

Дискриминантный анализ

Предварительный анализ включает:

Определение выборочных долей (р) и их ошибок (mp) при альтернативной группировки вариант

$$P \pm mp\%$$

$$65,0 \pm 15,0\%$$

На основании качественных данных составление кросс-таблиц, диаграмм

кросс-таблицы могут быть формата:

- 2 x 2
- 2 x n
- n x 2
- n x m

Дискриминантный анализ

Опять «2»? ?

Возраст	Кол-во (n=50)	%
<= 30	10	20,0%
30-40	15	30,0%
40-50	15	30,0%
50-60	10	20,0%
>= 60	5	10,0%

Жалобы	Кол-во	%
Боль	5	25,0%
Сонливость	3	15,0%
Тошнота	5	25,0%
Рвота	2	10,0%
Дискомфорт	5	25,0%
Итого	20	100%



Дискриминантный анализ

Таблица 2x2

Тест		Заболевание	
		Есть	Нет
Признак	Есть	a	b
	Нет	c	d

Дискриминантный анализ

Критерий согласия Пирсона – χ^2

<i>Chi-Square Tests</i>			
	<i>Value</i>	<i>df</i>	<i>Asymp. Sig. (2-sided)</i>
<i>Pearson Chi-Square</i>	<i>9,693^a</i>	<i>3</i>	<i>,027</i>
<i>Likelihood Ratio</i>	<i>11,563</i>	<i>3</i>	<i>,009</i>
<i>Linear-by-Linear Association</i>	<i>4,179</i>	<i>1</i>	<i>,041</i>
<i>N of Valid Cases</i>	<i>36</i>		

a. 5 cells (62,5%) have expected count less than 5. The minimum expected count is ,33.

Дискриминантный анализ

Многопольные таблицы

	гр1	гр2	гр3	гр4	...
пр1					
пр2					
пр3					
пр4					
пр5					
...					

Статистические данные по таблице 2x2

- ✓ Исход
- ✓ AP, ER (абсолютный риск)
- ✓ APK, CER абсолютный риск в группе контроля
- ✓ APЛ, EER абсолютный риск в группе лечения
- ✓ ДИ, CI (доверительный интервал)
- ✓ OR, RR (относительный риск)
- ✓ COP, RRR (снижение относительного риска)
- ✓ CAP, ARR (снижение абсолютного риска)
- ✓ ПОР, IRR (повышение относительного риска)
- ✓ ПАР, IAR (повышение абсолютного риска)
- ✓ ИПВ (индекс потенциального вреда)
- ✓ КПВ, NNT (количество подлежащих воздействию)
- ✓ Шанс, Odds
- ✓ ОШ, OR (отношение шансов)
- ✓ Чувствительность, Sensitivity
- ✓ Специфичность, Specificity
- ✓ Прогностическая ценность теста
- ✓ Прогностическая ценность отрицательного результата теста
- ✓ Прогностическая ценность положительного результата теста
- ✓ ОП (отношение правдоподобия)
- ✓ ОП+ (отношение правдоподобия положительного результата теста)
- ✓ ОП- (отношение правдоподобия отрицательного результата теста)
- ✓ и др.

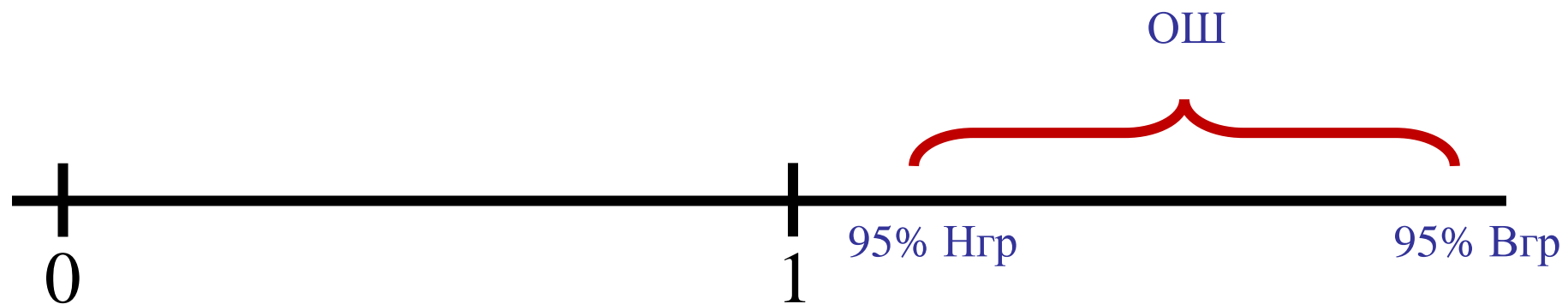
Дискриминантный анализ

Тест		Группы	
		Основная	Контроль
Признак	Есть	a	b
	Нет	c	d

$$Ш = \frac{p}{1-p}$$

$$ОШ (OR) = \frac{ad}{bc}$$

$$95\% \text{ ДИ (CI)} = e^{\ln OR \pm 1,96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$



