

Математические методы обработки информации

Меркулова Ольга Петровна

доцент кафедры психологии образования и развития

Olga.Merkulova@list.ru
Merkulova-OP@yandex.ru

Тема 3. Проверяющая статистика

Литература

- **Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие. СПб.: Речь, 2007. Главы 7, 8, 12.**
- **Сидоренко Е.В. Методы математической обработки в психологии. СПб.: Речь, 2010.**
- Сорокова М. Г. Методы математической статистики в психологии: Учебное пособие. — Saarbrücken: Palmarium Academic Publishing, 2014. — 405 с. — URL: <http://psychlib.ru/inc/absid.php?absid=180853>
- Сорокова М.Г. Математические методы в психологии: непараметрическая статистика. Учебное пособие. — М.: МГППУ, 2011. — 281 с. — URL: <http://psychlib.ru/inc/absid.php?absid=97964>
- Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии /Пер. с англ. С общ. Ред. Ю.П. Адлера. М.: Прогресс, 1976.

Проверяющая статистика

- *Что проверяет?*
 - статистические гипотезы
- *С помощью чего проверяет?*
 - с помощью специальных методов, включающих расчет статистических критериев
- *Для чего это нужно?*
 - чтобы отделить случайные результаты от неслучайных

Пример. Результаты ответов на вопрос «Обеспечила ли средняя школа Вам (Вашему ребенку) достаточную подготовку для поступления в вуз?»

- Различия в полученных результатах можно объяснить:
 - случайными причинами
 - закономерностью
- Очевидные объяснения:
 - если различия **очень малы** – случайностью
 - если различия **очень велики** – закономерностью

Статистический критерий
позволяет принять
неочевидное решение

2012 год	Студенты	Родители
Среднее значение	4,26	4,51
Стандартное отклонение	1,17	1,29

2003 год	Студенты	Родители
Среднее значение	3,80	3,93
Стандартное отклонение	1,41	1,45

Статистическая гипотеза

- Статистическая гипотеза
 - это предположение о сходствах или различиях каких-либо характеристик распределений признаков (случайных величин).

2003 год	Студенты	Родители
Среднее значение	3,80	3,93
Стандартное отклонение	1,41	1,45

- Пример

- **H_0** . Средние значения, полученные на выборках студентов и родителей существенно не различаются (различия можно объяснить случайными причинами)

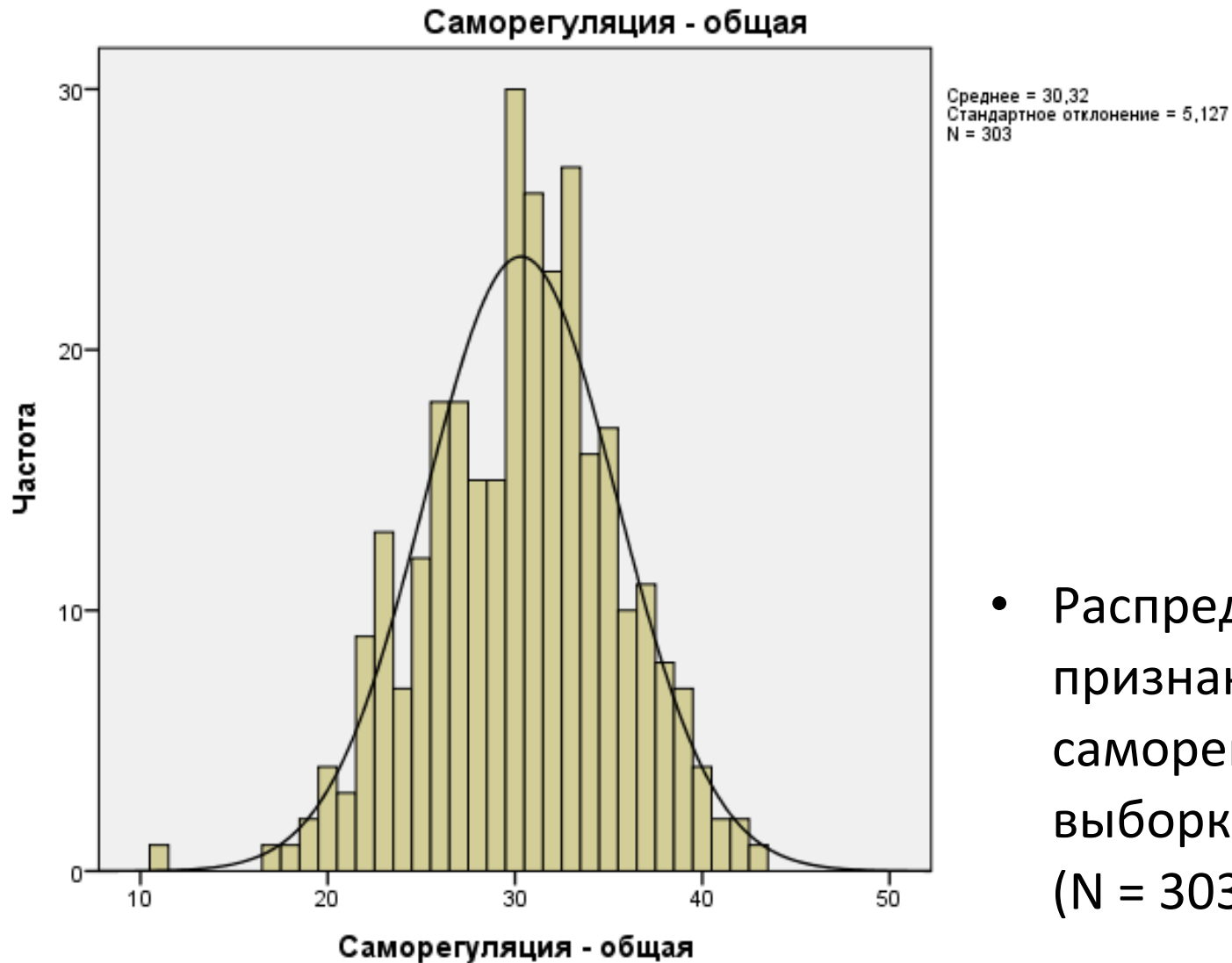
- **H_1** . Средние значения, полученные на выборках студентов и родителей существенно различаются (различия нельзя объяснить случайными причинами)

Статистическая гипотеза

Нулевая (H_0) – о том, что различия (связь) не существенны

Альтернативная (H_1) – о том, что различия (связь) существенны

Пример. Случайное деление выборки на 2 группы

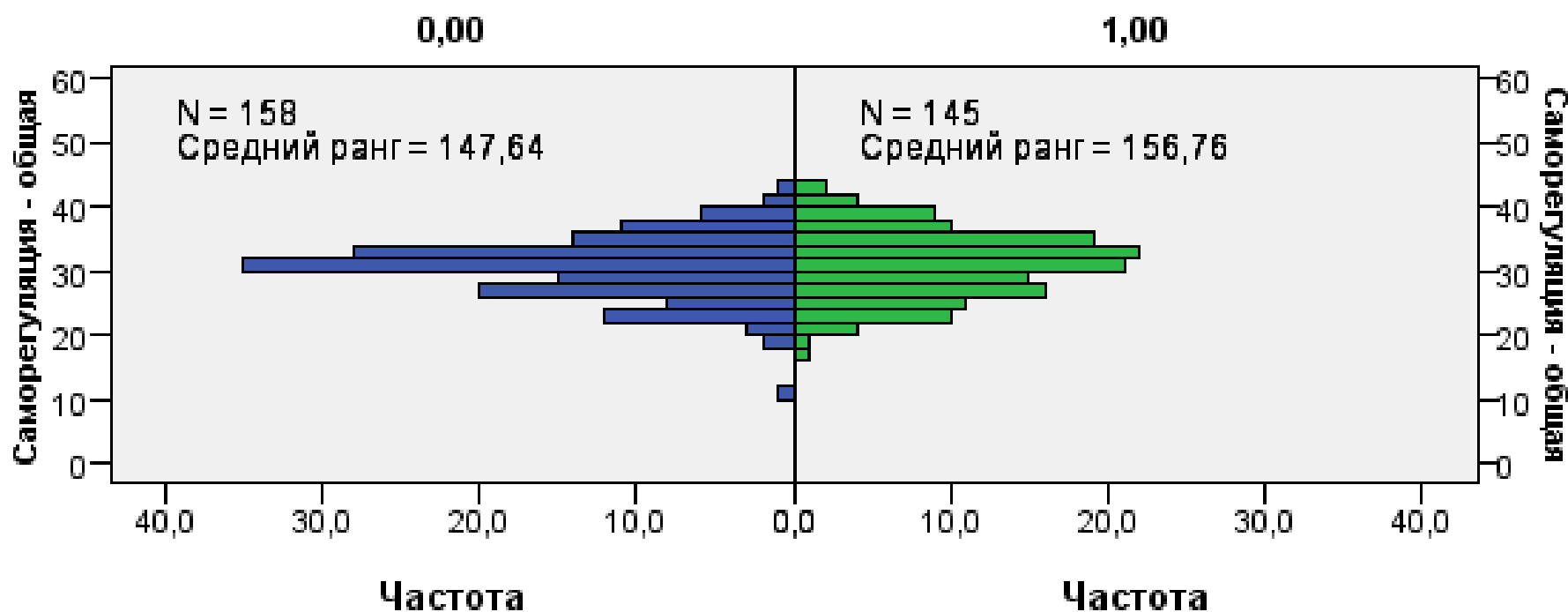


- Распределение признака «Общая саморегуляция» на выборке в целом (N = 303)

