

Раздел 2. «Анализ распределений в социологическом/маркетинговом исследовании».

Тема 2.3. Корреляционный анализ.

Корреляция – статистическая связь двух переменных (количественных или порядковых), показывающая, что большему значению одной величины в определенной части случаев соответствует большее (в случае положительной, прямой корреляции) или меньшее (в случае отрицательной, обратной корреляции) значение другой величины.

Корреляционный анализ (correlation analysis) – статистический метод изучения взаимосвязи между двумя и более случайными величинами. В качестве случайных величин в эмпирических исследованиях выступают значения переменных, измеряемые свойства исследуемых объектов наблюдения.

Корреляционный анализ – статистический метод, позволяющий с использованием коэффициентов корреляции определить, существует ли зависимость между переменными и насколько она сильна.

Суть корреляционного анализа заключается в расчете коэффициентов корреляции. Коэффициенты корреляции могут принимать, как правило, положительные и отрицательные значения. Знак коэффициента корреляции позволяет интерпретировать направление связи, а абсолютное значение – силу связи.

Способ расчета коэффициентов корреляции зависит от шкал измерения переменных, между которыми исследуется взаимосвязь. Для переменных, измеряемых в количественной шкале (интервальной шкале или шкале отношений), рассчитывают ковариацию или корреляционный момент, а на его основе линейный коэффициент корреляции (коэффициент корреляции Пирсона).

Для оценки силы и направления связи между переменными, измеренными в порядковой шкале, используются непараметрические ранговые коэффициенты корреляции: коэффициент ранговой корреляции Кендалла и коэффициент корреляции Спирмена.

Корреляционный анализ используется в экономике, социологии и психологии, медицине, управления качеством, биометрии и других сферах. Популярность корреляционного анализа объясняется тем, что коэффициенты корреляции относительно просты в расчете, и их применение не требует специальной математической подготовки. С другой стороны – коэффициенты корреляции легко интерпретировать.

Однако корреляционный анализ имеет свою специфику и методику. Очень важно использование этого метода только при соблюдении предпосылок расчета того, или иного, коэффициента корреляции. Методика корреляционного анализа предполагает, не просто расчет коэффициентов корреляции, но и обязательную проверку их значимости, в основе которой лежит принцип проверки статистических гипотез, построение интервальных оценок коэффициентов корреляции.

Нередки случаи возникновения так называемых «ложных корреляций», приводящим к ложным выводам. В этом случае при анализе взаимосвязи между количественными переменными рассчитывают и анализируют частные коэффициенты корреляции.

Корреляционный анализ решает две основные задачи:

1) Первая задача заключается в определении формы связи, т. е. в установлении математической формы, в которой выражается данная связь. Это очень важно, так как от правильного выбора формы связи зависит конечный результат изучения взаимосвязи между признаками.

2) Вторая задача состоит в измерении тесноты, т. е. меры связи между признаками с целью установить степень влияния данного фактора на результат. Она решается математически путем определения параметров корреляционного уравнения. Затем проводятся оценка и анализ полученных результатов при помощи специальных показателей корреляционного метода (коэффициентов детерминации, линейной и множественной корреляции и т. д.), а также проверка существенности связи между изучаемыми признаками. Для количественной оценки тесноты связи служит выборочный коэффициент корреляции.

Корреляционная зависимость – это согласованные изменения двух (парная корреляционная связь) или большего количества признаков (множественная корреляционная связь). Суть ее заключается в том, что при изменении значения одной переменной происходит закономерное изменение (уменьшение или увеличение) другой(-их) переменной(-ых).

Коэффициент корреляции – двумерная описательная статистика, количественная мера взаимосвязи (совместной изменчивости) двух переменных.

При положительной линейной корреляции более высоким значениям одного признака соответствуют более высокие значения другого, а более низким значениям одного признака – низкие значения другого. •

При отрицательной линейной корреляции более высоким значениям одного признака соответствуют более низкие значения другого, а более низким значениям одного признака – высокие значения другого.

