

Методы организации, планирования и обработки результатов инженерного эксперимента

Всего 144 ч., ауд. – 18 ч., самостоятельная работа - 126 ч.
Зачёт

Стукач Олег Владимирович

Каф. СУМ, пр. Ленина, д. 2, оф. 204

 (3822)-701777*2754

tomsk@ieee.org

<http://ieee.tpu.ru/pages/stukach.htm>

<http://portal.main.tpu.ru:7777/SHARED/o/OS>



Цель и задачи

- ☺ Дать обзор статистических методов, реализованных в программных комплексах «STATISTICA», «Labview», относящихся к инженерному эксперименту
 - ☺ Сформировать у слушателей понимание методов разведочного анализа данных, необходимых им для дальнейшей работы над своими задачами
 - ☺ Получить навыки анализа данных
 - ☺ Решить полностью или частично подзадачу из своей квалификационной работы
-

Статистическое мышление - это:

- 🚩 Способность мыслить системными категориями
- 🚩 Понимание того, что все процессы для нас носят статистический, вероятностный характер в силу ограниченности восприятия информации человеком
- 🚩 Знание методов математической статистики
- 🚩 Умение пользоваться статистическими методами в своей профессиональной деятельности
- 🚩 Умение оценивать риски в принятии решений, особенно управленческого характера
- 🚩 Способность осознавать себя частью целого и не замыкаться на частностях

Вы обладаете статистическим мышлением, если:

- ☺ Доверяете только фактам, а не абстрактным идеям
- ☺ Выражаете факты только в количественной форме, а не с помощью словесных внушений
- ☺ Осознаёте, что наблюдения ведутся над частью целого и их результаты всегда содержат ошибки и погрешности
- ☺ Принятые вами решения на основе статистических данных научно обоснованы

Вероятностные распределения

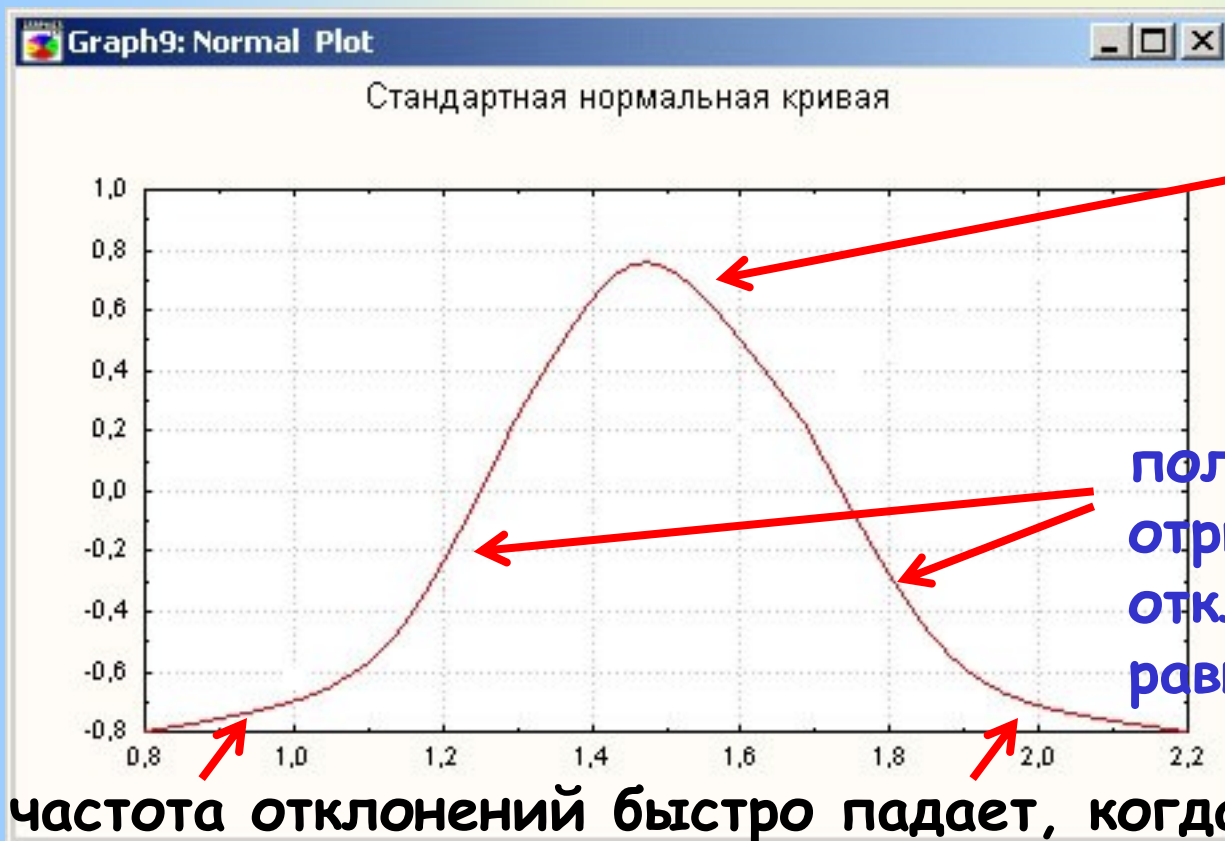
Если вы имеете некоторую переменную, принимающую значения случайным образом, то вам хотелось бы знать, с какими вероятностями эта переменная принимает определенные значения. Совокупность этих вероятностей как раз и задает распределение вероятностей.

Рассмотрим вероятностные законы, наиболее часто возникающие на практике.

Нормальное распределение

Нормальное распределение вероятностей особенно часто используется в статистике.

Нормальное распределение дает хорошую модель для реальных явлений, в которых:



имеется сильная тенденция данных группироваться вокруг центра

положительные и отрицательные отклонения от центра равновероятны

частота отклонений быстро падает, когда отклонения от центра становятся большими

Нормальное распределение

Характерное свойство нормального распределения состоит в том, что 68% из всех его наблюдений лежат в диапазоне 1 (стандартное отклонение от среднего), а диапазон 3 стандартных отклонений включает 99% значений. Другими словами, при нормальном распределении стандартизованные наблюдения, меньшие -3 или большие +3, имеют относительную частоту менее 1% (стандартизованное наблюдение означает, что из исходного значения вычтено среднее и результат поделен на стандартное отклонение.) Это и есть знаменитое правило 3-сигма, чрезвычайно популярное на практике.

Плотность нормального распределения определяется следующим образом:

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

где a -среднее;

σ -стандартное отклонение.

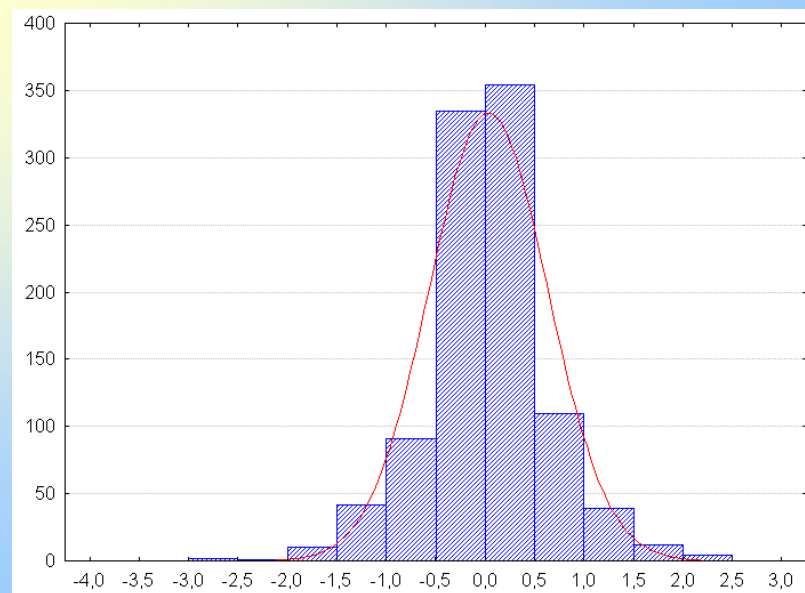
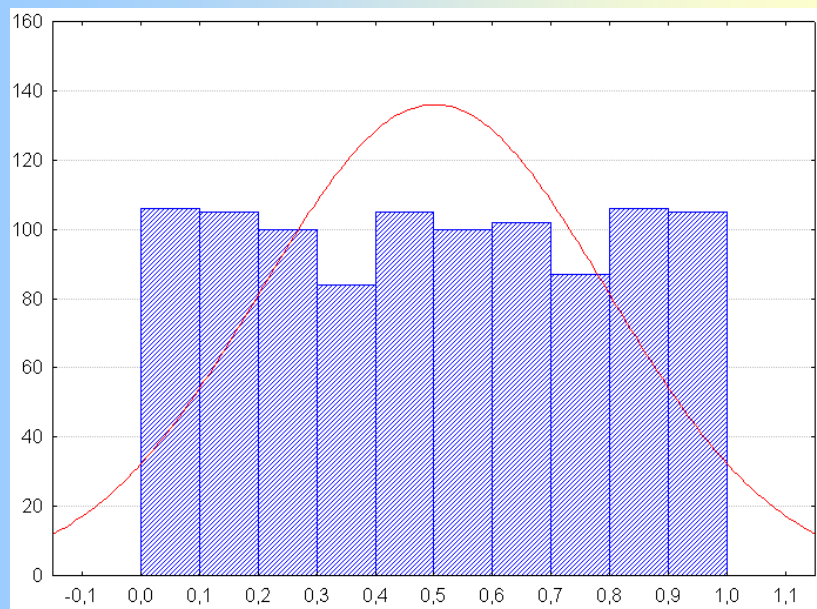
Нормальное распределение

