ВЫЧИСЛЕНИЕ ПОКАЗАТЕЛЕЙ СРЕДНЕЙ ТЕНДЕНЦИИ

В.С. Аванесов
testolog@mtu-net

В статье излагаются на÷альные сведения по статисти÷еским методам, применяемым для проведения педагоги÷еских измерений. Эта ÷асть авторского курса, расс÷итанного на на÷инающих тестологов и преподавателей, представляющих преимущественно гуманитарные у÷ебные дисциплины.

Статистика определяется как наука о массовых явлениях, с помощью которой можно получить обобщённые данные об изучаемых явлениях и совокупностях, рассчитать показатели связи и влияния, обнаружить закономерности в развитии изучаемых процессов, обосновать результаты эксперимента.

Исходным понятием статистики является понятие *совокупность,* объединяющее обычно какое-либо множество испытуемых по одному или нескольким интересующим признакам. Главное требование к выделению изучаемой совокупности — это её качественная однородность по интересующим признакам. Члены совокупности сравниваются между собой в отношении интересующего качества. При этом обычно абстрагируются от других качеств. Так, если исследователя интересует интеллект или успеваемость учащихся, то он не принимает во внимание, как правило, их рост, вес и другие характеристики, не относящиеся непосредственно к изучаемому вопросу.

Уместность применения того или иного статистического метода зависит от способа образования исследуемой совокупности и от количества испытуемых. Применение большинства статистических методов основано на идее использования небольшой случайной совокупности испытуемых из общего числа тех, на которых можно было бы распространить (генерализовать) выводы, полученные в результате изучения совокупности. Эта небольшая совокупность в статистике называется *выборочной совокупностью,* или короче, выборкой.

Главный принцип формирования выборки — это случайный отбор испытуемых из мыслимого множества учащихся (студентов), называемого *генеральной совокупностью*. Как по анализу элементов, содержащихся в капле крови, медики нередко судят о составе всей крови человека, так и по выборочной совокупности испытуемых изучаются признаки, характерные для всей генеральной совокупности. Генеральную совокупность составляют те лица, на которых можно распространить (генерализовать) выводы, полученные по выборочным данным.

Соответственно, объём генеральной совокупности зависит от уровня «претенциозности» исследователя. Однако с увеличением объёма генеральной совокупности возникают трудности организационного порядка. Для постановки строгого эксперимента, например в педагогике, каждый обучающийся, скажем, республики должен иметь шанс попасть в выборочную совокупность экспериментатора (в экспериментальную или контрольную группу), что, конечно, практически трудно сделать. Поэтому генеральная совокупность нередко ограничивается испытуемыми города (если он небольшой) или района (если исследование не проводится на республиканской или всероссийской выборке).

Но здесь возникает другой вопрос: какая, вообще говоря, науке польза, если данные эксперимента ограничиваются местными масштабами? Такого рода вопрос обычно решается компромиссом в сторону увеличения генеральной совокупности до пределов, теоретически допускающих включение каждого члена генеральной совокупности в выборочную совокупность.

3.Построение гистограммы

После ввода данных в ЭВМ вначале ставится задача, образно говоря, увидеть данные, т. е. определить, как распределены результаты. При использовании качественных показателей педагогического процесса в большинстве случаев существует так называемое *нормальное распределение* результатов, математическим (идеальным) аналогом которого является *кривая нормального распределения*.

Отклонения эмпирических данных от идеальной модели бывают, как правило, всегда. Вопрос в том, насколько существенны такие отклонения и можно ли применять для обработки на ЭВМ стандартные методы, основанные на предположении о нормальности распределения?

Некоторые суждения о характере распределения данных можно сделать после построения гистограммы.

Гистограммой называется столбчатая диаграмма (геометрическая фигура), дающая наглядное представление о распределении данных. Получить такое представление важно потому, что иногда вследствие неудачно подготовленного или выбранного теста (или в силу других причин), распределение эмпирических данных принимает различный характер. В этой связи потребуется и другая интерпретация результатов.

Для определения характера распределения рекомендуется построение гистограммы. Делается это на примере.

При тестировании остроты зрения у курсантов лётных училищ были получены следующие данные[1]:

59 48 53 47 57 64 62 62 65 57 57 81 83*
48 65 76 53 61 60 37* 51 51 63 81 60 77
71 57 82 66 54 47 61 76 50 57 58 52 57
40 53 66 71 61 61 55 73 50 70 59 50 59
83 69 67 66 47 56 60 43 54 47 81 76 69

1 *Guilford, J.P.* Fundamental Statiatics in Psychology and Education. N-Y, McGraw-Hill Book Co, 1956.

