

Немного о современной теории создания тестов

- Преимущества современной теории создания тестов.
- Понятие о параллельности вариантов.
- Надежность и валидность.

Item Response Theory (IRT), понимаемая как современная теория создания тестов, в основе которой применение математико-статистических моделей измерения для оценки латентных параметров испытуемых и параметров заданий теста.

Модели педагогических измерений в рамках IRT являются частным примером более общей теории латентно-структурного анализа LSA (Latent Structure Analyses).

Латентно-структурный анализ нацелен на выявление внутренних, скрытых качеств и факторов поведения (деятельности) учащихся посредством математико-статистических моделей измерения и обработки результатов.

Основные преимущества IRT:

- Устойчивость и объективность оценок параметра, характеризующего уровень подготовки испытуемых.
- Устойчивость и объективность оценок параметра трудности заданий, их независимость от свойств выборки испытуемых.
- Возможность измерения уровня подготовленности учащихся и трудности заданий на одной шкале (шкале логитов).

Перечисленные преимущества очень важны для развития процессов тестирования в любой стране. С помощью IRT можно предсказать вероятность правильного выполнения заданий теста любым испытуемым в выборке до предъявления теста учащимся. А также выявить эффективность различных по трудности заданий для тестируемых групп учащихся с различной подготовкой. Преимущества IRT подкреплены соответствующим аппаратом, для которого характерно строгое формализованное представление и анализ эмпирических данных.

Трудность заданий

β_j

β

Уровни подготовки
испытуемых

θ_i

θ

$|\theta_i - \beta_j|$ - это расстояние,
на котором находится
испытуемый с уровнем
подготовки θ_i , от задания
с трудностью β_j .

Большая и отрицательная разность
говорит о том, что ученик не сможет
его выполнить. Большая
положительная разность – очень легкое
задание.

X_{ij}

В зависимости от числа оцениваемых параметров тестируемых различают несколько видов математических моделей, используемых для анализа качества современных тестов.

Однопараметрическая модель Раша (G. Rasch) устанавливает зависимость между уровнем подготовленности испытуемого (θ) и трудностью (β) заданий. Вероятность правильного ответа на задания выражается посредством логистической функции, после введения которой симметрично возникла математическая модель, описывающая вероятность правильного ответа в зависимости от трудности заданий, строятся характеристические кривые. Если используются тест с известными устойчивыми значениями параметра трудности, то задача сводится к оценке значений уровня подготовленности испытуемых.

$$P_j(\theta) = \frac{e^{1,7(\theta - \beta_j)}}{1 + e^{1,7(\theta - \beta_j)}};$$

$$P_i(\beta) = \frac{e^{1,7(\theta_i - \beta)}}{1 + e^{1,7(\theta_i - \beta)}};$$

где θ и β - независимые переменные для первой и второй функции.

График функции P_j – характеристическая функция задания, график функции P_i – индивидуальная кривая испытуемого.

