МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

А. Н. Леонов, М. М. Дечко, В. Б. Ловкис

ОСНОВЫ МОДЕЛИРОВАНИЯ

*Рекомендовано Учебно-методическим объединением
по аграрному техническому образованию
в качестве учебно-методического пособия
для студентов учреждений высшего образования
по группе специальностей 74 06 Агроинженерия
и специальности 1-36 12 01 «Проектирование и производство
сельскохозяйственной техники»*

Минск
БГАТУ
2020

УДК 631.15:001.891.54(07)
ББК 22.144я7
Л47

Рецензенты:
кафедра технологии машиностроения БНТУ
(член-корреспондент НАН Беларуси, доктор технических наук,
профессор, заведующий кафедрой *В. К. Шелег*);
доктор технических наук, профессор, главный научный сотрудник
лаборатории механизации приготовления концентрированных кормов
«НПЦ НАН Беларуси по механизации сельского хозяйства» *В. И. Передня*

Леонов, А. Н.
Л47 Основы моделирования : учебно-методическое пособие / А. Н. Леонов, М. М. Дечко, В. Б. Ловкис ; под ред. А. Н. Леонова. – Минск : БГАТУ, 2020. – 160 с.
ISBN 978-985-25-0033-3.

Содержит теоретические сведения и примеры решения прикладных задач по трем темам: детерминированное моделирование технических систем, включающее алгоритм корректного оформления результатов инженерных и научных расчетов; оценка параметров генеральной совокупности выборочным методом, включающим оценку качества серийной и однотипной продукции; стохастическое моделирование и оптимизация однофакторных технических систем.

Рекомендовано для студентов I ступени высшего образования, обучающихся по группе специальностей 74 06 Агроинженерия, специальности 1-36 12 01 «Проектирование и производство сельскохозяйственной техники» и по специальностям 1-53 01 01-09 «Автоматизация технологических процессов и производство (по сельскому хозяйству)», 1-54 01 01-09 «Метрология, стандартизация и сертификация» (определенных протоколом компетенций), преподавателей учреждений высшего образования инженерного профиля.

УДК 631.15:001.891.54(07)
ББК 22.144я7

ISBN 978-985-25-0033-3

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Математическое моделирование и оптимизация являются эффективным методом решения прикладных задач. Благодаря широкому внедрению систем компьютерной математики в настоящее время появились новые возможности для разработки инновационных технологий и устройств, а также ускоренного внедрения их в народное хозяйство.

Математическое моделирование технических систем строится на двух принципиально различных подходах: детерминированном и стохастическом.

Детерминированное моделирование технических систем базируется на знании фундаментальных законов природы. Отличительная особенность детерминированного моделирования заключается в том, что заданному комплексу управляющих факторов всегда со 100%-й вероятностью соответствует только одно значение параметра оптимизации.

Стохастическое моделирование базируется на теории вероятностей и математической статистики. Отличительная особенность стохастического моделирования заключается в том, что заданному комплексу управляющих факторов соответствуют различные значения параметра оптимизации, причем каждое значение реализуется с некоторой вероятностью.

Стохастическое моделирование позволяет описывать функционирование сложных технических систем, в которых происходящие явления затруднительно описать конечным набором фундаментальных законов природы. Причина – влияние большого количества факторов различной природы, неконтролируемо воздействующих на техническую систему. Стохастические модели строятся путем обработки экспериментальных данных.

Научно-техническая литература, посвященная стохастическому моделированию, требует знаний специальных областей математики, которые изучаются на механико-математических факультетах университетов. В предлагаемом учебно-методическом пособии изложение основ прикладной математической статистики базируется на знании алгебры в объеме программы средней школы и математического анализа на уровне первого курса технического университета.

Решение прикладных задач методами математического моделирования и оптимизации включает в себя следующие этапы: выбор параметра оптимизации, характеризующего функционирование технической системы; выбор комплекса управляющих факторов, существенно влияющих на параметр оптимизации; выбор математической модели для описания функционирования технической системы; построение плана эксперимента; обработка экспериментальных данных методами математической статистики и регрессионного анализа; оптимизация. Сочетание перечисленных этапов позволяет решать прикладные задачи, направленные на изучение функционирования изучаемой технической системы, а также нахождение оптимального комплекса управляющих факторов, предопределяющих функционирование технической системы, например, при минимально возможных материальных затратах.

Учебно-методическое пособие содержит основные понятия и положения учебной дисциплины «Основы моделирования», цель которой – дать студентам I ступени высшего образования знания по основам детерминированного и стохастического моделирования; изложить теорию и алгоритмы, необходимые для решения прикладных задач, в том числе и в области АПК.

Пособие состоит из трех модулей, каждый из которых содержит теорию, алгоритмы решения прикладных задач, численное решение типовой задачи, вопросы для самопроверки с ответами. Пособие охватывает относительно небольшой класс прикладных задач, однако приобретенные студентами теоретические знания и практические навыки будут полезны при изучении теории, необходимой для решения более сложных задач.

В основу издания положен курс лекций, читавшийся авторами по учебной дисциплине «Основы моделирования» в БГАТУ в течение 10 лет.

1. ДЕТЕРМИНИРОВАННОЕ МОДЕЛИРОВАНИЕ ТЕХНИЧЕСКИХ СИСТЕМ

В результате изучения темы студенты должны:

– **усвоить** концептуальное различие детерминированных и стохастических закономерностей при моделировании технических систем;

– **уметь** оценивать абсолютную и относительную погрешность параметров, подчиняющихся детерминированным закономерностям;

– **уметь** корректно оформлять результаты инженерных расчетов и научных исследований, в том числе в области АПК.

1.1. Предварительная обработка экспериментальных данных

1.1.1. Детерминированное и стохастическое моделирование

Познание окружающего мира человечеством осуществляется на основе наблюдений повторяющихся явлений с наложением на них некоторых концепций философского характера. Математическое описание технических систем исторически началось с изучением явлений, подчиняющихся детерминированным закономерностям, основанным на следующем постулате: если эксперимент выполнен при одном и том же комплексе управляющих факторов, то его результат всегда со 100%-й вероятностью одинаковый. Примером детерминированного описания технических систем являются классическая механика и электродинамика, расцвет которых пришелся на XIX в.

По мере накопления знаний о поведении технических систем, подчиняющихся детерминированным закономерностям, были получены знания о функционировании систем, подчиняющихся стохастическим закономерностям. Поведение технических систем основано в этом случае на следующем постулате: если эксперимент выполнен при одном и том же комплексе управляющих факторов, то, несмотря на это, его результат неоднозначен и может быть спрогнозирован методами математической статистики. Неоднозначность параметра оптимизации технической системы, подчиняющегося стохастическим закономерностям, предопределяется влиянием на техническую систему большого количества случайных неуправляемых факторов, относительно небольших по величине по сравнению с управляющими факторами.

В качества примера рассмотрим техническую систему «Технологический процесс изготовления антифрикционных втулок из порошка медно-оловянной бронзы методом порошковой металлургии», включающий следующие операции: получение металлического порошка; прессование заготовок; спекание в атмосфере водорода, калибровку. В качестве параметра, который характеризует состояние технической системы, выберем, например, прочность втулок при сжатии. Комплекс факторов при серийном изготовлении один и тот же для всех втулок: состав 95 % Cu и 5 % Sn; диаметр частиц 100 мкм; давление прессования 150 МПа; температура спекания при 800 °С; калибровка одним и те же штампом. Несмотря на фиксированные значения факторов, значения прочности при сжатии 95 % втулок оказались в интервале 436–481 МПа, который в рамках выбранного технологического процесса изменить невозможно.

В научной литературе можно встретить понятия «хорошо организованные» (детерминированные) и «плохо организованные» (стохастические) технические системы. Особенность научного познания XX в. заключается в появлении новой методологии, дополняющей изучение «хорошо организованных» технических систем изучением «плохо организованных». Несмотря на концептуальное различие детерминированных и стохастических закономерностей в описании исследуемых технических систем, общим у них является наличие некоторой организации, постоянной в пространстве и времени. Современная наука приступила к созданию методов моделирования технических систем, внутренняя структура которых со временем усложняется в результате эволюционной самоорганизации.

1.1.2. Абсолютная и относительная погрешность параметра, подчиняющегося детерминированным закономерностям

Моделирование технических систем, подчиняющихся детерминированным закономерностям, включает в себя расчет параметров по фундаментальным законам природы. В этом случае детерминированный параметр определяется с абсолютной погрешностью, обусловленной погрешностью измерительной системы $\Delta Y_{\text{пр}}$. Следует отметить, что для технических систем, подчиняющихся детерминированным закономерностям, использование более точного измерительного инструмента позволяет уменьшить абсолютную погрешность параметра, характеризующего функционирование технической системы. Например, определение линейных размеров может быть выполнено с различной абсолютной погрешностью: 10 мм (швейный метр); 1 мм (металлическая линейка); 0,1 мм (штангенциркуль); 0,01 мм (микрометр); 0,001 мм (оптический микроскоп).

В отличие от технических систем, подчиняющихся детерминированным закономерностям, абсолютная погрешность параметра Y, подчиняющегося стохастическим закономерностям, обусловлена стохастической природой объекта $\Delta Y_{\text{ст}}$. Следует отметить, что для технических систем, подчиняющихся стохастическим закономерностям, использование более точного измерительного инструмента не позволяет уменьшить абсолютную погрешность параметра, характеризующего функционирование технической системы.

В общем случае абсолютная погрешность параметра, характеризующего функционирование технической системы, математически записывается следующим образом:

$$\Delta Y^2 = \Delta Y_{\text{пр}}^2 + \Delta Y_{\text{ст}}^2 . \tag{1}$$

Если $\Delta Y_{\text{пр}} >> \Delta Y_{\text{ст}}$, то $\Delta Y \approx \Delta Y_{\text{пр}}$. В этом случае стохастической природой изучаемой технической системы можно пренебречь, так как система с достаточной точностью подчиняется детерминированным закономерностям.

Если $\Delta Y_{\text{ст}} >> \Delta Y_{\text{пр}}$, то $\Delta Y \approx \Delta Y_{\text{ст}}$. В этом случае погрешностью измерительного инструмента можно пренебречь, так как изучаемая техническая система с достаточной точностью подчиняется стохастическим закономерностям.

Изложим метод определения абсолютной ΔY и относительной погрешности $\Delta Y/Y$ рассчитанного параметра Y, характеризующего функционирование технической системы, подчиняющейся детерминированным закономерностям. Пусть задана некоторая аналитическая функция $Y = Y(x, z)$ от двух детерминированных факторов x и z, которые определены с абсолютной погрешностью $x \pm \Delta x$ и $z \pm \Delta z$ (аналитическая функция и ее производные являются непрерывными функциями). Очевидно, что в этом случае рассчитанный параметр Y также будет вычислен с абсолютной погрешностью $Y \pm \Delta Y$.

Прежде чем вывести формулы для определения абсолютной и относительной погрешности детерминированного параметра Y, следует помнить, что если $Y = Y(x, z)$, то дифференциал функции двух переменных рассчитывается по уравнению

$$\mathrm{d}Y(x, z) = Y'_x(x, z)\,\mathrm{d}x + Y'_z(x, z)\,\mathrm{d}z, \tag{2}$$

где Y'_x (Y'_z) – частные производные параметра Y по факторам x и z.

Пусть параметр Y является суммой или разностью двух функций $f(x)$ и $g(z)$, причем факторы x и z определяются с абсолютными погрешностями $x \pm \Delta x$ и $z \pm \Delta z$. Для оценки абсолютной погрешности параметра Y рассчитаем дифференциал dY, используя уравнение (2):

$$Y(x, z) = f(x) \pm g(z), \tag{3}$$

$$\mathrm{d}Y = \left[f(x) \pm g(z) \right]'_x \mathrm{d}x + \left[f(x) \pm g(z) \right]'_z \mathrm{d}z = f'(x)\,\mathrm{d}x \pm g'(z)\,\mathrm{d}z\,. \tag{4}$$

По определению dx = Δx, dz = Δz. Кроме того, если выполняется условие $\Delta x << x$ и $\Delta z << z$, то можно утверждать, что

$$\mathrm{d}Y » \Delta Y. \tag{5}$$

Заменив дифференциалы соответствующими приращениями, получим уравнение для расчета абсолютной погрешности параметра Y:

$$\Delta Y = \left| f'(x) \right| \Delta x + \left| g'(z) \right| \Delta z\,. \tag{6}$$

Следует обратить внимание на то, что знак «±» в уравнении (4) заменен на знак «+» в уравнении (6), а значения производных взяты по модулю. Это продиктовано необходимостью рассчитать максимально возможную абсолютную погрешность.

Пусть параметр Y является произведением двух функций $f(x)$ и $g(z)$, причем факторы x и z определяются также с абсолютными погрешностями $x \pm \Delta x$ и $z \pm \Delta z$. Для оценки относительной погрешности параметра Y рассчитаем дифференциал dY, используя уравнение (2):

$$Y(x, z) = f(x)\, g(z), \tag{7}$$

$$\mathrm{d}Y = \left[f(x)\, g(z) \right]'_x \mathrm{d}x + \left[f(x)\, g(z) \right]'_z \mathrm{d}z\,, \tag{8}$$

$$\mathrm{d}Y = f'(x) g(z)\,\mathrm{d}x + f(x)\, g'(z)\,\mathrm{d}z\,. \tag{9}$$

Разделив уравнение (9) на уравнение (7) и заменив дифференциалы соответствующими приращениями (6), получим уравнение для определения относительной погрешности параметра *Y*, который равен произведению двух функций:

$$\frac{\Delta Y}{Y}=\left|\frac{f'(x)}{f(x)}\right|\Delta x+\left|\frac{g'(z)}{g(z)}\right|\Delta z\ . \qquad (10)$$

Используя такую же процедуру, найдем предельную относительную погрешность параметра *Y*, который является отношением двух функций $f(x)$ и $g(z)$, причем факторы x и z также определяются с абсолютными погрешностями $x \pm \Delta x$ и $z \pm \Delta z$:

$$Y(x, z)=\frac{f(x)}{g(z)}, \qquad (11)$$

$$\mathrm{d}Y=\left[\frac{f(x)}{g(z)}\right]_x'\mathrm{d}x+\left[\frac{f(x)}{g(z)}\right]_z'\mathrm{d}z\ , \qquad (12)$$

$$\mathrm{d}Y=\frac{f'(x)}{g(z)}\mathrm{d}x-\frac{f(x)g'(z)}{g^2(z)}\mathrm{d}z\ , \qquad (13)$$

так как

$$\left[\frac{1}{g(z)}\right]_z'=\left[g^{-1}(z)\right]_z'=(-1)g^{-2}(z)g'(z)=-\frac{g'(z)}{g^2(z)}. \qquad (14)$$

Разделив уравнение (13) на уравнение (11), получим:

$$\frac{\mathrm{d}Y}{Y}=\frac{f'(\mathrm{x})}{f(x)}\mathrm{d}x-\frac{g'(z)}{g(z)}\mathrm{d}z\ . \qquad (15)$$

Заменив дифференциалы соответствующими приращениями (см. уравнение (5)) с учетом пояснений, данных после уравнения (6), получим:

$$\frac{\Delta Y}{Y}=\left|\frac{f'(x)}{f(x)}\right|\Delta x+\left|\frac{g'(z)}{g(z)}\right|\Delta z\ . \qquad (16)$$

Уравнения (10) и (16) тождественны, следовательно, относительная погрешность параметра Y рассчитывается по одной и той же формуле независимо от того, является ли параметр Y произведением или отношением двух функций $f(x)$ и $g(z)$ (уравнения (7) и (11)).

Абсолютная погрешность ΔY параметра Y с учетом уравнений (10), (16) рассчитывается следующим способом:

$$\Delta Y = \left[\left|\frac{f'(x)}{f(x)}\right|\Delta x + \left|\frac{g'(z)}{g(z)}\right|\Delta z\right] Y\,. \tag{17}$$

1.1.3. Алгоритм округления результата инженерных и научных расчетов

При решении практических задач параметры Y, ΔY и $\Delta Y/Y$ могут быть записаны в виде бесконечной десятичной дроби с любым количеством цифр, например, $\rho = M/V$ = 7,834 609 948 500 71 г/см^3. Однако приборы, применяемые для технических измерений, обычно обеспечивают относительную погрешность ≈ 5 %. Поэтому для корректной записи рассчитанного параметра необходимо ввести алгоритм округления результатов расчета.

Для изложения алгоритма корректного оформления результата расчета напомним несколько известных положений из школьного курса математики.

1. Первая цифра, находящаяся слева, отличная от нуля, называется первой значащей цифрой. Все остальные цифры после первой значащей цифры также являются значащими цифрами. (*Обратите внимание*! *Все цифры после значащей, в том числе и нули.*) Например, в числе 0,004 567 – первая значащая цифра 4, в числе 567,258 – первая значащая цифра 5; в числе 231 400 – значащих цифр шесть; в числе 1,000 значащих цифр четыре; в числе 0,003 79 значащих цифр три.

2. Если в округляемом числе задана цифра, до которой его следует округлить, то все цифры после округляемой заменяются нулями, причем: если следующая цифра после округляемой равна 0, 1, 2, 3, 4, то округляемая цифра остается без изменения; если после округляемой следующая цифра 5, 6, 7, 8, 9, то округляемая цифра увеличивается на единицу. Нули после округляемой цифры, если они стоят в правой части после запятой, не пишутся.

Рассмотрим несколько примеров (**округляемая цифра выделена**).

5**4**23,056 $\to$ 5**4**00,000 $\to$ 5**4**00;

5**4**83,056 $\to$ 5**5**00,000 $\to$ 5**5**00;

1,9**8**32 $\to$ 1,9**8**00$\to$1,9**8**;

1,**9**832 $\to$ 2,**0**000$\to$2,**0**;

0,000 **9**83 12 $\to$ 0,001 **0**00 0 $\to$ 0,001**0**.

В общем случае округляемая цифра выбирается либо исследователем на основании опыта и интуиции, либо определяется на основе абсолютной погрешности рассчитанного параметра. Первый случай распространяется, как правило, на округление констант, например: π = 3,141 592 654…, e = 2,718 281 828… Второй случай требует более детального рассмотрения, поэтому проиллюстрируем его на следующем примере.

Рассчитаем скорость V велосипедиста, если расстояние L = (1278 ± 6) м он проехал за время t = (103 ± 0,7) с:

$$V = L \,/\, t = 1278 \,/\, 103 = 12{,}407\ 77\ldots \text{ м/с}.$$

Убедимся, что рассчитанный параметр записан с избыточной точностью, для чего рассчитаем его минимальное и максимальное значения:

$$V_{\min} = L_{\min} \,/\, t_{\max} = (1278 - 6) \,/\, (103{,}0 + 0{,}7) = 12{,}266\ 15\ldots$$

$$V_{\max} = L_{\max} \,/\, t_{\min} = (1278 + 6) \,/\, (103{,}0 - 0{,}7) = 12{,}551\ 32\ldots$$

Сравнивая предельные значения скорости, можно определить, что

$$\Delta V = (V_{\max} - V_{\min}) \,/\, 2 = (12{,}55132 - 12{,}26615) \,/\, 2 = 0{,}142\ 58\ldots \quad (18)$$

Проанализируем полученное выражение. Во-первых, можно утверждать, что достоверно значащими цифрами в рассчитанном параметре V являют только первые две цифры – 1 и 2. Следующие цифры после достоверных цифр позволяют определить интервал, в котором может находиться истинное значение рассчитанного параметра. Во-вторых, абсолютную погрешность нецелесообразно записывать, например, до четырех и более значащих цифр, если уже ее первая значащая цифра свидетельствует о порядке ошибки.