1. При построении гистограммы вначале определяется число испытуемых. Это число обычно обозначается латинской буквой N. В данном случае $N = 65$.

Если группа делится на подгруппы, то число испытуемых в каждой подгруппе выражается символом n_j, где индекс j означает номер подгруппы, а n – число испытуемых в ней.

2. Затем находятся максимальное и минимальное значения: максимальное значение теста равно 83, минимальное = 37. Эти значения иногда называют *лимитами,* $X_{\min}$ и $X_{\max}$. Последние полезно пометить звёздочкой, что и сделано в данном примере. Лимиты показывают на пределы варьирования результатов в данной выборке.

Варьирование эмпирических данных отражает фундаментальный принцип разнообразия (изменчивости) изучаемых объектов (или испытуемых по интересующему признаку. Будучи измеренным у каждого испытуемого, этот признак рассматривается как переменная величина.

3. Для того чтобы разделить выборку испытуемых на небольшие группы (классы) определяется значение *классового интервала* (i). Для этого используется формула:

$$i = \frac{(X_{\max} - X_{\min})}{\text{предполагаемое число классов}}. \qquad (1)$$

При выборе числа классов в знаменателе этой формулы заложен элемент произвольности (в смысле выбора числа классов). Так бывает нередко при применении статистических методов в силу вероятностного характера данных и ожидаемых выводов. Предполагаемое число классов берётся в зависимости от числа испытуемых, а также из эстетических соображений; желательно, чтобы гистограмма хорошо смотрелась. Чем больше испытуемых, тем большим может быть и число классов. При $N = 30–50$ человек это число обычно берётся равным 6–8, при $N>50$ число 8-10, при $N>200$ число классов может возрасти до 12. При $N>1000$ оно может возрасти до 20 и более.

В данном примере число классов можно взять равным 10. Тогда:

$$i = \frac{(83 - 37)}{10} = 4{,}6. \qquad (2)$$

Полученное значение классового интервала обычно округляется до целого числа. Таким образом, для данного примера i равно пяти. Полезно заметить, что в зависимости от числа испытуемых и тестовых баллов в качестве значений классовых интервалов считаются удобными числа 1, 2, 5, 10, 20, 50, 100 или более. Например, в тысячебалльной тестовой шкале значение классового интервала результатов испытуемых может приниматься равным 100.

4. Далее строится вспомогательная таблица. Для удобства расчётов наименьшая граница класса принимается равной числу, ближайшему к $X_{\min}$ и кратному 5. В качестве такового нижний предел равняется 35 (хотя такого значения тестового балла в выборке нет). Принимая расстояние между классами равное 5, строится вспомогательная таблица, где первый столбец – классы, по которым надо будет разносить данные примера 1; второй столбец – частоты, указывающие, как часто те или иные данные попадают в определённый класс.

Таблица 1.
Вспомогательная

№ п/п	*X* классы	*f* частоты
1	35–39	1
2	40–44	2
3	45–49	6
4	50–54	11
5	55–59	12
6	60–64	11
7	65–69	8
8	70–74	4
9	75–79	4
10	80–84	6
$\sum$		65

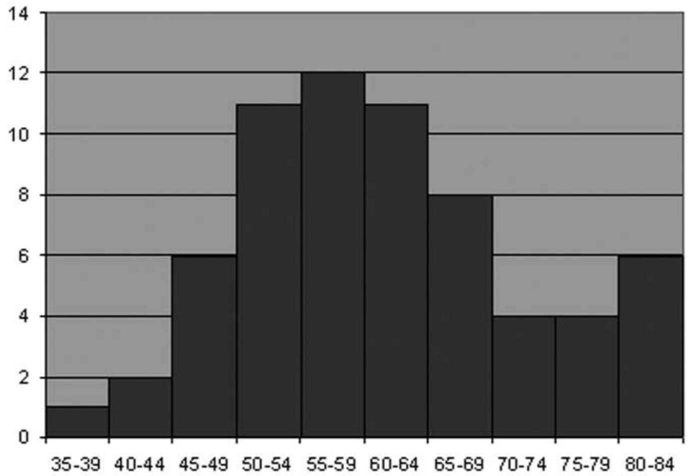
5. По результатам вспомогательной таблицы строится гистограмма. В ней по оси абсцисс отложены значения классов тестовых баллов (второй столбец табл. 1), а по оси ординат – частоты, с которыми в данной выборке представлены классы значений (третий столбец значений этой же таблицы) (см. рисунок).