Распределение признака в 2 группах при **случайном** разделении (вариант 1)

Случайное деление на 2 группы	Ср. знач.	N	Стд. откл.
0	30,02	158	4,931
1	30,64	145	5,330
Итого	30,32	303	5,127

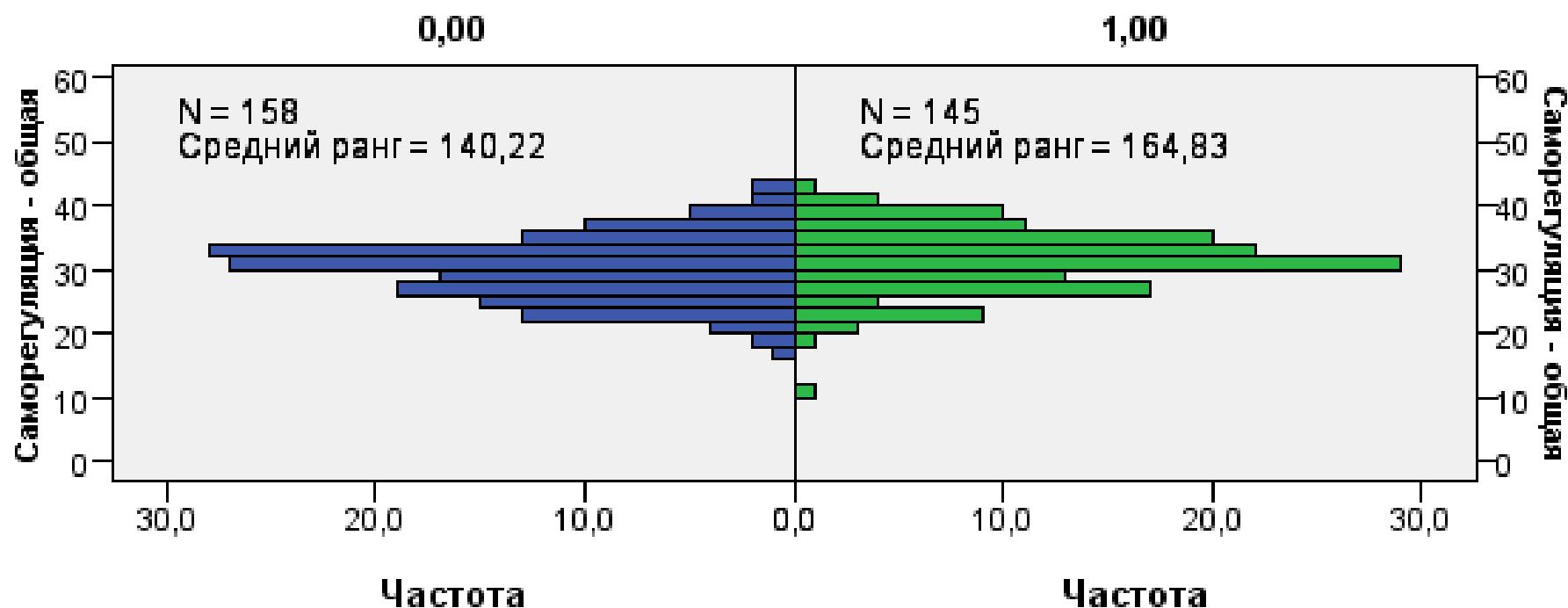
Случайное деление на 2 группы



Распределение признака в 2 группах при **случайном** разделении (вариант 2)

Случайное деление на 2 группы (в. 2)	Ср. знач.	N	Стд. откл.
0	29,66	158	5,042
1	31,03	145	5,141
Итого	30,32	303	5,127

Случайное деление на 2 группы - вариант 2

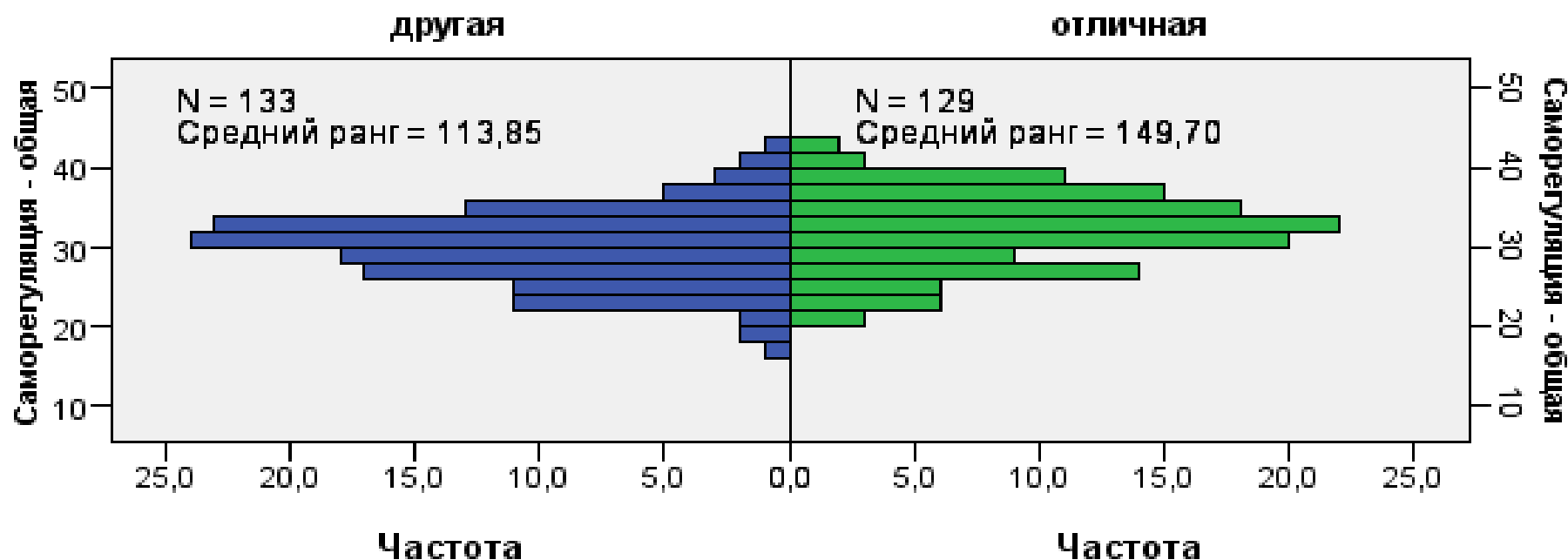


Распределение признака в 2 группах при разделении **по признаку** **успеваемости**

Группа по успеваемости	Ср. знач.	N	Стд. откл.
отличная	31,79	129	5,020
другая	29,46	133	4,698
Итого	30,61	262	4,988

Критерий U Манна-Уитни для независимых выборок

Оценка успеваемости



Статистический критерий

- **Критерий**

- (гр. *kriterion* — признак для суждения) — признак, основание, правило принятия решения по оценке чего-либо на соответствие предъявленным требованиям (мере) (Википедия).




- **Статистический критерий:**

- правило, на основании которого делается вывод *о выборе между нулевой и альтернативной статистическими гипотезами*, включает в себя:
 - а) способ расчета количественного показателя, который задает оценку конкретных данных с точки зрения вопроса гипотезы, а крайние значения которого соответствуют ситуациями очевидного решения о выборе в пользу гипотезы H_0 или H_1 ;
 - б) способ расчета или статистические таблицы, в которых всем возможным значениям критерия (количественного показателя – см. выше) сопоставлены вероятности их встречаемости при условии выполнения гипотезы H_0 ;
 - с) правило выбора в пользу гипотезы H_0 или H_1 на основе оценки вероятности появления эмпирического значения критерия, соответствующего заданной ситуации.

Пример. Критерий знаков G

- Задача
 - В группе из 10 (N) человек провели оценку изменения какого-либо признака по прошествии некоторого интервала времени (с помощью экспертного оценивания или дважды провели измерение признака и нашли разность). При этом для каждого человека определен положительный (+) или отрицательный (–) сдвиг.
 - Для конкретного числа k «нетипичных» сдвигов необходимо оценить, можно ли считать преобладание «типичного» (преобладающего) сдвига случайным или закономерным.

Например...

