УДК 378.147

Ким Владимир Сергеевич,
кандидат физико-математических наук, доцент,
Дальневосточный федеральный университет,
г. Уссурийск, Россия,
kim.vs@dvfu.ru

К МЕТОДИКЕ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ «СОВРЕМЕННЫЕ СРЕДСТВА ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ»

Аннотация. *В дисциплине «Современные средства оценивания результатов обучения» важное место отводится тесту, как одному из средств оценивания учебных достижений. Надежность является одной из характеристик определяющих качество теста. Для определения надежности вычисляется коэффициент корреляции между индивидуальными баллами испытуемых из двух параллельных тестов, либо в расщепленном тесте.*

Поэтому педагог, разрабатывающий качественный тест, должен уметь определять как надежность теста, так и наличие внутренних противоречий в тесте.

В работе даются рекомендации по изложению учебного материала по расчету коэффициента корреляции, а также рекомендации по методике выполнения практических занятий.

Ключевые слова: *диагностика, тест, мониторинг, бинарная матрица, коэффициент корреляции, дихотомические данные.*

Abstract. *In the training course "Modern means of assessing learning outcomes," an important place is given to the test, as one of the means of assessing learning achievements. Reliability is one of the characteristics that determine the quality of the test. To determine the reliability, the correlation coefficient between the individual scores of the subjects from the two parallel tests is calculated.*

Therefore, a teacher who develops a quality test should be able to determine both the reliability of the test and the presence of internal contradictions in the test.

The work gives recommendations on the presentation of educational material on the calculation of the correlation coefficient, as well as recommendations on the methodology for doing practical exercises.

Keywords: *diagnostics, test, monitoring, binary matrix, correlation coefficient, dichotomous data.*

Введение

Эффективный учебный процесс может быть организован только в рамках управляемой образовательной системы. Управляемость же системы обеспечивается наличием подсистемы обратной связи.

За входную информацию в этой подсистеме обратной связи отвечает диагностика учебных достижений, в которой часто используются тесты как объективное и технологичное средство оценивания учебных достижений в педагогике [1, 2].

Тест представляет собой систему тестовых заданий, которые часто называют вопросами теста, что не совсем точно. Выполнив тестовое задание (ответив на вопрос), испытуемый получает 1 балл, если задание выполнено верно и 0 баллов, если не верно. В этом случае мы имеем с последовательностью единиц и нулей, называемой профилем испытуемого или вектором-строкой бинарной матрицы, о которой более подробно мы скажем ниже. Профиль представляет собой результаты тестирования, которые можно рассматривать как дихотомические данные. Сумма единиц и нулей в профиле дает индивидуальный балл испытуемого, который очень важен для расчета надежности теста.

Очень редко используются и другие способы оценивания, когда профиль содержит, так называемые политомические данные, которые могут иметь значения, например, от 0 до 9. В данной работе мы не будем рассматривать политомический случай.

Можно рассматривать и другую последовательность единиц и нулей – результаты выполнения данного тестового задания всеми испытуемыми, это так называемый вектор-столбец бинарной матрицы.

Таким образом, в процессе анализа результатов тестирования педагог получает большой массив дихотомических данных, число элементов в котором может достигать 2000 – 3000. Визуально проанализировать такой массив данных не представляется возможным, поэтому используются математические операции, позволяющие «свернуть» информацию, в результате чего создается значительно меньший массив выходных данных.

Математическая «свертка» результатов тестирования происходит в процессе вычисления коэффициента корреляции между дихотомическими данными по каким-либо двум векторам-столбцам.

Полученные значения коэффициентов корреляции между всевозможными парами векторов-столбцов сводятся в корреляционную матрицу, которая позволяет судить о качестве отдельных тестовых заданий, об их избыточности.

После редактирования некачественных тестовых заданий выполняется повторная апробация теста с последующим расчетом корреляционной матрицы и т.д. Таким образом, создание качественного теста это многоэтапный итерационный процесс, предполагающий расчеты корреляционной матрицы на каждом этапе.

После получения непротиворечивых результатов тестирования можно определить надежность теста в целом – вычислить коэффициент корреляции индивидуальных баллов испытуемых, полученных в двух параллельных тестах либо в расщепленном тесте.

Вышеописанные процедуры анализа дихотомических данных имеют большое значение, что обусловлено широким применением тестового контроля знаний, как наиболее объективного и технологичного средства контроля и диагностики учебных достижений [3].

Для осмысленного использования современных, научно обоснованных систем диагностики студенты педагогических вузов изучают, например, дисциплину «Современные средства оценивания результатов обучения». В процессе преподавания этой дисциплины возникают определенные методические проблемы, связанные с разрешением парадоксов возникающих при анализе дихотомических данных, в частности аномальные скачки значения коэффициента корреляции между векторами-столбцами бинарной матрицы.

