

Медицинская статистика-2

Докладчик:

д.ф.ф.-м.н. Исмаил Гафаров

Вариационная статистика

Основные средние показатели:

N	– объем выборки
M	– среднее значение
Dx	– дисперсия
σ (Sx)	– среднеквадратическое отклонение
m	– стандартная ошибка
min (max)	– минимальное (максимальное) значение выборки
R	– размах вариации
CV	– коэффициент вариации
95%ДИ	– 95% доверительный интервал
Me	– медиана
Mo	– мода
As	– асимметрия
Ex	– эксцесс
K_0, K_2, \dots, K_4	– квартили (квантили)
P_1, P_2, \dots, P_{99}	– перцентили (процентили)

Вариационная статистика

А вот, незадача ?

Средний возраст детей равняется ($M \pm \sigma$)

3,6 ± 4,2 года

Средний уровень креатинина в крови (M ; 95%ДИ)

158,0 (-23,0; 339,0)

Значит, что-то неправильно в нашей выборке !

Причины:

- ✓ Расчет неправильный – пересчитать заново!*
- ✓ Недостаточное кол-во данных в выборке –
увеличить кол-во наблюдений!*
- ✓ В выборке имеются «выскакивающие» величины –
очистить выборку от них !*

Вариационная статистика

Как избавиться от «выскакивающих» величин ?

1. Перепроверить работу оператора
2. Перепроверить первоисточник
3. Уточнить методику измерения
4. Применить статистические критерии:
 - Критерий σ
 - Критерий Романовского
 - Критерий Шарля
 - Критерий Диксона
 - Процентильный способ
 - И .т.д

$$r_{max} = \frac{x_{max} - M}{\sigma_x}; r_{min} = \frac{M - x_{min}}{\sigma_x}$$

Следующий этап включает «разведывательный анализ»

- Проверить выборки на рандомизированность
- Проверить выборки на повторность
- Проверить выборки на «опустошенность»
- Сравнить средние данные
- Сравнить медианы
- Просмотреть процентильные распределения
- На основании ДИ сделать предварительные заключения
- Если нужно, пересмотреть:
 - ✓ перераспределение данных,
 - ✓ перераспределение групп,
 - ✓ классификацию
 - ✓ критерии исключения и т.д.
- Формировать «нулевую гипотезу»
- Планировать примерные направления статистического анализа

Разведывательный анализ:

Как проверить выборки на рандомизированность ?

Если не являются критериями разделения то, показатели сравниваемых групп статистически

не должны различаться:

- ✓ по полу;
- ✓ по возрасту;
- ✓ по месту жительства;
- ✓ по социальному статусу и .т.д.
- ✓ .
- ✓ .
- ✓ .
- ✓ и т.д.
- ✓ в продольных исследованиях: исходные клинические и лабораторные показатели

Разведывательный анализ

Как проверить выборки на опустошенность?

До	7 дней	30 дней
11,2	10,6	9,3
8,3	8,5	7,3
9,6	8,8	9,2
8,1	9,6	11,3
-	9,6	10,3
4,2	-	5,5
5,3	-	6,9
4,5	-	-
5,5	-	-
6,9	8,6	7,5
5,5	12,2	-
4,2	9,6	-
-	-	9,8
-	10,3	10,3
-	-	11,5

Разведывательный анализ: Как сравнить средние данные?

$$M \pm m$$

$$65,0 \pm 5,0$$

«Волшебная формула» наших любимых профессоров !

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

$$M = (\min + \max)/2 ?$$

M_1

M_2

$M_1 - m_1$

$M_1 + m_1$

$M_2 - m_2$

$M_2 + m_2$

Все-таки t-критерий Стьюдента универсален !

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

При обязательных условиях:

