

# **Выявление различий в распределении признака**

Лекция № 9

# План

1. Обоснование задачи сравнения распределений признака
2. критерий Пирсона
  - 2.1. Назначения критерия
  - 2.2. Описание критерия
  - 2.3. Гипотезы
  - 2.4. Графическое представление критерия
  - 2.5. Ограничения критерия
  - 2.6. Алгоритм расчета критерия

# Литература

1. Загвязинский, В.И. Методология и методы психолого-педагогического исследования : учеб. пособие для студентов пед. вузов по спец.031000 – Педагогика и психология / В.И. Загвязинский. – М. : Академия, 2001. – 202 с.
2. Сидоренко, Е.В. Методы математической обработки в психологии / Е.В. Сидоренко. – СПб.: Речь, 2010. – С. .

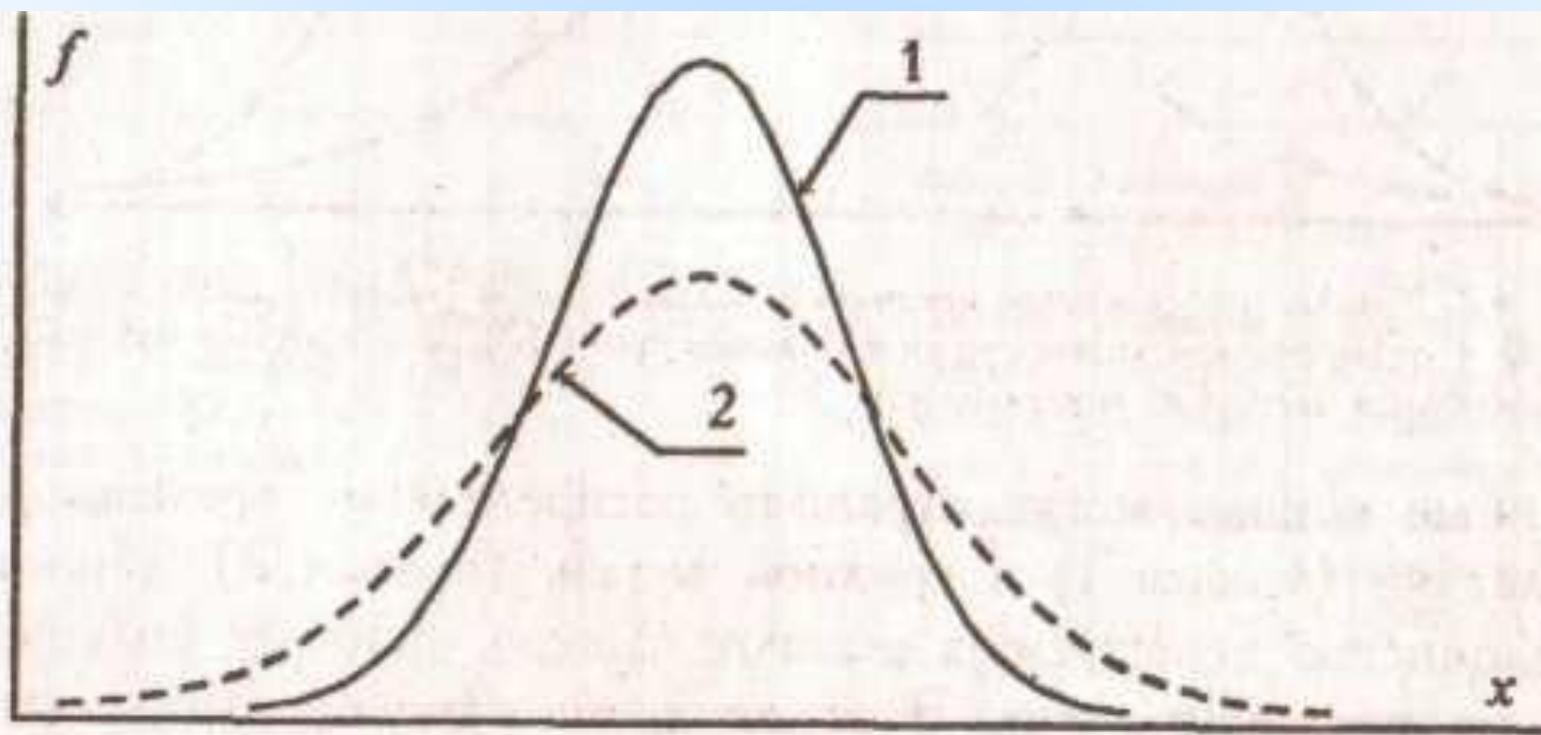


Рис. 4.1. Кривые распределения признака с меньшим диапазоном вариативности признака (1) и большим диапазоном распределения признака (2);  $x$  - значения признака;