Дискриминантный анализ

Тест		Заболевание	
		Есть	Нет
Признак	Есть	a (ИП)	b (ЛП)
	Нет	c (ЛО)	d (ИО)

$$\text{Чувствительность (Sn)} = \frac{\text{ИП}}{\text{ИП} + \text{ЛО}} \times 100\%$$

$$\text{Специфичность (Sp)} = \frac{\text{ИО}}{\text{ИО} + \text{ЛП}} \times 100\%$$

Дискриминантный анализ

$$pPV = \frac{ИП}{ИП + ЛП} \times 100\%$$

$$nPV = \frac{ИО}{ИО + ЛО} \times 100\%$$

Тест		Заболевание	
		Есть	Нет
Признак	Есть	a (ИП)	b (ЛП)
	Нет	c (ЛО)	d (ИО)

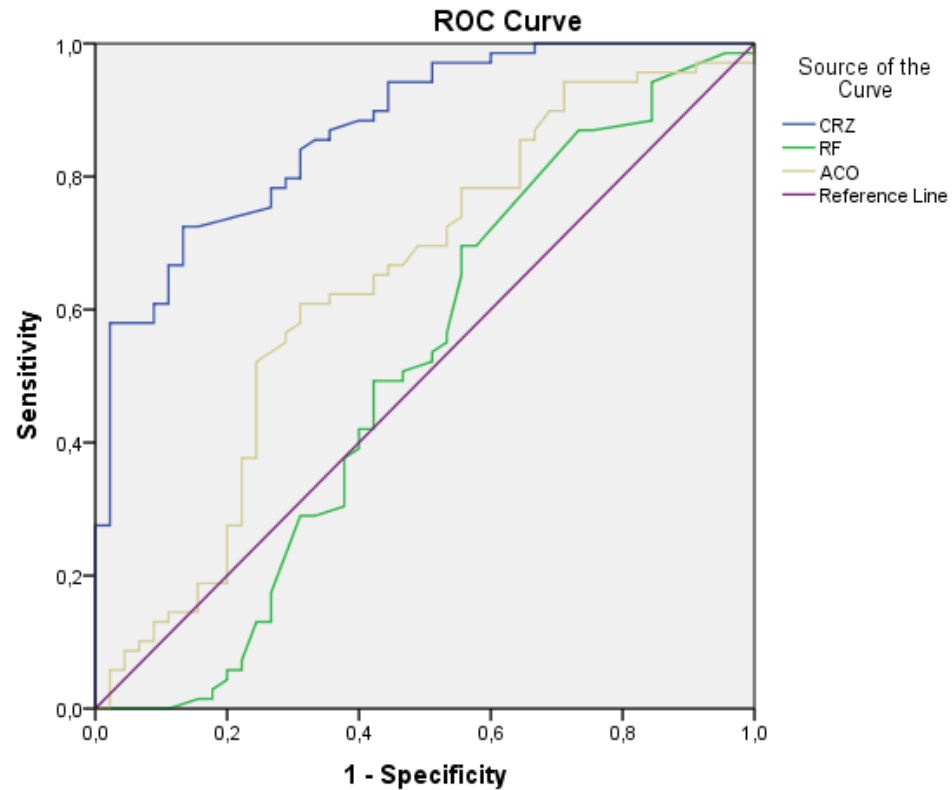
Практичность	LR-	LR+
Отлично	< 0,1	≥ 10
Хорошо	[0,1-0,2)	[5-10)
Посредственно	[0,2-0,5)	[2-5)
Не пригоден	≥ 0,5	< 2

$$LR+ = \frac{\text{Чувствительность}}{100 - \text{Специфичность}}$$

$$LR- = \frac{100 - \text{Чувствительность}}{\text{Специфичность}}$$

$$ДК = \frac{ИП + ИО}{ИП + ЛП + ЛО + ИО} \times 100\%$$

ROC-анализ



Diagonal segments are produced by ties.

Area Under the Curve

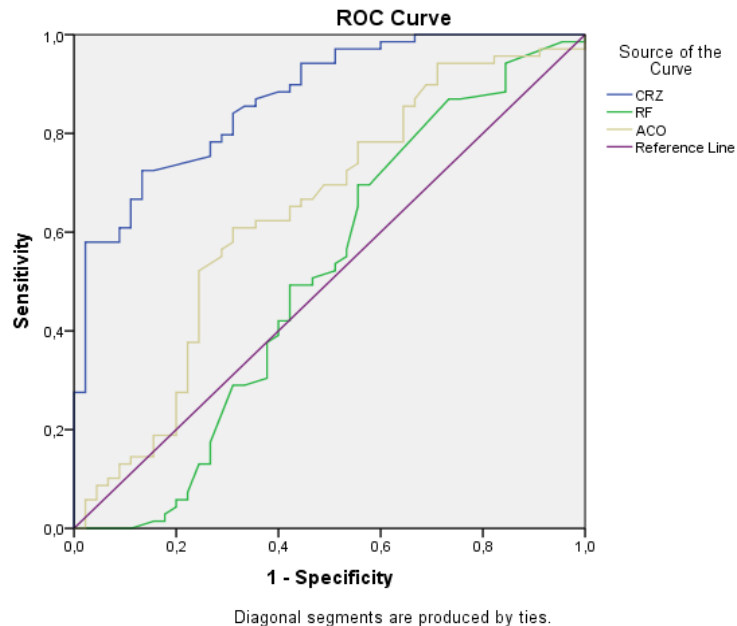
Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
CRZ	,871	,032	,000	,808	,934
RF	,511	,060	,837	,394	,628
ACO	,639	,055	,013	,531	,746

