Белорусский государственный технологический университет

Кафедра «Информационных систем и технологий»

Лабораторная работа №2

**Корреляционный анализ**

Выполнил студент

3 курса 3 группы

Процукович К.М.

Проверил

Колесников В. Л.

Минск 2018

1. **Краткое описание объекта исследования**

Корреляция - это связь между двумя переменными. Корреляционный анализ – совокупность методов обнаружения корреляционной зависимости между случайными величинами или признаками по их абсолютным значениям. Его применение возможно в случае наличия достаточного количества наблюдений из более чем одной переменной. Сравниваются коэффициенты корреляции между одной парой или множеством пар признаков, для установления между ними статистических взаимосвязей.

Популярность метода обусловлена двумя моментами: коэффициенты корреляции относительно просты в подсчете, их применение не требует специальной математической подготовки.

К недостаткам корреляционного анализа относится предположение о линейной зависимости наблюдаемых переменных до проверки значений.

В общем виде формула расчета коэффициента корреляции такова:

Корреляционная плеяда – эта группа признаков, связанных друг с другом сильнее, нежели с признаками других групп.

Корреляционные плеяды используются как наглядный способ представления результатов корреляционного анализа. Это не отдельный статистический метод, а лишь способ графически представить корреляционные взаимосвязи.

1. **Описание проблемы**

Для оптимальной и корректной работы производства необходимо обладать информацией о зависимости качества продукции от производственных факторов.

Современное предприятие должно стремиться к максимизации прибыли с производства, чего можно добиться, модернизировав производственный процесс. И самым успешным предприятием станет то, которое сумеет максимально успешно проводить модернизацию, а что именно необходимо модернизировать, можно разузнать используя методы корреляционного анализа. Использовав их, можно будет увидеть зависимость параметров производственного процесса друг от друга.

1. **Варианты решения проблемы**

Для решения поставленной задачи я воспользуюсь корреляционным анализом матрицы зависимости, а также построю корреляционную плеяду и минимальный дендрит.

Корреляционный анализ позволит определить производственные факторы, которые более всего влияют на основные выходные параметры производства:

* Прочность.
* Пластичность.
* Влагопрочность.
* Загрязнение воды.
* Загрязнение атмосферы.
* Затраты на производство.

Для каждого из выходных параметров будут обозначены два производственных фактора, оказывающих наибольшее влияние на результат.

Корреляционная плеяда будет построена для визуального отображения зависимостей выходных параметров от производственных факторов.

1. **Обработка результатов**

С прошлой лабораторной работы мы имеем базу данных (рис. 1) на основе которой мы и будем проводить корреляционный анализ.

