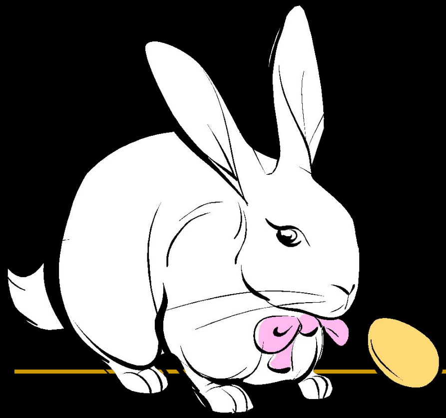
Занятие 6

Непараметрические критерии.

Частотный анализ



Трансформация данных

Повторение из предыдущих занятий

Особенности выборки, необходимые для проведения **параметрических** тестов

1. Случайность измерений (*randomness*)
2. Независимость измерений (*independence*)
3. **Гомогенность** дисперсии (*homogeneity* = *homoscedasticity*)
4. Соответствие **нормальному распределению**
5. Для факторной ANOVA – аддитивность (пояснить с табличкой)

Трансформация данных

Параметрические тесты:

нулевая гипотеза формулируется о конкретных ПАРАМЕТРАХ РАСПРЕДЕЛЕНИЯ и/или эти параметры входят в формулу статистики критерия.
Параметры: среднее значение, стандартное отклонение, дисперсия…

Почему при проведении параметрических тестов важно соблюдать условия?

Нарушим условие соответствия выборки нормальному распределению и проведём одновыборочный t-тест!

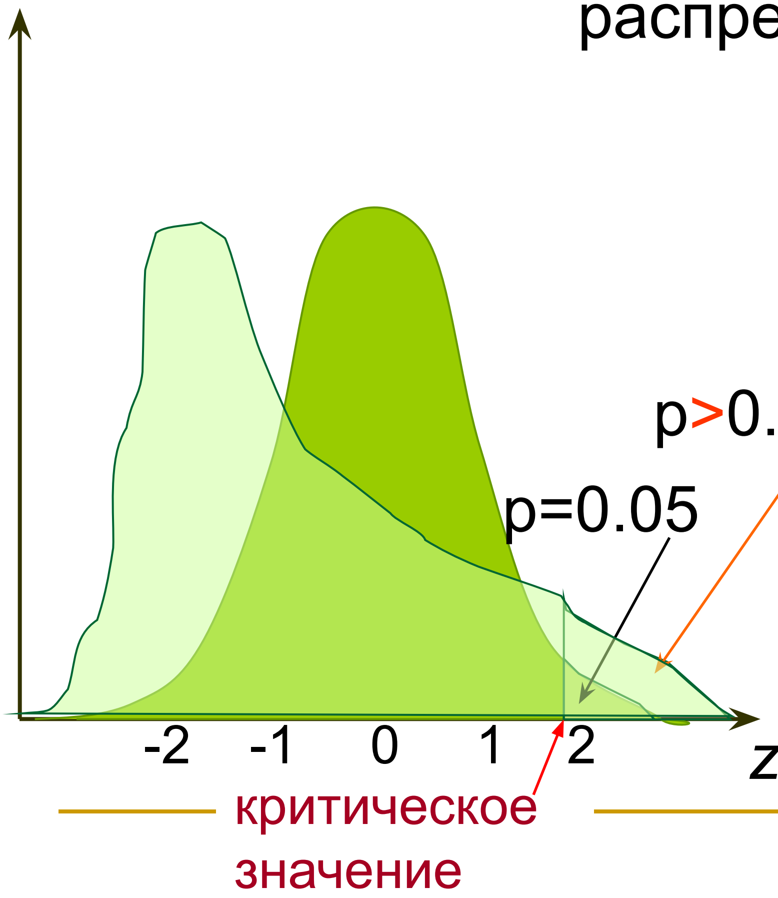
Трансформация данных

H₀

H_0: $\mu \leq 90$ г;
H_1 : $\mu > 90$ г
Пусть $\boldsymbol{\sigma}$ известна.

*Распределение **статистики критерия** не будет нормальным, если в выборке не нормальное распределение.*

Пусть наше распределение <u>скошено</u>. Z-распределение тоже будет скошено!



Вероятность, что среднее в выборке попадёт в критическую область (рассчитанную для нормального распределения), будет <u>выше</u>, чем 0.05 – увеличится ошибка 1-го рода!

Трансформация данных

Основной вывод:
пренебрежения условиями использования параметрических тестов может **увеличивать ошибку 1-го рода.**

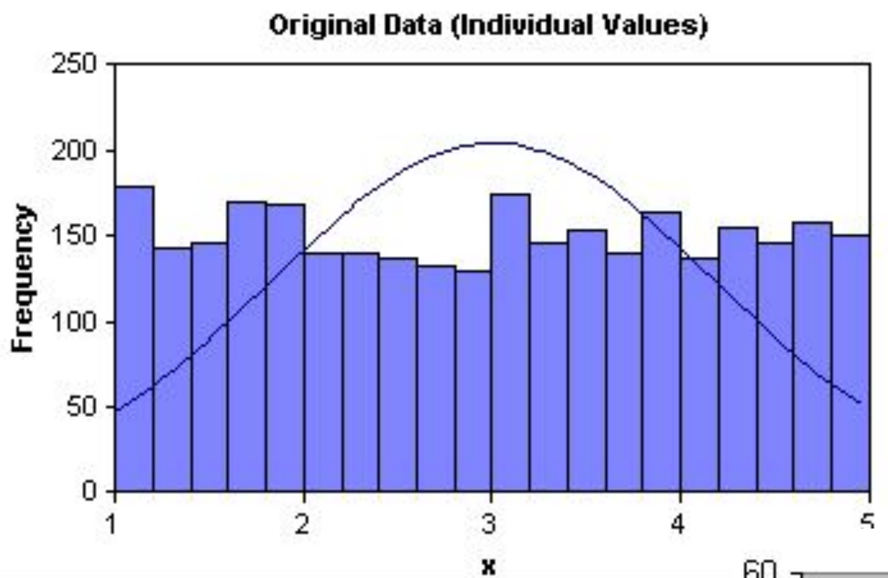
(Неизвестно, насколько)

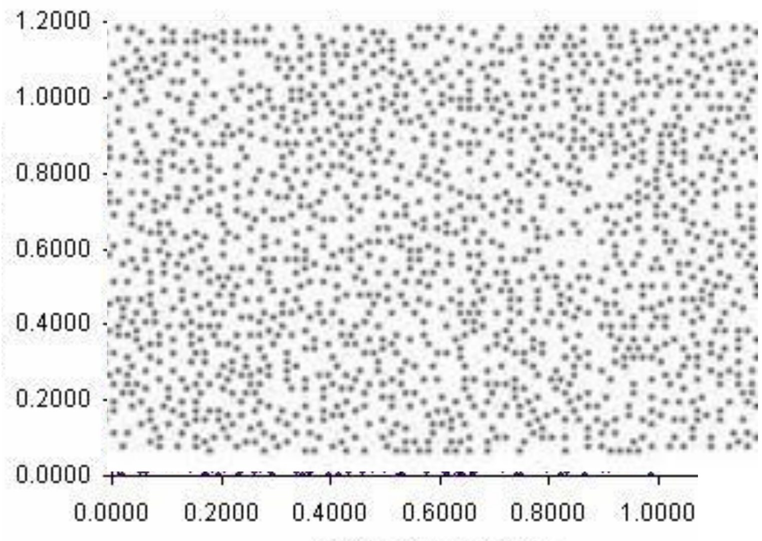
Примечание: слабые отклонения от нормального распределения не очень страшны (в силу Центральной предельной теоремы), а **для больших выборок ими можно пренебречь.**
ANOVA устойчива к отклонениям от нормального распределения, особенно если выборки одинаковы по размеру.

Трансформация данных

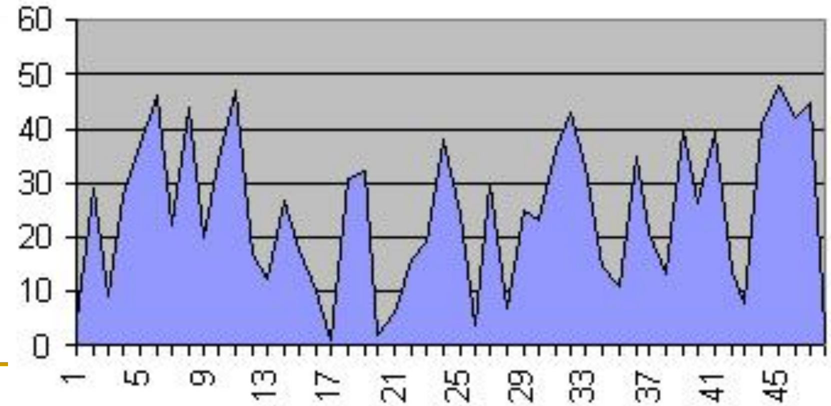
Какие бывают распределения:

1. Равномерное (uniform)





2. Случайное (random)



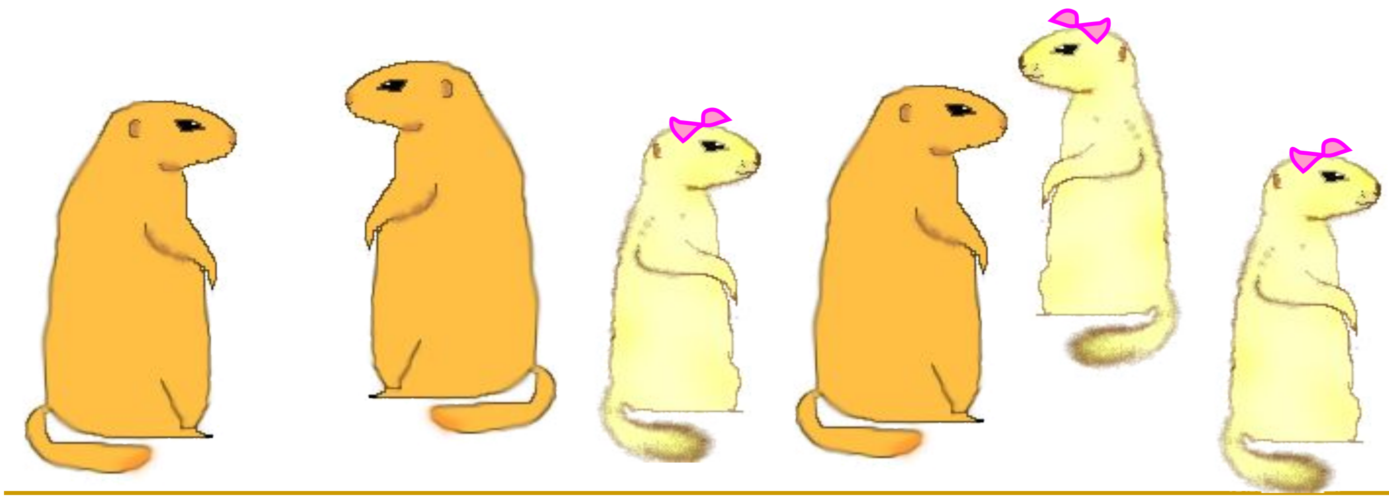
Могут быть и дискретными, и непрерывными

Трансформация данных

3. Биномиальное распределение (дискретное).

Пример: рассмотрим выводки из 6 детёнышей каждый.
Возможное соотношение самцов и самок в выводке:

6:0; 5:1; 4:2; 3:3; 2:4; 1:5; 0:6

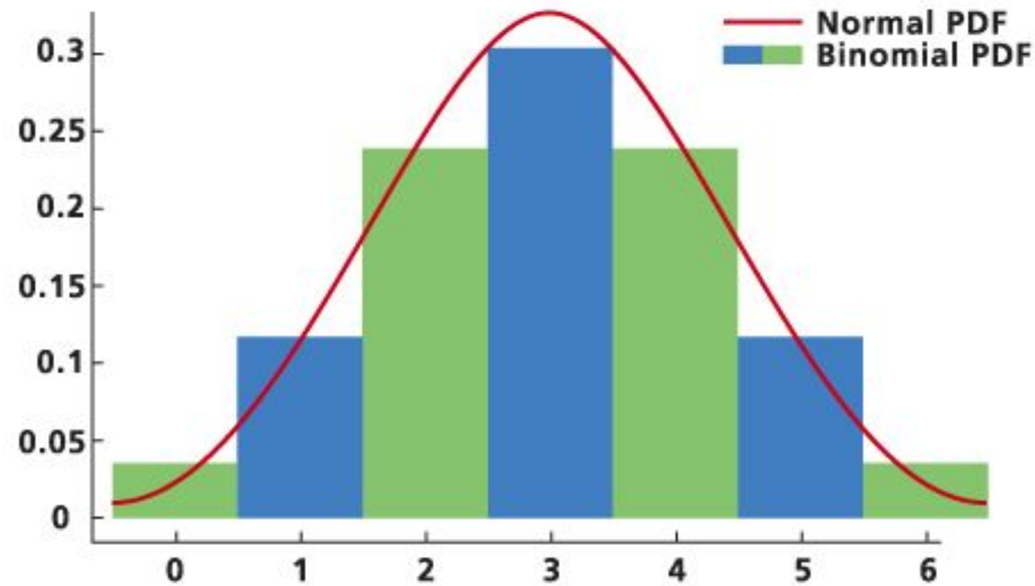


Трансформация данных

Биномиальное распределение

распределение количества «успехов» *(самцов)* в последовательности из N независимых случайных экспериментов, таких что вероятность «успеха» *(рождения самца)* в каждом из них постоянна и равна p.

Вероятность такого выводка



Количество самцов в выводке из 6 зверьков

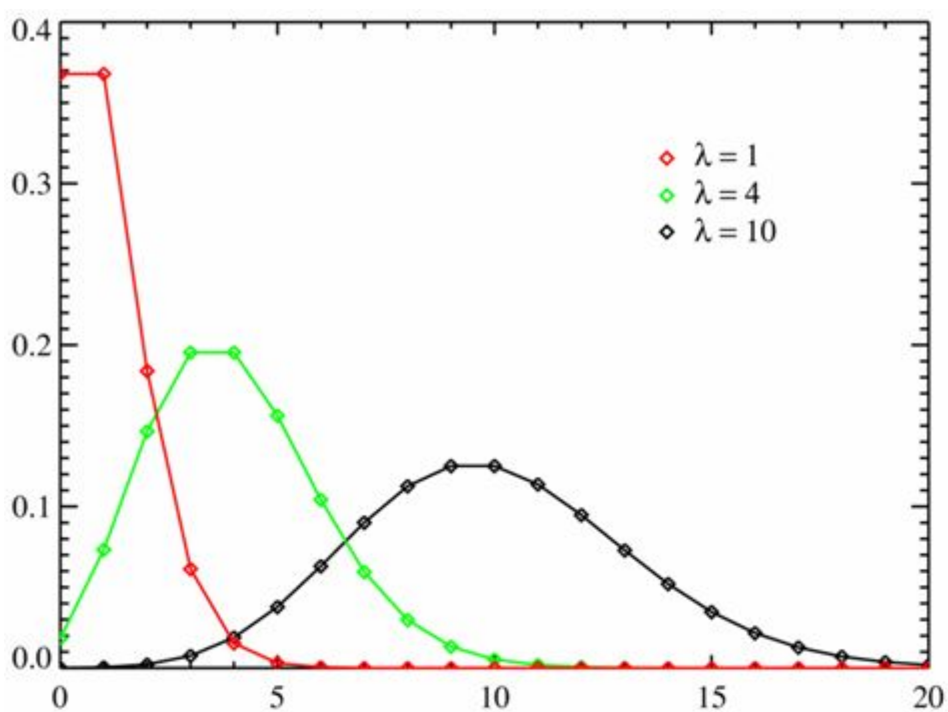
Трансформация данных
Распределение Пуассона

Показывает вероятность того или иного количества независимых друг от друга событий (особей, контактов, мутаций и пр.) на заданном интервале времени (участке пространства, объёме...). События должны быть редкие и случайные.

λ – ожидаемое среднее число событий

$$\mu = \sigma^2$$

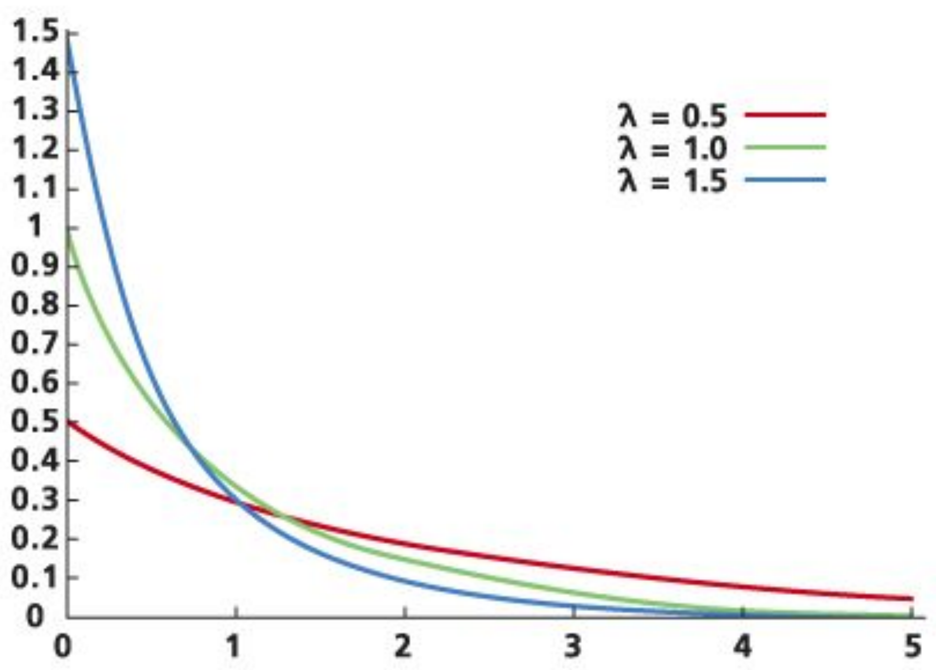
Siméon Denis Poisson



При больших n приближается к нормальному

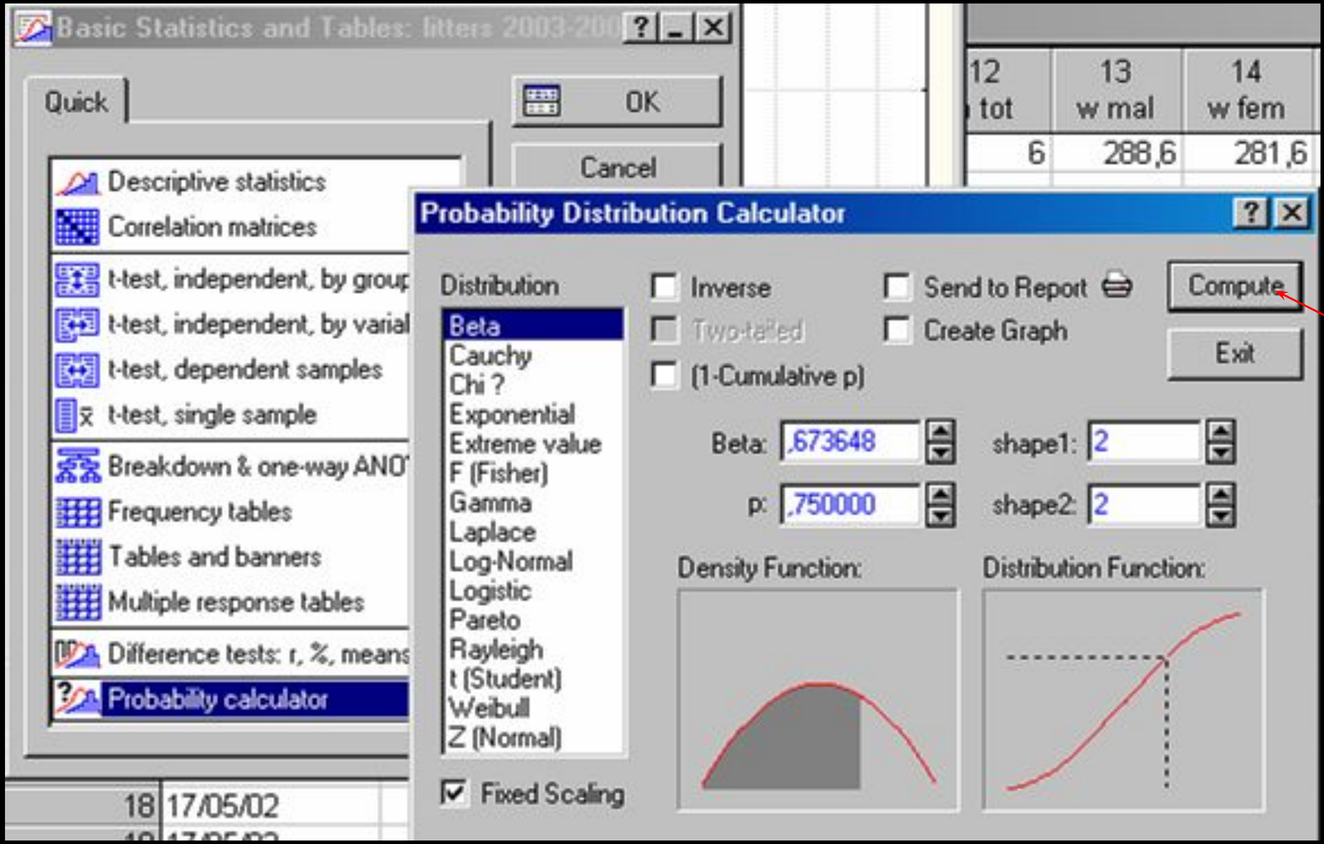
Трансформация данных
Экспоненциальное распределение

