КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Лабораторная работа

1. Функциональная и корреляционная зависимость

- Функциональная зависимость отражает чёткую однозначную зависимость, при которой изменение какого-либо одного фактора неизбежно приводит к однозначному изменению другого.

- Подобные связи характерны для «точных» наук (математика, химия, физика) и задаются, как правило, в виде формул, таблиц, графиков:

- Известно, что повышение температуры на 10^{o} C ускоряет химическую реакцию в два раза;

- Площадь круга однозначно определяется величиной его радиуса по формуле: $S = \pi r^2$

- В педагогике и биологии подобные связи если и наблюдаются, то в самых общих вариантах и в пределах определённых границ.

- Известно, например, что между ростом (длиной тела) и массой тела существует положительная связь: более высокие индивиды имеют обычно и большую массу, чем индивиды низкого роста. Тоже наблюдается и в отношении качественных признаков: блондины, как правило, имеют голубые, а брюнеты – карие глаза.

- Причина таких «исключений» в том, что каждый биологический признак, выражаясь математическим языком, является функцией многих переменных: на его величине сказывается влияние и генетических и средовых факторов, в том числе и случайных, что вызывает варьирование признаков.

- Если из множества значений аргумента X одному значению соответствуют множество значений Y на конечном интервале значений, то такая взаимосвязь называется **корреляционной.**

- Термин «корреляция» (от лат. Correlatio – соотношение, связь) впервые применил Ж. Кювье в труде «Лекции по сравнительной анатомии» (1806 г.).
- Математическое обоснование метода было дано в 1846 году другим французским учёным Огюстом Браве

Различают корреляции нескольких направлений:

- Прямая положительная корреляция, при которой увеличение причинного фактора вызывает увеличение следственного фактора; например, увеличение силы мышц разгибателей ног положительно сказывается на росте результатов в прыжках в высоту с разбега.

- Прямая отрицательная корреляция, при которой уменьшение причинного фактора вызывает уменьшение следственного фактора; например, уменьшение длины дистанции приводит к сокращению времени её преодоления.

- Обратная положительная корреляция, при которой уменьшение причинного фактора вызывает увеличение следственного фактора; например, уменьшение длины дистанции приводит к увеличению скорости бега.

- Обратная отрицательная корреляция, при которой увеличение причинного фактора вызывает уменьшение следственного; например, увеличение силы мышц может привести к уменьшению скорости их сокращения.

2. Корреляционное поле

- Графическое представление о корреляционной зависимости называется **корреляционным полем**.

- Для построения корреляционного поля в обычной системе координат наносятся точки с координатами (X, Y) в соответствии с исходными данными.

Рост, см	175	183	188	192	180	178	185	190	180	186
Вес, кг	74	72	74	76	75	77	70	82	65	80

