**ОСНОВНЫЕ ПОНЯТИЯ, ИСПОЛЬЗУЕМЫЕ В МАТЕМАТИЧЕСКОЙ ОБРАБОТКЕ ДАННЫХ**

**1.1. Признаки и переменные**

Признаки и переменные ˗ это измеряемые явления. Такими явлениями могут быть время решения задачи, количество допущенных ошибок, уровень тревожности, показатель интеллектуальной лабильности, интенсивность агрессивных реакций, угол поворота корпуса в беседе, показатель социометрического статуса и множество других переменных.

Понятия признака и переменной могут использоваться как взаи­мозаменяемые. Они являются наиболее общими. Иногда вместо них используются понятия показателя или уровня, например, уровень на­стойчивости, показатель вербального интеллекта и др. Понятия показателя и уровня указывают на то, что признак может быть измерен количественно, так как к ним применимы определения "высокий" или "низкий", например, высокий уровень интеллекта, низкие показатели тревожности и др.

Переменные являются случайными величинами, поскольку заранее неизвестно, какое именно значение они примут. Математическая обработка ˗ это оперирование со значениями признака, полученными в исследовании. Такие результаты называют также "наблюдениями", "наблюдаемыми значениями", "вариантами", "датами", и др. Значения признака определяются при помощи специальных шкал измерения.

**1.2. Шкалы измерения**

Измерение ˗ это приписывание числовых форм объектам или событиям в соответствии с определенными правилами. С.Стивенсом предложена классификация из 4 типов шкал измерения:

1) номинативная, или номинальная, или шкала наименований;

2) порядковая, или ординальная, шкала;

3) интервальная, или шкала равных интервалов;

4) шкала равных отношений.

**Номинативная шкала** ˗ это шкала, классифицирующая по названию: *потеп* (лат.) - имя, название. Название же не измеряется количественно, оно лишь позволяет отличить один объект от другого или одного субъекта от другого. Номинативная шкала ˗ это способ классификации объектов или субъектов, распределения их по ячейкам классификации. Простейший случай номинативной шкалы ˗ дихотомическая шкала, состоящая всего лишь из двух ячеек, например: "имеет братьев и сестер - единственный ребенок в семье"; "иностранец ˗ соотечественник"; "проголосовал "за" ˗ проголосовал "против"" и т.п.

Признак, который измеряется по дихотомической шкале наименований, называется альтернативным. Он может принимать всего два значения. При этом исследователь зачастую заинтересован в одном из них, и тогда он говорит, что признак “проявился”, если тот принял интересующее его значение, и что признак “не проявился”, если он принял противоположное значение. Например: "Признак леворукости проявился у 8 испытуемых из 20". В принципе номинативная шкала может состоять из ячеек "признак проявился ˗ признак не проявился .

Более сложный вариант номинативной шкалы ˗ классификация из трех и более ячеек, например: "экстрапунитивные ˗ интрапунитивные ˗импунитивные реакции" или "выбор кандидатуры А ˗ кандидатуры Б -˗ кандидатуры В ˗ кандидатуры Г" или "старший ˗ средний ˗ младший ˗ единственный ребенок в семье" и др.

Расклассифицировав все объекты, реакции или всех испытуемых по ячейкам классификации, мы получаем возможность от наименований перейти к числам, подсчитав количество наблюдений в каждой из ячеек.

Как уже указывалось, наблюдение ˗ это одна зарегистрированная реакция, один совершенный выбор, одно осуществленное действие или результат одного испытуемого.

Допустим, мы определим, что кандидатуру А выбрали 7 испытуемых, кандидатуру Б ˗ 11, кандидатуру В ˗ 28, а кандидатуру Г ˗ всего 1. Теперь мы можем оперировать этими числами, представляющими собой частоты встречаемости разных наименований, то есть частоты принятия признаком "выбор" каждого из 4 возможных значений. Далее мы можем сопоставить полученное распределение частот с равномерным или каким-то иным распределением. Таким образом, номинативная шкала позволяет нам подсчитывать частоты встречаемости разных "наименований", или значений признака, и затем работать с этими частотами с помощью математических методов.

Единица измерения, которой мы при этом оперируем ˗ количество наблюдений (испытуемых, реакций, выборов и т. п.), или частота.

**Порядковая шкала** ˗ это шкала, классифицирующая по принци­пу "больше - меньше". Если в шкале наименований было безразлично, в каком порядке мы расположим классификационные ячейки, то в порядковой шкале они образуют последовательность от ячейки "самое малое значение" к ячейке "самое большое значение" (или наоборот). Ячейки теперь уместнее называть классами, поскольку по отношению к классам употребимы определения "низкий", "средний" и "высокий" класс, или 1-й, 2-й, 3-й класс, и т.д.

В порядковой шкале должно быть не менее трех классов, например "положительная реакция ˗ нейтральная реакция ˗ отрицательная реакция" или "подходит для занятия вакантной должности ˗ подходит с оговорками ˗ не подходит" и т. п.

В порядковой шкале мы не знаем истинного расстояния между классами, а знаем лишь, что они образуют последовательность. Например, классы "подходит для занятия вакантной должности" и "подходит с оговорками" могут быть реально ближе друг к другу, чем класс "подходит с оговорками" к классу "не подходит".

От классов легко перейти к числам, если мы условимся считать, что низший класс получает ранг 1, средний класс ˗ ранг 2, а высший класс ˗ ранг 3, или наоборот. Чем больше классов в шкале, тем больше у нас возможностей для математической обработки полученных данных и проверки статистических гипотез.

Например, мы можем оценить различия между двумя выборками испытуемых по преобладанию у них более высоких или более низких рангов или подсчитать коэффициент ранговой корреляции между двумя переменными, измеренными в порядковой шкале, допустим, между оценками профессиональной компетентности руководителя, данными ему разными экспертами.

Все методы, использующие ранжирование, построены на применении шкалы порядка. Если испытуемому предлагается упорядочить 18 ценностей по степени их значимости для него, проранжировать список личностных качеств социального работника или 10 претендентов на эту должность по степени их профессиональной пригодности, то во всех этих случаях испытуемый совершает так называемое принудительное ранжирование, при котором количество рангов соответствует количеству ранжируемых субъектов или объектов (ценностей, качеств и т.п.).

Независимо от того, приписываем ли мы каждому качеству или испытуемому один из 3˗4 рангов или совершаем процедуру принудительного ранжирования, мы получаем в обоих случаях ряды значений, измеренные по порядковой шкале. Правда, если у нас всего 3 возможных класса и, следовательно, 3 ранга, и при этом, скажем, 20 ранжируемых испытуемых, то некоторые из них неизбежно получат одинаковые ранги. Все многообразие жизни не может уместиться в 3 градации, поэтому в один и тот же класс могут попасть люди, достаточно серьезно различающиеся между собой. С другой стороны, принудительное ранжирование, то есть образование последовательности из многих испытуемых, может искусственно преувеличивать различия между людьми. Кроме того, данные, полученные в разных группах, могут оказаться несопоставимыми, так как группы могут изначально различаться по уровню развития исследуемого качества, и испытуемый, получивший в одной группе высший ранг, в другой получил бы всего лишь средний, и т.п.

Выход из положения может быть найден, если задавать достаточно дробную классификационную систему, скажем, из 10 классов, или градаций, признака. В сущности, подавляющее большинство психологических методик, использующих экспертную оценку, построено на измерении одним и тем же "аршином" из 10, 20 или даже 100 градаций разных испытуемых в разных выборках.

Итак, единица измерения в шкале порядка ˗ расстояние в 1 класс или в 1 ранг, при этом расстояние между классами и рангами может быть разным (оно нам неизвестно). К данным, полученным по порядковой шкале, применимы все описанные в данной книге критерии и методы.

**Интервальная шкала** ˗ это шкала, классифицирующая по принципу "больше на определенное количество единиц ˗ меньше на определенное количество единиц". Каждое из возможных значений признака отстоит от другого на равном расстоянии.

Можно предположить, что если мы измеряем время решения задачи в секундах, то это уже явно шкала интервалов. Однако на самом деле это не так, поскольку психологически различие в 20 секунд между испытуемым А и Б может быть отнюдь не равно различию в 20 секунд между испытуемыми Б и Г, если испытуемый А решил задачу за 2 секунды, Б ˗ за 22, В ˗ за 222, а Г ˗ за 242.

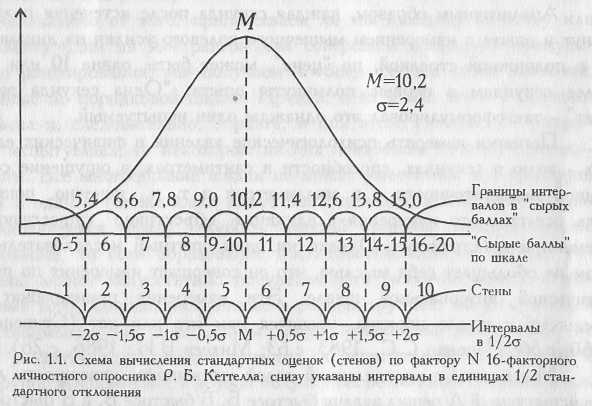
Аналогичным образом, каждая секунда после истечения полутора минут в опыте с измерением мышечного волевого усилия на динамометре с подвижной стрелкой, по "цене", может быть, равна 10 или даже более секундам в первые полминуты опыта. "Одна секунда за год идет" ˗ так сформулировал это однажды один испытуемый.

