



«Математические методы в психологии»

«Коэффициенты корреляции»

*Подготовил: ст. преподаватель
Дмитриева С.Ю.*



Коэффициент корреляции r - Пирсона



Критерий корреляции Пирсона – это метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость.



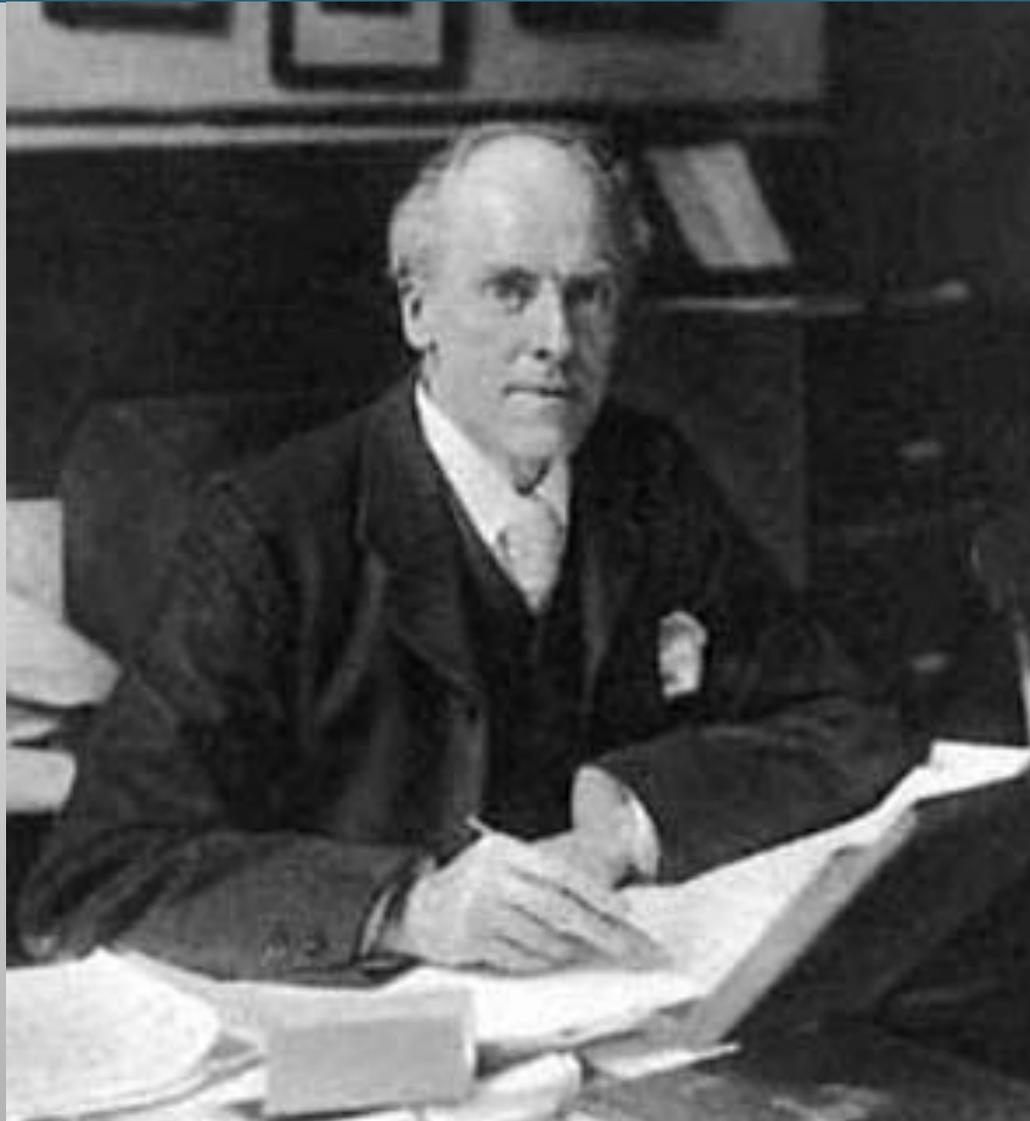
Другими словами, критерий корреляции Пирсона позволяет определить, есть ли линейная связь между изменениями значений двух переменных. В статистических расчетах и выводах коэффициент корреляции обычно обозначается как r_{xy} или R_{xy} .