При суммировании большого числа независимых величин получаются именно нормально распределенные величины. И это происходит независимо от распределения исходных величин. Иными словами, если на некоторую переменную воздействует множество факторов, эти воздействия независимы, относительно малы и слагаются друг с другом, то получаемая в итоге величина имеет нормальное распределение.

Множество величин на практике имеют нормальное распределение, например, распределение приращений индексов развитых стран, курсы акций и т. д.

Нормальное распределение

Нормальное распределение (кривая Гаусса) важно по многим причинам. Распределение большого числа статистик является нормальным или может быть получено из нормального с помощью некоторых преобразований.



Нормальное распределение

Variable 1

Font: Arial, Size: 10, Bold, Italic, Underline, x₂, x²

Name: Var1 Type: Double OK

MD code: -9999 Length: 8 Cancel

Display format

- General
- Number
- Date
- Time
- Scientific
- Currency
- Percentage
- Fraction
- Custom

Long name (label or formula with Functions): Function guide

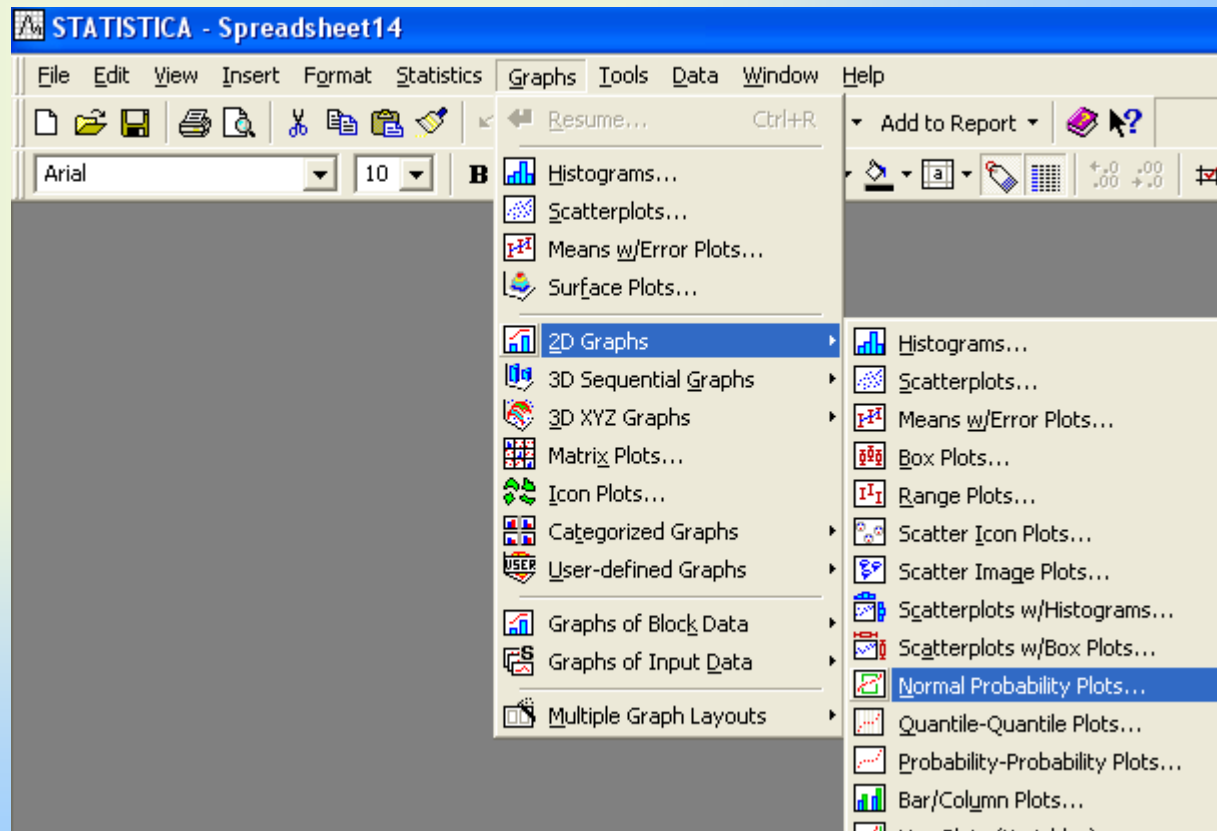
=VNormal(Rnd(1);0;1)

Labels: use any text. Formulas: use variable names or v1, v2, ..., v0 is case #.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after:)

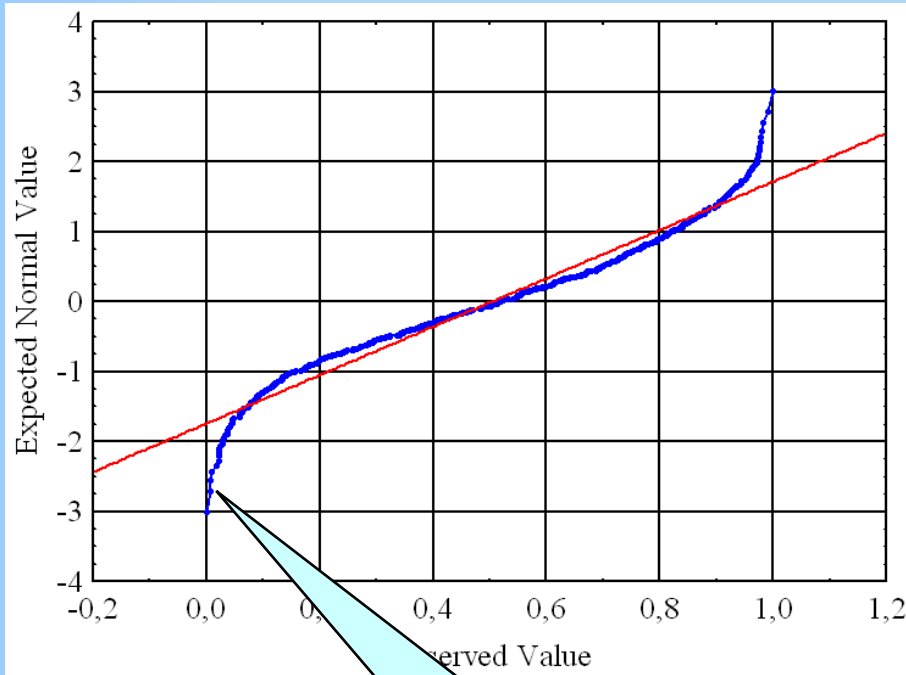
Можно сгенерировать нормально распределенные случайные числа в STATISTICA, используя датчик равномерно распределенных случайных чисел RND(1) и обратную нормальную функцию распределения VNormal

Нормальное распределение

Еще один простой и эффектный способ убедиться в нормальности распределения значений переменной



