Лекция 6

Проверка гипотез

Основная и альтернативная гипотезы. Статистический критерий.

Основная гипотеза $\mathbf{H_0}$ — утверждение, которое мы не хотим отвергать без явных свидетельств его ошибочности,

Альтернативная, или конкурирующая, гипотеза $\mathbf{H_A}$ — то отклонение от основной гипотезы, которое нам важно выявить (если оно имеет место).

Пример. При испытании лекарства основная гипотеза — лекарство недейственно (действует не лучше плацебо или другого доступного способа лечения), альтернативная — лекарство способствует улучшению состояния пациента в большей мере, чем плацебо или принятое лечение.

Точнее, это лишь на одном из этапов испытания...

Формально, статистическая гипотеза — множество функций распределения.
Простая гипотеза — та, которая состоит из одной функции распределения.
Сложная — более чем из одной.

$\mathbf{H_0}$ и $\mathbf{H_A}$ должны не пересекаться.

Статистический критерий (тест, решающее правило) — правило, по которому принимается решение, отвергнуть основную гипотезу в пользу альтернативной или нет, в зависимости от результатов наблюдений.

Обычно критерии выглядят как-то так...

Берётся некая функция от случайной выборки $T=f(X_1,\dots,X_n)$ — мы будем её называть *статистикой*.

Статистика выбирается так, чтобы при верной основной гипотезе её распределение было нам в точности известно (например, N(0,1) или хи-квадрат с известным числом степеней свободы).

Затем задаётся *критическая область* — область значений статистики, при попадании в которую мы отвергаем основную гипотезу.

Как правило, критическая область берётся так, чтобы вероятность попасть туда при верной основной гипотезе была низкой, а при верной альтернативной — выше.

ОШИБКИ ПРИ ПРОВЕРКЕ ГИПОТЕЗ

нежелательны

неизбежны

Выбрана	**Верна**	
H_0	H_0	
H_A	H_0	— ошибка I рода
H_0	H_A	— ошибка II рода
H_A	H_A	

Что такое ошибки I и II рода в примере с лекарством?

Уровень значимости — допустимая вероятность отвергнуть основную гипотезу, когда она верна.

Мощность — вероятность отвергнуть основную гипотезу, когда верна альтернативная.

Функция мощности — функция, связывающая вероятность отвержения основной гипотезы с параметрами генеральной совокупности.

Мы не приписываем вероятностей гипотезам — они не случайны.
Случайность возникает при использовании критерия, потому что он опирается на случайную выборку.

Кар Карыч рассказывает Лосяшу про нулевую гипотезу

Пример

Известно, что $X \sim R(0, a)$. Исследователь проверяет гипотезу H_0: $a=10$ против H_A: $a>10$ с помощью следующего критерия: отвергнуть H_0 в пользу H_A, если $X>c$. Каким должно быть число c, чтобы обеспечить уровень значимости 10%? При таком c найдите мощность критерия как функцию от a.

Решение.
Для определения уровня значимости предполагаем, что основная гипотеза верна.
Тогда $X \sim R(0, 10)$. Ищем вероятность, с которой мы отвергнем H_0.

$$\begin{array}{ll} c \le 0: & P(X > c) = 1. \\ 0 < c < 10: & P(X > c) = 1-(c/10). \\ c \ge 10: & P(X > c) = 0. \end{array}$$

Определяем c, соответствующее уровню значимости 10%:

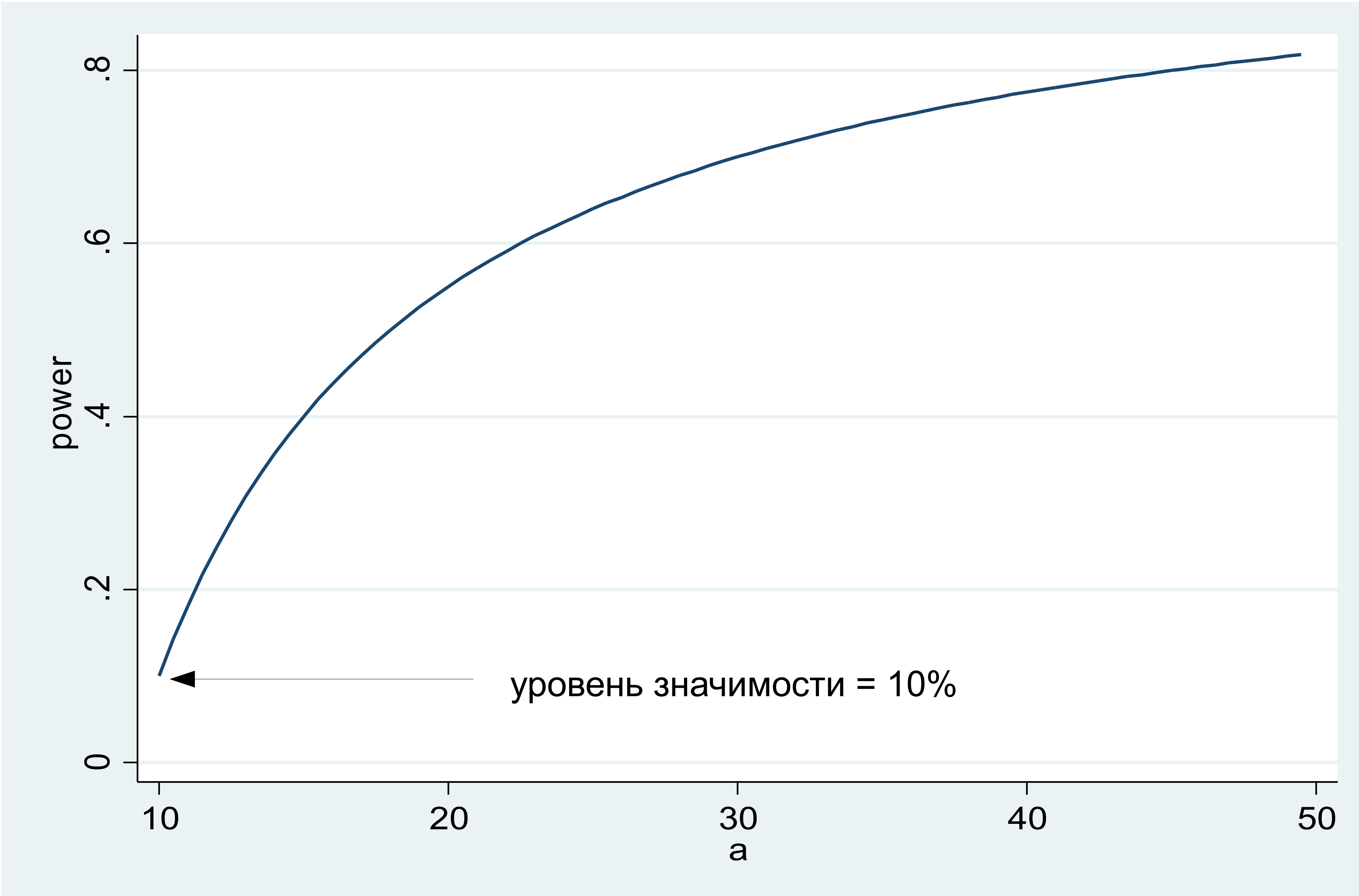
$$1 - (c/10) = 0.1 \Rightarrow c = 9.$$

Теперь ищем функцию мощности, связывающую вероятность отвержения H_0 со значением параметра a.

$$\text{Power}(a) = P_a(X > c) = P_a(X > 9) = 1 - (9/a).$$

Здесь учитывается то, что $a \ge 10$.

Функция мощности



Обычно уровень значимости выбирает сам исследователь.
Традиционные уровни значимости: 10%, 5%, 1%, 0.1%.

Проверка гипотез о параметрах нормального распределения

Вспомним предыдущие лекции.

Пусть $X_1, \ldots, X_n$ независимы, $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Тогда:

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1},$$

$$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Это нам пригодится!

Проверка гипотезы о среднем в нормальной генеральной совокупности

Пусть $X_1, \ldots, X_n$ независимы, $X_i \sim \mathrm{N}(\mu, \sigma^2)$.

$$\mathbf{H_0}: \mu = \mu_0$$

В качестве альтернативы обычно выбирается один из трёх вариантов в зависимости от цели исследования (выбираем то отклонение от H_0, которое нам важно выявить):

