

Устойчивые меры центральной тенденции: взвешивание как возможная альтернатива усечению данных при анализе времен ответов

Ю. С. Додонов*, Ю. А. Додонова

Московский городской психолого-педагогический университет, Москва, Россия

Поступила в редакцию 30 июня 2011 г. Дата публикации: 21 октября 2011 г.

Рекомендованная ссылка:

Додонов Ю.С., Додонова Ю.А. Устойчивые меры центральной тенденции: взвешивание как возможная альтернатива усечению данных при анализе времен ответов [Электронный ресурс] // Психологические исследования: электрон. науч. журн. 2011. N 5(19). URL: <http://psystudy.ru> (дата обращения: чч.мм.гггг).

Р Е З Ю М Е

Проблема устойчивости мер центральной тенденции рассматривается в данной статье в контексте исследований скорости переработки информации, где первичные данные для каждого испытуемого всегда представляют собой набор времен ответов, на основании которых должен быть рассчитан единственный индивидуальный показатель, характеризующий местоположение индивидуального распределения. Статья состоит из трех частей. В первой части предлагается обзор различных мер центральной тенденции. Подробно рассматриваются оценки центральной тенденции, основанные на усечении данных, поскольку именно мерам этой группы в современной литературе чаще всего отдается предпочтение как наиболее устойчивым к смещенности распределения и наличию в нем выбросов. Обсуждается другой возможный подход к получению устойчивых мер центральной тенденции, а именно взвешивание данных; предлагаются две авторские оценки центральной тенденции, основанные на взвешивании. Во второй части статьи на основании результатов, полученных в эмпирическом исследовании, и компьютерной симуляции данных демонстрируется, что выбор той или иной меры центральной тенденции в целях описания индивидуального времени ответа в скоростной задаче может существенным образом повлиять на вывод о значимости взаимосвязи данного скоростного показателя с внешней переменной. В третьей части статьи анализируемые меры центральной тенденции сравниваются на основании компьютерной симуляции данных, имитирующих времена ответа испытуемых в реальном эксперименте. В приложении приводятся алгоритмы для вычисления всех обсуждаемых мер центральной тенденции в программной среде R.

Ключевые слова: меры центральной тенденции, усеченное среднее, взвешенное среднее, скорость переработки информации, распределение времен ответов

В любом учебнике по статистическому анализу данных подчеркивается, что каждый метод применим тогда и только тогда, когда выполняются определенные требования – среди них наиболее частым является, например, условие нормальности распределения данных. Существенным является тот факт, что это требование относится не только к тем методам анализа эмпирических данных, которые используются при изучении взаимосвязи нескольких переменных – таким, как регрессионный, дисперсионный или факторный анализ.

Даже в случае работы с одной переменной и необходимости «простого» определения «среднего» значения для имеющегося ряда данных корректность количественной оценки местоположения распределения связана с его формой, а использование той или иной меры центральной тенденции требует выполнения определенных допущений о виде распределения.

При этом отметим, что задача анализа локализации

(местоположения) распределения данных не является редкой для психологических исследований и может возникать в различных контекстах. Так, в фокусе внимания может оказываться вопрос о равенстве средних значений какой-либо величины в экспериментальной и контрольной группах. Другой пример – анализ достоверности эффекта какого-либо воздействия (например, обучающего). Конечно, при решении каждой из этих задач проблему подсчета конкретного значения для оценки локализации распределения можно обойти – например, с помощью использования непараметрических критериев, нечувствительных к нарушению требования нормальности распределения исходных данных.

Однако в психологических исследованиях существуют и такие задачи, которые требуют именно вычисления мер центральной тенденции и их дальнейшего анализа. Наиболее остро, пожалуй, проблема оценивания местоположения распределения встает в исследованиях скорости переработки информации. Скоростные когнитивные задачи, используемые в таких исследованиях, включают

* Контакты.

E-mail address: ys.dodonov@gmail.com (Юрий Додонов)

множество однотипных предъявлений стимулов. Таким образом, для каждого испытуемого «сырые» данные представляют собой целый ряд зарегистрированных времен ответов. При этом для дальнейшего анализа для испытуемого должен быть рассчитан один показатель, репрезентирующий его «среднее» время ответа.

Данная задача, простая на первый взгляд, в действительности не является тривиальной. Во-первых, многие исследования показывают, что индивидуальное распределение времен ответов практически всегда является положительно смещенным [Heathcote et al., 1991; Hockley, 1984; Ratcliff, 1978, 1979, 1993; Ulrich, Miller, 1994; и др.]. Здесь и далее, следуя общепринятой терминологии, мы будем называть положительно смещенным такое распределение, для которого среднее значение больше медианы. Различные меры центральной тенденции для такого распределения не совпадают, как это происходит в случае симметричного распределения. А поскольку смещенность распределения не обязательно одинакова для всех испытуемых, то выбор той или иной меры центральной тенденции для расчета индивидуального показателя времени ответа в скоростной задаче оказывается не безразличным для дальнейшего анализа.

Во-вторых, данные времен ответов в реальном эмпирическом исследовании с той или иной, никогда не известной, вероятностью содержат выбросы. При прохождении скоростной задачи испытуемый может в некоторых случаях давать опережающие ответы, а может, напротив, задерживаться с ответом в силу различных причин, таких как естественные колебания внимания, случайные помехи в окружающей обстановке, и т.п. Как в случае опережающих реакций, так и в случае задержек выбросы будут идентифицируемы при анализе распределения только в том случае, если они действительно находятся далеко за пределами разумных значений. Проблема идентификации выбросов является предметом самостоятельных исследований [Barnett, Lewis, 1994; Lovie, 1986; и др.] и будет лишь косвенным образом затронута в данной работе. Отметим только, что в большинстве случаев исследователь не может быть уверен в том, является ли конкретное значение времени ответа выбросом или оно происходит из того же распределения, что и остальные значения. Таким образом, возможное наличие выбросов – это еще одна проблема, которая должна приниматься во внимание при подсчете меры центральной тенденции для индивидуальных данных времен ответов.

Наконец, анализ местоположения распределения времен ответов в большинстве случаев существенно затрудняется в силу небольшого количества предъявлений в типичной скоростной задаче. Безусловно, в рамках экспериментальной психологии, когда в задачи исследователя входит моделирование некоторого когнитивного процесса, число предъявлений может достигать нескольких сотен или даже тысяч, как, например, в исследованиях Р.Ратклиффа [Ratcliff, 1978, 1979; Ratcliff, McKoon, 2008; и др.]. Однако в этом случае тестирование каждого испытуемого должно осуществляться в рамках нескольких многочасовых сессий, что накладывает очевидные ограничения на возможный объем выборки исследования и делает этот подход практически

неприменимым в рамках психологии индивидуальных различий.

В большинстве современных исследований в рамках последнего подхода типичная скоростная задача включает небольшое количество предъявлений, варьирующееся обычно от 20 до 40. Таким образом, при анализе данных индивидуальных времен ответов исследователь имеет дело с небольшим количеством значений и не может с уверенностью говорить о том, из какого распределения они происходят. Более того, даже если предположить, что при повторном прохождении скоростной задачи тем же испытуемым времена ответов будут происходить из того же индивидуального распределения, конкретный набор полученных значений с неизбежностью будет другим. Таким образом, мера центральной тенденции, используемая для расчета индивидуального показателя скорости ответа, должна не просто быть устойчивой к смещенности данных и наличию выбросов, но и адекватным образом описывать лежащее в основе распределение в условиях колебания конкретных имеющихся значений.

Подчеркнем, что речь идет именно о двух различных источниках нестабильности меры центральной тенденции (обсуждение этой проблемы может быть найдено также в работе [Ratcliff, 1993]). Первый возможный источник нестабильности меры центральной тенденции – это появление выбросов в исходных данных и изменение их количества. Именно стабильность этого рода чаще всего анализируется в современной литературе, и именно для стабильности меры к появлению выбросов и изменению их количества чаще всего используется термин «устойчивость», который иногда прямо переводится на русский язык как «робастность». В качестве показателя, характеризующего робастность меры центральной тенденции, традиционно рассматривается так называемая пороговая точка (на русский язык термин *breakdown point* переводят также как «точка срыва», «точка излома» или «точка неустойчивости»), то есть оценка процента некорректных наблюдений (например, очень больших значений), при которых мера центральной тенденции не смещается существенным образом. Существуют и более сложные алгоритмы оценивания робастности, основанные, например, на анализе так называемой функции влияния.

Одновременно второй возможный источник нестабильности меры центральной тенденции гораздо реже оказывается в фокусе внимания исследователей. Речь идет о колебаниях в данных, в действительности происходящих из одного и того же распределения. Действительно, при анализе скорости переработки информации в конкретной задаче исследователь может ожидать (например, на основании теоретической модели), что индивидуальные времена ответов будут распределены определенным образом. Однако в реальности исследователь всегда имеет дело не собственно с распределением, а с достаточно небольшим (в силу ограниченности количества предъявлений стимулов) набором зарегистрированных значений. При том же исходном распределении набор конкретных значений мог оказаться иным, однако исследователь вправе рассчитывать, что используемая им мера центральной тенденции будет слабо чувствительной к подобного рода колебаниям, то есть

будет адекватным образом давать представление именно о местоположении лежащего в основе распределения данных. Всегда ли такого рода стабильность сочетается с робастностью, то есть устойчивостью к появлению выбросов? Правомерно ли анализировать робастность меры вне контекста ее стабильности? Этим вопросам будет уделено особое внимание в данной работе.