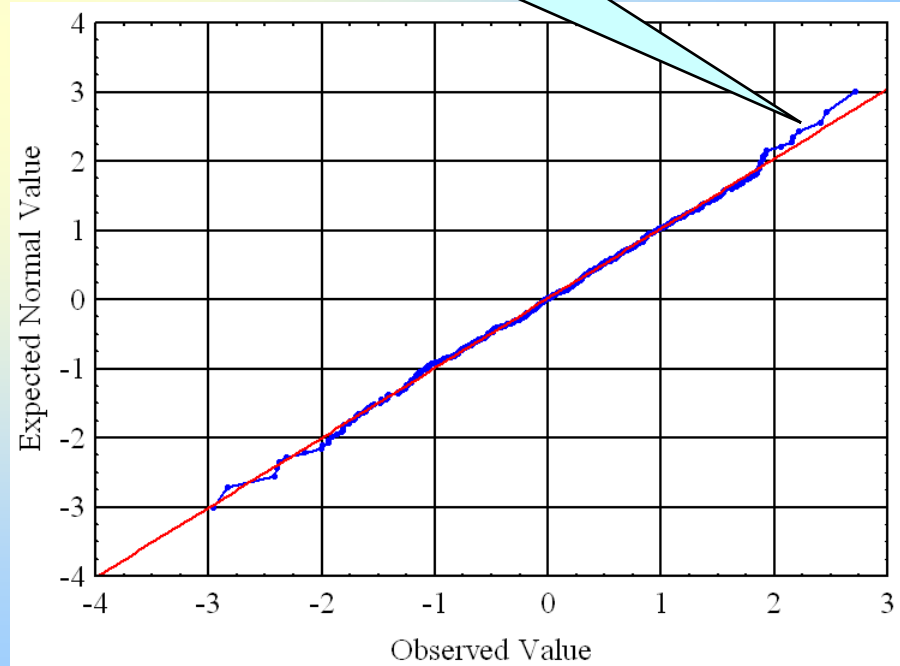
Вероятностный график



Равномерное
распределение

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Нормальное
распределение



Проверка нормальности распределения

Statistics Graphs Tools Data Window

Resume... Ctrl+R

Basic Statistics/Tables

Multiple Regression

ANOVA

Nonparametrics

Distribution Fitting

Advanced Linear/Nonlinear Models ▶

Multivariate Exploratory Techniques ▶

Industrial Statistics & Six Sigma ▶

Power Analysis

Data-Mining ▶

Statistics of Block Data ▶

STATISTICA Visual Basic

Probability Calculator ▶

Fitting Continuous Distributions: gipoteza.sta

Distribution: Normal ▼

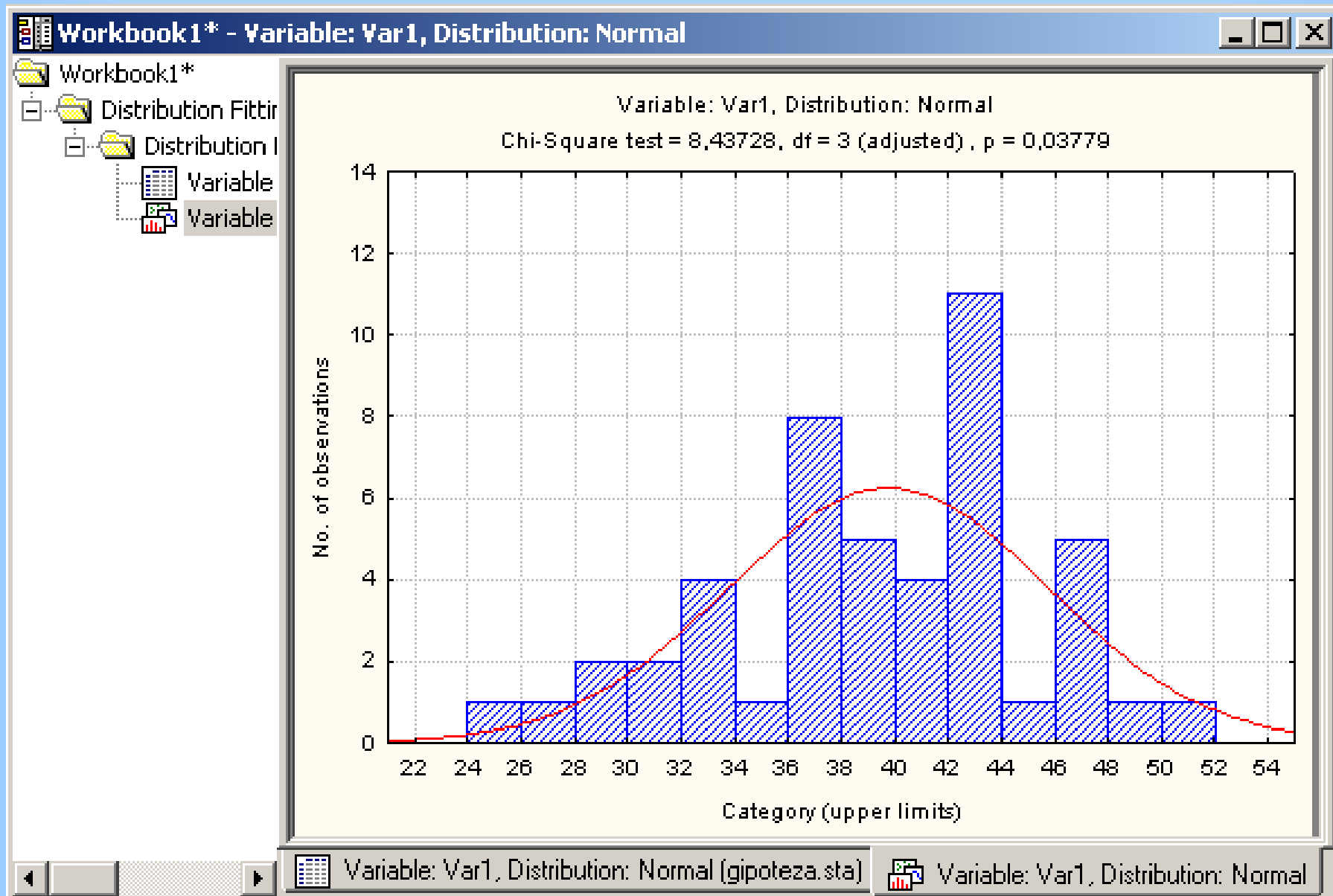
Variable: Var1

Quick Parameters Options

Summary: Observed and expected distribution

Plot of observed and expected distribution

Проверка нормальности распределения



Проверка нормальности распределения

Variable: Var1, Distribution: Normal (gipoteza.sta) Chi-Square = 8,43728, df = 3 (adjusted) , p = 0,03779										
Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected	
<= 24,00000	0	0	0,00000	0,0000	0,205079	0,20508	0,43634	0,4363	-0,20508	
26,00000	1	1	2,12766	2,1277	0,312575	0,51765	0,66505	1,1014	0,68743	
28,00000	1	2	2,12766	4,2553	0,664961	1,18261	1,41481	2,5162	0,33504	
30,00000	2	4	4,25532	8,5106	1,267655	2,45027	2,69714	5,2133	0,73234	
32,00000	2	6	4,25532	12,7660	2,165568	4,61584	4,60759	9,8209	-0,16557	
34,00000	4	10	8,51064	21,2766	3,315207	7,93105	7,05363	16,8746	0,68479	
36,00000	1	11	2,12766	23,4043	4,547992	12,47904	9,67658	26,5511	-3,54799	
38,00000	8	19	17,02128	40,4255	5,591140	18,07018	11,89604	38,4472	2,40886	
40,00000	5	24	10,63830	51,0638	6,159618	24,22979	13,10557	51,5528	-1,15962	
42,00000	4	28	8,51064	59,5745	6,081076	30,31087	12,93846	64,4912	-2,08108	
44,00000	11	39	23,40426	82,9787	5,379975	35,69085	11,44676	75,9380	5,62002	
46,00000	1	40	2,12766	85,1064	4,265330	39,95618	9,07517	85,0131	-3,26533	
48,00000	5	45	10,63830	95,7447	3,030375	42,98655	6,44761	91,4607	1,96962	
50,00000	1	46	2,12766	97,8723	1,929345	44,91590	4,10499	95,5657	-0,92935	
52,00000	1	47	2,12766	100,0000	1,100756	46,01665	2,34203	97,9078	-0,10076	
< Infinity	0	47	0,00000	100,0000	0,983347	47,00000	2,09223	100,0000	-0,98335	



[English](#) [Конференции](#) [О группе](#) [Курсы](#) [Книги](#) [Статьи](#) [Фотографии](#) [Вступить](#) [Контакты](#) [Домой](#)

[Главная страница публикаций](#)

[Информация об Институте IEEE](#)

[Деятельность Института IEEE](#)

[Отчёты о конференциях в Томске и его окрестностях](#)

[Креатив, поездки и мероприятия](#)

[Публикации, напоминающие научные](#)

[Всякое разное](#)

[ХУмор](#)

[Шаблон для вёрстки статей в сборниках конференций IEEE в Томске, PDF](#)

[Шаблон для вёрстки статей в сборниках конференций IEEE в Томске, DOC in ZIP](#)

