

Лекция 1. Статистические методы обработки информации в нефтегазовом деле.

Составитель ст. преп. каф. БНГС СамГТУ, магистр Никитин В.И.

1. ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.1. СТАТИСТИЧЕСКИЙ ХАРАКТЕР ЭМПИРИЧЕСКИХ ДАННЫХ

На результаты наблюдений и измерений, проводимых в процессе бурения (или экспериментирования), влияют различные случайные воздействия, которые неконтролируемо вносятся в процесс измерений, а также несовершенство экспериментальной методики.

Для технологического процесса бурения скважины характерны следующие основные особенности:

- 1) большое число случайных факторов, изменяющихся во времени и пространстве и влияющих на качество и технико-экономические показатели работ;
- 2) разнообразие геолого-технических условий бурения, невозможность целенаправленно управлять большинством параметров в связи с их случайным характером изменения;
- 3) искажение полезного сигнала (о нагрузке на крюке, крутящем моменте, затратах мощности, механической скорости бурения и др.), используемого для определения параметров режима бурения.

В связи с этим показатели бурения (скорость бурения, проходка на коронку или долото, удельный расход алмазов, выход керна) следует рассматривать как случайные отклонения (переменные величины), а процессы—как случайные функции. Различные горные породы по-разному влияют на износостойкость породоразрушающих инструментов, что приводит к возникновению вероятностного взгляда на проблему их выбора. При взаимодействии породоразрушающего инструмента с различными по физико-механическим свойствам породами изменяется направление скважины.

Неопределенность материалов (глины, химические реагенты, цементы и добавки к ним и др.), широко применяемых в разведочном бурении, большое

число воздействующих факторов (температура, давление, долевой состав смесей, минерализация и др.) и сложность физико-химических процессов, протекающих в дисперсных системах в период структурообразования и твердения, обуславливают статистический характер выходных параметров и промывочных жидкостей, и тампонажных смесей.

Результаты всех измерений, как бы тщательно и на каком бы научном уровне они не выполнялись, искажаются ошибками двоякого рода. рода, систематическими и случайными. Это отражается на многочисленных измерениях одного и того же параметра. Случайные ошибки возникают по многим причинам; они искажают результаты последовательно повторенных измерений. Анализируя разброс результатов многократных измерений статистическими методами, можно получить достоверную оценку случайной ошибки.

Если рассеяние результатов, возникающее в процессе самого измерения может трактоваться как мера погрешностей, допускаемых в измерениях, то неопределенность значений, связанная с природой исследуемого процесса позволяет судить лишь о статистических закономерностях этого явления* Результаты, полученные экспериментально, также характеризуются определенной величиной вероятности, т. е. являются случайными. В любом эксперименте средние значения наблюдаемых величин меняются в связи с изменением основных факторов (качественных и количественных) определяющих условия опыта, а также случайных факторов.

Даже при проведении специально запланированных экспериментов полученные показатели подвергаются посторонним влияниям, причем некоторые из них невозможно предвидеть. По этой причине очень редко можно дважды получать тот же самый результат при повторении эксперимента. Так, прочность образцов одинаковых размеров, изготовленных из кернов, выбуренных из одного и того же блока породы, различна. Наличие в породе одного названия различных видов строения и минералогического состава (анизотропии) приводит к тому, что на каком-то одном образце изучаемый

параметр не будет представителен для всей данной породы и выводится как среднестатистическая величина.

Таким образом, анализ эмпирических данных должен базироваться на вероятностных представлениях. Для эффективного управления процессом бурения необходим количественный анализ, т. е. оценка действующих факторов и параметров на основе обработки поступающей информации. Обработка опытных данных, полученных в процессе бурения или при проведении специальных экспериментов становится возможной только с помощью методов математической статистики.

1.2. Наблюдение и эксперимент. Генеральная совокупность и выборка.

Наблюдение - метод исследования предметов и явлений реальности в том виде, в каком они существуют и происходят в природе и обществе. Наблюдение отличается от простого восприятия информации наличием цели и активной позицией наблюдателя. Наблюдение отличается и от *эксперимента* отсутствием активного управляющего воздействия на явление или процесс. Наблюдение - это фиксация характерных признаков предмета или развития явления в пространстве и/или во времени.

Источником опытной основы инженерной и научной деятельности служит *наблюдение (эксперимент)*, которое в широком смысле этого слова можно разделить на два основных типа. Первый тип - это сведения о поведении какой-либо совокупности большого числа однородных объектов, а второй - информация о свойстве отдельного объекта в течение продолжительного промежутка времени. В статистике для описания любого множества объектов используется понятие *совокупности*. При этом *генеральной совокупностью* называют множество объектов, из которых проводится отбор в процессе конкретизации наблюдений. Отобранные для наблюдений объекты представляют собой *выборочную совокупность*, или *выборку*, а число этих объектов называют ее *объемом*.

Основой любых выводов о вероятностных свойствах генеральной совокупности X , т.е. статистических выводов, является так называемый *выборочный метод*. Суть выборочного метода заключается в том, что свойства случайной величины X устанавливаются на основании изучения случайной выборки.

