ОДНОВЫБОРОЧНЫЕ КРИТЕРИИ СТЬЮДЕНТА

ВЕС ДЕТЕЙ ПРИ РОЖДЕНИИ

- Средний вес детей при рождении — 3.3 кг, у женщин, живущих за чертой бедности — 2.8 кг.
- 25 женщин, живущих за чертой бедности, участвовали в экспериментальной программе ведения беременности. Средний вес их детей при рождении составил 3075 г, стандартное отклонение 500 г.

ВЕС ДЕТЕЙ ПРИ РОЖДЕНИИ

- 25 женщин, живущих за чертой бедности, участвовали в экспериментальной программе ведения беременности. Средний вес их детей при рождении составил 3075 г, стандартное отклонение 500 г.

- Эффективна ли программа?

Z-КРИТЕРИЙ

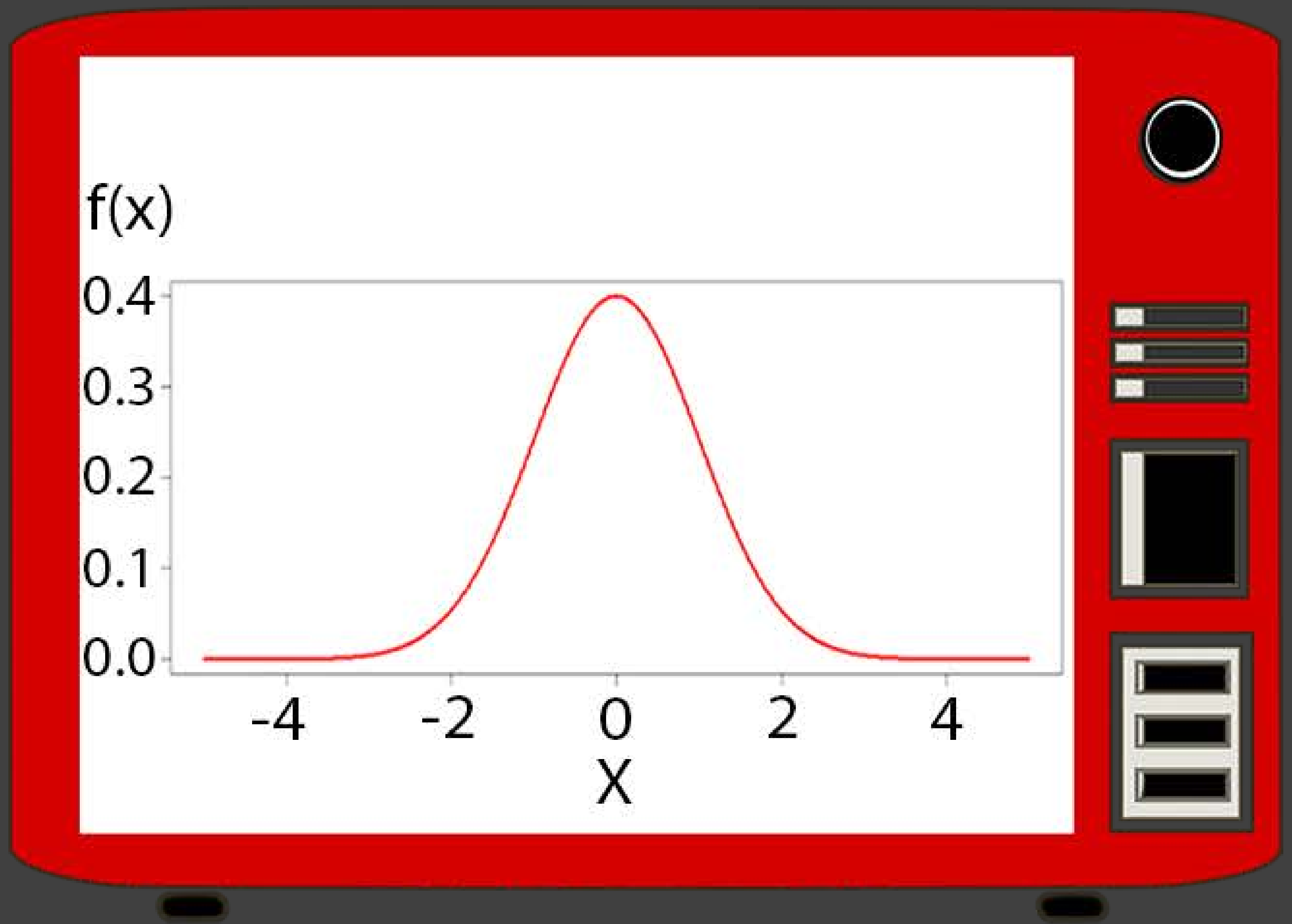
выборка: $X^n = (X_1, \ldots, X_n)$,
$X \sim N(\mu, \sigma^2)$, σ известна

нулевая гипотеза: $H_0\colon \mu = \mu_0$

альтернатива: $H_1\colon \mu <\neq> \mu_0$

статистика: $Z(X^n) = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

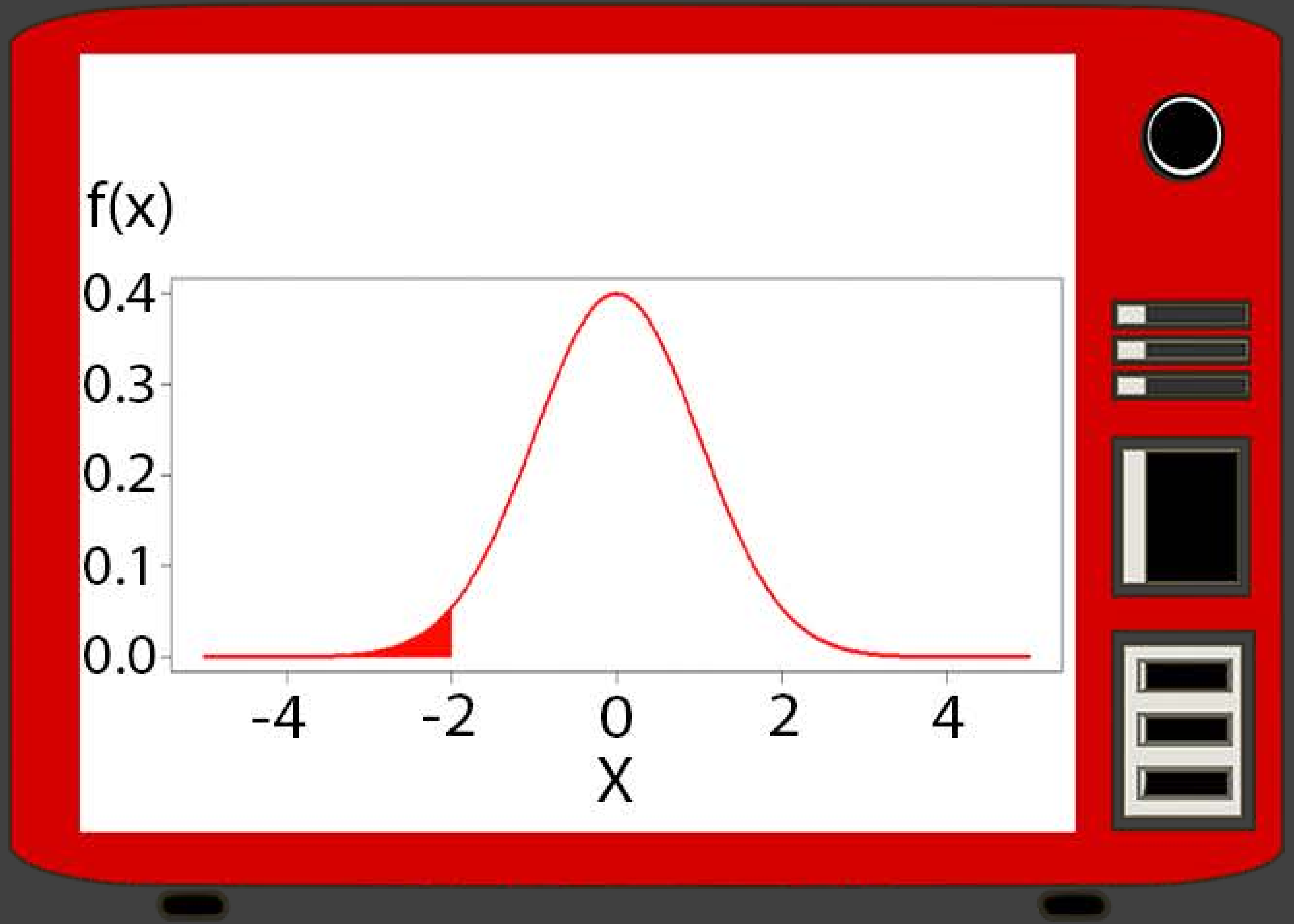
нулевое распределение: $Z(X^n) \sim N(0,1)$