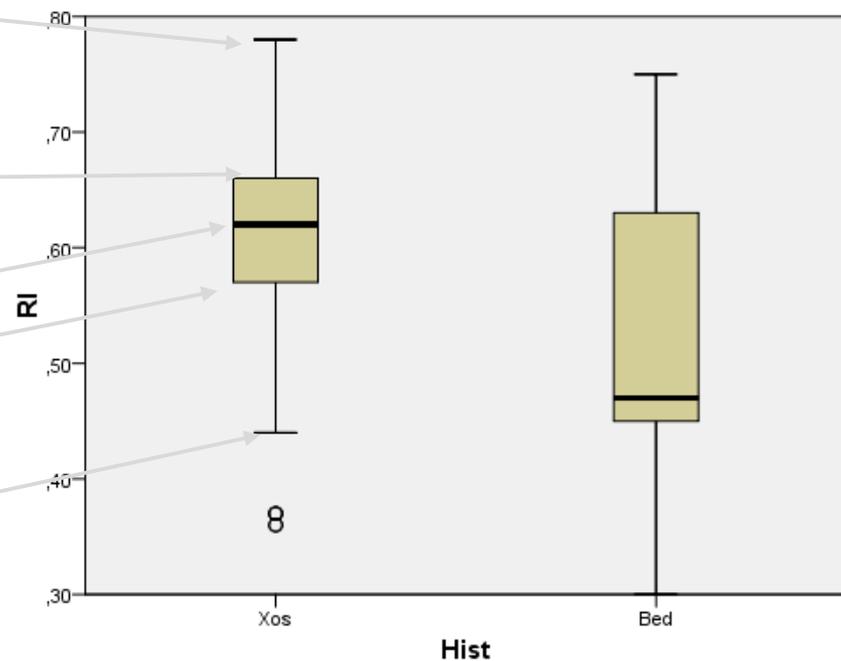
1. Обе выборки должны принадлежать одной и той же генеральной совокупности
2. Распределение выборок должны удовлетворять нормальному закону
3. Среднеквадратические отклонения должны быть примерно одинаковыми ($\sigma_1 \approx \sigma_2$)
4. Количество выборок должны быть одинаковыми ($n_1 = n_2$)

При невыполнении только 4-го условия ($n_1 \neq n_2$) можно применить следующую формулу:

$$t = \frac{|M_1 - M_2| \sqrt{\frac{n_1 n_2}{n_1 + n_2}}}{\sqrt{\frac{m_1^2 n_1 (n_1 - 1) + m_2^2 n_2 (n_2 - 1)}{n_1 + n_2 - 2}}}$$

Разведывательный анализ: Процентильные распределения

p1	4,20		
p5	4,20		
p10	5,57		
p25	8,17	K1	
p50	12,92	K2	Me
p75	20,52	K3	
p90	27,52		
p95	32,50		
p99	53,79		



Оставшиеся этапы разведывательного анализа

- Проверить выборки на рандомизированность
- Проверить выборки на повторность
- Проверить выборки на «опустошенность»
- Сравнить средние данные
- Сравнить медианы
- Просмотреть процентильные распределения
- На основании ДИ сделать предварительные заключения
- Если нужно, пересмотреть:
 - ✓ перераспределение данных,
 - ✓ перераспределение групп,
 - ✓ классификацию
 - ✓ критерии исключения и т.д.
- Формировать «нулевую гипотезу»
- Планировать примерные направления статистического анализа