Публикации, напоминающие научные

«Scientific Write and Hike»

- 🔗 [Американский изобретатель](#)
- 🔗 [Аспирантский нетворкинг](#)
- 🔗 [Девяностый хеппибэзник отрицательной обратной связи в усилителях](#)
- 🔗 [Да здравствуют конференции](#)
- 🔗 [Как писать аннотацию](#)
- 🔗 [Как писать диссертацию](#)
- 🔗 [Как написать статью, напоминающую научную \(чтение в постели или на заседаниях для тех, кто не может заснуть\)](#)
- 🔗 [Заочное участие в конференции](#)
- 🔗 [Навыки выживания при написании статей, напоминающих научные](#)
- 🔗 [Нетворкинг как в 1959-м](#)
- 🔗 [ПАТ-Инфо и Патинформбюро](#)
- 🔗 [Про квест](#)
- 🔗 [Работа и диссертация](#)
- 🔗 [Регистрационный взнос](#)
- 🔗 [Учёная степень и творческая активность](#)
- 🔗 [Nuts ans bolts of IEEEExplore](#)

Нанисты

- [Пизастор таинственный](#)

Томская группа и студенческое отделение IEEE помогает с приобретением и поиском статей обществ и конференций IEEE, находящимся в сфере её профессиональной деятельности. За отдельную небольшую плату возможна запись на компакт-диски и поиск статей по ключевому слову, а также ретроспективный поиск на глубину до пяти лет. В наличии имеются материалы обществ теории и техники СВЧ, связи, образования. Заявки направляйте по e-mail tomsk@ieeetpu.org.



ДИССЕРТАРИУМ

ИНФОХЕЛП

ПОДСОБКА

ЧТИВО!

МИМОЛЁТНОЕ

УСЛУГИ

КТО ЕСТЬ WHO



Диссертариум

Диссертариум

В этой рубрике размещаются материалы прямо или косвенно связанные с диссертацией, как основным продуктом жизнедеятельности аспирантов.

Жорес Иванович «жжот»

(Или Lewis Carroll о "защищающихся")

автор: Карнышев В.И. | 10.10.2016



Собираясь сегодня утром на работу, в четверть-уха слушал теле-интервью с Жоресом Ивановичем Алфёровым. ...

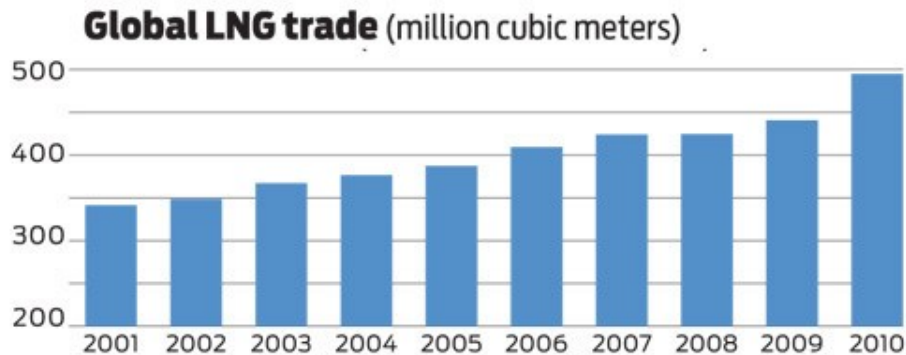
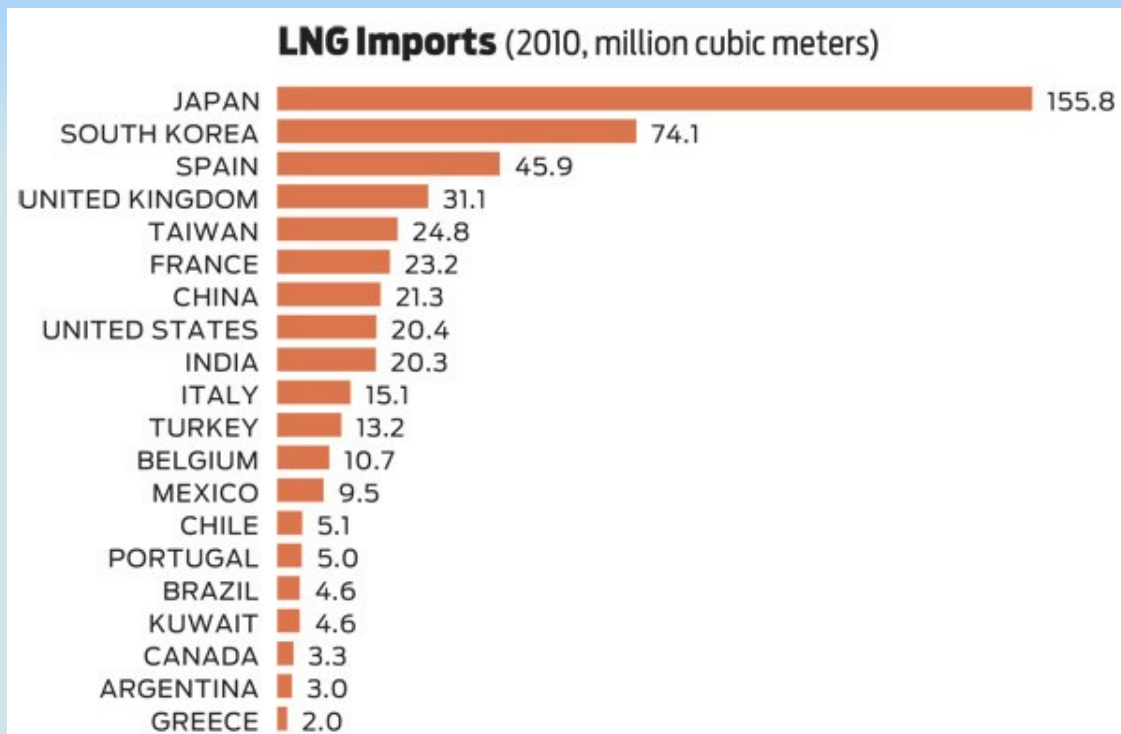
Поддержать сайт

Слабо?

Путеводитель по сайту

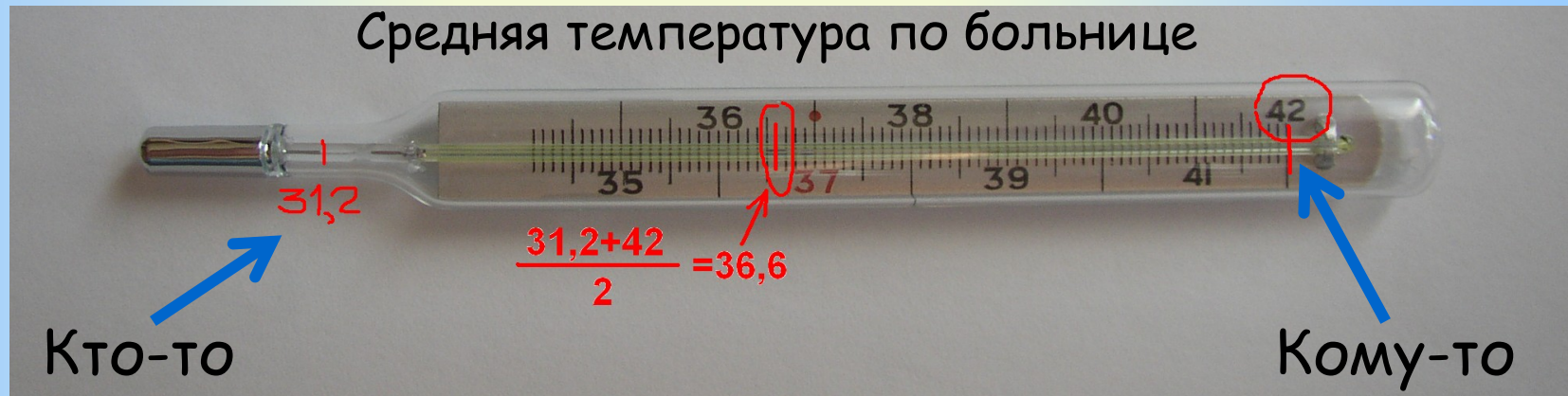
[Открыть](#)

Технология статистического обмана



Технология статистического обмана

Отличительной особенностью средних является то, что в них взаимно погашаются и уничтожаются индивидуальные отклонения различающихся между собой величин одного и того же признака.



