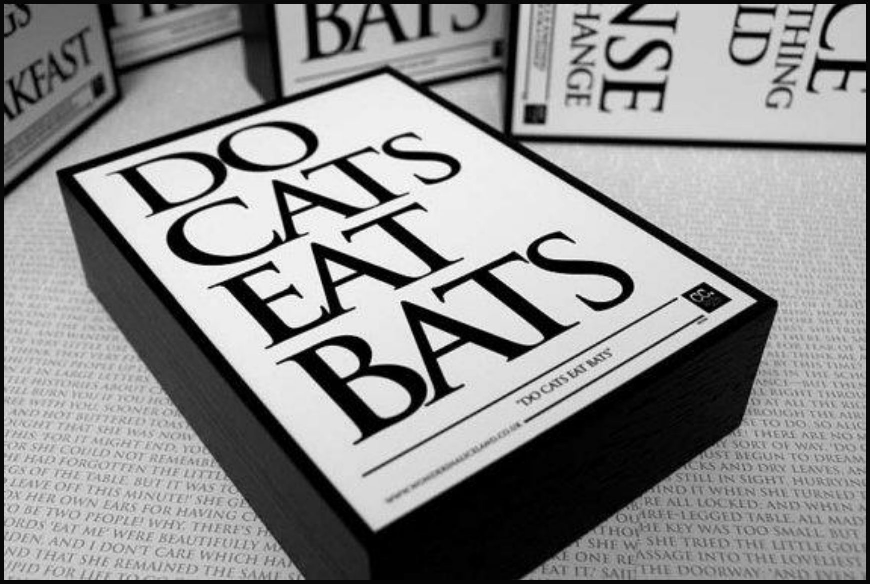
Занятие 2

Тестирование гипотез в статистике.

Критерии Стьюдента



Описательная статистика и оценка параметров позволяют количественно охарактеризовать данные. Но хочется большего!

Хочется получать ответы на вопросы вроде:

Кто жирнее, самцы или самки?
Как влияет плотность особей на дистанции расселения?
Худеют ли люди, если сидят на кефирной диете? …

Это вопросы о свойствах и взаимосвязях параметров в популяциях!

Мы **оценивали** параметры; теперь мы будем **тестировать гипотезы** об этих параметрах.

Анатомия научного исследования

Исходя из какой-л. Теории высказываем предположение – «научную гипотезу»

↓

Высказываем статистическую гипотезу о параметрах популяций

↓

Собираем данные, тестируем её (эксперимент или наблюдения)

Аспиранты ИПЭЭ отличаются от остальных людей по уровню IQ

↓

IQ аспирантов ИПЭЭ ≠ 100

↓

Устраиваем аспирантам экзамен и выясняем, так ли это

Тестирование гипотез в статистике

Статистическая гипотеза – **предположение** о свойстве **популяции,** которое выражено с помощью **параметров**.

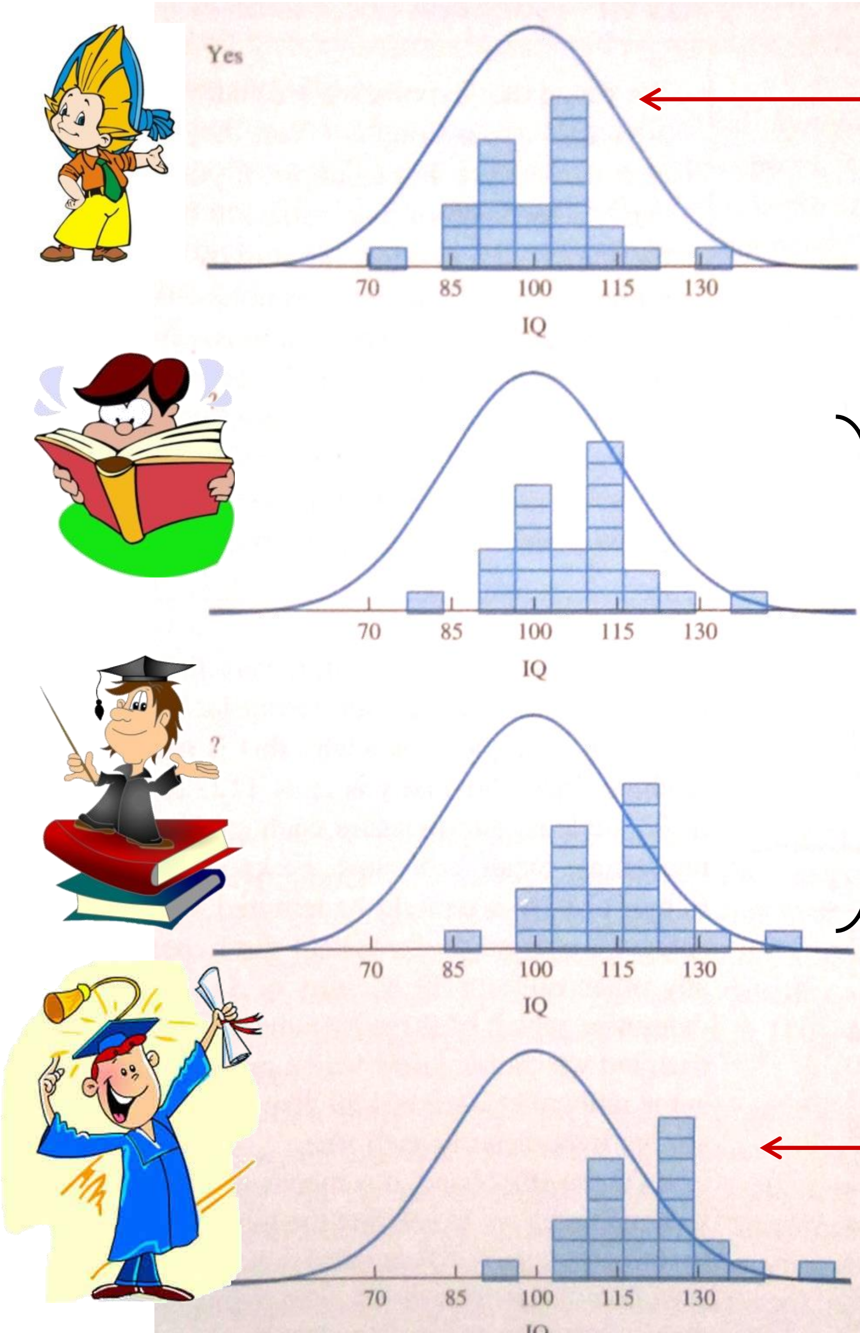
Тестирование гипотезы (hypothesis testing) – процедура, в которой мы решаем, принять гипотезу («accept») или отвергнуть (reject).

Итак:

Мы хотим знать, являются ли аспиранты ИПЭЭ случайной выборкой из популяции с IQ μ=100.

?

Аспиранты ИПЭЭ – случайная выборка из популяции с IQ = 100?



Скорее всего, ДА

Надо бы придумать какой-то критерий, когда сказать ДА, когда - НЕТ

Различия неочевидны

Скорее всего, НЕТ

Тестирование гипотез в статистике

Как проверить гипотезу? Гипотезу на основе эмпирических данных гораздо проще опровергнуть, чем подтвердить! (о лебедях)

Поэтому придумали сформулировать гипотезу, противоположную той, которую мы проверяем, и постараться опровергнуть её. («от противного»)

Нулевая гипотеза – null hypothesis, H_0

Тестирование гипотез в статистике

Гипотеза формулируется о ПАРАМЕТРАХ ПОПУЛЯЦИИ (предположения о выборке легко проверить прямыми расчётами, но эти выводы нельзя будет распространить за пределы выборки).

Формулируем ДВЕ взаимоисключающие гипотезы:

H_0 (**нулевая гипотеза**, null hypothesis) – её мы собираемся опровергать; обычно говорит, что нет различий, нет эффекта, нет изменений…

H_1 (**альтернативная гипотеза**, alternative hypothesis) – её мы примем, если удастся отвергнуть H_0.

Теперь на основе выборки надо бы что-то посчитать, потом посмотреть на то, что получилось и решить: отвергнуть ***H_0*** *или нет.*

Тестирование гипотез в статистике

Мы хотим узнать, отличается ли IQ аспирантов ИПЭЭ от 100. Аспиранты ИПЭЭ – наша популяция.

H_0: $\mu = 100$;

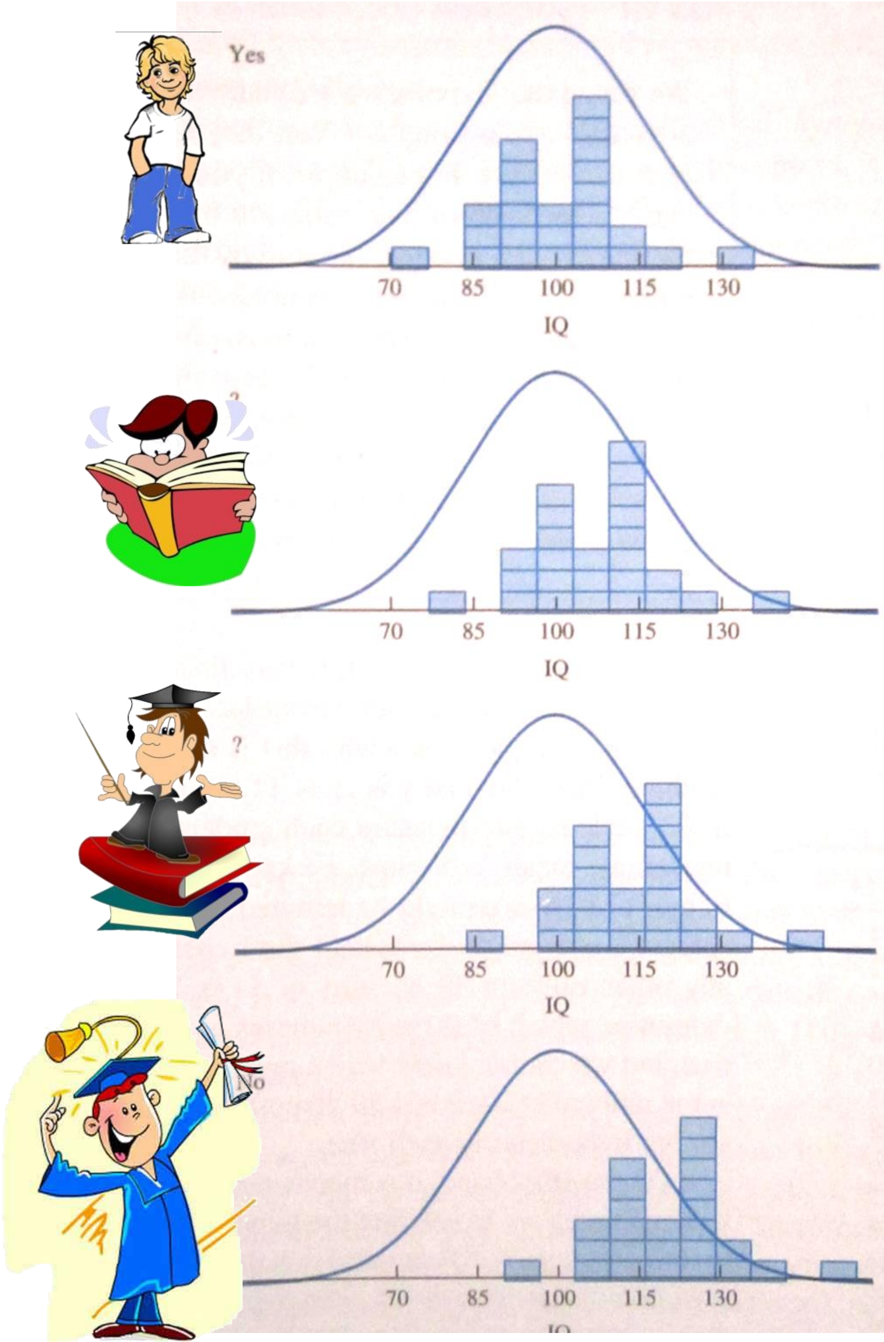
H_1 : $\mu \neq 100$

μ – среднее значение в <u>популяции</u> аспирантов ИПЭЭ.

