МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федерально автономное образовательное учреждение высшего образования

«Севастопольский государственный университет»

кафедра Информационных систем

Куркчи Ариф Эрнестович

Институт информационных технологий и управления в технических системах

курс 3 группа ИС/б-31-о

09.03.02 Информационные системы и технологии (уровень бакалавриата)

ОТЧЕТ

по лабораторной работе №1

по дисциплине «Интеллектуальный анализ данных»

на тему «Корреляционный и регрессионный анализ данных. Множественный регрессионный анализ»

Отметка о зачете \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_

(дата)

Руководитель практикума

ассистент Токарев А.И.

(должность) (подпись) (инициалы, фамилия)

Севастополь 2017

1.Цель работы

Закрепить теоретические знания и приобрести практические навыки в построении регрессионных моделей объекта по экспериментальным данным с использованием программы Statistica.

2.Постановка задачи

По результатам наблюдений за функционированием объектов получены экспериментальные данные. Построить регрессионные модели объектов по заданным экспериментальным данным.

Решение общей задачи разбивается на несколько этапов:

* предварительная обработка данных с целью стандартизации результатов наблюдения;
* оценка параметров регрессионных моделей;
* проверка значимости коэффициентов регрессии;
* оценка точности регрессионных моделей;
* выводы о возможности применения составленных регрессионных моделей.

3.Ход работ

Для начала анализа в программу Statistica ввели по 125 значений трех видов: евро, доллары, гривны. Получили дискриптивные статистики по каждому признаку.

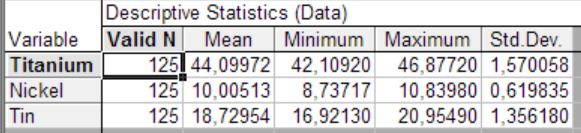
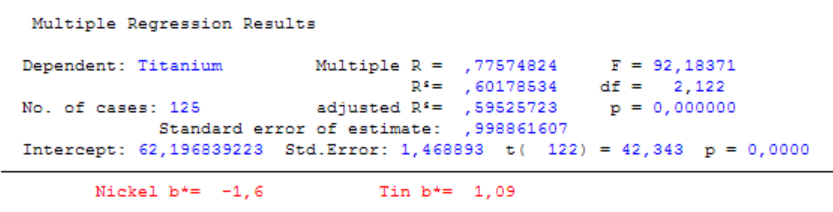


Рисунок 1 – Дискриптивные статистики трех признаков

Проведем множественную регрессию и оценим результаты



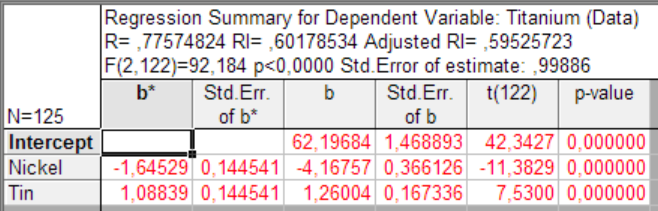


Рисунок 2 – Множественная регрессия

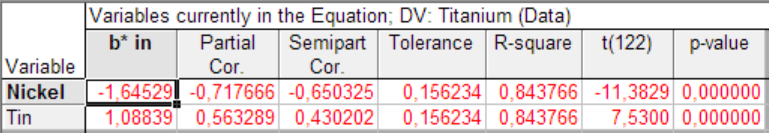
Так как значение коэффициента множественно корреляции (multiple R) стремиться к 1, то взаимосвязь между результативной и независимыми переменными средняя.

По полученным коэффициентам уравнения (столбец b) моем построить уравнение множественной регрессии.

Ti = 62,197-4,168\*Ni+1,26\*Sn

Это уравнение объясняет 60,17% вариации зависимой переменной.

Проанализируем линейные коэффициенты парной и частной корреляции. Коэффициент корреляции - это показатель, оценивающий тесноту связи между признаками. Так как коэффициенты корреляции олово-никель, никель-олово имеют значение больше, чем 0,91, то связь оценивается как очень тесная. Связь титан-никель оценивается как значительная, титан-олово как умеренная (взаимосвязи обратно пропорциональны).

Рисунок 3 – Линейные коэффициенты частной корреляции

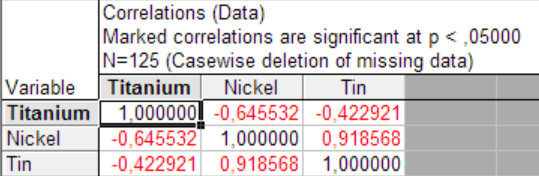


Рисунок 4 – Линейные коэффициенты парной корреляции

С помощью F-критерия Фишера оценим статистическую надежность уравнения регрессии в целом. F-критерий используется как общий критерий, подтверждающий или опровергающий влияние фактора на общую вариацию признака. В нашем случае для никеля вероятность нулевой гипотезы позволяет ее опровергнуть и говорить о достоверном влиянии цены титана на цену олова.

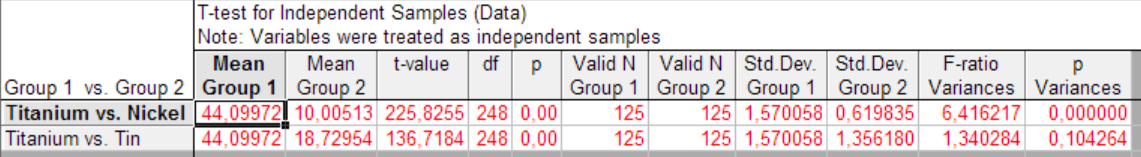


Рисунок 5– F-критерий Фишера

Выводы

В ходе лабораторной работы была изучена программа Statistica, с помощью которой были построены регрессионные модели объекта по экспериментальным данным.