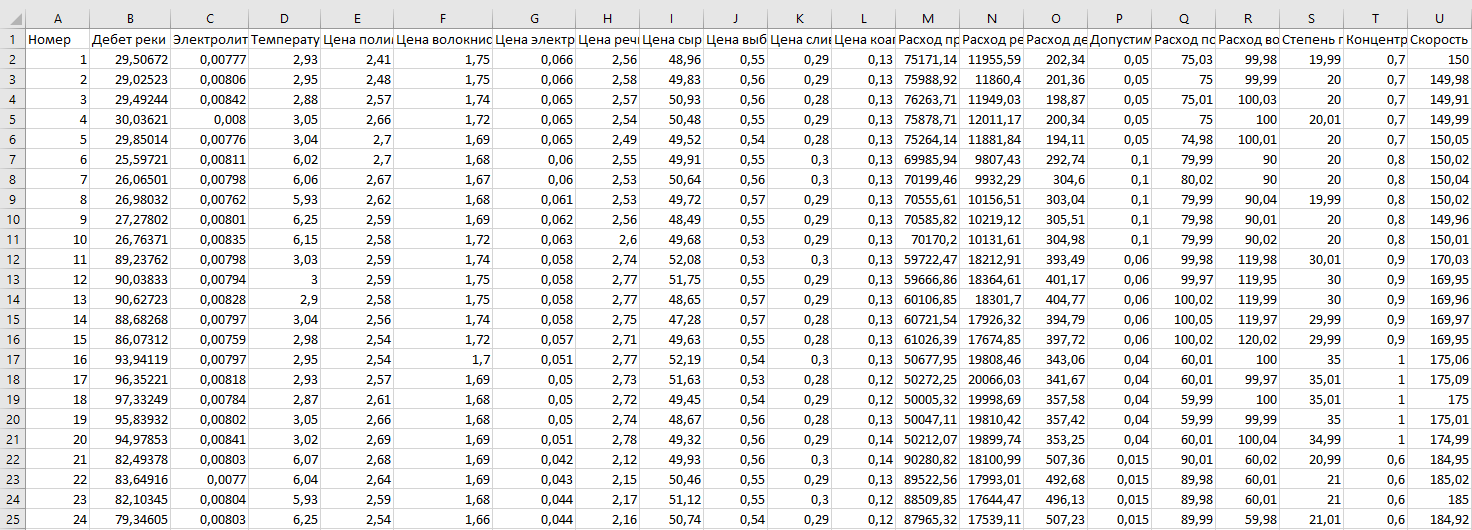


Рисунок 1 – Часть базы данных

Во-первых нам необходимо, взяв за основу нашу базу данных, построить корреляционную матрицу (рис. 2). Строить я её буду там же, где и база данных, в Excel.

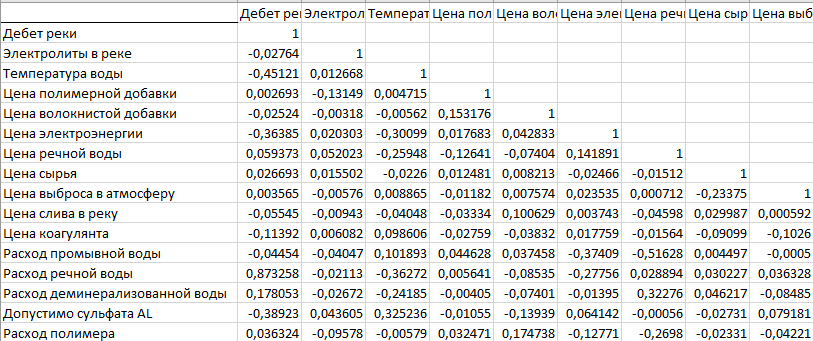


Рисунок 2 ­– часть корреляционной матрицы

В итоге мы получаем треугольную матрицу, но для дальнейшей работы нам её необходимо преобразовать в более удобный вид.

И первым шагом надо будет уменьшить размерность марицы оставив лишь строки выходных параметров и удалив такие столбцы (рис. 3), что позволит нам сократить ей более чем в три раза.



Рисунок 3 – уменьшенная матрица

Следующим нашим шагом будет оставить следующие уровни значений и обозначим соответсвующим цветом (рис 4.):

* 0.2 – 0.3
* 0.3 ­– 0.5
* 0.5 – 1

Такие пороговые значения позволят получить достаточно данных для построения плеяды. Также отбросим те столбцы, которые после преобразования остались без связей со строками.



Рисунок 4 – итоговая матрица

Как мы можем увидеть здесь есть место быть многофакторному воздействию на выходные параметры. Конечно, есть факторы, которые не попали в итоговою матрицу, потому что анализ показал, что взаимосвязи между ними и выходными параметрами найдено не было.

Затем на основе имеющих данных я построю корреляционную плеяду, на которую перенесу то, что мы имеем в матрицы, а именно факторы, влияющие на производство, они будут иметь прямоугольный вид и выходные параметры, они будут иметь овальный вид. А также укажу связи между ними, где в зависимости от её величины будут меняться цвет и толщина связи (рис. 5).