Попытки измерять психологические явления в физических единицах ˗ волю в секундах, способности в сантиметрах, а ощущение собственной недостаточности ˗ в миллиметрах и т. п., конечно, понятны, ведь все-таки это измерения в единицах "объективно" существующего времени и пространства. Однако ни один опытный исследователь при этом не обольщает себя мыслью, что он совершает измерения по психологической интервальной шкале. Эти измерения принадлежат по-прежнему к шкале порядка. Мы можем с определенной долей уверенности утверждать лишь, что испытуемый А решил задачу быстрее Б, Б быстрее В, а В быстрее Г.

Аналогичным образом, значения, полученные испытуемыми в баллах по любой нестандартизованной методике, оказываются измеренными лишь по шкале порядка. На самом деле равноинтервальными можно считать лишь шкалы в единицах стандартного отклонения и процентильные шкалы, и то лишь при условии, что распределение значений в стандартизующей выборке было нормальным.

Принцип построения большинства интервальных шкал построен на известном правиле "трех сигм": примерно 97,7-97,8% всех значений признака при нормальном его распределении укладываются в диапазоне М±3σ[[1]](#footnote-2) Можно построить шкалу в единицах долей стандартного отклонения, которая будет охватывать весь возможный диапазон изменения признака, если крайний слева и крайний справа интервалы оставить открытыми.

Р.Б. Кеттелл предложил, например, шкалу стенов ˗ "стандартной десятки". Среднее арифметическое значение в "сырых" баллах принимается за точку отсчета. Вправо и влево отмеряются интервалы, равные 1/2 стандартного отклонения. На Рис. 1.2 представлена схема вычисления стандартных оценок и перевода "сырых" баллов в стены по шкале N 16-факторного личностного опросника Р. Б. Кеттелла.



Справа от среднего значения будут располагаться интервалы, равные 6, 7, 8, 9 и 10 стенам, причем последний из этих интервалов открыт. Слева от среднего значения будут располагаться интервалы, равные 5, 4, 3, 2 и 1 стенам, и крайний интервал также открыт. Теперь мы поднимаемся вверх, к оси "сырых баллов", и размечаем границы интервалов в единицах "сырых" баллов. Поскольку М=10,2; σ=2,4, вправо мы откладываем 1/2σ, т.е. 1,2 "сырых" балла. Таким образом, гра­ница интервала составит: (10,2 + 1,2) = 11,4 "сырых" балла. Итак, границы интервала, соответствующего 6 стенам, будут простираться от 10,2 до 11,4 баллов. В сущности, в него попадает только одно "сырое" значение ˗ 11 баллов. Влево от средней мы откладываем 1/2 σ и получаем границу интервала: 10,2-1,2=9. Таким образом, границы интервала, соответствующие 9 стенам, простираются от 9 до 10,2. В этот интервал попадают уже два "сырых" значения ˗ 9 и 10. Если испытуемый получил 9 "сырых" баллов, ему начисляется теперь 5 стенов; если он получил 11 "сырых" баллов ˗ 6 стенов, и т. д.

Мы видим, что в шкале стенов иногда за разное количество "сырых" баллов будет начисляться одинаковое количество стенов. Например, за 16, 17, 18, 19 и 20 баллов будет начисляться 10 стенов, а за 14 и 15 - 9 стенов и т. д.

В принципе, шкалу стенов можно построить по любым данным, измеренным по крайней мере в порядковой шкале, при объеме выборки п>200 и нормальном распределении признака[[2]](#footnote-3).

Другой способ построения равноинтервальной шкалы ˗ группировка интервалов по принципу равенства накопленных частот. При нормальном распределении признака в окрестности среднего значения группируется большая часть всех наблюдений, поэтому в этой области среднего значения интервалы оказываются меньше, уже, а по мере удаления от центра распределения они увеличиваются, (см. Рис. 1.2). Следовательно, такая процентильная шкала является равноинтервальной только относительно накопленной частоты.



*Построение шкал равных интервалов по данным, полученным по шкале порядка, напоминает трюк с веревочной лестницей, на который ссылался С. Стивене. Мы сначала поднимаемся по лестнице, которая ни на чем не закреплена, и добираемся до лестницы, которая закреплена. Однако каким путем мы оказались на ней? Измерили некую психологическую переменную по шкале порядка, подсчитали средние и стандартные отклонения, а затем получили, наконец, интервальную шкалу. "Такому нелегальному использованию статистики может быть дано известное прагматическое оправдание; во многих случаях оно приводит к плодотворным результатам".*

Многие исследователи не проверяют степень совпадения полученного ими эмпирического распределения с нормальным распределением, и тем более не переводят получаемые значения в единицы долей стандартного отклонения или процентили, предпочитая пользоваться "сырыми" данными. "Сырые" же данные часто дают скошенное, срезанное по краям или двухвершинное распределение. На Рис. 1.3 представлено распределение показателя мышечного волевого усилия на выборке из 102 испытуемых. Распределение с удовлетворительной точностью можно считать нормальным (χ2=12,7, при v=9, M=89,75, σ= 25,1).



На Рис. 1.4 представлено распределение показателя самооценки по шкале методики Дж. Менестера - Р.Корзини "Уровень успеха, которого я должен был достичь уже сейчас" (n=356). Распределение значимо отличается от нормального (χ2=58,8, при v=7; *p<*0,01; М=80,64; σ =16,86).



С такими "ненормальными" распределениями приходится встречаться очень часто, чаще, может быть, чем с классическими нормальными. И дело здесь не в каком-то изъяне, а в самой специфике психологических признаков. По некоторым методикам от 10 до 20% испытуемых получают оценку "ноль" - например, в их рассказах не встречается ни одной словесной формулировки, которая отражала бы мотив "надежда на успех" или "боязнь неудачи" (методика Хекхаузена). То, что испытуемый получил оценку "ноль", нормально, но распределение таких оценок не может быть нормальным, как бы мы ни увеличивали объем выборки.

**Шкала равных отношений** ˗ это шкала, классифицирующая объекты или субъектов пропорционально степени выраженности измеряемого свойства. В шкалах отношений классы обозначаются числами, которые пропорциональны друг другу: 2 так относится к 4, как 4 к 8. Это предполагает наличие абсолютной нулевой точки отсчета. В физике абсолютная нулевая точка отсчета встречается при измерении длин отрезков или физических объектов и при измерении температуры по шкале Кельвина с абсолютным нулем температур. Считается, что в психологии примерами шкал равных отношений являются шкалы порогов абсолютной чувствительности (Стивене С, 1960; Гайда В. К., Захаров В. П., 1982). Возможности человеческой психики столь велики, что трудно представить себе абсолютный нуль в какой-либо измеряемой психологической переменной. Абсолютная глупость и абсолютная честность ˗ понятия скорее житейской психологии.

То же относится и к установлению равных отношений: только метафора обыденной речи допускает, чтобы Иванов был в 2 раза (3, 100, 1000) умнее Петрова или наоборот.

Абсолютный нуль, правда, может иметь место при подсчете количества объектов или субъектов. Например, при выборе одной из 3 альтернатив испытуемые не выбрали альтернативу А ни одного раза, альтернативу Б ˗ 14 раз и альтернативу В ˗ 28 раз. В этом случае мы можем утверждать, что альтернативу В выбирают в два раза чаще, чем альтернативу Б. Однако при этом измерено не психологическое свойство человека, а соотношение выборов у 42 человек.

По отношению к показателям частот возможно применять все арифметические операции: сложение, вычитание, деление и умножение. Единица измерения в этой шкале отношений ˗ 1 наблюдение, 1 выбор, 1 реакция и т. п. Мы вернулись к тому, с чего начали: к универсальной шкале измерения в частотах встречаемости того или иного значения признака и к единице измерения, которая представляет собой 1 наблюдение. Расклассифицировав испытуемых по ячейкам номинативной шкалы, мы можем применить потом высшую шкалу измерения ˗ шкалу отношений между частотами.

**1.3. Распределение признака. Параметры распределения**

Распределением признака называется закономерность встречаемости разных его значений. В психологических исследованиях чаще всего ссылаются на нормальное распределение. Нормальное распределение характеризуется тем, что крайние значения признака в нем встречаются достаточно редко, а значения, близкие к средней величине ˗ достаточно часто. Нормальным такое распределение называется потому, что оно очень часто встречалось в естественнонаучных исследованиях и казалось "нормой" всякого массового случайного проявления признаков. Это распределение следует закону, открытому тремя учеными в разное время: Муавром в 1733 г. в Англии, Гауссом в 1809 г. в Германии и Лапласом в 1812 г. во Франции. График нормального распределения представляет собой привычную глазу исследователя так называемую колоколообразную кривую.

Параметры распределения ˗ это его числовые характеристики, указывающие, где "в среднем" располагаются значения признака, на­сколько эти значения изменчивы и наблюдается ли преимущественное появление определенных значений признака. Наиболее практически важными параметрами являются математическое ожидание, дисперсия, показатели асимметрии и эксцесса.

В реальных исследованиях мы оперируем не параметрами, а их приближенными значениями, так называемыми оценками параметров. Это объясняется ограниченностью обследованных выборок. Чем больше выборка, тем ближе может быть оценка параметра к его истинному значению.