№	Изм. 1	Изм. 2	Сдвиг	
1	11	13	2	
2	10	12	2	
3	3	5	2	
4	6	11	5	
5	5	9	4	
6	11	9	-2	
7	5	7	2	
8	8	12	4	
9	8	7	-1	
10	7	10	3	

Пусть в опыте получено
8 положительных и
2 отрицательных
сдвига

Это может
объясняться
случайностью

H_0 : преобладание
«типичного» сдвига
является случайным

Это может
объясняться
закономерностью

H_1 : преобладание
«типичного» сдвига
является
неслучайным

Идея статистического вывода для критерия знаков

- **Предполагается**, что закономерности, приводящей к «типичному» (преобладающему) сдвигу нет - **оба варианта сдвига равновероятны** (верна гипотеза H_0 - преобладание «типичного» сдвига является **случайным**)
- Оценивается вероятность, с которой такое (или меньшее) число нетипичных сдвигов может появиться при этом предположении - когда верна гипотеза H_0 .
- **Принимается, что:**

вероятность,
с которой такое число
«нетипичных» сдвигов может
встретиться при случайной
модели
(допущении, что верна H_0)

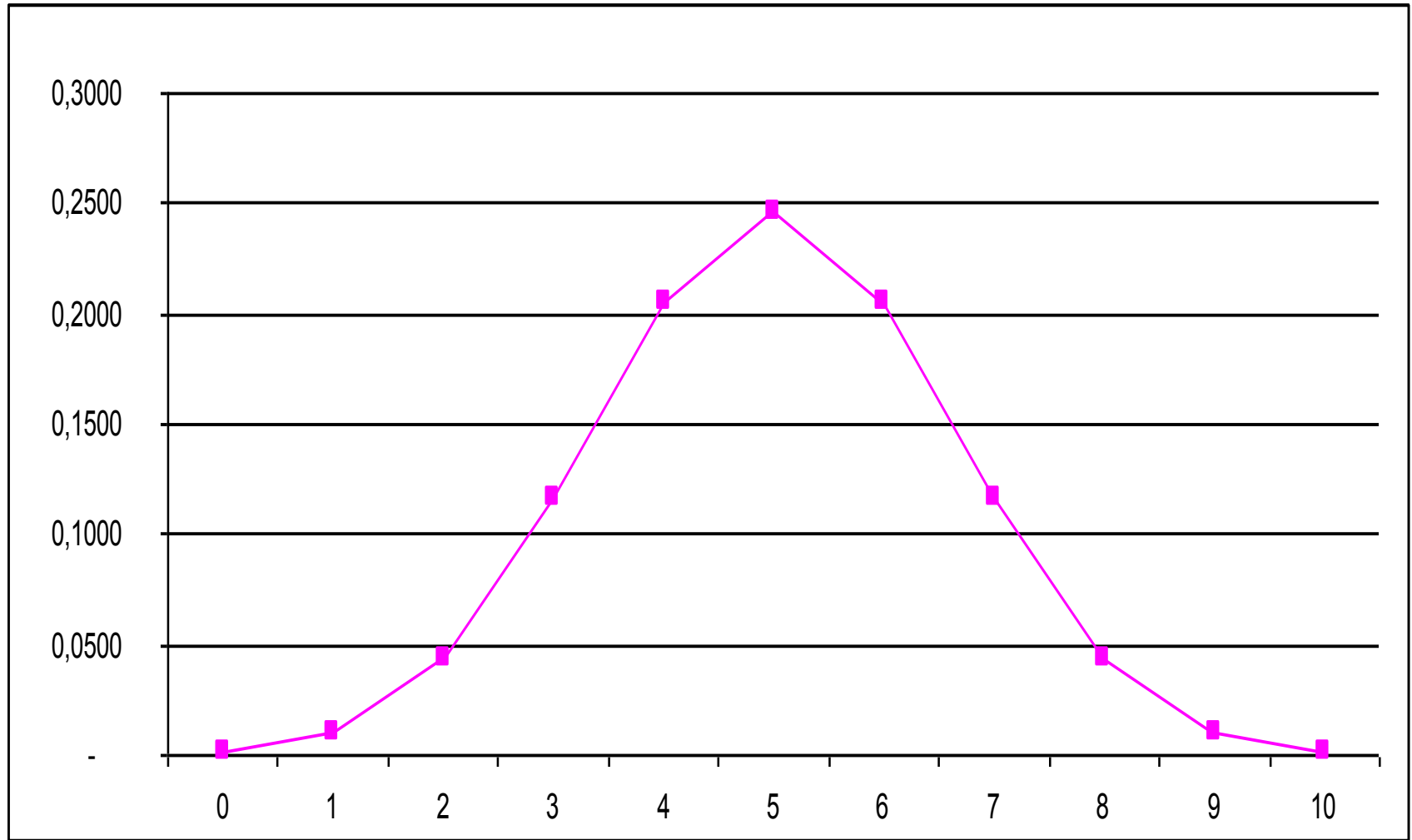


вероятности того,
что мы совершим ошибку,
приняв гипотезу H_1

Распределение вероятностей (для примера)

число нетипичных сдвигов из 10	вероятность появления данного числа сдвигов	вероятность появления числа нетипичных сдвигов, не больше данного (накопленная вероятность)
0	0,1%	0,1%
1	1,0%	1,1%
2	4,4%	5,5%
3	11,7%	17,2%
4	20,5%	37,7%
5	24,6%	62,3%
6	20,5%	82,8%
7	11,7%	94,5%
8	4,4%	98,9%
9	1,0%	99,9%
10	0,1%	100,0%

Распределение вероятностей (для примера)



Интерпретация (к примеру)

- При случайном сочетании положительных и отрицательных сдвигов (когда они равновероятны)
 - не более 1-го нетипичного сдвига ($k=1$) из 10 можно получить примерно в **1,1 %** случаев (чуть чаще, чем в среднем 1 раз из 100);
 - не более 2-х нетипичных сдвигов ($k=2$) из 10 можно получить примерно в **5,5 %** случаев (примерно 5-6 раз из 100);.
- Т.е. если мы посчитаем,
 - что при появлении **1 нетипичного сдвига** преобладание типичных сдвигов закономерно, то мы **ошибемся** с вероятностью примерно **1,1 %** .
 - что при появлении **2 нетипичных сдвигов** преобладание типичных сдвигов закономерно, то мы **ошибемся** с вероятностью примерно **5,5 %** .

Ошибки и вероятности при применении статистических критериев

	Решение	
	принять H_0	принять H_1
Верна H_0	Правильное с вероятностью $(1-\alpha)$	Ошибочное с вероятностью α (ошибка I рода)
Верна H_1	Ошибочное с вероятностью β (ошибка II рода)	Правильное с вероятностью $(1-\beta)$

- Если требуется подтверждение гипотезы H_1 , важно контролировать ошибку I рода. Вероятность α , в пределах которой должна быть ошибка I рода называют **уровнем значимости** при проверке статистической гипотезы (чем меньше α , тем достовернее выводы).
- Контроль ошибки II рода - **мощность критерия** (чем меньше β , тем мощнее критерий).

Принятые в психолого-педагогических исследованиях уровни значимости

- 5% ($p \leq 0,05$)
- 1% ($p \leq 0,01$)
- 0,1% ($p \leq 0,001$)

— где p — вероятность ошибки I рода

Уровень значимости показывает вероятность ошибиться, посчитав различия (взаимосвязь) не случайными (существенными, значимыми)

Возвращаясь к примеру

- Пусть в опыте получено 2 отрицательных сдвига. Необходимо принять одну из двух гипотез:
 - H_0 : преобладание «типичного» сдвига является случайным;
 - H_1 : преобладание «типичного» сдвига является неслучайным.

Эмпирическое значение критерия $G_{\text{эмп}}$ (для данного критерия - число «нетипичных» сдвигов) = 2

При использовании справочника нужно по таблице определить критические значения

Таблица V

Критические значения критерия знаков G для уровней статистической значимости $p \leq 0,05$ и $p \leq 0,01$ (по Оуэну Д.Б., 1966)

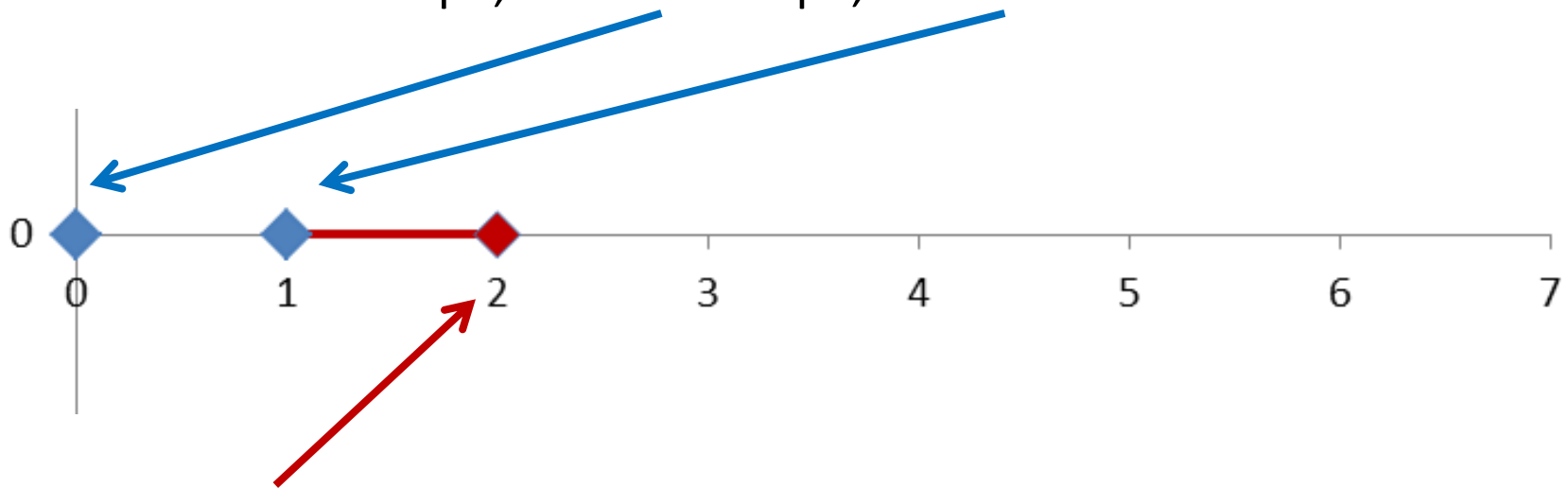
Преобладание "типичного" сдвига является достоверным, если $G_{\text{эмп}}$ ниже или равен $G_{0,05}$, и тем более достоверным, если $G_{\text{эмп}}$ ниже или равен $G_{0,01}$.

n	p		n	p		n	p		n	p	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
5	0	-	27	8	7	49	18	15	92	37	34
6	0	-	28	8	7	50	18	16	94	38	35
7	0	0	29	9	7	52	19	17	96	39	36
8	1	0	30	10	8	54	20	18	98	40	37
9	1	0	31	10	8	56	21	18	100	41	37
10	1	0	32	10	8	58	22	19	110	45	42
11	2	1	33	11	9	60	23	20	120	50	46
12	2	1	34	11	9	62	24	21	130	55	51
13	3	1	35	12	10	64	24	22	140	59	55
14	3	2	36	12	10	66	25	23	150	64	60
15	3	2	37	13	10	68	26	23	160	69	64
16	4	2	38	13	11	70	27	24	170	73	69

Для $N=10$ $G_{\text{кр}0,05} = 1$; $G_{\text{кр}0,01} = 0$

Сравнение эмпирического и критических значений критерия на «оси значимости»

- Для $N=10$ $G_{кр0,01} = 0$; $G_{кр0,05} = 1$



- $G_{эмп} = 2$

Вывод: принимается гипотеза H_0

Пример. Результаты ответов на вопрос «Обеспечила ли средняя школа Вам (Вашему ребенку) достаточную подготовку для поступления в вуз?»

