

Коэффициент корреляции и корреляционный анализ

Выполнил:

Толстопятов Д.

Е.

Паначевный В.

М.

Основные понятия

- **Коэффициент корреляции** - это статистический показатель зависимости двух случайных величин.
- **Корреляционный анализ** - метод, позволяющий обнаружить зависимость между несколькими случайными величинами.

Расчёт коэффициента корреляции

Есть массив из n точек $\{x_{1,i}, x_{2,i}\}$

Рассчитываются средние значения для каждого параметра:

$$\bar{x}_1 = \frac{\sum x_{1,i}}{n}, \quad \bar{x}_2 = \frac{\sum x_{2,i}}{n}$$

И коэффициент корреляции:

$$r = \frac{\sum (x_{1,i} - \bar{x}_1) \cdot (x_{2,i} - \bar{x}_2)}{\sqrt{\sum (x_{1,i} - \bar{x}_1)^2} \cdot \sqrt{\sum (x_{2,i} - \bar{x}_2)^2}}$$

r изменяется в пределах от -1 до 1. В данном случае это линейный коэффициент корреляции, он показывает линейную взаимосвязь между x_1 и x_2 : r равен 1 (или -1), если связь линейна.

Методами корреляционного анализа решаются следующие задачи:

1) Взаимосвязь.

Есть ли взаимосвязь между параметрами?

2) Прогнозирование.

Если известно поведение одного параметра, то можно предсказать поведение другого параметра, коррелирующего с первым.

3) Классификация и идентификация объектов.

Корреляционный анализ помогает подобрать набор независимых признаков для классификации.

Свойства коэффициента корреляции:

- r изменяется в интервале от -1 до $+1$.
- Знак r означает, увеличивается ли одна переменная по мере того, как увеличивается другая (положительный r), или уменьшается ли одна переменная по мере того, как увеличивается другая (отрицательный r).
- Величина r указывает, как близко расположены точки к прямой линии. В частности, если $r = +1$ или $r = -1$, то имеется абсолютная (функциональная) корреляция по всем точкам, лежащим на линии (практически это маловероятно); если $r = 0$, то линейной корреляции нет (хотя может быть нелинейное соотношение). Чем ближе r к крайним точкам (± 1), тем больше степень линейной связи.
- Коэффициент корреляции r безразмерен, т. е. не имеет единиц измерения.

Свойства коэффициента корреляции:

- Величина r обоснована только в диапазоне значений x и y в выборке. Нельзя заключить, что он будет иметь ту же величину при рассмотрении значений x или y , которые значительно больше, чем их значения в выборке.
- x и y могут взаимозаменяться, не влияя на величину.

Значения коэффициента корреляции

Охарактеризовать силу корреляционной связи можно прибегнув к шкале Челдока, в которой определенному числовому значению соответствует качественная характеристика. В случае положительной корреляции при значении:

- 0-0,3 – корреляционная связь очень слабая;
- 0,3-0,5 – слабая;
- 0,5-0,7 – средней силы;
- 0,7-0,9 – высокая;
- 0,9-1 – очень высокая сила корреляции

Расчёт коэффициента корреляции в Excel

Алгоритм:

1. Выбрать пустую ячейку, в которую будет выведен результат расчетов.
2. Нажать в главном меню Excel пункт «Формулы».
3. Среди кнопок, сгруппированных в «Библиотеку функций», выбрать «Другие функции».
4. В выпадающих списках выбрать функцию расчета корреляции (Статистические — КОРРЕЛ).
5. В Excel откроется панель «Аргументы функции». «Массив 1» и «Массив 2» — это диапазоны сравниваемых данных. Для автоматического заполнения этих полей можно просто выделить нужные ячейки таблицы.
6. Нажать «ОК», закрыв окно аргументов функции. В ячейке появится подсчитанный коэффициент корреляции.

Вставить Буфер обмена Шрифт Выравнивание Число

Общий

Условное форматирование

Форматировать как таблицу

Стили ячеек

КОРРЕЛ X ✓ fx =КОРРЕЛ(A2:A13;B2:B13)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	у	х															
2	23	10															
3	25,5	11,25															
4	26	11,5															
5	28	12,5															
6	33	15															
7	34	15,5															
8	36	19															
9	35	18,5															
10	32	17															
11	30	16															
12	30	16															
13	29	14															

$$r_{xy} = \frac{\sum (x_i - x_{\text{средн}})(y_i - y_{\text{средн}})}{\sqrt{\sum (x_i - x_{\text{средн}})^2 * \sum (y_i - y_{\text{средн}})^2}}$$

Аргументы функции

КОРРЕЛ

Массив1 A2:A13 = {23;25,5;26;28;33;34;36;35;32;30;...}

Массив2 B2:B13 = {10;11,25;11,5;12,5;15;15,5;19;18,...

= 0,931682953

Возвращает коэффициент корреляции между двумя множествами данных.

Массив2 второй диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.

Значение: 0,931682953

[Справка по этой функции](#)

OK

Отмена

B2:B13)

Расчёт коэффициента корреляции в Mathcad

1. Задаём 2 массива (x и y)

```
i := 0 .. 11
```

$z_i :=$

4
3
4
3
4
3
3
4
3
5
3
3

$y_i :=$

4
3
4
3
4
3
3
3
3
4
5
3
3

2. Вычислим средние оценки и стандартное отклонение:

```
Zmean := mean(z)
```

```
Zmean = 3.5
```

```
Ymean := mean(y)
```

```
Ymean = 3.5
```

```
Yst := stdev(y)
```

```
Yst = 0.64
```

3. Для вычисления коэффициента корреляции имеется встроенная функция

```
PhysMathCorr := corr(y, z)
```

```
PhysMathCorr = 0.8
```

Близкий к 1 коэффициент корреляции свидетельствует о высокой зависимости между успеваемостью по этим предметам.