**Раздел 3. СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ**

*ММНИмаг-18 Лекция №4=2ч.*

**Тема 3.1. Обработка результатов экспериментальных исследований**

*3.1.1. Методы статистической обработки данных*

*3.1.2. Основы теории случайных ошибок и методов оценки случайных погрешностей в измерениях*

*3.1.3. Методы графической обработки результатов измерений*

*3.1.4. Оформление результатов научного исследования*

***3.1.1. Методы статистической обработки данных***

Методами статистической обработки результатов экспери­мента называются математические приемы, формулы, способы количественных расчетов, с помощью которых показатели, по­лучаемые в ходе эксперимента, можно обобщать, приводить в си­стему, выявляя скрытые в них закономерности.

Речь идет о та­ких закономерностях статистического характера, которые су­ществуют между изучаемыми в эксперименте переменными ве­личинами.

*Данные* – это основные элементы, подлежащие классифицированию или разбитые на категории с целью обработки.

Некоторые из методов математико-статистического анализа позволяют вычислять так называемые элементарные математические статистики, характеризующие выборочное распреде­ление данных, например:

- выборочное среднее,

- выборочная диспер­сия,

- мода,

- медиана и ряд других.

Иные методы математической статистики позволяют судить о динамике изменения отдельных статис­тик выборки, например:

- дисперсионный анализ,

- регрессионный ана­лиз.

С помощью третьей группы методов выборочных данных, можно достоверно судить о статистических связях, существующих между переменными величинами, кото­рые исследуют в данном эксперименте:

- кор­реляционного анализа;

- факторного анализа;

- методов сравнения.

Все методы математико-статистического анализа условно де­лятся на первичные и вторичные27.

Первичными называют мето­ды, с помощью которых можно получить показатели, непосред­ственно отражающие результаты производимых в эксперимен­те измерений.

Вторичными называются методы статистической обработки, с помощью которых на базе первичных данных выявляют скры­тые в них статистические закономерности.

К первичным методам статистической обработки относят, на­пример:

- определение выборочной средней величины;

- выбороч­ной дисперсии;

- выборочной моды;

- выборочной медианы.

В чис­ло вторичных методов обычно включают:

- корреляционный ана­лиз;

- регрессионный анализ;

- методы сравнения первичных ста­тистик у двух или нескольких выборок.

Рассмотрим методы вычисления элементарных математичес­ких статистик, начав с выборочного среднего.

*Среднее арифметическое значение* ***–*** это отношение суммы всех значений данных к числу слагаемых.

Среднее значение как статистический показатель представляет собой среднюю оценку изучаемого в эксперименте психологического качества.

Эта оценка характеризует степень его развития в целом у той группы испытуемых, которая была под­вергнута психодиагностическому обследованию. Сравнивая не­посредственно средние значения двух или нескольких выборок, мы можем судить об относительной степени развития у людей, составляющих эти выборки, оцениваемого качества.

Выборочное среднее определяется при помощи следующей формулы:

 (1)

где хср —выборочная средняя величина или среднее арифметичес­кое значение по выборке;

п — количество испытуемых в выбор­ке или частных психодиагностических показателей, на основе ко­торых вычисляется средняя величина;

xk — частные значения по­казателей у отдельных испытуемых. Всего таких показателей п, поэтому индекс k данной переменной принимает значения от 1 до п;

∑ — принятый в математике знак суммирования величин тех переменных, которые находятся справа от этого знака.

*Дисперсия* – это мера разброса данных относительно среднего значения30.

Чем больше дисперсия, тем больше отклонения или разброс данных. Ее определяют для того, чтобы можно было отличать друг от друга величины, име­ющие одинаковую среднюю, но разный разброс.

Дисперсия определяется по следую­щей формуле:

 (2)

где — выборочная дисперсия, или просто дисперсия;

— выражение, означающее, что для всех xk от перво­го до последнего в данной выборке необходимо вычислить раз­ности между частными и средними значениями, возвести эти раз­ности в квадрат и просуммировать;

п — количество испытуемых в выборке или первичных зна­чений, по которым вычисляется дисперсия.

*Медианой* называется значение изучаемого признака, кото­рое делит выборку, упорядоченную по величине данного призна­ка, пополам.

Знание медианы полезно для того, чтобы установить, явля­ется ли распределение частных значений изученного признака симметричным и приближающимся к так называемому нормаль­ному распределению. Средняя и медиана для нормального рас­пределения обычно совпадают или очень мало отличаются друг от друга.

Если выборочное распределение признаков нормаль­но, то к нему можно применять методы вторичных статистичес­ких расчетов, основанные на нормальном распределении данных. В противном случае этого делать нельзя, так как в расчеты могут вкрасться серьезные ошибки.

*Мода* еще одна элементар­ная математическая статистика и характеристика распределе­ния опытных данных. Модой называют количественное зна­чение исследуемого признака, наиболее часто встречающееся в выборке.

Для симметричных распределений признаков, в том числе для нормального распределения, значения моды совпадают со значений среднего и медианы. Для других типов распре­делений, несимметричных, это не характерно.

