

Оценка достоверности сдвига в значении исследуемого признака

Лекция № 7

План

1. Обоснование задачи исследований изменений
2. G- критерий знаков
 - 2.1. Назначение критерия G
 - 2.2. Описание критерия G
 - 2.3. Графическое представление критерия знаков
 - 2.4. Ограничения критерия знаков
 - 2.5. Алгоритм расчета критерия знаков G
3. T - критерий Вилкоксона
 - 3.1. Назначение критерия
 - 3.2. Описание критерия T
 - 3.3. Графическое представление критерия T
 - 3.4. Ограничения в применении критерия T Вилкоксона
 - 3.5. Алгоритм подсчета критерия T Вилкоксона

Литература

1. Загвязинский, В.И. Методология и методы психолого-педагогического исследования : учеб. пособие для студентов пед. вузов по спец.031000 – Педагогика и психология / В.И. Загвязинский – М. : Академия, 2001. – 202 с.
2. Сидоренко Е.В. Методы математической обработки в психологии / Е. В. Сидоренко – СПб.: Речь, 2010. – С. 11 – 34.

Обоснование задачи исследований изменений

В психологических исследованиях часто бывает важно доказать, что в результате действия каких-либо факторов произошли достоверные изменения ("сдвиги") в измеряемых показателях. К числу таких факторов должен быть отнесен прежде всего фактор времени. Сопоставлен показателей, полученных у одних и тех же испытуемых по одним и же методикам, но в разное время, дает нам *временной сдвиг*.

Многokратные обследования одних и тех же лиц на протяжении достаточно длительного отрезка их жизненного пути, измеряемого иногда десятками лет, представляет собой так называемое *лонгитюдинальное исследование*. Этот метод позволяет определить генетические связи между фазами психического развития и дать научно обоснованный прогноз дальнейшего психического развития.

Сопоставление показателей, полученных по одним и тем же методикам, но в разных условиях измерения (например, "покоя" - "стресса"), дает нам *ситуационный* сдвиг. Условия измерения могут изменяться не только реально, но и умозрительно. Например, мы можем попросить испытуемого "представить себе", что он оказался других условиях измерения: в будущем, в позиции других людей, которые оценивают его как бы со стороны, в состоянии разгневанного и т. п. Сопоставляя показатели, измеренные в обычных и воображаемых условиях, мы получаем *умозрительный* сдвиг.

Мы можем создать специальные экспериментальные условия предположительно влияющие на те или иные показатели, и сопоставить замеры, произведенные до и после экспериментального воздействия. Если сдвиги окажутся статистически достоверными, это позволит утверждать, что экспериментальные воздействия были существенными, эффективными.

Во всех этих случаях мы говорим о *сдвиге под влиянием* контролируемых или не контролируемых воздействий.

И здесь мы наталкиваемся на методическую трудность, которую оказывается возможным преодолеть только путем введения контрольной группы, которая не испытывала бы на себе воздействия данного экспериментального фактора. Если нет контрольной группы, то сдвиг в экспериментальной группе может объясняться действием самых разных причин: временем суток, в которое производились замеры, важным для испытуемых событием, которое произошло между 1-м и 2-м замерами и по мощности воздействия значительно перекрыло экспериментальный фактор и т. п. Если в экспериментальной группе сдвиги окажутся достоверными, а в контрольной группе - недостоверными, то это, действительно, может свидетельствовать об эффективности воздействий. При отсутствии контрольной группы мы констатируем, что сдвиг произошел, но не имеем права приписать его именно данным, изучаемым нами, факторам воздействия.

Бывают случаи, когда мы не располагаем контрольной группой, зато в нашем распоряжении есть 2 или более экспериментальных групп, различающихся по условиям и способам воздействия на них. Это могут быть, помимо экспериментальных, и разнообразные естественные условия жизни, обучения, работы, общения и даже питания, водоснабжения, географического расположения и т. д. Сопоставление различающихся по этим признакам, позволит нам уточнить специфическое действие экспериментальных или естественно действующих факторов, хотя при этом нам следует помнить, что воздействие неучтенных факторов может оказаться еще более мощным.

В выводах мы все-таки будут ограничены, если не проверили свои результаты на контрольной группе, в которой измерения производились параллельно.

Помимо рассмотренных сдвигов: временных, ситуационных, умозрительных и сдвигов под влиянием, - можно рассмотреть еще особую категорию *структурных* сдвигов.

Мы можем сопоставлять между собой разные показатели одних и тех же испытуемых, если они измерены в одних и тех же единицах, по одной и той же шкале. Например, мы можем исследовать перепад между вербальным и невербальным интеллектом, измеренными по методике Д. Векслера, или сопоставлять экспертные оценки эмпатичности и наблюдательности, измеренные по одинаковой 10-балльной шкале, или время решения двух задач, измеренное в секундах, или экзаменационную успешность по разным дисциплинам и т.п.

Перед нами - зависимые ряды значений, поскольку они измерены на одних и тех же испытуемых, поэтому будет более обоснованным использовать критерии оценки достоверности сдвигов для связанных выборок.

Исключение представляют случаи, когда мы сопоставляем величины сдвигов в двух независимых группах испытуемых, например экспериментальной и контрольной. Допустим, если мы установили, что положительный сдвиг в сторону улучшения взаимопонимания наблюдается и в экспериментальной, и в контрольной группах, мы можем попробовать доказать, что в экспериментальной группе этот сдвиг достоверно больше, чем в контрольной, и что, следовательно, экспериментальное воздействие все-таки существенно.

Важный вопрос касается того, должны ли мы всегда производить оба замера на одной и той же выборке, или "сдвиг" можно изучать на сходных, так называемых "уравновешенных" выборках, совпадающих друг с другом по полу, возрасту, профессии и другим значимым для исследователя характеристикам.

