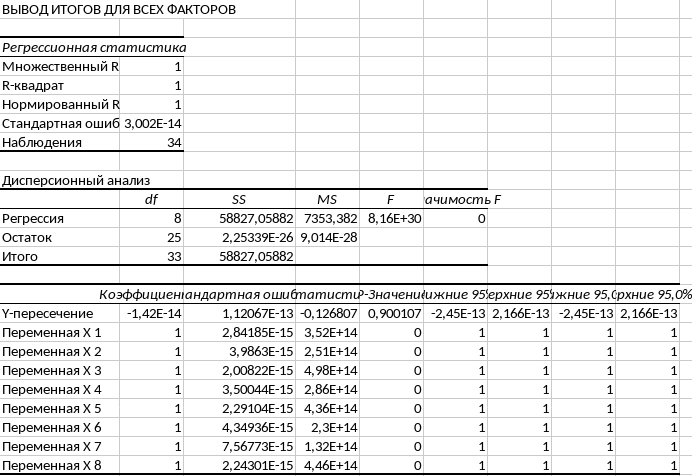
Значение статистики Стьюдента для коэффициента равно 2,19 на уровне значимости большем, чем 5%, следовательно, не значимо отличается от нуля.

Значение статистики Стьюдента для коэффициента равно 7,40 на уровне значимости большем, чем 5%, следовательно, не значимо отличается от нуля.

Средняя относительная ошибка аппроксимации для линейного уравнения со всеми множествами факторов составила 11,86%. Она превосходит 10%, следовательно, считается недопустимой.

Исходя из расчетов трёх статистических показателей можно сделать вывод, что модель множественной линейной регрессии всех факторов не применима.

Построим регрессионный анализ для всех факторов (рис. 10). После построения можно наблюдать, что линейная модель для всех факторов неприменима в виду полученных коэффициентов.



*Рисунок 10 – Регрессионный анализ всех факторов*

### Оценка структурной стабильности используемых исходных данных

Значимость структурных изменений можно оценить с помощью статистического критерия Грегори Чоу. Выдвинем гипотезу о структурной стабильности двух частей ряда (т.е. об отсутствии смены тенденции) и конкурирующую гипотезу о том, что структурные изменения в двух частях ряда статистически значимы (т.е. о наличии смены тенденции).

Для проверки критерия Грегори Чоу построим уравнения линейных трендов и найдем остаточные суммы квадратов.

Остаточная сумма квадратов кусочно-линейной модели равна

где , , – теоретические значения уровней, найденные соответственно по уравнениям (I), (II) и (III).

Изменение остаточной дисперсии при переходе от единого уравнения тренда к кусочно-линейной модели определяется как разность:

Расчетное значение F-критерия:

Оно сравнивается с табличным , найденным по таблице критических точек распределения Фишера для уровня значимости и числа степеней свободы , . Если , то гипотеза о структурной стабильности отклоняется, и влияние структурных изменений на динамику изучаемого показателя считается значимым.

Таблица 1 – Расчетная таблица критерия Грегори Чоу

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер уравнения | Уравнение тренда | Длина ряда n | Остаточная сумма квадратов | Число параметров уравнения тренда |
| (I) |  | 11 | 204,54 | 2 |
| (II) |  | 11 | 268,18 | 2 |
| (III) |  | 22 | 545,45 | 2 |

Все уравнения тренда линейные, число параметров всех уравнений

Вычислим сумму:

Найдем разность:

Определяем фактическое значение F-критерия:

Определяем критическое значение:

Структурные изменения считаются значимыми, следовательно, гипотеза о стабильности коэффициентов регрессии не отвергается.