Лекция 7

Проверка гипотезы о доле.

Критерий согласия хи-квадрат.

Напоминалка

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).

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

Критерий состоит из *статистики* и *критической области*.

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

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

Мощность — вероятность отвергнуть основную гипотезу, когда верна альтернативная.

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

Проверка гипотезы о доле

Пусть $X_1, \ldots, X_n$ независимы,

$$X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}.$$

Распределение выборочной доли:

$$\hat{p} \overset{\text{app}}{\sim} N\left(p, \frac{p(1-p)}{n}\right). \qquad \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{asy}}{\sim} N(0,1).$$

Проверяем основную гипотезу $\mathbf{H_0}$: $p = p_0$

против одной из трёх альтернатив: $\mathbf{H_A}$: $p \neq p_0$

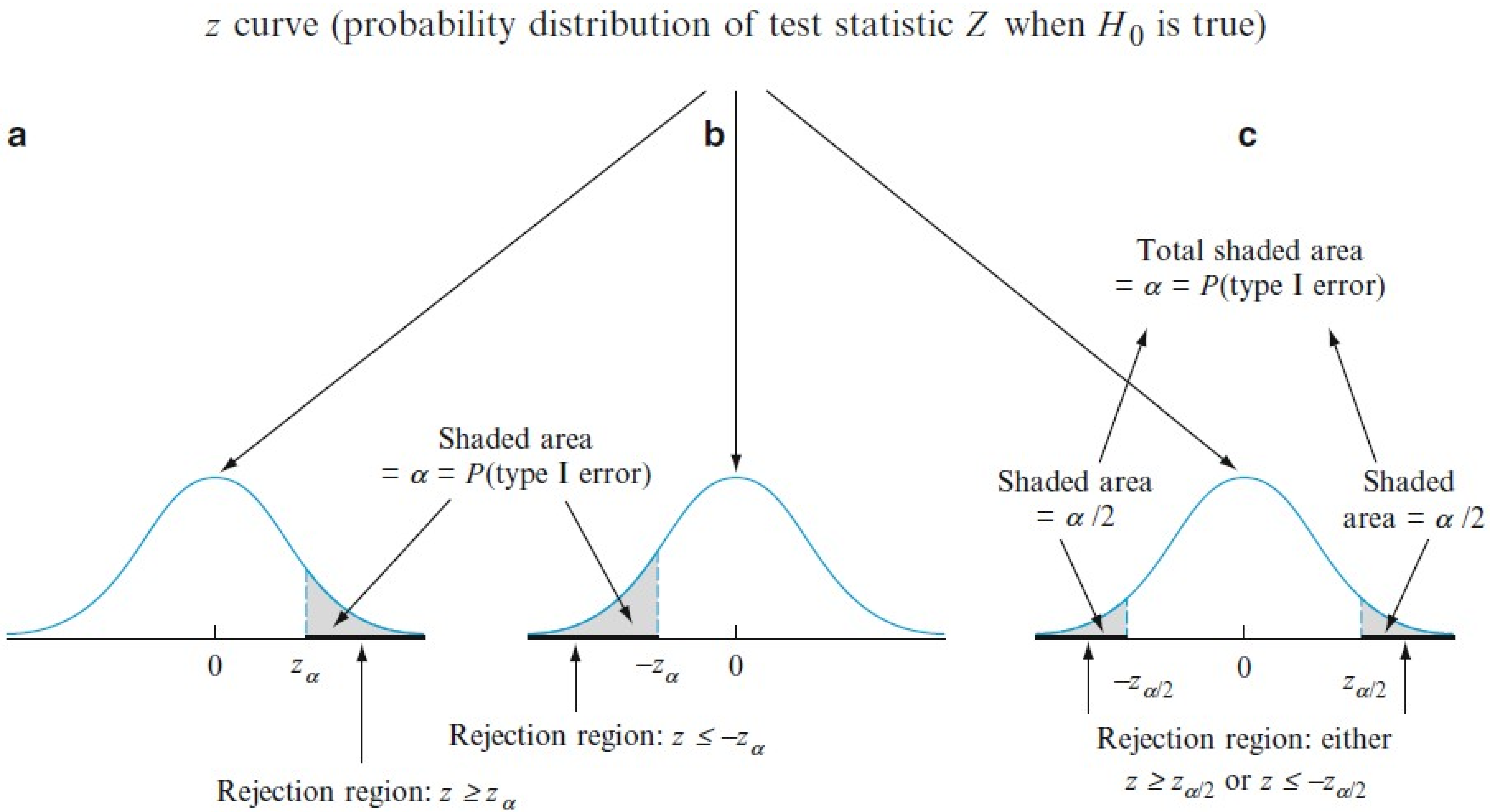
$\mathbf{H_A}$: $p > p_0$

$\mathbf{H_A}$: $p < p_0$

Критическая статистика :

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

Критическая область



H$_A$: $p > p_0$ **H**$_A$: $p < p_0$ **H**$_A$: $p \neq p_0$

Исторический пример №1

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз.
2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Будем использовать уровень значимости 5%.

