Санкт-Петербургский Национальный Исследовательский Университет Информационных технологий, механики и оптики

**Лабораторная работа №7**

**Построение и исследование моделей регрессионного анализа**

Выполнил: Долматов

Дмитрий Алексеевич

Проверила: Казанова

Полина Петровна

Санкт-Петербург

2021

**Цель работы:**

Изучить средства программы *Microsoft Excel* для регрессионного анализа данных.

**Задачи:**

1. Определить, чему равен коэффициент корреляции двух случайных величин;
2. Построить линейную модель на основе данных;
3. Определить параметры уравнения линейной регрессии.

**Ход работы:**

Существует 2 главных подразделения регрессии: линейная/нелинейная; а также в зависимости от факторов: парная (простую, при m – 1) и множественную (многофакторную, m > 1).

**Упражнение 1:**

Для исходных данных, представленных на рисунке 7.1.1, вычислим корреляцию (связь между признаками) попарно между признаками, применяю функцию *КОРРЕЛ(массив1;массив2)* и функцию *Корреляция*.

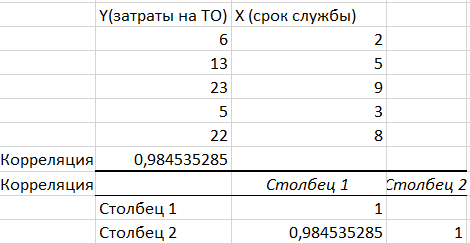


Рисунок 7.1.1 – Корреляция

Таким образом, коэффициент корреляции почти равен единице, что говорит о линейной зависимости между случайными величинами. Результаты нахождения корреляции двумя способами совпали.

**Упражнение 2:**

В данном упражнении необходимо определить коэффициент корреляции двух случайных величин, представленных на рисунке 7.2.1.

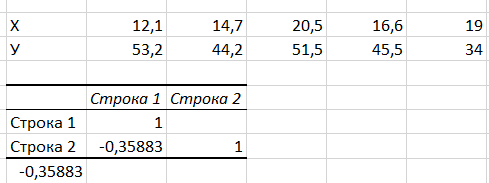


Рисунок 7.2.1 – Корреляция случайных величин

Как можно увидеть, в данном примере корреляция оказалась менее линейной, чем в прошлом примере, точнее, линейность ее прослеживается более слабо.

**Упражнение 3:**

На основе данных из рисунка 7.3.1 построим линейную модель и проведем ее анализ.



Рисунок 7.3.1 – Исходные данные

Используя функцию *Корреляция* проводим линейный анализ. Множественный *R* – коэффициенту корреляции, *R-квадрат* – коэффициенту детерминации, *нормированный R-квадрат* – нормированное значение коэффициента корреляции, *стандартная ошибка* – стандартное отклонение для остатков, а *наблюдение* – количество исходных наблюдений, *df* – число степеней свободы (количество факторных признаков), *SS* – сумма квадратов отклонений, *MS* – дисперсия, *F* – отношение факторной дисперсии к остаточной, а значимость *F* – значение уровни значимости, соответствующее вычисленному значению *F*. Результаты привидены на рисунке 7.3.2.