Все вышесказанное свидетельствует в пользу того, что при изучении дисциплин, содержащих элементы классической теории тестов, необходимо особое внимание уделять теме, связанной с обработкой дихотомических данных.

Надежность тестов

Очень важной характеристикой теста является его надежность. Если тест обладает низкой надежностью, то его нельзя использовать для целей диагностики учебных достижений [4, 5].

Определение надежности теста основано на анализе так называемых бинарных матриц [6,7]. Бинарная матрица представляют собой таблицу, столбцы которой соответствуют номерам тестовых заданий в тесте, а строки – номерам испытуемых, которые выполняли задания теста. На пересечении строк и столбцов находится результат выполнения задания: «1» - если задание выполнено верно и «0», если неверно. Как уже указывалось выше, горизонтальная последовательность единиц и нулей образует вектор-строку, или профиль испытуемого. По профилю испытуемого вычисляется его индивидуальный балл, который далее используется при расчете надежности теста в целом.

Вертикальная последовательность единиц и нулей для какого-либо тестового задания образует вектор-столбец. Если вычислить коэффициенты корреляции между парами векторов-столбцов, то получаем корреляционную матрицу, облегчающую анализ результатов тестирования.

Например, если вектор-столбец состоит сплошь из единиц, то это означает, что с соответствующим заданием справились все испытуемые, т.е. это задание очень легкое. Если целью теста является ранжирование испытуемых, то такое задание «не работает», поскольку оно не ранжирует испытуемых. В этом случае задание рекомендуется к удалению из теста. Аналогично рекомендуются к удалению задания с векторами-столбцами, состоящими только из нулей. Эти задания слишком трудны для данной группы испытуемых, и не позволяют их ранжировать.

Случаи, когда вектор-строка состоит только из единиц, означает, что для данного испытуемого все тестовые задания слишком легкие и его рекомендуется перевести в группу «сильных» испытуемых, с дальнейшим повторным тестированием с помощью более трудного теста. Аналогично, если вектор-строка содержит только нули, то для данного испытуемого тест очень труден. Здесь рекомендуется перевести испытуемого в группу «слабых» испытуемых, с дальнейшим повторным тестированием с помощью более легкого теста.

Если трудность теста и уровень знаний испытуемых соответствуют друг другу, то векторы-столбцы и векторы-строки содержат как единицы, так и нули. Само чередование единиц и нулей, например в векторе-строке, согласно классической теории тестов, позволяет получить дополнительную информацию об испытуемых, но рассмотрение этого вопроса находится за пределами тематики данной работы.

При анализе бинарных матриц проверяются внутренние противоречия в полученных результатах. В частности последовательность единиц и нулей дает информацию о систематизированности знаний испытуемого, о нарушениях в процедуре тестирования и т.п. Проверка корреляции между векторами-столбцами бинарной матрицы позволяет судить о гомогенности теста, о возможных ошибках в построении тестового задания.

Остановимся на вопросе – «каков же педагогический смысл корреляции между векторами-столбцами бинарной матрицы?». Высокое значение коэффициента корреляции между двумя величинами позволяет предсказать значение второй величины, если известна первая. Значение коэффициента корреляции R, как известно, может принимать значения от –1 до +1, то есть $-1 \leq R \leq +1$ [8].

При $R = +1$ или $R = -1$ мы имеем очень сильную корреляцию, когда по известным значениям одной величины можно совершенно точно предсказать значение другой величины. При $R = 0$ корреляция отсутствует полностью.

Что для педагога, проводившего тестирование, будет означать $R = +1$ например, для первого и второго тестового заданий? Это означает, что зная набор единиц и нулей для вектора-столбца первого задания, мы можем точно предсказать набор единиц и нулей для вектора-столбца второго задания. Тогда одно из этих заданий лишнее и его можно удалить из теста, поскольку оно не дает новой информации в дополнение к другому заданию.

С другой стороны, если корреляция отсутствует ($R = 0$), то это тоже плохо. Как отмечалось выше, тест это не просто некоторое множество, а система тестовых заданий, где все задания имеют внутренние связи между собой. Эти связи, например, обусловлены тем, что для обеспечения гомогенности теста, цели каждого тестового задания должны совпадать с общей целью теста или быть его подцелями. Это, казалось бы, очевидное требование на практике соблюдается далеко не всегда. Например, в тесте имеются задания, которые имеют не одну, а две цели. Тогда корреляция между этим заданием и другими будет очень низкой. Рекомендуется оставлять в тесте задания, у которых коэффициент корреляции с другими заданиями не менее 0,3. Если же задание имеет отрицательный коэффициент корреляции с индивидуальным баллом испытуемого, то такое задание рекомендуется удалить из теста.

