

Кафедра

педагогических измерений

ПРОВЕДЕНИЕ СТАТИСТИЧЕСКОГО АНАЛИЗА КАЧЕСТВА ЗАДАНИЙ В СРЕДЕ MS EXCEL

Роман Дубинка

г. Барановичи, Республика Беларусь

xxxroman@bk.ru

В статье рассмотрен процесс расчёта основных показателей качества теста¹ посредством стандартных встроенных функций Microsoft Excel 2007.

Ключевые слова: статистический анализ, Microsoft Excel, мастер функций, коэффициент корреляции.

Процесс проверки качества заданий предполагает проведение множества сложных математических расчётов, сделать это с помощью только ручки и бумаги сложно, даже с использованием калькулятора это может занять не один час. Интенсивное развитие компьютерных технологий, а в частности программного обеспечения, позволяет значительно ускорить данный процесс, сократив объём производимых операций.

Существует множество специализированных программ, так называемых статистических пакетов, которые позволяют глубоко и многосторонне обрабатывать математические данные, примером

1
По материалам сайта
www.testolog.narod.ru

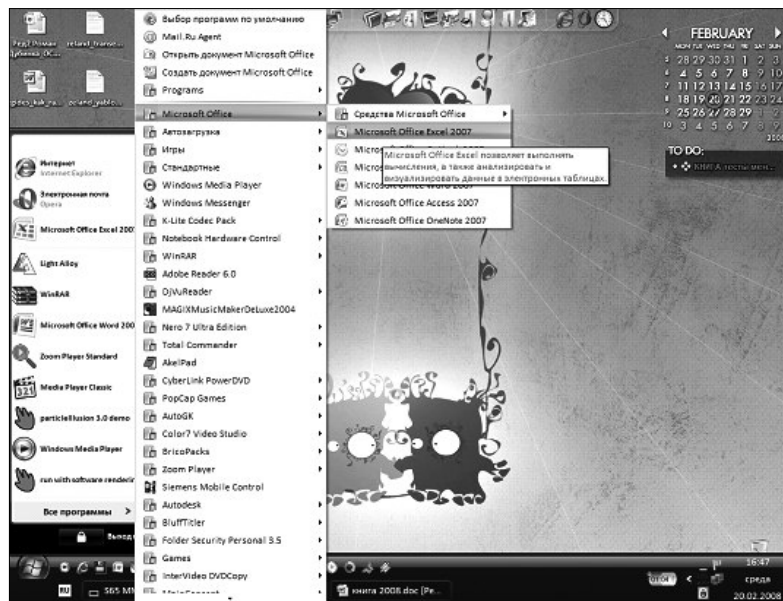
ПЕД
измерения

таких программ является SPSS, STATISTICA, STATGRAPHICS. Однако в силу того, что большинство статистических пакетов относится к специализированным программам и недоступны для обычного пользователя, появляется необходимость прибегать к помощи универсальных программ, способных проводить сложные математические расчёты и статистический анализ данных.

На сегодняшний день трудно найти компьютер, на котором не был бы установлен пакет программ Microsoft Office, в том числе, входящий в его стандартный набор, табличный процессор Excel². Он является

наиболее доступным и обладает набором необходимых характеристик необходимыми для организации эмпирического анализа теста.

