

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**
**ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ) ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО
БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

У Т В Е Р Ж Д А Ю
Зам. директора по УР

_____ В.Л. Бибик

" ____ " _____ 2014 г.

**СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ
ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ**

Методические указания к выполнению лабораторной работы по курсу «Математическое моделирование процессов сварки, пайки и наплавки» для студентов, обучающихся по специальности 150202 «Оборудование и технология сварочного производства» и направлению 150700 «Машиностроение», профиль «Оборудование и технология сварочного производства» очной и очно-заочной форм обучения

Юрга 2014

УДК 681.332

Статистический анализ результатов экспериментальных исследований: Методические указания к выполнению лабораторной работы по курсу «Математическое моделирование процессов сварки, пайки и наплавки» для студентов, обучающихся по специальности 150202 «Оборудование и технология сварочного производства» и направлению 150700 «Машиностроение», профиль «Оборудование и технология сварочного производства» очной и очно-заочной форм обучения. – Юрга: Изд. ЮТИ ТПУ, 2014. – 24 с.

Составитель: к.т.н., доцент Д.А. Чинахов

Рецензент к.т.н., доцент А.В. Проскоков

Методические указания рассмотрены и рекомендованы к изданию на заседании кафедры сварочного производства « ____ » _____ 2014 г.

Зав. кафедрой
к.т.н., доцент _____ Е.А. Зернин

ВВЕДЕНИЕ

Математическая статистика – наука о математических методах систематизации и использовании статистических данных для научных и практических выводов. Статистические методы позволяют количественно определять характеристики явлений и процессов, служат эффективным инструментом планирования, анализа и прогнозирования показателей.

Между различными показателями могут быть две формы зависимости и связи: функциональная и корреляционная. *Функциональной* называется зависимость, которая точно проявляется в каждом отдельном случае и подчинена принципу строго определенного соответствия между количественными признаками. Подобные зависимости выражаются в виде уравнений, таблиц или графиков. Например, зависимости, выраженные в законах Ньютона, Ома и др.

Корреляционной называется такая зависимость между явлениями и показателями, которая проявляется только в среднем, в массе наблюдений. Теория и методы корреляционного анализа используются для изучения основных связей между явлениями и признаками. Корреляционные зависимости занимают промежуточное положение между функциональными зависимостями и полной независимостью между переменными.

В сварочном производстве закономерности не проявляются так же точно и неизменно, как, например, в физике или химии, поэтому корреляционный анализ широко используется при установлении взаимосвязи технологических показателей.

Статистические методы чрезвычайно разнообразны – от элементарных статистических приемов до сложнейших современных способов и моделей. Область применения этих способов, моделей и приемов различна: в некоторых случаях для познания явления и решения задачи достаточно применить относительно простые методы, в других – специальные, весьма сложные, требующие определенных знаний от исполнителей.

1 СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ЗАКОНЫ ИХ РАСПРЕДЕЛЕНИЯ

Случайная величина является количественной характеристикой результата опыта и может принимать различные числовые значения, заранее неизвестные и зависящие от случайных причин, которые не могут быть учтены.

Примерами случайных величин являются долговечность образцов при усталостных и длительных статических испытаниях, пределы текучести и прочности, относительное удлинение, твердость, ударная вязкость и другие характеристики механических свойств материалов.

Случайная величина характеризуется областью возможных значений, которые она может принимать в результате опыта, и вероятностью приобретения этих значений.

Существуют случайные величины двух типов: *дискретные* (прерывные) и

непрерывные. Дискретная случайная величина может принимать изолированные одно от другого значения, которые можно заранее перечислить. Непрерывная случайная величина может принимать любые значения из некоторого конечного или бесконечного промежутка. Например, n образцов испытывают при идентичных условиях на длительную статическую прочность. Испытание образцов прекращают, если он не разрушается за базовое время. В этом опыте число неразрушенных образцов m является дискретной случайной величиной, которая может принимать все целые значения от нуля до n . Время до разрушения образцов является непрерывной случайной величиной и может принимать целые и дробные положительные значения в бесконечном или конечном интервале.

Все характеристики механических свойств материалов и деталей являются непрерывными случайными величинами, поэтому в книге им уделено основное внимание.

Наиболее полно случайные величины (обычно обозначают прописными буквами латинского алфавита, а их возможные значения – соответствующими строчными) могут быть охарактеризованы с помощью функции распределения $F(x)$, представляющей собой вероятность появления значения $X < x$:

$$P(X < x) = F(x) \quad (1)$$

Функция распределения $F(x)$ является неубывающей функцией X (рис. 1), т.е. для любых двух чисел x_1 и x_2 при $x_1 < x_2$ удовлетворяется условие $F(x_1) \leq F(x_2)$.

Вероятность обнаружения случайной величины X в интервале $x_1 < X \leq x_2$ равна

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1). \quad (2)$$

Функция распределения удовлетворяет условиям

$$F(-\infty) = 0 \text{ и } F(\infty) = 1. \quad (3)$$

Для непрерывных случайных величин функция распределения имеет производную. Первая производная функции распределения называется *плотностью вероятности*.

$$\varphi(x) = \frac{dF(x)}{dx} \quad (4)$$

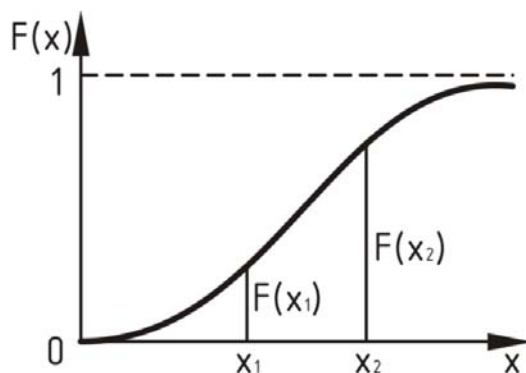


Рис. 1. График функции распределения случайной величины

Плотность вероятности удовлетворяет условию $\varphi(x) \geq 0$.

