**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ**

**1.3.1. Парная корреляция**

Предварительную характеристику корреляционной связи между случайными величинами x и y можно дать **путем построения так называемого корреляционного поля, т.е. графика зависимости  с нанесением на него всех экспериментальных точек.**

**О наличии связи между двумя СВ можно судить по тесноте группирования точек на корреляционном поле вокруг условной прямой или кривой линии.** В качестве примера приведем ряд корреляционных полей различной формы.

*Рис.1. Корреляционные поля различной конфигурации*

Так, из рисунка 1 а, в и г видно, что между х и у определенная связь существует, а вот по данным, приведенным на рисунке б, связь между х и у отсутствует. **По форме корреляционного поля можно судить и о предполагаемой форме связи между двумя СВ, которая может быть**:

* линейной (рис. 1 а, в);
* нелинейной (рис.1 г);
* прямой (рис. 1 а);
* обратной (рис, 1 в).

Кроме этого степень разбросанности точек на корреляционном поле в определенной мере свидетельствует и о силе связи между **х** и **у**.

Так, очевидно, что для данных, приведенных на рисунке а, связь между **х** и **у** слабая, тогда как для данных, показанных на рисунках в и г – связь между **х** и **у** – достаточно сильная.

Однако такая визуальная и качественная оценка, хотя и дает определенную информацию, но не может заменить количественной оценки существования связи между **х** и **у**, а также оценки формы и силы этой связи.

**Сила связи между двумя случайными величинами оценивается величиной коэффициента парной корреляции** или просто коэффициента корреляции, определяемого по следующей формуле:

 , **(37)**

* где: n – число пар наблюдений (измерений);  - средние арифметические значения х и у ; **σх, σу** – среднеквадратические отклонения **х** и **у**, рассчитываемые по формулам **(25)**
* дисперсию − D

; (25)

и **(11)**

* среднее квадратичное отклонение σ:

.

Значения коэффициента корреляции **rух** изменяются в пределах от –1 до +1, т.е. **–1 ≤ rух ≤ +1.**

**Если с ростом значения х значение у вырастает, то rух будет иметь знак плюс (положительная или прямая связь), а если уменьшается, то – знак минус (отрицательная или обратная связь).**

**Чем ближе абсолютное значение rух к 1,тем сильнее значения одной СВ зависят от того, какие значения принимает другая СВ, т.е. тем сильнее связь между ними.** .

Тесноту связи между х и у обычно считают:

* удовлетворительной при **rух ≥ |0,5|** ;
* хорошей ‑ при **rух = |0,8 ÷ 0,85|.**

Следует помнить о том, что rух является СВ, т.е. может принимать различные значения при повторных измерениях. Кроме этого величина rух зависит от числа пар наблюдений. С уменьшением и достоверность выводов, формируемых после определения rух, снижается.

**При rух = ±1 – две случайных величины связаны линейной функциональной связью, т.е. каждому конкретному значению х соответствует только одно строго определенное значение у.**

При **rух = 0** СВ называют **некоррелированными (независимыми).** Однако обратное утверждение, что СВ независимы, если **rух = 0**, несправедливо, так как **rух как мера тесноты связи имеет четкий математический смысл только при линейной зависимости между СВ и при нормальном их распределении**.

Поэтому значение **rух может быть равным нулю, когда СВ связаны нелинейной связью, а следовательно, зависимы друг от друга.**

Достоверность коэффициента корреляции оценивают критерием надежности (критерием Стьюдента):

**** **(39)**

где . **(40)**

Если расчетное значение Qr выше табличного, то можно сделать заключение о том, что величина коэффициента корреляции является значимой. Табличные значения находят по таблице значений критериев Стьюдента. При этом учитываются количество степеней свободы *(m = п-1*)и уровень значимости (обычно 0,05 или 0,01). Если фактическое Qr выше табличного, связь между показателями является надежной, а величина коэффициентов корреляции - значимой. В нашем примере Qr > 2,26 с доверительной вероятностью равной 0,95 можно утверждать о значимости найденного коэффициента корреляции **rух**, т.е. о существовании между х и у линейной связи.

По известным значениям величин **rух, σх и , σу** несложно определить линейное уравнение регрессии, описывающее связь между **х** и **у**, т.е.

, (41)

где  (42)

 (43)

**После нахождения линейной математической модели следует оценить возможность улучшения описания связи между х и у путем перехода к нелинейной модели.**

Для этого необходимо вычислить корреляционное отношение по следующей формуле:

 (44)

* где  - значение выходного параметра в i- м опыте, рассчитанное по найденной нелинейной модели.
* yi − фактическое значение параметра в i - ом опыте.

Корреляционное отношение **ηу** характеризует силу (степень тесноты) связи между двумя СВ **при отсутствии между ними линейной зависимости, т.е. связанными не линейно.**

Значения **ηу** могут находится в пределах от 0 до 1.

Для некоррелированных (независимых) СВ **ηу = 0**, а в случае функциональной зависимости между ними **ηу = 1.**

Если связь между двумя СВ линейна, то корреляционное отношение равно абсолютному значению коэффициента корреляции, т.е. **ηу = │ rух │.**

Следует отметить, что значимое различие значений **ηу и rух** проявляется только при достаточно большом числе пар измерений.

Достоверность корреляционного отношения оценивается по критерию его надежности.

. **(45)**

**При Θr > 2,6 с доверительной вероятностью равной 0,95 можно утверждать, что найденное корреляционное отношение значимо.**

По известным значениям ηу иrух оценивают степень нелинейности:

. **(46)**

Если **n02 < (12/n),** то переход к нелинейной модели не улучшит связи между х и у, а в противном случае – может привести к лучшим результатам.

**1.3.1.1. Применение корреляционного анализа для уменьшения**

**числа параметров (факторов)**

Очевидно, что если две случайные величины являются коррелированными т.е. зависимыми друг от друга, о чем свидетельствует значимость коэффициента корреляции **rух,** то любая из них (х или у) может быть исключена из рассмотрения.

Для сокращения числа параметров в случае одномерно‑многомерного объекта исследований или числа факторов в случае многомерно‑одномерного объекта исследований рассчитывают значения коэффициента корреляции между всеми возможными парами параметров (факторов), а так же в зависимости от схемы объекта исследований – между параметрами и входным фактором или входными факторами и параметром.

На основе расчетов составляют так называемую **нормированную корреляционную матрицу**, пример которой приведен ниже.

**Таблица 3.** Корреляционная матрица

|  |  |
| --- | --- |
| **Параметры** | **Значения коэффициента корреляции** |
| **х** | **у1** | **у2** | **у3** | **у4** |
| **х** | **1** | **ry1x\*** | **ry2x\*** | **ry3x\*** | **ry4x\*** |
| **у1** |  | **1** | **ry2y1\*** | **ry3y1\*** | **ry4y1\*** |
| **у2** |  |  | **1** | **ry3y2\*** | **ry4y2** |
| **у3** |  |  |  | **1** | **ry4y3\*** |
| **у4** |  |  |  |  | **1** |

В этой матрице значимые значения коэффициента корреляции принято обозначать звездочками. Из приведенной корреляционной матрицы следует, что незначимым является лишь коэффициент корреляции между **у2**и **у4**.

Отсюда следует, что при исследовании влияния фактора х на параметры **у1**,**у2**, **у3**, **у4** вместо четырех параметров можно ограничиться двумя - **у2**и **у4**.