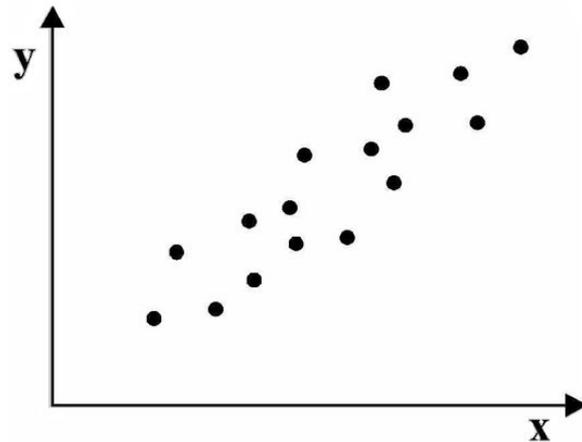


*

Классная работа

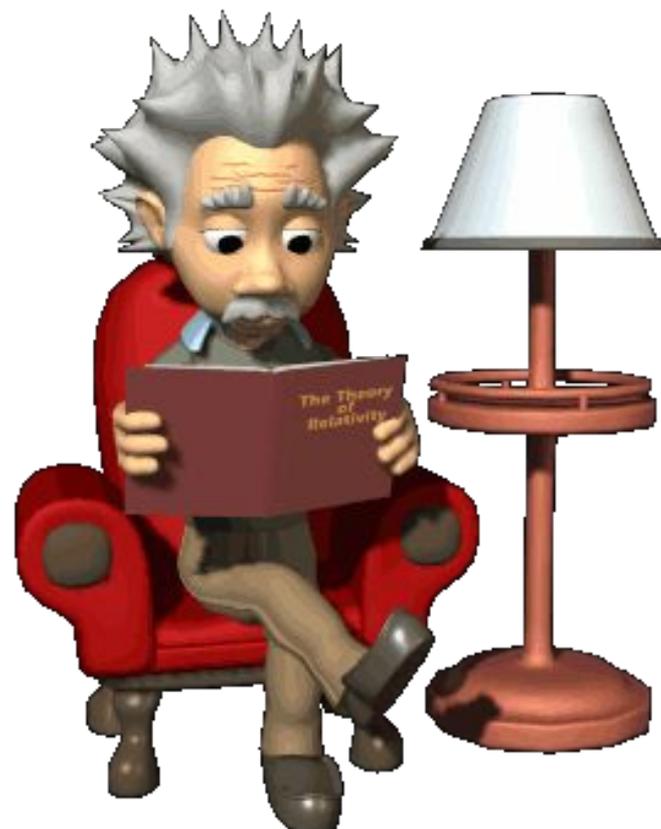
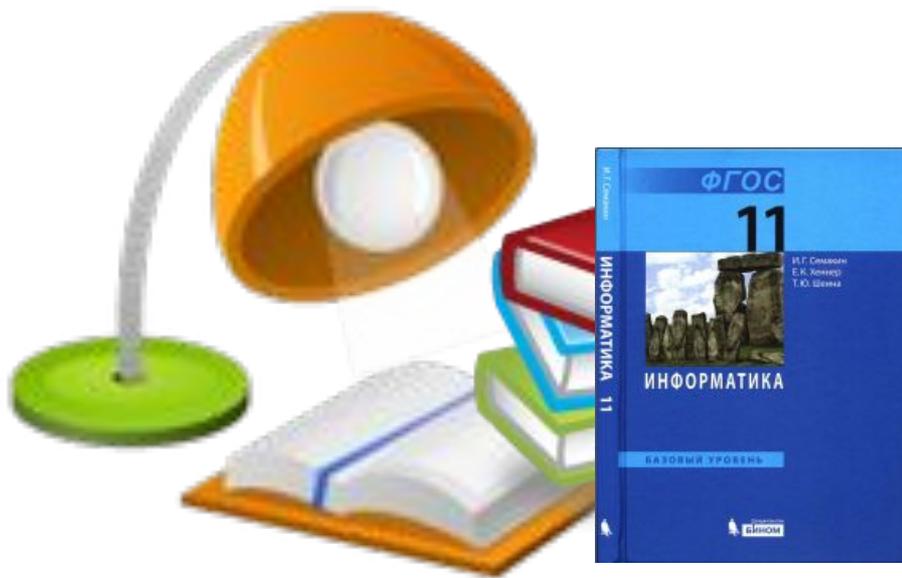
Корреляционный анализ



Урок 45

Домашнее задание

§19 (стр. 121–126).
Вопросы 1, 2 – устно.