Среднее арифметическое (оценка математического ожидания) вы­числяется по формуле: 

где xi ˗ каждое наблюдаемое значение признака;

*i* ˗ индекс, указывающий на порядковый номер данного значения признака;

*п* ˗ количество наблюдений;

нак суммирования.

Оценка дисперсии определяется по формуле: 

где xi ˗ каждое наблюдаемое значение признака;

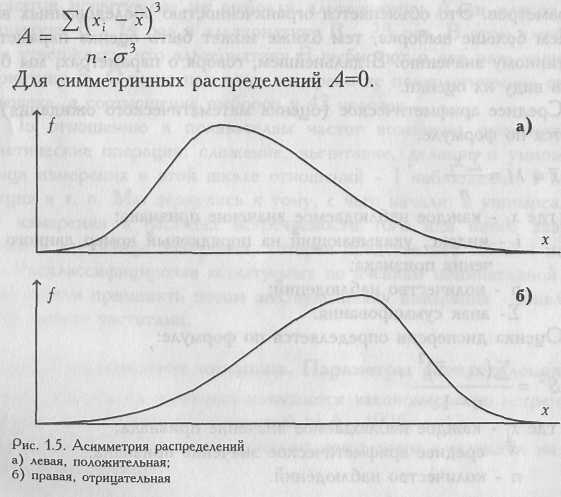
 ˗ среднее арифметическое значение признака;

n ˗ количество наблюдений.

Величина, представляющая собой квадратный корень из несмещенной оценки дисперсии (S), называется средним квадратическим отклонением. Для большинства исследователей привычно обозначать эту величину греческой буквой σ (сигма), а не S. σ ˗ это СКО в генеральной совокупности, a S ˗ несмещенная оценка этого параметра в исследованной выборке.

В тех случаях, когда какие-нибудь причины благоприятствуют более частому появлению значений, которые выше или, наоборот, ниже среднего, образуются асимметричные распределения. При левосторонней, или положительной, асимметрии в распределении чаще встречаются более низкие значения признака, а при правосторонней, или отрицательной ˗ более высокие (см. Рис. 1.5).

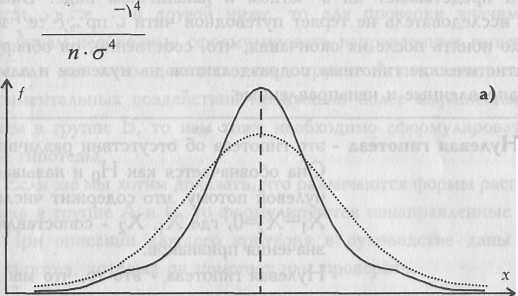
Показатель асимметрии (A)вычисляется по формуле:



В тех случаях, когда какие-либо причины способствуют преиму­щественному появлению средних или близких к средним значений, об­разуется распределение с положительным эксцессом. Если же в рас­пределении преобладают крайние значения, причем одновременно и более низкие, и более высокие, то такое распределение характеризуется отрицательным эксцессом и в центре распределения может образоваться впадина, превращающая его в двувершинное (см. Рис. 1.6).

Показатель эксцесса (E) определяется по формуле:





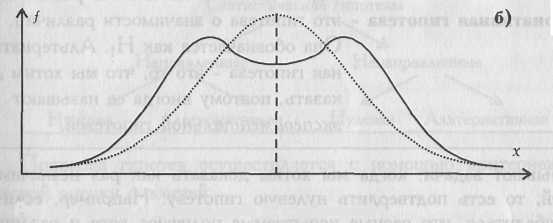


Рис. 1.6. Эксцесс: а) положительный; 6) отрицательный

В распределениях с нормальной выпуклостью E=0.

Параметры распределения оказывается возможным определить только по отношению к данным, представленным по крайней мере в интервальной шкале. Физические шкалы длин, времени, углов являются интервальными шкалами, и поэтому к ним применимы способы расчета оценок параметров, по крайней мере, с формальной точки зрения.

**1.4. Статистические гипотезы**

Формулирование гипотез систематизирует предположения иссле­дователя и представляет их в четком и лаконичном виде. Благодаря гипотезам исследователь не теряет путеводной нити в процессе расчетов и ему легко понять после их окончания, что, собственно, он обнаружил.

Статистические гипотезы подразделяются на нулевые и альтернативные, направленные и ненаправленные.

**Нулевая гипотеза** ˗ это гипотеза об **отсутствии различий**. Она обозначается как H0 и называется нулевой потому, что содержит число 0: X1 - Х2=0, где X1, X2 - сопоставляемые значения признаков. Нулевая гипотеза ˗ это то, что мы хотим опровергнуть, если перед нами стоит задача доказать значимость различий.

**Альтернативная гипотеза** ˗ это гипотеза о **значимости различий**. Она обозначается как H1. Альтернативная гипотеза ˗ это то, что мы хотим доказать, поэтому иногда ее называют *экспериментальной* гипотезой.

Бывают задачи, когда мы хотим доказать как раз незначимость различий, то есть подтвердить нулевую гипотезу. Например, если нам нужно убедиться, что разные испытуемые получают хотя и различные, но уравновешенные по трудности задания, или что экспериментальная и контрольная выборки не различаются между собой по каким-то значимым характеристикам. Однако чаще нам все-таки требуется доказать *значимость различий,* ибо они более информативны в поиске нового. Нулевая и альтернативная гипотезы могут быть **направленными и ненаправленными.**

**Направленные гипотезы**

H0: X1 не превышает Х2

H1: X1 превышает Х2

**Ненаправленные гипотезы**

H0: X1 не отличается от Х2

Н1: Х1 отличается от Х2

В одной из групп индивидуальные значения испытуемых по какому-либо признаку, например по социальной смелости, выше, а в другой ниже, то для проверки значимости этих различий нам необходимо сформулировать направленные гипотезы.

Если мы хотим доказать, что в группе А под влиянием каких-то экспериментальных воздействий произошли более выраженные изменения, чем в группе Б, то нам тоже необходимо сформулировать направленные гипотезы.

Если же мы хотим доказать, что различаются формы распределения признака в группе А и Б, то формулируются ненаправленные гипотезы.

При описании каждого критерия в руководстве даны формулировки гипотез, которые он помогает нам проверить.

Построим схему ˗ классификацию статистических гипотез.



Проверка гипотез осуществляется с помощью критериев статистической оценки различий.

**1.5. Статистические критерии**

Статистический критерий ˗ это решающее правило, обеспечивающее надежное поведение, то есть принятие истинной и отклонение ложной гипотезы с высокой вероятностью. Статистические критерии обозначают также метод расчета определенного числа и само это число.

Когда мы говорим, что достоверность различий определялась по критерию χ2, то имеем в виду, что использовали метод χ2 ˗для расчета определенного числа. Когда мы говорим, далее, что χ2=12,676, то имеем в виду определенное число, рассчитанное по методу χ2. Это число обозначается как эмпирическое значение критерия.

По соотношению эмпирического и критического значений критерия мы можем судить о том, подтверждается ли или опровергается нулевая гипотеза. Например, если χ2эмп> χ2кр, H0 отвергается.

В большинстве случаев для того, чтобы мы признали различия значимыми, необходимо, чтобы эмпирическое значение критерия превышало критическое, хотя есть критерии (например, критерий Манна-Уитни или критерий знаков), в которых мы должны придерживаться противоположного правила. Эти правила оговариваются в описании каждого из представленных в руководстве критериев.

В некоторых случаях расчетная формула критерия включает в себя количество наблюдений в исследуемой выборке, обозначаемое как *п*. В этом случае эмпирическое значение критерия одновременно является тестом для проверки статистических гипотез. По специальной таблице мы определяем, какому уровню статистической значимости различий соответствует данная эмпирическая величина. Примером такого критерия является критерий φ\*, вычисляемый на основе углового преобразования Фишера.

В большинстве случаев, однако, одно и то же эмпирическое значение критерия может оказаться значимым или незначимым в зависимости от количества наблюдений в исследуемой выборке (n) или от так называемого количества степеней свободы, которое обозначается как *v* или как *df.* Число степеней свободы *v* равно числу классов вариационного ряда минус число условий, при которых он был сформирован. К числу таких условий относятся объем выборки (n), средние и дисперсии.

Если мы расклассифицировали наблюдения по классам какой-либо номинативной шкалы и подсчитали количество наблюдений в каждой ячейке классификации, то мы получаем так называемый частотный вариационный ряд. Единственное условие, которое соблюдается при его формировании ˗ объем выборки *п.* Допустим, у нас 3 класса: "Умеет работать на компьютере ˗ умеет выполнять лишь определенные операции ˗ не умеет работать на компьютере". Выборка состоит из 50 человек. Если в первый класс отнесены 20 испытуемых, во второй ˗ тоже 20, то в третьем классе должны оказаться все остальные 10 испытуемых. Мы ограничены одним условием ˗ объемом выборки. Поэтому даже если мы потеряли данные о том, сколько человек не умеют работать на компьютере, мы можем определить это, зная, что в первом и втором классах ˗ по 20 испытуемых. Мы не свободны в определении количества испытуемых в третьем разряде, "свобода" простирается только на первые две ячейки классификации: *v*= *c*-l = 3-1 = 2

Аналогичным образом, если бы у нас была классификация из 10 разрядов, то мы были бы свободны только в 9 из них, если бы у нас было 100 классов ˗ то в 99 из них и т. д. Существуют способы более сложного подсчета числа степеней свободы при двухмерных классификациях, например критерии χ2и дисперсионном анализе.

