

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ДЛЯ РАЗРАБОТКИ ТЕСТОВ

Бул макалада жогорку окуу билим берүүдө компьютердик маалыматтарды кантип пайдаланууга керек экенин жана негизги маселелерди чечүүгө статистикалык ыкмалардын пайдаланышы көрсөтүлгөн.

В данной статье рассмотрены важнейшие задачи процесса вхождения высшей школы в мировое образовательное пространство, требующие совершенствования, а также серьезную переориентацию компьютерно-информационной составляющей, в котором информационный взрыв породил множество проблем, важнейшей из которых является проблема обучения.

Abstract: In this connection, the major problem of a present situation becomes process of the higher school in world educational space which demands perfection, and also serious reorientation of a computer-information component. Information explosion has generated set of problems, major of which is the problem of training.

Последние годы уходящего тысячелетия характеризовались сменой представлений о трансформации индустриального общества в постиндустриальное, новыми концептуальными декларациями построения информационного общества. В нем информация становится одним из основных экономических ресурсов и структурирующим социальным фактором, под воздействием которого меняются формы экономической деятельности, виды и типы предприятий и организаций, социальные взаимоотношения. Сохранение, развитие и рациональное использование этого стратегического ресурса будущего имеет огромное значение для любого общества и государства.

Информационные методы и технологии все глубже проникают практически во все виды деятельности. Информатизация, конвергенция информационных, коммуникационных технологий (ИКТ) и мультимедиа, переход к применению современных информационных систем в сфере науки и образования обеспечивают принципиально новый уровень получения и обобщения знаний, их распространения и использования.

Общепризнанно, что наиболее эффективно информационно-библиотечное обслуживание для нужд науки, культуры и образования сегодня (и, тем более завтра) достигается путем создания электронных библиотек (digital libraries), не столько состоящих из коллекций электронных документов (как оцифрованных традиционных материалов, так и созданных сразу в электронном виде), сколько представляющих из себя систему, реализующую унифицированный подход к

производству, хранению и организации разнообразной информации с целью поиска, анализа и доступа к ней с использованием глобальных компьютерных сетей.

Применение электронного учебника можно организовать как обычную книгу. В начале учебника идет описание навигационных средств учебника, таких, как кнопки перехода на страницы, кнопки вызова помощи, содержания и оглавления. Затем на странице “содержание” содержатся пункты основных разделов материала, представленного в учебнике, в виде “горячих” слов, нажатие мышкой на которые происходит перемещение на страницу, указанную в содержании. В конце каждого раздела помещаются контрольные вопросы или различного рода тестовые задания по пройденному материалу, оценка ответов, которые показывают, насколько максимально был усвоен предложенный материал. В конце электронного пособия возможно также обобщить все полученные оценки и выставить итоговую по пройденному курсу. В зависимости от требований, предъявляемых к каждому конкретному студенту, т.е. от степени интеллектуального развития, психологической устойчивости и различных других личностных факторов, преподаватель или ассистент курса могут предложить повторить те разделы, по которым оценка может считаться неудовлетворительной. На этих основаниях мы должны предлагать более подходящие тесты к пройденному материалу.

Существуют различные виды тестов. Так, для тестирования можно применить так называемую закрытую форму тестов. Пользователю представлен вопрос или утверждение, а также варианты ответов. Причем количество самих ответов и правильных из них может быть неограничено.

Тест может быть охарактеризован как эффективный при условии, что он удовлетворяет определенным требованиям. Основные требования следующие:

- надежность;
- валидность;
- дискриминативность.

Обсудим и определим все те термины, в которых следует досконально разобраться, чтобы тесты были не только правильно сконструированы, но и соответствующим образом использовались.

Надежность теста – это характеристика методики, отражающая точность психодиагностических измерений, а также устойчивость результатов теста к действию посторонних случайных факторов. При этом надежность бывает нескольких видов:

1) ре-тестовая надежность – когда рассматриваются показатели при повторном исследовании испытуемых с помощью одного и того же теста по прошествии времени;

2) надежность частей теста – получается путем анализа устойчивости результатов отдельных совокупностей тестовых задач или отдельных частей теста.

Ре-тестовая надежность. Когда тест не может дать тот же самый результат для некоторого испытуемого (при условии, что этот испытуемый не изменился) в различных условиях, – значит, не все в порядке. Способ измерения ре-тестовой надежности очень прост. Вычисляется корреляция показателей для выборки испытуемых, протестированных в двух случаях. Удовлетворительным наименьшим значением для ре-тестовой надежности является 0,7. Указанный предельный

коэффициент надежности в известной мере условен. Для проективных и некоторых других тестов личности показатель ре-тестовой надежности может быть ниже, при этом диагностическая ценность методики не снижается.

При более низком значении использование теста становится вряд ли целесообразным, поскольку стандартная погрешность получаемых по нему показателей будет настолько велика, что интерпретация показателей станет сомнительной.

Коэффициент надежности. Средняя величина корреляции одного теста или задания со всем тестами или заданиями из генеральной совокупности называется коэффициентом надежности. Квадратный корень из коэффициента надежности является корреляцией данного теста или задания с истинным показателем. Однако на практике невозможно точно вычислить это теоретическое значение надежности \bar{r}_{ii} , потому что количество разработанных нами заданий и тестов не является бесконечным. Это означает, что надежность r_{ii} некоторого теста можно оценить лишь приблизительно.

Таким образом, на практике коэффициенты надежности основаны на корреляции одного теста с другими, и эта оценка может быть не очень точной. Это означает, что имеющая более существенное значение корреляция теста или задания с истинным показателем тоже может быть оценена неточно.