История разработки критерия корреляции

Критерий корреляции Пирсона был разработан командой британских ученых во главе с **Карлом Пирсоном** (1857-1936) в 90-х годах 19-го века, для упрощения анализа ковариации двух случайных величин. Помимо Карла Пирсона над критерием корреляции Пирсона работали также **Фрэнсис Эджуорт** и **Рафаэль Уэлдон**.

PSYCHOLOGIE





Для чего используется критерий корреляции Пирсона?

Критерий корреляции Пирсона позволяет определить, какова теснота (или сила) корреляционной связи между двумя показателями, измеренными в количественной шкале. При помощи дополнительных расчетов можно также определить, насколько статистически значима выявленная связь.



Например, при помощи критерия корреляции Пирсона можно ответить на вопрос о наличии связи между температурой тела и содержанием лейкоцитов в крови при острых респираторных инфекциях, между ростом и весом пациента, между содержанием в питьевой воде фтора и заболеваемостью населения кариесом.



Условия и ограничения применения критерия хи-квадрат Пирсона

1. Сопоставляемые показатели должны быть измерены в *количественной шкале* (например, частота сердечных сокращений, температура тела, содержание лейкоцитов в 1 мл крови, систолическое артериальное давление).
2. Посредством критерия корреляции Пирсона можно определить лишь *наличие и силу линейной взаимосвязи* между величинами. Прочие характеристики связи, в том числе направление (прямая или обратная), характер изменений (прямолинейный или криволинейный), а также наличие зависимости одной переменной от другой - определяются при помощи регрессионного анализа.
3. Количество сопоставляемых величин должно быть равно двум. В случае анализ взаимосвязи трех и более параметров следует воспользоваться методом *факторного анализа*.



4. Критерий корреляции Пирсона является *параметрическим*, в связи с чем условием его применения служит *нормальное распределение* сопоставляемых переменных. В случае необходимости корреляционного анализа показателей, распределение которых отличается от нормального, в том числе измеренных в порядковой шкале, следует использовать коэффициент ранговой корреляции Спирмена.

5. Следует четко различать понятия зависимости и корреляции. Зависимость величин обуславливает наличие корреляционной связи между ними, но не наоборот.



Пример

Например, рост ребенка зависит от его возраста, то есть чем старше ребенок, тем он выше. Если мы возьмем двух детей разного возраста, то с высокой долей вероятности рост старшего ребенка будет больше, чем у младшего. Данное явление и называется *зависимостью*, подразумевающей причинно-следственную связь между показателями. Разумеется, между ними имеется и *корреляционная связь*, означающая, что изменения одного показателя сопровождаются изменениями другого показателя.



В другой ситуации рассмотрим связь роста ребенка и частоты сердечных сокращений (ЧСС). Как известно, обе эти величины напрямую зависят от возраста, поэтому в большинстве случаев дети большего роста (а значит и более старшего возраста) будут иметь меньшие значения ЧСС. То есть, *корреляционная связь* будет наблюдаться и может иметь достаточно высокую тесноту. Однако, если мы возьмем детей *одного возраста*, но *разного роста*, то, скорее всего, ЧСС у них будет различаться несущественно, в связи с чем можно сделать вывод о *независимости* ЧСС от роста.



Приведенный пример показывает, как важно различать фундаментальные в статистике понятия *связи* и *зависимости* показателей для построения верных выводов.



Как рассчитать коэффициента корреляции Пирсона?

Коэффициент линейной корреляции
Пирсона

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_i (x_i - M_x)^2 \times \sum_i (y_i - M_y)^2}}$$

Назад

Продолжить



Как интерпретировать значение коэффициента корреляции Пирсона?

Значения коэффициента корреляции Пирсона интерпретируются исходя из его абсолютных значений. Возможные значения коэффициента корреляции варьируют от 0 до ± 1 .