Зная n и/или число степеней свободы, по специальным таблицам можно определить критические значения критерия и сопоставить с ними полученное эмпирическое значение. Обычно это записывается так: "при n=22 критические значения критерия составляют ..." или "при *v*=2 критические значения критерия составляют ..." и т.п.

Критерии делятся на параметрические и непараметрические.

**Параметрические критерии**

Критерии, включающие в формулу расчета параметры распределения, то есть средние и дисперсии (*t* - критерий Стьюдента, критерий F и др.)

**Непараметрические критерии**

Критерии, не включающие в формулу расчета параметров распределе­ния и основанные на оперировании частотами или рангами (критерий Q Розенбаума, критерий Т Вилкоксона и др.)

И те, и другие критерии имеют свои преимущества и недостатки. На основании нескольких руководств можно составить таблицу, позволяющую оценить возможности и ограничения тех и других.

*Таблица 1.1*

Возможности и ограничения параметрических и непараметрических критериев

|  |  |
| --- | --- |
| **ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ** | **НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ** |
| 1. Позволяют прямо оценить различия в средних, полученных в двух выборках (t-критерий Стьюдента). | Позволяют оценить лишь средние тенденции, например, ответить на вопрос, чаще ли в выборке А встречаются более высокие, а в выборке Б ˗ более низкие значения признака (критерии Q, U, φ\* и др.). |
| 2. Позволяют прямо оценить различия в дисперсиях (критерий Фишера). | Позволяют оценить лишь различия в диапазонах вариативности признака (критерий φ\*). |
| 3. Позволяют выявить тенденции изменения признака при переходе от условия к условию (дисперсионный однофакторный анализ), но лишь при условии нормального распределения признака. | Позволяют выявить тенденции изменения признака при переходе от условия к условию при любом распределении признака (критерии тенденций L и S). |
| 4. Позволяют оценить взаимодействие двух и более факторов в их влиянии на изменения признака (двухфакторный дисперсионный анализ). | Эта возможность отсутствует. |
| 5. Экспериментальные данные должны отвечать двум, а иногда трем, условиям: а) значения признака измерены по интервальной шкале; б) распределение признака является нормальным; в) в дисперсионном анализе должно соблюдаться требование равенства дисперсий в ячейках комплекса. | Экспериментальные данные могут не отвечать ни одному из этих условий: а) значения признака могут быть представлены в любой шкале, начиная от шкалы наименований; б) распределение признака может быть любым и совпадение его с каким-либо теоретическим законом распределения необязательно и не нуждается в проверке; в) требование равенства дисперсий отсутствует. |
| 6. Математические расчеты довольно сложны. | Математические расчеты по большей части просты и занимают мало времени (за исключением критериев χ2и λ). |
| 7. Если условия, перечисленные в п.5, выполняются, параметрические критерии оказываются несколько более мощными, чем непараметрические. | Если условия, перечисленные в п.5, не выполняются, непараметрические критерии оказываются более мощными, чем параметрические, так как они менее чувствительны к "засорениям'. |

Из табл. 1.1 мы видно, что параметрические критерии могут оказаться несколько более мощными[[3]](#footnote-4), чем непараметрические, но только в том случае, если признак измерен по интервальной шкале и нормально распределен. С интервальной шкалой есть определенные проблемы. Лишь с некоторой натяжкой мы можем считать данные, представленные не в стандартизованных оценках, как интервальные. Кроме того, проверка распределения "на нормальность" требует достаточно сложных расчетов, результат которых заранее неизвестен. Может оказаться, что распределение признака отличается от нормального, и нам так или иначе все равно придется обратиться к непараметрическим критериям.

Непараметрические критерии лишены всех этих ограничений и не­требуют таких длительных и сложных расчетов. По сравнению с пара­метрическими критериями они ограничены лишь в одном ˗ с их помощью невозможно оценить взаимодействие двух или более условий или факторов, влияющих на изменение признака. Эту задачу может решить только дисперсионный двухфакторный анализ.

**1.6. Уровни статистической значимости**

Уровень значимости ˗- это вероятность того, что мы сочли различия существенными, а они на самом деле случайны.

Когда мы указываем, что различия достоверны на 5%-ом уровне значимости, или при *р<*0,05, то мы имеем виду, что вероятность того, что они все-таки недостоверны, составляет 0,05. Когда мы указываем, что различия достоверны на 1%-ом уровне значимости, или при *р<*0,01, то мы имеем в виду, что вероятность того, что они все-таки недостоверны, составляет 0,01.

Таким образом, уровень значимости ˗ это вероятность отклонения нулевой гипотезы, в то время как она верна.

**Ошибка, состоящая в том, что мы отклонили нулевую гипотезу, в то время как она верна, называется ошибкой 1 рода.**

Вероятность такой ошибки обычно обозначается как α. В сущности, мы должны были бы указывать в скобках не р≤0,05 или р≤0,01, а α≤0,05 или α≤0,01.

Если вероятность ошибки ˗ это α, то вероятность правильного решения: 1 ˗ α. Чем меньше α, тем больше вероятность правильного решения.

Исторически сложилось так, что принято считать низшим уровнем статистической значимости 5%-ый уровень (*р<*0,05): достаточным ˗ 1%-ый уровень (*р<*0,01) и высшим 0,1%-ый уровень (*р<*0,001), поэтому в таблицах критических значений обычно приводятся значения критериев, соответствующих уровням статистической значимости *р<*0,05 и *р<*0,01, иногда ˗ *р<*0,001. Для некоторых критериев в таблицах указан точный уровень значимости их разных эмпирических значений. Например, для φ\*=1,56 *р=*0,06.

До тех пор, однако, пока уровень статистической значимости не достигнет *р=*0,05, мы еще не имеем права отклонить нулевую гипотезу. Будем придерживаться следующего правила отклонения гипотезы об отсутст­вии различий (H0) и принятия гипотезы о статистической достоверно­сти различий (Н1).

**Правило отклонения H0 и принятия H1**

Если эмпирическое значение критерия равняется критическому значе­нию, соответствующему *р<*0,05 или превышает его, то H0 отклоняется, но мы еще не можем определенно принять H1. Если эмпирическое значение критерия равняется критическому значению, соответствующему *р<*0,01 или превышает его, то H0 отклоняется и принимается H1.

*Исключения:* критерий знаков G, критерий Т Вилкоксона и критерий U Манна-Уитни. Для них устанавливаются обратные соотношения.

Для облегчения процесса принятия решения можно всякий раз вычерчивать "ось значимости".



Критические значения критерия обозначены как Q0,05 и Q0,01, эмпирическое значение критерия как Qэмп. Оно заключено в эллипс.

Вправо от критического значения Q0,01 находится "зона значимости" ˗ сюда попадают эмпирические значения, превышающие Q0,01 и, следовательно, безусловно значимые. Влево от критического значения Q0,05 простирается "зона незначимости", ˗ сюда попадают эмпирические значения Q, которые ниже Q0,05, и, следовательно, безусловно незначимы.

Мы видим, что Q0,05=6; Q0,01=9; Qэмп=8

Эмпирическое значение критерия попадает в область между Q0,05 и Q0,01- Это зона "неопределенности": мы уже можем отклонить гипотезу о недостоверности различий (H0), но еще не можем принять гипотезы об их достоверности (H1).

Практически, однако, исследователь может считать достоверными уже те различия, которые не попадают в зону незначимости, заявив, что они достоверны при *р<*0,05, или указав точный уровень значимости полученного эмпирического значения критерия, например: *р=*0,02. С помощью таблиц это можно сделать по отношению к критериям Н Крускала-Уоллиса, χ2*,* Фридмана, L Пейджа, φ\* Фишера, А, Колмогорова.

Уровень статистической значимости или критические значения критериев определяются по-разному при проверке направленных и не­направленных статистических гипотез.

При направленной статистической гипотезе используется одно­сторонний критерий, при ненаправленной гипотезе ˗ двусторонний критерий. Двусторонний критерий более строг, поскольку он проверяет различия в обе стороны, и поэтому то эмпирическое значение критерия, которое ранее соответствовало уровню значимости *р<*0,05, теперь соответствует лишь уровню *р<*0,10.

**1.7. Мощность критериев**

Мощность критерия ˗ это его способность выявлять различия, если они есть. Иными словами, это его способность отклонить нулевую гипотезу об отсутствии различий, если она неверна.

**Ошибка, состоящая в том, что мы приняли нулевую гипотезу, в то время как она неверна, называется ошибкой II рода.**

Вероятность такой ошибки обозначается как β. Мощность критерия ˗ это его способность не допустить ошибку II рода, поэтому: Мощность=1˗β

Мощность критерия определяется эмпирическим путем. Одни и те же задачи могут быть решены с помощью разных критериев, при этом обнаруживается, что некоторые критерии позволяют выявить различия там, где другие оказываются неспособными это сделать, или выявляют более высокий уровень значимости различий. Возникает вопрос: а зачем же тогда использовать менее мощные критерии? Дело в том, что основанием для выбора критерия может быть не только мощность, но и другие его характеристики, а именно:

а) простота;

б) более широкий диапазон использования (например, по отношению к данным, определенным по номинативной шкале, или по отношению к большим *n*);

в) применимость по отношению к неравным по объему выборкам;

г) большая информативность результатов.