Регрессионная модель

- описывает зависимость между количественными характеристиками.
- строится когда известно, что зависимость между двумя факторами существует.

Новые задачи моделирования

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 ?

Рассмотрим школу

Пусть

- фактор A – средняя успеваемость учащихся школы,
- фактор 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.

Хозяйственные расходы → Успеваемость

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.

Статистические данные и точечная диаграмма

(итоги сбора данных по 20 школам)



Корреляция

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**.

Корреляционный анализ

- 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).

Функция КОРРЕЛ

- В Excel имеется функция вычисления коэффициента корреляции **КОРРЕЛ**, входит в группу статистических функций.

Для вычисления коэффициента корреляции необходимо:

- на том же листе Excel, где находится таблица, надо установить курсор на любую свободную ячейку и запустить функцию **КОРРЕЛ**.
- указать два диапазона значений по которым устанавливается зависимость
- после их ввода выведется ответ. Эта величина говорит о среднем уровне корреляции

Определение коэффициента корреляции



КОРРЕЛ =КОРРЕЛ(B2:B21;C2:C21)

	A	B	C	D	E	F	G	H	I	J	K	L
	№ п/п	Затраты (руб./чел.)	Успеваемость (средний балл)									
1												
2	1	50	3,81									
3	2	345	4,13									
4	3	79	4,3									
5	4	100	3,96									
6	5	203	3,87									
7	6	420	4,33									
8	7	210	4									
9	8	137	4,21									
10	9	463	4,4									
11	10	231	3,99									
12	11	134	3,9									
13	12	100	4,07									
14	13	294	4,15									
15	14	396	4,1									
16	15	77	3,76									
17	16	480	4,25									
18	17	450	3,88									
19	18	496	4,5									
20	19	102	4,12									
21	20	150	4,32									

Аргументы функции

КОРРЕЛ

Массив1 B2:B21 = {50;345;79;100;203}

Массив2 C2:C21 = {3,81;4,13;4,3;3,96}

= 0,500273843

Возвращает коэффициент корреляции между двумя множествами данных.

Массив1 первый диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.

[Справка по этой функции](#) Значение: 0,500273843

Пример

Рассмотрим другой пример, зависимость успеваемости учащихся старших классов от двух факторов:

- обеспеченности школьной библиотеки учебниками
- 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.

