

Корреляционный анализ

Корреляция

(от лат. *correlatio* «взаимосвязь»)
или корреляционная зависимость —
статистическая взаимосвязь двух или
более случайных величин (либо
величин, которые можно с некоторой
допустимой степенью точности считать
таковыми)

Корреляционный анализ

Линейная или нелинейная корреляция

А	Б
2	7
4	10
5	9
6	12
10	16
14	20
20	25
22	27
29	35
32	36
40	45

А	Б
2	5
4	15
5	25
6	35
10	100
14	200
20	402
22	450
29	950
32	1100
40	1550

Корреляционный анализ

Прямая (+) или обратная (-) корреляция

Уровень гемоглобина	Содержание железа
10,2	4
10,3	4,1
10,3	4,2
10,5	4,3
10,5	4,5
10,6	4,6
10,6	4,6
10,7	4,6
10,9	4,8
11	4,9
11,2	5
11,3	5,1
11,5	5,2

$$r = 0,094$$

Возраст новорожденного, дни	ЧСС
1	175
2	170
3	165
5	160
8	160
10	158
12	155
15	149
29	135
32	132
40	135

$$r = - 0,096$$

Корреляционный анализ

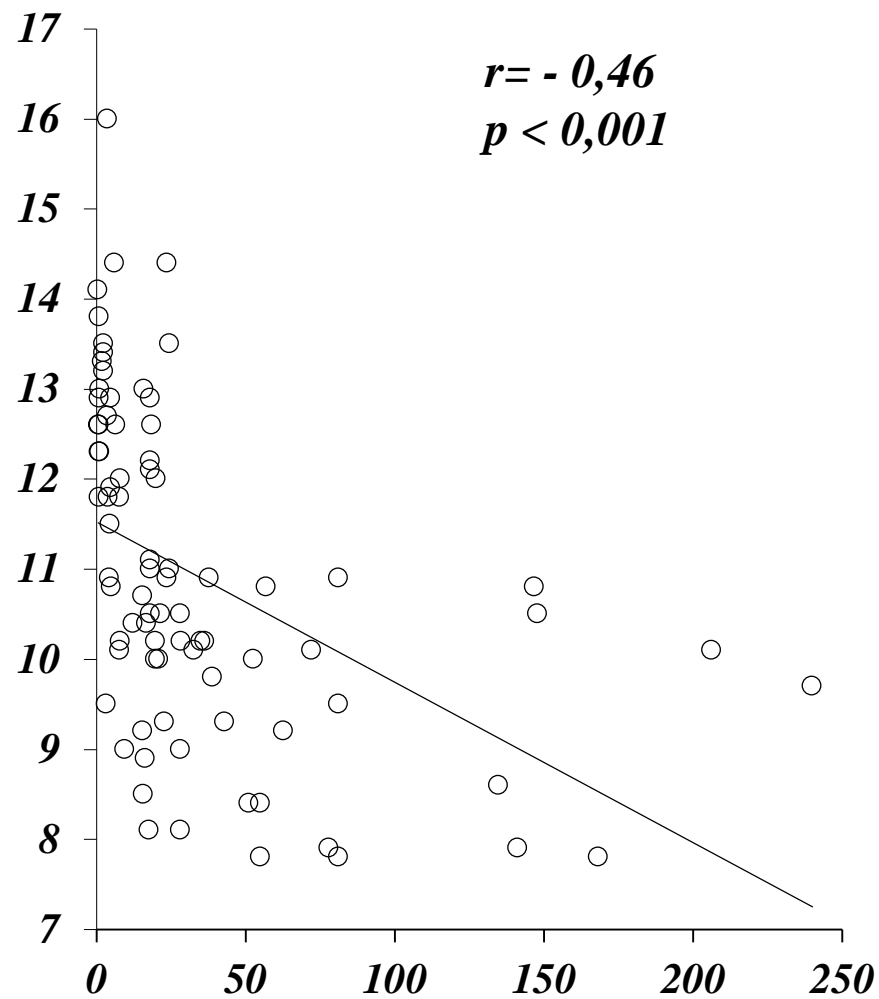
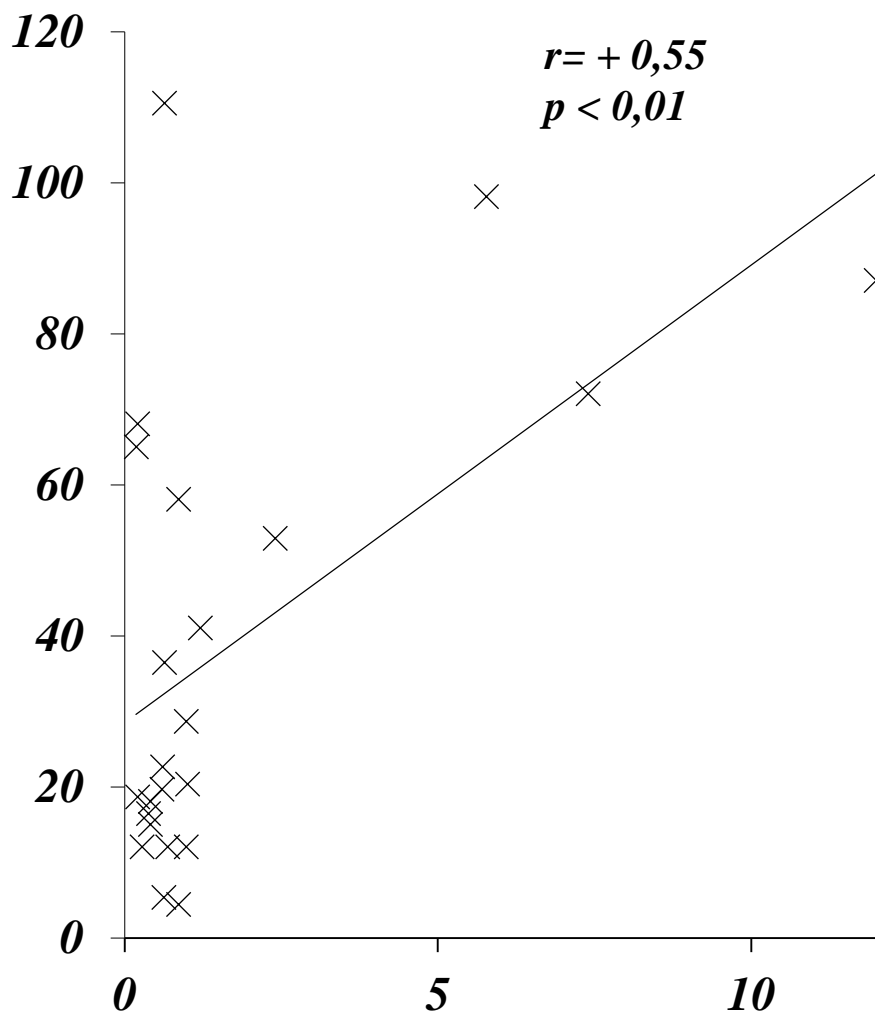
Количественная или качественная корреляция

