**Статистика в обработке материалов психологических исследований**

Статистические методы применяются при обработке материалов психологических исследований для того, чтобы извлечь из тех коли­чественных данных, которые получены в экспериментах, при опросе и наблюдениях, возможно больше полезной информации. В частности, в обработке данных, получаемых при испытаниях по психологиче­ской диагностике, это будет информация об индивидуально-психоло­гических особенностях испытуемых. Психологические исследования обычно строятся с опорой на количественные данные.

Вот пример.

К школьному психологу обратился шестиклассник Саша Ю. с прось­бой испытать его двигательный темп. Его очень интересовал баскетбол, и он собирался вступить в баскетбольную команду, а баскетболист, не­сомненно, должен иметь высокий двигательный темп. Психолог разра­ботал план небольшого исследования. Он начал с того, что попросил Сашу так быстро, как он только может, ставить точки в центре кружков, нарисованных на листке бумаги. За одну минуту мальчик поставил 137 то­чек. Насколько этот темп характерен для него? Чтобы установить это, психолог попросил Сашу повторить эту пробу 25 раз. Действительно, некоторые результаты превышали первоначально полученное число, но некоторые оказались и поменьше. Психолог просуммировал все полу­ченные за 25 проб результаты, а сумму разделил на 25 — таким путем он получил среднее арифметическое по всем пробам. Это среднее ариф­метическое составило 141. Таков по этой пробе максимальный темп это­го мальчика. Можно ли считать этот темп высоким? Потребовался еще один шаг в исследовании. Психолог сформировал группу из 50 шести­классников, не отличающихся от Саши и друг от друга по возрасту более чем на полгода. С этими ребятами психолог также провел сначала по несколько тренировочных проб, чтобы получить надежные данные об их темпе, и, наконец, последнюю пробу для обработки.

Все эти данные в виде средних арифметических были построены в один порядковый ряд, который был разбит по десяткам (по децилям).

Сашины данные вошли в первый десяток с наиболее быстрыми резуль­татами. По этим количественным данным психолог сделал вывод о том, что мальчик обладает сравнительно высоким двигательным темпом, о чем и было ему сообщено.

Современная математическая статистика представляет собой боль­шую и сложную систему знаний. Нельзя рассчитывать на то, что каж­дый психолог овладеет этими знаниями. Между тем статистика нужна психологу постоянно в его повседневной работе. Специалисты-стати­стики разработали целый комплекс простых методов, которые совер­шенно доступны любому человеку, не забывшему то, что он выучил еще в средней школе.

В зависимости от требований, которые предъявляют к статистике различные области науки и практики, создаются пособия по геологи­ческой, медицинской, биологической, психологической статистике '.

В этом приложении даются простейшие методы статистики для пси­хологов. Все необходимые для их применения вычисления можно вы­полнять вручную или на компьютере. Уместное грамотное применение этих методов позволит практику и исследователю, во всяком случае проведя начальную обработку, получить общую картину того, что дают количественные результаты его исследований, оперативно проконт­ролировать ход исследований. В дальнейшем, если возникнет такая необходимость, материалы исследований могут быть переданы для более глубокой разработки специалисту-статистику на большой компьютер.

**Статистические шкалы**

Применение тех или других статистических методов определяется тем, к какой статистической шкале относится полученный материал. С. Стивенс предложил различать четыре статистические шкалы:

1. шкалу наименований (или номинальную);
2. шкалу порядка;
3. шкалу интервалов;
4. шкалу отношений.

Зная типические особенности каждой шкалы, нетрудно установить, к какой из них следует отнести подлежащий статистической обработ­ке материал.

**Шкала наименований.** К этой шкале относятся материалы, в ко­торых изучаемые объекты отличаются друг от друга по их качеству.

При обработке таких материалов нет никакой нужды в том, чтобы располагать эти объекты в каком-то порядке, исходя из их характери­стик. В принципе, объекты можно располагать в любой последователь­ности.