Жорж-Луи Леклерк, граф де Бюффон.

Исторический пример №1

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз. 2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Будем использовать уровень значимости 5%.

Решение. Для начала сформулируем гипотезы.

$$H_0: p = 0.5$$
$$H_A: p \neq 0.5$$

Жорж-Луи Леклерк, граф де Бюффон.

Исторический пример №1

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз. 2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Жорж-Луи Леклерк, граф де Бюффон.

Будем использовать уровень значимости 5%.

Решение. Для начала сформулируем гипотезы.

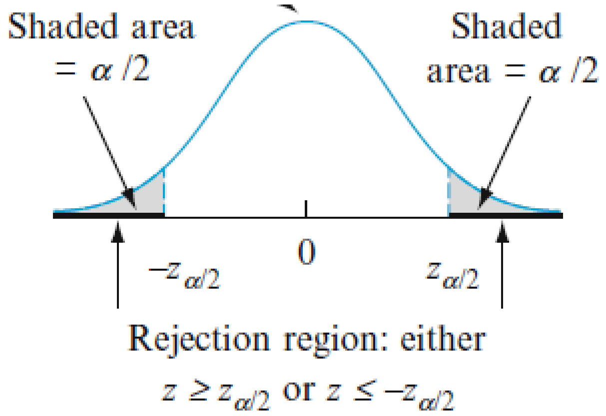
$$H_0: p = 0.5$$
$$H_A: p \neq 0.5$$

Двусторонняя альтернатива => критерий такой:

отвергнуть H_0, если $|Z| > z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}}$, где

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5\times(1-0.5)}{4040}}}.$$

Критическое значение: $z_{\frac{0.05}{2}} = 1.96$.



Исторический пример №1

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз. 2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Жорж-Луи Леклерк, граф де Бюффон.

Будем использовать уровень значимости 5%.

Решение. Для начала сформулируем гипотезы.

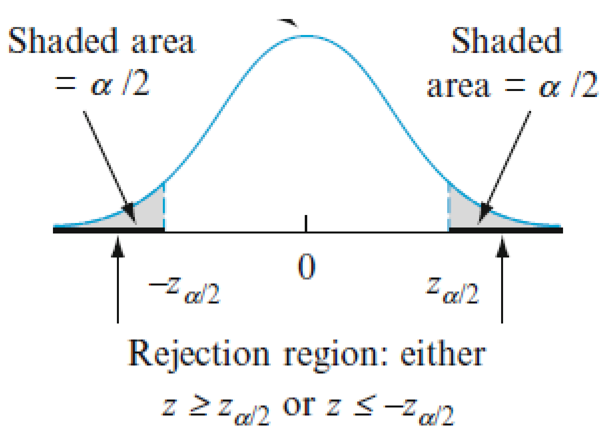
$$H_0: p = 0.5$$
$$H_A: p \neq 0.5$$

Двусторонняя альтернатива => критерий такой:

отвергнуть H_0, если $|Z| > z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = 1.96$.

Рассчитываем статистику:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{2048}{4040} - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{4040}}} \approx 0.88.$$



Исторический пример №1

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз. 2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Жорж-Луи Леклерк, граф де Бюффон.

Будем использовать уровень значимости 5%.

Решение. Для начала сформулируем гипотезы.

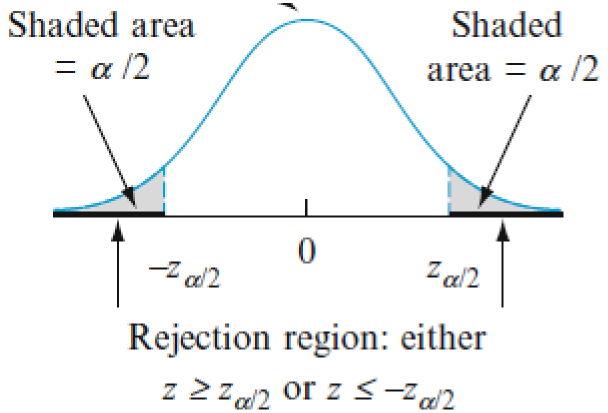
$$H_0: p = 0.5$$
$$H_A: p \neq 0.5$$

Двусторонняя альтернатива => критерий такой:

отвергнуть H_0, если $|Z| > z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = 1.96$.