Данная статья состоит из трех частей. В первой части статьи предлагается обзор различных мер центральной тенденции. Во-первых, перечисляются классические меры, называемые также пифагорейскими средними. Во-вторых, подробно рассматриваются те оценки центральной тенденции, которые основаны на усечении данных, поскольку именно мерам этой группы в современной литературе чаще всего отдается предпочтение как наиболее устойчивым к смещенности распределения и наличию в нем выбросов. В-третьих, обсуждается другой возможный подход к получению устойчивых мер центральной тенденции, а именно взвешивание данных; предлагаются две авторские меры центральной тенденции, основанные на взвешивании. Во второй части статьи на основании результатов, полученных в эмпирическом исследовании, и компьютерной симуляции данных демонстрируется, что выбор той или иной меры центральной тенденции в целях описания индивидуального времени ответа в скоростной задаче может существенным образом повлиять на вывод о значимости взаимосвязи данного скоростного показателя с внешними переменными. Наконец, в третьей части статьи обсуждается проблема критериев предпочтительности той или иной меры центральной тенденции; анализируемые в данной работе меры сравниваются на основании компьютерной симуляции данных, имитирующих времена ответа испытуемых в реальном эксперименте. В приложении к данной статье приводятся алгоритмы, которые могут быть использованы для вычисления всех обсуждаемых мер центральной тенденции в программной среде R.

Меры центральной тенденции

Анализируемые меры центральной тенденции объединены в статье в три группы. Для каждой меры приводится краткое описание, основные формулы и максимально простой числовой пример.

Классические меры центральной тенденции

Три меры этой группы были предложены еще пифагорейцами, поэтому в англоязычной литературе для них иногда используется термин «классические пифагорейские средние». Во-первых, это арифметическое среднее (M):

$$M = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1).$$

Алгоритм расчета арифметического среднего прост, и в целом его подсчет не представляет сложностей даже для неискушенного пользователя. Почему же могут быть необходимы другие меры центральной тенденции? В учебниках по статистике ответ на этот вопрос чаще всего иллюстрируется наглядным

числовым примером. Предположим, имеется ряд данных:

2, 3, 4, 4, 5, 5, 6, 6, 80.

Среднее арифметическое в данном примере равняется 12,10 – однако действительно ли данное число адекватным образом репрезентирует совокупность имеющихся значений? Наглядно видно, что в данном случае среднее существенно смещено в сторону завышения в силу наличия в ряду исходных данных единственного выбивающегося значения. В этом случае вопрос об адекватности оценки центральной тенденции для приведенного ряда данных путем простого усреднения выглядит вполне уместным.

Вторая классическая мера центральной тенденции – это гармоническое среднее (hM), которое рассчитывается следующим образом:

$$hM = \frac{n}{1/x_1 + 1/x_2 + \dots + 1/x_n} = n \cdot \left[\sum_{i=1}^n \frac{1}{x_i} \right]^{-1} \quad (2).$$

Очевидно, что на получаемую величину гармонического среднего наибольшее влияние оказывают небольшие значения из ряда данных, тогда как эффект значений, находящихся в правом хвосте распределения, существенно меньший. Для приведенного выше примера гармоническое среднее равняется 4,45. В целом для массива неодинаковых значений гармоническое среднее всегда будет меньше арифметического среднего. Больше гармонического, но по-прежнему меньше арифметического среднего оказывается и геометрическое среднее (gM) – еще одна классическая мера, которая может быть рекомендована к применению в случае правого смещения распределения данных:

$$gM = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (3);$$

или,

$$gM = \exp \left[\frac{1}{n} \sum_{i=1}^n \log(x_i) \right] \quad (4).$$

Формула (3) позволяет наглядно понять содержательный смысл геометрического среднего, однако ее использование для расчетов в программных средах бывает затруднительным в случае больших массивов данных, по причине значительных вычислительных нагрузок. Поэтому использование формулы (4) является более предпочтительным. Для приведенного выше примера геометрическое среднее равняется 5,78.

Как видно из приведенного описания, при расчете классических мер центральной тенденции используются все полученные данные. Другой подход к анализу локализации распределения, который в современной литературе рекомендуется к использованию в условиях смещенности распределений и наличия в них выбросов, – это расчет среднего значения после предварительного удаления (или, как принято говорить, усечения) части данных.

Меры центральной тенденции, основанные на усечении данных

Данная группа объединяет меры центральной тенденции, вычисление которых предполагает предварительное удаление части данных. Логика этого подхода состоит в следующем: если данные с большой вероятностью включают значения, выбивающиеся из общего ряда (выбросы), удаление некоторого количества крайних значений и последующее усреднение оставшихся данных позволит получить более устойчивую оценку центральной тенденции, чем в случае работы с полным массивом данных.

К мерам этой группы может быть отнесена в первую очередь медиана (Md). Для упорядоченного ряда значений $x_1 \leq \dots \leq x_n$, $Md = x_{k+1}$, если $n = 2k + 1$, или $Md = (x_k + x_{k+1})/2$, если $n = 2k$. В приведенном выше числовом примере количество значений четное, поэтому медиана подсчитывается как среднее значение для двух центральных наблюдений и оказывается равной 5,00. В целом из определения медианы видно, что при ее расчете усечению подвергаются все значения, кроме одного или двух, и только оставшиеся одно или два значения влияют на величину этой меры центральной тенденции. Таким образом, медиана гораздо менее чувствительна к появлению выбросов в хвостах распределения данных, чем большинство других мер центральной тенденции. Одновременно эта мера в высокой степени чувствительна к величине тех конкретных значений, которые окажутся «в середине», что может быть проблематично, в частности, при анализе времен ответов.

В исследованиях скорости переработки информации широкое распространение получило использование среднего значения с предварительным удалением данных, лежащих за пределами двух стандартных отклонений. Алгоритм расчета этой меры предполагает вычисление среднего значения и стандартного отклонения для исходного, полного набора имеющихся данных, идентификацию и удаление тех значений, которые лежат за пределами двух стандартных отклонений от среднего, и последующий расчет среднего значения для оставшихся данных. Для приведенного выше числового примера эта мера центральной тенденции равняется 4,56. Хотя в эмпирических исследованиях чаще всего в качестве граничного значения используются именно два стандартных отклонения, в качестве порога удаления данных может быть определено любое количество стандартных отклонений. В любом случае очевидная особенность этой меры состоит в том, что данные, подлежащие удалению, определяются относительно среднего значения, рассчитанного по полным данным, – а эта оценка, как обсуждалось выше, может значительно смещаться в присутствии выбросов.

Проблема привязывания значений, подлежащих усечению, к арифметическому среднему частично преодолевается в двух других алгоритмах оценивания центральной тенденции, а именно при подсчете усеченного среднего и винсоризованного среднего [Wilcox, 2001, 2003; Keselman, 2005]. Для подсчета $k\%$ усеченного среднего в упорядоченном ряду данных удаляется $k\%$ наименьших значений и $k\%$ наибольших значений, после чего рассчитывается

среднее для оставшихся значений. Так, для приведенного выше примера ряда данных 2, 3, 4, 4, 5, 5, 6, 6, 6, 80 в расчете 20% усеченного среднего участвуют следующие значения:

4, 4, 5, 5, 6, 6.

В данном примере величина 20% усеченного среднего равняется 5,00. Процедура расчета винсоризованного среднего отличается от расчета усеченного среднего тем, что $k\%$ наибольших и $k\%$ наименьших значений не просто удаляется, а заменяется наименьшими и наибольшими значениями из оставшегося массива данных. Так, для рассматриваемого примера в расчете винсоризованного среднего будут участвовать следующие значения:

4, 4, 4, 4, 5, 5, 6, 6, 6, 6.

В данном примере значения усеченного и винсоризованного среднего совпадают, что, конечно, не является общим правилом. Кроме того, значения усеченного и винсоризованного среднего будут изменяться в зависимости от того, какой процент данных отсекается. В зависимости от целей усечения данных, исследования с использованием компьютерных симуляций рекомендуют использование 10–15% усеченного среднего [Keselman et al., 2004], 15% усеченного среднего [Othman et al., 2004], 20% усеченного среднего [Keselman et al., 2008] или 20–25% усеченного среднего [Rocke, Downs, Rocke, 1982] (более подробное обсуждение данной проблемы может быть найдено, например, в работе [Wilcox, 2005]). При этом очевидно, что определение процента усечения данных в каждом конкретном случае остается на усмотрение исследователя. Кроме того, объем удаляемых значений при таком алгоритме никак не зависит от реального распределения эмпирических данных и наличия в них выбросов, а усечение традиционно является симметричным (одинаковым для левого и правого хвостов распределения).