Допускается сопоставление показателей разных выборок, уравнишенных по всем значимым для исследования признакам. Иными словами, можно уровень тревоги или объем внимания до экзамена измерять у одной подгруппы, а после экзамена - у другой подгруппы, если они "уравнишены". Опыт показывает, однако, что создать "уравнишенные" подгруппы практически невозможно. Мы всегда упираемся в факт существования различий между выделенными подгруппами, которые могут в значительной степени повлиять на результат. В итоге окажется, что мы исследовали не влияние экзаменационного пресса на уровень тревоги или объем внимания, а различия по этому показателю между двумя выделенными подгруппами. К сожалению, в значительной степени это относится и к проблеме сопоставления экспериментальной и контрольной групп: мы почти никогда не можем быть уверены, что выявленные различия объясняются действием исследуемых факторов, а не различиями между двумя выборками.

Многие исследователи обходят эту проблему самым простым образом: они вообще не заботятся о контрольной группе. Сдвиг есть - значит, воздействие эффективно. И действительно, при отсутствии контрольной выборки тоже можно порассуждать на тему о том, какими причинами, кроме предполагаемой, могут объясняться полученные сдвиги.

Другой вариант "уравновешивания" - введение параллельных форм теста. В тех случаях, когда на результатах повторных замеров могут сказаться эффекты научения, приходится "до" измерять реакции испытуемого с помощью одного инструмента, а "после" - с помощью другого. В результате на измерениях может отразиться и действие фактора времени, и различия в параллельных формах теста, и непонятно что еще.

Создать параллельную форму методики не менее трудно, чем подобрать "уравновешенную" группу испытуемых. И все же, в тех случаях, когда нет другого выхода, приходится прибегать к этому способу.

Классификация сдвигов и критериев оценки их статистической достоверности

В тех случаях, когда мы хотим оценить различия в интенсивности сдвига в двух группах испытуемых (контрольной и экспериментальной или двух экспериментальных), мы можем использовать различные варианты сопоставлений:

- 1) производить сопоставления отдельно в двух группах, используя критерии $L_{\chi^2_r}$;
- 2) сопоставлять показатели сдвига в двух группах (*Сдвиг* - это разность между вторым и первым замерами. Сначала вычисляются разности отдельно для каждой из групп, а уж затем проводятся сопоставления двух рядов разностей (сдвигов), полученных в разных группах). Поскольку группы независимы, значения сдвигом также независимы, и мы можем применять по отношению к ним уже известные нам критерии Q Розенбаума, U Манна-Уитни и ϕ^* угловое преобразование Фишера.

***G*- критерий знаков**

Назначение критерия G

Критерий знаков G предназначен для установления общего *направления* сдвига исследуемого признака.

Он позволяет установить, в какую сторону в выборке в целом изменяются значения признака при переходе от первого измерения ко второму: изменяются ли показатели в сторону улучшения, повышения или усиления или, наоборот, в сторону ухудшения, понижения или ослабления.

С математической точки зрения критерий знаков является частным случаем биномиального критерия для двух равновероятных альтернатив. При вероятности каждой из альтернатив $P=Q=0,50$ критерий знаков является зеркальным отражением биномиального критерия.

В некоторых руководствах критерий знаков называют критерием Мак-Немара.

Описание критерия G

Критерий знаков применим и к тем сдвигам, которые можно определить *качественно* (например, изменение отрицательного отношения к чему-либо на положительное), так и к тем сдвигам, которые могут быть измерены *количественно* (например, сокращение времени работы над заданием после экспериментального воздействия).

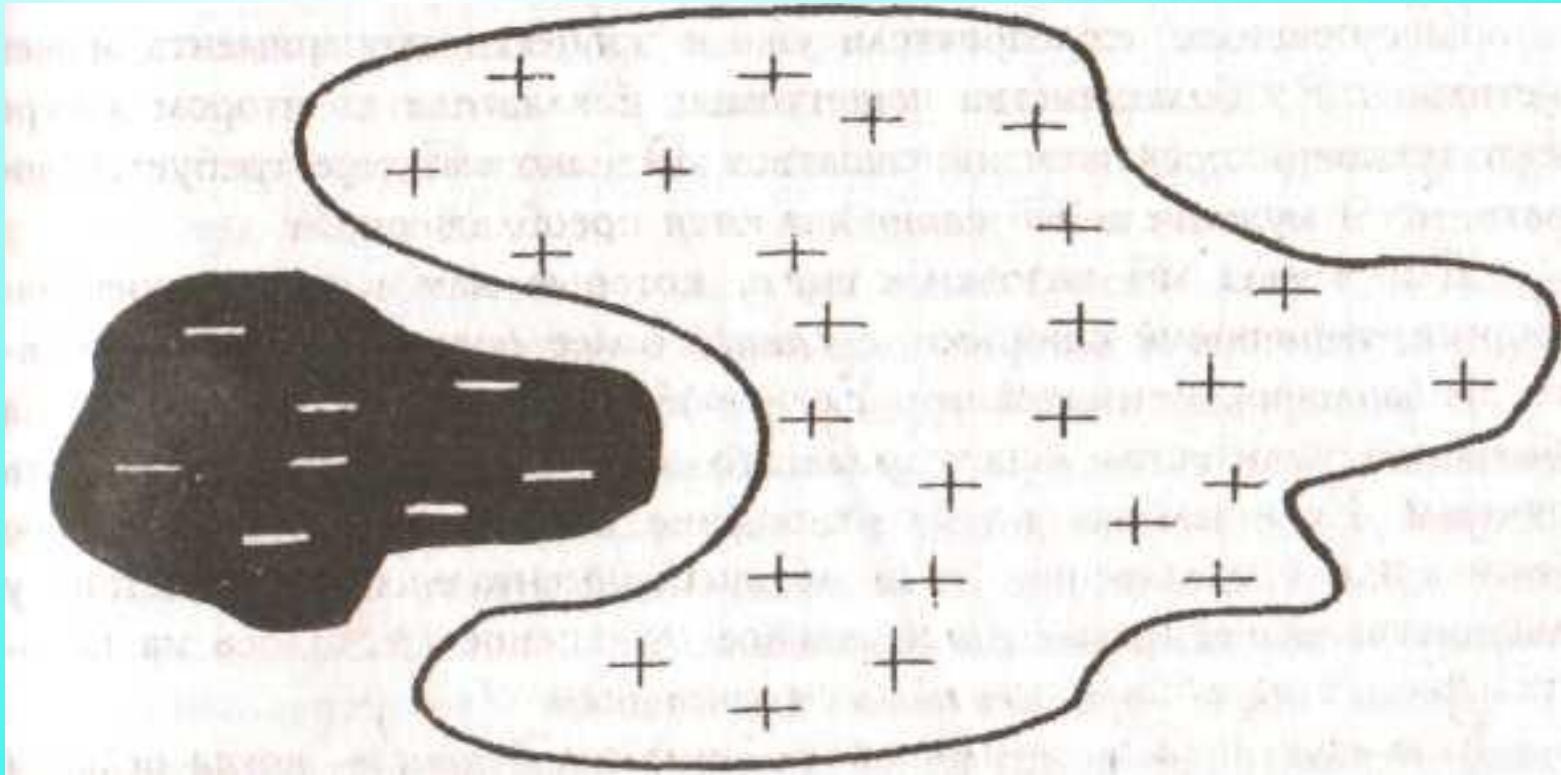
Во втором случае, однако, если сдвиги варьируют в достаточно широком диапазоне, лучше применять критерий T Вилкоксона. Он учитывает не только направление, но и интенсивность сдвигов и может оказаться более мощным в определении достоверности сдвигов, критерий знаков.

