МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВНУТРІШНІХ СПРАВ

Кафедра інформаційних технологій факультету № 4

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни «Математичні методи в психології»
обов’язкових компонент
освітньої програми першого рівня вищої освіти
053 "Психологія"

за темою – «G-критерій знаків. Т-критерій Вілкоксона»

Харків 2018

ЗАТВЕРДЖЕНО
Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 20.09.2018 № 9

СХВАЛЕНО
Вченою радою факультету № 4
Протокол від 19.09.2018 № 13

ПОГОДЖЕНО
Секцією Науково-методичної ради
ХНУВС з технічних дисциплін

Протокол від 20.09.2018 № 9

Розглянуто на засіданні кафедри інформаційних технологій *(протокол від 05.09.18 № 13)*

Розробники:
1. *Доцент, к.т.н., доцент Шеховцов С.Б.*

Рецензенти:
1. *Зав. кафедри, к.т.н., доцент Гнусов Ю.В.*

План лекції

1. Виявлення розходжень у рівні досліджуваної ознаки
2. Обґрунтування задачі зіставлення і порівняння
3. Q - критерій Розенбаума
4. G-критерий знаков
5. *Т*-критерий Вилкоксона

Література:

Основна:

1. Ермолаев О.Ю. Математическая статистика для психологов. –М.: Московский психолого-социальный институт: Флинта, 2004. – 336 с
2. Сидоренко Е.В. Методы математической обработки в психологии. СПб.: ООО «Речь», 2001. -350с.
3. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. - М.: Изд-во «Прогресс», 1976. – 496 с.
4. Конспект лекцій

Додаткова:

1. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. – К.: МОРИОН, 2001.– 408 с.
2. Кремер Н.Ш. Теория вероятностей и математическая статистика. -М.: Юнити-Дана, 2001. – 543 с.

Текст лекції

ТЕМА№2. Виявлення розходжень у рівні досліджуваної ознаки

2.1. Обґрунтування задачі зіставлення і порівняння

Дуже часто перед дослідником у психології стоїть задача виявлення розходжень між двома, трьома і більш вибірками випробуваних. Це може бути, наприклад, задача визначення психологічних особливостей хронічно хворих дітей у порівнянні зі здоровими, юних правопорушників у порівнянні з законослухняними однолітками чи розходжень між працівниками державних підприємств і приватних фірм, між людьми різної національності чи різної культури і, нарешті, між людьми різного віку в методі "поперечних зрізів".

Іноді по виявленим у дослідженні статистично достовірним розходженням формується «груповий профіль» чи «усереднений портрет» людини тієї чи іншої професії, статусу ін. ..

В останні роки всі частіше встає задача виявлення психологічного портрета фахівця нових професій: "успішного менеджера", "успішного політика", "успішного торгового представника", "успішного комерційного директора" і ін. Такого роду дослідження не завжди мають на увазі участь двох чи більш вибірок. Іноді обстежиться одна, але досить представницька вибірка кількістю не менш 60 чоловік, а потім усередині, цієї вибірки виділяються групи більш і менш успішних фахівців, і їхні дані по дослідженим перемінним порівнюються між собою. У найпростішому випадку критерієм для поділу вибірки на "успішних" і "неуспішних" буде середня величина по показнику успішності. Однак такий розподіл є досить грубим: особи, що одержали близькі оцінки по успішності, можуть виявитися в протилежних групах, а особи, що помітно розрізняються по оцінках успішності, - в одній групі. Це може викривити результати порівняння груп чи зробити розходження між групами менш помітними.

Щоб уникнути цього, можна спробувати виділити групи «успішних» і «неуспішних» фахівців більш строго, включаючи в першу з них тільки тих, чиї значення *перевищують* середню величину не менш чим на 1/4 стандартного відхилення, а в другу групу - тільки тих, чиї значення не менш чим на 1/4 стандартного відхилення *нижче* середньої величини. При цьому усі, хто виявляється в зоні середніх величин, $\bar{x} \pm \frac{1}{4}\sigma$, випадають з подальших зіставлень. Якщо розподіл близький до нормального, то випаде приблизно 19,8% випробуваних. Якщо розподіл відрізняється від нормального, то таких випробуваних може бути і більше. Щоб уникнути утрат, можна зіставляти не двох, а три групи випробуваних: з високою, середньою і низькою професійною успішністю.