Хорошо описывает распределение промежутков времени (расстояний) между случайными событиями с заданной средней частотой событий.



Другие распределения

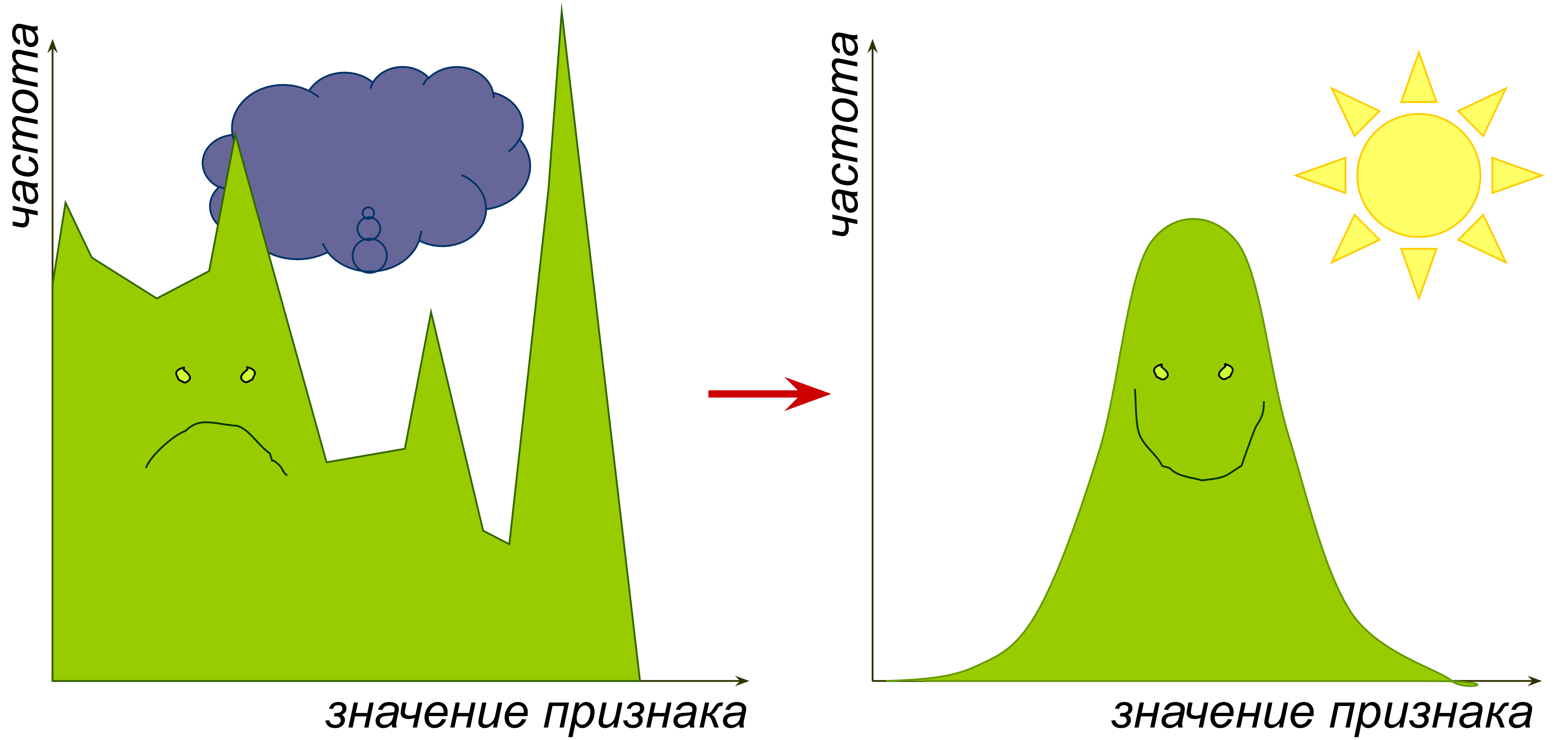
Логнормальное, Гамма, геометрическое, отрицательное биномиальное, гипергеометрическое и др.



Можно посчитать критические значения для любых распределений

Трансформация данных

Если распределение отлично от нормального (выборки не гомогенны, факторы мультипликативны), можно **ТРАНСФОРМИРОВАТЬ данные**

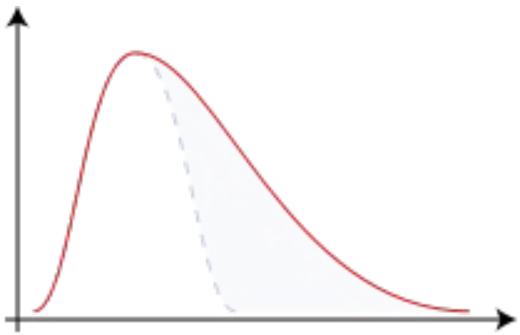


Трансформация данных

1. <u>Логарифмическая трансформация</u> *(logarithmic transformation):*

- Делает симметричным скошенное вправо (positively skewed) распределение.
- Используется в случае, когда чем больше среднее в группе, тем больше стандартное отклонение.

$$\boxed{X_i' = \lg X_i} \qquad \boxed{X_i' = \lg\left(X_i + 1\right)}$$



Если в результате логарифмирования получилось нормальное распределение, исходное распределение было **логнормальным.**

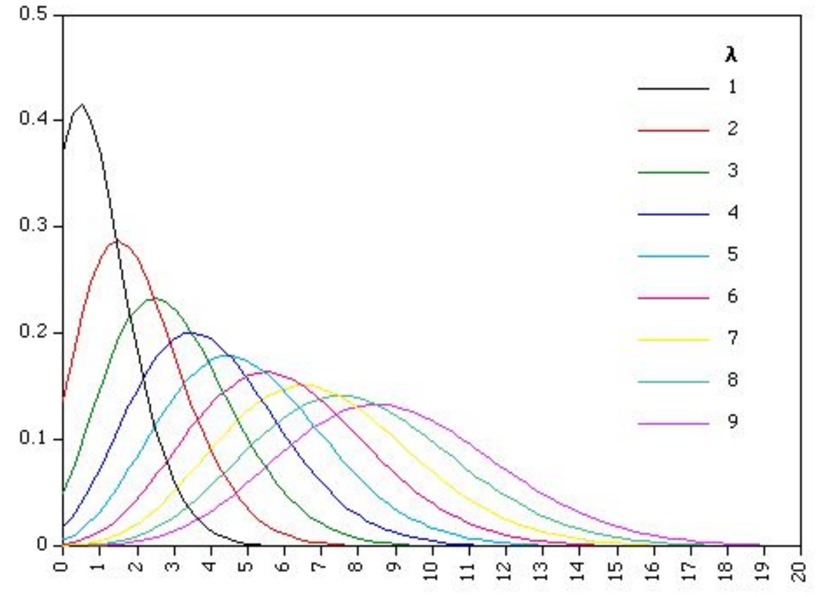
Трансформация данных

<u>2. Извлечение квадратного корня</u> *(square root transformation)*

- Используется, когда чем больше среднее в группе, тем больше дисперсия.
- обычно такое явление свойственно выборкам из распределения Пуассона (т.е., данные представляют собой количества случайных событий, объектов…)

$$X_i' = \sqrt{X_i}$$

$$X_i' = \sqrt{X_i + 0,5}$$



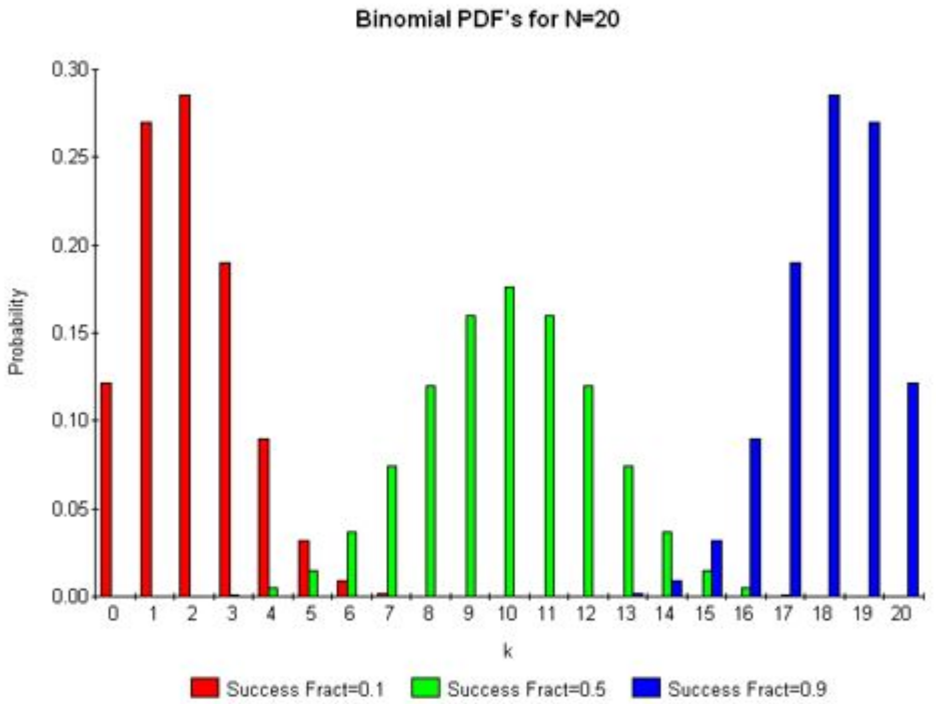
Например, количество социальных контактов в час.

Трансформация данных

3. Арксинусная трансформация (*arcsine transformation*)

- применяется для процентов и долей $(X_i \leq 1)$, которые обычно формируют биномиальное распределение.

$$X_i' = \arcsin\sqrt{X_i}$$



Например, мы исследуем долю самцов или долю переживших зиму детёнышей в выводках сурков.

Прочие трансформации см. Zar, 2010 (1999)

Непараметрические методы

Принципиально не годятся параметрические методы, если данные РАНГОВЫЕ:
мы не знаем, насколько одно значение отличается от другого.

Тут не спасёт никакая трансформация.

Если наше распределение не удовлетворяет условиям параметрических тестов и ни одна трансформация не помогает, наш выбор -

Непараметрические методы (nonparametric methods)

= “distribution-free” tests

- Свойства распределения неизвестны, и параметры распределения (среднее, дисперсию и т. п.) мы использовать не можем

- Основной подход – ранжирование (*ranking*) наблюдений (выстраиваем их по порядку от самого маленького значения к наибольшему).

- подразумевается, что сравниваемые распределения имеют одинаковую форму и дисперсию.

Непараметрические методы

Мы исследуем два редких вида сумчатых. Нам важно узнать, различаются ли виды по тому, какую освещённость местообитаний они предпочитают. Освещённость мы оценивали на глаз по 100-бальной шкале.

Фактор – вид. Группы: 1. длинноухие; 2. пятнистые

длинноухий

пятнистый

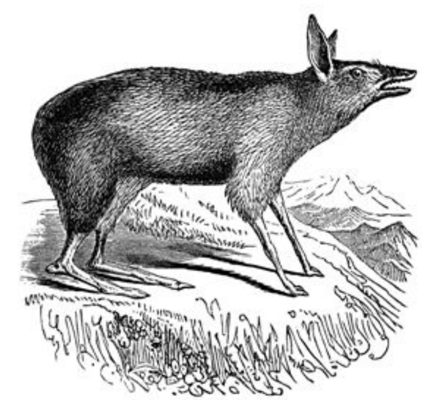
Непараметрические методы

Сравнение 2-х независимых групп:
Манн-Уитни тест (*Mann-Whitney U-test*)

H_0: распределение в популяции, из которой мы получили выборку длинноухих, такое же, как и в популяции, из которой выборка пятнистых.
H_1: распределения не одинаковые.

Мы ничего не говорим про параметры распределений!

Тест Манна-Уитни можно использовать и для ранговых, и для непрерывных переменных.



Непараметрические критерии

длинноухие		пятнистые	
свет	ранг	свет	ранг
8	15.5	4	5
7	13	7	13
4	5	5	8.5
7	13	8	15.5
9	17.5	3	2
3	2	3	2
5	8.5	5	8.5
6	11	4	5
9	17.5		
5	8.5		
	111.5		59.5

Это непараметрический аналог двухвыборочного t-теста.

Ранжируем данные от меньшего к большему (игнорируя деление на группы).

Число 3 встретилось трижды:
ранги у них будут одинаковы = (1+2+3)/3=2

Непараметрические методы

<u>Статистика критерия:</u>

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

n_1 и n_2 – размер выборок,
R_1 и R_2 – суммы рангов в выборках.

Статистикой критерия U_{obs} будет **меньшее** из этих двух значений. Причём H_0 мы отвергнем в случае, если оно будет <u>МЕНЬШЕ</u> критического значения U_{cv}. (т.е., это исключение среди прочих критериев).

Непараметрические критерии

Если размеры выборок больше 20, распределение статистики U приближается к **нормальному** со средним

$$\mu_U = \frac{n_1 n_2}{2}$$

Поэтому считается значение $$z_{obs} = \frac{U_{obs} - \mu_U}{\sigma_U}$$

И сравнивается с критическим значением для нормального распределения **Z** (наблюдаемое z должно быть по модулю больше критического).
Поэтому для маленьких выборок в статье можно приводить только U, а для больших выборок нужно приводить и U, и z.

Тест может быть <u>односторонним</u> и <u>двусторонним</u>

Непараметрические критерии

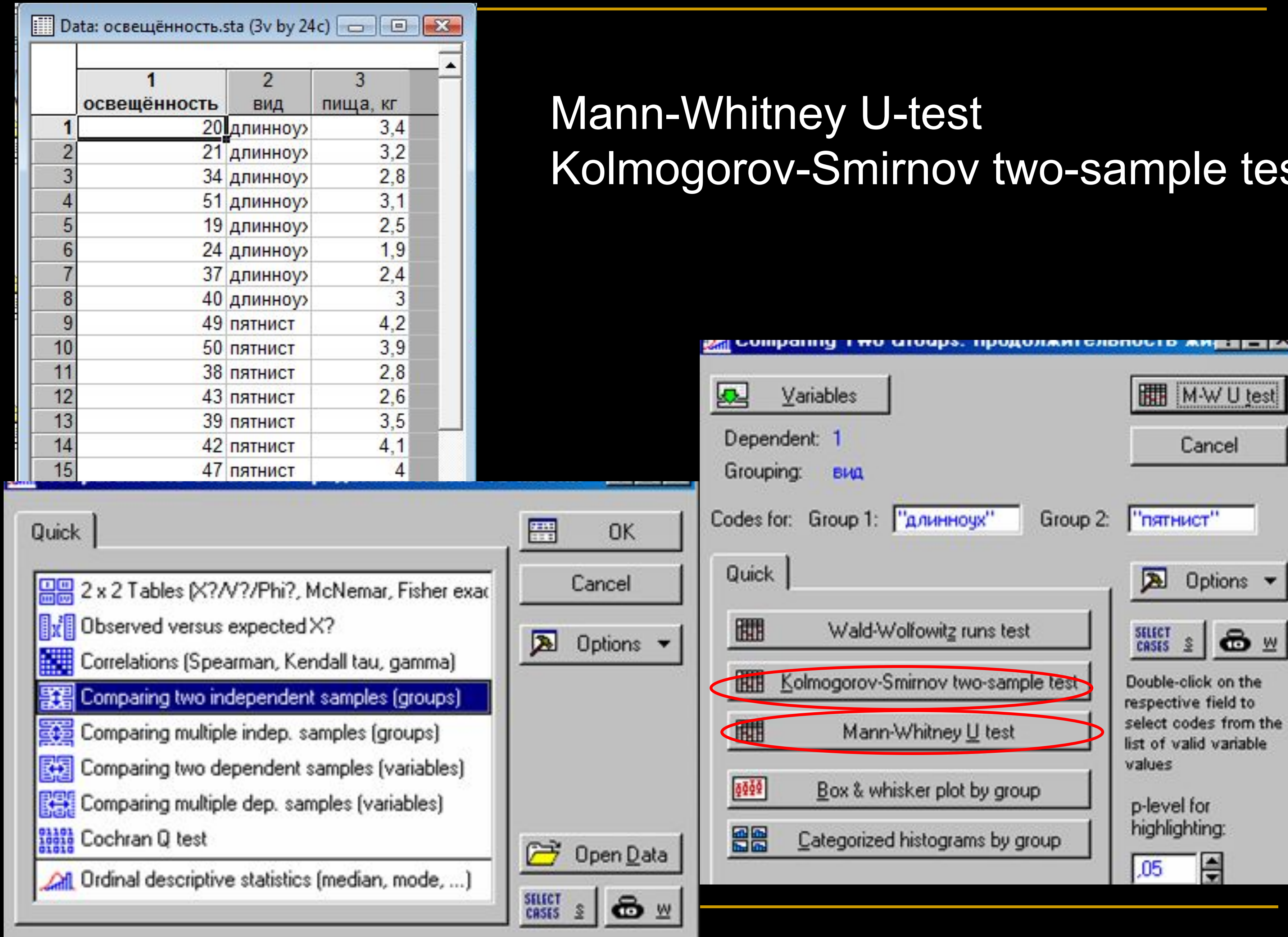
Сравнение 2-х независимых групп:

Тест Колмогорова-Смирнова (*Kolmogorov-Smirnov two-sample test*)

Отличается от теста Манн-Уитни тем, что М-У более чувствителен к различиям средних значений, медианы и т.п., а К-С тест более чувствителен к различиям распределений по форме.