f(x)
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
X

Z-КРИТЕРИЙ

- Достигаемый уровень значимости:
  - При $H_1: \mu < \mu_0$

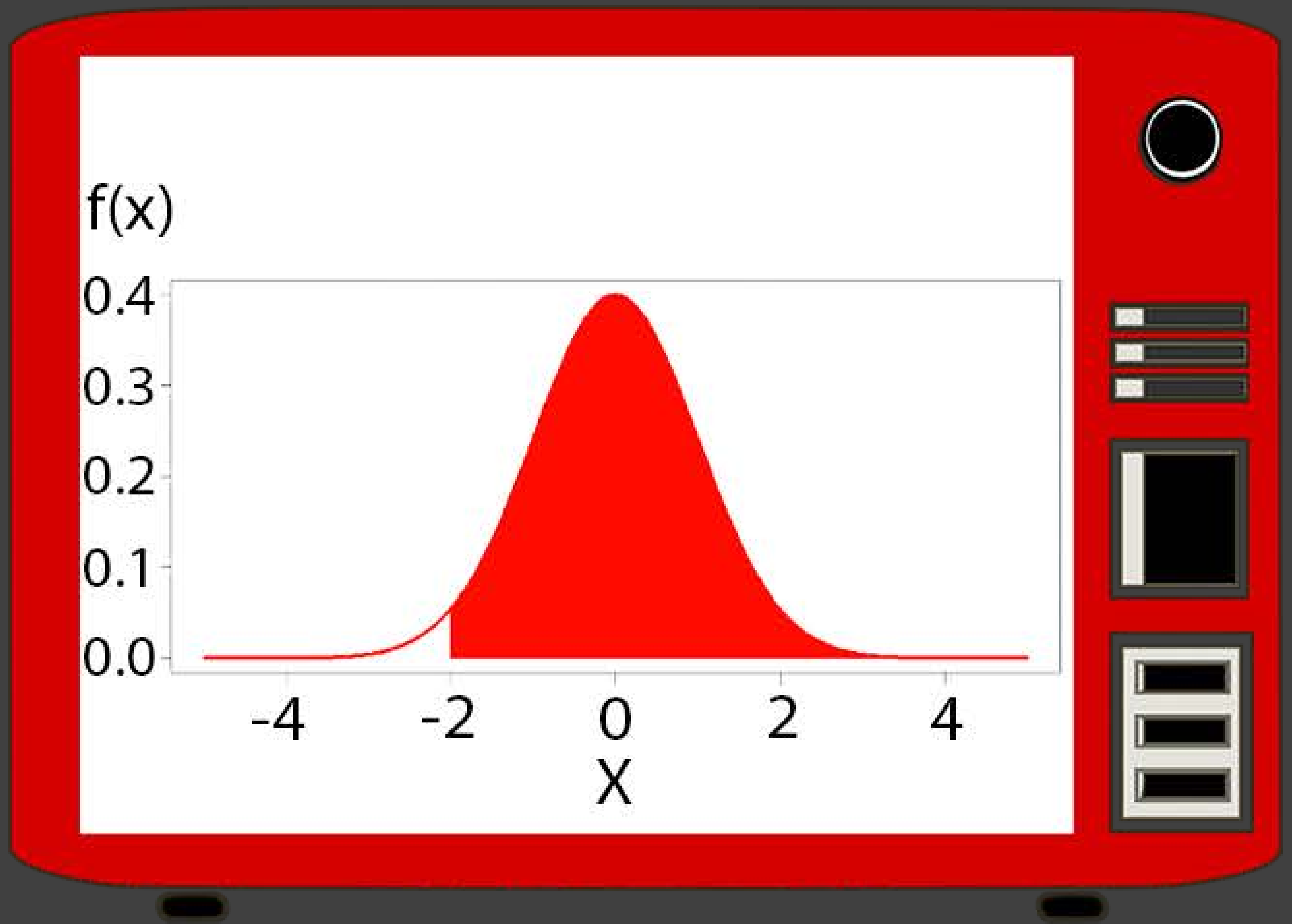
$$p = F_{N(0,1)}(z)$$


f(x)
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
X

Z-КРИТЕРИЙ

- Достигаемый уровень значимости:
  - При $H_1: \mu > \mu_0$

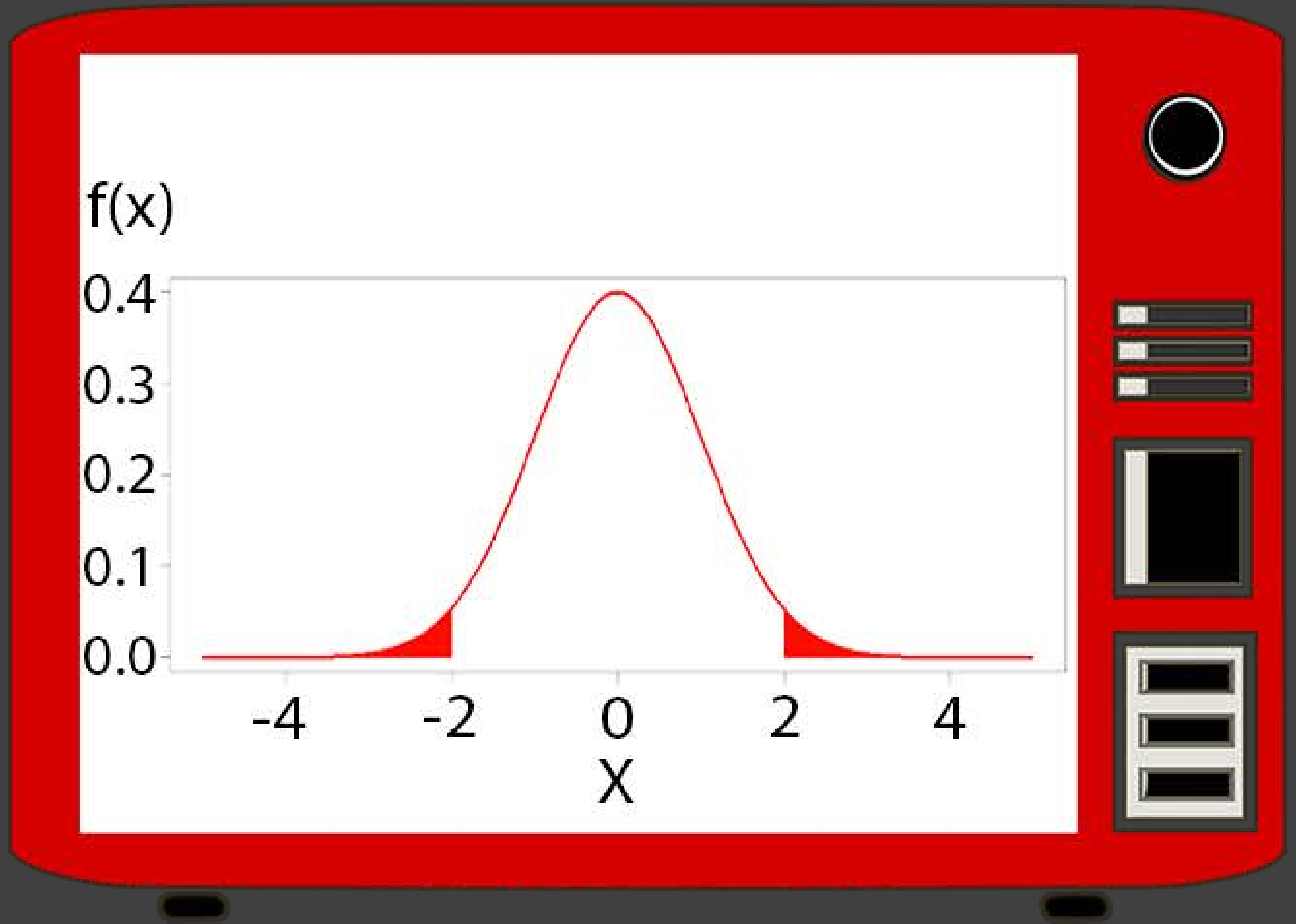
$$p = 1 - F_{N(0,1)}(z)$$


f(x)
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
X

Z-КРИТЕРИЙ

- Достигаемый уровень значимости:
  - При $H_1: \mu \neq \mu_0$

$$p = 2\left(1 - F_{N(0,1)}\left(|z|\right)\right)$$


f(x)
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
X

t-КРИТЕРИЙ

выборка: $X^n = (X_1, \ldots, X_n)$,

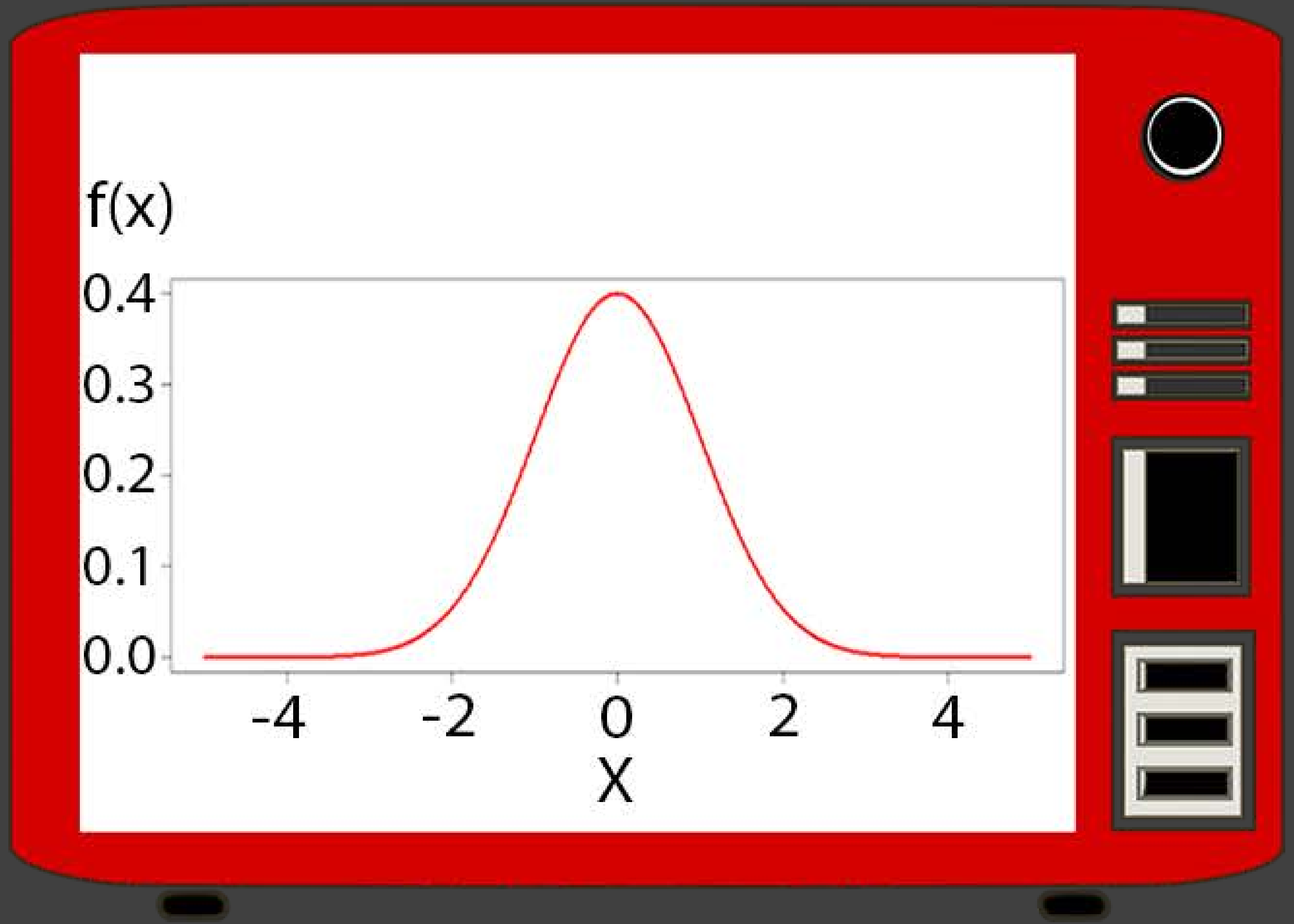
$X \sim N(\mu, \sigma^2)$, σ неизвестна

нулевая гипотеза: $H_0\colon \mu = \mu_0$

альтернатива: $H_1\colon \mu <\neq> \mu_0$

статистика: $T(X^n) = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

нулевое распределение: $T(X^n) \sim St(n-1)$


f(x)
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
X

t-КРИТЕРИЙ

- Достигаемый уровень значимости:

$$p = \begin{cases} F_{St(n-1)}(t), & H_1: \mu < \mu_0 \\ 1 - F_{St(n-1)}(t), & H_1: \mu > \mu_0 \\ 2\left(1 - F_{St(n-1)}(|t|)\right), & H_1: \mu \neq \mu_0 \end{cases}$$

t-КРИТЕРИЙ

- С ростом объёма выборки разница между t- и Z-критериями уменьшается.

