# Лекция №4.

# <u>Проверка статистических гипотез.</u>

***Статистическая гипотеза*** -- это предположение о генеральной совокупности, высказанное на основании статистических выборочных данных.

***Статистическая проверка гипотез*** -- это процедура обоснованного сопоставления высказанной гипотезы с имеющимися выборочными данными.

*Например:* исследуем влияние нового лекарственного препарата на снижение артериального давления.

**$X\{x_1, x_2, \dots x_{n1}\}$** -- контрольная группа (выборка, объёмом $n_1$)

**$Y\{y_1, y_2, \dots y_{n2}\}$** -- опытная группа (выборка объёмом $n_2$)

*Высказываются две альтернативные гипотезы:*

**$H_0$:** -- различия между выборками ***статистически не значимы*** (***не достоверны*** т.е. носят случайный характер).

**$H_1$:** -- различия между выборками ***статистически значимы*** (***достоверны*** т.е. влияние препарата достоверно (эффективно))

Чтобы принять или опровергнуть эти предположения, используют *статистические критерии или критерии достоверности.*

***Статистический критерий*** -- это случайная величина, закон распределения которой известен, т.е. каждому значению критерия поставлена в соответствие вероятность ***p***, с которой он эти значения принимает.

Т.к. решение об отклонении или принятии статистической гипотезы принимаются по выборочным данным, то возможны ***ошибочные решения***.

**Ошибка 1-го рода:** *отвергают* нулевую гипотезу, когда она *правильна* (истинна), и делают вывод, что *имеется эффект*, когда в действительности *его нет.*
Вероятность допустить ошибку 1-го рода обозначается **α** (альфа). Это *уровень значимости критерия*. (Обычно α = 0,05 ; 0,01; 0,005; 0,001).

**Ошибка 2-го рода:** *принимают* нулевую гипотезу, когда она *не правильна*, и делают вывод, что *нет эффекта*, тогда как в действительности *он существует*.

Вероятность возникновения ошибки 2-го рода обозначается **β** (бета); а величина **(1-β)** называется *мощностью критерия*.
Чем больше мощность критерия, тем вероятность ошибки II рода меньше.

Следовательно, ***мощность критерия***— это вероятность обнаружить реальный эффект лечения в выборке данного объема как статистически значимый.

Одновременное уменьшение ошибок 1-го и 2-го рода возможно лишь при увеличении объёма выборок. Поэтому обычно при заданном уровне значимости α отыскивается критерий с наибольшей мощностью.

Для каждого критерия существует ***таблица***, в которой содержатся *критические значения критерия*. Каждое критическое значение соответствует определённому *уровню значимости* ***α*** *и числу степеней свободы* **ν**(*ню*) ***(или к)***

$\nu(k) = n - a$ где ***a*** -- число наложенных связей или ограничений

Сравнение значения критерия, вычисленного по выборке, с табличным (критическим) значением критерия, позволяет сделать вывод о правомерности выдвигаемой гипотезы для данного уровня значимости.

***Например:*** Хотим доказать достоверность различия между выборками $\mathbf{X\{x_1, x_2, \ldots x_{n1}\}}$ и $\mathbf{Y\{y_1, y_2, \ldots y_{n2}\}}$ с $P_Д$=0,95 (это значит, что влияние препарата достоверно (эффективно) на 95%).

. Если в результате проверки выяснилось, что вычисленному значению критерия соответствует вероятность ***p*** большая, чем заданный уровень значимости (α=1-0,95=0,05), то нулевая гипотеза принимается.

# Основные этапы проверки статистических гипотез.

1).Выдвигается гипотеза $H_{0.}$

2).Выбирается величина уровня значимости α ($\alpha=1-P_Д$).

3).По заданному *α и числу степеней свободы v(или к)* в таблице находим критическое (табличное) значение критерия.

4).Подсчитывается экспериментальное значение критерия по имеющимся выборкам (для каждого критерия существует *формула* для определения значения критерия).

5).С помощью сравнения экспериментального и критического значений делается вывод о правомерности гипотезы $H_0$.