# Два распределения, различающиеся по знаку асимметрии

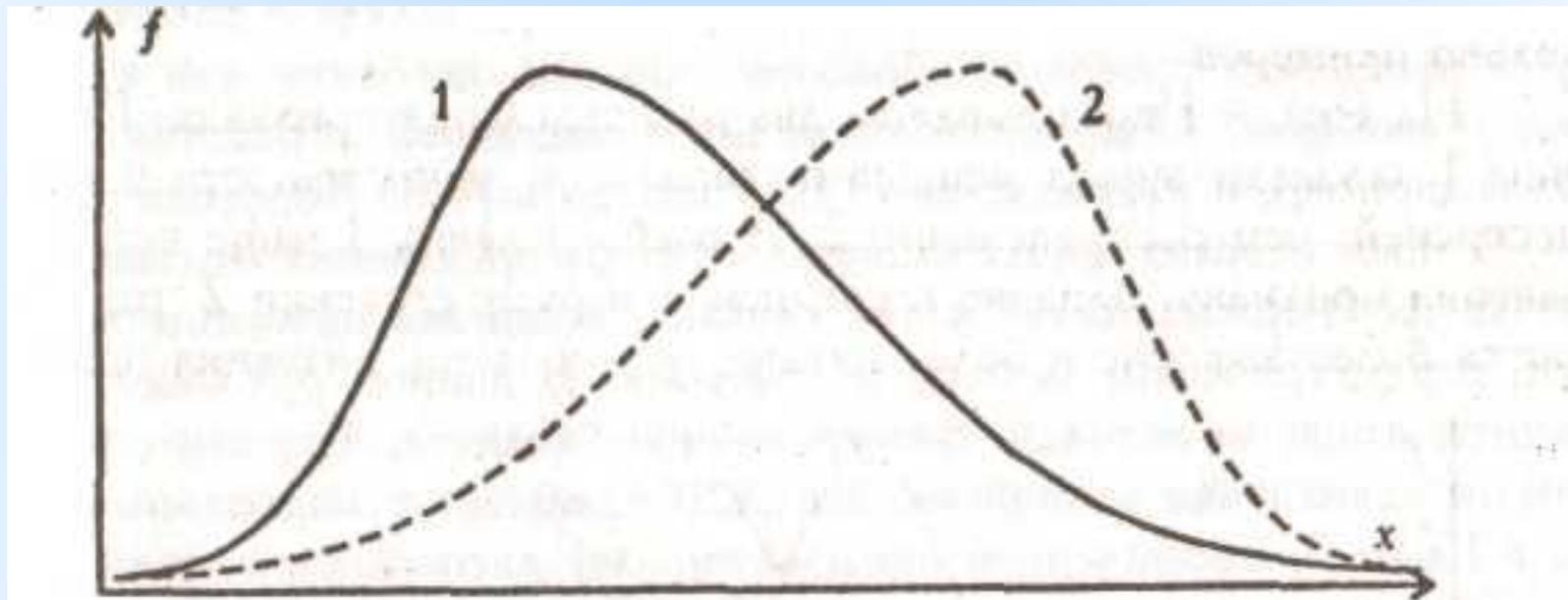


Рис. 4.2. Кривые распределения признака с положительной (левосторонней) асимметрией (1) и отрицательной (правосторонней) асимметрией (2);  $x$  - значения признака;  $f$  - относительная частота их встречаемости

Бывает полезно также сопоставить полученное эмпирическое распределение с теоретическим распределением. Традиционные для отечественной математической статистики критерии определения расхождения или согласия распределений - это метод  $\chi^2$  К. Пирсона и критерий  $\lambda$  Колмогорова-Смирнова.

# Методы не заменимы в следующих случаях

- в задачах, требующих доказательства неслучайности предпочтений при выборе из нескольких альтернатив;
- в задачах, требующих обнаружения точки максимального расхождения между двумя распределениями, которая затем используется для перегруппировки данных с целью применения критерия \* (углового преобразования Фишера).

# Критерий Пирсона

## Назначения критерия

Критерий  $\chi^2$  применяется в двух целях:

- 1) для сопоставления *эмпирического* распределения признака с *теоретическим* - равномерным, нормальным или каким-то иным;
- 2) для сопоставления *двух, трех или более эмпирических* распределений одного и того же признака.

# Описание критерия

Критерий  $\chi^2$  отвечает на вопрос о том, с одинаковой ли частотой встречаются разные значения признака в эмпирическом и теоретическом распределениях или в двух и более эмпирических распределениях.

Преимущество метода состоит в том, что он позволяет сопоставить распределения признаков, представленных в любой шкале, начиная со шкалы наименований. В самом простом случае альтернативного распределения "да - нет", "допустил брак - не допустил брака", "решил задачу - не решил задачу" и т. п. мы уже можем применить критерий  $\chi^2$ .

Допустим, некий наблюдатель фиксирует количество пешеходов, выбравших правую или левую из двух симметричных дорожек на пути на точки А в точку Б (см. Рис. 4.3).

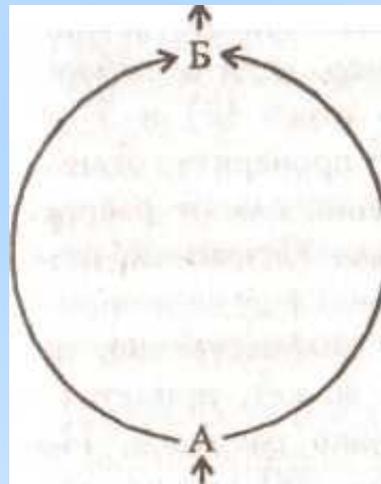


Рис. 4.3. Иллюстрация к примеру о теоретически равновероятном выборе из двух альтернатив

Допустим, в результате 70 наблюдений установлено, что 51 человек выбрали правую дорожку, и лишь 19 - левую. С помощью критерия<sup>2</sup> мы можем определить, отличается ли данное распределение выборов от равномерного распределения, при котором обе дорожки выбирал бы с одинаковой частотой. Это вариант сопоставления полученного *эмпирического* распределения с *теоретическим*.

Но представим себе, что наблюдатель решает совершенно другую задачу. Совпадение полученного распределения с равномерным его интересует гораздо меньшей степени, чем совпадение или несовпадение его данных с данными других исследователей. Ему известно, что люди с преобладанием правой ноги, склонны делать круг против часовой стрелки, а люди с преобладанием левой ноги - круг по ходу часовой стрелки.

С помощью метода он может сопоставить два эмпирического распределения: соотношение 51:19 в собственной выборке и соотношение 74:26 в выборке других исследователей.

Это вариант *сопоставления двух эмпирических распределений по простейшему альтернативному признаку* (конечно, простейшему математической точки зрения, а отнюдь не психологической).