Вот пример: изучается состав международной научной конференции. Среди участников есть французы, англичане, датчане, немцы и русские. Имеет ли значение порядок, в котором будут расположены участники при изучении состава конференции? Можно расположить их по алфавиту, это удобно, но ясно, что никакого принципиального значения в этом распо­ложении нет. При переводе этих материалов на другой язык (а значит и на другой алфавит) этот порядок будет нарушен. Можно расположить национальные группы по числу участников. Но при сравнении этого ма­териала с материалом другой конференции найдем, что вряд ли этот порядок окажется таким же. Отнесенные к шкале наименований объек­ты можно размещать в любой последовательности в зависимости от цели исследования.

При статистической обработке такого рода материалов нужно счи­таться с тем, каким числом единиц представлен каждый объект. Име­ются весьма эффективные статистические методы, позволяющие по этим числовым данным прийти к научно значимым выводам (напри­мер, метод хи-квадрат).

**Шкала порядка.** Если в шкале наименований порядок следования изучаемых объектов практически не играет никакой роли, то в шкале порядка — это видно из ее названия — именно на эту последователь­ность переключается все внимание.

*К этой шкале в статистике относят такие исследовательские ма­териалы, в которых рассмотрению подлежат объекты, принадлежа­щие к одному или нескольким классам, но отличающиеся при их сравне­нии одного с другим — «больше-меньше», «выше-ниже»- и т. п.*

Проще всего показать типические особенности шкалы порядка, если об­ратиться к публикуемым итогам любых спортивных соревнований. В этих итогах последовательно перечисляются участники, занявшие соответ­ственно первое, второе, третье и следующие по порядку места. Но в этой информации об итогах соревнований нередко отсутствуют или отходят на второй план сведения о фактических достижениях спортсменов, а на первый план ставятся их порядковые места.

Допустим, шахматист Д. занял в соревнованиях первое место. Како­вы же его достижения? Оказывается, он набрал 12 очков. Шахматист Е. занял второе место. Его достижение — 10 очков. Третье место занял Ж. с восемью очками, четвертое — 3. с шестью очками и т. д. В сообщениях о соревновании разница в достижениях при размещении шахматистов отходит на второй план, а на первом остаются их порядковые места. В том, что именно порядковому месту отводится главное значение, есть свой смысл. В самом деле, в нашем примере З. набрал шесть, а Д. — 12 очков. Это абсолютные их достижения — выигранные ими партии. Если попытаться истолковать эту разницу в достижениях чисто арифме­тически, то пришлось бы признать, что 3. играет вдвое хуже, чем Д. Но с этим нельзя согласиться. Обстоятельства соревнований не всегда про­сты, как не всегда просто и то, как провел их тот или другой участник. Поэтому, воздерживаясь от арифметической абсолютизации, ограничи­ваются тем, что устанавливают: шахматист 3. отстает от занявшего пер­вое место Д. на три порядковых места.

**Шкала интервалов.** *К ней относятся такие материалы, в которых дана количественная оценка изучаемого объекта в фиксированных еди­ницах.*

Вернемся к опытам, которые провел психолог с Сашей. В опытах учиты­валось, сколько точек могут поставить, работая с максимально доступ­ной им скоростью, сам Саша и каждый из его сверстников. Оценочными единицами в опытах служило число точек. Подсчитав их, исследователь получил то абсолютное число точек, которое оказалось возможным по­ставить за отведенное время каждому участнику опытов. Главная труд­ность при отнесении материалов к шкале интервалов состоит в том, что нужно располагать такой единицей, которая была бы при всех повтор­ных измерениях тождественной самой себе, т. е. одинаковой и неизмен­ной. В примере с шахматистами (шкала порядка) такой единицы вообще не существует.

В самом деле, учитывается число партий, выигранных каждым участ­ником соревнований. Но ясно, что партии далеко не одинаковы. Воз­можно, что участник соревнований, занявший четвертое место — он выиграл шесть партий, — выиграл труднейшую партию у самого лидера! Но в окончательных итогах как бы принимается, что все выигранные партии одинаковы. В действительности же этого нет. Поэтому при рабо­те с подобными материалами уместно их оценивать в соответствии с требованиями шкалы порядка, а не шкалы интервалов. Материалы, соответствующие шкале интервалов, должны иметь единицу измерения.