Назовем сдвиги, которые нам кажутся преобладающими, *типичными сдвигами*, а сдвиги более редкого, противоположного направления, *нетипичными*. Если значения показателя повышаются у большего количества испытуемых, то этот сдвиг мы будем считать типичным. Если мы исследуем отношение испытуемых к какому-либо событию или предложению, и после экспериментальных воздействий большинства испытуемых отрицательное отношение сменилось на положительное, то этот сдвиг мы назовем типичным.

Есть еще, правда, возможность "нулевых" сдвигов, когда реакция не изменяется или показатели не повышаются и не понижаются, а остаются на прежнем уровне. Однако такие "нулевые" сдвиги в критерии знаков исключаются из рассмотрения. При этом количество сопоставляемых пар уменьшается на число таких "нулевых" сдвигов.

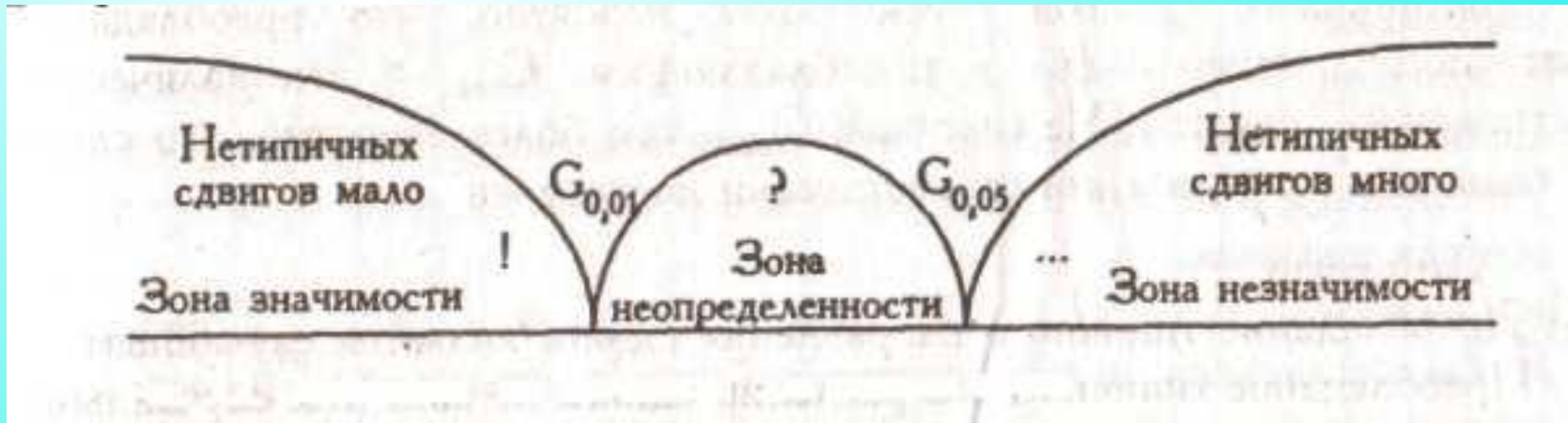
Суть критерия знаков состоит в том, что он определяет, слишком ли много наблюдается "нетипичных сдвигов", чтобы сдвиг в "типичном" направлении считать преобладающим? Ясно, что чем меньше "нетипичных сдвигов", тем более вероятно, что преобладание "типичного" сдвига является преобладающим. Гэмп - это количество "нетипичных" сдвигов. Чем меньше Гэмп, тем более вероятно, что сдвиг в "типичном" направлении статистически достоверен.

