УДК 159.9.072

Маничев С.А.[a], Федоров С.И.[a], Быков Р.О.[b]

[a]Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
[b]ООО «НН Формат», Санкт-Петербург, Россия

Взаимосвязи Закона сравнительных суждений Терстоуна и некоторых других моделей одномерного шкалирования

The Relationship Between the Thurstonian Law of Comparative Judgment and Other Models of Psychological Unidimensional Scaling

Аннотация

В работе рассматриваются некоторые подходы и модели психометрики, оказавшие значительное влияние на развитие методов шкалирования психологических величин и обсуждаются связи и взаимовлияние между ними. Сравниваются две процедуры получения эмпирических данных: парные сравнения (сравнение респондентом двух объектов) и прямые ответы (выбор респондентом ответа на вопрос о единичном объекте). Далее сравниваются две математические модели шкалирования, применимые к данным парных сравнений: модель Терстоуна (и вытекающая из нее модель Бредли-Терри-Люса) и модель развертывания Кумбса. Обсуждается применимость этих моделей и к данным прямых ответов (работы Андрича).

Ключевые слова: Закон сравнительных суждений Терстоуна, Метод одномерного развертывания, психологическое шкалирование, метод случайных величин, латентная переменная

Abstract

In this paper we evaluate a number of models and approaches in psychometrics, which had influenced greatly over the development and evolution of the theory of psychological scaling. We also discuss their mutual interference and relationships. We consider and compare the two procedures of empirical data collection: pair comparison procedure (a respondent compares a pair of objects with respect to the certain feature/attribute) and the direct response procedure (a respondent answers to a question about a single object). Later on we compare the two mathematical models of psychological scaling that may be applied to the pair comparison data: the Thurstone model (and implied Bradly-Terry-Luce model) and Coombs' unfolding model. The applicability of these models to direct response data is further discussed (see papers by David Andrich).

Keywords: Law of Comparative Jugments, unfolding, psychological scaling, method of random variables, latent variable

Введение

В этой работе авторы хотели бы привлечь внимание к некоторым известным математическим моделям в психологии, оказавшим значительное влияние на становление и развитие психометрики в целом.

Это модели:

1) Закон Сравнительных Суждений Терстоуна (в дальнейшем – ЗСС), который теперь часто называют Моделью Парных Сравнений Терстоуна и модели, ставшие его развитием;

2) Модель разворачивающихся шкал Кумбса – модель одномерного развертывания (Coombs' unfolding model);

3) Подход Дэвида Андрича к Модели Терстоуна и IRT-моделированию, основанный на сравнении различных процедур получения эмпирических данных.

Следует сразу отметить, что цель работы не столько в том, что бы еще раз напомнить об этих моделях по отдельности, а скорее в том, чтобы попытаться заново взглянуть на некоторые связи – формальные и неформальные – между ними.

При шкалировании объектов/стимулов традиционно используется схема сбора данных в виде парных сравнений – сравнения респондентами пар объектов, причем даже в этой схеме есть различные подходы: схема сравнения и схема предпочтения. В первой респондентам дается инструкция делать выбор между двумя объектами в общем, используя указанный признак, абстрагируясь от собственного «идеального субъективного представления о признаке» (см., например, Thurstone, 1928), во второй респондент основывает свой выбор уже на своем субъективном предпочтении объектов. Так, в модели Терстоуна предполагается схема сравнения, а в Методе Кумбса – схема предпочтения с учетом субъективного параметра. Парные сравнения являются ресурсозатратной процедурой, ибо каждый из респондентов должен провести сравнение для каждой пары объектов. В IRT (Item Responce Theory) же, обычно применяется схема сбора данных прямого ответа – выбор респондентом ответа на вопрос о единичном объекте. И количество ответов каждого респондента, соответственно, меньше количества парных сравнений в число раз, равное количеству объектов. Ниже, при обсуждении взаимосвязей между моделями, в том числе, рассматривается возможность применения схемы прямого ответа к этим моделям шкалирования.

Закон сравнительных суждений Терстоуна

В 1927 году Л.Л. Терстоун опубликовал свою знаменательную работу (Thurstone, 1927; см. также Thurstone, 1959; Thurstone, 1929), оказавшую вместе с более ранними работами Фехнера и Вебера (см. Fechner, 1966) существенное влияние на становление и развитие современной психометрики. Обсуждению важности этой работы Терстоуна посвящено, начиная с середины XX века и по сей день, множество публикаций в мировой научной печати (см., например, обсуждения и библиографии в Орлов, 2004; Давид, 1978; Luce, 1977; Luce, 1994). Так, в своей работе Р.Д. Люс (Luce, 1994) отмечает, как минимум, два момента во влиянии работ Терстоуна 20-х годов на современную психометрику и психофизику – развитие идей психологического континуума (на современном языке – шкалирования) и представления стимула на этом континууме в виде дискриминативного процесса, что на современном языке означает представление стимулов с помощью случайных величин. В работе J.-C. Falmagne (Falmagne, 2002) этот подход прямо называется методом случайных величин Терстоуна. Отметим, что использование случайных величин для представления стимулов имеет отношение не только к описанию дискриминативных процессов (процессов различения), но и к соответствующей процедуре принятия решения в процессе различения.

