

Анализ статистической информации

Анализ одномерного массива данных

- представить в виде ряда распределения путем ранжирования (в порядке возрастания или убывания анализируемого количественного признака),
- дать характеристику этой совокупности, указав:
 - центральные значения ряда (среднее арифметическое, медиану, моду),
 - размах варьирования (максимум, минимум),
 - частотное распределение в процентах,
 - форму кривой распределения.

Анализ двумерных данных

1-й вариант: изучать каждое измерение по отдельности как часть одномерной совокупности данных.

2-й вариант: при совместном изучении обоих параметров появляется возможность выявить *взаимосвязь* между ними.

При анализе статистических данных приходится решать проблему и более высокого уровня выявления функциональной зависимости между *воздействующим фактором* и *регистрируемой (изучаемой) величиной*.

Зависимость одной случайной величины от значений, которые принимает другая случайная величина, в статистике называется *регрессией*. Если этой зависимости придан аналитический вид, то такую форму представления изображают *уравнением регрессии*.

Процедура поиска предполагаемой зависимости

1. Установление значимости связи между числовыми совокупностями.
2. Возможность представления этой зависимости в форме математического выражения (уравнения регрессии).

1-й этап

Выявление так называемой *корреляции* или *корреляционной зависимости*.

Корреляция рассматривается как признак, указывающий на *взаимосвязь* ряда числовых последовательностей.

Иначе говоря, корреляция характеризует *силу взаимосвязи* в данных.

Если это касается взаимосвязи двух числовых массивов x и y , то такую корреляцию называют *парной*.

На этом этапе *не* ставится задача определить, является ли одна из этих случайных величин *функцией*, а другая – *аргументом*.

2-й этап

Регрессионный анализ - отыскание количественной зависимости между ними в форме конкретного аналитического выражения $y=f(x)$

Различие

Корреляционный анализ позволяет сделать вывод о силе взаимосвязи между парами данных x и y , а *регрессионный анализ* используется для *прогнозирования* одной переменной (y) на основании другой (x).

Виды зависимостей

1. *Функциональная* зависимость - При наличии функциональной связи каждому значению воздействующего фактора (аргумента) соответствует строго определенная величина другого показателя (функции), т.е. изменение результативного признака всецело обусловлено действием факторного признака.
2. *Статистическая* - случайная. Значению одного фактора соответствует какое-то приближенное значение исследуемого параметра, его точная величина является непредсказуемой и поэтому получаемые показатели оказываются случайными величинами. Возможно воздействие и иных факторов

Статистическая зависимость

Двумерные данные можно анализировать с использованием *диаграммы рассеяния* в координатах "**x** - **y**", которая дает визуальное представление о взаимосвязи исследуемых совокупностей.

Корреляция

Для количественной оценки существования связи между изучаемыми совокупностями случайных величин используется специальный статистический показатель – *коэффициент корреляции* r . Если предполагается, что эту связь можно описать линейным уравнением типа $y = a + bx$ (где a и b – константы), то принято говорить о существовании линейной корреляции.

- Коэффициент r - это безразмерная величина, она может меняться от 0 до 1. Чем ближе значение коэффициента к единице (неважно, с каким знаком), тем с большей уверенностью можно утверждать, что между двумя рассматриваемыми совокупностями переменных существует линейная связь. Иными словами, значение какой-то одной из этих случайных величин (y) существенным образом зависит от того, какое значение принимает другая (x).
- Если окажется, что $r=1$ (или -1), то имеет место классический случай чисто функциональной зависимости (т.е. реализуется идеальная взаимосвязь).

Сила связи

Зная коэффициент корреляции, можно дать качественно-количественную оценку тесноты связи.

Величина коэффициента парной корреляции Характеристика силы связи

- | | |
|------------|-------------------------|
| • До 0,3 | Практически отсутствует |
| • 0,3-0,5 | Слабая |
| • 0,5-0,7 | Заметная |
| • 0,7-0,9 | Сильная |
| • 0,9-0,99 | Очень сильная |