- Проверка различий в средних с помощью t-критерия Стьюдента.

- Различия между оценками студентов и родителей в 2003 году не значимы ($p = 0,343$).
- Различия между оценками студентов и родителей в 2012 году не значимы ($p = 0,644$).
- Различия между оценками родителей в 2003 и 2012 годах значимы на уровне $p \leq 0,05$ ($p = 0,033$)
- Различия между оценками студентов в 2003 и 2012 годах не значимы ($p = 0,090$)

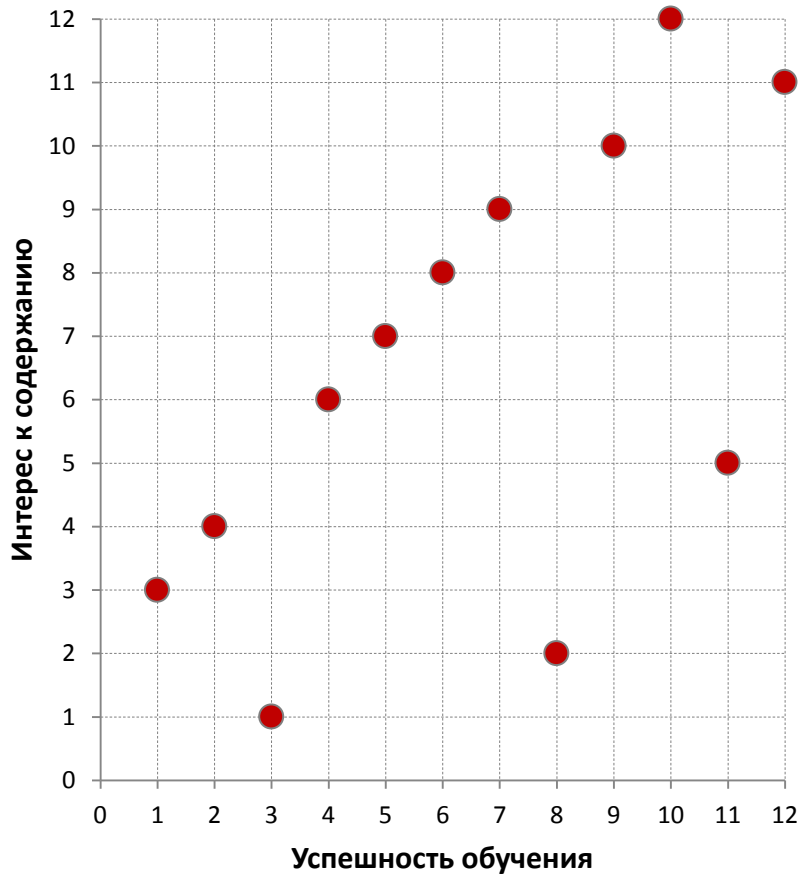
2003 год	Студенты	Родители
Среднее значение	3,80	3,93
Стандартное отклонение	1,41	1,45

2012 год	Студенты	Родители
Среднее значение	4,26	4,51
Стандартное отклонение	1,17	1,29

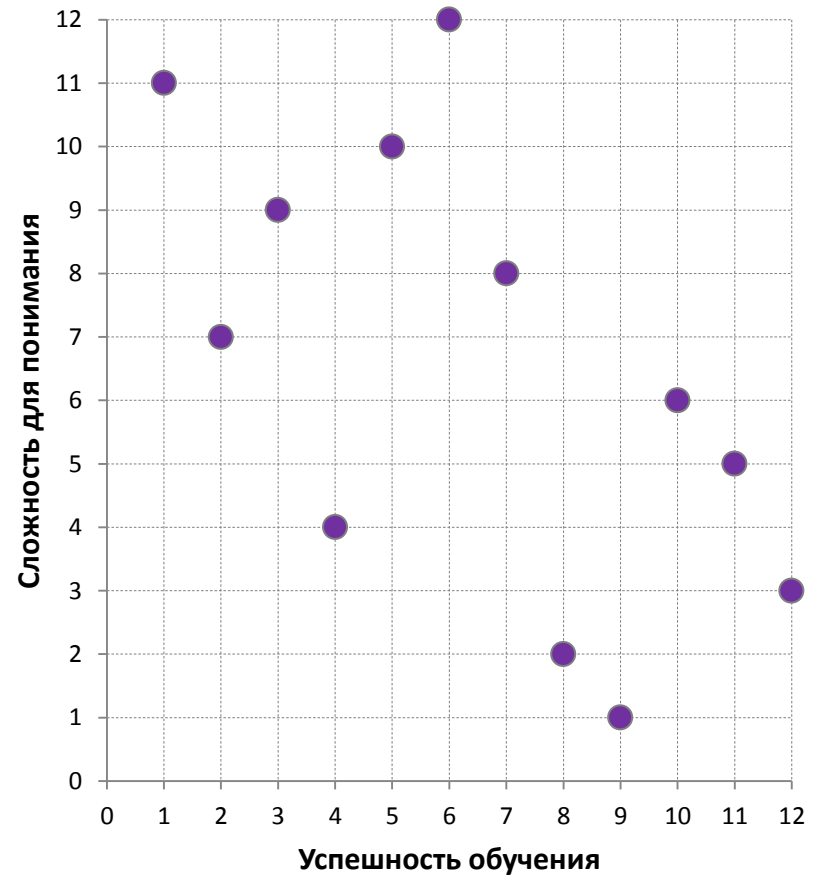
Корреляционный анализ связей между несколькими переменными

Предмет	Успешность обучения	Интерес к содержанию	Понимание значимости	Затраты времени, сил	Сложность для понимания
Иностранный язык	2	4	9	6	7
Русский язык и культура речи	5	7	8	9	10
Право	10	12	10	7	6
Социология	6	8	11	12	12
Основы педиатрии и гигиены	7	9	4	5	8
Общая и экспериментальная психология (с практикумом)	3	1	2	2	9
Теория обучения и воспитания	11	5	6	10	5
Психология развития	8	2	1	1	2
Социальная педагогика	12	11	12	11	3
Качественные и количественные методы психолого-педагогических исследований	4	6	3	4	4
Профессиональная этика в психолого-педагогической деятельности	9	10	7	8	1
Методика посреднической деятельности (СПБ)	1	3	5	3	11

Диаграммы рассеяния и коэффициенты корреляции для отдельных пар переменных оценивания



● $r = 0,629$

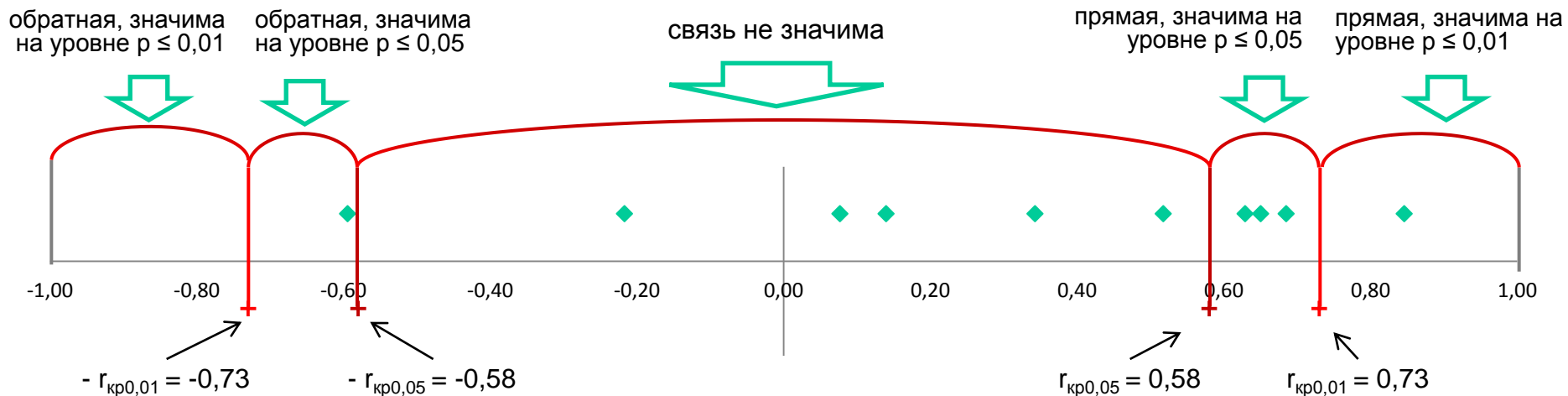


● $r = -0,594$

Матрица интеркорреляций

	Успешность обучения	Интерес к содержанию	Понимание значимости	Затраты времени, сил	Сложность для понимания
Успешность обучения	1	0,629*	0,343	0,517	-0,594
Интерес к содержанию	0,629*	1	0,685*	0,650*	-0,217
Понимание значимости	0,343	0,685*	1	0,846**	0,140
Затраты времени, сил	0,517	0,650*	0,846**	1	0,077
Сложность для понимания	-0,594	-0,217	0,140	0,077	1