Манн-Уитни тест более мощный.

Mann-Whitney U-test
Kolmogorov-Smirnov two-sample test



В обоих тестах отвергаем H_0: оба теста показали, что освещённость, в которой обитают звери разных видов неодинаковая

nn-Whitney U Test (освещённость)

Mann-Whitney U Test (освещённость)
By variable вид
Marked tests are significant at p <,05000

variable	Rank Sum длинноух	Rank Sum пятнист	U	Z	p-level	Z adjusted	p-level	Valid N длинноух	Valid N пятнист	2*1sided exact p
освещённость	46,00000	90,00000	10,00000	-2,31046	0,020863	-2,31046	0,020863	8	8	0,020668

ogorov-Smirnov Test (освещённость)

Kolmogorov-Smirnov Test (освещённость)
By variable вид
Marked tests are significant at p <,05000

variable	Max Neg Differnc	Max Pos Differnc	p-level	Mean длинноух	Mean пятнист	Std.Dev. длинноух	Std.Dev. пятнист	Valid N длинноух	Valid N пятнист
освещённость	-0,750000	0,125000	p < .025	30,75000	44,25000	11,58509	4,464143	8	8

Сравнение 2-х связанных групп

Критерий Вилкоксона (*Wilcoxon matched pair test*)

Изучаем утконосов, и хотим знать – различается ли отношение самки к самцу и самца к самке в парах

Мы считаем частоту дружелюбных контактов со стороны самки к самцу и наоборот. У *каждого самца* есть по *жене*, а у каждой *самки* – по *мужу*!

Непараметрические методы

H_0: распределение контактов в популяции, из которой мы получили выборку самцов, такое же, как и в популяции, из которой выборка самок.

H_1: распределения не одинаковые.

Фактор – пол. (1. самцы; 2. самки)

Непараметрические методы

	самец	самка
1 пара	356	363
2 пара	351	361
3 пара	353	358
4 пара	355	356
5 пара	354	359
6 пара	355	355

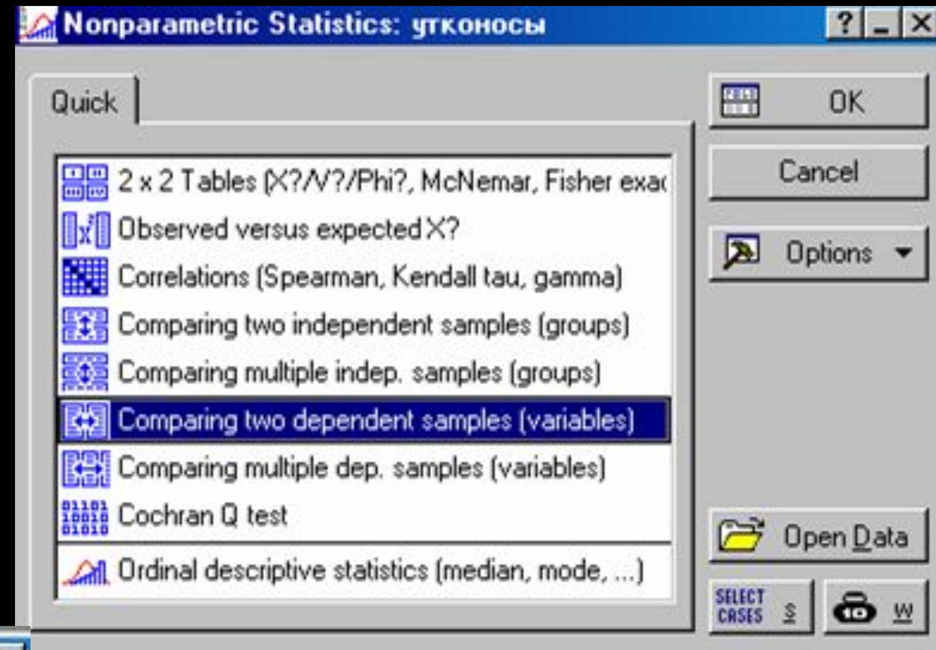
$$D_i = X_{i1} - X_{i2}$$

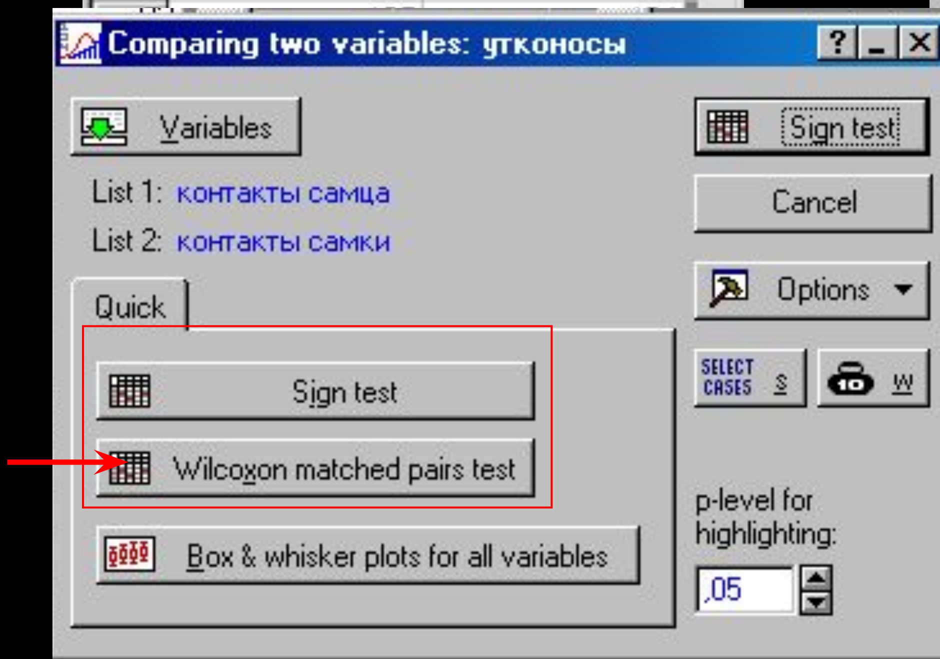
Предполагается, что распределение этих разностей симметрично относительно медианы

1. Считают <u>разности</u> между значениями в парах;
2. <u>исключают</u> нулевые <u>разности</u>;
3. присуждают абсолютным значениям (<u>по модулю</u>) разностей <u>ранги</u>;
4. <u>суммируют</u> отдельно **ранги** положительных и отрицательных разностей;
5. <u>Наименьшая</u> из этих сумм - статистика **T**.
6. Отвергаем H_0, если T меньше $T_{cv.}$

Аналог t-теста для двух связанных выборок. При числе пар >100 T апроксимируется нормальным распределением.

Wilcoxon matched pair test





Число дружелюбных контактов у самцов и самок в парах было неодинаковым

Непараметрические критерии

Сравнение 2-х связанных групп
Знаковый тест (*Sign test*)

Считают разности в парах, но не ранжируют их, а просто определяют число положительных и отрицательных разностей (нули исключают). Сравнивают их соотношение с 1:1. (биномиальным тестом)

Подходит для случаев, когда точные значения переменной не известны.

Имеет низкую мощность, поэтому применяется только в больших выборках (больше 20 пар).

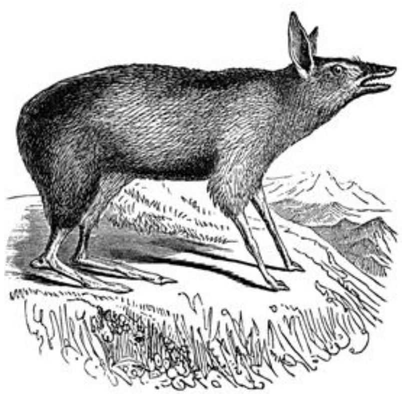
Непараметрические критерии

Сравнение ≥3-х независимых групп

Тест Крускала-Уоллиса (*Kruskal-Wallis test*)

Мы получили возможность включить в работу третий, особенно редкий вид сумчатого. Теперь нас интересует, различается ли доля растительной пищи, которую съедают за день особи этих видов.

Фактор – вид. Группы: 1. длинноухие; 2. пятнистые; 3. хвостатые



Непараметрические критерии

Критерий Крускал-Уоллиса (*Kruskal-Wallis test*)

- ✔ Непараметрический аналог One-way ANOVA
- ✔ на 95% настолько же мощный, как и ANOVA;
- ✔ для 2-х групп идентичен Манн-Уитни тесту;
- ✔ подразумевает сходство форм и дисперсий в распределениях (хотя бы на глаз)

Непараметрические критерии

1. все значения ранжируются от меньшего к большему (игнорируя деление на группы);
2. Считается сумма рангов в каждой группе;
3. считается статистика H(*df, N*).

H_0: распределение в популяциях, из которых мы получили выборки, одинаковое.
H_1: распределения не одинаковые.

сумма рангов в каждой группе

$$H = \frac{12}{N(N+1)} \sum \frac{R_j^2}{n_j} - 3(N+1)$$

размер группы

общий размер выборки

Непараметрические критерии

Критерий Крускал-Уоллиса (*Kruskal-Wallis test*)

При маленьких выборок и 3-5-и групп считается H-статистика.

Для больших выборок (или >5-и групп) H апроксимируется распределением χ^2.

Сравнение ≥2-х независимых групп

Медианный тест (*Median test*)

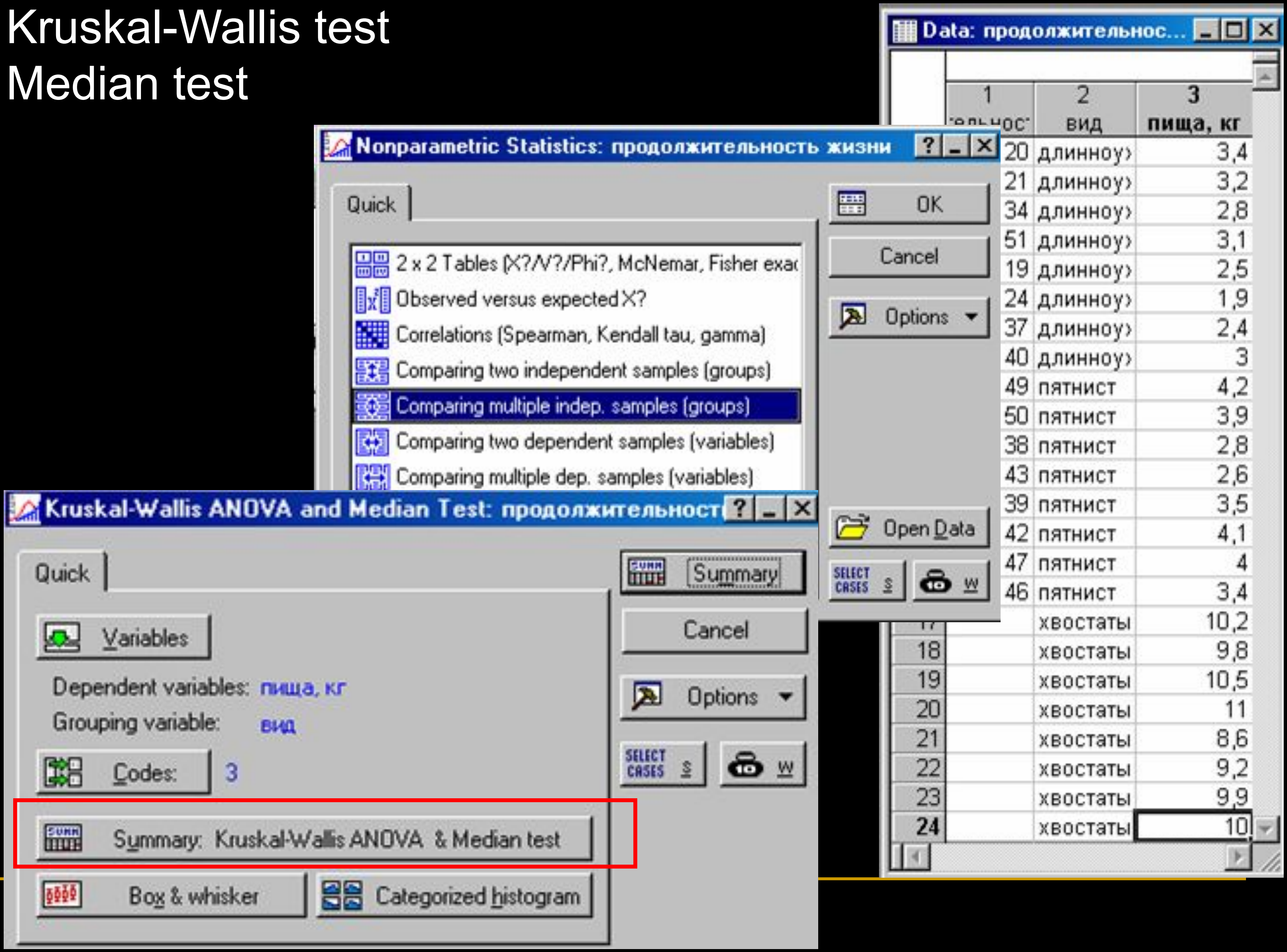
Считается **общая медиана** для всех групп (получается, что это не непараметрический тест, а distribution-free).

Затем критерием χ^2 (см. Частотные критерии) сравнивают числа значений, которые больше и которые меньше общей медианы в каждой из групп (табличка 2 x *k*).

Подходит для выборок, в которых часть наблюдений выходит за пределы шкалы (или их точные значения неизвестны).

Но имеет очень низкую мощность – лишь 67% мощности Манн-Уитни теста или теста Крускалла-Уоллеса.

Kruskal-Wallis test
Median test



Data: Kruskal-Wallis ANOVA by Ranks; пища, кг (продолжительность жи...

Kruskal-Wallis ANOVA by Ranks; пища, кг (продолжительность жизн
Independent (grouping) variable: вид
Kruskal-Wallis test: H (2, N= 24) =17,79547 p =,0001

Depend.: пища, кг	Code	Valid N	Sum of Ranks
длинноух	101	8	46,0000
пятнист	102	8	90,0000
хвостатые	103	8	164,0000

Data: Median Test, Overall Median = 3,70000; пища, кг (прод...

Median Test, Overall Median = 3,70000; пища,
Independent (grouping) variable: вид
Chi-Square = 16,00000, df = 2, p = ,0003

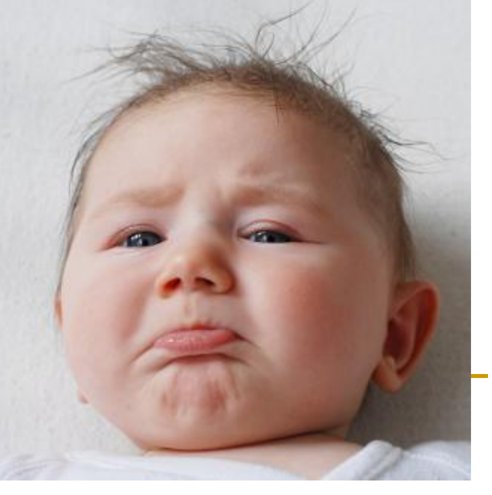
Dependent: пища, кг	длинноух	пятнист	хвостатые	Total
<= Median: observed	8,00000	4,000000	0,00000	12,00000
expected	4,00000	4,000000	4,00000	
obs.-exp.	4,00000	0,000000	-4,00000	
> Median: observed	0,00000	4,000000	8,00000	12,00000
expected	4,00000	4,000000	4,00000	
obs.-exp.	-4,00000	0,000000	4,00000	
Total: observed	8,00000	8,000000	8,00000	24,00000

Доля растительной пищи отличалась между разными видами

Критерий Крускал-Уоллиса (*Kruskal-Wallis test*)

Хотелось бы провести после сравнения нескольких групп **пост-хок тест** (апостериорное сравнение), по аналогии с тестом Тьюки.

Такие тесты существуют – Nemenyi test, Dunn’s test (Zar, 1999 или 2010).
Только в Statistica их нет, поэтому можно их считать вручную, либо задавать формулу в Statistica и какой-л. другой программе



Непараметрические критерии

Сравнение ≥**3** связанных групп

Критерий Фридмана (*Friedman ANOVA*)

У утконосов родились детёныши, и мы хотим знать, изменялось ли физическое состояние самок после беременности и после выкармливания потомства (мы оценивали его в баллах по упитанности и состоянию шерсти).

1. состояние до беременности;
2. после рождения детей;
3. после выкармливания детёнышей

Непараметрические критерии

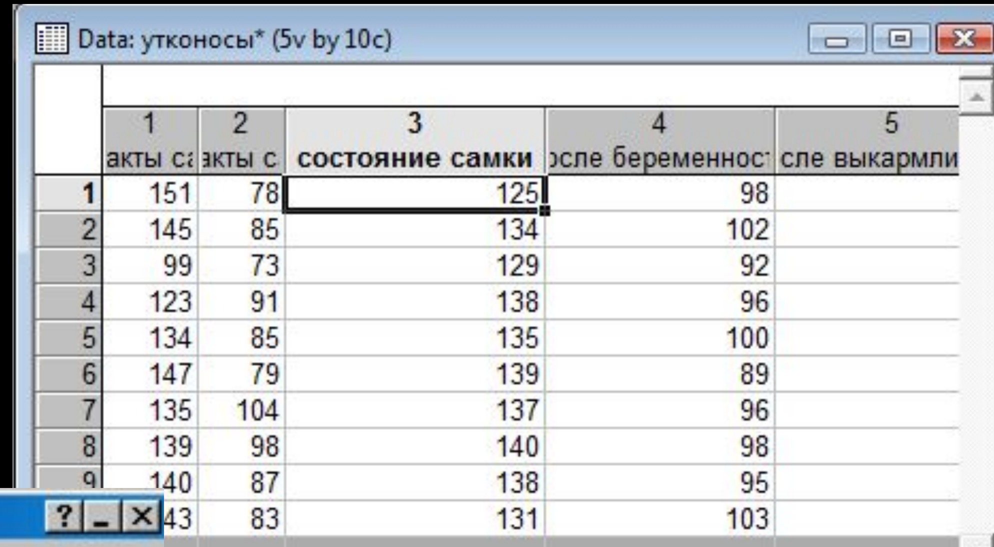
Критерий Фридмана (*Friedman ANOVA*)

- ✔ для двух групп эквивалентен Знаковому тесту (sign test);
- ✔ по сравнению с аналогичными параметрическими тестами, для 2-х групп имеет всего 64% мощности, для 3-х – 72%, для 100 стремится к 95%.