В своей основополагающей работе Л.Л. Терстоун (Thurstone, 1927) сформулировал и вывел математически свой знаменитый «закон сравнительных суждений». Суть этой модели в следующем: рассматриваются N объектов $A_1, A_2, \ldots A_N$, которые при соотнесении их с некоторым признаком (не только физической, но, может, и субъективной природы) имеют на шкале признаков определенные значения «ценности» / «привлекательности» / или «значимости» $a_1, a_2, \ldots a_N$. Некоторое значительное число M субъектов (скажем, «экспертов») проводит парные сравнения каждого объекта со всеми остальными по этому признаку. Каждый из «экспертов» (скажем, k-ый) при сравнении i-ого объекта с j-тым присваивает переменной X^k_{ij} значение 1, если по его мнению $a_i > a_j$, и присваивает 0, если $a_i < a_j$. Исходя из этих данных, требуется определить положения шкальных значений «ценностей» объектов на латентной шкале признаков.

Для каждой пары объектов A_i A_j в результате суммирования X^k_{ij} по всем «экспертам» подсчитывается относительная частота π_{ij} числа тех сравнений, в которых объект A_i был признан более «ценным» чем объект A_j , то есть эмпирическая вероятность того, что $a_i > a_j$. Далее, Терстоун, учитывая возможные ошибки и различия оценок «экспертов», вносит в модель элемент случайности, предполагая, что «эксперты» сравнивали не истинные значения «ценностей» объектов, а стохастические «ценности». Терстоун предполагает, что оцениваемые «ценности» объектов A_i есть случайные величины U_i, распределенные по нормальному закону с дисперсиями σ_i^2 и средними, равными «истинным ценностям» a_i. Таким образом, результат «эмпирического» дискриминативного процесса попарных сравнений объектов всеми «экспертами» приравнивается результату «теоретического» дискриминативного процесса – вероятность π_{ij} того, что i-ый объект ценнее j-того равняется $Pr(U_i - U_j > 0)$ – вероятности того, что случайная величина U_i больше случайной величины U_j и, следовательно, для каждой пары i и j выполняется уравнение (1):

$$\pi_{ij} = Pr\,(U_i - U_j)\,, \quad 1 \le i,j \le N \tag{1}$$

Для простоты изложения мы далее будем следовать предположениям V случая модели закона сравнительных суждений (см., например, Andrich, 1978 или Thurstone, 1927). А именно, предположения Терстоуна в этом случае заключаются в следующем: случайные величины U_i предполагаются попарно не коррелирующими, а их дисперсии равными ($2\sigma_i^2 = 2\sigma_j^2 = \sigma^2$ для всех i и j, меняющихся от 1 до N). Тогда, используя выражение для функции распределения разности двух нормально распределенных случайных величин, получается, что распределение для U_i - U_j нормально с дисперсией σ^2 и средним равным a_i - a_j – разности «истинных ценностей» i-ого и j-ого объектов. Далее, обозначая, как обычно, через Φ(x) стандартную функцию нормального распределения с нулевым средним и единичной дисперсией, сразу получаем, что $Pr\,(U_i - U_j > 0) = \Phi\,((a_i - a_j)/\sigma)$ и, следовательно, учитывая результаты дискриминативного процесса (парных сравнений), получаем уравнение (2):

$$\pi_{ij} = \Phi\left(\frac{a_i - a_j}{\sigma}\right), \quad 1 \le i,j \le N \tag{2}$$

Таким образом, мы, практически, закончили вывод «закона сравнительных суждений» на современном языке, в варианте, близком к изложению в (Andrich, 1978). Надо отметить, что можно дополнительно без хлопот избавиться от константы σ в этом уравнении – просто надо равномерно уменьшить масштаб шкалы ценностей в σ раз, то есть совершить преобразование a_j в a_j/σ и, после этого, уравнение (2) примет вид (3):

$$\pi_{ij} = \Phi\,(a_i - a_j)\,, \quad 1 \le i,j \le N \tag{3}$$

Поскольку Φ(x) – известная монотонно возрастающая функция и, следовательно, имеет хорошо определенную обратную функцию Φ^{-1}, то

$a_j - a_i = \Phi^{-1}(\pi_{ij})$ и , учитывая, что π_{ij} нам известны из опыта, мы моментально можем определить разность $a_j - a_i$. Однако, нахождение всех попарных разностей между значениями ценностей далеко от решения проблемы об определении самих шкальных значений ценностей объектов.

Следует отметить, особо не вдаваясь в подробности, что, если матрица эмпирических значений вероятностей сравнения π_{ij} не транзитивна, то система уравнений из (3) не имеет решения. Нетранзитивность, в частности, может означать, что мнения «экспертов» не одномерны, то есть при вынесении решения о различении они используют не одномерный признак, что нарушает правомерность использования метода случайных величин. Но, даже в случае, если нет нарушения транзитивности, система (3) переопределена – число уравнений существенно больше числа неизвестных. В этой ситуации, обычно, применяются приближенные методы решения (с последующей процедурой *model fit*), из которых наиболее популярны и разработаны метод наименьших квадратов и метод наибольшего правдоподобия (см., например Орлов, 2004; Давид, 1978; Гусев, 1998).

Полезно будет добавить к сказанному, что использованное при выводе соотношение (1), где эмпирически полученная вероятность приравнивается к теоретической вероятности того, что соответствующая случайная величина больше нуля, и послужила основой для вышеупомянутого наименования подхода Терстоуна методом случайных величин Терстоуна (см. Falmagne, 2002).