«Ось значимости» для приведенных коэффициентов корреляции



Критические значения приведены для N=12

Виды статистических критериев

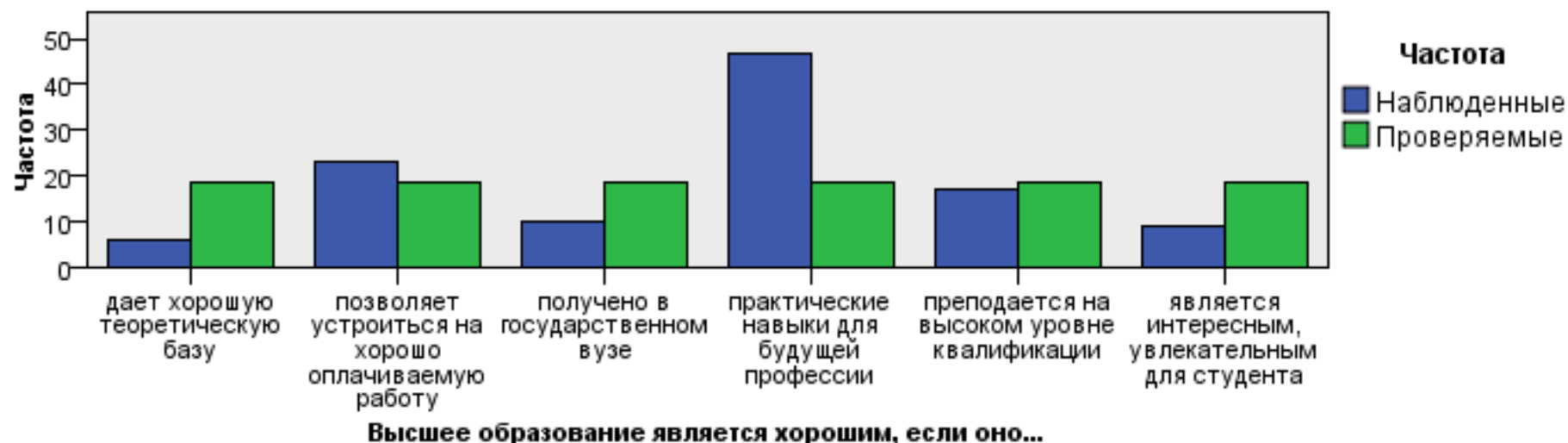
- По особенностям вычислений и требованиям к исходным данным (**разные критерии**):
 - параметрические
 - непараметрические
- По формулировке статистических гипотез (**разные варианты одного критерия**):
 - направленные (односторонние – сравнение «больше – меньше»)
 - ненаправленные (двусторонние сравнение «равно – не равно»)

Обзор статистических критериев в зависимости от задач и исходных данных

Изучение одного признака на одной (однородной) выборке

Тип данных (шкала)	Номина- тивные	Порядковые	Интерваль- ные
Описание данных	Частоты Мода	Частоты Медиана	Ср. арифм., Ст. откл.
Диаграммы	Круговая, столбчатая	Столбчатая, линейная	Гистограмма
Характер рас- пределения	Критерий χ^2 Пирсона	Критерии χ^2 Пирсона, λ Колмогорова–Смирного	
Сравнение среднего с ожидаемым	-	-	t критерий Стъюдента (одновыбо- рочный)

Одновыборочный критерий хи-квадрат

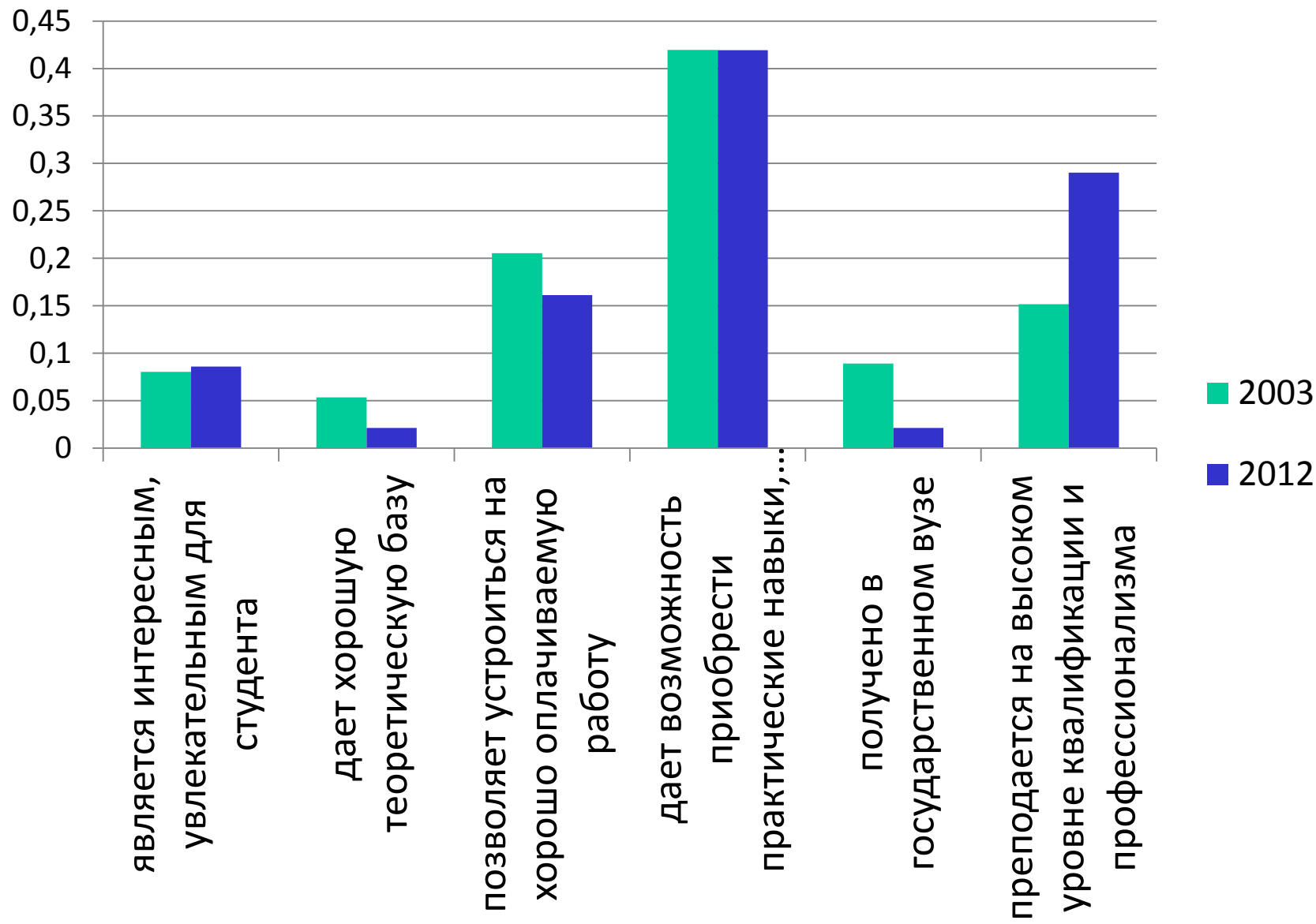


Всего N	112
Статистика критерия	61,786
Степени свободы	5
Асимптотическая знч. (2-сторонний критерий)	,000

1. Есть 0 ячеек 0 с ожидаемыми значениями меньше чем 5. Минимальное ожидаемое значение равно 18,667.

Сравнение одного признака на двух выборках (в двух разных группах)

Тип данных (шкала)	Номина- тивные	Порядковые	Интерваль- ные
Сравнение вида рас- пределения	Критерий χ^2 Пирсона	Критерии χ^2 Пирсона, λ Колмогорова–Смирного	
Сравнение уровня признака	-	U критерий Манна-Уитни	t критерий Стьюдента (для незави- симых выборок)

