ПСИХОЛОГИЧЕСКИЙ ЖУРНАЛ, 2011, том 32, № 1, с. 97–110

МАТЕМАТИЧЕСКАЯ ПСИХОЛОГИЯ

УСЛОВИЯ ПРИМЕНИМОСТИ КРИТЕРИЕВ СТЬЮДЕНТА И МАННА–УИТНИ

А. А. Корнеев[*], **А. Н. Кричевец**[**]

[*] *Кандидат психологических наук, младший научный сотрудник, факультет психологии МГУ имени М.В. Ломоносова, Москва; e-mail: korneeff@gmail.com*

[**] *Кандидат физико-математических наук, доктор философских наук, профессор, там же; e-mail: ankrich@mail.ru.*

Исследуются условия применимости критериев Стьюдента и Манна–Уитни – наиболее распространенных методов оценки сдвига центральных тенденций выборок. Последствия нарушений данных условий оцениваются методом Монте-Карло, т.е. моделированием случайных величин с соответствующими распределениями, формированием выборок с помощью их испытаний и оценки частот ошибок первого рода при использовании критериев. Даны рекомендации по выбору подходящего критерия.

Ключевые слова: статистические методы, условия применимости, критерий Стьюдента, критерий Манна–Уитни.

Наиболее часто встречающаяся в психологических исследованиях задача проверки статистических гипотез – это сравнение центральных тенденций двух независимых выборок. Наиболее часто встречающиеся методы ее решения – критерий Стьюдента (t-критерий) и непараметрический метод Манна–Уитни. Несмотря на то, что все пособия, учебники и справочники, ориентированные на психологов, описывают оба метода, рекомендации к употреблению того или другого довольно сильно разнятся. В любом руководстве читатель найдет упоминание о том, что t-критерий применим для нормально распределенных выборок, но почти нигде не упоминаются условия применимости критерия Манна–Уитни. Судя по распространенной рекомендации (сначала проверить подходящим методом, можно ли считать распределение тестируемых выборок нормальным[1], в случае положительного ответа применить критерий Стьюдента, в случае отрицательного – критерий Манна–Уитни), последний считается применимым при любых обстоятельствах (свободным от распределения, как иногда пишут [4, с. 68]), но это совсем не так.

В данной работе были поставлены следующие *задачи:*

1. Проанализировать логическую структуру проверок распределения, независимо от того, что и как проверяется;
2. Обсудить условия применимости критериев Стьюдента и Манна–Уитни. Сравнить, какие из параметров распределений и как влияют на вероятности ошибок первого рода для обоих критериев;
3. Обсудить понятие центральной тенденции в свете полученных результатов;
4. Предложить практические рекомендации по выбору критерия.

МЕТОДИКА

Техническим инструментом для нашего исследования служит программа, написанная в системе *MatLab*, которая позволяет порождать пары выборок, имеющих произвольный объем и распределения. Таким образом, мы сможем обрабатывать наиболее яркие примеры методом Монте-Карло, т.е., в данном случае, генерируя достаточное количество выборок с заданными параметрами и проверяя, насколько частоты попадания результатов в критическую область обсуждаемого критерия близки к теоретически ожидаемым частотам.

[1] Некоторые авторы рекомендуют прямую проверку соответствия эмпирического распределения нормальному закону [1, 9], другие указывают, что наибольшую опасность представляет асимметрия [14], а также асимметрия и эксцесс вместе [5, 7, 9, 10].

Мы проверяем частоты для наиболее употребительных в психологических исследованиях уровней значимости – 0.05 и 0.01. Нарушения условий применимости критериев для меньших значений могут влиять существенно иным образом[2]. Для психологов же наибольший интерес представляют именно эти значения, поскольку на этом уровне и происходит принятие решений об успешности/неуспешности исследования.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

1. Логическая структура проверок нормальности

Рассмотрим пример: биномиально распределенная случайная величина с вероятностью успеха p = 0.5 и количеством испытаний, равным 5 (эксцесс распределения приблизительно равен –0.4).

На рис. 1 представлено теоретическое распределение данной случайной величины вместе с кривой нормального распределения, имеющей такие же математическое ожидание и дисперсию. Выборочные гистограммы при больших размерах выборки будут практически повторять эту диаграмму распределения. Однако тот факт, что случайная величина может принимать только целые значения от нуля до пяти, в конце концов приведет к тому, что разность между непрерывной теоретической функцией нормального распределения и ступенчатой функцией накопленных частот, ступеньки которой не становятся меньше при увеличении выборки, станет для критерия Колмогорова–Смирнова все более "заметной".

Мы смоделировали выборки из испытаний данной случайной величины с помощью стандартной функции пакета *SPSS*. При размере выборки, равном 12, проверка нормальности критерием Колмогорова–Смирнова (*SPSS*) дает уровень значимости 0.163, что обычно считается свидетельством высокого сходства тестируемого распределения с нормальным. При увеличении размера выборки до 15 указанный уровень значимости становится равным 0.1. Дальнейшее увеличение приводит к быстрому уменьшению уровня значимости: при n = 25 p = 0.011, а при n = 50 p становится меньше 0.001, и гипотеза нормальности должна быть отвергнута.

[2] Исследование в более узкой области, но с большей точностью и для всех уровней значимости проведено, например, в работе [6]. Там же анализируется мощность критериев, которая в нашей статье не рассматривается.

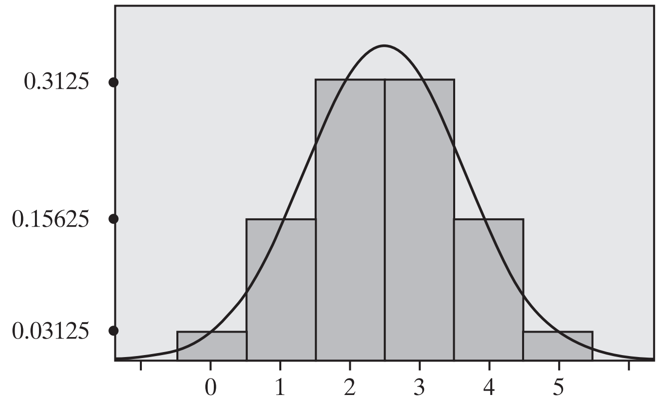

Рис. 1. Диаграмма распределения биномиальной случайной величины (5 испытаний, вероятность успеха равна 0.5).

Вторая половина задачи – проверка возможности использования t-критерия и критерия Манна–Уитни для оценки гипотез о равенстве средних значений двух выборок из испытаний данной биномиальной случайной величины. Мы решали ее методом Монте-Карло по следующей схеме:

1) Генерируются две (псевдо)случайные выборки, проводится их сравнение по критерию Манна–Уитни и t-критерию и определяется уровень значимости различия выборок; результат запоминается;

2) Эксперимент повторяется 1000 раз. Вычисляется частота попадания статистики Манна–Уитни и t-статистики в зону выше соответствующего верхнего 0.05-квантиля (0.01-квантиля) и ниже соответствующего нижнего 0.05-квантиля (0.01-квантиля) и берется их среднее арифметическое[3]. Такая серия повторяется 10 раз (результаты в первых 10 строках табл. 1).