Решение о том, принять или отвергнуть гипотезу принимается на основе <u>статистики критерия</u> (test statistic), которая считается на основе характеристик выборки (statistics).

Для каждого вида нулевой гипотезы уже придумана своя статистка критерия.

Тестирование гипотез в статистике



Итак, позволяют ли наши данные отвергнуть H_0?

Это мы решаем на основе СТАТИСТИКИ КРИТЕРИЯ (test statistic).

Статистика критерия рассчитывается на основе характеристик ВЫБОРКИ, и её распределение известно (и соотношение площадей под ним!).

Одновыборочные критерии сравнивающие среднее значение с заданным числом.

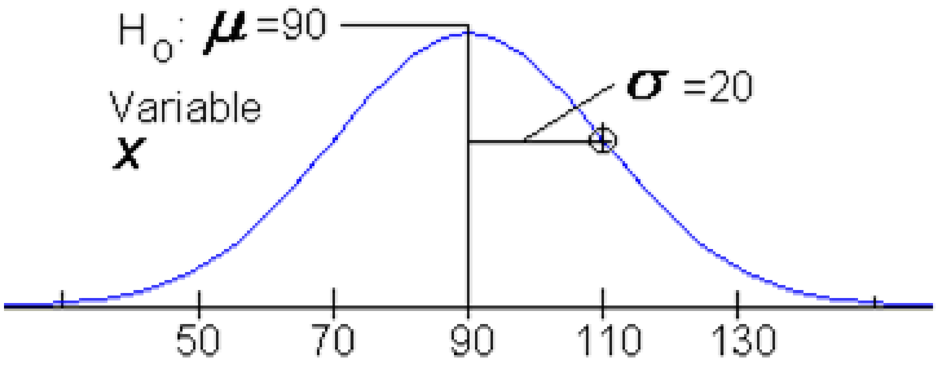
Мы хотим узнать, отличается ли средняя масса землероек, живущих на острове, от массы гигантских землероек = 90 г, заявленной в Mammalian species. Подготовимся к исследованию!

Мы знаем, что μ=90, σ=20; собираемся поймать 25 зверьков.

1. Формулируем H_0 и H_1 :

$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$



Имея в голове H_0, сидим дома и рисуем 3 распределения:

- масса землероек со средним=90 г;
- распределение выборочных средних для выборок N=25 для μ=90;
- распределение статистики критерия *z*.

Тестирование гипотез в статистике

Общий принцип формирования статистики простых критериев:

$$\text{Статистика} = \frac{\text{параметр } \mathbf{выборки} - \text{параметр } \mathbf{популяции}}{\text{стандартная } \mathbf{ошибка} \text{ параметра выборки}}$$

Параметр популяции – определяется <u>гипотезой H_0.</u>

Характеристика выборки – <u>оценка</u> этого параметра.

Стандартная ошибка характеристики выборки. Она определяет, насколько большими могут быть СЛУЧАЙНЫЕ отличия между характеристикой выборки и параметром популяции.

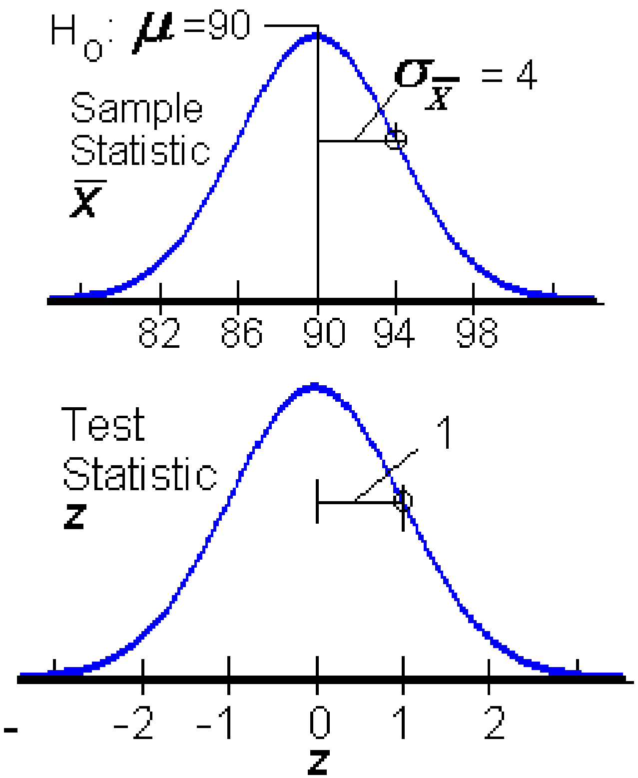
Одновыборочные критерии

$$\text{Статистика} = \frac{\text{параметр выборки} - \text{параметр популяции}}{\text{стандартная ошибка параметра выборки}}$$

разность <u>выборочного</u> среднего и популяционного

$$z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

ошибка среднего



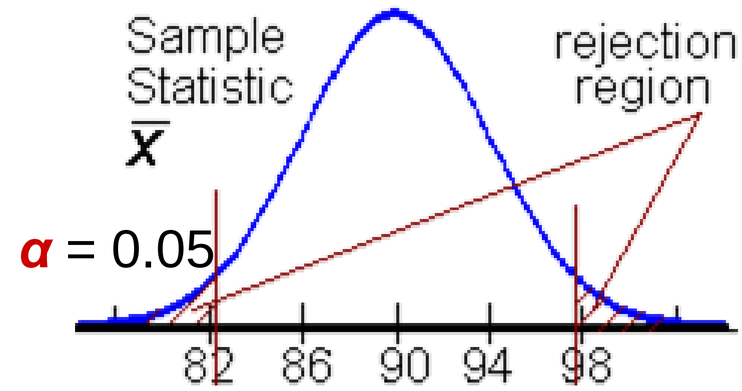
Это распределение <u>выборочных средних</u> (нормальное!) и построенное для него z-распределение.

Всё это мы производим ДО взвешивания землероек! Это пока распределения, определённые нулевой гипотезой.

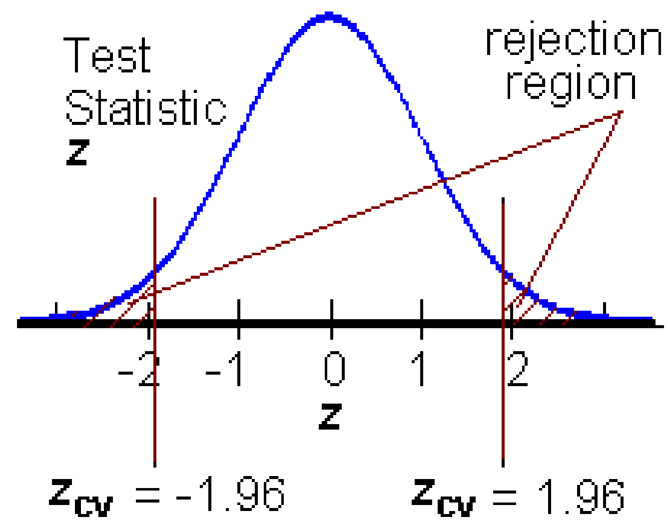
Одновыборочные критерии

2. Устанавливаем условия, при которых мы отвергнем H_0

Насколько далеко выборочное среднее может отклониться от популяционного, если выборка – случайная? Куда попадёт большинство (скажем, 95%) выборочных средних?



Можем легко посчитать благодаря волшебным свойствам нормального распределения!



*Договорились, что если выборочное среднее попадает **в 5%-ные крайние интервалы**, это уже **слишком далеко** от среднего, установленного H_0 .*

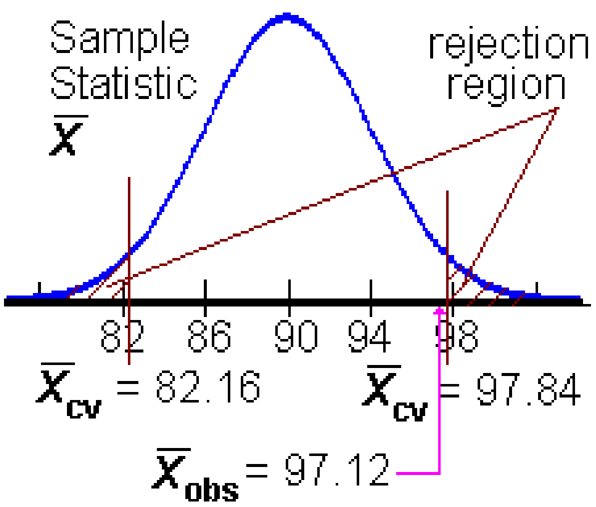
5% - условность, называется «уровень значимости»:

***α** = 0.05.*

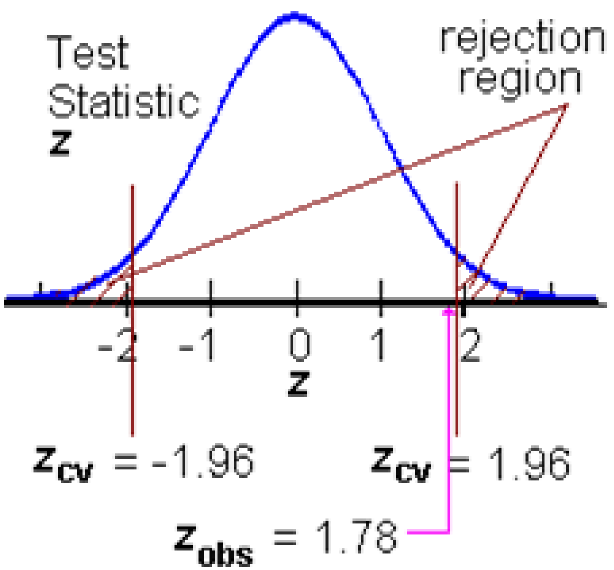
Одновыборочные критерии

3. Считаем реальные $\bar{X}$ и z

Наконец-то пора наловить и взвесить землероек!
Теперь мы **считаем среднее значение** и смотрим, куда оно попадёт.



Критическая область (rejection region) – часть оси переменной, над которой находится 5% «крайних» значений.



Критическое значение – начало критической области (здесь их два – справа и слева от среднего).

Здесь выборочное среднее в критическую область не попало - **H_0 не отвергаем**.

Беда в том, что: среднее случайной выборки из популяции, где H_0 верна, запросто (в 5% случаях) может попасть в критическую область; и, наоборот, среднее из популяции, где H_0 не верна, может попасть в интервал 95% (неизвестно, с какой вероятностью).

	Истинное (но неизвестное нам) положение дел	
	Верна H_0	Верна H_1
Мы «приняли» H_0	ПРАВИЛЬНО! 1-α (чувствительность критерия)	β ОШИБКА 2-го рода
Мы отвергли H_0	ОШИБКА 1-го рода (уровень значимости) α	ПРАВИЛЬНО! 1-β (мощность критерия)

Заметим: ошибку 1-го рода можно сделать только отвергая H_0, а ошибку 2-го рода – только «принимая» H_0 (невозможно сделать одновременно обе ошибки).

ОШИБКА 1 рода: *вероятность найти различия, где их нет.*
(IQ аспирантов ИПЭЭ такое же, как у всех; землеройки, в среднем, весят 90 г. Но нам показалось, глядя на выборку, что различия есть).
Это – нездоровые сенсации, которые могут принести большой вред.

ОШИБКА 2 рода: *вероятность не увидеть различий, где они есть.*
(На самом деле аспиранты ИПЭЭ намного умнее других! А землеройки – крупнее 90 г. Но мы были слишком строги к себе и посчитали, что этих различий недостаточно).
Это «близорукость», или «слепота» критерия, вред от неё не очень большой. Её контролировать напрямую мы не можем, только косвенным путём.