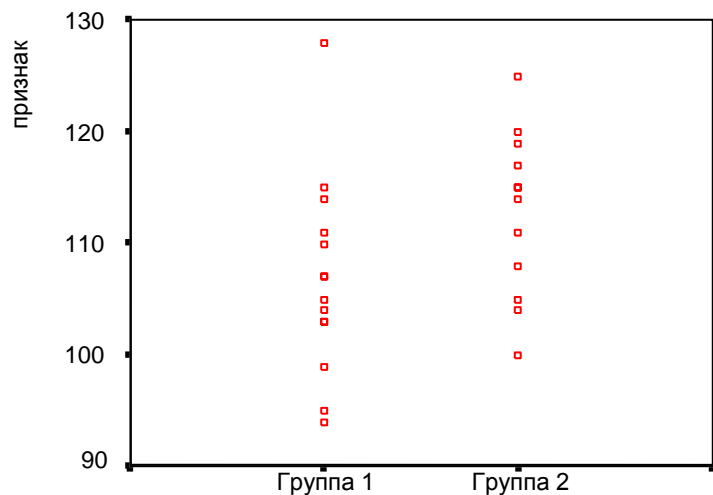
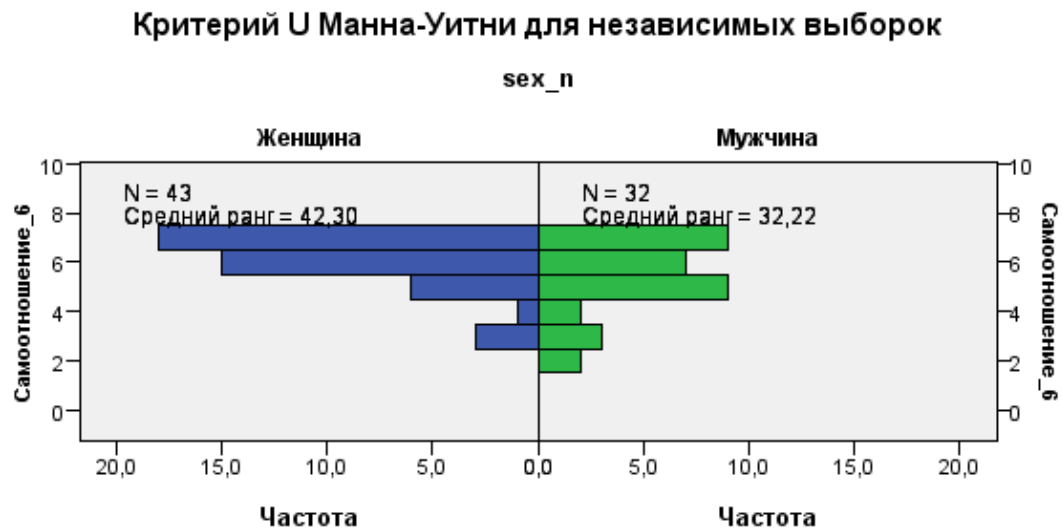
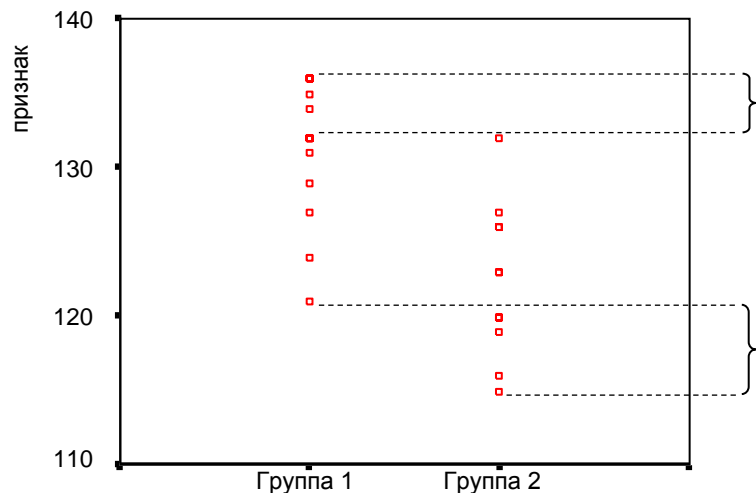


Пример

Высшее образование является хорошим, если оно... * ГОД Crosstabulation

			ГОД		Total
			2003	2012	
Высшее образование является хорошим, если оно...	является интересным, увлекательным для студента	Count % within ГОД	9 8,0%	8 8,6%	17 8,3%
	дает хорошую теоретическую базу	Count % within ГОД	6 5,4%	2 2,2%	8 3,9%
	позволяет устроиться на хорошо оплачиваемую	Count % within ГОД	23 20,5%	15 16,1%	38 18,5%
	практические навыки для будущей профессии	Count % within ГОД	47 42,0%	39 41,9%	86 42,0%
	получено в государственном вузе	Count % within ГОД	10 8,9%	2 2,2%	12 5,9%
	преподается на высоком уровне	Count % within ГОД	17 15,2%	27 29,0%	44 21,5%
Total		Count % within ГОД	112 100,0%	93 100,0%	205 100,0%

- Различия в распределении ответов, данных в 2003 и 2012 годах не значимы ($p = 0,064$) по критерию χ^2 Пирсона

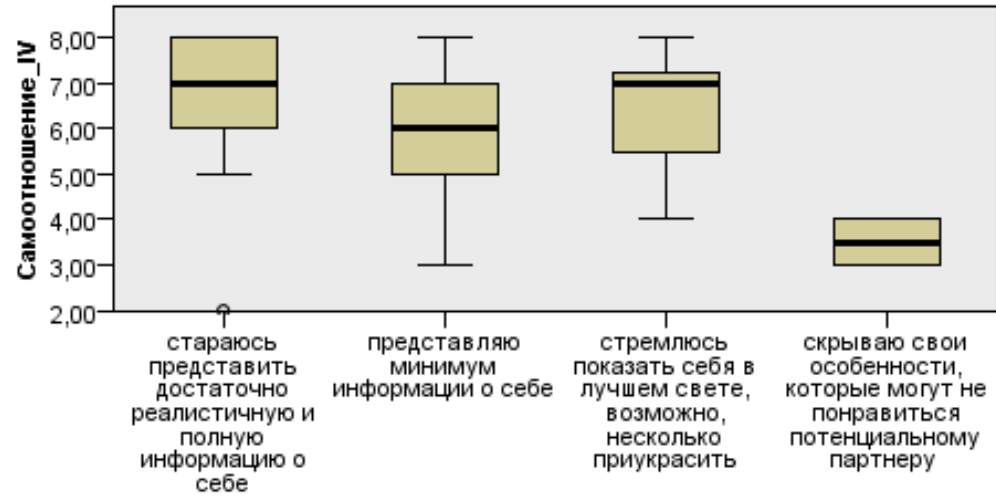


Всего N	75
U Манна-Уитни	503,000
W Вилкоксона	1 031,000
Статистика критерия	503,000
Стандартная ошибка	89,525
Стандартизованная статистика критерия	-2,066
Асимптотическая знч. (2-сторонний критерий)	,039

Сравнение одного признака на трех и более выборках (в трех и более группах)

Тип данных (шкала)	Номина- тивные	Порядковые	Интерваль- ные
Сравнение вида рас- пределения	Критерий χ^2 Пирсона		
Сравнение уровня признака	-	Н критерий Крускала- Уоллиса	Однофактор- ный дисперсион- ный анализ (ANOVA)

Критерий Краскала-Уоллиса для независимых выборок



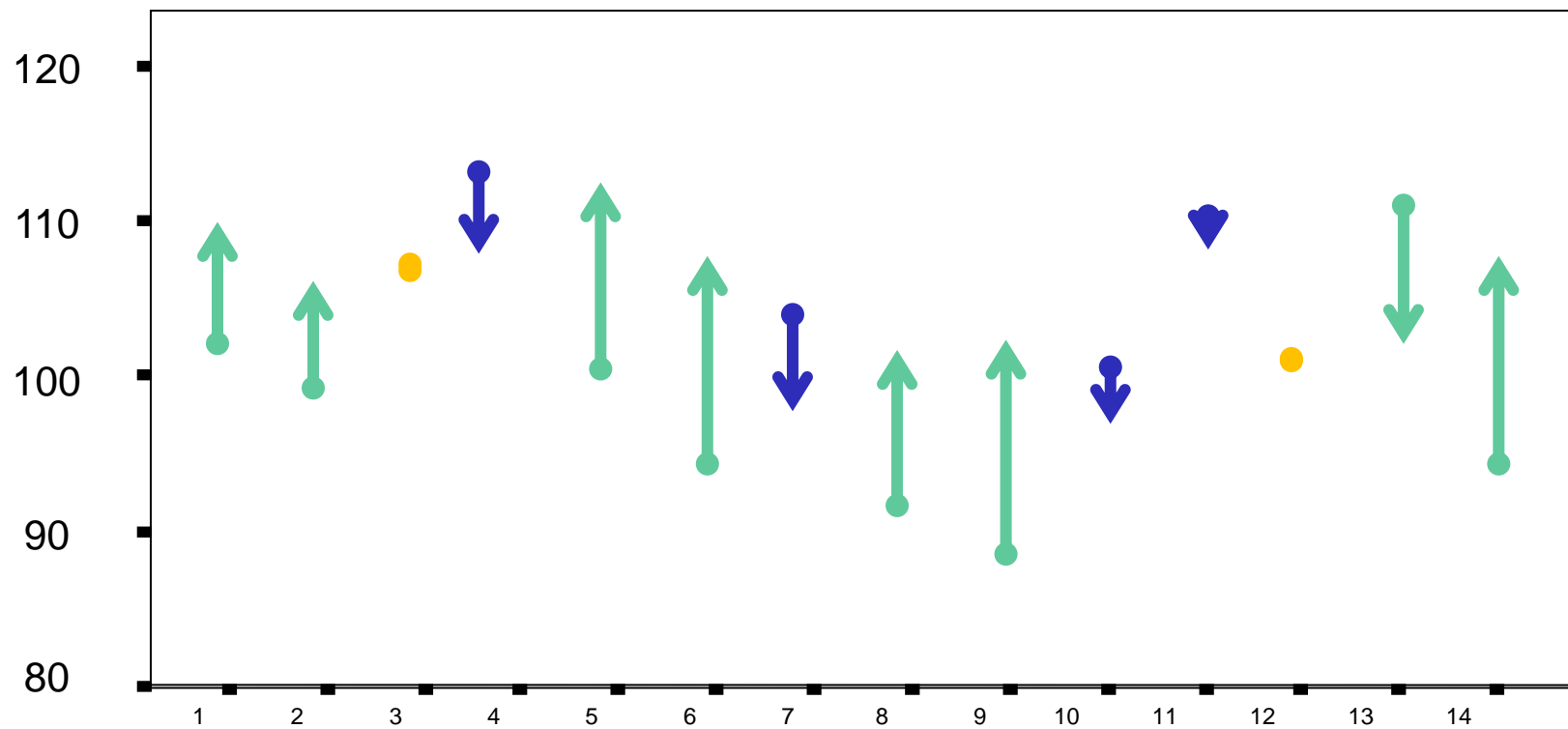
Как Вы подходите к заполнению СВОЕЙ анкеты на сайте знакомств?