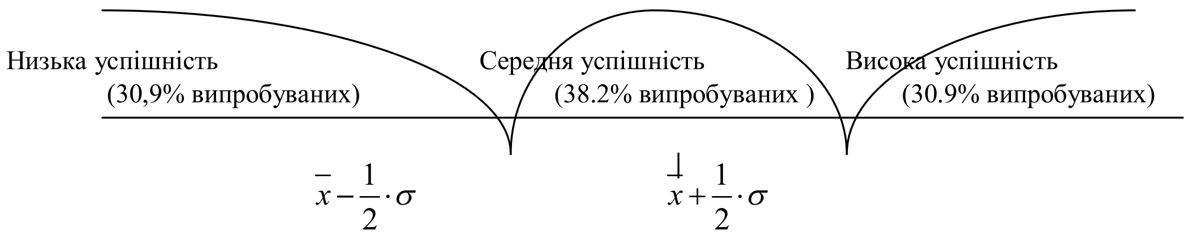


Рис 2.1. Схематичне зображення процесу поділу вибірки на групи з низькою, середньою і високою професійною успішністю

На Рис. 2.1 представлено схему поділу вибірки на групи з низькою, середньою і високою професійною успішністю за критерієм відхилення значень від середньої величини на 1/2 стандартного відхилення. При такому строгому критерії в "середню" групу попадають (при нормальному розподілі) близько 38,2% усіх випробуваних, а в крайніх групах виявляється по 30,9% випробуваних.

Чим менше випробуваних виявляється в групах, тим менше в нас можливостей для виявлення достовірних розходжень, тому що критичні значення більшості критеріїв при малих n суворіше, ніж при великих n.

Таким чином, при нестрогому поділі випробуваних на групи ми втрачаємо в точності, а при строгому - у кількості випробуваних. При рішенні задач виявлення розходжень у рівневих показниках варто пам'ятати, що "усереднений профіль успішного фахівця" повинний розглядатися скоріше як дослідницький результат, що дозволяє сформулювати гіпотези для подальших досліджень, а не як підстава для професійного добору.

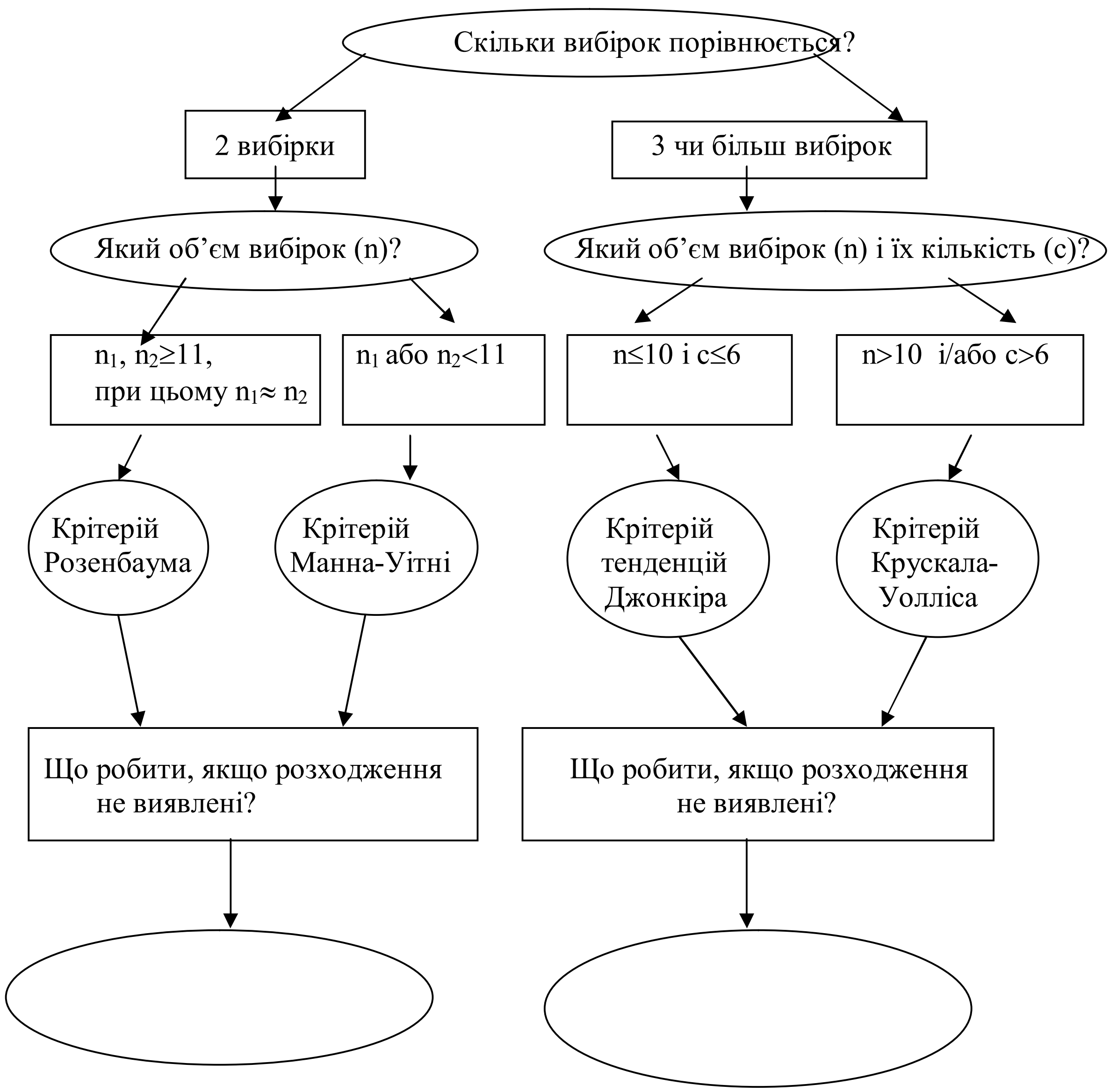
Тому є дві причини. По-перше, у жодного з успішних фахівців може не спостерігатися "усереднений профіль" - він, по суті, є абстрактним узагальненням; по-друге, у професійній діяльності наявність власного індивідуального стилю важніше відповідністі середньому профілю.