Разведывательный анализ

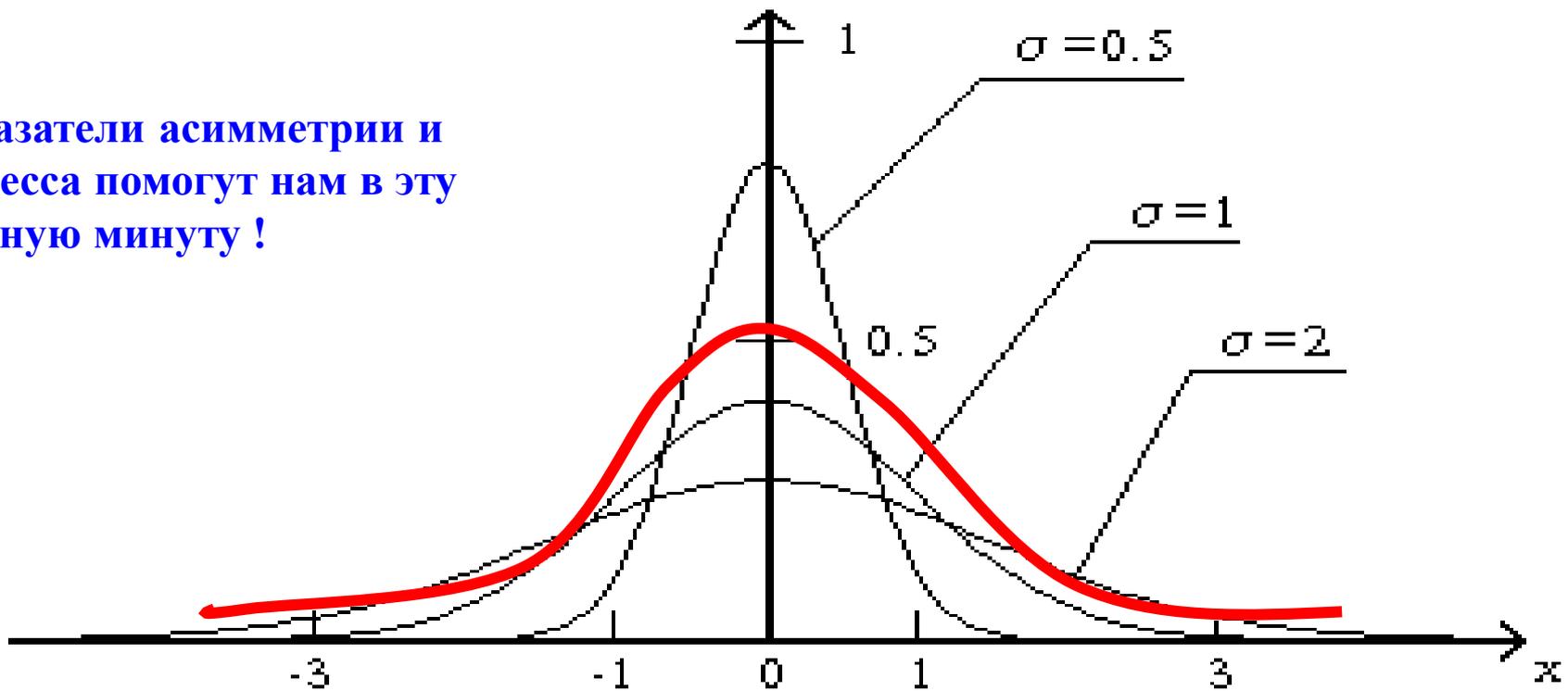
Как дальше быть?

Выбирать

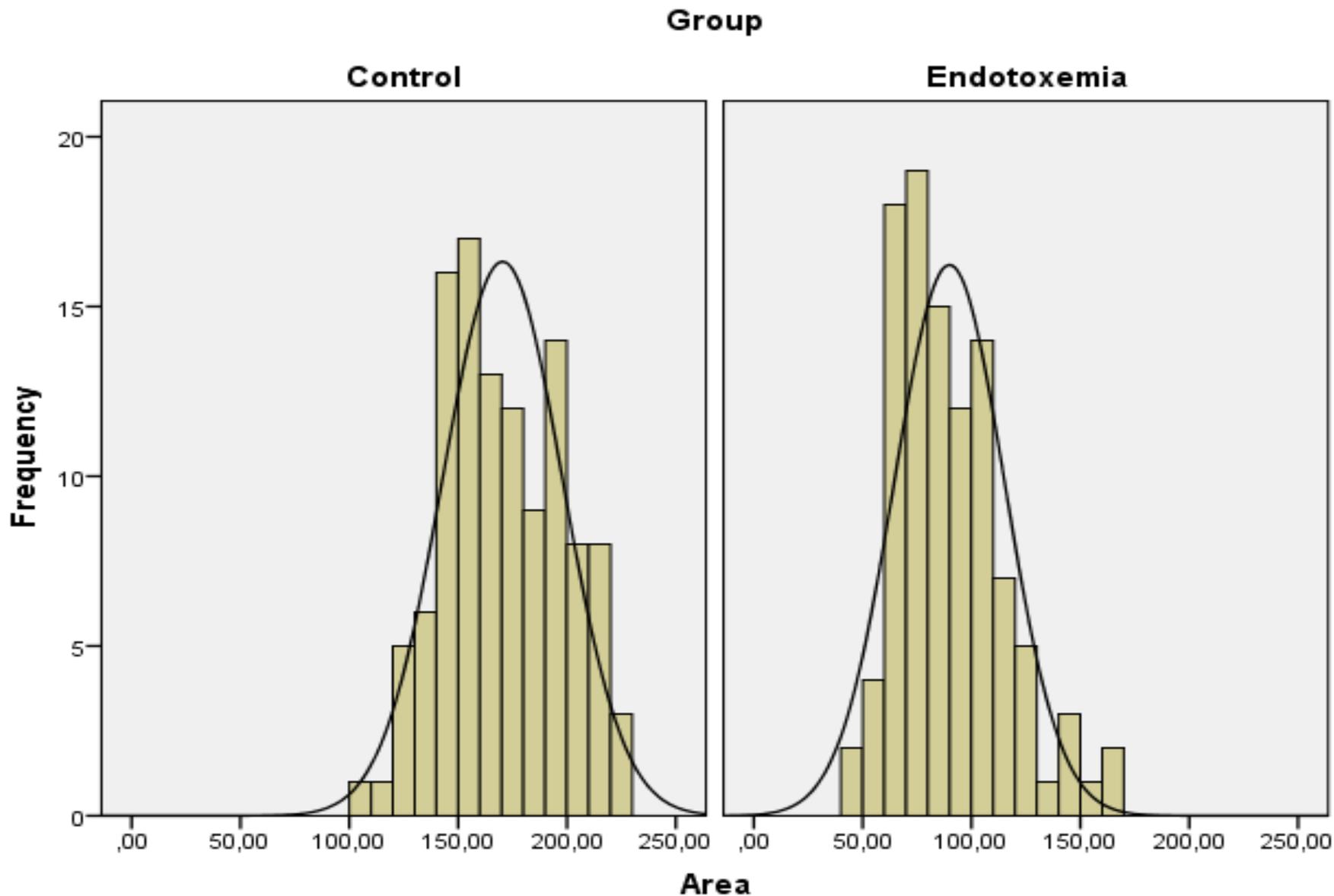
параметрические или непараметрические
методы анализа ?