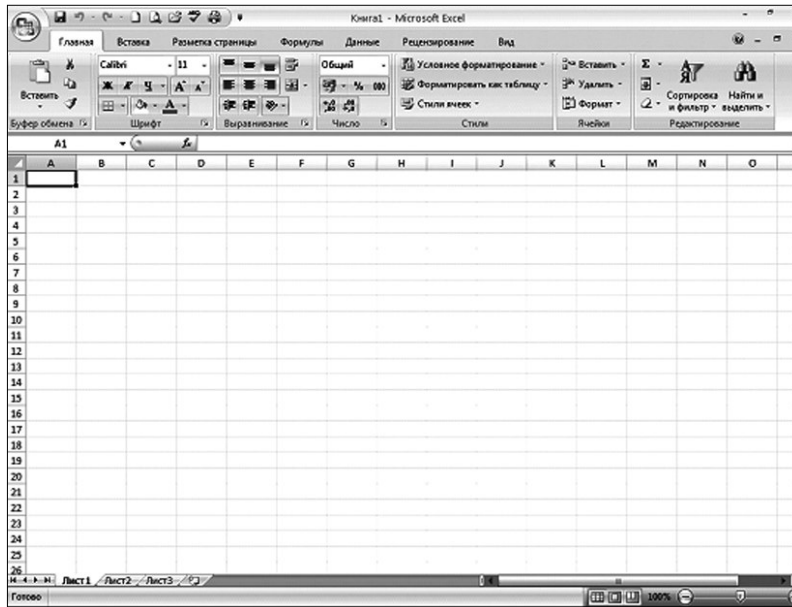
Рассмотрим проведение статистического анализа на примере. Предположим, имеется 10 заданий, а тестирование проводилось в группе, состоящей из 20 испытуемых. Проведение эмпирического анализа начинается с занесения результатов в матрицу и с упорядочения данных. Для начала работы открываем MS Excel, при обычной установке Пуск / Программы / Microsoft Office / Microsoft Office Excel 2007.



2
Эмпирический анализ проводился в последней версии Microsoft Excel 2007 на момент написания статьи.

В рабочей области приступим к заполнению. В первую

очередь заполним шапку, а затем баллы испытуемых.



Кафедра педагогических измерений
ИЗВЕЩЕНИЕ
О РЕЗУЛЬТАТАХ
КЭФЭИ

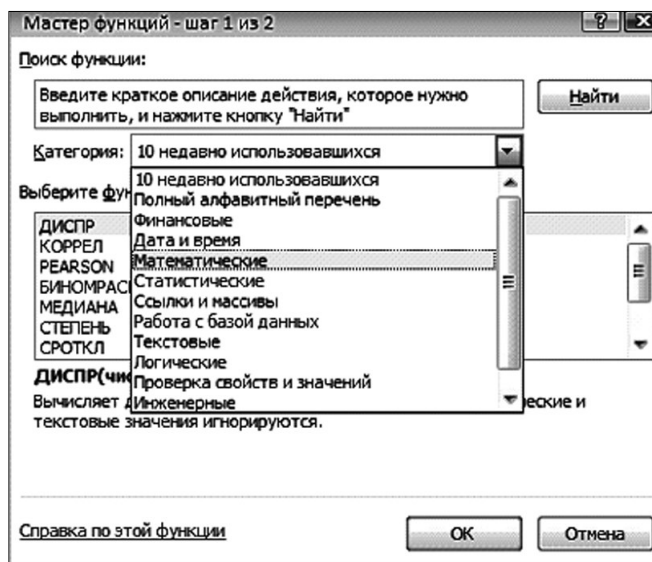
Соответственно, ввод в таблицу начинается с ячейки A2 вниз по столбцу, в ячейки (B1:K1) вводятся номера заданий в тестовой форме. Группа ячеек (B2:K21) заполняется результатами тестирования: нулями и единицами.

	A	B	C	D	E	F	G	H	I	J	K
1	задание	1	2	3	4	5	6	7	8	9	10
2	Испытуемый 1	1	0	1	0	0	1	0	0	0	0
3	Испытуемый 2	0	0	0	1	0	1	0	1	1	0
4	Испытуемый 3	1	0	1	0	0	0	0	1	0	1
5	Испытуемый 4	1	1	1	1	1	1	1	1	0	0
6	Испытуемый 5	0	0	0	0	0	0	1	0	1	0
7	Испытуемый 6	1	1	1	1	0	1	1	1	1	1
8	Испытуемый 7	1	1	1	1	1	1	1	1	1	1
9	Испытуемый 8	1	1	1	1	1	1	1	1	1	0
10	Испытуемый 9	1	1	1	1	0	1	1	1	1	1
11	Испытуемый 10	1	1	1	1	1	1	1	1	1	1
12	Испытуемый 11	0	1	1	1	1	1	0	1	1	1
13	Испытуемый 12	0	0	1	1	1	1	1	1	1	1
14	Испытуемый 13	0	1	1	1	1	1	0	1	0	0
15	Испытуемый 14	0	1	1	1	1	1	1	0	1	0
16	Испытуемый 15	0	1	1	0	1	1	1	0	1	1
17	Испытуемый 16	0	0	1	0	0	1	1	1	0	0
18	Испытуемый 17	0	1	1	0	1	1	1	0	0	0
19	Испытуемый 18	0	0	1	0	1	1	0	1	0	1
20	Испытуемый 19	1	0	1	0	1	1	1	0	1	0
21	Испытуемый 20	1	0	1	0	1	1	1	0	0	1