Затем вычисляется среднее арифметическое 10 результатов (среднее по столбцу) и стандартное отклонение (последние две строки табл. 1). Таким образом, итогом является среднее количество попаданий в критическую область – область отвержения нулевой гипотезы. Для уровня 0.05 ожидается, что количество попаданий составит 50 из тысячи, для 0.01–10 из тысячи. Мы приводим также для сравнения результаты аналогичной проверки критерия Манна–Уитни. Для малых выборок вычисление статистики проводилось нашей собственной программой, затем результат сравнивался с табличными квантилями, для выборок (n = 50) использовалась аппрокси-

[3] Усредняя частоты попадания в верхние и нижние критические области, мы несколько увеличивали точность при том же количестве экспериментов.

Таблица 1. Результаты сравнения случайных биномиальных выборок (n = 10) по двум критериям

Номер эксперимента	Манна–Уитни		t-критерий	
	$p < .05$	$p < .01$	$p < .05$	$p < .01$
1	38	9	56	13
2	40	4	50	7
3	40	8	63	16
4	32	6	55	12
5	36	7	48	11
6	32	5	47	8
7	33	6	49	9
8	37	4	49	7
9	40	9	61	13
10	34	3	52	6
Среднее	36.2	6.1	53	10.2
Ст. отклон.	3.29	2.13	5.58	3.29

мация системы *MatLab*[4]. Вычисляемое стандартное отклонение позволяет при необходимости оценить доверительный интервал. Если в испытаниях частота попадания в 5%-ю критическую область оказывается, например, 100 из 1000, то это означает, что вероятность ошибки первого рода для подобных распределений на самом деле равна примерно 0.1, и достоверность выводов сильно завышается.

В нашем случае проверка показывает, что уже сравнение двух выборок по 10 испытаний в каждой дает частоты ошибок первого рода, вполне согласующиеся с априорными оценками, сделанными по распределению Стьюдента. Увеличение размера выборок, разумеется, не ухудшает соответствие. В табл. 2 приводятся только средние значения 10 экспериментов по 1000 сравнений и в скобках соответствующее стандартное отклонение[5].

Таблица 2. Результаты сравнения случайных биномиальных выборок (n_1 = 20, n_2 = 50) по двум критериям

Размер выборки	Манна–Уитни		t-критерий	
	$p < .05$	$p < .01$	$p < .05$	$p < .01$
20	41.2 (3.39)	6.7 (2.39)	51.1 (5.36)	10.3 (3.80)
50	40.1 (2.64)	7.7 (1.56)	46.1 (5.27)	9.6 (2.76)

В итоге мы получаем логическую проблему: при увеличении объема выборки *применение t-критерия становится все более точным* [13, с. 380], *а проверка нормальности все более уверенно нам это применение запрещает.*

Этот парадокс присущ *логической схеме* проверки, а вовсе не конкретному примеру, связанному с нормальностью распределения. Парадоксальность ситуации может быть преодолена, например, разработкой таблицы, в которой для каждого размера выборки n указывался бы свой уровень значимости, на котором следует отвергать гипотезу нормальности распределения при данном размере выборки, причем при увеличении n уровень значимости должен был бы быстро стремиться к нулю. Это, однако, не имеет смысла делать по двум причинам: во-первых, у процедуры проверки есть и другие недостатки; во-вторых, как мы обсудим далее, предлагаемая альтернатива – критерий Манна–Уитни – также имеет условия применимости, которые проверить еще труднее.

В аналогичную ситуацию попадают и те, кто предлагает проверять специальные характеристики распределений – асимметрию и эксцесс. В одной из весьма популярных книг для психологов [10] в качестве рецепта проверки применимости t-критерия предлагается инструкция по проверке асимметрии и эксцесса из книги Н.А. Плохинского [8, с. 110]. Похоже, однако, что биолог Плохинский ставит иную цель этой проверки. Его интересует отражение в статистических показателях важных для биологической науки характеристик популяции[6]. В этом контексте выглядит вполне логично, что, как и в нашем предыдущем примере, любой сколь угодно незначительный эксцесс (асимметрия) данного распределения будет выявлен рано или поздно при увеличении выборки. Фрагмент текста Плохинского приведен на рис. 2.

[4] Стоит отметить здесь и далее систематическую переоценку вероятности ошибки первого рода тестом Манна–Уитни на малых выборках. Однако она не так велика, как может показаться. В таблицах распределения Манна–Уитни для двух одинаковых выборок (n = 10) приведен 0.05-квантиль 27 и 0.01-квантиль 19. Оценка по Монте-Карло показывает, что точное значение уровня значимости для граничного значения 27 составляет примерно 0.044 (а не 0.05), а для граничного значения 19 – примерно 0.0094 (а не 0.01). Если таблица содержит квантили фиксированного уровня значимости для дискретного распределения, то подобные неточности неизбежны. Иногда точные вероятности вписывают в клетки таблицы мелким шрифтом, но таких таблиц для распределения Манна–Уитни нам найти не удалось, поэтому мы и прибегли здесь к методу Монте-Карло. Если учесть сказанное, то тест Манна–Уитни демонстрирует для выборок с n = 10 менее значительное отклонение.

[5] Далее почти все таблицы будут содержать только средние и стандартные отклонения.

[6] Хотя эксцесс появляется в книге непосредственно после страниц, посвященных проверкам того типа, который нас интересует, но рассмотренные в связи с эксцессом примеры говорят о смене интереса.

Для выяснения достоверности того, что изучаемое распределение отличается от нормального именно в сторону асимметрии или эксцесса, применяют обычный в биометрии метод сравнения показателей с их ошибками репрезентативности.

Показатели асимметрии и эксцесса с их ошибками репрезентативности определяются по следующим формулам:

$$A = \frac{\Sigma D^3}{n\sigma^3}; \quad m_A = \sqrt{\frac{6}{n}}; \quad t_A = \frac{A}{m_A} \geqslant 3;$$

$$E = \frac{\Sigma D^4}{n\sigma^4} - 3; \quad m_E = 2\sqrt{\frac{6}{n}}; \quad t_E = \frac{A}{m_E} \geqslant 3,$$

где A – показатель асимметрии;
$\Sigma D^3 = \Sigma(W - M)^3$ – сумма кубов отклонений средин классов от средней арифметической (центральных отклонений);
σ^3 – среднее квадратическое отклонение, возведенное в третью степень;
E – показатель эксцесса;

Рис. 2. Фрагмент работы Н.А. Плохинского, к которому ведут ссылки по данной теме [8, с. 110].

Вслед за Н.А. Плохинским, автор руководства для психологов предлагает отвергать гипотезу нормальности распределения, а тем самым и возможность применения t-критерия, если отношение эксцесса E (асимметрии A) к соответствующей “ошибке репрезентативности” m_E (m_A) больше трех [10, с. 232]. Из приведенного фрагмента видно, что выборочная оценка эксцесса (асимметрии), которая при увеличении выборки сходится к реальному эксцессу (асимметрии) генеральной совокупности, умножается на константу, пропорциональную корню из n. Понятно, что даже небольшие отклонения эксцесса или асимметрии от нуля, будучи умножены на корень из n, при достаточно большом n превзойдут граничное значение, равное 3 единицам, в то время как влияние этого эксцесса или асимметрии на применимость t-критерия, говоря максимально осторожно, в пределе не растет с увеличением выборки, а скорее всего, убывает. Эта схема проверки страдает тем же самым недостатком, что и проверка соответствия по Колмогорову–Смирнову.