Р.Б. Кеттелл, з огляду на це, пропонував при дослідженні професійної успішності включати в розгляд індивідуальні профілі видатних представників тієї чи іншої професії .

Зіставлення рівневих показників у різних вибірках може бути необхідною частиною комплексних діагностичних, навчальних, психокоррекційних і інших програм. Воно допомагає нам звернути увагу на ті особливості обстежених вибірок, що повинні бути враховані і використані при адаптації програм до даної групи в процесі їхнього конкретного втілення.

Рішення про вибір того чи іншого критерію приймається на основі того, скільки вибірок зіставляється і який їхній обсяг (див. Алгоритм).

Алгоритм вибору критерію оцінки вірогідності розходжень
між незалежними вибірками по рівню ознаки



Використовувати кутове перетворення Фішера

Використовувати кутове перетворення Фішера зіставляючи по дві групи

2.2. Q - критерій Розенбаума

Призначення критерію

Критерій використовується для оцінки розходжень між двома вибірками за рівнем будь-якої ознаки, кількісно обмірюваній. У кожній з вибірок повинне бути не менш 11 випробуваних.

Опис критерію

Це дуже простий непараметричний критерій, що дозволяє швидко оцінити розходження між двома вибірками за якою-небудь ознакою. Однак якщо критерій Q не виявляє достовірних розходжень, це ще не означає, що їх дійсно немає.

У цьому випадку варто застосувати критерій φ^* Фішера. Якщо ж Q-критерій виявляє достовірні розходження між вибірками з рівнем значимості $p \leq 0,5$, можна обмежитися тільки їм і уникнути труднощів застосування інших критеріїв.

Критерій застосовується в тих випадках, коли дані представлені принаймні у порядковій шкалі. Ознака повинна варіювати в якомусь діапазоні значень, інакше зіставлення за допомогою Q-критерію просто неможливі. Наприклад, якщо в нас тільки 3 значення ознаки, 1, 2 і 3, - нам дуже важко буде встановити розходження. Метод Розенбаума вимагає, отже, досить тонко обмірюваних ознак.