Зависимость успеваемости от обеспечения уч. процесса

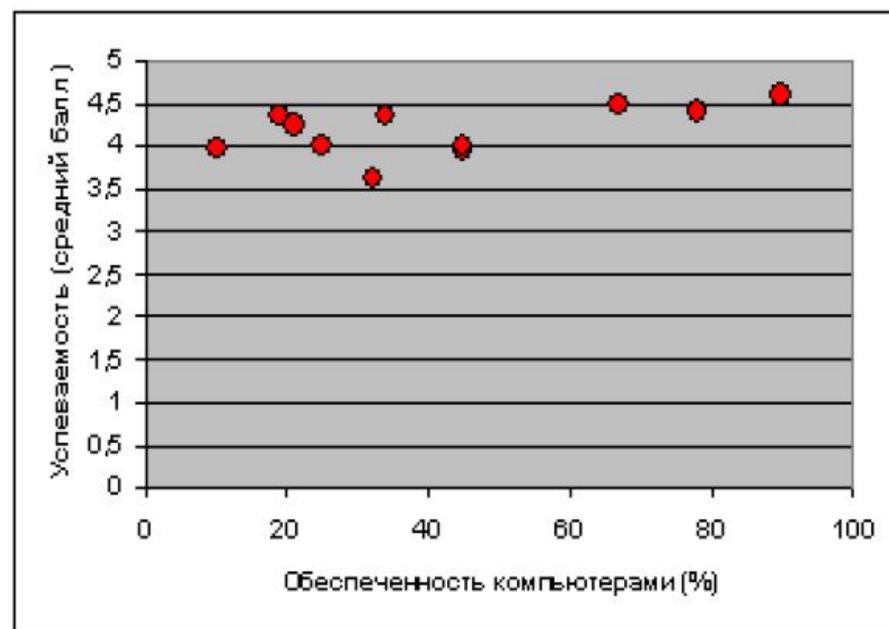
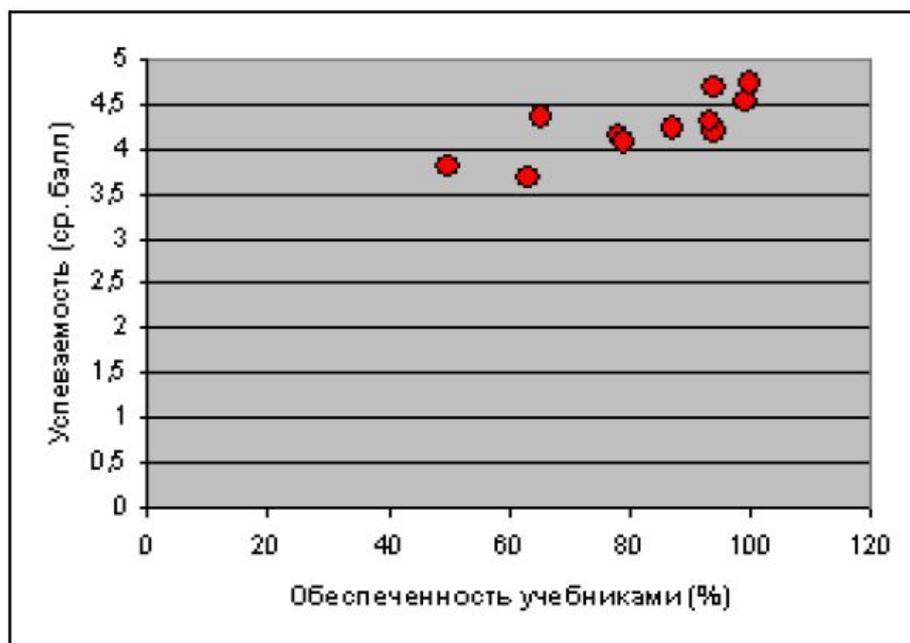
(результаты измерения обоих факторов в 11 разных школах)



	A	B	C	D	E
1	Обеспечение учебного процесса				
2	№	Обеспеченность учебниками (%)	Успеваемость (средний балл)	Обеспеченность компьютерами (%)	Успеваемость (средний балл)
3	1	50	3,81	10	3,9
4	2	78	4,15	25	4,0
5	3	94	4,69	19	4,3
6	4	65	4,37	78	4,4
7	5	99	4,53	45	3,9
8	6	87	4,23	32	3,6
9	7	100	4,73	90	4,1
10	8	63	3,69	21	4,2
11	9	79	4,08	34	4,3
12	10	94	4,2	45	3,9
13	11	93	4,32	67	4,1
14	$\rho =$	0,780931217		0,572465336	



Точечные диаграммы



Результат

- 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

С помощью корреляционного анализа можно решить следующие задачи:

- определить, оказывает ли один фактор существенное влияние на другой фактор;
- из нескольких факторов выбрать наиболее существенный.

Задание

Задание **3** (стр. **126**).

а) Для данных из таблицы, представленной на рис. **3.9**, **постройте** две линейные регрессионные модели.

б) Для этих же данных **вычислите** коэффициенты корреляции . **Сравните** с приведенными на рис. 3.9 результатами.

Работаем за компьютером

