Статистические критерии различий

- **Статистический критерий** – это решающее правило, обеспечивающее принятие истинной и отклонение ложной гипотезы с высокой вероятностью.

  Статистический критерий подразумевает также метод расчета определенного числа - **эмпирического значения критерия** ($Ч_{эмп}$).

Критерии имеют свою специфику и различаются между собой **по различным основаниям:**

1. Тип измерительной шкалы.
2. Зависимость или независимость выборок.
3. Количество сравниваемых выборок.
4. Совпадение (несовпадение) объемов сравниваемых выборок.
5. Ограничения по объему охватываемой выборки.
6. Мощность (способность выявлять различия между выборками).

Критерий различия называют ***параметрическим,*** *если он основан на конкретном типе распределения генеральной совокупности (как правило, нормальном) или использует параметры этой совокупности (средние, дисперсии и т.д.) в формуле расчета.*

Критерий различия называют ***непараметрическим,*** *если он не базируется на предположении о типе распределения генеральной совокупности и не использует параметры этой совокупности.*

По- другому: «критерий, свободный от распределения».

Эти критерии основаны на оперировании частотами и рангами.

Параметрические критерии:

1) при нормальном распределении генеральной совокупности обладают большей мощностью по сравнению с непараметрическими (т.е. они способны с большей достоверностью отвергать нулевую гипотезу, если она неверна);

2) им следует отдавать предпочтение когда выборки взяты из нормально распределенных генеральных совокупностей.

Понятие нормы в психологии многозначно.

1. *Норма* понимается как норматив, т.е. как эталон, на который необходимо равняться, оценивая по нему свое индивидуальное поведение (нормы питания, спортивные нормы и т.д.). Такие нормы (нормативы) являются условными и имеют значение только в определенной системе отсчета.
2. *Норма* также понимается как функциональный оптимум, подразумевающий протекание всех процессов в системе с наиболее возможной слаженностью, эффективностью и экономичностью.

Функциональная норма всегда индивидуальна, в ней лежит представление о неповторимости пути развития каждого человека, и ее нарушение определяется функциональными последствиями.

3. Третьей системой отсчета является норма, понимаемая как статистически среднее, наиболее часто встречающееся, массовое в явлениях.

«Нормальное» в статистическом смысле включает не только среднестатистическую величину, но и серию отклонений от нее в известном диапазоне.

Нормальный закон распределений лежит в основе измерений, разработки тестовых шкал и методов проверки гипотез.

Нормальное распределение играет большую роль в математической статистике, так как многие статистические методы предполагают, что анализируемые данные распределены нормально.

Нормальное распределение часто встречается в природе. Нормальное распределение характеризует такие случайные величины, на которые воздействует большое количество разнообразных факторов (ошибки, возникающие при измерениях, отклонения при стрельбе).

Например, если у испытуемых, выбранных случайным образом, измерять их рост, вес, интеллект, какие-либо свойства личности, а затем построить график частоты встречаемости показателей любой из этих величин, то мы получим распределение, у которого крайние значения встречаются редко, а от крайних значений к середине частота повышается.

График нормального распределения имеет вид симметричной, колоколообразной кривой.

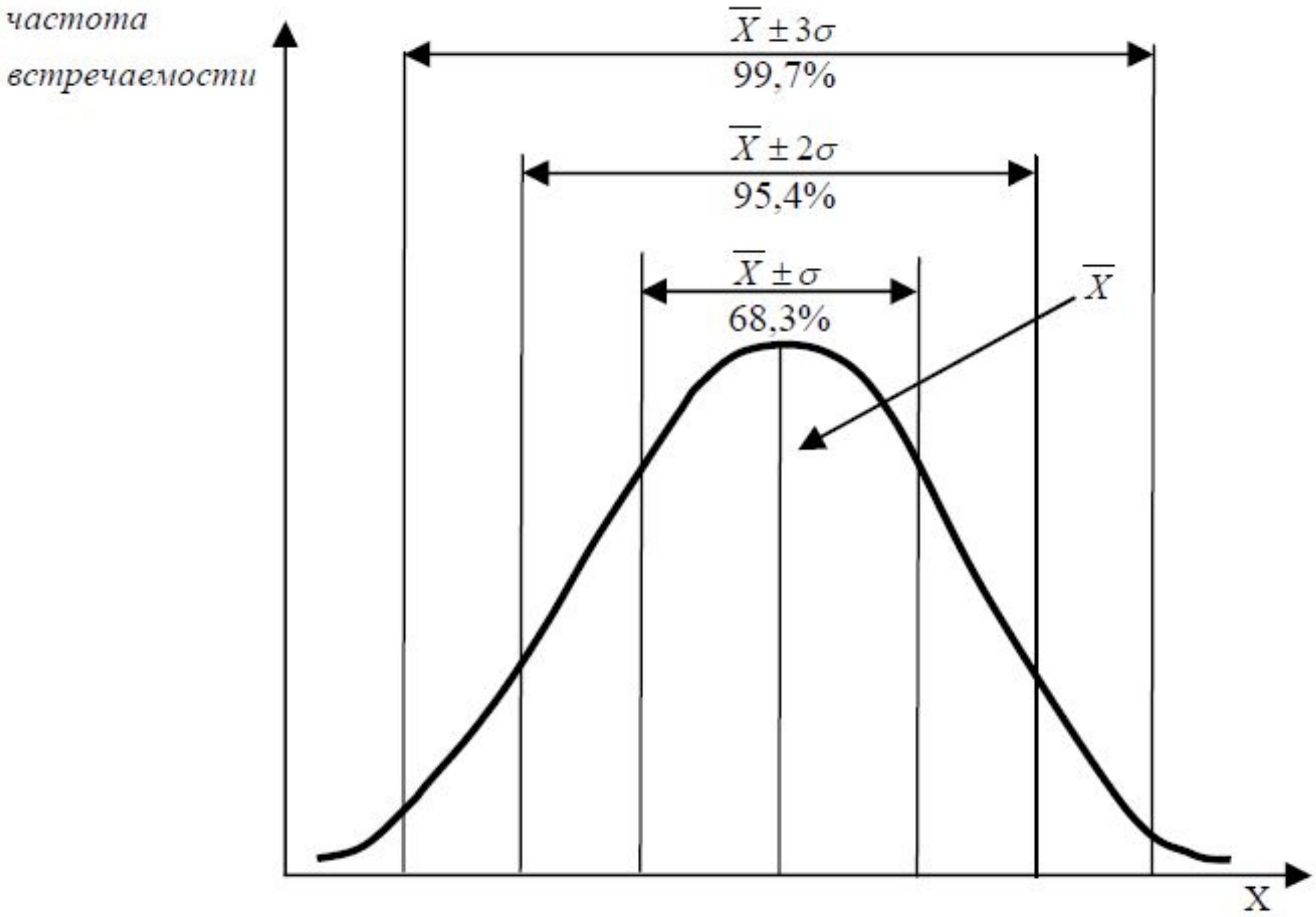


Рис. 1. График нормального распределения

В психологических исследованиях нормальное распределение используется при разработке и применении тестов интеллекта.