Вероятность попадания случайной величины в интервал $x_1 < X \leq x_2$ может быть найдена через плотность вероятности

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} \varphi(x) dx \quad (5)$$

Функция распределения $F(x)$ данной случайной величины связана с ее плотностью вероятности $\varphi(x)$ соотношением

$$F(x) = \int_{-\infty}^x \varphi(x) dx \quad (6)$$

2 ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

В практических задачах вместо задания функций распределения случайной величины бывает достаточно указать некоторые их *числовые характеристики*, называемые *статистиками*.

Характеристики центра распределения

В качестве числовых характеристик положения центра группирования случайных величин используют *математическое ожидание* или *среднее значение*, *моду* и *медиану* случайной величины.

Математическое ожидание или среднее значение случайной величины X обозначают через MX или a и определяют по формуле

$$a = MX = \int_{-\infty}^{\infty} x \varphi(x) dx \quad (7)$$

Модой случайной величины X является такое значение MoX , в котором плотность вероятности имеет максимальное значение.

Медианой случайной величины X служит значение MeX , которое соответствует условию

$$P(X < MeX) = P(X > MeX) = 0,5 \quad (8)$$

Геометрически медиана представляет абсциссу точек прямой, которая делит площадь, ограниченную кривой плотности вероятности, пополам.

Характеристики рассеивания случайных величин

Одной из основных характеристик рассеивания случайной величины X около центра распределения служит *дисперсия*, которая обозначается через DX или σ^2 и определяется по формуле

$$DX = \sigma^2 = \int_{-\infty}^{\infty} (x - a)^2 \varphi(x) dx \quad (9)$$

Часто вместо дисперсии за меру рассеивания случайной величины используют положительное значение квадратного корня из дисперсии, которое называется *средним квадратическим отклонением* или *стандартным отклонением*

$$\sigma = \sqrt{\sigma^2} \quad (10)$$

В практике широко применяют также характеристику рассеивания, называемую *коэффициентом вариации* γ , который представляет отношение среднего квадратического отклонения к математическому ожиданию,

$$\gamma = \frac{\delta}{a} \cdot 100\% \quad (11)$$

Коэффициент вариации показывает, насколько велико рассеивание по сравнению со средним значением случайной величины.

3 ПОНЯТИЕ О ВЫБОРКЕ

Механические свойства материала обычно изучают путем испытаний ограниченного числа образцов. В связи с неоднородностью конструкционных материалов найденные таким образом числовые характеристики механических свойств в большей или меньшей степени отличаются от так называемых *генеральных характеристик*, которые могут быть определены по результатам испытаний бесконечно большого числа образцов. Эта неограниченно большая воображаемая совокупность образцов (результатов испытаний), которые могут

быть выделены из исследуемого материала или полуфабриката, называется *генеральной совокупностью*. Ограниченная совокупность образцов (результатов испытаний), являющаяся частью *генеральной совокупности*, называется *выборкой*.

Отмеченная разница в выборочных и генеральных характеристиках механических свойств зависит от объема выборки. С увеличением числа испытаний в связи с проявлением закона больших чисел выборочные характеристики сходятся по вероятности к генеральным характеристикам, т.е. вероятность события, заключающегося в том, что разница между указанными характеристиками не будет превышать сколь угодно малую величину, при увеличении объема выборки неограниченно приближается к единице.

На практике на основании выборочных характеристик судят об уровне генеральных характеристик. Широко распространены следующие обозначения выборочных характеристик:

\bar{X} – выборочное среднее значение случайных величин X ;

s^2 – выборочная дисперсия;

s – выборочное среднее квадратическое отклонение;

v – выборочный коэффициент вариации.

Моменты выборочного распределения и показатели асимметрии и эксцесса будем обозначать теми же буквами, что и для теоретического (генерального) распределения.

Вычисление выборочных характеристик при малом объеме выборки ($n < 60$).

Выборочное среднее значение механической характеристики вычисляется по формуле

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (12)$$

где x_i – значение механической характеристики отдельных образцов; n – число испытанных образцов (объем выборки).

Выборочная дисперсия характеристики механических свойств

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (13)$$

Выборочное среднее квадратическое отклонение и выборочный коэффициент вариации определяют по формулам:

$$S = \sqrt{S^2} \quad (14)$$

$$v = \frac{S}{\bar{x}} \cdot 100\% \quad (15)$$

4 ОПРЕДЕЛЕНИЕ ГРАНИЦ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ДЛЯ ЧИСЛОВЫХ ХАРАКТЕРИСТИК

Выборочные числовые характеристики являются надежными количественными оценками генеральных характеристик лишь при большом объеме выборки. При ограниченных объемах испытаний необходимо указать степень точности и надежности оценок генеральных характеристик.