Аналогичным образом мы можем сопоставлять распределение выборов из трех и более альтернатив. Например, если в выборке из человек 30 выбрали ответ (а), 15 человек - ответ (б) и 5 человек ответ (в), то мы можем с помощью метода проверить, отличается это распределение от равномерного распределения или от распределения ответов в другой выборке, где ответ (а) выбрали 10 человек, ответ 25 человек, ответ (в) - 15 человек.

В тех случаях, если признак измеряется количественно, скажем, в баллах, секундах или миллиметрах, нам, быть может, придется объединить все обилие значений признака в несколько разрядов. Например, если время решения задачи варьирует от 10 до 300 секунд, то мы можем ввести 10 или 5 разрядов, в зависимости от объема выборки. Например, это будут разряды: 0-50 секунд; 51-100 секунд; 101-150 секунд и т. д. Затем мы с помощью метода будем сопоставлять частоты встречаемости разных разрядов признака, но в остальном принципиальная схема не меняется.

*При сопоставлении эмпирического распределения с теоретическим* определяем степень расхождения между эмпирическими и теоретическими частотами.

*При сопоставлении двух эмпирических распределений* определяем степень расхождения между эмпирическими частотами и теоретическими частотами, которые наблюдались бы в случае совпадения двух этих эмпирических распределений.

*Чем больше расхождение между двумя сопоставляемыми распределениями, тем больше эмпирическое значение  $\chi^2$*

# *Гипотезы*

## *Первый вариант*

$H_0$  : Полученное эмпирическое распределение признака не отличается от теоретического (например, равномерного) распределения.

$H_1$  : Полученное эмпирическое распределение признака отличается от теоретического распределения.

# *Гипотезы*

## *Второй вариант*

$H_0$  : Эмпирическое распределение 1 не отличается от эмпирического распределения 2.

$H_1$  : Эмпирическое распределение 1 отличается от эмпирического распределения 2.

# *Гипотезы*

## *Третий вариант*

$H_0$ : Эмпирические распределения 1, 2, 3, ... не различаются между собой.

$H_1$ : Эмпирические распределения 1, 2, 3, ... различаются между собой.

# Графическое представление критерия (на примере с выбором правой или левой дорожек на пути из точки А в точку Б )

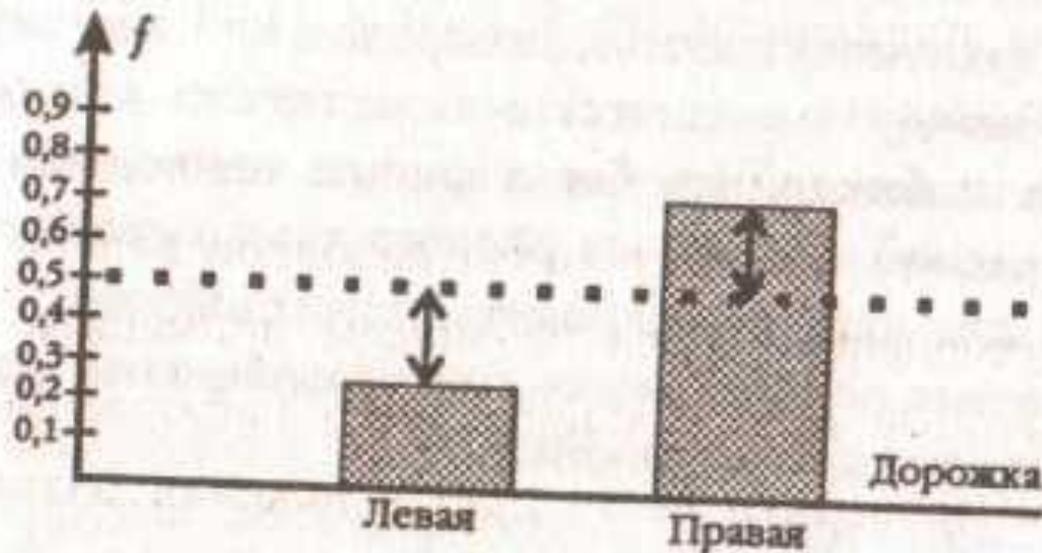


Рис. 4.4. Частоты выбора левой и правой дорожек; теоретическая частота представлена в виде горизонтальной планки, стрелками обозначены области расхождения между эмпирическими и теоретическими частотами

На Рис. 4.4 частота выбора левой дорожки представлена левым столбиком, а частота выбора правой дорожки - правым столбиком гистограммы.

На оси ординат отмеряются относительные частоты выбора, то есть частоты выбора той или иной дорожки, отнесенные к общему количеству наблюдений. Для левой дорожки относительная частота, которая называется также частота составляет  $19/70$ , то есть  $0,27$ , а для правой дорожки  $51/70$ , то  $0,73$ .

Если бы обе дорожки выбирались равновероятно, то половина испытуемых выбрала бы правую дорожку, а половина - левую. Вероятность выбора каждой из дорожек составляла бы  $0,50$ .

Отклонения эмпирических частот от этой величины довольно значительны. Возможно, различия между эмпирическим теоретическим распределением окажутся достоверными.