**1.8. Классификация задач и методов их решения**

Множество задач исследований предполагает те или иные сопоставления. Мы сопоставляем группы испытуемых по какому-либо признаку, чтобы выявить различия между ними по этому признаку. Мы сопоставляем то, что было "до" с тем, что стало "после" наших экспериментальных или любых иных воздействий, чтобы определить эффективность этих воздействий. Мы сопоставляем эмпирическое распределение значений признака с каким-либо теоретическим законом распределения или два эмпирических распределения между собой, с тем, чтобы доказать неслучайность выбора альтернатив или различия в форме распределений.

Далее можно сопоставлять два признака, измеренные на одной и той же выборке испытуемых, для того, чтобы установить степень согласованности их изменений, их сопряженность, корреляцию между ними.

Наконец, можно сопоставлять индивидуальные значения, полученные при разных комбинациях каких-либо существенных условий, с тем, чтобы выявить характер взаимодействия этих условий в их влиянии на индивидуальные значения признака.

Краткая классификация задач и методов дана в Таблице 1.2.

*Таблица 1.2*

|  |  |  |
| --- | --- | --- |
| Классификация задач | | и методов их решения |
| **Задачи** | **Условия** | **Методы** |
| 1.Выявление различий в уровне исследуемого признака | а) 2 выборки испытуемых | **Q**-критерий Розенбаума;  **U**-критерий Манна-Уитни;  ***φ\****-критерий (угловое преобразование Фишера) |
| б) 3 и более выборок испытуемых | **S**-критерий тенденций Джонкира;  **Н**-критерий Крускала-Уоллиса. |
| 2. Оценка сдвига значений исследуемого признака | а) 2 замера на одной и той же выборке испытуемых | **Т**-критерий Вилкоксона;  **G**-критерий знаков;  ***φ\****-критерий (угловое преобразование Фишера). |
| б) 3 и более замеров на одной и той же выборке испы­туемых | **χл2**-критерий Фридмана;  **L**-критерий тенденций Пейджа. |
| 3. Выявление различий в распределении | а) при сопоставлении эмпирического признака распределения с теоретическим | **χ2**-критерий Пирсона;  **λ***-*критерий Колмогорова-Смирнова;  **m**-биномиальный критерий. |
| б) при сопоставлении двух эмпириче­ских распределений | **χ2***-*критерий Пирсона;  **λ**-критерий Колмогорова-Смирнова;  ***φ\****-критерий (угловое преобразование Фишера). |
| 4.Выявление степени согласованности изменений | а) двух признаков | ***rs***-коэффициент ранговой корреляции Спирмена. |
| б) двух иерархий или профилей | ***rs***- коэффициент ранговой корреляции Спирмена. |
| 5. Анализ изменений признака под влиянием контролируе­мых условий | а) под влиянием одного фактора | **S**-критерий тенденций Джонкира;  **L**-критерий тенденций Пейджа; однофакторный дисперсионный анализ Фишера. |
| б) под влиянием двух факторов одновременно | Двухфакторный дисперсионный анализ Фишера. |

**1.9. Принятие решения о выборе метода математической обработки**

Если данные уже получены, то предлагается следующий алгоритм определения задачи и метода.

**АЛГОРИТМ 1**

**Принятие решения о задаче и методе обработки на стадии, когда данные уже получены**

1. По первому столбцу табл. 1.2 определить, какая из задач стоит в вашем исследовании.

2. По второму столбцу табл. 1.2 определить, каковы условия решения вашей задачи, например, сколько выборок обследовано или на какое количество групп вы можете разделить обследованную выборку.

3. По алгоритму принятия решения о выборе критерия определить, какой именно метод или критерий целесообразно использовать.

На стадии планирования исследования лучше заранее подобрать математическую модель, которую будете в дальнейшем использовать. Особенно необходимо планирование в тех случаях, когда в перспективе предполагается использование критериев тенденций или (в еще большей степени) дисперсионного анализа. В этом случае алгоритм принятия решения следующий:

**АЛГОРИТМ 2**

**Принятие решения о задаче и методе обработки на стадии планирования исследования**

1. Определите, какая модель наиболее подходящая для доказательства научных предположений.

2. Ознакомьтесь с описанием метода.

3. Рассмотрите "Ограничения критерия" и решите, сможете ли собрать данные, которые будут отвечать этим ограничениям (большие объемы выборок, наличие нескольких выборок, монотонно различающихся по какому-либо признаку, например, по возрасту и т.п.).

4. Проводите исследование, а затем обрабатывайте полученные данные по заранее выбранному алгоритму, если удалось выполнить ограничения.

5. Если ограничения выполнить не удалось, обратитесь к алгоритму 1.

В описании каждого критерия сохраняется следующая последова­тельность изложения:

* назначение критерия;
* описание критерия;
* гипотезы, которые он позволяет проверить;
* графическое представление критерия;
* ограничения критерия;
* пример или примеры.

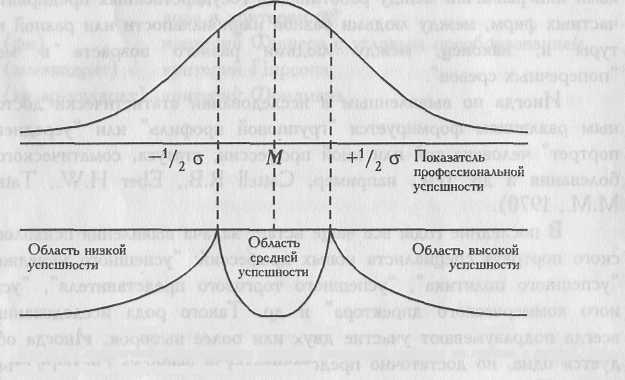
**ВЫЯВЛЕНИЕ РАЗЛИЧИЙ В УРОВНЕ ИССЛЕДУЕМОГО ПРИЗНАКА**

**2.1. Обоснование задачи сопоставления и сравнения**

Очень часто перед исследователем стоит задача выявления различий между двумя, тремя и более выборками испытуемых. Это может быть, например, задача определения психологических особенностей хронически больных детей по сравнению со здоровыми, юных правонарушителей по сравнению с законопослушными сверстниками или различий между работниками государственных предприятий и частных фирм, между людьми разной национальности или разной культуры и, наконец, между людьми разного возраста в методе "поперечных срезов".

Иногда по выявленным в исследовании статистически достоверным различиям формируется "групповой профиль" или "усредненный портрет" человека той или иной профессии, статуса, соматического заболевания и др. В последние годы все чаще встает задача выявления психологического портрета специалиста новых профессий: "успешного менеджера", "успешного политика", "успешного торгового представителя", "успешного коммерческого директора" и др. Такого рода исследования не всегда подразумевают участие двух или более выборок. Иногда обследуется одна, но достаточно представительная выборка численностью не менее 60 человек, а затем внутри, этой выборки выделяются группы более и менее успешных специалистов, и их данные по исследованным переменным сопоставляются между собой. В самом простом случае критерием для разделения выборки на "успешных" и "неуспешных" будет средняя величина по показателю успешности. Однако такое деление является довольно грубым: лица, получившие близкие оценки по успешности, могут оказаться в противоположных группах, а лица, заметно различающиеся по оценкам успешности, ˗ в одной и той же группе. Это может исказить результаты сопоставления групп или по крайней мере сделать различия между группами менее заметными.

Чтобы избежать этого, можно попробовать выделить группы "успешных" и "неуспешных" специалистов более строго, включая в первую из них только тех, чьи значения *превышают* среднюю величину не менее чем на 1/4 стандартного отклонения, а во вторую группу ˗ только тех, чьи значения не менее чем на 1/4 стандартного отклонения *ниже* средней величины. При этом все, кто оказывается в зоне средних величин, М±1/4 σ, выпадают из дальнейших сопоставлений. Если распределение близко к нормальному, то выпадет примерно 19,8% испытуемых. Если распределение отличается от нормального, то таких испытуемых может быть и больше. Чтобы избежать потерь, можно сопоставлять не две, а три группы испытуемых: с высокой, средней и низкой профессиональной успешностью.



(30,9% испытуемых) (38,2% испытуемых) (30,9% испытуемых)

Рис 2.1. Схематическое изображение процесса разделения выборки на группы с низкой, средней и высокой профессиональной успешностью

На Рис. 2.1 представлена схема разделения выборки на группы с низкой, средней и высокой профессиональной успешностью по критерию отклонения значений от средней величины на 1/2 стандартного отклонения. При таком строгом критерии в "среднюю" группу попадают (при нормальном распределении) около 38,2% всех испытуемых, а в крайних группах оказывается по 30,9% испытуемых.

Чем меньше испытуемых оказывается в группах, тем меньше у нас возможностей для выявления достоверных различий, так как критические значения большинства критериев при малых *п* строже, чем при больших *п.*

Таким образом, при нестрогом разделении испытуемых на группы мы теряем в точности, а при строгом ˗ в количестве испытуемых.

При решении задач выявления различий в уровневых показателях следует помнить, что "усредненный профиль успешного специалиста" должен рассматриваться скорее как исследовательский результат, по­зволяющий сформулировать гипотезы для дальнейших исследований, а не как основание для профессионального отбора. Тому есть две причины. Во-первых, ни у одного из успешных специалистов может не наблюдаться "усредненный профиль" ˗ он, в сущности, является отвлеченным обобщением; во-вторых, в профессиональной деятельности наличие собственного индивидуального стиля важнее соответствия "среднегрупповому" профилю. Недостаток в тех качествах, которые могут казаться важными, компенсируется другими качествами. У каждого успешного специалиста его психологические свойства создают не­повторимый ансамбль, который при усреднении данных теряется.