*6).Если $H_0$ принимается*, следовательно гипотеза $H_1$ (о достоверности различий) не верна.

*Если $H_0$ отвергается*, следовательно верна гипотеза $H_1$ ($H_0$ и $H_1$ -- противоположные события).

Критерии достоверности подразделяются на *параметрические и непараметрические.*

***Параметрические*** критерии для вычисления экспериментального значения используют статистические параметры: $\bar{x}, S_n^2, S_n\ S_{\bar{x}}$ . Они могут использоваться только для выборочных совокупностей, ***распределённых по закону близкому к нормальному (Гаусса).***

***Непараметрические критерии*** не используют статистические параметры (следовательно и не оценивают их), требуют большего объёма выборок, они менее точны, дают более грубую оценку, чем параметрические критерии, но:

1). Их можно применять к выборкам, закон распределения которых неизвестен (***не обязательно нормальное распределение***).

2). Они проще и позволяют быстрее производить проверку рассматриваемых гипотез.

## •1. Критерии отклонения распределения от нормального.

Очень многие статистические совокупности, встречающиеся в биологической практике, имеют нормальное, или почти нормальное распределение. Вместе с тем, нередки случаи, когда распределение не является нормальным даже приблизительно.

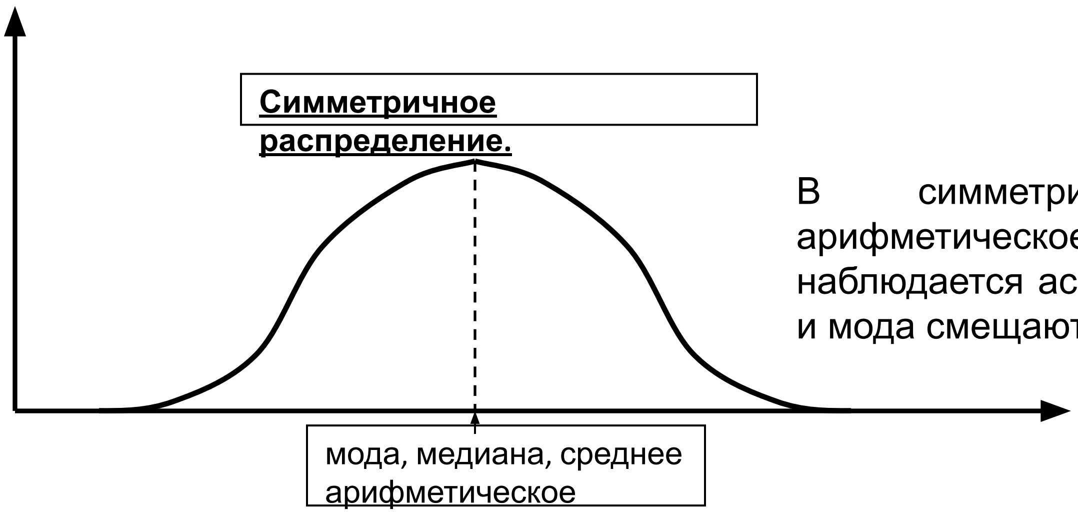
***Асимметрия и эксцесс*** – основные показатели, наиболее чувствительные к отклонению от нормальности.

### •1.1.Коэффициент асимметрии.

Кроме ***среднего арифметического***, существуют такие статистические характеристики совокупности как ***медиана и мода***.

***Медиана*** — это такое значение признака, которое разделяет ранжированный ряд распределения на две равные части — со значениями признака меньше медианы и со значениями признака больше медианы. Для нахождения медианы, нужно отыскать значение признака, которое находится на середине упорядоченного ряда.