Второй заслуживающий внимания вопрос к логической структуре проверки нормальности состоит в следующем: если мы *очень* боимся ошибиться в уровне значимости нашего вывода о средних значениях двух выборок в сторону увеличения вероятности ошибки первого рода (то есть переоценки достоверности вывода), то при проверке нормальности (или иного интересующей нас в этом смысле показателя) нам следует брать уровень значимости сильно превышающий привычные 0.05 – например, 0.3. Упомянутый Н.А. Плохинский (один из немногих, если не единственный автор, который обращает внимание на это обстоятельство) пишет следующее: “Чем **выше** ответственность, тем при **меньшем** расхождении распределений различие уже считается достоверным, и, наоборот, чем **менее** ответственно исследование, тем при **большем** расхождении распределений различие между ними все еще может считаться недостоверным” [8, с. 106]. Действительно, если ответственность невысока, то будет ли реальная ошибка первого рода совпадать с полученной t-критерием – не представляется очень уж важным вопросом. Напротив, если ответственность высока, то, чтобы гарантировать ту самую надежность вывода, которую декларирует t-критерий, мы должны обезопасить себя, отвергая выборки, которые вызывают даже минимальное подозрение. Действуя по этой логике, в жизненно важных вопросах нам следует отвергать как не соответствующие нормальному распределению от 30 до 90% выборок, которые, возможно, реально являются испытаниями нормально распределенных величин (если мы устанавливаем, что будем отвергать нормальность по какому-то выбранному нами критерию на уровне значимости, например 0.3, то это и означает, что, получая нормально распределенные выборки, мы будем по нашему критерию отбрасывать 30% из них, т.е. и будем совершать ошибку первого рода с вероятностью 0.3).

Такого рода увеличение строгости проверки имело бы смысл, если бы после отвержения гипотезы нормальности мы имели безупречную альтернативу, но это не соответствует действительности, как мы покажем дальше.

2. Анализ часто упоминаемых препятствий к применению *t*-критерия и влияние этих препятствий на применимость критерия Манна–Уитни

2.1. Неоднородность дисперсии. Рассмотрим, прежде всего, вопрос о влиянии на изменения вероятностей ошибки первого рода неравенства дисперсий двух выборок. Этот вопрос обсуждался в классической книге Г. Шеффе “Дисперсионный анализ” [13], к которой, в основном, и ведут ссылки по этой теме [2, 3]. Далее мы будем писать “t-критерий”, имея в виду, что дисперсионный анализ (*ANOVA*) для двух групп дает эквивалентные t-критерию результаты и все сказанное может быть отнесено и к дисперсионному анализу.

Очевидно, что если дисперсии различны и выборка с меньшей дисперсией превосходит по объему выборку с большей дисперсией, то внутригрупповая дисперсия (в формуле t-статистики она стоит в знаменателе под корнем) будет недооценена, что приведет к переоценке t-статистики, т.е. как раз к тому “опасному” варианту, когда реальная вероятность ошибки первого рода для такой конфигурации выборок может оказаться заметно больше, чем дают соответствующие табличные значения. Если же выборка с меньшей дисперсией имеет меньший объем, то ситуация будет обратной и реально значимое различие может быть оценено t-критерием как незначимое. Эти данные приведены на с. 386 книги Шеффе в таблице, которую мы воспроизводим на рис. 3. Отметим, что в столбце, помеченном цифрой 5, допущена опечатка – вместо 0.750 следует читать 0.050.

Влияние неравенства дисперсий ошибок и неравенства объемов групп на истинную вероятность того, что 95% доверительный интервал для $\mu_1 - \mu_2$
не покрывает истинное значение при больших n
θ – отношение дисперсий, R – отношение объемов групп

R \ θ	0^*	$\frac{1}{5}$	$\frac{1}{2}$	1	2	5	∞^*
1	0.050	0.050	0.050	0.050	0.050	0.750	0.050
2	0.17	0.12	0.080	0.050	0.029	0.014	0.006
5	0.38	0.22	0.12	0.050	0.014	0.002	$1 \cdot 10^{-5}$
∞^*	1.00	0.38	0.17	0.050	0.006	$1 \cdot 10^{-5}$	0

* Недостижимые предельные случаи показывают границы изменения вероятности.

Рис. 3. Фрагмент работы Г. Шеффе, в котором приведены реальные уровни значимости для дисперсионного анализа при неравных дисперсиях выборок [13, с. 386].

Из таблицы видно, что если большая дисперсия наблюдается в выборке, содержащей большее число членов (правый нижний угол), то значимость различия более или менее сильно недооценивается, т.е. надежность отвержения нулевой гипотезы в таком случае даже выше, чем декларируемая t-критерием.

Вывод о том, что “*ANOVA* с постоянными эффектами удивительно нечувствителен к отклонениям от нормальности, а когда n равны, то и к влиянию неоднородности дисперсий” [2, с. 337] повторяют многие авторы литературы по математическим методам в психологии [3, с. 13].

Доступная в *SPSS* модификация t-критерия позволяет производить сравнение выборок с неравными дисперсиями [1]. В окне вывода результатов сравнения имеется таблица с двумя строками. Пользователю предлагается самому решить, какую строку таблицы взять. Выбор предлагается сделать на основании результатов теста *Levene*, проверяющего однородность дисперсии. Как и в разобранных выше примерах, здесь очень трудно подобрать разумный уровень отвержения гипотезы о равенстве выборочных дисперсий. На наш взгляд, разумнее всего вообще не предполагать равенства дисперсий, если только нет серьезных априорных соображений в пользу этого равенства. В случаях, когда дисперсии приблизительно равны, столь же похожи и результаты вариантов t-теста в обеих строках. В случае, когда дисперсии сильно отличаются, разумеется, надо использовать модифицированный вариант.

Непараметрическая альтернатива t-критерия. В книге М. Холлендера и Д. Вулфа “Непараметрические методы статистики” приводится следующая вероятностная модель для альтернативы H_1 при использовании критерия Манна–Уитни: распределение случайной величины, испытания которой дают первую выборку, отличается от распределения случайной величины, испытания которой дают вторую выборку, только сдвигом [12, с. 85]. Тюрин и Макаров добавляют к этому, что распределение должно быть непрерывным [11, с. 112][7]. Рис. 4 дает представление об этом условии. Поскольку в расчете статистики используется только отношение порядка, а порядковые шкалы допускают монотонные деформации, то условие можно ослабить: существует монотонное преобразование данной шкалы, в котором распределения отличаются только сдвигом. Можно доказать для непрерывных случайных величин, что если отношения плотностей первого распределения ко второму – монотонная функция, то такая монотонная деформация шкалы существует. Однако для распределений, изображенных на рис. 4, это условие не выполняется, хотя условие сдвига выполнено. Это значит, что условие монотонности отношения плотностей является достаточным, но не необходимым. Необходимым является условие, что предел отношения плотностей слева равен бес-

[7] Наши эксперименты показывают, что повторы значений в выборках, которые наблюдаются тем чаще, чем меньшее количество возможных значений имеет случайная величина, приводят к завышению вероятности ошибки первого рода и занижению достоверности вывода критерием Манна–Уитни.