Вес детей при рождении

- $H_0:$ программа неэффективна, $\mu = 2800$.
- $H_1:$ программа как-то влияет на вес детей, $\mu \neq 2800$.
- t-критерий: $p = 7.1 \times 10^{-13}$, средний вес детей увеличивается на 275 г (95% доверительный интервал — $[233.7, 316.3]$)

ВЕС ДЕТЕЙ ПРИ РОЖДЕНИИ

- $H_0:$ программа неэффективна, $\mu = 2800$
- $H_1:$ программа как-то влияет на вес детей, $\mu > 2800$
- t-критерий: $p = 3.55 \times 10^{-13}$, средний вес детей увеличивается на 275 г (нижний 95% доверительный предел — 240.7 г)

АЛЬТЕРНАТИВА

- Одностороннюю альтернативу можно использовать, если знак изменения среднего известен заранее.

- Альтернатива должна выбираться до получения данных!

РЕЗЮМЕ

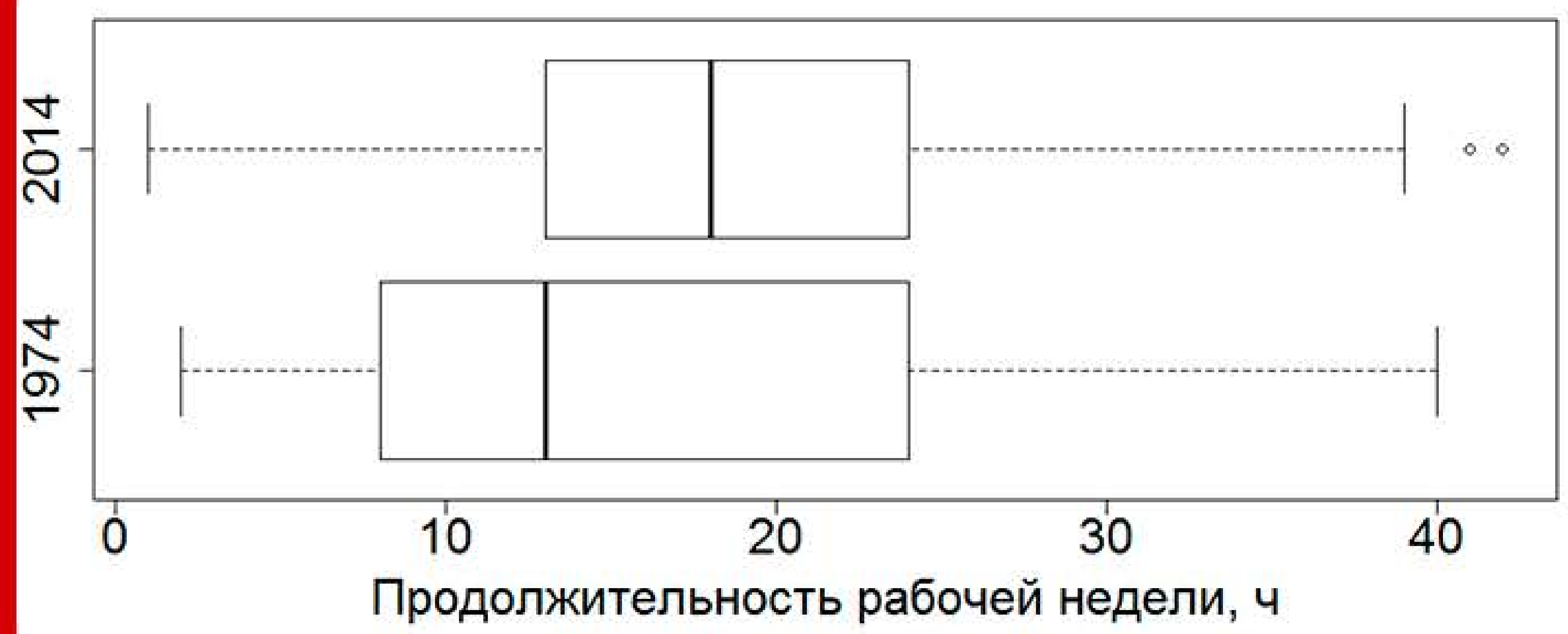
- Одновыборочные Z- и t-критерии Стьюдента
- Односторонняя и двусторонняя альтернатива
- Далее: двухвыборочные критерии Стьюдента

ДВУХВЫБОРОЧНЫЕ КРИТЕРИИ СТЬЮДЕНТА

ПРОДОЛЖИТЕЛЬНОСТЬ РАБОЧЕЙ НЕДЕЛИ

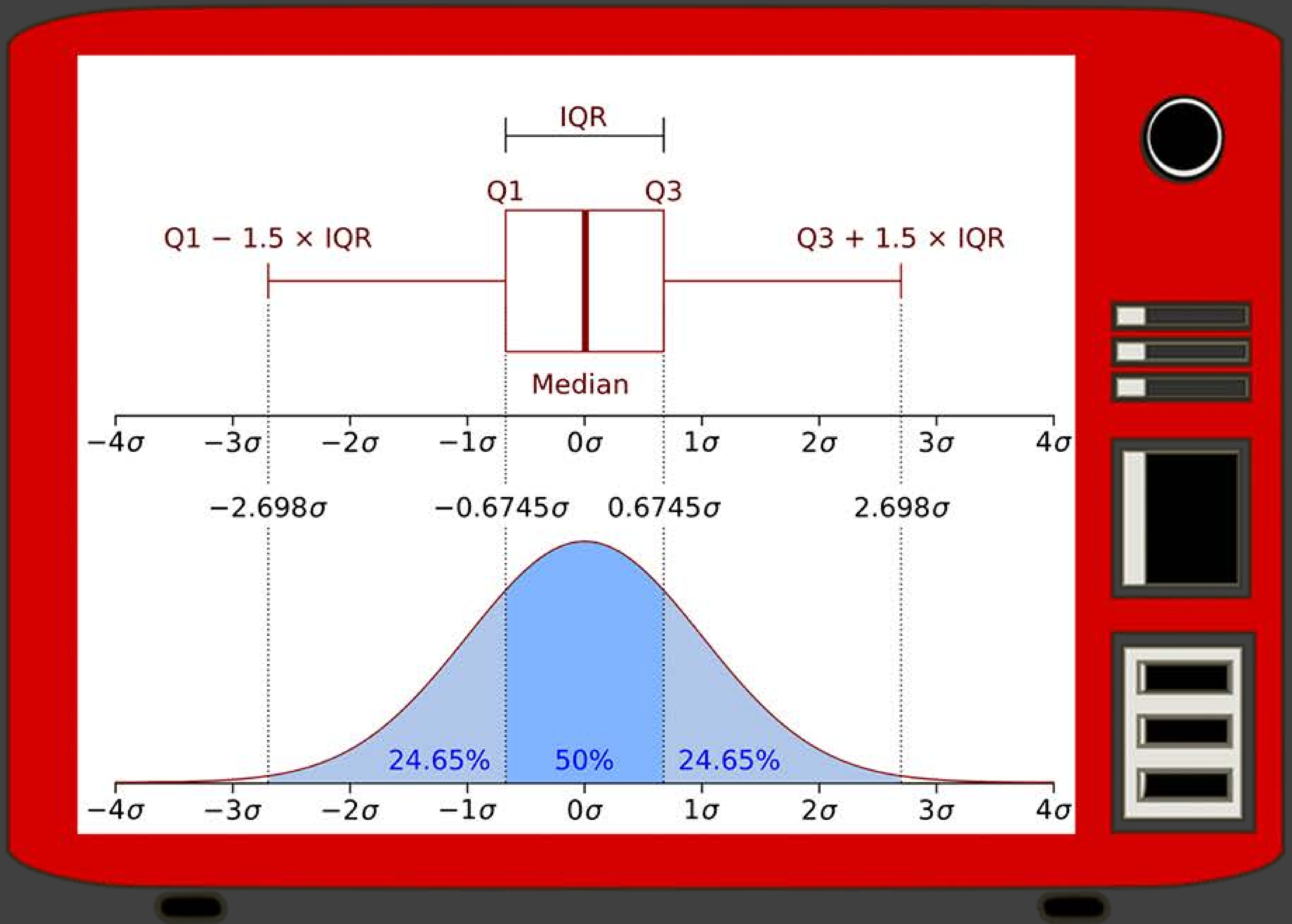
- В 1974 году 108 респондентов GSS работали неполный день, в 2014 — 196. Для каждого из них известно количество рабочих часов за неделю, предшествующую опросу

- Изменилось ли среднее время работы у работающих неполный день?


2014
1974
0
10
20
30
40
Продолжительность рабочей недели, ч

BOXPLOT

- Ящик с усами — способ визуализации основных характеристик распределения
- Длина усов отличается в разных реализациях


IQR
Q1
Q3
Q1 − 1.5 × IQR
Q3 + 1.5 × IQR
Median
−4σ −3σ −2σ −1σ 0σ 1σ 2σ 3σ 4σ
−2.698σ
−0.6745σ
0.6745σ
2.698σ
24.65%
50%
24.65%
−4σ −3σ −2σ −1σ 0σ 1σ 2σ 3σ 4σ

Z-КРИТЕРИЙ

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})$

$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})$

$X_1 \sim N\left(\mu_1, \sigma_1^2\right), X_2 \sim N\left(\mu_2, \sigma_2^2\right)$