***Мода*** — *значение признака*, имеющее наибольшую частоту в статистическом ряду распределения.



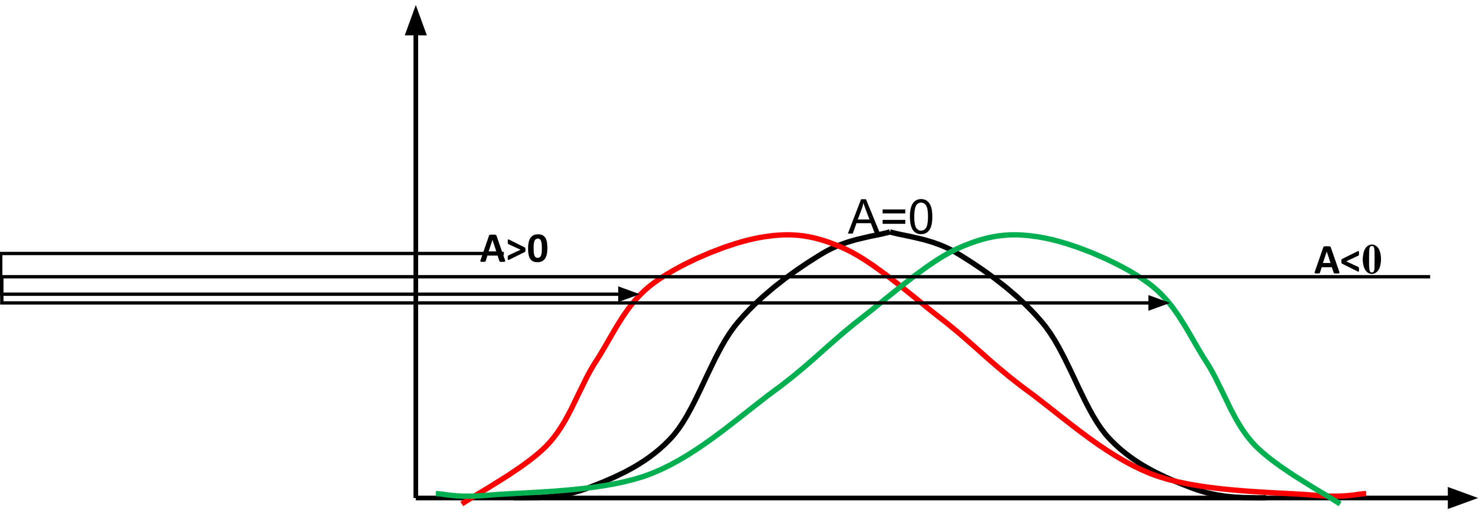

В симметричном распределении среднее арифметическое, медиана и мода совпадают, если же наблюдается асимметрия, то среднее арифметическое и мода смещаются относительно медианы.

Асимметрию оценивают по формуле:

$$A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot \sigma^3} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 \cdot p_i}{\sigma^3}$$

*где К – количество интервалов*
*n – объём выборки*

Знак при коэффициенте асимметрии указывает на направление асимметрии.




Если **A<0,** то это означает перевес наблюдений в правой части, а хвост кривой вытянут *влево.* Это ***левосторонняя асимметрия***.

Если **A>0,** то это означает перевес наблюдений в левой части, а хвост кривой вытянут *вправо*. Это ***правосторонняя асимметрия***.

Если **A=0**, то распределение ***симметрично***.

При большой асимметрии коэффициент асимметрии может быть и больше единицы.

$H_0$: Отличие коэффициента асимметрии от нуля носит случайный характер, то есть распределение нормально по асимметрии.

Вычисляем коэффициент асимметрии по экспериментальным данным по формуле:

$$A_{эксп} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 \cdot p_i}{\sigma^3}$$

***где К – количество интервалов***

***Таблица значений асимметрии***

| *N* | *α=0,05* | *α=0,01* |
|---|---|---|
| 10 | 1,13 | 1,49 |
| 20 | 0,92 | 1,21 |
| 30 | 0,79 | 1,05 |
| 40 | 0,71 | 0,93 |
| 50 | 0,63 | 0,84 |
| 60 | 0,59 | 0,78 |
| 80 | 0,52 | 0,68 |
| 100 | 0,47 | 0,62 |

Сравниваем $A_{эксп}$ с табличным (критическим) значением, которое находим в таблице критерия асимметрии для заданного уровня значимости ***ά.***

*Если* $|A_{эксп}| \leq A_{крит} \Rightarrow$ $H_0$ принимаем

***Вывод:*** экспериментальное распределение соответствует нормальному по асимметрии.

*Если* $|A_{эксп}| > A_{крит} \Rightarrow$ $H_0$ отвергаем.