Отклонения показателей интеллекта следуют закону нормального распределения.

Применительно к другим психологическим категориям и сферам (личностная, мотивацион-ная) применение закона нормального распределения является *дискуссионным.*

Существует множество критериев проверки соответствия изучаемого распределения нормальному.

Наиболее простой критерий: если мода, медиана и среднее арифметическое равны, то ряд имеет нормальное распределение.

Наиболее эффективным критерием при проверке нормальности распределения считается критерий Колмогорова-Смирнова.

Преимущества непараметрических критериев:

- *при оценке различий в распределениях, далеких от нормального, непараметрические критерии могут выявить значимые различия, в то время как параметрические критерии таких различий не обнаружат;*
- *непараметрические критерии выявляют значимые различия и в том случае, если распределение близко к нормальному;*
- *при вычислениях вручную непараметрические критерии являются значительно менее трудоемкими, чем параметрические;*
- подавляющее большинство данных, получаемых в психологических экспериментах, ***не распределены нормально.***

Критерий включает в себя:

- формулу расчета эмпирического значения критерия ($Ч_{эмп}$) по выборочным данным;
- правило (формулу) определения числа степеней свободы;
- теоретическое распределение для данного числа степеней свободы;
- правило соотнесения эмпирического значения критерия с теоретическим распределением для определения вероятности того, что проверяемая гипотеза верна.

Число степеней свободы – количество возможных направлений изменчивости признака.

Нахождение числа степеней свободы для каждого признака имеет свои специфические особенности.

Как правило, число степеней свободы линейно зависит от объема выборки, от числа признаков или их градаций: чем больше эти показатели, тем больше число степеней свободы.

Каждая формула для расчета эмпирического значения критерия обязательно сопровождается правилом (формулой) для определения числа степеней свободы.

Этапы подготовки исследования:

1. Определить, является ли выборка связной (зависимой) или несвязной (независимой).
2. Определить однородность–неоднородность выборки.
3. Оценить объем выборки и, зная ограничения каждого критерия по объему, выбрать соответствующий критерий.
4. Целесообразнее всего начинать работу с выбора наименее трудоемкого критерия.
5. Если используемый критерий не выявил различия – следует применить более мощный, но одновременно и более трудоемкий критерий.

6. Если в распоряжении исследователя имеется несколько критериев, то следует выбирать те из них, которые наиболее полно используют информацию, содержащуюся в экспериментальных данных.

7. При малом объеме выборки следует увеличивать величину уровня значимости (не менее 1%), так как небольшая выборка и низкий уровень значимости приводят к увеличению вероятности принятия ошибочных решений.

Оценка достоверности сдвига:

- G-критерий знаков;
- парный Т-критерий Вилкоксона;
- критерий χ^2 - Фридмана;
- *L*- критерий Пейджа;
- t-критерий Стьюдента для зависимых выборок.

Критерий знаков G

***<u>Назначение критерия</u>*.**

- Он предназначен для установления общего направления сдвига исследуемого признака. Позволяет установить, в какую сторону в выборке в целом изменяются значения признака при переходе от первого измерения ко второму.

Пример. Будет ли тренинг способствовать повышению показателей по методике «Шкала социального интереса»?

Результаты диагностики «до» и «после» воздействия:

№	Имя	До	После
1	Ира А.	3	8
2	Наташа О.	5	8
3	Оля Е.	5	9
4	Аня К.	8	9
5	Лида Д.	6	7
6	Максим У.	4	8
7	Ольга А.	8	9
8	Аня И.	3	3
9	Вера П.	5	6
10	Маша И.	5	8

Определим «сдвиг»,как разность между показателями каждого участника «после» и «до» тренинга: «после»- «до»

№	Имя	До	После	Сдвиг
1	Ира А.	3	8	+5
2	Наташа О.	5	8	+3
3	Оля Е.	5	9	+4
4	Аня К.	8	9	+1
5	Лида Д.	6	7	+1
6	Максим У.	4	8	+4
7	Ольга А.	8	9	+1
8	Аня И.	3	3	0
9	Вера П.	5	6	+1
10	Маша И.	5	8	+3

Подсчитаем общее число нулевых, положительных и отрицательных сдвигов:

- общее число нулевых сдвигов – 1;
- общее число положительных сдвигов – 9;
- общее число отрицательных сдвигов – 0.

- Преобладающие сдвиги назовем **типичными сдвигами;** их количество обозначается буквой ***n***.
- Сдвиги более редкого, противоположного направления – **нетипичными сдвигами;** их количество обозначается как $\mathbf{G_{эмп}}$.

В нашем случае количество типичных сдвигов $\boldsymbol{n = 9}$, а нетипичных сдвигов –

$\mathbf{G_{эмп} = 0}$.

Сформулируем статистические гипотезы:

- H_0 – преобладание типичного направления сдвига является случайным.
- H_1 - преобладание типичного направления сдвига не является случайным.

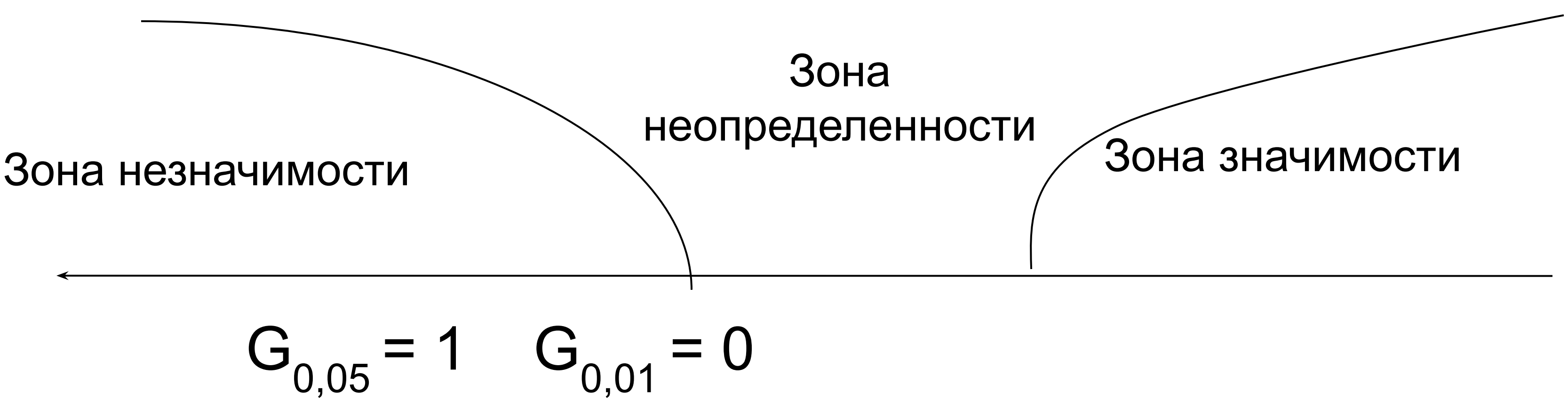
Оценка статистической достоверности сдвига по критерию G-знаков производится по таблице 1 Приложения.

