***Калдыбаев С.К.***

# ВЕРОЯТНОСТНЫЕ МОДЕЛИ ТЕОРИИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Исследование по теории и методологии педагогических измерений показало наличие ряда недостатков в статистической ее теории [1; 2; 4; 5].

1. Оценка трудности заданий педагогического теста зависит от конкретной выборки испытуемых. Чем выше общий уровень подготовленности испытуемых в выборке, тем больше верно выполненных заданий теста, что предполагает особые требования к созданию выборки.

2. Чем выше первичный балл испытуемого, тем выше его уровень подготовленности.

3. Задания с меньшими дисперсиями (близкие к нулю) должны быть удалены из состава теста. Но коэффициент корреляции иногда показывает достаточно сильную связь задания с тестом. Следовательно, во многих случаях выявляются несколько искаженные результаты.

4. В основном используется шкала порядка, где заложен принцип ранжирования объектов измерений в порядке возрастания (или убывания). Знания и уровень подготовленности испытуемых оцениваются с точки зрения отношения равенства между ними или отношения «больше-меньше», а расстояния между ними, т.е. насколько один испытуемый лучше другого, не имеют значения.

Идеи создания интервальной шкалы для измерения уровня подготовленности обучаемых и геометрических образов задания стимулировали научный поиск ученых исследователей. Этими проблемами занимались зарубежные исследователи А.Бирнбаум, В.Фергюссон, Н.Галликсен, Ф.Лорд, Д.Лоули, Г.Раш, М.Ричардсон и др.

В рамках методологии латентно-структурного анализа (LSA) были созданы и закреплены первые модели измерения IRT (Item Response Theory) как структурные построения, позволяющие выявить отношение между латентными и наблюдаемыми переменными [1; 2; 3; 6].

Свой систематизированный вид IRT получил в результате исследований датского ученого-математика Георга Раша. Эмпирическое исследование позволило Г.Рашу выразить в 1958 году функцию логистического вида:

$$P_j\left\{x_{ij}=1\middle|\beta_j\right\}=\frac{e^{(\theta-\beta_j)}}{1+e^{(\theta-\beta_j)}} \quad (1)$$

где, $x_{ij}$ө1, если испытуемый правильно выполняет j –тое задание, θ – уровень знаний, $\beta_j$ – уровень трудности j – го задания.

Функция (1) названа вероятностью правильного ответа любого испытуемого на задание j.

В дальнейшем Г.Рашем поставлен и другой вопрос: какова вероятность правильного ответа некоторого испытуемого с уровнем знаний $\theta_i$ на задание различного уровня трудности заданий? В ответ появилась следующая логистическая функция:

$$P_i\left\{x_{ij}=1\middle|\theta_i\right\}=\frac{e^{(\theta_i-\beta)}}{1+e^{(\theta_i-\beta)}} \quad (2)$$

Функции (1) и (2) взаимосвязаны, поскольку устанавливается связь между латентными переменными (параметры) θ и β. Во взаимодействиях этих переменных порождаются наблюдаемые результаты. А через наблюдаемые результаты ставится задача оценить значения параметров θ и β. Отношения между ними выражаются через их разность θ-β. Как оказалось, в этой разности кроется множество ценных информаций.

1. Казалось бы, уровень подготовленности испытуемого $\theta_i$ и уровень трудности $\beta_j$ несравнимы. Если в качестве основания брать внутренние свойства (знания испытуемого и изоморфные им элементы знаний в тестовых заданиях), то можно ставить вопрос об их сравнении. Они станут сравнимыми в единицах, так называемой, интервальной шкале логита.

2. Разность $|\theta_i-\beta_j|$ означает расстояние между уровнем знаний i – того испытуемого и уровнем трудности j – того задания.

3. Если $\theta_i$ ө $\beta_j$, совпадают, то вероятность $P_{ij}$ ө $\frac{1}{2}$ (равным образом испытуемый может выполнить или не выполнить данное задание).

4. Если $\theta_i-\beta_j>0$, то по формуле (1) вероятность $P_{ij}>\frac{1}{2}$. Если эта разница намного больше, то значение $P_{ij}$ станет ближе к единице. Это означает, что испытуемый обладает большими способностями, он может успешно вы-

полнить данное задание.

5. Если разность отрицательна, то значение $P_{ij} < \frac{1}{2}$. Если разность имеет большее отрицательное значение, то $P_{ij}$ стремится к нулю. Это означает, что испытуемый со слабым знанием, вероятно, не сможет решить задание j.