Застосування критерію починаємо з того, що упорядковуємо значення ознаки в обох вибірках по наростанню (чи убуванню) ознаки. Найкраще, якщо дані кожного випробуваних представлені на окремій картці. Тоді нічого не варто упорядкувати два ряди значень по цікавлячому нас ознаці, розкладаючи картки на столі. Так ми відразу побачимо, чи збігаються діапазони значень, і якщо ні, то наскільки один ряд значень "вище" (S_1), а другий - "нижче" (S_2). Для того, щоб не заплутатися, у цьому й у багатьох інших критеріях рекомендується першим рядом (вибіркою, групою) вважати той ряд, де значення вище, а другим рядом - той, де значення нижче.

Гіпотези

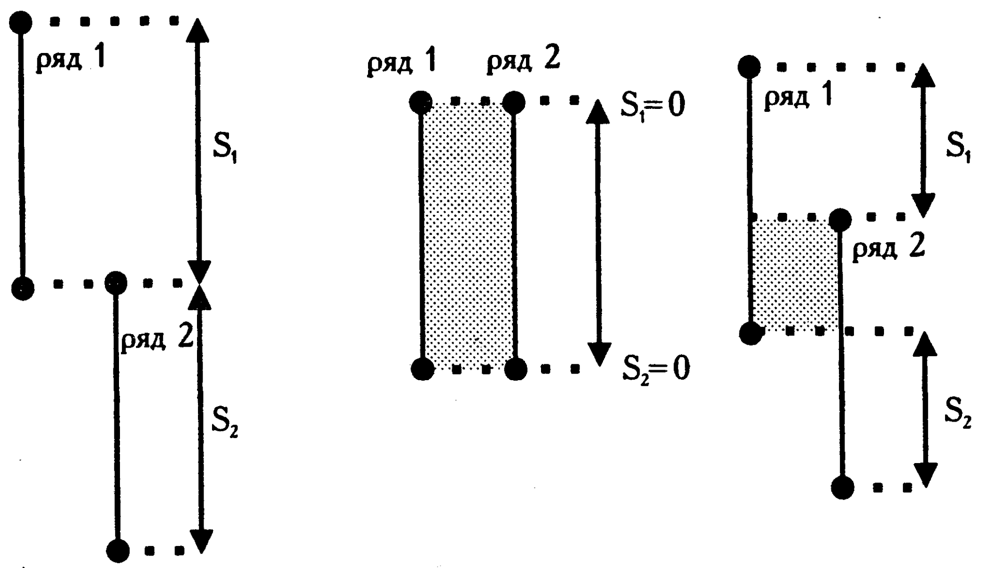
H_0: Рівень ознаки у вибірці 1 не перевищує рівня ознаки у вибірці 2.

H_1: Рівень ознаки у вибірці 1 перевищує рівень ознаки у вибірці 2.

Графічне представлення критерію Q

На Рис. 2.2. представлені три варіанти співвідношення рядів значень у двох вибірках. У варіанті (а) усі значення першого ряду вище всіх значень другого ряду. Розходження, безумовно, достовірні, при дотриманні умови, що $n_1, n_2 \geq 11$.

У варіанті (б), навпроти, обидва ряди знаходяться на тому самому рівні: розходження недостовірні. У варіанті (в) ряди частково перехрещуються, але все-таки перший ряд виявляється набагато вище другого, чи Досить великі зони S_1 і S_2, у сумі складові Q, можна визначити по Таблиці I Додатка 1, де приведені критичні значення Q для різних n. Чим величина Q більше, тим більше достовірні розходження ми зможемо констатувати.



а) б) в)

Рис.2.2. Три варіанти співвідношення рядів значень у двох вибірках.

Обмеження критерію Q

1. У кожній з вибірок, що зіставляються, повинне бути не менш 11 спостережень. При цьому обсяги вибірок повинні приблизно збігатися.

Існують наступні правила:

а) якщо в обох вибірках менше 50 спостережень, те абсолютна величина різниці між n_1 і n_2 не повинна бути більше 10 спостережень;

б) якщо в кожній з вибірок більше 51 спостереження, але менше 100, те абсолютна величина різниці між n_1 і n_2 не повинна бути більше 20 спостережень;

в) якщо в кожній з вибірок більше 100 спостережень, то допускається, щоб одна з вибірок була більше інший не більш ніж у 1,5-2 рази.