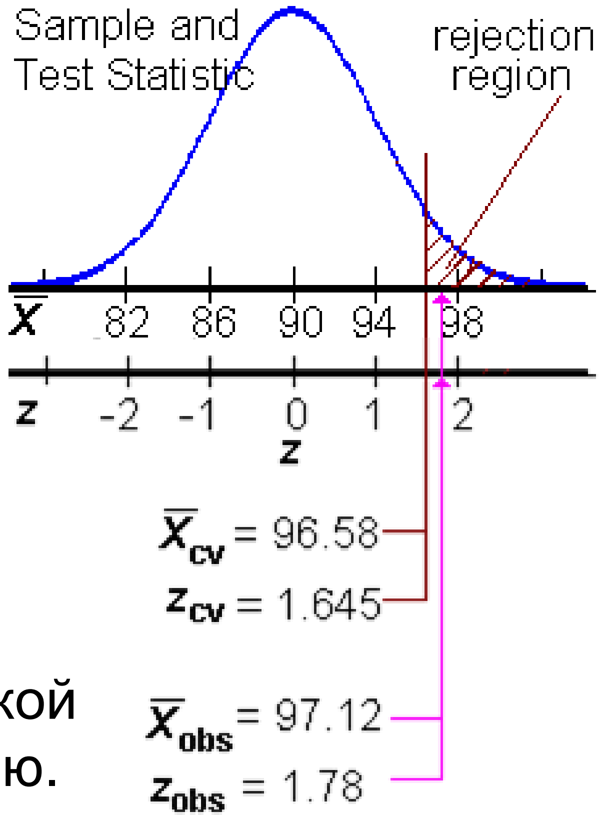
МОЩНОСТЬ (power): *вероятность найти различия, когда они есть*, что сродни мощности микроскопа; особенно интересует нас, если мы не отвергли H_0.

Когда отклонение от H_0 возможно только в одном направлении:

Недоброжелатель хочет доказать, что производитель не доливает Кока-колу в бутылки.

Он не допускает мысли, что производитель может её переливать.

Двусторонняя альтернатива (two-tailed hypothesis)	Односторонняя альтернатива (one-tailed hypothesis)
H_0: $\mu = 500$;	H_0: $\mu > 500$ г;
H_1 : $\mu \neq 500$	H_1 : $\mu \leq 500$ г



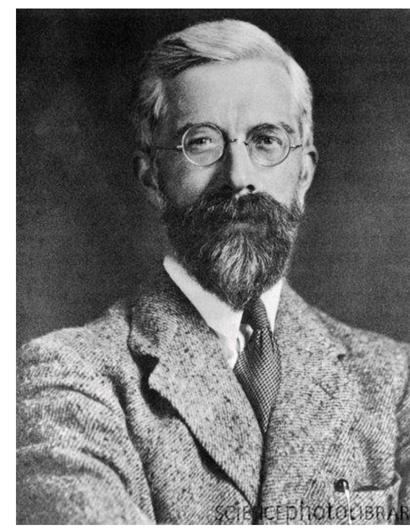
Вся критическая область (все 5% площади) отрезается с одного конца распределения статистики критерия, поэтому граница критической области оказывается ближе к среднему значению.

Применение односторонней альтернативы должно быть <u>теоретически обосновано</u> исходя из свойств переменной до сбора данных. На практике так не бывает почти никогда.

Немного об ошибке 1-го рода

Нейман и Пирсон (Neyman & Pearson): авторы концепции Ошибки 1-го рода и фиксированного уровня значимости. Это вероятность ошибочно отвергнуть H_0, когда она верна: если ***α*** = *0.05,* это значит, что если мы будем раз за разом повторять тест с разными случайными выборками из H_0-популяции, то в 5% тестах мы отвергнем H_0.

Фишер (Fisher): предлагал вместо заданного ***α*** = *0.05* считать значение статистики критерия и смотреть, какую часть площади распределения оно отрезает: оценивать ТОЧНУЮ вероятность ошибки 1-го рода – *P.* И пусть каждый сам решает, отвергать H_0 или нет.



Немного об ошибке 1-го рода

Компромисс между этими подходами: считаем точное значение *P* (сумма площадей, которые с двух сторон распределения отрезают ± рассчитанное значение статистики критерия), но всё-таки сравниваем с уровнем значимости ***α*** = *0.05*, и тогда уж делаем вывод, отвергать гипотезу или нет. *P* и приводим в статьях!

Z=2.0; значит, отрежем «хвост» z>2.0 и «хвост» z<-2.0

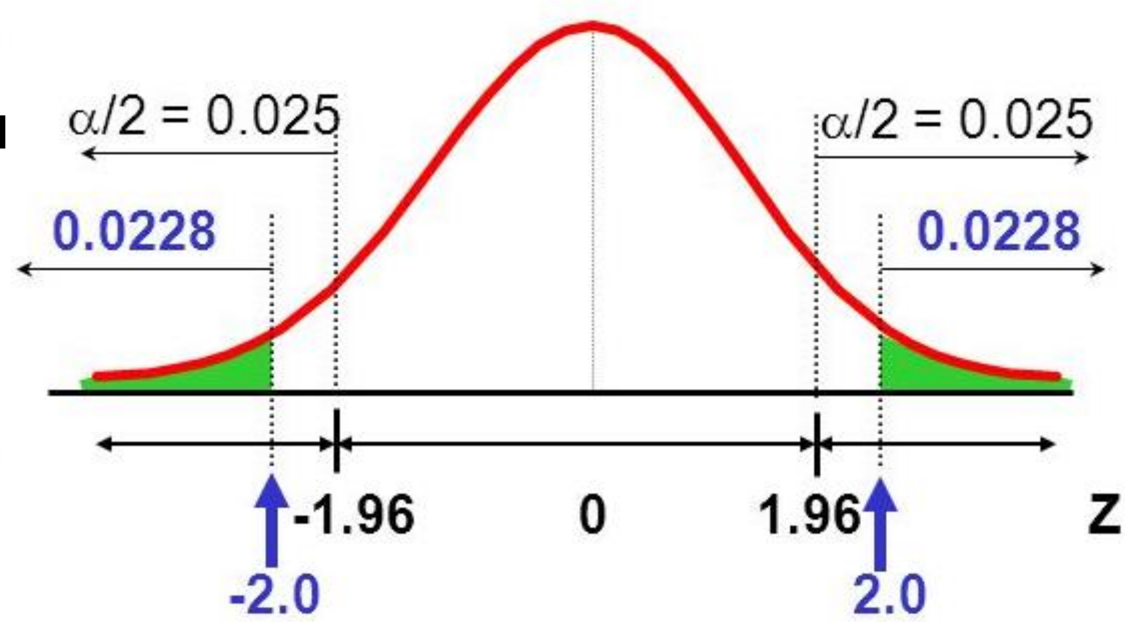


Схема тестирования статистической гипотезы

1. Формулируем H_0 и H_1. и подбираем подходящую статистику критерия. Для этой H_0 и предполагаемого размера выборки уже виртуально существуют:
   - распределение исследуемой переменной;
   - распределение статистики критерия.
2. Устанавливаем условия, при которых мы отвергнем H_0:
   - уровень значимости (α = 0.05 по умолчанию), критические области, критические значения статистики критерия.
3. Собираем данные, формируем выборку и считаем значение статистики критерия; сравниваем его с критическими значениями для H_0 – это делает программа.
4. если значение статистики попало в критическую область (т.е., $P<\alpha$), отвергаем H_0; если не попало ($P\geq \alpha$) – не отвергаем («недостоверный результат»).
5. Если P близка α , выводы делать не рекомендуется – дополнительное исследование.

Для публикаций

Для любого статистического теста приводим:

1) значение статистики критерия;

2)точное значение *P* (т.е., оценку вероятности ошибки 1-го рода). Тогда читатель может сам выбирать уровень значимости

3) Обязательно N – размер выборки!

Можно обозначить на картинке уровень *P* звёздочками:
* - достоверные различия – $p<0.05$
** - высокодостоверные различия - $p<0.01$
*** - $p<0.001$

Если вероятность ошибки близка к α, лучше всего провести дополнительные исследования и не делать окончательных выводов! (Zar, 2010)

«…All tests were two-tailed with a significance level of 0.05».

Одновыборочный t-критерий (one-sample t-test)

Обычно **дисперсия** в популяции неизвестна - распределение выборочных средних должно быть **шире**, чем нормальное!

Отличается ли масса землероек на острове от 90г? Согласно H_0, μ=90. На этот раз σ неизвестно. Предполагаем поймать 25 зверьков (n=25).

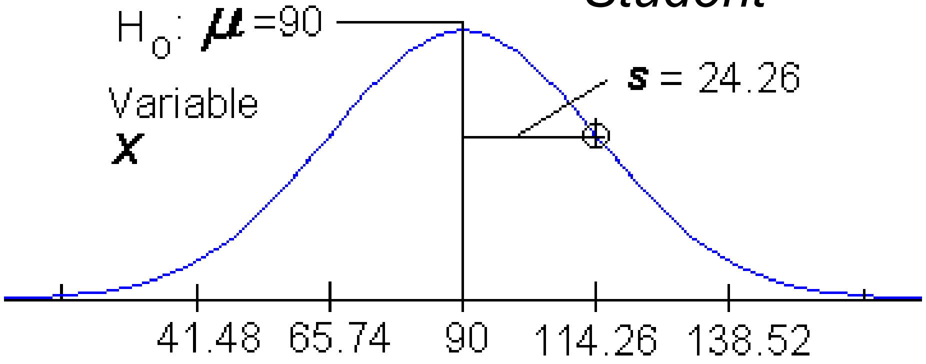
***William Sealy Gosset** (1876–1937) = “Student”*

Формулируем H_0 и H_1

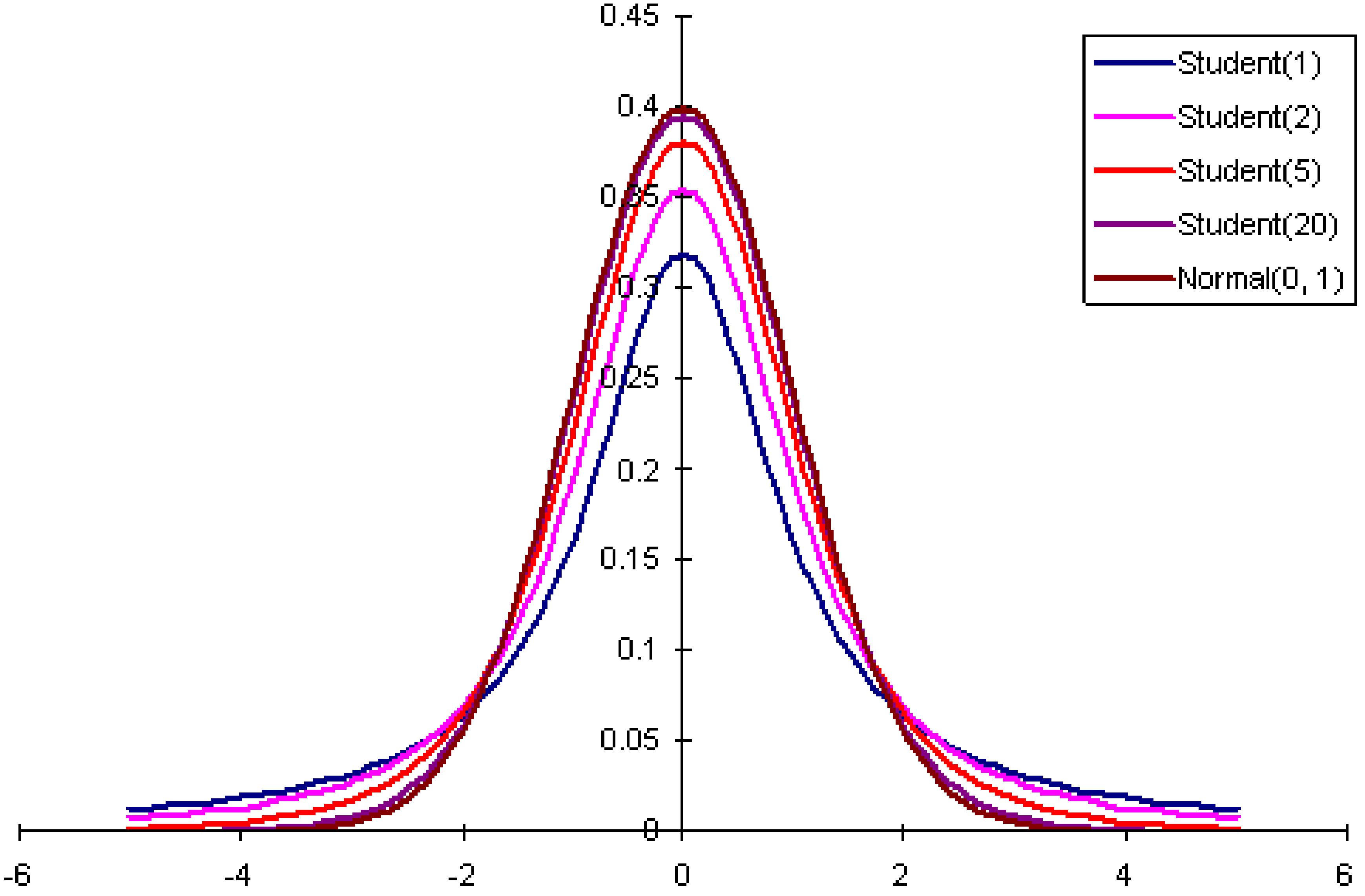
$$H_0 : \mu = 90$$

$$H_1 : \mu \neq 90$$

Дисперсию в популяции мы оцениваем дисперсией в выборке



Теперь распределение выборочных средних не будет нормальным, это – *t*-распределение. Статистика критерия – *t*.



t-распределение приближается к нормальному при df>30; чем меньше df, тем толще хвосты и ниже купол.