Надежность и величина теста. Таким образом, надежность возрастает с величиной теста. С точки зрения разработчика тестов важной является быстрота возрастания надежности с возрастанием количества заданий. Всегда трудно разработать большое количество валидных заданий (например, таких, которые принадлежат именно нужной генеральной совокупности); следовательно, если мы хотим продемонстрировать, что скажем, надежность двадцати пяти заданий (с заданной средней корреляцией) является высокой, то достижение этой цели будет иметь смысл.

Коэффициент ассоциации для оценки надежности теста

Тесноту связи между качественными признаками X и Y измеряют с помощью коэффициента ассоциации. Где X – вид теста, а Y – результаты тестирования. В простейшем виде формула, по которой рассчитывается этот показатель выглядит следующим образом:

$$r_a = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \quad (1)$$

где a, b, c, d – численности коррелируемых групп.

Коэффициент ассоциации, как и пирсоновский коэффициент корреляции, изменяется от -1 до +1. Значимость r_a можно проверить с помощью t-критерия Стьюдента. Нулевую гипотезу, которая сводится к предложению, что в генеральной совокупности этот показатель r_a равен нулю, отвергают, если

$$t_f = \frac{r_a \sqrt{n-2}}{\sqrt{1-r_a^2}} \geq t_{st}, \quad (2)$$

где n – количество тестируемых, r_a – коэффициент ассоциации, t_{st} – t-критерий Стьюдента, для принятого уровня значимости (α) и числа степеней свободы $k = n - 2$.

Так как коэффициент ассоциации имеет прямое отношение к пирсоновскому критерию χ^2 , на котором он основан, то распределение вероятных значений критерия χ^2 является непрерывным. Качественные же признаки дискретны, их числовое значение не распределяются непрерывно. Учитывая эту особенность, в формулу (1) принято вносить поправку Йейтса на непрерывность вариации, равную половине объема выборки. И формула (1) принимает следующий вид:

$$r_a = \frac{|ad - bc| - 0.5n}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \quad (3)$$

Валидность. Кратко рассмотрим природу валидности, второй из основных характеристик эффективных тестов. Тест называется валидным, если он измеряет то, для измерения чего он предназначен. Однако такое определение не разъясняет удовлетворительно значения валидности. В этом случае возникает новый вопрос: как мы узнаем, что тест измеряет то, для чего он предназначен? В действительности, существует много различных способов доказательства валидности тестов, и каждый из них соответствует разным аспектам этого значения. Говорят, что тест является, очевидно, валидным, если о нем складывается впечатление, что он измеряет именно то, что подразумевается, особенно с точки зрения испытуемых

Конкурентная валидность – эта валидность оценивается по корреляции результатов данного теста с результатами других тестов. Так, если мы пытаемся установить конкурентную валидность некоторого теста интеллекта, мы будем изучать его корреляцию с другими тестами, валидность которых установлена.

Содержательная валидность. Этот термин применяется, в основном, по отношению к тестам достижений и может быть просто объяснен следующим образом. Если можно показать, что задания теста отражают все аспекты исследуемой области поведения, то тест является, по существу, валидным, при условии, что инструкции изложены ясно. Содержательная валидность не сводится к простой очевидной валидности, которая связана с внешним видом заданий теста. Если в тесте математических навыков мы хотим протестировать умение перемножать выражения в скобках и имеем задания вида $(y + 2k)(2y - 3x) = ?$, то трудно оспаривать валидность этого задания. Очевидно, содержательная валидность полезна только для тех тестов, для которых, как в данном случае, смысл измеряемого параметра полностью ясен.

Были описаны разнообразные способы установления валидности тестов, некоторые из них коренным образом отличаются друг от друга. Из обсуждения должно быть также понятно, что не

может быть какого-либо единственного показателя, демонстрирующего валидность теста. Для полной ее проверки следует учитывать множеств получаемых показателей. Более того, как это станет ясно из дальнейшего изложения, валидность теста может фактически гарантироваться логически обоснованными методами конструирования тестов.

Дискриминативность – это способность отдельных заданий теста и теста в целом дифференцировать обследуемых относительно “максимального” и “минимального” результата теста. Еще одной особенностью эффективных тестов является дискриминативность. Действительно, достижение удовлетворительного распределения показателей является одной из целей разработчика теста. При помощи тщательного конструирования теста можно обеспечить соответствующий уровень дискриминативности, а это именно то, в чем тесты значительно выигрывают по сравнению с другими формами испытаний.

Список литературы

1. Ефимова М.Р. и др. Общая теория статистики: Учебник. – М.: Финансы и статистика, 2009. – С.206.
2. Теория статистики: Учебник / Под ред. Р.А.Шмойловой. – М.: Финансы и статистика, 2008. – С.147.
3. Практикум по теории статистики: Учебное пособие / Под ред. Р.А.Шмойловой. – М.: Финансы и статистика, 2008. – С. 407.
4. Курс социально-экономической статистики /Под ред. Назарова М.П. – М.: Финстатинформ, 2008. – С.296.
5. Экономическая статистика: Учебник для вузов / Под. ред. Ю.Н. Иванова. – М.:Инфра – М., 2009. – С.241.
6. Социальная статистика /Под ред. И.И.Елисейевой – М.: Финансы и статистика, 2005.
7. Практикум по социальной статистике /Под ред. И.И. Елисейевой. – М.: Финансы и статистика. 2009. – С.112.
8. Салин В.Н., Медведев В.Г. и др. Макроэкономическая статистика. – М.: Дело, 2009. – С.56.
9. Вахромеева М.А. Статистика финансов: Учебник для вузов. – М.: Финансы и статистика, 2008. – С.662.