***Вывод:*** экспериментальное распределение не соответствует нормальному по асимметрии.

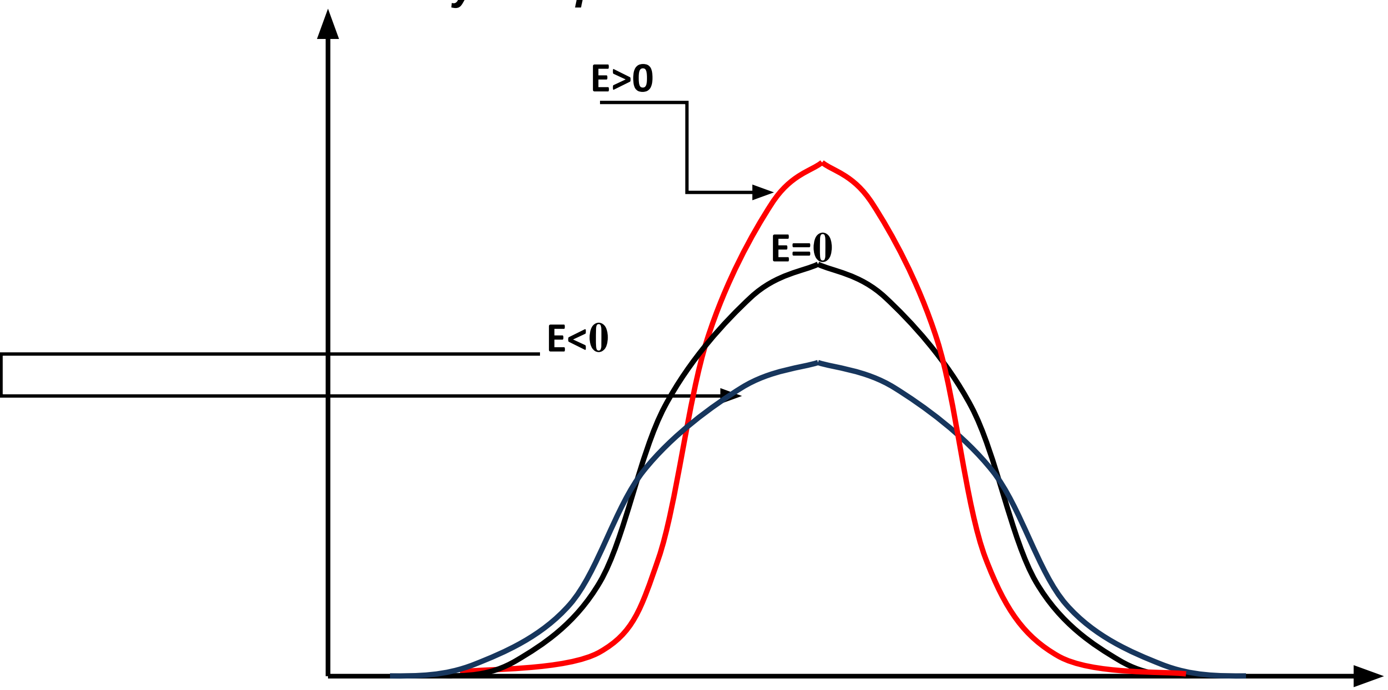
# 1.2 Эксцесс.

Иногда этот показатель называют крутостью кривой. Эксцесс вычисляется по формуле:

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \cdot \sigma^4} - 3 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 \cdot p_i}{\sigma^4} - 3$$

*где К – количество интервалов*

Если E>0 , *то кривая называется* ***островершинной,***
если E <0 ***туповершинной.***



$H_0$: Отличие эксцесса от нуля носит случайный характер, то есть распределение ***нормально по эксцессу***.

