

УДК 616-072.87(07)

ИНВАРИАНТНОСТЬ ПСИХОМЕТРИЧЕСКИХ МОДЕЛЕЙ

Ассанович М.А.

УО «Гродненский государственный медицинский университет», Гродно, Беларусь

Инвариантность является свойством стабильности результатов научного измерения. Проводилась сравнительная оценка инвариантности показателей классической теории тестов и модели Раша. Параметры классической теории тестов в значительной степени зависят от особенностей выборки. Модель Раша инвариантна независимо от выборки.

Ключевые слова: психометрия, классическая теория тестов, модель Раша, инвариантность.

Инвариантность – важнейшее свойство научного измерения. Инвариантность означает неизменность параметров измерения при изменении условий его проведения. В психодиагностическом смысле это значит, что при одинаковом уровне выраженности психологического конструкта результаты исследования разных людей будут также одинаковы [1, 3]. На допущении условия инвариантности построены нормативные критерии оценки выраженности психологического конструкта в классической теории тестов. Считается, что критерии выраженности измеряемого свойства, разработанные на одной популяции испытуемых, будут неизменно подходить и к другим популяциям. На основе этого допущения используется большинство психодиагностических методик [1, 2, 3].

Однако данные исследований и обзоров последних лет показывают, что психодиагностические тесты, используемые в рамках классической теории тестов (КТТ), на самом деле не обладают инвариантностью. В качестве действительно инвариантной психометрической модели указывается однопараметрическая логистическая модель Раша [1, 2].

В настоящее время практически все русскоязычные психодиагностические методики используются на основе классической теории тестов. Подавляющее большинство из них содержит нормативные критерии, полученные более 10-20 лет назад в популяциях, никакого отношения не имеющих к современному населению Республики Беларусь. В отдельных исследованиях, проводимых в РБ, используются методики, напрямую заимствованные из зарубежных иноязычных источников. Эти методики вообще никакой психометрической калибровки в белорусской популяции не проходили. В некоторых исследованиях предпринимаются попытки адаптации зарубежных методик на ограниченных выборках испытуемых. Все перечисленные моменты еще больше усугубляют проблему инвариантности психодиагностических методик, применяющихся в РБ.

На сегодняшний день особенно актуальна необходимость построения психодиагностических тестов на основе вероятностных метрических моделей, инвариантность которых может не зависеть от распределения тестовых оценок в популяции. Наиболее перспективной моделью в этом плане является модель Раша [2, 3].

Суть настоящей работы состоит в том, чтобы оценить, насколько отличаются инвариантности классической теории тестов и модели Раша при получении тестовых оценок на разных выборках.

Цель исследования: сравнительная оценка стабильности параметров классической теории тестов и модели Раша при исследовании выборок с разными особенностями распределения оценок.

Материал и методы исследования. В качестве психодиагностической методики, параметры которой оценивались в рамках двух психометрических мо-

делей, была выбрана шкала ипохондрии Миннесотского многофазного личностного опросника (ММРП). Выбор этой шкалы обусловлен тем, что шкала ипохондрии имеет достаточно простую и понятную эмпирическую структуру. Она состоит из 33 пунктов, на которые предоставляются дихотомические ответы «верно» или «неверно». Каждый ключевой ответ оценивается в 1 балл. Таким образом, диапазон сырых тестовых оценок по шкале составляет от 0 до 33 баллов.

В соответствии с целью исследования были сформированы 6 выборок протоколов исследования ММРП. Условно они получили кодовые названия НР1, НР2, НС, ВС, НЗ, ВЗ. Данные выборки были образованы из общей базы протоколов исследования 2148 пациентов с невротическими расстройствами, депрессиями легкой и умеренной степени выраженности и расстройствами личности. Характеристики всех 6 выборок представлены в таблице 1.

Таблица 1 – Общая характеристика выборок протоколов ММРП

Характеристика	Выборки протоколов					
	НР1	НР2	НС	ВС	НЗ	ВЗ
Количество протоколов	277	249	247	255	313	302
Характер распределения оценок	Близкое к нормальному	Близкое к нормальному	Смещенное в сторону низких оценок	Смещенно в сторону высоких оценок	Только низкие оценки	Только высокие оценки
Диапазон оценок (в баллах)	0-33	0-33	0-19	12-33	0-7	23-33

Как следует из таблицы 1, первые две выборки (НР1 и НР2) включали весь диапазон оценок, распределение которых было близким к нормальному. Следующие две выборки (НС и ВС) содержали преимущественно низкие - средние и высокие-средние оценки. Оставшиеся две выборки (НЗ и ВЗ) носили маргинальный характер в виде только низких или только высоких тестовых оценок.