Всего N	75
Статистика критерия	8,099
Степени свободы	3
Асимптотическая знч. (2-сторонний критерий)	,044

1. Статистика критерия скорректирована на наличие связей.

Оценка изменения признака на одной выборке

Тип данных (шкала)	Номина- тивные	Порядковые	Интерваль- ные
Оценка изменения признака	Анализ таблиц сопряжен- ности	G критерий знаков, T критерий Вилкоксона	t критерий Стъюдента (для связанных выборок)



Испытуемые

	Состояние: <i>до</i> экспериментального воздействия	Экспериментальное воздействие	Состояние: <i>после</i> экспериментальног о воздействия
Кон- трольная группа	И _{К1} (измерение в контрольной группе до воздействия)	нет воздействия	И _{К2} (измерение в контрольной группе после воздействия)
		\longleftrightarrow (2) ? – произошли ли значимые изменения в одной группе	
	\updownarrow (1) ? – значимы ли различия между различными группами		\updownarrow (3) ? – значимы ли различия между различными группами
Экспери- менталь- ная группа	И _{Э1} (измерение в экспериментальной группе до воздействия)	есть воздействие	И _{Э2} (измерение в экспериментально й группе после воздействия)
		\longleftrightarrow (4) ? – произошли ли значимые изменения в одной группе	

Выявление взаимосвязи между признаками

	Номинативный	Порядковый	Интервальный
Номинативный	Анализ таблиц сопряженности – критерий χ^2 Пирсона	U критерий Манна-Уитни, H критерий Крускала-Уоллиса	t критерий Стьюдента (для независимых выборок), Однофакторный дисперсионный анализ
Порядковый		Коэффициент ранговой корреляции Спирмена	«Понижение» типа шкалы: порядковой до номинативной или интервальной до порядковой
Интервальный			Коэффициент линейной корреляции Пирсона



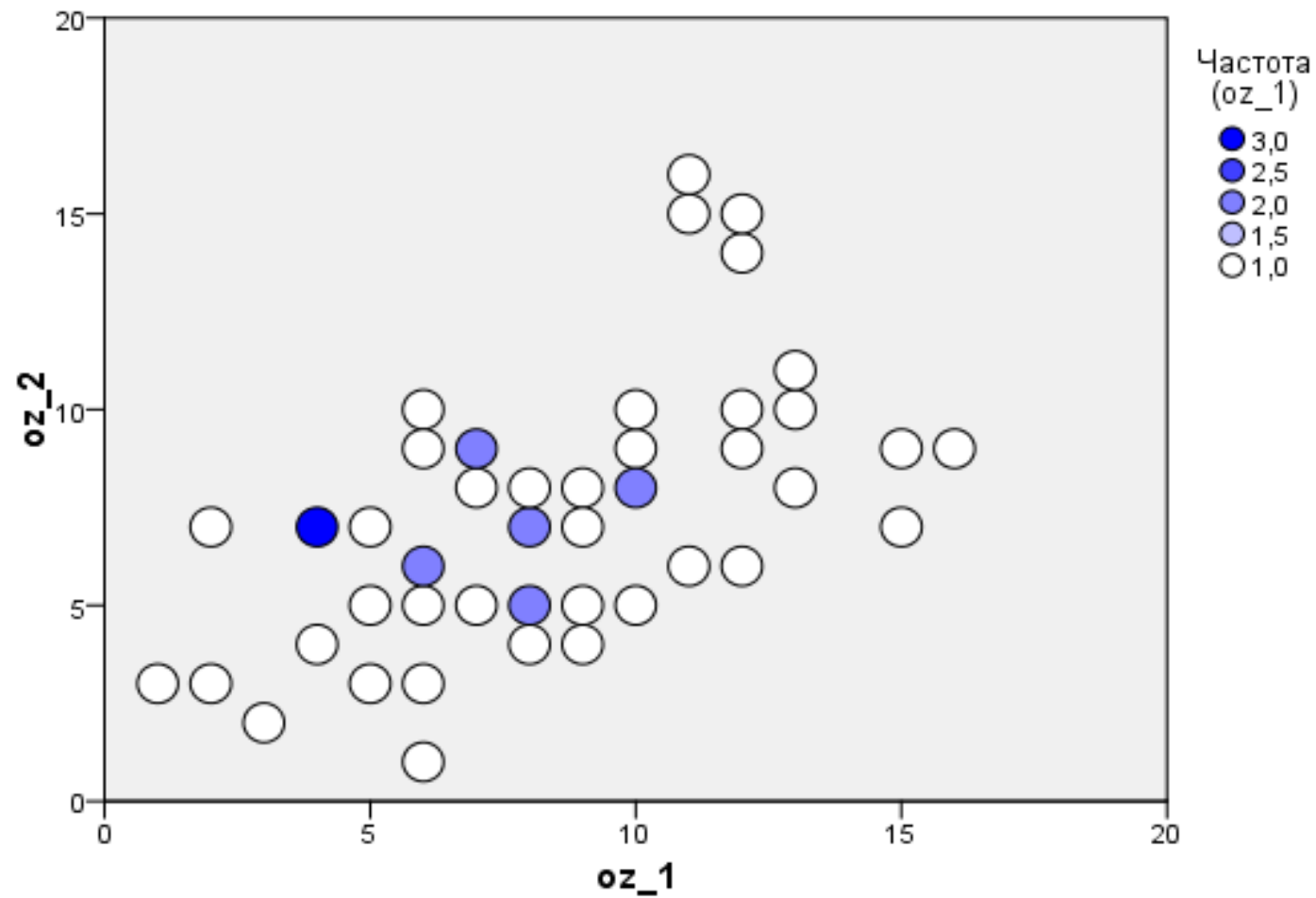
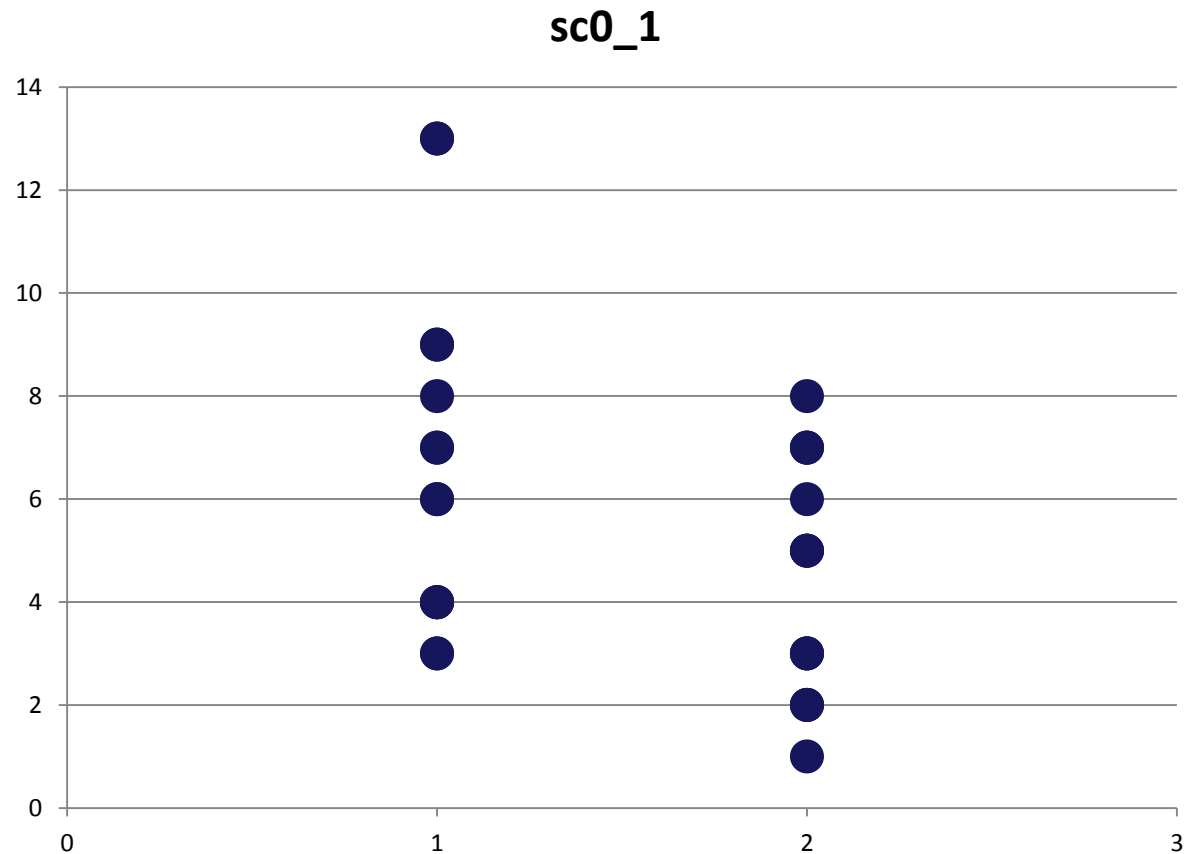


