Реферат

Корреляционные зависимости в физическом эксперименте

Введение

математический статистика корреляционный

Современные научные исследования и производственная практика требуют широкого применения математической статистики для анализа закономерностей массовых явлений во всех отраслях промышленности.

Предметом настоящей реферативной работы является относительно простой и, тем не менее, достаточно эффективный метод, известный как корреляционный анализ. В настоящее время наряду с другими элементами статистического анализа физических процессов он успешно используется для решения задач исследования закономерностей процессов в широких интервалах изменения параметров, поиска оптимальных технологических режимов и конструктивных элементов оборудования, а также различных задач оптимального автоматического управления и регулирования.

В связи со стремительным развитием электронно-вычислительной техники и соответствующего программного обеспечения инструментарий анализа корреляций становится ещё более простым в использовании и доступным для исследователя.

В данной реферативной работе

· приводятся основные теоретические выкладки корреляционного анализа

· рассматривается применение его инструментария в контексте металлургической промышленности с использованием программного средства Statistica 6.

1. Постановка задачи

При анализе физических процессов часто приходится решать задачи о степени связи, а также выражать в математической форме зависимость между двумя или более переменными.

.1 Эмпирические данные

Величины, между которыми устанавливается связь, (количественные характеристики изучаемого явления) являются результатами наблюдений (регистрации) и называются эмпирическими данными.

Эмпирические данные содержат ошибки и случайные колебания, обусловленные множеством неучтённых факторов, которые чаще всего входят аддитивно (добавляются к истинным значениям или вычитаются из них). Так или иначе данные можно рассматривать как сумму регулярной (детерминированной) и случайной составляющих, которые явно не выделены.

. Регулярная составляющая эмпирических данных является его закономерной частью, которая отражает сущность изучаемого явления (его истинную величину).

Регулярная составляющая однозначно определяется учитываемыми причинно-следственными связями с другими величинами, и остаётся неизменной при независимых повторных измерениях эмпирического значения.

В связи с этим методика наблюдений часто предусматривает независимые многократные измерения. При этом в используемом среднем значении уменьшается доля случайного, возрастает доля и надёжность регулярной части.

. Случайная составляющая эмпирических данных складывается из случайных отклонений от регулярной составляющей. Случайные отклонения порождаются множеством неучтённых связей и погрешностями измерений эмпирического значения. Отклонение от истинного значения происходит с определённой вероятностью, то есть данная составляющая является статистически устойчивой и, соответственно, подчиняется некоторому закону распределения. Наиболее часто значения случайной составляющей подчиняются нормальному закону с нулевым математическим ожиданием.

. Эмпирические данные в целом - случайные величины. Закон их статистического распределения в целом определяется случайной составляющей - чаще всего нормальный, но с математическим ожиданием, равным среднему значению регулярной составляющей, и дисперсией, складывающейся из дисперсий регулярной и случайной составляющих.

.2 Стохастическая эмпирическая зависимость

Зависимость между случайными величинами называется стохастической зависимостью. Она проявляется в изменении закона распределения одной из них (зависимой переменной) при изменении других (аргументов).

Графически стохастическая эмпирическая зависимость, в системе координат *зависимая переменная - аргументы*, представляет собой множество случайно расположенных точек, которое отражает общую тенденцию поведения зависимой переменной при изменении аргументов.

Стохастическая эмпирическая зависимость от одного аргумента называется парной зависимостью, если аргументов более одного - многомерной зависимостью. Пример парной линейной зависимости приведён на рис. 1.([3])



Рис. 1. (1- случайные значения зависимой переменной, 2 - тенденция поведения зависимой переменной при изменении аргумента)

В отличие от обычной функциональной зависимости, в которой изменениям значения аргумента (или нескольких аргументов) отвечает изменение детерминированной зависимой переменной, в стохастической зависимости при этом происходит изменение статистического распределения случайной зависимой переменной, в частности, математического ожидания.

.3 Задача математического моделирования (аппроксимации)

Построение стохастической зависимости иначе называется математическим моделированием (аппроксимацией) или приближением и состоит в нахождении её математического выражения (формулы).