Показатели асимметрии и эксцесса помогут нам в эту трудную минуту !



Разведывательный анализ



Вариационная статистика

Критерии сравнения количественных значений:

Для несвязанных выборок

- t – критерий Стьюдента
- U – критерий Уилкоксона (Манна-Уитни)
- W – критерий Уайта
- X – критерий Ван дер Вардена
- Z – критерий Колмогоров-Смирнова
- VV – критерий Вальда-Вольфовича
- KW – критерий Краскала-Уоллиса
- ST – критерий Джонкхир-Терпстр
- Me – медианный критерий
- M – критерий экстремальной реакции Мозеса
- ... и др.

Вариационная статистика

Критерии сравнения количественных значений:

Для связанных выборок

- d – критерий Стьюдента
- T – критерий Уилкоксона
- N – критерий Мак Немара
- z – критерий знаков
- XL – критерий Ходжеса-Лемана
- W – критерий Кендалла
- F – критерий Фридмана
- ... и др.



Дискриминантный анализ

Предварительный анализ включает:

Определение выборочных долей (p) 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
- ✓ APL, EER абсолютный риск в группе лечения
- ✓ ДИ, CI (доверительный интервал)
- ✓ OP, 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.

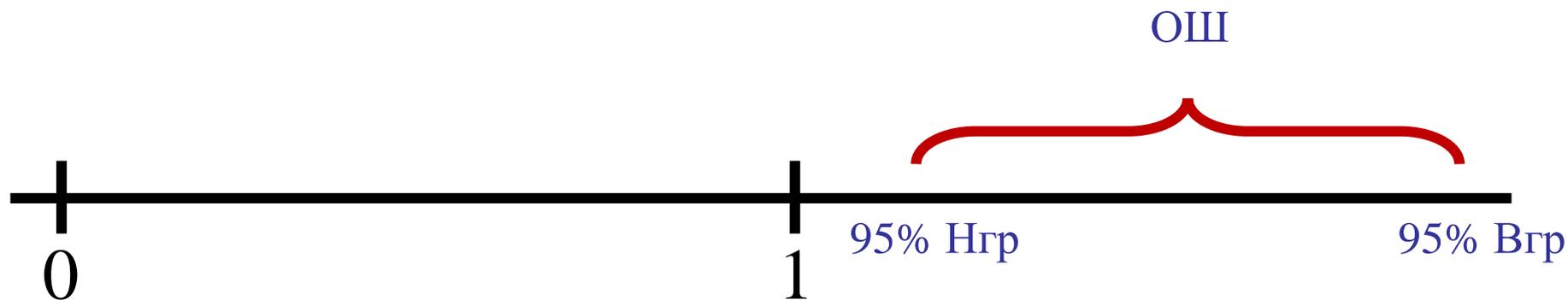
Дискриминантный анализ

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)