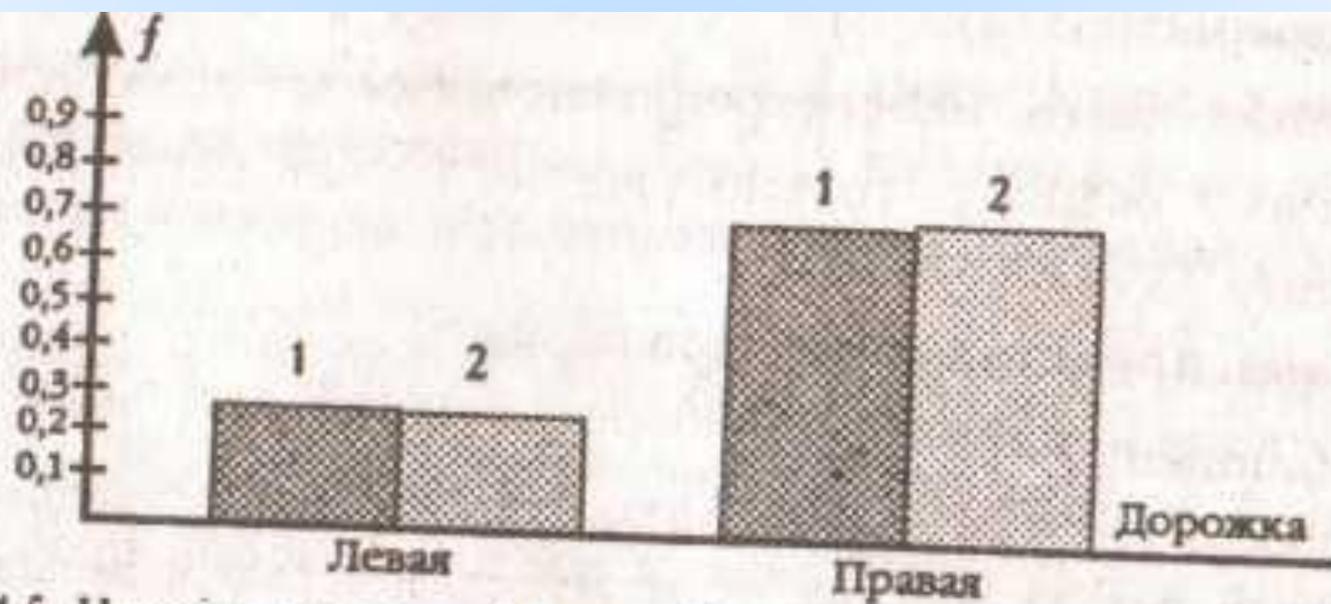


Рис. 4.5. Частоты выбора левой и правой дорожек в двух выборках испытуемых  
 1 – Выборка наблюдателя;  
 2 – Выборка других исследователей

На Рис. 4.5 фактически представлены две гистограммы, но столбики сгруппированы так, что слева сопоставляются частоты предпочтения левой дорожки в выборе наблюдателя (1) и в выборке Т.А. Доброхотовой и Н.Н. Брагиной (2), а справа - частоты предпочтения правой дорожки в этих же двух выборках.

Видно, что расхождения между выборками очень незначительны. Критерий  $\chi^2$ , скорее всего, подтвердит совпадение двух распределений.

## Ограничения критерия

1. Объем выборки должен быть достаточно большим:  
 $n \geq 30$  При  $n < 30$  критерий дает весьма приближенные значения. Точность критерия повышается при больших  $n$ .
2. Теоретическая частота для каждой ячейки таблицы не должна быть меньше 5:  $f \geq 5$ . Это означает, что если число разрядов задано заранее и не может быть изменено, то мы не можем применять метод  $\chi^2$ , не накопив определенного минимального числа наблюдений.

3. Выбранные разряды должны "вычерпывать" все распределение, то есть охватывать весь диапазон вариативности признаков. При этом группировка на разряды должна быть одинаковой во всех сопоставляемых распределениях.
4. Разряды должны быть неперекрещивающимися: если наблюдение отнесено к одному разряду, то оно уже не может быть отнесено ни к какому другому разряду. Сумма наблюдений по разрядам всегда должна быть равна общему количеству наблюдений.

Что считать числом наблюдений - количество выборов, реакций, действий или количество испытуемых, которые совершают выбор, проявляют реакции или производят действия?

Если испытуемый проявляет несколько реакций, и все они регистрируются, то количество испытуемых не будет совпадать с количеством реакций. Мы можем просуммировать реакции каждого испытуемого и сравнивать распределения индивидуальных сумм реакций в нескольких выборках.

В этом случае числом наблюдений будет количество испытуемых.

Если же подсчитываем частоту реакций определенного типа в целом по выборке, то получаем распределение реакций разного типа, и в этом случае количеством наблюдений будет общее количество зарегистрированных реакций, а не количество испытуемых.

С математической точки зрения правило независимости рядов соблюдается обоим случаям: одно наблюдение относится к одному и только одному ряду распределения.

## *Шутливый пример*

В комедии Н.В. Гоголя "Женитьба" у купеческой дочери Агафьи Тихоновны было пятеро женихов. Одного она сразу исключила из рассмотрения. А из остальных она не знала, кого выбрать: "Уж как трудно решиться, так просто рассказать нельзя, как трудно. Если бы Никанора Ивановича да приставить к носу Ивана Кузьмича, да сколько-нибудь развязности, какая у Балтазара Балтазарыча, да, пожалуй, прибавить к этому еще дородности Ивана Павловича, я бы тотчас решилась. А теперь поди подумай! просто голова даже стала болеть. Я думаю, лучше всего кинуть жребий"

И вот Агафья Тихоновна положила бумажки с четырьмя именами в ридикюль, пошарила рукою в ридикюле и вынула вместо одного — всех!

Ей хотелось, чтобы жених совмещал в себе достоинства всех четверых, и, вынимая все бумажки вместо одной, она бессознательно совершала процедуру выведения средней величины. Но вывести среднюю величину из четверых людей невозможно, и Агафья Тихоновна в смятении.