В нашем примере n = 9, поэтому наша часть таблицы выглядит следующим образом:

n	*p*	
	0,05	0,01
9	1	0

Построим «ось значимости», на которой расположим критические значения

$G_{0,05} = 1$, $G_{0,01} = 0$ и эмпирическое значение $G_{эмп} = 0$. (В G-критерии ось перевернута!)



$G_{эмп}$ совпало с критическим значением зоны значимости $G_{0,01} = 0$.

Выводы:

1.Гипотеза H_0 отклоняется и принимается гипотеза H_1 о том, что сдвиг показателей после тренинга является не случайным.

2.Полученный в результате эксперимента сдвиг показателей статистически значим на уровне Р = 0,01.

3. Тренинг способствовал увеличению показателей по методике «Шкала социального интереса» статистически достоверно.

Условия применимости G-критерия:

1. Измерение может быть проведено в шкале порядка, интервалов и отношений.
2. Выборка должна быть однородной и связной.
3. Объем сравниваемых выборок должен быть одинаковым.
4. G-критерий знаков может применяться при величине типичного сдвига от 5 до 300.
5. G-критерий знаков достаточно эффективен при больших объемах выборок.
6. При равенстве количества типичных и нетипичных сдвигов критерий знаков неприменим.

Парный критерий Т-Вилкоксона

Назначение критерия

- Критерий Т-Вилкоксона применяется для оценки различий экспериментальных данных, полученных в двух разных условиях на одной и той же выборке испытуемых.
- Он позволяет выявить не только направленность изменений, но и позволяет установить насколько сдвиг показателей в каком-то одном направлении является более интенсивным, чем в другом.

- **Пример.** Способствовала ли коррекционная работа снижению реактивной тревожности участников эксперимента?

Показатели реактивной тревожности по методике Ч.Д. Спилбергера

№	Имя	До	После
1	Саша К.	69	51
2	Лена Р.	73	76
3	Ваня Е.	56	45
4	Оля С.	63	51
5	Оля А.	71	63
6	Даша К.	69	42
7	Алина Л.	69	57
8	Вова П.	71	63
9	Коля М.	70	61
10	Ира В.	71	60
11	Ваня Б.	67	68
12	Максим С.	54	49

Сформулируем статистические **гипотезы**:

- H_0 – интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении .
- H_1 - интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении .

№	Имя	До	После	Сдвиг	Модуль сдвига	Ранги модуля сдвига	Символ нетипич. сдвига
1	Саша К.	69	51	-18	18	11	
2	Лена Р.	73	76	+3	3	2	*
3	Ваня Е.	56	45	-11	11	(7) 7,5	
4	Оля С.	63	51	-12	12	(9) 9,5	
5	Оля А.	71	63	-8	8	(5) 4,5	
6	Даша К.	69	42	-27	27	12	
7	Алина Л.	69	57	-12	12	(10) 9,5	
8	Вова П.	71	63	-8	8	(4) 4,5	
9	Коля М.	70	61	-9	9	6	
10	Ира В.	71	60	-11	11	(8) 7,5	
11	Ваня Б.	67	68	+1	1	1	*
12	Маша С.	54	49	-5	5	3	
Сумма							

- Сдвиг в более часто встречающемся направлении назовем **типичным сдвигом**.
- Сдвиг в противоположном направлении – **нетипичным**.

- $T_{эмп}$ численно равно **сумме рангов нетипичных сдвигов**.
- В нашем случае нетипичных сдвигов два: +3 и +1. Их ранги равны 2 и 1 соответственно. Следовательно,

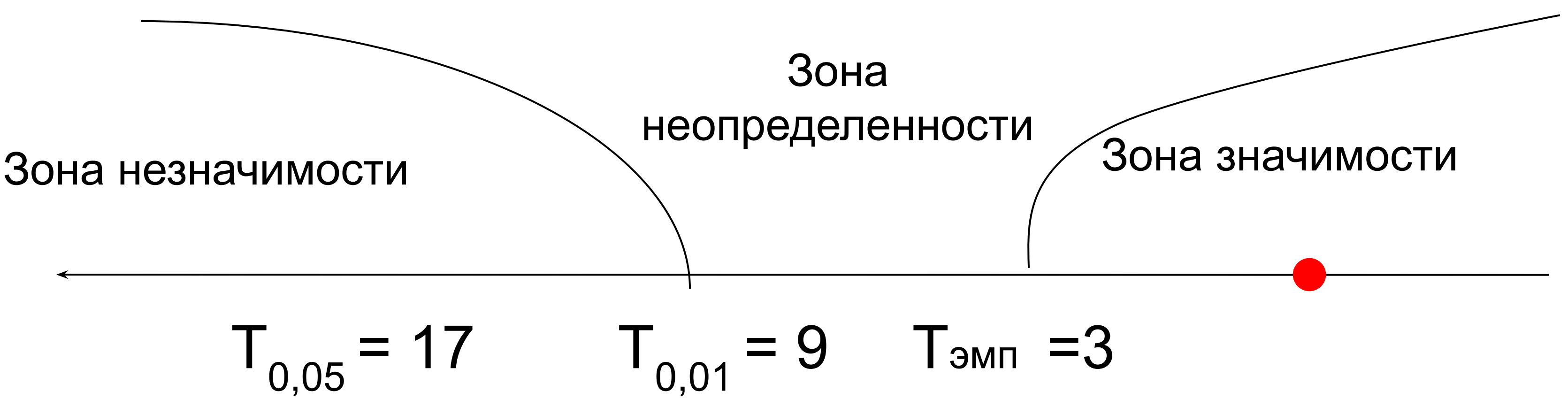
$$T_{эмп} = 2 + 1 = 3.$$

- Оценка статистической достоверности сдвига по Т-критерию производится по таблице 2 Приложения.
- Поиск критических величин по таблице ведется по числу испытуемых. В нашем примере n = 12, поэтому наша часть таблицы выглядит следующим образом

n	***p***	
	0,05	0,01
12	17	9

Построим «ось значимости», на которой расположим критические значения

$T_{0,05} = 17$, $T_{0,01} = 9$ и эмпирическое значение $T_{эмп} = 3$. (В Т-критерии ось перевернута!)