ПЕД
измерения

Посредством встроенной функции *СУММ* подсчитываем количество правильных ответов по каждому из заданий. Для реализации данной функции можно использовать два способа. Первый – когда формула напрямую вводится в нужную ячейку. Вторым способом используется в случае обработки большого количества

данных или слабого опыта работы. Он заключается в использовании мастера функций. Что бы запустить мастер функции, необходимо «кликнуть» по его значку на панели инструментов. В отрывшемся диалоговом окне из раскрывающегося списка *Категория*, необходимо выбрать *Математические*.



Из списка «Выберите функцию», указать *СУММ*, ниже появиться расшифровка функции: *Суммирует аргументы*. «Кликаем» *OK*. На следующем шаге мастера функции выбираем суммируемые аргументы. Для этого помещаем курсор в поле *Число1* и «кликаем» по значку.

Выделяем область с баллами по первому заданию (*B2:B21*) и вновь возвращаем

окно мастера функции к исходному размеру. Для завершения создания функции «кликаем» *OK*. После этого растягиваем содержимое ячейки по строке до ячейки *K22*. Для этого выделяем ячейку с имеющейся функцией и наводим курсор на правый нижний угол рамки, которая выделяет ячейку, когда курсор примет форму черного крестика кликаем левой клавишей мыши и, не отпуская её,

двигаем курсор вправо по строке до последней ячейки, в которой должна быть функция.

Количество неправильных ответов определяем как разницу между общим числом испытуемых и числом уже найденных правильных ответов. Сле-

	A	B	C	D	E	F	G	H	I	J	K
22	R	10	11	18	11	13	18	14	13	12	10
23	W	10	9	2	9	7	2	6	7	8	10
24	R\п=p	0,5	0,6	0,9	0,6	0,7	0,9	0,7	0,7	0,6	0,5
25	W\п=q	0,5	0,5	0,1	0,5	0,4	0,1	0,3	0,4	0,4	0,5

После перевода формулы на язык MS Excel для 1-го задания она примет вид: «=B24*B25», это произведение помещается в ячейку B26 и растягивается по строке на все задания, т.е. до ячейки K26 включительно.

В ячейки таблицы MS Excel вносятся следующие формулы: L2 «=СУММ(B2:K2)»; M2 «=10-L2»; N2 «=L2/10»; O2 «=1-N2», а затем выделяется массив (L2:O2) с введёнными в него формулами и растягивается на ячейки (L2:O21), тем самым пропадает необходимость вводить заново эти же формулы для каждого испытуемого.

Следующим элементом расчёта является дисперсия баллов испытуемых и средне-квадратическое отклонение, эти показатели рассчитываются как для заданий, так и для испытуемых. В последнюю свободную строку вносим простейшую функцию умножения доли правильных и неправильных

дующие две строки включают в себя доли правильных и неправильных ответов, для этого, как упоминалось выше, необходимо разделить поочередно значение ячеек (B22:K23) на 20 и поместить полученные цифры в следующие две строки.

ответов: ячейке B26 присваивается значение «=B24*B25».

В современных технологиях адаптивного обучения и контроля используется другая мера трудности задания, равная $\ln q_j/p_j^3$, а так же логит уровня знаний, равный $\ln p_j/q_i$.