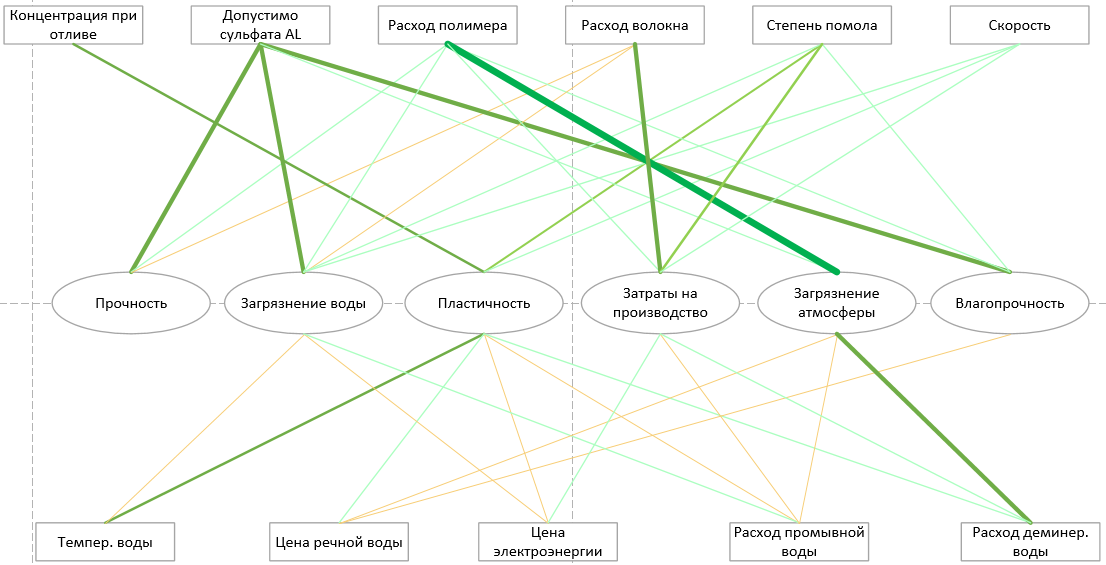


Рисунок 5 – корреляционная плеяда

Взглянув на плеяду можно четко увидеть, что каждый выходной параметр связан сильной связью (>0.5) с одним или более входным параметром. И присутствует даже одна связь равная 0.993 между загрязнением атмосферы и расходом полимера, что указывает на несомненную зависимость.

После построения плеяды я взялся за постройку дендрита (рис. 6)

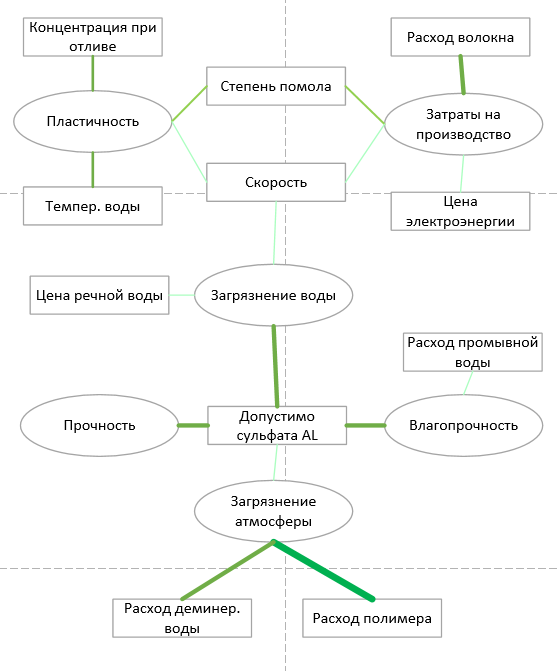


Рисунок 6 – Дендрит

Построив дендрит можно увидеть, что все выходные параметры связаны друг с другой в той или иной мере. Присутствуют как сильные, так и не очень взаимосвязи.

1. **Вывод**

Взяв базу данных из предыдущей лабораторной мною был проведен корриляционный анализ этой базы данных, составлена корриляционная матрица, затем она была отредактирована и по ней строилась корриляционная плеяда и дендрит, показывающие взаимосвязи между входными и выходними параметрами. В следсвии чего, мы могли наглядно увидеть сами взаимосвязи и их силу.

Показатель который влиял сильно (> 0.5; < 0.9) и на максимально колличество выходных параметров ­– «Допустимо cульфата алюминия». Как можно видеть он сильно влияет на три параметра и средне еще на один за другими показателями такого замечено не было. Так же есть очень сильная (>0.9) взаимосвязь, замеченная ранее, между загрязнением атмосферы и расходом полимера. В заключении выделю следующие входные показатели максимально влияющие на выходные параметры, это:

* Концентрация при отливе
* Расход волокна
* Расход деминерализованной воды
* Расход полимера
* Расход промывочной воды
* Цена электроэнергии
* Температура воды
* Цена речной воды
* Скорость
* Степень помола
* Допустимо сульфата алюминия

Из двадцати выходных параметров у нас осталось лишь одиннадцать, что чуть больше половины. Из этого делаем вывод, что оставшиеся параметры почти никак не влияют на производственный процесс и нам удалось выделить основные параметры производственного процесса влияющие на конечный результат.