

Понятие психометрии

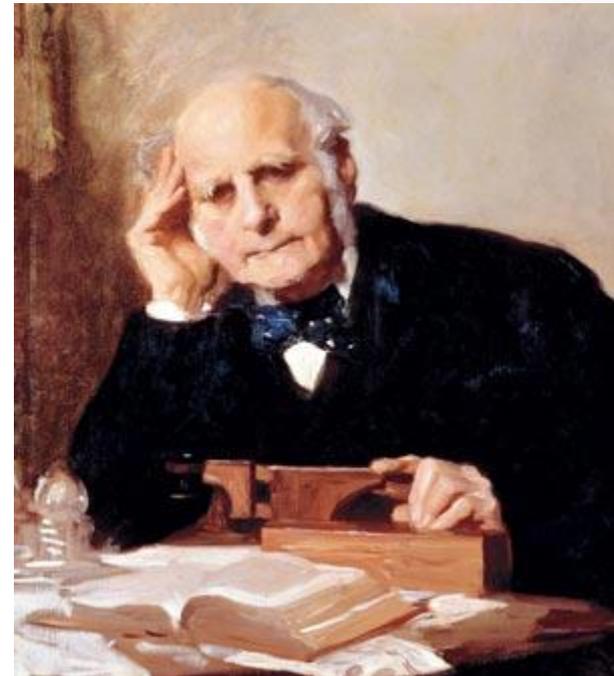
Выполнила студентка 4 курса
специальности Специальная психология
Панькова Анастасия Владимировна

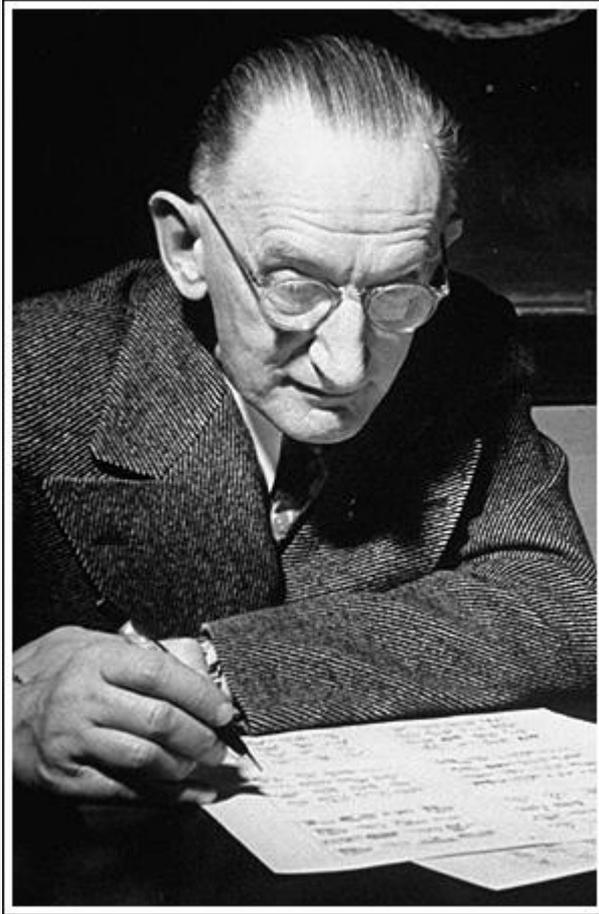
Что такое психометрия?

- Психометрия (психометрика) — дисциплина, изучающая теорию и методику психологических измерений, включая измерение знаний, способностей, взглядов и качеств личности. Психометрия является разделом прикладной статистики [источник не указан 831 день]. В первую очередь, эта область касается создания и валидации измерительных инструментов, таких как опросники, тесты и методики описания (оценки) личности. Она включает в себя две основные исследовательские задачи, а именно: 1) создание инструментов и построение процедур измерения; 2) развитие и усовершенствование теоретических подходов к измерению. Те, кто занимаются психометрией, могут быть как клиническими психологами, так и специалистами в области развития или HR. В любом случае, конкретной, отдельной квалификации по психометрии не требуется. В США преподавание психометрии ведётся на уровне бакалавриата, магистратуры и докторантуры.

Начало психометрии

- Большая часть ранних исследований в области психометрии была основана на стремлении измерить интеллект. Фрэнсис Гальтон, известный как «отец психометрии», включил ментальные измерения в антропометрические данные. Зарождение психометрии также связано с психофизикой. Два других начинателя психометрии Джеймс Маккин Кеттел и Чарльз Спирмен получили докторские звания в Лейпцигской лаборатории психофизики Вильгельма Вундта.