Интервал, в котором находится рассчитанный параметр, зависит от абсолютной погрешности определения факторов. Неизвестно, чему равно истинное расстояние, которое проехал велосипедист. Известно только, что

$$1272 \le L \le 1284 \text{ м.}$$

То же самое относится и ко времени. Мы не знаем, чему равно истинное время, за которое велосипедист проехал свой путь. Знаем только, что

$$102{,}3 \le t \le 103{,}7 \text{ с.}$$

В силу неопределенности интервал скорости велосипедиста может определяться одним из следующих выражений (см. уравнение (18)):

$$V \pm \Delta V_1 = 12{,}4 \pm 0{,}1 \text{ или } 12{,}3 \le V \le 12{,}5;$$

$$V \pm \Delta V_2 = 12{,}41 \pm 0{,}14 \text{ или } 12{,}27 \le V \le 12{,}55;$$

$$V \pm \Delta V_3 = 12{,}408 \pm 0{,}143 \text{ или } 12{,}265 \le V \le 12{,}551.$$

Округленная абсолютная погрешность $\Delta V_1 = 0{,}1$ (одна значащая цифра) отличается от «истинной» $\Delta V = 0{,}142\,58\ldots$ на 30 %, $\Delta V_2 = 0{,}14$ (две значащие цифры) – на 2 %, $\Delta V_3 = 0{,}143$ (три значащие цифры) – на 0,3 %. Поэтому в качестве абсолютной погрешности необходимо взять величину $\Delta V_2 = 0{,}14$ с двумя значащими цифрами, так как $\Delta V_1 = 0{,}1$ отличается от истинного значения ΔV на достаточно большую величину – 30 %, что является недопустимо большим отклонением; $\Delta V_3 = 0{,}143$ отличается от истинного значения ΔV на 0,3 %, однако такая точность является избыточной, так как используемые приборы не могут обеспечить такую точность.

Если в результате промежуточных расчетов получено значение

$$V \pm \Delta V = (12{,}407\,77 \pm 0{,}142\,58) \text{ м/с},$$

то, применив сформулированные выше рекомендации, окончательный ответ следует записать в следующем виде:

$$V \pm \Delta V = (12{,}41 \pm 0{,}14) \text{ м/с.}$$

Определение величины абсолютной погрешности путем предварительного расчета минимального и максимального значения рассчитанного параметра – достаточно длительная операция, особенно если учесть, что изучаемый параметр может задаваться сложной алгебраической формулой. Для расчета абсолютной погрешности проще сначала рассчитать относительную погрешность по уравнению (16):

$$\frac{\Delta V}{V} = \frac{\Delta L}{L} + \frac{\Delta t}{t} = \frac{6}{1278} + \frac{0,7}{103,0} = 0,011\,491... \qquad (19)$$

потом – абсолютную погрешность по уравнению (10):

$$\Delta V = \left(\frac{\Delta V}{V}\right)V = 0,011491 \cdot 12,407\,77 = 0,142\,58... \qquad (20)$$

Обратите внимание! *Абсолютная погрешность, определенная экспериментально (см. уравнение (18)), с большой точностью совпадает с абсолютной погрешностью, определенной теоретически по уравнениям (19), (20).*

Таким образом, алгоритм округления детерминированного параметра существует только при следующих условиях: либо субъективно задана округляемая цифра и ее разряд, либо известна абсолютная погрешность рассчитанного параметра.

Если известна абсолютная погрешность, то алгоритм округления рассчитанного параметра, подчиняющегося детерминированным закономерностям, может быть сформулирован следующим образом.

Алгоритм округления результата инженерных и научных расчетов:

– начинаем округление с абсолютной погрешности;

– если абсолютная погрешность ΔY начинается с цифр 1, 2, 3, 4, 5 («легкая» погрешность), то ее следует округлить до двух значащих цифр;

– если абсолютная погрешность ΔY начинается с цифр 6, 7, 8, 9 («тяжелая» погрешность), то ее следует округлить до одной значащей цифры;

– после округления абсолютной погрешности ΔY рассчитанный параметр Y следует округлить до значащих цифр округленной абсолютной погрешности;

– алгоритм округления относительной погрешности $\Delta Y/Y$ такой же, как и алгоритм округления абсолютной погрешности ΔY.

Округление ΔY и Y производится лишь в окончательном ответе. Все промежуточные результаты расчетов следует округлять *минимум* до четырех значащих цифр, что гарантирует абсолютную погрешность менее 0,1 %.

Следует отметить два «подводных камня», при работе с алгоритмом корректного округления рассчитанного параметра, подчиняющегося стохастическим закономерностям:

1) абсолютную погрешность (например, $\Delta Y = 0{,}596\,82$) следует корректно округлить до первых двух значащих цифр, так как она «легкая», то есть $DY = 0{,}596\,82 \approx 0{,}60$. Однако после округления ΔY из «легкой» превратилась в «тяжелую». В этом случае ΔY следует повторно округлить до одной значащей цифры, так как ΔY теперь стала «тяжелой», то есть $DY = 0{,}596\,82 \approx 0{,}60 \approx 0{,}6$.

2) абсолютную погрешность (например, $\Delta Y = 0{,}966\,82$) следует корректно округлить только до одной значащей цифры, так как ΔY является «тяжелой», то есть $\Delta Y = 0{,}966\,82 \approx 1$. Однако после округления ΔY из «тяжелой» превратилась в «легкую». В этом случае ΔY следует повторно округлить до двух значащих цифр, так как ΔY превратилась в «легкую», то есть $\Delta Y = 0{,}966\,82 \approx 1 \approx 1{,}0$.

Приведем несколько примеров корректной записи результатов расчета:

а) 19,5687 ± 0,012 357 → 19,5**69** ± 0,0**12**;

б) 4,3251 ± 0,196 206 → 4,**33** ± 0,**20**;

в) 6,3555 ± 0,594 200 → 6,**36** ± 0,**59**;

г) 6,3555 ± 0,597 22 → 6,**36** ± 0,**60** → 6,**4** ± 0,**6**;

д) 4,355 ± 0,725 006 → 4,**4** ± 0,**7**;

е) 54,325 ± 0,098 544 → 54,**3** ± 0,**1**→ 54,**33** ± 0,**10**;

ж) 1931,62 ± 48,363 82 → 1 9**32** ± **48**;

з) 7249,92 ± 592,3634 → 7 **25**0 ± **59**0;

и) 5939,92 ± 598,3609 → 5 **9**00 ± **6**00;

к) 4 987 456,92 ± 8597,364 70 → 4 98**7** 000 ± **9**000;

л) 5675,456 40 ± 0,960 56 → 567**5** ± **1**→ 567**5,5** ± **1,0**;

м) 978 256 ± 98 432 → **98**0 000 ± **10**0 000;

н) 998 256 ± 95 555 → 1 **00**0 000 ± **10**0 000.

Обратите внимание! *В примерах г), и) «легкая» погрешность при повторном округлении переходит в «тяжелую». В примерах е), л), м), н) «тяжелая» погрешность при повторном округлении переходит в «легкую».*

1.1.4. Алгоритм корректного оформления результата инженерных и научных расчетов

1. Если рассчитан параметр $Y = f(x) \pm g(z)$, причем факторы определены с абсолютной погрешностью $x \pm \Delta x$ и $z \pm \Delta z$, при условии $\Delta x / x << 1$ и $\Delta z / z << 1$, то абсолютная погрешность ΔY рассчитывается по следующему уравнению:

$$\Delta Y = \left| f'(x) \right| \Delta x + \left| g'(z) \right| \Delta z .$$

2. Если параметр Y рассчитан по уравнению $Y = f(x)\, g(z)$ или $Y = \dfrac{f(x)}{g(z)}$, причем факторы определены с абсолютной погрешностью $x \pm \Delta x$ и $z \pm \Delta z$, при условии $\Delta x / x << 1$ и $\Delta z / z << 1$, то относительная $\Delta Y / Y$ и абсолютная погрешность ΔY рассчитываются по следующим уравнениям:

$$\frac{\Delta Y}{Y} = \left| \frac{f'(x)}{f(x)} \right| \Delta x + \left| \frac{g'(z)}{g(z)} \right| \Delta z,$$

$$\Delta Y = \left(\left| \frac{f'(x)}{f(x)} \right| \Delta x + \left| \frac{g'(z)}{g(z)} \right| \Delta z \right) Y.$$

3. Алгоритм корректного оформления рассчитанного параметра Y, подчиняющегося детерминированным закономерностям:

– начинаем округление с абсолютной погрешности;

– если абсолютная погрешность ΔY начинается с цифр 1, 2, 3, 4, 5 («легкая» погрешность), то ее следует округлить до двух значащих цифр;

– если абсолютная погрешность ΔY начинается с цифр 6, 7, 8, 9 («тяжелая» погрешность), то ее следует округлить до одной значащей цифры;

– после округления абсолютной погрешности ΔY рассчитанный параметр Y следует округлить до значащих цифр округленной абсолютной погрешности;

– алгоритм округления относительной погрешности $\Delta Y / Y$ такой же, как и алгоритм округления абсолютной погрешности ΔY.

4. Округление ΔY и Y производится лишь в окончательном ответе. Все промежуточные результаты расчетов следует округлять *минимум* до четырех значащих цифр.

1.1.5. Типовая задача: простейшие детерминированные модели

На токарном станке выточили две заготовки $\varnothing 40 \times 20$ из сталей двух марок: сталь Р18 и Сталь 45. Знание марки необходимо, так как детали из стали Р18, полученные путем закалки, обладают бóльшей прочностью и твердостью по сравнению с деталями из Стали 45. К сожалению, заготовки не были промаркированы и их перепутали. Возникла необходимость идентифицировать материал заготовок.

Существует относительно большое количество различных способов определения марки стали. На данном производстве в наличии имелся достаточно широкий набор весов и инструментов для определения линейных размеров. Поэтому было принято решение определить марку стали достаточно и дешевым способом – по плотности заготовок. По справочным данным известно, что ρ_{18} = 8,800 г/см^3, а ρ_{45} = 7,826 г/см^3.

Допустим, что справочные данные по плотности рассматриваемых марок стали приведены с относительной погрешностью в 1 % (однако есть предположение, что их относительная погрешность не превышает 0,3 %). Тогда справочные данные можно переписать следующим образом:

$$\rho_{18} = \rho_{18} \pm \Delta\rho_{18} = 8{,}800 \pm 8{,}800 \cdot 0{,}01 = (8{,}80 \pm 0{,}09)\ \text{г/см}^3\ ,$$

$$\rho_{45} = \rho_{45} \pm \Delta\rho_{45} = 7{,}826 \pm 7{,}826 \cdot 0{,}01 = (7{,}83 \pm 0{,}08)\ \text{г/см}^3\ .$$

В наихудшем случае, при котором найденные плотности дадут наиболее близкие значения, плотность обеих марок стали может быть равна следующим величинам (марка стали Р18 будет определена меньше номинального значения, плотность Стали 45 – больше номинала).

$$\rho_{18} = 8{,}80 - 0{,}09 = 8{,}71\ \text{г/см}^3\ ,$$

$$\rho_{45} = 7{,}86 + 0{,}08 = 7{,}91\ \text{г/см}^3\ .$$

Массу заготовок можно определить на весах различного класса точности с абсолютной погрешностью 10; 1; 0,1; 0,01; 0,001; 0,0001 г. Линейные размеры можно определить также различными инструментами с абсолютной погрешностью 1; 0,1; 0,01; 0,001 мм. Следует отметить, что чем точнее инструмент, тем он дороже и тем труднее найти его в повседневной практике. Поэтому главная цель – подобрать дешевый и доступный инструмент, который позволил бы определить массу и линейные размеры заготовок с такой точностью, чтобы рассчитанная по результатам измерений плотность заготовок позволила бы со 100%-й вероятностью идентифицировать марку стали.

Расчет плотности заготовок производится по формуле

$$\rho = \frac{M}{V} = \frac{4M}{\pi D^2 H}, \qquad (21)$$

где ρ – плотность заготовки, г/см^3; M – масса заготовки, г; V – объем заготовки, см^3; D – диаметр заготовки, см; H – высота заготовки, см; π = 3,141 592 654... (число π приведено с большой точностью, так как пока неизвестно, до какого знака его следует округлить).

Измерив массу, диаметр и толщину заготовок, а затем рассчитав их плотность по формуле (21), можно было бы идентифицировать марку стали каждой заготовки. Однако в связи с тем, что эти три фактора определяются экспериментально с абсолютной погрешностью $M \pm \Delta M$, $D \pm \Delta D$, $H \pm \Delta H$, то и рассчитанная плотность также будет определена с абсолютной погрешностью $\rho \pm \Delta\rho$. Если абсолютная погрешность плотности заготовок превысит некоторую критическую величину, то идентифицировать заготовки по плотности со 100%-й вероятностью будет невозможно. Поясним это утверждение на конкретном примере.

Рассчитаем относительную погрешность плотности по формуле (см. уравнение (10)), причем первые три слагаемых правой части уравнения имеет смысл выбрать таким образом, чтобы они были равны между собой:

$$\frac{\Delta\rho}{\rho} = \frac{\Delta M}{M} = 2\frac{\Delta D}{D} = \frac{\Delta H}{H} = 3x, \qquad (22)$$

где $x = \frac{\Delta M}{M} = 2\frac{\Delta D}{D} = \frac{\Delta H}{H}$.

Допустим, что относительная погрешность равна x = 0,02. В этом случае

$$\frac{\Delta\rho}{\rho} = 3x = 0,06\,,$$

$$\Delta\rho = 0,06\rho\,,$$

$$\Delta\rho_{18} = 0,06\rho_{18} = 0,06 \cdot 8,71 = 0,5226 = 0,52 \text{ г/см}^3,$$

$$\Delta\rho_{45} = 0,06\rho_{45} = 0,06 \cdot 7,91 = 0,4746 = 0,47 \text{ г/см}^3.$$

Если x = 0,02, то наихудший случай:

$$(\rho_{18})_{\min} = \rho_{18} - \Delta\rho_{18} = 8,71 - 0,52 = 8,19 \text{ г/см}^3,$$

$$(\rho_{45})_{\max} = \rho_{45} + \Delta\rho_{45} = 7,91 + 0,47 = 8,38 \text{ г/см}^3,$$

что не позволяет со 100%-й вероятностью идентифицировать марку стали, так как при реализованной относительной погрешности определения массы, диаметра и толщины заготовок расчетное значение плотности Стали Р18 может оказаться меньше плотности Стали 45, что противоречит справочным данным.

Приведенный пример показывает, что допустимая относительная погрешность в определении плотности заготовок должна определяться неравенством

$$\rho_{18} - \Delta\rho_{18} > \rho_{45} + \Delta\rho_{45}\,. \tag{23}$$

Рассчитаем допустимое значение относительной погрешности x:

$$\Delta\rho_{18} = 3x\rho_{18}\,,$$

$$\Delta\rho_{45} = 3x\rho_{45}\,.$$

Из неравенства (23) следует:

$$\rho_{18} - 3x\rho_{18} > \rho_{45} + 3x\rho_{45}\,,$$

$$\rho_{18} - \rho_{45} > 3x(\rho_{18} + \rho_{45})\,,$$

$$x < \frac{\rho_{18} - \rho_{45}}{3(\rho_{18} - \rho_{45})} = \frac{8,71 - 7,91}{3(8,71 + 7,91)} = 0,016\ 04 = 0,016\ .$$

Для надежности возьмем абсолютную погрешность x на 20 % меньше допустимого значения, то есть примем, что x = 0,013. Чтобы округление числа π не повлияло на точность определения плотности, примем его равным π = 3,141 59. В этом случае его относительная погрешность

$$\frac{\Delta\pi}{\pi} = \frac{3{,}141\ 592\ 654 - 3{,}141\ 59}{3{,}141\ 592\ 654} = 8 \cdot 10^{-7}\ ,$$

что почти на 5 порядков меньше 3x, и она не может повлиять на абсолютную погрешность определения плотности заготовок.

Проверим, что при x = 0,013 определить марку стали можно будет в любом случае со 100%-й вероятностью:

$$\frac{\Delta\rho}{\rho} = \frac{\Delta M}{M} = 2\frac{\Delta D}{D} = \frac{\Delta H}{H} = 3x = 3 \cdot 0,013 = 0,039\ ,$$

$$\Delta\rho_{18} = 0,039\rho_{18} = 0,039 \cdot 8,71 = 0,3397 = 0,34 \text{ г/см}^3\ ,$$

$$\Delta\rho_{45} = 0,039\rho_{45} = 0,039 \cdot 7,94 = 0,3085 = 0,31 \text{ г/см}^3\ .$$

В случае, если x = 0,013, то наихудшем случае

$$(\rho_{18})_{min} = \rho_{18} - \Delta\rho_{18} = 8,71 - 0,34 = 8,37 \text{ г/см}^3\ ,$$

$$(\rho_{45})_{max} = \rho_{45} + \Delta\rho_{45} = 7,91 + 0,31 = 8,22 \text{ г/см}^3\ ,$$

что позволяет со 100%-й вероятностью идентифицировать марку стали, так как в наихудшем случае расчетное значение плотности стали Р18 больше плотности Стали 45.

Чтобы реализовать указанную относительную погрешность измерений, необходимо подобрать соответствующий инструмент.

Номинальные значения диаметра и высоты заготовок: D = 40 мм, H = 20 мм. Номинальные значения массы заготовок, определенные по справочным данным:

$$M_{18} = \frac{\pi\rho_{18} D_{18}^2 H_{18}}{4} = \frac{3{,}141\,59 \cdot 8{,}800 \cdot 4^2 \cdot 2}{4} = 221{,}1679 = 221{,}17 \text{ г},$$

$$M_{45} = \frac{\pi\rho_{45} D_{45}^2 H_{45}}{4} = \frac{3{,}141\,59 \cdot 7{,}826 \cdot 4^2 \cdot 2}{4} = 196{,}6887 = 196{,}69 \text{ г}.$$

Абсолютные погрешности экспериментально определенных величин факторов:

$$\Delta M_{18} \le 0{,}013 \cdot M = 0{,}013 \cdot 221{,}17 = 2{,}8752 = 2{,}9 \text{ г},$$

$$\Delta M_{18} \le 0{,}013 \cdot M = 0{,}013 \cdot 196{,}69 = 2{,}5570 = 2{,}6 \text{ г},$$

$$\Delta D_{18} = \Delta D_{45} \le \frac{0{,}013}{2} \cdot D = \frac{0{,}013}{2} \cdot 40 = 0{,}26 \text{ мм},$$

$$\Delta H_{18} = \Delta H_{45} \le 0{,}013 \cdot H = 0{,}013 \cdot 20 = 0{,}26 \text{ мм}.$$

Проанализировав требуемые абсолютные погрешности для измерения массы, диаметра и высоты заготовок можно дать следующие рекомендации: для измерения массы стальных заготовок следует использовать электронные технические весы с погрешностью определения массы ± 1 г, для измерения диаметра и высоты заготовок следует использовать штангенциркуль с погрешностью определения линейных размеров ± 0,1 мм. Указанные инструменты являются достаточно простыми, относительно дешевыми, и их легко найти на любых промышленных и сельскохозяйственных предприятиях или в торговой сети.