4. Вопросы проведения эксперимента

Испытуемые часто не равны по уровню подготовленности, и потому методы, эффективные, скажем, в одной группе испытуемых (или опрошенных), нередко могут оказаться неэффективными в другой группе.

Школьный класс, на котором проводится большинство педагогических экспериментов, не является, строго говоря, выборочной совокупностью. Это, скорее, группы учащихся, набранные по территориальному или иному признаку. Методы, эффективные, скажем, в одном классе школы, могут оказаться неэффективными в другом классе.

Поэтому специфической особенностью экспериментов в педагогических науках является то, что в них почти никогда не выдерживается требование случайности отбора по вполне понятной причине: в процессе учёбы трудно отбирать в случайном порядке



учащихся из разных классов и разных школ и формировать затем из них экспериментальную и контрольную группы. Это обстоятельство даёт некоторые основания относить эксперименты, проводимые в школьных классах, к числу так называемых *квазиэкспериментов*, т. е. нестрогих экспериментов.

Если нет возможности отобрать учащихся в случайном порядке, то надо отбирать в таком порядке целые классы учащихся. И обучать одни классы по экспериментальной программе, другие — по традиционной (контрольной). Такова реальность, с которой необходимо считаться, но которая не всегда учитывается в статистическом анализе данных, проводимом из обычного предположения о строго случайном отборе испытуемых.

Наиболее простой и часто применяемый вид эксперимента — это исследование результатов экспериментальной и контрольной групп. Из числа членов генеральной совокупности в случайном порядке формируются две группы учащихся. Случайный отбор лучше проводить по таблице случайных чисел и тому подобных методов, обеспечивающих равный шанс каждому попасть в выборочную совокупность.

В случайном порядке учащиеся разбиваются на экспериментальную и контрольную группы. Если требование случайности отбора строго выдержано, то, как правило, обе группы оказываются примерно одинаковыми по уровню начальной подготовленности и по другим признакам, влияющим на усвоение предмета. Тем не менее, до начала эксперимента целесообразно проверить, насколько правомерно это предположение.

5. Вычисление основных статистик

Статистические результаты, которые получаются в процессе изучения выборочных совокупностей, называются выборочными статистиками, или чаще *статистиками* В числе основных статистик – средняя арифметическая, значение дисперсии и стандартного отклонения, коэффициенты корреляции и др.

Например, в двух группах учащихся — экспериментальной и контрольной — получены следующие тестовые баллы:

Таблица 2.
Результаты эксперимента

Первая группа (экспериментальная), человек (n =11)	Вторая группа (контрольная), человек (n = 9)
12, 14, 13, 16, 11, 9, 13, 15, 15, 18, 14	13, 9, 11, 10, 7, 6, 8, 10, 11

5.1. Расчёт средних арифметических

Расчёт средних арифметических в каждой группе испытуемых производится по формуле:

$$M = \frac{\sum X_i}{N}, \qquad (3)$$

где X_i — тестовый результат любого (говорят i-го) испытуемого данной группы; Σ — знак суммирования результатов всех испытуемых, начиная от первого, с номером 1, до последнего, с номером N. В первой группе $\Sigma = (12 + 14 + 13 + 16 + 11 + 9 + + 13 +15 + 15+ + 18 +14 = 150$.

N – число испытуемых, в экспериментальной группе оно равно 11. Подставляя эти данные в формулу (1), получаем среднюю арифметическую для первой группы (M_1):

$$M_1 = \frac{150}{11} = 13{,}636.$$

Аналогично поступаем с результатами во второй группе испытуемых. Здесь Σ= (13 + 9 + 11 + + 10 + 7 + 6 + 8 + + 10 + 11) = 85. Число испытуемых (N) – девять человек. Подставляя эти данные в формулу (1) получаем значение средней арифметической для второй группы:

$$M_2 = \frac{85}{9} = 9{,}444.$$

Тестологи часто имеют дело с тестовыми матрицами. Матрица – это форма компактного и удобного представления результатов тестирования.

Пример тестовой матрицы[2] (см. стр. 127). В этой матрице по строкам предполагается расположение номеров испытуемых, а по столбцам – номера заданий. В явном виде это расположение выражено в табл. 3. С данными этой таблицы можно произвести ряд статистических расчётов.