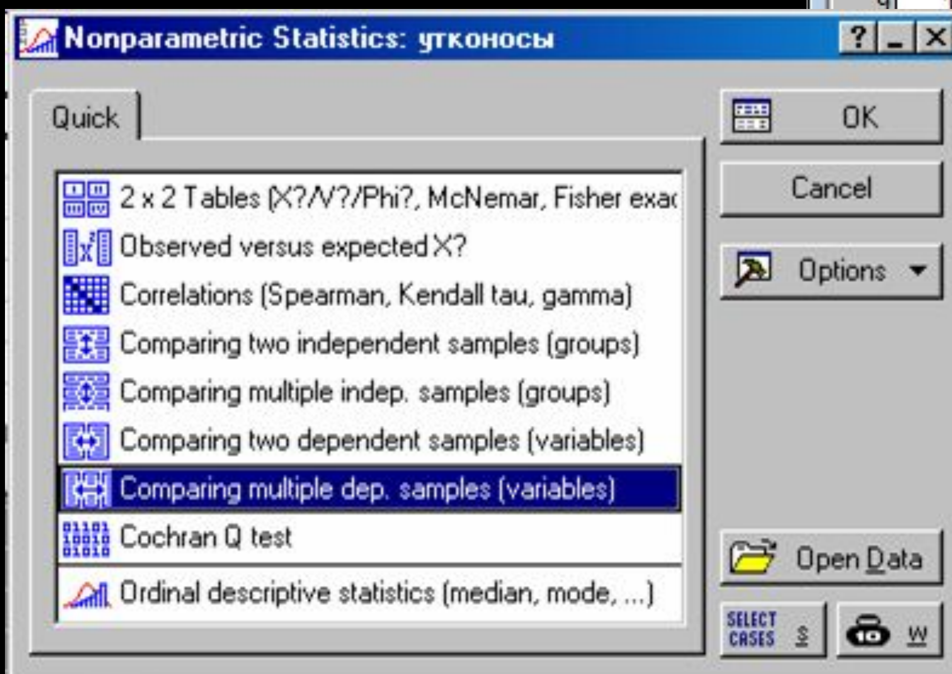
Основан на том, что значения ранжируются меньшего к большему внутри каждой **строки**. Потом суммируют ранги для каждого <u>столбца</u> и считают статистику χ^2_r, которая имеет распределение χ^2.

Нулевая и альтернативная гипотезы - по аналогии с предыдущими тестами, о сходстве выборок.

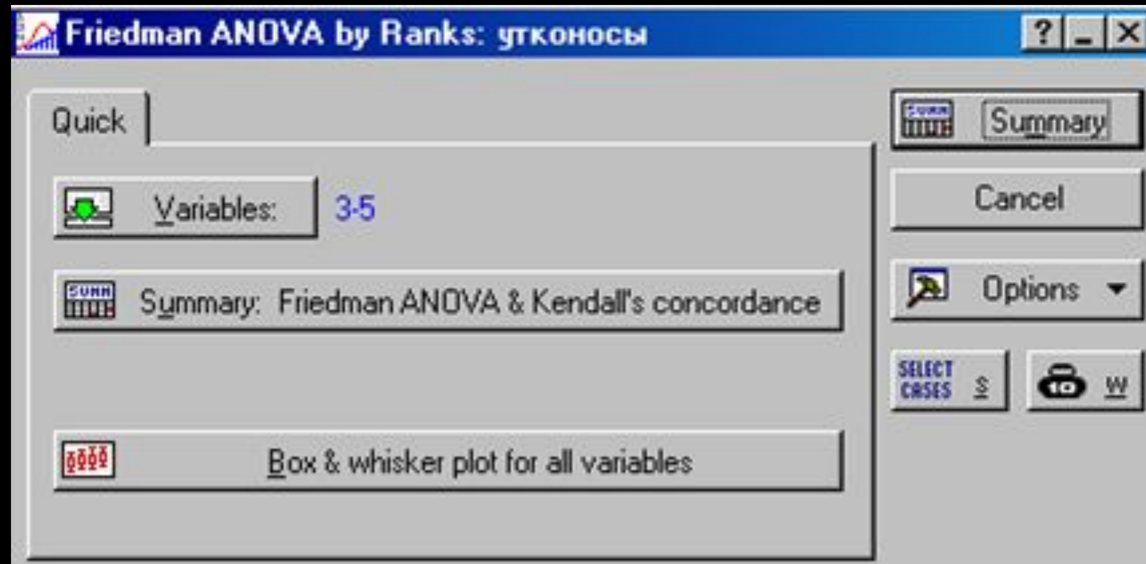
Friedman ANOVA

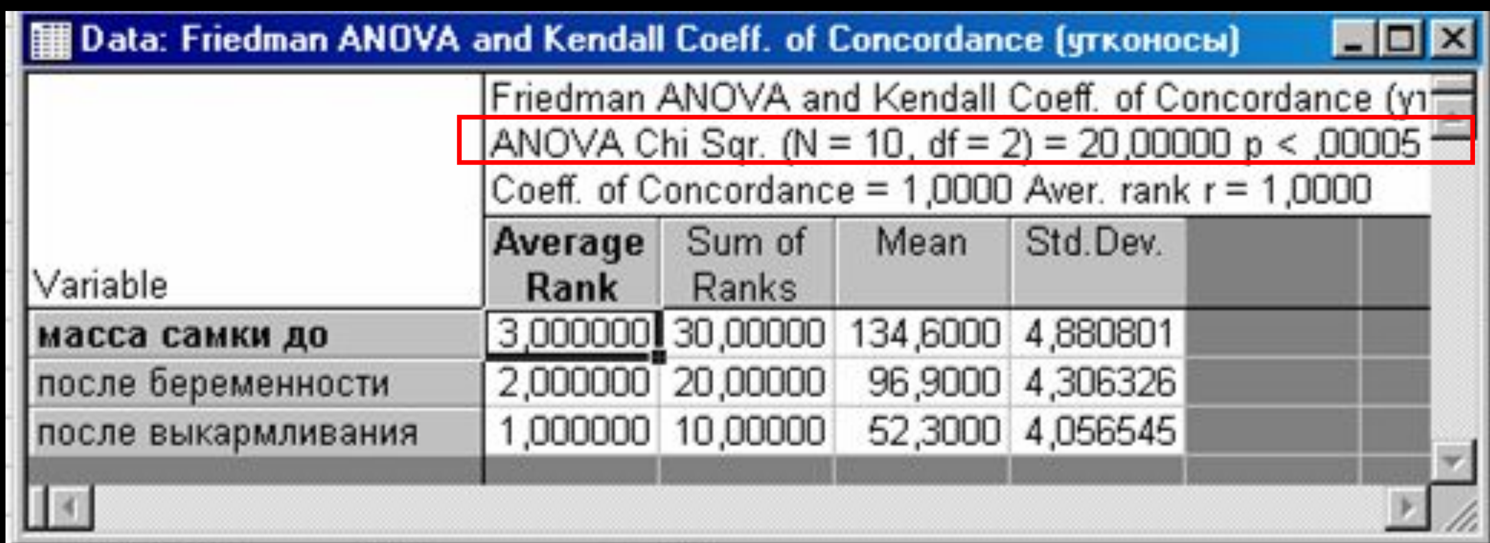


	1 акты са	2 акты с	3 состояние самки	4 осле беременност	5 сле выкармли
1	151	78	125	98	
2	145	85	134	102	
3	99	73	129	92	
4	123	91	138	96	
5	134	85	135	100	
6	147	79	139	89	
7	135	104	137	96	
8	139	98	140	98	
9	140	87	138	95	
	43	83	131	103	



Отвергаем H_0 – состояние самок изменялось




Data: Friedman ANOVA and Kendall Coeff. of Concordance (утконосы)

Friedman ANOVA and Kendall Coeff. of Concordance (ут
ANOVA Chi Sqr. (N = 10, df = 2) = 20,00000 p < ,00005
Coeff. of Concordance = 1,0000 Aver. rank r = 1,0000

Variable	Average Rank	Sum of Ranks	Mean	Std.Dev.
масса самки до	3,000000	30,00000	134,6000	4,880801
после беременности	2,000000	20,00000	96,9000	4,306326
после выкармливания	1,000000	10,00000	52,3000	4,056545

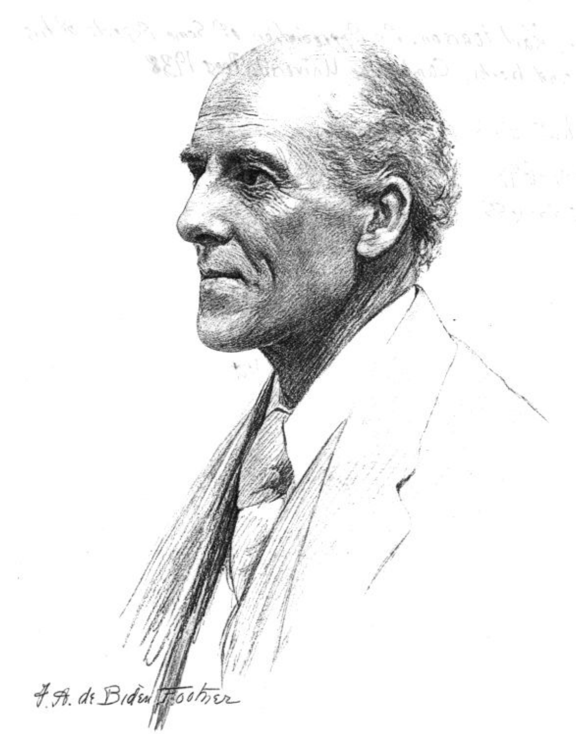
Итак, при выборе теста важно, что:

1. Параметрические тесты более <u>мощные,</u> чем непараметрические;
2. Непараметрические <u>безопаснее</u> в плане ошибки 1-го рода;
3. Чем больше <u>размер</u> выборки, тем <u>менее критичны</u> требования к распределению (по Центральной предельной теореме); для выборок N **≥ 100** используют параметрические тесты даже при больших отклонениях от нормального распределения.
4. АНОВА не очень чувствительна к отклонениям от нормального распределения (для одинаковых по размеру групп).

Частотные критерии

У нас есть выборка. Данные – качественные.
Вопрос: соответствует ли распределение в популяции, из которой получена выборка, теоретическому распределению? (которое мы сами определяем).

Ответ дадут **КРИТЕРИИ СОГЛАСИЯ** (*Tests for goodness of fit*).



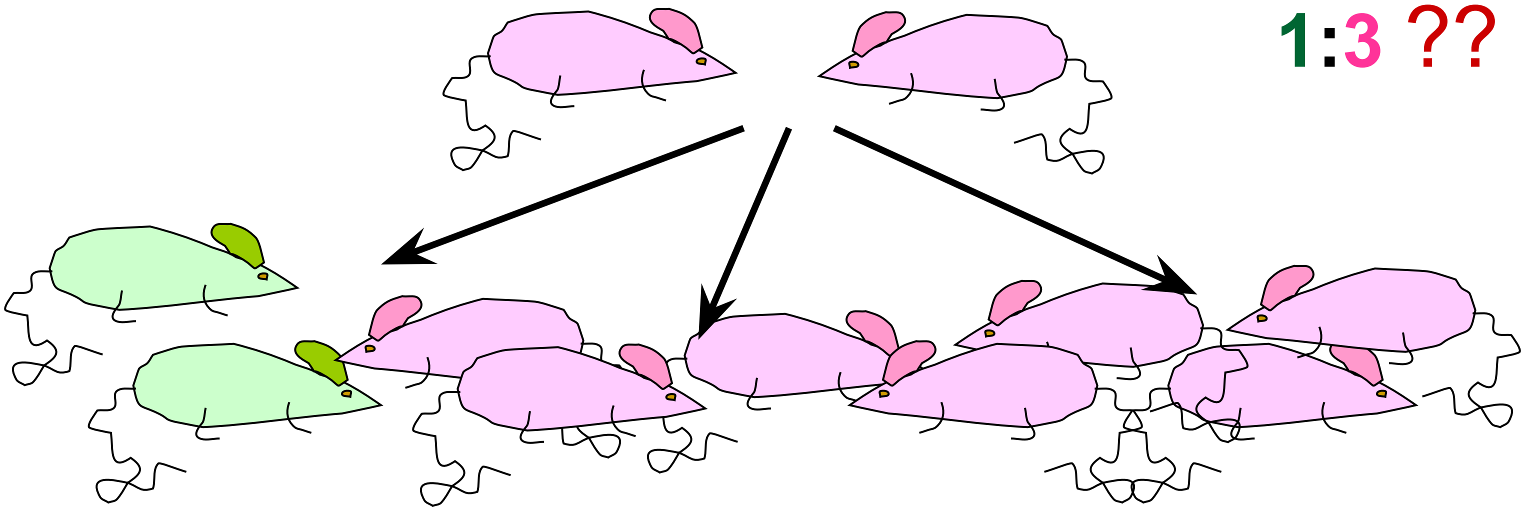
Придумал χ^2 статистику ещё в 1900 году!

Пример с игральной костью: как проверить, не кривая ли она? Очевидно, что бросая её 120 раз маловероятно получить ровно по 20 бросков на каждую сторону. Насколько же допустимы различия?

Критерии согласия

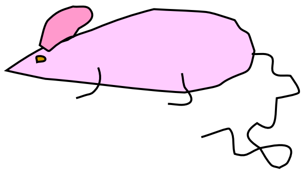
Родились:
84 розовых мыши и 16 зелёных.

H_0: выборка получена из популяции, где соотношение розовых и зелёных – 3:1.
H_1: выборка получена из популяции, где соотношение розовых и зелёных не равно 3:1



Заметим, что речь идёт только о частотах, но не о параметрах распределения.

Критерии согласия



	розовые	зелёные	всего
O_i	84	16	100
E_i	75	25	

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(84-75)^2}{75} + \frac{(16-25)^2}{25} =$$

$$= 1.080 + 3.240 = 4.320$$

$\boxed{df = k-1=1}$

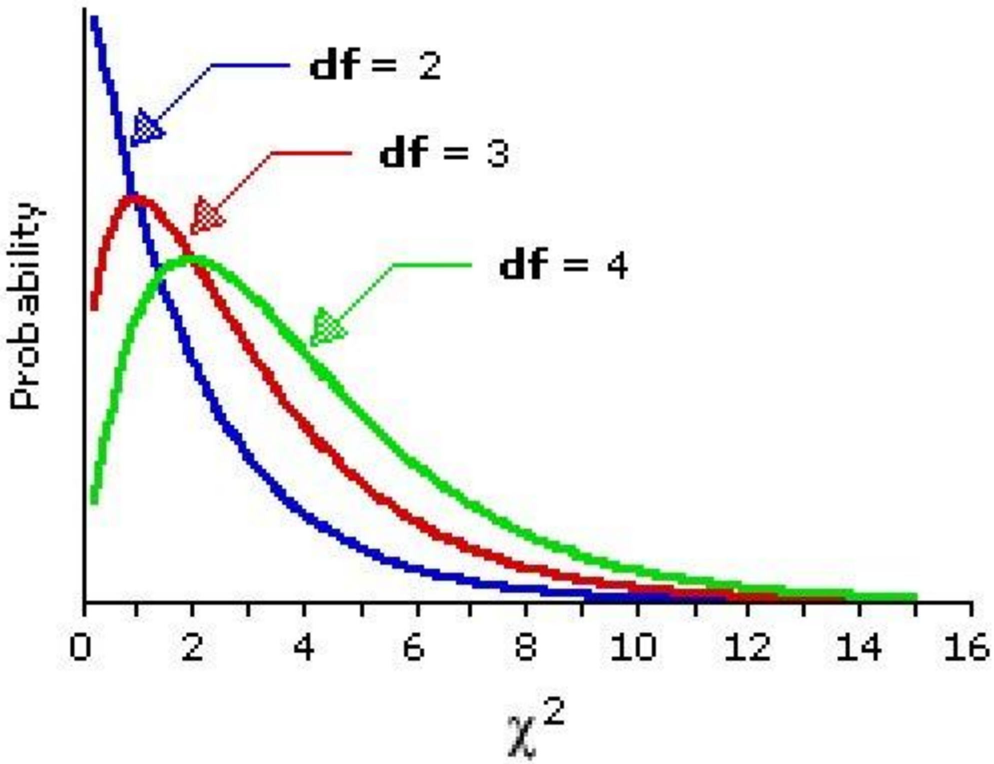
$\chi^2_{cv} = 3.841 < 4.320$

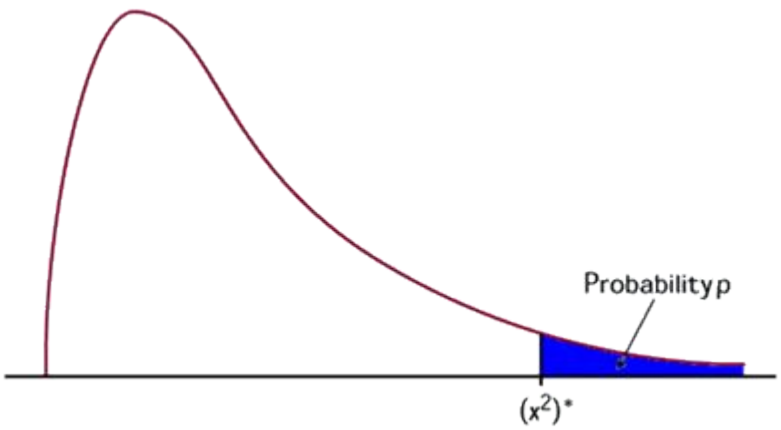
$p=0.038$

Чем больше значение χ^2,тем хуже наши данные соответствуют теоретическому распределению – тем меньше *p*