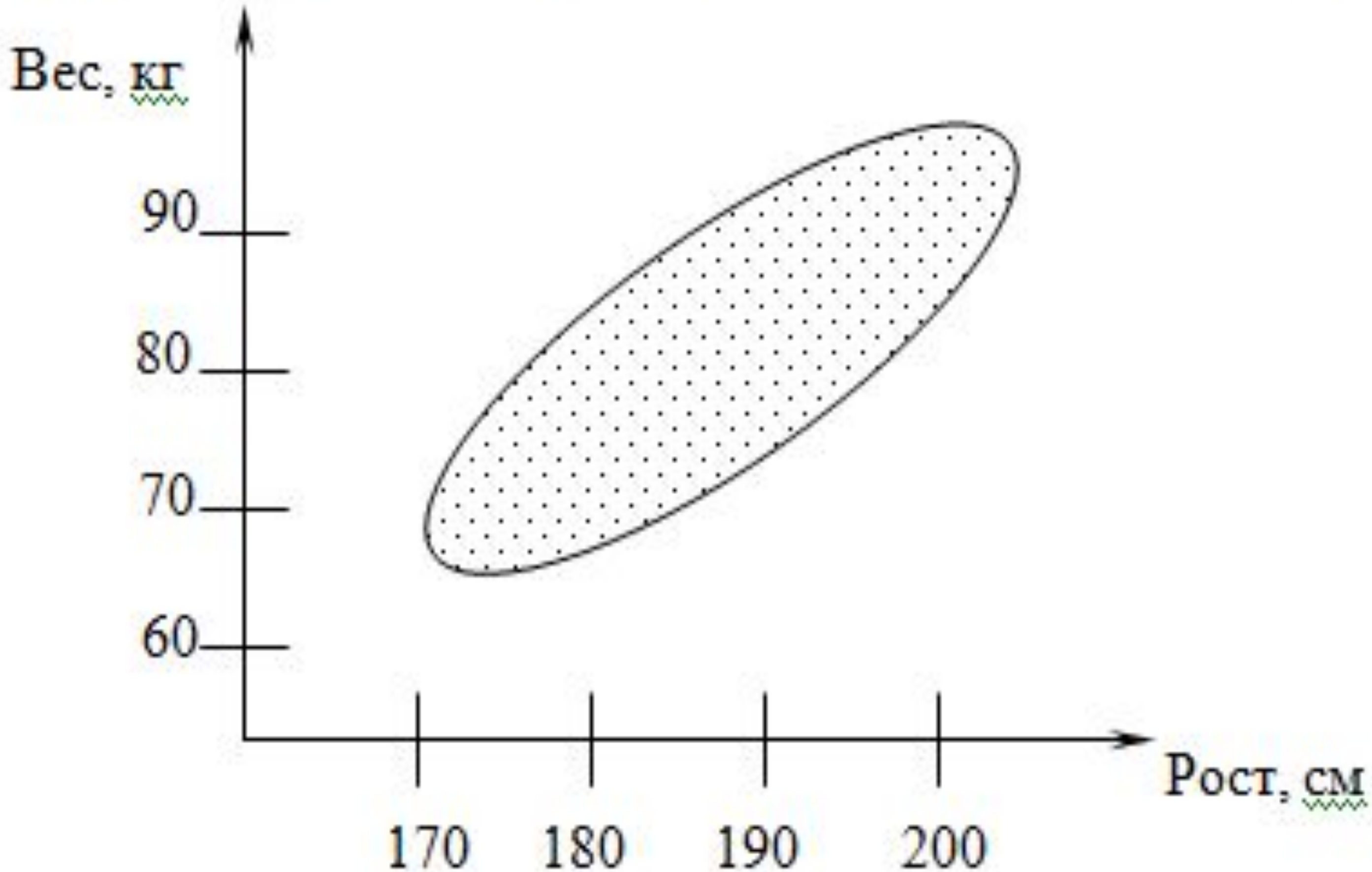

Рис. 1. График корреляционного поля, показывающего зависимость между значениями роста и веса тела студентов

3. Коэффициент корреляции

- Коэффициент корреляции (r)– показатель тесноты взаимосвязи между парой показателей, получивший широкое применение в практике.

- Количественную меру коэффициента корреляции принято различать по нескольким уровням:

Слабая связь – при /r/ < /0,30/

Средняя связь – при /0,31/ < /r/ < /0,69/

Сильная связь – при /0,70/ < /r/ < /0,99/

- Качественный анализ коэффициента корреляции принято различать по характеру взаимосвязи:

  Отрицательная связь – при $r < 0$

  Положительная связь – $0 < r$

  При $r=0$ – взаимосвязь отсутствует.