**Шкала отношений.** *К этой шкале относятся материалы, в которых учитываются не только число фиксированных единиц, как в шкале ин­тервалов, но и отношения полученных суммарных итогов между собой.* Чтобы работать с такими отношениями, нужно иметь некую абсолют­ную точку, от которой и ведется отсчет. При изучении психологиче­ских объектов эта шкала практически неприменима.

**О параметрических и непараметрических методах статистики**

Приступая к статистической обработке своих исследований, психо­лог должен решить, какие методы ему более подходят по особенностям его материала — параметрические или непараметрические. Раз­личие между ними легко понять.

Ранее уже говорилось об измерении двигательной скорости детей-шес­тиклассников.

Как обработать эти данные?

Нужно записать все произведенные измерения — в данном случае это будет число точек, поставленных каждым испытуемым, — затем вычис­лить для каждого испытуемого среднее арифметическое по его резуль­татам. После этого расположить все данные в их последовательности, например начиная с наименьших к наибольшим. Для облегчения обозри­мости этих данных их обычно объединяют в группы; в этом случае можно объединить по 5-9 измерений в группе. Вообще же при таком объеди­нении желательно, если общее число случаев не более ста, чтобы общее число групп было порядка двенадцати.

Далее нужно установить, сколько раз в опытах встретились числовые значения, соответствующие каждой группе. Сделав это, для каждой группы записать ее численность. Полученные в такой таблице данные носят назва­ние распределения численностей или частот. Рекомендуется предста­вить это распределение в виде диаграммы, на которой изображается по­лигон распределения, или гистограмма распределения. Контуры этого полигона помогут решить вопрос о статистических методах обработки.

Нередко эти контуры напоминают контуры колокола, с наивысшей точкой в центре полигона и с симметричными ветвями, отходящими в ту и другую сторону. Такой контур соответствует кривой нормально­го распределения. Это понятие было введено в математическую ста­тистику К. Ф. Гауссом (1777-1855), поэтому кривую именуют также **кривой Гаусса**. Он же дал математическое описание этой кривой. Для построения кривой Гаусса (или кривой нормального распределения) теоретически требуется бесчисленное количество случаев. Практиче­ски же приходится довольствоваться тем фактическим материалом, который накоплен в исследовании. Если данные, которыми распола­гает исследователь, при их внимательном рассмотрении или после пе­реноса их на диаграмму лишь в незначительной степени расходятся с кривой нормального распределения, то это дает право исследователю применять в статистической обработке параметрические методы, ис­ходные положения которых основываются на нормальной кривой рас­пределения Гаусса.

Нормальное распределение называют параметрическим потому, что для построения и анализа кривой Гаусса достаточно иметь всего два параметра: среднее значение, которое должно соответствовать высоте перпендикуляра, восстановленного в центре кривой, и так называемое среднее квадратическое, или стандартное, отклонение величины, ха­рактеризующей рассеивание значений вокруг среднего значения; о спо­собах вычисления той и другой величины будет рассказано ниже.

Параметрические методы обладают для исследователя многими преимуществами, но нельзя забывать о том, что применение их право­мерно только тогда, когда обрабатываемые данные показывают рас­пределение, лишь несущественно отличающееся от гауссовского.

При невозможности применить параметрические надлежит обра­титься к **непараметрическим методам**. Эти методы успешно разраба­тывались в последние 3-4 десятилетия, и их разработка была вызвана прежде всего потребностями ряда наук, в частности психологии. Они показали свою высокую эффективность. Вместе с тем они не требуют сложной вычислительной работы.

Современному психологу-исследователю нужно исходить из того, что «...имеется большое количество данных, которые либо вообще не поддаются анализу с помощью кривой нормального распределения, либо не удовлетворяют основным предпосылкам, необходимым для ее использования».

**Генеральная совокупность** и **выборка**. Психологу постоянно при­ходится иметь дело с этими двумя понятиями.

**Генеральная совокупность, или просто совокупность, — это мно­жество достаточно большого объема, все элементы которого обла­дают какими-то общими признаками.**

Так, все подростки-шестиклассники 12 лет (от 11,5 до 12,5) образу­ют совокупность. Дети того же возраста, но не обучающиеся в школе или же обучающиеся, но не в шестых классах, не подлежат включению в эту совокупность.