Эмпирически установленная формула (функция), которая отражает не всегда известную, но объективно существующую истинную зависимость и отвечает основному, устойчивому, повторяющемуся отношению между предметами, явлениями или их свойствами, рассматривается как математическая модель.

Устойчивое отношение вещей и их истинная зависимость. моделируется она или нет, существует объективно, имеет математическое выражение, и рассматривается как закон или его следствие.

Если подходящие закон или следствие из него известны, то их естественно рассматривать в качестве искомой аналитической зависимости. Например, эмпирическая зависимость силы тока *I* в цепи от напряжения *U* и сопротивления нагрузки *R* следует из закона Ома:

(1.1)

К сожалению, истинная зависимость переменных в подавляющем большинстве случаев априорно неизвестна, поэтому возникает необходимость её обнаружения, исходя из общих соображений и теоретических представлений, то есть построения математической модели рассматриваемой закономерности. При этом учитывается, что заданные переменные и их приращения на фоне случайных колебаний отражают математические свойства искомой истинной зависимости(поведение касательных, экстремумы, корни, асимптоты и т.п.)

Подбираемая, так или иначе, аппроксимирующая функция сглаживает (усредняет) случайные колебания исходных эмпирических значений зависимой переменной и, подавляя тем самым случайную составляющую, является приближением к регулярной составляющей и, стало быть, к искомой истинной зависимости.

Математическая модель эмпирической зависимости имеет теоретическое и практическое значение:

· позволяет установить адекватность экспериментальных данных тому или иному известному закону и выявить новые закономерности;

· решает для зависимой переменной задачи интерполяции внутри заданного интервала значений аргумента и прогнозирования (экстраполяции) за пределами интервала.

Однако, несмотря на большой теоретический интерес нахождения математической формулы для зависимости величин, на практике часто достаточно лишь определить, есть ли между ними связь и какова её сила.

.4 Задача корреляционного анализа

Методом изучения взаимосвязи между изменяющимися величинами является корреляционный анализ.

Ключевым понятием корреляционного анализа, описывающим связь между переменными является корреляция (от английского *correlation - согласование, связь, взаимосвязь, соотношение, взаимозависимость*).

Корреляционный анализ используется для обнаружения стохастической зависимости и оценки её силы (значимости) по величине коэффициентов корреляции и корреляционного отношения.

Если связь между переменными обнаружена, то говорят, что корреляция присутствует или что переменные коррелированны.

Показатели тесноты связи (коэффициент корреляции, корреляционное отношение) по модулю изменяются от 0(при отсутствии связи) до 1(при вырождении стохастической зависимости в функциональную).

Стохастическая связь полагается значимой (реальной), если абсолютная оценка коэффициента корреляции (корреляционного отношения) значима, то есть в 2-3 превышает стандартное отклонение оценки коэффициента.

Отметим, что в некоторых случаях связь может быть обнаружена между явлениями, не находящимися в очевидных причинно-следственных отношениях.

Например, для некоторых сельских районов выявлена прямая стохастическая связь между числом гнездящихся аистов и рождающихся детей. Весенний подсчёт аистов позволяет предсказывать, сколько в этом году родится детей, но зависимость, конечно, не доказывает известное поверье, и объясняется параллельными процессами:

· рождению детей обычно предшествует образование и обустройство новых семей с обзаведением сельскими домами и подворьями;

· расширение возможностей гнездования привлекает птиц и увеличивает их количество.

Подобная корреляция между признаками называется ложной(мнимой) корреляцией, хотя она может иметь прикладное значение.

2. Общее понятие об оценке реальности связи и её тесноты

Рассмотрим общий анализ связи на примере линейной зависимости двух переменных *x* и *y*.

.1 Случайное рассеяние и неопределённость связи

Причиной случайного рассеяния эмпирических данных является влияние множества неучитываемых факторов и ошибок измерений.

Случайное рассеяние при линейной зависимости проявляется в том, что каждое допустимое значение аргумента *x* обуславливает не определённую величину зависимой переменной *y(x)* а множество её случайных значений (точек в системе координат *x0y*).

