

Методы математической статистики:

Дисперсионный анализ
Корреляционный анализ
Регрессионный анализ

Статистические методы

- Метод корреляции
 - Метод регрессии
- Дисперсионный анализ

Корреляционный анализ

ГДЕ ИСПОЛЬЗУЕТСЯ КОРРЕЛЯЦИОННЫЙ АНАЛИЗ?

Например, можно ответить на вопрос о наличии связи между температурой тела и содержанием лейкоцитов в крови при острых респираторных инфекциях, между ростом и весом пациента, между содержанием в питьевой воде фтора и заболеваемостью населения кариесом.

КОРРЕЛЯЦИОННАЯ СВЯЗЬ

Общепринятым способом
выявления взаимосвязи между
переменными является расчет
коэффициента корреляции (r).

Значение коэффициента корреляции указывает, насколько тесно связаны между собой параметры.

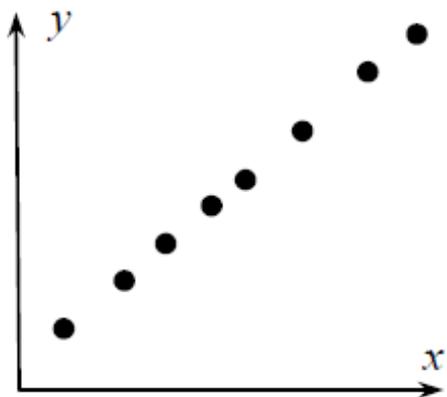
Условно выделяют следующие уровни корреляционной связи:

- $r = 1$ – функциональная зависимость;
- $0,95 \leq r < 1$ – связь очень сильная;
- $0,75 \leq r < 0,95$ – связь тесная (сильная);
 - $0,5 \leq r < 0,75$ – связь средняя (умеренная);
 - $0,2 \leq r < 0,5$ – связь слабая;
 - $0 < r < 0,2$ – очень слабая;
 - $r = 0$ – связь отсутствует.

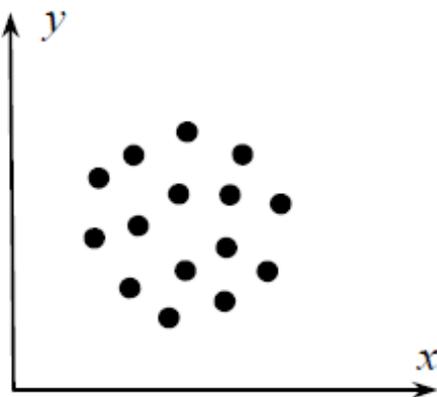
**Величина коэффициента
корреляции лежат в пределах от
-1 до +1.**

По **направлению** корреляционная
связь может быть положительной $+r$
("прямой") и отрицательной $-r$
("обратной").