Графическое представление критерия знаков



Графическое представление положительных и отрицательных сдвигов в форме облаков: светлое облако - положительные сдвиги, темное облако - отрицательные

Зона значимости критерия G



Зона значимости простирается влево, в сторону более низких значений, поскольку чем меньше "нетипичных" знаков, тем достовернее "типичный" сдвиг. Зона незначимости, напротив, простирается вправо, в сторону более высоких значений G . Постепенно "нетипичных" сдвигов становится так много, что теряется само ощущение какого-то преобладания в направленности, сдвигов. Зона незначимости характеризует ситуацию, когда сдвиги обоих направлений перемешаны.

Ограничения критерия знаков

Количество наблюдений в обоих замерах - не менее 5 и не более 300.

Пример

В исследовании Г.А. Бадасовой (1994) изучались личностные факторы суггестора, способствующие его внушающему воздействию на аудиторию. В эксперименте участвовало 39 слушателей колледжа и спецфакультета практической психологии Санкт-Петербургского университета, 9 мужчин и 30 женщин в возрасте от 18 до 39 лет, средний возраст 23,5 года. Испытуемые выступали в качестве суггерендов, т.е. лиц, по отношению к которым оказывалось внушающее воздействие.

В экспериментальной группе ($n_1=16$) испытуемые просматривали видеозапись речи суггестора о целесообразности применения физических наказаний в воспитании детей, а в контрольной группе ($n_2=23$) испытуемые просто читали про себя письменный текст. Содержание суггестора и текста полностью совпадали/

До и после предъявления видеозаписи (в экспериментальной группе) и текста (в контрольной группе) испытуемые отвечали на вопросы, оценивая степень согласия с их содержанием по 7-балльной шкале:

Не согласен 1 2 3 4 5 6 7 Согласен

1. Я считаю возможным иногда шлепнуть своего ребенка за дело, если он этого заслужил.
2. Если, придя домой, я узнаю, что кто-то из близких, бабушка дедушка, шлепнул моего ребенка за дело, то я буду считать, что нормально.
3. Если мне станет известно, что воспитательница детского сада или учительница в школе шлепнула моего ребенка за дело, то я восприму это как должное.
4. Я бы согласился отдать своего ребенка в школу, где применяется система физических наказаний по итогам недели.

Оценки степени согласия с утверждениями о допустимости телесных наказаний до и после предъявления видеозаписи.

Вопросы

1. Можно ли утверждать, что после просмотра видеозаписи о пользе телесных наказаний наблюдается достоверный сдвиг в сторону большего принятия их в экспериментальной группе?
2. Достоверны ли различия по выраженности положительного сдвига между экспериментальной и контрольной группами?
3. Является ли достоверным сдвиг оценок в контрольной группе?

Расчет количества положительных, отрицательных и нулевых сдвигов в двух группах суггерендов

Кол-во сдвигов в группах	Шкалы				
	«Я сам»	«Бабушка»	«Воспитатель»	«Школа»	Суммы
1. Экспериментальная группа					
а) положительных	8	9	2	3	22
б) отрицательных	0	0	3	1	4
в) нулевых	8	7	11	12	38
Суммы	16	16	16	16	64
2. Контрольная группа					
а) положительных	4	4	4	4	16
б) отрицательных	4	4	2	0	10
в) нулевых	15	15	17	19	66
Суммы	23	23	23	23	92

Из таблицы видим, что наиболее типичными являются "нулевые" сдвиги, то есть отсутствие сдвига в оценках после предъявления видеозаписи или письменного текста. Все же в экспериментальной группе по шкале "Я сам наказываю" и "Бабушка наказывает" положительные сдвиги наблюдаются примерно в половине случаев.

Нам необходимо учитывать только положительные и отрицательные сдвиги, а нулевые отбрасывать. Количество сопоставляемых пар значений при этом уменьшается на количество этих нулевых сдвигов. Теперь для шкалы "Я сам" $n=8$; для шкалы "Бабушка" $n=9$; шкалы "Воспитатель" $n=5$ и шкалы "Школа" $n=4$. По отношению к последней шкале критерий знаков вообще неприменим, так как количество сопоставляемых пар значений меньше 5.

Сумма положительных и отрицательных сдвигов по 4 шкалам составляет: $n=8+9+5+4=26$.

Гипотезы

H_0 : Сдвиг в сторону более снисходительного отношения к телесным наказаниям после внушения является случайным.

H_1 : Сдвиг в сторону более снисходительного отношения к телесным наказаниям после внушения является неслучайным.

Критические значения критерия знаков G определяем по таблице критических значений. Это максимальные количества "нетипичных", менее часто встречающихся, знаков, при которых сдвиг в "типичную" сторону еще можно считать существенным

1) Шкала "Я сам наказываю"

$$n = 8$$

Типичный сдвиг – положительный

Отрицательных сдвигов нет.

$$G_{кр} = \begin{cases} 1 (p \leq 0,05) \\ 0 (p \leq 0,01) \end{cases}$$

$$G_{эмп.} = 0$$

$$G_{эмп.} \leq G_{кр}$$

Н₀ отклоняется. Принимается Н₁ ($p \leq 0,01$)

2) Шкала "Бабушка наказывает"

$$n = 9$$

Типичный сдвиг – положительный

Отрицательных сдвигов нет.

$$G_{кр} = \begin{cases} 1 (p \leq 0,05) \\ 0 (p \leq 0,01) \end{cases}$$

$$G_{эмп.} = 0$$

$$G_{эмп.} \leq G_{кр}$$

Но отклоняется. Принимается H_1 ($p \leq 0,01$)

3) Шкала "Воспитательница наказывает"

$$n = 5$$

Типичный сдвиг – положительный

Отрицательных сдвигов нет.