Рассчитываем статистику:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{2048}{4040} - 0.5}{\sqrt{\frac{0.5 \times (1-0.5)}{4040}}} \approx 0.88.$$



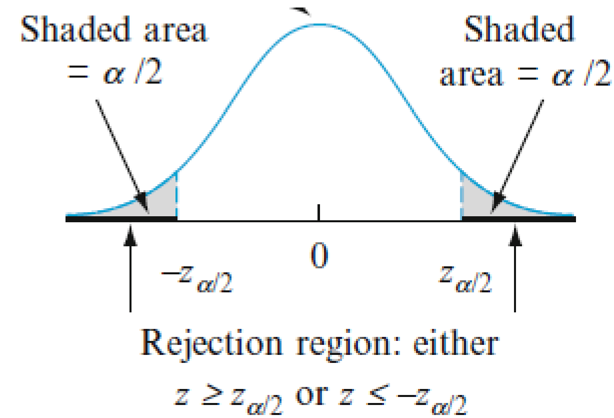
$|Z| < 1.96$ => не отвергаем H_0

Вывод: нет оснований отвергнуть H_0 и считать, что монетка неправильная.

Исторический пример №1: теперь с *p*-значением

В XVIII веке естествоиспытатель Жорж-Луи Бюффон подбросил монетку 4040 раз. 2048 раз она выпала «орлом» вверх, а 1992 раза - «решкой». Есть ли основание считать, что монетка неправильная?

Будем использовать уровень значимости 5%.



Решение. Гипотезы:

$$H_0 : p = 0.5$$
$$H_A : p \neq 0.5$$

Статистика: $Z \approx 0.88$.

Рассчитаем *p*-значение — наименьший уровень значимости, при котором H_0 отвергается.

Найдём уровень значимости, при котором критическое значение равно 0.88.
Это и будет p-value.

$$\underbrace{P\left(|Z| > 0.88\right) = 0.3789}_{p\text{-значение}} \quad \Rightarrow \quad z_{\frac{0.3789}{2}} = 0.88.$$

Основная гипотеза отвергается на уровнях значимости от 37.89%.
Значит, на уровне 5% она не отвергается — нет оснований считать, что монетка неправильная.

Правило: p-value $< \alpha$ => H_0 отвергается в пользу H_A.
p-value $> \alpha$ => нет оснований отвергнуть H_0.

Расчёт *p*-значений

картинка из (Devore, Berk)