1.2. Вопросы и задания для самопроверки

1. В чем концептуальное отличие детерминированных закономерностей от стохастических?

2. Какова природа погрешности рассчитанного параметра Y для детерминированных моделей?

3. Что такое «абсолютная погрешность»?

4. Что такое «относительная погрешность»?

5. Сформулируйте алгоритм корректного округления рассчитанных параметров для детерминированной модели.

6. Какими уравнениями определяется в общем случае абсолютная погрешность определения параметра Y ($\Delta Y_{\text{пр}}$ – погрешность измерительных приборов, $\Delta Y_{\text{ст}}$ – случайная погрешность, вызванная стохастической природой исследуемого объекта)?

а) $\pm\Delta Y = \pm\Delta Y_{\text{пр}} \pm \Delta Y_{\text{ст}}$; б) $\pm\Delta Y = \Delta Y_{\text{пр}} + \Delta Y_{\text{ст}}$;

в) $\Delta Y = \Delta Y_{\text{пр}} + \Delta Y_{\text{ст}}$; г) $\sqrt{\Delta Y^2} = \sqrt{\Delta Y_{\text{пр}}^2} + \sqrt{\Delta Y_{\text{ст}}^2}$.

7. Факторы x и z в детерминированном эксперименте определены с абсолютной погрешностью Δx и Δz. Если $Y = f(x)\, g(z)$, то:

а) $\dfrac{\Delta Y}{Y} = |f'(x)\, f(x)|\Delta x + |g'(z) g(z)|\Delta z$; б) $\dfrac{\Delta Y}{Y} = \dfrac{f'(x)}{f(x)}\Delta x + \dfrac{g'(z)}{g(z)}\Delta z$;

в) $\dfrac{\Delta Y}{Y} = \left|\dfrac{f(x)}{f'(x)}\right|\Delta x + \left|\dfrac{g(z)}{g'(z)}\right|\Delta z$; г) $\dfrac{\Delta Y}{Y} = \left|\dfrac{f'(x)}{f(x)}\right|\Delta x + \left|\dfrac{g'(z)}{g(z)}\right|\Delta z$.

8. Факторы x и z в детерминированном эксперименте определены с предельной абсолютной погрешностью Δx и Δz. Если $Y = f(x)/g(z)$, то:

а) $\dfrac{\Delta Y}{Y} = \left|\dfrac{f'(x)}{f(x)}\right|\Delta x + \left|\dfrac{g'(z)}{g(z)}\right|\Delta z$; б) $\dfrac{\Delta Y}{Y} = \left|\dfrac{f'(x)}{f(x)}\right|\Delta x - \left|\dfrac{g'(z)}{g(z)}\right|\Delta z$;

в) $\dfrac{\Delta Y}{Y} = \dfrac{f'(x)}{f(x)}\Delta x + \dfrac{g'(z)}{g(z)}\Delta z$; г) $\dfrac{\Delta Y}{Y} = \dfrac{f(x)\Delta x}{x} + \dfrac{g(z)\Delta z}{z}$.

9. В какой формуле округление выполнено корректно?

а) 29,735 ± 0,0778 → 29,74 ± 0,08;

б) 29,745 ± 0,0778 → 29,74 ± 0,08;

в) 29,734 ± 0,0778 → 29,7 ± 0,1;

г) 29,7345 ± 0,0778 → 29,74 ± 0,07

д) 68,535 31 ± 3,596 → 68,53 ± 3,59;

е) 68,535 31 ± 3,596 → 68,54 ± 3,60;

ж) 68,535 31 ± 3,596 → 68,5 ± 3,6;

з) 68,535 31 ± 3,596 → 68,54 ± 3,59.

и) $699{,}5452 \pm 53{,}0561 \to 699{,}545 \pm 53{,}056$;

к) $699{,}5452 \pm 53{,}0561 \to 699{,}5 \pm 53{,}0$;

л) $699{,}5452 \pm 53{,}0561 \to 700 \pm 53$;

м) $699{,}5452 \pm 53{,}0561 \to 700 \pm 50$.

н) $61\,161{,}2353 \pm 97{,}59061 \to 61\,160 \pm 100$;

о) $61\,161{,}2353 \pm 97{,}59061 \to 61\,161 \pm 98$;

п) $61\,161{,}2353 \pm 97{,}59061 \to 61\,161{,}23 \pm 97{,}59$;

р) $61\,161{,}2353 \pm 97{,}59061 \to 61\,160 \pm 90$.

10. Мощность двигателя $W = I \cdot U$, где I – ток, U – напряжение. Рассчитать абсолютную погрешность при определении мощности двигателя, если погрешность определения тока $\Delta I > 0$, погрешность напряжения $\Delta U > 0$.

Ответ: $\Delta W = IU\left(\dfrac{\Delta I}{I} + \dfrac{\Delta U}{U}\right)$.

11. Дано $Y = x^5 e^{2z}$, причем факторы x и z определены с абсолютной погрешностью $\Delta x > 0$ и $\Delta z > 0$. Доказать, что

$$\frac{\Delta Y}{Y} = 5\frac{\Delta x}{|x|} + 2\Delta z\ .$$

12. Дано $Y = \ln(\beta x)\cos(\gamma z)$, причем факторы x и z определены с абсолютной погрешностью $\Delta x > 0$ и $\Delta z > 0$. Доказать, что

$$\frac{\Delta Y}{Y} = \frac{\Delta x}{|x\ln(\beta x)|} + |\gamma\,\mathrm{tg}(\gamma z)|\,\Delta z\ .$$

13. Дано $Y = \cos x \big/ \sqrt[3]{z}$, причем факторы x и z определены с абсолютной погрешностью $\Delta x > 0$ и $\Delta z > 0$. Доказать, что

$$\frac{\Delta Y}{Y} = |\mathrm{tg}\,x|\,\Delta x + \frac{1}{3}\frac{\Delta z}{|z|}\,.$$

14. Вольтметром с относительной погрешностью 3 % измерено напряжение электрической сети 225 В. Омметром с относительной погрешностью 4 % измерено сопротивление электродвигателя

50 Ом. Определить мощность электродвигателя и ее абсолютную погрешность.

Ответ: (1060 ± 110) Вт, или (1,06 ± 0,11) кВт.

15. Автомобиль массой 4,8 т движется равномерно и прямолинейно со скоростью 72 км/ч. Абсолютная погрешность определения массы автомобиля – 150 кг, скорости – 1,5 м/с. Определить кинетическую энергию автомобиля и абсолютную погрешность.

Ответ: (960 000 ± 170 000) Дж, или (0,96 ± 0,17) МДж.

16. Дано $Y = f(x)\, g(z)\, h(w)$, причем факторы x, z, и wв детерминированном эксперименте определены с абсолютной погрешностью $\Delta x > 0$, $\Delta z > 0$, $\Delta w > 0$. Доказать, что

$$\frac{\Delta Y}{Y} = \left|\frac{f'(x)}{f(x)}\right| \Delta x + \left|\frac{g'(z)}{g(z)}\right| \Delta z + \left|\frac{h'(w)}{h(w)}\right| \Delta w\ .$$

2. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ СТОХАСТИЧЕСКОГО ЭКСПЕРИМЕНТА МЕТОДАМИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В результате изучения темы студенты должны

знать:

– нормальный закон распределения и его параметры;

– концептуальное различие между генеральной совокупностью и выборкой;

– выборочный метод определения параметров генеральной совокупности;

– выборочный метод оценки качества серийной продукции;

– выборочный метод сравнения однотипной серийной продукции по качеству;

уметь:

– проводить анализ случайных значений выборки на промах по критерию Смирнова-Граббса и принадлежность их к нормальному распределению по критерию Гири;

– рассчитывать по выборочным данным доверительные интервалы параметров генеральной совокупности по критериям Стьюдента и Пирсона;

– проводить анализ данных двух выборок на однородность выборочных дисперсий по критерию Фишера и на существенное различие выборочных средних по критерию Стьюдента.

2.1. Нормальный закон распределения и его значение для моделирования технических систем

Техническая система, описываемая детерминированной моделью, при заданном значении управляющего фактора всегда характеризуется единственно возможным значением измеряемого параметра со 100%-й вероятностью (в пределах погрешности измерительного инструмента). В отличие от этого, измеряемый параметр технической системы, описываемой стохастической моделью, характеризуется при фиксированном значении управляющего фактора множеством значений измеряемого параметра, причем каждое из них может иметь разную вероятность. Следует отметить, что одна и та же техническая система может моделироваться или детерминированными, или стохастическими моделями в зависимости от того, какой параметр этой системы изучается. Приведенное утверждение проиллюстрируем на примерах экспериментов с монетой.

Первый эксперимент состоит в определении массы монеты путем взвешивания на весах. В данном случае изучаемый параметр – масса монеты. Пусть масса монеты (г) – ($m \pm \Delta m$), где Δm – погрешность используемых весов при условии $\Delta m << m$. Сколько бы опытов не проводилось, в пределах погрешности наблюдаем одно и то же значение массы, так как никакие случайные факторы в условиях данного эксперимента не могут повлиять на результат измерения. Результаты этого эксперимента позволяют утверждать со 100%-й вероятностью, что масса монеты (г) равна ($m \pm \Delta m$), следовательно, состояние объекта в данном примере описывается детерминированной моделью.

Во втором эксперименте будем бросать ту же монету с некоторого расстояния на прямую, проведенную на земле. Изучаемый параметр – расстояние центра монеты от линии. Отклонение центра монеты от линии при перелете будем считать положительным, при недолете – отрицательным. Результат каждого опыта в этом эксперименте непредсказуем, и потому состояние объекта в этом эксперименте описывается стохастической моделью.

В общем случае для объектов, описываемых стохастическими моделями, результат опыта предсказывается только с некой вероятностью, характеризующей состояние объекта в условиях данного эксперимента. Во втором примере, из общих соображений, можно ожидать, что расстояния до прямой, близкие к нулю, будут наиболее вероятны, а большие расстояния – менее вероятны.

2.1.1. Генеральная совокупность, нормальный закон распределения

Состояние технических систем характеризуется параметрами оптимизации, зависящими от управляющих факторов. В отличие от технических систем, описываемых детерминированными моделями, состояние технических систем, описываемых стохастическими моделями, характеризуется значениями параметров оптимизации, которые реализуются с той или иной вероятностью. Математически отражение вероятностного поведения параметра оптимизации задается функцией плотности вероятности.

В технических системах, изучаемых в науке и технике, распределение параметров достаточно часто подчиняется нормальному закону распределения (распределение Гаусса). Функция плотности вероятности с таким распределением имеет колоколообразный вид (рис. 1).

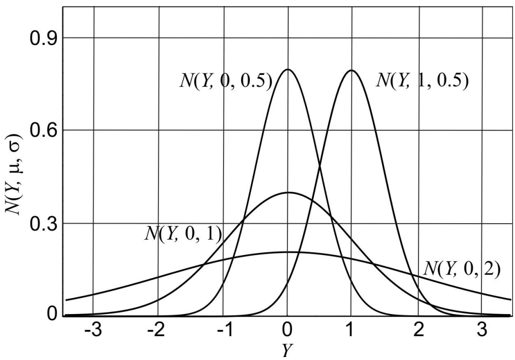


Рис. 1. Графики функции плотности вероятности нормального распределения

Теоретическая функция плотности вероятности нормального распределения однозначно задается двумя параметрами μ и σ:

$$N(Y,\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(Y-\mu)^2}{2\sigma^2}}, \qquad (24)$$

где Y – изучаемый параметр оптимизации; μ – генеральное среднее; σ – генеральное стандартное отклонение; σ^2 – генеральная дисперсия.

Физический смысл функции плотности вероятности заключается в том, что произведение $N(Y, \mu, \sigma)\Delta Y$ равно числу состояний случайного параметра в интервале $[Y; Y + \Delta Y]$.

Число состояний технической системы, выражаемое полным множеством случайных значений параметра оптимизации при всех возможных комбинациях детерминированных факторов, является теоретически бесконечной величиной, распределение которой описывается уравнением (24).

Положение вершины кривой, описываемой уравнением (24), относительно начала координат определяется параметром μ (генеральное среднее), рассеяние случайных значений параметра Y относительно μ определяется параметром σ (генеральное стандартное отклонение) (см. рис. 1). Параметры Y, μ, σ имеют одинаковую размерность. Нормальный закон распределения обозначают следующим образом: $N(Y, 0, 1)$, где N – нормальный закон распределения случайной величины, Y –случайная величина, распределение которой описывается нормальным законом с параметрами: генеральное среднее $\mu = 0$ и генеральное стандартное отклонение $\sigma = 1$.

Нормальный закон распределения случайных величин имеет фундаментальное значение для решения широкого круга народнохозяйственных задач, поскольку адекватно описывает распределение случайных отклонений исследуемого параметра, обусловленных большим количеством мало влияющих случайных независимых факторов. Он нашел широкое распространение при моделировании технических систем в АПК с использованием методов математической статистики.

В пособии ограничимся стохастическим моделированием технических систем, для которых случайные значения исследуемых параметров подчиняются только нормальному закону распределения.

Уравнение (24) позволяет рассчитать вероятность попадания случайных значений параметра Y в некоторый интервал. Для приведенных ниже интервалов доля всех состояний случайных нормально распределенных величин Y равна:

$$p\,(-\sigma \le Y \le \sigma) = 0{,}6827\,, \tag{25}$$

$$p(-2\sigma \le Y \le 2\sigma) = 0{,}9545\,, \tag{26}$$

$$p(-3\sigma \le Y \le 3\sigma) = 0{,}9973\,. \tag{27}$$

Таким образом, практически все значения параметра Y, подчиняющиеся нормальному закону распределения, находятся в интервале $\mu \pm 3\sigma$ (вероятность этого утверждения равна 0,9973). Однако для анализа технических систем используют менее жесткое утверждение: принято считать, что все случайные значения параметра Y, подчиняющиеся нормальному закону распределения, находятся в интервале $\mu \pm 2\sigma$, так как вероятность этого утверждения равна 0,9545. Остальные 5 % в заданный интервал не попадают.

Требование к доле случайных значений, которые не попадают в заданный интервал, для технических систем различное. Если функционирование технической системы связано с большим риском для жизни людей, то целесообразно, чтобы вероятность желаемого события составляла не менее 0,999 99 (99,999 %). Однако создание систем с таким высоким уровнем надежности требует больших материальных затрат. В тех случаях, когда материальные ресурсы ограничены, а возможный ущерб от неисполнения желаемого события невелик, то целесообразно ограничиться вероятностью 0,95 (95 %), что характерно для большинства технических систем АПК. Поэтому в дальнейшем при статистической обработке экспериментальных данных будет использоваться доверительная вероятность $p = 0{,}95$.

Чтобы экспериментально определить генеральные параметры μ и σ, необходимо количественно измерить все случайные значения бесконечно большой генеральной совокупности, затем рассчитать параметры μ и σ (по уравнениям (28)–(31)). Очевидно, что бесконечное количество измерений невозможно реализовать на практике. Но концептуальная заслуга математической статистики именно в том, что ее методы позволяют оценивать генеральные параметры μ и σ, используя для этого конечное, относительно небольшое количество случайных величин, принадлежащих генеральной совокупности, называемой выборкой. Однако, «платой» за такую «экономию» будет некоторая погрешность в определении генеральных параметров μ и σ, причем тем бóльшая, чем меньше объем выборки.

На практике под генеральной совокупностью, состоящей из бесконечного количества значений параметра, подразумевается генеральная совокупность из относительно большого, но конечного количества значений (например, генеральная совокупность цыплят на птицеферме насчитывает десятки тысяч особей).

2.1.2. Выборка и определение ее параметров

Из определения генеральной совокупности следует, что определить ее параметры μ и σ практически невозможно. Однако значения этих параметров необходимы для описания функционирования технических систем, и поэтому в математической статистике разработан метод оценки генеральных параметров по относительно небольшому количеству значений. Название этого метода – выборочный метод оценки генеральных параметров.

Расчет выборочных параметров. Рассмотрим некоторую выборку случайных величин $Y_1,\ldots,\ Y_n$, принадлежащую генеральной совокупности с параметрами μ и σ. Оценка параметров μ и σ производится на основе выборки объемом n: по выборочному среднему $\overline{Y}$, выборочной дисперсии S^2 или выборочному стандартному отклонению S, числу степеней свободы выборочной дисперсии f_S :

$$\overline{Y} = \frac{Y_1 + \ldots + Y_n}{n} = \frac{\sum_{i=1}^{n} Y_i}{n}, \tag{28}$$

$$S^2 = \frac{(Y_1 - \overline{Y})^2 + \ldots + (Y_n - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}, \tag{29}$$

$$S = \sqrt{S^2}, \tag{30}$$

$$f_S = n - 1. \tag{31}$$

В общем случае число степеней свободы рассчитанного выборочного параметра равно объему выборки за вычетом числа параметров B, рассчитанных предварительно по этой же выборке, то есть $f = n - B$. Например, в случае расчета выборочной дисперсии по той же выборке было рассчитано значение выборочного среднего, поэтому $f_S = n - 1$.

Проверка случайных значений выборки на промах по критерию Смирнова–Граббса. Среди случайных значений выборки может содержаться грубая ошибка (промах), который возникает при ошибочном действии экспериментатора или обусловлен каким-нибудь непредсказуемым внешним воздействием. Очевидно, что промах искажает выборочные параметры $\overline{Y}$, S, которые искажают оценку генеральных параметров μ и σ. Поэтому, прежде чем приступить к оценке параметров генеральной совокупности выборочным методом, следует провести статистический анализ экспериментальных данных, направленный на исключение промаха.

Алгоритм проверки случайных значений выборки на промах с помощью критерия Смирнова–Граббса включает следующие операции:

– определение кандидата $Y_{\text{канд}}$. Кандидат – случайное значение выборки, которое находится на максимальном расстоянии от выборочного среднего, то есть удовлетворяет следующему условию:

$$\left|Y_{\text{канд}} - \overline{Y}\right| = \max; \tag{32}$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{\text{э}} = \frac{\left|Y_{\text{канд}} - \overline{Y}\right|}{S}; \tag{33}$$

– определение критического значение критерия Смирнова–Граббса $\tau_{f_\tau;\, p}$, где $f_\tau = n - 2$ – число степеней свободы предварительно были рассчитаны два параметра: $\overline{Y}$, S) при доверительной вероятности $p = 0{,}95$ (прилож. 1),

– принятие решения:

– $Y_{\text{канд}}$ не является промахом, если

$$\tau_{\text{э}} < \tau_{n-2;\ 0{,}95}, \tag{34}$$

в этом случае его следует оставить в выборке;

– $Y_{\text{канд}}$ является промахом, если

$$\tau_{\text{э}} \geq \tau_{n-2;\ 0{,}95}, \tag{35}$$

в этом случае его следует исключить из выборки.

После исключения промаха следует убедиться, что в оставшейся выборке нет промаха. В противном случае методы математической статистики, излагаемые далее, неприменимы к выборкам, которые имеют два и более промаха. Следует отметить, что критерий Смирнова-Граббса имеет ограниченное применение, так как с его помощью выборки на промах можно проверить случайные значения, объем которой $n \geq 6$ при доверительной вероятности 0,95.