Одновыборочный t-критерий

$$\text{Статистика} = \frac{\text{параметр выборки} - \text{параметр популяции}}{\text{стандартная ошибка параметра выборки}}$$

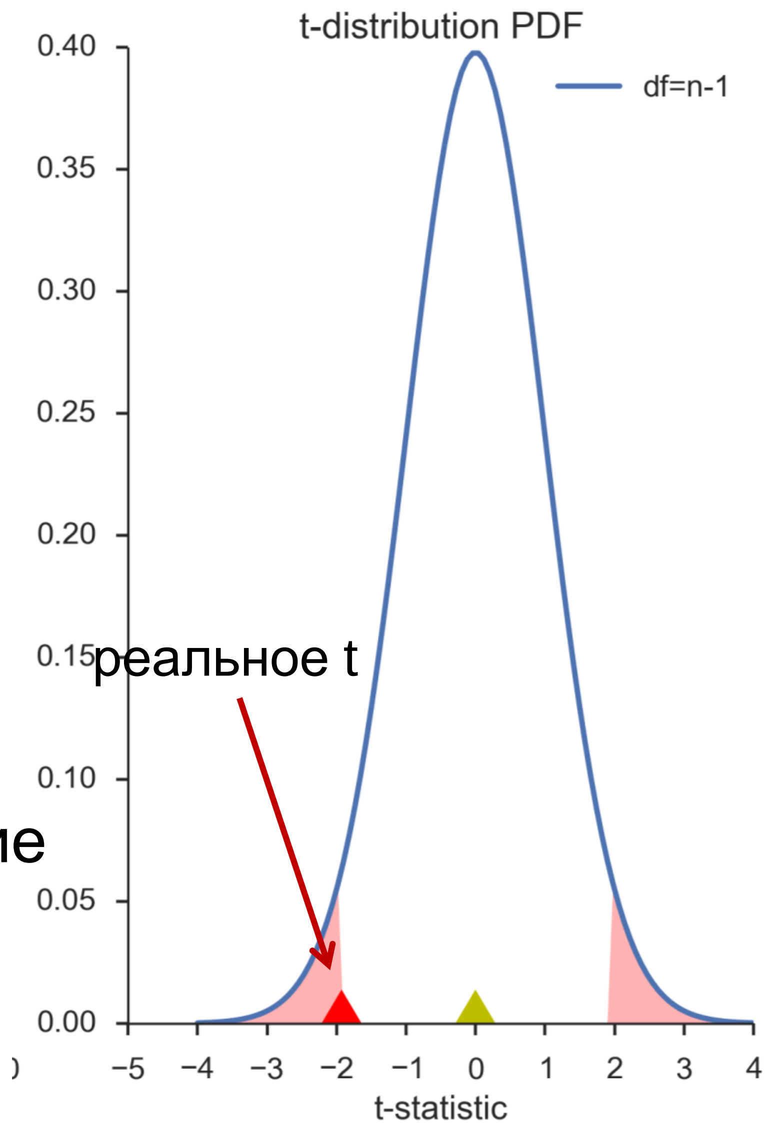
разность <u>выборочного</u> среднего и популяционного

$$\boxed{t = \frac{\overline{X} - \mu}{s_{\overline{X}}}}$$

ошибка среднего

$$s_{\overline{X}} = SE = \frac{s}{\sqrt{n}}$$

$\boxed{df = \text{n-1}}$



Соотношение площадей для t-распределения известно; критические значения дальше от среднего, чем для z.

Мы <u>не отвергаем H_0</u>!

Связь одновыборочных критериев с доверительными интервалами

выборочное среднее

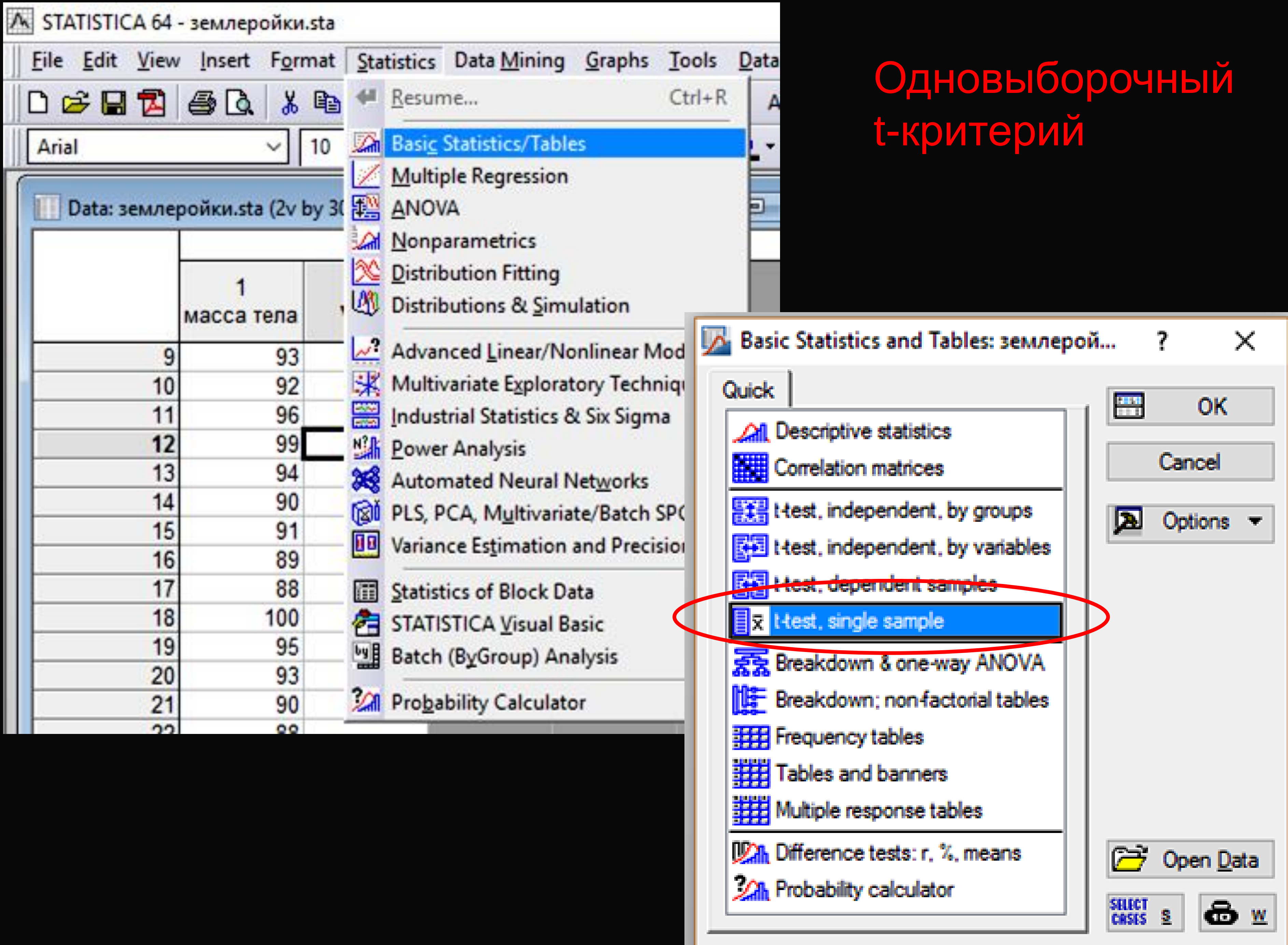
Дисперсия известна: $\mu \in \left[\overline{X} \pm z \cdot \sigma_{\overline{X}}\right]$

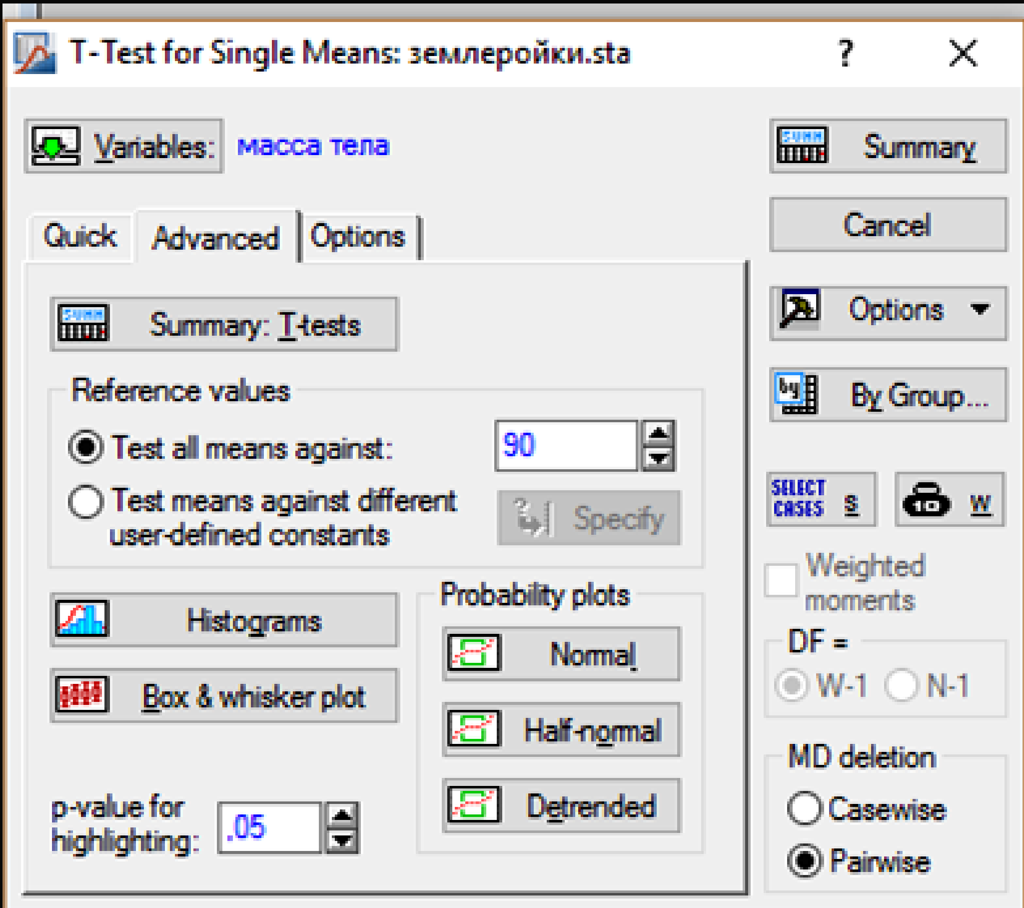
Дисперсия неизвестна: $\mu \in \left[\overline{X} \pm t \cdot s_{\overline{X}}\right]$

с вероятностью 95%

Если мы сравниваем *μ* с числом, то попадание этого числа в доверительный интервал соответствует ситуации, когда мы НЕ отвергаем H_0, а попадание за его пределы – ситуации, когда мы H_0 отвергаем.