- Полученная величина Тэмп попала в зону значимости.
- Гипотеза H_0 отклоняется и принимается гипотеза H_1 о том, что сдвиг показателей после коррекционной работы является не случайным.
- Полученный в результате эксперимента сдвиг показателей статистически значим на уровне $p < 0,01$.
- Коррекционная работа способствовала снижению реактивной тревожности участников эксперимента статистически достоверно.

Условия применимости критерия Т-Вилкоксона:

1. Измерение может быть проведено во всех шкалах, кроме номинальной.
2. Выборка должна быть связной.
3. Объем сравниваемых выборок должен быть одинаковым.
4. Критерий Т- Вилкоксона может применяться при численности выборки от 5 до 50.

Оценка достоверности различий:

- Q- критерий Розенбаума;
- U- критерий Манна-Уитни;
- ϕ -критерий (угловое преобразование Фишера).

Критерий Q Розенбаума

Назначение критерия

- Критерий используется для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного.
- В каждой из выборок должно быть не менее 11 испытуемых.

Замечание.

1. Если критерий не выявляет достоверных различий, это еще <u>не означает</u>, <u>что их действительно нет.</u>
2. Если же Q- критерий выявляет достоверные различия между выборками с уровнем значимости $P \geq 0.01$, можно ограничиться только им и избежать трудностей применения других критериев.

Критерий применяется в тех случаях, когда данные представлены, по крайней мере, в порядковой шкале. Признак должен варьировать в каком-то диапазоне, иначе сопоставления с помощью Q-критерия просто невозможны.

- Применение критерия начинают с того, что упорядочивают значения признака в обеих выборках по нарастанию (или убыванию) признака.
- При этом рекомендуется первым рядом (выборкой, группой) считать тот ряд, где значения выше, а вторым рядом – тот, где значения ниже.

.

Гипотезы:

- H_0: Уровень признака в выборке 1 не превышает уровня признака в выборке 2.
- H_1 : Уровень признака в выборке 1 превышает уровень признака в выборке 2.

Условия использования критерия:

1) Измерение может быть проведено в шкале порядка, интервалов и отношений.
2) Выборки должны быть **<u>независимыми.</u>**
3) В каждой из выборок должно быть не меньше 11 испытуемых.
4) Приведенная в пособии таблица ограничивает верхний предел выборки 26 испытуемыми.
5) При числе наблюдений n_1, $n_2 \geq 26$ можно пользо-ваться следующими величинами *:*

   $Q_{кр1} = 8$ *если* $P \leq 0{,}05$ *;* $Q_{кр2} = 10$ *если* $P \leq 0{,}01$ *.*

6. Принципиальным условием, дающим возможность применять критерий, является наличие **«хвостов»** в сравниваемых рядах .

Замечание. В случае расположения выборок следующим образом (один из двух рядов имеет два «хвоста»):

x x x x x x x x x x x x x x

y y y y y y y

критерий Q-Розенбаума   неприменим!.

Работа с критерием Розенбаума предполагает подсчет так называемых **«хвостов».** Потому этот критерий имеет также название — «критерий хвостов».

$$|t\ \ t\ \ t\ \ t|\ t\ \ t\ \ t\ \ t\ \ t\ \ t$$

$$z\ \ z\ \ z\ \ z\ \ |z\ z\ z\ z|$$

$S_1=6$, $S_2=4$

$Q_{эмп}= S_1 +S_2$

Алгоритм *подсчета критерия* Q *Розенбаума*:

1.Проверить, выполняются ли ограничения: n_1, $n_2 \geq 11$

$n_1 \approx n_2$.

2. Упорядочить значения отдельно в каждой выборке по степени возрастания признака. Считать выборкой 1 ту выборку, значения в которой предположительно выше (правее), а выборкой 2 – ту, где значения предположительно ниже (левее).
3. Определить самое высокое (максимальное) значение в выборке 2.
4. Подсчитать количество значений в выборке 1, которые выше максимального значения в выборке 2. Обозначить полученную величину как S_1 .

5. Определить самое низкое (минимальное) значение в выборке 1.

6. Подсчитать количество значений в выборке 2, которые ниже минимального значения выборки 1. Обозначить полученную величину как S_2 .

7. Посчитать по формуле: $Q_{эмп} = S_1 + S_2$

8. По таблице 8 Приложения определить Qкр для данных n_1 и n_2 . Если $Q_{эмп} \geq Q_{кр\ 0,05}$, то H_0 - отвергается.

9. При n_1, $n_2 \geq 26$ сопоставить полученное $Q_{эмп}$ с

$Q_{кр1}=8$ если $P \leq 0,05$;

$Q_{кр2}=10$ если $P \leq 0,01$.

Если $Q_{эмп}$ превышает или, по крайней мере, равняется *$Q_{кр1}=8$* , то H_0 - отвергается.

Задача. Будут ли обнаружены статистически достоверные различия в показателях ситуативной тревожности между подростками с делинквентным (асоциальным, противоправным) поведением и подростками без отклоняющегося поведения?

№	Ситуативная тревожность	
	Подростки с делинквентным поведением	Подростки без отклоняющегося поведения
1	36	38
2	36	40
3	39	41
4	32	36
5	34	37
6	40	42
7	42	45
8	42	48
9	24	36
10	53	54
11	44	46

1. Упорядочим числа в порядке возрастания.
2. Разместим два сравниваемых ряда таким образом, чтобы равные элементы находились друг под другом

			36	36	37	38		40	41	42			45	46	48		54
24	32	34	36	36			39	40		42	42	44				53	

Сформулируем **статистические гипотезы**:

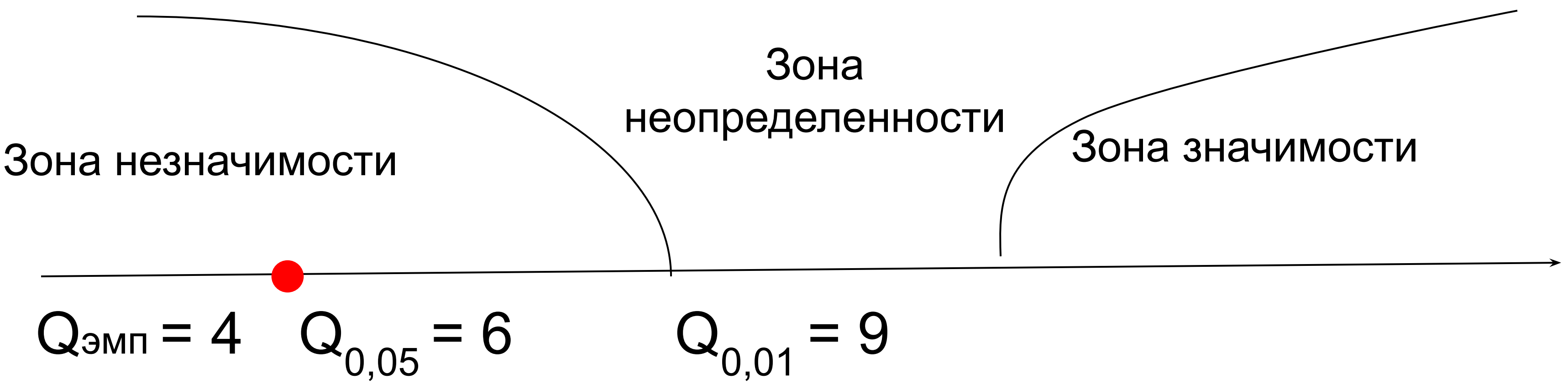
- H_0 – отсутствуют статистически достоверные различия между группами.
- H_1 – существуют статистически достоверные различия между группами.