Проверка случайных значений выборки на принадлежность их к нормальному закону распределения по критерию Гири. После проверки выборки на промах и исключения его, если он был обнаружен, необходимо проверить гипотезу о принадлежности случайных значений выборки к нормальному закону распределения. Это обусловлено тем, что математический аппарат, применяемый в пособии, основан на предположении принадлежности выборки к нормальному закону распределения.

Алгоритм проверки выборки на принадлежность к нормальному закону с помощью критерия Гири включает следующие операции:

– расчет среднего абсолютного отклонения (САО):

$$\text{САО} = \frac{\sum_{i=1}^{n} \left| Y_i - \bar{Y} \right|}{n}; \tag{36}$$

– расчет экспериментального значения критерия Гири:

$$\theta_{э} = \left| \sqrt{\frac{n}{n-1}} \cdot \frac{\text{САО}}{S} - 0,7979 \right|; \tag{37}$$

– определение критического значения критерия Гири $\theta_{f_\theta; p}$ при числе степеней свободы $f_\theta = n$ и доверительной вероятности p = 0,95:

$$\theta_{f_\theta; p} = \frac{0,4}{\sqrt{n}}; \tag{38}$$

– принятие решения:

если

$$\theta_{э} < \theta_{n; 0,95}, \tag{39}$$

то случайные значения выборки подчиняются нормальному закону распределения,

если

$$\theta_{э} \geq \theta_{n; 0,95}, \tag{40}$$

то случайные значения выборки не подчиняются нормальному закону распределения.

Выборку, которая не содержит промаха и подчиняется нормальному закону распределения, будем называть «очищенной» (далее математические формулы справедливы только для «очищенных» выборок).

Следует отметить, что критерий Гири имеет ограниченное применение, так как с его помощью можно проверить принадлежность случайных значений выборки к нормальному закону распределения, объем которой $n \geq 8$ при доверительной вероятности 0,95.

2.2. Выборочный метод оценки параметров генеральной совокупности

Убедившись, что выборка является «очищенной», выполним на ее основе оценку параметров генеральной совокупности. Так как выборка имеет конечный объем, то параметры генеральной совокупности μ и σ оцениваются с некоторой погрешностью, определяемой посредством доверительных интервалов, и чем больше объем выборки, тем точнее будут определены параметры генеральной совокупности.

Доверительный интервал для генерального среднего μ определяется по критерию Стьюдента:

$$\begin{cases} \bar{Y} - \Delta\bar{Y} < \mu < \bar{Y} + \Delta\bar{Y}, \\ \Delta\bar{Y} = \dfrac{t_{f_s;1-p}\,S}{\sqrt{n}} \end{cases}, \tag{41}$$

где $t_{f_S;1-p}$ – критическое значение критерия Стьюдента при числе степеней свободы $f_S = n - 1$ и доверительной вероятности $p = 0{,}95$ (прилож. 2).

Доверительный интервал для генеральной дисперсии σ^2 и генерального стандартного отклонения определяется по критерию Пирсона:

$$\begin{cases} \dfrac{(n-1)S^2}{\chi^2_{f_S;(1+p)/2}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{f_S;(1-p)/2}}, \\ \sqrt{\dfrac{(n-1)S^2}{\chi^2_{f_S;(1+p)/2}}} < \sigma < \sqrt{\dfrac{(n-1)S^2}{\chi^2_{f_S;(1-p)/2}}}, \end{cases} \tag{42}$$

где $\chi^2_{f_S;(1+p)/2}$, $\chi^2_{f_S;(1-p)/2}$ – верхнее и нижнее критические значения критерия Пирсона при числе степеней свободы $f_S = n - 1$ и доверительной вероятности $p = 0{,}95$ (прилож. 3).

2.2.1. Алгоритм решения прикладных задач выборочным методом

При серийном производстве 100%-й контроль качества продукции может быть невозможен по материальным или временным причинам. Особенно это актуально, если используются разрушающие методы контроля. Поэтому на практике часто используют выборочный контроль. Для этого из партии изделий определенного размера, предназначенной для отправки потребителю, случайным образом отбирают относительно небольшое количество изделий для контроля, например: если партия, подготовленная к отправке, состоит из 1000 штук, контроль производится, например, по 20 штукам. На основе статистической обработки результатов выборочного контроля требуемого параметра составляется сертификат качества отправляемой потребителю партии изделий.

Алгоритм выборочного контроля выборки $Y_1, \ldots, Y_i, \ldots, Y_n$ объемом n включает следующие операции.

1. Расчет выборочных параметров:

– выборочное среднее:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n};$$

– выборочная дисперсия, выборочное стандартное отклонение и число степеней свободы выборочной дисперсии:

$$S^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1},$$

$$S = \sqrt{S^2},$$

$$f_S = n - 1.$$

2. Проверка случайных значений выборки на промах по критерию Смирнова–Граббса:

– определение $Y_{\text{канд}}$, который находится от выборочного среднего на максимальном расстоянии, то есть удовлетворяет следующему условию:

$$\left|Y_{\text{канд}} - \bar{Y}\right| = \max;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{э} = \frac{\left|Y_{\text{канд}} - \bar{Y}\right|}{S},$$

– определение критического значения критерия Смирнова–Граббса $\tau_{f_\tau;\, p}$, где $f_t = n - 2$ – число степеней свободы при доверительной вероятности $p = 0{,}95$ (прилож. 1);

– принятие решения о наличии промаха:

если $\tau_{э} < \tau_{n-2;\; 0,95}$, то $Y_{\text{канд}}$ не является промахом, в этом случае его следует оставить в выборке,

если $\tau_{э} \geq \tau_{n-2;\; 0,95}$, то $Y_{\text{канд}}$ является промахом, в этом случае его следует исключить из выборки.

После исключения промаха необходимо убедиться, что в оставшейся выборке нет промаха. В противном случае для такой выборки неприменимы формулы и методы, излагаемые в пп. 3–5.

3. Проверка случайных значений выборки на принадлежность их к нормальному закону распределения по критерию Гири:

– расчет среднего абсолютного отклонения:

$$\text{САО} = \frac{\sum_{i=1}^{n}\left|Y_i - \bar{Y}\right|}{n};$$

– расчет экспериментального значения критерия Гири:

$$\theta_{э} = \left|\sqrt{\frac{n}{n-1}} \cdot \frac{\text{САО}}{S} - 0{,}7979\right|;$$

– определение критического значения критерия Гири $\theta_{f_\theta;\, p}$ при числе степеней свободы $f_\theta = n$ и доверительной вероятности $p = 0{,}95$:

$$\theta_{f_\theta;\, p} = \frac{0{,}4}{\sqrt{n}};$$

– принятие решения о принадлежности случайных значений выборки к нормальному закону распределения:

если $\theta_э < \theta_{n;\,p}$, то выборка подчиняется нормальному закону распределения,

если $\theta_э \geq \theta_{n;\,p}$, то выборка не подчиняется нормальному закону распределения.

Выборка, которая не содержит промаха и подчиняется нормальному закону распределения, называется «очищенной». В пособии математические формулы справедливы только для «очищенных» выборок.

4. Расчет доверительного интервала генерального среднего μ по критерию Стьюдента:

$$\begin{cases} \bar{Y} - \Delta\bar{Y} < \mu < \bar{Y} + \Delta\bar{Y}, \\ \Delta\bar{Y} = \dfrac{t_{f_s;1-p}\,S}{\sqrt{n}} \end{cases},$$

где $t_{f_s;1-p}$ – критическое значение критерия Стьюдента при числе степеней свободы $f_S = n - 1$ и доверительной вероятности $p = 0{,}95$ (прилож. 2).

5. Расчет доверительного интервала генеральной дисперсии σ^2 по критерию Пирсона:

$$\begin{cases} \dfrac{(n-1)S^2}{\chi^2_{f_S;(1+p)/2}} < \sigma^2 < \dfrac{(n-1)S^2}{\chi^2_{f_S;(1-p)/2}}, \\ \sqrt{\dfrac{(n-1)S^2}{\chi^2_{f_S;(1+p)/2}}} < \sigma < \sqrt{\dfrac{(n-1)S^2}{\chi^2_{f_S;(1-p)/2}}}, \end{cases}$$

где $\chi^2_{f_S;(1+p)/2}$, $\chi^2_{f_S;(1-p)/2}$ – верхнее и нижнее критические значения критерия Пирсона при числе степеней свободы $f_S = n - 1$ и доверительной вероятности $p = 0{,}95$ (прилож. 3).

Так как на основе анализа выборки невозможно точно определить значения генерального среднего μ и генерального стандартного отклонения σ, то заключение о качестве партии изделий может быть сделано на основе доверительных интервалов их значений.

2.2.2. Типовая задача: определение качества серийной продукции выборочным методом

Цель – определение качества серийной продукции выборочным методом.

Поясним методику решения типовой задачи по описанному алгоритму на следующем примере: в вольерах птицефабрики содержится приблизительно 50 000 цыплят одного возраста. Требуется оценить суммарный вес цыплят с предельной относительной погрешностью 1 %.

Сформулируем задачу в терминах математической статистики. Генеральная совокупность состоит из ≈ 50 000 цыплят. Необходимо определить генеральные параметры μ и σ.

Определение массы цыплят – операция не затратная, поэтому объем исходной выборки возьмем побольше: $n_1 = 12$. Экспериментальные данные: 565, 860, 882, 893, 893, 902, 909, 909, 923, 937, 945, 998.

Обработка данных исходной выборки, $n_1 = 12$

Напоминаем, что все предварительные результаты расчета должны быть выполнены с точностью не менее четырех значащих цифр.

1. Расчет выборочных параметров исходной выборки:

– выборочное среднее:

$$\bar{m}_1 = \frac{\sum_{i=1}^{n_1} m_{1i}}{n_1} = \frac{\sum_{i=1}^{12} m_{1i}}{12} =$$

$$= \frac{565+860+882+893+893+902+909+909+923+937+945+998}{12} = 884,7\,;$$

– выборочная дисперсия, стандартное отклонение и число степеней свободы выборочной дисперсии:

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (m_{1i} - \bar{m}_1)^2}{n_1 - 1} = \frac{\sum_{i=1}^{12} (m_{1i} - \bar{m}_1)^2}{12 - 1} =$$

$$= \frac{(565-884,7)^2 + (860-884,7)^2 + (882-884,7)^2 + (893-884,7)^2 + (893-884,7)^2}{11} +$$

$$+ \frac{(902-884,7)^2 + (909-884,7)^2 + (909-884,7)^2 + (923-884,7)^2 + (937-884,7)^2}{11} +$$

$$+\frac{(945-884,7)^2+(988-884,7)^2}{11}=11\,374;$$

$$S_1=\sqrt{S_1^2}=\sqrt{11\,374}=106,7;$$

$$f_S = n_1 - 1 = 12 - 1 = 11.$$

2. Проверка случайных значений исходной выборки на промах:

– определение кандидата $m_{1\text{канд}}$ на промах, которым может быть минимальное ($m_{1,1} = 565$) или максимальное ($m_{1,12} = 998$) значение выборки. Так как

$$|565 - 884,7| > |998 - 884,7|,$$

то кандидатом на промах является значение $m_{1\text{канд}} = 565$;

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{э_1}=\frac{\left|m_{1\text{канд}}-\overline{m}_1\right|}{S_1}=\frac{\left|565-884,7\right|}{106,7}=2,997;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{f_{\tau_1};\,0,95} = 2,387$ при числе степеней свободы $f_{\tau_1} = n_1 - 2 = 10$, при доверительной вероятности 0,95 (прилож. 1);

– так как экспериментальное значение критерия Смирнова–Граббса больше критического: $\tau_{э_1} = 2,997 > \tau_{10;\,0,95} = 2,387$, то в исходной выборке значение $m_{1\text{канд}} = 565$ является промахом.

После исключения выявленного промаха из исходной выборки, проверим на наличие промаха оставшиеся случайные значения обновленной выборки 860, 882, 893, 893, 902, 909, 909, 923, 937, 945, 998 объемом $n_2 = 11$.

Обработка данных обновленной выборки, $n_1 = 11$

1. Расчет выборочных параметров обновленной выборки:

– выборочное среднее:

$$\overline{m}_2=\frac{\sum_{i=1}^{n_2}m_{2i}}{n_2}=\frac{\sum_{i=1}^{11}m_{2i}}{11}=$$

$$=\frac{860+882+893+893+902+909+909+923+937+945+998}{11}=913,7;$$

– выборочная дисперсия, стандартное отклонение и число степеней свободы выборочной дисперсии:

$$S_2^2=\frac{\sum_{i=1}^{n_2}(m_{2i}-\bar{m}_2)^2}{n_2-1}=\frac{\sum_{i=1}^{11}(m_{2i}-\bar{m}_2)^2}{11-1}=$$

$$=\frac{(860-913,7)^2+(882-913,7)^2+(893-913,7)^2+(893-913,7)^2}{10}+$$

$$+\frac{(902-913,7)^2+(909-913,7)^2+(909-913,7)^2+(923-913,7)^2}{10}+$$

$$+\frac{(937-913,7)^2+(945-913,7)^2+(1215-913,7)^2}{10}=1364,$$

$$S_2=\sqrt{S_2^2}=\sqrt{1364}=36,94,$$

$$f_S = n_2 - 1 = 12 - 1 = 10.$$

2. Проверка случайных значений обновленной выборки на промах:

– определение кандидата $m_{2\text{канд}}$ на промах, которым может быть минимальное ($m_{2,1}$ = 860) или максимальное ($m_{2,11}$ = 998) значение выборки. Так как

$$|860 - 913,7| < |998 - 913,7|,$$

то кандидатом на промах является значение $m_{2\text{канд}}$ = 998;

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{э_2}=\frac{\left|m_{2\,\text{канд}}-\bar{m}_2\right|}{S_2}=\frac{\left|998-913,7\right|}{36,94}=2,282;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{f_{\tau_2};\,0,95}$ = 2,343 при числе степеней свободы $f_{\tau_2} = n_2 - 2$ = 11 – 2 = 9 при доверительной вероятности 0,95 (прилож. 1);

– так как экспериментальное значение критерия Смирнова–Граббса меньше критического: $\tau_{э_2} = 2{,}282 < \tau_{9,\,0{,}95} = 2{,}343$, то в обновленной выборке промаха нет, то есть значение $m_{2канд} = 998$ не является промахом.

3. Проверка случайных значений обновленной выборки m_{2i} на принадлежность их к нормальному закону распределения по критерию Гири:

– расчет среднего абсолютного отклонения:

$$\text{САО}_2 = \frac{\sum_{i=1}^{n_2} \left| m_{2i} - \bar{m}_2 \right|}{n_2} = \frac{\sum_{i=1}^{11} \left| m_{2i} - \bar{m}_2 \right|}{11} =$$

$$= \frac{|860-913{,}7| + |882-913{,}7| + |893-913{,}7| + |893-913{,}7| + |902-913{,}7|}{11} +$$

$$+ \frac{|909-913{,}7| + |909-913{,}7| + |923-913{,}7| + |937-913{,}7|}{11} +$$

$$+ \frac{|923-913{,}7| + |937-913{,}7| + |945-913{,}7| + |998-913{,}7|}{11} = 26{,}93;$$

– расчет экспериментального значения критерия Гири:

$$\theta_{э_2} = \left| \frac{\text{САО}_2}{S_2} \sqrt{\frac{n_2}{n_2 - 1}} - 0{,}7979 \right| = \left| \frac{26{,}93}{36{,}94} - 0{,}7979 \right| = 0{,}033\,32;$$

– определение критического значения критерия Гири $\theta_{f_\theta;\,p}$ при числе степеней свободы $f_\theta = n_2$ и доверительной вероятности $p = 0{,}95$:

$$\theta_{n_2:\,0{,}95} = \frac{0{,}4}{\sqrt{n_2}} = \frac{0{,}4}{\sqrt{11}} = 0{,}1206;$$

– так как экспериментальное значения критерия Гири меньше критического: $\theta_{э_2} = 0{,}03332 < \theta_{11;\,0{,}95} = 0{,}1206$, то случайные значения обновленной выборки подчиняются нормальному закону распределения. Таким образом, обновленная выборка является «очищенной».

4. Расчет параметров генеральной совокупности:

– абсолютная погрешность оценки выборочного среднего $\overline{m}_2$:

$$\Delta\overline{m}_2 = \frac{t_{n_2-1;\,0,95}\,S_2}{\sqrt{n_2}} = \frac{t_{10;\,0,95}\,S_2}{\sqrt{n_2}} = \frac{2,228\cdot 36,94}{\sqrt{11}} = 24,81 \approx 25\,,$$

где $t_{10;\,0,95} = 2{,}228$ – критическое значение критерия Стьюдента при числе степеней свободы $f_{S_2} = n_2 - 1 = 10$ и доверительной вероятности 0,95 (прилож. 2).

Рассчитанную абсолютную погрешность $\Delta\overline{m}_2 = 24{,}81$ округлим до двух значащих цифр («легкая» погрешность), то есть $\overline{m}_2 = 25$;

– доверительный интервал генерального среднего μ:

$$\overline{m}_2 - \Delta\overline{m}_2 < \mu < \overline{m}_2 + \Delta\overline{m}_2\,,$$

$$914 - 25 < \mu < 914 + 25\,,$$

$$889 < \mu < 939\,,$$

так как при расчете доверительного интервала выборочное среднее $\overline{m}_2 = 913,7 \approx 914$ округляется до значащих цифр доверительного интервала;

– относительная погрешность определения выборочного среднего:

$$\frac{\Delta\overline{m}_2}{\overline{m}_2} = \frac{25}{914} = 0,027\,35 \approx 0,027\ \ (2,7\,\%)\,;$$

– доверительный интервал генерального стандартного отклонения:

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;\,(1+p)/2}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1;\,(1-p)/2}}}\,,$$

$$\sqrt{\frac{10\cdot 1364}{20,483}} < \sigma < \sqrt{\frac{10\cdot 1394}{3,247}},$$

$$26 < \sigma < 65,$$

где $\chi^2_{10;\,0,975} = 20,483, \chi^2_{10;\,0,025} = 3,247$ – критические значения критерия Пирсона при числе степеней свободы $f_{S_2} = n_2 - 1 = 10$ и доверительной вероятности 0,95 (прилож. 3).

Рассчитанные параметры генеральной совокупности позволяют решить целый ряд прикладных задач, в частности – по составлению бизнес-плана. Например, в качестве одного из пунктов бизнес-плана спрогнозируем выручку птицефермы при продаже 50 000 цыплят, с учетом того, что наиболее вероятным значением генерального среднего является выборочное среднее $\mu \approx \overline{m}_2$ а наиболее вероятным значением генерального стандартного отклонения является выборочное стандартное отклонение $\sigma \approx S_2$.

Исходные данные для прогнозирования: стоимость цыплят высшего сорта – $P_{\text{в}}$= 3,40 руб./кг, первого сорта – P_1= 2,80 руб./кг, второго сорта – P_2 = 2,40 руб./кг. Анализируя результаты типовой задачи, можно сделать вывод, что наиболее вероятные параметры генеральной совокупности, соответственно, равны: $\mu \approx \overline{m}_2 = 0,914$ кг; $\sigma \approx S_2 = 0,037$ кг.