Подмножество случайных значений *y(x)* для каждого *x* образует статистическое распределение, а для последовательности *x* - семейство распределений.

Неопределённость стохастической связи в математической статистике понимается как показатель рассеяния (разброса) случайных величин, отсутствия у них общей тенденции.

Графически, в системе декартовых координат, рассеяние случайных величин отображается множеством точек с общим центром. Чем хаотичнее разброс множества точек, чем менее оно подчинено общей тенденции, тем неопределеннее связь и, соответственно слабее корреляция. По смыслу неопределённость противоположна понятиям реальности связи и её силы, как поясняется на рис 2.([3])



Рис. 2А отвечает рассеянию переменных x и y относительно центра  при отсутствии общей тенденции группирования точек. Здесь нельзя указать линию, проходящую через центр  , которая отвечает тенденции упорядочения точек, поэтому неопределённость рассеяния максимальна, корреляция отсутствует, а также задача линейной аппроксимации не имеет решения.

Рис. 2В отражает противоположный случай, когда нет рассеяния точек - все они подчиняются общей тенденции (принадлежат одной и той же прямой), то есть стохастическая связь вырождается в функциональную, и неопределённость отсутствует.

Рис. 2Б иллюстрирует общий случай линейной стохастической связи, когда рассеяние точек есть, но оно имеет общую тенденцию, и точки группируются в области, вытянутой в некотором направлении, вдоль прямой, проходящей через центр  и отвечающей линейной стохастической зависимости.

Одной из оценок характера связи является коэффициент неопределённости - это доля рассеяния зависимой переменной *y* относительно модели в общем рассеянии зависимой переменной *у*.

Иначе, коэффициент неопределённости - это отношение сумм квадратов:

 (2.1)

Величину и смысл коэффициента неопределённости можно понять из показанных на рис 2. случаев рассеяния.

При отсутствии связи (рис. 2А) отсутствует общая тенденция группирования точек. Они оказываются одинаково рассеянными относительно любой линии, проходящей через центр, в том числе линии средних значений , поэтому коэффициент неопределённости достигает максимально возможного значения - 1, переменные не коррелированны.

Если точки группируются в области, вытянутой в некотором направлении, вдоль прямой, проходящей через центр  и отвечающей линейной стохастической зависимости, то рассеяние *y* относительно неё меньше, чем относительно среднего значения  (рис. 2Б), и коэффициент неопределённости меньше 1, переменные коррелированны.

При полном отсутствии неопределённости (рис. 2В) стохастическая связь вырождается в функциональную зависимость, поэтому все точки принадлежат модели , то есть относительно неё рассеяния *y* нет, и коэффициент неопределённости равен 0.

.2 Корреляционное отношение

В качестве показателя тесноты стохастической связи при решённой, либо решаемой задаче аппроксимации, используется величина, противоположная коэффициенту неопределённости:

 (2.2)

Такая величина называется корреляционным отношением. Она является приближенной оценкой тесноты связи, поскольку, как и коэффициент неопределённости, не учитывает числа степеней свободы у используемых сумм квадратов разностей. Первая из них (в числителе) имеет *n - 2* степеней свободы, так как линейная зависимость накладывает две связи, отвечающие двум параметрам *a* и *b*. Вторая сумма имеет *n - 1* степень свободы, поскольку накладывается одна связь, определяемая средним. В итоге данная оценка оказывается смещённой (несколько завышенной), чем обычно пренебрегают, особенно при большом объёме выборки. Отметим, что программные средства обычно выводят не *R*, а *R2* и её несмещённую величину (*Adjusted R2*).

Корреляционное отношение *R* равно 0 при отсутствии связи (рис. 2А), когда коэффициент неопределённости равен 1. При функциональной связи корреляционное отношение максимально и достигает 1. В общем случае корреляционное отношение удовлетворяет неравенству 0 < *R* < 1.

