



Корреляционн ые зависимости

Регрессионная модель

- это функция, описывающая зависимость между количественными характеристиками сложных систем.
- Регрессионные математические модели строятся в тех случаях, когда известно, что зависимость между двумя факторами существует и требуется получить ее математическое описание.

■ Рассмотрим задачи другого рода. Пусть важной характеристикой некоторой сложной системы является фактор A . На него могут оказывать влияние одновременно многие другие факторы: B , C , D и так далее. Мы рассмотрим два типа задач:

- 1) требуется определить, оказывает ли фактор B какое-либо заметное регулярное влияние на фактор A ?
- 2) Какие из факторов B , C , D и так далее оказывают наибольшее влияние на фактор A ?

- В качестве примера сложной системы будем рассматривать школу.
- Пусть для первого типа задач фактором A является средняя успеваемость учащихся школы, фактором B — финансовые расходы школы на хозяйственные нужды: ремонт здания, обновление мебели, эстетическое оформление помещения и т. п.
- Здесь влияние фактора B на фактор A не очевидно. Наверное, гораздо сильнее на успеваемость влияют другие причины: уровень квалификации учителей, контингент учащихся, уровень технических средств обучения и другие.

- Специалисты по статистике знают, что для того, чтобы выявить зависимость от какого-то определенного фактора, нужно максимально исключить влияние других факторов.
- Проще говоря, собирая информацию из разных школ, нужно выбирать такие школы, в которых приблизительно одинаковый контингент учеников, квалификация учителей и пр., но хозяйственные расходы школ разные (у одних школ могут быть богатые спонсоры, у других — нет).

- Итак, пусть хозяйственные расходы школы выражаются количеством рублей, отнесенных к числу учеников в школе (руб./чел.), потраченных за определенный период времени (например, за последние 5 лет).
- Успеваемость же пусть оценивается средним баллом учеников школы по результатам окончания последнего учебного года.

А	В	С
№ п/п	Затраты (руб/чел.)	Успеваемость (ср.балл)
1	50	3,81
2	345	4,13
3	79	4,30
4	100	3,96
5	203	3,87
6	420	4,33
7	210	4
8	137	4,21
9	463	4,4
10	231	3,99
11	134	3,9
12	100	4,07
13	294	4,15
14	396	4,1
15	77	3,76
16	480	4,25
17	450	3,88
18	496	4,50
19	102	4,12
20	150	4,32

На рисунке
представлены
итоги сбора
данных
по 20 школам

На рисунке приведена точечная диаграмма, построенная по этим данным.



- Значения обеих величин: финансовых затрат и успеваемости учеников имеют значительный разброс и, на первый взгляд, взаимосвязи между ними не видно.
- Зависимости между величинами, каждая из которых подвергается не контролируемому полностью разбросу, называются корреляционными зависимостями.
- Мерой корреляционной зависимости является величина, которая называется *коэффициентом корреляции*.

- Раздел математической статистики, который исследует такие зависимости, называется корреляционным анализом.
- Корреляционный анализ изучает усредненный закон поведения каждой из величин в зависимости от значений другой величины, а также меру такой зависимости.

Коэффициент корреляции

- Коэффициент корреляции обозначается греческой буквой ρ («ро»).
- Это число, заключенное в диапазоне от -1 до +1 (если это число по модулю близко к 1, то имеет место сильная корреляция, если к 0, то слабая).
- Близость ρ к +1 означает, что возрастанию одного набора значений соответствует возрастание другого набора, близость к -1 означает обратное.
- Значение ρ легко найти с помощью Excel без всяких формул (разумеется, потому, что в Excel они встроены).

- В Excel функция вычисления коэффициента корреляции называется **КОРРЕЛ** и входит в группу статистических функций.
- Покажем, как ей воспользоваться. На том же листе Excel, где находится таблица, представленная на рис. 2.16, надо установить курсор на любую свободную ячейку и запустить функцию КОРРЕЛ. Она запросит два диапазона значений. Укажем B2:B21 и C2:C21. После их ввода выведется ответ; $r=0,500273843$. Эта величина говорит о среднем уровне корреляции

- Рассмотрим другой пример, в котором проводится исследование по определению зависимости успеваемости учащихся старших классов от двух факторов: обеспеченности школьной библиотеки учебниками и обеспеченности школы компьютерами.
- Нормой обеспеченности учебниками является их полный комплект, то есть такое количество, когда каждому ученику выдаются из библиотеки все нужные ему для учебы книги. Нормой обеспеченности компьютерами будем считать такое их количество, при котором на каждые четыре старшеклассника в школе приходится один компьютер. Предполагается, что компьютерами ученики пользуются не только на информатике, но и на других уроках, а также во внеурочное время.

**В таблице, приведены результаты измерения
обоих факторов в 11 разных школах.**

<i>Обеспечение учебного процесса</i>				
№	Обеспеченность учебниками (%)	Успеваемость (ср.балл)	Обеспеченность компьютерами (%)	Успеваемость (ср.балл)
1	50	3,81	10	3,98
2	78	4,15	25	4,01
3	94	4,69	19	4,34
4	65	4,37	78	4,41
5	99	4,53	45	3,94
6	87	4,23	32	3,62
7	100	4,73	90	4,6
8	63	3,69	21	4,24
9	79	4,08	34	4,36
10	94	4,2	45	3,99
11	93	4,32	67	4,5
$\rho = 0,780931$			$\rho = 0,572465$	

- Для обеих зависимостей получены коэффициенты линейной корреляции.
- Как видно из таблицы, корреляция между обеспеченностью учебниками и успеваемостью сильнее, чем корреляция между компьютерным обеспечением и успеваемостью (хотя и тот и другой коэффициенты корреляции не очень большие).
- Отсюда можно сделать вывод, что пока еще книга остается более значительным источником знаний, чем компьютер.

ВЫВОДЫ:

- Зависимости между величинами, каждая из которых подвергается не контролируемому полностью разбросу, называются корреляционными.
- С помощью корреляционного анализа можно решить следующие задачи; определить, оказывает ли один фактор существенное влияние на другой фактор; из нескольких факторов выбрать наиболее существенный.

Выводы:

- Количественной мерой корреляции двух величин является коэффициент корреляции.
- Значение коэффициента корреляции лежит между -1 и +1. Чем значение ближе по модулю к 0, тем корреляция (связь) сильнее.
- В MS Excel для определения коэффициента корреляции используется функция КОРРЕЛ из группы статистических данных.