ROC-анализ: определение точки отсечения (COP)

Coordinates of the Curve

Test Result Variable(s)	Positive if Greater Than or Equal To	Sensitivity	1 - Specificity
	,5500	1,000	,978
	,7000	1,000	,956
	,8500	,982	,889
	,9500	,966	,867

	187,3000	,029	,015
	223,1500	,014	,007
	241,0000	,000	,000



$$\max R: \left\{ R = \sqrt{Sn^2 + Sp^2} \right\}$$



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

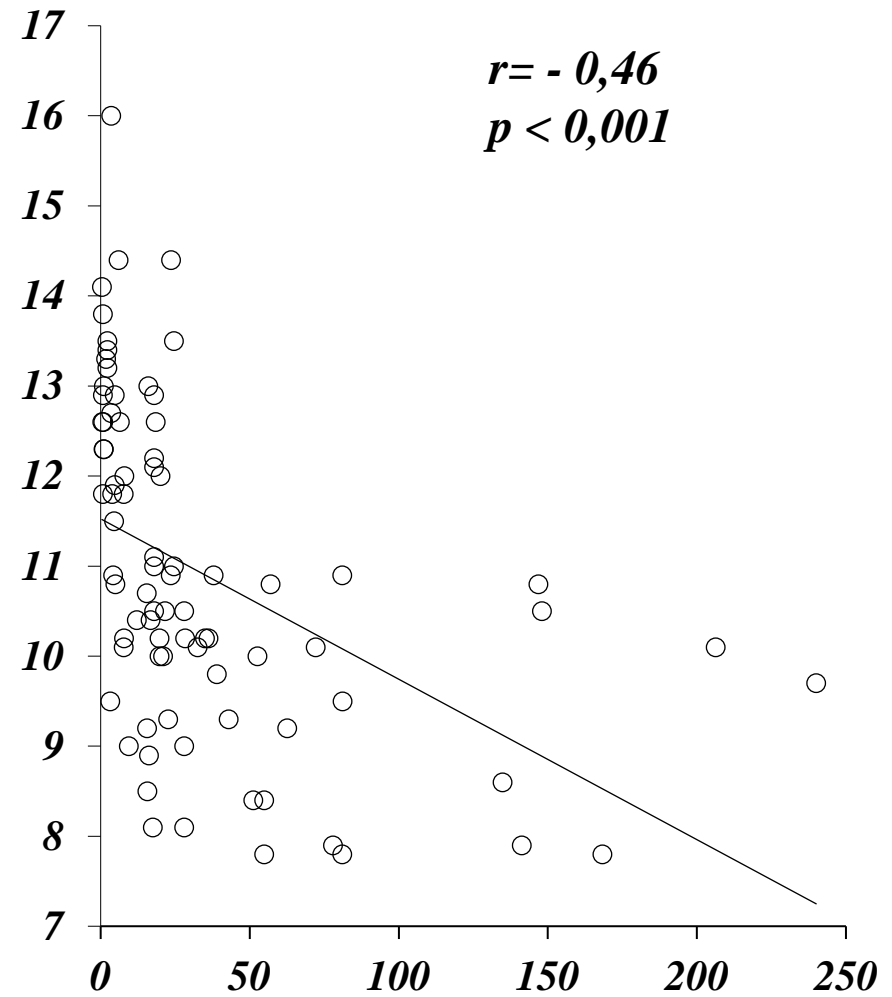
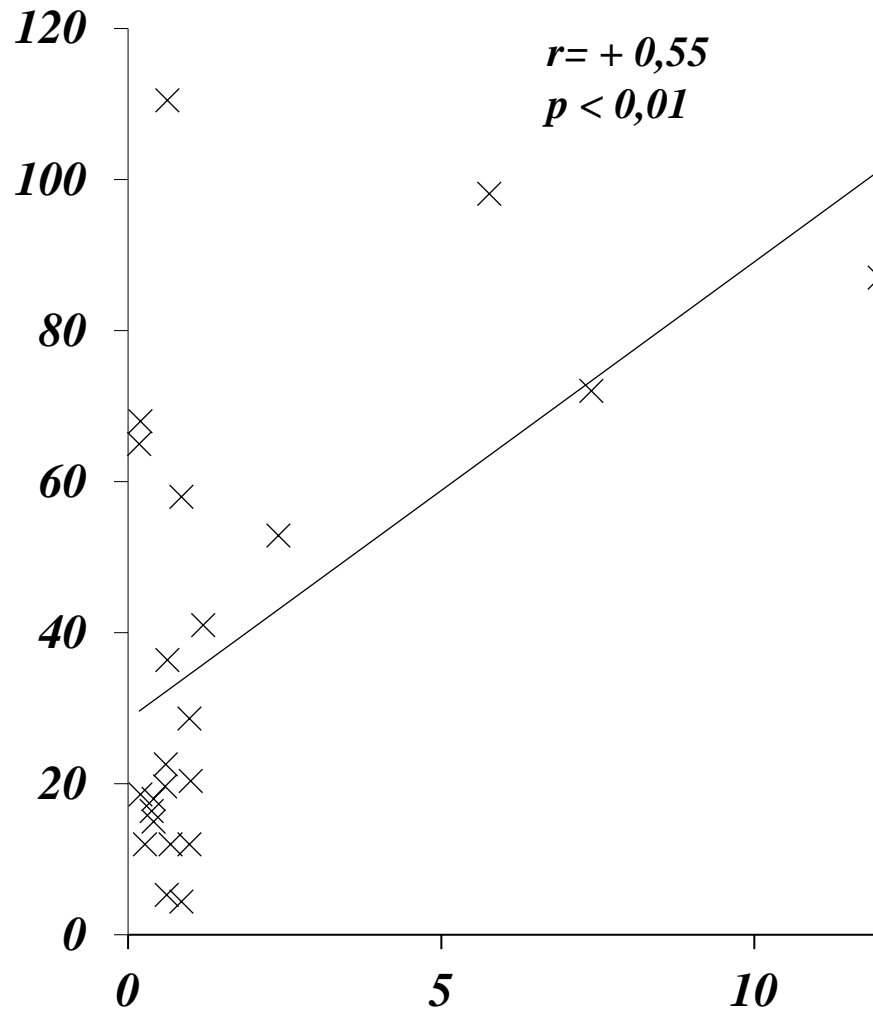
Количественная или качественная корреляция