На практике, для упрощения вычислений по формуле (3), часто заменяют нормальную функцию распределения Ф(x) более простой по виду, но очень близкой к ней, логистической функцией (4):

$$L(x) = \frac{\exp(x)}{1 + \exp(x)}, \tag{4}$$

Известно, что Φ(x) можно аппроксимировать с достаточно хорошей точностью (обычно достаточной для приближенных вычислений) с помощью L(x) для всех значений x, а именно, для всех x верно, что -0,01 < Φ(x) - L(1,7x) < 0,01 (см. Johnson, Kotz, 1972). Поэтому выражение (3) может быть приближенно заменено на выражение (5):

$$\pi_{ij} = L\left(1,7(a_i - a_j)\right), \quad 1 \leq i, j \leq N \tag{5}$$

которое для простоты использования, как и в выше описанном случае с переходом от (2) к (3), растяжением масштаба шкалы в 1,7 раз, то есть преобразованием a_i в 1,7 a_i может быть приведено к виду (6):

$$\pi_{ij} = L(a_i - a_j) = \frac{\exp\left(a_i - a_j\right)}{1 + \exp\left(a_i - a_j\right)}, \quad 1 \leq i, j \leq N \tag{6}$$

В таком виде закон сравнительных суждений прямо соответствует модели парных сравнений Бредли-Терри-Люса (см. Bredly, Terry, 1952; Luce, 1959), в которой, правда, выражение в правой части равенств (6) не выводится, в отличие от закона сравнительных суждений, а просто постулируется.

Метод Одномерного Развертывания Кумбса

В 50-70-е и в 80-е годы прошлого столетия появилось несколько подходов и моделей, предложивших дополнить подход Терстоуна учетом субъективного фактора, то есть особенностей «экспертов», и всю «вину» или, хотя бы, часть «вины» за стохастичность результатов парных сравнений переложить на них. Иными словами, предполагается, что стохастика оценок вызвана не только значениями «ценностей» объектов, но и мнениями «экспертов». Первые примеры разработок в этом направлении, о которых мы собираемся упомянуть, – это метод одномерного развертывания Кумбса (Coombs, 1964; Coombs, 1950; Coombs, 1952) и наиболее значимые частные подходы и модели в его развитии (см., например, Bechtel, 1968; Coombs, Greenberg, Zinnes, 1961; Greenberg, 1965; Sixtl, 1973).

Кумбс предположил, что каждый из «экспертов» имеет свое собственное «идеальное мнение» о «ценностях» объектов и может сопоставить его с «ценностями» объектов: определить в любом наборе объектов тот объект, который наиболее близок к индивидуальному представлению «эксперта» о ценности. Таким образом, участие «эксперта» в процессе парных сравнений заключается в определении в каждой паре объектов того объекта, который ближе к идеалу «эксперта». Но это – эмпирическая часть парных сравнений.

На математическом языке идеалы «»экспертов» представляются на шкале ценности объектов некой латентной переменной *y*, называемой идеалом. Тогда, легко понять, что из двух объектов A_i и A_j со значениями ценностей a_i и a_j ближайшим к идеалу *y* будет та из ценностей объектов, на стороне которой по отношению к середине интервала между ними – $a_{ij} = (a_i + a_j)/2$ – находится идеал *y*. Это математическое утверждение точно соответствует сказанному выше словами описанию действий при парных сравнениях «эксперта» с идеалом *y* – он из пары объектов выбирает тот, чья ценность ближе к его идеалу. При формулировке своего подхода Кумбс изначально не делал никаких ограничительных метрических и стохастических предположений ни о распределении идеалов, ни о распределениях ценностей «объектов». Такие ограничения, помогающие построить общую шкалу идеалов и признаков – «ценностей» объектов – стали накладывать в частных моделях подхода Кумбса.

Уточним, что означает термин «развертывание» в названии подхода Кумбса. Дело в том, что сравнение идеала с «ценностью» объекта не имеет направления – наиболее близкий к идеалу объект на абсолютной для всех «экспертов» шкале ценностей (в терминологии Кумбса – J-шкале, *Joint scale*) может иметь «ценность» как выше, так и ниже значения идеала, важно лишь то, что его «ценность» – ближайшая к идеалу. Поэтому индивидуальные шкалы каждого из «экспертов» (в терминологии Кумбса – I-шкалы, *Individual scales*), могут существенно отличаться друг от друга. Так, если идеал на общей шкале стоит ниже упорядоченных в порядке возрастания «ценностей» всех объектов, то на его I-шкале порядок ранжирования объектов будет тот же, но если идеал стоит выше «ценностей» всех объектов, то порядок ранжирования для этого «эксперта» изменится на обратный. Причем, изменения порядка ранжирования объектов при движении точки, соответствующей значению идеала на оси, слева направо будут происходить именно при пересечении точкой положения идеала серединных точек $a_{ij} = (a_i + a_j)/2$.

Таким образом, I-шкалы оказываются свернутыми по отношению к общей J-шкале и, наоборот, общая J-шкала оказывается развернутой в ряд I-шкал. Причем, как видно из уже обсужденного, порядок на J-шкале прямо определяет порядок на I-шкале идеала *y*, но не наоборот – порядок на I-шкале не определяет напрямую порядка на J-шкале. Собственно, разные модели в методе Кумбса и предлагают разные подходы для восстановления J-шкалы из I-шкал для каждого из «экспертов», то есть предлагают разные подходы и алгоритмы для развертывания I-шкал в общую J-шкалу.