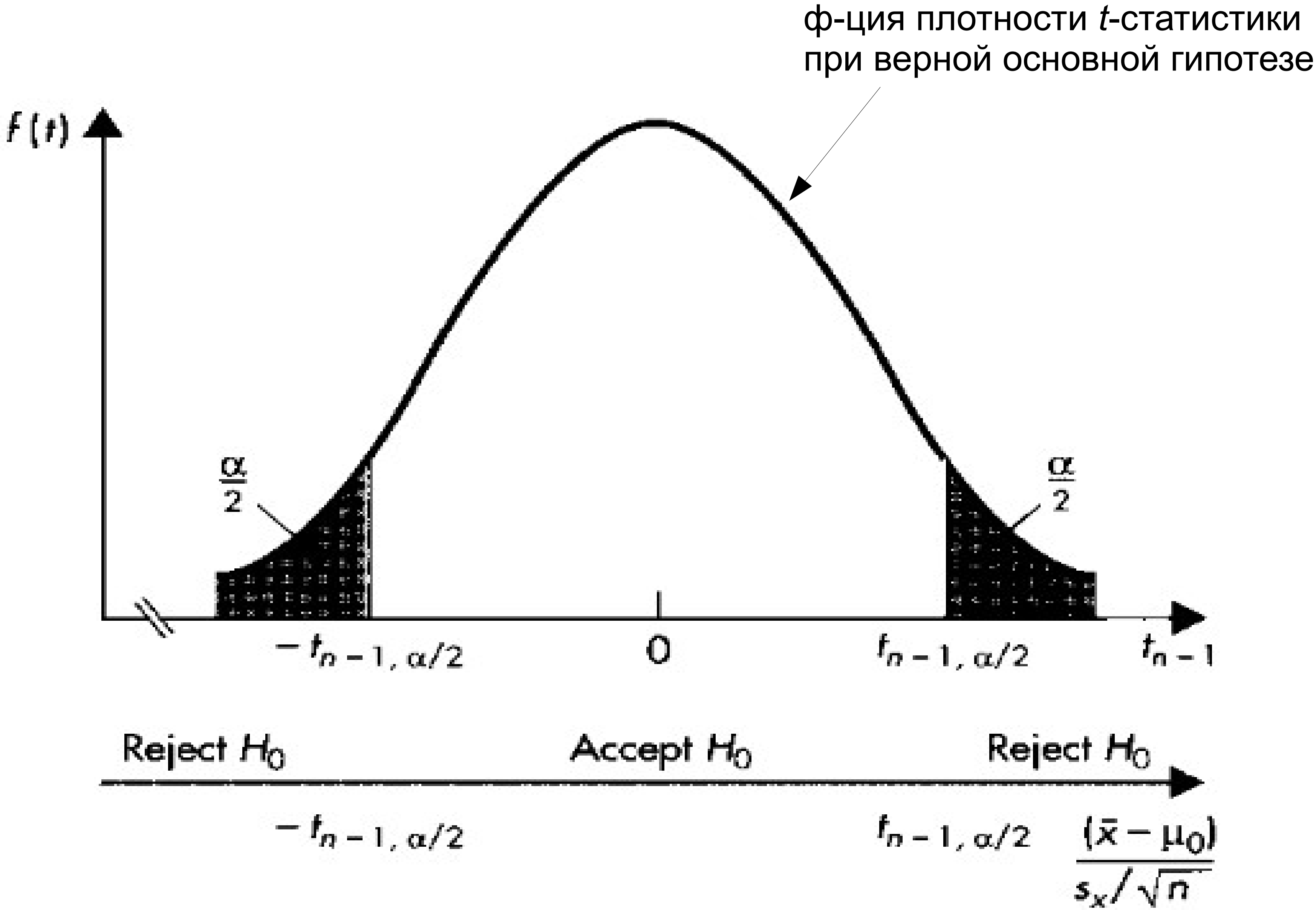
$$\mathbf{H_A}: \mu \neq \mu_0$$
$$\mathbf{H_A}: \mu > \mu_0$$
$$\mathbf{H_A}: \mu < \mu_0$$

Мы сами выбираем уровень значимости α.

Критическая статистика: $t = \dfrac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}.$