Оценка инвариантности психометрических моделей включала сравнение параметров пунктов, распределений шкальных оценок, показателей надежности шкалы в указанных выше выборках.

Результаты исследования и их обсуждение

1. Оценка инвариантности психометрических моделей по показателям параметров пунктов шкалы ипохондрии.

Сравнивались такие параметры пунктов, как трудность и качество. В классической теории тестов качество пункта оценивается по значению коэффициента бисериальной корреляции (КБК) [1]. В модели Раша качество пункта определяется по значениям индексов среднеквадратических остатков. Выделяются два типа индексов: взвешенный (WMS) и невзвешенный (UMS) [2].

Результаты сравнительного анализа параметров пунктов в выборках представлены в таблице 2.

Таблица 2 – Совпадения оценок параметров пунктов в психометрических моделях

Пары выборок	Классическая теория тестов		Модель Раша		
	Совпадения оценок трудности пунктов	Совпадения значений КБК	Совпадения оценок трудности пунктов	Совпадения значений UMS	Совпадения значений WMS
HP1-HP2	100%	70%	100%	79%	100%
HC-BC	64%	61%	97%	91%	100%
H3-B3	6%	0%	49%	85%	100%

Анализ совпадений показывает, что параметры пунктов в формате КТТ сохраняют относительную инвариантность только в выборках со сходными распределениями оценок. В выборках со смещенными распределениями, и особенно с маргинальной разницей в распределениях, степень рассогласования в оценках параметров значительно велика. Модель Раша обнаруживает рассогласование в оценках трудностей только в маргинальных выборках, при этом сохраняя инвариантность у половины пунктов. Индексы качества пунктов в модели Раша отличаются значительно большей устойчивостью к различиям в распределении оценок по сравнению с коэффициентами полисерийной корреляции КТТ. При этом абсолютной инвариантностью характеризуется индекс взвешенных среднеквадратичных остатков.

2. Оценка инвариантности шкал нормализованных тестовых оценок.

Результаты сравнительного анализа обоих психометрических подходов приведены в таблице 3.

Таблица 3 – Параметры нормализованных шкал в психометрических моделях

Выборки и диапазон оценок	Классическая теория тестов		Модель Раша	
	Диапазон шкальных оценок	Средний уровень по шкале	Диапазон шкальных оценок	Средний уровень по шкале
HP1 (0-33)	0-32	15	0-33	16
HP2 (0-33)	0-30	16	0-33	16
HC (0-19)	0-19	13	0-33	16
BC (12-33)	12-33	18	0-33	16
H3 (0-7)	0-7	5	0-33	16
B3 (23-33)	23-33	24	0-33	16

Анализ представленных в таблице данных демонстрирует главное преимущество модели Раша перед классической теорией тестов. Если психодиагностическая шкала нормализуется в рамках КТТ, то нормализованные оценки будут полностью зависеть от особенностей распределения сырых оценок в выборке. Данные в таблице показывают, что даже в выборках со сходными распределениями, охватывающими весь диапазон возможных тестовых баллов, диапазоны нормализации и средние уровни полностью не совпадают. Этим недостатком полностью лишена модель Раша. Независимо от распределения сырых оценок в выборке при построении шкалы фиксируется средний уровень, обусловленный не распределением тестовых баллов, а количеством пунктов в шкале. Вне зависимости от ограничений диапазона сырых оценок шкальная метрическая структура включает все возможные тестовые оценки.

3. Оценка инвариантности показателей надежности.

В таблице 4 представлены значения коэффициентов надежности по данным исследования каждой выборки. В формате КТТ рассчитывался коэффициент альфа Кронбаха [1]. В формате модели Раша рассчитывались индексы надежности отдельно для пунктов и испытуемых, психометри-

чески эквивалентные коэффициенту альфа [2, 3].