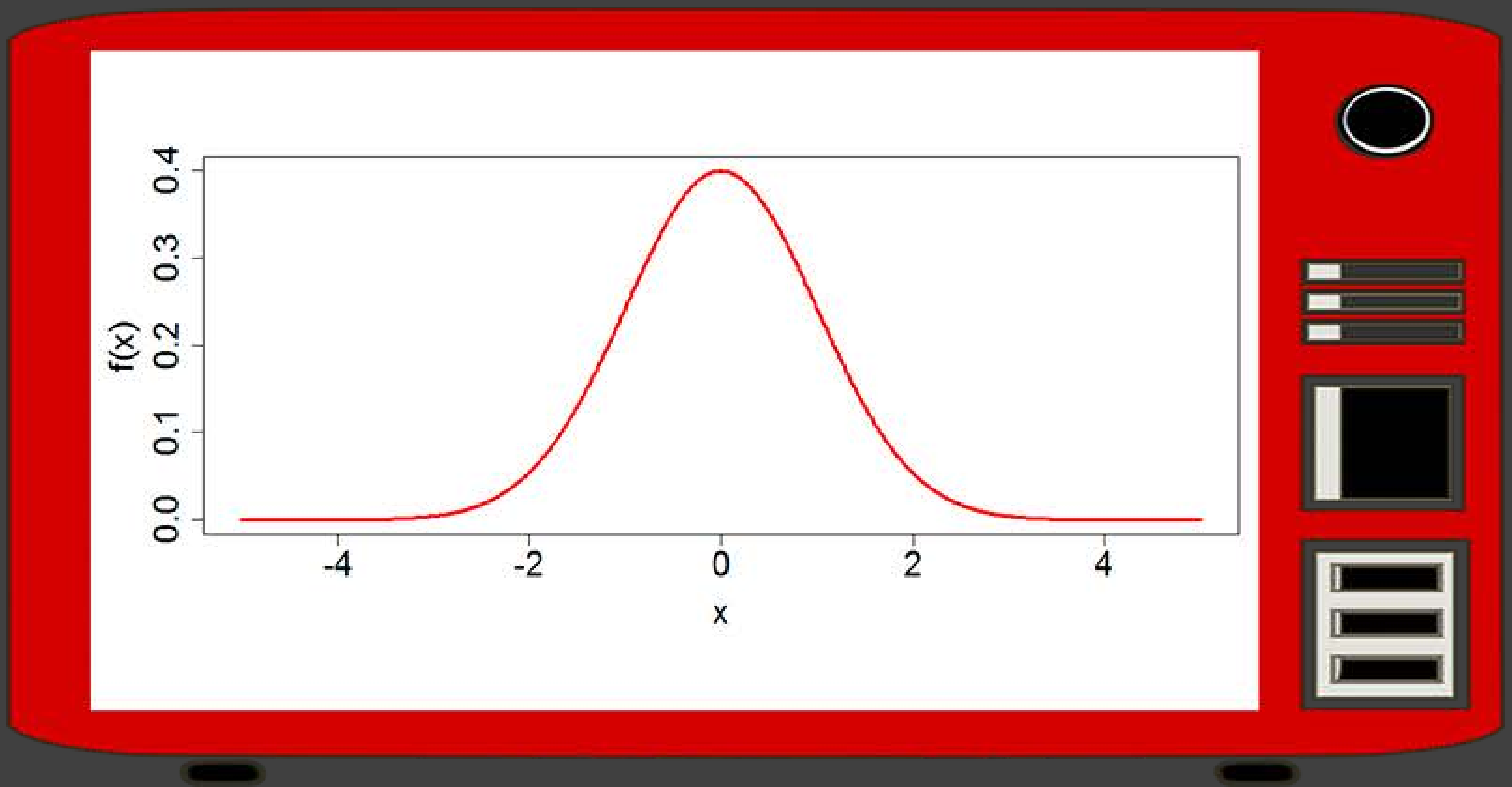
σ_1, σ_2 известны

нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

альтернатива: $H_1\colon \mu_1 < \neq > \mu_2$

статистика: $Z\left(X_1^{n_1}, X_2^{n_2}\right) = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

нулевое распределение: $Z\left(X_1^{n_1}, X_2^{n_2}\right) \sim N(0, 1)$



f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

t-КРИТЕРИЙ

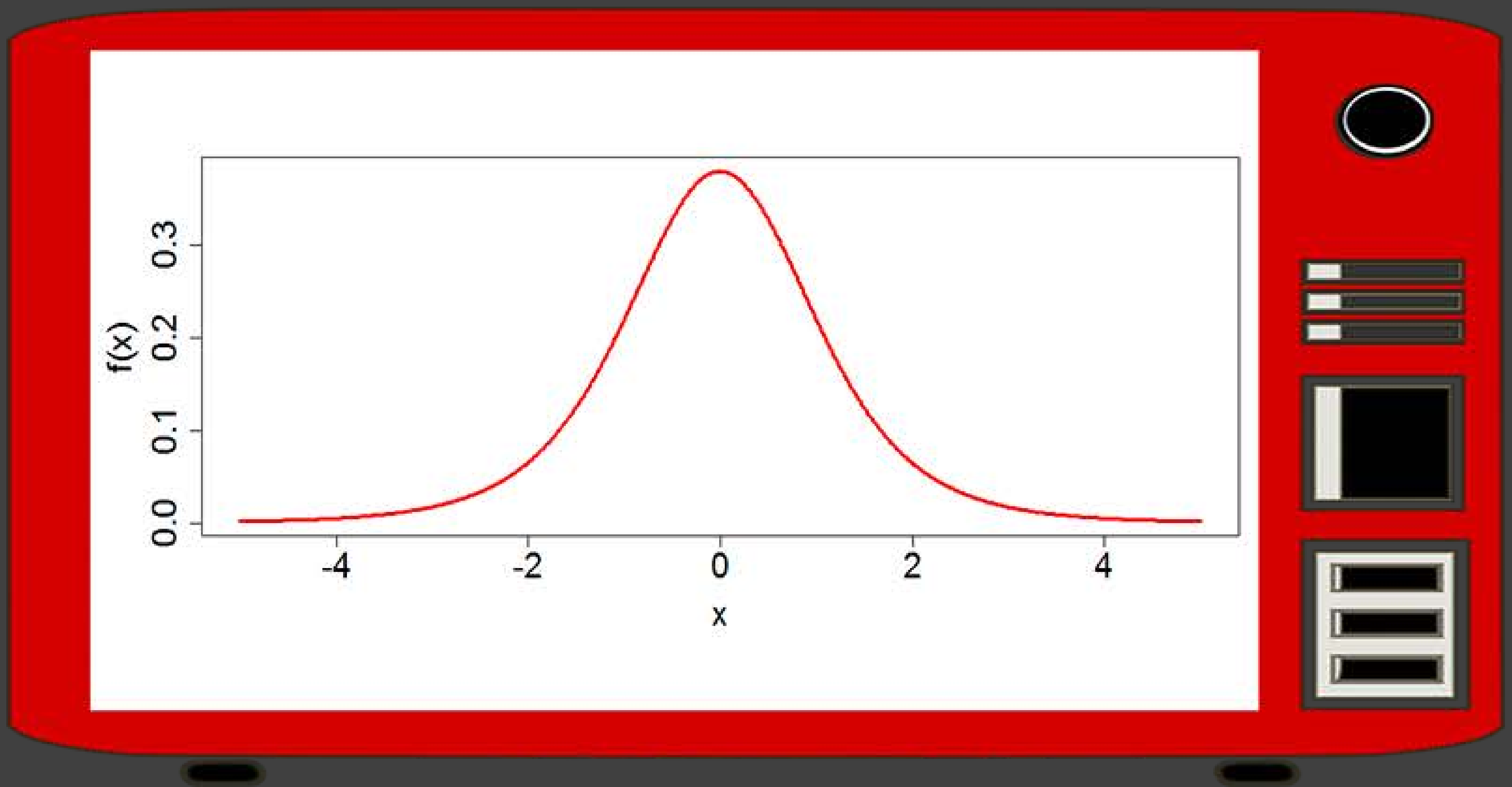
выборки: $X_1^{n_1} = (X_{11}, \ldots, X_{1n_1})$
$X_2^{n_2} = (X_{21}, \ldots, X_{2n_2})$
$X_1 \sim N\left(\mu_1, \sigma_1^2\right)$, $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$
σ_1, σ_2 неизвестны

нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

альтернатива: $H_1\colon \mu_1 <\neq> \mu_2$

статистика: $$T\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

нулевое распределение: $T\left(X_1^{n_1}, X_2^{n_2}\right) \approx\sim St(\nu)$


f(x)
0.0
0.1
0.2
0.3
-4
-2
0
2
4
x

t-КРИТЕРИЙ

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

- Нулевое распределение приближённое, а не точное. Точного решения не существует! (проблема Беренца-Фишера)

t-КРИТЕРИЙ

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

- Приближение достаточно точно при $n_1 = n_2$ или $[n_1 > n_2] = [\sigma_1 > \sigma_2]$

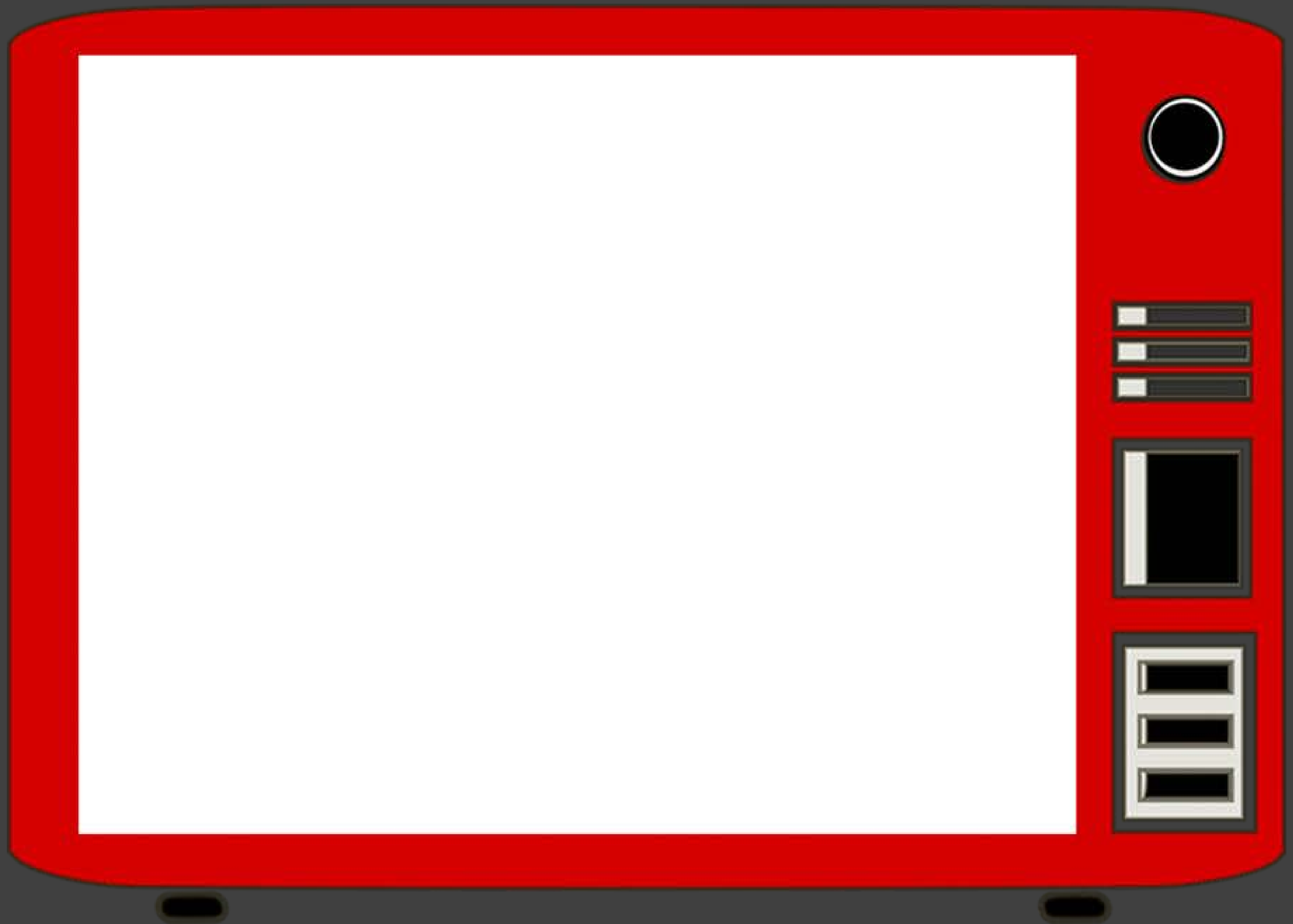
ПРОДОЛЖИТЕЛЬНОСТЬ РАБОЧЕЙ НЕДЕЛИ

- $\boldsymbol{H_0:}$ среднее время работы не изменилось, $\boldsymbol{\mu_1 = \mu_2}$
  $\boldsymbol{H_1:}$ среднее время работы изменилось, $\boldsymbol{\mu_1 \neq \mu_2}$

- t-критерий: $\boldsymbol{p = 0.02707}$, средняя продолжительность рабочей недели увеличилась на $\mathbf{2.57}$ часов ($\mathbf{95\%}$ доверительный интервал — $\mathbf{[0.29, 4.85]}$ ч).