- Психометрист Луис Тёрстоун, основатель и первый президент Психометрического общества, в 1936 году разработал теоретический подход к измерению, который известен как закон сравнительных суждений. Этот подход тесно связан с психофизическими теориями Эрнста Вебера и Густава Фехнера. Также, Спирмен и Тёрстоун внесли большой вклад в развитие факторного анализа. Карл Пирсон, Генри Кайзер, Джордж Раш, Джонсон О'Коннор, Фредерик Лорд, Ледьярд Тюкер, Артур Дженсен также внесли большой вклад в развитие психометрии.

Область психометрии

- Область психометрии связана с количественным подходом к анализу тестовых данных. Психометрическая теория обеспечивает исследователей и психологов математическими моделями, используемыми при анализе ответов на отдельные задания или пункты тестов, тесты в целом и наборы тестов. Прикладная психометрия занимается применением этих моделей и аналитических процедур к конкретным тестовым данным. Четырьмя областями психометрического анализа являются нормирование и приравнивание, оценка надежности, оценка валидности и анализ заданий. Каждая из этих областей содержит набор определенных теоретических положений и конкретные процедуры, используемые при оценке качества работы теста в каждом отдельном случае.

Определение понятия «измерение» в социальных науках

- Определение измерения в социальных науках имеет долгую историю. В настоящее время широкое определение, предложенное Стэнли Смит Стивенсом (1946), гласит, что измерение «приписывание чисел объектам или событиям по некоторому правилу». Это определение было представлено в работе, в которой Стивенс предложил четыре уровня измерения. Хотя это определение имеет широкое распространение, оно отличается от более классического определения измерения, принятого в физике, которое гласит, что измерение — это численная оценка и выражение одной величины по отношению к другой.



- Действительно, определение Стивенса было выдвинуто в ответ Британскому Комитету Фергюсона, председатель которого, А. Фергюсон, был физиком. Комитет был назначен в 1932 году Британской ассоциацией для содействия развитию науки в исследовании возможности количественной оценки сенсорных восприятий. Хотя её председатель и другие члены были физиками, комитет также включал нескольких психологов. Доклад Комитета подчеркнул важность определения измерения. В то время, как ответ Стивенса заключался в том, чтобы предложить новое определение, которое окажет значительное влияние на эту область, это был не единственный ответ на доклад. Другой, кардинально отличающийся, ответ призывал принять классическое определение, как это отражено в следующем заявлении: "Измерение в психологии и физике ни в каком смысле не различны. Физики могут проводить измерения тогда, когда они могут найти операции, с помощью которых можно обнаружить необходимый критерий. Психологи могут не беспокоиться о таинственных различиях в значении «измерения» в двух науках".
- Эти различные точки зрения отражены в альтернативных подходах к измерению. Например, методы, основанные на ковариационной матрице, как правило, используют числа, такие как сырые баллы, как измерение. Такой подход неявно влечет за собой определение Стивенса, которое требует только то, что номера присваиваются по некоторому правилу. Таким образом, основной задачей исследований, как правило, считается открытие связей между показателями, и факторов, положенных в основу этих связей.
- С другой стороны, когда используется измерительная модель такая, как модель Раша, номера не присваиваются на основе правил. Вместо этого, в соответствии с заявлением Риз выше, конкретные критерии для измерения указаны, а цель состоит в построении процедур или операций, которые предоставляют данные, отвечающие соответствующим критериям. Измерения оцениваются на основе моделей, и испытания проводятся для того, чтобы удостовериться в том, были ли выполнены соответствующие критерии.

Нормирование тестов

- Нормирование тестов — составная часть их стандартизации, обычно включает проведение обследования репрезентативной выборки лиц, определение различных уровней выполнения тестов и перевод сырых тестовых оценок в общую систему показателей. Тесты иногда приравнивают, когда существуют различные формы того же самого теста. Приравнивание приводит оценки по всем формам к общей шкале. Существуют 4 основные стратегии приравнивания.
- Первый метод предполагает проведение каждой формы теста на эквивалентной (например, случайной отобранной) группе респондентов, а затем оценки по этим различным формам устанавливаются т. о., чтобы равные оценки имели равные процентильные ранги (та же самая пропорция респондентов получает ту же или более низкую оценку). При более точном методе все респонденты заполняют все формы теста, и для определения эквивалентности показателей используются уравнения. Третий часто используемый метод связан с проведением общего теста или части теста со всеми респондентами. Эта общая оценочная процедура служит в качестве «связывающего» теста, который позволяет все последующие измерения привязывать к единой шкале. При проведении обследования с использованием различных форм одного и того же теста в каждую включаются несколько «анкерных заданий», выполняющих функцию такого «связывающего» теста.

