

Точность оценок

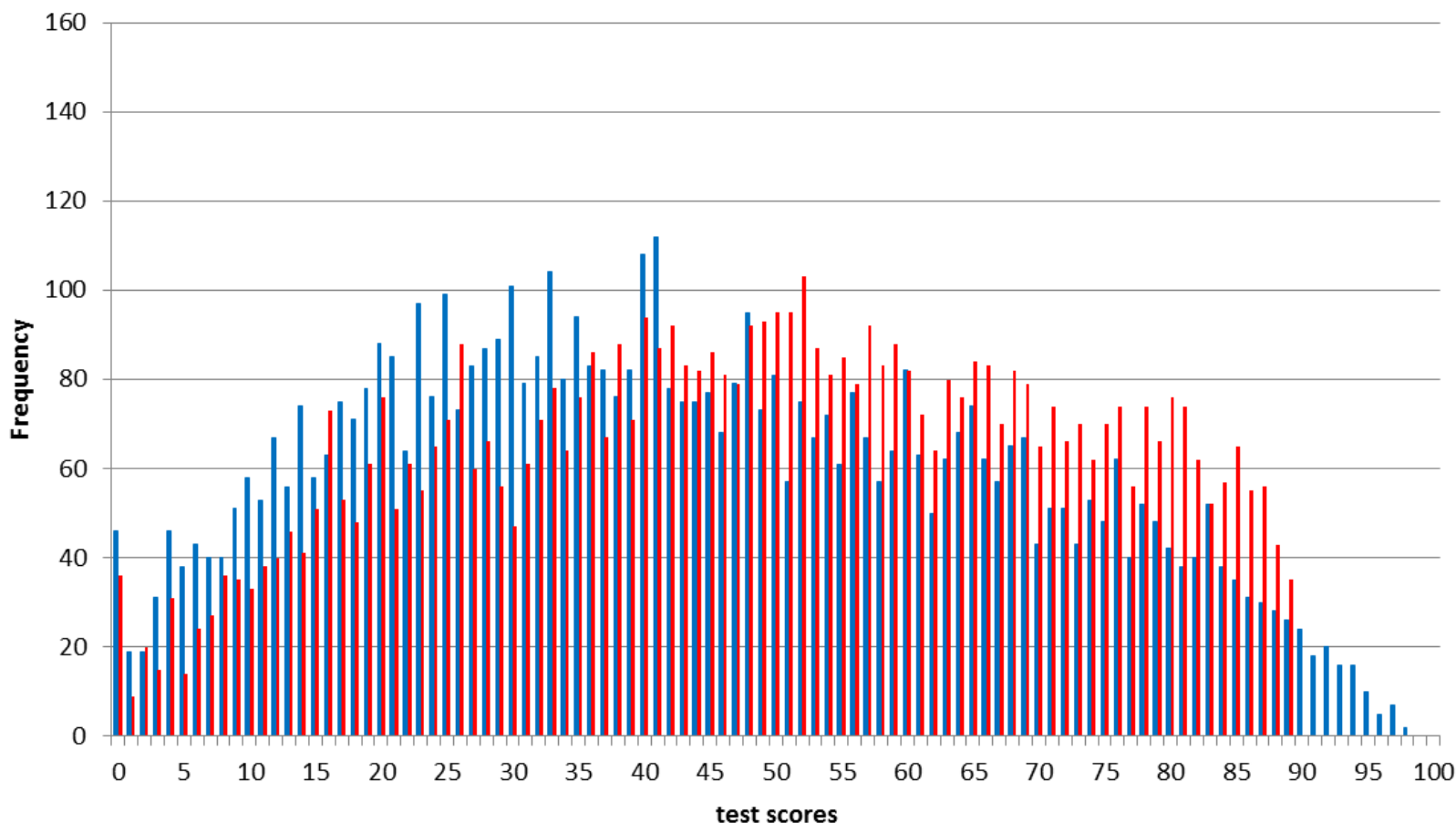
Лекция 2.1

Калашнов Георгий, Сучкова Ольга

go9513@gmail.com suchkovaolga.91@mail.ru

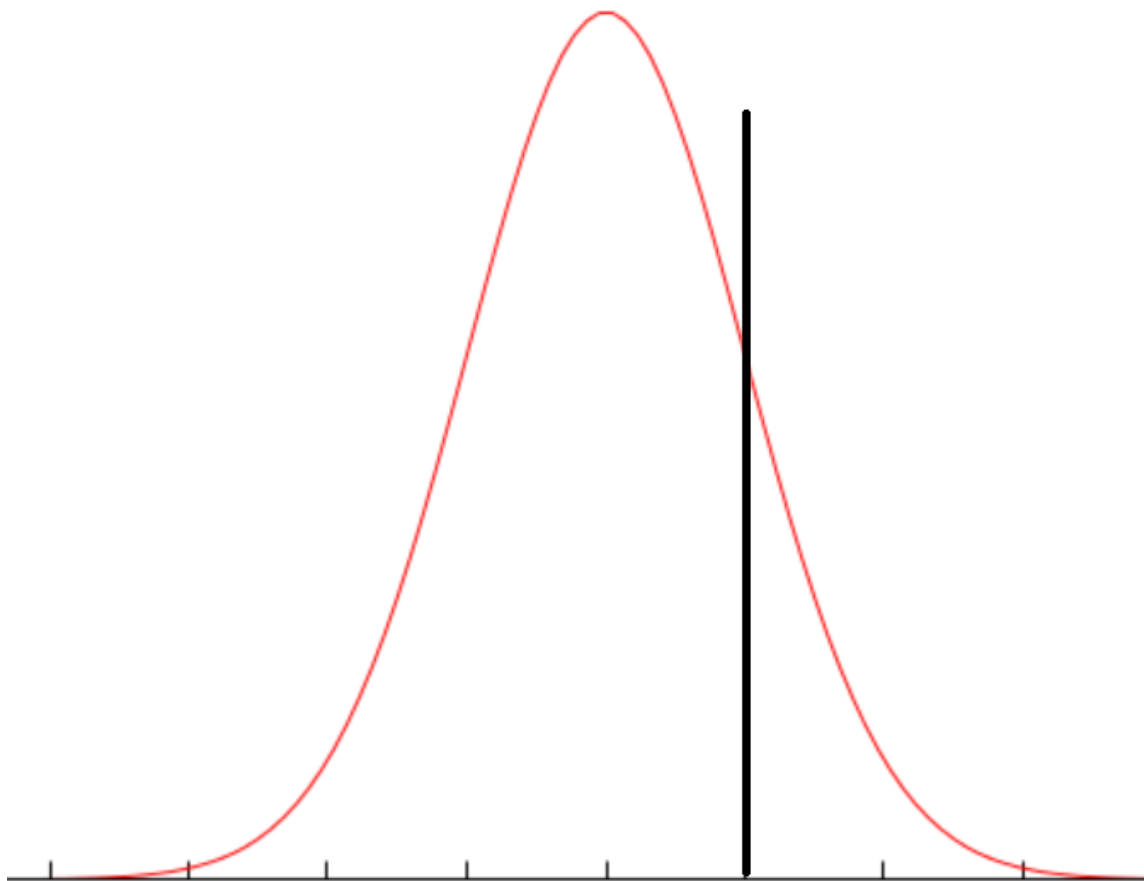
26 февраля 2020

Оценки школьников в Мумбаи: $ATE^{\wedge} = 6$. Можно ли доверять результату?



Здесь и далее графики – см из п.2 списка литературы на последнем слайде

Точечная и интервальная оценка. Доверительный интервал



«Презумпция нуля»

- «Бремя доказательства» лежит на исследователе: важно доказать, что эффект значим.
- Аналогично именно на исследователе лежит бремя доказательства, что оценка показывает именно заявленный эффект, что отсутствует «влияние третьих факторов» и т.д.