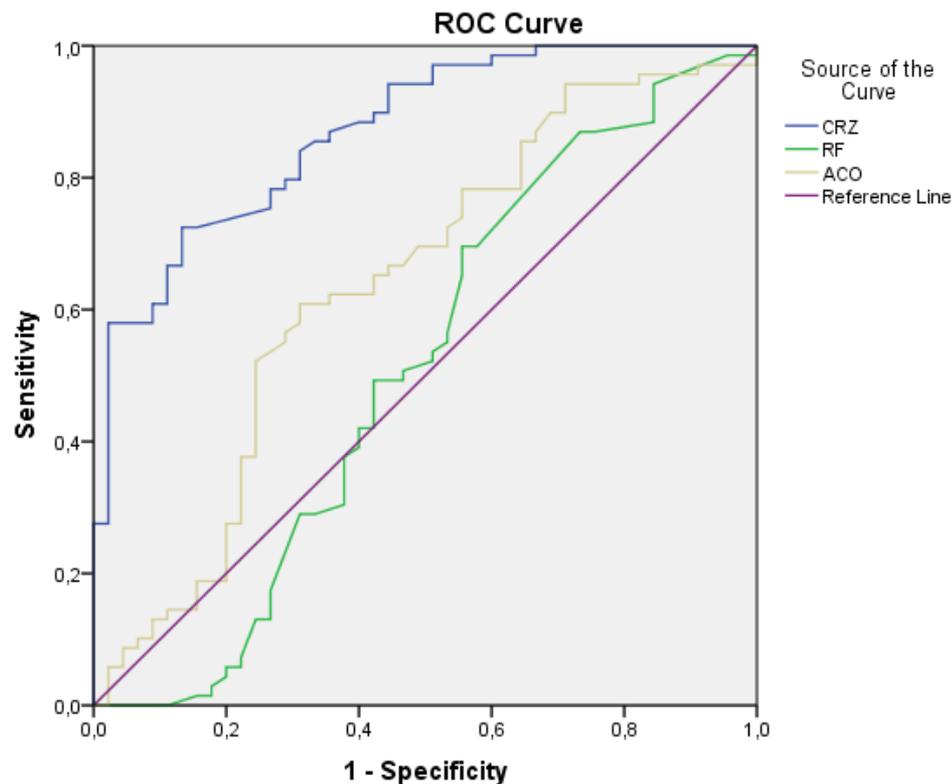
Практичность	LR-	LR+
Отлично	< 0,1	≥ 10
Хорошо	[0,1-0,2)	[5-10)
Посредственно	[0,2-0,5)	[2-5)
Не пригоден	≥ 0,5	< 2

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

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

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

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)

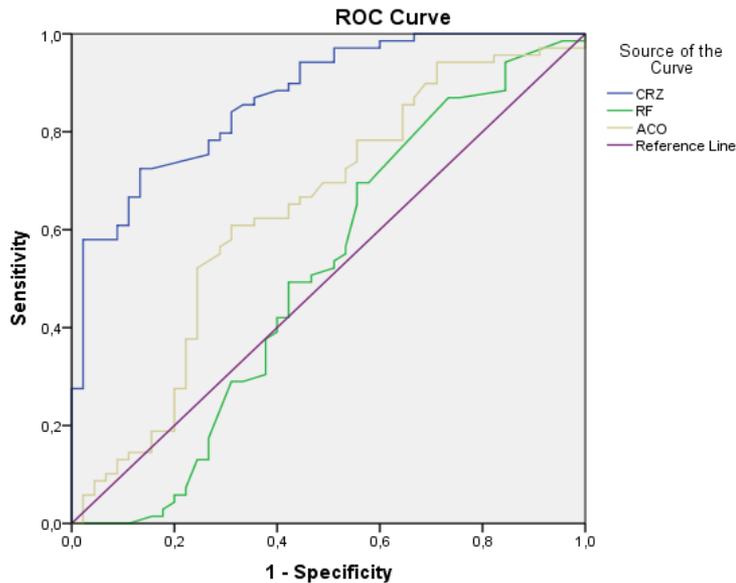
Coordinates of the Curve

Test Result Variable(s)

Positive if
Greater Than
or Equal To

Sensitivity

1 -
Specificity



Diagonal segments are produced by ties.

,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

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