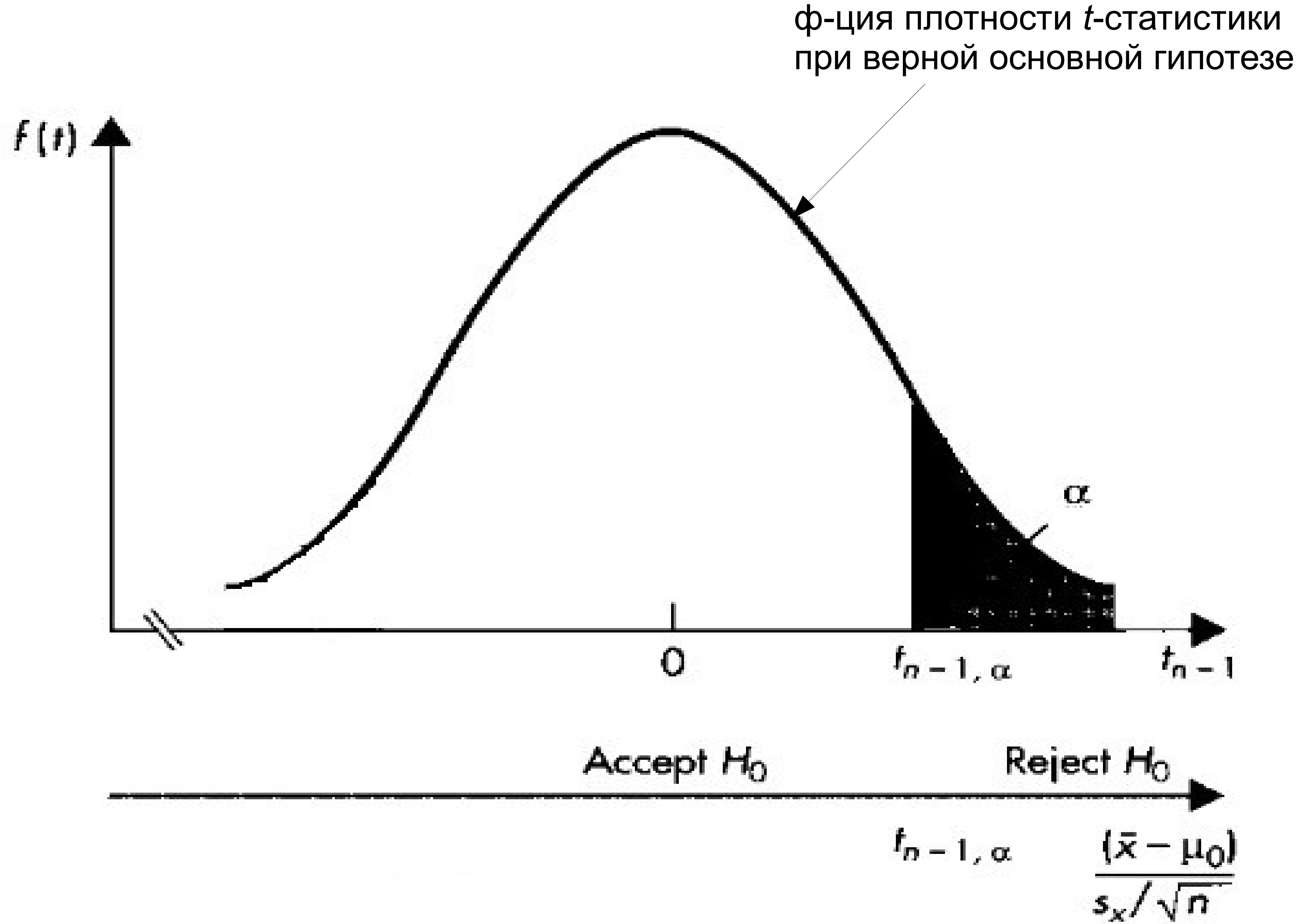
Как выбрать критическую область?

Критическая область при **H_A**: $\mu \neq \mu_0$



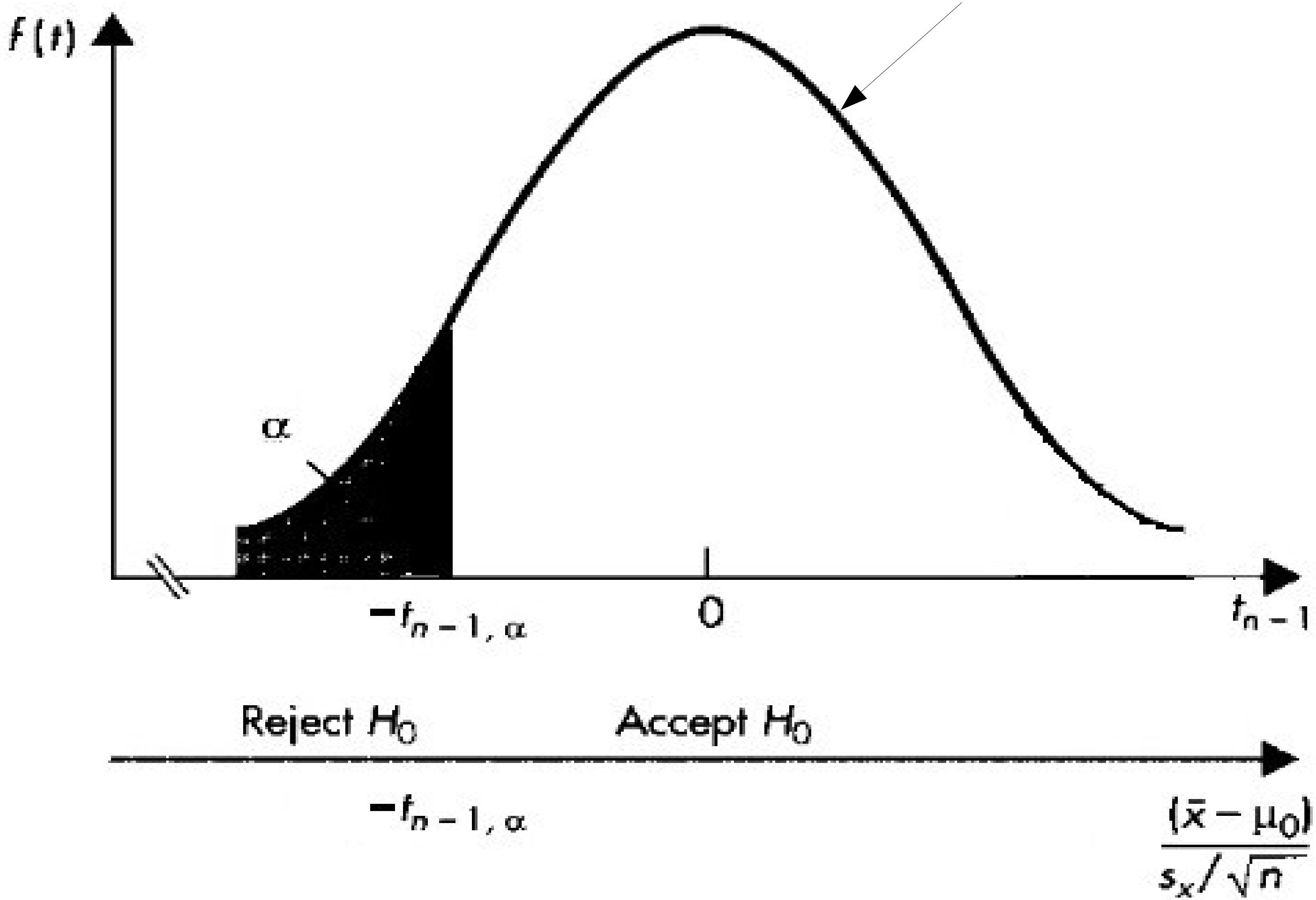
Критерий: основная гипотеза отвергается при $|t| > t_{n-1,\frac{\alpha}{2}}$.

