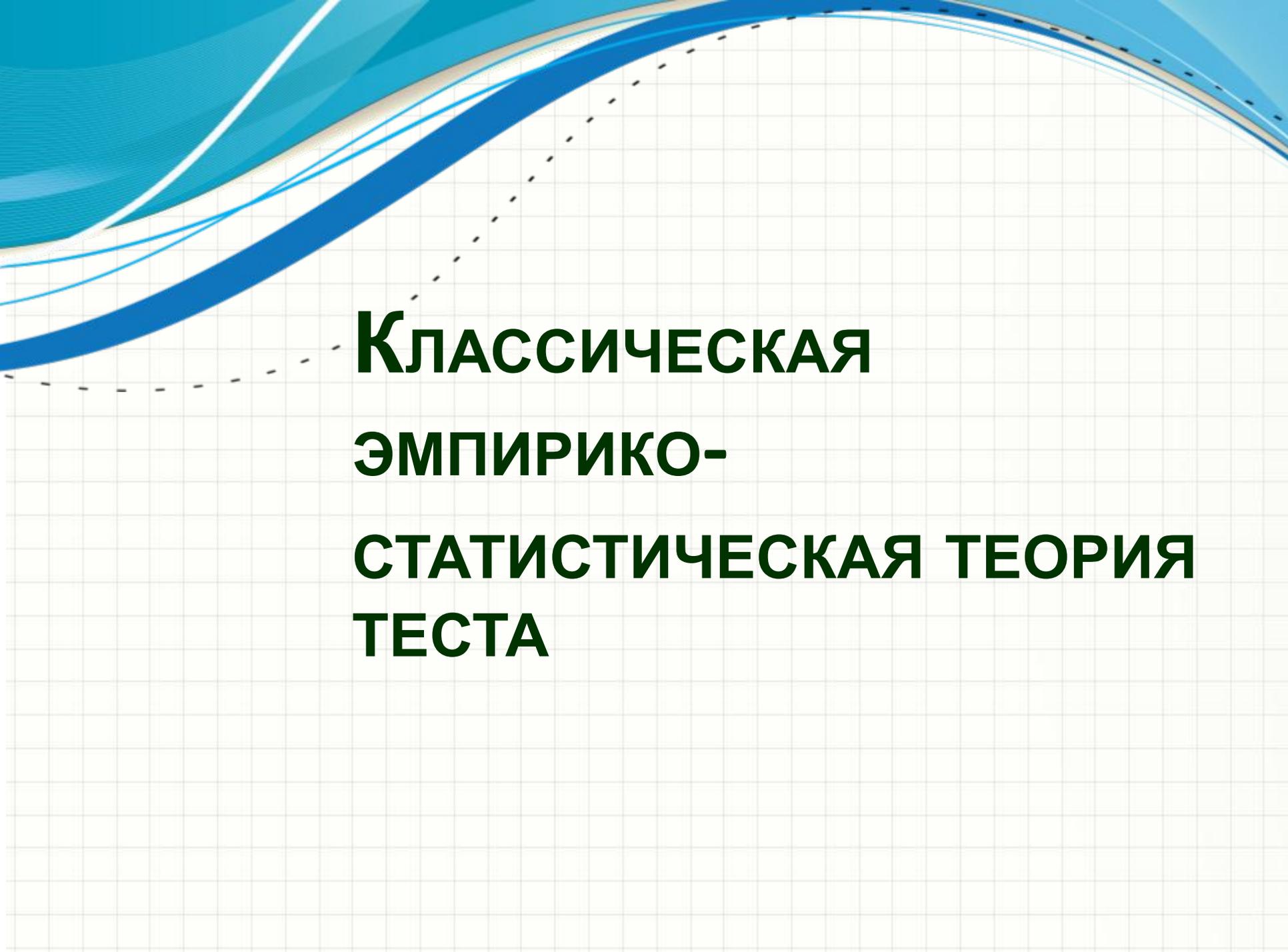




ОБЩИЙ ПСИХОЛОГИЧЕСКИЙ ПРАКТИКУМ: ТЕОРИЯ ИЗМЕРЕНИЙ

Куликов И.А.

ТГУ 2015



**КЛАССИЧЕСКАЯ
ЭМПИРИКО-
СТАТИСТИЧЕСКАЯ ТЕОРИЯ
ТЕСТА**

Классическая теория теста

лежит в основе современной
дифференциальной психометрики.

Психометрика*

(или психометрия) — совокупность теоретико-математических моделей и процедурно-методических правил организации сбора и обработки эмпирических данных, позволяющих выразить психические свойства и параметры психических процессов в числовой или квазичисловой (ранговой, категориальной) форме.