Отметим возможность применения данной величины для многомерной и нелинейной зависимости, например, в случае



выражение для корреляционного отношения примет вид:

 (2.3)

К недостаткам оценки силы связи с помощью корелляционного отношения *R* следует отнести необходимость предварительного построения модели для определения постоянных, входящих в формулу его вычисления.

В современных пакетах программ, ориентированных на статистический анализ данных, в том числе Statistica 6, уже встроена линейная, параболическая, логарифмическая и другие виды аппроксимации, что позволяет активно использовать корреляционное отношение в качестве оценки силы связи как при наличии одного, так и нескольких аргументов.

.3 Коэффициент детерминации

Реальность и теснота стохастической связи характеризуется показателем определённости, или коэффициентом детерминации, определяемым как отношение дисперсии зависимой переменной *y*, объяснённой моделью, к общей дисперсии этой переменной.

Иными словами, коэффициент детерминации есть доля дисперсии, объяснённой моделью, в общей дисперсии зависимой переменной и выражается в %.

Формально коэффициент детерминации равен квадрату несмещённого значения корреляционного отношения, то есть в случае рассмотренной выше линейной зависимости  коэффициент детерминации может быть представлен как

(2.4)

3. Парная линейная корреляция

Центральное место в корреляционном анализе занимает парная линейная корреляция. Как было отмечено выше, если имеется пара переменных, то корреляция между ними - это мера связи (зависимости) именно между этими переменными.

На первый взгляд большинство нелинейных парных связей, то есть связей, не удовлетворяющих формуле  можно, трансформируя переменные, заменить линейными зависимостями. В этом случае они стали бы доступными для простого в использовании инструментария, применяемого только для исследования линейных корреляций.

Скажем, нелинейную зависимость типа



можно преобразовать при помощи логарифмирования так:





если  , то получим линейную зависимость следующего вида. Аналогично нелинейную функцию можно выразить с помощью логарифмов так:



Если , то получим зависимость  , которая также линейна.

Однако следует учитывать, что необходимым условием истинности линейной связи (и её оптимальности) является адекватность математическим свойствам эмпирической зависимости. Практически это означает, что область определения и нулевые значения линейной модели должны соответствовать искомой истинной зависимости и их проявлениям в эмпирических данных, которые могут иметь своей асимптотой только саму аппроксимирующую прямую.

Нетрудно заметить, что при переходе к логарифмам данное условие не выполняется, а значит, исследование нелинейных и многомерных корреляций требует своих, обычно более сложных методов.

Анализ же линейной корреляции между двумя переменными опирается на следующие инструменты математической статистики.

.1 Ковариация

Ковариация является вторым смешанным центральным моментом случайных величин *x* и *y*, который характеризует их связь.

Ковариация или коэффициент ковариации определяется как

(3.1)

где *M* - оператор математического ожидания.

Так как математическое ожидание  и, аналогично,  , то правую часть ковариации можно упростить:

 (3.2)

Смешанные произведения в обеих формулах могут иметь разные знаки соответственно монотонному возрастанию или убыванию зависимости:

· знак плюс, когда знаки сомножителей одинаковы (зависимость между переменными является монотонно возрастающей)

· знак минус при разных знаках (зависимость между переменными является монотонно убывающей).

Если переменные *x* и *y* независимы, то справедлива следующая теорема.

Теорема. Математическое ожидание произведения независимых случайных величин *x* и *y* равно произведению их математических ожиданий, то есть



Примем без доказательства.

На основании этой теоремы ковариация двух независимых величин *x* и *y* равна нулю.

Очевидно, что если , то случайные величины *x* и *y* зависимы:

· при  зависимость умеет вид монотонного возрастания;

· при  зависимость умеет вид монотонного убывания.

Ковариацию иначе называют корреляционным моментом или моментом связи, она является признаком существования зависимости между случайными величинами и её вида. Однако на практике использование ковариации неудобно, поскольку она зависит от единиц измерения случайных величин *x* и *y*.