С помощью критерия  $\chi^2$  можно было бы попробовать установить, в кого больше влюблена Агафья Тихоновна. Но для этого нам не нужно измерять губы Никанора Ивановича или нос Ивана Кузьмича, или объем талии дородного экзекутора Ивана Павловича; не нужно нам и пускаться на какие-нибудь опасные эксперименты, чтобы определить, насколько далеко простирается развязность Балтазара Балтазарыча. Мы эти их достоинства принимаем *как* данность потому лишь, что они нравятся Агафье Тихоновне. Мы принимаем их за разряды одного и того же признака, например, направленности взгляда Агафьи Тихоновны: сколько раз она взглянула на губы Никанора Ивановича? На нос Ивана Кузьмича? Благосклонно взирала на дородного Ивана Павловича или развязного Балтазара Балтазаровича? Внимательная сваха или тетушка вполне могла бы этот признак наблюдать.

*Распределение взгляда Агафьи Тихоновны  
между 4 женихами*

Женихи	Никанор Иванович	Иван Кузьмич	Иван Павлович	Балтазар Балтаза- рыч	Всего взглядов
Кол-во взглядов	14	5	8	5	32

## *Гипотезы*

- $H_0$  : Распределение взглядов Агафьи Тихоновны между женихами не отличается от равномерного распределения.
- $H_1$  : Распределение взглядов Агафьи Тихоновны между женихами отличается от равномерного распределения.

Теперь нам нужно определить теоретическую частоту взгляда при равномерном распределении. Если бы все взгляды невесты распределялись равномерно между 4-мя женихами, то, по-видимому, каждый из них получил бы по  $1/4$  всех ее взглядов.

Теоретическая частота при сопоставлении эмпирического распределения равномерным определяется по формуле:

$$f_{\text{теор.}} = n/k$$

где  $n$  - количество наблюдений;

$k$  - количество разрядов признака.

В рассматриваемом примере признак - взгляд невесты, направленный на кого-либо из женихов; количество разрядов признака - 4 направленных взгляда, по количеству женихов; количество наблюдений - 32.

Итак, в нашем случае:  $f_{\text{теор.}} = 32/4 = 8$ .

На Рис. 4.6 сопоставления эмпирических частот с теоретической представлены графически. Похоже, что области расхождений достаточно значительны. Однако для того, чтобы доказать неравномерность получения эмпирического распределения, нам необходимо произвести ТС

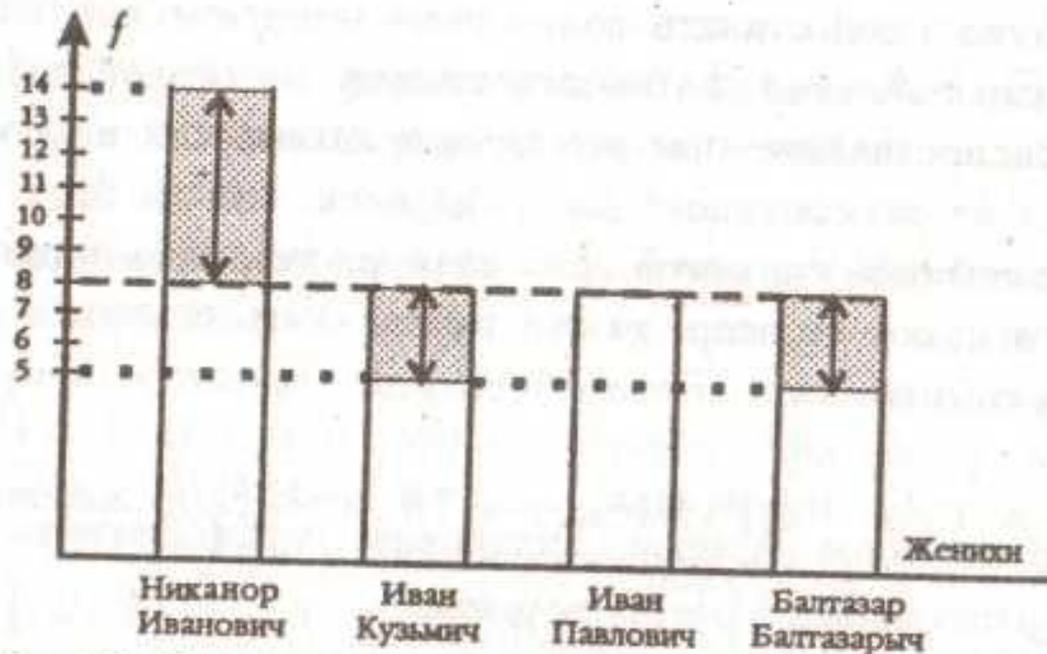


Рис. 4.6. Сопоставление эмпирических частот взгляда Агафьи Тихоновны на каждого из женихов (столбики гистограммы) с теоретической частотой (горизонтальная плавка); темной штриховкой отмечены области расхождений между эмпирическими и теоретическими частотами.

## *Алгоритм расчета критерия $\chi^2$*

1. Занести в таблицу наименования разрядов и соответствующие им эмпирические частоты (первый столбец).
2. Рядом с каждой эмпирической частотой записать теоретическую частоту (второй столбец).
3. Подсчитать разности между эмпирической и теоретической частотой по каждому разряду (строке) и записать их в третий столбец.
4. Определить число степеней свободы по формуле:  $v=k-1$ , где  $k$  - количество разрядов признака.
5. Возвести в квадрат полученные разности и занести их в четвертый столбец.

6. Разделить полученные квадраты разностей на теоретическую частоту и записать результаты в пятый столбец.
7. Просуммировать значения пятого столбца. Полученную сумму обозначить как  $\chi^2_{\text{эмп.}}$ .
8. Определить по таблице критические значения для данного числа степеней свободы  $\nu$ .