2. Діапазони розкиду значень у двох вибірках повинні не збігатися між собою, у противному випадку застосування критерію безглуздо. Тим часом, можливі випадки, коли діапазони розкиду значень збігаються, але, унаслідок різнобічної асиметрії двох розподілів, розходження в середніх величинах ознак істотний (Рис. 2.3., 2.4).

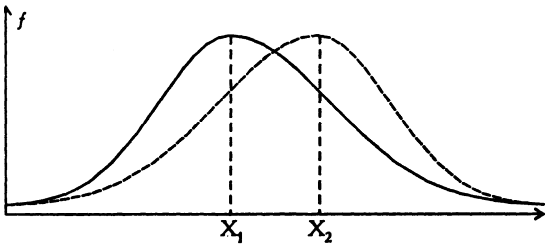


Рис. 2.3. Варіант співвідношення розподілів ознаки в двох вибірках, при якому критерій Q безпомічний

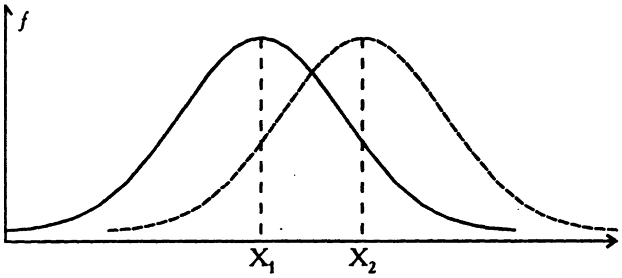


Рис. 2.4. Варіант співвідношення розподілів ознаки в двох вибірках, при якому критерій Q може бути могутнім

Підрахунок критерію Q Розенбаума

1. Перевірити, чи виконуються обмеження: $n_1, n_2 \geq 11$, $n_1 \approx n_2$
2. Упорядкувати значення окремо в кожній вибірці по ступені зростання ознаки. Вважати вибіркою 1 ту вибірку, значення в який приблизно вище, а вибіркою 2 - ту, де значення приблизно нижче.
3. Визначити найвище (максимальне) значення у вибірці 2.
4. Підрахувати кількість значень у вибірці 1, що вище максимального значення у вибірці 2. Позначити отриману величину як S_1.
5. Визначити найнижче (мінімальне) значення у вибірці 1.
6. Підрахувати кількість значень у вибірці 2, що нижче мінімальні значення вибірки 1. Позначити отриману величину як S_2.
7. Підрахувати емпіричне значення Q по формулі: $Q=S_1+S_2$.
8. По Табл. 2.1 визначити критичні значення Q для даних n_1 і n_2. Якщо Qэмп. дорівнює $Q_{0,05}$ чи перевищує його, H_0 відкидається.
9. При $n_1, n_2 > 26$ зіставити отримане емпіричне значення з $Q_{кр}=8$ ($p \leq 0.05$) і $Q_{кр}=10$ ($p \leq 0.01$). Якщо Qэмп, чи перевищує принаймні дорівнює $Q_{кр}=8$, H_0 відкидається.

Таблиця 2.1

Критичні значення критерію Q Розенбаума для рівнів статистичної значимості $p \leq 0,5$ і $p \leq 0,1$

n	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
$p=0,05$																
11	6															
12	6	6														
13	6	6	6													
14	7	7	6	6												
15	7	7	6	6	6											
16	8	7	7	7	6	6										
17	7	7	7	7	7	7	7									
18	7	7	7	7	7	7	7	7								
19	7	7	7	7	7	7	7	7	7							
20	7	7	7	7	7	7	7	7	7	7						
21	8	7	7	7	7	7	7	7	7	7	7					
22	8	7	7	7	7	7	7	7	7	7	7	7				
23	8	8	7	7	7	7	7	7	7	7	7	7	7			
24	8	8	8	8	8	8	8	8	8	8	7	7	7	7		
25	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7	
26	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7
$p=0,01$																
11	9															
12	9	9														
13	9	9	9													
14	9	9	9	9												
15	9	9	9	9	9											
16	9	9	9	9	9	9										
17	10	9	9	9	9	9	9									
18	10	10	9	9	9	9	9	9								
19	10	10	10	9	9	9	9	9	9							
20	10	10	10	10	9	9	9	9	9	9						
21	11	10	10	10	9	9	9	9	9	9	9					
22	11	11	10	10	10	9	9	9	9	9	9	9				
23	11	11	10	10	10	10	9	9	9	9	9	9	9			
24	12	11	11	10	10	10	10	9	9	9	9	9	9	9		
25	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9	
26	12	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9