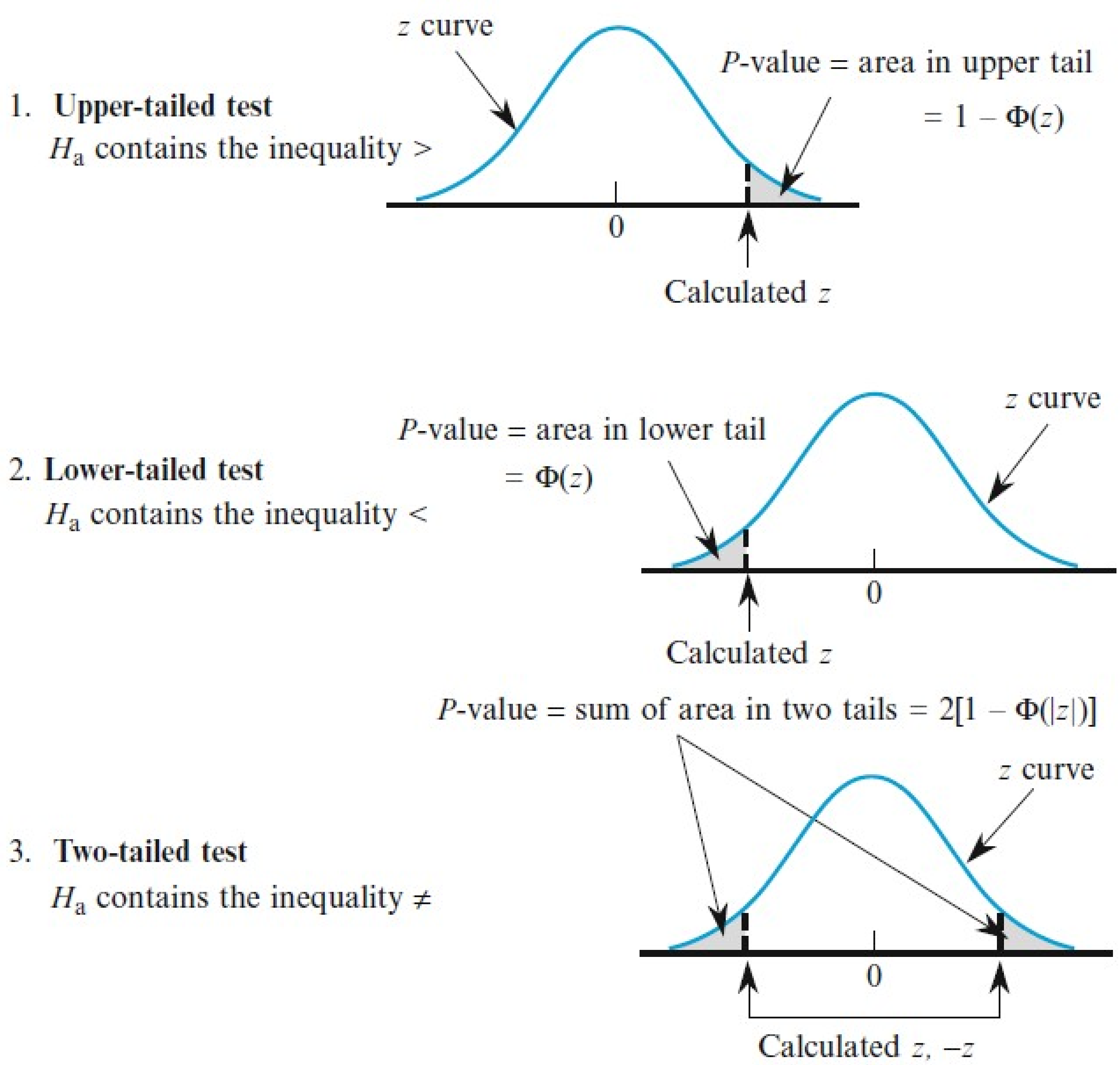


Figure 9.7 Determination of the *P*-value for a *z* test

Исторический пример №2

Кто выгрывает рэп-баттлы?

Обращаемся к заметке Андрея Зубанова.

Изучая статью Зубанова, обратите внимание:

► Автор использует немножно другую статистику — в знаменателе фигурирует не предполагаемая вероятность p_0, а выборочная доля $\hat{p}$. Так тоже можно.

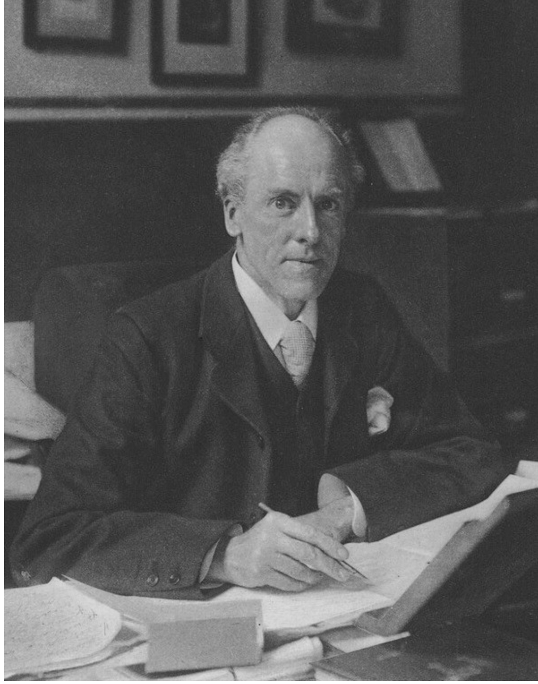
► Андрей неуместно увязывает *z*-статистику с распределением Стьюдента. Так не надо делать. Там лучше будет смотреться нормальное, хотя 28 наблюдений — маловато.

► 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).

Критерий согласия хи-квадрат

(критерий согласия Пирсона)



Карл Пирсон

Пусть $X_1, \ldots, X_n$ независимы,

$$X_i \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_k \\ p_1 & p_2 & \ldots & p_k \end{pmatrix}.$$

Основная гипотеза в точности задаёт *всё* распределение признака X:

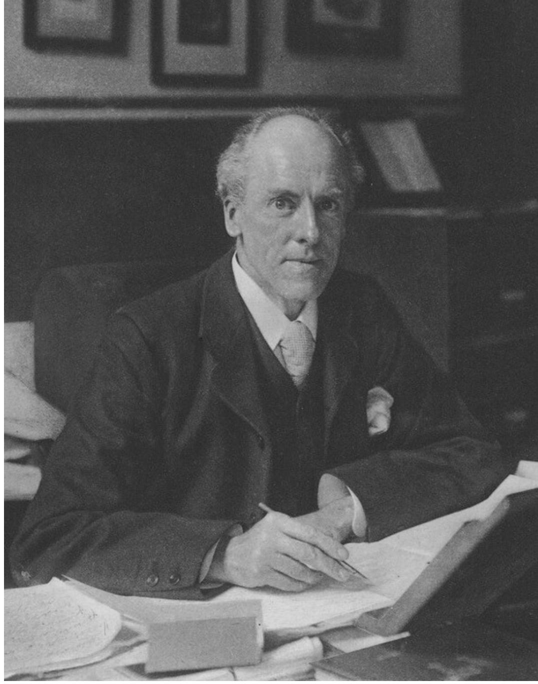
$$H_0 : p_1 = p_1^0, \quad p_2 = p_2^0, \ldots, \quad p_k = p_k^0.$$

Альтернативная — любое отклонение от H_0:

$$H_A : (\exists\, j : p_j \neq p_j^0).$$

Пример *(Демешев)***.** Вася Сидоров утверждает, что в кино он ходит в два раза чаще, чем в спортзал, а в спортзал в два раза чаще, чем в театр. За последние полгода он 10 раз был в театре, 17 раз — в спортзале и 39 раз — в кино. Правдоподобно ли Васино утверждение?