Представление об уровне точности и надежности оценок дают *доверительные интервалы*. Для любого малого уровня значимости $\alpha > 0$ можно указать значение $\varepsilon = |\theta - \bar{\theta}|$, при котором

$$P(\bar{\theta} - \varepsilon < \theta < \bar{\theta} + \varepsilon) = 1 - \alpha, \quad (16)$$

где $\bar{\theta}$ является оценкой для параметра θ .

Если многократно повторять выборки и каждый раз находить доверительные интервалы, то в $P = (1 - \alpha) - 100\%$ случаев доверительные интервалы накроют истинное значение параметра. Вероятность $P = 1 - \alpha$, с которой доверительный интервал при многократном повторении опыта накрывает истинное значение параметра, называется *доверительной вероятностью*.

При определении доверительных интервалов уровни доверительной вероятности принимают равными 0,9; 0,95 и 0,99.

Доверительный интервал для математического ожидания определяют на основании выборочных значений \bar{x} и s из выражения

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha, k} < a < \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha, k} \quad (17)$$

где $t_{\alpha, k}$ – критерий Стьюдента, который определяется по таблицам [2] для уровня значимости α и числа степеней свободы $k = n - 1$.

Доверительный интервал для среднего квадратического отклонения вычисляют по формуле

$$sz_1 \sqrt{\frac{n-1}{n}} < \sigma < sz_2 \sqrt{\frac{n-1}{n}}, \quad (18)$$

где z_1 и z_2 – коэффициенты, зависящие от уровня доверительной вероятности и числа степеней свободы $k = n - 1$ [2].

Границы доверительных интервалов для дисперсии могут быть найдены путем возведения в квадрат значения границ доверительных интервалов для

среднего квадратического отклонения.

5 ПАРНАЯ КОРРЕЛЯЦИЯ

Исследование связей между двумя или несколькими показателями требует статистических наблюдений за всеми изучаемыми показателями. При этом возможно изучение взаимосвязей и взаимозависимостей между показателями. Для этого применяется корреляция (от английского *correlation* – соотношение, соответствие, взаимозависимость, взаимосвязь), которая дает возможность установить, насколько средняя величина одного из показателей меняется в зависимости от другого.

Парная корреляция позволяет выяснить зависимость между двумя показателями. Если имеются наблюдения за одним показателем $y_{1,1}, y_{1,2}, \dots, y_{1,n}$, а за вторым – $y_{2,1}, y_{2,2}, \dots, y_{2,n}$, то коэффициент парной корреляции будет рассчитываться по формуле

$$r_{1,2} = \left[\sum_{j=1}^n (y_{1j} - \bar{y}_1)(y_{2j} - \bar{y}_2) \right] / \sqrt{\sum_{j=1}^n (y_{1j} - \bar{y}_1)^2 \sum_{j=1}^n (y_{2j} - \bar{y}_2)^2} \quad (19)$$

где \bar{y}_j – среднее арифметическое для i -го показателя ($j = 1, 2$).

В общем случае данную формулу можно записать следующим образом:

$$r_{i,k} = \left[\sum_{j=1}^n (y_{ij} - \bar{y}_i)(y_{kj} - \bar{y}_k) \right] / \sqrt{\sum_{j=1}^n (y_{ij} - \bar{y}_i)^2 \sum_{j=1}^n (y_{kj} - \bar{y}_k)^2} \quad (20)$$

Коэффициент парной корреляции изменяется в пределах от -1 до $+1$, причем, чем ближе коэффициент к $+1$, тем сильнее между показателями связь, выражающаяся в прямой зависимости, т.е. при увеличении одного показателя увеличивается и второй, и наоборот. Чем ближе коэффициент парной корреляции к -1 , тем сильнее связь между двумя показателями, выражающаяся в обратной зависимости, т.е. при увеличении одного показателя другой уменьшается, и наоборот. При приближении коэффициента к нулю связь между показателями ослабевает, имеет место предположение о независимости показателей.

6 СУЩНОСТЬ РЕГРЕССИОННОГО АНАЛИЗА

Еще одним весьма мощным инструментом анализа результатов экспериментальных исследований и средством планирования, который предоставляет математическая статистика, является регрессионный анализ, регрессионные зависимости. Если необходимо установить функциональную зависимость одного показателя, то пользуются уравнениями регрессии, выражающими зависимость показателя-результата от показателя-фактора или функции от аргумента. *Регрессия* — линия, вид зависимости среднего результативного признака от факторного. Этот термин был введен английским статистиком Ф. Гальтоном при изучении наследственности. В случае если изучается зависимость показателя-результата от одного показателя-фактора, то такая регрессионная зависимость называется *однофакторной*. Если же изучается зависимость показателя-результата от нескольких показателей-факторов, то такого типа зависимости называются *многофакторными*.

Функциональная зависимость отыскивается следующим образом.

Задается вид функциональной зависимости, например линейная зависимость $y=a_0 + a_1x$ или квадратичная $y=a_0 + a_1x + a_2x^2$ для функции одной переменной. Выбранную зависимость подставляют в следующую функцию, подлежащую минимизации, например:

$$\sum_{j=1}^n (y_j - a_0 - a_1x_j)^2 \rightarrow \min \quad (21)$$

В этой функции рассчитывается квадрат отклонения значений аналитической кривой регрессии от значений y_j ($j= 1, 2 \dots n$). Дифференцируя по коэффициентам уравнения регрессии и приравнявая полученные значения нулю, получаем систему уравнений, из которой легко найти коэффициенты регрессионной функции. Описанный метод называется методом *наименьших квадратов*, поскольку обеспечивает минимум суммы квадратов отклонения $\sum_{j=1}^n \Delta_j^2 \rightarrow \min$ для

всех $j= 1, 2 \dots n$.