1. Подсчитаем правый (S_1) и левый (S_2) «хвосты». Величина S_1 равна числу элементов первого ряда, которые находятся справа и не имеют совпадающих элементов второго ряда. Величина S_2 – числу элементов второго ряда, находящихся слева и не имеющих совпадающих элементов первого ряда.
2. В нашем случае $S_1 = 1$, а $S_2 = 3$.
3. $Q_{эмп} = S_1 + S_2 = 1+3=4$

- Критические значения для критерия Q-Розенбаума находим по таблице 8 Приложения.
- Поиск критических величин ведется по числу испытуемых n_1=11, n_2=11. Определяем что $Q_{0,05}$ = 6; $Q_{0,01}$ = 9.

Построим «ось значимости», на которой расположим критические значения

$Q_{0,05} = 6$, $Q_{0,01} = 9$ и эмпирическое значение $Q_{эмп} = 4$.



- Полученная величина $Q_{эмп}$ попала в зону незначимости.
- Принимается гипотеза H_0 о том, что отсутствуют статистически достоверные различия между группами.
- Статистически достоверные различия в показателях ситуативной тревожности между подростками с делинквентным поведением и подростками без отклоняющегося поведения не выявлены.

Критерий *U* Вилкоксона-Манна-Уитни

Назначение критерия

- Критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного.
- Он позволяет выявлять различие между малыми выборками, когда $n_1, n_2 \geq 3$ или $n_1 = 2, n_2 \geq 5$ и является более мощным, чем критерий Розенбаума.

Гипотезы:

- H_0: Уровень признака в группе 2 не ниже уровня признака в группе1.
- H_1 : Уровень признака в группе 2 ниже уровня признака в группе1.

(1-*м рядом, выборкой, группой называется ряд значений, в котором значения, по предварительной оценке, выше*, *а 2-м рядом – тот, где они предположительно ниже*).

Условия применимости U-критерия:

1) Измерение должно быть проведено в шкале интервалов и отношений.
2) Выборки должны быть **независимыми.**
3) Нижняя граница применимости критерия $n_1, n_2 \geq 3$ или $n_1 = 2, n_2 \geq 5$.
4) Верхняя граница применимости критерия: $n_1, n_2 \leq 60$.

Алгоритм *подсчета U-критерия* :

1. Исходные данные расположить в таблице в двух столбцах в порядке возрастания (с пропусками). Количество строк в таблице n_1+n_2 .

X	Y	Ранги R(X)	Ранги R(Y)
		Сумма	Сумма

2. Проранжировать данные двух столбцов как одного, записывая ранги чисел первого столбца в столбец R(X), а ранги 2-го столбца – в столбец R(Y).
3. По каждому столбцу в отдельности подсчитать суммы рангов.

4. Проверить правильность ранжирования.
5. Наибольшая по величине ранговая сумма обозначается как R_{max} .
6. Определить значение $U_{эмп}$ по формуле:

$$U_{эмп} = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - R_{\max}$$

где n_1 - численное значение первой выборки,

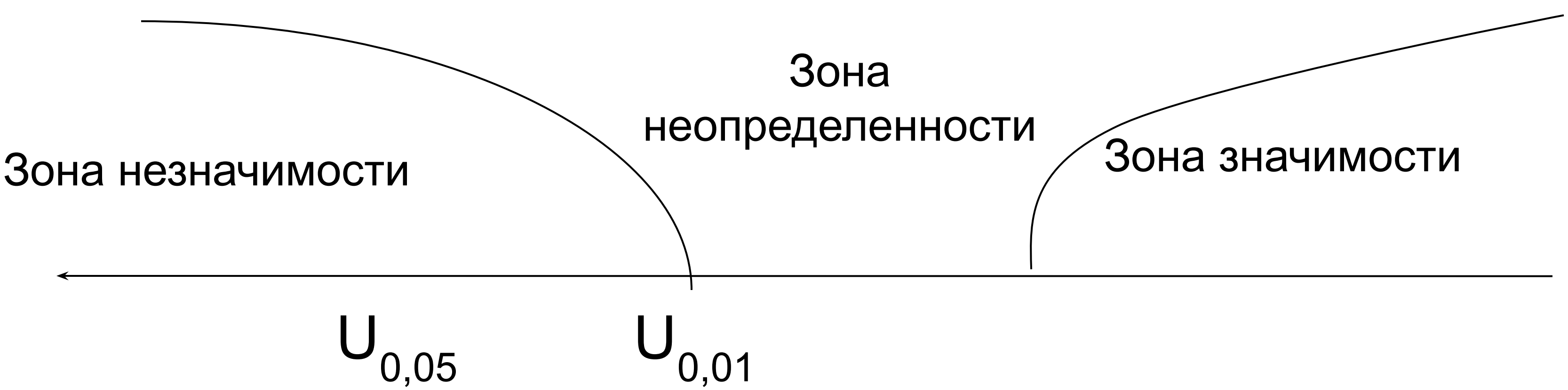
n_2 - численное значение второй выборки,

R_{max} - наибольшая по величине сумма рангов,

n_x- количество испытуемых в группе с большей суммой рангов.

7. Определить критические значения $U_{кр\ 0,05}$ и $U_{кр\ 0,01}$ по таблице 7 Приложения 1.
8. Построить «ось значимости», на которой расположить критические значения $U_{кр\ 0,05}$, $U_{кр\ 0,01}$ и эмпирическое значение $U_{эмп}$.

   (В U-критерии ось перевернута!)



9. Если $U_{эмп} > U_{кр\ 0,05}$ принимается гипотеза H_0. Если $U_{эмп} \leq U_{кр\ 0,05}$, то H_0 отвергается.

Чем меньше $U_{эмп}$, тем достоверность различий выше.

Задача:

Две неравные по численности группы испытуемых решали техническую задачу. Показателем успешности служило время решения. Испытуемые меньшей по численности группы получали дополнительную мотивацию в виде денежного вознаграждения. Психолога интересует вопрос — влияет ли вознаграждение на успешность решения задачи?

Психологом были получены следующие результаты времени решения технической задачи в секундах: в первой группе — с дополнительной мотивацией — 39, 38, 44, 6, 25, 25, 30, 43; во второй группе — без дополнительной мотивации — 46, 8, 50, 45, 32, 41, 41, 31, 55. Число испытуемых в первой группе обозначается как $n1$ и равно 8, во — второй как $n2$ и равно 9.