**Если  $\chi^2_{\text{эмп.}}$  меньше критического значения, расхождения между распределениями статистически недостоверны.**

**Если  $\chi^2_{\text{эмп.}}$  равно критическому значению или превышает его, расхождения между распределениями статистически достоверны.**

Расчет критерия  $\chi^2$  при сопоставлении эмпирического распределения взгляда Агафьи Тихоновны между женихами с равномерным распределением

Разряды - женихи	Эмпирическая частота взгляда ( $f_{эi}$ )	Теоретическая частота ( $f_T$ )	$f_{эi} - f_T$	$(f_{эi} - f_T)^2$	$(f_{эi} - f_T)^2 / f_T$
1 Никанор Иванович	14	8	+6	36	4,500
2 Иван Кузьмич	5	8	-3	9	1,125
4 Иван Павлович	8	8	0	0	0
5 <u>Балтазар Балтазарыч</u>	5	8	-3	9	1,125
Суммы	32	32	0		6,750

Необходимо всякий раз убеждаться в том, что сумма разностей между эмпирическими и теоретической частотами (сумма по третьему столбцу) равна 0.

**Алгоритм вычислений выражается формулой:**

$$\chi^2 = \sum_{j=1}^k \frac{f_{эj} - f_T}{f_T}$$

где  $f_{эj}$  - эмпирическая частота по  $j$ -тому разряду признака;

$f_T$  - теоретическая частота;

$j$  - порядковый номер разряда;

$k$  - количество разрядов признака.

В данном случае:

$$\chi^2 = \frac{(14 - 8)^2}{8} + \frac{(5 - 8)^2}{8} + \frac{(8 - 8)^2}{8} + \frac{(5 - 8)^2}{8} = 6,75$$

Чтобы установить критические значения, нам нужно определить число степеней свободы  $\nu$  по формуле:  $\nu = k - 1$ ,

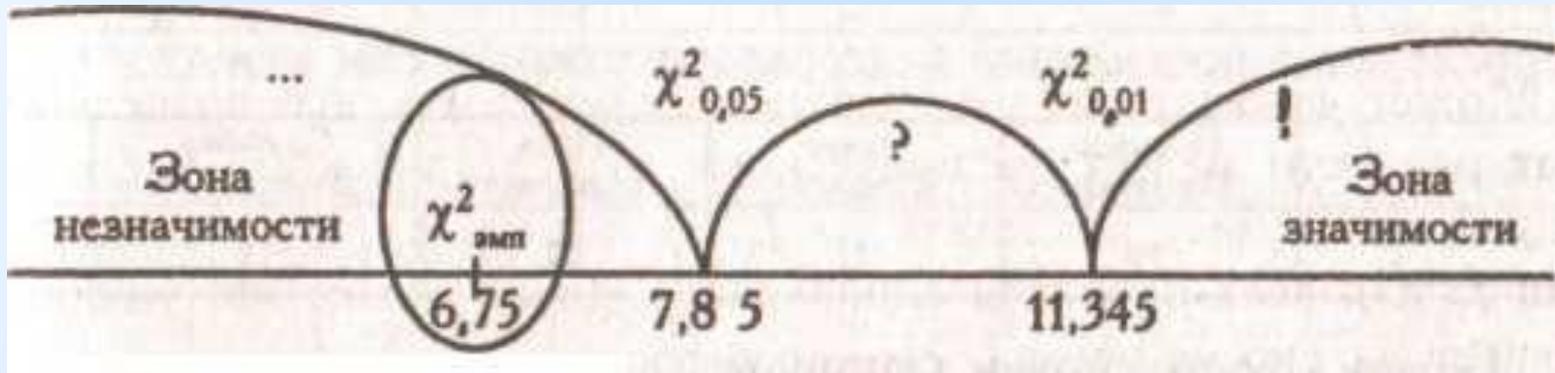
где  $k$  - количество разрядов.

В нашем случае  $\nu = 4 - 1 = 3$ .

По таблице определяем:

$$\chi^2 = \begin{cases} 7,815 & (p \leq 0,05) \\ 11,343 & (p \leq 0,01) \end{cases}$$

## Ось значимости



Чем больше отклонения эмпирических частот от теоретической, тем больше будет величина  $\chi^2$ .

Поэтому зона значимости располагается справа  $\chi^2$ , а зона незначимости – слева.

В данном случае  $\chi^2_{\text{эмп}} < \chi^2_{\text{кр}}$ . Следовательно, принимается гипотеза  $H_0$ . Распределение взгляда Агафьи Тихоновны между женихами не отличается от равномерного распределения

Допустим, тетушка Агафьи Тихоновны на этом не успокоилась. Она стала внимательно следить за тем, сколько раз племянница упомянет в разговоре каждого из женихов. ею получено следующее распределение упоминаний Агафьей Тихоновной женихов и их достоинств: Никанор Иванович - 15 раз, Иван Кузьмич - 6 раз, Иван Павлович - 9 раз, Балтазар Балтазарыч - 6 раз.

Тетушка уже видит, что похоже, Никанор Иванович ("уж такой деликатный, а губы, мать моя, - малина, совсем малина") пользуется большей благосклонностью Агафьи Тихоновны, чем все остальные женихи. У нее есть два пути, чтобы это доказать статистически. 1) Суммировать все проявления благосклонности со стороны невесты: взгляды + упоминания в разговоре, - и сопоставить полученное распределение с равномерным. Поскольку количество наблюдений возросло, есть шанс, что различия окажутся достоверными. 2) Сопоставить два эмпирических распределения - взгляда и упоминаний в разговоре, - с тем, чтобы показать, что они совпадают между собой, то есть и во взглядах, и в словах Агафья Тихоновна придерживается одинаковой системы предпочтений.