H$_0$ отвергаем – соотношение мышей не соответствует ожидаемому

Критерии согласия





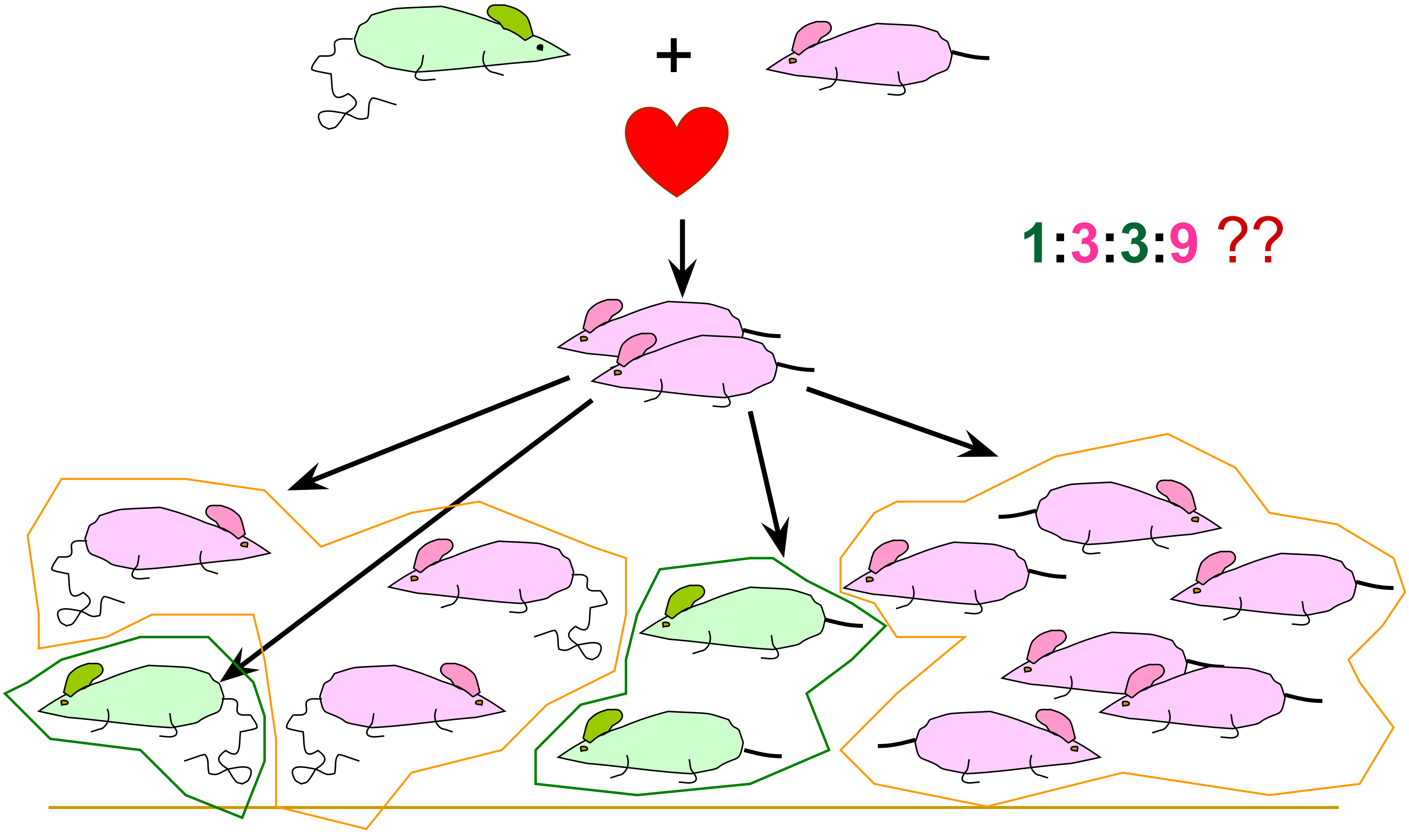
Критерии согласия

Категорий может быть сколько угодно.

Родились:
152 розовых мыши с острым хвостом; 39 розовых с курчавым хвостом; 53 зелёных с острым, 6 зелёных с курчавым.

H_0: выборка получена из популяции, где соотношение фенотипов – 9:3:3:1.
H_1: выборка получена из популяции, где соотношение фенотипов не равно 9:3:3:1

Критерии согласия



Критерии согласия

Важное замечание:

В всех критериях согласия H_0 гипотеза – о том, что форма распределений ОДИНАКОВА.

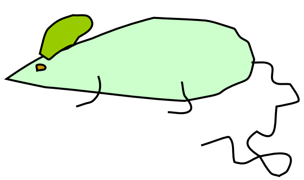
То есть, когда мы ищем подтверждение тому, что наши данные удовлетворяют некоторому распределению, мы должны радоваться, получив *p*>>0.05!

Критерии согласия

Zar, 1999:

Если мы сравнили распределение с теоретическим, <u>получили отличия (!)</u>, а теперь хотим показать, из-за какой именно категории эти отличия возникли, можно отдельно сравнить с теоретическим распределением остальные категории, а затем – отношение этой категории к остальным.

Т.е., если нам кажется, что всё портят зелёные мыши с курчавыми хвостами, сравним:



1. соотношение остальных мышей с 9:3:3;
2. отношение зелёных-курчавых к остальным с 1:15.

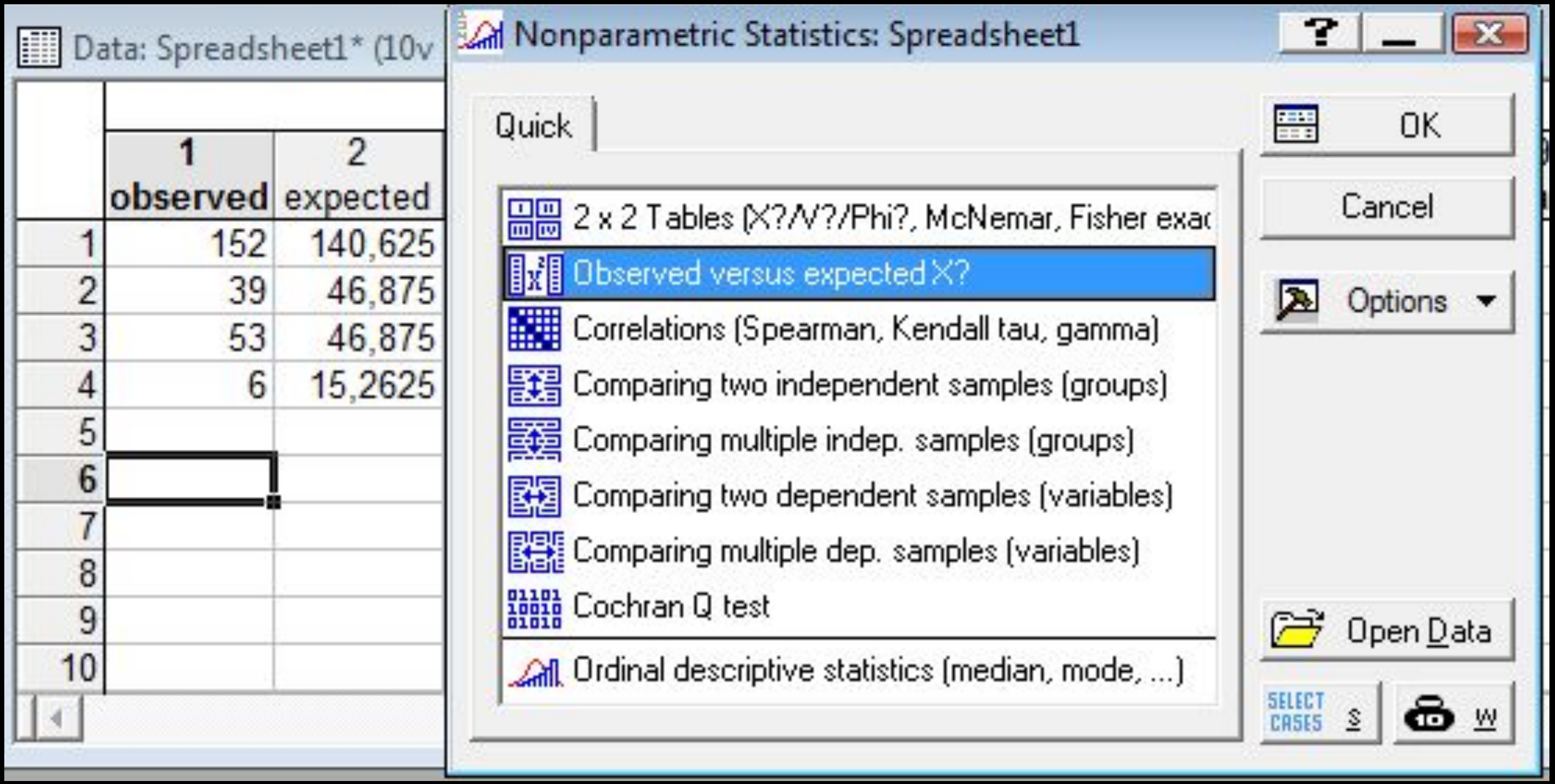
Критерии согласия

Итак:

1. у нас **одна выборка**
2. Переменная **качественная**
3. мы сравниваем наблюдаемые частоты с ожидаемыми (**observed and expected**)

Критерий χ^2 Пирсона (**Pearson Chi-square test**)

Сравнение ***нашего распределения с теоретическим*** (нужна таблица с посчитанными частотами)



результаты

Data: Observed vs. Expected Frequencies (Spreadsheet1)

Observed vs. Expected Frequencies (Spreadsheet1)

Chi-Square = 8,664667 df = 3 p < ,034100

NOTE: Unequal sums of obs. & exp. frequencies

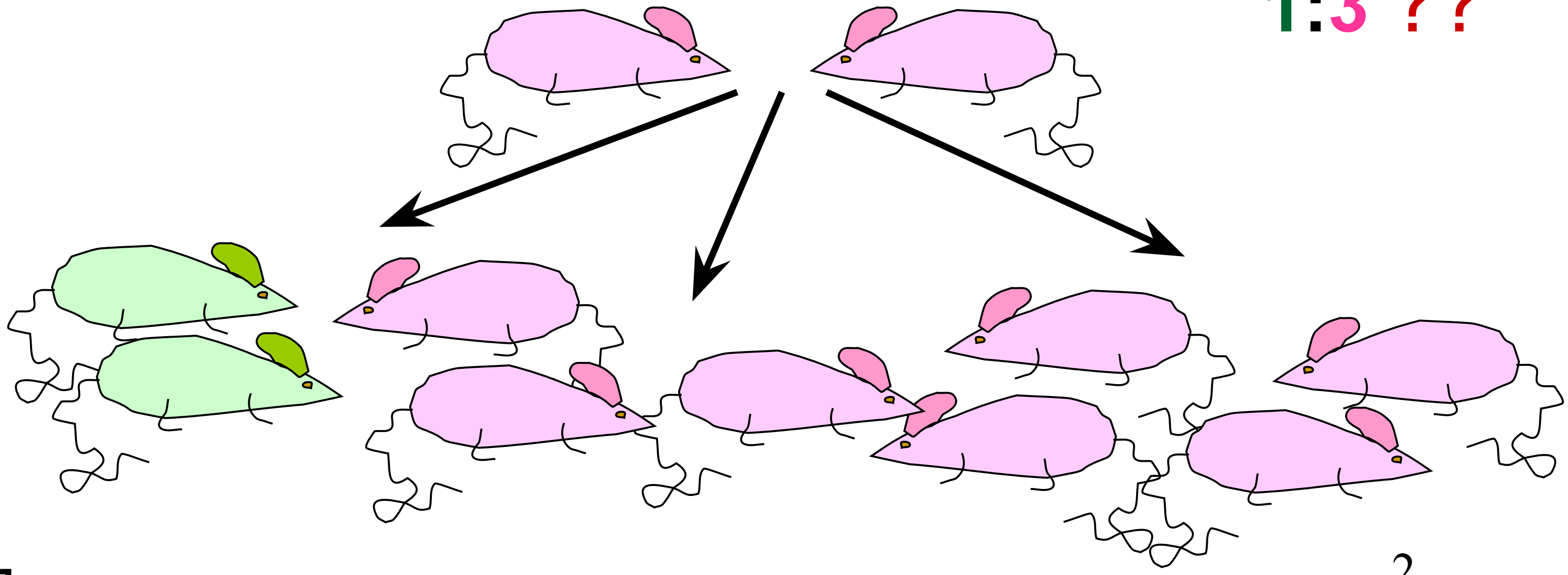
Case	observed observed	expected expected	O - E	(O-E)**2 /E
C: 1	152,0000	140,6250	11,37500	0,920111
C: 2	39,0000	46,8750	-7,87500	1,323000
C: 3	53,0000	46,8750	6,12500	0,800333
C: 4	6,0000	15,2625	-9,26250	5,621222
Sum	250,0000	249,6375	0,36250	8,664667

Критерии согласия

Если у нас только 2 проявления признака

Поправка Йейтса для критерия χ^2 (Yates correction for continuity)

1:**3** ??



Для заданного теоретического распределения χ^2 может принимать только строго определённые значения для разных наблюдаемых распределений.

Критерии согласия

Например: если ожидаемые частоты – 75 и 25, то значения χ^2 будут

для 84 и 16 – 4.32,
для 83 и 17 – 3.14,
для 82 и 18 – 2.61

промежуточных значений не может быть для данных ожидаемых частот

Но χ^2 распределение непрерывное. И для заданного уровня значимости *p* мы не найдём точно соответствующего ему значения χ^2.

χ^2 с поправкой Йейтса:

(для больших N не нужен)

Делает тест более консервативным.

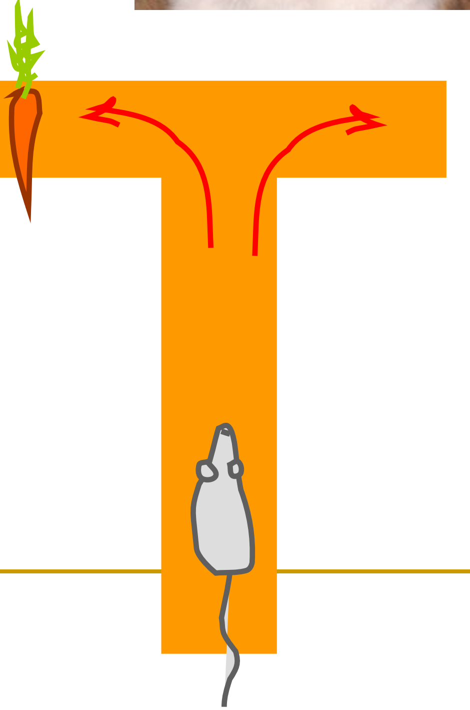
$$\chi^2 = \sum_{i=1}^{k} \frac{\left(\left|O_i - E_i\right| - 0.5\right)^2}{E_i}$$

Биномиальный тест

Элементарный тест для сравнения двух частот с теоретическими (для маленьких выборок, легко считать вручную). Большие выборки – задача для теста χ^2.

Пример с котом Гусом: у нас есть подозрение, что он правша. Мы дали ему игрушку на резинке, он ударил по ней 10 раз: 8 - правой, 2 – левой. Справедливо ли наше подозрение?

Пример с Т-образным лабиринтом: 10 мышей пошли налево, 3 – направо.



Источники:
Zar, 2010 (1999).
http://udel.edu/~mcdonald/statexactbin.htm

Критерии согласия

Замечательный тест Колмогорова-Смирнова для ранговых данных (Kolmogorov-Smirnov goodness of fit for discrete ordinal scale data).

35 кошек выбирают из 5 типов корма, различающихся по влажности. Случаен ли выбор или есть предпочтения?

То есть, 5 типов корма можно <u>проранжировать</u> от самого влажного к самому сухому, это не просто качественные признаки. Мощность такого теста выше, чем χ^2, но его нет в Staristica.

Zar, 2010 (1999).

Критерии согласия

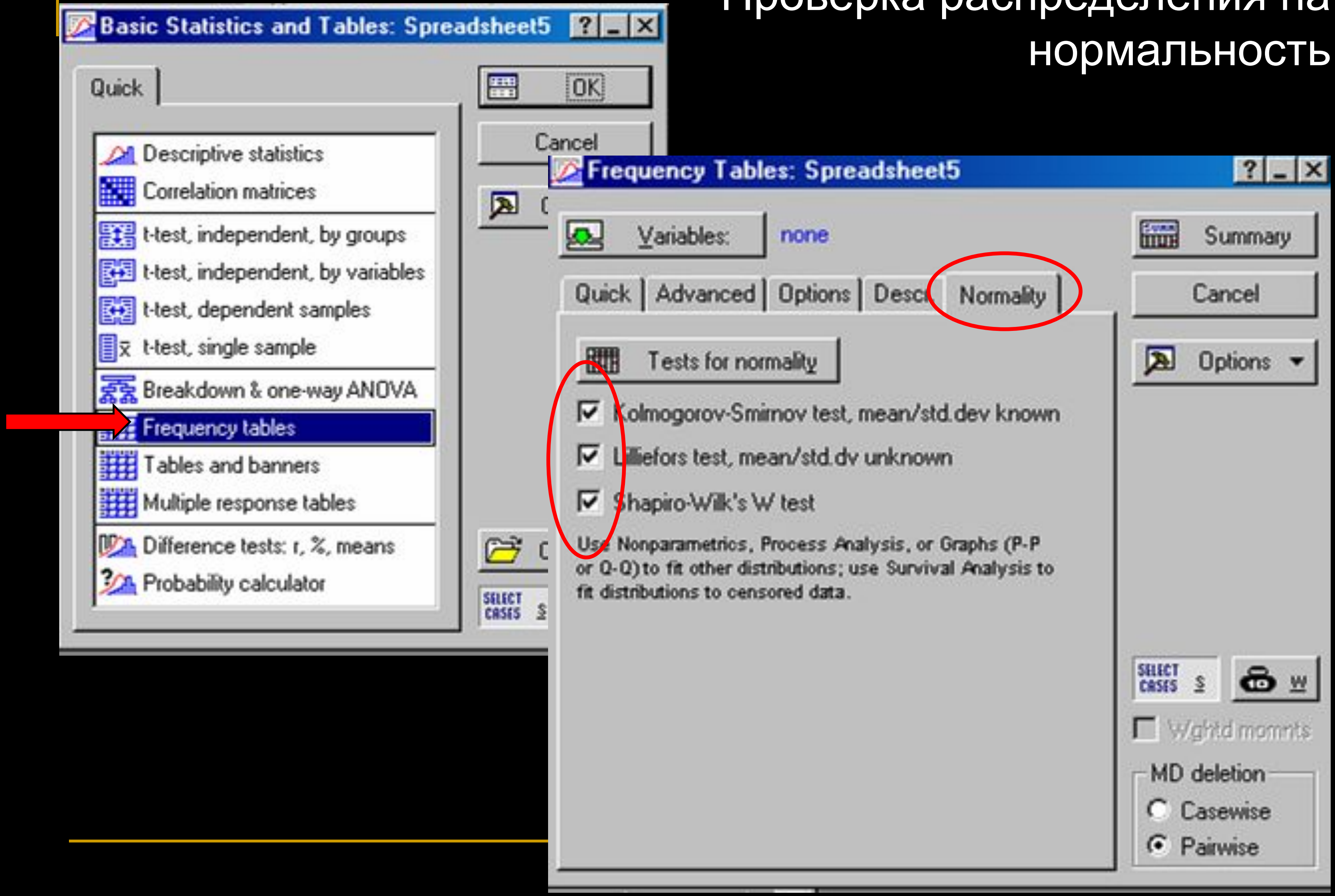
Тесты на соответствие непрерывным распределениям

Соответствует ли распределение мотыльков на дереве **НОРМАЛЬНОМУ РАСПРЕДЕЛЕНИЮ**?