Вычисляем ***эксцесс*** по экспериментальным данным по формуле:

*где К – количество интервалов*

$$E_{эксп} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 \cdot p_i}{\sigma^4} - 3$$

Сравниваем $E_{эксп}$ с табличным (критическим) значением, которое находим в таблице критерия эксцесса для заданного уровня значимости ***ά.***

***Таблица значений эксцесса***

| *N* | *α=0,05* | *α=0,01* |
|---|---|---|
| 10 | 1,43 | |
| 20 | 1,41 | 1,95 |
| 30 | 1,31 | 1,78 |
| 40 | 1,19 | 1,62 |
| 50 | 1,11 | 1,50 |
| 60 | 1,05 | 1,42 |
| 80 | 0,94 | 1,25 |
| 100 | 0,85 | 1,14 |

*Если* $|E_{эксп}| \le E_{крит} \Rightarrow$ $H_0$ принимаем.

***Вывод:*** экспериментальное распределение соответствует нормальному по эксцессу.

*Если* $|E_{эксп}| > E_{крит} \Rightarrow$ $H_0$ отвергаем.

***Вывод:*** экспериментальное распределение не соответствует нормальному по эксцессу.

# Проверка гипотез о законе распределения.

## Критерий Пирсона ($\chi^2$).

Критерий χ2 применяется в двух целях:
**1)**.для сопоставления *эмпирического* распределения признака *с теоретическим* – равномерным, нормальным или каким-то иным.
**2)** для сопоставления *двух или более эмпирических* распределений одного и того же признака.

**Сопоставление *эмпирического* распределения признака с *теоретическим.***

**$H_0$** заключается в том, что различие между наблюдаемыми экспериментальными частотами $m_i$ попадания вариант выборки в интервалы вариационного ряда от вычисленных теоретических частот $m_{i\ теор}=n\cdot P_{i\ теор}$ не достоверно (т.е. носит случайный характер). Другими словами:

**$H_0$:** экспериментальные данные соответствуют предложенному теоретическому закону распределения.

Экспериментальное значение критерия вычисляется по формуле:

$$P_i = \frac{m_i}{n} \Rightarrow m_i = n \cdot P_i$$

$$\boxed{\chi^2_{эксп} = \sum_{i=1}^{k} \frac{\left(m_{i\ эксп} - m_{i\ теор}\right)^2}{m_{i\ теор}} = \sum_{i=1}^{k} \frac{\left(m_{i\ эксп} - n\cdot P_{i\ теор}\right)^2}{n\cdot P_{i\ теор}}}$$

где $\sum_{i=1}^{k} m_i = n$ -- объём выборки, **к** -- количество интервалов,

$P_{i\ теор}$ -- вероятность попадания в интервал для теоретического распределения.

Затем, по таблице критерия Пирсона для ***заданного уровня значимости α и числа степеней свободы*** $\nu = к - a$

где а -- ***число наложенных связей***, находим

$$\chi^2_{критич} = \chi^2_{табличное}$$

если теоретическое распределение *произвольное*, то *а=1,* $\nu = к - 1$

если теоретическое распределение *распределено по нормальному закону Гаусса,* то *а=3* -- числу наложенных связей, необходимых для вычисления вероятности: n,M[X],и σ[X],. $\nu = к - 3$

***Если*** $\chi^2_{эксп} \le \chi^2_{крит.} \Rightarrow$ ***аем***.

*Вывод:* экспериментальное распределение соответствует теоретическому.

***Если*** $\chi^2_{эксп} > \chi^2_{крит.} \Rightarrow$ ***аем***

.

*Вывод:* экспериментальное распределение не соответствует теоретическому.