Вышеупомянутую математическую форму описания действий «эксперта» при парных сравнениях (определение ближайшего к его идеалу объекта) можно переформулировать. Предположим, что объекты имеют определенные фиксированные ценности на общей J-шкале (что, естественно, например, для объектов физической природы, сравниваемым по физическим признакам) и они упорядочены на шкале «ценностей» в порядке возрастания $a_1 < a_2 < . . . < a_N$. В этом случае в каждом сравнении объектов A_i и A_j (при упомянутом условии, что $a_i < a_j$) объект A_i предпочитается (ближе к идеалу *y*)

ISSN: 2225-7527

объекту A_i тогда и только тогда, когда для идеала *y* выполняется неравенство (7):

$$y < \frac{a_i + a_j}{2} = a_{ij}, \quad 1 \leq i, j \leq N \tag{7}$$

Более того, если предположить, что переменная идеалов *y* имеет определенное распределение на шкале ценностей, функция распределения которого F(x) строго монотонно возрастает, то оказывается (см., например, Sixtl, 1973), что эмпирические значения вероятностей парных сравнений π_{ij} связаны не с попарными разностями ценностей (как у Терстоуна), а с этими самыми серединами интервалов между «ценностями» $a_{ij} = (a_i + a_j)/2$, а именно (8):

$$\pi_{ij} = F(a_{ij}) = F\left(\frac{a_i + a_j}{2}\right), 1 \leq k \leq K, 1 \leq j \leq N \tag{8}$$

Выражение (8) в методе Кумбса, конечно же, играет роль, аналогичную выражению (3) в исходном подходе Терстоуна, и далее, основной трудностью для решения проблемы парных сравнений будет определение шкальных значений объектов в подходе Кумбса. Если функция распределения идеалов F(x), например, нормальная, с известными средним и дисперсией (как в модели Бечтеля (Bechtel, 1968)), то каждая средняя точка $a_{ij} = (a_i + a_j)/2$ между парой «ценностей» вполне определяется соответствующей «эмпирической» вероятностью π_{ij} и, если разобраться предварительно с опять же могущей возникнуть нетранзитивностью, далее для решения (как и в случае системы, определяемой выражением (3)) идут в ход приближенные методы решения переопределенных систем уравнений (у Бечтеля метод наименьших квадратов и неизбежная процедура *model-fit*). В конце своей работы Бечтел при обсуждении результатов анализирует степень ограничительности предположений своей модели и приходит к заключению, что предположение о нормальности распределения идеалов менее ограничительно и более соответствует практике, чем предположение о существовании общей J-шкалы для объектов с единственным естественным порядком объектов, приводящее к выражению (7). Он замечает, что в практических случаях иногда приходится рассматривать несколько вариантов (если не все) шкал и порядков на них и выбирать наилучший из них по результатам проверки статистической транзитивности и проведения процедуры *model-fit*. Однако, по его мнению, наличие такой процедуры последовательных проверок шкал и порядков может иметь и положительный эффект – может, в принципе, привести к нахождению новых признаков и характеристик, управляющих процедурой предпочтения.

В своей работе Сикстлу удалось в своем подходе (Sixtl, 1973), который он назвал вероятностным развертыванием, избежать дополнительного предположения Бечтеля о конкретизации вида распределения идеалов, более того, оценки распределения идеалов получаются у него как следствие соответствия и применимости (*model-fit*) его модели к данным.

Подход Д. Андрича к модели Терстоуна и методу развертывания в IRT

Второй вариант введения субъективного параметра (характеристики «эксперта») был предпринят конце 70-х годов прошлого столетия известным австралийским специалистом в области IRT-шкалирования Д. Андричем в работе (Andrich, 1978). Следует, забегая вперед, предположить по результатам анализа его публикаций до и после этой работы, что он, скорее всего, не был достаточно глубоко знаком на момент ее написания с методом Кумбса и основными моделями по его применению.

В этой работе Андрич, проанализировав работу Терстоуна о законе сравнительных суждений и переизложив достаточно изящно и неформально математическую часть ее вывода на современном языке, замечает, что он удивлен отсутствием в подходе Терстоуна учета субъективного параметра, относящегося к характеристике «экспертов», ибо включение этого параметра, по его мнению, прямо и естественно ложится на канву вывода, проведенного Терстоуном. По сути, субъективный параметр, о котором говорит Андрич, это прямой аналог идеала в Методе Кумбса – внутреннего представления «эксперта» об «идеальной ценности».

Однако, по Андричу, применение этого идеала в процедуре парных сравнений идет иначе, чем в методе Кумбса. «Эксперты» получают в схеме Кумбса и в схеме Андрича разные «инструкции» по применению идеала в процессе парных сравнений.

Об «инструкции» в методе Кумбса мы уже говорили – найти в каждой паре объектов тот объект, который ближе к идеалу.

«Инструкция» Андрича – сравнить каждый объект с идеалом и определить выше (ответ – 0) или ниже (ответ – 1) идеала ценность объекта. Вместо матрицы парных сравнений в этой схеме эмпирические данные представляют собой матрицу из нулей и единиц, где строкам соответствует респондент, а столбцам – оцениваемый объект.

Каждый из К «экспертов» или групп «близких по мнению экспертов» (то есть имеющих одинаковый идеал) имеет идеал y_k – нормально распределенную случайную величину, со средним t_k и дисперсией σ_k^2, где k соответствует номеру «эксперта» или группы. Далее Андрич накладывает ограничения, вполне в духе V случая закона сравнительных суждений: все σ_k равны между собой, $2\sigma_k^2 = 2\sigma_l^2 = \sigma^2$ для всех индексов k и l между 1 и К , а σ – некоторое фиксированное число. Следуя своему варианту вывода закона сравнительных суждений, Андрич получает выражение для Pr (y_k – a_j > 0) – теоретической вероятности того, что k-ый «эксперт» оценил свой идеал стоящим выше ценности j-ого объекта:

$$Pr(y_k - aj > 0) = \Phi(t_k - aj)), \quad 1 \leq k \leq K, 1 \leq j \leq N,$$

ISSN: 2225-7527

и опять, используя уже объясненную нами выше связь между функцией стандартного нормального распределения Ф(x) и логистической функцией L(x), в том же ключе, как было выведено ранее выражение (6) из (3), выводит представление вероятности Pr(y_k - a_j >0) в терминах логистической функции L(x) (9):

$$Pr(y_k - a_j > 0) = \frac{\exp(t_k - a_j)}{1 + \exp(t_k - a_j)}, \quad 1 \leq k \leq K, 1 \leq j \leq N \tag{9}$$