Сопоставление уровневых показателей в разных выборках может быть необходимой частью комплексных диагностических, учебных, пси-хокоррекционных и иных программ. Оно помогает обратить внимание на те особенности обследованных выборок, которые должны быть учтены и использованы при адаптации программ к данной группе в процессе их конкретного воплощения.

Критерии предполагают, что сопоставляются так называемые независимые выборки, то есть две или более выборки, состоящие из разных испытуемых. Тот испытуемый, который входит в одну выборку, уже не может входить в другую. В противоположность этому, если мы обследуем одну и ту же выборку испытуемых, несколько раз подвергая ее аналогичным измерениям ("замерам"), то перед нами ˗ так называемые связанные, или за­висимые, выборки данных.

Решение о выборе того или иного критерия принимается на основе того, сколько выборок сопоставляется и каков их объем.

**Критерий знаков.**

**G- критерий знаков**

**Назначение критерия G**

Критерий знаков[[4]](#footnote-5) 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.

**Описание критерия G**

Критерий знаков применим и к тем сдвигам, которые можно определить лишь *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 Т Вилкоксона. Он учитывает не только направление, но и интенсивность сдвигов и может оказаться более мощным в определении достоверности сдвигов, чем критерий знаков.

Как правило, исследователь уже в процессе эксперимента может заметить, что у большинства испытуемых показатели во втором замере имеют тенденцию, скажем, повышаться. Однако ему еще требуется доказать, что положительный сдвиг является преобладающим.

Сдвиги, которые нам кажутся преобладающими, называются 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 у большинства испытуемых отрицательное отношение сменилось на положительное, то этот сдвиг 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, не 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 превалирующим. Gэмп ˗ это количество "нетипичных" сдвигов. Чем меньше Gэмп, тем более вероятно, что сдвиг в "типичном" направлении статистически достоверен.

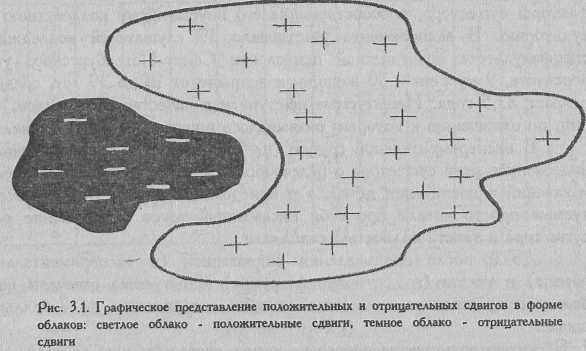
**Гипотезы**

Н0: Преобладание типичного направления сдвига является случайным.

H1: Преобладание типичного направления сдвига не является случайным.

**Графическое представление критерия знаков**

На Рис. 3.1"типичные" сдвиги изображены в виде светлого облака, а нетипичные сдвиги ˗ темного облака. Мы видим, что на рисунке темное облако значительное меньше. Допустим, после выступления оратора большинство слушателей изменили свое отрицательное отношение к какому-то предложению на положительное. Вместе с тем, часть слушателей изменила свое положительное отношение на отрицательное, проявив "нетипичную" реакцию. Критерий знаков позволяет определить, не слишком ли значительная часть слушателей нетипично прореагировала на выступление оратора? Поглощает ли масса светлого облака небольшое темное облако?



В Таблице V Приложения 1 даны критические значения критерия знаков для разных *п*. Поскольку критерий знаков представляет собой одно из трех исключений из общего правила, представим обобщенную "ось значимости" для этого критерия графически (Рис. 3.2)



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 (n1=16) испытуемые просматривали видеозапись речи суггестора о целесообразности применения физических наказаний в воспитании детей, а в контрольной группе (n2=23) испытуемые просто читали про себя письменный текст. Содержание речи суггестора и текста полностью совпадали. До и после предъявления видеозаписи (в экспериментальной группе) и текста (в контрольной группе) испытуемые отвечали на 4 вопроса, оценивая степень согласия с их содержанием по 7-балльной шкале:

1. Я считаю возможным иногда шлепнуть своего ребенка за дело, если

он этого заслужил:

Несогласен 1 2 3 4 5 6 7 Согласен

2. Если, придя домой, я узнаю, что кто-то из близких, бабушка или дедушка, шлепнул моего ребенка за дело, то я буду считать, что это нормально:

Несогласен 1 2 3 4 5 6 7 Согласен

3. Если мне станет известно, что воспитательница детского сада или учительница в школе шлепнула моего ребенка за дело, то я восприму это как должное:

Несогласен 1 2 3 4 5 6 7 Согласен

4. Я бы согласился отдать своего ребенка в школу, где применяется

система физических наказаний по итогам недели:

Несогласен 1 2 3 4 5 6 7 Согласен

Суггестор был подобран по признакам, которые были выявлены в пилотажном исследовании. Результаты двух замеров по обеим группам представлены в Табл. 3.2 и Табл. 3 3

*Таблица 3.2*

Оценки степени согласия с утверждениями о допустимости телесных наказаний до и после предъявления видеозаписи в экспериментальной группе (n1=16)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | Оценки и сдвиги оценок ("после" - "до") по шкалам | | | | | | | | | | | |
| "Я сам" | | | "Бабушке" | | | "Воспитатель" | | | Школа | | |
| до | после | сдвиг | до | после | сдвиг | до | после | сдвиг | до | после | сдвиг |
| 1 | 4 | 4 | 0 | 2 | 4 | +2 | 1 | 1 | 0 | 1 | 1 | 0 |
| 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 3 | 5 | 5 | 0 | 4 | 4 | 0 | 4 | 4 | 0 | 1 | 1 | 0 |
| 4 | 4 | 5 | +1 | 3 | 3 | 0 | 2 | 3 | +1 | 1 | 2 | +1 |
| 5 | 3 | 3 | 0 | 3 | 4 | +1 | 2 | 3 | +1 | 1 | 1 | 0 |
| 6 | 4 | 5 | +1 | 5 | 5 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 7 | 3 | 3 | 0 | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 8 | 5 | 6 | +1 | 5 | 6 | +1 | 3 | 3 | 0 | 2 | 1 | -1 |
| 9 | 6 | 7 | +1 | 5 | 7 | +2 | 3 | 3 | 0 | 1 | 2 | +1 |
| 10 | 2 | 3 | +1 | 2 | 3 | +1 | 2 | 1 | -1 | 1 | 1 | 0 |
| 11 | 6 | 6 | 0 | 3 | 3 | 0 | 2 | 1 | -1 | 1 | 1 | 0 |
| 12 | 5 | 5 | 0 | 3 | 5 | +2 | 4 | 4 | 0 | 1 | 1 | 0 |
| 13 | 7 | 7 | 0 | 5 | 5 | 0 | 4 | 4 | 0 | 1 | 1 | 0 |
| 14 | 5 | 6 | +1 | 5 | 6 | +1 | 2 | 2 | 0 | 1 | 2 | +1 |
| 15 | 5 | 6 | +1 | 5 | 6 | +1 | 4 | 3 | -1 | 2 | 2 | 0 |
| 16 | 6 | 7 | +1 | 6 | 7 | +1 | 4 | 4 | 0 | 2 | 2 | 0 |

*Таблица 3.3*

Оценки степени согласия с утверждениями о допустимости телесных наказаний до и после предъявления письменного текста в контрольной группе (n2=23)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| №  п/п | Оценки и сдвиги оценок (после - "до") по шкалам | | | | | | | | | | | |
| "Я сам" | | | "Бабушка | | | "Воспитатель" | | | Школа | | |
| до | после | сдвиг | до | после | сдвиг | до | после | сдвиг | до | после | сдвиг |
| 1 | 4 | 4 | 0 | 5 | 5 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 2 | 7 | 7 | 0 | 7 | 7 | 0 | 7 | 7 | 0 | 4 | 4 | 0 |
| 3 | 2 | 2 | 0 | 1 | 1 | 0 | 3 | 1 | -2 | 1 | 1 | 0 |
| 4 | 4 | 3 | -1 | 3 | 2 | -1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 5 | 3 | 5 | +2 | 5 | 5 | 0 | 3 | 3 | 0 | 1 | 1 | 0 |
| 6 | 2 | 1 | -1 | 2 | 1 | -1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 7 | 5 | 5 | 0 | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 8 | 2 | 2 | 0 | 2 | 3 | +1 | 1 | 3 | +2 | 1 | 3 | +2 |
| 9 | 3 | 4 | +1 | 3 | 4 | +1 | 1 | 1 | 0 | 1 | 6 | +5 |
| 10 | 5 | 5 | 0 | 5 | 5 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 11 | 5 | 5 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 12 | 2 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 13 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | +1 | 6 | 7 | +1 |
| 14 | 4 | 3 | -1 | 7 | 5 | -2 | 2 | 4 | +2 | 1 | 1 | 0 |
| 15 | 3 | 4 | +1 | 2 | 3 | +1 | 1 | 2 | +1 | 1 | 1 | 0 |
| 16 | 4 | 4 | 0 | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 17 | 3 | 3 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 18 | 6 | 6 | 0 | 6 | 6 | 0 | 6 | 6 | 0 | 1 | 3 | +2 |
| 19 | 2 | 2 | 0 | 2 | 1 | -1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 20 | 1 | 2 | +1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 21 | 2 | 2 | 0 | 2 | 2 | 0 | 2 | 1 | -1 | 1 | 1 | 0 |
| 22 | 6 | 6 | 0 | 6 | 6 | 0 | 3 | 3 | 0 | 1 | 1 | 0 |
| 23 | 3 | 2 | -1 | 1 | 2 | +1 | 1 | 1 | 0 | 1 | 1 | 0 |

*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?