**Пример:** Изучался рост 50 человек. В таблице приведены экспериментальные частоты попадания в интервал $m_i$ и теоретические частоты, рассчитанные из вероятностей попадания в интервал для распределения Гаусса. *К=5 , n=50.v=5-3=2,*

| № интервала | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $m_i$ практические | 5 | 9 | 22 | 8 | 6 |
| $m_i$ теоретические $m_{i\ \text{теор}} = n \cdot P_{i\ \text{теор}}$ | 5 | 10 | 20 | 10 | 5 |
| $P_{i\ \text{теор}}$ | 0,1 | 0,2 | 0,4 | 0,2 | 0,1 |

**$H_0$:** Экспериментальное (практическое) распределение соответствует распределению Гаусса. (То есть различие между частотами не достоверно, носит случайный характер).

Из таблицы для ***v=5-3=2 и ά=0,05*** находим $\chi^2_{\text{крит.}}$ =5,99

$$\chi^2_{\text{эксп}} = \frac{(5-5)^2}{5} + \frac{(9-10)^2}{10} + \frac{(22-20)^2}{20} + \frac{(8-10)^2}{10} + \frac{(6-5)^2}{5} = 0 + \frac{1}{10} + \frac{4}{20} + \frac{4}{10} + \frac{1}{5} = \frac{1+2+4+2}{10} = \mathbf{0,9}$$

Т.к. $\underset{0,9}{\chi^2_{\text{эксп}}} < \underset{5,99}{\chi^2_{\text{крит}}} \Rightarrow$ $H_0$ принимаем.

.

***Вывод:*** практическое распределение соответствует распределению Гаусса.

***Значения критерия Пирсона (критерия $\chi^2$)***

| Число степеней свободы, ν | ***α=0,05*** | ***α=0,01*** |
|---|---|---|
| 1 | 3,84 | 6,63 |
| 2 | 5,99 | 9,21 |
| 3 | 7,81 | 11,3 |
| 4 | 9,49 | 13,3 |
| 5 | 11,1 | 15,1 |
| 6 | 12,6 | 16,8 |
| 7 | 14,1 | 18,5 |
| 8 | 15,5 | 20,1 |
| 9 | 16,9 | 21,7 |
| 10 | 18,3 | 23,2 |

## *Сопоставления двух или более эмпирических распределений.*

Выборки X и Y – *дихотомические признаки*, то есть могут принимать только две категории значений (например ДА-Нет, **+** или **—**, получил – не получил и т.д.).

***Четырехпольная таблица сопряженности выглядит следующим образом:***

***Таблица 1.***

| | Да (1) | Нет (0) | Всего |
|---|---|---|---|
| **Да (1)** | **A** | **B** | **A + B** |
| **Нет (0)** | **C** | **D** | **C + D** |
| **Всего** | **A + C** | **B + D** | **A + B + C + D** |

Эта таблица содержит *экспериментальные частоты*, например А – это количество (т.е частота) комбинаций (1,1) и т.д.

*Теоретические (ожидаемые) частоты* рассчитываются следующим образом:

***Таблица 2.***

| | **Исход есть (1)** | **Исхода нет (0)** | **Всего** |
|---|---|---|---|
| **Да (1)** | **(A+B)*(A+C) / (A+B+C+D)** | **(A+B)*(B+D)/ (A+B+C+D)** | **A + B** |
| **Нет(0)** | **(C+D)*(A+C)/ (A+B+C+D)** | **(C+D)*(B+D)/ (A+B+C+D)** | **C + D** |
| **Всего** | **A + C** | **B + D** | **A+B+C+D** |

Значение ***χ2****эксп* подсчитывается по знакомой формуле. Для этого из величин, представленных в ячейках *таблицы 1* вычитаются соответствующие величины из *таблицы2* (**m**-номер столбца, **n**-номер строки*)* *,* ***k=m·n:***

*Международное обозначение частоты* ***f*** *(frequency).*

$$\chi^2_{\text{эксп}} = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(m_{ij\text{эксп}} - m_{ij\text{теор}}\right)^2}{m_{ij\text{теор}}} = \boxed{\chi^2_{\text{эксп}} = \sum_{i=1}^{k}\frac{\left(m_{i\,\text{эксп}} - m_{i\,\text{теор}}\right)^2}{m_{i\,\text{теор}}}}$$