Альтернативный подход к усечению данных состоит в предварительном выявлении выбросов и их количества. В современной статистике существует несколько мер этой группы (примеры могут быть найдены в работе [Rousseeuw, Croux, 1993]). В данном исследовании будут анализироваться две меры – M -оценка (иногда называемая также одношаговой M -оценкой Хабера, по фамилии предложившего ее автора [Huber, 1981]) и модифицированная M -оценка.

Первый шаг при расчете этих двух мер – идентификация выбросов. Стандартный алгоритм расчета M -оценки предполагает вычисление для каждого значения в ряду данных абсолютной величины его отклонения от медианы. Таким образом, для набора данных x_1, x_2, \dots, x_n может быть подсчитано множество абсолютных отклонений $|x_1 - Md(x)|, |x_2 - Md(x)|, \dots, |x_n - Md(x)|$. Для данного множества абсолютных отклонений рассчитывается медиана (MAD), после чего полученное значение шкалируется путем деления на фиксированную величину:

$$MADN = \frac{MAD}{0,675} \quad (5).$$

В соответствии с описываемым алгоритмом расчета в качестве выбросов в исходном ряду данных в дальнейшем рассматриваются все значения, для которых выполняется условие:

$$\frac{|x_i - Md(x)|}{MADN} > K \quad (6),$$

где K – это некоторая постоянная величина, в стандартном алгоритме расчета одношаговой M -оценки принимаемая равной 1,28.

Таким образом, пусть a_1 – количество наблюдаемых значений, для которых выполняется:

$$\frac{x_i - Md(x)}{MADN} < -1,28,$$

a_2 – количество наблюдаемых значений, для которых выполняется:

$$\frac{x_i - Md(x)}{MADN} > 1,28,$$

тогда одношаговая M -оценка (OSE) вычисляется по формуле:

$$OSE = \frac{1,28 \cdot MADN \cdot (a_2 - a_1) + \sum_{i=a_1+1}^{n-a_2} x_i}{n - (a_2 + a_1)} \quad (7).$$

Модифицированная одношаговая M -оценка (MOSE) предполагает более простой алгоритм расчета, а именно простое удаление всех значений, идентифицированных в качестве выбросов, и последующий расчет среднего значения для оставшихся данных. В качестве «порогового» для удаления выбросов в данном алгоритме традиционно рассматривается условие:

$$\frac{|x_i - Md(x)|}{MADN} > 2,24.$$

При этом очевидно, что в качестве пороговой величины вместо традиционно используемого значения 2,24 может быть использовано другое значение, соответствующее более строгому или более мягкому условию для идентификации выбросов.

Для приведенного выше числового примера значение одношаговой M -оценки составляет 4,87, значение модифицированной одношаговой M -оценки – 4,56. Нетрудно видеть, что последнее значение в данном примере совпадает с величиной среднего, подсчитанного после удаления выбросов, лежащих за пределами двух стандартных отклонений, поскольку оба алгоритма идентифицировали одно и то же единственное значение в ряду данных в качестве выброса. Очевидно при этом, что это совпадение не является общим правилом.

Взвешивание данных как возможный подход к получению устойчивых мер центральной тенденции

Выше рассматривалось два принципиально различных подхода к анализу локализации распределения данных. В рамках первого подхода при расчете меры центральной тенденции учитываются все значения исходного ряда данных, второй подход предполагает предварительное удаление определенного количества выбросов. Оба подхода, наряду с несомненными плюсами, имеют и очевидные минусы. Алгоритмы, предполагающие работу с полным массивом данных, несомненно, более просты, однако получаемые оценки локализации распределения могут оказаться смещенными в том случае, если часть данных

лежит далеко от основного массива. Алгоритмы, основанные на усечении данных, позволяют получить оценки, более устойчивые к наличию выбросов, однако вопрос о количестве удаляемых данных неизбежно остается на усмотрение исследователя. Кроме того, в основе такого подхода лежит предположение о том, что заведомо большие или маленькие значения являются исключительно случайными и должны быть удалены для наиболее адекватного описания локализации распределения данных.

Однако возможен и еще один подход к расчету мер центральной тенденции, который не требует столь жесткой классификации части данных как заведомо случайных выбросов. Действительно, вместо простого удаления некоторых значений в хвостах распределения такие значения могут быть учтены в анализе, при этом им может придаваться меньший вес, чем тем значениям, которые лежат в рамках основного массива данных.

На наш взгляд, является удивительным тот факт, что идея взвешивания значений относительно всего массива данных до сих пор не нашла применения при анализе локализации распределений, в частности при работе с данными времен ответов. Действительно, практически в любом учебнике по статистике приводится общая формула взвешенного среднего:

$$WM = \frac{\sum_{i=1}^n (w_i \cdot x_i)}{\sum_{i=1}^n w_i} \quad (8).$$

Однако в анализе эмпирических данных взвешивание до сих пор используется крайне редко. Возможная проблема состоит в том, как определять весовые коэффициенты для имеющих значения. Действительно, в приведенную выше общую формулу могут быть подставлены любые весовые коэффициенты w , однако выработка алгоритма расчета этих коэффициентов сама по себе представляет отдельную исследовательскую задачу.

В данной работе будут описаны разработанные нами возможные алгоритмы взвешивания данных. В двух представленных алгоритмах расчет весовых коэффициентов не требует среднего значения или иных параметров распределения в качестве входной информации, при этом взвешивание каждого имеющегося значения осуществляется относительно всего массива данных.

Первый, наиболее простой, алгоритм взвешивания заключается в расчете оценки, взвешенной по расстояниям (DWE). Пусть имеется набор данных x_1, x_2, \dots, x_n . Весовой коэффициент для x_i рассчитывается как значение, обратное среднему расстоянию x_i от остальных значений множества данных:

$$w_i = (n - 1) \cdot \left(\sum_{j=1}^n \text{abs}(x_i - x_j) \right)^{-1} \quad (9),$$

после чего полученные весовые коэффициенты подставляются в приведенную выше общую формулу (8). Заметим, что расчет весовых коэффициентов возможно проводить и без операции усреднения, но в

некоторых случаях (при большой сумме расстояний) это может привести к занижению точности вычислений.

Для наглядности рассмотрим простой числовой пример, где ряд исходных данных состоит из четырех значений. Имеем набор данных $x_1 = 2, x_2 = 3, x_3 = 5, x_4 = 12$ ($n = 4$). Весовые коэффициенты для x_i рассчитываются как

$$w_1 = 3/(|2 - 3| + |2 - 5| + |2 - 12|) = 3/14,$$

$$w_2 = 3/(|3 - 2| + |3 - 5| + |3 - 12|) = 3/12,$$

$$w_3 = 3/(|5 - 2| + |5 - 3| + |5 - 12|) = 3/12,$$

$$w_4 = 3/(|12 - 2| + |12 - 3| + |12 - 5|) = 3/26.$$

В соответствии с приведенными выше формулами:

$$DWE = \frac{w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4}{w_1 + w_2 + w_3 + w_4} \approx 4,59.$$

Легко убедиться, что для числового примера, рассмотренного в предыдущих разделах данной статьи, величина оценки, взвешенной по расстояниям, составляет 5,69.

Еще один возможный алгоритм взвешивания данных – это вычисление оценки, взвешенной на скалярное произведение единичных векторов (SPWE, [Dodonov, 2010]). Пусть имеется набор данных x_1, x_2, \dots, x_n (исходный массив X). Для расчета SPWE исходный массив преобразуется во вторичный массив Y путем попарного усреднения всех имеющихся значений:

$$y_k = \frac{x_i + x_j}{2} \Big|_{i < j} \quad (10).$$

Для расчета весовых коэффициентов для массива Y исходный массив X также представляется в виде множества единичных векторов l с углами поворота относительно оси абсцисс от 0 до $\pi/2$:

$$\theta(\hat{l}_i) = \frac{\pi}{2} \cdot \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (11).$$

Весовые коэффициенты массива Y представляют собой скалярные произведения векторов массива l:

$$w_k = \hat{l}_i \cdot \hat{l}_j = \cos(\lambda[\hat{l}_i, \hat{l}_j]) = \cos\left[\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right] \quad (12).$$

Таким образом, общая формула SPWE:

$$SPWE = \frac{\sum_{i=1, j=1, i < j}^n \left[(x_i + x_j) \cdot \cos\left(\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right) \right]}{\sum_{i=1, j=1, i < j}^n \cos\left(\frac{\pi}{2} \cdot \frac{x_i - x_j}{x_{max} - x_{min}}\right)} \quad (13).$$

Для наглядности рассмотрим числовой пример. Как и в предыдущем случае, имеем набор данных: $x_1 = 2, x_2 = 3, x_3 = 5, x_4 = 12$ ($n = 4$). Вторичный массив данных Y, полученный путем усреднения каждой пары значений, представляет собой:

$$y_1(x_1, x_2) = 2,5, y_2(x_1, x_3) = 3,5, y_3(x_1, x_4) = 7;$$

$$y_4(x_2, x_3) = 4, y_5(x_2, x_4) = 7,5, y_6(x_3, x_4) = 8,5.$$

Одновременно представим исходные значения в виде единичных векторов с углами поворота

относительно оси абсцисс от 0 до $\pi/2$, как показано на рисунке 1.