Функции (1) и (2) могут быть изображены графически, по которым можно произвести визуальную интерпретацию об уровнях знаний испытуемого и об уровне трудности задания. График функции $P_j$ носит название характеристической кривой задания j, а график функции $P_i$ – индивидуальной кривой i-го испытуемого.

На рисунке 1 изображена характеристическая кривая задания j.

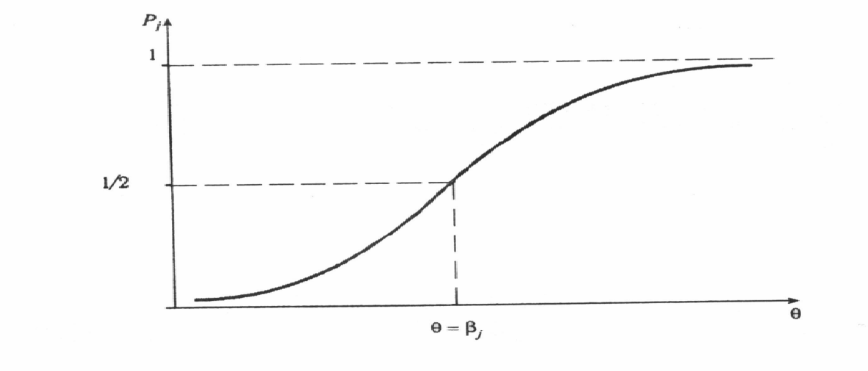


***Рис. 1. Характеристическая кривая задания j.***

Значения θ колеблется от $-\infty$ до $+\infty$, а значения $P_j$: $0<P_j<1$. По рисунку видно, что при $\theta ϴ \beta_j$, вероятность $P_j ϴ \frac{1}{2}$. При $\theta<\beta_j$, $P_j$ стремится к 0, а при $\theta>\beta_j$ $P_j$ стремится к 1. Чем ниже уровень знаний испытуемых, тем меньше вероятность правильного ответа на задание j.

На рисунке 2 изображена индивидуальная кривая i-го испытуемого:

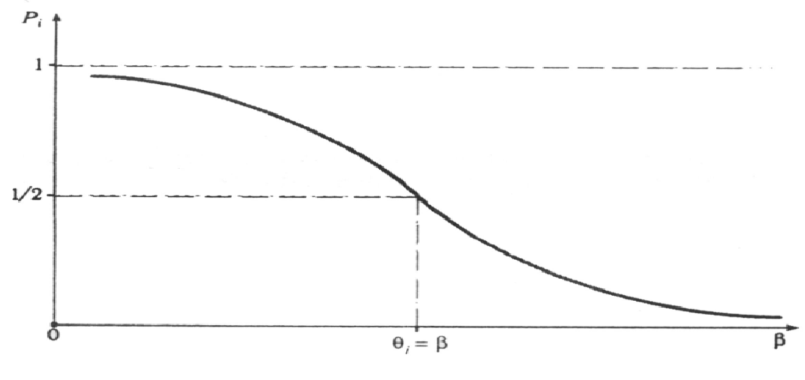


***Рис. 2. Индивидуальная кривая испытуемого i.***

Рисунок информирует, что, чем труднее задания, тем меньше вероятность испытуемого с уровнем знаний $\theta_i$.

Г.Рашем было установлено еще одно утверждение. $P_{ij}$ – вероятность правильного ответа i-го испытуемого на j-е задание. $Q_{ij}$ – вероятность того же испытуемого допустить ошибку в выполнении того же задания. Отношение $\frac{P_{ij}}{Q_{ij}}$ означает шанс i-го испытуемого ответить на задание j правильно.