Таким образом, корреляционная матрица в свернутом виде содержит всю важнейшую информацию из трудно обозримой бинарной матрицы, содержащей результаты тестирования. Собственно расчет корреляционной матрицы на персональном компьютере в настоящее время не представляет особых трудностей. Трудности возникают при интерпретации полученных значений. Это обусловлено тем, что процесс расчета коэффициента корреляции скрыт от педагога. В компьютер загрузили исходные данные, он выполнил расчеты и выдал значение коэффициента корреляции.

Раньше, когда подобные вычисления проводились вручную, педагог-исследователь хорошо представлял себе, что происходит во время вычислений, поскольку сам их выполнял. Теперь же, оторванность от процесса вычислений, может приводить к ситуациям, когда полученные корреляционные матрицы могут вызывать удивление у педагога-исследователя.

Почему это происходит? Дело в том, что поскольку столбцы бинарной матрицы содержат только единицы и нули, то вычисленные значения коэффициента корреляции R по формуле Пирсона [8] могут вести себя парадоксальным образом.

Наглядно это можно продемонстрировать графически. Если коэффициент корреляции Пирсона близок к единице, то соответствующая корреляционная зависимость представляется прямой линией. Для задания линии, как минимум, необходимы две точки, каждая из которых задается двумя координатами – x и y. Поскольку обе эти переменные принимают значения только «1» или «0», то существую всего четыре возможные точки: 0,0; 0,1; 1,0; 1,1.

Иными словами, если мы на листе бумаги поставим две точки и с помощью линейки соединим их, то получим отрезок прямой линии. Схематически это показано на чертеже Fig.1а, где показаны отрезки прямой линии для различных пар точек из четырех возможных.

Если же мы поставим три точки и соединим их, то получим два отрезка прямой линии – между первой и второй и между второй и третьей точками. В целом мы получим ломаную линию (Fig.1б,). Аналогично мы получим три отрезка прямой линии, если поставим четыре точки (Fig.1в).

Таким образом, включая в рассмотрение 2, 3 или 4 точки, мы получим различные графические представления, как показано на Fig.1.

Всего получим три варианта графического представления:

а) использование двух точек;

б) использование трех точек;

в) использование четырех точек.

Подпись: Fig.1. Типы линий для разных наборов точек.

Легко видеть, что только в случае а) мы получаем линейную зависимость, для которой коэффициент корреляции Пирсона равен 1 (R = 1). В случаях б) и в) мы имеем ломаные линии, для которых коэффициент корреляции Пирсона будет значительно меньше единицы (R << 1).

Из приведенных графиков следует, что добавление одной или двух точек, относящихся к другому набору, приводит к переходу от случая а) к случаям б) или в). В результате коэффициент корреляции резко изменяется.

Совершенно иная картина наблюдается, когда мы имеем дело с корреляционными переменными, определенных на множестве вещественных чисел. В этом случае переменные могут принимать любое значение из неограниченного числа точек. Добавление одной или двух точек практически не повлияет на значение коэффициента корреляции.

Поскольку в учебной литературе понятие коэффициента корреляции вводится на основе именно вещественных переменных, то студенты привыкают к мысли, что изменение нескольких точек в большом массиве данных не может заметно повлиять на коэффициент корреляции.

При анализе же бинарных матриц, студенты сталкиваются с парадоксальным поведением коэффициента корреляции, когда последний может резко изменяться при добавлении всего одной эмпирической точки. Подобное поведение коэффициента корреляции при анализе дихотомических данных вызывает недоумение студентов и требует дополнительных разъяснений. Эти разъяснениям и посвящена настоящая работа.

Расчеты

Продемонстрируем вышеприведенные рассуждения практическими расчетами.

Вариант 1. 10 точек на непрерывной шкале.

X	0,0	0,1	0,2	0,4	0,5	0,6	0,7	0,8	0,9	1,1
Y	0,0	0,3	0,6	0,8	1,2	1,6	1,8	2,0	2,2	2,4

R = 0,989

Вариант 2. 11 точек на непрерывной шкале.

X	0,0	0,1	0,2	0,4	0,5	0,6	0,7	0,8	0,9	1,1	0,3
Y	0,0	0,3	0,6	0,8	1,2	1,6	1,8	2,0	2,2	2,4	1,1

R = 0,981

После добавления 11-й точки коэффициент корреляции уменьшился на 0,8%, т.е. практически не изменился.

Вариант 3. 10 точек на дихотомической шкале

Задание 1	X	0	1	1	1	0	0	1	0	0	0
Задание 2	Y	0	1	1	1	0	0	1	0	0	0

R = 1,000

Вариант 4. 11 точек на дихотомической шкале

Задание 1	X	0	1	1	1	0	0	1	0	0	0	1
Задание 2	Y	0	1	1	1	0	0	1	0	0	0	0

R = 0,828

После добавления 11-й точки коэффициент корреляции уменьшился на 17%, т.е. сильно изменился.