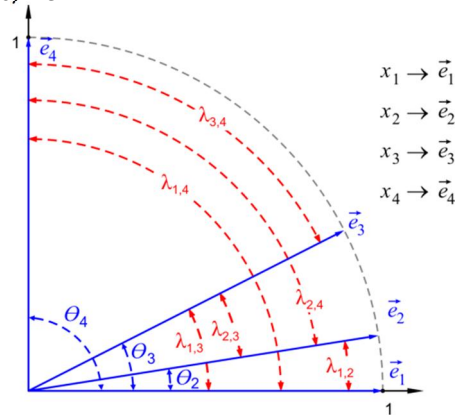


Рис. 1. Представление исходного ряда данных в виде единичных векторов.

При таком представлении данных в качестве меры расстояния между двумя исходными значениями может рассматриваться угол между соответствующими единичными векторами. Таким образом, весовой коэффициент для пары исходных значений (то есть для соответствующего элемента во вторичном массиве данных Y) рассчитывается как скалярное произведение соответствующих единичных векторов (то есть косинус угла между ними). В данном примере весовые коэффициенты w_k для y_k равны:

$$w_1[y_1] \approx 0,988; w_2[y_2] \approx 0,891; w_3[y_3] = 0;$$

$$w_4[y_4] \approx 0,951; w_5[y_5] \approx 0,156; w_6[y_6] \approx 0,454.$$

Видно, что наибольшие веса имеют y_k для пар наиболее близких исходных значений. В данном случае пара наиболее близких значений – это $x_1 = 2$ и $x_2 = 3$, поэтому y_1 , полученное при их усреднении, получает наибольший вес. Чем дальше отстоят друг от друга два значения в исходном ряду данных, тем меньший вес будет иметь соответствующее y_k . Подставляя w_k и y_k в общую формулу (8), получим для нашего примера ($x_1 = 2, x_2 = 3, x_3 = 5, x_4 = 12$) $SPWE \approx 4,193$.

Для рассмотренного выше ряда данных из десяти значений (2, 3, 4, 4, 5, 5, 6, 6, 6, 80) значение оценки, взвешенной по скалярному произведению единичных векторов, составляет 5,04.

Вычисления, необходимые для расчета последней из перечисленных мер центральной тенденции, выглядят объемными, однако необходимые алгоритмы легко имплементируются в программные среды, доступные любому исследователю – от стандартной версии Excel до программной среды R. В целом же при современной мощности вычислительной техники объем необходимых операций (в разумных, конечно, пределах) вряд ли может быть достаточным критерием для предпочтения того или иного алгоритма при анализе локализации распределения данных. Содержательные же преимущества и недостатки тех или иных мер центральной тенденции будут обсуждаться в следующих разделах.

Значимость проблемы выбора меры центральной тенденции: пример эмпирических данных времен ответа

Простейший числовой пример, использованный в предыдущем разделе, наглядно демонстрирует, что оценка локализации распределения данных может существенно различаться при использовании разных мер центральной тенденции. Применительно к исследованию скорости переработки информации в простейших когнитивных задачах это означает, что в зависимости от выбора той или иной меры центральной тенденции для каждого испытуемого могут быть получены разные величины, характеризующие местоположение его индивидуального распределения времен ответов. Может ли предпочтение конкретной меры центральной тенденции стать решающим в дальнейшем анализе взаимосвязи времени ответа в скоростной задаче с внешними переменными? Обсуждению этой проблемы на примере реальных эмпирических данных посвящен следующий раздел данной статьи.

Методы

В данном разделе будут рассматриваться результаты выполнения скоростной задачи различения стимулов, полученные в более широком контексте ранее проведенного исследования [Dodonova, Dodonov, in press]. Объем рассматриваемой в данной статье выборки составил 231 человек (58% женского пола), средний возраст 15,64 лет (стандартное отклонение 0,70).

Скоростная задача, которая будет рассматриваться в данной статье, требовала от испытуемых как можно быстрее ответить, является ли предъявляемая на экране монитора фигура треугольником. В рамках тестирования на белом фоне могла появляться одна из пяти фигур черного цвета (приблизительный размер стимулов составлял 35 × 35 мм). Испытуемые получали инструкцию как можно быстрее нажимать клавишу «1» в случае, если фигура являлась треугольником, и клавишу «0» для всех остальных фигур. Задача состояла из 20 предъявлений, количество ответов «Да» и «Нет» было уравниваем.

Времена неправильных ответов были исключены из дальнейшего анализа. После этого на основании зарегистрированных времен ответа для каждого испытуемого в отдельности рассчитывалось 11 возможных индивидуальных показателей скорости

различения стимулов – в соответствии с количеством мер центральной тенденции, описанных в предыдущей части данной статьи. Расчеты производились в программной среде R с использованием алгоритмов, приведенных в приложении.

В иллюстративных целях в качестве внешней «критериальной» переменной использовался искусственно сгенерированный ряд данных. Генерация данных выполнялась в программной среде R таким образом, чтобы корреляция между искусственно созданной внешней переменной и реальными показателями скорости различения стимулов была «пограничной» с точки зрения значимости. Очевидно, что именно в таких случаях выбор той или иной меры центральной тенденции для подсчета скоростного показателя может оказаться критичным для вывода о значимости наблюдаемой взаимосвязи с внешней переменной.

Результаты

Как и следовало ожидать, индивидуальные показатели скорости различения стимулов, подсчитанные с помощью различных мер центральной тенденции, были высоко связанными. Корреляции между индивидуальными показателями, подсчитанными разными способами, варьировали от 0,949 (между простым средним и модифицированной одношаговой M-оценкой) до 0,999 (между усеченным средним и оценкой, взвешенной по расстояниям).

Коэффициенты корреляции анализируемых индивидуальных показателей скорости различения стимулов с искусственно сгенерированной внешней переменной приведены в таблице 1.

Строго говоря, различия между коэффициентами корреляции, приведенными в таблице 1, статистически незначимы (критерий Фишера для сравнения выборочных коэффициентов корреляции, $p > 0,05$ в каждом случае). Однако для шести мер из одиннадцати взаимосвязь с внешней переменной может быть интерпретирована как статистически достоверная, тогда как при использовании остальных пяти мер центральной тенденции результаты говорят о статистической недостоверности связи скорости различения стимулов с рассматриваемой внешней переменной.

Обсуждение

Скоростная задача, описанная выше, в целом типична для исследований индивидуальных различий в скорости переработки информации и их взаимосвязи с

Таблица 1

Взаимосвязь показателей скорости различения стимулов, вычисленных с помощью различных мер центральной тенденции, и искусственно сгенерированной внешней переменной

	<i>M</i>	<i>hM</i>	<i>gM</i>	<i>Md</i>	<i>M(2SD)</i>	<i>TrimM</i>	<i>WinsM</i>	<i>OSE</i>	<i>MOSE</i>	<i>DWE</i>	<i>SPWE</i>
<i>r</i>	0,157	0,174	0,167	0,072	0,159	0,125	0,149	0,126	0,096	0,129	0,152
<i>p</i>	0,017	0,008	0,011	0,275	0,015	0,057	0,023	0,057	0,145	0,050	0,021

Примечания. *r* – коэффициент корреляции Пирсона, *p* – уровень значимости; *M* – арифметическое среднее; *hM* – гармоническое среднее; *gM* – геометрическое среднее; *Md* – медиана; *M(2SD)* – среднее после отсеечения данных, лежащих за пределами двух стандартных отклонений; *TrimM* – 20% усеченное среднее, с удалением 20% данных в каждом хвосте распределения; *WinsM* – винсоризованное среднее после предварительного удаления 20% данных в каждом хвосте распределения; *OSE* – одношаговая M-оценка; *MOSE* – модифицированная одношаговая M-оценка; *DWE* – оценка, взвешенная по расстоянию; *SPWE* – оценка, взвешенная по скалярному произведению единичных векторов.

другими когнитивными показателями. Задача включала относительно небольшое количество предъявлений, для каждого испытуемого имелся набор зарегистрированных времен ответов, на основании которого необходимо было получить один показатель, характеризующий скорость переработки информации.

В целом индивидуальные показатели скорости ответа, полученные с помощью различных мер центральной тенденции, оказались высоко связанными между собой. Однако приведенные данные демонстрируют, что даже при незначительных, на первый взгляд, различиях между этими показателями выбор той или иной меры центральной тенденции при анализе индивидуальных времен ответа может оказываться критическим при принятии решения о статистической достоверности взаимосвязи скорости переработки информации с внешней переменной. Так, в приведенном анализе для одних и тех же исходных данных времен ответов, в зависимости от меры центральной тенденции, использованной при подсчете скоростного показателя, мог быть сделан вывод как о высокой достоверности связи скорости ответа с внешней переменной (например, в случае использования гармонического среднего), так и об отсутствии взаимосвязи (например, в случае использования медианы).