Критическая область при **H$_A$**: $\mu > \mu_0$



Критерий: основная гипотеза отвергается при $t > t_{n-1,\alpha}$.

Критическая область при **H**$_A$: $\mu < \mu_0$



Критерий: основная гипотеза отвергается при $t < -t_{n-1,\alpha}$.

Пример (Devore, Berk). Один из источников загрязнения городских ливневых вод — выброшенные батарейки. В пригородах Кливленда была собрана 51 батарейка Panasonic AAA. Средняя масса цинка в батарейке оказалась равна 2.06 грамма, выборочное стандартное отклонение — 0.141 грамма. Дают ли сделанные наблюдения основание считать, что средняя масса цинка среди всех выбрасываемых батареек Panasonic AAA больше 2 грамм?

Решение.

H_0: $\mu = 2$

H_A: $\mu > 2$

Используем уровень значимости 5%.

Из условия: $n=51\,;\quad \bar{X}=2.06\,;\quad \hat{\sigma}=0.141\,.$

Рассчитываем критическую статистику: $t = \dfrac{\bar{X}-\mu_0}{\hat{\sigma}/\sqrt{n}} = \dfrac{2.06-2}{0.141/\sqrt{51}} = 3.04.$

Критическое значение: $t_{n-1,\ \alpha} = t_{50,\ 0.05} = 1.676.$

t-статистика попала в критическую область ($t>t_{50,\ 0.05}$), так что H_0 отвергается.

Вывод: есть основание считать, что средняя масса цинка среди выбрасываемых батареек Panasonic AAA превышает 2 грамма.

Вывод зависит от того, какой уровень значимости мы берём.

Чем меньше уровень значимости, тем менее мы склонны отвергать H_0 (т. е. тем более веское свидетельство против основной гипотезы мы требуем для её отвержения).

Чтобы понять, насколько наш вывод чувствителен к выбору уровня значимости, можно рассчитать *p-значение*.

p-значение *(probability value, p-value) — наименьший уровень значимости, при котором основная гипотеза отвергается.*

Как его рассчитать? Нужно найти такой уровень значимости, при котором *t*-статистика в точности совпадёт с критическим значением — это и будет *p*-значение.

Рассчитаем p-значение для батареек.

Критическая статистика: t=3.04.
При каком уровне значимости 3.04 будет критическим значением?

Возьмём с.в. $U \sim t_{50}$.

Критическое значение $t_{50,\alpha}$ находится из условия $P\left(U > t_{50,\alpha}\right) = \alpha$.

Следовательно, искомый уровень значимости равен $P\left(U > 3.04\right) = 0.0019$.

Итак, H_0 будет отвергаться при любом уровне значимости, не меньшем 0.0019.
Обычно это считается веским доводом против основной гипотезы.