Приклад

У учасників психологічного експерименту, що моделює діяльність повітряного диспетчера, був обмірюван рівень вербального і невербального інтелекту за допомогою методики Д. Векслера. Було обстежено 26 юнака у віці від 18 до 24 років (середній вік 20,5 років). 14 з них були студентами фізичного факультету, а 12 - студентами психологічного факультету Ленінградського університету (Сидоренко Е.В., 1978). Показники вербального інтелекту представлені в Табл. 2.2.

Чи можна затверджувати, що одна з груп перевершує іншу за рівнем вербального інтелекту?

Таблиця 2.2

Індивідуальні значення вербального інтелекту у вибірках студентів фізичного (n_1=14) і психологічного (n_2=12) факультетів

Студенти-фізики			Студенти - психологи		
	П.І.	Показник вербального інтелекту		П.І.	Показник вербального інтелекту
1.	И.А	132	1.	Н.Т.	126
2.	К.А.	134	2.	О.В.	127
3.	К.Е.	124	3.	Е.В.	132
4.	П.А.	132	4.	Ф.О.	120
5.	С.А.	135	5.	И.Н.	119
6.	Ст.А.	132	6.	И.Ч.	126
7.	Т.А.	131	7.	И.В.	120
8.	Ф.А.	132	8.	К.О.	123
9.	Ч.И.	121	9.	Р.Р.	120
10	Ц.А	127	10	Р.И.	116
11	См.А.	136	11.	О.К.	123

Студенти-фізики			Студенти - психологи		
	П.І.	Показник вербального інтелекту		П.І.	Показник вербального інтелекту
12.	К.Ан.	129	12.	Н.К.	115
13.	Б.Л.	136			
14	Ф.В.	136			

Упорядкуємо значення в обох вибірках, а потім сформулюємо гіпотези:

H_0; Студенти-фізики не перевершують студентів-психологів за рівнем вербального інтелекту.

H_1: Студенти-фізики перевершують студентів-психологів за рівнем вербального інтелекту.

Таблиця 2.3

Упорядковані ряди індивідуальних значень вербального інтелекту у двох студентських вибірках

1 ряд- студенти-фізики		2 ряд- студенти-психологи	
1. См. А.	136 --		
2. Б.Л.	136		
3. Ф.В.	136 S1		
4. С.А.	135		
5. К.А.	134		
6. И.А.	132	1. Е.В.	132
7. П.А.	132		
8. Ст.А.	132		
9. Ф.А. 10. Т.А.	132 131		
11. К.Ан.	129		
12. Ц.А.	127	2. О.В.	127
		3. Н.Т.	126
		4. И.Ч.	126
13 ДО.Е.	124		
		5. К.О.	123
		6. О.К.	123
14. Ч.И.	121		
		7. Ф.О.	120
		8. И.В.	120
		9. Р.Р.	120
		S2 10. И.Н.	119
		11. Р.И.	116
		12. Н.К.	115

Як видно з Табл. 2.3, ми правильно позначили ряди: перший, той, що "вище" - ряд фізиків, а другий, той, що "нижче" - ряд психологів.

По Табл. 2.3 визначаємо кількість значень першого ряду, що більше максимального значення другого ряду: S1=5.

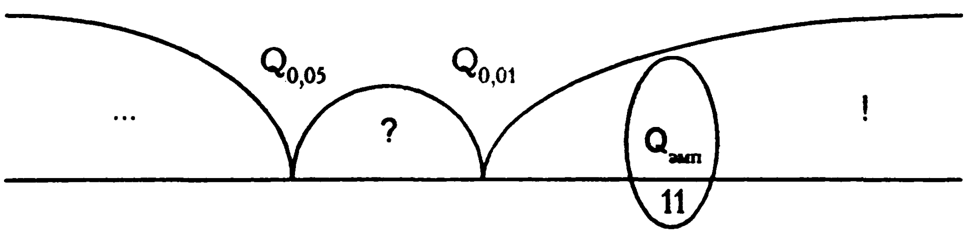
Тепер визначаємо кількість значень другого ряду, що менше мінімального значення першого ряду: S2=6.