И, в дополнение к этому, Андрич выводит из (9) выражение (6), соответствующее выражению вероятности в модели Бредли-Терри-Люса, используя при этом известный в IRT моделировании математический трюк (см., например, Rasch, chapter 10, 1980; Fisher, Molenaar, chapter 3, 1995), позволяющий, во вполне определенных случаях, при шкалировании объектов / стимулов / вопросов исключить из рассмотрения субъективные параметры.

Для этого, в данном контексте, рассматривается матрица ответов «экспертов» $\{X_{kl}\}$, где k – номер «эксперта» (или группы экспертов с тем же средним значением t_k идеала), а i – номер объекта, $1 \leq 1 \leq N$. Выберем пару объектов с номерами i и j, сравнение которых подлежит изучению. В матрице ответов i-ый и j-ый столбцы содержат ответы всех «экспертов» на сравнение их идеалов с ценностями этих объектов. Для перехода от этой матрицы прямых ответов к парным сравнениям надо рассмотреть только тех «экспертов», идеал которых находится между a_i и a_j – ценностями рассматриваемой пары объектов. Понятно, что для этих «экспертов» в их строчках матрицы ответов либо X_{ki}=1, а X_{kj}=0, либо X_{ki}=0, а X_{kj}=1. Отсюда сразу вытекает, что эти «эксперты» описываются условием $X_{ki} + X_{kj}$=1. Вычислим теперь вероятность π^k_{ij} того, что для таких «экспертов» ценность i-ого объекта лежит не ниже идеала (X_{ki}=0), а ценность j-ого объекта лежит ниже идеала (X_{kj}=1). Иными словами нам надо посчитать условную вероятность $Pr(X_{kj}=1 \mid X_{ki}+ X_{kj}=1)$ – это и будет вероятность того, что для идела y_k этого «эксперта» выполнено неравенство $a_j \leq y_k < a_i$, и с такой вероятностью парное сравнение, вытекающее из его прямых ответов, будет в пользу a_i. Она нетрудно вычисляется с помощью выражения (9): используя (9) и учитывая, что $Pr(X_{kl}=0)=1- Pr(X_{kl}=1)$ для всех номеров l от 1 до N и, что ответы в разных столбцах независимы, сразу можем записать, что (10):

$$Pr(X_{ki}=0, X_{kj}=1) = Pr(X_{ki}=0)Pr(X_{kj}=1) \quad (10)$$
$$= \frac{\exp(t_k - aj)}{(1+\exp(t_k - a_j))} \cdot \frac{1}{(1+\exp(t_k - a_i))} = \frac{\exp(t_k - aj)}{(1+\exp(t_k - a_j))(1+\exp(t_k - a_i))}$$

и (11):

$$Pr(X_{ki}=1, X_{kj}=0) = Pr(X_{ki}=1)Pr(X_{kj}=0) \quad (11)$$
$$= \frac{\exp(t_k - a_i)}{(1+\exp(t_k - a_1))} \cdot \frac{1}{(1+\exp(t_k - a_j))} = \frac{\exp(t_k - ai)}{(1+\exp(t_k - a_i))(1+\exp(t_k - a_j))}$$

Далее, заметим, что сложение двух предыдущих выражений дает (12):

$$Pr(X_{ki}+X_{kj}=1) = Pr(X_{ki}=0, X_{kj}=1) + Pr(X_{ki}=1, X_{kj}=0) \quad (12)$$
$$= \frac{\exp(t_k - a_j) + \exp(t_k - a_i)}{(1+\exp(t_k - a_j))(1+\exp(t_k - a_i))}$$

и, что по (12):

$$\pi_{ij}^k = Pr(X_{kj}=1 | X_{ki}+X_{kj}=1) = \frac{Pr(X_{ki}=0, X_{kj}=1)}{Pr(X_{ki}=0, X_{kj}=1) + Pr(X_{ki}=1, X_{kj}=0)}$$

Мы можем подставить в последнее выражение результаты (10), (11) и (12) и получить искомое выражение для вероятности π^k_{ij}. Мы не будем приводить здесь соответствующие громоздкие формулы и проводить над ними простые, по сути, но «многоэтажные» алгебраические упрощения и приведем только окончательный результат (13):

$$\pi_{ij}^{k}=\frac{\exp(t_k-aj)}{\exp\left(t_k-a_j\right)+\exp\left(t_k-a_i\right)}=\frac{\exp(a_i-aj)}{1+\exp\left(a_i-a_j\right)} \quad (13)$$

Осталось заметить, что из (13) сразу следует, что левая часть этой цепочки равенств не зависит от субъективного параметра, поскольку от него не зависит правая. Субъективный параметр был исключен в процессе вычисления соответствующей условной вероятности (*subject parameter has been conditioned out*). Правая же часть (13) представляет собой в точности выражение (6) из модели Бредли-Терри-Люса.

Таким образом, Андрич следовал своему варианту вывода закона сравнительных суждений, но всю стохастичность в вынесении оценок объектов переложил на субъектов («экспертов»): идеалы являются случайными величинами, а ценности объектов фиксированы и имеют вполне определенные (латентные) значения a_1, a_2,...a_N.