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.**



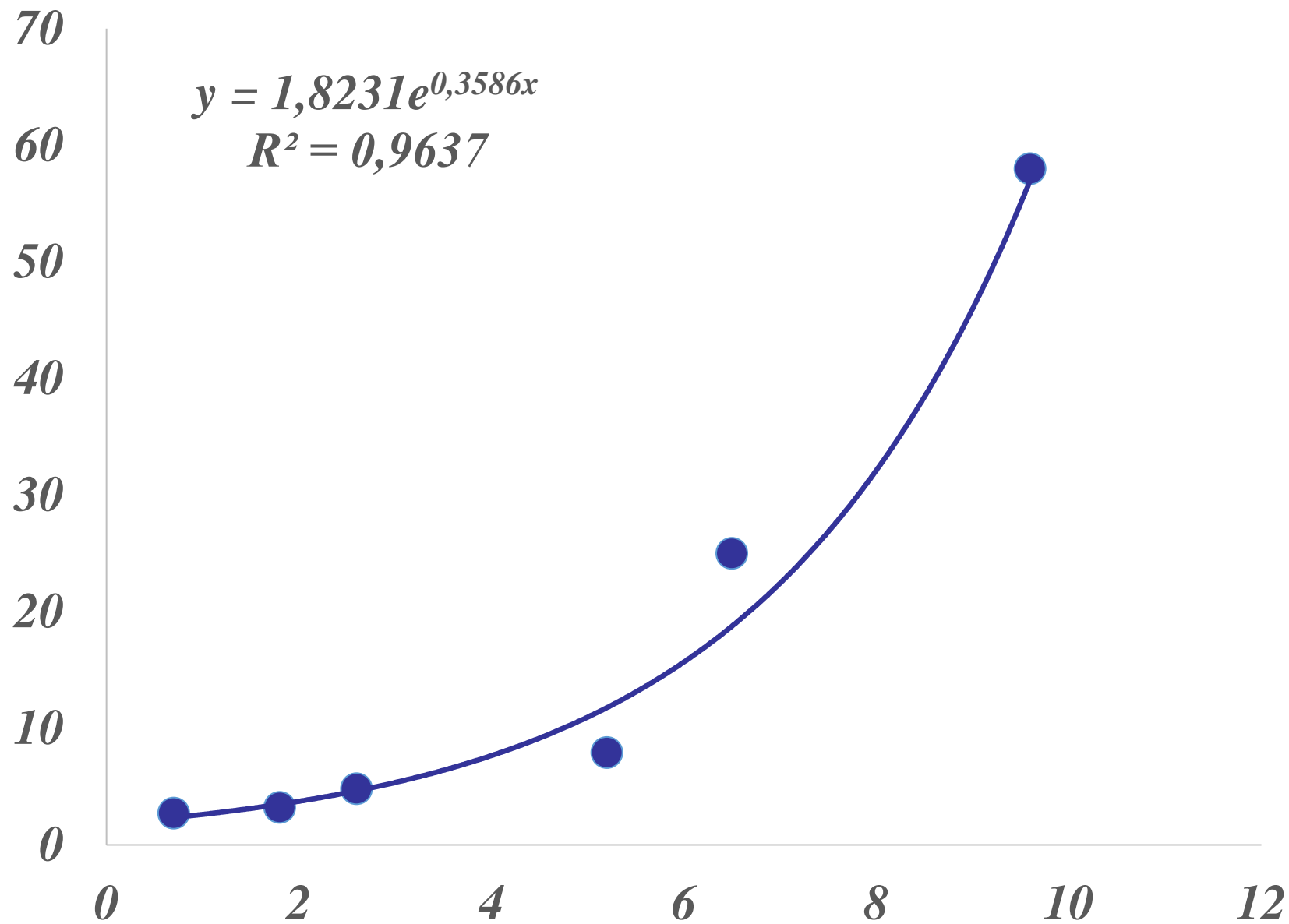
Регрессионный анализ

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

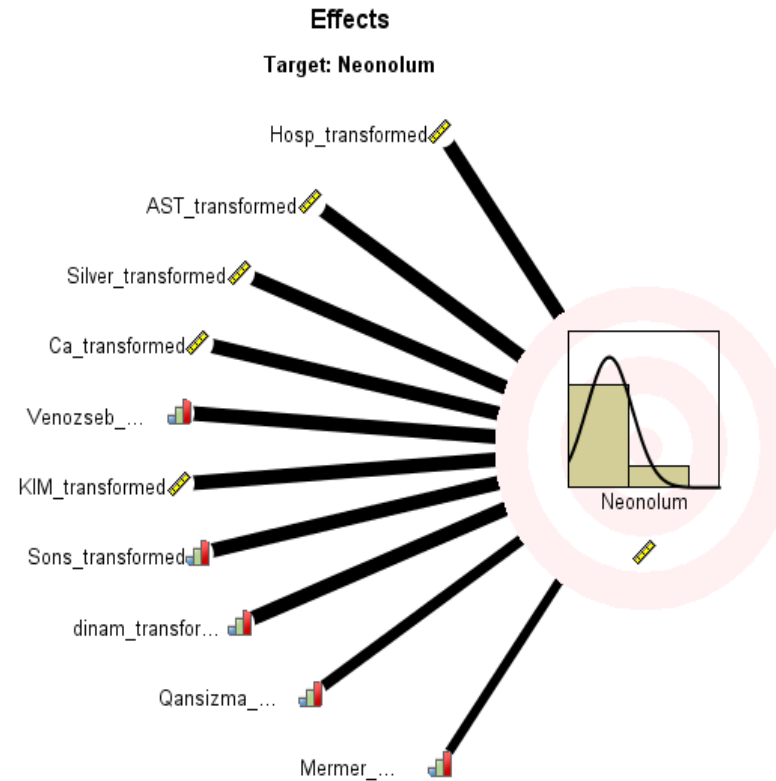
Виды регрессионных анализов

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