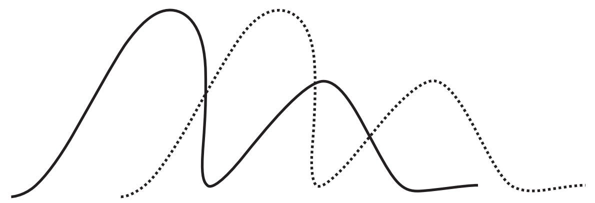

Рис. 4. Два распределения, отличающиеся только сдвигом.

конечности, а справа – нулю (с уточнениями, связанными с неопределенностью вида 0/0). Это условие для выборочных распределений проверить не представляется возможным. Однако легко увидеть, что для плотностей нормальных распределений с неравными дисперсиями оно явно нарушено.

К чему приводит это нарушение, мы проверили методом Монте-Карло, сопоставив результаты применения t-критерия и критерия Манна–Уитни в ситуациях неравенства выборочных дисперсий. Нам могут возразить, что понятие выборочной дисперсии нерелевантно порядковым критериям. Однако когда речь идет о последовательности операций “проверка нормальности – отвержение гипотезы нормальности – применение критерия Манна–Уитни”, мы имеем дело с интервальными шкалами, и понятие дисперсии здесь вполне уместно.

С помощью нашей программы в системе *MatLab* были смоделированы различные соотношения размеров выборок и их дисперсий. Например, пары выборок с $n = 10$ каждая, распределения которых описывались законами $N(0;1)$, $N(0;4)$ соответственно (нормальные распределения с нулевым средним и дисперсией единица для первой и четыре для второй выборки). Напомним схему эксперимента. Для каждой такой пары рассчитывались t-статистика и статистика Манна–Уитни, которые сравнивались с соответствующими критическими значениями на уровнях значимости 0.05 и 0.01. В рамках одной серии операция была повторена 1000 раз. В каждой строке табл. 3 приведены количества наблюдений из этой тысячи, когда значение соответствующей статистики превышало соответствующее критическое значение. Этот расчет был повторен 10 раз, каждая строка отражает одну из таких серий. Ожидаемые значения должны колебаться, как мы уже говорили, в районе 50 = 1000*0.05 и 10 = 1000*0.01 в соответствующих столбцах.

В табл. 4 собраны результаты экспериментов с этими и другими значениями объемов выборок и отношения дисперсий. Приводятся только средние значения десяти серий при соответствующих условиях (в скобках стандартное отклонение).

Мы видим, что чем больше различаются дисперсии, тем сильнее вероятность ошибки первого рода у статистики Манна–Уитни отличается от декларируемой вероятности 0.05 на больших выборках (причем с увеличением выборки ошибка только возрастает)[8], в то время как t-критерий показывает заметные отклонения только на малых выборках.

[8] С учетом замечания в сноске 4 переоценка надежности оказывается еще больше.

Таблица 3. Результаты испытаний пар нормально распределенных случайных величин с отношением дисперсий 1:4 и $n = 10$

Номер эксперимента	Критерий Манна–Уитни		t-критерий Стьюдента	
	Количество значений $p < .05$	Количество значений $p < .01$	Количество значений $p < .05$	Количество значений $p < .01$
1	54	15	60	15
2	41	9	39	7
3	48	11	49	9
4	53	11	51	8
5	46	7	40	6
6	53	13	57	14
7	55	15	63	12
8	46	11	56	12
9	56	15	59	18
10	48	10	53	11
Средние (ст. отклонение)	50 (4.90)	11.7 (2.75)	52.7 (8.12)	11.2 (3.79)

Таблица 4. Результаты сравнения пар выборок с разными соотношениями дисперсий и разными объемами

Условия		Критерий Манна–Уитни		t-критерий Стьюдента	
Размер выборки	Соотношение дисперсий	Количество значений $p < .05$	Количество значений $p < .01$	Количество значений $p < .05$	Количество значений $p < .01$
$n = 10$	1:1	45.35 (4.32)	8.00 (1.64)	48.00 (6.83)	8.80 (2.74)
	1:4	50.00 (4.90)	11.70 (2.75)	52.70(8.12)	11.20 (3.79)
	1:16	58.90 (7.77)	14.30 (2.51)	62.80 (9.47)	15.70 (3.46)
$n = 20$	1:1	46.95 (4.73)	9.50 (1.31)	48.20 (8.48)	10.20 (2.34)
	1:4	54.55 (3.83)	12.55 (3.09)	53.30 (7.93)	11.50 (4.03)
	1:16	65.20 (5.27)	16.95 (1.98)	56.40 (7.72)	12.40 (4.99)
$n = 50$	1:1	50.55 (4.09)	10.10 (1.98)	51.00 (7.39)	10.90 (3.28)
	1:4	55.90 (2.59)	12.75 (1.94)	50.90 (6.91)	13.30 (3.83)
	1:16	66.10 (3.58)	16.60 (3.71)	48.70 (6.63)	11.30 (3.80)

Таблица 5. Результаты сравнения пар выборок испытаний дискретной случайной величины с двумя равновероятными значениями

Размер выборки	Критерий Манна–Уитни		t-критерий Стьюдента	
	Количество значений $p < .05$	Количество значений $p < .01$	Количество значений $p < .05$	Количество значений $p < .01$
10	22.05 (2.32)	1.25 (0.63)	44.40 (4.72)	13.30 (2.11)
20	18.65 (3.11)	3.05 (1.50)	40.80 (7.16)	9.30 (3.13)
50	27.95 (3.13)	3.25 (1.88)	55.90 (6.28)	11.70 (4.22)

2.2. Эксцесс. Чаще всего в работах встречается утверждение о том, что отрицательный эксцесс свидетельствует о “плоской”, “тупой”, а в предельном случае бимодальной (“двухвершинной” или “двугорбой”) форме графика плотности распределения в районе математического ожидания [5, с. 31]. “Двугорбость” считается опасной для применения t-критерия. Минимальным (максимальным по модулю отрицательным) эксцессом, равным минус двум, обладает максимально “двугорбое” распределение: распределение дискретной случайной величины с двумя значениями, выпадающими с равной вероятностью.

Мы проверили для такой случайной величины, насколько реальные частоты ошибок первого рода при применении t-критерия (в сопоставлении с критерием Манна–Уитни) отличаются от теоретически предсказанных. Для сравнения выборок с n = 10, 20 и 50 результаты представлены в табл. 5.

Как видно, критерий Манна–Уитни показывает серьезную переоценку вероятности ошибки первого рода, т.е. недооценку значимости вывода, а t-критерий оценивает ее достаточно адекватно[9].

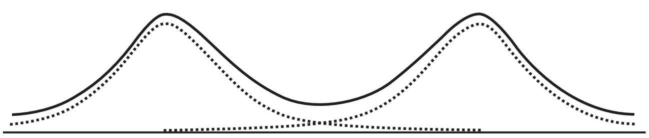

Рис. 5. Бимодальное распределение, “склеенное” из двух нормальных распределений (верхний огибающий график) с расстоянием между центрами, равным шести среднеквадратическим отклонениям.

Далее мы рассмотрели пример непрерывного распределения с близким к минимальному значением эксцесса, равным –1.62. Плотность этого распределения представляет собой среднее арифметическое плотностей двух стандартных нормальных распределений с расстоянием между центрами в 6 среднеквадратических отклонений (рис. 5, на рисунке для наглядности показана сумма плотностей, а не среднее арифметическое)[10]. Приведены средние значения по 10 экспериментам по 1000 пар выборок в каждом с разными объемами выборок. Результаты представлены в табл. 6.