Приведенные результаты очевидным образом демонстрируют, что выбору меры центральной тенденции должно уделяться особенно пристальное внимание, поскольку уже на этапе подсчета индивидуального показателя времени ответа предпочтение того или иного алгоритма может повлиять на результаты дальнейшего более сложного анализа. Конечно, приведенный пример иллюстрировал крайний случай, когда достоверность взаимосвязи неочевидна и вывод о ее наличии зависит от выбора конкретной меры центральной тенденции. Однако такого рода «пограничные» коэффициенты корреляции не являются редкими для эмпирических данных в психологии, в том числе для исследований скорости переработки информации.

Например, А.Дженсен приводит данные о среднем уровне корреляции простой скорости ответа с интеллектом порядка $-0,10$, времени различения стимулов с интеллектом – порядка $-0,20$ [Jensen, 1998, p. 211]. Л.Шеппард и Ф.Вернон на основании метаанализа данных из 195 отдельных исследований говорят о корреляции между скоростью реакции и, например, кристаллизованным интеллектом порядка $-0,17$ [Sheppard, Vernon, 2008]. Заслуживают ли внимания исследователя такого рода невысокие коэффициенты корреляции? Безусловно, обсуждение этого вопроса выходит далеко за рамки данной работы. Однако в случае скорости реакции и интеллекта взаимосвязь, хоть и слабая, действительно является достоверной и воспроизводимой.

Так, в упомянутом выше метааналитическом исследовании Л.Шеппард и Ф.Вернон анализируют значимость приведенного коэффициента корреляции между скоростью реакции и кристаллизованным интеллектом с использованием метода, предложенного Розенбергом [Rosenberg, 2005]. Авторы говорят о том, что для того, чтобы полученный средний коэффициент корреляции $-0,17$ оказался незначимым на уровне $0,05$, необходимо предположить существование еще 58 тысяч неопубликованных

исследований или исследований, сообщающих незначимую взаимосвязь между скоростью реакции или интеллектом. Альтернативно, необходимо получить незначимую взаимосвязь между скоростью реакции и интеллектом в одном исследовании с объемом выборки не менее 58 158 испытуемых. Оба варианта, конечно, крайне маловероятны, и в целом накопленные за последние десятилетия данные свидетельствуют в пользу достоверности этой слабой по величине взаимосвязи.

Наличие существенного объема непротиворечивых эмпирических данных, конечно, делает доводы в пользу существования той или иной взаимосвязи более весомыми. Однако является ли достоверной корреляция скорости ответа испытуемого и внешней переменной в отдельном исследовании – как, например, в приведенных выше смоделированных данных? Результаты, приведенные в таблице 1, позволяют говорить о том, что вывод о достоверности взаимосвязи в отдельном исследовании может меняться в зависимости от выбора той или иной меры центральной тенденции при подсчете индивидуальных времен ответа.

Однако приведенные результаты ничего не говорят о том, какая мера является наиболее предпочтительной и корректной при анализе данных. Для ответа на этот вопрос поведение различных мер центральной тенденции в условиях, имитирующих реальные эмпирические данные времен ответов, моделировалось в рамках серии компьютерных симуляций. Результаты сравнительного анализа приведены в следующем разделе.

Поведение мер центральной тенденции: компьютерная симуляция данных времен ответов

Данный раздел состоит из трех частей. В первой части подробно описываются методы исследования, а именно использованный алгоритм компьютерной симуляции данных времен ответов и наличия в них выбросов, а также те критерии сопоставления мер центральной тенденции, которые будут рассматриваться в дальнейшем анализе. Во второй части приводятся результаты, полученные в реализованной серии компьютерных симуляций. Обсуждение, представленное в третьей части, касается поведения тех или иных мер центральной тенденции в смоделированных условиях.

Методы

Для симуляции данных времен ответа в данной работе было выбрано экс-гауссовское распределение, которое является конволюцией нормального и экспоненциального распределений (одно из первых обсуждений данного распределения может быть найдено в работе [Luce, 1986]). Математически функции плотности вероятности для экс-гауссианы выглядит следующим образом:

$$f(x|\mu, \sigma, \tau) = \frac{1}{\tau\sqrt{2\pi}} \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{x-\mu}{\tau}\right) \cdot \int_{-\infty}^{\left\lfloor \frac{(x-\mu)}{\sigma} \right\rfloor - \left(\frac{\sigma}{\tau}\right)} \exp\left(-\frac{y^2}{2}\right) dy \quad (14).$$

В иллюстративных целях на рисунке 2 показаны исходные распределения – экспоненциальное и гауссовское – и полученное на их основе экс-гауссовское распределение.

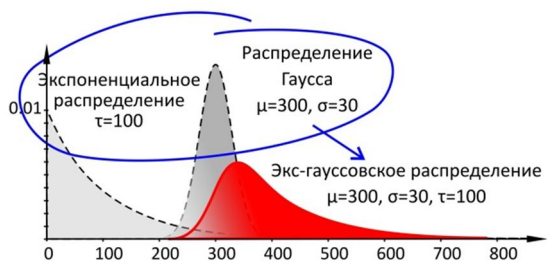


Рис. 2. Экспоненциальное и гауссовское распределения и полученное на их основе экс-гауссовское распределение.

Таким образом, два параметра экс-гауссовского распределения (14) – μ и σ – соответствуют среднему значению и стандартному отклонению гауссовского компонента. Третий параметр τ – это среднее значение экспоненциального компонента.

Использование экс-гауссовского распределения для описания данных времен ответов имеет давнюю историю в рамках экспериментальной психологии. Впервые содержательное рассмотрение параметров экс-гауссовского распределения в контексте описания когнитивных процессов, требующихся для решения скоростной задачи, было предложено Р.Хёле [Hohle, 1965]. Описывая получаемые времена ответа с помощью экс-гауссовского распределения, Р.Хёле предложил рассматривать экспоненциальный компонент как репрезентирующий собственно время когнитивных процессов, требующихся для принятия решения, а гауссовский компонент – как отражение времени сенсомоторных процессов, необходимых для выполнения скоростной задачи.

Оставляя за рамками данной статьи вопрос содержательной интерпретации параметров экс-гауссианы, отметим, что в многочисленных исследованиях была показана адекватность применения данного распределения для описания экспериментальных данных времен ответов в различных задачах [Heathcote et al., 1991; Hockley, 1982, 1984; Hohle, 1965; Ratcliff, 1978, 1979; Ratcliff, Murdock, 1976; Ulrich, Miller, 1994]. Именно поэтому для компьютерной симуляции времен ответов в рамках данной работы было выбрано экс-гауссовское распределение, параметры которого были подобраны таким образом, чтобы правдоподобно имитировать распределение времен ответа в простой скоростной задаче (например, времени реакции выбора или времени различения стимулов, в миллисекундах): $\mu = 400$, $\sigma = 20$, $\tau = 100$.

В программной среде R генерировалось 30 значений из экс-гауссовского распределения с указанными параметрами. Число генерируемых значений намеренно было небольшим, поскольку, как обсуждалось выше, типичная скоростная задача в реальном эмпирическом исследовании включает порядка 20–40 предъявлений, и именно таково количество значений, с которым приходится иметь дело при анализе данных времен ответов. Для генерированных значений подсчитывалась каждая из обсуждаемых в данной статье мер центральной

тенденции. Описанная процедура повторялась 50 000 раз.

Таким образом, в каждой из пятидесяти тысяч генераций данные происходили из одного и того же распределения с указанными параметрами, хотя конкретный набор получаемых значений, очевидно, был различным. Логично, что если мера центральной тенденции чувствительна к колебаниям в экспериментальных данных, значения, подсчитанные для этой меры в каждой из пятидесяти тысяч генераций, будут существенно варьировать, несмотря на то, что данные происходят из одного распределения. С другой стороны, если мера является стабильной и может адекватным образом репрезентировать локализацию лежащего в основе распределения, ее колебания в отдельных генерациях будут незначительными. Иными словами, в качестве показателя устойчивости меры центральной тенденции может рассматриваться величина стандартного отклонения этой меры для ряда генераций.

Выше была описана процедура компьютерной симуляции распределения времен ответов в «идеальных» условиях отсутствия выбросов. Однако в реальных эмпирических условиях, как правило, присутствуют выбросы, то есть такие значения, которые не ложатся в общее распределение данных. При этом такого рода «случайные» значения совершенно не обязательно появляются в хвостах распределения времен ответов. Случайные задержки ответа могут быть незначительными, а превосходящие ответы не очевидно неправдоподобными – иными словами, «шум» может появляться в произвольной части распределения. Чтобы смоделировать наличие выбросов в данных, полученных с помощью компьютерной симуляции, могут быть использованы различные подходы.

Во-первых, некоторая величина может прибавляться к выбранным случайным образом значениям из основного распределения. Например, в одном из вариантов компьютерных генераций Р.Ратклифф [Ratcliff, 1993] генерирует времена реакции из экс-гауссовского распределения, после чего к некоторым значениям прибавляет случайным образом значения от 0 до 2000 мс, сгенерированные из равномерного распределения. Очевидно, однако, что при описанном подходе симулируются именно случайные задержки ответа, тогда как реальные данные могут включать и превосходящие ответы. Чтобы смоделировать данные с такого рода выбросами, может быть использовано не прибавление, а именно замена отдельных значений из основного распределения на сгенерированные некоторым образом выбросы. Например, в той же работе [Ratcliff, 1993] Р.Ратклифф демонстрирует эффект наличия выбросов только слева и только справа следующим образом: 80% данных генерируется из основного экс-гауссовского распределения, после чего оставшиеся 20% генерируются из экс-гауссовского распределения, параметр μ которого сильно уменьшен (увеличен) по сравнению с основным распределением.