*Решение*

Подсчитаем сначала количество положительных, отрицательных и нулевых сдвигов по каждой шкале в каждой из выборок. Это необхо­димо для выявления "типичных" знаков изменения оценок и значительно облегчит нам дальнейшие расчеты и рассуждения.

*Таблица 3.4*

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Кол-во сдвигов | Шкалы | | | | |
| в группах | "Я сам" | "Бабушка" | "Воспитатель" | "Школа" | Суммы |
| 1. Экспериментальная группа | | | | | |
| а) положительных б) отрицательных  в) нулевых | 8  0  8 | 9  0  7 | 2  3  11 | 3  1  12 | 22  4  38 |
| Суммы | 16 | 16 | 16 | 16 | 64 |
| 2. Контрольная группа | | | | | |
| а) положительных 6) отрицательных  в) нулевых | 4  4  15 | 4  4  15 | 4  2  17 | 4  0  19 | 16  10  66 |
| Суммы | 23 | 23 | 23 | 23 | 92 |

Из табл. 3.4 видим, что наиболее типичными являются "нулевые" сдвиги, то есть отсутствие сдвига в оценках после предъявления видеозаписи или письменного текста. И 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. Мы видим, что по отношению к последней шкале критерий знаков вообще неприменим, так как количество сопоставляемых пар значений меньше 5.

Можно сразу же проверить и гипотезу о преобладании положительного сдвига в ответах по сумме 4 шкал. Сумма положительных и отрицательных сдвигов по 4 шкалам составляет: *n=*8+9+5+4=26.

Сформулируем гипотезы.

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

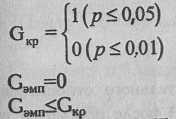
H1: Сдвиг в сторону более снисходительного отношения к телесным наказаниям после внушения является неслучайным.

По табл. определяем критические значения критерия знаков G. 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.

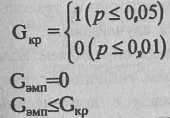


Н0 отклоняется. Принимается H1 (p<0,01).

*2) Шкала "Бабушка наказывает" n=*9

Типичный сдвиг ˗ положительный.

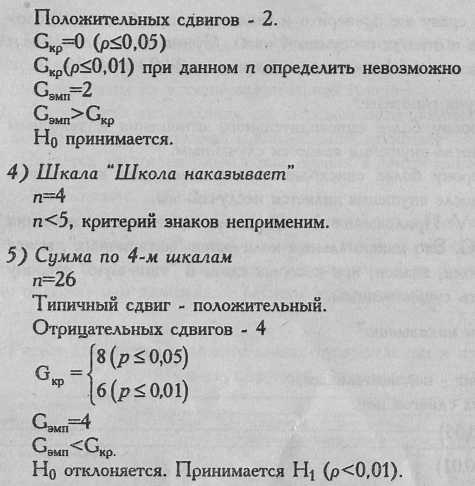
Отрицательных сдвигов нет.



Н0 отклоняется. Принимается H1 (p<0,01).

*3) Шкала "Воспитательница наказывает" n=*5

Типичный сдвиг ˗ отрицательный.



*Ответ:* Сдвиг в сторону более снисходительного отношения к телесным наказаниям в экспериментальной группе после просмотра видеозаписи является неслучайным для шкал "Я сам наказываю", "Бабушка наказывает" и по сумме четырех шкал (р<0,01 во всех случаях).

Сформулируем гипотезы для контрольной группы.

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

H1: 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 и количество нетипичных сдвигов (Gэмп) сопоставляем с критическими значениям G, определяемыми по табл.

*1) Шкала "Я сам наказываю" n*=8

Положительных сдвигов ˗ 4, отрицательных сдвигов ˗ 4.

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

Н0 принимается.

*2) Шкала "Бабушка наказывает" n=*8

Положительных сдвигов ˗ 4, отрицательных сдвигов ˗ 4.

Н0 принимается по тем же основаниям, что и для предыдущей шкалы.

*3) Шкала "Воспитательница наказывает" n=*6

Типичный сдвиг ˗ положительный.

Отрицательных сдвигов ˗ 2.

Gкp=0 (p≤0,05)

Gкр(p≤0,01) при данном *п* определить невозможно.

Gэмп*=*2

Gэм*n>*Gкp

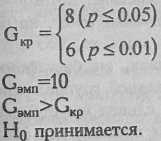
Н0 принимается.

*4) Шкала "Школа наказывает"*

Поскольку *n<*5, критерий знаков неприменим.

*5) Сумма по 4-м шкалам n=*26

Типичный сдвиг ˗ положительный. Количество отрицательных сдвигов ˗ 10.



*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 по выраженности положительного сдвига между экспериментальной и контрольной группами?

Дело в том, что 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 двух группах с помощью критериев для сравнения независимых выборок ˗ Q-критерия Розенбаума, U-критерия Манна-Уитни и критерия φ\* Фишера (см. табл. 3.1). 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, выраженнее (см. Задачу 1). В данном же случае нами доказано, что в контрольной выборке не произошло сколько-нибудь значимых изменений, и мы можем этим удовлетвориться.

Казалось бы, мы доказали все, что необходимо: в экспериментальной группе испытуемые стали снисходительнее относиться к телесным наказаниям, а в контрольной группе достоверных сдвигов не обнаружено. Похоже, суггестор, отобранный по выявленным качествам, действительно повлиял на изменение оценок, и притом именно он, что-то в его личности оказало это воздействие, потому что контрольной группе предъявлялся тот же по содержанию текст, но без суггестора. Однако, на самом деле 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 8**

**Расчет критерия знаков G**

1. Подсчитать количество нулевых реакций и исключить их из рас­смотрения.

В результате *п* 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 табл. определить критические значения G для данного п.

5. Сопоставить Gэмп с Gкр. Если Gэмп меньше Gкр или, по крайней мере, равен ему, сдвиг в типичную сторону может считаться достоверным.

**Проверка на нормальность путем оценки разброса через количество СКО.**

**Критерий Стьюдента для оценки расхождения средних арифметических выборок (однородности).**

**…………**

**Определение наличия корреляции и линейная аппроксимация.**

Критерий Манна-Уитни (*U*-критерий) для оценки различий между двумя малыми выборками

**Назначение критерия**

Критерий предназначен для оценки различий между *двумя* выборками по *уровню* какого-либо признака, количественно измеренного. Он позволяет выявлять различия между *малыми* выборками, когда *n1•n2≥*3 или *n1=2, n2≥5,* и является более мощным, чем критерий Розенбаума.

**Описание критерия**

Существует несколько способов использования критерия и несколько вариантов таблиц критических значений, соответствующих этим способам.

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами. Мы помним, что 1-м рядом (выборкой, группой) мы называем тот ряд значений, в котором значения, по предварительной оценке, выше, а 2-м рядом ˗ тот, где они предположительно ниже.

Чем меньше область перекрещивающихся значений, тем более вероятно, что различия достоверны. Иногда эти различия называют различиями в *расположении* двух выборок.

Эмпирическое значение критерия U отражает то, насколько велика зона совпадения между рядами. Поэтому *чем меньше* Uэмп, *тем более* вероятно, что различия *достоверны.*

**Гипотезы**

Н0: Уровень признака в группе 2 не ниже уровня признака в группе 1.

H1: Уровень признака в группе 2 ниже уровня признака в группе 1.

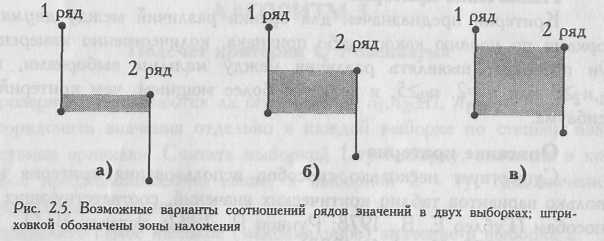
**Графическое представление критерия U**

На рис. 2.5. представлены три из множества возможных вариантов соотношения двух рядов значений.

В варианте (а) второй ряд ниже первого, и ряды почти не пере­крещиваются. Область наложения слишком мала, чтобы скрадывать различия между рядами. Есть шанс, что различия между ними достоверны. Точно определить это мы сможем с помощью критерия U.

В варианте (б) второй ряд тоже ниже первого, но и область пе­рекрещивающихся значений у двух рядов достаточно обширна. Она может еще не достигать критической величины, когда различия придется признать несущественными. Но так ли это, можно определить только путем точного подсчета критерия U.

В варианте (в) второй ряд ниже первого, но область наложения настолько обширна, что различия между рядами скрадываются.



**Ограничения критерия U**

1. В каждой выборке должно быть не менее 3 наблюдений: *n1•n2≥3;* допускается, чтобы в одной выборке было 2 наблюдения, но тогда во второй их должно быть не менее 5.

2. В каждой выборке должно быть не более 60 наблюдений; *n1•n2≤60.* Однако уже при *n1•n2>20* ранжирование становится достаточно трудоемким.