Для того, чтобы иметь дело с безразмерным показателем отклонения случайных величин от своих средних, ковариация нормируется на стандартные (среднеквадратичные отклонения). Вместо ковариации в виде  берётся математическое ожидание нормированных величин  и , где (3.3)

.2 Коэффициент парной линейной корреляции (Пирсона)

Коэффициент корреляции  двух переменных, измеренных в интервальной шкале называется коэффициентом корреляции Пирсона, а также линейной корреляцией, так как отражает степень линейной связи между переменными. Этот коэффициент представляет собой ковариацию нормированных величин   :

 (3.4)

Поскольку величина  имеет дискретное равномерное распределение, то её математическое ожидание равно среднему арифметическому всех принимаемых значений. Учитывая это, коэффициент корреляции может быть представлен как:

(3.5)

. Коэффициент корреляции независимых случайных величин *x* и *y* равен нулю, так как в этом случае .

Случайные величины, для которых ковариация, и, следовательно, коэффициент корреляции равны нулю, называют линейно некоррелированными (линейно не связанными).

Иными словами, если случайные величины независимы, то всегда , однако обратное вообще говоря неверно - из равенства коэффициента корреляции нулю не следует независимость случайных величин. Можно говорить лишь об отсутствии между ними линейной связи. В этом легко убедиться на примере:



Пусть переменные связаны функциональной зависимостью  , график которой приведён на рис. 3. ([3]) Мы видим, что  .

Вследствие симметрии каждому отклонению от оси абсцисс от среднего со знаком плюс отвечает такое же отклонение со знаком минус с одними и теми же отклонениями от среднего ординат, поэтому математическое ожидание смешанных произведений в формулах ковариации и коэффициента корреляции равно нулю. Следовательно,  , хотя переменные связаны функциональной зависимостью.

. Коэффициент корреляции линейно связанных случайных величин *x* и *y* отличается от нуля, но находится в некоторых границах.

Существование границ значений коэффициента корреляции следует из дисперсии суммы зависимых случайных величин

 (3.6)

Дисперсия всегда положительна, значит 

Из выражения для коэффициента корреляции следует, что . С учётом этого из равенства (3.6) после упрощений можно получить неравенство  или 

Как было показано для ковариации, её знак, и, стало быть, знакуказывает:

· знак плюс - на возрастание линейной стохастической зависимости

· знак минус - на убывание линейной стохастической зависимости.

Сами граничные значения  отвечают вырождению линейной стохастической зависимости в функциональную. Очевидно,  соответствует линейно возрастающей функциональной зависимости с угловым коэффициентом *a* > 0,  - линейно убывающей функциональной зависимости с угловым коэффициентом *a* < 0.

.3 Интерпретация линейной корреляции

Линейную стохастическую связь можно полагать реальной, если абсолютная величина коэффициента корреляции в 2-3 раза превышает стандартное отклонение оценки коэффициента. Невыполнение данного условия указывает на отсутствие линейной связи, однако не исключает линейной стохастической зависимости тех же переменных.

Значимая абсолютная величина коэффициента корреляции, в 2-3 раза превышающая стандартное отклонение оценки коэффициента, свидетельствует о значимом проявлении линейной составляющей стохастической связи, однако не исключает более тесной нелинейной стохастической зависимости.

Стохастическая связь (линейная или нелинейная), если она реальна, сама по себе не указывает на причинно-следственную связь переменных, даже при надёжном предсказании одной переменной по значениям другой. Здесь требуются дополнительные основания для выяснения, какой из признаков является причиной другого.

Вследствие этого, корреляционные зависимости приято подразделять на действительные (истинные) и мнимые (ложные).

Действительные корреляционные связи отражают причинные отношения между зависимой и независимой переменными, при этом различаются причинности:

· непосредственные, например, в зависимости *Z* - числа междугородних телефонных переговоров от *X* - количества АТС, то есть  ;

· через промежуточные переменные (одну или несколько) - в примере выше это может быть число телефонов *Y* в АТС .

Ложные корреляционные зависимости могут возникать между переменными, которые не находятся между собой в причинно-следственной связи.