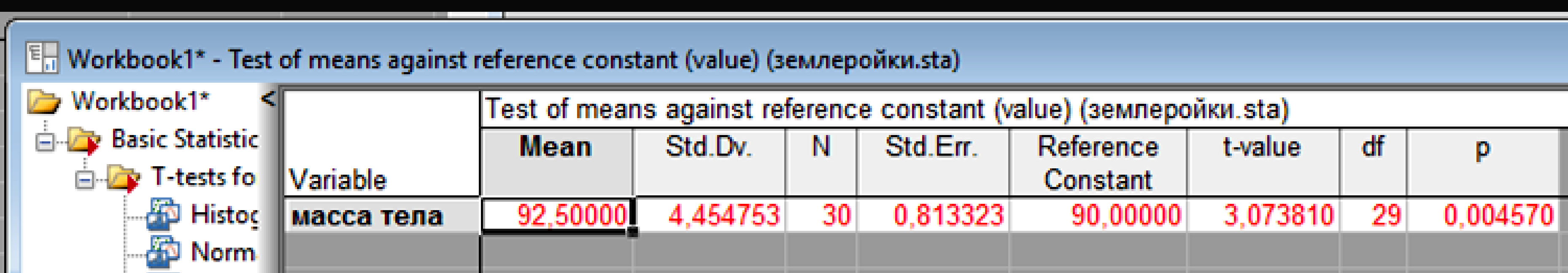
Одновыборочный t-критерий





Одновыборочный t-критерий

В данном случае мы отвергаем гипотезу о том, что масса тела у землероек = 90г.



	Test of means against reference constant (value) (землеройки.sta)							
Variable	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
масса тела	92,50000	4,454753	30	0,813323	90,00000	3,073810	29	0,004570

Одновыборочные критерии проверяют гипотезы относительно равенства числу (Zar, 2010):

- ✓Медианы
- ✓Дисперсии
- ✓Коэффициента вариации
- ✓Симметрии
- ✓Эксцесса

Но мы эти критерии рассматривать не будем, на практике это надо немногим.

Сравнение между собой средних значений 2-х выборок

До сих пор мы работали с переменными по одиночке; пора рассмотреть взаимосвязи между переменными!

Зависимая переменная – собственно та, которая нас интересует (dependent variable).

Независимая – определяет нахождение в той или иной группе. В статистике – grouping variable = factor.

Двухвыборочные критерии.

Различаются ли *по массе* тигры-самцы и тигры-самки? Сравниваем средние массы наших зверьков.

Мы анализируем влияние пола на массу тигров.
Зависимая переменная – масса.
Независимая (группирующая) – пол (два уровня: 1. самцы; 2. самки).

самка

самец

Критерий Стьюдента для **независимых** выборок
(t-test for independent samples)

Общий вопрос: получены ли выборки из одной популяции?
Частный вопрос: равны ли **средние** значения между собой?

Сформулируем гипотезы:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Двухвыборочные критерии.

$$\text{Статистика} = \frac{\text{параметр выборки} - \text{параметр популяции}}{\text{стандартная ошибка параметра выборки}}$$

$$H_0 : \mu_1 = \mu_2 \quad \longrightarrow \quad H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2$$

разность выборочных средних

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \boxed{\frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}}$$

«ошибка»

$$\boxed{df = n_1 + n_2 - 2}$$

Ошибка считается из средних квадратов стандартных отклонений в выборках

Основное распределение - *t*-распределение (Стьюдента)

Это статистика для двустороннего критерия; про односторонний говорить не будем.

Двухвыборочные критерии.

Стандартная ошибка РАЗНОСТИ между средними:

$$s_{\overline{X}_1 - \overline{X}_2}$$

Взвешенная по размерам выборок средняя дисперсия

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s^2_{pooled}}{n_1} + \frac{s^2_{pooled}}{n_2}}$$

Идея в том, что стандартная ошибка разности средних определяется **дисперсиями** в обеих выборках и **размерами** этих выборок.

$$s^2_{pooled} = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

Двухвыборочные критерии.

От чего будет зависеть, отвергнем ли мы гипотезу H_0 или нет?

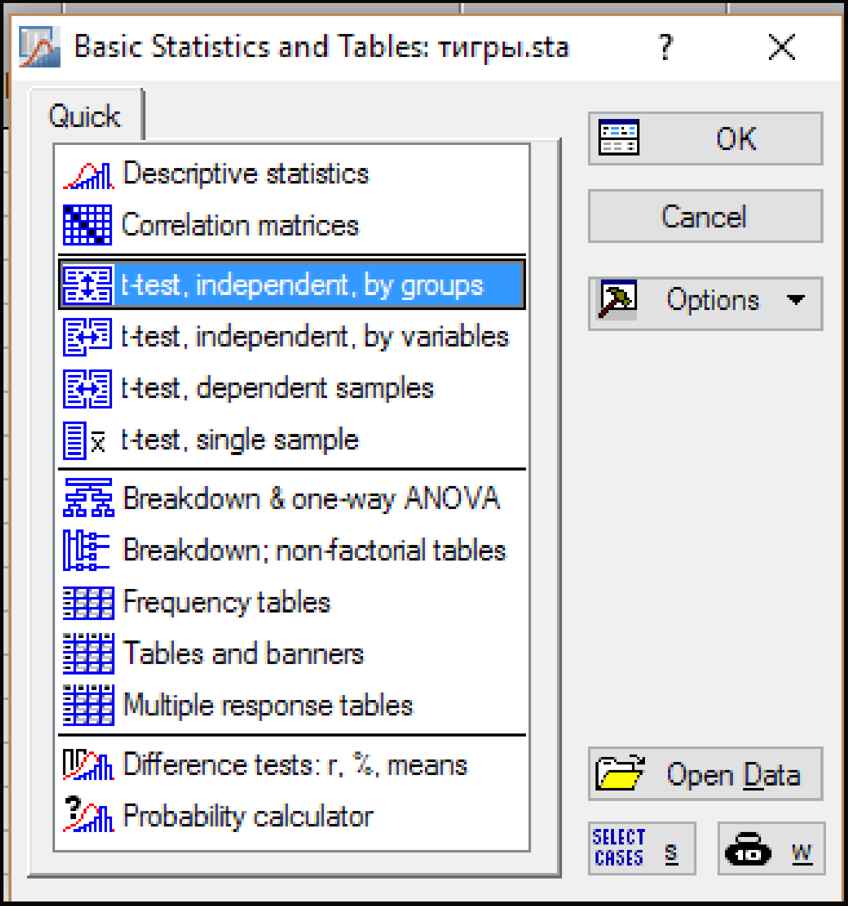
От того, насколько большое (по модулю) t у нас получится; а оно зависит от:

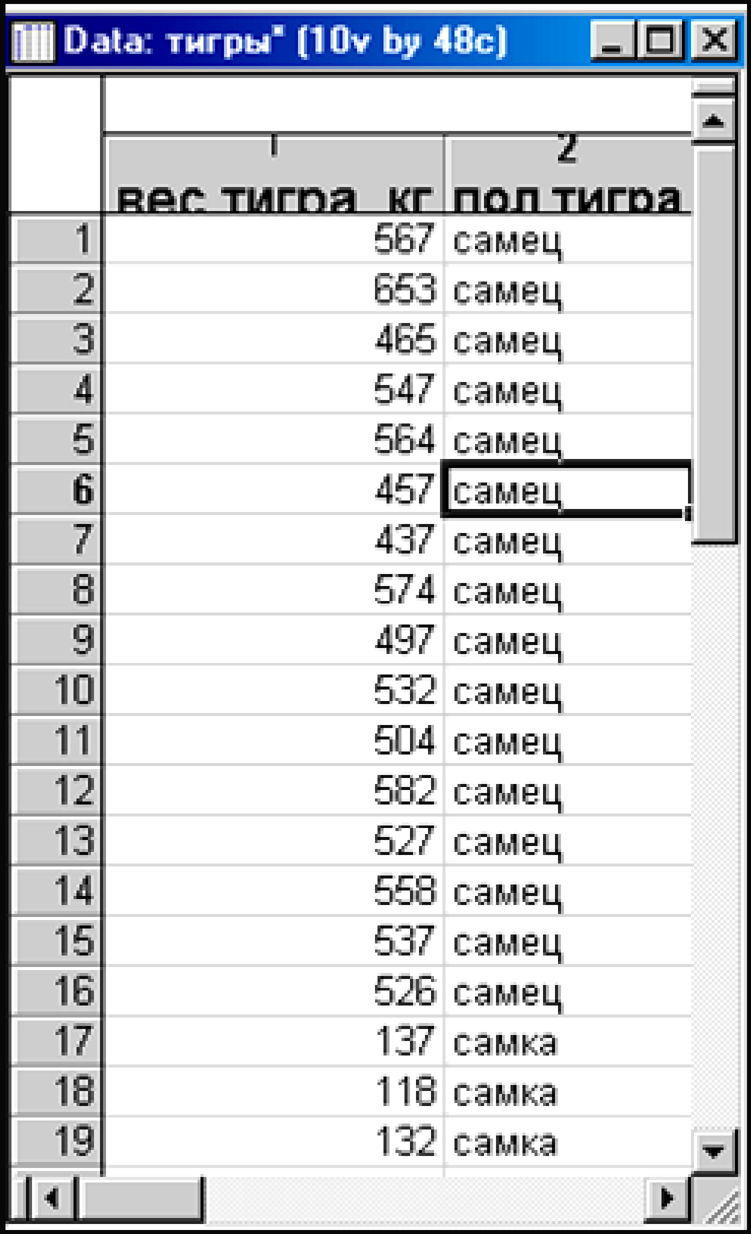
$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

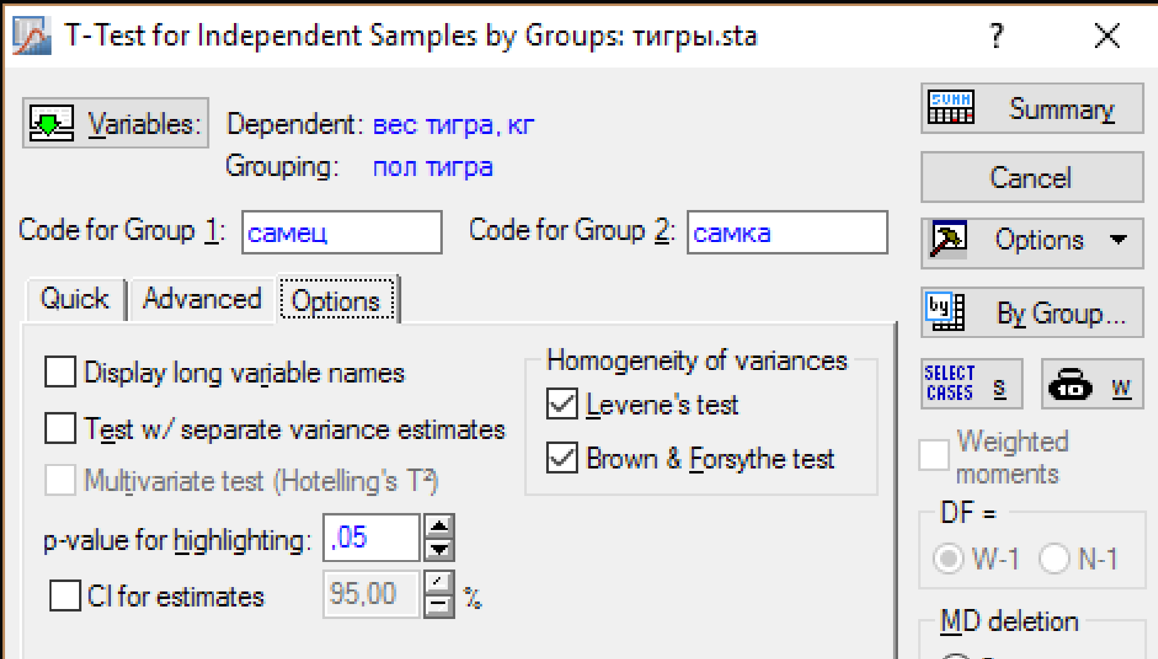
1. Непосредственно от различий между средними значениями;
2. От изменчивости внутри групп;
3. От размера выборок.

