ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Пусть некоторый объект характеризуется двумя признаками Х и Y.

Между X и Y могут существовать различные виды зависимостей:

**Функциональная** – каждому значению признака X соответствует единственное значение признака Y. 

**Статистическая** – каждому значению признака Х соответствует статистическое распределение признака Y ( корреляционная таблица).

**Корреляционная**– частный случай статистической, когда каждому значению признака х соответствует среднее значение признака Y: ,связь между ними достаточно хорошо описывается функцией , называется уравнением регрессии у по х.

Основные задачи теории корреляции:

* 1. Оценить силу (тесноту) связи между признаками X и Y.
	2. Найти вид этой связи в виде уравнения прямой регрессии 
1. Пример 3.1: По данным корреляционной таблицы найти условные средние . Оценить тесноту связи и составить уравнение регрессии у по х. Сделать чертёж.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| х у | 5 | 7 | 9 | 11 | nx |
| 1,8 | - | - | 2 | 5 | 7 |
| 4,4 | - | 5 | 5 | - | 10 |
| 7 | - | 8 | 3 | 2 | 13 |
| 9,6 | 3 | 11 | 1 | 2 | 17 |
| 12,2 | 2 | 1 | - | - | 3 |
| nу | 5 | 25 | 11 | 9 | 50 |

Частота nху показывает сколько раз наблюдалась каждая пара значений.

Например: пара значений (9,6; 7) наблюдалась 11 раз.

В таблице каждому значению х соответствует статистическое распределение признака у.

Например для х=7

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| у | 5 | 7 | 9 | 11 |
|  nху | 0 | 8 | 3 | 2 |

 Σnxy=13
Среднее значение у при х=7 или условная средняя 

Аналогично, каждому значению у можно найти статистическое распределение по х.

Найдём условные варианты:

Ui=(xi-с1)/h1

 νi=(уi-c2)/h2

c1=7, h1=4,4-1,8=2,6 c2=7, h2=7-5=2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ui  Vi | -1 | 0 | 1 | 2 | nu | nu∙u |
| -2 | - | - | 2 | 5 | 7 | -14 |
| -1 | - | 5 | 5 | - | 10 | -10 |
| 0 | - | 8 | 3 | 2 | 13 | 0 |
| 1 | 3 | 11 | 1 | 2 | 17 | +17 |
| 2 | 2 | 1 | - | - | 3 | 6 |
| nv | 5 | 25 | 11 | 9 | 50 | ∑nu∙u=-1 |
| nv∙v | -5 | 0 | 11 | 18 | ∑nu∙u=24 |

$$\overbar{u}=\frac{\sum\_{}^{}n\_{u}u}{n}=\frac{-2∙7-1∙10+0∙13+1∙17+2∙3}{50}=-0.02$$

$$\overbar{v}=\frac{\sum\_{}^{}n\_{v}∙v}{n}=\frac{-1∙5+0∙25+1∙11+2∙9}{50}=0.48$$

Найдём вспомогательные величины









∑nnvu∙v найдём из таблицы:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| u v | -1 | 0 | 1 | 2 | V=∑nnvv | u∙v |
| -2 | - | - | 2 | 5 | 12 | -24 |
| -1 | - | 5 | 5 | - | 5 | -5 |
| 0 | - | 8 | 3 | 2 | 7 | 0 |
| 1 | 3 | 11 | 1 | 2 | 2 | 2 |
| 2 | 2 | 1 | - | - | -2 | -4 |
| u=∑nnvu | 7 | 8 | -8 | -8 |  | ∑=-31 |
| v∙u | -7 | 0 | -8 | -16 | ∑=-31 |  |

∑u∙v=∑v∙u=-31=∑nuvuv



Искомый выборочный коэффициент корреляции



Пояснения к расчётной таблице

1. В каждой строчке, где частота nuv≠0 в правом верхнем углу стоит произведение частоты nuv на варианту v.
2. Складываются все числа, полученные в правых углах одной строки, записываются в столбец v.
3. Варианту u умножают на v.
4. Складывают все числа столбца uv, получают

 ∑uv=∑nuvuv

Аналогично (для контроля) вычисляют ∑vu=∑nuvuv по столбцам.

|  |
| --- |
| Оценка тесноты связи |
| Значение ׀rв׀ | 0-0,1 | 0,1-0,3 | 0,3-0,5 | 0,5-0,7 | 0,7-0,9 | 0,9-0,99 | 1 |
| Теснота линейной связи | нет | слабая | умеренная | заметная | высокая | очень высокая | функциональная |



Т.к. rв=-0,58<0, связь обратная, т.е с ростом х убывает y.

׀rв׀Є(0,5-0,7)=>связь заметная.

Строим график функции  и отмечаем условные средние:



Замечание: Если расчёты и построения выполнены верно, то прямая регрессии y по х должна пройти вблизи всех точек , отмеченных \*



**Самостоятельно**. По данной таблице найти выборочное уравнение прямой линии регрессии Y на Х. Оценить тесноту связи. Построить эмпирическую и теоретическую линии регрессии.

|  |  |
| --- | --- |
| Y | X |
| 15 | 20 | 25 | 30 | 35 | 40 | ny |
| 30 | 1 | 6 |  |  |  |  | 7 |
| 40 |  |  | 4 |  |  | 5 | 9 |
| 50 |  | 4 | 7 | 30 | 9 |  | 50 |
| 60 |  |  | 2 | 10 | 8 |  | 20 |
| 70 | 5 |  |  |  | 6 | 3 | 14 |
| nx | 6 | 10 | 13 | 40 | 23 | 8 | n = 100 |