.4 Расчёт коэффициента корреляции

Коэффициент корреляции  часто рассчитывается по приведённой выше формуле:

(3.7)

Данное выражение можно преобразовать алгебраически, в результате чего оно становится более удобным для числового расчета:

(3.8)

При небольшом количестве наблюдений ( ) расчёт производится непосредственно по этой формуле.

В случае большого числа наблюдений, вводим новые переменные:



После внесения результатов в корреляционную таблицу и введения новых переменных довольно легко рассчитать величину коэффициента корреляции по формуле:

 (3.9)

где *r* - общее количество интервалов группирования переменной *x*;

*s* - общее количество интервалов группирования переменной *y*;

*nik* - количество наблюдений, значение переменной *x* которых лежит в *i*-том интервале группирования переменной *x*, а величина переменной y лежит в *k*-том интервале группирования переменной *y*.

3.5 Пример расчёта коэффициента корреляции при малом числе наблюдений

Расчёт коэффициента корреляции при малом количестве наблюдений покажем на примере исследования зависимости между содержанием серы (S) в стали одной марки и прочностью на растяжение ( ). Всего было выбрано 13 плавок, из стали которых по одинаковой технологии изготовили стержни одного номинального диаметра. Расчёт необходимых величин



приведен в таблице 1. Прежде всего, рассчитывают среднеарифметическое и основное отклонение для обеих переменных:

Имея все необходимые величины, определим коэффициент корреляции:

Таблица 1. Расчёт коэффициента корреляции при малом числе наблюдений

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S(%) |  [кГ/мм2]  |  |  |  |
| 0,025 0,030 0,032 0,040 0,046 0,048 0,050 0,054 0,056 0,060 0,070 0,072 0,072  | 29,0 29,5 29,0 31,0 32,0 31,5 32,3 33,0 32,4 34,5 33,0 33,8 35,5  | 0,000625 0,000900 0,001024 0,001600 0,002116 0,002304 0,002500 0,002916 0,003136 0,003600 0,004900 0,005184 0,005184  | 841,00 870,25 841,00 961,00 1024,00 992,25 1043,29 1089,00 1049,76 1190,25 1089,00 1142,44 1260,25  | 0,7250 0,8850 0,9280 1,2400 1,4720 1,5120 1,6150 1,7820 1,8144 2,0700 2,3100 2,4336 2,5560  |
|  |  |  |  |  |

Мы видим, что величина коэффициента корреляции положительная и близкая к единице; это значит, что в данном случае существует тесная взаимосвязь между содержанием в стали серы и прочностью на растяжение.

Однако следует помнить, что величина коэффициента корреляции была установлена на основании определённой выборки, в данном случае относительно небольшого объёма. Следовательно, коэффициент корреляции можно считать значимым только предположив, что распределение эмпирических данных является нормальным.

При проведении более требовательного статистического анализа на основе малой выборки, то есть когда нельзя проверить, подчиняется ли распределение нормальному закону, переходят к альтернативному показателю тесноты связи, как это будет показано далее.

4. Статистический анализ данных вискозиметрического эксперимента[5,6]

Для определения оптимального режима выплавки металлических сплавов актуальна информация о структурном состоянии жидкого металла. Одним из наиболее распространенных косвенных способов исследования структурного состояния металлических расплавов является измерение их свойств и, в частности, вязкости. Исследователи часто отмечают расхождения политерм, полученных в ходе нагрева и последующего охлаждения образца, повышенный статистический разброс значений вязкости, наличие максимумов, минимумов, точек перегиба на кривых и т.д. Как правило, с температурами, соответствующими этим точкам, они связывают изменения в структуре металлического расплава. В большинстве случаев особые точки и отвечающие им температуры определяются при визуальном анализе. Для получения более точной информации необходимо применение специальных методов анализа опытных данных с использованием компьютера, в частности, методов математической статистики.

В работе [5] приведены данные вискозиметрического исследования расплава алюминия марки А-999 в интервале температур от ликвидуса до 11000C в режиме нагрева и последующего охлаждения образца, и результаты их статистического анализа. Целью авторов было получение наиболее полной и объективной информации о характере температурной зависимости вязкости жидкого алюминия. Для анализа результатов эксперимента использовалось программное средство Statistica 6.0 [4], встроенные функции которого позволяют осуществлять статистический анализ экспериментальных данных.