Ошибки первого и второго рода

	Тритмент-эффект есть	Тритмент-эффекта нет
Тест в пользу H1 Оценка тритмент- эффекта значимая	Ок	Ошибка 1 рода (вероятность = альфа)
Тест в пользу H0 Оценка тритмент- эффекта незначимая	Ошибка 2 рода (вероятность = κ)	Ок

Уровень значимости

- Уровень значимости 5% означает, вероятность случайно получить отличный от нуля результат составляет 5%.

Мощность

- Это вероятность того, что если истинный эффект - размера b , то наш эксперимент будет в состоянии разграничить оценку этого эффекта и ноль.
- Это вероятность избежать ошибки 2 рода ($= 1 - k$).

Связь: мощность и...

- предполагаемый размер эффекта
- вариативность показателя
- размер выборки
- пропорция между тритмент- и контрольной группой

MDE (minimum detectable effect) size

The diagram shows the formula for Minimum Detectable Effect (MDE) size, enclosed in a rectangular box. Red arrows point from labels to specific parts of the formula: 'Effect Size' points to the entire formula; 'Power' points to the $t_{(1-\kappa)}$ term; 'Significance Level' points to the t_α term; 'Proportion in Treatment' points to the P in the denominator of the first square root; 'Variance' points to the σ^2 in the numerator of the second square root; and 'Sample Size' points to the N in the denominator of the second square root.

$$EffectSize = \left(t_{(1-\kappa)} + t_\alpha \right) * \sqrt{\frac{1}{P(1-P)}} * \sqrt{\frac{\sigma^2}{N}}$$

Источник: Glennerster, Takavarasha "Running randomized evaluations. A practical Guide", ch 6.

Пример: проект NERICA

- L. Casaburi, R. Glennerster, S. Kamara, T. Suri (2014) "Providing collateral and improving product market access for smallholder farmers in Sierra Leone"
- Части фермеров в Сьерра Леоне раздали новый сорт риса на посев.
- Выросла ли урожайность и выручка фермеров? – Опрос фермеров, но затраты на дорогу!
- Размер эффекта – по результатам «лабораторных испытаний» - как должна отличаться урожайность (97 кг с гектара)
- Уровень значимости и мощность – заданы
- Расчёт необходимого количества наблюдений

MDE: аналогия

$$EffectSize = \left(t_{(1-\kappa)} + t_{\alpha} \right) * \sqrt{\frac{1}{P(1-P)}} * \sqrt{\frac{\sigma^2}{N}}$$

Стандартная ошибка
оценки коэффициента

- Аналогия: расчётная t-статистика в тесте на значимость коэффициента показывает, на сколько стандартных ошибок оценка отличается от нуля.

Minimum detectable effect: аналогия

=>

Чтобы быть значимым на 5% уровне, коэффициент должен быть как минимум в “t_табличное” раз больше, чем его стандартная ошибка