В ходе конкретизации проблем своего исследования психологу не­избежно придется обозначить границы изучаемой им совокупности.

Следует ли включать в изучаемую совокупность детей того же воз­раста, но обучающихся в колледжах, гимназиях, лицеях и других по­добных учебных заведениях?

В ответе на этот и другие такие же вопросы может помочь статистика.

В подавляющем большинстве случаев исследователь не в состоя­нии охватить в изучении всю совокупность. Приходится, хотя это и связано с некоторой утратой информации, взять для изучения лишь часть совокупности, ее и называют выборкой. Задача исследователя заключается в том, чтобы подобрать такую выборку, которая репре­зентировала бы, представляла совокупность; другими словами, при­знаки элементов совокупности должны быть представлены в выборке. Это достигается, прежде всего, использованием случайной выборки из совокупности. Составить такую выборку, в точности повторяющую все разнообразные сочетания признаков, которые имеются в элемен­тах совокупности, вряд ли возможно. Поэтому некоторые потери в информации оказываются неизбежными. Важно, чтобы были сохра­нены в выборке существенные с точки зрения данного исследования признаки совокупности. Возможны случаи, и для их обнаружения есть статистические методы, когда задачи исследования требуют создания двух выборок одной совокупности; при этом нужно установить, не взя­ты ли выборки из равных совокупностей. Эти и другие подобные ка­зусы нужно иметь в виду психологу при обработке результатов выбо­рочных исследований.

Следует рассмотреть типы задач, с которыми чаще всего имеет дело психолог. Соответственно приводятся и статистические методы, которые приложимы для обработки психологических материалов, на­правленных на решение этих задач.

**Первый тип задач.** Данный тип задач представлен в ситуации, когда психологу *нужно дать сжатую и достаточно информативную харак­теристику психологических особенностей какой-то выборки*, например школьников определенного класса. Чтобы подойти к решению этой задачи, необходимо располагать; результатами диагностических испы­таний; эти испытания, разумеется, следует заранее спланировать так, чтобы они давали информацию о тех особенностях группы, которые в этом конкретном случае интересуют психолога. Это могут быть осо­бенности умственного развития, психофизиологические особенности, данные об изменении работоспособности и т. д.

Получив все экспериментальные результаты и материалы наблю­дений, следует подумать о том, как их подать пользователю в компакт­ном виде, чтобы при этом свести к минимуму потерю информации. В перечне статистических методов, используемых при решении подоб­ных задач, обычно находят свое место и параметрические, и непара­метрические методы; о возможностях применения тех и других, как было сказано выше, судят по самому полученному материалу. Об этих статистических методах и их использовании пойдет речь далее.

**Второй тип задач.** Это, пожалуй, наиболее часто встречающиеся задачи в исследовательской и практической деятельности психолога: *сравниваются между собой несколько выборок, чтобы установить, яв­ляются ли выборки независимыми или принадлежат одной и той же совокупности.* Так, проведя эксперименты в восьмых классах двух раз­личных школ, психолог сравнивает эти выборки между собой.

К этому же типу относятся *задачи с определением тесноты связи двух рядов показателей, полученных на одной и той же выборке*; в та­кой обработке чаще всего применяют метод корреляций.

**Третий тип задач.** Это *задачи, в которых обработке подлежат вре­менные ряды, ряды, в которых расположены показатели, меняющиеся во времени; их называют также динамическими рядами.* В предшеству­ющих типах задач фактор времени не принимался во внимание, и ма­териал анализировался так, как будто он весь поступил в руки иссле­дователя в одно и то же время. Такое допущение можно оправдать тем, что за тот короткий период времени, который был затрачен на собира­ние материала, он не претерпел существенных перемен. Но психологу приходится работать и с таким материалом, в котором наибольший интерес представляют как раз его изменения во времени. Допустим, психолог намерен изучить изменение работоспособности школьников в течение учебной четверти. В этом случае информативными будут показатели, по которым можно судить о динамике работоспособнос­ти. Берясь за такой материал, психолог должен понимать, что при ана­лизе динамических рядов нет смысла пользоваться средним арифме­тическим ряда, так как среднее арифметическое замаскирует нужную информацию о динамике.