Без знания психометрических правил современный специалист (не только исследователь, но и практик) не может корректно применять разнообразные измерительные

МЕТОДЫ В ПСИХОЛОГИИ

Краткий психологический словарь. — Ростов-на-Дону: «ФЕНИКС». Л. А. Карпенко, А. В. Петровский, М. Г. Ярошевский. 1998.

Психометрика

```
graph TD; A[Психометрика] --> B[Общая]; A --> C[Дифференциальная];
```

Общая

Устанавливает числовые соответствия между каким-то свойством стимула (например, интенсивностью света, звука) и соответствующим ощущением определенной силы (яркостью, громкостью и т. п.)

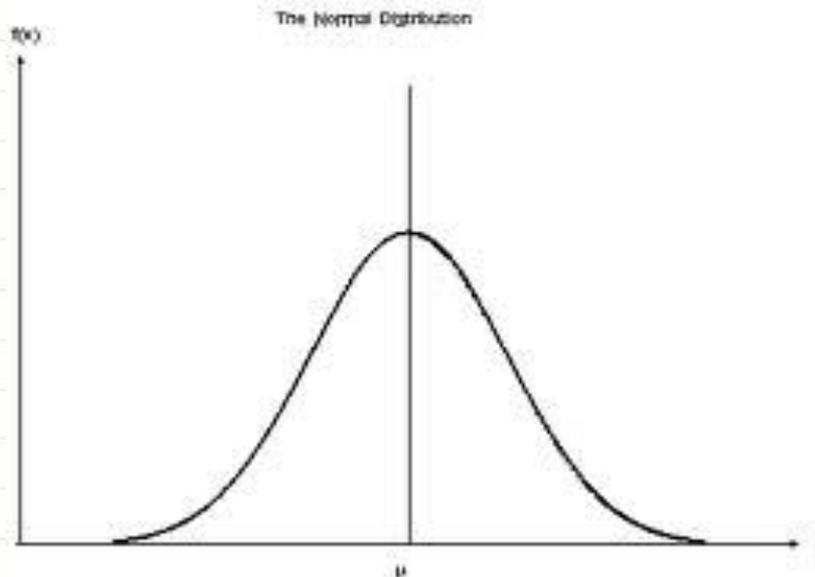
Дифференциальная

Выражает в количественной форме индивидуальные психологические различия между людьми.

1. Основные положения теории психологического тестирования

- *Конструирование тестов для изменения психологических свойств и состояний основано на шкале интервалов.*

Измеряемое психическое свойство считается линейным и одномерным. Предполагается также, что распределение совокупности людей, обладающих данным свойством, описывается кривой нормального распределения.

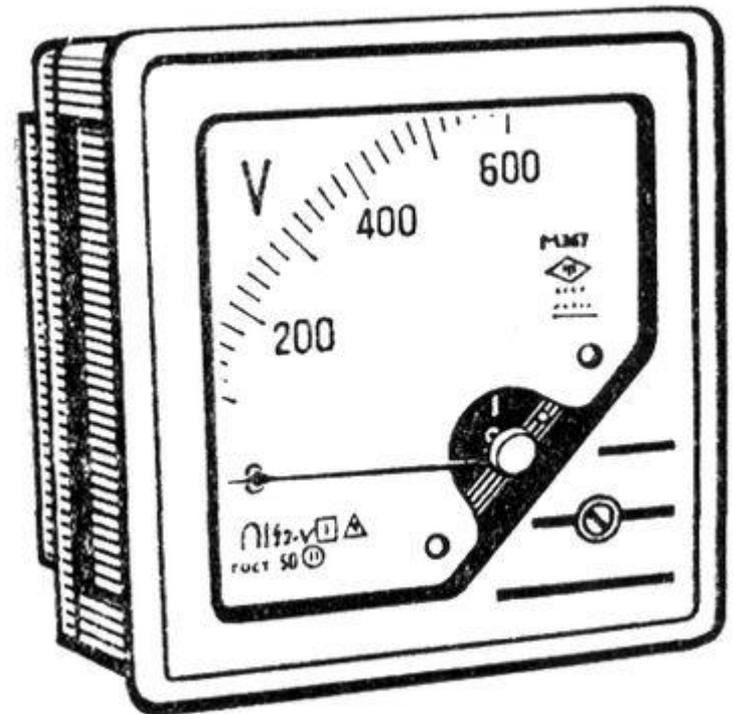


1. Основные положения теории психологического тестирования

- *В основе тестирования лежит классическая теория погрешности измерений; она полностью заимствована из физики.*

Считается, что тест — измерительный прибор. Любое свойство личности имеет «истинный» показатель, а показания по тесту отклоняются от истинного на величину случайной погрешности.