ТИХО
ОСТЫВАЕТ

А в среднем - все
здоровы, лечения
не требуется

очень
плохо

Оценка по средним значениям признаков имеет недостатки, которые связаны с потерей значимой для управления информации

«Средняя зарплата»



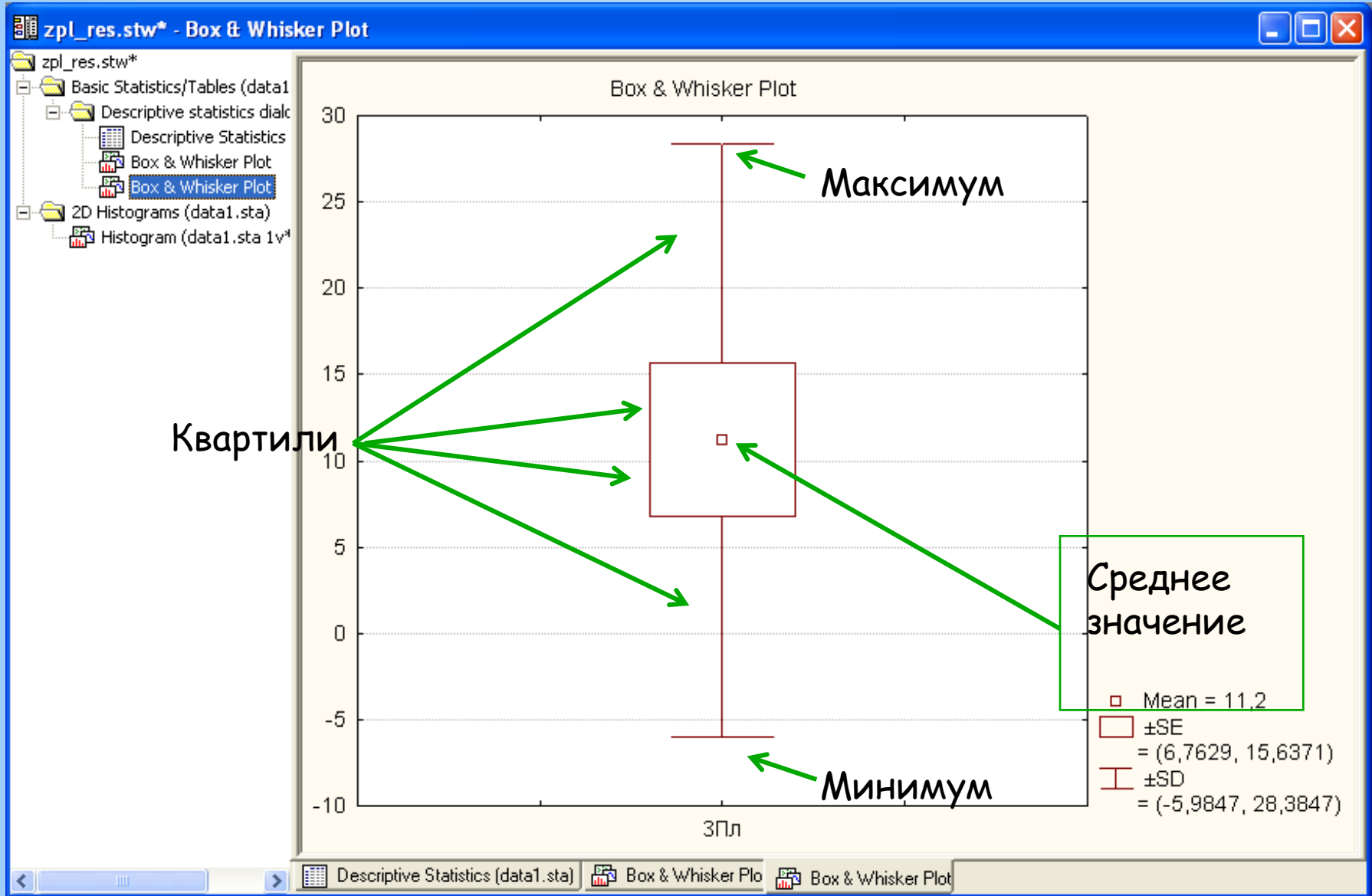
	1 ЗПл
1	3
2	2
3	3
4	4
5	3
6	2
7	4
8	5
9	6
10	7
11	9
12	8
13	10
14	37
15	65

Низкий доход

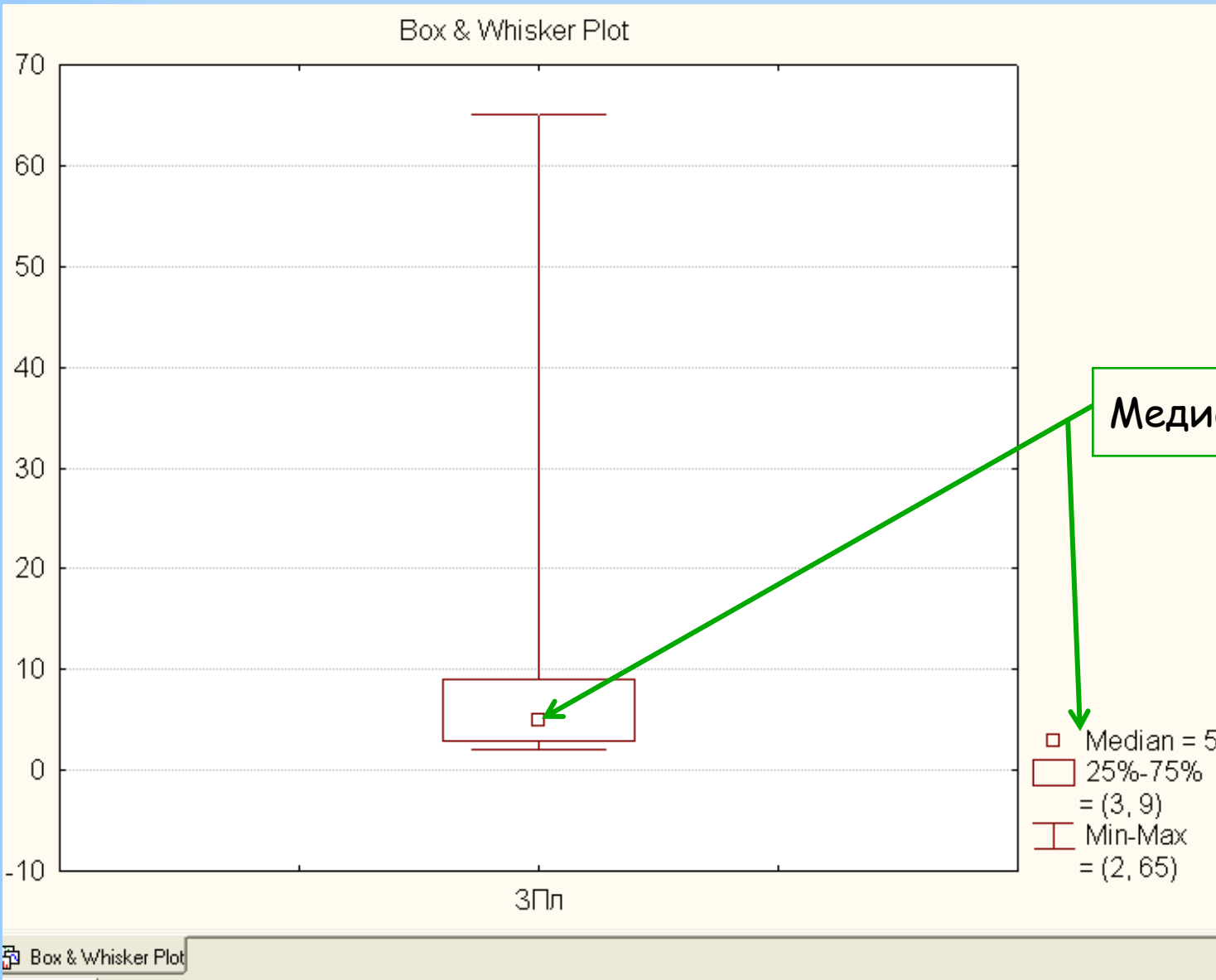
Средний
доход

Высокий
доход

Боксплот, ящик с усами (среднее)