В общем случае, если t — значение критической статистики, а $U \sim t_{n-1}$,
рассчитываем p-значение так:

$\mathbf{H_A}: \mu \neq \mu_0$	p-значение = $P(\lvert U\rvert > \lvert t\rvert)$
$\mathbf{H_A}: \mu > \mu_0$	p-значение = $P(U>t)$
$\mathbf{H_A}: \mu < \mu_0$	p-значение = $P(U<t)$

Картинка из (Devore, Berk)

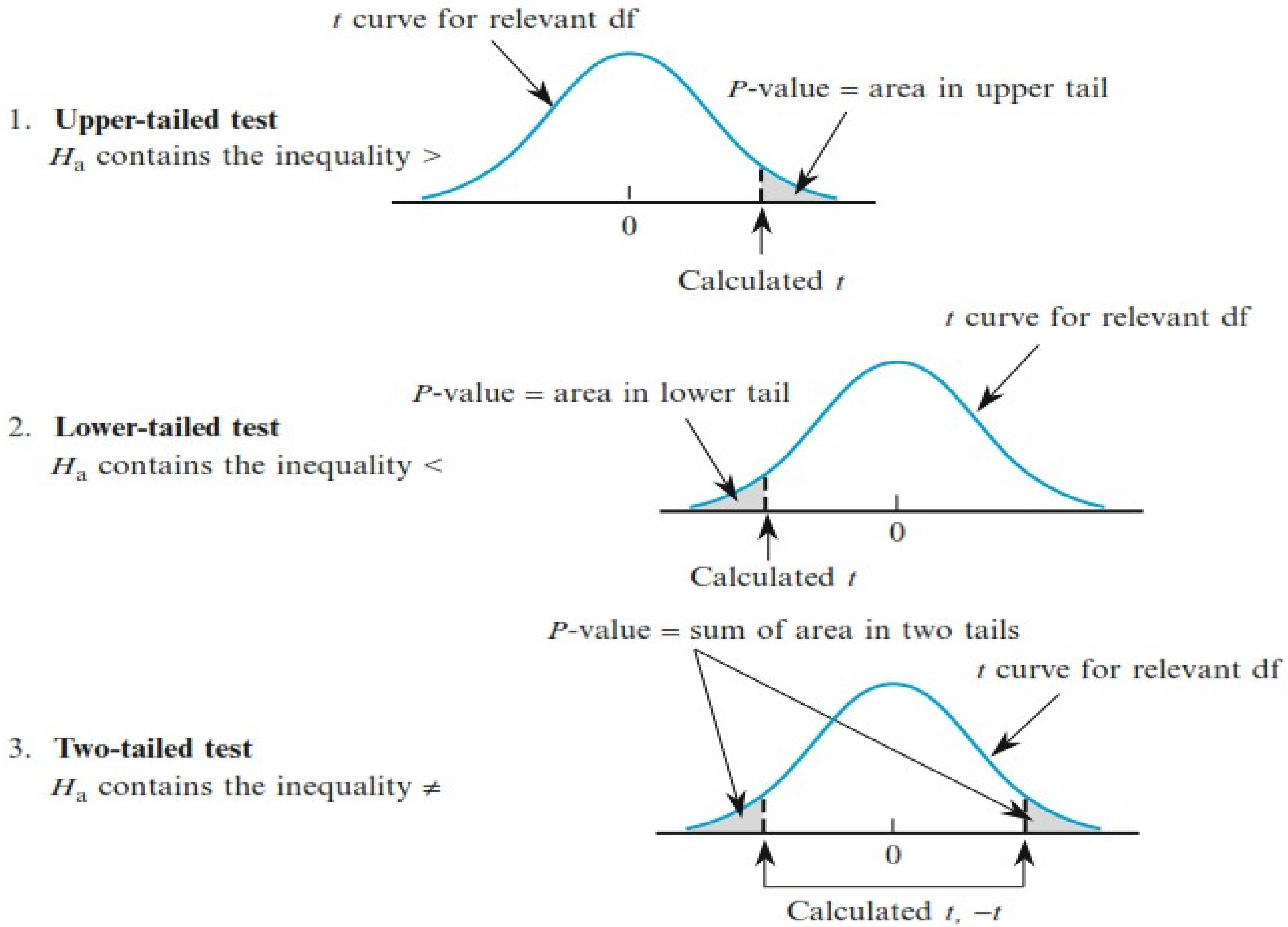


Figure 9.8 *P*-values for *t* tests

Важно помнить

Р-значение — это не вероятность того, что основная гипотеза верна, и не вероятность ошибки I рода.

Это и не вероятность ошибки II рода, но с ней реже путают.

Проверка гипотезы о дисперсии в нормальной генеральной совокупности

Пусть $X_1, \ldots, X_n$ независимы, $X_i \sim \mathrm{N}(\mu, \sigma^2)$.

$$\mathbf{H_0}: \sigma^2 = \sigma^2{}_0.$$

Сдвинутый нолик — не специально, просто не хотелось набивать это в редакторе формул.