Переменная – высота от земли в метрах

- ✔ Тест Колмогорова-Смирнова (Kolmogorov-Smirnov test) (если известны дисперсия и среднее в популяции) D-статистика.
- ✔ Lilliefors test – если НЕизвестны дисперсия и среднее в популяции – «улучшенный К-С тест»
- ✔ **Shapiro-Wilk's W test** (самый мощный, размер выборки до 5000) – наиболее предпочтительный.

Проверка распределения на нормальность



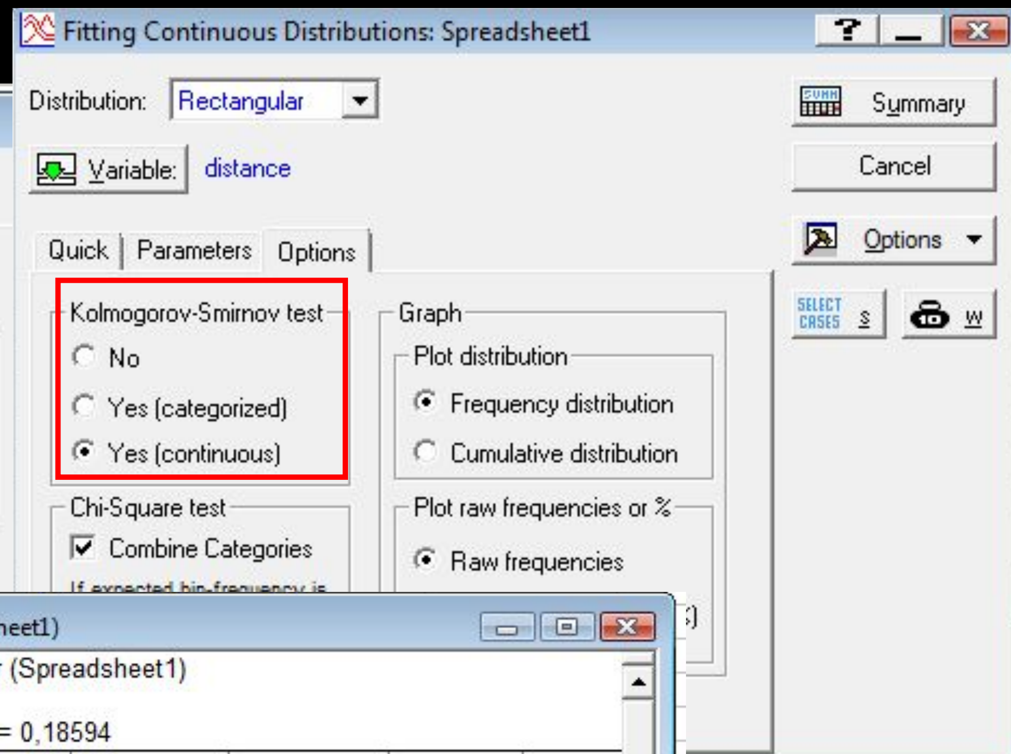
Data: Tests of Normality (зубы)

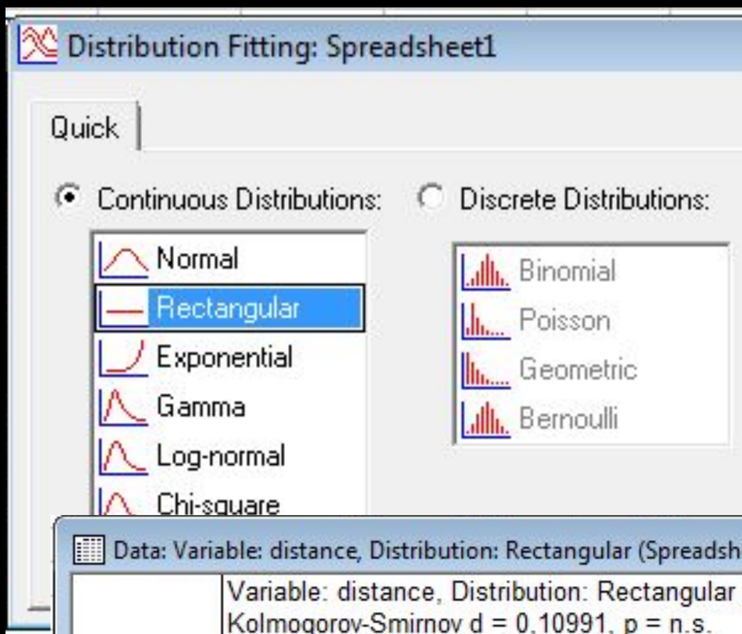
Variable	Tests of Normality (зубы)					
	N	max D	K-S p	Lilliefors p	W	p
weight	186	0,074337	p > .20	p < ,05	0,952717	0,000007

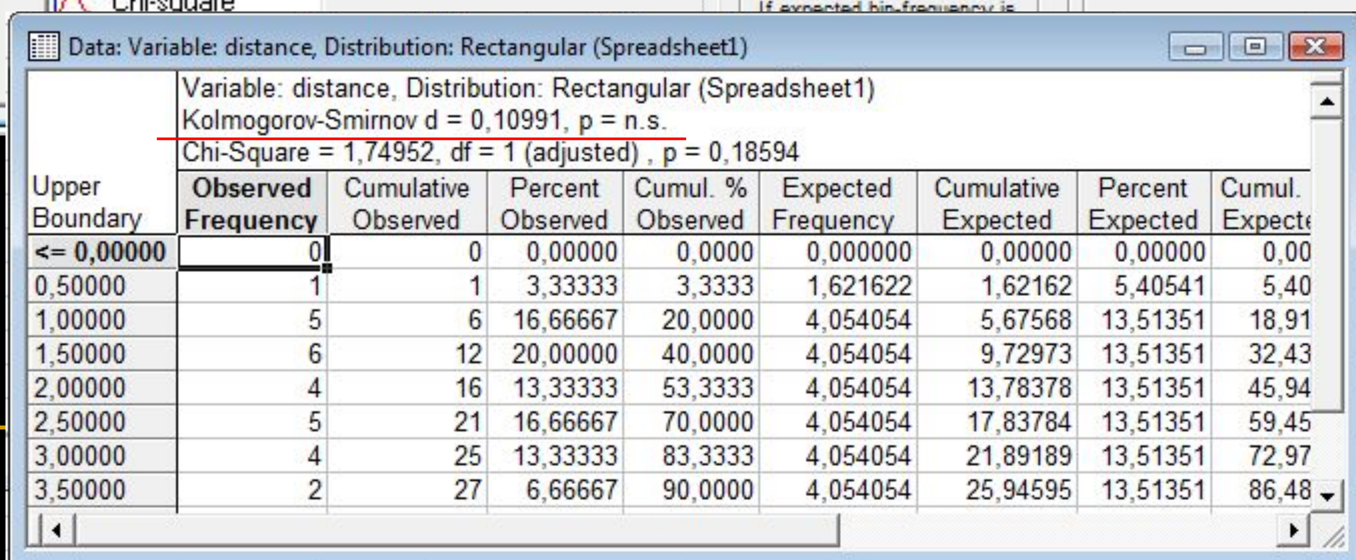
маленькое *p* говорит о том, что данные не соответствуют нормальному распределению.

Сравнение с другими теоретическими распределениями:

Тест Колмогорова-Смирнова для непрерывных распределений







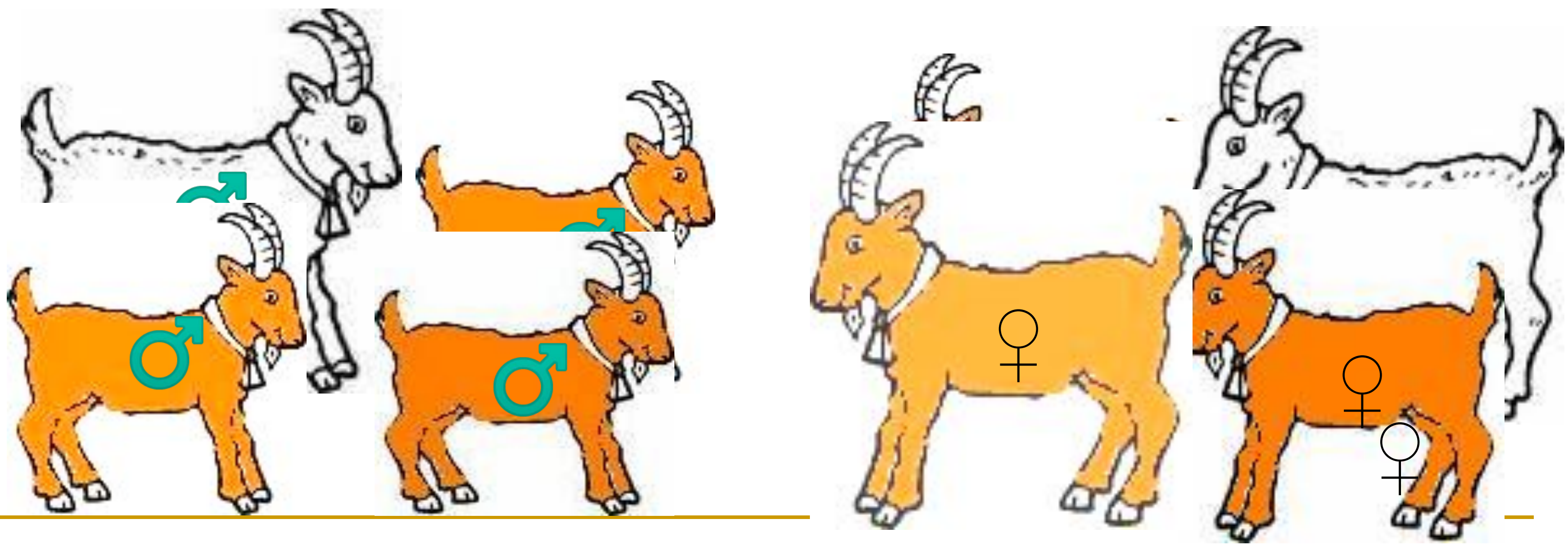
Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. Expecte
<= 0,00000	0	0	0,00000	0,0000	0,000000	0,00000	0,00000	0,00
0,50000	1	1	3,33333	3,3333	1,621622	1,62162	5,40541	5,40
1,00000	5	6	16,66667	20,0000	4,054054	5,67568	13,51351	18,91
1,50000	6	12	20,00000	40,0000	4,054054	9,72973	13,51351	32,43
2,00000	4	16	13,33333	53,3333	4,054054	13,78378	13,51351	45,94
2,50000	5	21	16,66667	70,0000	4,054054	17,83784	13,51351	59,45
3,00000	4	25	13,33333	83,3333	4,054054	21,89189	13,51351	72,97
3,50000	2	27	6,66667	90,0000	4,054054	25,94595	13,51351	86,48

Частотный анализ

Сравниваем **независимые** выборки, причём все переменные (≥2) **категориальные.**

Tests of independence – проверяют, зависит ли форма распределения одной переменной от значений другой переменной (переменных).

Критерий χ^2 (χ^2 analysis of contingency tables = χ^2 test of independence)



Связаны ли пол и цвет у коз?

Частотный анализ

пол	белые	красные	жёлтые	серые	Всего
самцы самки	32 55	43 65	16 64	9 16	100 200
всего	87	108	80	25	300

Таблицы вида a × b. Общая H_0 гипотеза: частоты в строчках не зависят от частот в столбцах.

H_0: цвет меха не зависит от пола в популяции коз;
H_1: цвет меха зависит от пола в популяции коз.

Как и в корреляции, здесь не идёт речь о причинно-следственной связи, табличку всегда можно перевернуть.

<u>Пример из жизни сусликов:</u>
Связаны ли категории социальных контактов (как контактирует) с полом партнёра?

Частотный анализ

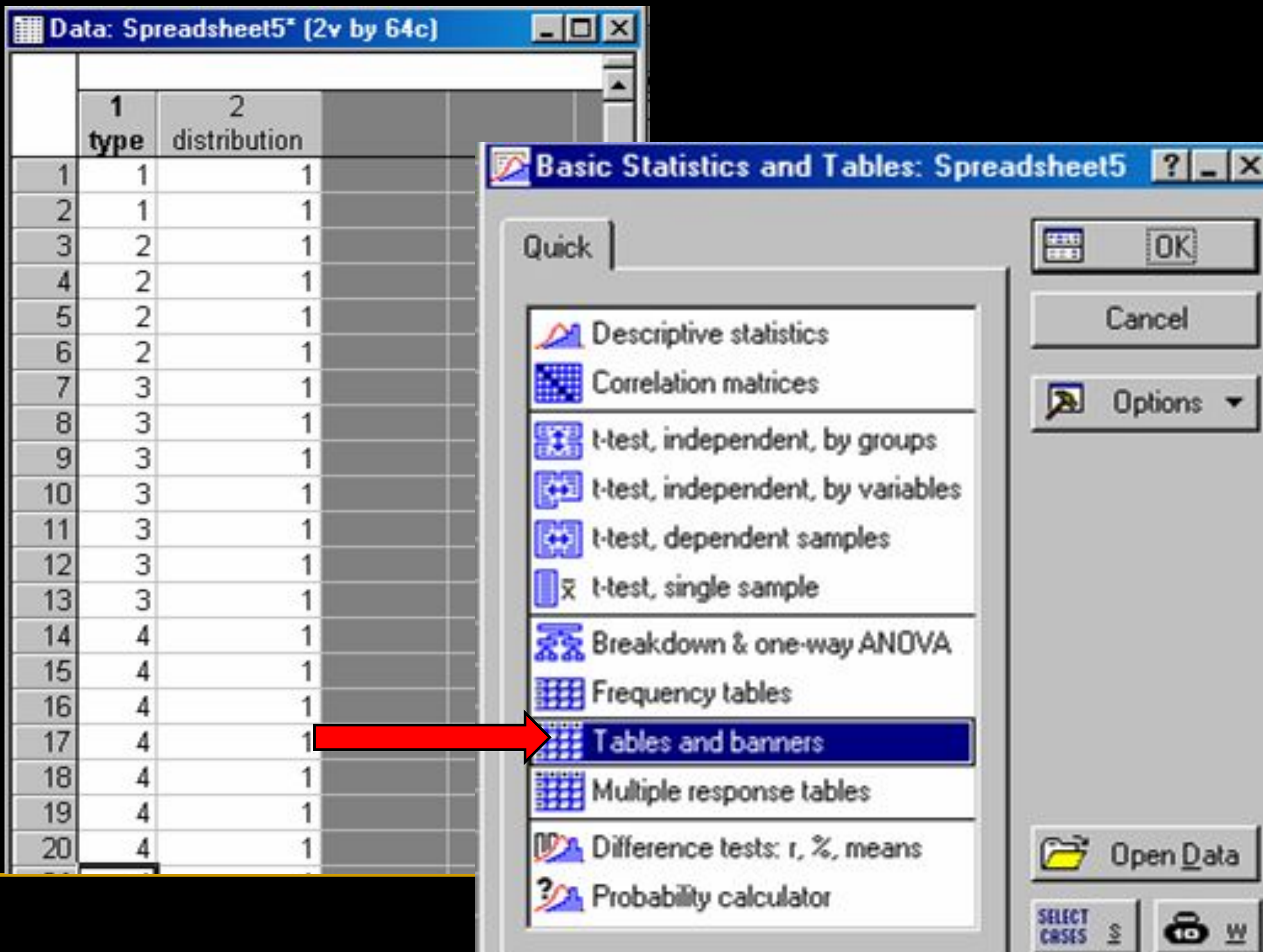
пол	белые	красные	жёлтые	серые	Всего
самцы	32	43	16	9	100
самки	55	65	64	16	200
всего	87	108	80	25	300

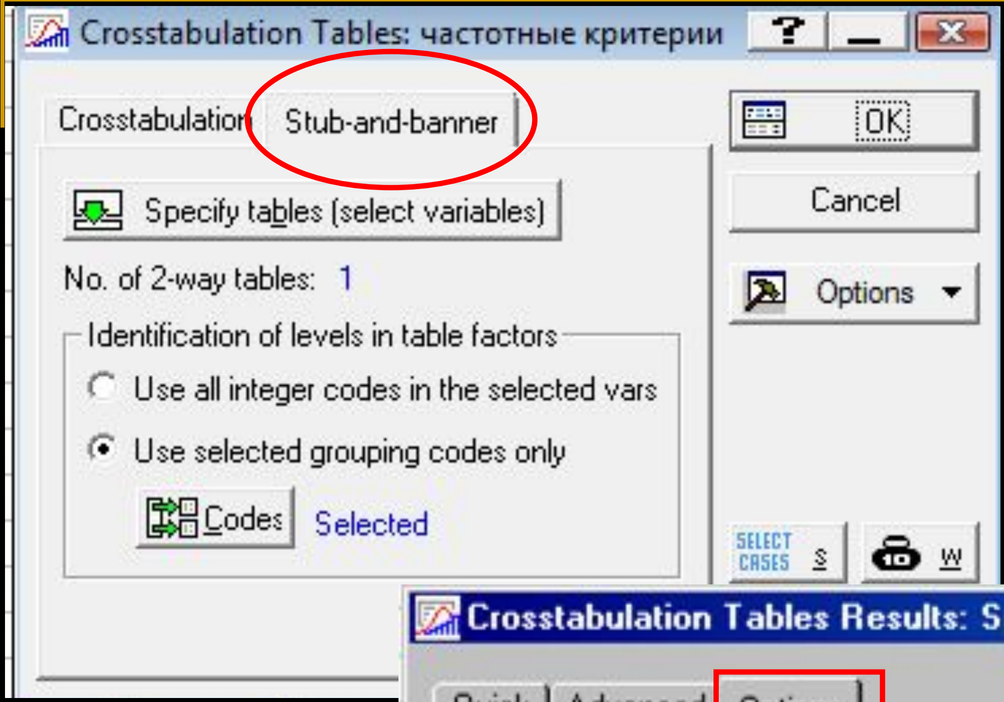
Мы для каждой ячейки рассчитываем ожидаемую частоту (на основе общих частот для столбцов и строк).