Безусловно, предпочтение того или иного способа симуляции выбросов не может быть однозначным и во многом зависит от целей конкретного исследования. Поскольку в данной работе в качестве выбросов интерес представляли любые данные, вносящие «шум» в основное распределение, выбросы генерировались

случайным образом из равномерного распределения от 0 до 2000 мс, как показано на рисунке 3.

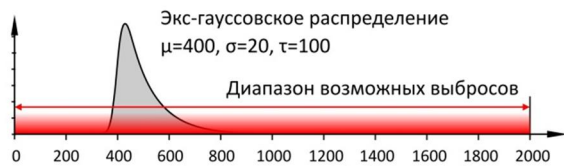


Рис. 3. Распределение, из которого генерировались значения, и диапазон возможных выбросов.

Таким образом, в дополнение к описанному ранее условию отсутствия выбросов симулировались еще три условия: данные включают один выброс, два выброса или три выброса. Для условий с наличием выбросов, как и для описанного выше условия с отсутствием выбросов, общее количество значений в каждой генерации составляло 30, меры центральной тенденции подсчитывались с использованием описанных выше алгоритмов. Для каждого условия генерация повторялась 50 000 раз. По пятидесяти тысячам генераций рассчитывалось среднее значение для каждой меры центральной тенденции и ее стандартное отклонение. Алгоритм генерации и последующих расчетов (для условия наличия двух выбросов) приведен в приложении.

Расчет средних значений каждой из мер центральной тенденции в отдельности в четырех моделируемых условиях позволил получить еще одну характеристику мер центральной тенденции, а именно оценить их устойчивость к появлению выбросов (то есть стабильность при переходе от одного моделируемого условия к другому). В качестве такого показателя рассчитывалось стандартное отклонение для четырех средних значений меры, полученных в четырех моделируемых условиях. Заметим, что только этот показатель, по существу, позволяет получить представление об устойчивости меры центральной тенденции в том смысле, в каком этот термин в большинстве случаев употребляется при обсуждении так называемых робастных, то есть нечувствительных к наличию выбросов, статистических методов. Как отмечалось выше, в контексте данного исследования этот вид устойчивости будет лишь одним из анализируемых показателей.

Наконец, выполнялась еще одна серия из пятидесяти тысяч генераций, в которой количество выбросов не являлось постоянной величиной. В каждой генерации полученные 30 значений могли содержать произвольное количество выбросов от 0 до 3. Как и в предыдущих условиях, в каждой генерации рассчитывались все меры центральной тенденции, после чего для каждой меры были получены среднее значение и стандартное отклонение по пятидесяти тысячам генераций. Из логики описанного алгоритма понятно, что стандартное отклонение меры центральной тенденции в данном случае будет отражать оба источника нестабильности меры, обсуждавшихся выше, – и неустойчивость к колебаниям в наборе данных, и отсутствие робастности к появлению выбросов.

Результаты

В таблице 2 для каждого из моделируемых условий приведены среднее значение и стандартное

отклонение для каждой меры центральной тенденции по пятидесяти тысячам генераций. С целью упрощения восприятия таблицы средние значения округлялись до целых чисел. Для каждого условия жирный курсив использован, чтобы выделить три меры с наибольшей стабильностью по пятидесяти тысячам генераций.

В столбце, озаглавленном SD*, для каждой меры приведено стандартное отклонение, рассчитанное по приведенным четырем средним значениям, соответствующим четырем условиям с фиксированным количеством выбросов. Жирный курсив использован, чтобы выделить три меры, наиболее устойчивые к появлению выбросов в исходных данных.

Наконец, в двух последних столбцах приведены среднее значение и стандартное отклонение мер центральной тенденции по пятидесяти тысячам генераций, в которых количество выбросов в распределении данных произвольно варьировалось от 0 до 10%. Как и в предыдущих случаях, жирный курсив использован для обозначения наиболее стабильных мер.

Обсуждение

Как и ожидалось, среднее арифметическое оказалось менее стабильной мерой центральной тенденции, чем большинство других мер, как в случае отсутствия выбросов, так и в случае их наличия. С увеличением количества выбросов поведение этой меры центральной тенденции ухудшается. Логичным образом, для медианы наблюдалась обратная картина. В случае наличия большого количества выбросов (в данном случае – 10%) медиана оказывалась одной из наиболее стабильных мер. Однако в случае отсутствия выбросов или их небольшого и изменяющегося количества медиана была менее устойчива, чем многие другие меры центральной тенденции.

В целом в случае отсутствия выбросов наиболее стабильными мерами центральной тенденции для анализируемого смещенного распределения оказываются гармоническое среднее, геометрическое среднее и предложенная в данной работе оценка, взвешенная по расстояниям. Однако существенная особенность гармонического среднего и геометрического среднего состоит в том, что поведение этих мер резко ухудшается в условиях наличия в распределении выбросов. В случае, если имеющееся распределение включает выбросы, гармоническое среднее, геометрическое среднее и арифметическое среднее формируют тройку наименее стабильных мер, чувствительных к колебаниям в исходных данных. При этом взвешенная по расстоянию оценка продолжает оставаться одной из трех наиболее стабильных мер в каждом из условий с наличием выбросов в распределении данных.

Арифметическое среднее, подсчитанное после удаления данных, лежащих за пределами двух стандартных отклонений, в целом оказывается более стабильной мерой, чем простое арифметическое среднее. Однако ни в одном из моделируемых условий эту меру нельзя было считать предпочтительной. Этот результат в целом можно назвать предсказуемым в силу специфики алгоритма подсчета данной меры центральной тенденции. Действительно, речь идет о предварительном удалении данных, лежащих за пределами двух стандартных отклонений от

арифметического среднего, рассчитанного для исходного распределения. А поскольку, как уже подчеркивалось выше, само это исходное среднее значение может быть существенно смещенным, такая процедура усечения данных не может быть названа эффективной. Следует отметить, однако, что этот алгоритм удаления выбросов по-прежнему остается одним из наиболее популярных в исследованиях скорости переработки информации. В этом контексте демонстрируемые результаты могут служить еще одним напоминанием для исследователей о том, что в арсенале современных статистических методов существуют доступные и гораздо более эффективные способы анализа местоположения распределения в условиях наличия выбросов, чем привычная процедура удаления данных, лежащих за пределами двух стандартных отклонений.

В целом полученные результаты подтверждают, что 20% усеченное среднее действительно может рассматриваться в качестве одной из наиболее предпочтительных мер центральной тенденции. Эта мера оказалась слабо подверженной колебаниям во всех случаях, когда распределение данных включало выбросы. Кроме этого, усеченное среднее оказалось одной из мер, наиболее устойчивых к появлению выбросов и изменению их количества.

Винсоризованное среднее также было одной из наиболее предпочтительных мер в двух моделируемых условиях из четырех, а также в условии переменного количества выбросов. Однако следует отметить, что в основе подсчета этой меры лежит, по существу, тот же

алгоритм, что и для усеченного среднего. Как и в случае 20% усеченного среднего, для каждого хвоста распределения удалялось 20% данных, однако после этого выполнялась дополнительная процедура – на место удаленных значений подставлялись, соответственно, максимальные и минимальные величины из оставшихся данных. В целом, поскольку во всех рассмотренных случаях винсоризованное среднее оказалось менее устойчивой мерой, чем усеченное среднее, можно говорить о том, что полученные нами результаты не демонстрируют необходимости и целесообразности процедуры винсоризации при расчете меры центральной тенденции.

Наконец, достаточно неожиданными могут быть названы результаты, касающиеся поведения одношаговой *M*-оценки и модифицированной одношаговой *M*-оценки. Эти меры центральной тенденции, как и некоторые другие меры семейства *M*-оценок, в современной литературе часто рекомендуются в качестве статистик, устойчивых к появлению выбросов и смещенности распределения. Действительно, модифицированная одношаговая *M*-оценка оказалась мерой, наименее изменяющейся при переходе от одного моделируемого условия к другому. Однако внутри каждого из условий с фиксированным количеством выбросов ни одна из рассматриваемых *M*-оценок не входила в число наиболее стабильных мер. Более того, рассматриваемые *M*-оценки не входили в число наиболее стабильных мер и в том случае, когда эксплицитно моделировались данные, в которых количество выбросов заранее неизвестно.