*Первый вариант развития шутливого  
примера: увеличение количества  
наблюдений*

Распределение проявлений благосклонности невесты  
между женихами

Женихи	Никанор Иванович	Иван Кузьмич	Иван Павлович	Балтазар Балтаза- рыч	Всего
Кол-во проявле- ний	29	11	17	11	68

$H_0$  : Распределение проявлений благосклонности невесты (взгляды и упоминания в разговоре) не отличается от равномерного распределения.

$H_1$  : Распределение проявлений благосклонности невесты отличается от равномерного распределения. Все расчеты произведем в таблице по алгоритму.

Расчет критерия  $\chi^2$  при сопоставлении проявлений благосклонности  
Агафьи Тихоновны с равномерным распределением

Разряды - женихи	Эмпирическая частота взгляда ( $f_{эj}$ )	Теоретическая частота ( $f_T$ )	$f_{эj} - f_T$	$(f_{эj} - f_T)^2$	$(f_{эj} - f_T)^2 / f_T$
1 Никанор Иванович	29	17	12	144	8,47
2 Иван Кузьмич	11	17	-6	36	2,12
3 Иван Павлович	17	17	0	0	0
4 <u>Балтазар Балтазарыч</u>	11	17	-6	36	2,12
Суммы	68	68	0		12,71

$$f_T = n/k = 68/4 = 17$$

$$v = k - 1 = 3$$

$$\chi^2 = \begin{cases} 7,815 & (p \leq 0,05) \\ 11,343 & (p \leq 0,01) \end{cases}$$

$$\chi^2_{\text{эмп.}} > \chi^2_{\text{кр.}}$$

Но отклоняется, принимается  $H_1$ . Распределение проявим благосклонности невесты между женихами отличается от равномерного распределения ( $p < 0,01$ ).

***Второй вариант развития шутливого  
примера: сопоставление двух эмпирических  
распределений***

**Вопрос:** одинакова ли система предпочтений проявляется во взгляде Агафьи Тихоновны и ее словах?

**Гипотезы:**

**H<sub>0</sub> :** Распределения невербально и вербально выражаемых предпочтений не различаются между собой.

**H<sub>1</sub> :** Распределения невербально и вербально выражаемых предпочтений различаются между собой.

Для подсчета теоретических частот составим специальную таблицу (Табл. 4.5). Ячейки в двух столбцах слева обозначим буквами. Для каждой из них теперь будет подсчитана особая, только к данной ячейке относящаяся, теоретическая частота. Это обусловлено тем, что количества взглядов и словесных отзывов невесты о женихах неравны; взглядов 32, а словесных отзывов - 36. Мы должны всякий раз учитывать эту пропорцию.

Таблица 4.5

Эмпирические и теоретические частоты взглядов и упоминаний о женихах

	Эмпирические частоты			Теоретические частоты	
	Взгляда	Упоминаний в <u>разговоре</u>		Взгляда	Упоминаний в <u>разговоре</u>
<i>Ник. Ив.</i>	14 <u>А</u>	15 <u>Б</u>	29	13,63 <u>А</u>	15,37 <u>Б</u>
<i>Ив. Куз.</i>	5 <u>В</u>	6 <u>Г</u>	11	5,17 <u>В</u>	5,83 <u>Г</u>
<i>Ив. Пав.</i>	8 <u>Д</u>	9 <u>Е</u>	17	7,99 <u>Д</u>	9,01 <u>Е</u>
<i>Бал. Бал.</i>	5 <u>Ж</u>	6 <u>З</u>	11	5,17 <u>Ж</u>	5,83 <u>З</u>
<i>Суммы</i>	32	36	68	32	36

Всего проявлений благосклонней отмечено 68, из них 32 - взгляды и 36 - словесные высказывания. Доля взглядов составит  $32/68=0,47$ ; доля упоминаний -  $36/68=0,53$ .

Итак, во всех строках взгляды должны были бы составлять 0,47 всех проявлений по данной строке, а упоминания в разговоре - 0,53 всех проявлений. Теперь, зная суммы проявлений по каждой строке, мы можем рассчитать теоретические частоты для каждой ячейки Табл.4.5.

$$f_{A\text{теор}}=29 \cdot 0,47=13,63$$

$$f_{B\text{теор}}=29 \cdot 0,53=15,37$$

$$f_{B\text{теор}}=11 \cdot 0,47=5,17$$

$$f_{Г\text{теор}}=11 \cdot 0,53=5,83$$

$$f_{д\text{теор}}=17 \cdot 0,47=7,99$$

$$f_{E\text{теор}}=17 \cdot 0,53=9,01$$

$$f_{ж\text{теор}}=11 \cdot 0,47=5,17$$

$$f_{З\text{теор}}=11 \cdot 0,53=5,83$$

*Общая формула подсчета  $f_{\text{теор}}$  для сопоставления двух или более эмпирических распределений*

$$f_{\text{теор}} = \frac{\left( \begin{array}{c} \text{Сумма частот по} \\ \text{соответствующей строке} \end{array} \right) \cdot \left( \begin{array}{c} \text{Сумма частот по} \\ \text{соответствующему столбцу} \end{array} \right)}{\left( \text{Общее количество наблюдений} \right)}$$

Расчет критерия  $\chi^2$  при сопоставлении распределений невербальных и вербальных признаков благосклонности невесты

ячейки частот	таблицы	Эмпирическая частота $f_{ai}$	Теоретическая частота $f_{T}$	$(f_{ai} - f_{T})$	$(f_{ai} - f_{T})^2$	$(f_{ai} - f_{T})^2 / f_{T}$
	А	14	13,63	+0,37	0,14	0,01
	Б	15	15,37	-0,37	0,14	0,01
	В	5	5,17	-0,17	0,03	0,01
	Г	6	5,83	+0,17	0,02	0,00
	Д	8	7,99	+0,01	0,00	0,00
	Е	9	9,01	-0,01	0,00	0,00
	Ж	5	5,17	-0,17	0,03	0,01
	З	6	5,83	+0,17	0,02	0,00
Суммы		68	68	0		0,04

Число степеней свободы при сопоставлении двух эмпирических определений определяется по формуле:

$$v=(k-1) \cdot (c-1).$$

где  $k$  - количество разрядов признака (строк в таблице эмпирических частот);

$c$  - количество сравниваемых распределений (столбцов в таблице эмпирических частот).