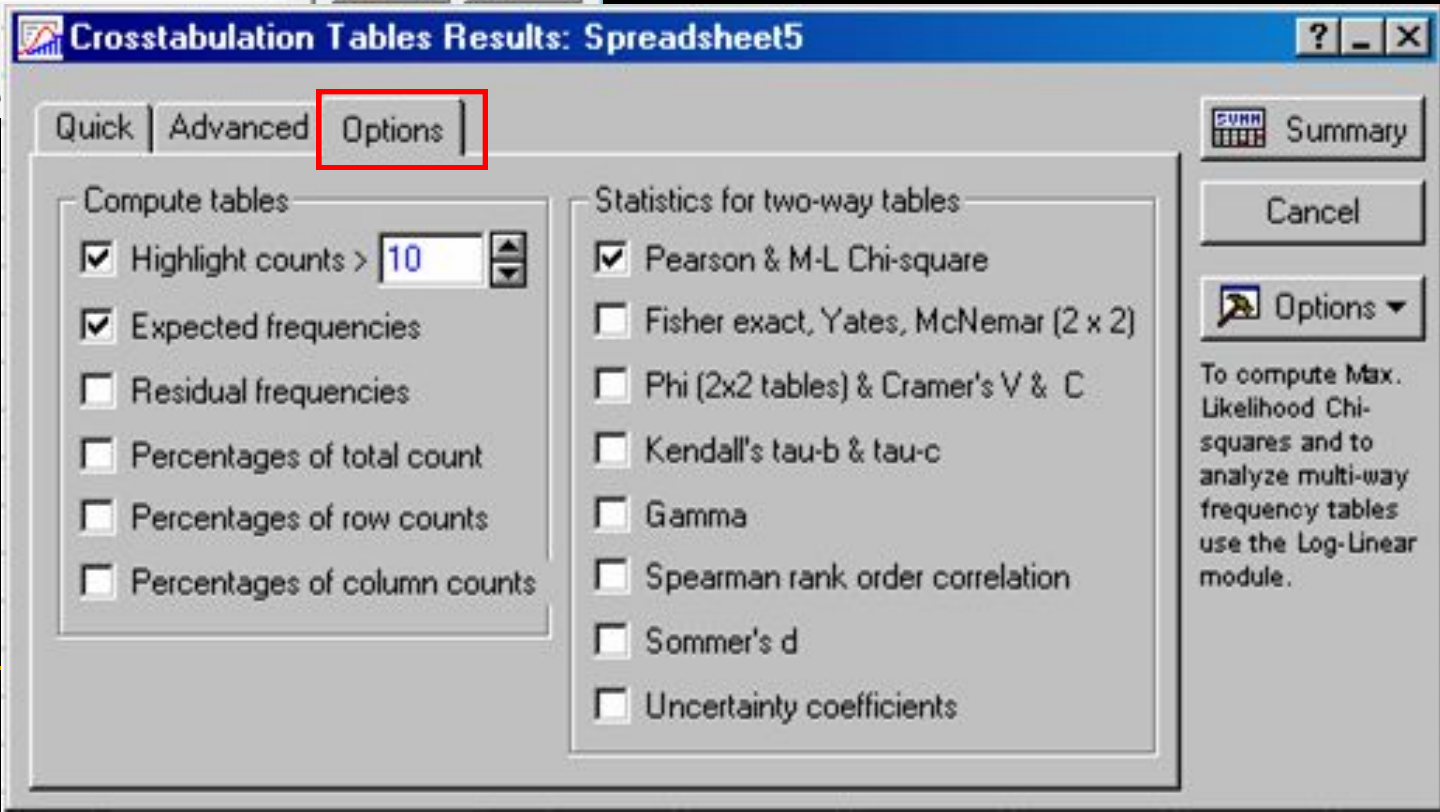
Потом считаем обыкновенную статистику χ^2 :

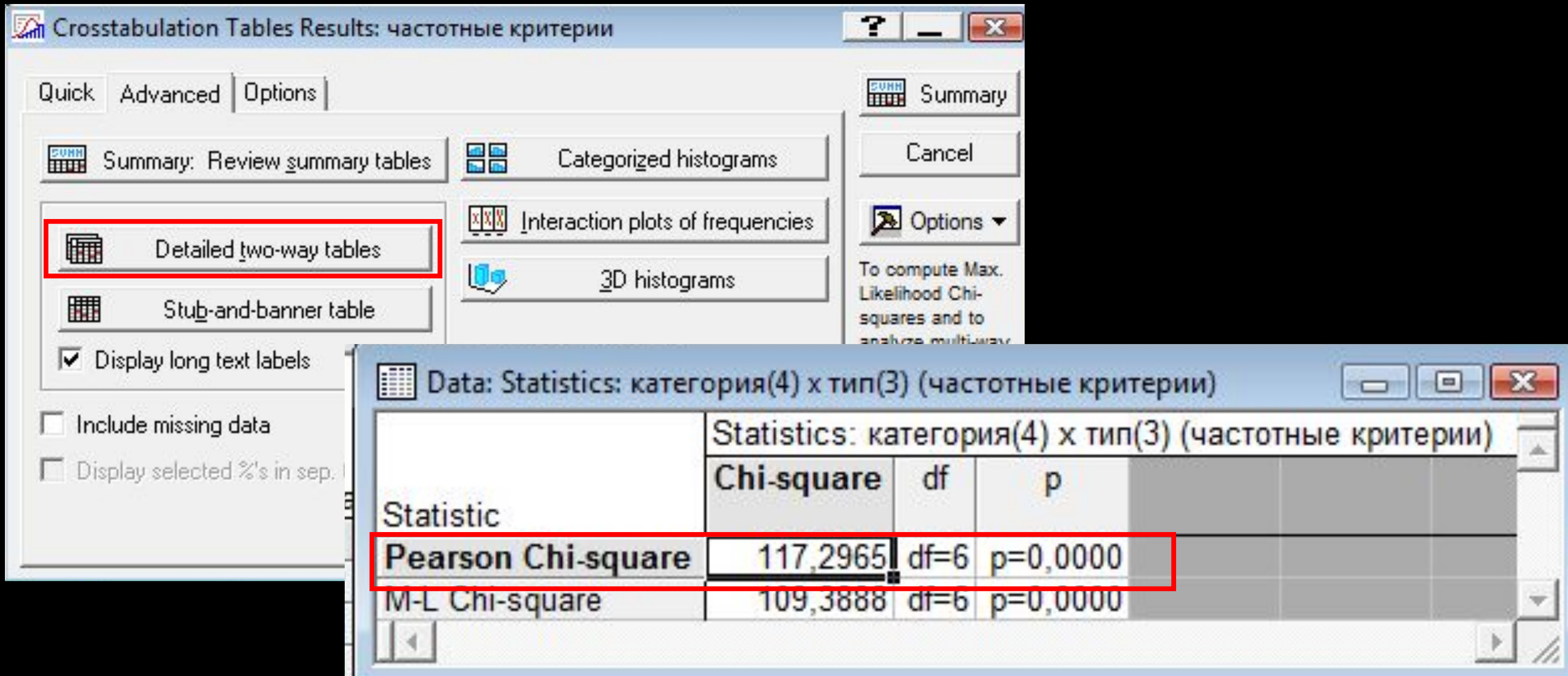
$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

в таблице должны быть сырые данные




Crosstabulation Tables: частотные критерии
Crosstabulation
Stub-and-banner
OK
Cancel
Specify tables (select variables)
No. of 2-way tables: 1
Options
Identification of levels in table factors
Use all integer codes in the selected vars
Use selected grouping codes only
Codes
Selected
SELECT CASES
S
W


Crosstabulation Tables Results: Spreadsheet5
Quick
Advanced
Options
Summary
Cancel
Options
Compute tables
Highlight counts > 10
Expected frequencies
Residual frequencies
Percentages of total count
Percentages of row counts
Percentages of column counts
Statistics for two-way tables
Pearson & M-L Chi-square
Fisher exact, Yates, McNemar (2 x 2)
Phi (2x2 tables) & Cramer's V & C
Kendall's tau-b & tau-c
Gamma
Spearman rank order correlation
Sommer's d
Uncertainty coefficients
To compute Max. Likelihood Chi-squares and to analyze multi-way frequency tables use the Log-Linear module.



Отвергаем нулевую гипотезу об отсутствии взаимодействия между переменными

В табличке с частотами вида a × b не должно быть значений меньше 5. Если это не так, следует объединить какие-нибудь проявления признака.

Частотный анализ

Zar, 1999:

Если вы не отвергли связь переменных (!), а теперь хотите показать, из-за какой именно категории есть связь, можно отдельно проверить связь переменных на остальных категориях, а затем – отношение этой категории к остальным.

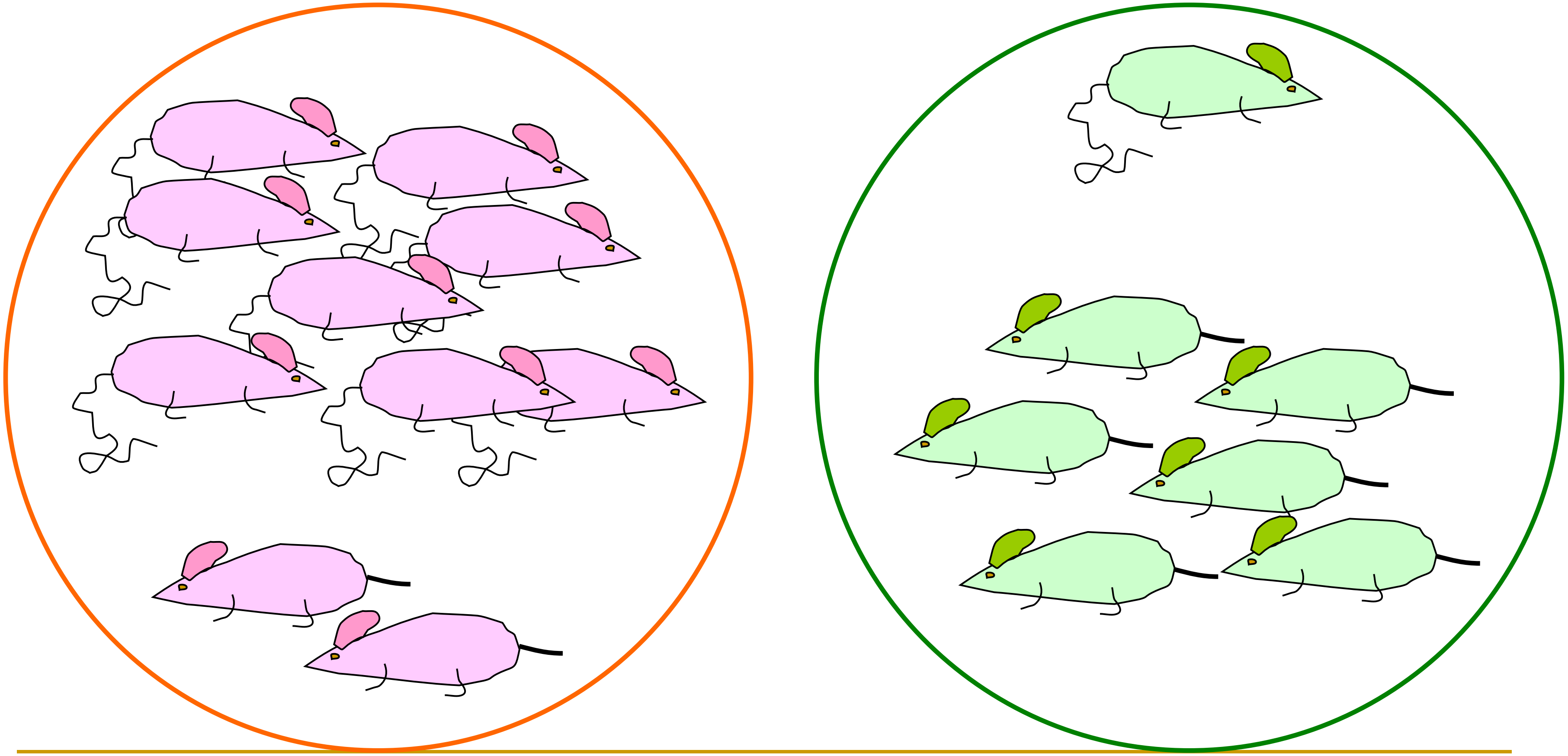
Например, если самцы и самки коз отличаются, по-видимому, только по соотношению белых коз, можно:

1. исключить белых, проверить связь пола и цвета для остальных;
2. проверить связь пола и присутствия белого цвета у козы.

Частотный анализ

Четырёхпольные таблицы (2 x 2 tables) для независимых выборок.

Есть только 2 фактора, у каждого – только по 2 проявления.



Связан ли цвет мышей с формой их хвостов??

Частотный анализ

Четырёхпольные таблицы (2 x 2 table)

	хвост	хвост	
роз	18	12	29
зел	11	26	38
	29	38	67

Модель 1: мы задаём только общий размер выборки

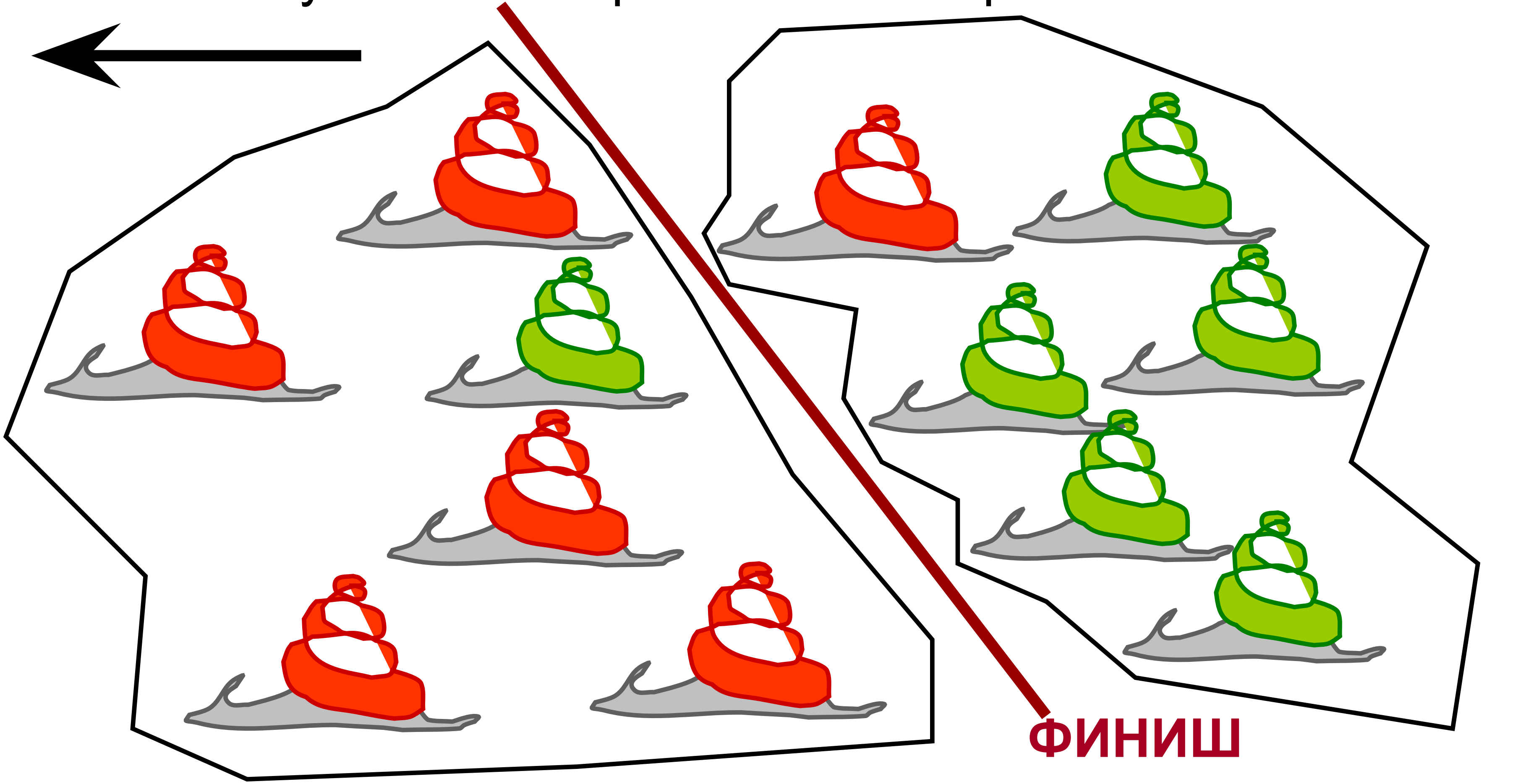
Модель 2: одна из сумм фиксирована (взяли поровну мальчиков и девочек и сравниваем долю левшей).

Модель 3: фиксированы обе суммы (про улиток)

Обычно мы имеем дело с моделями 1 и 2.

Частотный анализ

Пояснение к Модели 3 – красных и зелёных улиток по 6 штук, соревнование продолжалось до тех пор, пока половина улиток не перешла линию финиша



Критерий χ^2 (Chi-square) с поправкой Йейтса.

Если в табличке сырые данные, а не готовая четырёхпольная таблица – Tables and Banners.
Если готовая таблица – 2 х 2 tables.

Принцип введения поправки – тот же, что для сравнения наблюдаемых и ожидаемых частот, делает тест более консервативным.
Не нужна для больших выборок. В Statistica: поправку вводят, если хотя бы одна частота меньше 10.

Лучше всего подходит для модели 1.