№ 1	№ 2	№ 3	№ 4
Группа с дополнительной мотивацией *X* ($n1 = 8$)	Группа без дополнительной мотивации *Y* ($n2 = 9$)	Ранги *X* $R(x)$	Ранги *Y* $R(y)$
6	—	1	—
—	8	—	2
25	—	(3) 3,5	—
25	—	(4) 3,5	—
30	—	(5) 5,5	—
—	30	—	(6) 5,5
—	32	—	7
38	—	8	—
41	—	(9) 10,5	—
—	41	—	(10) 10,5
—	41	—	(11) 10,5
41	—	(12) 10,5	—
44	—	13	—
—	45	—	14
—	46	—	15
—	50	—	16
—	55	—	17
Суммы рангов		55,5	97,5

4. Проверим правильность ранжирования:

55,5+97,5=153

N=8+9=17. $N\cdot(N+1)/2$=17·18/2=153

5. Наибольшая по величине ранговая сумма R_{max} =97,5

6. Определим значение $U_{эмп}$ по формуле:

$$U_{эмп} = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - R_{\max}$$

где n_1 - численное значение первой выборки,

n_2 - численное значение второй выборки,

R_{max} - наибольшая по величине сумма рангов,

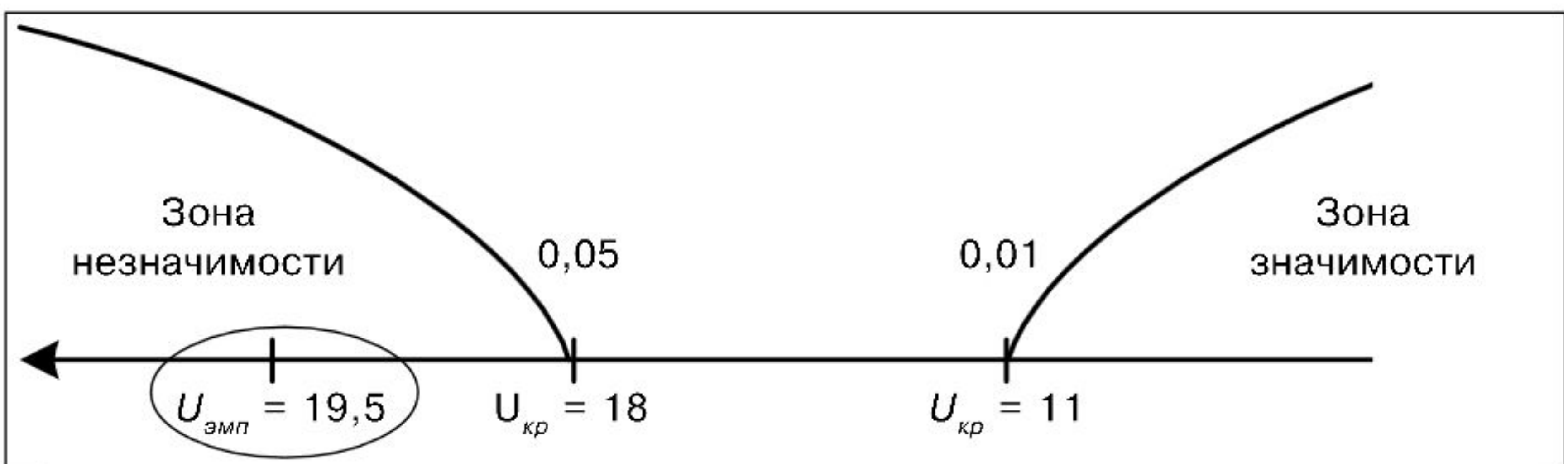
n_x- количество испытуемых в группе с большей суммой рангов.

Вычислим $U_{эмп}$.

$$U_{эмп} = (8 \cdot 9) + \frac{9 \cdot (9+1)}{2} - 97{,}5 = 19{,}5$$

Величины критических значений находим по таблице 7 Приложения.

Строим «ось значимости»:



Вывод:

1. Принимается гипотеза H_0 о сходстве.

H_0: Уровень признака в группе 2 не ниже уровня признака в группе1.

2. Психолог может утверждать, что мотивация не приводит к статистически значимому увеличению времени решения технической задачи.

Выявление различий в распределении признака:

При сопоставлении эмпирического распределения с теоретическим	критерий χ^2- Пирсона;
	λ- критерий Колмогорова - Смирнова;
	t-критерий Стьюдента.
При сопоставлении двух эмпирических распределений	критерий χ^2 - Пирсона;
	λ- критерий Колмогорова - Смирнова;
	критерий ϕ (угловое преобразование Фишера)

Критерий Пирсона χ^2 (*хи*-квадрат)

- один из наиболее часто использующихся в психологических исследованиях, поскольку он позволяет решать большое число разных задач, и, кроме того, исходные данные для него могут быть получены в любой шкале, начиная со шкалы наименований.

Назначение критерия хи-квадрат Пирсона

- Критерий отвечает на вопрос о том, с одинаковой ли частотой встречаются разные значения признака в эмпирическом и теоретическом распределениях или в двух и более эмпирических распределениях.

Критерий *хи*-квадрат используется в двух вариантах:

1) для расчета согласия эмпирического распределения и предполагаемого теоретического; в этом случае проверяется гипотеза об отсутствии различий между теоретическим и эмпирическим распределениями;

2) для расчета однородности двух независимых экспериментальных выборок; в этом случае проверяется гипотеза об отсутствии различий между двумя (тремя или более) эмпирическими (экспериментальными) распределениями одного и того же признака.

Критерий построен так, что при полном совпадении двух экспериментальных распределений величина $\chi^2_{эмп} = 0$,

и чем больше расхождение между сопоставляемыми распределениями, тем больше величина эмпирического значения *хи*-квадрат.

Гипотезы:

Первый вариант:

- H_0: Полученное эмпирическое распределение признака не отличается от теоретического (например, равномерного) распределения.
- H_1 : Полученное эмпирическое распределение признака отличается от теоретического распределения.

Второй вариант:

- H_0: Эмпирическое распределение 1 не отличается от эмпирического распределения 2.
- H_1 : Эмпирическое распределение 1 отличается от эмпирического распределения 2.

Третий вариант:

- H_0 : Эмпирические распределения 1, 2. 3. … не различаются между собой.
- H_1 : Эмпирические распределения 1, 2. 3. … различаются между собой

Условия применимости критерия χ^2 - Пирсона:

1. Измерение может быть проведено в любой шкале.
2. Выборки должны быть случайными и независимыми.
3. Желательно, чтобы объем выборки был ≥ 20. С увеличением объема выборки точность критерия повышается.
4. Теоретическая частота для каждого выборочного интервала <u>не должна быть меньше 5</u>.
5. Сумма наблюдений по всем интервалам должна быть равна общему количеству наблюдений.
6. Таблица критических значений критерия χ^2 рассчитана для числа степеней свободы v , которое каждый раз рассчитывается по определенным правилам;

Для таблиц число степеней свободы определяется по формуле:

v=(k-1)(c-1) , где *k* - число строк, *c* - число столбцов.