В основном тексте книги упоминалось о *лонгитюдинальном иссле­довании*, т. е. таком, в котором однообразный по содержанию психоло­гический материал по одной выборке собирается в течение длитель­ного времени. Показатели лонгитюда — это также динамические ряды, и при их обработке следует пользоваться методами, предназначенны­ми для таких рядов.

**Четвертый тип задач.** Это *задачи, возникающие перед психологом, за­нимающимся конструированием диагностических методик, проверкой и обработкой результатов их применения.* Отчасти об этих задачах уже говорилось в других главах, но не уделялось внимания специально ста­тистике. Психологическая диагностика, в особенности тестология, имеет целый ряд канонических правил, применение которых должно обеспечивать высокое качество информации, получаемой посредством диагностических методик. Так, методика должна быть *надежной, гомогенной, валидной.* По упрочившимся в тестологии правилам все эти свойства проверяются статистическими методами.

Выше были перечислены типы задач, с которыми чаще всего встре­чаются психологи. Теперь мы перейдем к изложению конкретных статистических методов, способствующих успешному решению пере­численных задач. Но прежде следует высказать некоторые соображе­ния о возможностях статистики в проведении психологического ис­следования.

*Назначение статистики состоит в том, чтобы извлечь из этих материалов боль­ше полезной информации. Вместе с тем статистика показывает, что эта информация не случайна и что добытые данные имеют определен­ную и значимую вероятность.*

Статистические методы раскрывают связи между изучаемыми яв­лениями. Однако необходимо твердо знать, что, как бы ни была высока вероятность таких связей, они не дают права исследователю признать их причинно-следственными отношениями. Статистика, например, утверждает, что существует значимая связь между двигательной ско­ростью и игрой в теннис. Но отсюда еще не вытекает, будто двигатель­ная скорость и есть причина успешной игры. Нельзя, по крайней мере в некоторых случаях, исключить и того, что сама двигательная ско­рость явилась следствием успешной игры.

Чтобы подтвердить или отвергнуть существование причинно-след­ственных отношений, исследователю зачастую приходится продумы­вать целые серии экспериментов. Если они будут правильно постро­ены и проведены, то статистика поможет извлечь из результатов этих экспериментов информацию, которая необходима исследователю, что­бы либо обосновать и подтвердить свою гипотезу, либо признать ее недоказанной.

**Статистические методы, примеры их применения для принятия решения**

**Первый тип задач**

Допустим, что школьному психологу нужно представить краткую информацию о развитии психомоторных функций учащихся шестых классов. В этих классах обучается 50 учеников. В процессе выполнения своей программы психолог провел диагностическое изучение дви­гательной скорости, применив ранее описанную методику (описание дано на первой странице данного раздела).

Для реализации своей программы психологу надлежало получить количественные характеристики, свидетельствующие о состоянии изучаемой функции — ее центральной тенденции, величины, показы­вающей размах колебания, в пределах которого находятся данные от­дельных учеников, и то, как распределяются эти данные. Какими ме­тодами вести обработку, зависит от того, в какой статистической шкале измерены значения исследуемого признака. Визуальное озна­комление с полученными данными показывает, что возможно вычис­ление **среднего арифметического**, *выражающего центральную тен­денцию*, и **среднеквадратического отклонения**, *показывающего размах и особенности варьирования экспериментальных результатов.*

Нельзя ограничиться вычислением только среднего арифметиче­ского, так как оно не дает полных сведений об изучаемой выборке.

Вот пример.

В одном купе вагона поместилась бабушка 60 лет с четырьмя внука­ми: один — 4 лет, двое — по 5 лет и один — 6 лет. Среднее арифметиче­ское возраста всех пассажиров этого купе 80/5= 16.

В другом купе расположилась компания молодежи: двое — 15-летних, один — 16-летний и двое — 17-летних. Средний возраст пассажиров это­го купе также равен 80/5= 16. Таким образом, по средним арифметическим пассажиры этих купе как бы и не отличаются. Но если обратиться к особенностям варьирования, то сразу можно установить, что в одном купе возраст пассажиров варьируется в пределах 56 единиц, а во вто­ром — в пределах 2.