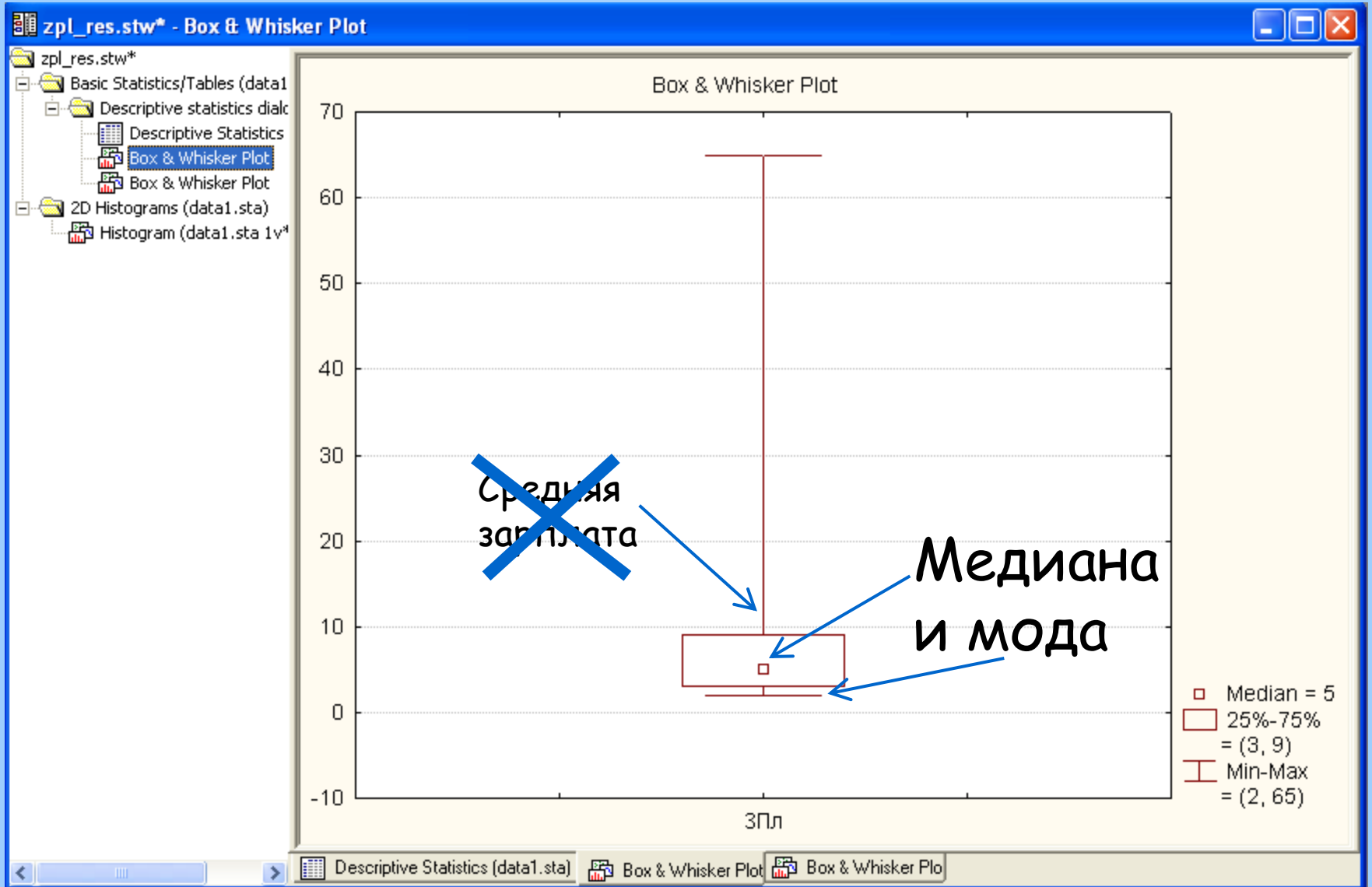
Боксплот, ящик с усами (медиана)



«Средняя зарплата»



«Средняя зарплата»



«Средняя зарплата»

Цифра:

Число занятых в научно технологическом (S&E, science and engineering) секторе США в 2008 г. достигло 5,8 млн человек. Занятые в этом секторе составляют в настоящее время 4,3% от всех занятых в стране, и эта цифра доходит до 7-15% всей рабочей силы в крупных метрополисах с высокой концентрацией S&E работодателей. С 2004 по 2008 г. занятость в разных профессиях росла в США в среднем на 1,3% ежегодно. Однако аналогичный рост в научно технологическом секторе достиг 3,3%, или 2,3% для всех профессий STEM (science, technology, engineering, mathematics). При медианной зарплате в стране в 2008 г. на уровне \$32.390 в год она составляла \$72.940 для S&E профессий и \$69.570 для STEM профессий.

А. К.

Простейшие описательные статистики

Так как значения переменных не постоянны, нужно научиться описывать их изменчивость.

Для этого придуманы описательные или дескриптивные статистики: минимум, максимум, среднее, дисперсия, стандартное отклонение, медиана, квартили, мода и т. д.

Идея этих статистик очень проста: вместо того чтобы рассматривать все значения переменной, а их может быть очень много (тысячи и миллионы), вначале стоит просмотреть описательные статистики. Они дают общее представление о значениях, которые принимает переменная.

Основные статистики

The screenshot displays the STATISTICA software interface. The main window title is "STATISTICA - zpl_res.stw - [Descriptive Statistics (data1.sta)]". The menu bar includes File, Edit, View, Insert, Format, Statistics, Graphs, Tools, Data, Workbook, Window, and Help. The "Statistics" menu is open, highlighting "Basic Statistics/Tables". Other menu items include Multiple Regression, ANOVA, Nonparametrics, Distribution Fitting, Advanced Linear/Nonlinear Models, Multivariate Exploratory Techniques, Industrial Statistics & Six Sigma, Power Analysis, Data-Mining, Statistics of Block Data, STATISTICA Visual Basic, and Probability Calculator.

In the background, a data table is visible with the following content:

	1
	3Пл
1	3
2	2
3	3
4	4
5	3
6	2
7	4
8	5
9	6
10	7
11	9
12	8
13	10
14	37
15	65

A text box with a green border contains the text: "Вы работаете в модуле «Basic Statistics / Tables» (Основные статистики и таблицы)".

In the foreground, a window titled "zpl_res.stw - Descriptive Statistics (data1.sta)" is open. It displays a table of descriptive statistics for the variable "3Пл":

Variable	Mean	Median	Mode	Frequency of Mode	Lower Quartile	Upper Quartile	Std.Dev.
3Пл	11,20000	5,000000	3,000000	3	3,000000	9,000000	17,18471

The window also shows a tree view of the project structure, including "Basic Statistics/Tables (data1)", "Descriptive statistics dial", "Descriptive Statistics", "Box & Whisker Plot", and "2D Histograms (data1.sta)".

Основные статистики

Медиана - делит
стат.ряд пополам

Нижний квартиль (25%)

СКО (σ)

The screenshot shows the 'Descriptive Statistics (data1.sta)' window in SPSS. The table below represents the data shown in the window. A blue box highlights the Lower and Upper Quartile columns, and a red arrow points to the Frequency of Mode column.