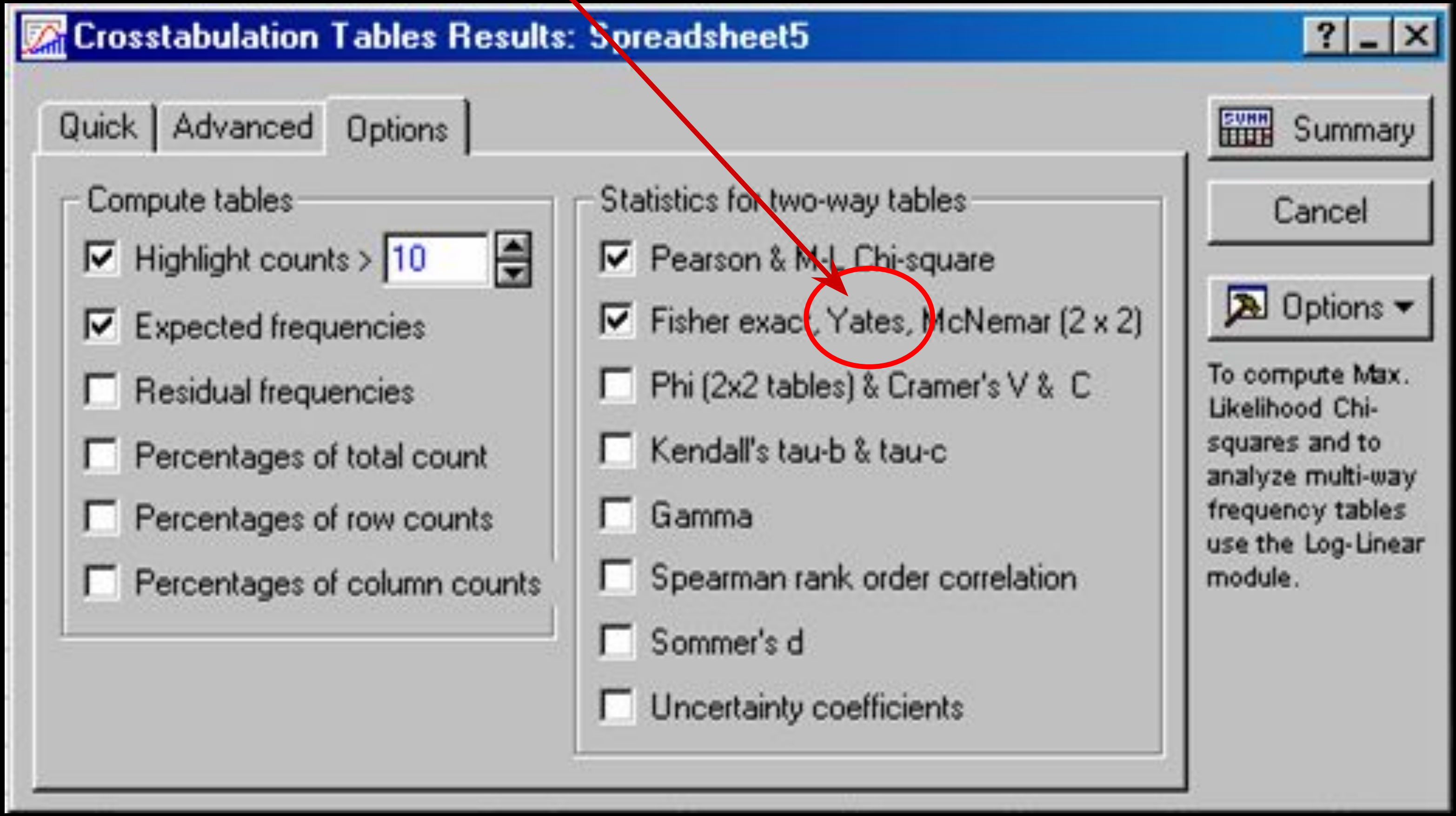
Частотный анализ

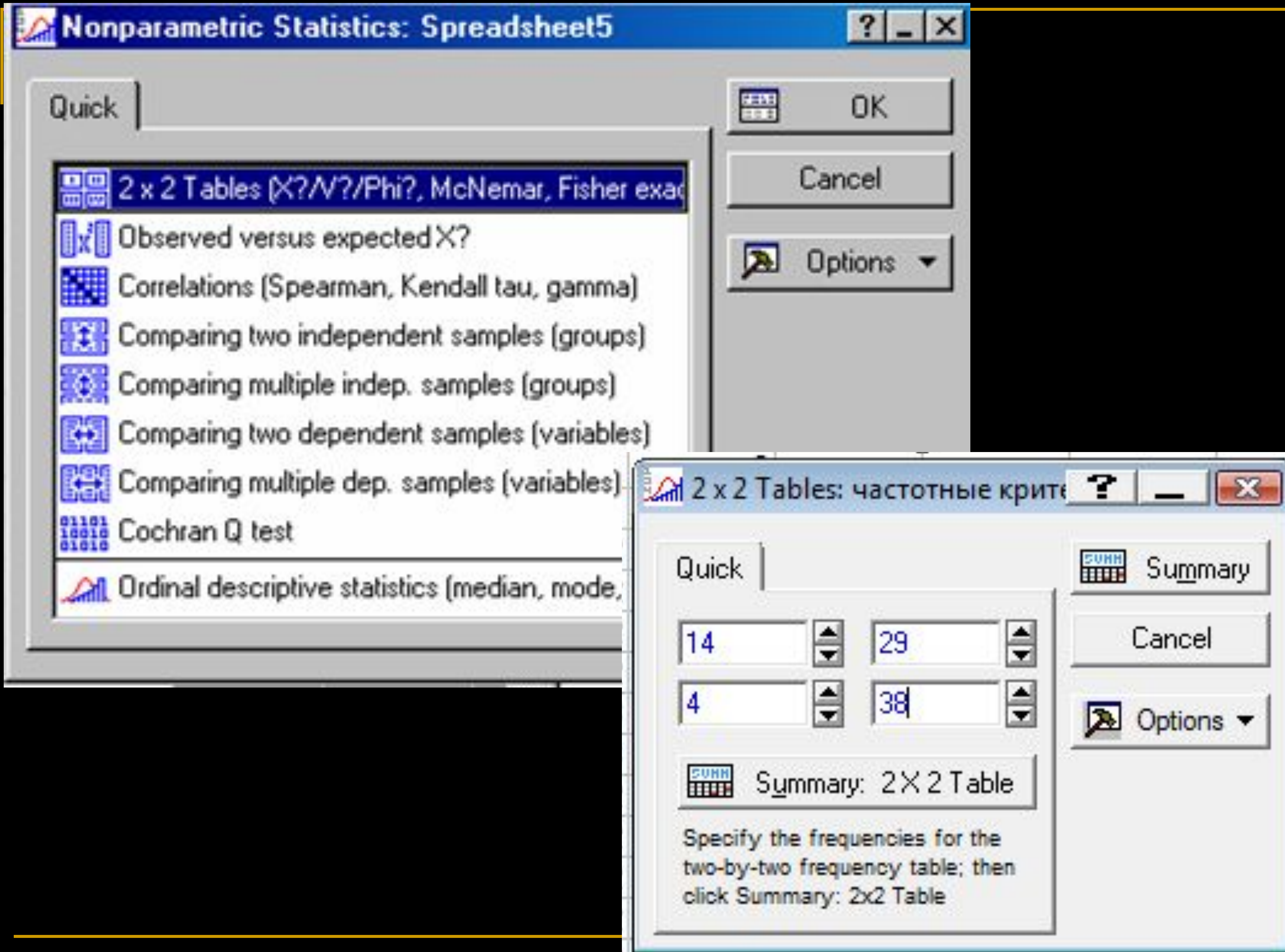
Точный критерий Фишера (Fisher exact test)

Годится, если одна из частот меньше 5 и вообще, для небольших выборок. Подходит для 3-й модели. Вообще, лучший из 2x2 тестов (Zar, 1999)

скунсы	***с бешенством***	***без бешенства***
восточные	14	29
западные	5	38

H_0: район, где живёт скунс, и заболеваемость не связаны друг с другом;
H_1 : между районом и заболеванием есть связь.


Crosstabulation Tables Results: Spreadsheet5
Quick
Advanced
Options
Summary
Cancel
Options
Compute tables
Highlight counts >
10
Expected frequencies
Residual frequencies
Percentages of total count
Percentages of row counts
Percentages of column counts
Statistics for two-way tables
Pearson & M-L Chi-square
Fisher exact, Yates, McNemar (2 x 2)
Phi (2x2 tables) & Cramer's V & C
Kendall's tau-b & tau-c
Gamma
Spearman rank order correlation
Sommer's d
Uncertainty coefficients
To compute Max. Likelihood Chi-squares and to analyze multi-way frequency tables use the Log-Linear module.


Nonparametric Statistics: Spreadsheet5
Quick
OK
Cancel
Options
2 x 2 Tables (X?/V?/Phi?, McNemar, Fisher exac
Observed versus expected X?
Correlations (Spearman, Kendall tau, gamma)
Comparing two independent samples (groups)
Comparing multiple indep. samples (groups)
Comparing two dependent samples (variables)
Comparing multiple dep. samples (variables)
Cochran Q test
Ordinal descriptive statistics (median, mode,
2 x 2 Tables: частотные крите
Quick
Summary
Cancel
Options
14
29
4
38
Summary: 2 X 2 Table
Specify the frequencies for the two-by-two frequency table; then click Summary: 2x2 Table

Data: 2 x 2 Table (частотные критерии)

	2 x 2 Table (частотные критерии)		
	Column 1	Column 2	Row Totals
Frequencies, row 1	14	29	43
Percent of total	16,471%	34,118%	50,588%
Frequencies, row 2	4	38	42
Percent of total	4,706%	44,706%	49,412%
Column totals	18	67	85
Percent of total	21,176%	78,824%	
Chi-square (df=1)	6,75	p= ,0094	
V-square (df=1)	6,67	p= ,0098	
Yates corrected Chi-square	5,44	p= ,0196	
Phi-square	,07946		
Fisher exact p, one-tailed		p= ,0089	
two-tailed		p= ,0155	
McNemar Chi-square (A/D)	10,17	p= ,0014	
Chi-square (B/C)	17,45	p= ,0000	

Отвергаем H_0

Скунсы из разных районов имеют разную заболеваемость.

Замечание: тест в данном случае <u>двусторонний</u>!!

Частотный анализ

Односторонний тест Фишера:

Для случаев, когда мы заранее знаем, куда может отклониться соотношение частот.

Например, мы даём лекарство больным зверям и сравниваем, сколько из них выздоровело по сравнению с контрольной группой. Предполагается, что лекарство не может ухудшить состояние зверей, а только может либо вылечить, либо нет.

Частотный анализ

Phi-square – показатель корреляции между качественными переменными.

V-square – разновидность χ^2 теста.

Все эти тесты подразумевали, что выборки <u>независимы</u> (например, каждая особь входит только в одну из ячеек).

Анализ 2-х связанных выборок:

Критерий Мак-Немара (McNemar Chi-square)

Мы провели в сентябре экзамен по математике. Из 100 учеников 36 сдали экзамен, остальные - провалили. Потом мы подвергли всех учеников интенсивным занятиям по математике.
Для тех же учеников мы провели экзамен во 2-й четверти.
Повлияли ли занятия на успеваемость?

По сути дела, это просто двухвыборочный тест для связанных выборок – аналог критерия Вилкоксона, только для качественных переменных

Требуется специальная организация таблицы

Частотный анализ

Экзамен второй	Экзамен первый		Всего
	Не сдали	Сдали	
Не сдали	12	6	18
Сдали	52	30	82
	64	36	

H_0: доля учеников, которые сдали экзамен в первый раз, такая же, как и во второй раз.

H_1 : эти доли различаются.

Рассчитываем ожидаемые частоты для «зелёных» ячеек и сравниваем их с наблюдаемыми частотами тестом χ^2. Нельзя менять порядок чисел, когда мы вносим их в Статистику!

Частотный анализ

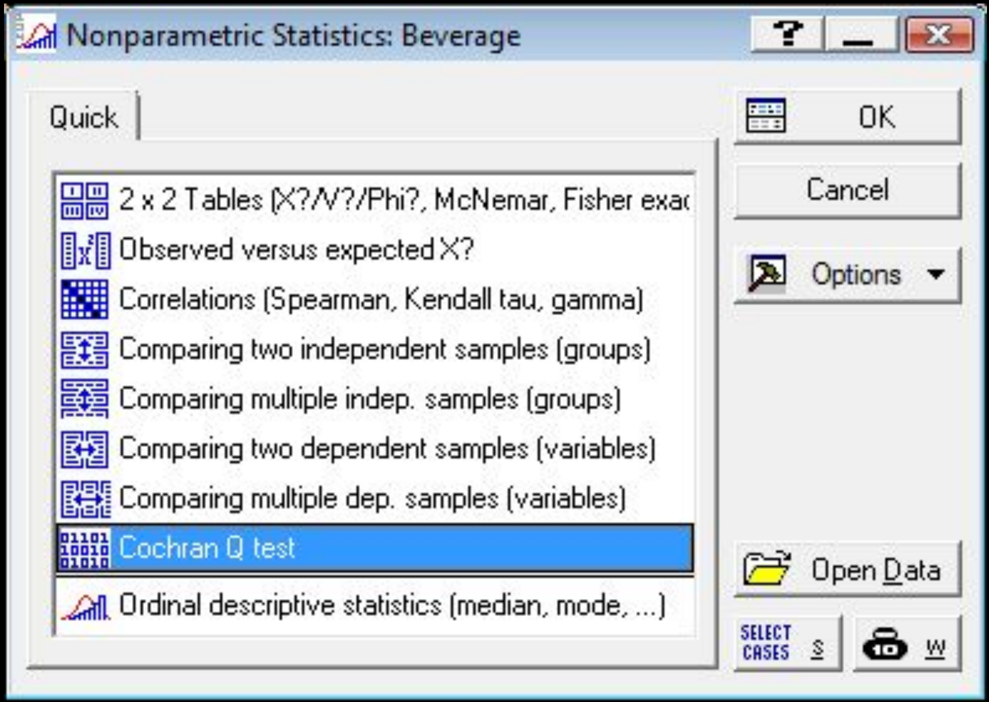
Анализ ≥3-х связанных выборок:
Cochran's Q test

Сравнивает несколько связанных измерений одной БИНАРНОЙ переменной.
(например, присутствие/отсутствие гельминтов у самок суслика сразу после спячки – во время беременности – во время лактации – перед спячкой).

Beverage (16v by 34c)

Beverage preferences (see Hoffman & Frank

	1 COKE_Y	2 COKE_N	3 DCOKE_Y	4 DCOKE_N
1	1	0	0	1
2	1	0	0	1
3	1	0	0	1
4	0	1	1	0
5	1	0	0	1
6	1	0	0	1
7	0	1	1	0
8	1	0	1	0
9	1	0	1	0
10	1	0	0	1
11	1	0	0	1
12	0	1	1	0
13	0	1	0	1
14	1	0	0	1
15	0	1	1	0
16	0	1	0	1
17	0	1	1	0
18	1	0	1	0
19	1	0	[illegible]	[illegible]



ran Q Test (Beverage)

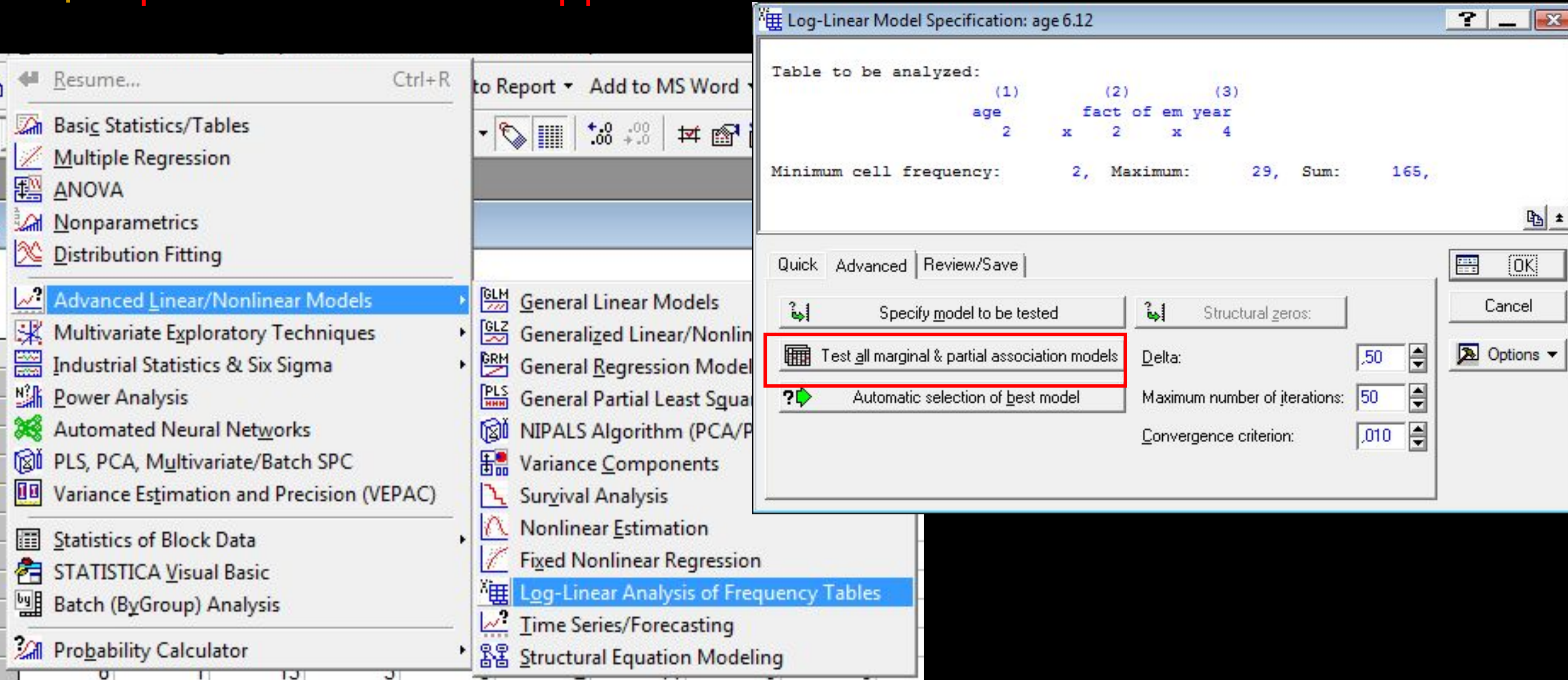
Cochran Q Test (Beverage)

Number of valid cases:34

Q = 1,588235, df = 3, p < ,662061

Variable	Sum	Percent 0's	Percent 1's
COKE_Y	20,00000	41,17647	58,82353
COKE_N	14,00000	58,82353	41,17647
DCOKE_Y	17,00000	50,00000	50,00000
DCOKE_N	17,00000	50,00000	50,00000

Частотные критерии для 3-х и более переменных, с оценкой их взаимодействия



of Marginal and Partial Association (age 6.12)

Tests of Marginal and Partial Association (age 6.12)

Effect	Degrs.of Freedom	Prt.Ass. Chi-sqr.	Prt.Ass. p	Mrg.Ass. Chi-sqr.	Mrg.Ass. p
1	1	6,33354	0,011848	6,33354	0,011848
2	1	14,07040	0,000176	14,07040	0,000176
3	3	7,86508	0,048884	7,86508	0,048884
12	1	2,36479	0,124101	3,62716	0,056844
13	3	14,00820	0,002894	15,27058	0,001599
23	3	14,98889	0,001826	16,25127	0,001007

Задания.

1. Хазел Нат продаёт смесь орехов. На упаковке написано, что в пачке содержится 30% кешью, 20% бразильских орехов, 20% грецких, 30% лесных. Мы хотим проверить, так ли это, взяли большую пачку и посчитали в ней разные орехи (200 орехов). H_0? Статистический критерий?
2. Мы хотим прививать детям Сибири бережное отношение к природе. Мы выбрали 100 первоклашек и спросили их, можно ли охотиться на кабаргу (78 ответили «да», 22 – «нет»). Потом им показывали фильмы и рассказывали о местной фауне весь год. Весной этих же детей спросили о том же. Из тех, кто был за охоту, 18 опять ответили «да», 60 – «нет». Из тех, кто был против – 2 ответили «да», 20 - «нет». H_0? Статистический критерий?
3. Издатели хотят узнать, насколько наличие цветных картинок в статье помогает воспринимать текст. Выбрали 13 студентов, и каждому дали два текста одинаковой сложности - с цветными и чёрно-белыми картинками. Потом попросили оценить сложность текста по 10-бальной шкале. Влияют ли цветные картинки на восприятие текста? H_0? Статистический критерий?

4. Проходят соревнования по фигурному катанию. Мы хотим узнать, влияет ли жанр исполняемой музыкальной композиции во время выступления на оценку фигуриста. 30 фигуристам случайным образом заранее предложили композиции на основе классической музыки, тяжёлого рока и поп-музыки (по 10 композиций на жанр). Жюри выставило оценки. Зависят ли они от музыкального жанра? H_0? Статистический критерий?

5. Мы хотим знать, зависит ли вероятность принести потомство от возраста самки у белок. Мы не знаем точный возраст зверьков, можем лишь отличить взрослых от годовалых. Мы исследовали 50 годовалых и взрослых самок, и выяснили, какие самки из них принесли выводки, какие – остались холостыми. H_0? Статистический критерий?