Критерий согласия хи-квадрат
(критерий согласия Пирсона)



Карл Пирсон

Пусть $X_1,\ldots,X_n$ независимы,

$$X_i \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_k \\ p_1 & p_2 & \ldots & p_k \end{pmatrix}.$$

Основная гипотеза в точности задаёт *всё* распределение признака X:

$$H_0: p_1 = p_1^0, \quad p_2 = p_2^0, \ldots, \quad p_k = p_k^0.$$

Альтернативная — любое отклонение от H_0:

$$H_A: (\exists\, j : p_j \neq p_j^0).$$

Пример *(Демешев)***.** Вася Сидоров утверждает, что в кино он ходит в два раза чаще, чем в спортзал, а в спортзал в два раза чаще, чем в театр. За последние полгода он 10 раз был в театре, 17 раз — в спортзале и 39 раз — в кино. Правдоподобно ли Васино утверждение?

Здесь

$$H_0: \ X_i \sim \begin{pmatrix} \text{кино} & \text{спортзал} & \text{театр} \\ 4/7 & 2/7 & 1/7 \end{pmatrix}$$

$$H_A: \ \text{любое другое распределение}$$

Критерий согласия хи-квадрат
(критерий согласия Пирсона)

Пример *(Демешев)*. Вася Сидоров утверждает, что в кино он ходит в два раза чаще, чем в спортзал, а в спортзал в два раза чаще, чем в театр. За последние полгода он 10 раз был в театре, 17 раз — в спортзале и 39 раз — в кино. Правдоподобно ли Васино утверждение?

$$H_0:\ X_i \sim \begin{pmatrix} \text{кино} & \text{спортзал} & \text{театр} \\ 4/7 & 2/7 & 1/7 \end{pmatrix}$$

Всего $n = 10 + 17 + 39 = 66$ наблюдений

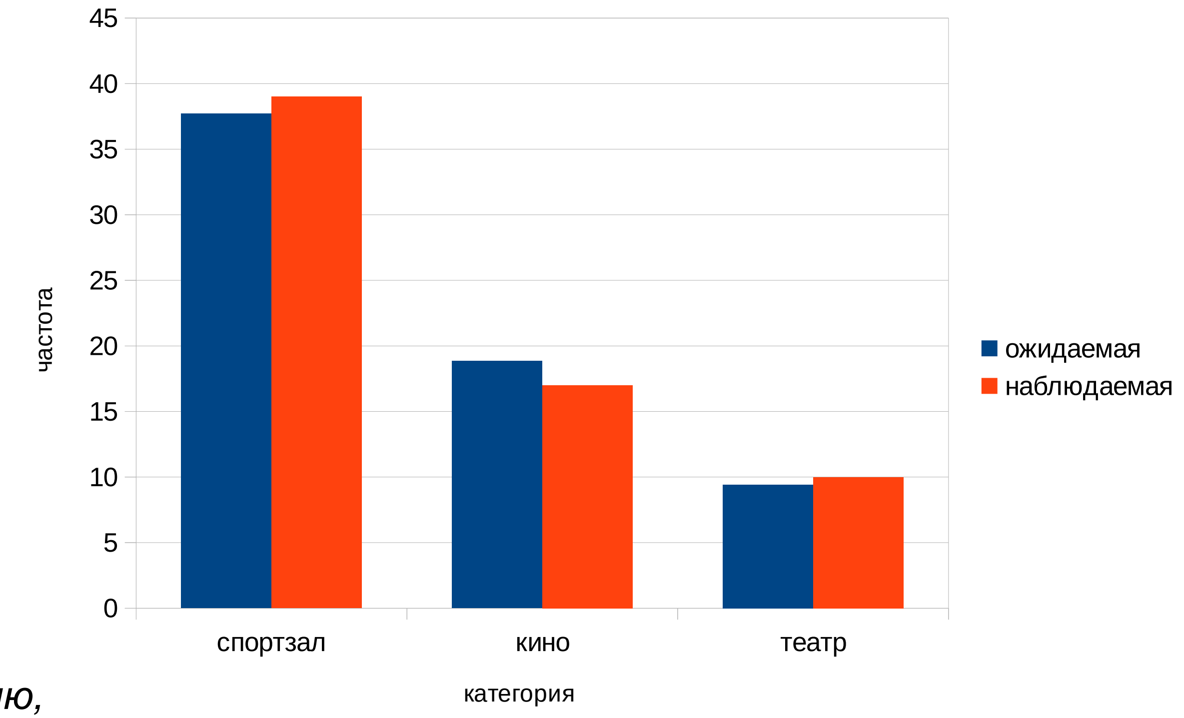
$$H_A:\ \text{любое другое распределение}$$

Попробуем наглядно представить разницу между основной гипотезой и данными.