РЕЗЮМЕ

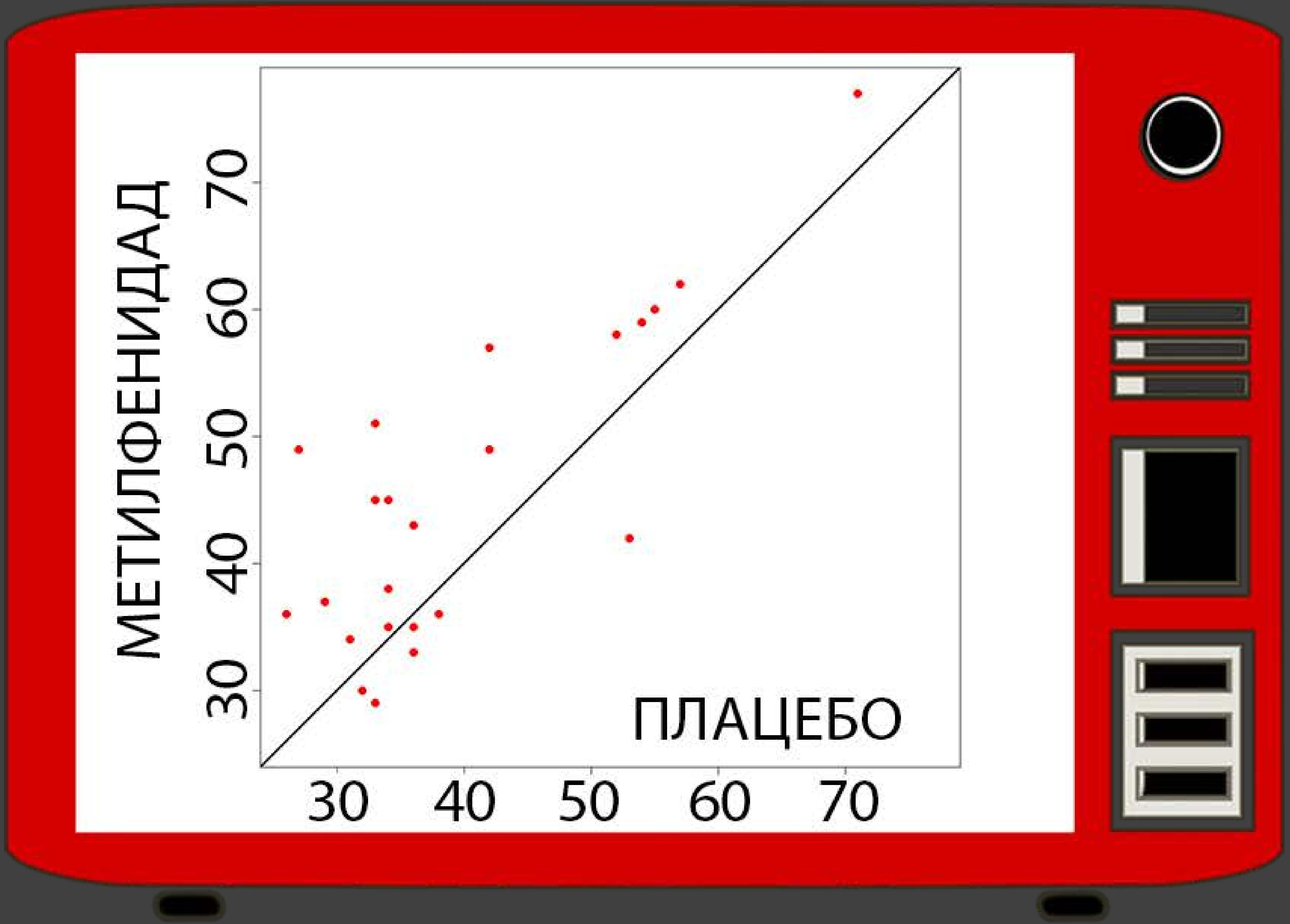
- Двухвыборочные Z- и t-критерии Стьюдента
- Проблема Беренца-Фишера
- Далее: связанные выборки



КРИТЕРИЙ СТЬЮДЕНТА ДЛЯ СВЯЗАННЫХ ВЫБОРОК

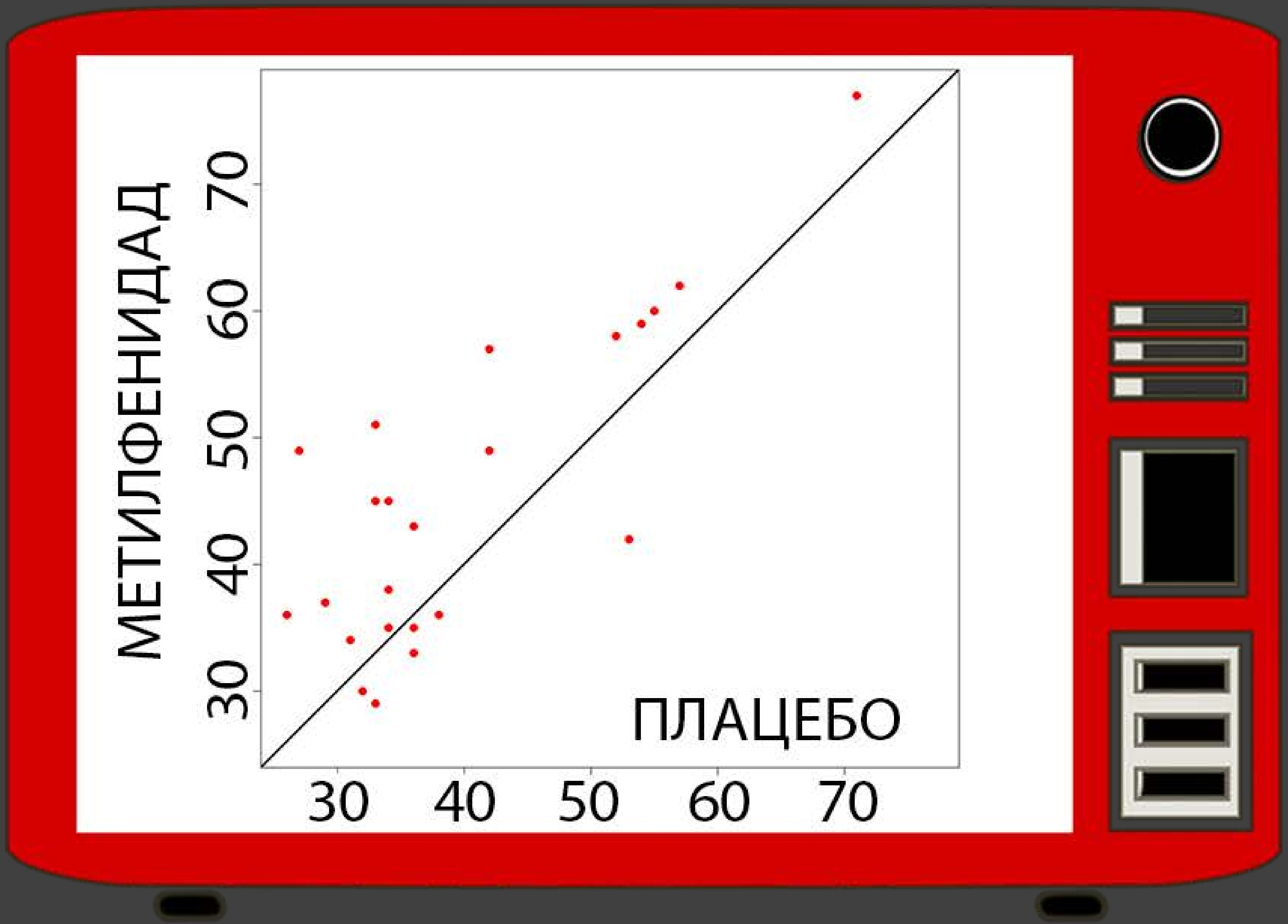
ЛЕЧЕНИЕ СДВГ

- 24 ребёнка прошли тест на способность к подавлению импульсивных поведенческих реакций после недели приёма метилфенидата и после недели приёма плацебо.


МЕТИЛФЕНИДАД
ПЛАЦЕБО
30
40
50
60
70
30
40
50
60
70

ЛЕЧЕНИЕ СДВГ

- Каков эффект препарата?