$$G_{кр} = \begin{cases} 1 (p \leq 0,05) \\ 0 (p \leq 0,01) \end{cases}$$

$$G_{эмп.} = 2$$

$$G_{эмп.} > G_{кр}$$

Но принимается

4) Шкала "Воспитательница наказывает"

$$n = 4$$

$n < 5$, критерий знаков неприменим

5) Сумма по 4-м шкалам

$$n = 26$$

Типичный сдвиг – положительный

Отрицательных сдвигов – 4.

$$G_{кр} = \begin{cases} 8 (p \leq 0,05) \\ 6 (p \leq 0,01) \end{cases}$$

$$G_{эмп.} = 4$$

$$G_{эмп} < G_{кр}$$

Н₀ отклоняется. Принимается Н₁ ($p \leq 0,01$)

Ответ

Сдвиг в сторону более снисходительного отношения телесным наказаниям в экспериментальной группе после просмотра видеозаписи является неслучайным для шкал "Я сам наказываю", "Бабушка наказывает" и по сумме четырех шкал (во всех случаях).

Гипотезы для контрольной группы

Н₀: Сдвиг в сторону более снисходительного отношения к телесным наказаниям после прочтения текста является случайным.

Н₁: Сдвиг в сторону более снисходительного отношения к телесным наказаниям после прочтения текста не является случайным.

Далее действуем по тому же принципу: вначале определяем количество сдвигов в ту или иную сторону (n), выявляем типичный сдвиг количество нетипичных сдвигов ($G_{эмп}$) сопоставляем с критическими значениям G , определенными по таблице критических значений.

1) Шкала "Я сам наказываю"

$$n = 8$$

Положительных сдвигов 4.

Отрицательных сдвигов 4.

Типичный сдвиг установить невозможно, т.к. положительных и отрицательных сдвигов поровну.

Но принимается.

2) Шкала "Бабушка наказывает"

$$n = 8$$

Положительных сдвигов 4.

Отрицательных сдвигов 4.

Но принимается.

3) Шкала "Воспитательница наказывает"

$$n = 6$$

Типичный сдвиг – положительный

Отрицательных сдвигов 2.

$$G_{кр} = 0 (p \leq 0,05)$$

$G_{кр} (p \leq 0,01)$ при данном n определить невозможно

$$G_{эмп.} = 2$$

$$G_{эмп.} > G_{кр.}$$

Но принимается

4) Шкала "Воспитательница наказывает"

$n < 5$, критерий знаков неприменим

5) Сумма по 4-м шкалам

$$n = 26$$

Типичный сдвиг – положительный

Отрицательных сдвигов – 10.

$$G_{кр} = \begin{cases} 8 (p \leq 0,05) \\ 6 (p \leq 0,01) \end{cases}$$

$$G_{эмп.} = 10$$

$$G_{эмп.} > G_{кр}$$

Но принимается

Ответ

Сдвиг в сторону более снисходительного отношения к телесным наказаниям в контрольной группе является случайным - и по каждой из шкал в отдельности, и по сумме шкал.

Мы можем определенно ответить на 1-ый вопрос задачи: да, можно утверждать, что после просмотра видеозаписи о пользе телесных наказаний наблюдается достоверный сдвиг в пользу большего принятия их в экспериментальной группе. Мы можем ответить и на 3-й вопрос задачи: нет, сдвиг оценок в контрольной группе недостоверен. Однако мы пока не ответили на 2-й вопрос - о том, достоверны ли различия выраженности положительного сдвига между экспериментальной и контрольной группами?

Дело в том, что нами был избран вариант сопоставлений, предполагающий сравнение значений "после" и "до" экспериментального воздействия отдельно в экспериментальной и контрольной выборках. Для того, чтобы ответить на вопрос 2, необходимо выбрать второй вариант сопоставлений, предусматривающий сравнение сдвигов в группах с помощью критериев для сравнения независимых выборок Q - критерия Розенбаума, U - критерия Манна-Уитни и критерия Фишера.

Однако такого рода сопоставления, как правило, проводятся только в том случае, если и в экспериментальной, и в контрольной группах выявлен достоверный однонаправленный эффект, и нужно доказать, что в экспериментальной выборке он достоверно больше. В данном же случае нами доказано, что в контрольной выборке не произошло сколько-нибудь значимых изменений, и мы можем этим удовлетвориться.

Мы установили лишь то, что в тех случаях, когда наблюдался какой-то сдвиг в оценках, он был скорее положительным, чем отрицательным в экспериментальной группе и скорее случайным в контрольной группе. Все нулевые сдвиги мы отбросили, а ведь они составляют от 43,8 до 50% по тем шкалам, где обнаружен положительный достоверный сдвиг в экспериментальной выборке. Похоже, что многие, очень многие испытуемые экспериментальной выборки просто проигнорировали выступление суггестора... Однако статистический критерий свидетельствует: положительный сдвиг в оценках достоверен, по крайней мере для первых двух шкал и для тех испытуемых, которые хоть как-то прореагировали на выступление суггестора.

Алгоритм расчета критерия знаков G

1. Подсчитать количество нулевых реакций и исключить их из рассмотрения. В результате n уменьшится на количество нулевых реакций.
2. Определить преобладающее направление изменений. Считать сдвиги в преобладающем направлении "типичными".
3. Определить количество "нетипичных" сдвигов. Считать это число эмпирическим значением G .
4. По таблице критических значений определить критические значения G для данного n .
5. Сопоставить $G_{\text{эмп}}$ с $G_{\text{кр}}$. Если $G_{\text{эмп}}$ меньше $G_{\text{кр}}$ или по крайней мере равен ему, сдвиг в типичную сторону может считаться достоверным.

***T* - критерий Вилкоксона**

Назначение критерия

Критерий применяется для сопоставления показателей, измеренных *в двух разных условиях на одной и той же выборке* испытуемых.