Таблица 2

Стабильность мер центральной тенденции

Мера центральной тенденции	Отсутствие выбросов		Один выброс		Два выброса		Три выброса		SD*	От нуля до трех выбросов	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>
<i>M</i>	500	18,580	517	26,609	533	32,598	550	37,787	21,57	525	35,151
<i>hM</i>	484	14,699	481	43,451	478	58,925	474	71,390	4,20	479	52,197
<i>gM</i>	491	16,319	498	23,043	505	28,160	513	32,991	9,13	502	27,072
<i>Md</i>	473	18,109	475	19,062	477	19,877	480	20,922	2,80	476	19,600
<i>M(2SD)</i>	484	17,280	492	19,022	497	20,410	501	23,164	7,11	494	21,008
<i>TrimM</i>	479	16,581	482	17,573	485	18,509	489	19,756	4,03	484	18,389
<i>WinsM</i>	484	17,164	488	18,356	491	19,587	496	21,253	5,07	490	19,593
<i>OSE</i>	484	17,543	487	18,740	490	19,851	494	21,375	4,35	488	19,700
<i>MOSE</i>	475	19,369	477	19,971	478	20,484	480	21,259	2,20	478	20,319
<i>DWE</i>	483	16,501	487	17,715	492	19,006	498	20,901	6,71	490	19,407
<i>SPWE</i>	490	17,066	498	19,407	507	21,882	517	25,079	11,61	503	23,275

Примечания. *r* – коэффициент корреляции Пирсона, *p* – уровень значимости; *M* – арифметическое среднее; *hM* – гармоническое среднее; *gM* – геометрическое среднее; *Md* – медиана; *M(2SD)* – среднее после отсекания данных, лежащих за пределами двух стандартных отклонений; *TrimM* – 20% усеченное среднее, с удалением 20% данных в каждом хвосте распределения; *WinsM* – винсоризованное среднее после предварительного удаления 20% данных в каждом хвосте распределения; *OSE* – одношаговая *M*-оценка; *MOSE* – модифицированная одношаговая *M*-оценка; *DWE* – оценка, взвешенная по расстоянию; *SPWE* – оценка, взвешенная по скалярному произведению единичных векторов; *SD* – стандартное отклонение значения меры центральной тенденции по 50 000 генерациям; *SD** – стандартное отклонение среднего значения меры центральной тенденции по четырем условиям наличия выбросов. В каждом столбце жирным курсивом выделены три меры с наименьшим стандартным отклонением.

В этом контексте представляется необходимым вернуться к обсуждению двух возможных источников нестабильности меры центральной тенденции, о которых говорилось выше. В современной литературе стабильность меры центральной тенденции чаще всего оценивается как ее устойчивость к появлению выбросов в конкретном наборе данных. Выше говорилось о том, что существуют различные подходы к оценке робастности мер центральной тенденции. В данной работе в качестве такого показателя рассматривалась стабильность меры при переходе от одного моделируемого условия к другому, при этом количество выбросов было фиксированным внутри каждого условия. Полученные результаты подтверждают, что такие меры, как медиана или одношаговая М-оценка, действительно являются наиболее устойчивыми к появлению выбросов и изменению их количества.

Однако из результатов очевидна и существующая проблема – наряду с невысокой изменчивостью этих мер при переходе от одного условия к другому внутри каждого условия они оказываются чувствительными к колебаниям в конкретном наборе данных, хотя моделируемые данные происходят из одного и того же распределения. Результаты, полученные в условии с переменным количеством выбросов, наглядно иллюстрируют, что в случае одновременного учета двух возможных источников нестабильности мер центральной тенденции ни медиана, ни рассматриваемые М-оценки не могут быть рекомендованы в качестве наиболее стабильных показателей. Эта особенность данных мер центральной тенденции делает проблематичным их применение для анализа локализации распределения эмпирических данных, особенно если набор имеющихся значений невелик.

Обобщая полученные результаты, можно говорить о том, что в рассматриваемых условиях наиболее предпочтительным оказалось введение двух мер центральной тенденции – 20% усеченного среднего и оценки, взвешенной по расстояниям. Однако необходимо еще раз подчеркнуть, что в рамках данной работы ставилась цель симуляции данных, правдоподобно имитирующих времена ответов, которые могут быть получены в реальной элементарной скоростной задаче. Поведение мер центральной тенденции, безусловно, будет изменяться при изменении смещения распределения исходных данных, а также при увеличении количества выбросов и изменении диапазона их возможных значений. Иными словами, корректный выбор меры центральной тенденции в каждом случае должен основываться на предварительном сравнительном анализе поведения различных мер в тех конкретных условиях, которые наилучшим образом описывают имеющиеся эмпирические данные.

Заключение и выводы

В данной работе рассматривалась проблема оценивания центральной тенденции в условиях смещенности распределения эмпирических данных и наличия в нем выбросов. Основное внимание было сфокусировано на проблеме получения индивидуального показателя скорости переработки информации в элементарной когнитивной задаче, то

есть в условиях, когда для каждого испытуемого имеется относительно небольшой набор значений, распределение которых смещено и может содержать выбросы.

Результаты, представленные в данной работе, позволяют говорить о том, что в качестве наиболее предпочтительных могут рассматриваться две меры центральной тенденции: 20% усеченное среднее и предложенная в данной работе оценка, взвешенная по расстояниям. Результаты, касающиеся усеченного среднего, в целом являются предсказуемыми – в современной литературе по статистике эта мера часто рекомендуется в качестве наиболее предпочтительной в условиях отсутствия нормальности распределения данных и возможного наличия в них выбросов.

Однако следует подчеркнуть, что в расчете этой меры центральной тенденции, например в случае 20% усечения, участвует только 60% полученных эмпирических значений. Эта особенность данной меры центральной тенденции, безусловно, должна четко осознаваться использующим ее исследователем. И речь идет не только о том, что, как справедливо отмечал Р.Вилкоккс, использование усеченных данных в целом противоречит интуиции и здравому смыслу [Wilcox, 2001]. Действительно, трудно принять, что удаление существенного количества информации может помочь получить более точные оценки, чем в случае использования всех имеющихся данных. Однако существует и более глубокая проблема, а именно вопрос о том, имеет ли исследователь право полностью исключать из анализа те или иные эмпирически зафиксированные значения.

Учебники по статистике, демонстрируя эффект влияния выбросов на меры центральной тенденции, традиционно используют подчеркнuto наглядные примеры – как тот пример из значений 2, 3, 4, 4, 5, 5, 6, 6, 6, 80, который был использован нами в первой части данной работы. Однако можно ли считать выбросом значения, которые далеко не так очевидно отстоят от основного массива данных?

Например, какие значения являются выбросами в ряду данных 260, 291, 303, 343, 403, 468, 494, 536, 548, 821 (приведены времена ответов реального испытуемого, в мс)? При этом вопрос состоит не только в точности идентификации выбросов, но и в правомочности их полного игнорирования при расчете показателя, характеризующего локализацию распределения имеющихся значений. Действительно ли полученное очень большое значение времени ответа является исключительно следствием внешних факторов или наличие больших времен тоже до некоторой степени отражает особенности протекания когнитивного процесса, интересующего исследователя?

На наш взгляд, основное преимущество мер центральной тенденции, основанных на взвешивании, состоит именно в отсутствии необходимости однозначного рассмотрения части значений в качестве выбросов, подлежащих удалению. Так, предложенная в данной статье оценка, взвешенная по расстояниям, позволяет работать с полным набором данных, взвешивая каждое значение в соответствии с его удаленностью относительно основного массива. Проведенная серия компьютерных симуляций позволяет говорить о том, что поведение этой меры центральной тенденции вполне сопоставимо с поведением 20% усеченного среднего. При этом

логично, что для моделируемых положительно смещенных распределений оценка, основанная на взвешивании, в каждом условии оказывалась несколько более высокой, чем усеченное среднее, поскольку в случае взвешивания более редкие большие значения в правом хвосте распределения не игнорировались, как при усечении, а все же учитывались при расчете показателя, характеризующего местоположение распределения.

Наконец, в заключение представляется необходимым еще раз подчеркнуть, что проблема выбора меры центральной тенденции не является специфичной для анализа данных, получаемых в скоростных когнитивных задачах. Действительно, анализ времен ответов является той областью, где проблема оценивания местоположения распределения данных встает с неизбежной остротой. В данной работе было показано, насколько может различаться поведение мер центральной тенденции в зависимости от конкретных экспериментальных условий, и как предпочтение той или иной меры может сказываться на общих результатах и выводах эмпирического исследования. Нет ни одной причины, почему описанная проблема может относиться только к данным времен ответов и не касаться других случаев, когда исследователь имеет дело с усреднением данных, полученных в эмпирическом исследовании. Иными словами, вне зависимости от того, подсчитывает ли исследователь некоторый балл на основании результатов заполнения опросника, обобщает ли данные наблюдения или количественно выраженные клинические данные, «простой» подсчет среднего – эта та операция, которая заслуживает специального внимания и предварительного анализа имеющихся эмпирических данных.

Завершая данную работу, представляется возможным кратко сформулировать следующие выводы.

1. Во многих случаях данные, с которыми работает исследователь, не являются нормально или даже симметрично распределенными, а количество имеющихся значений ограничено. В таких условиях арифметическое среднее не может рассматриваться в качестве надежной меры центральной тенденции в силу отсутствия устойчивости к смещенности распределения и наличию выбросов.