Для вычисления среднего арифметического применяется формула:

" х = ∑ х / n

а для среднеквадратического отклонения формула:

σ = √∑ (х - " х )2 / n

В этих формулах "х означает среднее арифметическое, х — каждую величину изучаемого ряда, ∑ означает сумму; σ означает среднеквадратическое отклонение; буквой n обозначают число членов изучаемо­го ряда.

Ниже представлен весь ход его обработки.

В опытах участвовало 50 испытуемых. Каждый из них выполнил 25 проб, по 1 мин каждая. Вычислено среднее для каждого испытуемого. Полу­ченный ряд упорядочен, и все индивидуальные результаты представле­ны в последовательности от меньшего к большему.

85-93-93-99-101-105-109-110-111-115-115-116-116-117-117-117-118-119-121-121-122-124-124-124-124-125-125-125-127-127-127-127-127-128-130-131-132-132-133-134-134-135-138-138-140-143-144-146-150-158.

Для удобства дальнейшей обработки эти первичные данные соеди­нены в группы. Благодаря группировке отчетливее выступает присущее данному ряду распределение величин и их численностей. Отчасти упро­щается и вычисление среднего арифметического и среднеквадратиче­ского отклонения. Этим компенсируется количественное искажение ин­формации, неизбежное при вычислениях на сгруппированных данных.

При выборе группового интервала следует принять во внимание такие соображения. Если ряд не очень велик, например содержит до 100 элементов, то и число групп не должно быть очень велико, напри­мер порядка 8-12. Желательно, чтобы при группировании начальная величина — при соблюдении последовательности от меньшей величи­ны к большей — была меньше самой меньшей величины ряда, а самая большая — больше самой большой величины изучаемого ряда. Если ряд, как в данном случае, начинается с 85, группирование нужно на­чать с меньшей величины, а поскольку ряд завершается числом 158, то и группирование должно завершаться большей величиной. В ряду, который нами изучается, с учетом высказанных соображений можно выбрать групповой интервал в 9 единиц и произвести разбивку ряда на группы, начав с 83. Тогда последняя группа будет завершаться ве­личиной, превышающей значение последней величины ряда (т. е. 159). Число групп будет равно 9. В табл. 1 представлены группы в их после­довательности и все другие величины для вычисления среднего ариф­метического и среднеквадратического отклонения. Таблица состоит из 8 столбцов.

1-й столбец — группы, полученные после разбиения изучаемого ряда.

2-й столбец — средние значения интервалов по каждой группе.

3-й столбец показывает результаты «ручной» разноски величин ряда или иксов (каждая величина занесена в соответствующую ее зна­чению группу в виде черточки).

4-й столбец — итог подсчета результатов разноски.

5-й столбец — произведения величин 2-го столбца на величины 4-го столбца по строчкам. Итоги 4-го и 5-го столбцов дают суммы, необхо­димые для вычисления среднего арифметического.

Таблица 1

**Вычисление среднего арифметического и среднеквадратического**

**отклонения**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Границы интерва­лов | Средние интер­валов х | Резуль­тат  разно­ски | Итоги  разно­ски | f \*х | х – "х | (х - " х )2 | f \*(х - "х)2 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 83-91 | 87 | I | 1 | 87 | -36 | 1296 | 1296 |
| 92-100 | 96 |  | 3 | 288 | -27 | 729 | 2187 |
| 101-109 | 105 |  | 3 | 315 | -18 | 324 | 972 |
| 110-118 | 114 |  | 10 | 1140 | -9 | 81 | 810 |
| 119-127 | 123 |  | 16 | 1968 | 0 | 0 | 0 |
| 128-136 | 132 |  | 9 | 1188 | 9 | 81 | 729 |
| 137-145 | 141 |  | 5 | 705 | 18 | 324 | 1620 |
| 146-154 | 150 |  | 2 | 300 | 27 | 729 | 1458 |
| 155-163 | 159 | I | 1 | 159 | 36 | 1296 | 1296 |

n = 50 ; ∑f \* х = 6150 ; ∑f \*(х - " х )2 = 10368

6-й столбец показывает построчные разности между значениями х 2-го столбца и средним арифметическим "х.