Чем больше абсолютное значение r_{xy} – тем выше теснота связи между двумя величинами. $r_{xy} = 0$ говорит о полном отсутствии связи. $r_{xy} = 1$ – свидетельствует о наличии абсолютной (функциональной) связи.

Если значение критерия корреляции Пирсона оказалось больше 1 или меньше -1 – в расчетах допущена ошибка.



Для оценки тесноты, или силы, корреляционной связи обычно используют общепринятые критерии, согласно которым абсолютные значения $r_{xy} < 0.3$ свидетельствуют о *слабой* связи, значения r_{xy} от 0.3 до 0.7 - о связи *средней* тесноты, значения $r_{xy} > 0.7$ - о *сильной* связи.



Более точную оценку силы корреляционной связи можно получить, если воспользоваться *таблицей Чеддока*:

бсолютное значение r_{xy}	Теснота (сила) корреляционной связи
менее 0.3	слабая
от 0.3 до 0.5	умеренная
от 0.5 до 0.7	заметная
от 0.7 до 0.9	высокая
более 0.9	весьма высокая



Оценка *статистической значимости* коэффициента корреляции r_{xy} осуществляется при помощи **t**-критерия, рассчитываемого по следующей формуле:

$$t_r = \frac{r_{xy} \sqrt{n - 2}}{\sqrt{1 - r_{xy}^2}}$$



Полученное значение t_r сравнивается с критическим значением при определенном уровне значимости и числе степеней свободы $n-2$. Если t_r превышает $t_{\text{крит}}$, то делается вывод о статистической значимости выявленной корреляционной связи.



Коэффициент корреляции r-Спирмена (ранговый)



Коэффициент ранговой корреляции Спирмена – это непараметрический метод, который используется с целью статистического изучения связи между явлениями.

В этом случае определяется фактическая степень параллелизма между двумя количественными рядами изучаемых признаков и дается оценка тесноты установленной связи с помощью количественно выраженного коэффициента.

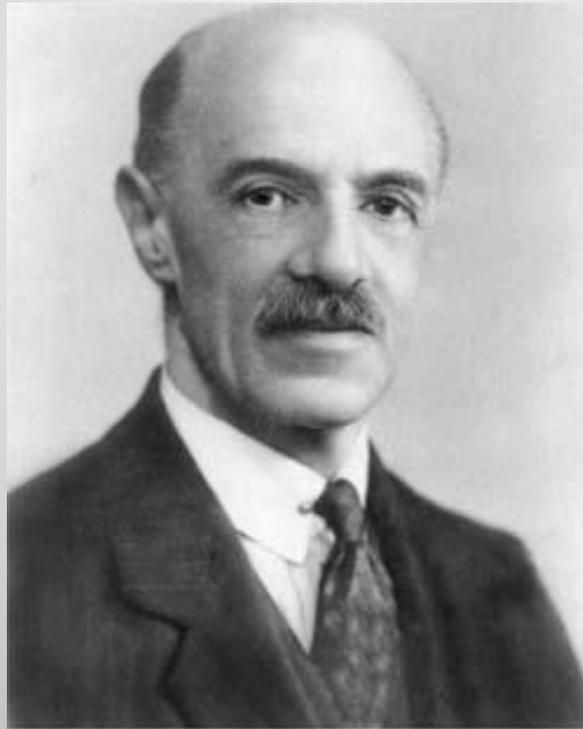


История разработки коэффициента ранговой корреляции

Данный критерий был разработан и предложен для проведения корреляционного анализа в 1904 году **Чарльзом Эдвардом Спирменом**, английским психологом, профессором Лондонского и Честерфилдского университетов.



Чарльз Эдвард Спирмен





Для чего используется коэффициент Спирмена?

Коэффициент ранговой корреляции Спирмена используется для выявления и оценки тесноты связи между двумя рядами сопоставляемых *количественных показателей*.