МЕТИЛФЕНИДАД
ПЛАЦЕБО
30
40
50
60
70
30
40
50
60
70

t-КРИТЕРИЙ ДЛЯ СВЯЗАННЫХ ВЫБОРОК

выборки: $X_1^n = (X_{11}, \ldots, X_{1n})\,, X_1 \sim N\left(\mu_1, \sigma_1^2\right)$

$X_2^n = (X_{21}, \ldots, X_{2n})\,, X_2 \sim N\left(\mu_2, \sigma_2^2\right)$

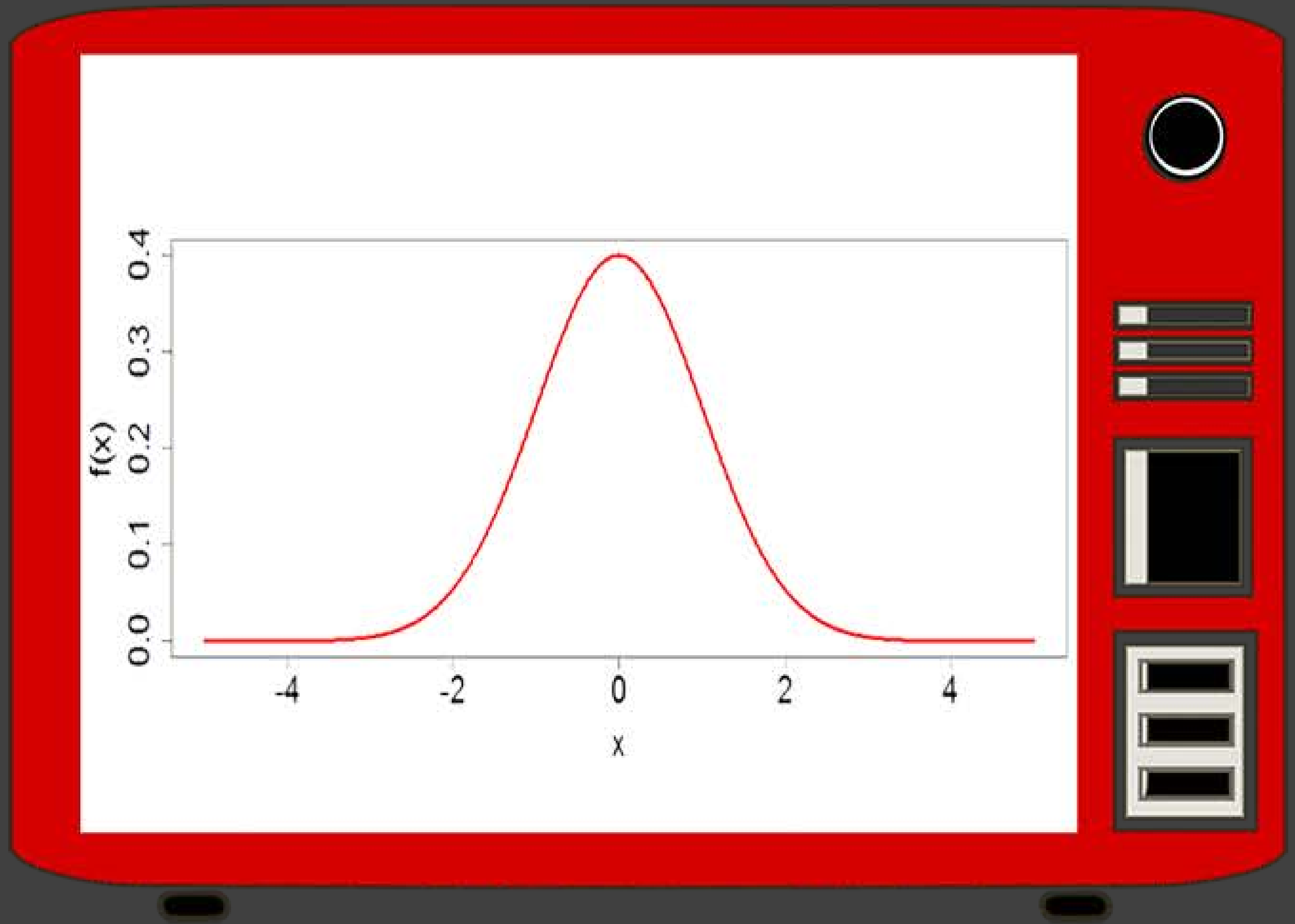
нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

альтернатива: $H_1\colon \mu_1 <\neq> \mu_2$

статистика: $T\left(X_1^n, X_2^n\right) = \dfrac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}}$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2, D_i = X_{1i} - X_{2i}$$

нулевое распределение: $T\left(X_1^n, X_2^n\right) \sim St(n-1)$



f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

t-КРИТЕРИЙ ДЛЯ СВЯЗАННЫХ ВЫБОРОК

- $\Leftrightarrow$ Переходим от пары связанных выборок к выборке их попарных разностей и применяем одновыборочный t-критерий.

ЛЕЧЕНИЕ СДВГ

- H_0 : способность к подавлению импульсивных поведенческих реакций не изменилась, $\mu_1 = \mu_2$.

- H_1 : способность к подавлению импульсивных поведенческих реакций изменилась, $\mu_1 \neq \mu_2$.

ЛЕЧЕНИЕ СДВГ

- t-критерий: $p = 0.00377$, средняя способность к подавлению импульсивных поведенческих реакций увеличилась на 4.95 пунктов (95% доверительный интервал — $[1.78, 8.14]$ пунктов).

Резюме

- Пять видов критериев Стьюдента
- Далее: нормальность выборок

НОРМАЛЬНОСТЬ ВЫБОРОК

КРИТЕРИЙ ХИ-КВАДРАТ

выборка: $X^n = (X_1, \ldots, X_n)$

нулевая гипотеза: $H_0\colon X \sim N(\mu, \sigma^2)$

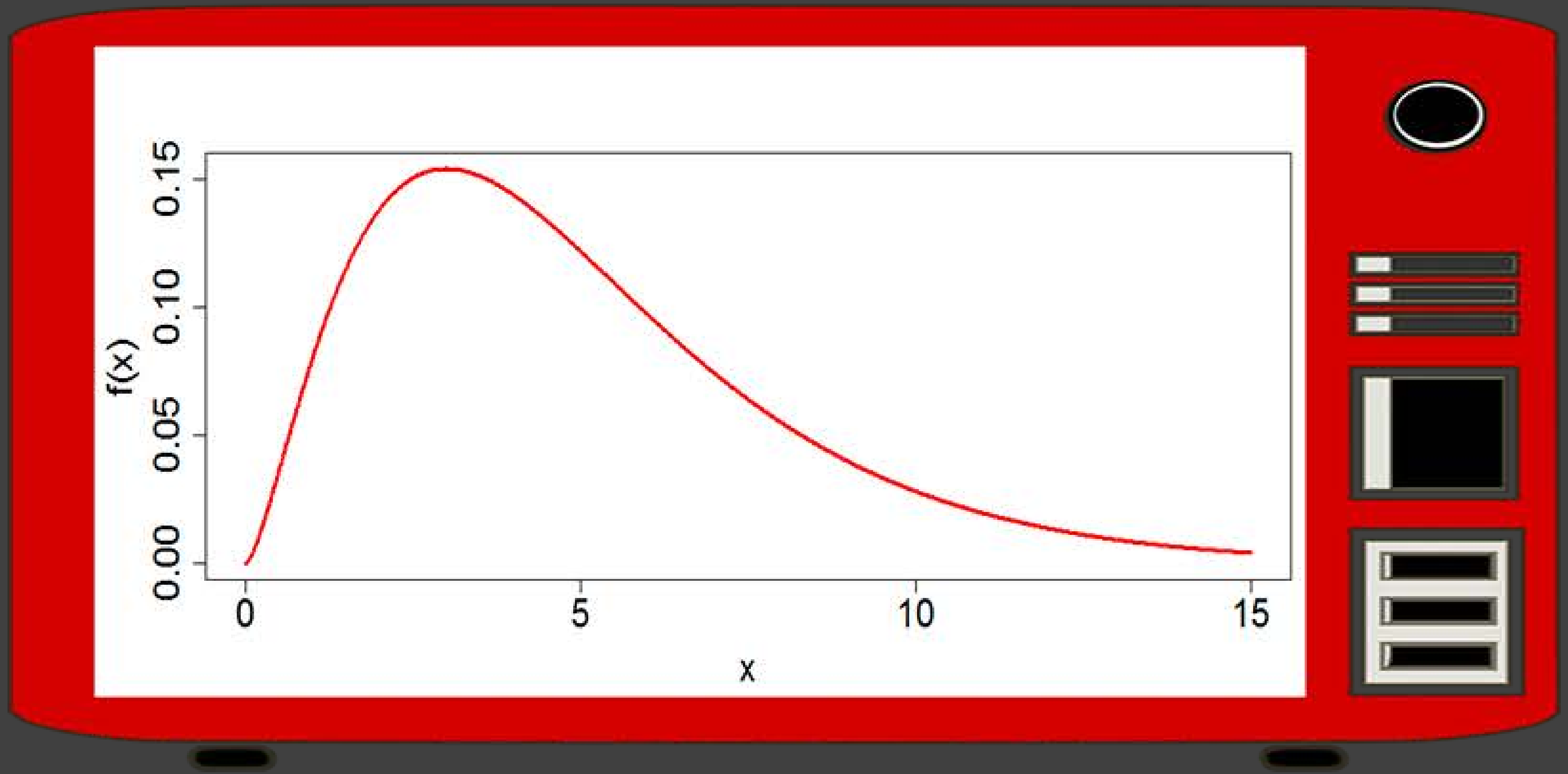
альтернатива: $H_1\colon H_0$ неверна

статистика: $\chi^2(X^n) = \sum_{i=1}^{K} \frac{(n_i - np_i)^2}{np_i}$

нулевое распределение: $\chi^2(X^n) \sim \begin{cases} \chi^2_{K-1}, & \mu,\sigma \text{ заданы} \\ \chi^2_{K-3}, & \mu,\sigma \text{ оцениваются} \end{cases}$