В том случае, если число ожидаемого явления *меньше 10 хотя бы в одной ячейке*, при анализе четырехпольных таблиц должен рассчитываться **критерий χ2 с поправкой Йейтса**

*Данная поправка позволяет уменьшить вероятность ошибки первого рода, т.е. обнаружения различий там, где их нет.*

$$\chi^2_{\text{эксп}} = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(\left|m_{ij\text{эксп}} - m_{ij\text{теор}}\right| - 0{,}5\right)^2}{m_{ij\text{теор}}} = \boxed{\chi^2_{\text{эксп}} = \sum_{i=1}^{k}\frac{\left(\left|m_{i\text{эксп}} - m_{i\text{теор}}\right| - 0{,}5\right)^2}{m_{i\,\text{теор}}}}$$

Затем, по таблице критерия Пирсона для заданного уровня значимости **α** и *числа степеней свободы* $\nu = (m-1)\cdot(n-1)$ находим $\chi^2_{\text{критич}}$ Для четырёхпольной таблицы ***m=2, n=2→γ=1.***

*Если* $\chi^2_{\text{эксп}} \le \chi^2_{\text{крит.}} \Rightarrow$ $H_0$ принимаем.

*Вывод:* различие между наблюдаемыми распределениями *не достоверно* (статистически не значимо)

*Если* $\chi^2_{\text{эксп}} > \chi^2_{\text{крит.}} \Rightarrow$ $H_0$ отвергаем.

*Вывод:* различие между наблюдаемыми распределениями *достоверно* (статистически значимо)

***Пример.*** Определим статистическую значимость влияния фактора курения на частоту случаев артериальной гипертонии по рассмотренной ниже таблице:

| | Артериальная гипертония есть (1) | Артериальной гипертонии нет (0) | **Всего** |
|---|---|---|---|
| Курящие (1) | 40 | 30 | **70** |
| Некурящие (0) | 32 | 48 | **80** |
| **Всего** | **72** | **78** | **150** |

$H_0$: Зависимость между курением и гипертонией не достоверна.

*Рассчитываем ожидаемые значения для каждой ячейки:*

| | Артериальная гипертония есть (1) | Артериальной гипертонии нет (0) | **Всего** |
|---|---|---|---|
| Курящие (1) | (70*72)/150 = **33.6** | (70*78)/150 = **36.4** | **70** |
| Некурящие (0) | (80*72)/150 = **38.4** | (80*78)/150 = **41.6** | **80** |
| **Всего** | **72** | **78** | **150** |

$$\chi^2_{\text{эксп}} = \chi^2_{\text{эксп}} = \sum_{i=1}^{k} \frac{(m_{i\,\text{эксп}} - m_{i\,\text{теор}})^2}{m_{i\,\text{теор}}} = \frac{(40-33{,}6)^2}{33{,}6} + \frac{(30-33{,}4)^2}{33{,}4} + \frac{(32-38{,}4)^2}{38{,}4} + \frac{(48-41{,}6)^2}{41{,}6} = 4{,}396$$

Число степеней свободы γ= (2-1)*(2-1) = 1. Находим по таблице критерия Пирсона, для α=0,05 и v=1 $\chi^2_{\text{крит.}} = 3{,}84$

$\chi^2_{\text{эксп}} > \chi^2_{\text{крит.}} \Rightarrow$ $H_0$ отвергаем. *Следовательно, зависимость частоты случаев артериальной гипертонии от наличия курения достоверна.*

$\chi^2_{эксп}$ *для двухпольных таблиц (2х2)можно рассчитать и по упрощенной формуле:*

$$\chi^2_{эксп} = \frac{(AD - BC)^2 \cdot (A + B + C + D)}{(A + B) \cdot (C + D) \cdot (A + C) \cdot (B + D)}$$