[9] Разумеется, теоретически более адекватным было бы здесь применение таблиц сопряженности.

[10] Эксцесс таких распределений стремится к –2 при увеличении расстояния между центрами.

Таблица 6. Результаты сравнения пар выборок испытаний случайной величины с бимодальным распределением

Объем выборки	Манна–Уитни		t-критерий	
	$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	44.8 (6.12)	9.7 (1.97)	50.3 (6.81)	12.4 (2.59)
20	45.6 (3.28)	9.8 (1.54)	48.8 (4.52)	10.5 (3.54)
50	49.1 (5.46)	8.9 (2.69)	49.0 (8.41)	9.0 (3.71)

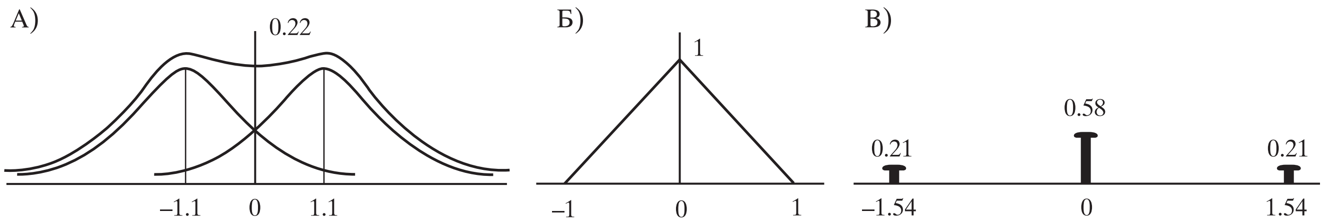

Рис. 6. Три распределения с одинаковым эксцессом, равным –0.6.

Небольшой отрицательный эксцесс не может быть интерпретирован в “визуальных” терминах типа двугорбости, плоскости вершины и т.п. Эксцесс, равный –0.6, присущ треугольному распределению, показанному на рис. 6Б. Для сравнения рядом представлено бимодальное распределение, имеющее такой же эксцесс, плотность которого представляет собой среднее арифметическое двух нормальных (как было описано выше) с расстоянием между центрами в 2.2 среднеквадратических отклонения (рис. 6А), и дискретное распределение с подобранными параметрами (рис. 6В), эксцесс которого также равен минус 0.6.

Тестирование по указанной схеме этих распределений дает результаты, приведенные в табл. 7, 8 и 9.

Как видно, отклонения частот от предсказываемых табличными значениями для t-критерия невелико, и значительно больше для критерия Манна–Уитни только в случае дискретного распределения.

Следует обратить внимание также на то, что во всех случаях оценка эксцесса оказывается заметно завышенной (для дискретного распределения на малых выборках она дает даже положительное значение) и только очень медленно сходится к теоретическому значению[11].

[11] Наша проверка показала, что при объеме выборок $n = 1000$ оценка эксцесса практически точна, вероятности ошибок первого рода (5%-е и 1%-е соответственно) оказываются в этом случае около 0.044 и 0.0085 для обоих проверяемых тестов.

Положительный эксцесс также считается препятствием для применения дисперсионного анализа. В работе Шеффе указано, что положительный эксцесс довольно существенно ухудшает выборочную оценку дисперсии, стоящую в знаменателе формулы дисперсионного анализа и t-статистики – хотя в среднем оценка точна, но ее дисперсия умножается на зависящий только от самого эксцесса коэффициент [13, с. 381]. Вследствие этого, как мы считаем, *возможно* искажение вероятностей ошибок первого рода. В работе [9, с. 66] авторы предлагают считать существенным значение эксцесса больше 0.5, впрочем, предлагая также и другой критерий, связанный со стандартной ошибкой эксцесса (см. также [7, с. 60]), которая незначительно отличается от ошибки репрезентативности, описанной в работах [8, 10].

Мы провели эксперимент с двумя распределениями, имеющими положительный эксцесс около 1.4 (рис. 7). Первое из них непрерывное (А), второе дискретное (Б). Результаты приведены в табл. 10 и 11.

Снова обращает на себя внимание заметное искажение табличной вероятности статистики Манна–Уитни на небольших дискретных выборках. Дело, по-видимому, в чрезвычайно малом наборе значений нашей случайной величины. С положительным эксцессом как таковым, возможно, связано уменьшение частоты попадания в 1%-ю критическую область, которую демонстрирует t-критерий в обоих случаях на малых выборках.

Таблица 7. Результаты сравнения пар выборок испытаний случайной величины с непрерывным бимодальным распределением

Объем выборки	Эксцесс (оценка)	Манна–Уитни ручной		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	–0.39	44.1 (5.13)	9.2 (1.36)	49.7 (4.45)	10.7 (2.41)
20	–0.5	47.2 (2.59)	8.4 (1.31)	49.2 (2.78)	8.9 (2.47)
50	–0.56	49.1 (4.94)	9.2 (2.34)	51.8 (10.09)	9.9 (3.21)

Таблица 8. Результаты сравнения пар выборок испытаний случайной величины с “треугольным” распределением

Объем выборок	Эксцесс (оценка)	Манна–Уитни ручной		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	–0.29	44.3 (4.27)	9.2 (2.02)	53.2 (7.00)	10.6 (3.50)
20	–0.4	47.9 (4.08)	9.9 (1.96)	51.1 (5.06)	11.8 (3.52)
50	–0.47	50.8 (4.25)	10.2 (2.98)	53.3 (6.98)	10.5 (5.62)

Таблица 9. Результаты сравнения пар выборок испытаний дискретной случайной величины с тремя значениями

Объем выборок	Эксцесс (оценка)	Манна–Уитни ручной		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	0.38	22.7 (1.81)	2.9 (1.50)	45.0 (6.00)	10.9 (4.48)
20	–0.17	28.6 (3.45)	4.0 (1.46)	49.6 (8.62)	10.5 (3.81)
50	–0.46	29.0 (3.39)	3.8 (0.85)	47.1 (5.43)	9.6 (2.37)

В целом наши опыты позволяют сделать следующие выводы: эксцесс не может служить единственной причиной серьезного искажения достоверности выводов, опирающихся на t-критерий (только на уровне 0.01 для малых выборок это искажение заметно). Для некоторых дискретных распределений критерий Манна–Уитни дает значительно большее искажение, к тому же более устойчивое к увеличению выборки[12]. Чтобы проверить гипотезу о том, что эти искажения дает не эксцесс, а малое число возможных значений случайной величины, мы провели эксперимент с трехзначным распределением, аналогичным приведенным на рис. 4В и 5В, но имеющим нулевой эксцесс. Результаты приведены в табл. 12.

Скорее всего, наша гипотеза верна, поскольку во всех трех испытаниях трехзначных распределений с весьма различными эксцессами получены похожие результаты.

Оценка эксцесса чрезвычайно неустойчива и смещена. Для малых выборок оценка всегда дает существенно большее значение, чем теоретическое. Вопрос о том, когда вообще можно применять выборочную оценку эксцесса, нуждается в дополнительном исследовании.

