Парный корреляционный анализ

Лектор: доцент Мирзаев С.С.

План:

1. Основные задачи статистического анализа
2. Коэффициент ковариации
3. Коэффициент корреляции
4. Коэффициент детерминации

Основные задачи статистического анализа

- Выявление наличия или отсутствия взаимосвязи между изучаемыми факторами (корреляционный анализ)
- Определение вида взаимосвязи между изучаемыми факторами (регрессионный анализ)
- Проверка гипотезы о виде взаимосвязи между факторами

Функциональная, статистическая и корреляционная зависимости

- Две случайные величины могут быть связаны либо ***функциональной*** зависимостью, либо зависимостью другого рода, называемой ***статистической***, либо быть ***независимыми***.
- *Функциональная* зависимость: Если каждому значению случайной величины ***X*** соответствует одно возможное значение случайной величины ***Y***, то ***Y*** называют функцией случайного аргумента ***X***:

$$\boldsymbol{Y = \varphi(X)}$$

Примеры (функциональная зависимость)

1. $y = ax + b$
2. $y = x^2$
3. $y = \sin(x)$

Статистическая зависимость

- *Статистической* называют зависимость, при которой изменение одной из величин влечет изменение ***распределения*** другой
- В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется ***среднее*** значение другой; в этом случае статистическую зависимость называют ***корреляционной***

Пример

Зависимость между урожайностью хлопка и количеством вносимых удобрений является статистической. Действительно, увеличение количества вносимых удобрений приводит к увеличению урожайности хлопка в среднем

Корреляционный анализ

Пусть изучается зависимость между факторами ***X*** и ***Y***. В результате ***n*** независимых опытов получены ***n*** пар чисел:

$$(x_1; y_1), (x_2; y_2), \ldots, (x_n; y_n)$$

X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

Коэффициент ковариации

Является мерой взаимосвязи между факторами:

$$\operatorname{cov}(x, y) = \frac{1}{n} \sum \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right)$$

Пример 1

Найти коэффициент ковариации:

Количество внесенных удобрений	150	160	180	200	250
Урожайность	25	24	26	27	28

$$\bar{x} = (150 + 160 + 180 + 200 + 250)/5 = 188$$

$$\bar{y} = (25 + 24 + 26 + 27 + 28)/5 = 26$$

$$\mathrm{cov}(x, y) = ((150-188)(25-26) + (160-188)(24-26) + (180-188)(26-26) +$$
$$(200-188)(27-26) + (250-188)(28-26))/5 = 46$$

Тесноту связи между двумя взаимозависимыми рядами характеризует *коэффициент линейной корреляции*, который показывает, существует ли и насколько велика связь изучаемых явлений

$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 \sum (y - \overline{y})^2}}$$

Пример 2

Найти коэффициент корреляции:

Количество внесенных удобрений	150	160	180	200	250
Урожайность	25	24	26	27	28

	Количество внесенных удобрений	Урожайность	x-xcp	y-ycp	(x-xcp)(y-ycp)	(x-xcp)^2	(y-ycp)^2	r	D
1	150	25	-38	-1	38	1444	1		
2	160	24	-28	-2	56	784	4		
3	180	26	-8	0	0	64	0		
4	200	27	12	1	12	144	1		
5	250	28	62	2	124	3844	4		
Сумма	940	130			**230**	**6280**	**10**		
Среднее	**188**	**26**			**46**		250,599	**0,92**	**0,84**