категория, j	кино	спортзал	театр
предполагаемая вероятность, p_j^0	4/7	2/7	1/7
ожидаемая частота, $E_j = np_j^0$	37.71	18.86	9.43
наблюдаемая частота, O_j	39	17	10

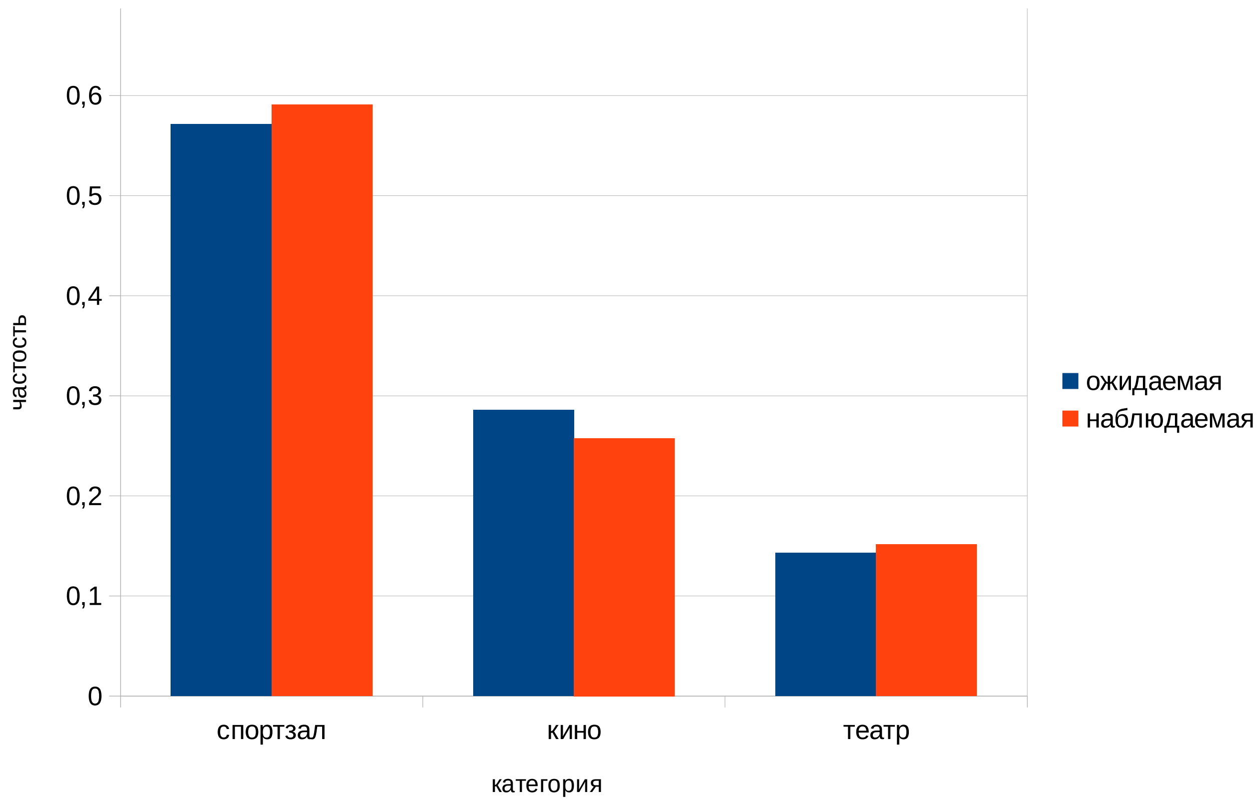
Это можно даже представить на графике.
Впрочем, и так видно, что существенных расхождений нет.

График частот:



Частота — число попадений в категорию, частость — доля попаданий.

График частостей (долей):



Критерий согласия хи-квадрат
(критерий согласия Пирсона)

Пусть $X_1,\ldots, X_n$ независимы,

$$X_i \sim \begin{pmatrix} x_1 & x_2 & \ldots & x_k \\ p_1 & p_2 & \ldots & p_k \end{pmatrix}.$$

Основная гипотеза в точности задаёт *всё* распределение признака X:

$$H_0 : p_1 = p_1^0, \quad p_2 = p_2^0, \ldots, \quad p_k = p_k^0.$$

Альтернативная — любое отклонение от H_0:

$$H_A : \left(\exists\, j : p_j \neq p_j^0\right).$$

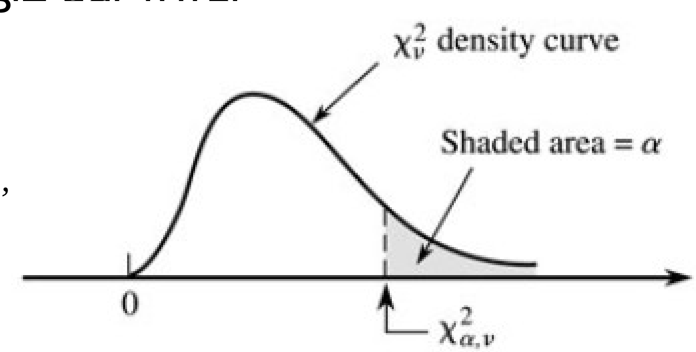
Статистика:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j} \overset{H_0}{\sim} \chi^2_{k-1}.$$