Решение задачи. Договор с организацией об оптовой закупке предусматривает, что к высшему сорту относят цыплят, масса которых находится в интервале 0,88–0,95 кг; к первому сорту – цыплят, масса которых не выходит за интервал высшего сорта более, чем на 0,04 кг; ко второму сорту – цыплят, масса которых выходит за интервал первого сорта. Сопоставление приведенных интервалов с полученными выборочными параметрами распределения цыплят по массе показывает, что границы разделения по сортам близки к рассчитываемым по формулам (25)–(27): $0,877 = \mu - \sigma \le M_{\text{вс}} \le \mu + \sigma = 0,951$ кг, $0,840 = \mu - 2\sigma \le M_1 \le \mu - \sigma = 0,877$ кг и $0,951 = \mu + \sigma \le M_1 \le \mu + 2\sigma =$ $= 0,988$ кг (рис. 2). Тогда ожидаемая выручка

$$V = \left[\,0,6834P_{\text{в}} + (0,95450 - 0,6834)P_1 + (1 - 0,9545)P_2\,\right]N =$$

$$= \left(0,6834\cdot 3,40 + 0,2711\cdot 2,80 + 0,0455\cdot 2,40\right)\cdot 5\cdot 10^4 \approx 160 \text{ тыс.руб.}$$

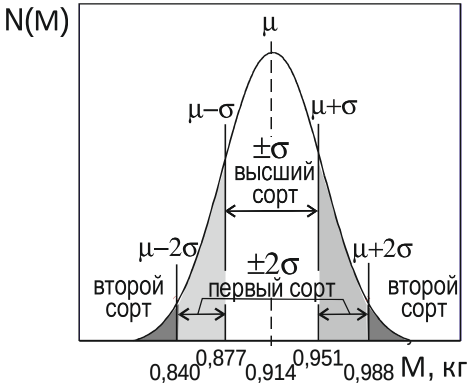


Рис. 2. График распределения продукции по качеству

Таким образом, определение доверительных интервалов генеральных параметров серийной продукции выборочным методом позволяет обосновать некоторые показатели бизнес-плана.

2.3. Сравнительная оценка параметров двух выборок

Итак, рассмотрена теория выборочного метода оценки параметров генеральной совокупности и решение на ее основе важной народнохозяйственной задачи – оценки качества больших партий серийной продукции (десятки и сотни тысяч изделий) на основе результатов контроля относительно небольшого количества изделий (10–20 шт.). Приведен математический аппарат обработки экспериментальных данных методами математической статистики, с помощью которого можно определить, является ли изучаемая выборка «очищенной».

В промышленности и сельском хозяйстве существует большое количество важных проблем, которые могут быть решены методами математической статистики. В частности, большое практическое значение имеет решение задачи сравнения серийной однотипной продукции по качеству. Под однотипной продукцией понимается продукция одного функционального назначения, качество которой характеризуется количественным критерием. Несмотря на то, что продукция является однотипной, она может производиться по разным технологиям.

Рассмотрим однотипную продукцию двух партий, характеризующихся следующими выборочными параметрами ($\overline{Y}_1$, S_1^2, n_1) и ($\overline{Y}_2$, S_2^2, n_2). Понятие однотипной продукции следует дополнить критерием, который определяется соотношением выборочных дисперсий S_1^2 и S_2^2. Практически всегда $S_1^2 \neq S_2^2$. При одном соотношении S_1^2 и S_2^2 рассмотренные партии являются однотипными, при другом соотношении S_1^2 и S_2^2 – не являются однотипными.

Рассмотрим две партии валов, имеющих различные доверительные интервалы диаметров (50 ± 0,010) мм и (50 ± 0,100) мм. Если критерием качества является разброс диаметров валов не более ± 0,015 мм, то эти партии не являются однотипными, так как вторая партия не может быть использована в производстве, в котором используется первая партия деталей. Доверительный интервал определяется выборочными дисперсиями S_1^2 и S_2^2. Очевидно, что в данном примере S_2^2 существенно больше, чем S_1^2. Поэтому такую продукцию нельзя считать однотипной. Если бы доверительные интервалы диаметров валов равнялись бы (50 ± 0,010) мм и (50 ± 0,015) мм, то являлась бы такая продукция однотипной? Для ответа на этот вопрос в математической статистике введено понятие «однородность выборочных дисперсий».

Так как практически всегда $S_1^2 \neq S_2^2$, то возможны два случая: 1) обе выборки относятся к генеральным совокупностям с одинаковой генеральной дисперсией $\sigma_1 = \sigma_2$, и в этом случае выборочные дисперсии S_1^2 и S_2^2 являются однородными, и сравниваемая продукция является однотипной; 2) обе выборки относятся к генеральным совокупностям с различными генеральными дисперсиями $\sigma_1 \neq \sigma_2$; в этом случае выборочные дисперсии S_1^2 и S_2^2 не являются однородными, и сравниваемая продукция не является однотипной.

Рассмотрим задачу: на некотором предприятии по производству льноволокна имеются 2 партии однотипной продукции (одинаковое функциональное назначение, однородные выборочные дисперсии). Является ли их различие по качеству существенным?

Критерий качества – прочность льноволокна. Технологический процесс его производства включает несколько операций. Технолог предлагает внести изменения в две операции, которые, как он считает, должны повысить прочность льноволокна. Для проверки этого утверждения выпущена по модернизированной технологии партия льноволокна. Последующий выборочный контроль показал, что среднее значение прочности льноволокна, выпущенного по новой технологии, оказалось выше.

Возникает вопрос – действительно ли «новое» льноволокно прочнее «старого»? В возможны два варианта:

1) «новое» льноволокно действительно существенно прочнее «старого», и поэтому все последующие выборки будут показывать среднюю прочность «нового» волокна выше «старого»;

2) «новое» льноволокно по прочности такое же, как и «старое», несмотря на то, что прочность «нового» волокна в контрольной выборке больше прочности «старого». Это различие случайно, так как все последующие результаты выборочного контроля будут показывать среднюю прочность «нового» волокна то выше, то ниже «старого».

Обобщая оба примера, рассматривающих качество однотипной продукции, характеризуемое выборочными параметрами ($\overline{Y}_1$, S_1^2, n_1) и ($\overline{Y}_2$, S_2^2, n_2), укажем, что обе выборки являются частью некоторых генеральных совокупностей. При этом возможны 4 случая: 1) $\mu_1 = \mu_2$ и $\sigma_1^2 = \sigma_2^2$; 2) $\mu_1 \neq \mu_2$ и $\sigma_1^2 = \sigma_2^2$; 3) $\mu_1 = \mu_2$ и $\sigma_1^2 \neq \sigma_2^2$; 4) $\mu_1 \neq \mu_2$ и $\sigma_1^2 \neq \sigma_2^2$.

Первые два случая будут использованы в дальнейшем для анализа серийной однотипной продукции на существенное или несущественное различие по качеству, так как их генеральные дисперсии равны. Первый случай относится к несущественному различию выборочных средних, второй – к существенному. Случаи 3) и 4) не рассматриваются, так как партии изделий с неоднородными дисперсиями не относятся к однотипной серийной продукции.

Алгоритм проверки выборочных дисперсий на однородность по критерию Фишера:

– расчет выборочных параметров обеих выборок $\overline{Y}_1$, S_1^2, f_{S_1} и $\overline{Y}_2$, S_2^2, f_{S_2}:

$$\overline{Y} = \frac{Y_1 + \ldots + Y_n}{n} = \frac{\sum_{i=1}^{n} Y_i}{n},$$

$$S^2 = \frac{(Y_1 - \overline{Y})^2 + \ldots + (Y_n - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1},$$

$$S = \sqrt{S^2},$$

$$f_S = n - 1;$$

– расчет экспериментального значения критерия Фишера:

$$F_э = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)}; \qquad (43)$$

– определение критического значения критерия Фишера $F_{f_{числ},\, f_{знам},\, p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии, а на втором – число степеней свободы *меньшей* дисперсии при доверительной вероятности p = 0,95 (прилож. 4):

если

$$F_э < F_{f_1;\, f_2;\, 0{,}95}, \qquad (44)$$

то выборочные дисперсии S_1^2 и S_2^2 однородны, то есть $\sigma_1^2 = \sigma_2^2$,

если

$$F_э \geq F_{f_1;\, f_2;\, 0{,}95}, \qquad (45)$$

то выборочные дисперсии S_1^2 и S_2^2 неоднородны, то есть $\sigma_1^2 \neq \sigma_2^2$.

Сравнительный анализ двух выборок с неоднородными дисперсиями не рассматривается, так как такая продукция не является однотипной.

Если выборочные дисперсии однородны ($\sigma_1^2 = \sigma_2^2$), то в этом случае (и только в этом случае) можно сравнивать выборочные средние $\overline{Y}_1$ и $\overline{Y}_2$ на существенное различие с помощью критерия Стьюдента по следующему алгоритму:

– расчет экспериментального значения критерия Стьюдента:

$$t_э = \frac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}; \qquad (46)$$

– определение критического значения критерия Стьюдента $t_{f_t, \text{р}}$ при числе степеней свободы $f_t = n_1 + n_2 - 2$ и доверительной вероятности p = 0,95 (прилож. 2):

если

$$t_э < t_{n_1+n_2-2;\,0,95}, \tag{47}$$

то различие между выборочными средними $\overline{Y}_1$ и $\overline{Y}_2$ несущественно. В этом случае обе выборки принадлежат к одной и той же генеральной совокупности, то есть $\mu_1 = \mu_2$, поэтому различие между выборочными средними $\overline{Y}_1$ и $\overline{Y}_2$ случайно, следовательно, качество изделий, характеризуемое параметрами $\overline{Y}_1$ и $\overline{Y}_2$, одинаково,

если

$$t_э \geq t_{n_1+n_2-2;\,0,95}, \tag{48}$$

то различие между $\overline{Y}_1$ и $\overline{Y}_2$ существенно. Это означает, что две выборки принадлежат к различным генеральным совокупностям $\mu_1 \neq \mu_2$, причем, если $\overline{Y_1} > \overline{Y_2}$, то $\mu_1 > \mu_2$, если $\overline{Y_1} < \overline{Y_2}$, то $\mu_1 < \mu_2$. Если $\overline{Y_1} > \overline{Y_2}$, то это означает, что изделия первой выборки лучше по качеству, чем изделия второй выборки (предполагается, что чем больше $\overline{Y}$, тем выше качество). Если $\overline{Y_1} < \overline{Y_2}$, то изделия второй выборки лучше по качеству, чем изделия первой выборки.

Таким образом, сделать заключение о качестве серийной однотипной продукции, характеризуемом параметрами $\overline{Y}_1$ и $\overline{Y}_2$ по критерию Стьюдента, можно только при условии, что выборочные дисперсии однородны по критерию Фишера.

2.3.1. Алгоритм сравнительной оценки параметров двух выборок

Алгоритм сравнение двух выборок нормально распределенных случайных величин включает следующие операции.

1. Расчет выборочных параметров для каждой выборки:

– выборочное среднее:

$$\overline{Y} = \frac{Y_1 + \ldots + Y_n}{n} = \frac{\sum_{i=1}^{n} Y_i}{n};$$

– выборочная дисперсия, выборочное стандартное отклонение и число степеней свободы выборочной дисперсии в каждой выборке:

$$S^2 = \frac{(Y_1 - \overline{Y})^2 + \ldots + (Y_n - \overline{Y})^2}{n-1} = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1},$$

$$S = \sqrt{S^2},$$

$$f_S = n - 1.$$

2. Проверка случайных значений обеих выборок на наличие промаха по критерию Смирнова–Граббса:

– определение «кандидата» $Y_{\text{канд}}$, который находится от выборочного среднего на максимальном расстоянии, то есть удовлетворяет следующему условию:

$$Y_{\text{канд}} = \left| Y_i - \overline{Y} \right| = \max;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{\text{э}} = \frac{\left| Y_{\text{канд}} - \overline{Y} \right|}{S};$$

– определение критического значения критерия Смирнова–Граббса $\tau_{f_\tau,\, p}$, где $f_\tau = n - 2$ – число степеней свободы при доверительной вероятности p = 0,95 (прилож. 1):

– $Y_{\text{канд}}$ не является промахом, если

$$\tau_{\text{э}} < \tau_{n-2;\, 0,95},$$

и в этом случае его следует оставить в выборке;

– $Y_{\text{канд}}$ является промахом, если

$$\tau_{\text{э}} \geq \tau_{n-2;\, 0,95},$$

и в этом случае его следует исключить из выборки.

3. Проверка случайных значений обеих выборок на принадлежность к нормальному закону распределения по критерию Гири:

– расчет среднего абсолютного отклонения:

$$\text{САО} = \frac{\sum_{i=1}^{n_j} \left| Y_i - \overline{Y} \right|}{n};$$

– расчет экспериментального значения критерия Гири:

$$\theta_э = \left| \sqrt{\frac{n}{n-1}} \cdot \frac{\text{САО}}{S} - 0,7979 \right|;$$

– определение критического значения критерия Гири:

$$\theta_{n;\,p} = \frac{0,4}{\sqrt{n}};$$

– случайные значения выборки принадлежат к нормальному закону распределения, если $\theta_э < \theta_{n_j\,;\,p}$;

– случайные значения выборки не принадлежат к закону нормального распределения, если $\theta_э \geq \theta_{n_j\,;\,p}$.

Выборки, случайные значения которых не содержат промаха и подчиняются нормальному закону распределения, называются очищенными. Дальнейший статистический анализ будет осуществляться только на очищенных выборках.

4. Проверка дисперсии двух выборок $(\overline{Y}_1,\, S_1^2,\, n_1)$ и $(\overline{Y}_2,\, S_2^2,\, n_2)$ на однородность по критерию Фишера:

– расчет экспериментального значения критерия Фишера:

$$F_э = \frac{\max(S_1^2,\, S_2^2)}{\min(S_1^2,\, S_2^2)};$$

– определение критического значения критерия Фишера $F_{f_{\text{числ}},\, f_{\text{знам}},\, p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии, на втором – число степеней свободы *меньшей* дисперсии при доверительной вероятности $p = 0{,}95$ (прилож. 4):

если

$$F_э < F_{f_1, f_2, p},$$

то выборочные дисперсии S_1^2 и S_2^2 однородны, то есть $\sigma_1^2 = \sigma_2^2$,

если

$$F_э \geq F_{f_1, f_2, p},$$

то выборочные дисперсии S_1^2 и S_2^2 неоднородны, то есть $\sigma_1^2 \neq \sigma_2^2$.

Если выборочные дисперсии S_1^2 и S_2^2 однородны, то $\sigma_1^2 = \sigma_2^2$, и только в этом случае можно сравнивать выборочные средние $\overline{Y}_1$ и $\overline{Y}_2$ на существенное различие по критерию Стьюдента по следующему алгоритму.

5. Сравнение серийной однотипной продукции по качеству, характеризуемую двумя выборками $(\overline{Y}_1, S_1^2, n_1)$ и $(\overline{Y}_2, S_2^2, n_2)$, $\overline{Y}_1 \neq \overline{Y}_2$ с однородными дисперсиями $\sigma_1^2 = \sigma_2^2$:

– расчет экспериментального значения критерия Стьюдента:

$$t_э = \frac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}};$$

– определение критического значения критерия Стьюдента $t_{f_S; p}$ при числе степеней свободы $f_S = n_1 + n_2 - 2$ и доверительной вероятности p = 0,95 (прилож. 2):

если

$$t_э < t_{n_1 + n_2 - 2;\, p},$$

то различие между выборочными средними $\overline{Y}_1$ и $\overline{Y}_2$ несущественно, то есть, несмотря на то, что $\overline{Y}_1 \neq \overline{Y}_2$, изделия обеих выборок обладают одинаковым качеством,

если

$$t_э \geq t_{n_1 + n_2 - 2;\, p},$$

то различие между выборочными средними $\overline{Y}_1$ и $\overline{Y}_2$ существенно. То есть, если $\overline{Y}_1 > \overline{Y}_2$, то изделия первой выборки лучше по качеству, чем изделия второй выборки (предполагается, чем больше $\overline{Y}$, тем выше качество). Если $\overline{Y}_1 < \overline{Y}_2$, то изделия второй выборки лучше по качеству, чем изделия первой выборки.

2.3.2. Типовая задача: сравнительная оценка однотипной продукции по качеству

Цель – сравнение однотипной серийной продукции по качеству.

Поясним методику решения типовой задачи по описанному алгоритму на следующем примере: в вольерах птицефабрики выращиваются цыплята одной породы. В процессе вскармливания предлагается ввести новую витаминную добавку в корм. Необходимо оценить, оказывает ли введение в корм витаминной добавки существенное влияние на привес цыплят.

Математическая постановка задачи – оценить по двум выборкам существенное влияние витаминной добавки на привес цыплят в процессе вскармливания.

Для проведения эксперимента были организованы 2 контрольных вольера, в каждый из которых поместили по 200 цыплят суточного возраста. Чтобы убедиться, что в начале эксперимента нет существенной разницы в стартовой массе цыплят, из вольера № 1 отобрали случайным образом 18 цыплят, а из вольера № 2 – 20 цыплят, после чего определили их массу (табл. 1).

Таблица 1

Стартовая масса цыплят

Вольер	Масса, г									
№ 1	50	48	49	49	50	52	50	51	51	50
	51	50	51	49	49	54	49	47	–	–
№ 2	52	49	48	48	50	50	45	51	50	49
	47	53	52	48	50	53	49	46	47	47

Для оценки различия выборочных средних значений стартовой массы цыплят в вольерах № 1 и № 2 выполним статистические расчеты в соответствии с алгоритмом (см. п. 2.2.1).