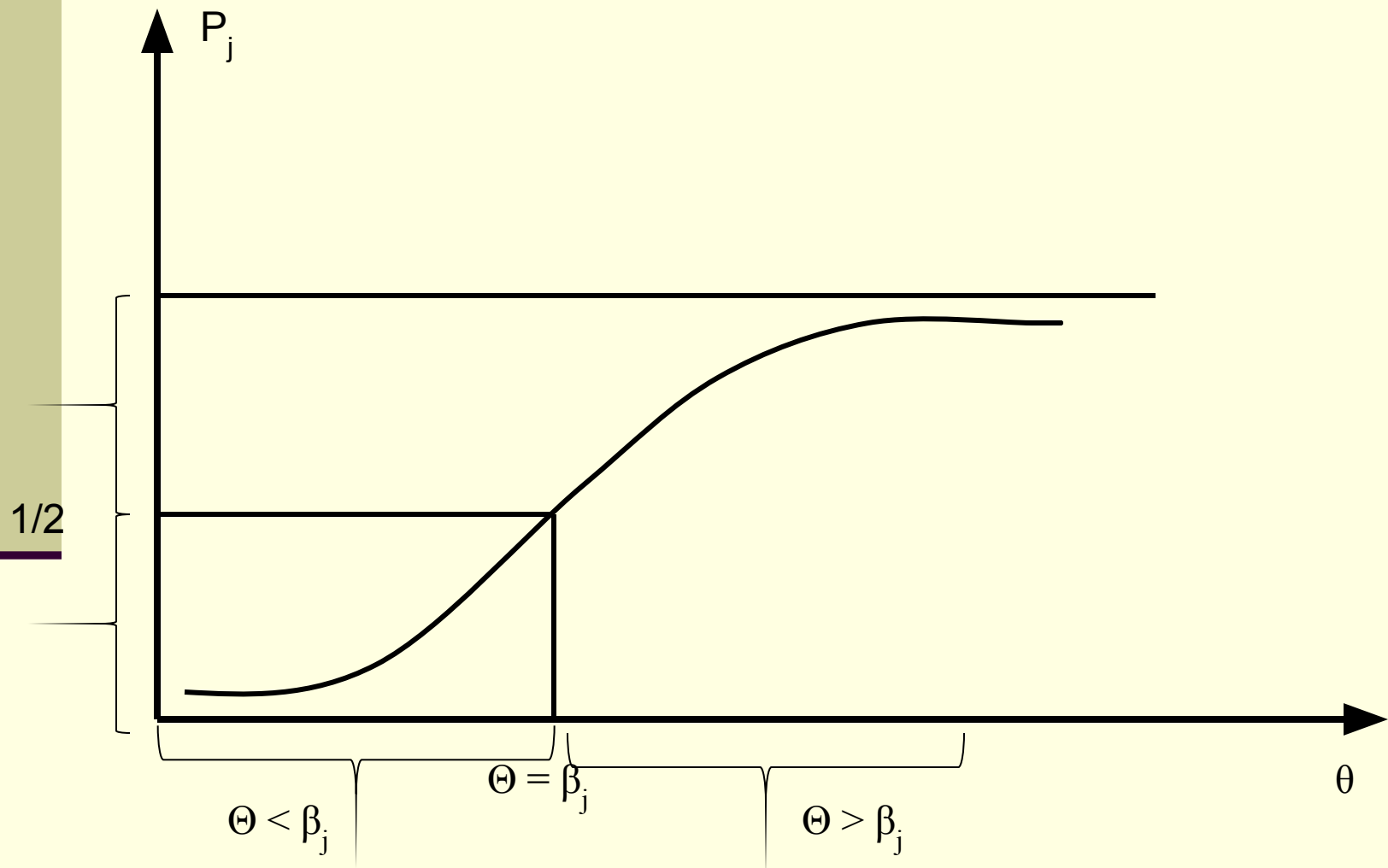
В предположении о нормальном распределении значений латентных переменных θ и β , получаем две функции: $\Psi(x)$ – логистическая кривая, $\Phi(x)$ – интегральная функция нормированного нормального распределения.

Поскольку для одних и тех же значений x ординаты точек графиков функций $\Phi(x)$ и $\Psi(1,7x)$ отличаются друг от друга достаточно мало

$$|\Phi(x) - \Psi(1,7x)| < 0,01,$$

то для практических задач предпочитают логистическую функцию $\Psi(1,7x)$.