В том случае, если ранги показателей, упорядоченных по степени возрастания или убывания, в большинстве случаев совпадают (большему значению одного показателя соответствует большее значение другого показателя - например, *при сопоставлении роста пациента и его массы тела*), делается вывод о наличии **прямой** корреляционной связи.



Если ранги показателей имеют противоположную направленность (большему значению одного показателя соответствует меньшее значение другого - например, *при сопоставлении возраста и частоты сердечных сокращений*), то говорят об **обратной** связи между показателями.



Коэффициент корреляции Спирмена обладает следующими свойствами:

1. Коэффициент корреляции может принимать значения от минус единицы до единицы, причем при $r_s=1$ имеет место строго прямая связь, а при $r_s=-1$ – строго обратная связь.
2. Если коэффициент корреляции отрицательный, то имеет место обратная связь, если положительный, то – прямая связь.
3. Если коэффициент корреляции равен нулю, то связь между величинами практически отсутствует.
4. Чем ближе модуль коэффициента корреляции к единице, тем более сильной является связь между измеряемыми величинами.



В каких случаях можно использовать коэффициент Спирмена?

В связи с тем, что коэффициент является методом *непараметрического анализа*, проверка на нормальность распределения не требуется.

Сопоставляемые показатели могут быть измерены как в *непрерывной шкале* (например, число эритроцитов в 1 мкл крови), так и в *порядковой* (например, баллы экспертной оценки от 1 до 5).



Эффективность и качество оценки методом Спирмена снижается, если разница между различными значениями какой-либо из измеряемых величин достаточно велика. Не рекомендуется использовать коэффициент Спирмена, если имеет место неравномерное распределение значений измеряемой величины.



Как рассчитать коэффициент Спирмена?

Расчет коэффициента ранговой корреляции Спирмена включает следующие этапы:

1. Сопоставить каждому из признаков их порядковый номер (*ранг*) по возрастанию или убыванию.
2. Определить *разности рангов* каждой пары сопоставляемых значений (*d*).
3. Возвести в *квадрат* каждую разность и *суммировать* полученные результаты.
4. Вычислить коэффициент корреляции рангов по формуле:



РАНГОВЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ СПИРМЕНА

Применяется в том случае, когда значения признаков заданы в порядковой шкале (оценки судей в баллах, места на соревнованиях и пр.)

$$\rho = 1 - 6 \cdot \frac{\sum_{i=1}^{i=n} D_i^2}{n \cdot (n^2 - 1)}$$

- где $D_i = (R_{x_i} - R_{y_i})$ – это разности рангов i -й пары значений X_i и Y_i ,
- n – объем выборки (количество пар)



Определить *статистическую значимость* коэффициента при помощи **t-критерия**, рассчитанного по следующей формуле:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$



Как интерпретировать значение коэффициента Спирмена?

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее - показателями слабой тесноты связи; значения более 0,4, но менее 0,7 - показателями умеренной тесноты связи, а значения 0,7 и более - показателями высокой тесноты связи.



Статистическая значимость полученного коэффициента оценивается при помощи t-критерия Стьюдента. Если рассчитанное значение t-критерия меньше табличного при заданном числе степеней свободы, статистическая значимость наблюдаемой взаимосвязи - отсутствует. Если больше, то корреляционная связь считается статистически значимой.

PSYCHOLOGIE



Коэффициент корреляции «тао» Кендалла



- Коэффициент корреляции рангов Кендалла τ : расчет также начинается с ранжирования значений признаков X и Y .
- Ранги $X (R_x)$ располагают строго в порядке возрастания и параллельно записывают соответствующее каждому R_x значение R_y .
- Каждому R_y определяют число следующих за ним рангов, превышающих его значение, и число рангов, меньших по значению.
- Первые («правильное» следование) учитываются как баллы со знаком «+», и их сумма обозначается буквой P .
- Вторые («неправильное» следование) учитываются как баллы со знаком «-», и их сумма обозначается буквой Q .
- Коэффициент корреляции рангов Кендала равен:

$$\tau = \frac{S}{n(n-1)/2} = \frac{2S}{n(n-1)} \quad S = P - Q$$