2.3. Асимметрия. По мнению некоторых авторов, асимметрия распределения также служит основанием для сомнений в применимости t-критерия по отношению к соответствующим выборкам [9, 10]. В учебнике [14, с. 132] к этому прибавляется,

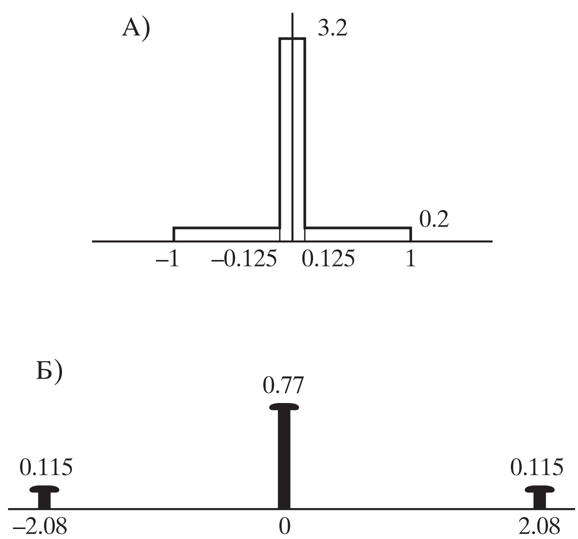

Рис. 7. Два распределения с эксцессом, равным 1.4.

[12] Возможно, искажение еще увеличивается, когда уровень значимости уменьшается.

Таблица 10. Результаты сравнения пар выборок испытаний непрерывной случайной величины с положительным эксцессом

Объем выборок	Эксцесс (оценка)	Манна–Уитни		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	2.20	44.5 (5.47)	8.8 (2.47)	49.0 (4.76)	6.5 (1.78)
20	2.12	49.1 (2.90)	9.9 (2.89)	52.5 (6.72)	9.1 (3.57)
50	1.70	51.4 (3.84)	10.3 (3.44)	53.6 (8.95)	8.7 (3.13)

Таблица 11. Результаты сравнения пар выборок испытаний дискретной случайной величины с положительным эксцессом

Объем выборок	Эксцесс (оценка)	Манна–Уитни		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	2.93	30.1 (4.12)	4.1 (1.02)	42.8 (5.22)	3.6 (1.58)
20	2.82	35.8 (3.32)	6.0 (1.89)	48.6 (6.83)	9.0 (1.41)
50	1.81	40.2 (5.38)	6.8 (1.92)	52.7 (7.56)	9.4 (3.41)

Таблица 12. Результаты сравнения пар выборок испытаний дискретной случайной величины с нулевым эксцессом

Объем выборок	Эксцесс (оценка)	Манна–Уитни		t-критерий	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	1.35	33.8 (2.62)	5.4 (1.14)	49.9 (6.69)	7.2 (2.97)
20	0.72	39.3 (4.59)	6.5 (1.39)	50.2 (5.75)	9.4 (2.72)
50	0.24	41.8 (3.96)	7.1 (1.12)	49.0 (4.55)	8.5 (2.01)

что особенно опасна при сравнении двух выборок асимметрия разнонаправленная – когда распределения выборок “скошены” в разные стороны.

Мы не будем приводить данные проверок сравнения двух выборок с большой, но одинаковой по знаку асимметрией, поскольку они не показали никаких заметных отклонений от табличных вероятностей для t-критерия, а для критерия Манна–Уитни однонаправленная асимметрия вообще не нарушает условий применимости (при равном разбросе выборок).

Что касается распределений с большой по модулю асимметрией, но имеющей при этом разные знаки, то этот случай оказывается чрезвычайно интересным и далеко выходящим за рамки обсуждения соответствия нормальному распределению.

Асимметрия случайной величины, распределение которой показано на рис. 8А, равна –1.95, а ее математическое ожидание равно нулю. В качестве “скошенного” в другую сторону мы взяли зеркальное отражение данного распределения относительно оси ординат – у него асимметрия равна +1.95 и также нулевое математическое ожидание (рис. 8Б).

Для t-критерия нулевая гипотеза утверждает равенство средних значений. Наша следующая задача состоит в том, чтобы оценить, насколько вероятности ошибок первого рода при условии истинности нулевой гипотезы (распределения имеют заданный вид и их средние значения совпадают) соответствуют вероятностям, предсказываемым квантилями распределения Стьюдента и Манна–Уитни. В табл. 13 приведены результаты. Поскольку симметрия нарушена, верхние и нижние хвосты рассматриваются отдельно.

Совсем не странно, что с большой вероятностью (и эта вероятность растет при росте объема выборок: 0.35, 0.62, 0.93) первая выборка, согласно критерию Манна–Уитни, лежит правее, чем вторая, поскольку вероятность выпадения значения больше нуля для первой случайной величины равна 0.8, а для второй 0.2. Ясно, что среднее значение не является адекватным понятием для порядкового критерия[13].

[13] Отдельно отметим чрезвычайно странное поведение оценок эксцесса: при росте объема выборок (10, 20, 50) средние

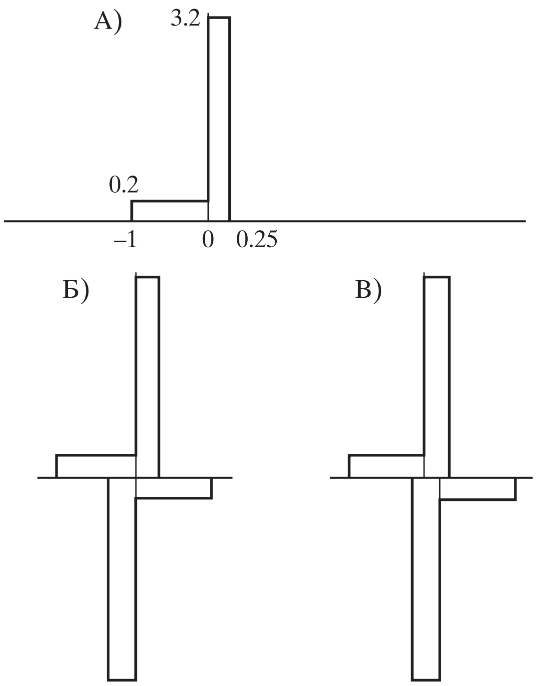

Рис. 8. Распределения с противоположными эксцессами. А) исходное; Б) выровненные по среднему значению распределения (для наглядности второе распределение отражено относительно оси абсцисс; В) выровненное по медиане распределение.

Объяснима и ошибка t-критерия. Здесь также наблюдается перекос, и в ту же сторону, что и в критерии Манна–Уитни. Объясняется это тем, что в случаях, когда в первую выборку попали положительные значения (в количестве выше ожидания), а во вторую – отрицательные, то в силу их большей "плотности" (чем у представителей "хвоста" распределений) оценка дисперсии будет заниженной, что приведет к завышению t-статистики[14]. Эта тенденция постепенно "сходит на нет" при росте объема выборок, но в интересующей нас области объемов дает вполне ощутимое искажение, что и означает, что t-критерий в данном случае не дает адекватной оценки достоверности.

Вернемся к критерию Манна–Уитни. Критики могут возразить, что для этого критерия "равные центральные тенденции" следовало бы интерпретировать как равные медианы. Это, однако, не совсем верно.

Если наши распределения выровнять по медиане[15], равной 3/32 для первого распределения (рис. 8В), то результаты оказываются следующими (табл. 14).