Графическая интерпретация метода дана на рис. 2.

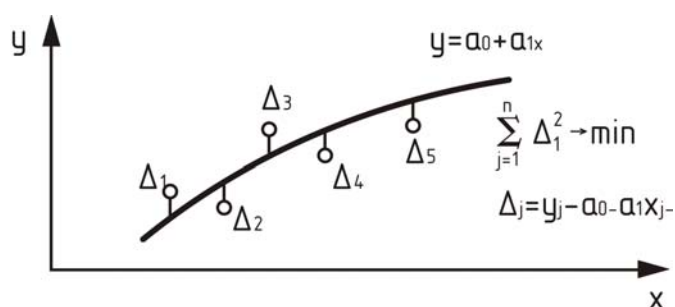


Рис. 2. Графическая интерпретация метода наименьших квадратов

7 СТАТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ СРАВНИТЕЛЬНЫХ ИСПЫТАНИЙ

В связи с тем, что выборочное среднее значение и выборочная дисперсия характеристик механических свойств являются случайными величинами, принимающими различные значения при повторных экспериментах, при исследовании влияния различных факторов на механические свойства материала или деталей при оптимизации технологии производства, выборе сплава и решении других задач возникает необходимость определения значимости или случайности в расхождениях выборочных статистик характеристик механических свойств с помощью специальных критериев.

Критерий равенства двух дисперсий

Две дисперсии сравнивают с помощью критерия F [2], для этого вычисляют отношение большей дисперсии к меньшей.

Если дисперсное отношение $F = \frac{s_1^2}{s_2^2}$ при $s_1^2 > s_2^2$ или $F = \frac{s_2^2}{s_1^2}$ при $s_1^2 < s_2^2$

больше $F_{1-\alpha/2}$ [2], принимают гипотезу о неравенстве двух генеральных дисперсий ($\sigma_1^2 > \sigma_2^2$ или $\sigma_1^2 < \sigma_2^2$). В случае соблюдения условия $F \leq F_{1-\alpha/2}$ принимают гипотезу о равенстве генеральных дисперсий ($\sigma_1^2 = \sigma_2^2$).

Критерий равенства двух средних значений

Средние значения нормально распределенных величин сравнивают с помощью t – критерия Стьюдента. Для этого в случае $\sigma_1^2 = \sigma_2^2$ вычисляют сводную дисперсию

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (25)$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (26)$$

Если $|t| \leq t_{\alpha, r}$ [2], то принимаем гипотезу о равенстве средних. В противном случае $a_1 \neq a_2$. Величину доверительной вероятности $P=1 - \alpha$ выбирают в пределах $0,90-0,99$, число степеней свободы определяют из выражения $k=n_1+n_2-2$.

В тех случаях, когда $\sigma_1^2 \neq \sigma_2^2$, проверку равенства двух средних производят с помощью приближенного t – критерия [2]. Величину t вычисляют на основании следующей формулы:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \quad (27)$$

Число степеней свободы определяют из выражения

$$\frac{1}{k} = \frac{c^2}{n_1 - 1} + \frac{(1 - c)^2}{(n_2 - 1)}, \quad (28)$$

где

$$c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \quad (29)$$

При $|t| \leq t_{\alpha, k}$ имеем $a_1 = a_2$. В противном случае $a_1 \neq a_2$.

Критерий однородности ряда дисперсий

Однородность (равенство) ряда дисперсий в случае равенства числа образцов во всех партиях оценивают с помощью критерия Кохрана, для чего вычисляют отношение

$$G_{\max} = \frac{\left[s_i^2 \right]_{\max}}{\sum_{i=1}^m s_i^2}, \quad (30)$$

где $\left[s_i^2 \right]_{\max}$ – наибольшая выборочная дисперсия; m – число партий.

Если $G_{\max} \leq G_a$ [2], то гипотеза однородности (равенства) ряда дисперсий принимается.

При неодинаковом числе образцов в отдельных партиях однородность дисперсий может быть проверена с помощью критерия Бартлета. В этом случае вычисляют величину

$$\chi^2 = \frac{2,3026}{c} \left[\lg s^2 \left(\sum_{i=1}^m n_i - m \right) - \sum_{i=1}^m (n_i - 1) \lg s_i^2 \right], \quad (31)$$

где

$$c = 1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^m n_i - m} \right), \quad (32)$$

$$s_2 = \frac{\sum_{i=1}^m (n_i - 1) s_i^2}{\sum_{i=1}^m n_i - m}, \quad (33)$$

и сравнивают с табличным значением, найденным для уровня значимости α и числа степеней свободы $k=m-1$ [2]. Если $\chi^2 \leq \chi_{\alpha}^2$, то гипотеза однородности ряда дисперсий подтверждается.

Критерий равенства ряда средних значений.

