

Вербальные шкалы

Надежность доверительных интервалов (ДИ)

Уровень значимости, α	Уровень доверия, $(1 - \alpha) \times 100\%$	Надежность интервальной оценки
0,05	95%	Приемлемая
0,01	99%	Удовлетворительная
0,001	99,9%	Высокая

Традиционная интерпретация P-значений (шкала Michelin)

P-значение	Статистическая значимость	Шкала Мишлена
$> 0,05$	Незначимо	
$0,05 - 0,01$	Умеренно значимо	*
$0,01 - 0,001$	Значимо	**
$< 0,001$	Высоко значимо	***

Калибровка P -значения

P -значение	Нижняя граница для вероятности нулевой гипотезы $P(H_0)$	Верхняя граница для вероятности воспроизведения P_{repr}
0,05	> 29%	< 50%
0,01	> 11%	< 73%
0,001	> 1,8%	< 90%

Для наглядности значения в таблице округлены до первой значащей цифры. Более точно значения для $P(H_0)$ (сверху вниз) равны 29%, 11% и 1,8%.

Chow SC, Shao J, and Wang, H. **Sample Size Calculations in Clinical Research**. Second edition, Chapman Hall/CRC Press, Taylor & Francis, New York, New York. P. 6, Table 1.1.2.

Идеальный, бесполезный и абсурдный тесты в терминах Se и Sp

- Если тест идеальный, то
 - $Se = 1$ и $Sp = 1$.
- Если
 - $Se = 0,5$ и $Sp = 0,5$,
- то тест бесполезен.
- Если
 - $Se < 0,5$ и $Sp = 0,5$,
- то тест следует признать абсурдным или даже «вредным».

Возможные словесные интерпретации для градаций Se и Sp

Se	Способность положительных результатов диагностического теста распознавать наличие болезни:
0,9 – 1,0	высокая
0,7 – 0,9	средняя
0,5 – 0,7	низкая
< 0,5	Абсурдная
Sp	Способность отрицательных результатов диагностического теста распознавать отсутствие болезни:
0,9 – 1,0	высокая
0,7 – 0,9	средняя
0,5 – 0,7	низкая
< 0,5	Абсурдная

Идеальный, бесполезный и абсурдный тесты в терминах $LR[+]$ и $LR[-]$

- Если тест идеальный, то
 - $LR[+] = \infty$ и $LR[-] = 0$.
- Если
 - $LR[+] = LR[-] = 1$,
- то тест бесполезен.
- Если
 - $LR[+] < 1$ и/или $LR[-] > 1$,
- то тест следует признать абсурдным или даже «вредным».

Принятые словесные интерпретации для градаций $LR[+]$

$LR[+]$	Повышение вероятности наличия болезни у субъекта после получения у него положительного результата проверяемого диагностического теста:
> 1000	Практически идеальное
100 - 1000	Очень высокое
33 – 100	Высокое
10 – 33	Удовлетворительное
3 – 10	Посредственное
1 – 3	Неудовлетворительное
< 1	Абсурдное

Принятые словесные интерпретации для градаций $LR[-]$

$LR[-]$	Повышение вероятности отсутствия болезни у субъекта после получения у него отрицательного результата проверяемого диагностического теста:
$> 1,0$	Абсурдное
$1 - 0,3$	неудовлетворительное
$0,3 - 0,1$	посредственное
$0,1 - 0,03$	удовлетворительное
$0,03 - 0,01$	высокое
$0,01 - 0,001$	очень высокое
$< 0,001$	Практически идеальное

Идеальный, бесполезный и абсурдный тесты в терминах PPV и NPV

- Тест идеален, когда:
 - $PPV = 1$ и $NPV = 1$.
- Когда
 - $PPV = P(D+)$
 - $NPV = P(D-)$,
- то тест бесполезен.
- Когда
 - $PPV < P(D+)$
 - $NPV < P(D-)$,
- то тест следует признать абсурдным или даже «вредным».

Общепринятые словесные интерпретации для градаций *AUC*

Интервалы значений <i>AUC</i>	Способность диагностического теста распознавать наличие или отсутствие болезни
0,9 – 1,0	Отличная
0,8 – 0,9	Хорошая
0,7 – 0,8	Удовлетворительная
0,6 – 0,7	Посредственная
0,5 – 0,6	Неудовлетворительная
< 0,5	Абсурдная («вредная»)

Интерпретация стандартизированного размера эффекта d_c

<http://www.sportsci.org/resource/stats/>

Размер эффекта, d_c	Градация эффекта
0 – 0,2	Ничтожный
0,2 – 0,6	Малый
0,6 – 1,2	Умеренный
1,2 – 2,0	Большой
2,0 – 4,0	Очень большой
4,0 - ∞	Исключительно большой

Словесные интерпретации для разности долей RD (или δ)

Интервалы значений RD	Практическая важность разности долей
0,0 – 0,1	Ничтожная
0,1 – 0,3	Малая
0,3 – 0,6	Умеренная
0,6 – 0,9	Большая
0,9 – 1,0	Очень большая

Словесная шкала градаций для размеров эффекта

Мера эффекта				Градация эффекта
<i>RD</i>	<i>RR</i>	<i>OR</i>	<i>NNT</i>	
0 – 0,1	1,0 – 1,2	1,0 – 1,5	∞	НИЧТОЖНЫЙ
0,1 – 0,5	1,2 – 3,0	1,5 – 9,0	> 5	слабый
0,5 – 0,9	3,0 - 20	9,0 - 360	2 - 5	сильный
0,9 – 1,0	> 20	> 360	< 2	очень сильный

Клиническая значимость коэффициента каппа Коуэна

Значения коэффициента каппа Коуэна	Интерпертация
- 1	Полная рассогласованность, противоречивость
< 0	Рассогласованность
0	Независимость
	Степень согласованности:
0 – 0,2	Плохая
0,2 – 0,4	Посредственная
0,4 – 0,6	Удовлетворительная
0,6 – 0,8	Хорошая
0,8 - 1,0	Отличная

Интерпретация убедительности Бейзовых факторов, BF_{10} и BF_{01}

BF_{10}	Свидетельство в пользу гипотезы H_1 против гипотезы H_0
>100	Убедительное
30 – 100	Очень сильное
10 – 30	Сильное
3 – 10	Умеренное
1 – 3	Пренебрежимо малое

BF_{01}	Свидетельство в пользу гипотезы H_0 против гипотезы H_1
<0,01	Убедительное
0,01 – 0,03	Очень сильное
0,03 – 0,1	Сильное
0,1 – 0,3	Умеренное
0,3 - 1	Пренебрежимо малое