7-й столбец — квадрат этих разностей.

8-й столбец показывает построчные произведения значений 4-го и 7-го столбцов. Суммирование величин этого столбца дает итог, не­обходимый для вычисления среднеквадратического отклонения.

Включение буквы f, означающей, насколько часто встречалась та или другая величина, ничего не изменяет в формулах среднего ариф­метического и среднеквадратического отклонения. Поэтому формулы

" х = ∑х/ n = ∑f \*х/ n

Как и формулы вполне тождественны.

σ = √∑ (х - " х )2 / n = √∑f \* (х - " х )2 / n

Остается показать, как вычисляются по формулам среднее арифме­тическое и среднеквадратическое отклонение. Обратимся к величи­нам, полученным в табл. 1:

" х = 6150/50 = 123

При составлении табл. 1 это число было заранее вычислено, без него нельзя было бы получить числовые значения 6, 7 и 8-го столбцов таблицы.

σ = √10368/50 = √207,3 = 14,4

При обработке изучаемого ряда оказалось возможным применение параметрического метода; визуально можно заметить, что распределе­ние численностей приближается к нормальному.

Нормальное распределение обладает некоторыми весьма полезны­ми для исследователя свойствами. Так, в границах "х ± σ находится примерно 68 % всего ряда или всей выборки. В границах "х ± 2σ нахо­дится примерно 95 %, а в границах "х ± 3σ - 99,7 % выборки. В практи­ке исследований часто берут границы "х ± 2/3σ. В этих границах при нормальном распределении будут находиться 50 % выборки; распре­деление это симметрично, поэтому 25 % окажутся ниже, а 25 % выше гра­ниц "х ± 2/3σ. Все эти расчеты не требуют никакой дополнительной проверки при условии, что изучаемый ряд имеет нормальное распре­деление, а число элементов в нем велико, порядка нескольких сотен или тысяч.

Для рассматриваемого примера необходимо также вычислить ко­эффициент вариации по формуле:

V = σ/ "х ·100 %.

В примере, который был рассмотрен выше,

V = 14,4/123 ·100% = 11,7%.

Выполнив все эти вычисления, психолог может представить инфор­мацию об изучении двигательной скорости с помощью примененной методики в шестых классах. Согласно результатам изучения в шестых классах, получены:

* + среднее арифметическое — 123;
  + среднеквадратическое отклонение — 14,4;
  + коэффициент вариации — 11,7 %.

Если значения изучаемого признака измерены в порядковой шкале, то в качестве меры центральной тенденции выступает медиана, а ха­рактеристикой диапазона варьирования выступает среднее кварталь­ное отклонение.

Вот пример.

После проведения диагностических испытаний уровня умственного развития учеников шестого класса все полученные данные были упоря­дочены, т. е. расположены в последовательности от меньшей величины к большей. Испытания проходили 18 учащихся. Буквами обозначены уча­щиеся, числами — полученные ими баллы по тесту, столбцы под буква­ми R — ранги (табл. 2).

Процедура ранжирования состоит в следующем. Все числа ряда в их последовательности получают по своим порядковым местам присва­иваемые им ранги. Если какие-нибудь числа повторяются, то всем по­вторяющимся числам присваивается один и тот же ранг — средний из общей суммы занятых этими числами мест. Так, числу «28» в изучаемом ряду присвоен ранг «2». Затем следуют трижды повторяющиеся числа «39». На них приходятся занятые ими ранговые места «3», «4», «5». По­этому этим числам присваивается один и тот же средний ранг, в данном случае — «4». Поскольку места до 5 включительно заняты, то следующее число получает ранг «6» и т. д.