Однофакторный дисперсный анализ

Равенство (однородность) ряда средних значений, т.е. незначимость влияния различных технологических факторов производства полуфабрикатов и деталей, конструктивных особенностей испытываемых элементов, влияния условий испытания и т.д. на средние значения характеристик механических свойств оценивают с помощью дисперсионного анализа результатов испытаний.

В основе дисперсионного анализа лежит предположение о нормальности закона распределения характеристик механических свойств и однородности дисперсий. Оценки параметров закона распределения механических характеристик могут быть найдены на основании указанного анализа.

Результаты испытаний для однофакторного дисперсионного анализа представлены в табл. 1, где показана также схема предварительных вычислений.

Первоначально для каждой партии вычисляют оценки среднего значения и дисперсии, после чего проверяют гипотезу об однородности ряда дисперсий. В случае подтверждения гипотезы определяют оценку общего среднего

$$\bar{x} = \frac{\sum_{i=1}^m \bar{x}_i n_i}{\sum_{i=1}^m n_i}, \quad (34)$$

где m – общее число партий; n_i – число образцов i -й партии; \bar{x}_i – оценка среднего значения характеристики механических свойств для i -й партии.

Ход дальнейших вычислений показан в табл. 2.

Дисперсия s_1^2 характеризует рассеивание по факторам, так как она обуславливается влиянием изучаемых факторов. Дисперсия s_2^2 характеризует внутреннее рассеивание, связанное с неоднородностью конструкционных материалов, случайными колебаниями условий испытаний и т.д., и носит название внутренней или остаточной дисперсии.

Равенство (однородность) средних проверяют с помощью критерия F .

Схема вычислений при однофакторном дисперсионном анализе

№ партии	Результаты испытаний	n_i	Сумма $S_i = \sum_{j=1}^{n_i} x_{ij}$	$\bar{x}_i = \frac{S_i}{n_i}$	Сумма $SS_i = \sum_{j=1}^{n_i} x_{ij}^2$	$s_i^2 = \frac{1}{n_i - 1} \left(SS_i - \frac{S_i^2}{n_i} \right)$
1	$x_{11}, x_{12}, x_{13}, \dots,$ x_{1j}, \dots, x_{1n_1}	n_1	S_1	\bar{x}_1	SS_1	s_1^2
2	$x_{21}, x_{22}, x_{23}, \dots,$ x_{2j}, \dots, x_{2n_2}	n_2	S_2	\bar{x}_2	SS_2	s_2^2
3	$x_{31}, x_{32}, x_{33}, \dots,$ x_{3j}, \dots, x_{3n_3}	n_3	S_3	\bar{x}_3	SS_3	s_3^2
.
.
.
i	$x_{i1}, x_{i2}, x_{i3}, \dots,$ $\dots, x_{ij}, \dots, x_{in_i}$	n_i	S_i	\bar{x}_i	SS_i	s_i^2
.
.
.
m	$x_{m1}, x_{m2}, x_{m3}, \dots,$ x_{mj}, \dots, x_{mn_m}	n_m	S_m	\bar{x}_m	SS_m	s_m^2

Если дисперсионное отношение

$$F = \frac{s_1^2}{s_2^2}$$

окажется меньше табличного значения $F_{1-\alpha}$, найденного для числа степеней свободы $k_1 = m - 1$, $k_2 = k_2 = \sum_{i=1}^m n_i - m$ и уровня значимости α [2], то исследуемые

изменения в режимах технологии производства отдельных партий, конструктивные факторы, направленные изменения условий испытаний и т. д. не оказывают значимого влияния на механические свойства материала или деталей. В этом случае все рассматриваемые результаты испытаний принадлежат одной генеральной совокупности, распределенной нормально с параметрами σ^2 и α .

Оценкой σ^2 служит выборочная полная (общая) дисперсия (табл. 2), а оценкой α – выборочное общее среднее \bar{x} .

Таблица 2

Схема однофакторного дисперсионного анализа

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат (дисперсия)
Между партиями	$Q_1 = \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2$	$k_1 = m - 1$	$s_1^2 = \frac{1}{m-1} \cdot \sum_{i=1}^m n_i (\bar{x}_i - \bar{x})^2$
Внутри партий (остаточная)	$Q_2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	$k_2 = \sum_{i=1}^m n_i - m$	$s_2^2 = \frac{1}{\sum_{i=1}^m n_i - m} \cdot \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$
Полная (общая)	$Q = Q_1 + Q_2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$k = \sum_{i=1}^m n_i - 1$	$s^2 = \frac{1}{\sum_{i=1}^m n_i - 1} \cdot \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$

Доверительные интервалы для α и σ^2 могут быть найдены из выражений:

$$\bar{x} - \frac{s}{\sqrt{\sum_{i=1}^m n_i}} t_{\alpha, k} < \alpha < \bar{x} + \frac{s}{\sqrt{\sum_{i=1}^m n_i}} t_{\alpha, k} \quad (35)$$

$$sz_1 \sqrt{\frac{\sum_{i=1}^m n_i - 1}{\sum_{i=1}^m n_i}} < \sigma < sz_2 \sqrt{\frac{\sum_{i=1}^m n_i - 1}{\sum_{i=1}^m n_i}} \quad (36)$$

для $k = \sum_{i=1}^m n_i - 1$ степеней свободы. Значения $t_{\alpha, k}$, z_1 и z_2 находят по таблицам [2]

в зависимости от числа степеней свободы k и выбранного уровня доверительной вероятности.