Регрессионный анализ



Регрессионный анализ



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

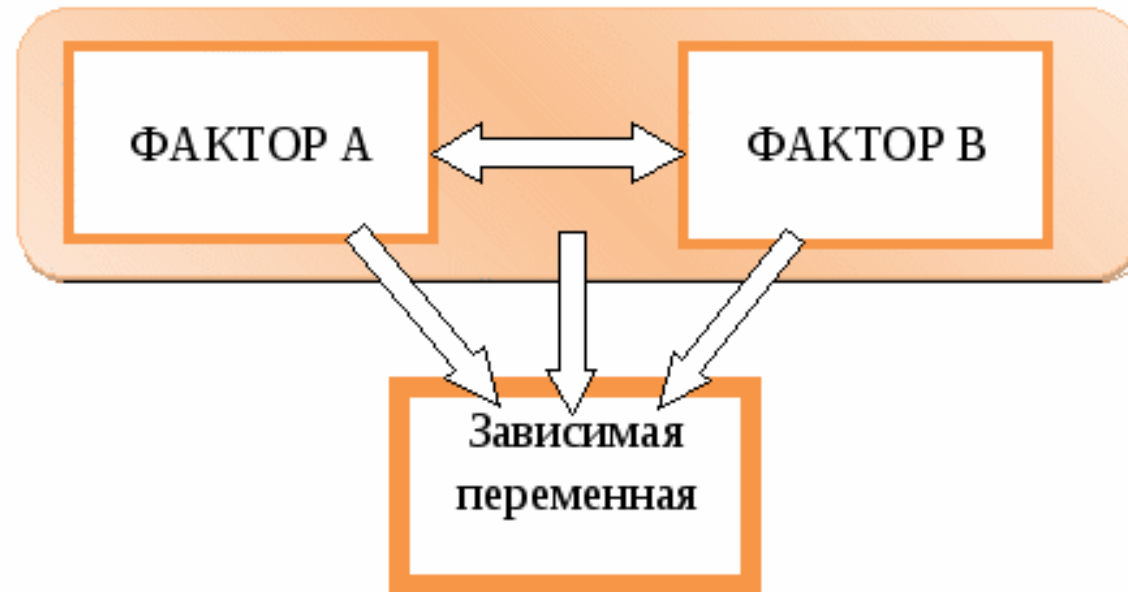
Далее полученная формула проходит «экзамен», оценивается специфичность, чувствительность и т.д. и естественно оценивается статистическая значимость результатов