Таблица 2

**Ранжирование результатов**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Учащиеся | Баллы по тесту | Ранг (R) | Учащиеся | Баллы по тесту | Ранг (R) |
| А | 25 | 1 | К | 68 | 10 |
| Б | 28 | 2 | Л | 69 | 11,5 |
| В | 39 | 4 | м | 69 | 11,5 |
| Г | 39 | 4 | н | 70 | 14,5 |
| д | 39 | 4 | О | 70 | 14,5 |
| Е | 45 | 6 | п | 70 | 14,5 |
| Ж | 50 | 7 | р | 70 | 14,5 |
| 3 | 52 | 8,5 | с | 74 | 17,5 |
| И | 52 | 8,5 | т | 74 | 17,5 |

При обработке ряда, не имеющего признаков нормального распре­деления, иначе — непараметрического ряда, — для величины, которая выражала бы его центральную тенденцию, более всего пригодна меди­ана, т. е. величина, расположенная в середине ряда. Ее определяют по срединному рангу по формуле.

Медиана ряда определяется по ранговой медиане:

MeR = (n +1)/2

где n — число членов ряда.

Возьмем, к примеру, ряд в семь членов: 3-5-6-7-9-10-11.

Проранжировав этот ряд, имеем:

1-2-3-4-5-6-7.

Ранговая медиана

MeR = (7 + 1)/2 = 4 ,

дает медиану рассматриваемого ряда Me = 7.

Возьмем ряд в восемь членов: 3-5-6-7-9-10-11-12.

Проранжировав этот ряд, имеем:

1-2-3-4-5-6-7-8.

Ранговая медиана в этом ряду равна:

MeR = (8+1)/2 = 4,5

Этому рангу соответствует середина между двумя величинами, име­ющими ранг 4 и ранг 5, т. е. между 7 и 9. Медиана этого ряда равна:

Me = (7 + 9)/2 =8

Следует обратить внимание на то, что величины 8 в составе ряда пет, но таково значение медианы этого ряда.

Вернемся к изучаемому ряду. Он состоит из 18 членов. Его ранго­вая медиана равна:

MeR = (18+1)/2= 9,5.

Она расположится между 9-й и 10-й величиной ряда. 9-я величина ряда - 52, 10-я величина ряда - 68. Медиана занимает срединное ме­сто между этими величинами, следовательно:

Me = (52 + 68)/2 = 60

По обе стороны от этой величины находится по 50% величин ряда. Характеристику распределения численностей в непараметрическом ряду можно получить из отношения его квартилей. Квартилью назы­вается величина, отграничивающая 1/4 всех величин ряда. Квартиль первая - ее обозначение Q1- вычисляется по формуле:

Q1 = R1 + Rn/2(лев) / 2

Это полусумма первого и последнего рангов первой, левой от меди­аны половины ряда; квартиль третья, обозначаемая Q3, вычисляется, по формуле:

Q3 = Rn/2 + Rn/2(прав) / 2

т. е. как полусумма первого и последнего рангов второй, правой от ме­дианы половины ряда. Берутся порядковые значения рангов по их пос­ледовательности в ряду. В обрабатываемом ряду

Q1 = (1+9)/2 = 5, Q3 = (10+18)/2 = 14

Рангу 5 в этом ряду соответствует величина 39, а рангу 14 - вели­чина 70.

Для характеристики распределения в непараметрическом ряду вы­числяется среднее квартальное отклонение, обозначаемое Q.

Формула для Q такова:

Q = (Q3 - Q1)/2

В обрабатываемом ряду Q3 = 70, a Q1 = 39, следовательно:

Q = (70 – 39)/2 =15,5.

Были рассмотрены статистическая обработка параметрического ряда ("х и σ) и статистическая обработка непараметрического ряда (Me и Q). Параметрический ряд относится к шкале интервалов, непараметричес­кий — к шкале порядка. Но встречаются также ряды, относящиеся к шкале наименований. Наиболее краткая, но малоинформативная ха­рактеристика такого ряда может быть получена с помощью моды — величины в ряду, имеющей наибольшую численность из числа п — чле­нов ряда. Следует заметить, что моду можно лишь условно считать вы­ражением центральной тенденции в ряду, относящемуся к шкале наи­менований. Она выражает наиболее типичную величину ряда.

Рассмотрим пример, где речь идет об участниках некой конференции; в их числе 3 англичанина, 2 датчанина, 5 немцев, 1 русский и 2 фран­цуза. Мода в данном ряду приходится на участников конференции — немцев. Число членов ряда — 13, а мода Мо = 5.