Двухвыборочные критерии.

t-test for independent samples







t-test for **independent** samples

ts; Grouping: пол тигра (тигры.sta)

T-tests; Grouping: пол тигра (тигры.sta)
Group 1: самец
Group 2: самка

Variable	Mean самец	Mean самка	t-value	df	p	Valid N самец	Valid N самка	Std.Dev. самец	Std.Dev. самка	F-ratio Variances	p Variances
вес тигра, кг	541,3125	132,4375	86,81657	30	0,00	16	16	15,00542	11,38987	1,735635	0,296639

Отвергаем H_0: средняя масса самцов и самок неодинакова

Levene F(1,df)	df Levene	p Levene	Brn-Fors F(1,df)	df Brn-Fors	p Brn-Fors
4,312109	30	0,046505	2,559995	30	0,120079

Двухвыборочные критерии.

Что ещё можно сравнить у 2-х выборок:

- ✓Медианы (занятие 6)
- ✓Дисперсии – вот это бывает действительно нужно;
- ✓Индексы разнообразия (занятие 6)
- ✓…

Двухвыборочные критерии.

Критерий Стьюдента для **связанных** выборок
(t-test for dependent samples)

К тиграм-самцам (в зоопарке) пришёл новый служитель, и возможно, они стали по-другому питаться. Мы хотим узнать, не изменилась ли их масса.

Мы анализируем влияние служителя на массу тигров-самцов.
Зависимая переменная – масса.
Независимая – уровни: 1. до нового служителя; 2. после)

Пример с левой и правой ногами

Двухвыборочные критерии.

Каждый тигр **два** раза участвует в наблюдениях: он входит в обе группы.

	ДО	ПОСЛЕ
1 тигр	356	363
2 тигр	351	361
3 тигр	353	358
4 тигр	355	356
5 тигр	354	359
6 тигр	355	355

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

Таких D столько, сколько пар.

Для этих разниц можно посчитать среднее.

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D \neq 0$$

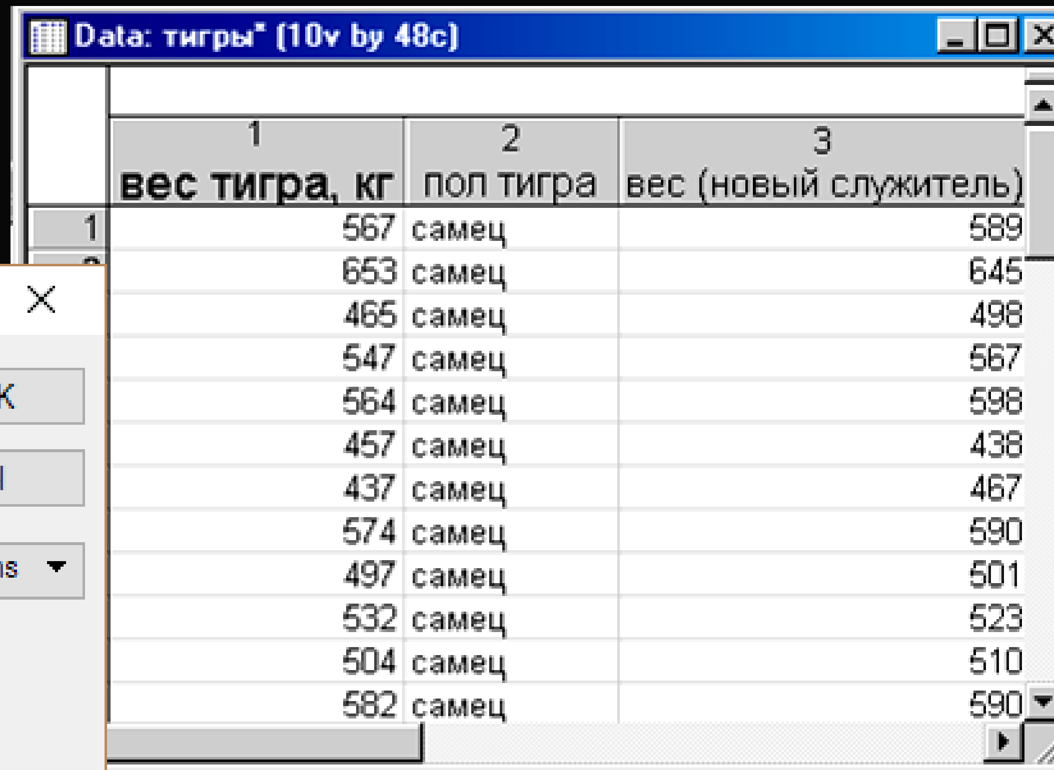
Идентично **одновыборочному *t*-критерию!**

Статистика: $$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}} \rightarrow \boxed{t = \frac{\overline{D}}{s_{\overline{D}}}}$$

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

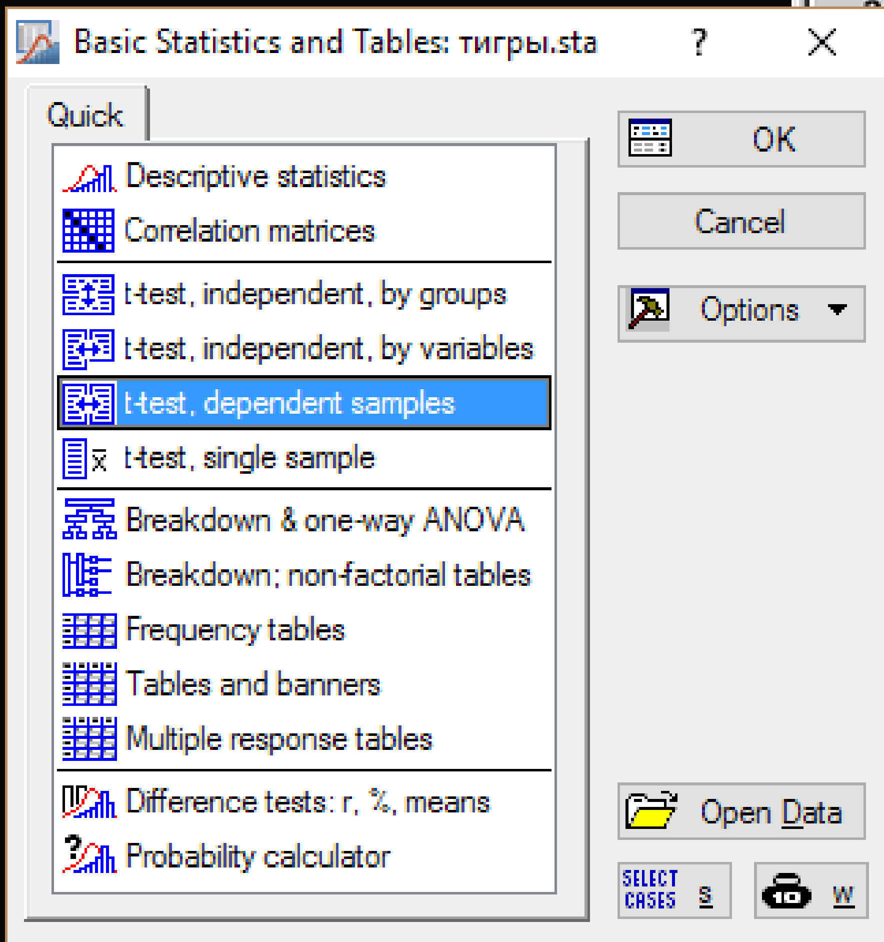
Двухвыборочные критерии.

t-test for dependent samples

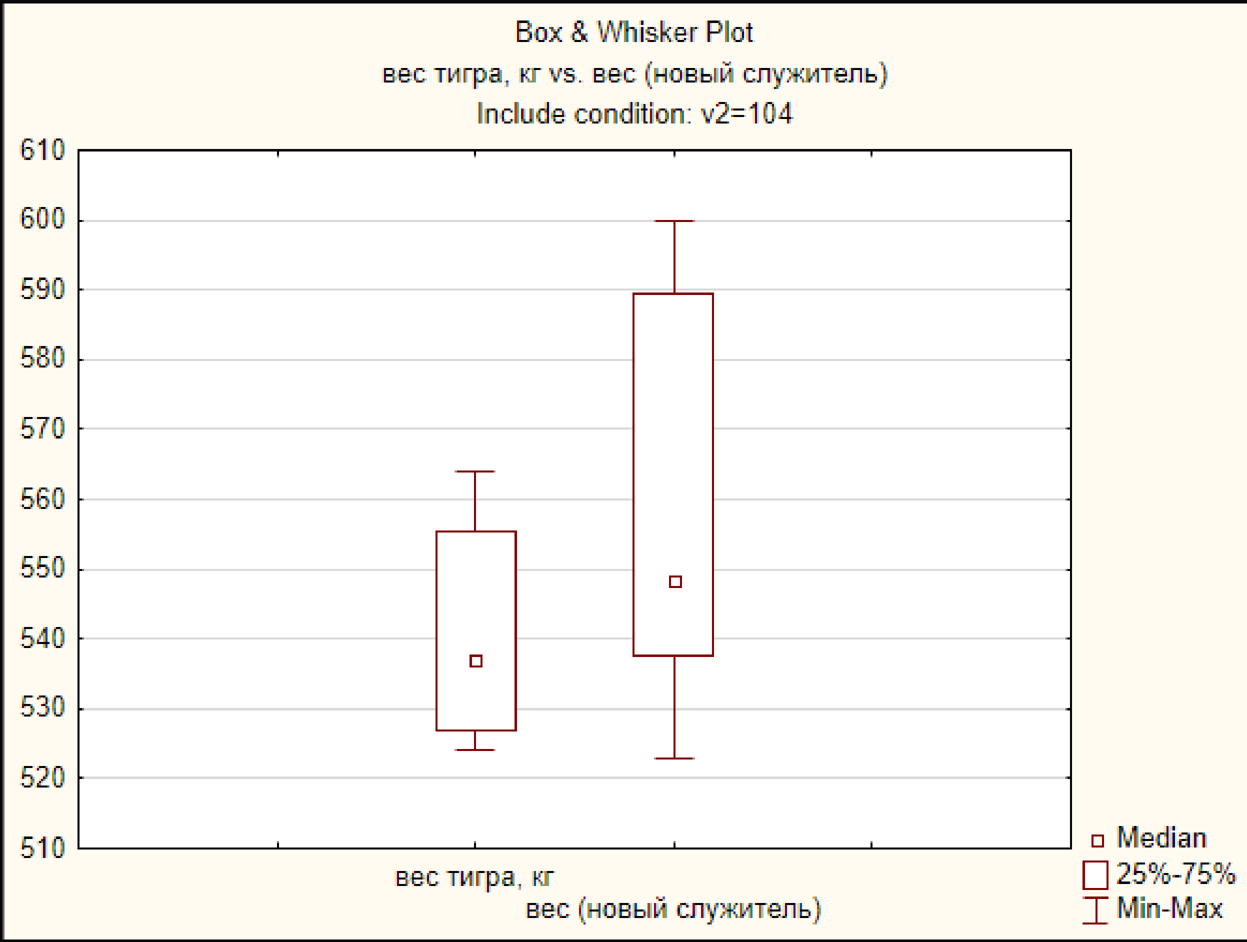


Data: тигры* (10v by 48c)

	1 вес тигра, кг	2 пол тигра	3 вес (новый служитель)
1	567	самец	589
	653	самец	645
	465	самец	498
	547	самец	567
	564	самец	598
	457	самец	438
	437	самец	467
	574	самец	590
	497	самец	501
	532	самец	523
	504	самец	510
	582	самец	590



t-test for dependent samples



Отвергаем H_0: Масса тигров в среднем достоверно увеличилась после прихода нового служителя.