1. Расчет выборочных параметров для каждой выборки:

– выборочные средние в каждой выборке:

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n_1} Y_{1i}}{n_1} = \frac{\sum_{i=1}^{18} Y_{1i}}{18} = \frac{50 + 48 + 49 + \mathbf{K} + 47}{18} = 50,00,$$

$$\overline{Y}_2 = \frac{\sum_{i=1}^{n_1} Y_{2i}}{n_2} = \frac{\sum_{i=1}^{20} Y_{2i}}{20} = \frac{52 + 49 + 48 + \mathbf{K} + 47}{18} = 49,20;$$

– выборочные дисперсии, выборочные стандартные отклонения и числа степеней свободы выборочной дисперсии в каждой выборке:

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (Y_{1i} - \overline{Y}_1)^2}{n_1 - 1} = \frac{(50-50,00)^2 + (48-50,00)^2 + \mathbf{K} + (47-50,00)^2}{18-1} = 2,685,$$

$$S_1 = \sqrt{S_1^2} = 1,620,$$

$$f_{S_1} = n_1 - 1 = 18 - 1 = 17,$$

$$S_2^2 = \frac{\sum_{i=1}^{n_1} (Y_i - \overline{Y}_2)^2}{n_2 - 1} = \frac{(52-49,20)^2 + (49-49,20)^2 + \mathbf{K} + (47-49,20)^2}{20-1} = 5,116,$$

$$S_2 = \sqrt{S_2^2} = 2,262,$$

$$f_{S_2} = n_2 - 1 = 20 - 1 = 19.$$

2. Проверка значений обеих выборок на наличие промаха по критерию Смирнова–Граббса:

– определение в 1-й выборке «кандидата» на промах $Y_{1\text{канд}}$, который находится на максимальном расстоянии от выборочного среднего, то есть удовлетворяет следующему условию:

$$Y_{\text{канд}} = \left| Y_i - \overline{Y} \right| = \max,$$

$$Y_{1\text{канд}} = 54;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{1э} = \frac{\left|Y_{1канд} - \overline{Y}_1\right|}{S_1} = \frac{\left|54 - 50{,}00\right|}{1{,}620} = 2{,}469\,;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{1\, f_{1\tau},\, p}$, где $f_{1\tau} = n_1 - 2 = 18 - 2 = 16$ – число степеней свободы при доверительной вероятности $p = 0{,}95$ (прилож. 1):

$$\tau_{1_{16,\,0{,}95}} = 2{,}577\,;$$

– $Y_{1канд} = 54$ не является промахом, так как

$$\tau_{1э} = 2{,}545 < 2{,}577 = \tau_{16,\,0{,}95}\,,$$

следовательно, 1-я выборка не содержит промаха;

– определение во 2-й выборке «кандидата» на промах $Y_{2канд}$, который находится на максимальном расстоянии от выборочного среднего, то есть удовлетворяет следующему условию:

$$Y_{канд} = \left|Y_i - \overline{Y}\right| = \max\,,$$

$$Y_{2канд} = 45;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{2э} = \frac{\left|Y_{2канд} - \overline{Y}_2\right|}{S_2} = \frac{\left|45 - 49{,}2\right|}{2{,}262} = 1{,}857\,;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{2\, f_{2\tau},\, \mathrm{p}}$, где $f_{2\tau} = n_2 - 2 = 20 - 2 = 18$ – число степеней свободы при доверительной вероятности $p = 0{,}95$ (прилож. 1):

$$\tau_{2_{18,\,0{,}95}} = 2{,}623;$$

– $Y_{2канд} = 45$ не является промахом, так как

$$\tau_{2э} = 1{,}857 < 2{,}623 = \tau_{18,\,0{,}95}\,,$$

следовательно, 2-я выборка не содержит промаха.

3. Проверка случайных значений обеих выборок на принадлежность к нормальному закону распределения по критерию Гири:

– расчет для 1-й выборки среднего абсолютного отклонения:

$$\text{САО}_1 = \frac{\sum_{i=1}^{n_1} \left| Y_{1i} - \overline{Y}_1 \right|}{n_1} = \frac{\left|50 - 50,00\right| + \left|48 - 50,00\right| + \mathbf{K} + \left|47 - 50,00\right|}{18 - 1} = 1{,}111;$$

– расчет для 1-й выборки экспериментального значения критерия Гири:

$$\theta_э = \left| \sqrt{\frac{n_1}{n_1 - 1}} \cdot \frac{\text{САО}_1}{S_1} - 0,7979 \right| = \left| \sqrt{\frac{18}{18 - 1}} \cdot \frac{1,111}{1,572} - 0,7979 \right| = 0,07051;$$

– определение для 1-й выборки критического значения критерия Гири:

$$\theta_{n_1;\,0,95} = \frac{0,4}{\sqrt{n_1}} = \frac{0,4}{\sqrt{18}} = 0,09428;$$

– 1-я выборка принадлежит к нормальному закону распределения, так как

$$\theta_{1э} = 0,070\ 51 < 0,094\ 28 = \theta_{18;0,95}.$$

Вывод: 1-я выборка является очищенной выборкой;

– расчет для 2-й выборки среднего абсолютного отклонения:

$$\text{САО}_2 = \frac{\sum_{i=1}^{n_1} \left| Y_{2i} - \overline{Y}_2 \right|}{n_2} = \frac{\left|52 - 49,20\right| + \left|49 - 49,20\right| + \mathbf{K} + \left|47 - 49,20\right|}{20 - 1} = 1{,}820;$$

– расчет для 2-й выборки экспериментального значения критерия Гири:

$$\theta_э = \left| \sqrt{\frac{n_2}{n_2 - 1}} \cdot \frac{\text{САО}_2}{S_2} - 0,7979 \right| = \left| \sqrt{\frac{20}{20 - 1}} \cdot \frac{1,820}{2,262} - 0,7979 \right| = 0,027\,67;$$

– определение для 2-й выборки критического значения критерия Гири:

$$\theta_{n_2;\,0,95} = \frac{0,4}{\sqrt{n_2}} = \frac{0,4}{\sqrt{20}} = 0,089\,44;$$

– 2-я выборка принадлежит к нормальному закону распределения, так как

$$\theta_{2э} = 0,027\ 67 < 0,089\ 44 = \theta_{n_2;0,95}.$$

Вывод: 2-я выборка является очищенной выборкой.

4. Проверка дисперсии двух выборок $(\overline{Y}_1,\ S_1^2,\ n_1)$ и $(\overline{Y}_2, S_2^2, n_2)$ на однородность по критерию Фишера:

– расчет экспериментального значения критерия Фишера:

$$F_э = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)} = \frac{5,116}{2,685} = 1,905;$$

– определение критического значения критерия Фишера $F_{f_{числ},\, f_{знам},\, p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии, а на втором – число степеней свободы *меньшей* дисперсии при доверительной вероятности p = 0,95 (прилож. 4):

$$F_{19;17;0,95} = 2,243;$$

– так как

$$F_э = 1,905 < 2,243 = F_{19;17;0,95},$$

то выборочные дисперсии S_1^2 и S_2^2 однородны, то есть $\sigma_1^2 = \sigma_2^2$.

Вывод: однородность выборочных дисперсий позволяет проверить выборочные средние на существенное различие по критерию Стьюдента.

5. Сравнение однотипной продукции (цыплята из вольеров № 1 и № 2) по качеству (стартовая масса цыпленка), характеризуемой двумя выборками $(\overline{Y}_1,\ S_1^2,\ n_1)$ и $(\overline{Y}_2, S_2^2, n_2)$, $\overline{Y}_1 \neq \overline{Y}_2$ с однородными дисперсиями $\sigma_1^2 = \sigma_2^2$:

– расчет экспериментального значения критерия Стьюдента:

$$t_э = \frac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}}\sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}} =$$

$$= \frac{\left|50,00 - 49,20\right|}{\sqrt{(18-1)\,2,685 + (20-1)\,5,116}}\sqrt{\frac{18\cdot 20(18+20-2)}{18+20}} = 1,252;$$

– определение критического значения критерия Стьюдента $t_{18+20-2;\,0,95} = 2,028$ (прилож. 2);

– так как

$$t_э = 1,252 < 2,028 = t_{18+20-2;\,0,95},$$

то различие между выборочными средними $\overline{Y}_1 = 50,00$ и $\overline{Y}_2 = 49,20$ несущественно.

Так как выборочные средние отражают среднюю массу цыплят суточного возраста, и различие между ними несущественно, то можно сделать вывод: различие по массе цыплят носит случайный характер и их стартовая масса одинакова.

Для исследования влияния витаминной добавки на массу цыплят суточного возраста ее в количестве 4 г/сут добавляли в корм для цыплят в вольере № 2 в течение двух недель, а потом из вольеров № 1 и № 2 случайным образом выбрали, соответственно, 24 и 26 цыплят и определили их массу (табл. 2).

Таблица 2

Масса цыплят 2-недельного возраста

Вольер	Масса, г												
№ 1	402	411	405	406	406	411	408	407	404	407	412	418	–
	407	409	412	411	412	405	405	405	408	404	397	400	–
№ 2	414	411	416	413	419	415	413	413	414	418	415	416	422
	413	414	412	408	416	415	418	426	420	420	420	420	419

Для сравнения выборочных средних значений массы цыплят 2-недельного возраста, получавших различный корм, выполним необходимые расчеты в соответствии с алгоритмом (см. п. 2.2.1).

1. Расчет выборочных параметров для каждой выборки:
– выборочные средние в каждой выборке:

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n_1} Y_{1i}}{n_1} = \frac{\sum_{i=1}^{24} Y_{1i}}{24} = \frac{402 + 411 + 405 + \mathbf{K} + 400}{24} = 407,2\,.$$

$$\overline{Y}_2 = \frac{\sum_{i=1}^{n_2} Y_{2i}}{n_2} = \frac{\sum_{i=1}^{20} \mathrm{Y}_{2i}}{20} = \frac{52 + 49 + 48 + \mathbf{K} + 47}{18} = 49,20;$$

– выборочные дисперсии, выборочные стандартные отклонения и числа степеней свободы выборочных дисперсий в каждой выборке:

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (Y_{1i} - \overline{Y}_1)^2}{n_1 - 1} =$$

$$= \frac{\left(402 - 407,2\right)^2 + \left(411 - 407,2\right)^2 + \mathbf{K} + \left(400 - 407,2\right)^2}{24 - 1} = 19,97,$$

$$S_1 = \sqrt{S_1^2} = 4,469\,,$$

$$f_{S_1} = n_1 - 1 = 24 - 1 = 23,$$

$$S_2^2 = \frac{\sum_{i=1}^{n_1} (Y_i - \overline{Y}_2)^2}{n_2 - 1} =$$

$$= \frac{\left(414 - 416,2\right)^2 + \left(411 - 416,2\right)^2 + \mathbf{K} + \left(419 - 416,2\right)^2}{26 - 1} = 15,26,$$

$$S_2 = \sqrt{S_2^2} = 3,906\,,$$

$$f_{S_2} = n_2 - 1 = 26 - 1 = 25.$$

2. Проверка случайных значений обеих выборок на наличие промахов по критерию Смирнова–Граббса:

– определение в 1-й выборке «кандидата» на промах $Y_{1канд}$, который находится на максимальном расстоянии от выборочного среднего, то есть удовлетворяет следующему условию:

$$Y_{канд} = \left| Y_i - \overline{Y} \right| = \max ,$$

$$Y_{1канд} = 418;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{1э} = \frac{\left| Y_{1канд} - \overline{Y}_1 \right|}{S_1} = \frac{\left| 418 - 407,2 \right|}{4,469} = 2,424 ;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{1\, f_{1\tau},\, p}$, где $f_{1\tau} = n_1 - 2 = 24 - 2 = 22$ – число степеней свободы при доверительной вероятности $p = 0,95$ (прилож. 1):

$$\tau_{1_{22,\, 0,95}} = 2,701;$$

– $Y_{1канд} = 418$ не является промахом, так как

$$\tau_{1э} = 2,424 < 2,701 = \tau_{22,\, 0,95} ,$$

следовательно, 1-я выборка не содержит промаха;

– определение во 2-й выборке «кандидата» на промах $Y_{2канд}$, который находится на максимальном расстоянии от выборочного среднего, то есть удовлетворяет следующему условию:

$$Y_{канд} = \left| Y_i - \overline{Y} \right| = \max ,$$

$$Y_{2канд} = 426;$$

– расчет экспериментального значения критерия Смирнова–Граббса:

$$\tau_{2э} = \frac{\left| Y_{2канд} - \overline{Y}_2 \right|}{S_2} = \frac{\left| 426 - 416,2 \right|}{3,906} = 2,521;$$

– определение критического значения критерия Смирнова–Граббса $\tau_{2\,f_{2\tau},\,p}$, где $f_{2\tau} = n_2 - 2 = 26 - 2 = 24$ – число степеней свободы при доверительной вероятности $p = 0{,}95$ (прилож. 1):

$$\tau_{2_{24,\,0,95}} = 2{,}734;$$

– $Y_{2\text{канд}} = 426$ не является промахом, так как

$$\tau_{2\,э} = 1{,}857 < 2{,}623 = \tau_{18,\,0,95},$$

следовательно, 2-я выборка не содержит промаха.

3. Проверка случайных значений обеих выборок на принадлежность к нормальному закону распределения по критерию Гири:

– расчет для 1-й выборки среднего абсолютного отклонения:

$$\text{САО}_1 = \frac{\sum_{i=1}^{n_1} \left|Y_{1i} - \bar{Y}_1\right|}{n_1} = \frac{\left|402 - 407{,}2\right| + \left|411 - 407{,}2\right| + \mathbf{K} + \left|400 - 407{,}2\right|}{24 - 1} = 3{,}361;$$

– расчет для 1-й выборки экспериментального значения критерия Гири:

$$\theta_э = \left|\sqrt{\frac{n_1}{n_1 - 1}} \cdot \frac{\text{САО}_1}{S_1} - 0{,}7979\right| = \left|\sqrt{\frac{24}{24 - 1}} \cdot \frac{3{,}361}{4{,}469} - 0{,}7979\right| = 0{,}02961;$$

– определение для 1-й выборки критического значения критерия Гири:

$$\theta_{n_1;\,0,95} = \frac{0{,}4}{\sqrt{\text{n}_1}} = \frac{0{,}4}{\sqrt{24}} = 0{,}081\,65;$$

– 1-я выборка принадлежит к нормальному закону распределения, так как

$$\theta_{1э} = 0{,}07051 < 0{,}09428 = \theta_{24;0,95}.$$

Вывод: 1-я выборка является очищенной выборкой;

– расчет для 2-й выборки среднего абсолютного отклонения:

$$\text{САО}_2 = \frac{\sum_{i=1}^{n_1} \left| Y_{2i} - \bar{Y}_2 \right|}{n_2} = \frac{|402-407,2| + |411-407,2| + \mathbf{K} + |400-407,2|}{24-1} = 3{,}117;$$

– расчет для 2-й выборки экспериментального значения критерия Гири:

$$\theta_э = \left| \sqrt{\frac{n_2}{n_2-1}} \cdot \frac{\text{САО}_2}{S_2} - 0,7979 \right| = \left| \sqrt{\frac{26}{26-1}} \cdot \frac{3,117}{3,906} - 0,7979 \right| = 0,01475;$$

– расчет для 2-й выборки критического значения критерия Гири:

$$\theta_{n_2;\,0,95} = \frac{0,4}{\sqrt{n_2}} = \frac{0,4}{\sqrt{26}} = 0,07845;$$

– 2-я выборка принадлежит к нормальному закону распределения, так как

$$\theta_{2э} = 0,014\,75 < 0,07845 = \theta_{26;0,95}.$$

Вывод: 2-я выборка является очищенной выборкой.

4. Проверка дисперсий двух выборок $(\bar{Y}_1,\, S_1^2,\, n_1)$ и $(\bar{Y}_2,\, S_2^2,\, n_2)$ на однородность по критерию Фишера:

– расчет экспериментального значения критерия Фишера:

$$F_э = \frac{\max(S_1^2,\, S_2^2)}{\min(S_1^2,\, S_2^2)} = \frac{19,97}{15,26} = 1,309;$$

– определение критического значения критерия Фишера $F_{f_{\text{числ}},\, f_{\text{знам}},\, p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии, на втором – число степеней свободы *меньшей* дисперсии при доверительной вероятности p = 0,95 (прилож. 4):

$$F_{23;25;0,95} = 1,974;$$

так как

$$F_э = 1,309 < 1,974 = F_{23;25;0,95},$$

то выборочные дисперсии S_1^2 и S_2^2 однородны, то есть $\sigma_1^2 = \sigma_2^2$.

Вывод: однородность выборочных дисперсий позволяет проверить выборочные средние на существенное различие по критерию Стьюдента.

5. Сравнение однотипной продукции (цыплята из вольеров № 1 и № 2) по качеству (масса цыпленка после 2-недельного вскармливания) по двум выборкам $(\overline{Y}_1, S_1^2, n_1)$ и $(\overline{Y}_2, S_2^2, n_2)$, $\overline{Y}_1 \neq \overline{Y}_2$ с однородными дисперсиями $\sigma_1^2 = \sigma_2^2$:

– расчет экспериментального значения критерия Стьюдента:

$$t_э = \frac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} =$$

$$= \frac{\left|407,2 - 416,2\right|}{\sqrt{(24-1)19,97 + (26-1)15,26}} \sqrt{\frac{24 \cdot 26(24 + 26 - 2)}{24 + 26}} = 7,586;$$

– определение критического значения критерия Стьюдента $t_{24+26-2;\,0,95} = 2,011$ (прилож. 2);

– так как

$$t_э = 7,586 > 2,011 = t_{24+26-2;\,0,95},$$

то различие между выборочными средними массы цыплят $\overline{Y}_1 = 407,2$ и $\overline{Y}_2 = 416,2$ существенно, то есть цыплята, получавшие витаминную добавку, набрали бóльшую массу за 2 недели вскармливания.

Таким образом, на основе статистического анализа выборочных данных, полученных при экспериментальном исследовании влияния витаминной добавки на массу цыплят при вскармливании, установлено, что витаминная добавка оказывает положительное биологическое воздействие на привес цыплят.

2.4. Вопросы и задания для самопроверки

1. Какими параметрами однозначно определяется нормальный закон распределения случайных величин?

2. В чем различие между понятиями «генеральная совокупность» и «выборка»?

3. Напишите формулу для расчета выборочного среднего.

4. Напишите формулу для расчета выборочной дисперсии.

5. Сформулируйте алгоритм нахождения промаха по критерию Смирнова–Граббса.

6. Сформулируйте критерий принадлежности случайных величин выборки к нормальному закону распределения.

7. Напишите формулу для расчета доверительного интервала генерального среднего.

8. Напишите формулу для расчета доверительного интервала генеральной дисперсии.

9. Сформулируйте алгоритм определения однородности дисперсий двух выборок. Как найти критическое значение критерия Фишера?

10. Обязателен ли одинаковый объем выборок $n_1 = n_2$ для сравнения двух выборочных дисперсий?

11. Приведите формулу для определения существенного различия между двумя выборочными средними. Как найти критическое значение критерия Стьюдента?

12. Укажите правильные формулы.

Выборочное среднее:

а) $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n-1}$; б) $\overline{Y} = \dfrac{Y_1 + \ldots + Y_n}{n}$; в) $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$; г) $\overline{Y} = \dfrac{\sum_{i=1}^{n} |Y_i|}{n}$.

Выборочная дисперсия:

а) $S^2 = \dfrac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n}$; б) $S^2 = \dfrac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1}$;

в) $S^2 = \dfrac{\sum_{i=1}^{n} Y_i^2}{n}$; г) $S^2 = \dfrac{\sum_{i=1}^{n} \left|Y_i - \overline{Y}\right|}{n-1}$.

Критерий Смирнова–Граббса:

а) $\dfrac{\left|Y_{\min} - Y_{\max}\right|}{S} < \tau_{n-1,\,0{,}95}$; б) $\dfrac{\left(Y_{\max} - \overline{Y}\right)}{S} < \tau_{n-2,\,0{,}95}$;

в) $\dfrac{\max\left|Y_i - \overline{Y}\right|}{S} < \tau_{n-2,\,0{,}95}$; г) $\dfrac{\max\left|Y_i - \overline{Y}\right|}{S} < \tau_{n-1,\,0.95}$.

Критерий принадлежности случайных значений выборки к нормальному закону распределения:

а) $\left|\dfrac{\sum\limits_{i=1}^{n}\left|Y_i - \overline{Y}\right|}{nS} - 0{,}7979\right| < \dfrac{2}{5\sqrt{n}}$; б) $\left|\dfrac{\sum\limits_{i=1}^{n}\left|Y_i - \overline{Y}\right|}{S^2} - 0{,}7979\right| < \dfrac{0{,}4}{\sqrt{n}}$;

в) $\left|\sqrt{\dfrac{n}{n-1}} \cdot \dfrac{\text{САО}}{S} - 0{,}7979\right| < \dfrac{0{,}4}{\sqrt{n}}$; г) $\left|\dfrac{\sum\limits_{i=1}^{n}\left|Y_i - \overline{Y}\right|}{nS} - 0{,}7979\right| < 0{,}4\sqrt{n}$.