В рассматриваемом примере количество разрядов - это количество женихов, поэтому  $k=4$ . Количество сопоставляемых распределений  $c=2$ . Итак, для данного случая,  $\nu=(4-1)(2-1)=3$

Определяем по таблице критические значения для  $\nu=3$

$$\chi^2 = \begin{cases} 7,815 & (p \leq 0,05) \\ 11,343 & (p \leq 0,01) \end{cases}$$

$$\chi^2_{\text{эмп.}} = 0,04 < \chi^2_{\text{кр.}}$$

Но принимается. Распределения невербально и вербально выражаемых невестой предпочтений не различаются между собой.

## *Третий вариант развития шутливого примера: сопоставление встречных выборов*

У Ивана Павловича, а, главное, у Никанора Ивановича, которому невестой отдается столь явное предпочтение, проскальзывают в разговоре по большей части как раз отрицательные и задумчиво-неодобрительные отзывы о невесте: "Нос велик... Нет, не то, не то... Я даже думаю, что вряд ли она знакома с обхождением высшего общества. Да и знает ли она еще по-французски".

Благосклонных отзывов ("А сказать правду - мне понравилась не потому, что полная женщина" и т. п.) поступило:

от Никанора Ивановича - ни одного; от Ивана Кузьмича - 15;  
от Ивана Павловича - 6, от Балтазара Балтазарыча - 18.

Попробуем ответить на вопрос: согласуются ли распределения благосклонных отзывов невесты о женихах и женихов о невесте?

Мы сопоставим два эмпирических распределения с совпадающей классификацией разрядов, но в одном случае это распределение реакций одного человека на четверых других, а в другом случае это реакции четырех человек на одного и того же человека.

### **Гипотезы**

**H<sub>0</sub>**: Распределение положительных отзывов невесты совпадает с распределением положительных отзывов женихов.

**H<sub>1</sub>**: Распределение положительных отзывов невесты не совпадает с распределением положительных отзывов женихов.

**Эмпирические и теоретические частоты положительных высказываний  
невесты о женихах и женихов о невесте**

Разряды женихи	Эмпирические частоты		Сум- мы	Теоретические частоты	
	Положитель- ных высказыва- ний невесты о женихах	Положитель- ных женихов о невесте		Положитель- ных высказываний невесты о женихах	Положитель- ных женихов о невесте
<i>Ник. Ив.</i>	15 А	0 Б	15	7,20 А	7,80 Б
<i>Ив. Куз.</i>	6 В	15 Г	21	10,08 В	10,92 Г
<i>Ив. Пав.</i>	9 Д	6 Е	15	7,20 Д	7,80 Е
<i>Бал. Бал.</i>	6 Ж	18 З	24	11,52 Ж	12,48 З
<b>Суммы</b>	<b>36</b>	<b>39</b>	<b>75</b>	<b>36</b>	<b>39</b>

Теоретические частоты рассчитываем по известной формуле:

$$f_{\text{теор}} = \frac{\left( \begin{array}{c} \text{Сумма частот по} \\ \text{соответствующей строке} \end{array} \right) \cdot \left( \begin{array}{c} \text{Сумма частот по} \\ \text{соответствующему столбцу} \end{array} \right)}{\text{(Общее количество наблюдений)}}$$

$$f_{\text{А теор}} = 15 \cdot 36 / 75 = 7,20$$

$$f_{\text{Б теор}} = 15 \cdot 39 / 75 = 7,80$$

$$f_{\text{В теор}} = 21 \cdot 36 / 75 = 10,08$$

$$f_{\text{Г теор}} = 21 \cdot 39 / 75 = 10,92$$

$$f_{\text{Д теор}} = 15 \cdot 36 / 75 = 7,20$$

$$f_{\text{Е теор}} = 15 \cdot 39 / 75 = 7,80$$

$$f_{\text{Ж теор}} = 24 \cdot 36 / 75 = 11,52$$

$$f_{\text{З теор}} = 24 \cdot 39 / 75 = 12,48$$

Расчет критерия  $\chi^2$  при сопоставлении распределений высказываний невесты о женихах и женихов о невесте

Ячейки таблицы частот		Эмпирическая частота	Теоретическая частота	$f_{3i} - f_T$	$(f_{3i} - f_T)^2$	$(f_{3i} - f_T)^2 / f_T$
1	А	15	7,20	+7,80	60,84	8,45
2	Б	0	7,80	-7,80	60,84	7,80
3	В	6	10,08	-4,08	16,65	1,65
4	Г	15	10,92	+4,08	16,65	1,52
5	Д	9	7,20	+1,80	3,24	0,45
6	Е	6	7,80	-1,80	3,24	0,42
7	Ж	6	11,52	-5,52	30,47	2,64
8	З	18	12,48	+5,52	30,47	2,44
Суммы		75	75	0		25,37

$$v=(k-1)-(c-1)=3$$

$$\chi^2 = \begin{cases} 7,815 & (p \leq 0,05) \\ 11,343 & (p \leq 0,01) \end{cases}$$

$$\chi^2 = 25,37$$

$$\chi^2_{\text{эмп}} > \chi^2_{\text{кр.}}$$

Ответ:  $H_0$  отвергается. Принимается  $H_1$ . Распределение положительных отзывов предпочтений невесты не совпадает с распределением положительных отзывов женихов ( $p < 0,01$ )