Далее Андрич показывает, что для шкалирования с помощью прямых ответов пригодны не только модели Терстоуна (ее вариант – модель Бредли-Терри-Люса), но и модель развертывания.

Для данных прямых ответов Андрич использовал вариант IRT моделирования (см. Andrich, 1989, 1993, 1995, 1998). В классическую IRT модель (см., например, Fisher, Molenaar, 1995; van der Linden, 2016) прямо входит разность (θ - β) между параметрами субъекта и характеристиками вопроса (то есть, упомянутого в вопросе стимула): вероятность ответа «да» полагается равной значению логистической функции L(x) (см. выражение (4)) для x, равного (θ - β), то есть выражению (14):

$$\frac{\exp\left(\theta-\beta\right)}{1+\exp\left(\theta-\beta\right)}, \quad (14)$$

В этой модели вероятность выбора ответа «да» монотонно возрастает в зависимости от величины разности.

В вероятностных моделях развертывания (Бетчел, Сикстл и др.) вероятность выбора зависит от близости объекта к идеалу, т.е. от абсолютной величины разности: с увеличением величины разности вероятность выбора падает, знак разности не имеет значения.

Поэтому Андрич вводит функцию вероятности правильного ответа, зависящую не от порядка расположения параметров субъекта и вопроса (стимула) на латентной шкале, а только от величины интервала между ними. В его работах (Andrich, 1995; Andrich, 1993) эта идея реализуется путем использования в качестве функции вероятности ответа функции гиперболического косинуса – четной функции, а в работах (Andrich, 1988; Andrich, 1989) заменой в выражении (14) разности (θ - β) на ее квадрат, взятый со знаком минус: $-(\theta - \beta)^2$. И в том, и в другом случае вероятность ответа – четная функция, на значение которой не влияет знак разности (θ - β).

Можно, конечно, задаться более практическим вопросом соответствия моделей реальности – какая из схем сбора данных, более естественна для «эксперта» – субъекта, проводящего сравнения? Насколько естественна предложенная Андричем в (Andrich, 1978) схема сравнения собственного идеала субъекта с каждым из объектов и, на основе полученного результата, выбора наибольшего по «ценности» объекта?

Складывается ощущение, что процедура сравнения идеала с одним объектом сильно напоминает *ad hoc* процедуру, и, что многие специалисты назовут более естественным парное сравнение, основываясь при этом на следующем неформальным соображении – сравнение объекта и идеала должно быть «одномерным», то есть происходить по вполне определенному признаку. Однако идеальное представление об объекте не обязательно «одномерно». В этом случае парные сравнения (использование пары объектов) сводят контекст сравнения к общему признаку пары объектов и «снимают» многомерность или многозначность идеала.

Обсуждение

1) Мы проследили за связью модели Терстоуна (ЗСС) с подходом Кумбса и с введением в рассмотрение в последнем субъективного параметра – идеала. Обе модели работают для парных сравнений (как схемы сбора данных).

2) Модели IRT работают с данными схемы прямых ответов. Д. Андрич показал, что через введение субъективного параметра – идеала, и по схеме прямого ответа можно получить результаты, схожие с заключительной формой модели Терстоуна и модели Бредли-Терри-Люса. Андрич также предложил IRT модели развертывания, что показывает, что модель развертывания работает и с данными схемы прямых ответов. Таким образом, (как это и было отмечено в (Andrich, 1989)) нет жесткой привязки схем сбора данных к используемым математическим моделям – обе модели – и модель Терстоуна, и модель развертывания, в принципе, могут работать и со схемой парных сравнений, и со схемой прямого ответа.

3) Во всех рассмотренных моделях для определения шкальных значений стимулов / вопросов были указаны соответствующие системы уравнений (3), (8), (13). Как уже отмечалось, все они оказываются переопределенными (число уравнений существенно превышает число неизвестных параметров, подлежащих определению). Поэтому решают эти системы хорошо разработанными приближенными методами с последующим обязательным применением процедуры *model-fit*, дающей статистическую оценку соответствия модели и полученных решений исходным эмпирическим данным и, в некоторой степени, позволяющей даже оценить, какой из приближенных методов лучше подходит для конкретной модели. *Model-fit* процедура предполагает использование одной или нескольких, так называемых, *Goodness-of-fit* статистик, созданных для тестирования конкретной модели или класса моделей. Однако, их применение может не гарантировать получения вразумительного ответа на вопрос о соответствии

модели данным даже в случае полного тестирования на *Overall-Goodness-of-fit*, предполагающего применение всех возможных доступных статистик, включая самые жесткие. Это полное тестирование само по себе ресурсозатратно в смысле компьютерных возможностей и не всегда возможно, особенно в случае редких / разреженных данных (*sparse data*).

В этой связи уместно привести в сокращенном виде разумную идею, высказанную одним из известных специалистов в *model-fit* тестировании А. Мейдеу-Оливаресом в работе (Maydeu-Olivares, 2015): всякая модель имеет некоторые определенные свойства и используется (и в том числе тестируется на соответствие данным) для некоторых определенных целей. Поэтому перед тем как использовать и, тем более, тестировать модель на соответствие данным, исследователю имеет смысл озаботиться вопросом о том, какие свойства модели и для каких целей он собирается применять.