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

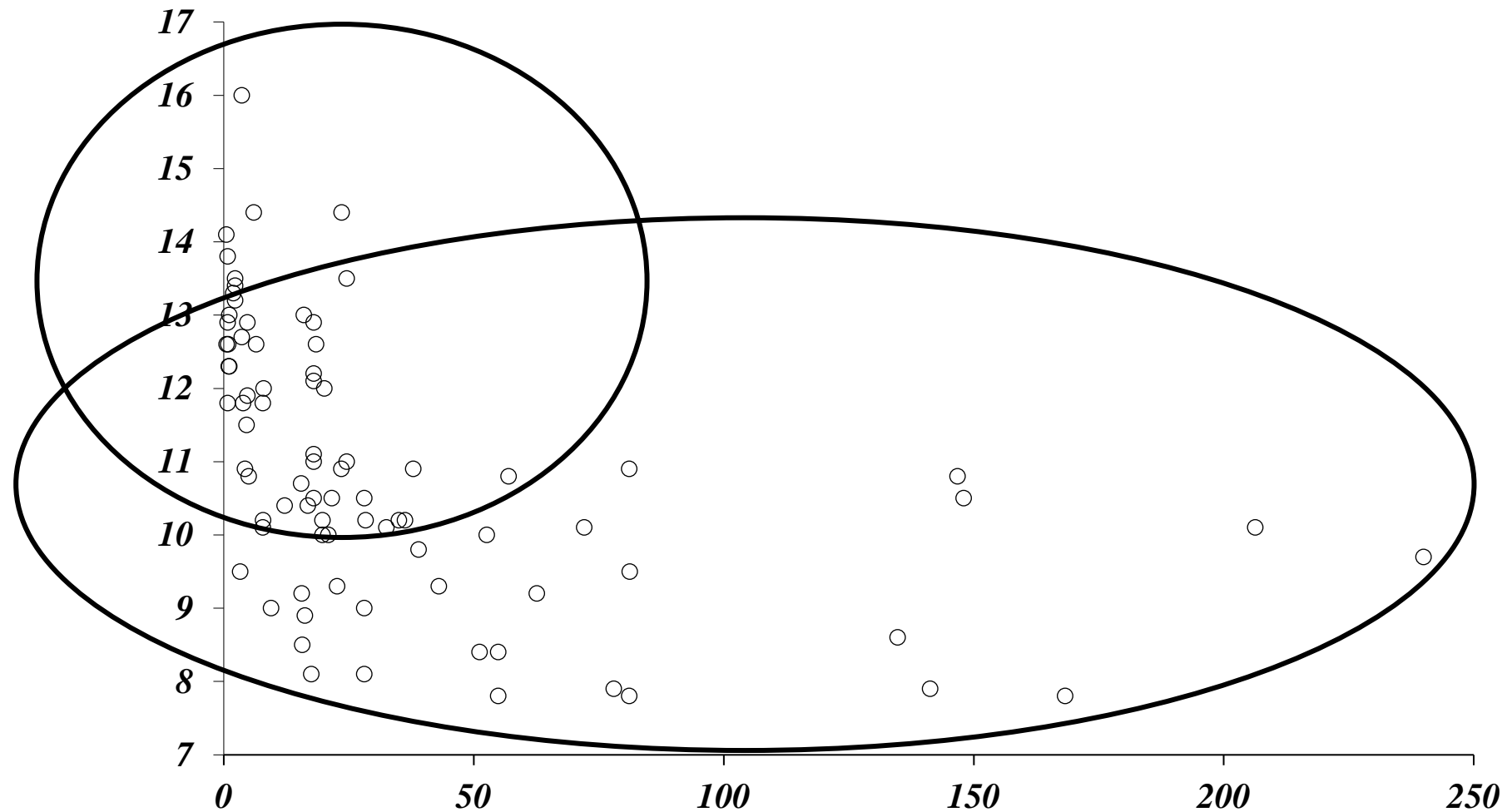
$$F_{расч} = \frac{d_{факт}}{d_{ост}}$$

Дисперсионный анализ (тест ANOVA)

