**Тема: Корреляционный анализ**

**Задание 1.** Для 20 работников предприятия были проанализированы следующие показатели:

* квалификация (разряд);
* трудовой стаж;
* оценка работника с точки зрения руководителя (по 5-бальной шкале);
* количество нарушений трудовой дисциплины в течение года;
* количество пропущенных по болезни дней;
* средняя заработная плата;
* процент брака.

Определить, имеется ли связь между средней заработной платой и процентом брака с квалификацией, стажем, оценкой, количеством нарушений и количеством пропущенных по болезни дней. Использовать процедуру ***Корреляция*** Пакета Анализа.

Исходные данные:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № | Квалифи-кация | Стаж | Оценка | Количество нарушений | Кол-во дней по болезни | Средняя заработная плата | Процент брака |
| 1 | 4 | 10 | 3 | 3 | 18 | 225 | 6,2 |
| 2 | 3 | 6 | 3 | 11 | 6 | 220 | 4,3 |
| 3 | 3 | 10 | 3 | 4 | 0 | 211 | 5,3 |
| 4 | 5 | 18 | 3 | 0 | 3 | 243 | 5,2 |
| 5 | 6 | 26 | 5 | 1 | 4 | 322 | 4,4 |
| 6 | 4 | 8 | 3 | 0 | 4 | 277 | 4,3 |
| 7 | 4 | 21 | 3 | 0 | 0 | 300 | 3,3 |
| 8 | 3 | 13 | 4 | 3 | 4 | 225 | 4,2 |
| 9 | 5 | 8 | 3 | 0 | 0 | 327 | 3,6 |
| 10 | 3 | 15 | 3 | 0 | 0 | 231 | 5,4 |
| 11 | 4 | 11 | 3 | 3 | 0 | 272 | 5,2 |
| 12 | 6 | 18 | 3 | 0 | 12 | 353 | 5,3 |
| 13 | 3 | 8 | 3 | 2 | 4 | 199 | 6,3 |
| 14 | 3 | 10 | 4 | 2 | 0 | 221 | 5,2 |
| 15 | 3 | 8 | 3 | 3 | 4 | 202 | 6,2 |
| 16 | 2 | 13 | 3 | 7 | 5 | 207 | 5,3 |
| 17 | 4 | 18 | 3 | 0 | 0 | 218 | 5,4 |
| 18 | 3 | 8 | 3 | 0 | 0 | 203 | 6,2 |
| 19 | 3 | 10 | 4 | 2 | 4 | 196 | 6,3 |
| 20 | 3 | 8 | 3 | 3 | 5 | 198 | 6,1 |

**Корреляционный и регрессионный анализ в пакете STATISTICA**

Для ответа на вопрос, имеется ли какая-либо статистическая связь между исследуемыми переменными и какова теснота связи, необходимо:

- выбрать с учетом специфики анализируемых переменных показатель статистической связи (парный, множественный или ранговый коэффициент корреляции);

- определить его числовое значение по имеющимся данным;

- проверить полученное значение на статистическую значимость.

Рассмотрим следующий пример: для 34 работников предприятия были проанализированы следующие показатели (Data2.sta):

Х1 – квалификация (разряд);

Х2 – трудовой стаж;

Х3 – оценка работника с точки зрения руководителя (по 5-бальной шкале);

Х4 – количество нарушений трудовой дисциплины в течение года;

Х5 – личное отношение к работе (1 – не нравится, 2 – нравится);

Х6 – семейное положение (1 – холост, 2 – женат);

Х7 – количество пропущенных по болезни дней;

Y1 – средняя заработная плата;

Y2 – выработка (в % к средней по цеху);

Y3 – процент брака.

Факторы Х1 – Х7 составляют первую группу (группу входных факторов), а факторы Y1 – Y3 вторую группу выходных параметров.

**1. Расчет корреляционной матрицы**

Построим корреляционную матрицу для данных. Для проведения корреляционного анализа необходимо выбрать *Статистика / Основная статистика/Таблицы*, раздел *Correlation matrices*.

Далее необходимо определить, какие переменные будут находиться в строках и столбцах (*One variable list*, все, кроме номера). Для представления в корреляционной матрице значений уровня значимости *р* для всех коэффициентов корреляции необходимо во вкладке Options выбрать позицию *Display r, p-levels*. После нажатия на *Summary* будет рассчитана матрица. Красным цветом будут выделены те парные коэффициенты корреляции, для которых уровень значимости не превышает 0,05. (Уровень значимости характеризует вероятность событий, условно принимаемых за невероятные, т.е. чем ниже его величина, тем достоверней результат.)

**Ответить на следующий вопрос: с какими фактами связаны в наибольшей степени средняя заработная плата, выработка и процент брака?**

Для построения графического отображения корреляционной взаимосвязи любых двух переменных необходимо выбрать переменные и нажать *2D scatterplots*. На рисунке будут представлены: функция, отражающая взаимосвязь переменных, ее графическое отображение, значение парного коэффициента корреляции, корреляционное поле точек и границы доверительного интервала.

**Получить графическое отображение корреляционной взаимосвязи средней заработной платы, выработки, процента брака и квалификации рабочего.**

**2. Анализ парных ранговых связей**

Под ранговой корреляцией понимается статистическая связь между ординальными (порядковыми) переменными. Методы ранговой корреляции широко используются при проведении различных экспертных исследований. Основной задачей ранговой корреляции является ответ на вопрос: есть ли связь между упорядочением (ранжировкой) анализируемых объектов по свойству Х1 и упорядочением тех же объектов по свойству Х2?. Степень взаимосвязи оценивают, как правило, по коэффициентам Спирмена или Кендала.

Пример. Десять показателей были проранжированы вначале по степени интенсивности, а затем – по эффективности их функционирования. В результате были получены следующие ряды рангов: Х1 = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) и Х2 = (2, 3, 1, 4, 6, 5, 9, 7, 8, 10).

**Создать файл с данными, затем выбрать *Статистика / Непараметрические данные*, далее *Correlations (Spearman …),* задать переменные, для которых необходимо оценить тесноту взаимосвязи и во вкладке *Advanced* определить требуемый коэффициент.**

В рассматриваемом примере корреляция по Спирмену дает коэффициент = 0,915 с высоким уровнем статистической значимости.