Если вручную вычисления данного показателя займет много времени, то с помощью встроенной функции

MS Excel данная процедура значительно упрощается. В ячейку

	L	M	N	O
1	R	W	p	q
2	3	7	0,3	0,7
3	4	6	0,4	0,6
4	4	6	0,4	0,6
5	8	2	0,8	0,2
6	2	8	0,2	0,8
7	9	1	0,9	0,1
8	10	0	1	0
9	9	1	0,9	0,1
10	9	1	0,9	0,1
11	10	0	1	0
12	8	2	0,8	0,2
13	8	2	0,8	0,2
14	6	4	0,6	0,4
15	7	3	0,7	0,3
16	7	3	0,7	0,3
17	4	6	0,4	0,6
18	5	5	0,5	0,5
19	5	5	0,5	0,5
20	6	4	0,6	0,4
21	6	4	0,6	0,4

Кафедра педагогических измерений

ИЗМЕРЕНИЙ
ЦЕНТРАЛЬНЫХ
КАФЕДРА

3

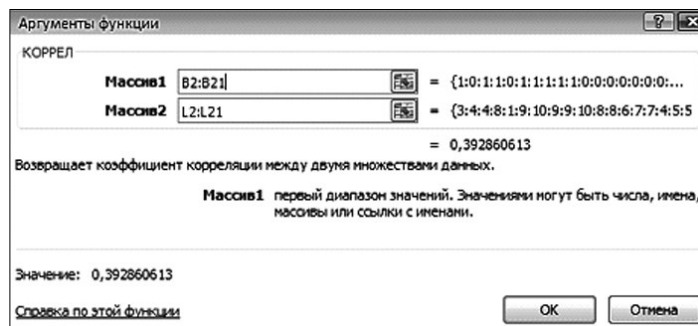
Эту меру трудности, получаемую в шкале натуральных логарифмов, называют логит трудности задания.

ПЕД
измерения

В27 вставляем математическую функцию вычисления логарифма LN и вводим выражение, из которого будем находить логарифм: « $=LN(B25/B24)$ ». Выделив ячейки B26-B27, растягиваем их на все задания. Логит уровня знаний рассчитывается по формуле « $=LN(N2/O2)$ », которая заносится в ячейку P2, где N2 — доли правильных ответов первого испытуемого по всему тесту; O2 — доли неправильных ответов.

Теперь переходим к расчёту ключевого показателя —

коэффициента корреляции ответов на задание с ответами на сумму баллов, потому что задание в тестовой форме нельзя называть тестовым, если оно не коррелирует с суммой баллов по всему тесту. Этот показатель будет заноситься в строку 28. «Кликнув» левой клавишей мыши по кнопке добавления функции, на вкладке категория выбираем «Статистические». В графе «Выберите функции», выбираем *КОРЕЛЛ*. Нажимаем на кнопку *OK*.



В окне два поля: *Массив1* и *Массив2*, в поля заносится массив с ячейками данных, между которыми устанавливается корреляционная зависимость. В качестве первого массива выбирается область ячеек с результатами испытуемых по первому заданию (B2:B21), в качестве второго массива — область ячеек с суммой баллов по всему тесту. Для того, чтобы не пришлось набирать функцию для каждой ячейки, необходи-

мо закрепить элементы *Массива1* и *Массива2*, т.е. вместо (L2:L21) заносим в поле ($\$L\$2:\$L\21), внесенные изменения позволят свободно растянуть значение ячейки, потому что при проведении данной операции значения *Массива1* будут изменяться автоматически, так же как и *Массива2*, но необходимо, чтобы изменялся только первый массив, поэтому с помощью символа \$ закрепляется значение *Массива2*. В этом же окне



можно посмотреть результат вычислений. Для завершения ввода функции нажимаем **OK!** Дублируем значение ячейки.

Попарная корреляция ответов на задания рассчитывается аналогично коэффициенту корреляции ответов на задание с ответами на сумму баллов, разница заключается в том, что при расчёте данного показателя устанавливается связь меж-

ду баллами по заданию и суммой баллов по остальным заданиям. При расчёте в поле *Массив2* вводим следующее ($L\$2:\$L\$21-B2:B21$), т.е. произвели вычитание баллов задания из общей суммы баллов. Результаты вычислений заносятся в *строку 29*.