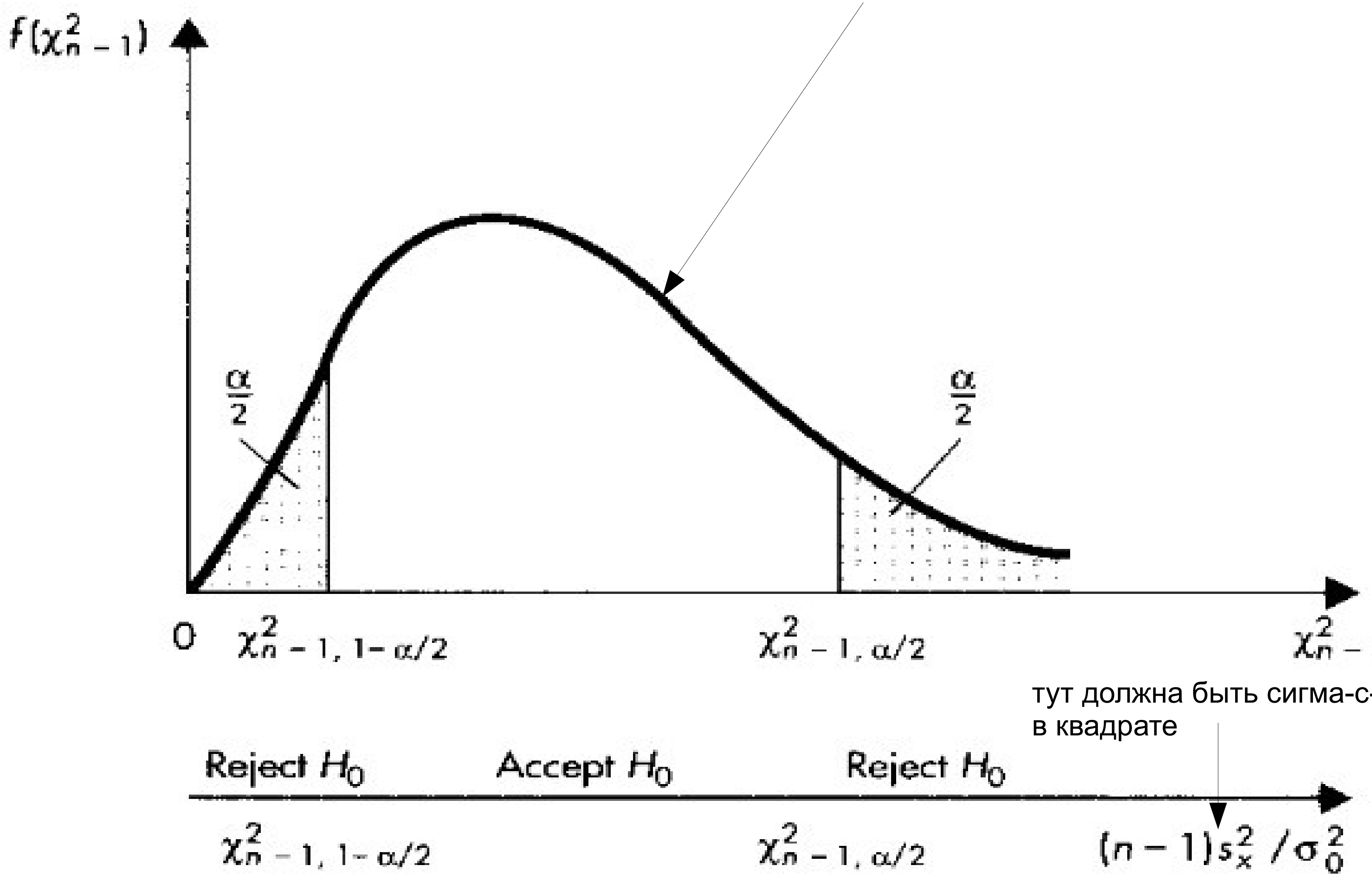
Альтернатива снова одна из трёх:

$$\mathbf{H_A}: \sigma^2 \neq \sigma^2{}_0$$
$$\mathbf{H_A}: \sigma^2 > \sigma^2{}_0$$
$$\mathbf{H_A}: \sigma^2 < \sigma^2{}_0$$

Критическая статистика: $\chi^2 = \dfrac{(n-1)\hat{\sigma}^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{n-1}.$

Критическая область при **H$_A$**: $\sigma^2 \neq \sigma_0^2$

ф-ция плотности χ^2-статистики
при верной основной гипотезе



тут должна быть сигма-с-крышкой
в квадрате

Критерий: основная гипотеза отвергается при $\chi^2 > \chi^2_{n-1,\frac{\alpha}{2}}$
или $\chi^2 < \chi^2_{n-1,1-\frac{\alpha}{2}}$.

Критические области для односторонних гипотез.

$$H_A: \sigma^2 > \sigma_0^2$$

основная гипотеза отвергается при $\chi^2 > \chi^2_{n-1,\alpha}$.

$$H_A: \sigma^2 < \sigma_0^2$$

основная гипотеза отвергается при $\chi^2 < \chi^2_{n-1,1-\alpha}$.

p-значение рассчитывается по тому же принципу: ищем уровень значимости, для которого статистика хи-квадрат совпадёт с критическим значением.