Элементы таблицы отражают результаты тестирования: за каждый правильный ответ на соответствующее задание студенты

[2] *Аванесов В.С.* Композиция тестовых заданий. М.: Центр тестирования, 2002. С. 160.

1	1	1	0	1	1	1	1	1	1
1	1	0	1	1	1	1	1	1	0
1	1	1	1	0	1	1	0	1	0
1	1	1	1	0	1	0	1	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	0	0	1	0	0	0
1	1	0	1	1	0	1	0	0	0
1	1	1	1	1	0	0	0	0	0
1	0	1	0	1	1	0	0	0	0
0	1	1	0	0	0	0	1	0	1
1	1	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0

получают один балл, за неправильный ответ — ноль. В таблице по строкам располагаются фамилии или номера испытуемых, по столбцам — номера заданий.

Сложение этих элементов по строкам и по столбцам имеет педагогический смысл. Из последнего столбца табли-

Таблица 3. **Тестовые результаты**

№№	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_i
1.	1	1	1	0	1	1	1	1	1	1	9
2.	1	1	0	1	1	1	1	1	1	0	8
3.	1	1	1	1	0	1	1	0	1	0	7
4.	1	1	1	1	0	1	0	1	0	0	6
5.	1	1	1	1	1	1	0	0	0	0	6
6.	1	1	1	1	0	0	1	0	0	0	5
7.	1	1	0	1	1	0	1	0	0	0	5
8.	1	1	1	1	1	0	0	0	0	0	5
9.	1	0	1	0	1	1	0	0	0	0	4
10.	0	1	1	0	0	0	0	1	0	1	4
11.	1	1	1	0	0	0	0	0	0	0	3
12.	1	1	0	0	0	0	0	0	0	0	2
13.	1	0	0	0	0	0	0	0	0	0	1
R_j	12	11	9	7	6	6	5	4	3	2	65
W_j	1	2	4	6	7	7	8	9	10	11	
p_j	0,923	0,846	0,692	0,538	0,462	0,462	0,385	0,308	0,231	0,154	5
q_j	0,077	0,154	0,308	0,462	0,538	0,538	0,615	0,692	0,769	0,846	
$p_j q_j$	0,071	0,130	0,213	0,248	0,248	0,248	0,236	0,213	0,178	0,130	

цы видно, что больше правильных ответов у первого испытуемого, а меньше — у последнего. Элементы этого столбца указывают на значение тестового балла X, полученного каждым (любым) испытуемым , что обозначается символом Y_i.

Применение формулы (1) для данных последнего столбца (Y) табл. 3 даёт средний тестовый балл испытуемых:

$$M_y = \frac{\sum Y_i}{N} = \frac{65}{13} = 5.$$

В практике разработки тестов особое внимание уделяется расчёту статистик p_j и q_j.

Часто бывает необходимо получать так называемую *частную среднюю арифметическую*. Это средняя, рассчитанная по отдельным группам испытуемых, или иначе, на одном или нескольких подмножествах испытуемых. Например, если есть множество испытуемых (N=100) и надо рассчитать среднюю арифметическую на подмножестве (в группе) первых двадцати трёх человек, то тогда складываются баллы только этих испытуемых, а полученная сумма делится, естественно, на 23.

$$M = \frac{\sum_{i=1}^{23} X_i}{23}.$$

Соответственно, для расчёта частных средних используется формула с несколько изменённой символикой, поясняющей происхождение интересующей суммы. В общем виде расчёт частной средней арифметической в подмножестве j в отдельных группах записывается так:

$$M = \frac{\sum_{i=1}^{n} X_i}{n_j}, \qquad (2)$$

где под и над знаком Σ появились индексы, ориентирующие, с какой частной суммой исследователь имеет дело.

Это пример математически строгой записи формулы расчёта средних арифметических любого подмножества. Значения 1 и n_j,

стоящие под и над знаком $\sum_{i=1}^{n}$,

называются *пределами* суммирования. Они указывают на номера испытуемых (i), включённые в расчёт.

Второй пример: если понадобится найти сумму баллов в другой группе испытуемых, где имеется 10 человек (n_2 = 10), с номерами с 1 по 6, то тогда знак

суммирования приобретет вид $\sum_{i=1}^{6}$,

что говорит о том, что для суммирования берутся баллы именно этих испытуемых. Соответственно, в знаменателе формулы (2) появится число 6, равное числу испытуемых, включённых в данную группу.