- Развитие теории корреляции и применение её к изучению наследственности, и изменчивости количественных признаков связанно с именами Ф. Гальтона, К. Пирсона и др. Формула Браве – Пирсона для расчёта прямолинейного, эмпирического коэффициента

$$r = \sum_{i=1}^{n}(x_i - x_{сред})(y_i - y_{сред}) \,/\, \Delta\sigma_x\,\sigma_y,$$

- Позднее формула была видоизменена Пирсоном, что упростило вычислительную работу:

$$r = \sum_{i=1}^{n}(x_i - x_{сред})(y_i - y_{сред}) \;/\; \sqrt{\sum(x_i - x_{сред})^2 \sum(y_i - y_{сред})^2}$$

- Функция расчёта параметрического коэффициента корреляции по формуле Браве – Пирсона: функция «CORREL» (OpenOffice.orgCalc), функция «КОРРЕЛ» (Microsoft Excel).

- Применение параметрического коэффициента корреляции Браве - Пирсона возможно при условии нормального распределения вариант X и Y(симметричные гистограммы).

Результат вычисления коэффициента корреляции позволяет отвечать на три вопроса:

- Имеется ли взаимосвязь между двумя величинами?
- Какова направленность этой взаимосвязи (прямо или обратно пропорциональная)?
- Какова теснота взаимосвязи?

- Цель корреляционного анализа – установить, можно ли значения одного показателя предсказывать по значениям другого.

Задачи корреляционного анализа:

- Установить, надёжны ли исходные данные при оценке корреляции.
- Установить, имеет ли она практическое значение.

Вывод

- Если величина коэффициента корреляции по модулю больше или ровна 0,7 , то говорят, что корреляция, имеет практическое значение, если значение меньше 0,7 , то корреляция не имеет практического значения.

Коэффициент ранговой корреляции

- Ранговая корреляция Спирмена (рангов) является одним из наиболее простых способов установления меры связи между факторами. Само название метода указывает на то, что связь определяется между рангами, т. е. рядами полученных количественных значений, ранжированных в убывающем или возрастающем порядке

- Ранжируя попарно связанные значения признаков, если возрастающим значениям одного признака (X) соответствуют значения другого (Y), то между ними существует положительная связь, если же при возрастающих значениях одного признака значения другого последовательно уменьшаются, это покажет на отрицательную связь между ними.

Коэффициент ранговой корреляции (Спирмена):

$$r = 1 - 6 \sum^{n}_{i=n} d_i^2 / n\,(n^2 - 1)$$

№	Рост, см	Вес , кг	Ранг роста, ν_I	Ранг веса, ν_i	Разность рангов $d_i = \nu_{I\,1} - \nu_{I\,2}$	Квадрат разности рангов d^2
1	2	3	4	5	6	7
1.	175	74	1	4,5	-3,5	12,25
2.	183	72	5	3	2	4
3.	188	74	8	4,5	3,5	12,25
4.	192	76	10	7	3	9
5.	180	75	3,5	6	-2,5	6,25
6.	178	77	2	8	-6	36
7.	185	70	6	2	-1	16
8.	190	82	9	10	-1	8
9.	180	65	3,5	1	2,5	6,25
10	186	80	7	9	-2	4

$\sum 107{,}00$

$$r = 1 - 6\sum d_i^2/n\ (n^2 - 1) = 10(100-1) = 1 - 0{,}6 = 0{,}4$$

- Функция определения рангов: функция «RANK» (OpenOffice.orgCalc), функция «РАНГ» (Microsoft Excel).

- Во-первых, ранговую корреляцию не рекомендуется проводить, если связанных пар меньше четырёх и больше двадцати;
- Во-вторых, ранговая корреляция позволяет устанавливать связь и в том случае, если значения носят, так сказать, полу количественный характер, т. е. не имея числовых выражений отражают чёткий порядок следования этих величин;
- В-третьих, ранговую корреляцию целесообразно применять в тех случаях, когда достаточно получить лишь приблизительную информацию.

- Таким образом, ранговый коэффициент корреляции – это непараметрический, а порядковый показатель, который позволяет измерить степень взаимосвязи между признаками независимо от закона распределении и формы связи.

- Коэффициент корреляции рангов позволяет измерить тесноту связи между такими признаками, которые не поддаются непосредственному количественному измерению, но могут быть выражены баллами и другими условными единицами, позволяющими ранжировать выборку.