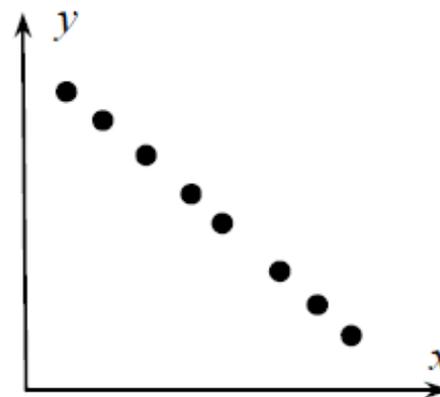
ВЕЛИЧИНЫ КОЭФФИЦИЕНТА ЛИНЕЙНОЙ КОРРЕЛЯЦИИ В РАЗЛИЧНЫХ СИТУАЦИЯХ



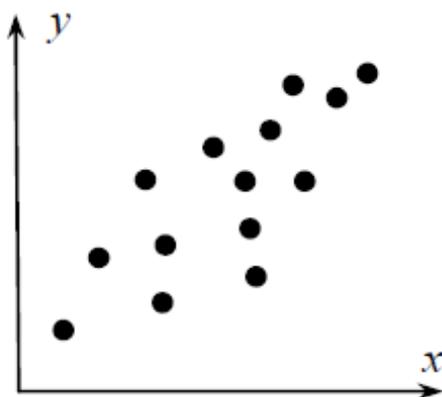
$$r(x, y) = 1$$



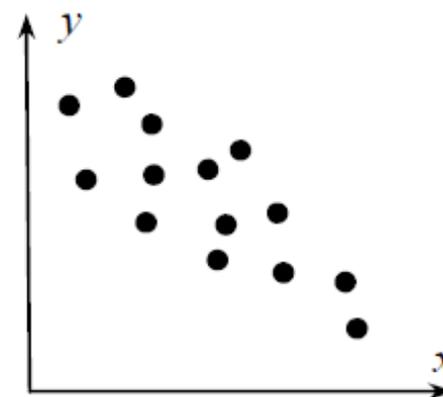
$$r(x, y) = 0$$



$$r(x, y) = -1$$



$$r(x, y) = 1/2$$



$$r(x, y) = -1/2$$

МЕТОДЫ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

✓ **Критерий корреляции Пирсона** – это метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость.

Обозначение: r_{xy}

В Excel для вычисления коэффициента корреляции Пирсона могут быть использованы следующие методы:

**ФОРМУЛЫ → ДРУГИЕ
ФУНКЦИИ → СТАТИСТИЧЕСКИЕ
→ КОРРЕЛ**

**Данные → Анализ данных →
Корреляция**

МЕТОДЫ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

✓ Коэффициент ранговой корреляции
Спирмена – это непараметрический
метод, который используется для
выявления и оценки тесноты связи между
двумя рядами сопоставляемых
количественных показателей.

Обозначение: ρ

Регрессионный анализ

-статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную.

**Задача регрессионного анализа:
установить, как количественно меняется
одна величина при изменении другой на
единицу .**

**Результатом регрессионного анализа
считается:**

- расчет коэффициента регрессии**
- построения теоретической линии регрессии**
- расчета уравнения регрессии**
- построение эмпирической линии регрессии**

ЛИНЕЙНАЯ РЕГРЕССИЯ

Общая формула уравнения линейной регрессии

$$y_x = a + bx$$

где y рассматривают в качестве зависимой переменной, или функции от другой – независимой – переменной величины x ,

a – первоначальное значение y ;

b – коэффициент пропорциональности, который показывает степень зависимости x от y и определяет наклон линии регрессии по отношению к осям прямоугольных координат (коэффициент регрессии)

В среде MS Excel для нахождения модели регрессии можно использовать несколько способов:

- 1) использовать встроенную функцию **ЛИНЕЙН**;
- 2) графический способ – построение линии тренда на диаграмме с показом уравнения регрессии;
- 3) инструмент **Регрессия** из Пакета анализа.

РЕГРЕССИОННЫЙ АНАЛИЗ В MS EXCEL

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Calibri 11

Вставить

Буфер обмена

Шрифт Выравнивание

Общий Вставить Σ

A16

	A	B	C	D
1	N	Время, мин	Толщина рубца, мм	
2	1	2,4	17	
3	2	0,6	3	
4	3	1,7	12	
5	4	1	5	
6	5	1,6	8	
7	6	1,5	9	
8	7	1,8	14	
9	8	3	20	
10				
11	Имеются результаты проводившиеся у 8 больных эффективной криодеструкции			
12	кожных рубцов различной толщины			
13				
14	Необходимо рассмотреть возможность на основании этих данных определять			
15	предполагаемое время криодеструкции			
16				

Лист1 Лист2 Лист3

Готово 100%

Линейный регрессионный анализ заключается в подборе графика и его уравнения для набора наблюдений

Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных

Книра1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Calibri 11

Общий

Вставить Удалить Формат

Буфер обмена Шрифт Выравнивание Число Ячейки

Сортировка и фильтр Найти и выделить

D1 Вывод ИТОВ

	A	B	C	D	E	F	G
3	2	0,6	3	<i>Регрессионная статистика</i>			
4	3	1,7	12	Множественный R	0,972338443		
5	4	1	5	R-квадрат	0,945442047		
6	5	1,6	8	Нормированный K-квадрат	0,936349055		
7	6	1,5	9	Стандартная ошибка	0,189278601		
8	7	1,8	14	Наблюдения	8		
9	8	3	20				
10				<i>Дисперсионный анализ</i>			
11					<i>df</i>	<i>SS</i>	<i>MS</i>
12				Регрессия	1	3,725041667	3,725
						0,214958333	0,035
						3,94	
						<i>Стандартная ошибка</i>	<i>t-статистика</i>
						0,150135837	2,199
						0,012217881	10,199

100%

Коэффициент детерминации (R-квадрат) оценивает «силу влияния» переменной x на y .

- R-квадрат $> 0,95$ – высокая точность аппроксимации;
- $0,8 < \text{R-квадрат} < 0,95$ – удовлетворительная;
- R-квадрат $< 0,6$ – точность не достаточна и модель требует улучшения.

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Разработчик

Получить внешние данные Обновить все Подключения

Сортировка Фильтр Сортировка и фильтр

Очистить Применить повторно Дополнительно

Текст по столбцам Удалить дубликаты Работа с данными

Структура Анализ данных Анализ

C1 *f* ВЫВОД ИТОГОВ

	B	C	D	E	F	G	H
4	5	Множественный R	0,972338443				
5	8	R-квадрат	0,945442047				
6	9	Нормированный R-квадрат	0,936349055				
7	14	Стандартная ошибка	0,189278601				
8	20	Наблюдения	8				
9							
10		Дисперсионный анализ					
11			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12		Регрессия	1	3,725041667	3,725041667	103,974801	5,18222E-05
13		Остаток	6	0,214958333	0,035826389		
14		Итого	7	3,94			
15							
16		<i>Коэффициент a</i>	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
17		Y-пересечение	0,329583333	0,150135837	2,195234256	0,070563987	-0,037785826
18		Переменная X 1	0,124583333	0,0349	5,18222E-05	0,094687255	
19							
20							
21							

Лист1 Лист2 Лист3

Готово Среднее: 5,017902256 Количество: 57 Сумма: 155,5549699 100%

Значимость F не должна превышать 0,05. Значимость F говорит о высокой достоверности результатов и отсутствии случайности и наличия оправданной в нашем исследовании закономерности.

С учетом коэффициентов уравнение регрессии будет иметь вид:

Коэффициент *a*

Коэффициент *b*

Время криодеструкции (*Y*) = 0,33 + 0,125 * *X* (толщина рубца в мм)