Вариант 5. 12 точек на дихотомической шкале

Задание 1	X	0	1	1	1	0	0	1	0	0	0	1	0
Задание 2	Y	0	1	1	1	0	0	1	0	0	0	0	1

R = 0,657

После добавления 12-й точки коэффициент корреляции уменьшился на 34%, т.е. изменился очень сильно.

Обсуждение результатов

Расчетные модельные варианты 1 и 2 относятся к корреляционным переменным, заданным на непрерывной шкале. Из приведенных расчетов видно, что добавление еще одной точки к 10-ти имеющимся, практически не изменяет коэффициент корреляции. Если бы у нас было не 10, а 100 или 1000 исходных точек, то коэффициент корреляции изменился бы еще меньше.

Расчетные модельные варианты 3, 4 и 5 относятся к корреляционным переменным, заданным на дихотомической шкале. Из приведенных расчетов видно, что добавление еще одной точки к 10-ти имеющимся, сильно изменяет коэффициент корреляции. Если бы у нас было не 10, а 100 или 1000 исходных точек, то коэффициент корреляции все равно бы, сильно изменился. Происходит это по причине того, что добавляется точка из другого набора данных. Таким образом, наблюдается парадоксально сильное изменение коэффициента корреляции при слабом изменении количественного состава анализируемых точек.

Методические рекомендации

При изложении темы, посвященной расчету корреляционной матрицы тестовых результатов, необходимо обратить внимание студентов на вышеуказанные парадоксальные скачки значения коэффициента корреляции.

Для разрешения парадокса следует проработать графики на Fig.1. Необходимо разъяснить, что кардинальное изменение вида линии при переходе от случая а) к случаям б) и в) происходит по причине того, что изменяется тип наборов данных.

Если у нас будет большое количество точек из какого-либо набора данных, то добавление еще одной точки из этого же набора не приведет к изменению коэффициента корреляции. Добавление же точки из другого набора данных приведет к скачкообразному изменению коэффициента корреляции.

Необходимо добиться полного понимания студентами этого утверждения. Для этого будет весьма полезно студентам выполнить практическую работу по расчету коэффициента корреляции для дихотомических данных.

1. Исходные данные должны содержать несколько сотен точек из первого набора (Fig.1a). Далее вычисляется коэффициент корреляции векторов-столбцов для двух заданий теста.

2. На следующем этапе практической работы добавляется одна точка из другого набора данных (Fig.1б). Вычисляется коэффициент корреляции и сравнивается с его предыдущим значением и делается вывод о его сильном изменении.

3. Для сравнения, в исходный набор данных (Fig.1a) добавляется одна точка из этого же набора данных. Сравнивается вновь полученный коэффициент корреляции с исходным и делается вывод о его неизменности.

Выводы

Таким образом, коэффициент корреляции тестовых заданий между собой может проявлять парадоксальные, на первый взгляд, изменения своего значения. Для разрешения этого парадокса необходимо еще раз разъяснить студентам сущность дихотомических данных и выполнить практическую работу по их анализу.

В заключение выражаю признательность Фалалеевой О.Н. за привлечение внимания к описанной проблеме, возникающей в процессе преподавания дисциплины «Современные средства оценивания результатов обучения».

Литература

1. Зайчикова Т.Н. Технология педагогического тестирования как средство эффективного управления функционированием и развитием образовательной системы региона // Дисс. на соискание ученой степени кандидата педагогических наук. - Нижний Новгород, 2003. -322 с.
2. Ким В.С. Компьютерное тестирование как элемент управления учебным процессом // Вестник Московского государственного областного университета. Серия: Педагогика. 2007. Т. 2. № 2. –С. 94-98.
3. Ким В.С. Матричное представление результатов тестирования. Вестник МГОУ, сер. Педагогика, 2012, №4. С.114-120.
4. Crocker Linda, Algina James. Introduction to Classical and Modern Test Theory. –New-York: Harcourt Brace Jovanovich, 1986. – 527 p.
5. Анастази А., Урбина С. Психологическое тестирование. – Спб.: Питер, 2006. -688 с.
6. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. - М., 1989. –167 с.
7. Ким В.С. Бинарные матрицы тестовых результатов // Инфо-Стратегия 2012: Общество. Государство. Образование. Сб.матер.конференции. – Самара, 2012. –С.233-237.
8. Glass G.V., Stanley J.C. Statistical Methods in Education and Psychology. - New Jersey: Prentice-Hall, Inc., Englewood Ciiffs, 1970. - 495 p.