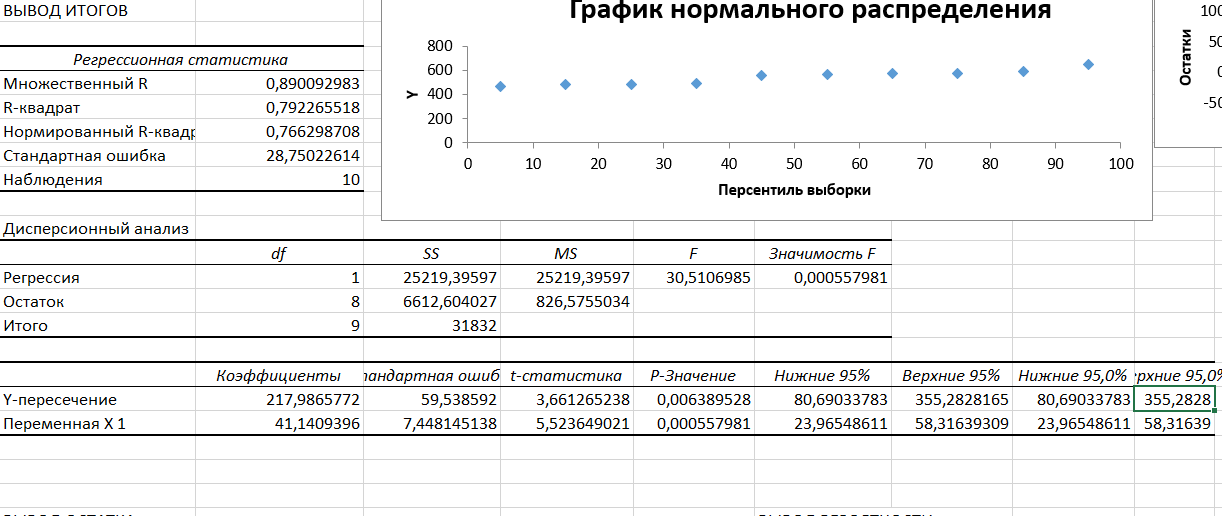


Рисунок 7.3.2 – Результаты линейного анализа

Анализируя данные результаты, придем к выводу, что модель получится следующего вида: Коэффициент *детерминации* – 0.79 и *множественный R* – 0.89, что говорит нам о линейности и прямой зависимости между количеством людей и объемом производства (что логично). Значимость F меньше F, тем самым дисперсия случайных величин признается одинаковым. Тем более, ее значение очень мало, значит модель точнее (предел 0.05 по сравнению с нашими 0.0005).

**Упражнение 4:**

Построим модель зависимости величины заработной платы от стажа работы и пола сотрудника. Исходные данные представлены на рисунке 7.4.1.

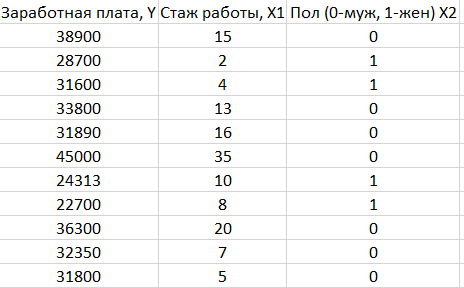


Рисунок 7.4.1 – Исходные данные

Проведем линейный анализ с помощью функции *Регрессия*. Результаты представлена на рисунке 7.4.2.

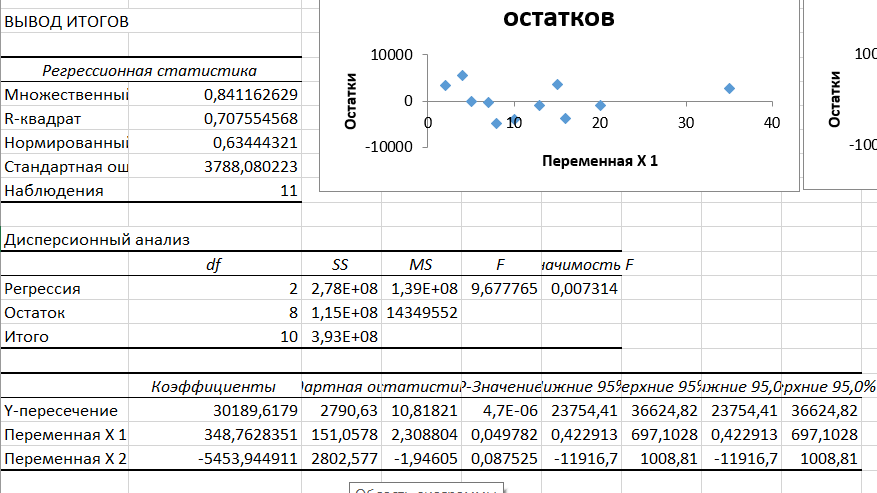


Рисунок 7.4.2 – Результаты линейного анализа

В результате мы имеем следующие результаты:

* уравнение функции регрессии имеет вид: ;
* коэффициент детерминации равен 0.7- – корреляция стажа и пола оказывают некоторое влияние на зарплату;
* значимость F – 0.007, что подтверждает значение коэффициента детерминации (меньше 0.05);
* P-значение подтверждает тот факт, что параметр, определяющий стаж работы, оказывает влияние на значение функции, а параметр, указывающий на принадлежность к полу, равный 0.0498 (что близко к 0.05), говорит о том, что данный параметр не оказывает существенного влияния на результат функции. Таким образом, гендерная дискриминация в данной задаче не прослеживается.

Перестроим нашу регрессию без учета пола сотрудников. Результаты представлены на рисунке 7.4.3.