Метод вторичной статистической обработки, по­средством которого выясняется связь или прямая зависимость между двумя рядами экспериментальных данных, носит назва­ние *метод корреляционного анализа.* Он показывает, каким образом одно яв­ление влияет на другое или связано с ним в своей динамике. По­добного рода зависимости существуют, к примеру, между вели­чинами, находящимися в причинно-следственных связях друг с другом. Если выясняется, что два явления статистически досто­верно коррелируют друг с другом и если при этом есть уверен­ность в том, что одно из них может выступать в качестве причи­ны другого явления, то отсюда определенно следует вывод о на­личии между ними причинно-следственной зависимости.

Имеется несколько разновидностей данного метода:

Линейный корреля­ционный анализ позволяет устанавливать прямые связи между переменными величинами по их абсолютным значениям. Эти связи графически выражаются прямой линией, отсюда название «линейный».

Коэффициент линейной корреляции определяется при по­мощи следующей формулы:

 (3)

где rxy *—* коэффициент линейной корреляции;

*х, у -* средние выборочные значения сравниваемых величин;

*хi,уi —* частные выборочные значения сравниваемых величин;

*п —* общее число величин в сравниваемых рядах показателей;

— дисперсии, отклонения сравниваемых величин от средних значений.

Ранговая корреляция определяет зависимость не между абсолютными значениями переменных, а между поряд­ковыми местами, или рангами, занимаемыми ими в упорядочен­ном по величине ряду. Формула коэффициента ран­говой корреляции следующая32:

 (4)

где Rs — коэффициент ранговой корреляции по Спирмену;

*di —* разница между рангами показателей одних и тех же ис­пытуемых в упорядоченных рядах;

*п —* число испытуемых или цифровых данных (рангов) в кор­релируемых рядах.

***3.1.2. Основы теории случайных ошибок и методов оценки случайных погрешностей в измерениях***

Исследователь должен одновременно с производством опытов и измерений проводить предварительную, а затем и окончательную обработку результатов измерений, их анализ, что позволяет корректировать эксперимент, контролировать и улучшать методику в ходе опыта. Анализ случайных погрешностей основывается на теории случайных ошибок. Он даёт возможность с определенной гарантией вычислить действительное значение измеренной величины и оценить возможные ошибки. Основу теории случайных ошибок составляют следующие предположения:

– большие погрешности встречаются реже, чем малые, так как вероятность появления погрешности уменьшается с ростом ее величины;

 – при большом числе измерений случайные погрешности одинаковой величины, но разного знака встречаются одинаково часто;

– при бесконечно большом числе измерений истинное значение измеряемой величины равно среднеарифметическому значению всех результатов измерений, а появление того или иного результата измерения как случайного события описывается нормальным законом распределения.

Совокупность измерений может быть генеральной и выборочной. *Генеральная совокупность* – это все множество возможных значений изменений хi или возможных значений погрешности Δ хi.

При *выборочной совокупности* число измерений n ограничено и в каждом случае строго определяется. Обычно считают, что если n > 30, то среднее значение совокупности измерений x достаточно точно приближается к истинному значению.

Теория случайных ошибок позволяет оценить точность и надежность измерения при данном количестве замеров или определить минимальное количество замеров, гарантирующее требуемую точность и надежность измерений. Также необходимо исключить возможность появления грубых ошибок и определить достоверность полученных результатов.

Обработка опытов, как правило, проводится в несколько этапов:

- предварительный, заключающийся в просмотре всех экспериментальных данных, при котором отмечают все резко отличающиеся величины, и убеждаются в том, что процесс не выходит за рамки регистрирующих устройств;

- основной, во время которого определяются точечные и интервальные оценки исследуемых величин и точность измерений, возможные взаимодействия факторов, коэффициенты уравнения;

- заключительный – позволяющий сформулировать статистические выводы о существенности отличий оценок измерений, проверке нулевых гипотез, адекватности исследуемых моделей.

Еще при постановке экспериментов необходимо учитывать, что обработка результатов в настоящее время производится с помощью компьютеров и, следовательно, объем обрабатываемой информации может быть большим, что позволяет снизить уровень возможной погрешности в оценке опытов.

 ***3.1.3. Методы графической обработки результатов измерений***

При обработке результатов измерений широко используют методы графического изображения. Такие методы дают более наглядное представление о результатах эксперимента, чем табличные данные. Поэтому чаще табличные данные обрабатывают графическими методами с использованием обычной прямоугольной системы координат. Чтобы построить график, необходимо хорошо знать ход исследования, течение исследовательского процесса, т.е. то, что можно взять из теоретических исследований. Экспериментальные точки на графике необходимо соединять плавной линией, чтобы она проходила как можно ближе ко всем экспериментальным точкам. Но могут быть исключения, так как иногда исследуют явления, для которых в определенных интервалах наблюдается быстрое скачкообразное изменение одной из координат.