Здесь O_j - наблюдаемые частоты, $E_j = np_j^0$ - ожидаемые частоты

Критическое правило: отвергнуть H_0, если $\chi^2 > \chi^2_{k-1,}$

где α — уровень значимости.



Критерий согласия хи-квадрат

(критерий согласия Пирсона)

Статистика:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j} \overset{H_0}{\sim} \chi^2_{k-1}.$$

Критическое правило: отвергнуть H_0, если $\chi^2 > \chi^2_{k-1,\alpha}$,
где α — уровень значимости.

Проверим Васино утверждение на уровне 10%.

категория, j	кино	спортзал	театр
ожидаемая частота, $E_j = np_j^0$	37.71	18.86	9.43
наблюдаемая частота, O_j	39	17	10

Статистика: $\chi^2 = \frac{(39-37.71)^2}{37.71} + \frac{(17-18.86)^2}{18.86} + \frac{(10-9.43)^2}{9.43} = 0.26.$

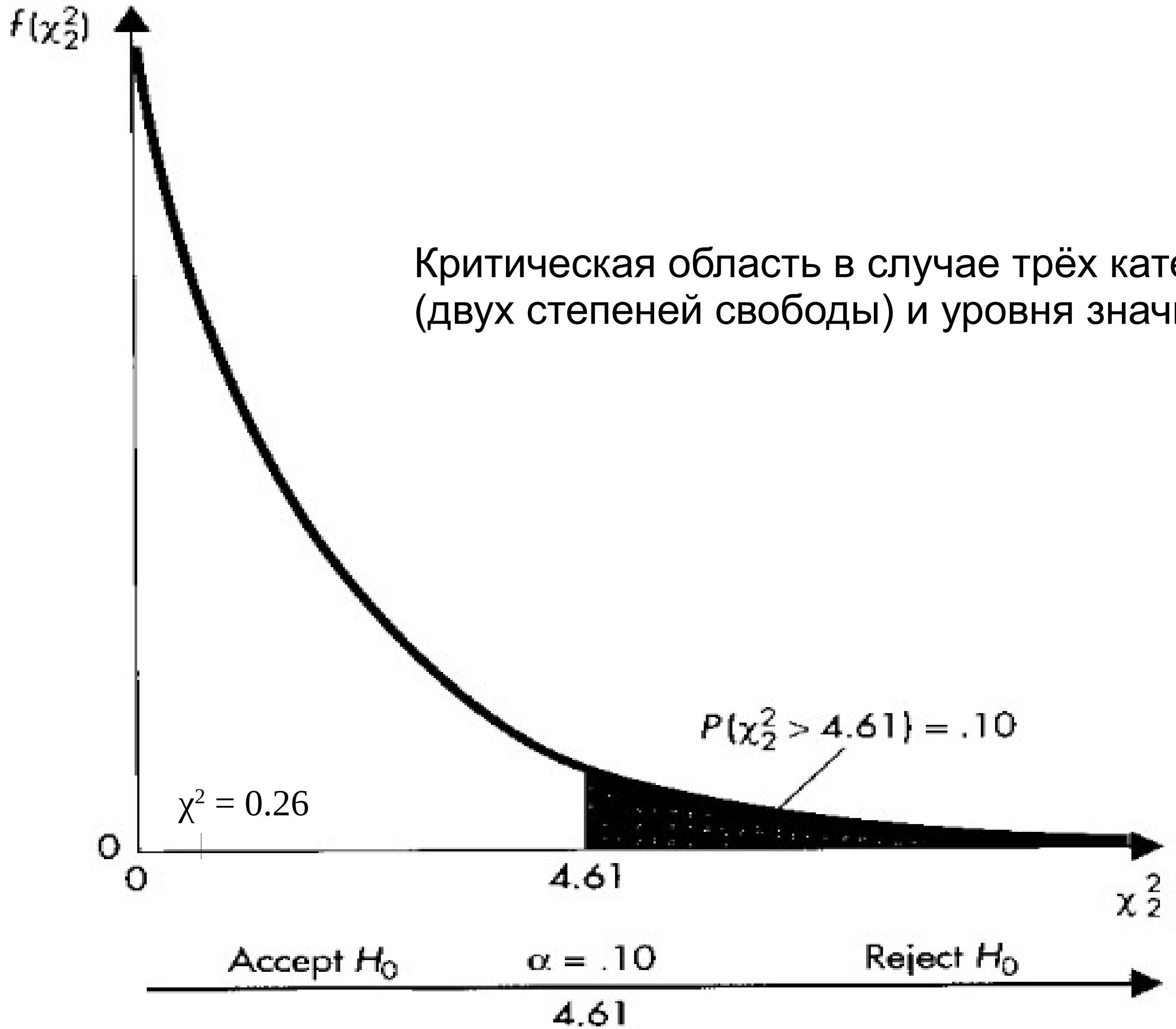
Критическое значение: $\chi^2_{2,0.1} = 4.61.$

0.26<4.61 => нет оснований отвергнуть H_0 и уличить Васю во лжи.