Диаграмма рассеяния для признаков при $r=0,553^{}$ (N=51)**

Пример из самостоятельной работы по применению проверяющей статистики

Применение критерия
Манна-Уитни для сравнения
уровня выраженности признака в
двух группах

группа	sc0_1	ранг
2	1	1
2	2	3
2	2	3
2	2	3
1	3	7
1	3	7
2	3	7
2	3	7
2	3	7
1	4	11,5
1	4	11,5
1	4	11,5
1	4	11,5
2	5	15
2	5	15
2	5	15
1	6	18
1	6	18
2	6	18
1	7	22
1	7	22
2	7	22
2	7	22
1	8	25,5
2	8	25,5
1	9	27,5
1	9	27,5
1	13	29,5
1	13	29,5

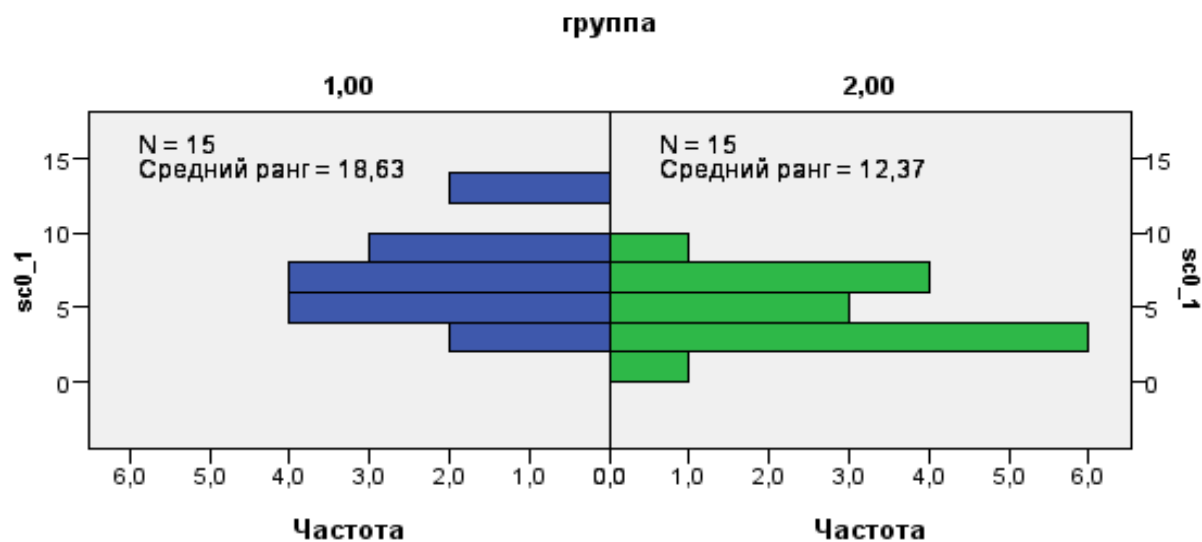


- H_0 : Уровень признака в группе 2 не ниже уровня признака в группе 1.
- H_1 : Уровень признака в группе 2 ниже уровня признака в группе 1.

- Для групп 1 и 2 отдельно подсчитаны ранговые суммы:
- $T_1 = 279,5$ $T_2 = 185,5$
- Бóльшая из ранговых сумм – T_1 .
- $U_{\text{эмп}} = (n_1 * n_2) + n_1 * (n_1 + 1)/2 - T_1 = (15 * 15) + 15 * 16 / 2 - 279,5 = 65,5$
- По таблицам [Сидоренко, 2010, с. 316] находим $U_{\text{кр } 0,05} = 72$, $U_{\text{кр } 0,01} = 56$ для $n_1 = 15$ и $n_2 = 15$. Для данного критерия, если $U_{\text{эмп}} < U_{\text{кр}}$ различия можно считать значимыми на соответствующем уровне.
- $U_{\text{эмп}} < U_{\text{кр } 0,05}$, но $U_{\text{эмп}} > U_{\text{кр } 0,01}$, значит, различия между группами можно считать значимыми на уровне $p \leq 0,05$. Принимается альтернативная гипотеза уровень признака в группе 2 ниже уровня признака в группе 1, вероятность ошибки решения принять альтернативную гипотезу находится в пределах 5%.
- Вывод, который можно записать в тексте с описанием результатов исследования: «Уровень рассматриваемого признака в группе 1 выше, чем в группе 2 (различия значимы на уровне $p \leq 0,05$ по критерию U Манна-Уитни)».

Результаты обработки данных примера в программе SPSS

Критерий U Манна-Уитни для независимых выборок



Всего N	30
U Манна-Уитни	65,500
W Вилкоксона	185,500
Статистика критерия	65,500
Стандартная ошибка	23,934
Стандартизованная статистика критерия	-1,964
Асимптотическая знч. (2-сторонний критерий)	,050
Точная знч. (2-сторонний критерий)	,050

Пример из самостоятельной работы по применению проверяющей статистики

Применение критерия
Вилкоксона для оценки изменений
признака в одной группе

№	sc0_1	sc0_2	сдвиг d	модуль d
1	8	5	-3	3
2	3	8	5	5
3	4	2	-2	2
4	6	5	-1	1
5	13	10	-3	3
6	4	6	2	2
7	9	6	-3	3
8	6	7	1	1
9	3	2	-1	1
10	9	9	0	0
11	4	7	3	3
12	4	3	-1	1
13	7	9	2	2
14	13	6	-7	7
15	7	7	0	0
16	3	5	2	2
17	4	7	3	3
18	5	6	1	1
19	2	2	0	0
20	3	3	0	0
21	8	6	-2	2
22	9	7	-2	2
23	5	6	1	1
24	9	7	-2	2
25	3	1	-2	2
26	3	7	4	4
27	3	3	0	0
28	6	7	1	1
29	9	4	-5	5
30	7	10	3	3

№	sc0_1	sc0_2	сдвиг d	модуль d	Ранг модуля d
10	9	9	0	0	
15	7	7	0	0	
19	2	2	0	0	
20	3	3	0	0	
27	3	3	0	0	
4	6	5	-1	1	4
8	6	7	1	1	4
9	3	2	-1	1	4
12	4	3	-1	1	4
18	5	6	1	1	4
23	5	6	1	1	4
28	6	7	1	1	4
3	4	2	-2	2	11,5
6	4	6	2	2	11,5
13	7	9	2	2	11,5
16	3	5	2	2	11,5
21	8	6	-2	2	11,5
22	9	7	-2	2	11,5
24	9	7	-2	2	11,5
25	3	1	-2	2	11,5
1	8	5	-3	3	18,5
5	13	10	-3	3	18,5
7	9	6	-3	3	18,5
11	4	7	3	3	18,5
17	4	7	3	3	18,5
30	7	10	3	3	18,5
26	3	7	4	4	22
2	3	8	5	5	23,5
29	9	4	-5	5	23,5
14	13	6	-7	7	25

T_{эмп} (сумма рангов нетипичных сдвигов) = **151,5**.

Для **N = 25** (объем выборки – нулевые сдвиги)

T_{кр 0,05} = 100,
T_{кр 0,01} = 76.

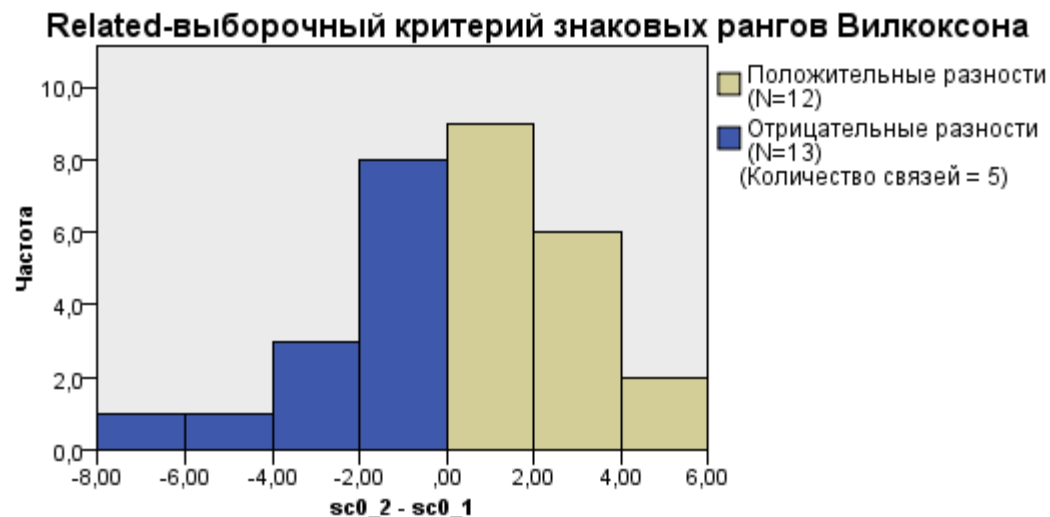
Для данного критерия, если **T_{эмп} < T_{кр}**, различия можно считать значимыми на соответствующем уровне.

Так как

T_{эмп} > T_{кр 0,05} > T_{кр 0,01},

принимается **нулевая гипотеза**, изменения в уровне признака от первого тестирования ко второму статистически не значимы.

Результаты обработки данных примера в программе SPSS



Всего N	30
Статистика критерия	151,500
Стандартная ошибка	36,868
Стандартизованная статистика критерия	-,298
Асимптотическая знч. (2-сторонний критерий)	,765