В случае, если *n1•n2>20,* лучше использовать другой критерий, а именно угловое преобразование Фишера в комбинации с критерием λ, позволяющим выявить критическую точку, в которой накапливаются максимальные различия между двумя сопоставляемыми выборками (см. п. 5.4). .Формулировка звучит сложно, но сам метод достаточно прост. Каждому исследователю лучше попробовать разные пути и выбрать тот, который кажется ему более подходящим.

**Пример**

Вернемся к результатам обследования студентов физического и психологического факультетов Ленинградского университета с помощью методики Д. Векслера для измерения вербального и невербального ин­теллекта. С помощью критерия Q Розенбаума мы в предыдущем параграфе смогли с высоким уровнем значимости определить, что уровень вербального интеллекта в выборке студентов физического факультета выше. Попытаемся установить теперь, воспроизводится ли этот результат при сопоставлении выборок по уровню невербального интеллекта. Данные приведены в табл. 2.3.

Можно ли утверждать, что одна из выборок превосходит другую по уровню невербального интеллекта?

*Таблица 2.3*

Индивидуальные значения невербального интеллекта в выборках студентов физического *(n1=\4)* и психологического (n2=12) факультетов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Студенты-физики | | | Студенты-психологи | | |
| Код имени  испытуемого | | Показатель невербального  интеллекта | Код  имени  испытуемого | | Показатель невербального  интеллекта |
| 1. | И.А. | 111 | 1. | Н.Т. | ИЗ |
| 2. | К.А. | 104 | 2. | О.В. | 107 |
| 3. | К.Е. | 107 | 3. | Е.В. | 123 |
| 4. | П.А. | 90 | 4. | Ф.О. | 122 |
| 5. | С.А. | 115 | 5. | И.Н. | 117 |
| 6. | Ст.А. | 107 | 6. | И.Ч. | 112 |
| 7. | Т.А. | 106 | 7. | И.В. | 105 |
| 8. | Ф.А. | 107 | 8. | К.О. | 108 |
| 9. | Ч.И. | 95 | 9. | P.P. | 111 |
| 10. | ЦА. | 116 | 10. | Р.И. | 114 |
| 11. | См.А. | 127 | 11. | O.K. | 102 |
| 12. | К.Ан. | 115 | 12. | Н.К. | 104 |
| 13. | Б.Л. | 102 |  |  |  |
| 14. | Ф.В. | 99 |  |  |  |

Критерий U требует тщательности и внимания. Прежде всего, необходимо помнить правила ранжирования.

**Правила ранжирования**

1. Меньшему значению начисляется меньший ранг. Наименьшему значению начисляется ранг 1.

Наибольшему значению начисляется ранг, соответствующий количеству ранжируемых значений. Например, если n=7, то наибольшее значение получит ранг 7, за возможным исключением для тех случаев, которые предусмотрены правилом 2.

2. В случае, если несколько значений равны, им начисляется ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны.

Например, 3 наименьших значения равны 10 секундам. Если бы мы измеряли время более точно, то эти значения могли бы различаться и составляли бы, скажем, 10,2 сек; 10,5 сек; 10,7 сек. В этом случае они получили бы ранги, соответственно, 1, 2 и 3. Но поскольку полученные нами значения равны, каждое из них получает средний ранг:



Допустим, следующие 2 значения равны 12 сек. Они должны были бы получить ранги 4 и 5, но, поскольку они равны, то получают средний ранг:



3. Общая сумма рангов должна совпадать с расчетной, которая опре­деляется по формуле:



где *N* ˗ общее количество ранжируемых наблюдений (значений). Несовпадение реальной и расчетной сумм рангов будет свидетельствовать об ошибке, допущенной при начислении рангов или их суммировании. Прежде чем продолжить работу, необходимо найти ошибку и устранить ее.

При подсчете критерия U легче всего сразу приучить себя действовать по строгому алгоритму.

**АЛГОРИТМ 4**

**Подсчет критерия U Манна-Уитни.**

1. Перенести все данные испытуемых на индивидуальные карточки.

2. Пометить карточки испытуемых выборки 1 одним цветом, скажем красным, а все карточки из выборки 2 ˗ другим, например, синим.

3. Разложить все карточки в единый ряд по степени нарастания при­знака, не считаясь с тем, к какой выборке они относятся, как если бы мы работали с одной большой выборкой.

4. Проранжировать значения на карточках, приписывая меньшему зна­чению меньший ранг. Всего рангов получится столько, сколько у нас *(n1+п2).*

*5.* Вновь разложить карточки на две группы, ориентируясь на цветные обозначения: красные карточки в один ряд, синие ˗ в другой.

6. Подсчитать сумму рангов отдельно на красных карточках (выборка 1) и на синих карточках (выборка 2). Проверить, совпадает ли общая сумма рангов с расчетной.

7. Определить большую из двух ранговых сумм.

8. Определить значение U по формуле:



где *n1* ˗количество испытуемых в выборке 1;

*n2* ˗ количество испытуемых в выборке 2;

*Тх* ˗большая из двух ранговых сумм;

nх ˗ количество испытуемых в группе с большей суммой рангов.

9. Определить критические значения U по табл. Если Uэмп.>Uкp 005, Но принимается. Если Uэмп*≤*Uкp\_005, Но отвергается. Чем меньше значения U, тем достоверность различий выше.

Теперь проделаем всю эту работу на материале данного примера. В результате работы по 1-6 шагам алгоритма построим таблицу.

*Таблица 2.4*

Подсчет ранговых сумм по выборкам студентов физического и психа-логического факультетов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студенты-физики (n1=14) | | | Студенты-психологи (n2=12) | |
| Показатель невербального  интеллекта | | Ранг | Показатель невербального  интеллекта | Ранг |
|  |
| 127 | | 26 |  |  |
|  | |  | 123 | 25 |
|  | |  | 122 | 24 |
|  | |  | 117 | 23 |
| 116 | | 22 |  |  |
| 115 | | 20,5 |  |  |
| 115 | | 20,5 |  |  |
|  | |  | 114 | 19 |
|  | |  | 113 | 18 |
|  | |  | 112 | 17 |
| 111 | | 15,5 | 111 | 15.5 |
|  | |  | 108 | 14' |
| 107 | | 11.5 | 107 | 11,5 |
| 107 | | 11,5 |  |  |
| 107 | | 11,5 |  |  |
| 106 | | 9 |  |  |
|  | |  | 105 | 8 |
| 104 | | 6.5 | 104 | 6,5 |
| 102 | | 4,5 | 102 | 4,5 |
| 99 | | 3 |  |  |
| 95 | | 2 |  |  |
| 90 | | 1 |  |  |
| Суммы | 1501 | 165 | 1338 | 186 |
| Средние | 107,2 |  | 111,5 |  |

Общая сумма рангов: 165+186=351. Расчетная сумма:



Равенство реальной и расчетной сумм соблюдено.

Видим, что по уровню невербального интеллекта более "высоким" рядом оказывается выборка студентов-психологов. Именно на эту выборку приходится большая ранговая сумма: 186.

Теперь мы готовы сформулировать гипотезы:

H0: Группа студентов-психологов не превосходит группу студентов-физиков по уровню невербального интеллекта.

Н1: Группа студентов-психологов превосходит группу студентов-физиков по уровню невербального интеллекта.

В соответствии со следующим шагом алгоритма определяем эмпи­рическую величину U:

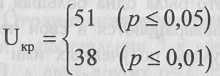


Поскольку в нашем случае *п\Фп2,* подсчитаем эмпирическую величину U и для второй ранговой суммы (165), подставляя в формулу соответствующее ей *пх:*



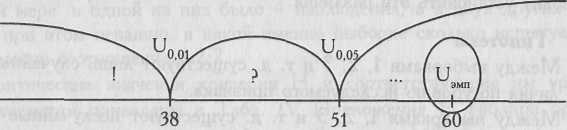
Такую проверку рекомендуется производить. Для сопоставления с критическим значением выбираем меньшую величину U: Uэмп=60.

По табл. определяем критические значения для *n1*=14, *n2=12.*



Помним, что критерий U является одним из двух исключений из общего правила принятия решения о достоверности различий, а именно можем констатировать достоверные различия, если Uэмп*≤*Uкp

Построим "ось значимости".



Uэмп**=**60

Uэмп**>**Uкp

*Ответ:* H0 принимается. Группа студентов-психологов не превосходит группы студентов-физиков по уровню невербального интеллекта.

Обратим внимание на то, что для данного случая критерий Q Розенбаума неприменим, так как размах вариативности в группе физиков шире, чем в группе психологов: и самое высокое, и самое низкое значение невербального интеллекта приходится на группу физиков (см. Табл. 2.4).

1. Определения и формулы расчета М и СТ даны в параграфе "Распределение при­знака. Параметры распределения". [↑](#footnote-ref-2)
2. О нормальном распределении см. Пояснения в п. 1.3. [↑](#footnote-ref-3)
3. О понятии мощности критерия см. ниже. [↑](#footnote-ref-4)
4. Критерий знаков с математической точки зрения является частным случаем биномиального критерий для двух равновероятных альтернатив. При вероятности каждой из альтернатив P=Q=0,50 критерий знаков является зеркальным отраже­нием биномиального критерия (см. параграф 5.3). В некоторых руководствах критерий знаков называют критерием Мак-Немара (McCall R., 1970; Рунион Р., 1982). [↑](#footnote-ref-5)