**Корреляция**

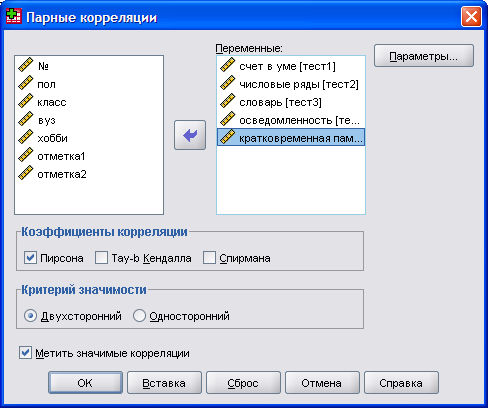
**Задание 1.** Необходимо установить, насколько значения переменных тест1 – тест5 зависимы друг от друга на основе данных файла ex01.sav.

Прежде чем оценивать корреляцию двух переменных, необходимо построить график зависимости между ними — график двумерного рассеяния.

В меню **Анализ** выберите команду **Корреляции>Парные**. На экране появится диалоговое окно Парные корреляции:

Окно Парные корреляции позволяет настраивать параметры вычисления корреляций. В левом списке содержатся имена всех переменных, имеющий числовой тип.

Сначала необходимо создать корреляционную матрицу для переменных тест1, ..., тест5, с вычислением корреляций Пирсона. Для этого переместите их в окно **Переменные**.

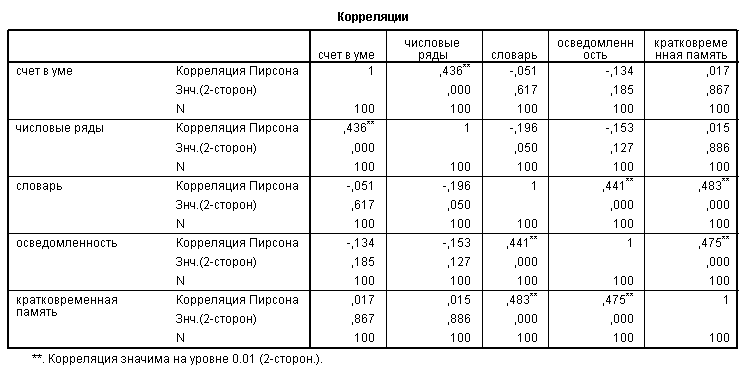


В группе *Коэффициенты корреляции* по умолчанию установлен флажок Пирсона.

В группе *Критерий* значимости по умолчанию установлен переключатель *Двухсторонний.*

Флажок *Метить значимые корреляции* по умолчанию установлен. Это означает, что корреляции, вычисленные с уровнем значимости от 0,01 до 0,05, будут помечены одной звездочкой (\*), а от 0 до 0,01 — двумя звездочками (\*\*).

Результаты в окне вывода представляются следующим образом:



Верхнее значение в каждой ячейке таблицы является коэффициентом корреляции между соответствующими двумя переменными, вычисленным с точностью до трех знаков после запятой.

Следующее значение показывает уровень значимости этого коэффициента корреляции. Нижнее значение равно числу объектов, участвовавших в вычислении коэффициента корреляции.

Надписи под таблицей поясняют смысл символов \* и \*\*, используемых в качестве меток для некоторых вычисленных коэффициентов корреляции, и содержат информацию об уровне значимости.

На главной диагонали таблицы лежат единичные значения коэффициентов корреляции для всех переменных. То есть, каждая переменная имеет строгую положительную корреляцию сама с собой.

**Задание 2.**

Эту же задачу решить в Pyton. Можно использовать файл данных ex01.xlsx

Построить корреляционную матрицу с помощью тепловой карты. Установить статистически значимую взаимосвязь между переменными и определить силу этой связи.

**Задание 3**

Определите отсутствие или наличие линейной корреляционной взаимосвязи между валовым региональным продуктом и уровнем безработицы. Для анализа используйте данные Федеральной службы государственной статистики (файлы trud.xlsx и vrp,xlsx).

Для расчетов используйте данные за один год по всем субъектам РФ в соответствии с вариантами (вариант- номер списка в журнале):

|  |  |
| --- | --- |
| **Вариант** | **Год** |
| 1 | 2001 |
| 2 | 2002 |
| 3 | 2003 |
| 4 | 2004 |
| 5 | 2005 |
| 6 | 2006 |
| 7 | 2007 |
| 8 | 2008 |
| 9 | 2009 |
| 10 | 2010 |
| 11 | 2011 |
| 12 | 2012 |
| 13 | 2013 |
| 14 | 2014 |
| 15 | 2015 |
| 16 | 2016 |
| 17 | 2017 |
| 18 | 2018 |
| 19 | 2019 |
| 20 | 2020 |
| 21 | 2021 |
| 22 | 2000 |

Инструмент анализа выбирается любой

Этапы выполнения задания:

1. Постройте диаграмму рассеяния и выдвинете гипотезу о характере связи между рассматриваемыми переменными.

2. Рассчитайте коэффициент корреляции.

3. Проверьте значимость коэффициента корреляции.

4. Сделайте выводы.

Сохраните результаты всех заданий в файле Word и опишите результаты.