На показания теста влияет и «систематическая» погрешность, но она сводится к прибавлению (вычитанию) константы к «истинной» величине параметра, что для интервальной шкалы значения не имеет.



1. Основные положения теории психологического тестирования

- *Если тест проводить много раз, то среднее будет характеристикой «истинной» величины параметра.*

Отсюда выводится понятие ретестовой надежности: чем теснее коррелируют результаты начального и повторного проведения теста, тем он надежнее

1. Основные положения теории психологического тестирования

- *Предполагается, что существует множество заданий, которые могут репрезентировать измеряемое свойство. Тест есть лишь выборка заданий из их генеральной совокупности.*

В идеале можно создать сколько угодно эквивалентных форм теста. Отсюда — определение надежности теста методами параллельных форм и расщепление его на эквивалентные равные части.

Задания теста должны измерять «истинное» значение свойства. Все задания одинаково скоррелированы друг с другом.

Пример расщепления шкалы (в социологии)

Пример расщепления шкалы для оценки уровня удовлетворенности специальностью

Уровень удовлетворенности специальностью	Пункты четной и нечетной половин испытываемой шкалы
(A) Специальность очень нравится	(a_1) Считаю свою специальность лучше всех других
(B) Специальность скорее нравится, чем не нравится	(a_2) Моя специальность одна из лучших (b_1) Считаю, что моя специальность не хуже многих других (b_2) Считаю, что есть специальности, которые гораздо хуже, чем моя
(C) К специальности относятся безразлично	(c_1) Моя специальность не хуже – не лучше других (c_2) Нравится или не нравится специальность – не знаю
(D) Специальность скорее не нравится	(d_1) Считаю, что есть много специальностей лучше моей (d_2) Моя специальность хуже, чем многие другие
(E) Специальность совершенно не нравится	(e_1) Это одна из худших специальностей (e_2) Большинство других специальностей лучше моей

1. Основные положения теории психологического тестирования

- *Поскольку в реальном монометрическом тесте число заданий ограничено (не более 100), то оценка надежности теста всегда приближительна.*

Так, определяемая надежность теста связана с однородностью, которая выражается в корреляциях между заданиями. Надежность возрастает с увеличением одномерности теста и числа его заданий, причем довольно быстро.

Стандартная надежность 0,02 соответствует тесту длиной в 10 заданий, а при 30 заданиях она равна 0,007

1. Основные положения теории психологического тестирования

- *В принципе классическая теория теста касается лишь проблемы надежности. Вся она базируется на том, что результаты выполнения разных заданий можно суммировать с учетом весовых коэффициентов.*

1. Основные положения теории психологического тестирования

- *Теоретически проблема валидности не решается. Надежность теста является приоритетной.*

Правило: валидность теста не может быть больше его надежности.



1. Основные положения теории психологического тестирования

- *Валидность.*

Валидность означает пригодность теста измерять то свойство, для измерения которого он предназначен.

Следовательно, чем больше на результат выполнения теста или отдельного задания влияет измеряемое свойство и чем меньше — другие переменные (в том числе внешние), тем тест валидней и, добавим, надежнее, поскольку влияние помех на деятельность испытуемого, измеряемую валидным тестом, минимально

1. Основные положения теории психологического тестирования

- *Валидность.*

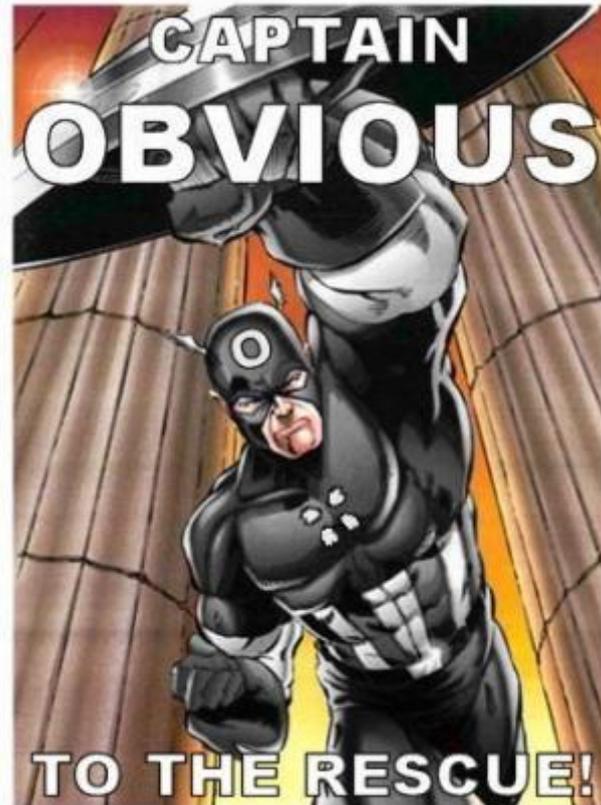
Это противоречит классической теории теста, основанной не на бихевиористской парадигме: стимул—ответ.