При этом стоит иметь ввиду, что имеют место случаи, когда применение хорошо изученных *model-fit*-процедур может дать не только отрицательный ответ о соответствии модели данным, но и ложный положительный ответ. Так в (Karabatsos, 2001) приводится один интересный пример, в котором хорошо известная *fit*-процедура применялась к модели с известным и интересующим исследователя свойством. Эмпирические данные же были сгенерированы с помощью другой модели, указанным свойством тестируемой модели не обладающим. Процедура дала положительный ответ о соответствии модели этим данным, в которых, по их построению, упомянутое свойство тестируемой модели не выполнялось. Из чего напрашивается вывод, что использованная *fit*-процедура на проверку соответствия данных (по меньшей мере) именно этому свойству модели не была настроена.

Выводы

Рассмотренные модели шкалирования – и модель Терстоуна, и метод развертывания – позволяют использовать сбор данных и по схеме парных сравнений, и по схеме прямого ответа. Таким образом, перед исследователем, в принципе, раскрывается возможность использования четырех вариантов дизайна исследования. Используемые приближенные процедуры нахождения шкальных оценок объектов / стимулов в обязательном порядке предполагают статистическую оценку соответствия результатов применения модели эмпирическим данным, которая получается с применением критериев *model-fit*. Однако исследователь должен постоянно иметь ввиду, что результаты этой статистической оценки могут как не гарантировать соответствия эмпирических данных математической модели с определенными, интересующими исследователя свойствами, так и давать ложный положительный результат относительно этого соответствия. Используемые процедуры / критерии *model-fit* могут не быть настроены на проверку соответствия данных интересующему исследователя свойству модели.

Список использованных источников

Орлов А.И. Прикладная статистика. М., Экзамен, 2004.

Дэвид Г. Метод парных сравнений. М., Статистика, 1978.

Гусев А.Н., Измайлов Ч.А., Михалевская М.Б. Измерение в психологии: общий психологический практикум. Серия «Практикум». Вып. 2. М., Смысл, 1998.

Andrich D. Relationships between the Thurstone and Rasch approaches to item scaling // Applied Psychological Measurement, 1978. Vol. 2. No. 3. Pp. 449-460.

Andrich D. The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude // Applied Psychological Measurement, 1988. Vol. 12. No. 1. Pp. 33-51.

Andrich D. A Probabilistic IRT Model for Unfolding Preference Data // Applied Psychological Measurement, 1989. Vol. 13. No. 2. Pp. 193-216.

Andrich D. Hyperbolic Cosine Latent Trait Models for Unfolding Direct Responses and Pairwise Preferences // Applied Psychological Measurement, 1995. Vol. 19. No. 3. Pp. 269-290.

Andrich D, Luo G. A Hyperbolic Cosine Latent Trait Model For Unfolding Dichotomous Single-Stimulus Responses // Applied Psychological Measurement, 1993. Vol. 17. No. 3. Pp. 253-276.

Andrich D., Styles I.M. The structural relationship between attitude and behavior statements from the unfolding perspective // Psychological Methods, 1998. Vol. 3(4). Pp. 454-469.

Bradley R.A., Terry M.E. Rank analysis of incomplete block designs, I. the method of paired comparisons // Biometrika, 1952. No. 39. Pp. 324-345.

Bechtel G.G. Folded and unfolded scaling from preferential paired comparisons // Journal of Mathematical Psychology, 1968. No. 5. Pp. 333-357.

Coombs C.H. A theory of data. New York, Wiley, 1964.

Coombs C.H. Psychological scaling without a unit of measurement // Psychological Review, 1950. Vol. 57(3). Pp. 145-158.

Coombs C.H. A theory of psychological scaling // Engineering Research Bulletin, University of Michigan Press, 1952. No. 34.

Coombs C.H., Greenberg M., Zinnes J.L. A double law of comparative judgment for the analysis of preferential choice and similarities data // Psychometrika, 1961. No. 26. Pp. 165-171.

Falmagne J.-C. Elements of psychophysical theory // Oxford psychology series. New York, Oxford University Press, 2002. No. 6.

Fechner G. Elements of Psychophysics. New York: Holt, Rinehart and Winston, 1966.

Fischer G.H., Molenaar I.W. (Eds.). Rasch Models.Foundations, Recent Developments, and Applications. Springer-Verlag, New York, 1995.

Greenberg M.G. A method of successive cumulations for the scaling of pair-comparison preference judgments // Psychometrika, 1965. No. 30. Pp. 441-448.

Johnson N.L., Kotz S. Distributions in statistics: Continuous univariate distributions. Vol. 3. New York, Wiley, 1972.

Karabatsos G. The Rasch Model, Additive Conjoint Measurement and New Models of Probabilistic Measurement Theory // Journal of Applied Measurement, 2001. No. 2(4). Pp. 389-423.

Luce R.D. Thurstone's discriminal processes fifty years later // Psychometrika, 1977. Vol. 42. Is. 4. Pp. 461-489

Luce R.D. Thurstone and Sensory Scaling: Then and Now // Psychological Review, 1994. Vol. 101. No. 2. Pp. 271-277.

Luce R.D. Individual Choice Behaviours: A Theoretical Analysis. New York: J. Wiley, 1959.

Maydeu-Olivares A. Evaluating fit in IRT models / Chapter 6 in Handbook of Item Response Theory Modeling: Applications to Typical Performance. Routledge, 2015. Pp. 111-127.

Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute of Educational Research, 1960 / Expanded edition. Chicago: The University of Chicago Press, 1980.

Sixtl F. Probabilistic unfolding // Psychometrica, 1973. Vol. 38. No. 2. Pp. 235-248.

Thurstone L.L. A law of comparative judgement // Psychological Review, 1927. Vol. 34. Pp. 273-286.

Thurstone L.L. The Measurement of Values. Chicago: The University of Chicago Press, 1959.