Подставляя значения $P_{ij}$ и $Q_{ij}$, получаем:

$$\frac{P_{ij}}{Q_{ij}} = \frac{e^{(\theta_i-\beta_j)}}{1+e^{(\theta_i-\beta_j)}} : \frac{1}{1+e^{(\theta_i-\beta_j)}} = e^{(\theta_i-\beta_j)} \quad (5)$$

Сравнение шансов двух испытуемых на задание j также дает полезную информацию:

$$\frac{e^{(\theta_1-\beta_j)}}{e^{(\theta_2-\beta_j)}}=\frac{e^{\theta_1}}{e^{\beta_j}}:\frac{e^{\theta_2}}{e^{\beta_j}}=e^{(\theta_1-\theta_2)} \quad (6)$$

Графически можно это изобразить следующим образом:

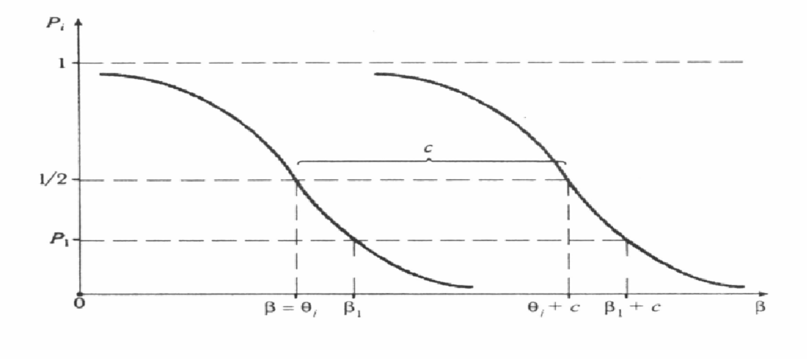


***Рис. 3. Индивидуальные кривые двух испытуемых.***

Сравнивая шансы одного испытуемого ответить на два задания, Г.Раш пришел к выводу, что $\frac{e^{(\theta_i-\beta_1)}}{e^{(\theta_i-\beta_2)}}=e^{(\beta_1-\beta_2)}$, т.е. шанс i-го испытуемого ответить правильно на два задания зависит от трудности этих заданий.

На рисунке 4 изображены характеристические кривые двух заданий.

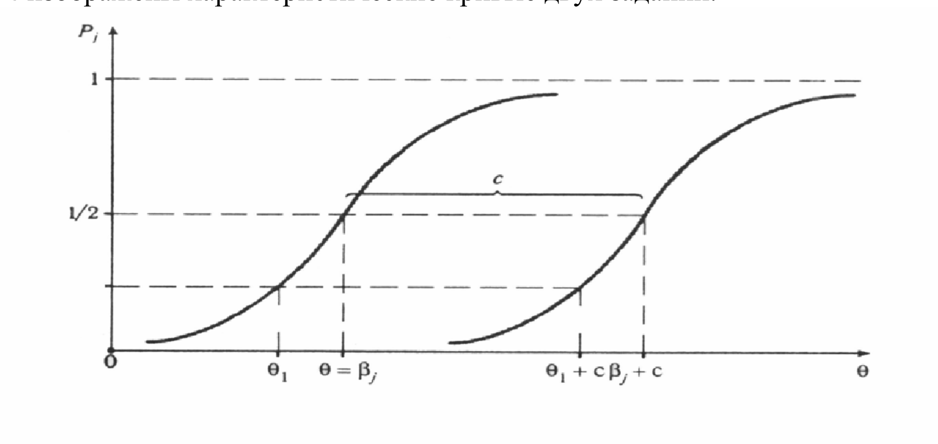


***Рис. 4. Характеристические кривые двух заданий.***

Дальнейшее исследование по улучшению модели Г.Раша было связано с характером и особенностью самих заданий, включенных в тест. Иногда при анализе эмпирических данных могут оказаться тестовые задания с одинаковым уровнем трудности. Но их крутизна кривой с параметром $a_j$ может быть разной. Данный параметр прямо пропорционален тангенсу угла наклона касательной к характеристической кривой задания в точке $\theta\theta\beta_j$. Из этого следует, что более крутые кривые (когда угол между касательной наклона и осью абсцисс больше, чем $45^0$) соответствуют большим значениям $a_j$, пологие кривые (угол наклона меньше, чем $45^0$) – малым значениям $a_j$. Соответственно, вероятность двух испытуемых с уровнем знаний $\theta_1$ и $\theta_2$ правильно выполнить задание с большим $a_j$ имеет существенную разницу, когда вероятность этих же испытуемых ответить на задание с малым $a_j$ близки между собой. Это видно из рисунка [6, с. 281]:

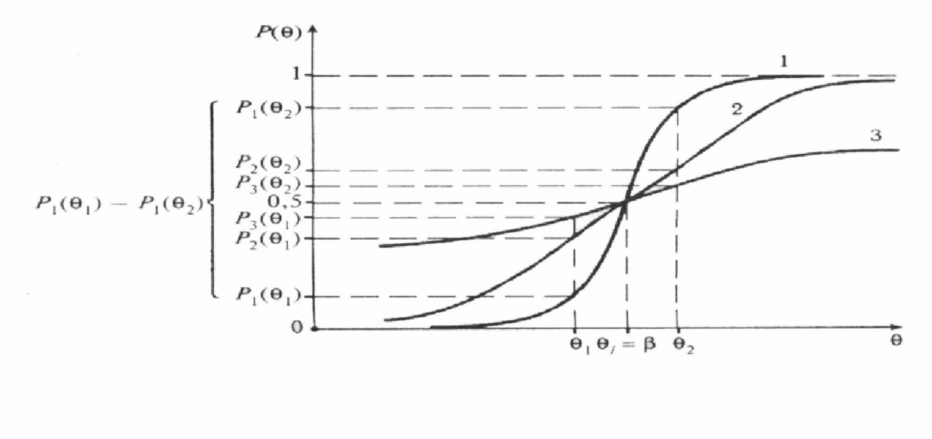


***Рис. 5. Характеристические кривые трех заданий одинаковой трудности.***

Исследованием ученых предложены следующие классы математических моделей IRT [1; 2; 6]. Причем основанием для деления на классы явилось число входящих в нее параметров.

1. Однопараметрическая модель Г.Раша:

$$P_j\left\{x_{ij} = 1 \middle| \beta_j\right\} = \frac{e^{1,7(\theta-\beta_j)}}{1+e^{1,7(\theta-\beta_j)}} \quad (7)$$

где θ – независимая переменная, $\beta_j$ – параметр уровня трудности задания j;

$$P_i\left\{x_{ij} = 1 \middle| \theta_i\right\} = \frac{e^{1,7(\theta_i-\beta)}}{1+e^{1,7(\theta_i-\beta)}} \quad (8)$$

где β – независимая переменная, $\theta_i$ –уровень знаний i-го испытуемого.

3. Двухпараметрическая модель А.Бирнбаума:
4.

$$P_j\left\{x_{ij} = 1 \middle| \beta_j, a_j\right\} = \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}} \quad (9)$$

где θ – независимая переменная, $\beta_j$ – параметр уровня трудности задания j, $a_j$ – параметр крутизны задания j;

$$P_i\left\{x_{ij} = 1 \middle| \theta_i, a_i\right\} = \frac{e^{1,7a_i(\theta_i-\beta)}}{1+e^{1,7a_i(\theta_i-\beta)}} \quad (10)$$

где β – независимая переменная, $\theta_i$ – уровень знаний i-го испытуемого, $a_i$ – мера структурированности знаний i-го испытуемого.

3. Трехпараметрическая модель А.Бирнбаума:

$$P_j\left\{x_{ij} = 1 \middle| \beta_j, a_j, c_j\right\} = c_j + (1-c_j)\frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}} \quad (11)$$

где $\beta_j$ – параметр уровня трудности задания j, $a_j$ – параметр крутизны задания j, $c_j$ – вероятность угадывания правильного ответа в заданиях с выбором правильного ответа. Если имеется 5 ответов для выбора, то $c_j$ө$\frac{1}{5}$ө0,2, если имеется 4 ответа для выбора, то $c_j$ө0,25, и т.д.

Математические модели педагогического измерения сыграли существенную роль в решении вопроса моделирования педагогического теста и точности измерения уровня подготовленности испытуемых. В свою очередь, математические модели оказались способными для преодоления тех трудностей и недостатков классической теории педагогических измерений.

**Литература:**


1. Аванесов В.С. ITEM RESPONSE THEORY: Основные понятия и положения. http://testolog.narod.ru
2. Аванесов В.С. Методологические и теоретические основы тестового педагогического контроля. Дисс. … д-ра пед. наук: 13.00.01. –Москва, 1994. -339с.
3. Звонников В.И., Найденова Н.Н., Никифоров С.В., Челышкова М.Б. Шкалирование и выравнивание результатов педагогических измерений. –М., 2003. -96с.
4. Карданова Е.Ю. Преимущества современной теории тестирования по сравнению с классической теории тестирования //Вопросы тестирования в образовании, №10, 2004. –с.7-34.
5. Орлов.А.И. Теория измерений и педагогическая диагностика// Педагогическое измерение, №1, 2004. -с.22-51.
6. Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. –М.: Логос, 2002. -432с.