Если же рассматривать тестирование как активное порождение испытуемым ответов на задания, то надежность теста будет функцией, производной от валидности:

- Тест валиден (и надежен), если на его результаты влияет лишь измеряемое свойство.
- Тест невалиден (и ненадежен), если результаты тестирования определяются влиянием нерелевантных переменных.

2. Виды валидности тестов:

1. *Очевидная валидность.* Тест считается валидным, если у испытуемого складывается впечатление, что он измеряет то, что должен



2. Виды валидности тестов:

2. Конкретная валидность, или конвергентная—дивергентная валидность. Тест должен хорошо коррелировать с тестами, измеряющими конкретное свойство либо близкое ему по содержанию, и иметь низкие корреляции с тестами, измеряющими заведомо иные свойства.

2. Виды валидности тестов

3. *Прогностическая валидность.* Тест должен коррелировать с отдаленными по времени внешними критериями: измерение интеллекта в детстве должно предсказывать будущие профессиональные успехи.

2. Виды валидности тестов

4. *Содержательная валидность.*

Применяется для тестов достижений: тест должен охватывать всю область изучаемого поведения.

2. Виды валидности тестов

5. *Конструктивная валидность.*

Предполагает:

- а) полное описание измеряемой переменной;
- б) выдвижение системы гипотез о связях ее с другими переменными;
- в) эмпирическое подтверждение (неопровержение) этих гипотез.

2. Валидность теста

С теоретической точки зрения, единственным способом установления «внутренней» валидности теста и отдельных заданий является метод факторного анализа (и аналогичные), позволяющий:

а) выявлять латентные свойства и вычислять значение «факторных нагрузок» — коэффициенты детерминации свойств тех или иных поведенческих признаков;

б) определять меру влияния каждого латентного свойства на результаты тестирования.

В классической теории теста не выявлены причинные связи факторных нагрузок и надежности теста.

3. Дискриминативность теста

Тест должен хорошо «различать» испытуемых с разными уровнями выраженности свойства.

Считается, что больше 9-10 градаций использовать не стоит.

4. Тестовые нормы

полученные в ходе стандартизации, представляют собой систему шкал с характеристиками распределения тестового балла для различных выборок.

Они не являются «внутренним» свойством теста, а лишь облегчают его практическое применение.

Стохастическая теория тестов

опирается на теорию измерения, — *Item Response Theory (IRT)*. Она основывается на теории латентно-структурного анализа (ЛСА), созданной П. Лазарсфельдом и его последователями.

Теория измерения латентных черт предполагает:

1. Существует одномерный континуум свойства — латентной переменной (x); на этом континууме происходит вероятностное распределение индивидов с определенной плотностью $f(x)$.

Теория измерения латентных черт предполагает:

2. Существует вероятностная зависимость ответа испытуемого на задачу (пункт теста) от уровня его психического свойства, которая называется характеристикой кривой пункта. Если ответ имеет две градации («да — нет», «верно — неверно»), то эта функция есть вероятность ответа, зависящая от места, занимаемого индивидом на континууме (x).

Теория измерения латентных черт предполагает:

3. Ответы испытуемого не зависят друг от друга, а связаны только через латентную черту. Вероятность того, что, выполняя тест, испытуемый даст определенную последовательность ответов, равна произведению вероятностей ответов на отдельные задания.

Модели ЛСА

- В модели латентного класса функция плотности распределения индивидов является точно-дискретной: все индивиды относятся к разным непересекающимся классам. Измерение производится при помощи номинальной шкалы.

Модели ЛСА

- В модели латентной дистанции постулируется, что вероятность ответа индивида на пункт текста является мультипликативной функцией от параметров задачи и величины свойства

Модели ЛСА

- Логистическая модель используется наиболее широко, так как она специально предназначена для тестов, где свойство измеряется суммированием баллов, полученных за выполнение каждого задания с учетом их весов.

База для IRT – модель латентной дистанции

- Предполагается, что и индивидов, и задания можно расположить на одной оси «способность — трудность» или «интенсивность свойства — сила пункта». Каждому испытуемому ставится в соответствие только одно значение латентного параметра («способности»).
- В общем виде вероятность ответа зависит от множества свойств испытуемого, но в моделях *IRT* рассматривается лишь одномерный случай.