Выборка должна быть *репрезентативной* (представительной), то есть такой, по которой можно с уверенностью определять интересующие нас характеристики генеральной совокупности. Другими словами, репрезентативность выборки заключается в том, что она достоверно и в полной мере отображает все характеристики генеральной совокупности, частью которой она является.

При использовании в статистических исследованиях выборочного метода возникают так называемые ошибки репрезентативности. Ошибки репрезентативности обусловлены тем, что выборочная совокупность не полностью воспроизводит генеральную и представляют собой расхождение между значениями параметров, полученными по выборке, и значениями соответствующих параметров генеральной совокупности.

В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществлять случайно, объем выборки должен быть достаточно большим.

Существует несколько способов отбора случайных величин для получения выборочной совокупности:

1. Отбор, не требующий расчленения генеральной совокупности на части:

- ☐ простой случайный бесповторный отбор;
- ☐ простой случайный повторный отбор.

2. Отбор, при котором производится расчленение генеральной совокупности на части:

- ☐ типический отбор;
- ☐ механический отбор;
- ☐ серийный отбор.

Выборка называется **повторной**, если объект возвращается в генеральную совокупность перед производением следующего отбора.

Бесповторной называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

При большом объеме генеральной совокупности различия между повторной и бесповторной выборкой незначительны.

Типический отбор – это отбор, при котором неоднородная генеральная совокупность разбивается на типические (однородные) группы, из которых затем и производится случайный отбор. Так, например, если продукция изготавливается на нескольких станках, то для проведения контроля качества естественно проводить выборку продукции, произведенной на каждом из станков.

Механический отбор производится через определенный интервал, например, выбирают каждую 100 деталь.

При **серийном** отборе – отбирают не отдельную единицу (например, произведенную деталь или прибор), а группу, серию, например, продукцию, изготовленную одним станком.

На практике часто применяют комбинированный отбор.

1.3. Погрешности измерений

Погрешности измерений (ошибки измерений) - отклонения результатов измерений от истинных значений измеряемых величин. Различают систематические, случайные и грубые погрешности измерений (последний вид погрешностей измерений иногда называется промахами). *Систематические* погрешности измерений обусловлены главным образом погрешностями средств измерений и несовершенством методов измерений; *случайные* - рядом неконтролируемых причин (незначительными изменениями условий измерений и др.); *промахи* - неисправностью средств измерений, неправильным

отсчитыванием показаний, резкими изменениями условий измерений и др. *Грубые* ошибки заметны, и при обработке результатов измерений их просто отбрасывают, влияние систематических ошибок стремятся уменьшить различными поправочными множителями, оценки случайных погрешностей измерений осуществляют методами математической статистики.

1.4. Случайные события и случайные величины

Событие - то или иное уникальное явление, случившийся факт. Событие становится *достоверным*, если в данном конкретном сочетании факторов оно *необходимо* осуществляется. Событие рассматривается *случайным*, если оно может осуществиться, а может и не осуществиться. Изолированных событий в природе нет, все события осуществляются или не осуществляются в той или иной определённой системе событий. Событие может быть причиной, может быть следствием. Событие как явление окружающего нас мира может пройти незамеченным, а может быть целью. В последнем случае можно говорить о событии как результате *наблюдения*.

Случайное событие - событие, наступление или не наступление которого в некотором испытании (эксперименте) зависит от множества случайных факторов и для которого постулируется определённая вероятность его наступления при данных условиях. Случайное событие - одно из основных понятий *теории вероятностей*.

Случайная величина – величина, значение которой невозможно предсказать исходя из условий *эксперимента* или *наблюдений*. Результаты измерений физических величин по существу случайные величины. Случайные величины могут изменяться *непрерывно* (температура, давление, концентрация, радиус частиц дисперсной фазы) или *дискретно* (число «очков» в азартной игре, число частиц, число дефектов, число отказов, аварийность). Например, если представляет интерес количество монет всё ещё находящихся в обращении, как функция их возраста, то можно реализовать два подхода к решению этой проблемы. Если в качестве случайной величины использовать год чеканки монеты, т.е. $X = \dots, 1999, 2000, 2001, 2002, 2003, \dots$, то анализировать придётся

дискретные случайные величины, а если массу монеты, - то непрерывные случайные величины. Между *непрерывными и дискретными* случайными величинами есть существенное отличие – равенство первых невозможно, вторых – возможно. Невозможность равенства непрерывных случайных величин обусловлена ограниченной точностью измерения и/или ограниченной точностью математической обработки данных. Таким образом, результаты наблюдений и экспериментов, по своей природе являются случайными величинами, при работе с которыми следует пользоваться методами математической статистики.

Основные задачи математической статистики

Прежде всего, следует отметить, что при решении любой задачи математической статистики, в распоряжении имеются два источника информации: результаты статистического эксперимента (то есть выборка из генеральной совокупности случайной величины X) и некоторая априорная информация об интересующих свойствах генеральной совокупности, известной к текущему моменту. Априорная информация учитывается в выбираемой статистической модели.

Перечислим некоторые наиболее часто встречающиеся задачи математической статистики:

1. Оценка неизвестных параметров.
2. Проверка статистических гипотез.
3. Установление формы и степени связи между случайными величинами.