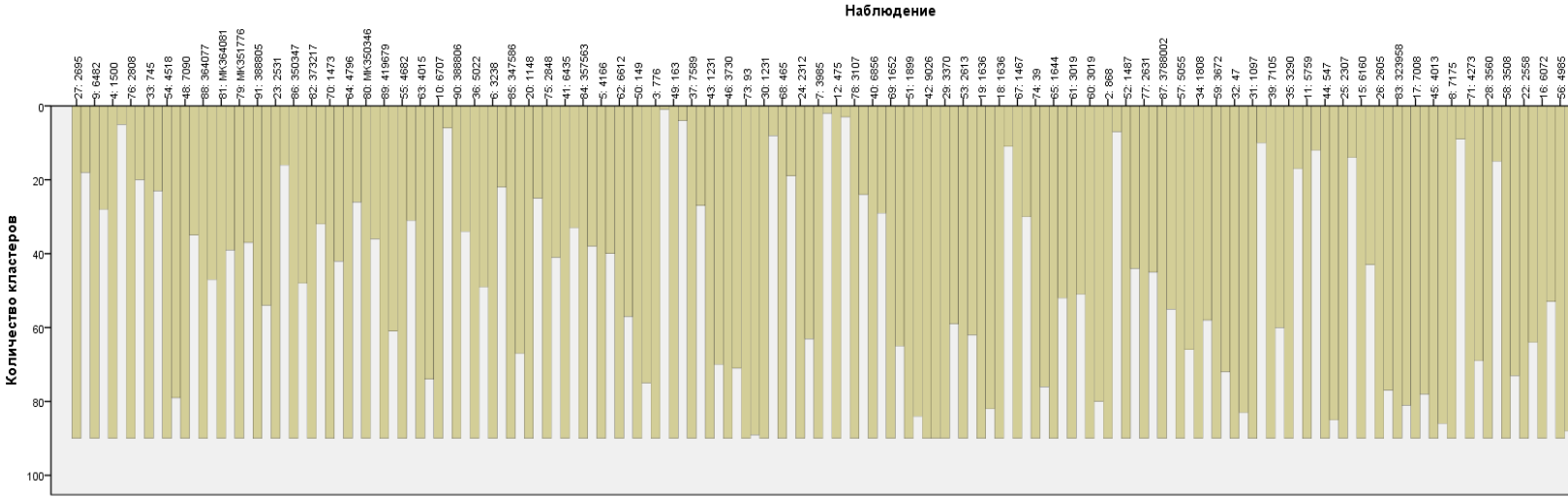
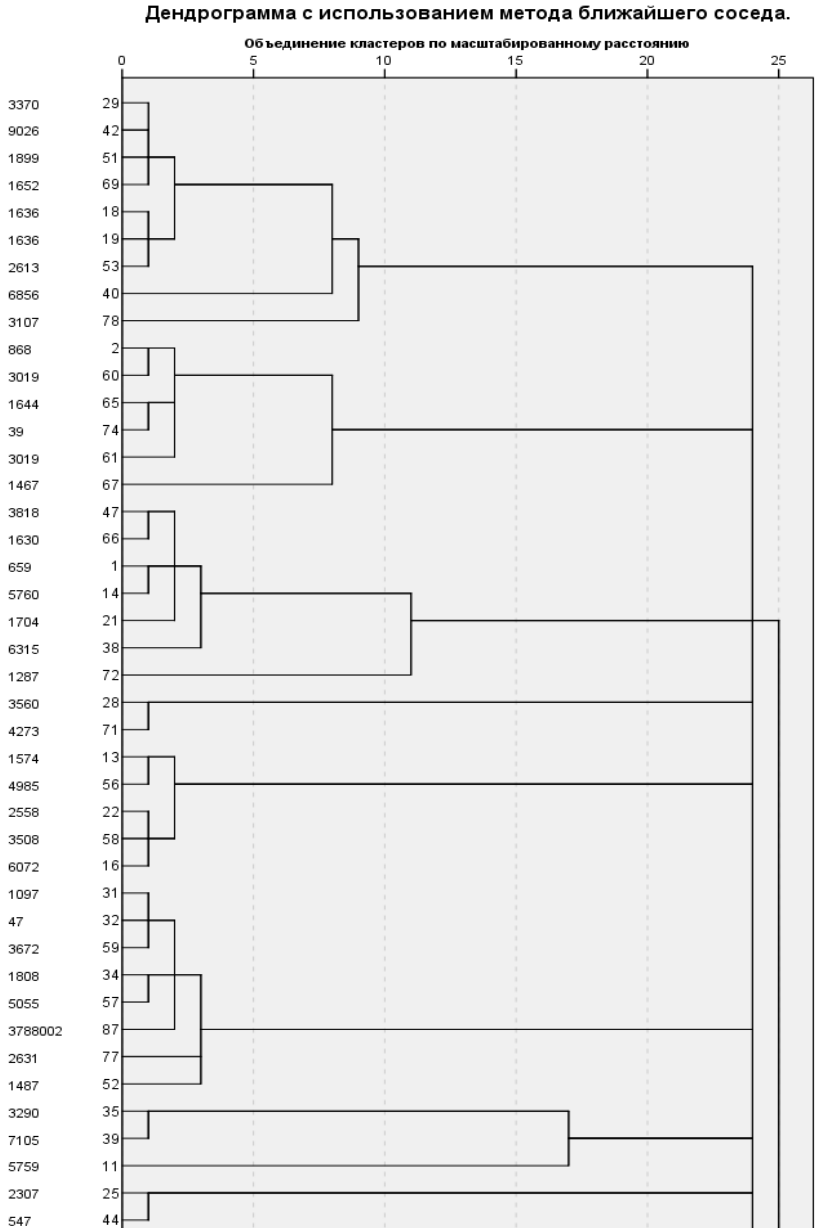
<i>ANOVA</i>						
		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
<i>Area</i>	<i>Between Groups</i>	349417,653	1	349417,653	1,842	,122
	<i>Within Groups</i>	150804,896	214	704,696		
	<i>Total</i>	500222,548	215			
<i>Feret Mean</i>	<i>Between Groups</i>	285,495	1	285,495	6,291	,023
	<i>Within Groups</i>	615,324	214	2,875		
	<i>Total</i>	900,819	215			
<i>Shape Factor</i>	<i>Between Groups</i>	5,309	1	5,309	12,688	,000
	<i>Within Groups</i>	,879	214	,004		
	<i>Total</i>	6,188	215			