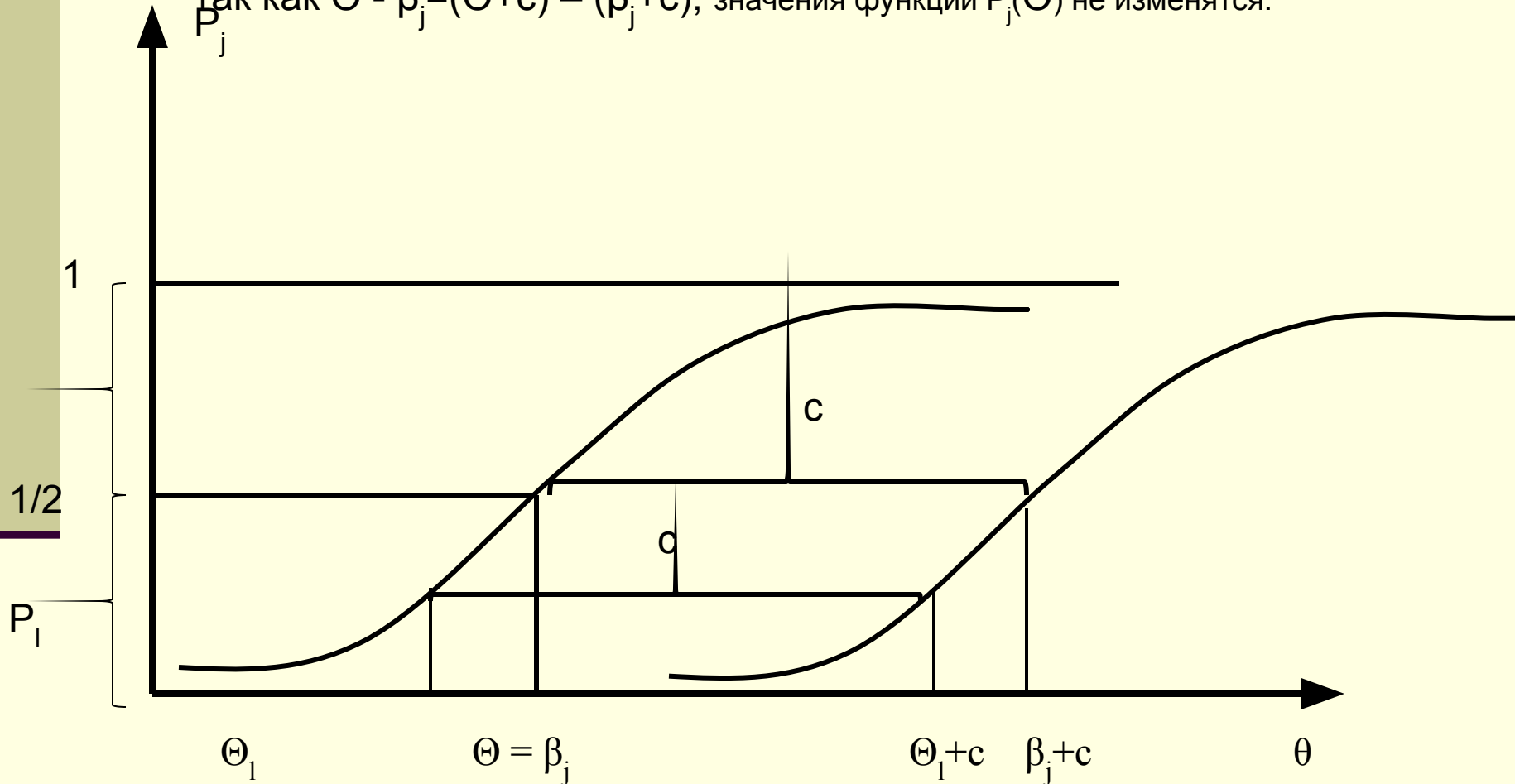
Характеристическая кривая для j-го задания теста



Характеристическая кривая для j-го задания теста

Увеличение трудности задания на константу c вызовет смещение кривой вправо. С прежней вероятностью на это более трудное задание будут отвечать испытуемые с уровнем подготовки $\Theta_1 + c$,

так как $\Theta - \beta_j = (\Theta + c) - (\beta_j + c)$, значения функции $P_j(\Theta)$ не изменятся.



Двухпараметрическая модель А. Бирнбаума (A. Birnbaum) позволяет добавить для анализа еще один параметр – дифференцирующую способность, позволяющая оценить меру структурированности знаний учащегося. Для трехпараметрической модели учитывается вероятность угадывания. Но из-за сложности обработки трехпараметрическую модель почти не используют.

Используемые модели IRT позволяют повысить точность измерений и оптимизировать процедуру контроля за счет адаптации теста к уровню подготовленности учащегося.

Возможность объединить на одной шкале измерения уровня подготовленности учащихся и трудности заданий серьезное преимущество IRT перед классической теорией.