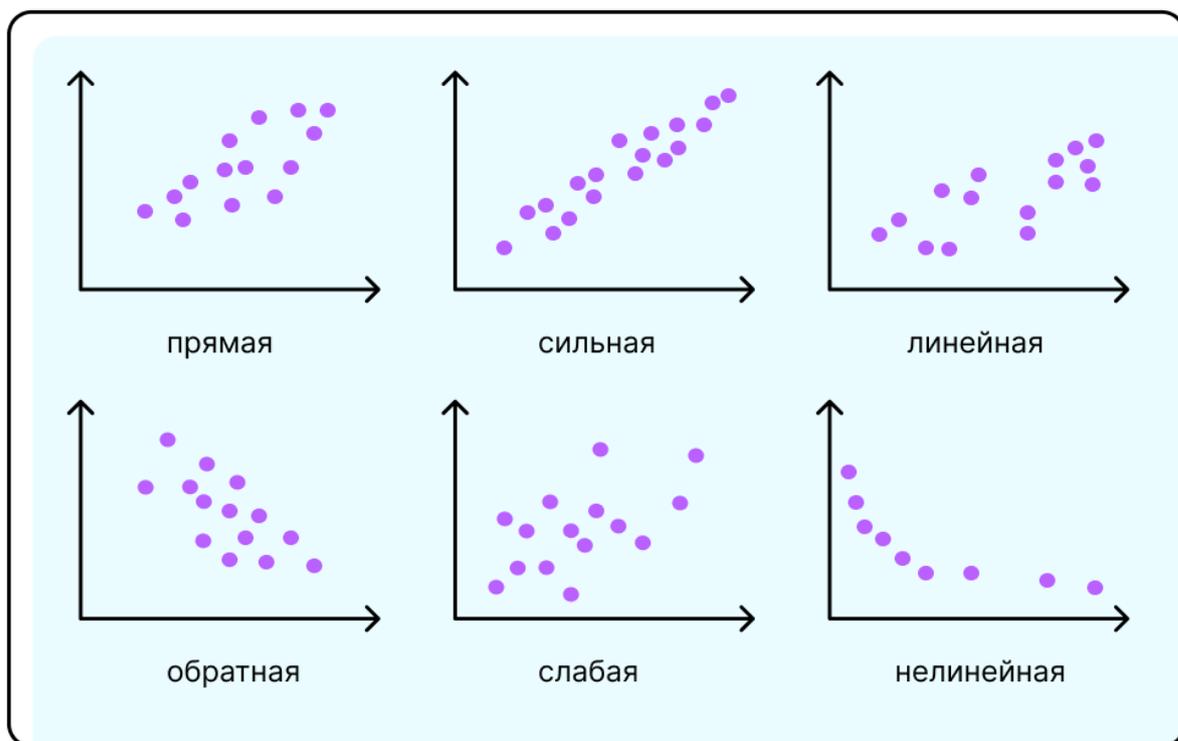
По сути, коэффициент корреляции означает степень, в которой две переменные движутся в тандеме друг с другом.

Положительный коэффициент, вплоть до максимального уровня «1», указывает, что движения двух переменных идеально выровнены и в одном направлении – если одна увеличивается, другая увеличивается на ту же величину.

Отрицательный коэффициент, вплоть до минимального уровня «-1», является прямо противоположным, указывая на то, что две величины движутся в противоположном направлении друг относительно друга.

После обнаружения стохастических связей между изучаемыми переменными величинами исследователь приступает к математическому описанию интересующих его зависимостей. Другими словами, необходимо перейти от корреляционного анализа к регрессионному анализу.

Корреляционный анализ не позволяет определить форму связи между переменными и предсказывать значения одной зависимой переменной по одной или нескольким независимым. Для этого, например, для количественных переменных применяется [линейный регрессионный анализ](#).

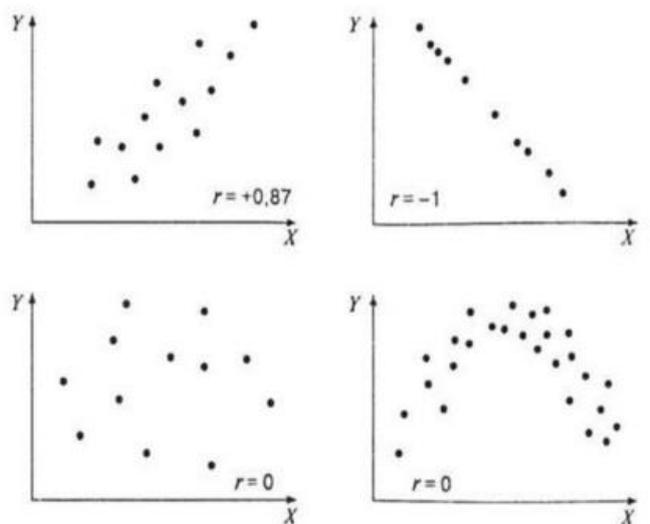




Корреляционный анализ

- **Корреляционный анализ** — это количественный метод определения тесноты и направления взаимосвязи между выборочными переменными величинами.

Примеры диаграмм
рассеивания и
коэффициентов
корреляции



МЕТОД КОРРЕЛЯЦИИ ПИРСОНА

$$r_{xy} = \frac{\Sigma (d_x \times d_y)}{\sqrt{(\Sigma d_x^2 \times \Sigma d_y^2)}}$$

R_{xy} – коэффициент корреляции

«x» и «y» – коррелируемые ряды (признаки, между которыми определяется связь)

d_x и d_y – отклонения каждого из чисел этих рядов от их средних

Σ – знак суммы