Thurstone L.L. The Measurement of Psychological Value / In T.V. Smith and W.K. Wright (Eds.), Essays in Philosophy by Seventeen Doctors of Philosophy of the University of Chicago. Chicago: Open Court, 1929.

Thurstone L.L. Attitudes can be measured // American Journal of Sociology, 1928. Vol. 33. No. 4. Pp. 529-554.

van der Linden W.J. Handbook of Item Response Theory. Volume One: Models / Statistics in the Social and Behavioral Sciences Series, CRC Press Taylor & Francis Group, 2016.

References

Orlov A.I. Prikladnaia statistika [Applied Statistics]. Moscow, Ekzamen Publ., 2004. (In Russian)

Devid G. Metod parnykh sravnenii [The method of paired comparisons]. Moscow, Statistika Publ., 1978. (In Russian)

Gusev A.N., Izmailov Ch.A., Mikhalevskaia M.B. Izmerenie v psikhologii: obshchii psikhologicheskii praktikum [The measumemt in psychology]. Seriia «Praktikum». Is. 2. Moscow, Smysl Publ., 1998. (In Russian)

Andrich D. Relationships between the Thurstone and Rasch approaches to item scaling // Applied Psychological Measurement, 1978. Vol. 2. No. 3. Pp. 449-460.

Andrich D. The Application of an Unfolding Model of the PIRT Type to the Measurement of Attitude // Applied Psychological Measurement, 1988. Vol. 12. No. 1. Pp. 33-51.

Andrich D. A Probabilistic IRT Model for Unfolding Preference Data // Applied Psychological Measurement, 1989. Vol. 13. No. 2. Pp. 193-216.

Andrich D. Hyperbolic Cosine Latent Trait Models for Unfolding Direct Responses and Pairwise Preferences // Applied Psychological Measurement, 1995. Vol. 19. No. 3. Pp. 269-290.

Andrich D, Luo G. A Hyperbolic Cosine Latent Trait Model For Unfolding Dichotomous Single-Stimulus Responses // Applied Psychological Measurement, 1993. Vol. 17. No. 3. Pp. 253-276.

Andrich D., Styles I.M. The structural relationship between attitude and behavior statements from the unfolding perspective // Psychological Methods, 1998. Vol. 3(4). Pp. 454-469.

Bradley R.A., Terry M.E. Rank analysis of incomplete block designs, I. the method of paired comparisons // Biometrika, 1952. No. 39. Pp. 324-345.

Bechtel G.G. Folded and unfolded scaling from preferential paired comparisons // Journal of Mathematical Psychology, 1968. No. 5. Pp. 333-357.

Coombs C.H. A theory of data. New York, Wiley, 1964.

Coombs C.H. Psychological scaling without a unit of measurement // Psychological Review, 1950. Vol. 57(3). Pp. 145-158.

Coombs C.H. A theory of psychological scaling // Engineering Research Bulletin, University of Michigan Press, 1952. No. 34.

Coombs C.H., Greenberg M., Zinnes J.L. A double law of comparative judgment for the analysis of preferential choice and similarities data // Psychometrika, 1961. No. 26. Pp. 165-171.

Falmagne J.-C. Elements of psychophysical theory // Oxford psychology series. New York, Oxford University Press, 2002. No. 6.

Fechner G. Elements of Psychophysics. New York: Holt, Rinehart and Winston, 1966.

Fischer G.H., Molenaar I.W. (Eds.). Rasch Models.Foundations, Recent Developments, and Applications. Springer-Verlag, New York, 1995.

Greenberg M.G. A method of successive cumulations for the scaling of pair-comparison preference judgments // Psychometrika, 1965. No. 30. Pp. 441-448.

Johnson N.L., Kotz S. Distributions in statistics: Continuous univariate distributions. Vol. 3. New York, Wiley, 1972.

Karabatsos G. The Rasch Model, Additive Conjoint Measurement and New Models of Probabilistic Measurement Theory // Journal of Applied Measurement, 2001. No. 2(4). Pp. 389-423.

Luce R.D. Thurstone's discriminal processes fifty years later // Psychometrika, 1977. Vol. 42. Is. 4. Pp. 461-489

Luce R.D. Thurstone and Sensory Scaling: Then and Now // Psychological Review, 1994. Vol. 101. No. 2. Pp. 271-277.

Luce R.D. Individual Choice Behaviours: A Theoretical Analysis. New York: J. Wiley, 1959.

Maydeu-Olivares A. Evaluating fit in IRT models / Chapter 6 in Handbook of Item Response Theory Modeling: Applications to Typical Performance. Routledge, 2015. Pp. 111-127.

Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute of Educational Research, 1960 / Expanded edition. Chicago: The University of Chicago Press, 1980.

Sixtl F. Probabilistic unfolding // Psychometrica, 1973. Vol. 38. No. 2. Pp. 235-248.

Thurstone L.L. A law of comparative judgement // Psychological Review, 1927. Vol. 34. Pp. 273-286.

Thurstone L.L. The Measurement of Values. Chicago: The University of Chicago Press, 1959.

Thurstone L.L. The Measurement of Psychological Value / In T.V. Smith and W.K. Wright (Eds.), Essays in Philosophy by Seventeen Doctors of Philosophy of the University of Chicago. Chicago: Open Court, 1929.

Thurstone L.L. Attitudes can be measured // American Journal of Sociology, 1928. Vol. 33. No. 4. Pp. 529-554.

van der Linden W.J. Handbook of Item Response Theory. Volume One: Models / Statistics in the Social and Behavioral Sciences Series, CRC Press Taylor & Francis Group, 2016.