Необходимо указать на сложность расплава алюминия как объекта экспериментального исследования. Трудности при проведении опытов обычно возникают из-за высокой окисляемости алюминия. Поэтому в зависимости от условий проведения измерений можно получить различные данные о вязкости этого металла. Противоречия в литературных данных о вязкости жидкого алюминия в основном связаны с тем, что не во всех работах указываются условия проведения эксперимента. В последние годы большинство исследователей перед измерениями рекомендуют проводить переплав металла в динамическом вакууме ниже 1 Па при температуре 950-1000ºС, в ходе которого оксид Al2O3 переходит в летучий субокисел AlO и удаляется из расплава. Этих рекомендаций придерживался при проведении опытов и автор работы [5].

Вязкость измерялась методом затухающих крутильных колебаний тигля с расплавом в режиме нагрева и последующего охлаждения образцов. Непосредственно перед измерением вязкости в установке образец нагревался до 9000C в вакууме с целью удаления оксидной пленки. Все опыты проводились при разрежении ниже 1 Па. Изотермические выдержки в точках отсчета составляли не менее 15 минут. Погрешность определения вязкости не превышала 3%. После охлаждения образца алюминия марки А 999 до комнатной температуры без разгерметизации установки в ходе последующего нагрева и охлаждения были получены политермы ν, приведённые на рис.4([5]).



Основная проблема при применении методов математической статистики [4] для анализа температурных, временных и концентрационных зависимостей кинематической вязкости - малый объем выборки. Как уже отмечалось ранее, дело в том, что когда число экспериментальных точек велико (100 или более опытов), можно считать при расчете случайной ошибки, что распределение экспериментальных данных является нормальным. При малом числе опытов нет способов проверить это предположение. Для анализа малых выборок применяют непараметрические методы. Эти методы и применяли авторы работы [5]. Выполненные процедуры попадают в одну из следующих категорий:

Задача 1. Оценка степени зависимости между переменными (выявление временной зависимости вязкости).

Задача 2. Определение критерия различия для зависимых выборок (сравнение значений вязкости при одинаковой температуре в режиме нагрева и охлаждения);

Задача 3. Определение критерия различия для независимых выборок (вязкость мало меняется с ростом температуры и необходимо сравнить как значимо различаются эти данные).



Авторы для оценки степени корреляции между значением кинематической вязкости и временем использовали непараметрическую альтернативу коэффициенту корреляции Пирсона - корреляцию Спирмена *Rxy*. Если опытные данные ранжировать соответственно номеру наблюдения в вариационном ряде, то есть каждому значению переменной присвоить ранг, то корреляцию Спирмена *Rxy* можно представить себе как вычисленную по рангам корреляцию Пирсона, т.е. в терминах доли изменения одной величины, связанной с изменением другой. Формально *ранговый коэффициент корреляции Спирмена* между переменными вычисляется следующим образом:

 (4.1)

где *Pi* - ранг наблюдения *xi*, где *Si* - ранг наблюдения *yi*.

Сравнив эту формулу с формулой корреляции Пирсона, нетрудно заметить, что корреляция Спирмена является прямым аналогом корреляции Пирсона.

Соответствующая опция модуля "Непараметрическая статистика» программного средства Statistica 6, используемого авторами в работе [6] позволяет вычислить три различные альтернативы коэффициенту корреляции Пирсона: корреляцию Спирмена *Rxy*, статистику (тау) Кендалла и статистику Гамма. Статистика (тау) Кендалла и статистика Гамма скорее оценивают вероятности, точнее, разность между вероятностью того, что наблюдаемые значения переменных имеют один и тот же порядок, и вероятностью того, что порядок различный.