Доверительный интервал генерального среднего $\overline{Y} - \Delta\overline{Y} < \mu < \overline{Y} + \Delta\overline{Y}$:

а) $\Delta\overline{Y} = t_{n-1;\,0{,}95}\, S(\overline{Y})$; б) $\Delta\overline{Y} = \dfrac{t_{n-1;\,0{,}95}\, S(Y)}{\sqrt{n}}$;

в) $\Delta\overline{Y} = \dfrac{t_{n-1;\,0{,}95}\, S(\overline{Y})}{\sqrt{n}}$; г) $\Delta\overline{Y} = \dfrac{t_{n;\,0{,}95}\, S}{\sqrt{n}}$.

Доверительный интервал генерального стандартного отклонения:

а) $\sqrt{\dfrac{(n-1)}{\chi^2_{n-1;\,0{,}025}}} < \sigma < \sqrt{\dfrac{(n-1)}{\chi^2_{n-1;\,0{,}975}}}$; б) $\dfrac{S\sqrt{n-1}}{\sqrt{\chi^2_{n-1;\,0{,}025}}} < \sigma < \dfrac{S\sqrt{n-1}}{\sqrt{\chi^2_{n-1;\,0{,}975}}}$;

в) $\sqrt{\dfrac{n}{\chi^2_{n-1;\,0{,}025}}} < \dfrac{\sigma^2}{S^2} < \sqrt{\dfrac{n}{\chi^2_{n-1;\,0{,}975}}}$; г) $\dfrac{(n-1)}{\chi^2_{n-1;\,0{,}025}} < \dfrac{\sigma}{S} < \dfrac{(n-1)}{\chi^2_{n-1;\,0{,}975}}$.

Какая часть случайных величин, подчиняющихся нормальному закону распределения, попадает в диапазон $-2\sigma \le \delta \le 2\sigma$:

а) 68,27 %; б) 95,44 %; в) 99,27 %; г) 95,0 %?

Две выборочные дисперсии однородны, то есть $\sigma_1^2 = \sigma_2^2$, при $S_1^2 > S_2^2$, если

а) $\dfrac{S_2^2}{S_1^2} < F_{n_2-1;\, n_1-1;\, 0,95}$; б) $\dfrac{S_1^2}{S_2^2} < F_{n_1-1;\, n_2-1;\, 0,95}$;

в) $\dfrac{S_1}{S_2} < F_{n_1-1;\, n_2-1;\, 0,95}$; г) $\dfrac{S_1^2}{S_2^2} < F_{n_1;\, n_2;\, 0,95}$.

Различие двух выборочных средних, для которых $\sigma_1^2 = \sigma_2^2$, несущественно, то есть $\mu_1 = \mu_2$ при $\overline{Y}_1 \neq \overline{Y}_2$, если:

а) $\dfrac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} < t_{n_1+n_2-2;\, p}$;

б) $\dfrac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1-1) S_1^2 + (n_2-1) S_2^2}} \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} < t_{n_1+n_2-2;\, p}$;

в) $\dfrac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1-1) S_1 + (n_2-1) S_2}} \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} < t_{n_1+n_2-2;\, p}$;

г) $\dfrac{\left|\overline{Y}_1 - \overline{Y}_2\right|}{\sqrt{(n_1-1) S_1^2 + (n_2-1) S_2^2}} \sqrt{\dfrac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} < t_{n_1+n_2;\, p}$.

13. Дана выборка: 5,2; 5,0; 5,4; 3,1; 6,4; 5,8; 4,4; 4,6; 4,6; 6,1. Рассчитать выборочные среднее, дисперсию и стандартное отклонение.

Ответ: $\overline{Y} = 5{,}1$; $S^2 = 0{,}92$; $S = 0{,}96$.

14. Есть ли в выборке задачи № 13 промах? Чему равно критическое значение критерия Смирнова–Граббса для этой выборки при доверительной вероятности $p = 0{,}95$?

Ответ: в этой выборке промаха нет. Критерий $\tau_{9,\, 0,95} = 2{,}294$.

15. Проверить, выполняется ли для случайных величин выборки, приведенной в задаче № 13, закон нормального распределения при доверительной вероятности $p = 0{,}95$.

Ответ: выполняется.

16. Рассчитать доверительный интервал генерального среднего μ по выборке задачи № 13 при доверительной вероятности $p = 0{,}95$.

Ответ: $4{,}4 < \mu < 5{,}8$.

17. Рассчитать доверительный интервал генерального стандартного отклонения σ по выборке задачи № 13 при доверительной вероятности $p = 0{,}95$.

Ответ: $0{,}7 < \sigma < 1{,}8$.

18. Есть ли в выборке 5,2; 5,0; 5,4; 2,2; 6,4; 5,8; 4,4; 4,6; 4,6; 6,1 промах при доверительной вероятности $p = 0{,}95$?

Ответ: значение 2,2 является промахом.

19. Проверить, выполняется ли для случайных величин выборки 5,2; 5,0; 5,4;2,9; 6,5; 6,4; 4,4; 4,6; 4,6; 7,1 закон нормального распределения при доверительной вероятности $p = 0{,}95$.

Ответ: промахов нет, но закон нормального распределения при доверительной вероятности $p = 0{,}95$ не выполняется.

20. Для сравнения двух крупных партий жемчуга по качеству (чем меньше разброс по диаметру жемчужин, тем выше качество и стоимость) произвели измерения отклонений диаметров от номинального размера в двух выборках: $n_1 = 18$, $S_1^2 = 0{,}015$ мм^2; $n_2 = 22$, $S_2^2 = 0{,}025$ мм^2. Можно ли утверждать, что качество обеих партий жемчуга одинаково?

Ответ: качество жемчуга в обеих партиях одинаковое ($\sigma_1 = \sigma_2$).

21. При производстве кирпичей возникло подозрение, что кирпичи, обожженные в двух разных печах, различаются по плотности. Выборочный контроль дал следующие результаты: $n_1 = 14$, $\overline{Y}_1 = 2{,}43$ г/см^3, $S_1^2 = 16{,}4$ (г/см^3)^2; $n_2 = 10$, $\overline{Y}_2 = 4{,}90$ г/см^3, $S_2^2 = 22{,}5$ (г/см^3)^2. Существенно ли различие кирпичей по плотности?

Ответ: статистическая обработка экспериментальных данных показала, что $\sigma_1 = \sigma_2$ и $\mu_1 = \mu_2$. Наблюдаемое различие по плотности несущественно.

22. Для обработки деталей применялись сверла, изготовленные из инструментальных сталей различных марок. Работоспособность сверл оценивалась временем их работы (в минутах) до переточки. Получены следующие результаты: $\overline{Y}_1 = 26{,}7$ мин, $S_1^2 = 48{,}2$ мин^2, $n_1 = 14$; $\overline{Y}_2 = 15{,}0$ мин, $S_2^2 = 53{,}3$ мин^2, $n_2 = 16$. Можно ли считать

различие выборочных средних значений случайным или это различие существенно?

Ответ: выборочные дисперсии сверл обеих партий однородны (генеральные дисперсии σ_1 и σ_2 равны), а различие в значениях средней стойкости существенно ($\overline{Y_1}$ существенно больше, чем $\overline{Y_2}$).

23. Проверить, выполняется ли для случайных величин выборки 5,2; 5,0; 5,2; 1,6; 6,3; 6,4; 4,4; 5,0; 4,6; 7,1 закон нормального распределения при доверительной вероятности $p = 0{,}95$. Рассчитать выборочные среднее, дисперсию и стандартное отклонение выборки для оценки доверительного интервала выборочного среднего с погрешностью $\Delta\overline{Y} = 0{,}25$ при доверительной вероятности $p = 0{,}95$.

Ответ: рассматривать выполнение закона нормального распределения для заданной выборки некорректно, так как значение 1,6 является промахом. После его исключения значения «очищенной» выборки 5,2; 5,0; 5,2; 6,3; 6,4; 4,4; 5,0; 4,6; 7,1 подчиняются нормальному закону распределения. Для нее $\overline{Y} = 5{,}5$; $S^2 = 0{,}84$; $S = 0{,}92$; $4{,}8 < \mu < 6{,}2$; $0{,}6 < \sigma < 1{,}8$.

24. Даны две выборки: 17,4; 16,3; 17,1; 15,2; 14,1; 14,8; 14,5; 12,9; 17,6; 17,4; 16,6; 16,3 и 14,6; 16,2; 16,2; 17,9; 15,4; 16,6; 18,3; 18,4;19,5; 16,2; 15,3; 17,7; 21,2; 15,9. Найти выборочные параметры обеих выборок $\overline{Y}$, S^2 и сделать вывод о соответствующих генеральных параметрах.

Ответ: в обеих выборках промаха нет, и обе они подчиняются нормальному закону распределения. Параметры выборок: $\overline{Y_1} = 15{,}85$, $S_1^2 = 2{,}32$, $n_1 = 12$; $\overline{Y_2} = 17{,}10$, $S_1^2 = 3{,}32$, $n_2 = 14$. Несмотря на то, что выборочные параметры различаются, генеральные параметры существенно не различаются: $\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$.

25. Даны две выборки: 17,2; 15,1; 17,2; 16,3; 14,2; 16,8; 16,5; 16,0; 15,2; 11,0; 12,9; 16,9 и 21,9; 20,8; 19,1; 18,0; 17,0; 19,1; 19,9; 20,0; 17,4; 20,3; 16,3; 19,4; 20,4; 18,2. Найти выборочные параметры обеих выборок $\overline{Y}$ и S^2 и сделать вывод о генеральных параметрах.

Ответ: в обеих выборках промахов нет, и они подчиняются нормальному закону распределения. Параметры выборок: $\overline{Y_1} = 15{,}44$, $S_1^2 = 3{,}66$, $n_1 = 12$; $\overline{Y_2} = 19{,}13$, $S_1^2 = 2{,}50$, $n_2 = 14$. Выборочные дисперсии однородны: $\sigma_1^2 = \sigma_2^2$. Различие выборочных средних существенно.

26. Даны две выборки: 17,4; 16,3; 17,1; 15,2; 14,1; 14,8; 14,5; 12,9; 17,6; 17,4; 16,6; 16,3 и 12,7; 18,8; 12,3; 15,3; 14,1; 13,3; 13,1; 9,8; 17,7; 14,8; 19,2; 12,7; 19,3; 10,9. Найти выборочные параметры обеих выборок $\overline{Y}$ и S^2 и сделать вывод об аналогичных генеральных параметрах.

Ответ: в обеих выборках промаха нет, и их значения подчиняются нормальному закону распределения. Параметры выборок: $\overline{Y_1} = 15{,}85$, $S_1^2 = 2{,}32$, $n_1 = 12$; $\overline{Y_2} = 14{,}57$, $S_1^2 = 9{,}57$, $n_2 = 14$. Выборочные дисперсии неоднородны. Выборочные средние сравнивать с помощью приведенных в алгоритме методов нельзя.

3. СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ ОДНОФАКТОРНЫХ ТЕХНИЧЕСКИХ СИСТЕМ ПЛАНИРОВАНИЕМ ЭКСПЕРИМЕНТА

В результате изучения темы студент должен

знать:

– алгоритм стохастического моделирования и оптимизации однофакторных технических систем;

уметь:

– применять метод наименьших квадратов (МНК);

– строить однофакторные ортогональные планы 2-го порядка;

– производить предварительную обработку экспериментальных данных однофакторного эксперимента методами математической статистики;

– моделировать однофакторные технические системы уравнениями регрессии 2-го порядка (расчет коэффициентов уравнения регрессии, проверка их на значимость, проверка уравнения регрессии на адекватность);

– оптимизировать технические системы на основе однофакторного уравнения регрессии 2-го порядка (расчет оптимального значения фактора, расчет экстремального значения параметра оптимизации, расчет абсолютной и относительной погрешности);

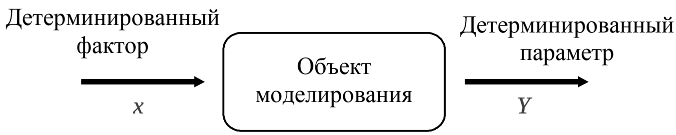
владеть:

– современными методами стохастического моделирования, прогнозирования и оптимизации для решения практических задач в области АПК.

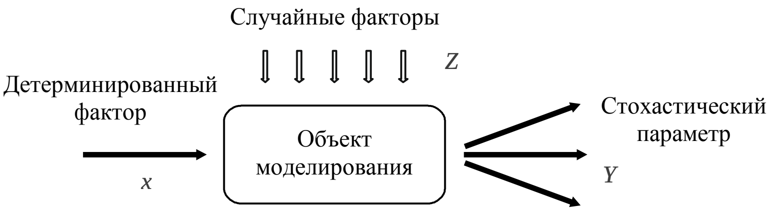
3.1. Стохастическое моделирование однофакторных технических систем

Решение задач агропромышленного комплекса в области животноводства, растениеводства, переработки сельхозпродукции, проектирования и производства машин и механизмов направлено на прогнозирование, управление и оптимизацию процессов и устройств. Современная методология таких решений базируется на математическом моделировании, которое позволяет описывать функционирование технических систем с помощью математических уравнений, построенных на основе фундаментальных законов (детерминированное моделирование) или на экспериментальных данных (стохастическое моделирование), и изучать техническую систему с помощью построенной модели.

Модель однофакторной технической системы схематично можно представить в виде некоторого объекта, состояние которого описывается параметром Y, который изменяется под воздействием детерминированного фактора x (рис. 3). Математическая модель, в которой со 100%-й вероятностью каждому значению фактора x соответствует только одно значение параметра оптимизации технической системы Y, называется детерминированной (рис. 3, *а*). Детерминированная модель строится на базе фундаментальных законов природы, используемых для моделируемой технической системы, что является достаточно сложной и не алгоритмизируемой задачей.



а



б

Рис. 3. Модель технической системы:
а – детерминированной; *б* – стохастической

Однако в реальных технических системах существенное влияние на параметр Y могут оказывать случайные незначительные факторы Z. В этом случае детерминированному значению фактора x соответствуют различные значения параметра оптимизации Y, причем каждое значение может реализоваться с различной вероятностью. Математическая модель, описывающая зависимость выходного параметра Y от входного фактора и учитывающая влияние случайных факторов на параметр Y, в этом случае называется стохастической (рис. 3, *б*). Априори случайные значения параметра оптимизации подчиняются нормальному закону распределения.

Единственным источником информации для построения стохастической модели является совокупность экспериментальных данных, полученных при исследовании технической системы. При этом, если применение фундаментальных законов природы, лежащих в основе функционирования технической системы, является достаточно сложной задачей, то построение стохастической модели является задачей относительно простой. В этом случае существует алгоритм решения, позволяющий на основе экспериментальных данных создать математическую модель в виде алгебраического полинома и использовать ее для оптимизации технической системы.

Любое экспериментальное изучение технической системы, направленное на получение уравнения, отражающего зависимость параметра оптимизации Y от некоторого фактора x, как правило, представляется исследователем в виде графика (рис. 4).

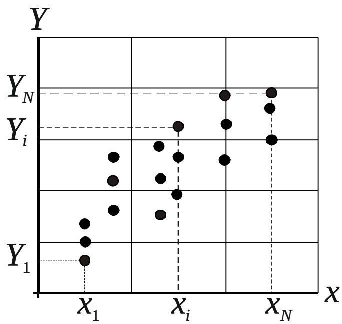


Рис. 4. График результатов наблюдений зависимости параметра Y от фактора x

Однако «облако» экспериментальных данных не позволяет выполнить полноценное моделирование и оптимизацию технической системы. Решение этих задач значительно упрощается, если известна математическая модель в виде аналитической функции $Y = f(x)$. В случае, когда построение детерминированной модели на основе фундаментальных законов природы является достаточно сложной задачей, математическую модель целесообразно построить методом стохастического моделирования. Известно, что любая аналитическая функция может быть представлена в виде ряда Тейлора – алгебраического степенного ряда m-степени, используемого для построения уравнения регрессии:

$$Y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_m x^m, \tag{49}$$

где b_0, b_1, b_2, b_3, …, b_m – регрессионные коэффициенты, рассчитываемые на базе экспериментальных данных.

На практике степень полинома выбирают по принципу разумной достаточности: сначала выбирают уравнение регрессии (например, первой степени), а в случае его неадекватности повышают степень.

Важным преимуществом стохастического моделирования и оптимизации является наличие алгоритма, позволяющего не задумываясь решить эту задачу, включающего 4 блока.

1. Планирование и проведение эксперимента:

– выбор параметра оптимизации, эффективно характеризующего функционирование моделируемой технической системы;

– выбор главного управляющего фактора, влияющего на параметр оптимизации, и интервала его варьирования в натуральных координатах;

– выбор порядка степенного ряда в качестве уравнения регрессии;

– выбор числа опытов и числа дублей в каждом опыте;

– построение матрицы планирования эксперимента, задающей натуральное значение фактора в каждом опыте;

– проведение эксперимента.

2. Предварительная обработка экспериментальных данных:

– расчет выборочного среднего в каждом опыте;

– расчет выборочной дисперсии в каждом опыте;

– проверка выборочных дисперсий на однородность по критерию Кохрена;

– расчет дисперсии воспроизводимости и числа ее степеней свободы.

3. Построение уравнения регрессии и проверка его на статистическое качество:

– перевод натуральных значений фактора в нормированные;

– построение матрицы моделирования, задающей нормированные значения фиктивного, главного и производных факторов в каждом опыте;

– расчет регрессионных коэффициентов методом наименьших квадратов;

– проверка регрессионных коэффициентов на значимость по критерию Стьюдента;

– проверка уравнения регрессии на адекватность по критерию Фишера.

4. Оптимизация технической системы:

– расчет оптимального значения фактора;

– расчет оптимального значения параметра, а также его абсолютной и относительной погрешности.

Результаты расчетов, выполненных в блоках 2 и 3, для наглядности целесообразно записать в виде таблиц перехода от натуральных значений управляющего фактора к нормированным, матрицы планирования и матрицы моделирования.

3.2. Моделирование технических систем однофакторными регрессионными уравнениями 1-го порядка

3.2.1. Матрица планирования и проведение эксперимента

Планирование эксперимента для построения регрессионных моделей начинается с неформализованного этапа, заключающегося в выборе параметра оптимизации Y и главного фактора x.

Выбор параметра оптимизации. Параметр оптимизации – наиболее важная характеристика, отражающая функционирование технической системы. Важно отметить, что параметр оптимизации имеет стохастическую природу, обусловленную непредсказуемым изменением большого количества относительно незначительных факторов Z, воздействие которых невозможно учесть заранее (см. рис. 3, *б*). Несмотря на большое разнообразие параметров оптимизации ограничимся рассмотрением только таких параметров, которые могут быть однозначно определены количественно и имеют наглядный физический смысл. При оптимизации технических систем в сельском хозяйстве (технологии и средства механизации) эффективными параметрами являются, например, уровень удельных затрат (Дж/кг, Дж/м^2, руб./кг, руб./м^2), производительность труда (т/чел./ч, Дж/чел./ч, га/чел./ч).