Двухпараметрическая модель А.Бирнбаума

$$P_j(\theta) = \frac{e^{1,7a_j(\theta - \beta_j)}}{1 + e^{1,7a_j(\theta - \beta_j)}};$$

$$P_i(\beta) = \frac{e^{1,7a_i(\theta_i - \beta)}}{1 + e^{1,7a_i(\theta_i - \beta)}};$$

А. Бирнбаум ввел параметр a_j для характеристики дифференцирующей способности задания при изменении различных значений θ , параметр a_i указывает на меру структурированности знаний ученика.

Проблема параллельных вариантов

Два варианта итогового теста можно считать параллельными, если задания проверяют достижение одинаковых требований обязательной подготовки и имеют одинаковые статистические характеристики. Иначе говоря, тесты считают параллельными, если они построены на основе одной спецификации, имеют одинаковое количество заданий попарно равной трудности с совпадающими характеристиками.

В зарубежной практике пользуются понятием «эквивалентные формы» с таким объяснением: *«любая из двух или более форм теста со сходным содержанием и сложностью, которые дают для данной группы очень сходные средние оценки и меры изменчивости».*

Показатели качества теста в целом в рамках классической теории тестов

- *Надежность* отражает точность тестовых измерений и устойчивость тестовых результатов к действию случайных факторов. Данная характеристика имеет два значения: точность измерения означает устойчивость к действию помех, всевозможных случайностей при проведении процедуры тестирования. Уменьшить случайные факторы можно, как говорилось ранее, грамотно составленной инструкцией, устойчивым тест считают, если получены близкие по значению результаты при повторном тестировании на параллельной форме.

Формула для расчета надежности, через дисперсию баллов

$$r_n = 1 - \frac{S_E^2}{S_x^2}$$

На надежность влияют:

- 1. Внутренняя согласованность теста (содержательная однородность)*
- 2. Длина теста.*

Формула Кьюдера-Ричардсона (KR-20)

$$(KR - 20) = \frac{k}{k - 1} \left(1 - \sum \frac{PQ}{S_y^2} \right)$$

Где P-доля учащихся, верно выполнивших задание,

Q=1-P; k – количество заданий, S_y^2 квадрат стандартного отклонения

Валидность – способность теста служить поставленной цели измерения

Виды:

- содержательная;
- критериальная;
- конструктивная.

Способы оценки валидности:

подсчет корреляции с внешним критерием;
экспертиза.

1. Информация это –

А) отображение реального мира с помощью знаков и сигналов

Б) то, что ученик узнал на уроке или можно узнать из газет, радио, телевидения

В) важные для человека сведения

Г) наблюдения, записанные и сохраненные на каких-либо носителях.

2. Виды информации, которые человек получает с помощью органов чувств называется:

А) органической;

Б) органолептической;

В) чувственной;

Г) чувствительной.

**Информацию, приводящую к
правильному пониманию и принятию
верного решения называют:**

- А) Полной
- Б) Понятной
- В) Актуальной
- Г) Достоверной

Под термином «канал связи» в информатике понимают:

- А) техническое устройство, обеспечивающее кодирование сигнала при передаче его от источника информации к приемнику информации
- Б) физическую линию (прямое соединение), телефонную, телеграфную или спутниковую линию связи и аппаратные средства, используемые для передачи данных (информации);
- В) устройство кодирования и декодирования информации при передаче сообщений;
- Г) совокупность технических устройств, обеспечивающих прием информации

При передаче информации в обязательном порядке предполагается наличие:

- А) источника и приемника информации, а также канала связи между ними;
- Б) двух людей;
- В) дуплексного канала связи;
- Г) избыточности передающейся информации.

Чему равен 1 байт?

- А) 12 бит
- Б) 8 бит
- В) 1 бит
- Г) 1 мегабайт

Процесс преобразования информации в машинный код называется:

- А) кодированием
- Б) декодированием
- В) перекодированием
- Г) научного определения для данного процесса нет

К какому информационному процессу относится измерение параметров окружающей среды (температуры воздуха, атмосферного давления, скорости ветра и т. п.)?

Ответ: получение.