n_i— число элементов выборки в $[a_i, a_{i+1}]$

$p_i = F_{N(\mu,\sigma^2)}(a_{i+1}) - F_{N(\mu,\sigma^2)}(a_i)$



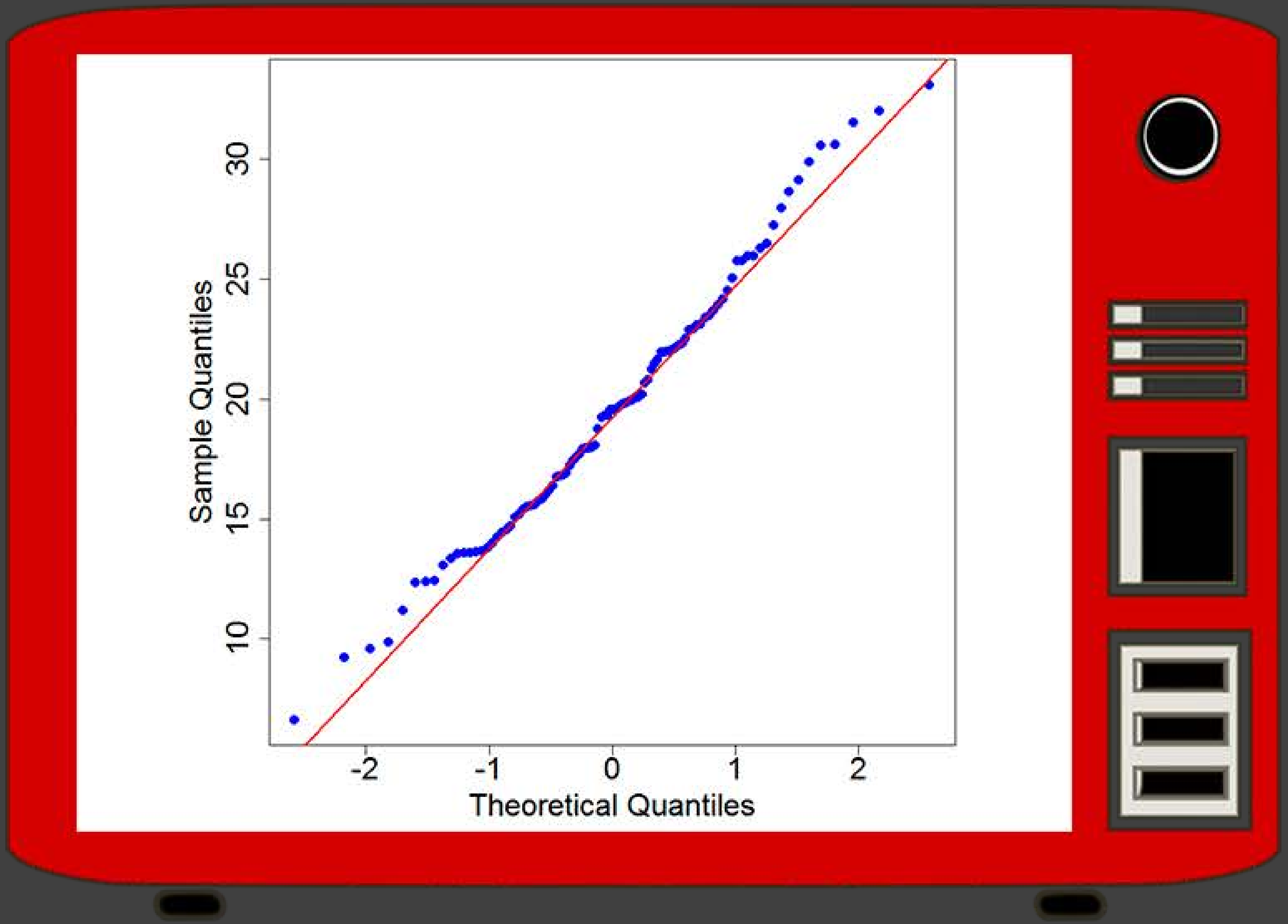
f(x)
0.00
0.05
0.10
0.15
0
5
10
15
x

КРИТЕРИЙ ХИ-КВАДРАТ

- Недостатки:
  - Разбиение на интервалы неоднозначно;
  - Требует больших выборок ($np_i > 5$ в 80% интервалов).

Q-Q PLOT

- Визуальный метод проверки согласия выборки и распределения — ку-ку график:



Sample Quantiles
10
15
20
25
30
-2
-1
0
1
2
Theoretical Quantiles

КРИТЕРИЙ ШАПИРО-УИЛКА

выборка: $X^n = (X_1, \ldots, X_n)$

нулевая гипотеза: $H_0\colon X \sim N\left(\mu, \sigma^2\right)$

альтернатива: $H_1\colon H_0$ неверна

статистика: $$W\left(X^n\right) = \frac{\left(\sum\limits_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum\limits_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

нулевое распределение: табличное

КРИТЕРИЙ ШАПИРО-УИЛКА

- a_i основаны на матожиданиях порядковых статистик нормального распределения и также табулированы

- Критерий проверяет, сильно ли точки на ку-ку графике отклоняются от прямой

ДРУГИЕ КРИТЕРИИ ДЛЯ ПРОВЕРКИ НОРМАЛЬНОСТИ

- Харке-Бера, Колмогорова (Лиллиефорса), Крамера-фон Мизеса, Андерсона-Дарлинга, . . .

- ???

ЗАЧЕМ ПРОВЕРЯТЬ НОРМАЛЬНОСТЬ?

- На маленьких выборках нормальность, скорее всего, не отвергается

- На больших выборках нормальность, скорее всего, отвергается

- Многие методы нечувствительны к отклонениям от нормальности (например, критерии Стьюдента)

ЗАЧЕМ ПРОВЕРЯТЬ НОРМАЛЬНОСТЬ?

- Многие методы нечувствительны к отклонениям от нормальности (например, критерии Стьюдента)
- «Все модели неверны, но некоторые полезны» (Джордж Бокс)

КАК ПРОВЕРЯТЬ НОРМАЛЬНОСТЬ?

- Если данные явно ненормальны (например, бинарны или дискретны), нужно выбрать метод, специфичный для такого распределения

- Если на ку-ку графике не видно существенных отклонений от нормальности, можно сразу использовать методы, устойчивые к небольшим отклонениям

КАК ПРОВЕРЯТЬ НОРМАЛЬНОСТЬ?

- Если на ку-ку графике не видно существенных отклонений от нормальности, можно сразу использовать методы, устойчивые к небольшим отклонениям

- Если метод чувствителен к отклонениям от нормальности, проверять её рекомендуется критерием Шапиро-Уилка

КАК ПРОВЕРЯТЬ НОРМАЛЬНОСТЬ?

- Если метод чувствителен к отклонениям от нормальности, проверять её рекомендуется критерием Шапиро-Уилка

- Если нормальность отвергается, чувствительные методы, предполагающие нормальность, использовать нельзя!

РЕЗЮМЕ

- Предположение нормальности и способы его проверки
- Далее: применение в Python

ГИПОТЕЗЫ О ДОЛЯХ

ОТБОР ПРИСЯЖНЫХ

- Известный педиатр Бенджамин Спок был арестован за участие в антивоенной демонстрации. Его дело должен был рассматривать суд присяжных.

- Присяжные назначаются с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек; среди них оказалось только 90 женщин.

ОТБОР ПРИСЯЖНЫХ

- Присяжные назначаются с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек; среди них оказалось только 90 женщин.

- Был ли отбор беспристрастным?

Z-КРИТЕРИЙ ДЛЯ ДОЛИ

выборка: $X^n = (X_1, \dots, X_n)$

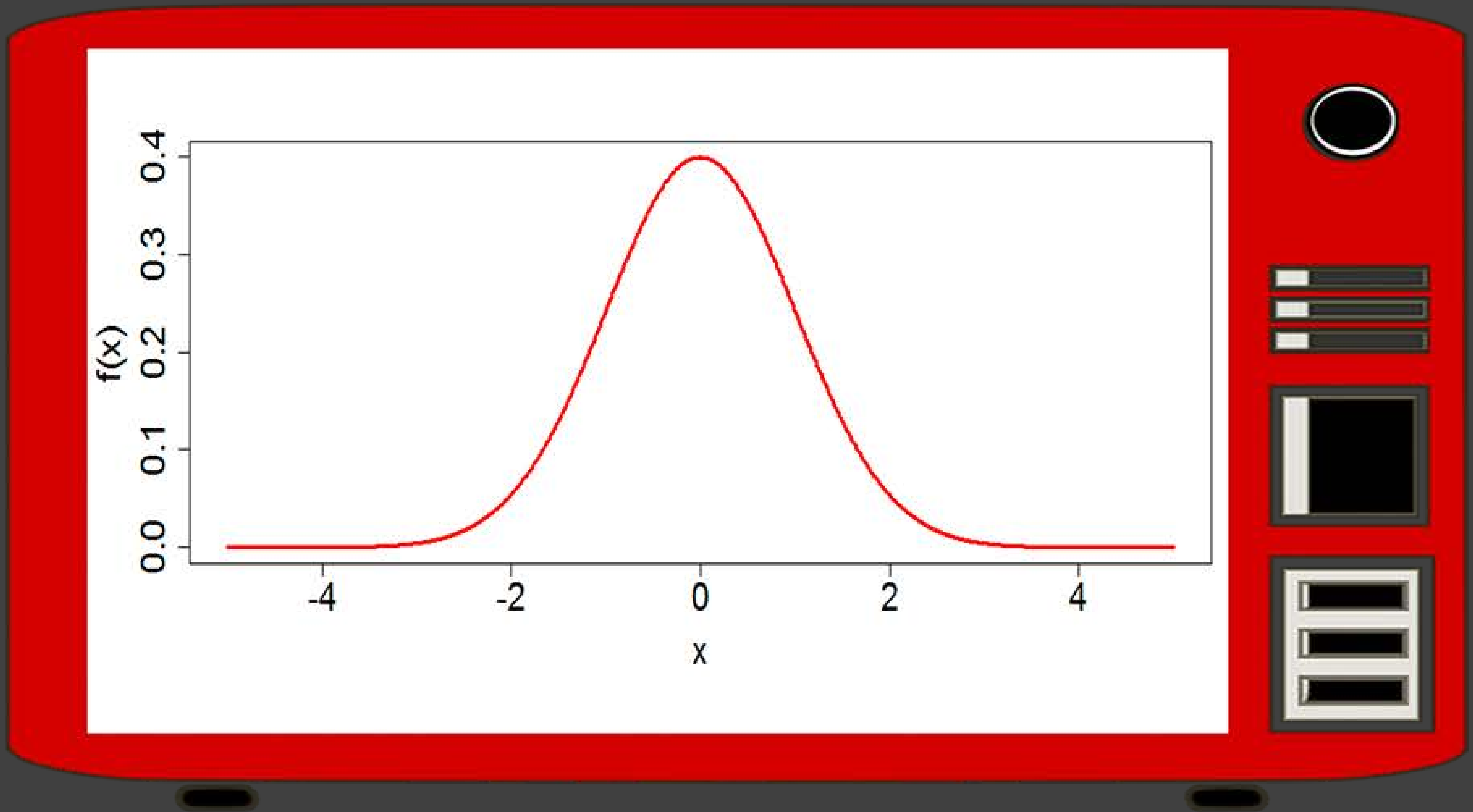
$X \sim Ber(p)$

нулевая гипотеза: $H_0 \colon p = p_0$

альтернатива: $H_1 \colon p <\neq> p_0$

статистика: $Z(X^n) = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \ \hat{p} = \bar{X}_n$

нулевое распределение: $Z(X^n) \sim N(0,1)$



f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

Отбор присяжных

- H_0 : процедура отбора была беспристрастной, женщины попадали в выборку с вероятностью 1/2
- H_1 : процедура отбора была смещённой
- Z-критерий: $p = 4.6 \times 10^{-12}$, женщины отбирались с вероятностью 0.3 (95% доверительный интервал — [0.248, 0.352])

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

- Из 1600 опрошенных граждан Великобритании, имеющих право голоса, 944 высказали одобрение деятельности премьер-министра.

- Через 6 месяцев одобрение высказали 880 из 1600 опрошенных.

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

- Через 6 месяцев одобрение высказали 880 из 1600 опрошенных.

Результат \ Опрос	I	II
+	$a = 944$	$b = 880$
-	$c = 656$	$d = 720$
$\sum$	$n_1 = 1600$	$n_2 = 1600$

- Изменился ли рейтинг?