Таблицы сопряженности могут иметь и более сложный вид, когда каждый признак имеет более двух градаций (многопольные таблицы). Например, таблица 3х4:

| | ***профессия*** | | | | |
|---|---|---|---|---|---|
| ***обращаемость к врачу*** | *строители* | *шахтеры* | *учителя* | *госслужащие* | ***всего*** |
| *до 3 в год* | *21* | *26* | *19* | *17* | *83* |
| *от 4 до 6 в год* | *9* | *15* | *12* | *6* | *42* |
| *более 6 в год* | *7* | *8* | *6* | *4* | *25* |
| ***всего*** | *37* | *49* | *37* | *27* | *150* |

*Анализ таких таблиц предпочтительно проводить с использованием компьютерных программ.*

Сопоставляемые группы должны быть *независимыми*, то есть критерий **χ2** не должен применяться при сравнении наблюдений "до-"после". В этих случаях проводится **тест Мак-Немара** (*при сравнении двух связанных совокупностей*) или рассчитывается **Q-критерий Кохрена** (*в случае сравнения трех и более групп*).

# ***Критерий Мак-Немара (McNemar's test).***

Используется для анализа *таблиц сопряженности размером 2х2* (для дихотомического признака, который имеет только две категории: да-нет). В отличие от критерия **χ2,** критерий Мак-Немара применяется, когда условие независимости наблюдений не выполняется, но, напротив, *учет признака выполняется на одних и тех же субъектах*.

| ДО / ПОСЛЕ | 1 | 0 | Всего |
|---|---|---|---|
| 1 | A | B | A + B |
| 0 | C | D | C + D |
| Всего | A + C | B + D | A + B + C + D |

***$H_0$*** **:**Различие значений исследуемого показателя до и после эксперимента не достоверно

Значение критерия вычисляется по формуле:

$$M_{эксп} = \frac{(B - C)^2}{B + C}$$

Для низкочастотных выборок (хотя бы в одной ячейке число наблюдений меньше 10) используют *поправку Йейтса:*

$$\mathrm{M}_{эксп} = \frac{(|B - C| - 0{,}5)^2}{B + C}$$

*Для достаточно больших выборок (B+C)>25 используют распределение* ***χ2*** *для числа степеней свободы γ=1.*

*Если* $\mathrm{M}_{эксп} \le \chi^2_{крит.} \Rightarrow$ $H_0$ принимаем.

*Вывод:* Различие значений исследуемого показателя до и после эксперимента *не достоверно*

*Если* $\mathrm{M}_{эксп} > \chi^2_{крит.} \Rightarrow$ $H_0$ отвергаем.

*Вывод:* Различие значений исследуемого показателя до и после эксперимента *достоверно*

***Пример.*** Учащиеся тестировались до и после проведения тренинга по повышению качества усвоения учебного материала. Экспериментальные данные, представляют итог прохождения теста: **+**– тест пройден успешно; **--** – тест не пройден.

| | *Второе тестирование* | |
|---|---|---|
| *Первое тестирование* | *справились* | *Не справились* |
| *справились* | ***A=50*** | ***B=19*** |
| *не справились* | ***C=31*** | ***D=20*** |

***$H_0$ :***Различие значений исследуемого показателя до и после эксперимента не достоверно

$$M_{эксп} = \frac{(B-C)^2}{B+C} = \frac{(19-31)^2}{19+31} = \frac{144}{50} = 2{,}88$$

$$\mathrm{M}_{\text{крит}} = \chi^2_{\text{крит}} = 3{,}84 \text{ для } \alpha = 0{,}05 \text{ и } \gamma = 1$$

$$\mathrm{M}_{\text{эксп}} \leq \chi^2_{\text{крит.}} \quad \Rightarrow$$ $H_0$ принимаем.

*Вывод:* Различие значений исследуемого показателя до и после эксперимента не достоверно.

# Контрольные вопросы.

1.Что такое статистическая гипотеза и критерии проверки статистических гипотез?

2. Основные этапы проверки статистических гипотез.

3. Критерий Асимметрии.

4. Критерий Эксцесса.

5. Критерий Пирсона ($\chi^2$).

6. Критерии Пирсона и Мак-Немара для таблиц сопряженности.