T-test for Dependent Samples (тигры.sta)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
вес тигра, кг	541,3125	15,00542								
вес (новый служитель)	558,1875	26,41772	16	-16,8750	22,69471	-2,97426	15	0,009455	-28,9682	-4,78184

Двухвыборочные критерии.

В принципе, можно использовать тест для независимых выборок и для связанных выборок.

Но мы рискуем **не увидеть** существующих различий, особенно при **большой изменчивости** в выборках!

Тесты для связанных выборок как раз для того и существуют, чтобы исключить из анализа внутригрупповую изменчивость.

Условия проведения критериев Стьюдента (assumptions)

Соотношение площадей под распределением статистики критерия может быть известно только если это распределение имеет определённую форму. А оно зависит от распределения в исследуемых популяциях!

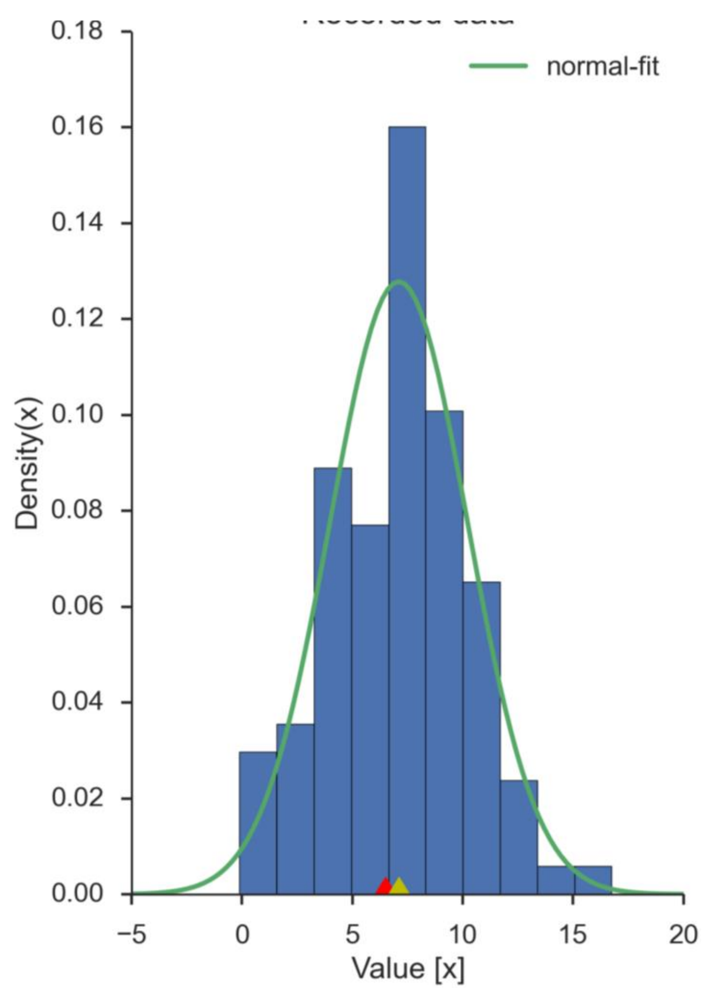
1. Выборки должны быть получены из популяций с нормальным распределением (небольшие отклонения допустимы, кроме сильно скошенных и мультимодальных распределений).
2. Для двухвыборочных тестов: популяции, из которых получены выборки должны иметь равные дисперсии (гомогенность дисперсий). Это очень важно! Особенно при разных размерах выборок; совсем плохо, если дисперсия больше в меньшей выборке. Разница в 2-3 раза - плохо.
3. Очень нежелательны аутлаеры.
4. Ограничение на размер выборки: $N \geq 10$ в каждой группе.
5. Надо бороться за равенство размеров выборок!
6. Выборки обязательно должны быть случайными.

Условия проведения критериев Стьюдента (assumptions)

Если условия нарушены, соотношения площадей будут другими, и границы критических областей сдвинутся непонятно куда – непредсказуемо изменится вероятность ошибки 1-го рода (Quinn, Keough, 2002; Sokal, Rohlf, 2009; Zar, 2010).

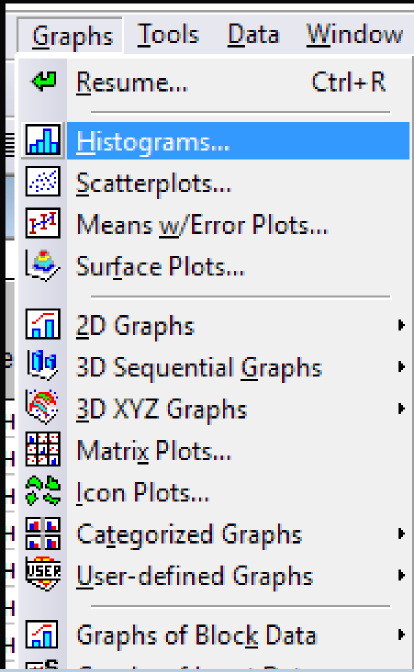
Что делать?

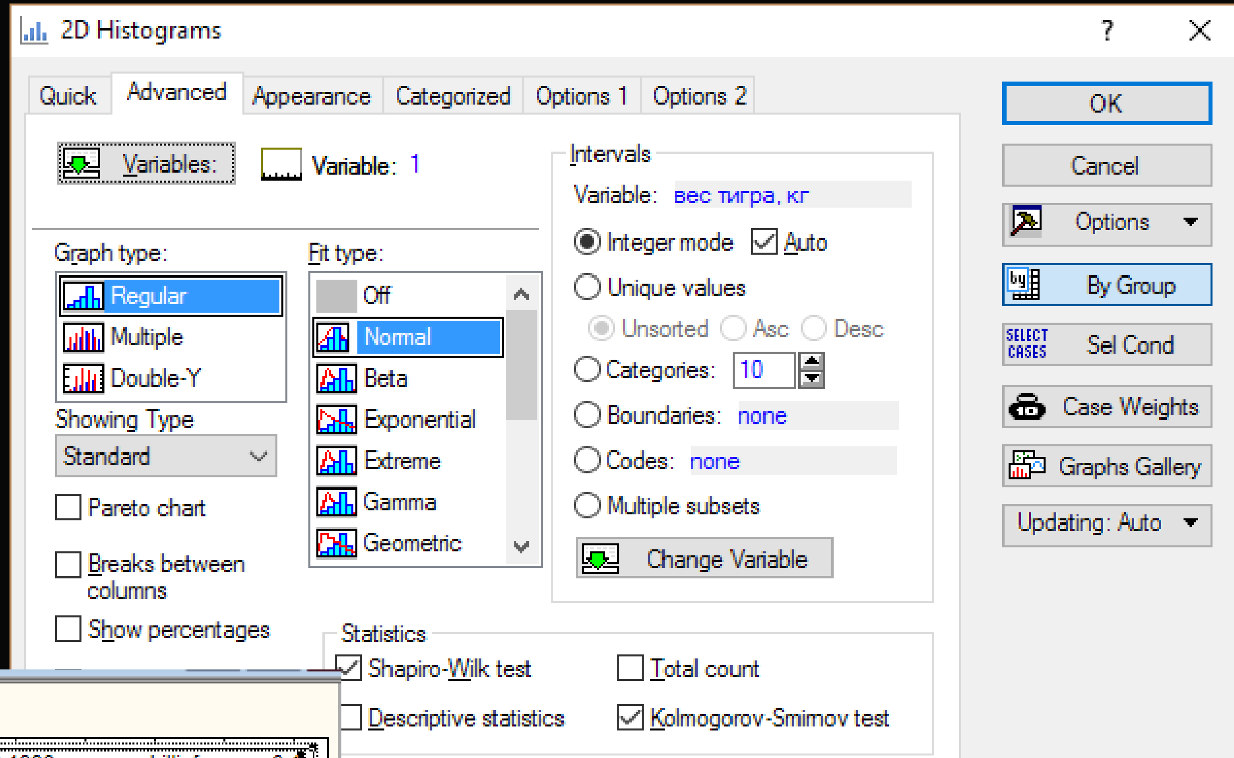
1. Обязательно построить 2 картинки - гистограмму и «усатый ящик». Глазами оценить: симметрию; наличие аутлаеров; равенство дисперсий.

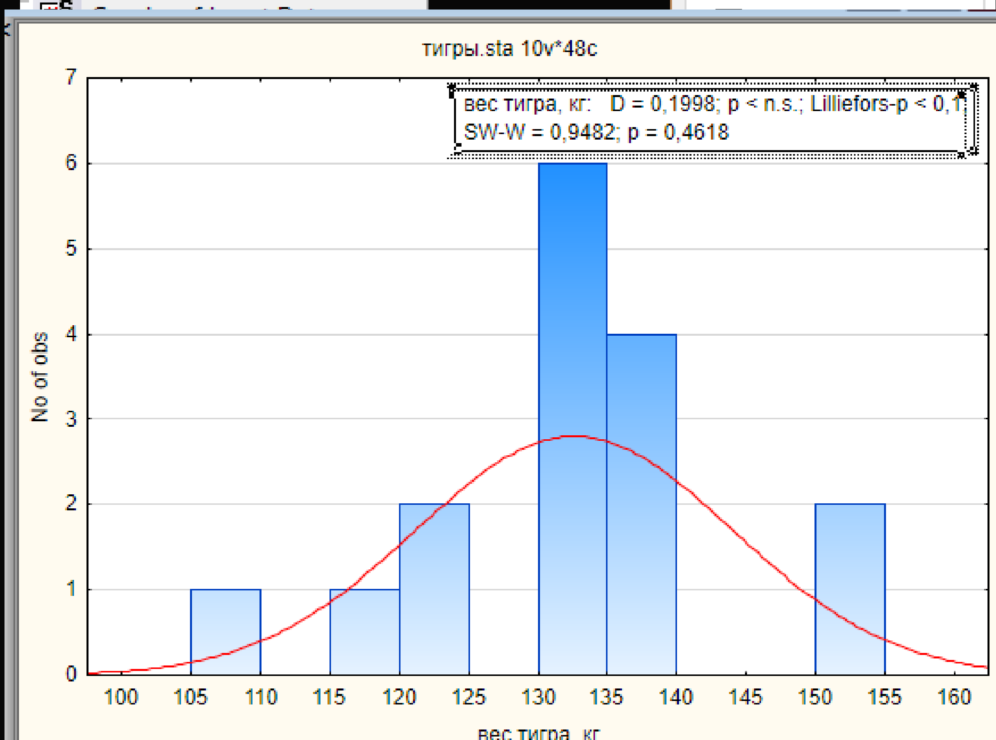


Условия проведения критериев Стьюдента (assumptions)

2. Проверить «нормальность» и гомогенность специальными тестами (но: Quinn, Keough, 2002)
3. Трансформация данных «лечит» часто и «ненормальность», и гетерогенность.
4. Бороться за равенство размеров выборок!
5. Чем больше N, тем менее страшна «ненормальность» (от 30 в каждой группе – повод не очень волноваться о ней).
6. Существуют непараметрические аналоги критериев Стьюдента, но они не спасают от аутлаеров и плохо спасают от неравенства дисперсий.
7. *Небольшое отклонение от какого-нибудь из требований компенсируется соблюдением остальных!*