Z-КРИТЕРИЙ ДЛЯ ДВУХ ДОЛЕЙ

выборки: $X_1^{n_1} = (X_{11}, \dots, X_{1n_1})\,, X_1 \sim Ber\,(p_1)$

$X_2^{n_2} = (X_{21}, \dots, X_{2n_2})\,, X_2 \sim Ber\,(p_2)$

выборки независимы

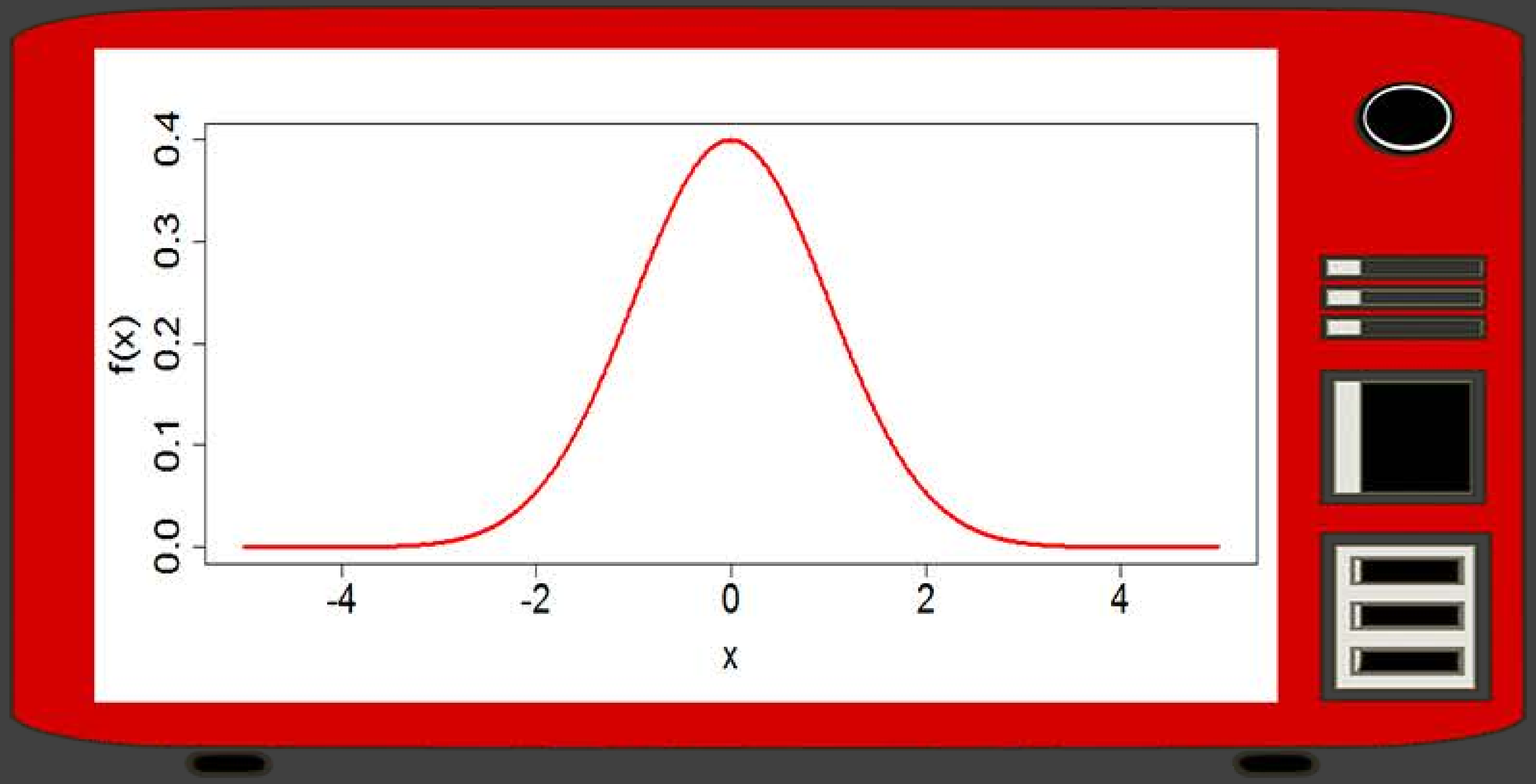
нулевая гипотеза: $H_0\colon p_1 = p_2$

альтернатива: $H_1\colon p_1 <\neq> p_2$

статистика: $$Z\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{P\,(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

нулевое распределение: $Z\left(X_1^{n_1}, X_2^{n_2}\right) \sim N(0,1)$



f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

Z-КРИТЕРИЙ ДЛЯ ДВУХ ДОЛЕЙ

Выборка / Исход	X_1	X_2
1	a	b
0	c	d
$\sum$	n_1	n_2

- При независимых выборках Z-критерий использует только первую строку таблицы:

$$\hat{p}_1 = \frac{a}{n_1}, \; \hat{p}_2 = \frac{b}{n_2}.$$

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

Результат \ Опрос	I	II
+	$a = 944$	$b = 880$
-	$c = 656$	$d = 720$
$\sum$	$n_1 = 1600$	$n_2 = 1600$

- H_0 : рейтинг не изменился
- H_1 : рейтинг изменился

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

- H_0 : рейтинг не изменился
- H_1 : рейтинг изменился
- Z-критерий: $p = 0.022$, рейтинг упал на 4% (95% доверительный интервал — [0.6, 7.4]%)

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

- На самом деле в двух опросах участвовали одни и те же люди:

I \ II	+	-	$\sum$
+	794	150	944
-	86	570	656
$\sum$	880	720	1600

- Изменился ли рейтинг?

Z-КРИТЕРИЙ ДЛЯ ДВУХ ДОЛЕЙ

X_1 \ X_2	1	0	$\sum$
1	e	f	$e+g$
0	g	h	$g+h$
$\sum$	$e+g$	$f+h$	n

Z-КРИТЕРИЙ ДЛЯ ДВУХ ДОЛЕЙ

выборки: $X_1^n = (X_{11}, \ldots, X_{1n})$, $X_1 \sim Ber(p_1)$

$X_2^n = (X_{21}, \ldots, X_{2n})$, $X_2 \sim Ber(p_2)$

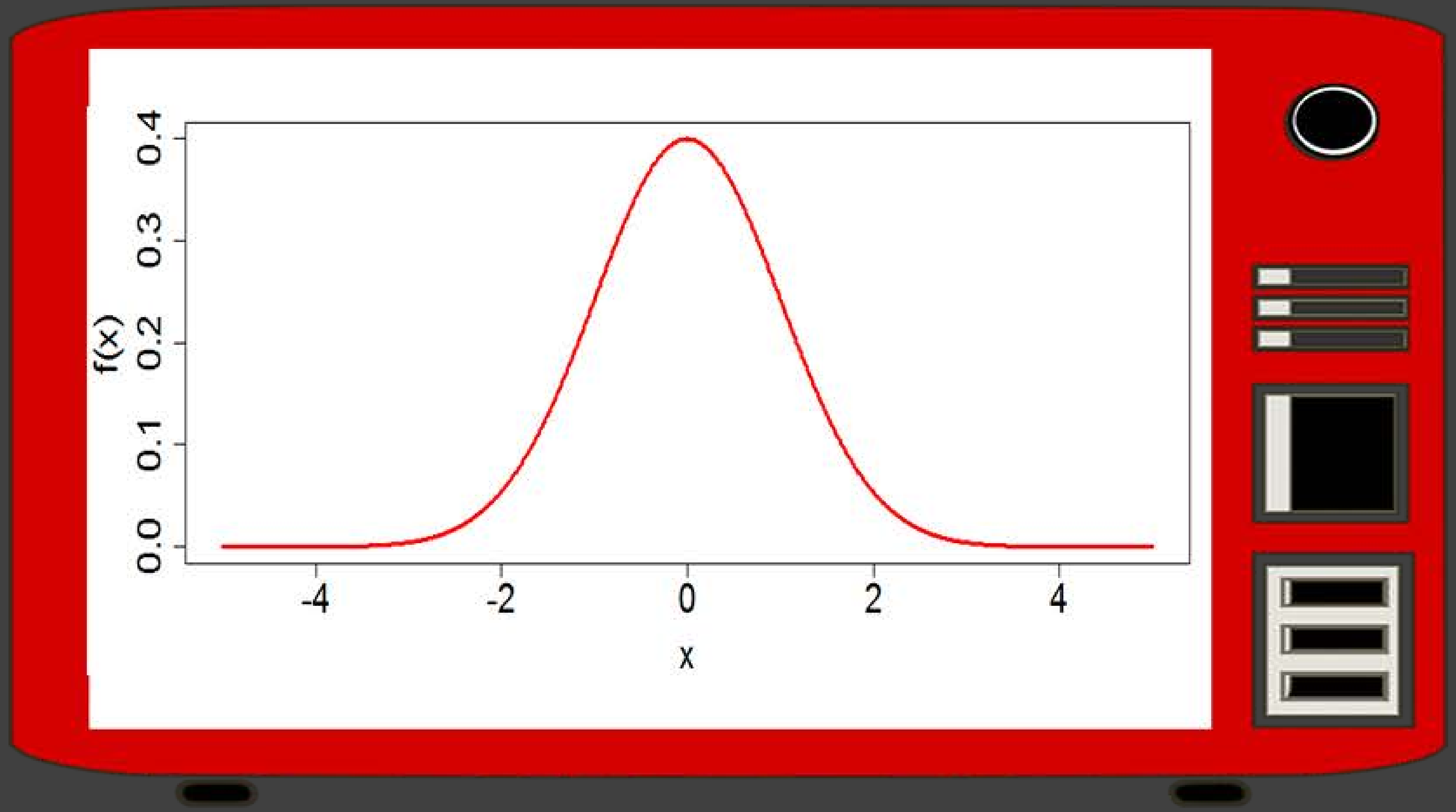
выборки связанные

нулевая гипотеза: $H_0 \colon p_1 = p_2$

альтернатива: $H_1 \colon p_1 <\neq> p_2$

статистика: $Z\left(X_1^n, X_2^n\right) = \dfrac{f-g}{\sqrt{f+g-\frac{(f-g)^2}{n}}}$

нулевое распределение: $Z\left(X_1^n, X_2^n\right) \sim N(0,1)$ при $H_0 = 0$



f(x)
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
x

Z-КРИТЕРИЙ ДЛЯ ДВУХ ДОЛЕЙ

X_1^n \ X_2^n	1	0	$\sum$
1	e	f	$e+f$
0	g	h	$g+h$
$\sum$	$e+g$	$f+h$	n

- При связанных выборках Z-критерий использует только внедиагональные элементы таблицы.

РЕЙТИНГ ПРЕМЬЕР-МИНИСТРА

I \ II	+	-	$\sum$
+	794	150	944
-	86	570	656
$\sum$	880	720	1600

- H_0 : рейтинг не изменился
- H_1 : рейтинг изменился

Рейтинг премьер-министра

- H_0 : рейтинг не изменился
- H_1 : рейтинг изменился
- Z-критерий: $p = 2.8 \times 10^{-5}$, рейтинг упал на 4% (95% доверительный интервал — [2.1, 5.8]%)

РЕЗЮМЕ

- Z-критерии для проверки гипотез о долях:
  - Одна выборка
  - Две независимые выборки
  - Две связные выборки

РЕЗЮМЕ

- Z-критерии для проверки гипотез о долях:
  - Одна выборка
  - Две независимые выборки
  - Две связные выборки

- Далее: применение в Python