Мы видим, что нулевая гипотеза "распределения имеют равные медианы"[16] для выборки с n = 50 будет отвергнута с вдвое большей частотой, чем теоретические 0.05 в верхнем хвосте и вдвое меньшей в нижнем хвосте. Для того чтобы частоты попадания в верхний и нижний хвосты распределений Манна–Уитни были равны, нам придется сдвинуть распределения обратно в сторону средних. Это значит, что центральная тенденция, измеряемая статистикой Манна–Уитни, в данном случае сдвинута с медианы в сторону среднего, хотя и ненамного. Ниже мы приведем примеры более "рельефные" и обсудим вопрос более подробно. В заключение параграфа отметим, что на

оценки таковы: 2.74, 3.97, 3.81. Разброс внутри групп минимален. Расчетный эксцесс распределения при этом равен 2.8.

[14] Моделирование методом Монте-Карло показывает, что имеется очень высокая корреляция (больше 0.8) между средним значением выборки и оценкой ее дисперсии для нашего скошенного вправо распределения, причем при разных объемах выборки. Вопрос нуждается в дальнейшем исследовании.

[15] То есть сдвинуть первый график влево на 3/32, а второй на столько же вправо.

[16] С t-критерием в данном случае все понятно. Поскольку средние значения двух случайных величин отличаются примерно на 0.2, то при росте объема выборок это различие фиксируется с растущей достоверностью.

Таблица 13. Результаты сравнения пар выборок испытаний случайных величин с асимметриями разных знаков

Объем выборки	Модуль асим.	Манна–Уитни, нижн. границы		Манна–Уитни, верх. границы		t-критерий нижн. границы		t-критерий верх. границы	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$	$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	1.3	360.5	206.5	1.2	0.1	132.7	80.6	13.8	1.5
	(0.03)	(11.53)	(9.72)	(0.91)	(0.32)	(10.68)	(5.85)	(4.42)	(1.43)
20	1.81	631.2	396.0	0	0	110.8	50.5	20.0	1.5
	(0.02)	(16.49)	(15.75)	(0.00)	(0.00)	(9.18)	(6.57)	(4.16)	(1.43)
50	1.97	929.8	798.5	0	0	93.7	36.3	26.1	2.9
	(0.01)	(7.21)	(13.86)	(0.00)	(0.00)	(6.43)	(5.87)	(3.03)	(1.59)

Таблица 14. Результаты сравнения пар выборок испытаний случайных величин с асимметриями разных знаков

Объем выборки	Модуль асим.	Манна–Уитни, нижн. границы		Манна–Уитни, верх. границы		t-критерий нижн. границы		t-критерий верх. границы	
		$p < .05$	$p < .01$	$p < .05$	$p < .01$	$p < .05$	$p < .01$	$p < .05$	$p < .01$
10	1.3	37.9	7.8	67.1	13.3	10.9	4.1	299.7	47.2
	(0.03)	(6.04)	(2.14)	(10.13)	(4.34)	(2.68)	(2.81)	(13.77)	(7.54)
20	1.81	33.3	6.2	77.6	16.5	1.1	0.4	631.5	247.2
	(0.03)	(6.60)	(3.26)	(9.15)	(2.87)	(0.74)	(0.52)	(10.01)	(22.19)
50	1.97	24.6	5.1	103.4	27.1	0	0	966.1	839.5
	(0.02)	(3.66)	(1.91)	(5.32)	(4.01)	(0)	(0)	(5.66)	(7.89)

других распределениях при разнонаправленных асимметриях получаются сходные результаты, хотя и менее резко выраженные.

3. Центральная тенденция

Выше мы пришли к выводу, что центральная тенденция по Манна–Уитни не есть медиана. Продолжим аргументацию. Пусть даны две выборки:

a: {3,4,8};

b: {2,6,7};

Вариационный ряд при сравнении ***a*** и ***b*:** *baabba*. Расчет статистики Манна–Уитни дает *Uab* = 4, *Uba* = 5, что показывает некоторое превосходство “центральной тенденции” выборки ***a*.** Однако медианы выборок находятся в противоположном отношении: медиана ***a*** равна 4, медиана ***b*** равна 6.

Наш пример можно последовательно усиливать: для конфигурации *bbbaaaabbbbaaa* *Uab* = 16, *Uba* = 33, для конфигурации *bbbbbbbaaaaaaaabbbbbbbbaaaaaaa* *Uab* = 64, *Uba* = 161, и различие становится значимым, в то время как медианы по-прежнему оппонируют статистике Манна–Уитни. Причина аномалии ясна: нарушены условия применимости этой статистики – отличие выборок не исчерпывается сдвигом (хотя в последнем примере этого почти не видно, но именно этого “почти” и хватает медиане).

Подобно тому, как, согласно крылатому выражению, “интеллект – это то, что измеряется тестом интеллекта”, наш пример склоняет нас к мнению, что центральная тенденция, которую оценивает статистика Манна–Уитни, это не математическое ожидание и не медиана, а “центральная тенденция Манна–Уитни”. Аналогично своеобразные центральные тенденции будут оценивать и другие критерии (например *Q*-критерий Розенбаума). Общая смысловая направленность всех этих критериев – выявить превосходство по какому-то параметру одной выборки над другой. Е.В. Сидоренко разъясняет на примере, в каких случаях *Q*-критерий работает [10, с. 44]. Но пример этот демонстрирует как раз сдвиг распределений (да и выглядят они при этом, как нормальные), а в этом случае может работать любой из наших методов, поскольку распределения отличаются как раз сдвигом. В более “экзотических” же случаях вместе с утратой применимости одних методов, возможно, утрачивается и осмысленность других, которые продолжают считаться применимыми. Небольшое дополнение к первому примеру как раз и создает такую абсурдную ситуацию.

Если к выборкам ***a***: {3,4,8} и ***b***: {2,6,7} добавить выборку ***c*:** {1, 5, 9}, то попарные сравнения дают следующие результаты:

1. *baabba: Uab* = 4, *Uba* = 5, ***a*** превосходит ***b*,**
2. *caacac: Uca* = 4, *Uac* = 5, ***c*** превосходит ***a*,**
3. *cbcbbc: Ubc* = 4, *Ucb* = 5, ***b*** превосходит ***c*.**

Мы видим, что при попарных сравнениях нарушена транзитивность отношения превосходства на “центральных тенденциях по Манну–Уитни”[17]. Это показывает, что при более серьезном нарушении условия применимости критерия само понятие “центральная тенденция по Манна–Уитни” некорректно.

ВЫВОДЫ

1. Наши данные показывают, что логическая форма проверки нормальности выборочного распределения в качестве условия применимости *t*-критерия не соответствует целям, которые перед ней ставятся.

[17] Размножая экземпляры, можно добиться значимого циклического превосходства.

Во-первых, один и тот же уровень значимости, устанавливаемый в качестве критического для принятия/отвержения гипотезы о нормальности, при различных объемах выборок задает различную чувствительность к отклонениям от нормальности, причем реальная пригодность *t*-критерия растет при росте объема выборок, в то время как пригодность с точки зрения проверки все более уверенно отвергается.

Во-вторых, используемый в процедуре проверки уровень значимости должен устанавливаться с учетом цены ошибок, а также метода, используемого в качестве альтернативы *t*-критерию в случае отвержения гипотезы о нормальности, а учет этих обстоятельств, как правило, никогда не проводится.