Если справедливо неравенство

$$F = \frac{s_1^2}{s_2^2} > F_{1-\alpha},$$

то гипотеза о равенстве средних значений механических свойств отвергается. Здесь имеется m нормально распределенных генеральных совокупностей с

общей дисперсией σ^2 и разными средними значениями α_i . Оценкой генеральной дисперсии σ^2 является величина s_2^2 , а оценками генеральных средних α_i – выборочные средние \bar{x}_i (табл. 1). Доверительные интервалы для σ^2 и α_i определяют на основании выражений:

$$\bar{x}_i - \frac{s_2}{\sqrt{n_i}} t_{\alpha, k} < \alpha_i < \bar{x}_i + \frac{s_2}{\sqrt{n_i}} t_{\alpha, k} \quad (37)$$

$$s_2 z_1 \sqrt{\frac{\sum_{i=1}^m n_i - m}{\sum_{i=1}^m n_i - m + 1}} < \sigma < s_2 z_2 \sqrt{\frac{\sum_{i=1}^m n_i - m}{\sum_{i=1}^m n_i - m + 1}} \quad (38)$$

для $k = \sum_{i=1}^m n_i - m$ – степеней свободы.

Пример 1. По результатам испытаний, приведенных в табл. 3 провести дисперсионный анализ с целью проверки равенства средних значений предела прочности алюминиевого сплава.

Таблица 3

Результаты статических испытаний на разрыв образцов из профиля 15 плавок алюминиевого сплава ($n=20$)

Номер плавки	σ_a кГ/мм ²	
	\bar{x}_i	s_a^2
1	40,32	0,202
2	41,22	0,196
3	40,31	0,047
4	40,60	0,219
5	40,00	0,065
6	40,73	0,201
7	40,54	0,466
8	40,17	0,076
9	40,26	0,267
10	40,05	0,534
11	40,38	0,149
12	39,93	0,494
13	40,84	0,184
14	40,14	0,426
15	40,60	0,156
Сумма	606,09	3,682

Учитывая, что число образцов для всех плавков одинаковое ($n=20$), строят критерий однородности дисперсий по формуле (30). Самая большая выборочная дисперсия наблюдается у десятой плавки

$$G_{\max} = \frac{0,534}{3,682} = 0,145.$$

В табл. [2] для $n - 1 = 19$ и $k = m = 15$ находят $G_{0,01} = 0,156$ и $G_{0,05} = 0,139$, что не отвергает гипотезу однородности дисперсий для различных плавков.

Результаты дисперсионного анализа приведены в табл. 4.

Таблица 4

Дисперсионный анализ влияния плавочных отклонений на предел прочности образцов из алюминиевого сплава

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия
Межплавочная	34,6498	14	$s_1^2 = 2,475$
Внутриплавочная (остаточная)	73,5301	285	$s_2^2 = 0,258$
Полная (общая)	108,1799	299	$s^2 = 0,362$

Дисперсионное отношение

$$F = \frac{s_1^2}{s_2^2} = 9,58$$

при $k_1 = 14$ и $k_2 = 285$ значительно превышает $F_{0,99} = 2,15$ [2]. Следовательно, колебания в режимах технологии производства полуфабрикатов оказывают значимое влияние на предел прочности алюминиевого сплава.

8 КОРРЕЛЯЦИОННЫЙ И ДИСПЕРСИОННЫЙ АНАЛИЗ

Понятие многомерной выборки. Во время статистических наблюдений для каждого объекта в ряде случаев можно измерить значения нескольких признаков. Таким образом, получается многомерная выборка. Если многомерную выборку обработать по значениям отдельного признака, то получится обычная обработка одномерной выборки.

Смысл обработки многомерных выборок состоит в том, чтобы установить связи между признаками. Связи могут быть функциональными, т.е. каждому значению одной величины соответствует определенное значение другой величины.

Связь между случайными величинами часто носит случайный характер. Такая связь называется *стохастической* или *статистической*, если изменение одной величины вызывает изменение распределения другой величины. Если среднее значение одной случайной величины функционально зависит от значений другой случайной величины, то такая статистическая зависимость называется *корреляционной*.

Далее будем рассматривать в основном двумерные выборки.

Эмпирическая формула. Величины X и Y могут быть функционально зависимы, но на процесс измерения их значений влияют разные, в основном случайные, факторы. Установить по результатам измерений вид фактической зависимости не так просто.

Результаты измерений или наблюдений фиксируют в таблице наблюдений (табл. 5).

Таблица 5

Результаты измерений или наблюдений					
X	x_1	x_2	x_3	...	x_n
Y	y_1	y_2	y_3	...	y_n

Представим эти результаты на координатной плоскости (корреляционном поле) в виде точек, координатами которых являются значения признаков X и Y одного объекта (x_i, y_i) , $i = 1, \dots, n$ (рис. 3).