Он позволяет установить не только *направленность* изменений, но и их *выраженность*. С его помощью мы определяем, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Описание критерия T

- Этот критерий применим в тех случаях, когда признаки измерены по крайней мере по шкале порядка, и сдвиги между вторым и первым замерами тоже могут быть упорядочены. Для этого они должны варьировать в достаточно широком диапазоне. В принципе, можно применять критерий T и в тех случаях, когда сдвиги принимают только три значения: — 1, 0 и +1, но тогда критерий T вряд ли добавит что-нибудь новое к тем выводам, которые можно было бы получить с помощью критерия знаков. Вот если сдвиги изменяются, скажем, от — 30 до +45, тогда имеет смысл их ранжировать и потом суммировать ранги.

Суть метода состоит в сопоставлении выраженности сдвигов в том и ином направлениях по абсолютной величине. Для этого сначала ранжируют все абсолютные величины сдвигов, а потом суммируют ранги. Если сдвиги в положительную и в отрицательную сторону происходят случайно, то суммы рангов абсолютных значений будут примерно равны. Если же интенсивность сдвига в одном из направлений перевешивает, то сумма рангов абсолютных значений сдвигов в противоположную сторону будет значительно ниже, чем это могло быть при случайных изменениях.

Первоначально мы исходим из предположения о том, что типичным сдвигом будет сдвиг в более часто встречающемся направлении, нетипичным, или редким, сдвигом - сдвиг в более редко встречающемся направлении.

Гипотезы

H_0 : Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

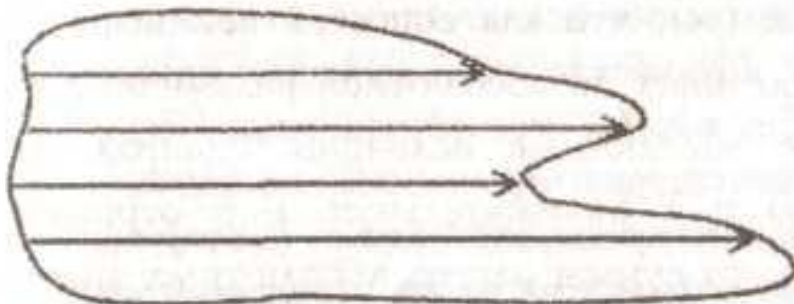
H_1 : Интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении.

Графическое представление критерия T

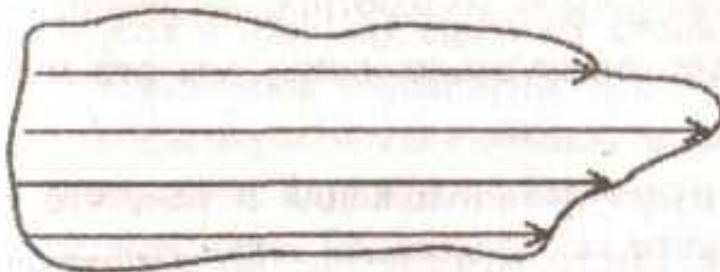
Сдвиги в противоположные стороны мы можем представить себе в виде двух облаков. Величина облака зависит не только от количества соответствующих сдвигов, но и от их интенсивности, отраженной в длине стрелок. В сущности, облака противостоят друг другу, как два воздушных фронта: они не просто соревнуются по величине, они меряются силами!

При определенных n , а именно при $n \geq 10$ вообще можем отказаться от понятия типичного сдвига. Сдвигов в ту и другую сторону может оказаться поровну, но если 9 меньших сдвигов будут относиться к одному направлению, а 9 больших сдвигов - к противоположному, то можно констатировать достоверное преобладание этого противоположного направления сдвигов.

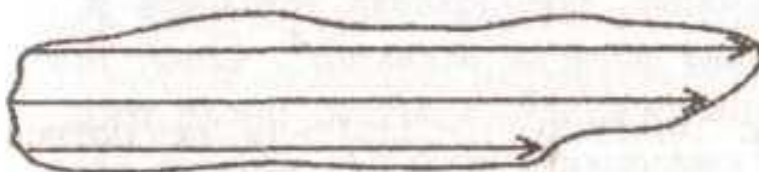
Критерий знаков в этом случае не выявил бы никаких достоверных различий.



а) "светлый фронт" преобладает над "темным фронтом" и по количеству сдвигов, и по их интенсивности



б) "светлый фронт" преобладает над "темным" только по интенсивности сдвигов, но по количеству сдвигов они равны



в) "светлый фронт" уступает "темному" по количеству сдвигов, но самые интенсивные сдвиги принадлежат "светлому фронту"

Рис. 3.3. Варианты соотношения "светлого" и "темного фронтов" - сдвигов двух разных направленностей

Ограничения в применении критерия Т Вилкоксона

1. Минимальное количество испытуемых, прошедших измерения в двух условиях - 5 человек. Максимальное количество испытуемых - 50 человек, что диктуется верхней границей имеющихся таблиц.
2. Нулевые сдвиги из рассмотрения исключаются, и количество наблюдений n уменьшается на количество этих нулевых сдвигов. Можно обойти это ограничение, сформулировав гипотезы, включающие отсутствие изменений, например: "Сдвиги сторону увеличения значений превышает сдвиг в сторону уменьшения значений и тенденцию сохранения их на прежнем уровне".

Пример

В выборке курсантов военного училища (юноши в возрасте от 18 до 20 лет) измерялась способность к удержанию физического волевого усилия на динамометре. Сначала у испытуемых измерялась максимальная мышечная сила каждой из рук, а на следующий день им предлагалось выдерживать, на динамометре с подвижной стрелкой мышечное усилие, равное $1/2$ максимальной мышечной силы данной руки. Почувствовав усталость, испытуемый должен был сообщить об этом экспериментатору, но не прекращать опыт, преодолевая усталость и неприятные ощущения - "бороться, пока воля не иссякнет".