Уровень гемоглобина	Содержание железа
10,2	4
10,3	4,1
10,3	4,2
10,5	4,3
10,5	4,5
10,6	4,6
10,6	4,6
10,7	4,6
10,9	4,8
11	4,9
11,2	5
11,3	5,1
11,5	5,2

Возрастная группа	Средний САД
1 (до 20 лет)	115,1
2 (21-30 лет)	120,2
3 (31-40 лет)	122,6
4 (41-50 лет)	128,3
5 (51-60 лет)	130,1
6 (61-70 лет)	132,3
7 (71-80 лет)	135,4
8 (> 80 лет)	139,9

Корреляционный анализ

Прямая (+) или обратная (-) корреляция



Корреляционный анализ

«Миф» о коэффициенте корреляции

Если $0,30 < r < 0,50$ – то слабая корреляция

Если $0,50 < r < 0,70$ – то умеренная корреляция

Если $0,70 < r$ – то сильная корреляция

Абсолютная ложь

Коэффициент корреляции должен сопровождаться статистической значимостью, которая оценивается односторонним или двусторонним критерием значимости:

Например:

при $n = 1000$; $r = 0,07$; $p = 0,027$ – стат. достоверная

при $n = 10$; $r = 0,70$; $p = 0,051$ – стат. недостоверная

Корреляционный анализ

- **Линейная корреляция (с Z-преобразованием Фишера)**
- **Нелинейная корреляция**
- **Корреляция Пирсона (для качественных данных)**
- **Корреляция Тай-б-Кэндала**
- **Корреляция Спирмена**
- **и т.д.**