Отличие IRT от КТТ

- Не ставятся и не решаются фундаментальные проблемы эмпирической валидности и надежности теста: задача априорно соотносится лишь с одним свойством, т.е. тест заранее считается валидным. Вся процедура сводится к получению оценок параметров трудности задания и к измерению «способностей» испытуемых (образованию «характеристических кривых»).
- В классической теории теста индивидуальный балл (уровень свойства) считается некоторым постоянным значением. В *IRT* латентный параметр трактуется как непрерывная переменная.

Модель Г. Раша

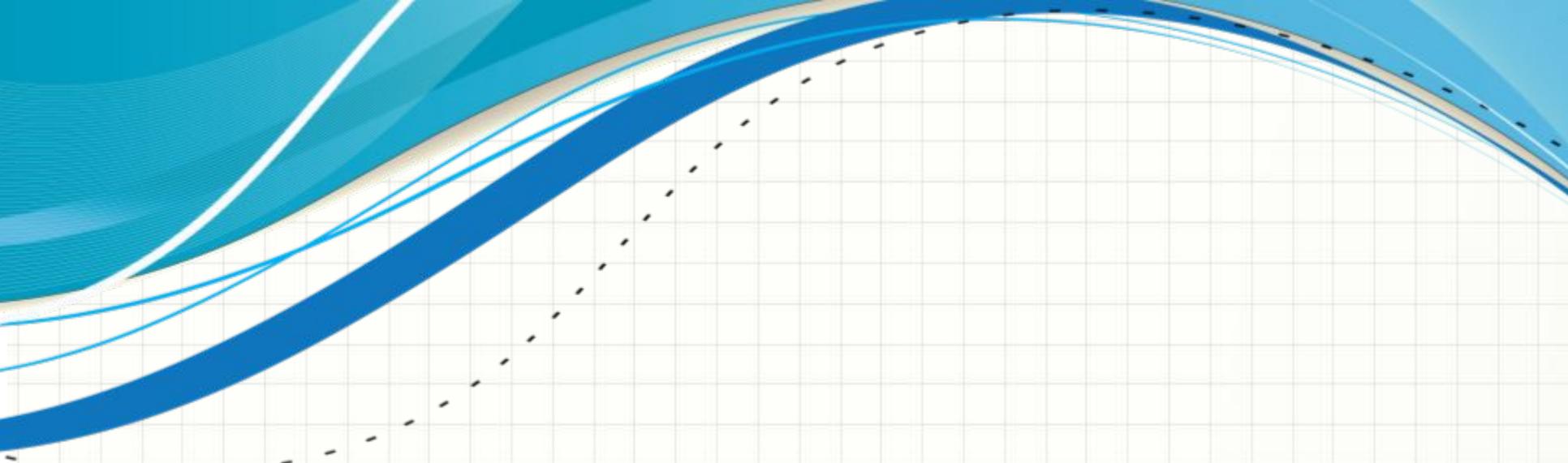
- разность уровня способности и трудности теста $x_i - \beta_j$, где x_i — положение i -го испытуемого на шкале, а β_j — положение j -го задания на той же шкале.
- Расстояние $(x_i - \beta_j)$ характеризует отставание способности испытуемого от уровня сложности задания. Если разница велика и отрицательна, то задание не может быть выполнено, так как для данного испытуемого оно слишком сложно. Если же разница велика и положительна, то задание также не информативно, ибо испытуемый заведомо легко и правильно его решает.

Модель Г. Раша

- В *IRT* функции x и $f(\beta)$ называются функциями выбора пункта. Соответственно первая является характеристической функцией испытуемого, а вторая — характеристической функцией задания.
- Считается, что латентные переменные x и β нормально распределены, поэтому для характеристически функций выбирают либо логистическую функцию, либо интегральную функцию нормированного нормального распределения (как мы уже отметили выше, они мало отличаются друг от друга).
- Поскольку логистическую функцию проще аналитически задавать, ее используют чаще, чем функцию нормального распределения.
- Кроме «свойства» и «силы пункта» (она же — трудность задания) в аналитическую модель *IRT* могут включаться и другие переменные. Все варианты *IRT* классифицируются по числу используемых в них переменных.

Ресурсы

- Дружинин В.Н. Экспериментальная психология: Учебник для вузов. 2-е изд. – СПб.: Питер, 2011.
- Наследов А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. 4-е изд. - СПб: «Речь», 2011.



Вопросы?