2. Условия, при которых применимость критерия Манна–Уитни не вызывает сомнений, – это наличие порядковой шкалы, эквивалентной исходной, в которой выборочные распределения отличаются только сдвигом.

2.1. Условия применимости критерия Манна–Уитни не выполняются в случае, если выборки заданы в интервальной шкале и их теоретические распределения имеют существенно разные дисперсии. В этом случае реальная вероятность ошибки первого рода существенно отличается от табличной[18] в сторону увеличения. При этом достоверность выводов завышается, а *t*-критерий демонстрирует близкие к табличным вероятностям значения частот ошибок первого рода.

2.2. В случае отклонений непрерывных распределений[19] от нормального для выборок равного объема существенные отклонения частот ошибок первого рода для *t*-критерия не наблюдаются при различных положительных и отрицательных значениях эксцесса. Тем более, эксцесс распределения не может влиять на применимость критерия Манна–Уитни.

Некоторые дискретные распределения демонстрируют существенное снижение частот попадания в критическую область по сравнению с табличными для критерия Манна–Уитни.

Во всех наших примерах, связанных с эксцессом, зафиксировано отклонение частот лишь в сторону занижения. Это значит, что критика работ, в которых указанные методы применялись без проверок, не может утверждать, что выводы недостоверны. Напротив, выводы в этих случаях более достоверны, чем можно заключить по приводимым уровням значимости.

Заметим, что оценка эксцесса является смещенной и ее динамика при увеличении объема выборок совершенно не изучена, поэтому принимать какие-либо решения на основании оценок эксцесса, особенно сделанных по небольшим выборкам, представляется неразумным.

2.3. Асимметрия выборочных распределений играет серьезную роль для обоих критериев, но лишь в том случае, если значения асимметрии имеют у выборок противоположные знаки. Этот случай представляет особый интерес. При применении *t*-критерия, если в качестве нулевой принимается гипотеза о равенстве средних значений, происходит существенное, причем несимметричное, отклонение частот ошибок первого рода от табличных вероятностей. При применении критерия Манна–Уитни возникает проблема формулировки нулевой гипотезы, связанная с неопределенностью центральной тенденции по Манна–Уитни.

Если в качестве таковой принимается равенство средних значений, то критерий Манна–Уитни дает совершенно неприемлемые результаты. Несколько лучше обстоит дело, если в качестве нулевой принимается гипотеза о равенстве медиан, однако и в этом случае наблюдается существенное несимметричное отклонение частот ошибок первого рода от табличных вероятностей.

3. Специальное исследование показывает, что центральная тенденция, оцениваемая статистикой Манна–Уитни, не определяется[20] иначе, как в терминах самой статистики Манна–Уитни, если нарушены условия ее применимости. В случае если распределения отличаются лишь сдвигом, в качестве оценки центральной тенденции могут быть приняты и среднее значение, и медиана, их сдвиг будет обнаружен с равным успехом.

4. Практические выводы таковы:

– Если шкалы, в которых произведены измерения, принципиально не могут претендовать на

[18] Напоминаем, что все проверки в данной статье выполнены методом Монте-Карло для уровней значимости 0.05 и 0.01, на которых обычно проводится оценка психологических исследований. Возможно, и даже очень вероятно, влияние нарушения условий применимости критериев для других уровней значимости будут иными.

[19] Между непрерывными и дискретными распределениями невозможно провести четкую границу. Всякое дискретное распределение можно приблизить с любой точностью к непрерывным, поэтому наши выводы касаются непрерывных распределений интуитивно привычного вида: их график не должен состоять из различимых волн с крутыми склонами (производная невелика по модулю и не меняет резко знак с плюса на минус).

[20] Имеется в виду определение-дефиниция.

интервальность, то применение t-критерия в той же степени принципиально некорректно;

– Если измерения проводятся в шкалах, которые можно считать интервальными, то в одних ситуациях разумнее применять критерий Манна–Уитни, в других – t-критерий;

– Проверка нормальности распределений вряд ли может быть признана разумным основанием для принятия решения;

– Целесообразно предпочесть t-критерий, если дисперсии симметричных распределений различаются в несколько раз;

– Если распределения имеют противоположную асимметрию, необходимо понять, какого рода центральная тенденция нас интересует, поскольку ни среднее значение, ни медиана не являются вполне адекватными экспликациями центральной тенденции по Манна–Уитни в подобных сложных случаях. При противоположной асимметрии выборок вопрос о пригодности критериев оказывается не столь существенным, как содержательный вопрос о том, что именно мы хотим оценить;

– Во всех остальных случаях универсальная рекомендация такова: применить оба метода. В случае, если результаты различаются несущественно, принять решение на основании этих результатов (в статьи и отчеты при этом включаются оба). В случае, если различия велики и существенны, для принятия решения необходимо разбираться в особенностях распределения и принимать решение после содержательного анализа причин расхождения. Следует иметь в виду, что выборки из дискретных распределений с малым набором значений могут давать наибольшие отклонения от табличных вероятностей.

СПИСОК ЛИТЕРАТУРЫ

1. *Бююль А., Цефель П.* SPSS: искусство обработки информации. М., СПб., Киев: DiaSoft, 2002.
2. *Глас Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М.: Прогресс, 1976.
3. *Гусев А.Н.* Дисперсионный анализ в экспериментальной психологии. М.: Психология, 2000.
4. *Ермолаев О.Ю.* Математическая статистика для психологов. М.: МПСИ – Флинта, 2002.
5. *Кутейников А.Н.* Математические методы в психологии. СПб.: Речь, 2008.
6. *Лемешко Б.Ю., Лемешко С.Б.* Об устойчивости и мощности критериев проверки однородности средних // Измерительная техника. 2008. №9. С. 23–28.
7. *Наследов А.Д.* Математические методы психологического исследования. СПб: Питер, 2008.
8. *Плохинский Н.А.* Биометрия. М., Изд-во МГУ, 1970.
9. *Рубцова Н.Е., Леньков С.Л.* Статистические методы в психологии. М.: УМК "Психология", 2005.
10. *Сидоренко Е.В.* Методы математической обработки в психологии. СПб.: Речь, 2000.
11. *Тюрин Ю.Н., Макаров А.А.* Статистический анализ данных на компьютере. М.: Инфра-М, 1998.
12. *Холлендер М., Вулф Д.* Непараметрические методы статистики. М.: Финансы и статистика, 1983.
13. *Шеффе Г.* Дисперсионный анализ. М.: Наука, 1980.
14. *Howitt D., Cramer D.* Introduction to Statistic in Psychology. L.: Financial Times, 2008.

CONDITIONS FOR STUDENT T-TEST AND MANN–WHITNEY U-TEST APPLICATION

A. A. Korneev*, A. N. Krichevets**

** PhD, junior research assistant, department of psychology, M. V. Lomonosov Moscow State University, Moscow*
***PhD (physico-mathematical), Sc.D (philosophy), the same place*

Conditions for Student T-Test and Mann–Whitney U-Test – the most popular methods of shift in central tendencies evaluation – application are analyzed in the article. Effects of violation of the given conditions are estimated by Monte-Carlo method, i.e. by random numbers with corresponding distribution modeling, forming of samples by means of theirs testing and evaluation of frequency of type I errors when using these tests. Recommendations on appropriate test choosing are given.

Key words: statistical methods, conditions of applicability, Student T-Test, Mann–Whitney U-Test