Критерий:

при $\chi^2 > \chi^2_{k-1,\alpha}$ H_0 отвергается,

при $\chi^2 \le \chi^2_{k-1,\alpha}$ нет оснований отвергнуть H_0.



Критическая область в случае трёх категорий (двух степеней свободы) и уровня значимости 10%

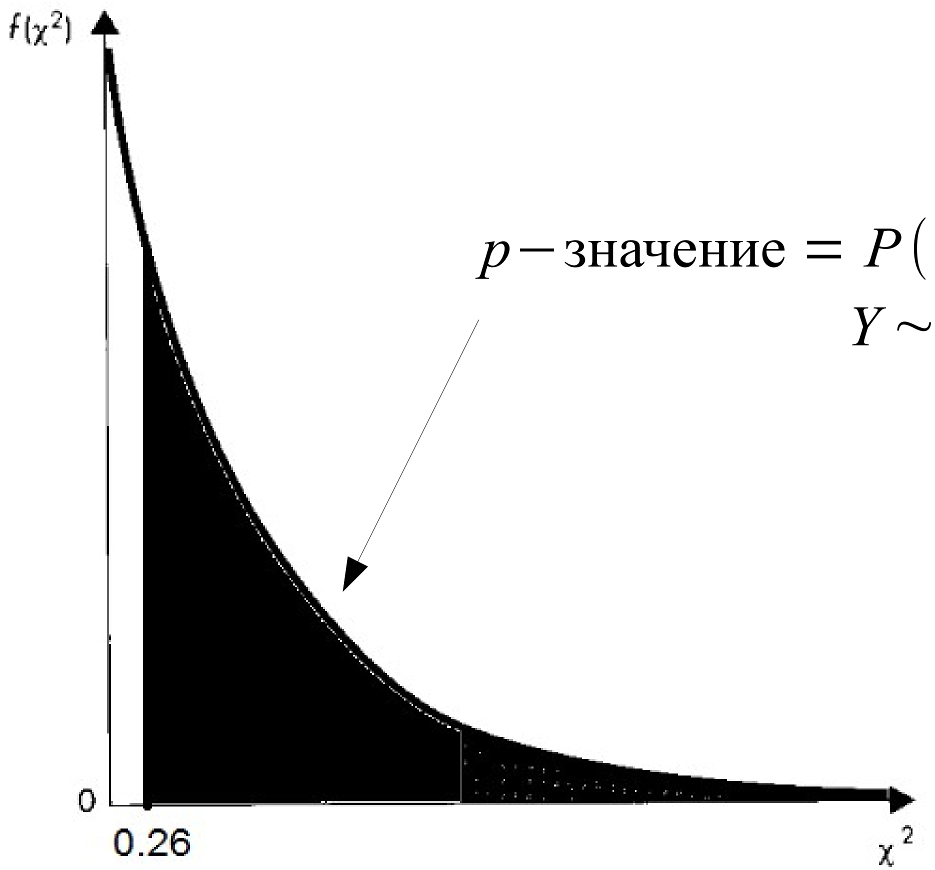
А теперь — с *p*-значением!

Статистика: $\chi^2 = 0.26.$

Рассчитаем *p*-значение.

Ищем такой уровень значимости, для которого критическое значение равно 0.26.

$$p - \text{значение} = P\left(Y > 0.26\right) = 0.878\,,$$
$$Y \sim \chi_2^2.$$



p-значение > 0.1 => не отвергаем основную гипотезу.

Критерий согласия хи-квадрат

(критерий согласия Пирсона)

На заметку:

► в рассмотренном варианте критерий хи-квадрат пригоден только для проверки простой основной гипотезы (нет неизвестных параметров распределения в H_0);

► критерий можно применять для любых типов признаков;

► он плохо работает, когда есть близкие к нулю ожидаемые частоты (и вообще работает неправильно, если есть $p_j^0 = 0$);

► при большом числе возможных значений эти значения приходится группировать (чтобы частоты не были близки к нулю), результат будет зависеть от группировки;

► для количественных признаков часто лучше применять другие критерии, чтобы избежать группировки;

► не ленитесь сверять предполагаемое распределение с наблюдаемым на графике.

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:

Анализ связанных пар