Кластерный анализ (метод ближайшего соседа)



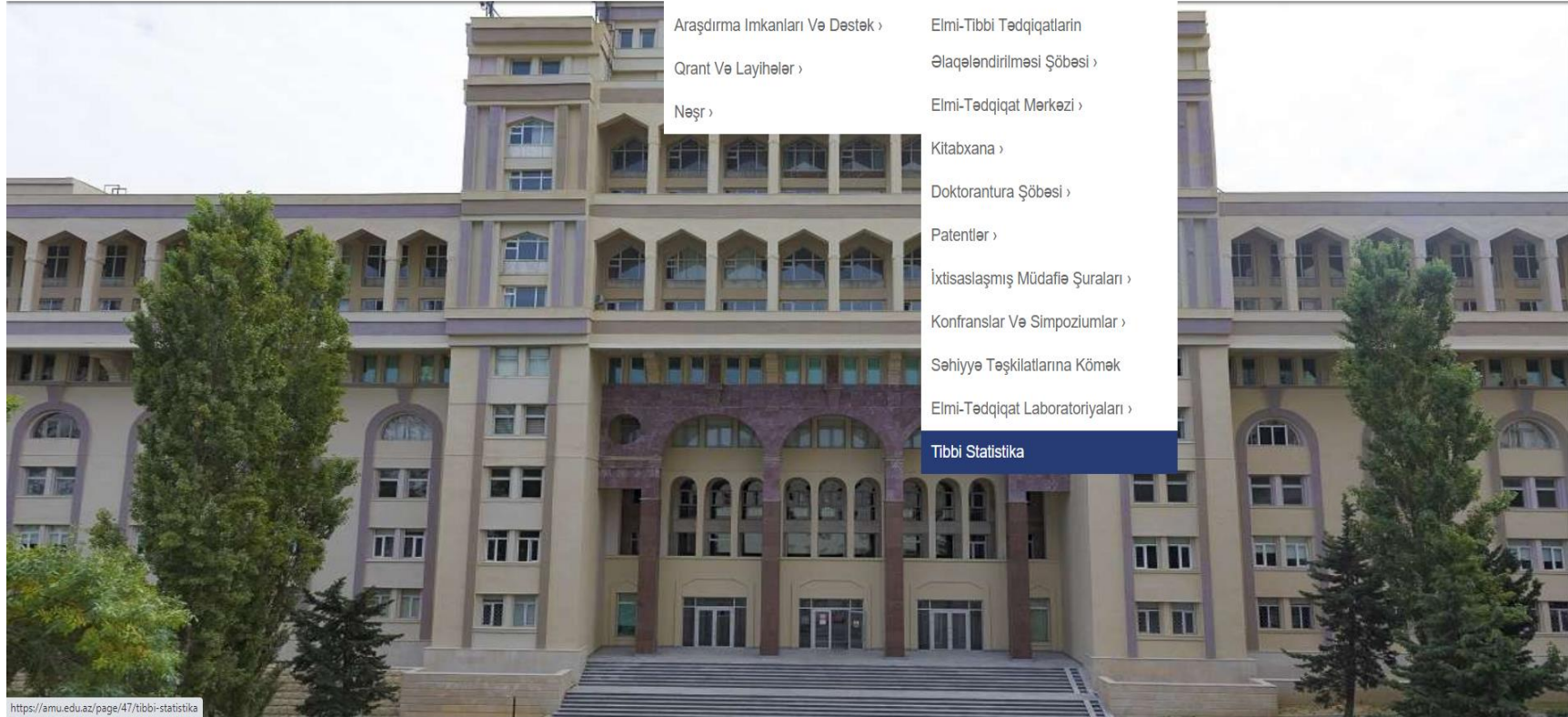
Кластерный анализ (метод ближайшего соседа)







- Elm >
- Araşdırma imkanları Və Dəstək >
- Grant Və Layihələr >
- Nəşr >
- Elmi Hissə >
- Elmi-Tibbi Tədqiqatların Əlaqələndirilməsi Şöbəsi >
- Elmi-Tədqiqat Mərkəzi >
- Kitabxana >
- Doktorantura Şöbəsi >
- Patentlər >
- İxtisaslaşmış Müdafiə Şuraları >
- Konfranslar Və Simpoziumlar >
- Səhiyyə Təşkilatlarına Kömək
- Elmi-Tədqiqat Laboratoriyaları >
- Tibbi Statistika**



Презентация окончена)

Спасибо за внимание!