Обчислюємо Qемп за формулою: Qемп = S1+ S2 =5+6 =11

За табл.2.1. визначаємо критичні значення Q для n_1=14, n_2=12: Qкр0,05=7 та Qкр0,01=9 .

Побудуємо "вісь значимості".

Зона незначимості (приймається H_0) | Зона невизначеності | Зона значимості (приймається H_1)



7 9

Qэмп>Qкр ($p \le 0{,}1$)

Відповідь: H_0 відхиляється.

Приймається H_1. Студента-фізики перевершують студентів-психологів за рівнем вербального інтелекту ($p \le 0{,}1$).

Відзначимо, що в тих випадках, коли емпіричне значення критерію виявляється на границі зони незначимості, ми маємо право стверджувати лише, що розходження достовірні при $p \le 0{,}5$, якщо ж воно знаходиться між двома критичними значеннями, то ми можемо стверджувати, що $p < 0{,}5$.

Якщо емпіричне значення критерію виявляється на границі, ми можемо стверджувати, що $p \le 0{,}1$, якщо воно попадає в зону значимості, ми можемо стверджувати, що $p \le 0{,}1$.

Оскільки рівень значимості виявлених розходжень досить високий ($p \le 0{,}1$), ми могли б на цьому зупинитися. Однак якщо дослідник сам психолог, а не фізик, навряд чи він на цьому зупиниться. Він може спробувати зіставити вибірки за рівнем невербального інтелекту, оскільки саме невербальний інтелект визначає рівень інтелекту в цілому і ступінь його організованості.

G-критерий знаков

Таблица 1

Показатели уровня аргументации до и после комплексной программы тренинга

Испытуемый	Уровень аргументации		Разность "после" - "до"
	до тренинга	после тренинга	
И.1	3	7	4
И.2	3	5	2
И.3	4	6	2
И.4	4	5	1
И.5	6	6	0
И.6	4	6	2
И.7	3	5	2
И.8	6	5	**-1**
И.9	6	5	**-1**
И.10	5	6	1
И.11	6	3	**-3**
И.12	6	5	**-1**

Типичный сдвиг – повышение уровня аргументации; *нетипичный* сдвиг – снижение уровня аргументации.

Этапы проверки гипотез.

1. Гипотеза H_0: преобладание типичного направления сдвига, т.е. сдвига в сторону повышения уровня аргументации испытуемых является ***случайным, незначимым***.

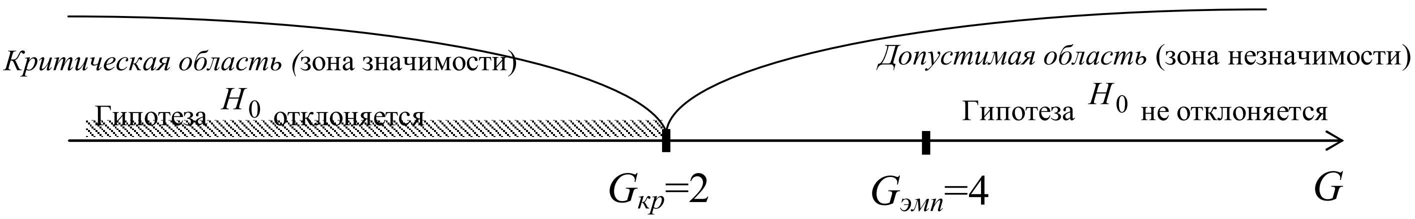
Гипотеза H_1: преобладание типичного направления сдвига, т.е. сдвига в сторону повышения уровня аргументации испытуемых не является ***случайным, значимым***.

2. G-критерий знаков. Уровень значимости α=0,05.

3. Эмпирическое значение критерия знаков $G_{эмп} = 4$ – количество **нетипичных** сдвигов.

4. Критическое значение критерия знаков $G_{кр} = G_{\alpha;n}$ для уровня значимости: α=0,05 и числа наблюдений $n = 11$ (**после исключения пар с нулевым сдвигом**) равно $G_{кр} = G_{0,05;11} = 2$.

5. Ось значимости, на которой изображается левосторонняя критическая область имеет вид:



6. **Принятие статистического решения**:

– если $G_{эмп} > G_{кр}$, то нулевая гипотеза H_0 не отклоняется;

– если $G_{эмп} \le G_{кр}$, то нулевая гипотеза H_0 отклоняется, принимается альтернативная гипотеза H_1.