Интерпретация коэффициента корреляции

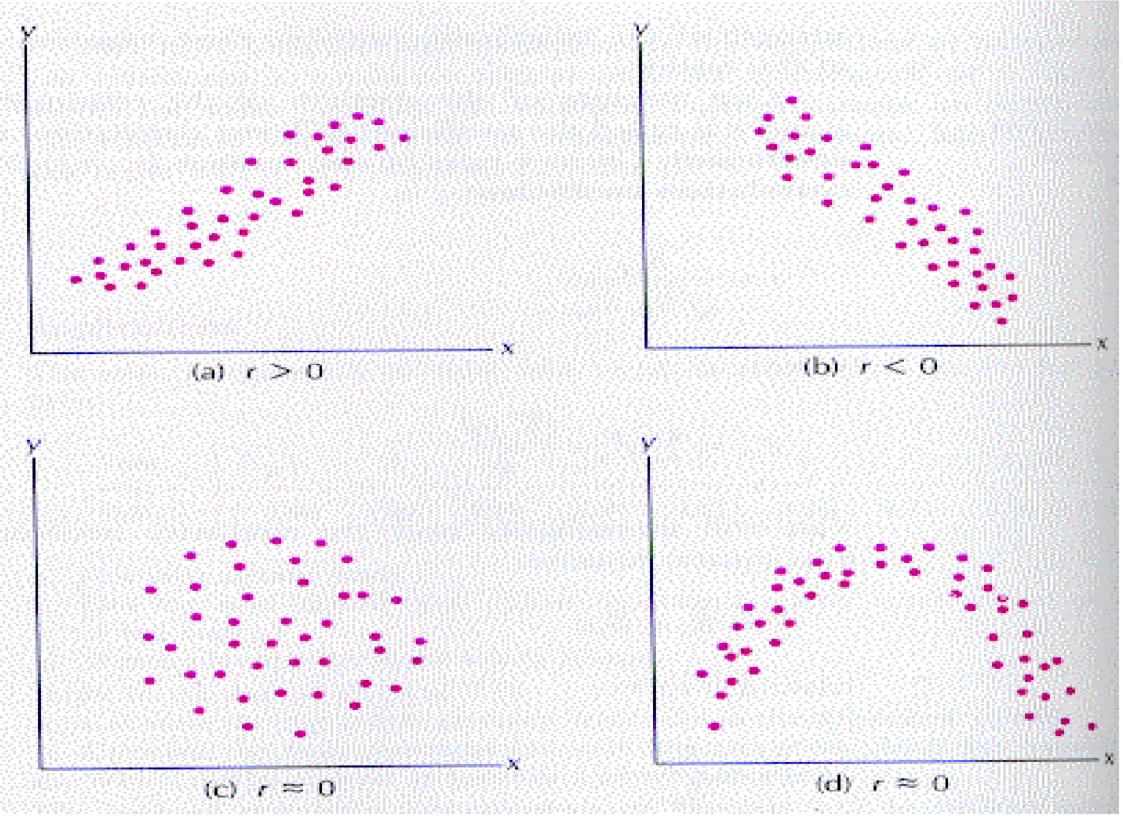
Коэффициент корреляции

- Коэффициент корреляции r может принимать значения от -1 до 1. Если $r<0$, то связь обратная, если $r>0$, то связь прямая

Коэффициент корреляции

Чем ближе значение коэффициента корреляции к ***-1*** или ***+1***, тем теснее связь между факторами, и наоборот, чем ближе значение коэффициента корреляции к ***0***, тем слабее связь между факторами

Свойства r



Коэффициент корреляции

- Коэффициент корреляции указывает следующую степень связи:
- $0 \div \pm 0,15$ – отсутствие связи
- $\pm 0,16 \div \pm 0,20$ – плохая
- $\pm 0,21 \div \pm 0,30$ – слабая
- $\pm 0,31 \div \pm 0,40$ – умеренная
- $\pm 0,41 \div \pm 0,60$ – средняя
- $\pm 0,61 \div \pm 0,80$ – высокая
- $\pm 0,81 \div \pm 0,9$ – очень высокая
- $\pm 0,91 \div \pm 1,0$ – полная

Коэффициент детерминации

При анализе взаимосвязи между факторами также вычисляют *коэффициент детерминации (r-квадрат)*

$$D = r^2$$

Коэффициент детерминации

Коэффициент детерминации показывает, какое влияние оказывают выбранные факторы на результативный показатель

Пример 3

Например, если ***r = 0,92*** (***r*** - коэффициент корреляции), то ***D=0,84***, то есть величина результативного показателя на ***84% (0,84·100)*** зависит от изменения исследуемых факторов и на ***16%*** от остальных факторов

При анализе тесноты связи между переменной и несколькими факторами определяют *коэффициент множественной корреляции* по формуле

$$R = \sqrt{1 - \frac{\sum_i (y_i - \overline{Y}_i)^2}{\sum_i (y_i - \overline{y}_i)^2}}$$

Задание

- Скачать файл «Практическое занятие 5» из MOODLE и выполнить приведенные задания
- Решение задания загрузить в MOODLE

СПАСИБО ЗА ВНИМАНИЕ!

☎ + 998 71 237 1948

✉ smirzaev@tiiame.uz