В таких случаях необходимо плавно соединять точки кривой. Общее «осреднение» всех точек плавной кривой может привести к тому, что скачок функции подменяется погрешностями измерений. Иногда исследуются явления, для которых в определенном интервале наблюдается скачкообразное изменение одной из координат, объясняемое сущностью физико-химического процесса. Если при построении графика появляются точки, которые резко удаляются от плавной кривой, необходимо проанализировать причину этого отклонения, а затем повторить измерение в диапазоне резкого отклонения точки. Повторные измерения могут подтвердить или отвергнуть наличие такого отклонения. Если измеряемая величина является функцией двух переменных параметров (х, у), то в одних координатах можно построить несколько графиков, разбив диапазон измерения одного из параметров на несколько отрезков у1, у2…уп.

Иногда при графическом изображении результатов эксперимента приходится иметь дело с тремя переменными b = f(x,y,z). В таком случае применяют метод разделения переменных. Одной из величин z в пределах интервала измерения z1 – zп задают несколько последовательных значений. Для двух остальных переменных x и y строят графики y = f1(x), при z1 = cоnst. В результате на одном графике получают семейство кривых y = f1(x) для различных значений z.

Также при графическом изображении результатов экспериментов существенную роль играет выбор системы координат или координатные сетки. Они бывают равномерными и неравномерными. У равномерных координатных сеток ординаты и абсциссы имеют равномерную шкалу. Например, в системе прямоугольных координат длина откладываемых единичных отрезков на обеих осях одинаковая. Неравномерные сетки бывают логарифмическими, полулогарифмическими, вероятностными. Их применяют для более наглядного представления изучаемой зависимости, например, спрямление криволинейных зависимостей.

Полулогарифмическая координатная сетка имеет равномерную ординату и логарифмическую абсциссу; логарифмическая координатная сетка имеет обе оси логарифмические; вероятностная координатная сетка имеет обычно равномерную ординату и вероятностную шкалу по оси абсцисс. Назначение неравномерных сеток бывает различным. В основном их применяют для наглядного изображения функций. Важное значение при графическом изображении экспериментальных данных имеет вероятностная сетка, применяемая в разных случаях, например, при определении расчетных характеристик (расчетных значений модуля упругости бетона, расчетной влажности щебня) или при обработке измерений для оценки точности.

Также при обработке экспериментальных данных графическим способом необходимо составить расчетные графики, которые ускоряют нахождение по одной переменной других. При этом повышаются требования к точности изображения функции на графике. При вычерчивании расчетных графиков необходимо в зависимости от числа переменных выбрать координатную сетку и определить вид графика. Это может быть одна кривая, семейство кривых или серия семейств. Большое значение имеет при построении графиков, особенно расчетных, выбор масштаба, что связано с размерами чертежа и, соответственно, с точностью снимаемых с него значений величин. Чем больше масштаб, тем выше точность снимаемых значений. Графики, как правило, не должны превышать размеров 20 х15 строительной механики. Графики с минимумом или максимумом необходимо особенно тщательно вычерчивать в области экстремума. Поэтому здесь экспериментальные точки должны быть чаще. Часто для систематических расчетов вместо сложных теоретических или эмпирических формул используют номограммы, которые строят, применяя равномерные или неравномерные координатные сетки.

***3.1.4. Оформление результатов научного исследования***

Когда сформулированы выводы и обобщения, продуманы доказательства и подготовлены все иллюстрации, наступает следующий этап – литературное оформление полученных результатов в виде отчета, статьи, доклада или презентации.

Литературное оформление результатов творческого труда предполагает знание и соблюдение определенных требований, предъявляемых к содержанию научной рукописи. В научных работах особенно важны ясность изложения, систематичность и последовательность представления материала. Текст научной рукописи следует делить на абзацы, то есть на части, начинающиеся с красной строки. Важно помнить, что правильная разбивка на абзацы облегчает чтение и усвоение содержания текста. Критерием такого деления является смысл написанного – каждый абзац должен включать самостоятельную мысль, содержащуюся в одном или нескольких предложениях. Также в рукописи следует избегать повторений, не допускать перехода к новой мысли, пока первая не получила полного законченного выражения. Писать текст нужно по возможности краткими и ясными для понимания предложениями. Текст лучше воспринимается, если в нем исключены частое повторение одних и тех же слов, и выражений, тавтологии, сочетания в одной фразе нескольких свистящих и шипящих букв. Изложение должно включать критическую оценку существующих точек зрения, высказанных по данному вопросу, даже если они не в пользу автора. В тексте нежелательно делать много ссылок на себя. При необходимости следует употреблять выражения в третьем лице, например, автор полагает или, по нашему мнению, и т.д. Не рекомендуется перегружать рукопись цифрами, цитатами, иллюстрациями, так как это отвлекает внимание читателя и затрудняет понимание содержания. Цитируемые в рукописи места (например, высказывания) должны иметь точные ссылки на источники. Необходимым условием является соблюдение единства условных обозначений и допускаемых сокращений слов, которые должны соответствовать принятым стандартам.