Результаты анализа корреляции значений кинематической вязкости расплава алюминия со временем, полученные в работах [5,6] представлены на взятых их них рис. 5. и в таб. 2. Целью анализа было выявление временной зависимости кинематической вязкости при данной температуре. Обнаружено, что зависимость кинематической вязкости от номера эксперимента, а, следовательно, и от времени уменьшается с повышением температуры. Следовательно, релаксационные процессы в расплаве более выражены при низких температурах. Степень недоверия к полученным результатам (p-level) составляет 0,5%.

Таблица 2 Корреляционный анализ вязкости в зависимости от номера эксперимента при постоянной температуре

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Т | Valid N | Spearman R | t(N-2) | p-level |
| 690 | 11 | -0,059914 | -0,180064 | 0,861091 |
| 750 | 11 | 0,529686 | 1,873460 | 0,093780 |
| 804 | 11 | -0,221738 | -0,682196 | 0,512288 |
| 862 | 11 | 0,547951 | 1,965131 | 0,080974 |
| 917 | 11 | 0,165914 | 0,504739 | 0,625868 |
| 983 | 11 | -0,162495 | -0,494052 | 0,633103 |
| 956 | 11 | -0,087566 | -0,263711 | 0,797941 |
| 893 | 11 | -0,060054 | -0,180488 | 0,860768 |
| 827 | 11 | 0,018392 | 0,055185 | 0,957196 |
| 778 | 11 | -0,314869 | -0,995229 | 0,345628 |
| 723 | 11 | -0,593613 | -2,21291 | 0,054187 |

Заключение

В данном реферате был подробно рассмотрен такой метод математической статистики, как корреляционный анализ. Особый интерес для автора представляло его прикладное значение в решении задач вискозиметрического эксперимента.

Однако, следует отметить, что на пути к широкому внедрению математической статистики в промышленности имеется ряд трудностей, к числу которых относится уже упомянутая необходимость использования большого числа отдельных наблюдений. Часто при текущем контроле и анализе производства технологических процессов инженерно-технические работники вынуждены пользоваться малым числом наблюдений (объектов, опытов) в основном из-за высокой цены и сложности физического эксперимента в промышленной отрасли. Поэтому перспективы дальнейшего развития прикладного корреляционного анализа автор данной работы видит в более широком использовании элементов непараметрической статистики, которое вполне способны обеспечить современные пакеты программ .

Список литературы

1. А.Г. Дьячко, Математические модели металлургических процессов Курс лекций ч. I и II. М.: Мисис 1974. - 157с.

. М. Кнотек, Р. Войта, И. Шерц, Анализ металлургических процессов методами математической статистики. М.: Металлургия 1968. - 212 с.

3. Компьютерный анализ и интерпретация эмпирических зависимостей. Учебник. М.: ООО"Бином-Пресс". 2009. - 336с. (под редакцией С.В. Поршнева)

. В. Боровиков Statistica: искусство анализа данных на компьютере. Для профессионалов. - СПб.: Питер, 2001. - 656 с.

. О.А. Чикова, С.С. Горшков, Статистический анализ данных вискозиметрического эксперимента с металлическими расплавами // Тезисы 3 Российской научно-технической конференции "Физические свойства металлов и сплавов". УГТУ-УПИ, 16-18 ноября 2005, г. Екатеринбург.

. О.А. Чикова, Микрорасслоение расплавов на основе алюминия и его влияние на структуру литого металла. Автореферат дис. канд. Физ.-мат. наук. Свердловск, УПИ им. С.М.Кирова, 1991.

. Е.И. Куликов, Прикладной статистический анализ - 2-е изд., перераб. и доп. - М.: Горячая линия - Телеком, 2008. - 464 с. (Учебное пособие для высших учебных заведений).

. Я.Б. Зельдович, А.Д. Мышкис, Элементы прикладной математики. М.: Наука, 1965. - 616 с.9. Элементы корреляционного и дисперсионного анализа: Метод. указ. к решению задач мат. статистики для студентов заочного обучения всех специальностей; Сост. Р.А. Вайсбурд, А.Б. Абрамова; Под ред. В.Б. Винокуровой. - Свердловск: УПИ, 1989. - 36 с.

.В.В. Налимов, Применение математической статистики при анализе вещества. - М.: Физматлит, 1960. - 431 с.