7. **Вывод**. Так как $G_{эмп} = 4 > 2 = G_{kp}$, то гипотеза H_0 о случайности преобладания типичного направления сдвига, т.е. сдвига в сторону повышения уровня аргументации испытуемых не отклоняется при уровне значимости $\alpha=0,05$, т.е. согласуется с опытными данными.

T-критерий Вилкоксона

Таблица 2

Показатели уровня эмоционального напряжения до и после комплексной программы тренинга (исключены испытуемые, у которых уровень эмоционального напряжения не изменился)

Испытуемый	Уровень эмоц. напряж.		Разность "после" - "до"	Абсолютное значение разности	Ранговый номер разности
	до тренинга	после тренинга			
И.1	6	5	-1	1	3
И.2	4	1	-3	3	10,5
И.3	7	4	-3	3	10,5
И.4	5	4	-1	1	3
И.5	5	4	-1	1	3
И.6	7	5	-2	2	7
И.7	8	5	-3	3	10,5
И.8	7	5	-2	2	7
И.9	6	7	1	1	**3**
И.10	7	6	-1	1	3
И.11	4	6	2	2	**7**
И.12	6	3	-3	3	10,5
Сумма					**78**

Типичный сдвиг – снижение уровня эмоционального напряжения; *нетипичный* сдвиг – повышение уровня эмоционального напряжения.

Этапы проверки гипотез.

1. Гипотеза H_0: интенсивность сдвигов в сторону снижения уровня эмоционального напряжения испытуемых ***не превышает*** интенсивность сдвигов в сторону повышения уровня эмоционального напряжения.

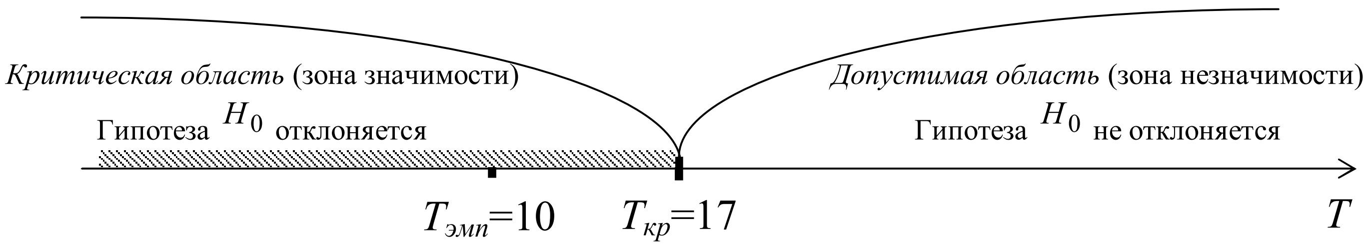
Гипотеза H_1: интенсивность сдвигов в сторону снижения уровня эмоционального напряжения ***превышает*** интенсивность сдвигов в сторону повышения уровня эмоционального напряжения.

2. *T*-критерий Вилкоксона. Уровень значимости: α=0,05.

3. Эмпирическое значение *T*–критерия Вилкоксона $T_{эмп} = 3 + 7 = 10$ – сумма рангов **нетипичных** сдвигов.

4. Критическое значение *T*–критерия Вилкоксона $T_{кр} = T_{\alpha;n}$ для уровня значимости: α=0,05 и числа наблюдений $n = 12$ (**после исключения пар с нулевым сдвигом**) равно $T_{кр} = T_{0,05;12} = 17$.

5. Ось значимости, на которой изображается левосторонняя критическая область имеет вид:



6. **Принятие статистического решения**:

– если $T_{эмп} > T_{кр}$, то нулевая гипотеза H_0 не отклоняется;

– если $T_{эмп} \le T_{кр}$, то нулевая гипотеза H_0 отклоняется, принимается альтернативная гипотеза H_1.

7. **Вывод**. Так как $T_{эмп} = 10 < 17 = T_{kp}$, то гипотеза H_0 отклоняется при уровне значимости α=0,05, т.е. интенсивность сдвигов в сторону снижения уровня эмоционального напряжения превышает интенсивность сдвигов в сторону повышения уровня эмоционального напряжения.