Опыт проводился дважды; вначале с обычной инструкцией, а затем, после того, как испытуемый заполнял опросник самооценки волевых качеств по методике А.Ц. Пуни, ему предлагалось представить себе, что он уже добился идеала в развитии волевых качеств, и продемонстрировать соответствующее идеалу волевое усилие. Подтвердилась ли гипотеза экспериментатора о том, что обращение к идеалу способствуя возрастанию волевого усилия?

Расчет критерия Т при сопоставлении замеров
физического волевого усилия (табл. 5)

Первый шаг в подсчете критерия Т - вычитание каждого индивидуального значения "до" из значения "после".

Из табл. 5 видно, что 8 полученных разностей - отрицательные и лишь 3 - положительные. Это означает, что у 8 испытуемых длительность удержания мышечного усилия во втором замере уменьшилась, а у 3 - увеличилась. Мы столкнулись с тем случаем, когда уже сейчас мы не можем сформулировать статистическую гипотезу, соответствующую первоначальному предположению исследователя. Предполагалось, что обращение к идеалу будет увеличивать длительность мышечного усилия, а экспериментальные данные свидетельствуют, что лишь в 3 случаях из 11 этот показатель действительно увеличился. Мы можем сформулировать лишь гипотезу, предполагающую несущественность сдвига этого показателя в сторону снижения.

Гипотезы

H_0 : Интенсивность сдвигов в сторону уменьшения длительности мышечного усилия не превышает интенсивности сдвигов в сторону ее увеличения.

H_1 : Интенсивность сдвигов в сторону уменьшения длительности мышечного усилия превышает интенсивность сдвигов в сторону ее увеличения

На следующем шаге все сдвиги, независимо от их знака, должны быть проранжированы по выраженности. В табл. 5 в четвертом слева столбце приведены абсолютные величины сдвигов, а в последнем столбце (справа) - ранги этих абсолютных величин. Меньшему значению соответствует меньший ранг. При этом сумма рангов равна 66, что соответствует расчетной:

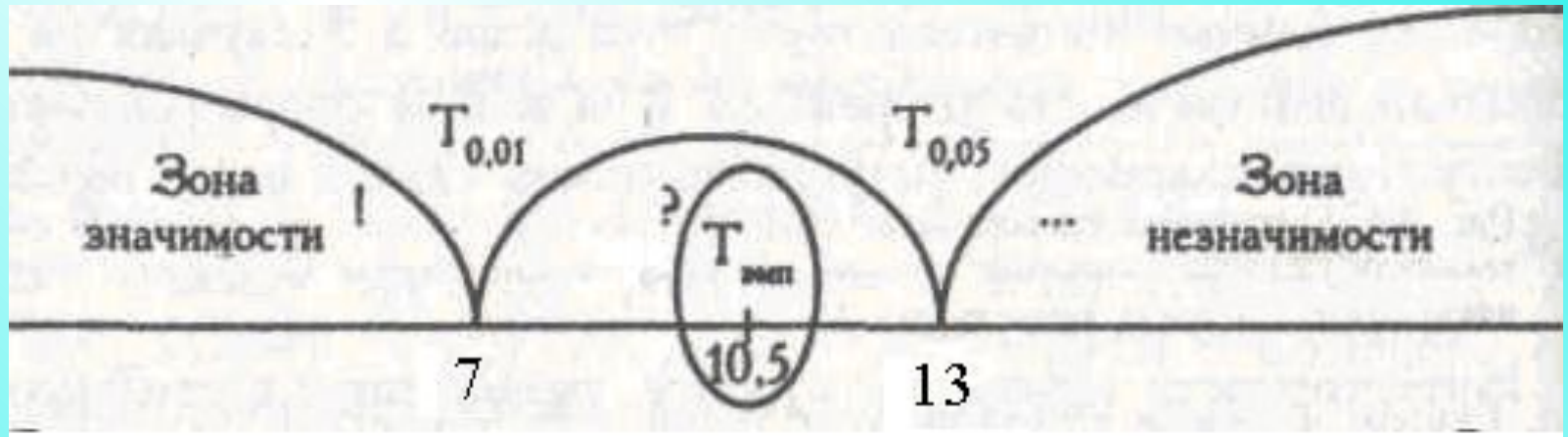
$$\sum R_i = \frac{N(N+1)}{2} = \frac{11 \cdot (11+1)}{2} = 66$$

Отметим те сдвиги, которые являются нетипичными, в данном случае - положительными. Сумма рангов этих "редких" сдвигов и составляет эмпирическое значение критерия $T = \sum R_r$,
где R_r - ранговые значения сдвигов с более редким знаком.
Итак, в данном случае, $T_{\text{эмп}}=1+2.5+7=10,5$.

По таблице критических значений найдем $T_{\text{кр}}$.

$$T_{\text{кр}} = \begin{cases} 13 & (p \leq 0,05) \\ 7 & (p \leq 0,01) \end{cases}$$