Пусть χ^2 — значение критической статистики, а $Y \sim \chi^2_{n-1}$. Тогда:

$$H_A: \sigma^2 \neq \sigma_0^2 \qquad p-значение = 2 \times min\{P(Y > \chi^2),\ P(Y < \chi^2)\}$$
$$H_A: \sigma^2 > \sigma_0^2 \qquad p-значение = P(Y > \chi^2)$$
$$H_A: \sigma^2 < \sigma_0^2 \qquad p-значение = P(Y < \chi^2)$$

Есть смысл обдумать это

Пример. В прошлом году Господа Состоятельные Кроты обследовали весь урожай свёклы в огороде Владислава Валерьевича и пришли к заключению, что дисперсия массы корнеплодов составила 600 г^2. В настоящем году крот Борислав измерил массу восьми случайно отобранных образцов свёклы:

157 201 194 197 204 152 199 143.

Есть ли основание считать, что дисперсия массы свёклы изменилась по отношению к предыдущему урожаю? Используйте уровень значимости 10%.

Решение.

H_0: $\sigma^2 = 600$

H_A: $\sigma^2 \neq 600$

Оцениваем дисперсию по выборке: $\hat{\sigma}^2 = 648.411$

Рассчитываем критическую статистику: $\chi^2 = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2} = \frac{(8-1)\times 648.411}{600} = 7.565.$

Критические значения: $\chi^2_{n-1,\ \frac{\alpha}{2}} = \chi^2_{7,\ 0.05} = 14.067;\ \ \chi^2_{n-1,\ 1-\frac{\alpha}{2}} = \chi^2_{7,\ 0.95} = 2.167.$

$2.167 < \chi^2 < 14.067$, так что нет оснований отвергнуть H_0 и считать, что дисперсия массы свёклы изменилась.

p-значение = 0.745 (по таблицам не найти, рассчитал в Stata).

Подумать и поболтать

- а правда ли мы верим, что дисперсия свёклы осталась той же, что в прошлом году?

- ловушки большой и малой выборки.

- статистическая и практическая значимость.

То, что мы не отвергаем основную гипотезу вовсе не означает, что она верна или что мы получили ей какое-либо подтверждение. Часто не отвержение основной гипотезы означает лишь то, что собранных данных недостаточно, чтобы эту гипотезу уверенно отвергнуть.

Проверка гипотезы о доле

$$X_1, \dots, X_n \;\; \textit{независимы}, \;\; X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

Из прошлой лекции: $\hat{p} \overset{asy}{\sim} N\left(p, \dfrac{p(1-p)}{n}\right);$

$$\frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \overset{asy}{\sim} N(0,1).$$

Проверяем основную гипотезу $\mathbf{H_0}$: $\mathrm{p} = \mathrm{p}_0$

против одной из трёх альтернатив: $\mathbf{H_A}$: $\mathrm{p} \neq \mathrm{p}_0$

$\mathbf{H_A}$: $\mathrm{p} > \mathrm{p}_0$

$\mathbf{H_A}$: $\mathrm{p} < \mathrm{p}_0$

Критическая статистика : $Z = \dfrac{\hat{p}-p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \overset{H_0,\, asy}{\sim} N(0,1).$

Дальше всё то же...

Критические области и p-значения

H_A: $p \neq p_0$

Отвергаем H_0, если $|Z|>z_{\alpha/2}$.

p-значение = $P(|\xi|>|Z|)$,

где $\xi \sim N(0, 1)$, Z — наблюдаемое значение критической статистики.

H_A: $p > p_0$

Отвергаем H_0, если $Z>z_{\alpha}$.

p-значение = $P(\xi>Z)$.

H_A: $p < p_0$

Отвергаем H_0, если $Z<-z_{\alpha}$.

p-значение = $P(\xi<Z)$.

Пример: заметка А. Зубанова «Кто выигрывает рэп-баттлы».

Изучая статью Зубанова, обратите внимание:

> Автор неуместно увязывает z-статистику с распределением Стьюдента. Так не надо делать. Там лучше будет смотреться нормальное, хотя 28 наблюдений — маловато.

> Оба раза автор неправильно рассчитал z-статистики. Правильные значения: 3.68 для порядка выступления, 1.58 для расположения. Поэтому в действительности выявляется связь только с порядком.

> Индекс 0.975 при критическом значении относится не к уровню значимости, а к порядку квантили (это не ошибка, просто другое обозначение).

А тут экскурс в историю: разбираем статью Джона Арбетнота.

An Argument for Divince Providence, taken from Constant Regularity observ'd in the Births of both Sexes

By Dr. John Arbuthnott, Physitian in ordinary to her Majesty, and Fellow of the College of Physitians and Royal Society. 1710

Следующая лекция:

Критерий согласия хи-квадрат.

Тестирование последовательностей случайных чисел.