После всех манипуляций таблица примет следующий вид.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	задание	1	2	3	4	5	6	7	8	9	10	R	W	p	q	ln(p/q)
2	Испытуемый 1	1	0	1	0	0	1	0	0	0	0	3	7	0,3	0,7	-0,847
3	Испытуемый 2	0	0	0	1	0	1	0	1	1	0	4	6	0,4	0,6	-0,405
4	Испытуемый 3	1	0	1	0	0	0	0	1	0	1	4	6	0,4	0,6	-0,405
5	Испытуемый 4	1	1	1	1	1	1	1	1	0	0	8	2	0,8	0,2	1,386
6	Испытуемый 5	0	0	0	0	0	0	0	0	1	0	1	9	0,1	0,9	-2,197
7	Испытуемый 6	1	1	1	1	0	1	1	1	1	1	9	1	0,9	0,1	2,197
8	Испытуемый 7	1	1	1	1	1	1	1	1	1	1	10	0	1	0	#####
9	Испытуемый 8	1	1	1	1	1	1	1	1	1	0	9	1	0,9	0,1	2,197
10	Испытуемый 9	1	1	1	1	0	1	1	1	1	1	9	1	0,9	0,1	2,197
11	Испытуемый 10	1	1	1	1	1	1	1	1	1	1	10	0	1	0	#####
12	Испытуемый 11	0	1	1	1	1	1	0	1	1	1	8	2	0,8	0,2	1,386
13	Испытуемый 12	0	0	1	1	1	1	1	1	1	1	8	2	0,8	0,2	1,386
14	Испытуемый 13	0	1	1	1	1	1	0	1	0	0	6	4	0,6	0,4	0,405
15	Испытуемый 14	0	1	1	1	1	1	1	0	1	0	7	3	0,7	0,3	0,847
16	Испытуемый 15	0	1	1	0	1	1	1	0	1	1	7	3	0,7	0,3	0,847
17	Испытуемый 16	0	0	1	0	0	1	1	1	0	0	4	6	0,4	0,6	-0,405
18	Испытуемый 17	0	1	1	0	1	1	1	0	0	0	5	5	0,5	0,5	0
19	Испытуемый 18	0	0	1	0	1	1	0	1	0	1	5	5	0,5	0,5	0
20	Испытуемый 19	1	0	1	0	1	1	1	0	1	0	6	4	0,6	0,4	0,405
21	Испытуемый 20	1	0	1	0	1	1	1	0	0	1	6	4	0,6	0,4	0,405
22	R	10	11	18	11	13	18	13	13	12	10					
23	W	10	9	2	9	7	2	7	7	8	10					
24	R\p	0,5	0,55	0,9	0,55	0,65	0,9	0,65	0,65	0,6	0,5					
25	W\p=q	0,5	0,45	0,1	0,45	0,35	0,1	0,35	0,35	0,4	0,5					
26	p=q	0,25	0,25	0,09	0,25	0,228	0,09	0,23	0,23	0,24	0,25					
27	ln p/q	0	-0,2	-2,2	-0,2	-0,62	-2,2	-0,6	-0,6	-0,4	0					
28	xy	0,39	0,71	0,54	0,71	0,483	0,34	0,61	0,44	0,43	0,48					
29	lr	0,2	0,58	0,45	0,58	0,311	0,45	0,47	0,26	0,26	0,29					

Использование Microsoft Excel при проведении эмпирического анализа теста позволяет:

1. Сократить затрачиваемое на вычисления время.
2. Снизить вероятность допущения ошибки при вычислениях.
3. Обрабатывать огромные массивы данных.

4. Легко переносить и публиковать результаты вычислений, как в виде таблиц, так и в виде графиков и диаграмм.

5. Импортировать в табличный процессор результаты тестирования из специальных тестовых оболочек и программ.