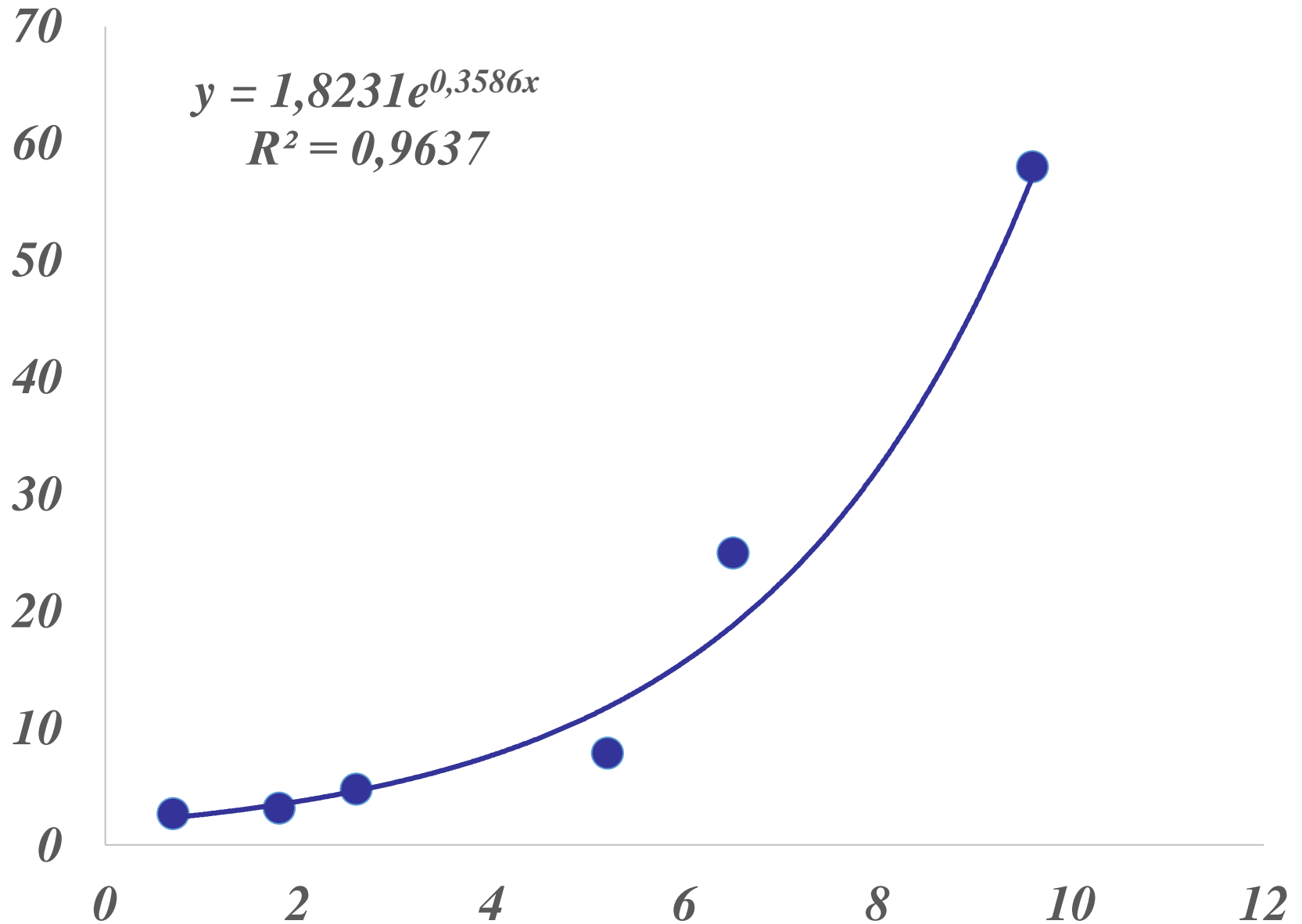
Регрессионный анализ

Регрессионный анализ статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную

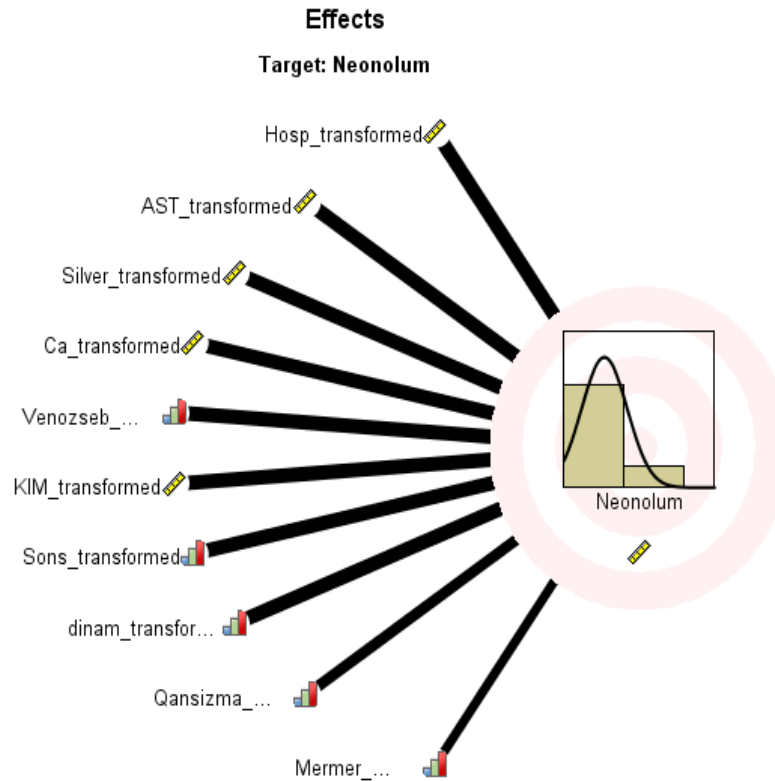
Виды регрессионных анализов

- Линейная
- Нелинейная
- Порядковая
- Категориальная
- Логистическая
- Мультиномиальная
- Метод наименьших квадратов
- Регрессия Кокса и т.д.

Регрессионный анализ



Регрессионный анализ



$$\text{Результат} = A_0 + A_1 * \text{Фактор}_1 + A_2 * \text{Фактор}_2 + \dots + A_k * \text{Фактор}_k$$

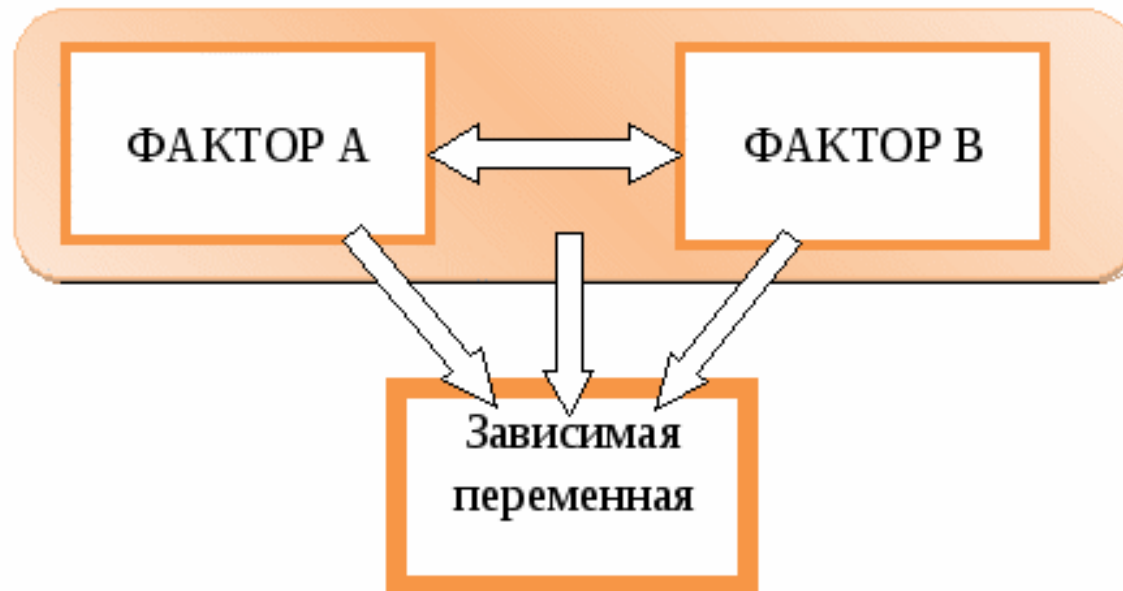
Далее полученная формула проходит «экзамен», оценивается специфичность, чувствительность и т.д. и естественно оценивается статистическая значимость результатов



Дисперсионный анализ

Задачей дисперсионного анализа является изучение влияния одного или нескольких факторов на рассматриваемый признак.

Иногда ее называют «тест ANOVA»



Дисперсионный анализ

- Однофакторный дисперсионный анализ
- Двухфакторный дисперсионный анализ
- Многофакторный дисперсионный анализ

$$F_{расч} = \frac{d_{факт}}{d_{ост}}$$

Презентация окончена)

Спасибо за внимание!