=>

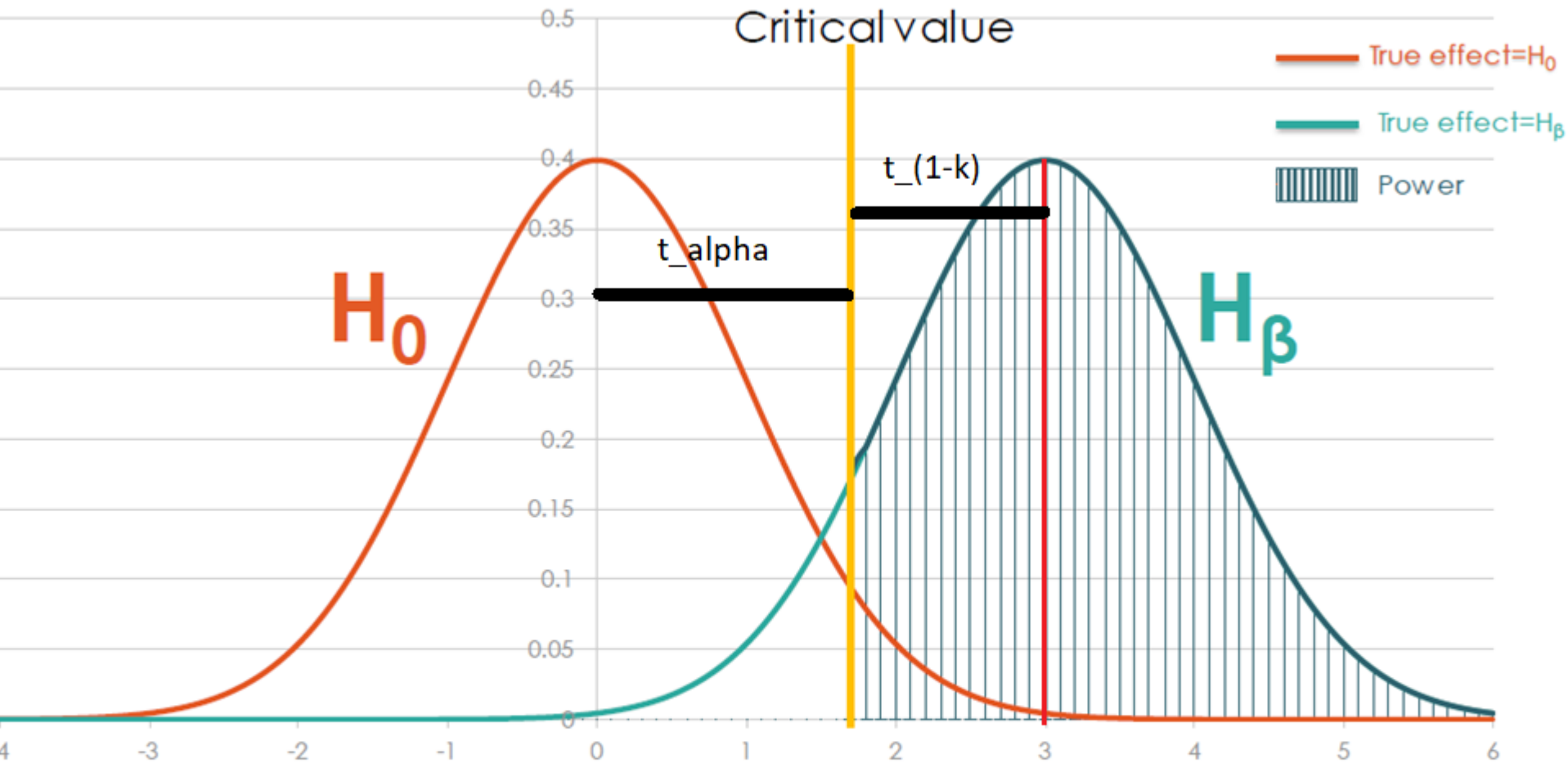
Минимально отличимый от нуля на alpha %-ном уровне коэффициент должен быть $= t_{\alpha/2} * s.e(b^{\wedge})$

Minimum detectable effect: наш случай

- Чтобы быть отличным от нуля при заданной мощности $(1-k)$ и уровне значимости (α) , коэффициент должен быть как минимум в $t_{(1-k)} + t_{\alpha}$ раз больше, чем s.e.

$$EffectSize = (t_{(1-\kappa)} + t_{\alpha}) * \sqrt{\frac{1}{P(1-P)}} * \sqrt{\frac{\sigma^2}{N}}$$

MDE графически



Что почитать

- Gerber, Green “Field experiments”, ch.4, 10
- Glennerster, Takavarasha “Running randomized evaluations. A practical Guide”, ch. 6
- Owen Ozier, лекция + выкладки + примеры
http://economics.ozier.com/owen/slides/ozier_powercalc_talk_20100914a.pdf
- Кратко + код: <https://egap.org/methods-guides/10-things-you-need-know-about-statistical-power>