Таблица 4 – Значения коэффициентов надежности в выборках

Выборки	Классическая теория тестов	Модель Раша	
	Коэффициент альфа Кронбаха	Индекс надежности пунктов	Индекс надежности испытуемых
HP1	0.85	0.97	0.85
HP2	0.87	0.97	0.86
HC	0.72	0.97	0.75
BC	0.62	0.98	0.65
H3	0.07	0.94	0.11
B3	0.96	0.97	0.01

Анализ данных в таблице оценок надежности раскрывает некоторые важные аспекты оценивания надежности в сравниваемых психометрических моделях. В классической теории тестов коэффициент альфа Кронбаха не дифференцирует, в какой степени оценка надежности характеризует психодиагностическую методику и ответы испытуемых из определенной выборки. Мы видим, что оценки надежности высоки в первых двух выборках и ниже в выборках со смещением. В маргинальных выборках значения альфа также носят маргинальный характер. При этом в выборке с высокими значениями (B3) получено самое высокое значение надежности. Скорее всего, это объясняется высокими дисперсиями ответов на пункты, что обусловлено преобладанием высоких значений тестовых оценок в данной выборке. Чем больше оценки испытуемых в выборке отличаются от средней оценки, тем выше дисперсия оценок и тем выше будут значения коэффициента надежности. Таким образом, в классической теории тестов оценки надежности в существенной степени зависят от распределения тестовых оценок в выборке. В модели Раша рассчитываются индексы двух видов надежности: пунктов и испытуемых. Представленные в таблице данные говорят о том, что оценки надежности пунктов, имеющие отношение непосредственно к конструкции шкалы, не зависят от особенностей выборки и носят действительный инвариантный характер. С другой стороны, оценки надежности испытуемых зависят от характера выборки и имеют более высокие значения в выборках с максимальным охватом возможных тестовых оценок. Собственно говоря, индекс надежности испытуемых в модели Раша и предназначен для того, чтобы определить, насколько данная выборка имеет целевое назначение для исследования данной шкалой в формате модели Раша [3].

Подводя итог проведенному анализу, можно сделать следующие выводы.

Психометрические характеристики методики, рассматриваемые в рамках классической теории тестов, обладают инвариантностью только при исследовании выборок со сходными распределениями и включающими весь диапазон возможных тестовых оценок. В практическом смысле это имеет отношение к нормализованным шкалам и нормативным критериям. Вследствие выраженной зависимости психометрических показателей от особенностей выборки недопустимо слепо использовать предлагаемые в руководстве к методике нормативные критерии оценки данных. Фактически, любая созданная в рамках классической теории тестов психодиагностическая методика нуждается в перерасчете психометрических показателей и повторной нормализации шкалы применительно к тем популяциям, в исследовании которых ее планируется использовать. Модель Раша

представляет собой альтернативную КТТ психометрическую модель, обладающую значительно более устойчивыми к особенностям выборки психометрическими свойствами. Неоспоримым преимуществом модели Раша перед КТТ является абсолютная инвариантность интервальной шкалы тестовых оценок вне зависимости от особенностей популяции и рас-

пределения тестовых баллов. Учитывая сказанное, следует считать крайне необходимым и целесообразным реконструирование используемых в Республике Беларусь психодиагностических методик в формате вероятностно-логистической психометрической модели Раша.

Литература

1. Crocher, L., Algina, J. Introduction to Classical and Modern Test Theory / L. Crocher, J. Algina. – Cengage Learning, 2008. – 527 p.
2. Embertson, S.E., Reise, S.P. Item Response Theory for

Psychologists /S.E. Embertson, S.P. Reise. – LEA, 2000. – 371 p.

3. Hambleton, R.K., Swaminathan, H., Jane Rogers, H. Fundamentals of Item Response Theory / R.K. Hambleton, H. Swaminathan, H. Jane Rogers. – Sage Publication, 1991. – 174 p.

INVARIANCE OF PSYCHOMETRIC MODELS

Assanovich M.A.

Educational Establishment "Grodno State Medical University", Grodno, Belarus

Invariance means stability of scientific measurement results. We conducted comparative assessment of the classical test theory and the Rasch model invariance. The classical test theory parameters considerably depend on the sample features. The Rasch model is invariant independently of the sample.

Key words: psychometrics, classical test theory, Rasch model, invariance.

Адрес для корреспонденции: e-mail: marat.as@tut.by

Поступила 12.11.2013