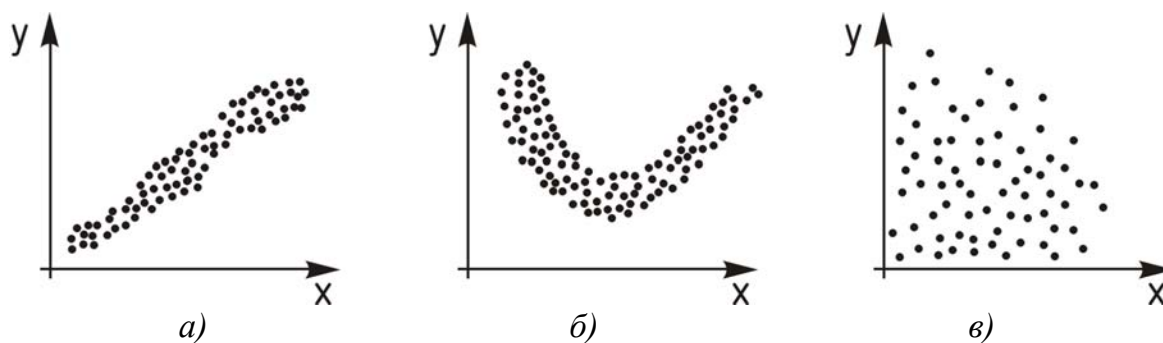


Рис. 3. Представление двумерной выборки на корреляционном поле

Из рис. 3 видно, что в случае а) следует искать линейную зависимость, в случае б) – нелинейную зависимость, а в случае в) вряд ли какая-то зависимость существует.

Конкретный вид функциональной зависимости между величинами X и Y , установленный по двумерной выборке, называют *эмпирической формулой*. Если построить график эмпирической формулы на корреляционном поле, то он должен пройти через все точки (x_i, y_i) выборки, а быть наилучшим приближением к этим точкам. Среднее расстояние этих точек от графика должно быть минимальным. При этом предпочтение отдается эмпирической формуле, имеющей более простой вид.

Простейшим видом эмпирической формулы является линейная функция

$$y = ax + b \quad (39)$$

Задача установления эмпирической формулы заключается в вычислении по выборке коэффициентов a и b в формуле (39). Аналогично можно получить и другие функции, например $y=ax^2+bx+c$ и т.д.

Нахождение линейной эмпирической формулы. Для получения линейной эмпирической формулы (39) имеется несколько способов: метод натянутой нити, метод сумм и метод наименьших квадратов.

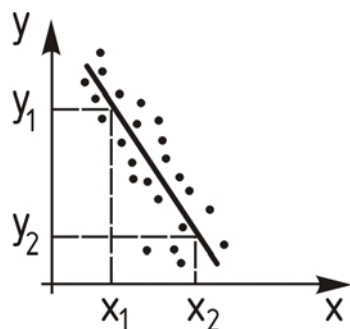


Рис. 4. Метод «натянутой нити»

В *методе натянутой нити* все результаты измерений изображают в виде точек на корреляционном поле (рис. 4). Поэтому следует мысленно натянуть между этими точками нить так, чтобы по обе стороны осталось примерно одинаковое количество точек (точнее, чтобы их суммарные отклонения в обе стороны были равными). Возьмем на прямой, совпадающей с направлением нити, две точки с координатами (x_1, y_1) и (x_2, y_2) , которые не обязательно должны присутствовать в выборке, но быть достаточно удаленными друг от друга (рис. 4). Подставив эти координаты в формулу (39), имеем систему линейных уравнений.

$$\begin{cases} y_1 = ax_1 + b; \\ y_2 = ax_2 + b. \end{cases} \quad (40)$$

где неизвестными являются коэффициенты a и b .

Решая систему (40), получаем эмпирическую формулу (39).

В *методе сумм* рассуждают следующим образом. Пусть имеется двумерная выборка, представленная в виде таблицы наблюдений (табл. 5). Предположим, что уже найдена эмпирическая формула (39). Подставим в нее значения случайной величины X из выборки и запишем соответствующие значения случайной величины Y в виде $\check{y} = ax_i + b, i=1, \dots, n$. Найдем отклонения между измеренными и вычисленными значениями Y :

$$\Delta_i = y_i - \check{y}_i = y_i - ax_i - b \quad (41)$$

Ставят условие, чтобы для всей выборки сумма этих отклонений была бы равна нулю:

$$\sum_{i=1}^n \Delta_i = 0. \quad (42)$$

Получено одно уравнение для определения коэффициентов a и b , чтобы иметь два уравнения, разделим таблицу наблюдений (табл. 5) на две части. Пусть в первой из них k наблюдений (табл. 5) на две части. Пусть в первой из них k наблюдений, тогда во второй части $n-k$ наблюдений. Потребуем, чтобы в

обеих частях выполнялось условие (42). В результате получаем следующую систему линейных уравнений:

$$\begin{cases} \sum_{i=1}^k (y_i - ax_i - b) = 0; \\ \sum_{i=k+1}^n (y_i - ax_i - b) = 0. \end{cases} \quad (43)$$

Если раскрыть скобки и просуммировать подобные члены, то получим

$$\begin{cases} \sum_{i=1}^k y_i - a \sum_{i=1}^k x_i - kb = 0; \\ \sum_{i=k+1}^n y_i - a \sum_{i=k+1}^n x_i - (n-k)b = 0. \end{cases} \quad (44)$$

Отсюда видно, что в обеих частях таблицы наблюдений нужно просуммировать значения x_i и y_i и затем решить систему (44).

Метод наименьших квадратов будет рассмотрен далее. Этот метод является самым точным, но и в тоже время и самым трудоемким. Метод *натянутой нити* самый простой, но самый неточный.

Линейная регрессия. Будем искать функцию регрессии в самом простом – линейном виде.