Зависимая переменная

Независимая переменная

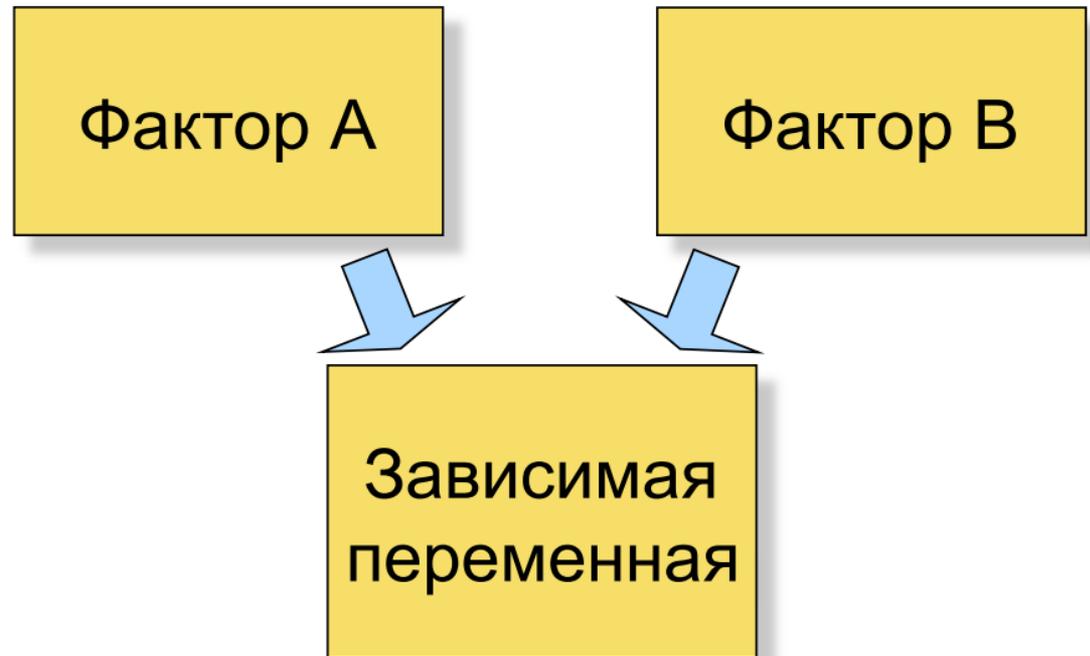
Дисперсионный анализ

**статистический метод,
применяемый для выявления
влияния отдельных факторов
(количественных, порядковых или
качественных) на изучаемый
признак и оценку степени этого
влияния.**

Однофакторный анализ



Двухфакторный анализ



Сущность дисперсионного анализа заключается в установлении роли отдельных факторов в изменчивости того или иного признака



Дисперсионный анализ позволяет учесть и при необходимости исключить случайные факторы

В среде MS Excel для выполнения дисперсионного анализа используют следующий метод:

**Данные → Анализ данных →
Однофакторный
дисперсионный анализ /
Двухфакторный
дисперсионный анализ**

ДИСПЕРСИОННЫЙ АНАЛИЗ В MS EXCEL

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Calibri 11

Ж К Ч А А

Вставить Буфер обмена Шрифт Выравнивание

	A	B	C	D	E
1		Без шума	6 дБ	12 дБ	
2	Процент правильных ответов	78,6	61,9	45,2	
3		95,2	97,6	97,6	
4		83,3	61,9	80,9	
5		85,7	73,8	62,4	
6		80,4	75,6	70,6	
7		90,2	68,8	69,2	
8					
9					
10					
11					
12					
13	В ходе исследования возрастных изменений слуховой функции у детей использовался тест определения				
14	эмоциональной составляющей речи в условиях маскировки шумом				
15	Необходимо определить влияет ли уровень шума на правильность распознавания эмоциональной				
16	составляющей речи у детей				

Лист1 Лист2 Лист3

Готово 100%

Дисперсионный анализ предназначен для исследования задачи о действии на измеряемую величину (отклик) одного или нескольких факторов

В одно-, двухфакторном и т.д. анализе влияющие на результат факторы известны, и речь идет только о выяснении существенности или оценке этого влияния

дисперсионный_пример.xls [Режим совместимости] - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Разработчик

Calibri 11

Общий

Вставить Удалить Формат Ячейки

Сортировка и фильтр Найти и выделить

Буфер обмена Шрифт Выравнивание Число

E2 fx Однофакторный дисперсионный анализ

	E	F	G	H	I	J	K
5	<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
6	Строка 1	3	185,7	61,9	278,89		
7	Строка 2	3	290,4	96,8	1,92		
8	Строка 3	3	226,1	75,36666667	137,45333333		
9	Строка 4	3	221,9	73,96666667	135,74333333		
10	Строка 5	3	226,6	75,53333333	24,01333333		
11	Строка 6	3	228,2	76,06666667	149,85333333		
12							
13							
14	Дисперсионный анализ						
15	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критичес</i>
16	Между группами	1902,022778	5	380,4045556	3,135747978	0,048634425	3,105875
17	Внутри групп	1455,746667	12	121,3122222			
18							
19	Итого	3357,769444	17				

В случаях, когда P-Значение < 0.05, критерий Фишера значим, и влияние исследуемого фактора можно считать доказанным.