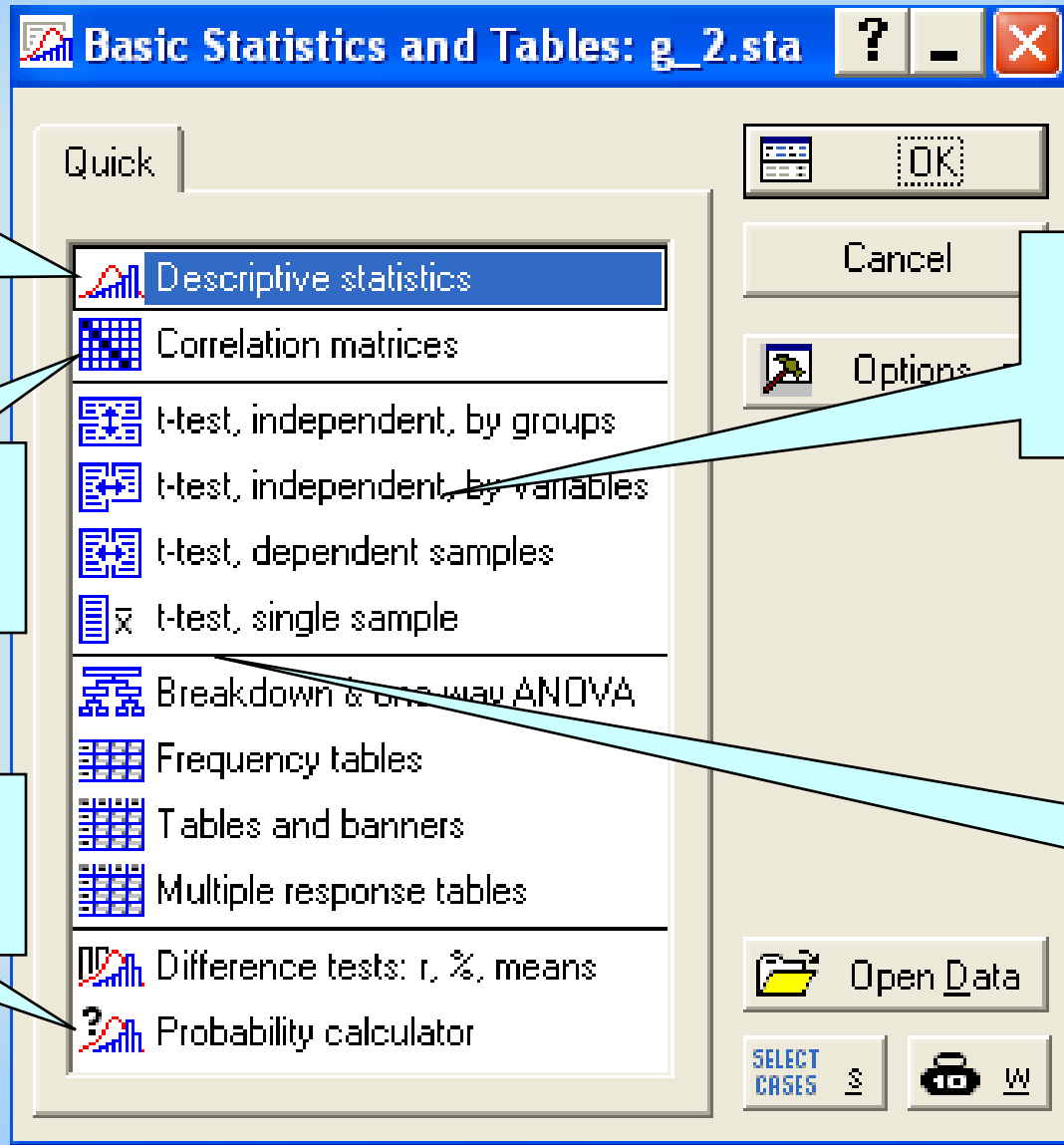
Variable	Mean	Median	Mode	Frequency of Mode	Lower Quartile	Upper Quartile	Std.Dev.
ЗПл	11,200000	5,000000	3,000000	3	3,000000	9,000000	17,18471

Среднее
значение

Мода - **часто**
встречающееся
значение

Верхний квартиль 25%

Простейшие описательные статистики



Описательные статистики

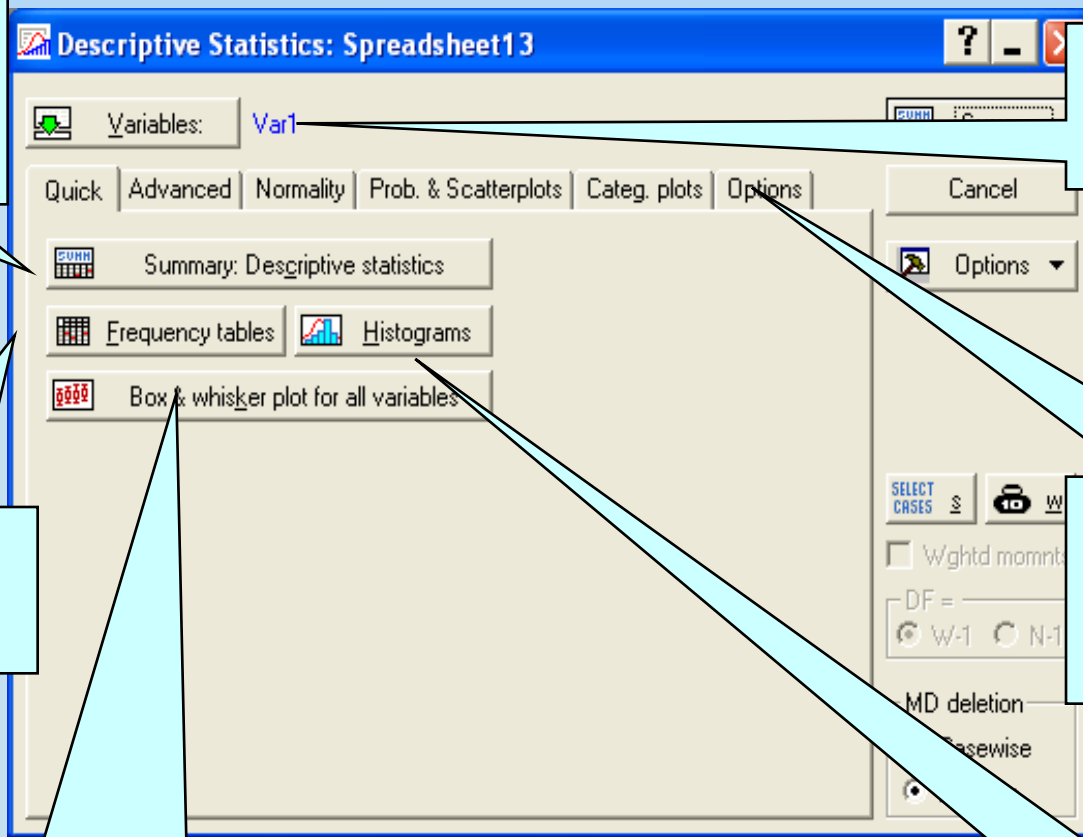
Корреляционные матрицы

Вероятностный калькулятор

Сравнения средних значений

Анализ вариаций

Простейшие описательные статистики



Расчёт
описательных
статистик

Выбор
переменных

Таблица
частот

Выбор
описательных
статистик

Визуализация
описательных статистик

Построение
гистограмм

Сравнение случайных величин

Data: Spreadsheet2* (4v by 50c)

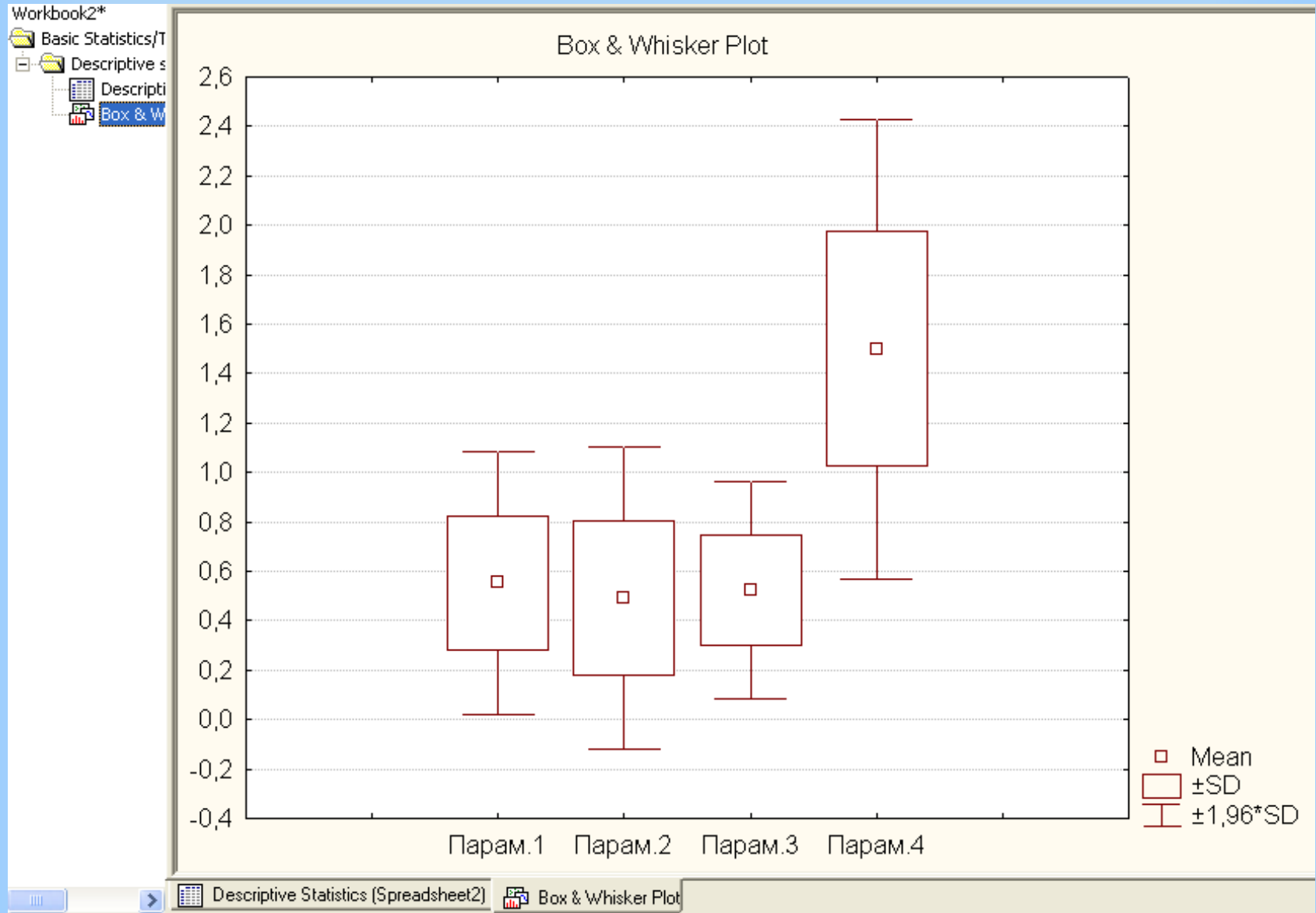
	1	2	3	4
	Парам.1	Парам.2	Парам.3	Парам.4
1	0,064952	0,379765	0,222358	1,259498
2	0,890727	0,653227	0,771977	2,280696
3	0,522733	0,906113	0,714423	2,232757
4	0,485761	0,685897	0,585829	1,293009
5	0,24773	0,341205	0,294467	1,085903
6	0,907793	0,734077	0,820935	1,73104
7	0,974603	0,317569	0,646086	1,464597
8	0,348712	0,081989	0,21535	0,589493
9	0,726354	0,858925	0,79264	2,020239
10	0,334436	0,500911	0,417674	1,629975
11	0,995497	0,555424	0,725464	1,499929