Имеем

$$y=f(x)=\alpha_1x+\alpha_0; \quad (45)$$

$$x=g(y)=\beta_1y+\beta_0. \quad (46)$$

Все рассуждения проводим для коэффициентов α_1 и α_0 первой функции. Формулы для вычисления β_1 и β_0 получаем из формул для α_1 и α_0 , заменяя x на y , а y на x .

В методе наименьших квадратов первоначально рассуждаем так же, как и в методе сумм. Находим отклонения между измеренными и вычисленными значениями Y :

$$\Delta_i = y_i - \tilde{y}_i = y_i - \alpha_1 x_i - \alpha_0, \quad i = 1, \dots, n \quad (47)$$

В методе наименьших квадратов коэффициенты α_1 и α_0 определяют исходя из требований, состоящего в том, чтобы сумма квадратов отклонений Δ_i , была минимальной:

$$T = \sum_{i=1}^n \Delta_i^2 = \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)^2 = \min. \quad (48)$$

Сумма (48) является функцией неизвестных коэффициентов α_1 и α_0 . Для нахождения минимума запишем частные производные:

$$\frac{\partial T}{\partial \alpha_1} = 2 \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)(-x_i); \quad (49)$$

$$\frac{\partial T}{\partial \alpha_0} = 2 \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)(-1). \quad (50)$$

Приравняв эти частные производные нулю, получим систему линейных уравнений для определения коэффициентов α_1 и α_0 . В результате несложных преобразований получаем решение системы.

$$\begin{aligned} \alpha_0 &= \bar{y} - \alpha_1 \bar{x}; \\ \alpha_1 &= \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{S}_X^2}, \end{aligned} \quad (51)$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$, $\overline{S}_X^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$.

Эти формулы можно использовать для вычисления коэффициентов функции линейной регрессии, но и последние часто вычисляют с помощью коэффициента линейной корреляции

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{S}_X \cdot \overline{S}_Y}, \quad (52)$$

где многие величины определены выше, а

$$\overline{S}_X = \sqrt{\overline{S}_X^2}, \quad \overline{S}_Y^2 = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2}. \quad (53)$$

Отсюда получаем

$$\alpha_0 = \bar{y} - \alpha_1 \bar{x}, \alpha_1 = r \cdot \frac{\overline{S}_Y}{\overline{S}_X}. \quad (54)$$

Заменяя в этих формулах величины, связанные с x , соответствующими величинами, зависящими от y , и наоборот, получим формулы для вычисления коэффициентов β_1 и β_0 функции линейной регрессии:

$$\beta_0 = \bar{x} - \beta_1 \bar{y}, \beta_1 = r \frac{\overline{S}_X}{\overline{S}_Y}. \quad (55)$$

9 ВАРИАНТЫ ЗАДАНИЙ

Цель работы: Научиться применять регрессионный и дисперсионный анализы для обработки результатов экспериментальных исследований.

Задание: Определить значения выборочных характеристик сварных образцов (σ_B , a_H). Вычислить доверительные интервалы для средних значений (σ_B , a_H), коэффициенты парных корреляций (для всех значений таблицы), критерии равенства и однородности (σ_B или a_H). Найти регрессионные однофакторные линейные зависимости $\sigma_B(I)$, $\sigma_B(V)$ методом сумм. Оценить влияние технологический факторов на результирующий фактор Z при помощи дисперсионного однофакторного анализа (при этом разделить данные на две партии $m = 2$, $n_1 = n_2 = 4$).

1-5 отнять от значений таблицы номер своего варианта, $Z = \sigma_B$

6-10 прибавить к значениям таблицы номер своего варианта, $Z = a_H$

11-20 прибавить к значениям 1 и 2 строки таблицы номер своего варианта, $Z = \sigma_B$

21-30 прибавить к значениям 2 и 4 строки таблицы номер своего варианта, $Z = a_H$

Наименование	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1. Временное сопротивление разрыву σ_B , МПа	700	750	737	721	754	730	706	770
2. Сила тока I, А	200	220	210	215	230	210	205	240
3. Скорость сварки V, м/ч	12	17	14	15	18	13	16	11
4. Ударная вязкость a_H , Дж/см ²	72	61	65	67	64	68	70	60

ЛИТЕРАТУРА

1. Колде Я.К. Практикум по теории вероятностей и математической статике: Учеб. Пособие для техникумов. – М.: Всш. Шк., 1991. – 157 с.: ил.
2. Степнов М.Н. Статистическая обработка результатов механических испытаний. М.: Машиностроение. 1972. – 232 с.
3. Турчак Л.И. Основы численных методов: Учеб. пособие. – М.: Гл. ред. Физ.-мат. лит. 1987. – 320 с.
4. Болгаев Ю. П. Вычислительная математика и программирование: Учеб. пособие для студентов вузов. – М.: Высш. шк. 1990. – 544 с.: ил.
5. <http://mathcad-2010.narod.ru/book.htm>

СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Методические указания

Составитель: доцент, к.т.н. Д.А. Чинахов

Подписано к печати _____

Формат 60×84/16. Бумага офсетная.

Плоская печать. Усл. печ. л. 1,5 . Уч. изд.л. 1,26

Тираж _____ экз. Заказ № _____.

ЮТИ ТПУ. Лицензия ПЛД № 44-55 от 04.12.97.

Ризограф ЮТИ ТПУ. 652055, Юрга, ул. Московская, 17.