Сравнение двух эмпирических распределений

Исходные данные двух эмпирических распределений для сравнения между собой могут быть представлены разными способами.

Наиболее простой из этих способов: так называемая «четырехпольная таблица». Она используется в тех случаях, когда в первой выборке имеются два значения (числа) и во второй выборке также два значения (числа).

- **Задача**. Одинаков ли уровень подготовленности учащихся в двух школах, если в первой школе из 100 человек поступили в вуз 82 человека и во второй школе из 87 человек поступили в вуз 44?
- ***Решение.*** Условия задачи можно представить в виде четырехпольной таблицы ячейки которой, обозначаются обычно как *A, B, C и D:*

Таблица 1

	Школа 1	Школа 2
Число поступивших в вуз	A 82	B 44
Число не поступивших в вуз	C 18	D 43
Сумма	100	87

Величину χ^2 подсчитаем по формуле:

$$\chi^2_{эмп} = \sum_{i=1}^{k} \frac{\left(f_э - f_m\right)^2}{f_m} \quad ,$$

где $f_э$ - эмпирическая частота,

f_m - теоретическая частота,

k- количество разрядов признака.

Согласно данным, представленным в таблице, в нашем случае имеется четыре эмпирические частоты, это соответственно 82, 44, 18 и 43.

Для того чтобы можно было использовать формулу расчета, необходимо для каждой из этих эмпирических частот найти соответственные «теоретические» частоты.

Из таблицы следует, что 18 и 43 человека из первой и второй школ соответственно не поступили в вуз.

Относительно этих величин подсчитывается величина *P*. Это так называемая ***доля признака***, или ***частота***.

В данном случае признаком явилось то, что выпускники не поступили в вуз.

Величина *P* подсчитывается по формуле

$$P = \frac{18+43}{100+87} = 0{,}33$$

Величина *P* позволяет рассчитать «теоретические» частоты для третьей строчки таблицы, которые обозначим как f_{m1} и f_{m2}.

Эти частоты показывают, сколько учащихся из первой и второй школ не должны были поступить в вуз. Они подсчитываются следующим образом:

для первой школы $f_{m1} = 0,33 \cdot 100 = 33$

для второй школы $f_{m2} = 0,33 \cdot 87 = 28,71$

Произведем расчет того, сколько учащихся должны были бы поступить в вуз из первой и второй школ с учетом полученных «теоретических» частот 33 и 28,71:

- для первой школы $f_{m3} = 100 - 33 = 67$
- для второй школы $f_{m4} = 87 - 28{,}71 = 58{,}29$

Запишем полученные «теоретические» частоты в новую таблицу

Таблица 2

	Школа 1	**Школа 2**
Число учащихся, которые должны были бы поступить в вуз	A $f_{m3} = 67$	B $f_{m4} = 58{,}29$
Число учащихся, которые не должны были бы поступить в вуз	C $f_{m1} = 18$	D $f_{m2} = 43$
Сумма	100	87

Вычислим χ^2 по формуле

$$\chi^2_{эмп} = \sum_{i=1}^{4} \frac{\left(f_{эi} - f_{mi}\right)^2}{f_{mi}} = \frac{\left(f_{э1} - f_{m1}\right)^2}{f_{m1}} + \frac{\left(f_{э2} - f_{m2}\right)^2}{f_{m2}} + \frac{\left(f_{э3} - f_{m3}\right)^2}{f_{m3}}$$

$$+ \frac{\left(f_{э4} - f_{m4}\right)^2}{f_{m4}} = \frac{(18-33)^2}{33} + \frac{(82-67)^2}{67} + \frac{(43-28{,}71)^2}{28{,}71} +$$

$$+ \frac{(44-58{,}29)^2}{58{,}29} = 20{,}9$$

из величин таблицы 1 вычитаются величины таблицы 2

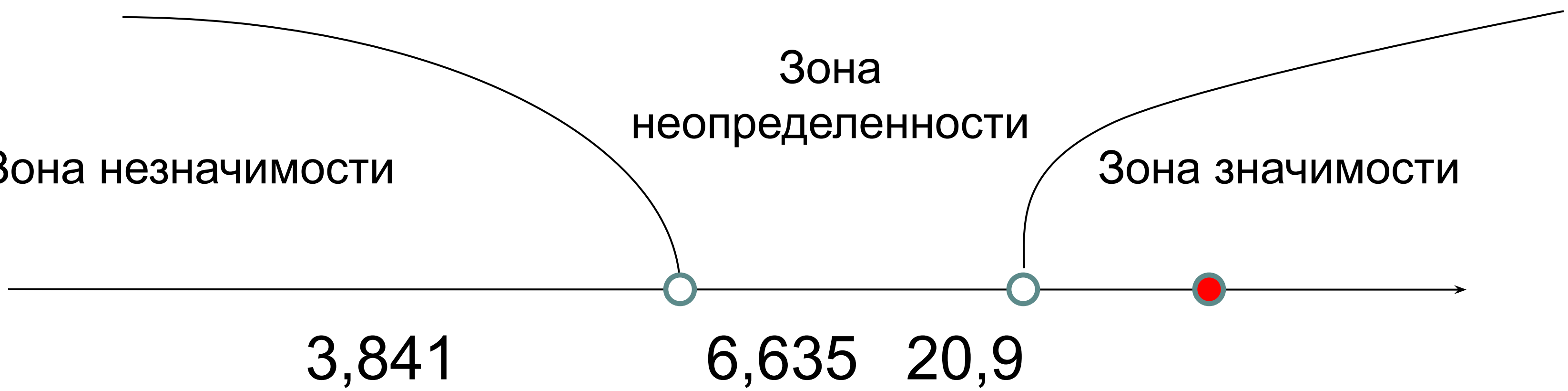
В данном случае число степеней свободы $v = (k\text{-}1)\cdot(c\text{-}1)$ подсчитывается как произведение числа столбцов минус 1 на число строк минус 1.

$v = (2\text{-}1)\Delta(2\text{-}1)=1$, поскольку у нас 2 строки и два столбца.

В соответствии с таблицей 12 Приложения 1 находим:

$$\chi^2_{кр1} = 3{,}841 \ (\mathrm{P} \le 0{,}05);$$
$$\chi^2_{кр2} = 6{,}635 \ (\mathrm{P} \le 0{,}01).$$

Построим «ось значимости», на которой расположим критические значения и полученное эмпирическое значение:



Полученная величина эмпирического значения *хи*-квадрат попала в ***зону значимости***.