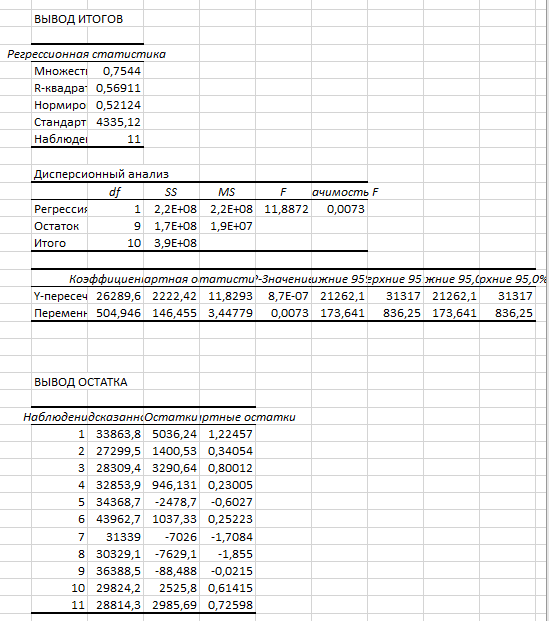


Рисунок 7.4.3 – Регрессионный анализ без учета пола

Вывод анализа:

* Уравнение функции регрессии имеет вид:
* Коэффициент детерминации – 0.569 и коэффициент линейности 0.75, что говорит о том, что прямая связь прослеживается не так сильно, как в прошлом варианте;
* Значимость F, меньшая, чем 0.05, подтверждает правильность значения коэффициента детерминации;
* P-значение подтверждает значимость стажа работы как параметра.

Таким образом, хоть и значение пола не играло важную роль, однако без учета пола линейность снизилась, что говорит о менее адекватной системе в нынешнем варианте решения (без учета).

**Упражнение 5:**

Определить по исходным данным уравнение линейной регрессии и провести его анализ. Используем функцию *Регрессия* и выводим результаты на рисунке 7.5.1.

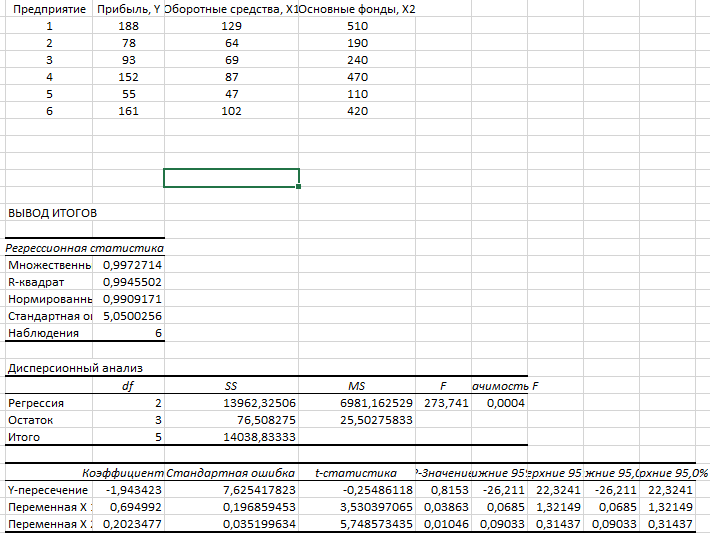


Рисунок 7.5.1 – Результаты линейного анализа

Вывод анализа:

* Уравнение регрессии имеет вид: 0.69;
* Коэффициент детерминации – 0.994 и коэффициент линейности 0.997, что говорит о том, что прибыль компании линейно зависит как от оборотных средств, так и от основного фонда;
* Значимость F, меньшая, чем 0.05, подтверждает правильность значения коэффициента детерминации;
* P-значение подтверждает значимость двух параметров.

Таким образом, влияние количества оборота и функциональности фондов оказывают линейную зависимость по отношению к прибыли предприятия.

**Вывод:**

В процессе выполнения данной лабораторной работы были изучены средства программы *Microsoft Excel* для регрессионного анализа данных. Определили, чему равен коэффициент корреляции двух случайных величин, построили линейную модель на основе исходных данных и определили параметры уравнения линейной регрессии.

**Контрольные вопросы**:

1. Регрессионная модель – это форма связи результативного признака области значения с факторами (параметрами) из области определения, которая подразделяется на линейную и нелинейную, парную и множественную;
2. Общий вид регрессионной модели представляет собой сумму произведений параметра и линейного члена (коэффициента): ;
3. Значимость коэффициента регрессии определяется с помощью нахождения *P-значения* регрессии (значение уровня значимости);
4. При определении коэффициентов регрессионной модели применяется метод наименьших квадратов (МНК) в функции *Регрессия*.