2. Подход, наиболее часто рекомендуемый в современной статистике для анализа локализации распределения эмпирических данных, основывается на предварительном усечении тех значений, которые расположены в хвостах распределений, и последующем усреднении оставшихся значений. Хотя существует множество алгоритмов идентификации значений, подлежащих удалению, устойчивая мера центральной тенденции может быть получена уже путем простого усечения фиксированного одинакового процента значений из каждого хвоста распределения данных. Вопрос о количестве значений, подлежащих усечению, должен решаться с учетом особенностей конкретного исследования, но в большинстве случаев усечение порядка 20% данных из каждого хвоста распределения можно считать оптимальным.

3. В качестве альтернативы усечению данных может рассматриваться подход, основанный на взвешивании. В рамках этого подхода в расчете меры центральной тенденции участвуют все полученные значения, однако

их относительный вес зависит от их удаленности от основного массива данных. Из двух взвешенных оценок центральной тенденции, предложенных в данной работе, по крайней мере одна оценка оказывается предпочтительной по сравнению с большинством других мер и вполне сопоставимой с 20% усеченным средним. При этом несомненное преимущество подхода, основанного на взвешивании, состоит в отсутствии необходимости удаления части имеющихся значений, что особенно актуально для реальных эмпирических данных, когда практически невозможно с уверенностью отнести то или иное значение к категории выбросов.

Приложение

Алгоритмы расчета мер центральной тенденции в программной среде R

```
# Арифметическое среднее:
mean(x)

# Гармоническое среднее: hM
hM = function(x,n) {y = x[!is.na(x)]
n = length(x); n/sum(1/x)}

# Геометрическое среднее: gM
gM = function(x) {y = x[!is.na(x)]; exp(mean(log(y)))}

# Медиана:
median(x)

# Среднее в двух стандартных отклонениях: m2sd
m2sd = function(x) {y = x[!is.na(x)]; a1 = mean(y) + 2*sd(y)
a2 = mean(y) - 2*sd(y); z = y[y > a2 & y < a1]; mean(z)}

# 20% усеченное среднее:
mean(x,trim = 0,2)

# Винсоризованное среднее: wins
wins = function(x,tr) {y = sort(x); n = length(x)
ibot = floor(tr*n) + 1; itop = length(x) - ibot + 1
xbot = y[ibot]; xtop = y[itop]; y = ifelse(y <= xbot, xbot, y)
y = ifelse(y >= xtop, xtop, y); mean(y)}

# Одношаговая M оценка
ose = function(x) {y = x[!is.na(x)]; m = median(y)
z = abs(y - m); md = median(z)/0.6745
a1 = m - 1.28*md; a2 = m + 1.28*md; d = y[y > a1 & y < a2]
i1 = length(y[y < a1]); i2 = length(y[y > a2])
(1.28*md*(i2 - i1) + sum(d))/length(d)}

# Модифицированная одношаговая M оценка
mose = function(x) {y = x[!is.na(x)]; m = median(y)
z = abs(y - m); md = median(z)/0.6745
a1 = m - 2.24*md; a2 = m + 2.24*md
d = y[y > a1 & y < a2]; mean(d)}

# Скалярно-взвешенная оценка
spwe = function(x) {y = x[!is.na(x)]
a = pi/2*(y - min(y))/(max(y) - min(y)); b = a
p = outer(b, a, function(b, a) abs(cos(b - a))); c = y
q = outer(c, y, function(c, y) (c + y)/2)
m1 = p - diag(diag(p)); m2 = q - diag(diag(q))
sum(m1*m2)/sum(m1)}

# Оценка, взвешенная по расстоянию
dwe = function(x) {y = x[!is.na(x)]; a = y; b = y
p = outer(b, a, function(b, a) abs((b - a)))
n = colSums(p)/length(y); w = 1/n; sum(y*w)/sum(w)}
```

Пример алгоритма компьютерной симуляции данных

Условие с двумя выбросами, описание параметров генерации приведено выше (стр. 10):

```
Mu = 400; Sigma = 20; Nu = 100
N = 30; out = 2
y = 1:50000; a1=y; a2 = y; a3 = y; a4 = y; a5 = y; a6 = y
a7 = y; a8 = y; a9 = y; a10 = y; a11 = y
for(i in 1:50000){
t = rnorm(N-out, Mu, Sigma) + Nu*rexp(N-out)
p = runif(out, 0, 2000)
x = c(t,p)
a1[i] = mean(x); a2[i] = median(x); a3[i] = hM(x)
a4[i] = gM(x); a5[i] = m2sd(x); a6[i] = ose(x)
a7[i] = mose(x); a8[i] = dwe(x); a9[i] = spwe(x)
a10[i] = mean(x,trim = 0.2); a11[i] = wins(x,0.2)}
mean(a1); mean(a2); mean(a3); mean(a4); mean(a5)
mean(a6); mean(a7); mean(a8); mean(a9); mean(a10)
mean(a11)
sd(a1); sd(a2); sd(a3); sd(a4); sd(a5); sd(a6)
sd(a7); sd(a8); sd(a9); sd(a10); sd(a11)
```

Литература

Barnett V., Lewis T. Outliers in statistical data. 3rd ed. New York: Wiley, 1994.

Dodonov Y.S. Response time and intelligence: problems of data weighting and averaging // Poster presented on the Eleventh Annual Conference of International Society for Intelligence Research. Alexandria, USA, 2010.

Dodonova Y.A., Dodonov Y.S. Speed of emotional information processing and emotional intelligence // International Journal of Psychology (in press).

Heathcote A., Popiel S.J., Mewhort D.J. Analysis of response time distributions: An example using the Stroop task // Psychological Bulletin. 1991. Vol. 109. P. 340–347.

Hockley W.E. Retrieval processes in continuous recognition // Journal of Experimental Psychology: Learning, Memory, & Cognition. 1982. Vol. 8. P. 497–512.

Hockley W.E. Analysis of response time distributions in the study of cognitive processes // Journal of Experimental Psychology: Learning, Memory, & Cognition. 1984. Vol. 10. P. 598–615.

Hohle R.H. Inferred components of reaction times as a function of foreperiod duration // Journal of Experimental Psychology. 1965. Vol. 69. P. 382–386.

Huber P.J. Robust statistics. New York: Wiley, 1981.

Jensen A. The g factor. London: Praeger, 1998.

Keselman H.J., Othman A.R., Wilcox R.R., Fradette K. The new and Improved two-sample t test // American Psychological Society. 2004. Vol. 15. P. 47–51.

Keselman H.J., Algina J., Lix L. M., Wilcox R.R., Deering

K. A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes // Psychological Methods. 2008. Vol. 13. P. 110–129.

Lovie P. Identifying outliers // A.D.Lovie (ed.). New developments in statistics for psychology and the social sciences. British Psychological Society: London, 1986. P. 44–69.

Luce R.D. Response times: Their role in inferring elementary organization. New York: Oxford University Press, 1996.

Othman A.R., Keselman H.J., Padmanabhan A.R., Wilcox R.R., Fradette K. Comparing measures of the 'typical' score across treatment groups // British Journal of Mathematical and Statistical Psychology. 2004. Vol. 57. P. 215–234.

Ratcliff R. A theory of memory retrieval // Psychological Review. 1978. Vol. 85. P. 59–108.

Ratcliff R. Group reaction time distributions and an analysis of distribution statistics // Psychological Bulletin. 1979. Vol. 86. P. 446–461.

Ratcliff R. Methods for dealing with reaction time outliers // Psychological Bulletin. 1993. Vol. 114. P. 510–532.

Ratcliff R., McKoon G. The diffusion decision model: Theory and data for two-choice decision tasks // Neural Computation. 2008. Vol. 20. P. 873–922.

Ratcliff R., Murdock B.B., Jr. Retrieval processes in recognition memory // Psychological Review. 1976. Vol. 83. P. 190–214.

Rocke D.M., Downs G.W., Rocke A.J. Are robust estimators really necessary? // Technometrics. 1982. Vol. 24. P. 95–101.

Rosenberg, M.S. The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis // Evolution. 2005. Vol. 59. P. 464–468.

Rousseeuw P.J., Croux C. Alternatives to the median absolute deviation // Journal of the American Statistical Association. 1993. Vol. 88. P. 1273–1283.

Sheppard L.D., Vernon P.A. Intelligence and speed of information-processing: A review of 50 years of research // Personality and Individual Differences. 2008. Vol. 44(3). P. 535–551.

Ulrich R., Miller J. Effects of outlier exclusion on reaction time analysis // Journal of Experimental Psychology: General. 1994. Vol. 123. P. 34–80.

Wilcox R.R. Fundamentals of modern statistical methods. New York: Springer, 2001.

Wilcox R.R. Applying contemporary statistical techniques. San Diego, CA: Academic Press, 2003.

Wilcox R.R. Introduction to robust estimation and hypothesis testing. 2nd ed. San Diego, CA: Academic Press, 2005.

Wilcox R.R., Keselman H.J. Modern robust data analysis methods: Measures of central tendency // Psychological Methods. 2005. Vol. 8. P. 254–274.

Рекомендованная ссылка:

Додонов Ю.С., Додонова Ю.А. Устойчивые меры центральной тенденции: взвешивание как возможная альтернатива усечению данных при анализе времен ответов [Электронный ресурс] // Психологические исследования: электрон. науч. журн. 2011. N 5(19). URL: <http://psystudy.ru> (дата обращения: чч.мм.гггг).