Следует принять гипотезу H_1, о наличии различий между двумя эмпирическими распределениями.

Таким образом, уровень подготовленности учащихся в двух школах оказался разным.

На основе эмпирических данных мы можем теперь утверждать, что уровень подготовленности учащихся в первой школе существенно выше, чем во второй.

Без использования критерия *хи*-квадрат такого вывода мы сделать бы не могли.

Замечание. С помощью этого критерия можно решать задачу, в которой сравни-ваются две выборки, имеющие более чем по два значения в каждой.

Задание. Самостоятельно рассмотрите случаи в учебнике, когда сравниваются две выборки, имеющие по четыре значения в каждой.

Число переменных в сравниваемых выборках может быть достаточно большим.

В этом случае целесообразно использовать прием разбиения группировки значений по интервалам.

Алгоритм подсчета эмпирического значения критерия хи-квадрат (2 вариант) :

1. Составить интервальный ряд.

2. Произвести предварительные расчеты, необходимые для вычисления эмпирического значения критерия *хи*-квадрат.

При условии разного числа испытуемых в первой и второй выборках вычисления проводятся по формуле:

$$\chi^2_{эмп} = \frac{N^2}{n_1 \cdot n_2} \cdot \left(\sum_{i=1}^{k} \frac{f_1^2}{f_1 + f_2} - \frac{n_1^2}{N} \right)$$

f_1 - частоты первого распределения,

f_2 - частоты второго;

n_1 и n_2 – объемы первой и второй выборок;

$N = n_1 + n_2$.

При условии одинакового числа испытуемых в первой и второй выборках вычисления проводятся по формуле:

$$\chi^2_{эмп} = 4\sum_{i=1}^{k} \frac{f_1^2}{f_1 + f_2} - 2N$$

f_1 - частоты первого распределения,

f_2 - частоты второго;

N– объем выборок ($n_1 = n_2 = N$) .

3. Рассчитать число степеней свободы
$v = (k - 1) \cdot (c - 1)$, где k - число интервалов разбиения, а c – число выборок (у нас $c=2$).

4. В соответствии с таблицей 12 Приложения 1 определить критические значения $\chi^2_{кр}$ соответствующие уровням значимости Р=0,05 и Р= 0,01 для данного числа степеней свободы.

5. Построить «ось значимости», на которой расположить критические значения $\chi^2_{кр}$ и эмпирическое значение $\chi^2_{эмп}$.

6. По расположению на оси значимости $\chi^2_{эмп}$ принять статистическое решение (принять или отклонить гипотезу H_1).

7. Сформулировать содержательный вывод.

- Задача. Психолог сравнивает два эмпирических распределения, в каждом из которых было обследовано по тесту интеллекта разное количество испытуемых. Вопрос : различаются ли между собой эти два распределения?

Уровни интеллекта	Частоты		$f_1 \cdot f_1$	$f_1 + f_2$	$\frac{f_1^2}{f_1 + f_2}$
	f_1	f_2			
60	1	0	1	1	1,0000
70	8	0	64	8	8,0000
80	23	1	529	24	22,0417
90	30	11	900	41	21,9512
100	38	18	1444	56	25,7857
110	12	14	144	26	5,5385
120	7	3	49	10	4,9000
130	4	4	16	8	2,0000
140	1	1	1	2	0,5000
150	0	1	0	1	0,0000
Сумма	124	53			91,7171

Расчет эмпирического значения критерия $\chi^2_{эмп}$ при условии разного числа испытуемых в первой и второй выборках производится по формуле:

$$\chi^2_{эмп} = \frac{N^2}{n_1 \cdot n_2} \cdot \left(\sum_{i=1}^{k} \frac{f_1^2}{f_1 + f_2} - \frac{n_1^2}{N} \right)$$

где f_1 – частоты первого распределения;
f_2 – частоты второго распределения;
$N = n_1+n_2$ – сумма числа элементов в двух выборках.

Из вспомогательной таблицы получаем: $\chi^2_{эмп} = 23{,}11$

Число степеней свободы v=(k-1)·(c-1),

где *k* - число интервалов разбиения,

c - число столбцов.

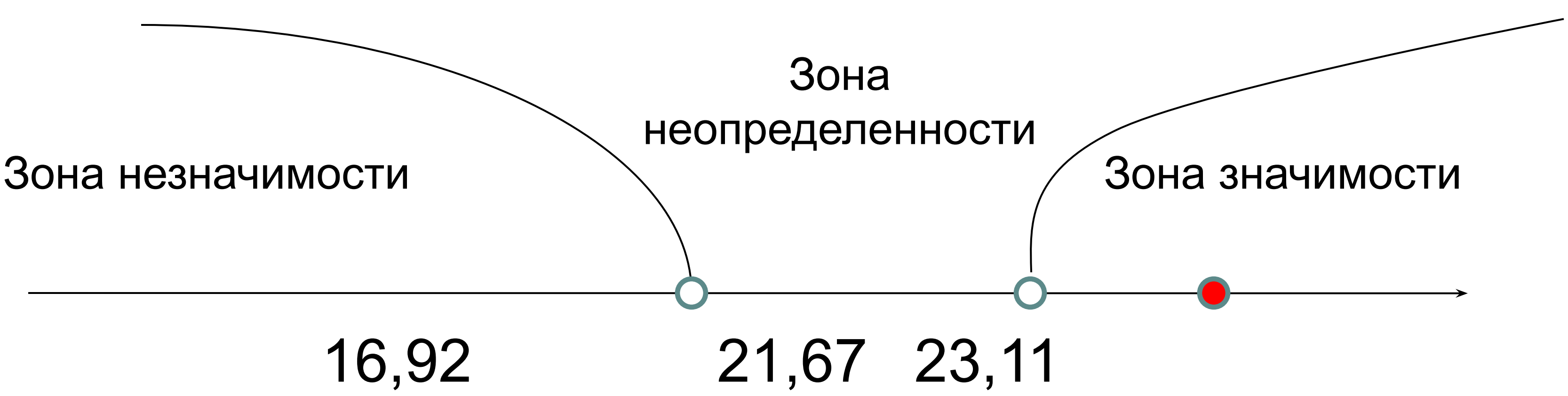
v=(10-1)·(2-1)=9.

В соответствии с таблицей 12 Приложения 1 находим:

$$\chi^2_{кр1} = 16{,}92 \ \ (\, P \le 0{,}05);$$

$$\chi^2_{кр2} = 21{,}67 \ \ (\, P \le 0{,}01).$$

Построим «ось значимости», на которой расположим критические значения и полученное эмпирическое значение:



Полученная величина эмпирического значения *хи*-квадрат попала в ***зону значимости***.

Следует принять гипотезу о том, что распределения уровней интеллекта в двух не равных по численности выборках статистически значимо отличаются между собой