Workbook2* - Descriptive Statistics (Spreadsheet2)

Variable	Descriptive Statistics (Spreadsheet2)									
	Valid N	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Lower Quartile	Upper Quartile	Std.Dev.
Парам.1	50	0,553831	0,569197	Multiple		0,011541	0,974603	0,311221	0,800318	0,270494
Парам.2	50	0,494847	0,486001	Multiple		0,015520	0,981356	0,217244	0,743759	0,312068
Парам.3	50	0,524339	0,551729	Multiple		0,079048	0,891407	0,340506	0,714423	0,223535
Парам.4	50	1,499914	1,470173	Multiple		0,519372	2,522353	1,199558	1,896710	0,473853

Descriptive Statistics (Spreadsheet2) Box & Whisker Plot

Сравнение случайных величин

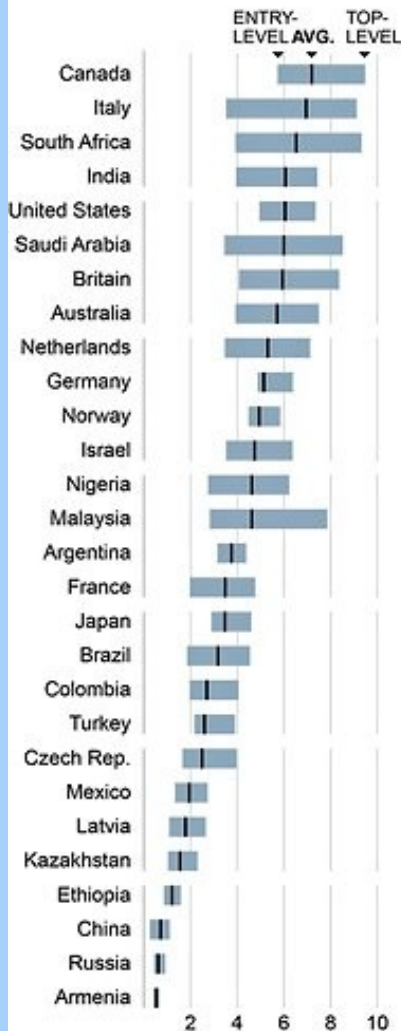


Teachers' Buying Power

A new book examines relative pay among public university teachers around the world.

Entry-level, average and top-level monthly salaries for teachers at public universities

In thousands of dollars, adjusted for purchasing power parity, before taxes



Source: "Paying the Professoriate, A Global Comparison of Compensation and Contracts," edited by Philip G. Altbach, Liz Reisberg, Maria Yudkevich, Gregory Androushchak and Iván F. Pacheco

How Much Is a Professor Worth?

Проверка статистических гипотез

Термин «гипотеза» означает предположение, которое не только вызывает сомнения, но и которое мы собираемся в данный момент проверить

Нулевая гипотеза H_0 – это гипотеза об отсутствии различий.

Это то, что мы хотим опровергнуть, если стоит задача доказать значимость различий

Она содержит число 0: $x_1 - x_2 = 0$, где x_1 и x_2 – сопоставляемые значения признаков.

Альтернативная гипотеза H_1 – это гипотеза о значимости различий.

Это то, что мы хотим доказать, поэтому иногда её называют экспериментальной гипотезой

Бывают задачи, когда мы хотим доказать незначимость различий, то есть подтвердить нулевую гипотезу. Однако чаще требуется доказать значимость различий, ибо они более информативны в поиске нового.

Статистический критерий

Статистический критерий – это правило, по которому принимается решение о принятии истинной и отклонении ложной гипотезы с высокой вероятностью

Критерии делятся на параметрические и непараметрические. Параметрические критерии – это критерии, включающие в формулу расчёта параметры распределения, то есть средние и дисперсии (t-критерий Стьюдента, критерий F и др.).

Схема проверки гипотез

1. Вычисляется эмпирическое (или фактическое, реальное) значение критерия $F_{\text{эмп}}$. Вычисляется число степеней свободы и уровень значимости.
2. По таблицам критических значений для выбранного критерия находится так называемая критическая точка (или критическое значение) $F_{\text{кр}}$.
3. По соотношению эмпирического и критического значений критерия судят о том, подтверждается или опровергается нулевая гипотеза. Например, если $F_{\text{эмп}} > F_{\text{кр}}$, гипотеза H_0 отвергается.

Уровень значимости

Уровень значимости – это вероятность отклонения нулевой гипотезы, в то время как она верна

В пакете Statistica значение задаваемого уровня значимости не используется. Как правило, в выходных данных содержатся выборочные значения статистики критерия и вероятность того, что случайная величина превышает это выборочное значение при условии, что верна гипотеза H_0 . Эта вероятность называется p -значением (p-level).

Ошибки при принятии гипотез

Ошибка, состоящая в том, что правильная гипотеза отклонена, в то время как она верна, называется ошибкой I рода

Ошибка, состоящая в том, что правильная гипотеза принята, в то время как она неверна, называется ошибкой II рода

Проверка гипотез об однородности выборок

Data: gipo_3s.sta				
	1	2	3	4
	P1	P2	P3	P4
Z1	82	535	1173	1714
Z2	63	429	995	1307

данные о наличии примесей (P1–P4) в углеродистой стали, выплавляемой двумя заводами Z1, Z2

Проверим гипотезу о том, что распределения содержания нежелательной примеси одинаковы на этих заводах

Statistics/ Advanced Linear/Nonlinear models/ Log-Linear Analysis of Frequency Tables.

Проверка гипотез об однородности выборок

В строке *Input file*: выбираем *Frequencies w/out coding variables* (частоты без кодирующих переменных). Кнопкой *Variables* вводим все переменные (*Select all*). Кнопкой *Specify Table* (спецификация таблицы) в ячейках *No. of levels*: вводим 4 и 2

Specify the dimensions of the... ? X

No. of levels:	Factor Name:
1: 4	P
2: 2	Z
3:	
4:	
5:	
6:	
7:	

OK

Cancel

For each factor in the table, specify the number of levels and the name

Проверка гипотез об однородности выборок

Дважды нажимаем *OK* и во вкладке *Advanced* получившегося окна выполним *Test all marginal & partial association models*.

В таблице *Results of Fitting all K-Factor Interactions* в последней строке получаем столбца значение статистики критерия χ^2 (*Chi-Square*), равное 3,59, число степеней свободы (*Degrs. of Freedom*) $df=3$ и уровень значимости 0,30887. Эта величина не больше критической.

Следовательно, гипотезу об одинаковом распределении содержания примеси в металле на двух заводах можно принять.

T - критерии

t-критерий является наиболее часто используемым методом, позволяющим выявить различие между средними двух выборок. Например, *t*-критерий можно использовать в промышленности для сравнения средних показателей в различных сменах, цехах, подразделениях.

В *STATISTICA* доступны:
t-критерий для зависимых выборок
t-критерий для независимых выборок