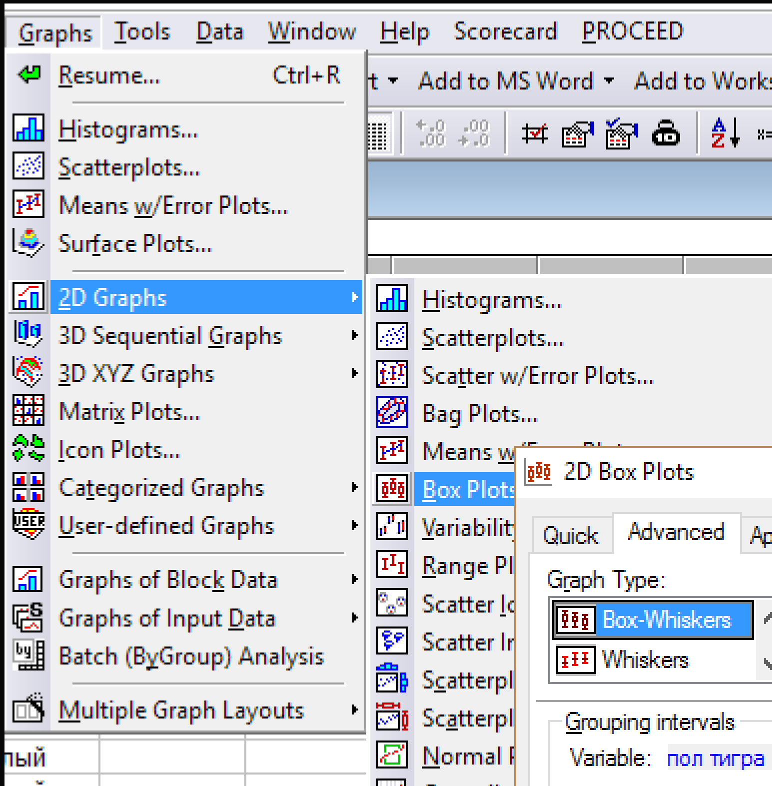


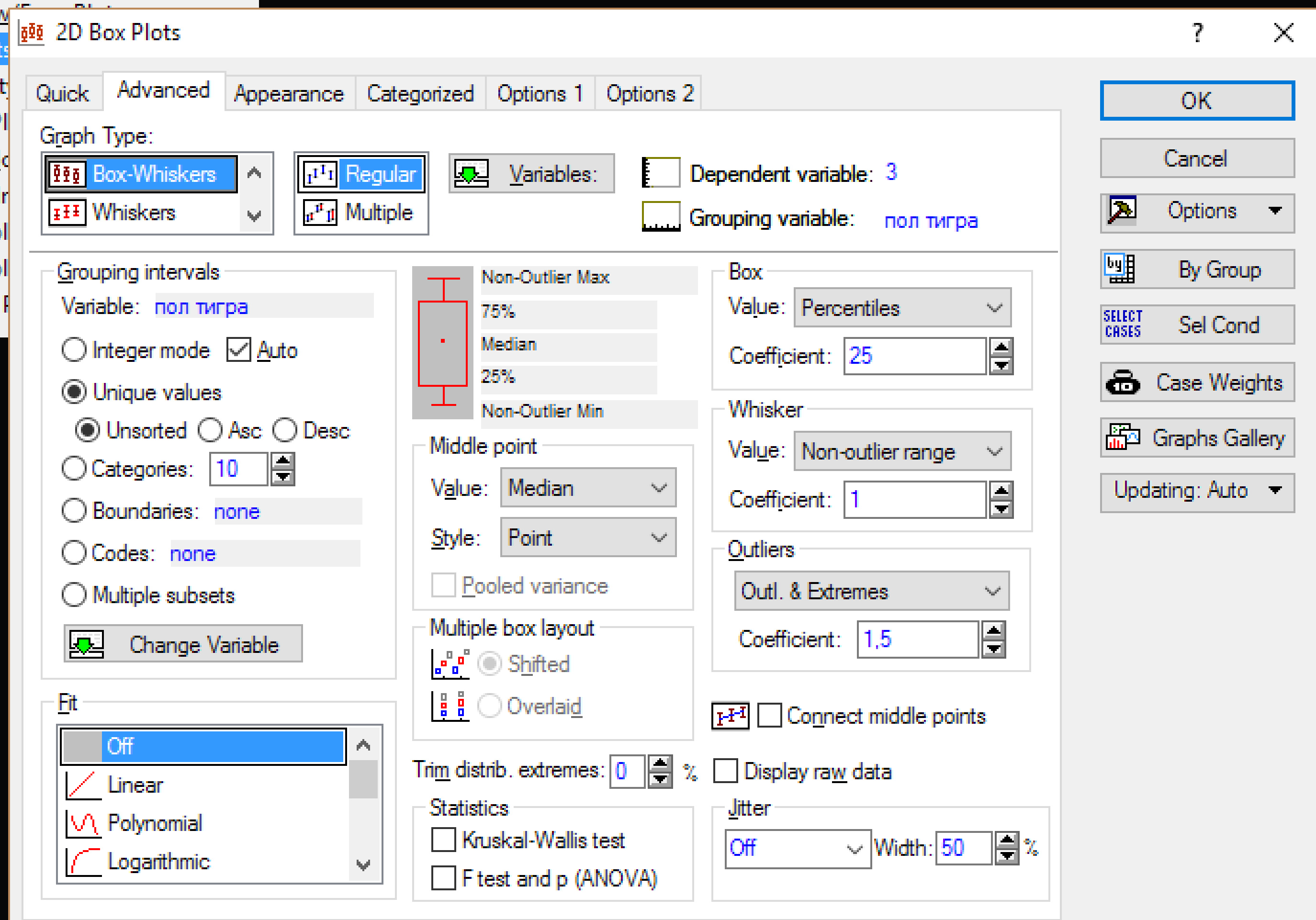




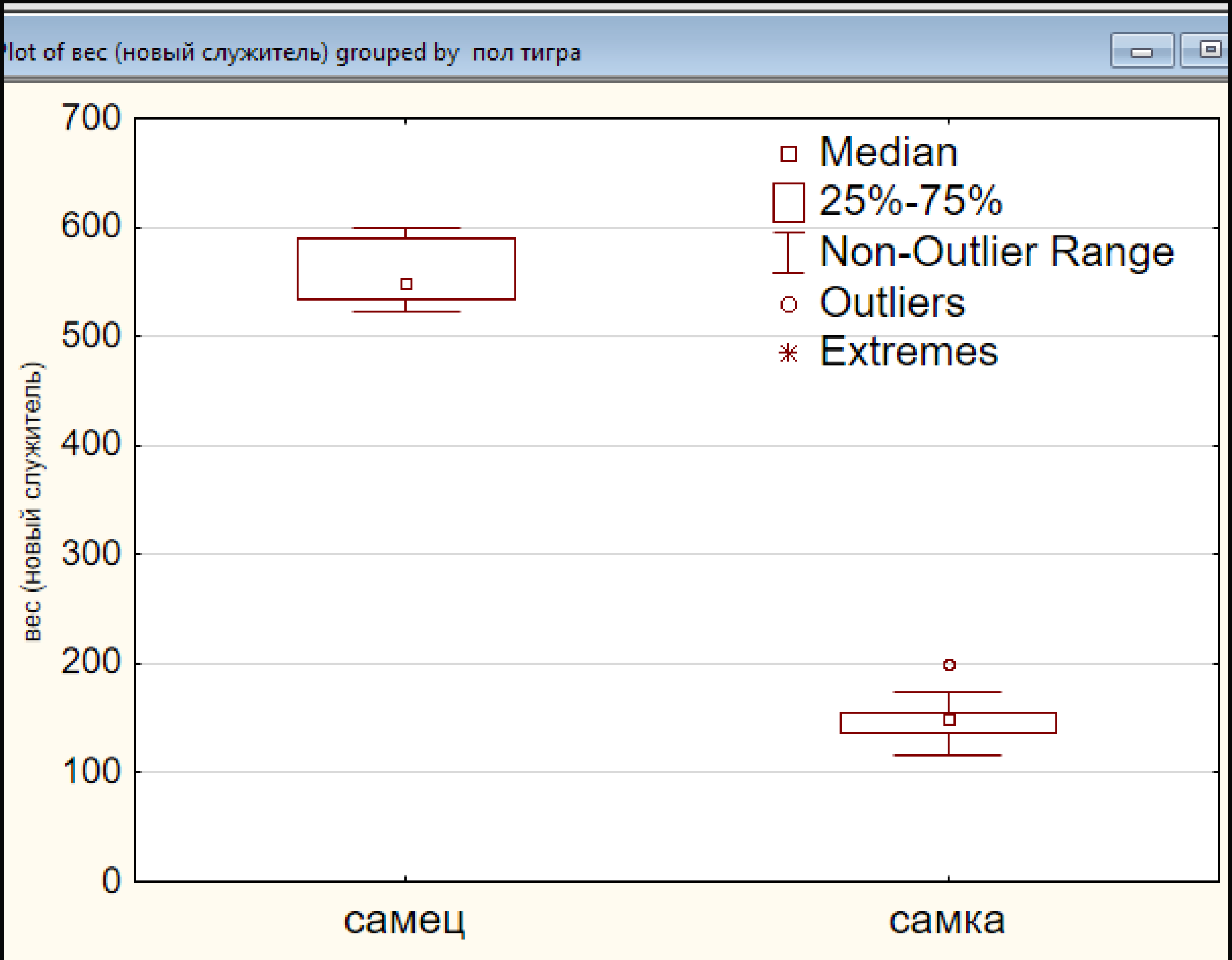
Графическое исследование данных: гистограмма

Графическое исследование данных: усатый ящик





Графическое исследование данных: усатый ящик



Аутлаеры (outliers) – значения, слишком отличающиеся от остальных. В разных книгах/программах это «слишком» - разное. Традиционно – отстоят более чем на 1.5 интерквартильных размаха от границы интерквартильного размаха (той или иной квартили).

Для публикаций

В секции **методов**:

1. Выборки удовлетворяли критериям нормального распределения и то, каким тестом это было установлено «… *all variables conformed to the assumptions of normality (Shapiro-Wilk's W test, p>0.05*»);
2. Выборки удовлетворяли условиям гомогенности дисперсий, и то, каким тестом это было установлено "*For all parametrical tests, the samples were homoscedastic (Levene's test, p>0.05*)".
3. Какой критерий использовали: *«we used Student's t-test for two-group comparisons»*
4. Односторонний или двусторонний тест использовался, и если односторонний – должно быть обоснование.

В **результатах**:

t, *df* (N), *p*, effect size index (*d*)

The body mass in females was smaller than in males (445 ± 93 g, N=58 vs 803 ± 125 g, N=56; t = 11.1, P < 0.0001).

Вопросы

1. охарактеризовать распределения:
   - ✓ Длины травинок на только что покошенном газоне
   - ✓ Массы детёнышей при рождении
   - ✓ Массы всех зверьков в популяции

2. В институте каждый год проводят экзамен по статистике, и средний результат = 80 баллов. Преподаватель решил удвоить количество домашних заданий и посмотреть, как это скажется на результатах учеников на экзамене.
   - ✓ Какими будут нулевая и альтернативная гипотезы?
   - ✓ Альтернатива в тесте односторонняя или двусторонняя?
   - ✓ Предположим, учитель проводит статистический анализ и отвергает нулевую гипотезу. Возможна ли при этом ошибка 1-го рода? 2-го рода? Что будут представлять из себя эти ошибки в данном случае?

2. Д-р Симонс решил узнать, как утренняя зарядка влияет на самочувствие людей. Он померил ЧСС у 52 человек и заставил их 8 недель каждое утро делать зарядку, после чего померил ЧСС вновь.

- ✓Какими будут нулевая и альтернативная гипотезы?
- ✓Какой статистический тест доктор будет использовать для анализа?

3. зоолог обнаружил две изолированные популяции белок – северную и южную. Ему кажется, что в северной популяции белки крупнее (различается их масса). Он хочет проверить своё предположение статистически.

- ✓Какими будут нулевая и альтернативная гипотезы?
- ✓Какой статистический тест учёный будет использовать для анализа?

Примерные темы занятий (2016 -2017)

1. Основные понятия. Описательная статистика
2. Тестирование гипотез в статистике. Критерии Стьюдента
3. Критика тестирования гипотез. Мощность статистического теста. Размер эффекта.
4. Дисперсионный анализ ANOVA
5. Дисперсионный анализ ANOVA (продолжение)
6. Корреляции. Регрессионный анализ
7. Трансформация данных. Непараметрические критерии.
8. Частотный анализ.
9. Основы многомерных методов анализа. Дискриминантный анализ
10. Факторный анализ.
11. Многомерное шкалирование. Кластерный анализ
12. Обобщённые линейные модели и модели смешанных эффектов.