Ось значимости



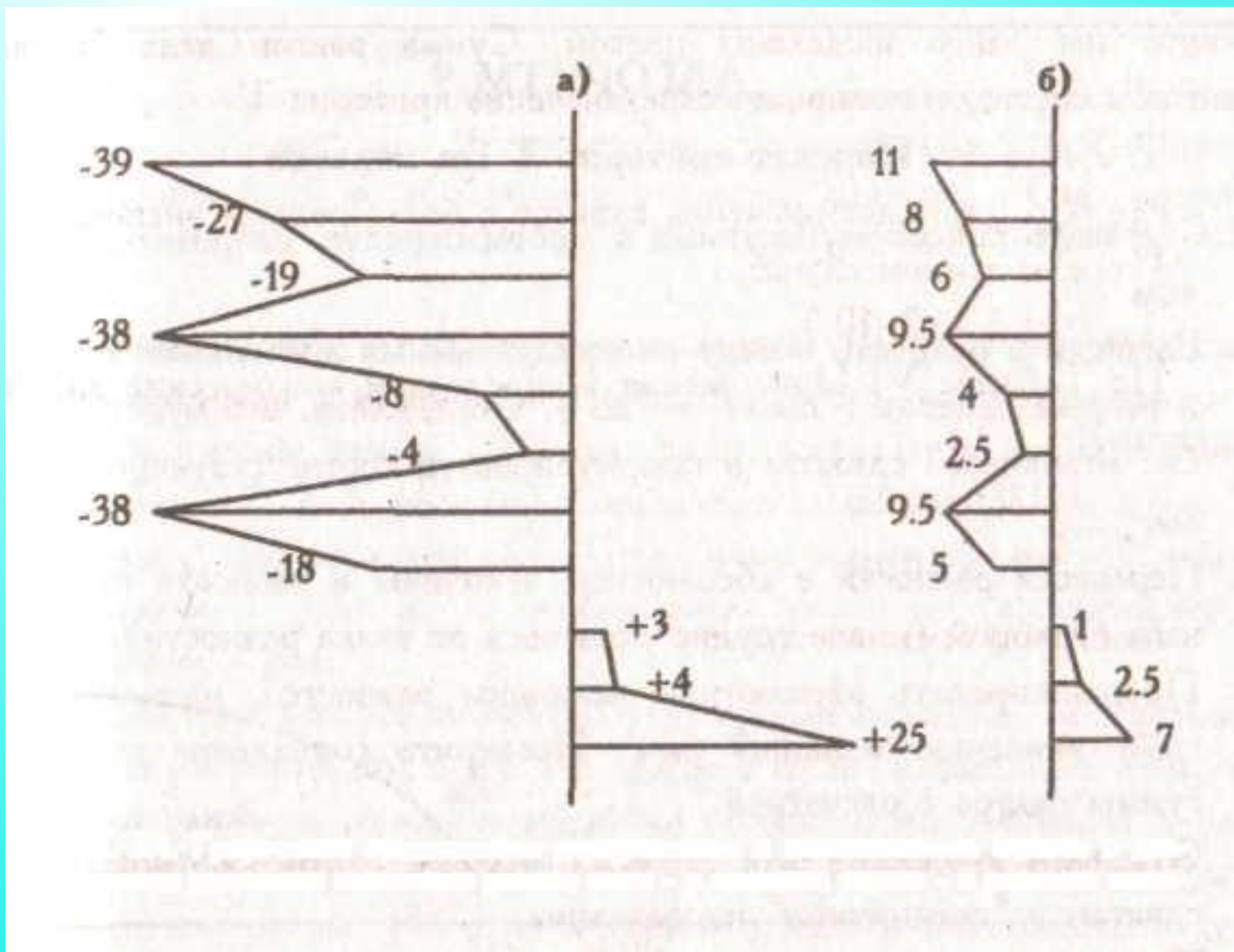
Зона значимости в данном случае простирается влево. Действительно, если бы "редких", в данном случае положительных, сдвигов было совсем, то и сумма их рангов равнялась бы нулю. В данном случае эмпирическое значение T попадает в зону неопределенности:

$$T_{\text{эмп}} < T_{\text{кр}} (0,05)$$

Ответ

Н₀ отвергается. Интенсивность отрицательного сдвига показателя физического волевого усилия превышает интенсивность положительного сдвига ($p < 0,05$).

Графическое представление отрицательных и положительных сдвигов в длительности удержания мышечного усилия;
слева - в секундах; справа - в ранговых значениях



Таким образом, исследователю придется признать, что продолжительность удержания мышечного волевого усилия во втором замере снижается, и этот сдвиг неслучаен. Инструкция, ориентирующая испытуемого на соответствие идеалу в развитии воли, оказалась гораздо менее мощным фактором, чем какая-то иная сила - возможно, мышечное утомление, может быть, разочарование в себе или в возможностях данной психологического эксперимента. А может быть, в момент второго замера просто перестает действовать какой-то мощный фактор, который был активен вначале? На все эти вопросы статистические методы не могут ответить, если в схему эксперимента не включена контрольная группа - в данном случае, выборка, уравновешенная с экспериментальной группой по всем значимым характеристикам (полу, возрасту, профессии, месту обучения), у которой просто измерили бы вторично волевое усилие через такой же промежуток времени, не призывая соответствовать идеалу в развитии воли.

Алгоритм подсчета критерия Т Вилкоксона

1. Составить список испытуемых в любом порядке, например, алфавитном.
2. Вычислить разность между индивидуальными значениями во втором и первом замерах ("после" - "до"). Определить, что будет считаться "типичным" сдвигом и сформулировать соответствующие гипотезы.
3. Перевести разности в абсолютные величины и записать их отдельным столбцом (иначе трудно отвлечься от знака разности).
4. Проранжировать абсолютные величины разностей, начисляя меньшему значению меньший ранг. Проверить совпадение полученных суммы рангов с расчетной.
5. Отметить кружками или другими знаками ранги, соответствующими сдвигам в "нетипичном" направлении.
6. Подсчитать сумму этих рангов по формуле:
$$T_{\text{эмп.}} = \sum R_r$$
где R_r - ранговые значения сдвигов с более редким знаком.
7. Определить критические значения T для данного n по таблице критических значений. Если $T_{\text{эмп.}}$ меньше или равен $T_{\text{кр.}}$, сдвиг в "типичную" сторону по интенсивности достоверно преобладает.

Спасибо за внимание!