Требования к тестированию

- Надежность и валидность имеют отношение к обобщаемости показателей тестов — определению того, какие выводы по тестовым показателям являются обоснованными. Надежность касается выводов о согласованности измерения. Согласованность определяется по-разному: как временная устойчивость, как сходство между предположительно эквивалентными тестами, как однородность в рамках одного теста или как сравнимость оценок, выносимых экспертами. При использовании метода «тест-ретест» надежность теста устанавливается путем повторного его проведения с той же группой спустя определенный промежуток времени. Затем два полученных набора показателей сравниваются с целью определения степени сходства между ними. При использовании метода взаимозаменяемых форм, на выборке обследуемых проводятся два параллельных измерения. Привлечение экспертов («оценщиков») к оценке качества параллельных форм теста дает меру надежности, наз. надежностью оценщиков. Этот метод часто применяют, когда есть необходимость в экспертной оценке. Валидность характеризует качество выводов, получаемых на основе результатов проведения измерительной процедуры. Валидность рассматривается как способность теста отвечать поставленным целям и обосновывать адекватность решений, принятых на основе результата. Недостаточно валидный тест не может считаться инструментом измерения и использоваться на практике, поскольку зачастую полученный результат может серьезно влиять на будущее тестируемого. Выделяется три вида валидности тестов.

Виды валидности тестов

- Конструктивная (концептуальная) валидность. Её требуется определить, если тест измеряет свойство, имеющее абстрактный характер, то есть не поддающееся прямому измерению. В таких случаях необходимо создание концептуальной модели, которая бы объясняла данное свойство. Эту модель и подтверждает или опровергает тест.
- Критериальная (эмпирическая) валидность. Показывает, насколько соотносятся результаты теста с неким внешним критерием. Эмпирическая валидность существует в двух видах: текущая критериальная валидность — корреляция результатов теста с выбранным критерием, существующим в настоящее время; прогностическая критериальная валидность — корреляция результатов с критерием, который появится в будущем. Определяет, насколько тест предсказывает проявление измеряемого качества в будущем, учитывая влияние внешних факторов и собственной деятельности тестируемого.

- Содержательная валидность. Определяет, насколько соответствует тест его предметной области, то есть измеряет ли он качество, для измерения которого предназначен, у репрезентативной выборки. Чтобы поддержать содержательную валидность теста, необходимы его регулярные проверки на соответствие, так как реальная картина проявления определённого качества может меняться у выборки с течением времени. Оценка содержательной валидности должна производиться экспертом в предметной области теста.



- Процесс валидизации теста должен представлять собой не сбор доказательств его валидности, а комплекс мер по повышению этой валидности. Большинство процедур анализа заданий предполагают: а) регистрацию числа испытуемых, давших правильный или неправильный ответ на определенное задание; б) корреляцию отдельных заданий с др. переменными; в) проверку заданий на систематическую ошибку (или «необъективность»). Долю испытуемых, справившихся с заданием теста, наз., возможно не вполне точно, трудностью задания. Способ улучшить задания — подсчитать процент выбора каждого варианта ответа на задание с множественным выбором; полезно также вычислить средний тестовый показатель испытуемых, выбравших каждый вариант. Эти процедуры позволяют контролировать, чтобы варианты ответов выглядели правдоподобными для неподготовленных испытуемых, но не казались правильными наиболее знающим. Отбор заданий, которые сильно коррелируют с показателем полного теста, максимизирует надежность как внутреннюю согласованность теста, тогда как отбор заданий, которые сильно коррелируют с внешним критерием, максимизирует его прогностическую валидность. Описательная аналоговая модель этих корреляций называется характеристической кривой задания; в типичных случаях — это график зависимости доли испытуемых, правильно отвечающих на вопрос, от их суммарного тестового показателя. Для эффективных заданий эти графики представляют собой положительные восходящие кривые, не снижающиеся по мере прироста способности.

Спасибо за внимание!