Общие требования к параметру оптимизации предусматривают, что он должен быть:

– количественным;

– экспериментально определяемым на всех уровнях управляющего фактора;

– непрерывным в исследуемом интервале варьирования фактора;

– имеющим наглядный физический смысл;

– существенным, то есть характеризовать важное свойство технической системы.

Выбор главного фактора. Главный управляющий фактор должен оказывать существенное влияние на параметр оптимизации. В отличие от параметра оптимизации, который имеет стохастическую природу, главный фактор имеет природу детерминированную. Главный фактор должен быть однозначно количественно определяемым, непрерывным. Значения главного управляющего фактора определяются планом эксперимента, разработанным исследователем. При оптимизации технических систем в сельском хозяйстве используются такие факторы, как температура окружающего воздуха, скорость агрегата, мощность трактора, время выполнения операции и др.

Выбор интервалов варьирования главного фактора – неформализуемая задача, и этот выбор может быть основан на знании закономерностей функционирования технической системы, на опыте эксплуатации технической системы и интуиции.

Общие требования к управляющему фактору предусматривают, что он должен быть:

– существенным, то есть вызывать значимое изменение параметра оптимизации;

– количественным;

– детерминированным, то есть его значения на всех уровнях варьирования должны фиксироваться с относительно незначимой приборной погрешностью;

– задаваемым однозначно.

Алгоритм построения стохастической модели проиллюстрируем на примере моделирования процесса сушки зерна. В этом случае в качестве параметра оптимизации целесообразно выбрать уровень удельных затрат Y (Дж/т), а в качестве главного управляющего фактора – температуру (°С) горячего воздуха x, используемого для сушки зерна. Исходя из имеющегося опыта, сушку зерна целесообразно проводить в температурном диапазоне, например, от 60 °С до 120 °С.

Выбор уравнения регрессии в виде степенного ряда. Для иллюстрации универсальности и мощности регрессионного анализа, не нарушая общности рассуждений, рассмотрим однофакторное уравнение регрессии 1-го порядка, которое является частным случаем уравнения регрессии (49):

$$Y = b_0 + b_1 x, \tag{50}$$

где b_0, b_1 – регрессионные коэффициенты, рассчитываемые на основе экспериментальных данных; x – главный фактор (общее количество коэффициентов регрессии $k = 2$).

Построение матрицы планирования начинается с выбора количества опытов N и количества дублей в каждом опыте n. К сожалению, это – неформализуемый этап. Чем больше опытов выполнено, тем с большей точностью сможем прогнозировать закономерность изменения параметра оптимизации. С другой стороны, увеличение количества опытов увеличивает материальные затраты на проведение эксперимента. Поэтому необходим некоторый компромисс в выборе N, который составляет $N > 2k$, где k – общее коэффициентов уравнения регрессии в однофакторном уравнении регрессии.

Существует математическая теория построения оптимальных планов эксперимента, и для начального этапа стохастического моделирования предлагается использовать равномерный симметричный план (РСП).

РСП состоит из равноотстоящих друг от друга N уровней фактора x (N опытов), значения которых рассчитываются следующим образом:

$$x_j = x_{\min} + \frac{j-1}{N-1}\left(x_{\max} - x_{\min}\right), \quad j = 1, \ldots, N. \tag{51}$$

Так как параметр оптимизации имеет стохастическую природу, то для повышения точности его определения необходимо дублирование каждого опыта. Для решения прикладных задач желательно выбрать $n \geq 4$. Кроме того, будем рассматривать только такие планы эксперимента, в которых число дублей в каждом опыте одинаково.

В общем случае матрица планирования эксперимента должна содержать столбцы с номером опыта j и со значениями уровней главного фактора в натуральных координатах x_j, результаты опытов Y_{ji} (параметр оптимизации в j-м опыте, $j = 1, \ldots, N$, i-м дубле,

i = 1, …, n), а также результаты предварительной обработки экспериментальных данных $\overline{Y}_j$, S_j^2, проверки однородности выборочных дисперсий по критерию Кохрена G, расчета дисперсии воспроизводимости $S^2_{\text{воспр}}$ и ее числа степеней свободы $f_{\text{воспр}}$.

В качестве примера приведена матрица планирования для РСП, в котором N = 5, n = 4, $x_{\min}$ = 60 °С, $x_{\max}$ = 120 °С (табл. 3).

Таблица 3

Матрица планирования эксперимента

j	x_{1j}	Y_{1j}	Y_{2j}	Y_{3j}	Y_{4j}	$\overline{Y}_j$	S_j^2
1	60	12,2	11,2	11,7	13,3	12,10	0,807
2	75	14,0	14,6	14,4	16,6	14,90	1,347
3	90	16,8	14,5	17,1	15,2	15,90	1,567
4	105	18,4	17,4	18,1	19,3	18,30	0,620
5	120	18,8	20,3	19,4	17,5	19,00	1,380
$G_{э}$ = 0,274 $G_{3;\,5;\,0,95}$ = 0,598 Выборочные дисперсии однородны						$S^2_{\text{воспр}}$ = 1,144 $f_{\text{воспр}}$ = 15	

Проведение эксперимента. При проведении j-го опыта неконтролируемые случайные факторы Z вызывают непредсказуемые отклонения результатов Y_{ji}, получаемых при фиксированном значении главного фактора x_j (см. рис. 3, *б*). Часто в условиях эксперимента могут возникать неконтролируемые факторы, связанные со временем. Например, температура и влажность нагреваемого в сушилке воздуха могут меняться в зависимости от времени суток. Если, например, все 4 дубля 1-го опыта выполнялись рано утром, а все 4 дубля 2-го опыта – в полдень, то различие температуры и влажности нагреваемого воздуха накладываются на эффект, связанный с изменением управляющего фактора x. Поэтому для устранения возникновения возможных неконтролируемых связей с главным фактором осуществляется рандомизация последовательности проведения опытов (Y_{ji}, j = 1, …, N; i = 1, …, n). Суть ее в случайном порядке выполнения всех 20-ти экспериментов матрицы планирования ($N\,n$ = 20).

3.2.2. Предварительная обработка экспериментальных данных

Предварительная обработка результатов экспериментальных данных необходима для корректного моделирования и оптимизации технической системы.

Предварительная обработка экспериментальных данных методами математической статистики включает в себя следующие операции:

– расчет выборочного среднего $\overline{Y}_j$ в каждом опыте:

$$\overline{Y}_j = \frac{Y_{j1} + \ldots + Y_{jn}}{n} = \frac{\sum_{i=1}^{n} Y_{ji}}{n}, \quad j = 1, \ldots, N; \tag{52}$$

– расчет выборочной дисперсии S_j^2 и ее числа степеней свободы в каждом опыте:

$$S_j^2 = \frac{(Y_{j1} - \overline{Y}_j)^2 + \ldots + (Y_{jn} - \overline{Y}_j)^2}{n-1} = \frac{\sum_{i=1}^{n} (Y_{ji} - \overline{Y}_j)^2}{n-1}, \quad j = 1, \ldots, N; \tag{53}$$

$$f_{S_j} = n - 1, \quad j = 1, \ldots, N; \tag{54}$$

– проверка выборочных дисперсий всех опытов на однородность по критерию Кохрена, который применяется при условии равенства числа дублей n в каждом опыте. Алгоритм проверки N выборочных дисперсий с n-дублями в каждой выборке на однородность включает следующие операции:

а) расчет экспериментального значения критерия Кохрена:

$$G_э = \frac{\max S_j^2}{\sum_{j=1}^{N} S_j^2}; \tag{55}$$

б) определение критического значения критерия Кохрена $G_{f_1; f_2; p}$, в котором на первом месте – число степеней свободы максимальной дисперсии $f_1 = n - 1$, на втором – число степеней свободы для суммы всех выборочных дисперсий $f_2 = N$ при доверительной вероятности $p = 0{,}95$ (прилож. 5);

в) если

$$G_э < G_{n-1;\,N;\,0,95}, \tag{56}$$

то все выборочные дисперсии однородны,

если

$$G_э \geq G_{n-1;\,N;\,0,95}, \tag{57}$$

то все выборочные дисперсии неоднородны.

В этом случае для дальнейшей корректной обработки экспериментальных данных необходимо повторить опыт, в котором обнаружено максимальное значение дисперсии. Если однородность выборочных дисперсий не достигается, то дальнейшее моделирование по изложенной методике не рассматривается;

если все выборочные дисперсии однородны, то есть $\sigma_1 = \sigma_2 = \ldots = \sigma_N$, то в этом случае можно рассчитать дисперсию воспроизводимости $S^2_{воспр}$ и ее число степеней свободы $f_{воспр}$, характеризующие точность эксперимента в целом:

$$S^2_{воспр} = \frac{\sum_{j=1}^{n} S_j^2}{N}, \tag{58}$$

$$f_{воспр} = N(n-1). \tag{59}$$

Дисперсия воспроизводимости играет ключевую роль в стохастическом моделировании, так как она отражает культуру и точность эксперимента. Следует отметить, что статистическая значимость любой выборочной дисперсии определяется ее числом степеней свободы. Поэтому статистическая значимость дисперсии воспроизводимости $S^2_{воспр}$ существенно больше статистической значимости выборочной дисперсии в каждом опыте S_j^2, так как $N(n-1) > (n-1)$.

Результаты предварительной обработки экспериментальных данных, выполненной по уравнениям (52)–(59), представлены в табл. 3.

3.2.3. Построение ортогонализированного уравнения регрессии 1-го порядка и проверка его на статистическое качество

Перевод натуральных значений фактора в нормированные. Следует отметить, что стохастическая модель в форме уравнения (49) имеет ряд недостатков. Один из них связан с тем, что использование натуральных значений факторов усложняет процедуру расчета для построения регрессионных моделей. При повышении степени уравнения регрессии (49) задача нахождения его коэффициентов становится достаточно громоздкой. Например, если в некотором опыте температура сушильного агента $x = 95$ °С, тогда $x^3 = 857\,375$ °С^3, из чего виден принципиально различный масштаб приведенных факторов, что потребует вычислений с повышенной точностью коэффициентов регрессии при высокостепенных факторах. Кроме того, задача усложняется, когда исследуемая техническая система управляется несколькими факторами, имеющими различную размерность.

Для устранения недостатков, связанных с различным масштабом и размерностью фактора в натуральных координатах x, применяется операция нормирования фактора. Переход от натуральных значений фактора x к нормированным X и обратно осуществляется по формулам:

$$X = \frac{x - x_0}{\Delta x}, \tag{60}$$

$$x_0 = \frac{x_{max} + x_{min}}{2}, \tag{61}$$

$$\Delta x = \frac{x_{max} - x_{min}}{2}, \tag{62}$$

$$x = x_0 + X \Delta x. \tag{63}$$

Уравнения (60)–(62) переводят значения натурального фактора $x \in [x_{min}, x_{max}]$ в нормированные значения фактора $X \in [-1, +1]$. Уравнения (61)–(63) переводят значения нормированного фактора $X \in [-1, +1]$ в натуральные значения фактора $x \in [x_{min}, x_{max}]$.

Преимущество нормированных значений фактора X заключается в том, что они являются безразмерными и по модулю меньше единицы при любых натуральных значениях фактора x. В дальнейшем все процедуры стохастического моделирования будут проводиться с фактором X, представляемым в нормированных координатах. Поэтому уравнение регрессии (49) после преобразования натурального фактора x в нормированный X будет иметь следующий вид:

$$Y = b_0 X_0 + b_1 X + b_2 X^2 + b_3 X^3 + ... + b_m X^m . \qquad (64)$$

В уравнении (64) представлены факторы трех типов: главный X, фиктивный X_0 и производные факторы $X^2, X^3, ...$

В уравнение (64) из соображений подобия для регрессионного коэффициента b_0 введен фиктивный фактор X_0, все значения которого $X_0 = 1$.

Переход от натуральных значений фактора x к нормированным X и обратно приведен в табл. 4.

Таблица 4

Переход от натуральных значений фактора x к нормированным X и обратно

Характеристики фактора	Натуральные значения фактора x	Нормированные значения фактора X
Основной уровень, x_0	90 °C	0
Интервал варьирования, Δx	30 °C	1
Верхний уровень, x_{max}	120 °C	+1
Нижний уровень, x_{min}	60 °C	–1
Формулы перехода	$X = \frac{x - 90}{30}$; $x = 90 + 30X$	

Для проведения экспериментов и предварительной обработки экспериментальных данных, была создана матрица планирования, на базе которой была выполнена предварительная обработка экспериментальных данных (блок 2) (см. табл. 4).

Построение матрицы моделирования. Следующий этап моделирования технической системы связан с построением матрицы моделирования, которая включает следующую информацию: столбец с номером опыта j, столбцы ортогональных нормированных факторов X_{0j}, X_j; выборочные средние в каждом опыте $\overline{Y}_j$, вспомогательные столбцы для расчета коэффициентов уравнения регрессии

$X_0\overline{Y_j}$, $X\overline{Y_j}$, столбцы для проверки уравнения регрессии на адекватность Y_j^p, $(\overline{Y_j} - Y_j^p)^2$ (табл. 5), а также результаты статистических расчетов.

Таблица 5

Матрица моделирования для построения однофакторного ортогонализированного уравнения регрессии 1-го порядка на базе РСП

<table>
<tr><td>j</td><td>X_{0j}</td><td>X_j</td><td>$\overline{Y_j}$</td><td>$X_0\overline{Y_j}$</td><td>$X\overline{Y_j}$</td><td>Y_j^p</td><td>$(\overline{Y_j} - Y_j^p)^2$</td></tr>
<tr><td>1</td><td>1</td><td>–1,0</td><td>12,10</td><td>12,10</td><td>–12,10</td><td>12,60</td><td>0,2500</td></tr>
<tr><td>2</td><td>1</td><td>–0,5</td><td>14,90</td><td>14,90</td><td>–7,45</td><td>14,32</td><td>0,3364</td></tr>
<tr><td>3</td><td>1</td><td>0</td><td>15,90</td><td>15,90</td><td>0</td><td>16,04</td><td>0,0196</td></tr>
<tr><td>4</td><td>1</td><td>0,5</td><td>18,30</td><td>18,30</td><td>9,15</td><td>17,76</td><td>0,2916</td></tr>
<tr><td>5</td><td>1</td><td>1,0</td><td>19,00</td><td>19,00</td><td>19,00</td><td>19,48</td><td>0,2304</td></tr>
<tr><td colspan="2">$N = 5$</td><td colspan="2">$\sum_{j=1}^{5} X_{0j}^2 = 5$</td><td colspan="2">$\sum_{j=1}^{5} X_{0j}\overline{Y_j} = 80,2$</td><td></td><td>$\varphi = 3,487$</td></tr>
<tr><td colspan="2">$n = 4$</td><td colspan="2">$\sum_{j=1}^{5} X_j^2 = 2,5$</td><td colspan="2">$\sum_{j=1}^{5} X_j\overline{Y_j} = 8,250$</td><td colspan="2">$S_{\text{ад}}^2 = 1,504$,
$f = 3$</td></tr>
<tr><td colspan="2"></td><td colspan="2">$b_0 = 16,04$</td><td colspan="2">$b_1 = 3,4$</td><td colspan="2" rowspan="4">$F_{\text{э}} = 1,315$;
$F_{3;\ 15:\ 0,95} = 3,287$
Модель адекватна
$Y = 16,04 + 3,4X$</td></tr>
<tr><td colspan="2" rowspan="2">$S_{\text{воспр}}^2 = 1,144$
$f_{\text{воспр}} = 15$</td><td colspan="2">$S^2(b_0) = 0,05720$</td><td colspan="2">$S^2(b_1) = 0,1144$</td></tr>
<tr><td colspan="2">$S(b_0) = 0,2392$</td><td colspan="2">$S(b_1) = 0,3382$</td></tr>
<tr><td colspan="2">$t_{15;\ 0,95} = 2,131$</td><td colspan="2">$\Delta b_0 = 0,51$
b_0 – значим</td><td colspan="2">$\Delta b_1 = 0,7$
b_1 – значим</td></tr>
</table>

Расчет коэффициентов уравнения регрессии методом наименьших квадратов. Построить уравнение (50) можно графически, что эквивалентно определению коэффициентов b_0, b_1. Для этого через «облако» экспериментальных точек произвольно проведем прямую (рис. 5). Очевидно, что построение прямой таким способом неоднозначно в случае $N > k$, где N – число опытов, k – число коэффициентов.

Для однозначного построения уравнения регрессии (50) по экспериментальным точкам, задаваемого коэффициентами b_0 и b_1, необходим критерий, по которому можно выбрать наилучшее решение из всех возможных. Для решения этой задачи наибольшее распространение получил метод наименьших квадратов (МНК).

Суть МНК заключается в том, что наилучшим уравнением регрессии является уравнение регрессии, для которого остаточная сумма квадратов разности расчетных и экспериментальных значений $\varphi\,(b_0, b_1)$ будет минимальна:

$$\varphi\left(b_0, b_1\right) = \sum_{j=1}^{N}\left(Y_j^p - \overline{Y}_j\right)^2 = \sum_{j=1}^{N}\left(b_0 X_{0j} + b_1 X_j - \overline{Y}_j\right)^2 \to \min . \quad (65)$$

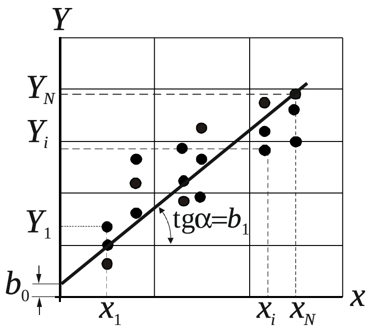


Рис. 5. Геометрический образ метода наименьших квадратов

Из математического анализа известно, что необходимым условием минимума функции $\varphi\,(b_0, b_1)$ является равенство нулю ее частных производных:

$$\begin{cases} \varphi'_{b_0} = \sum_{j=1}^{N} 2\left(b_0 X_{0j} + b_1 X_j - \overline{Y}_j\right) X_{0j} = 0; \\ \varphi'_{b_1} = \sum_{j=1}^{N} 2\left(b_0 X_{0j} + b_1 X_j - \overline{Y}_j\right) X_j = 0. \end{cases} \quad (66)$$

После преобразования система (66) приобретает следующий вид:

$$\begin{cases} b_0 \sum_{j=1}^{N} X_{0j}^2 + b_1 \sum_{j=1}^{N} X_{0j} X_j = \sum_{j=1}^{N} X_{0j} \overline{Y}_j ; \\ b_0 \sum_{j=1}^{N} X_{0j} X_j + b_1 \sum_{j=1}^{N} X_j^2 = \sum_{j=1}^{N} X_j \overline{Y}_j . \end{cases} \quad (67)$$

Система уравнений (67) достаточна для нахождения коэффициентов уравнения регрессии 1-го порядка b_0 и b_1 (два уравнения с двумя неизвестными).

Решение системы уравнений (67), полученной МНК, значительно упростилось бы, если бы

$$\sum_{j=1}^{N} X_{0j} X_j = 0 . \tag{68}$$

В этом случае система уравнений приобретет диагональный вид:

$$\begin{cases} b_0 \sum_{j=1}^{N} X_{0j}^2 + 0 = \sum_{j=1}^{N} X_{0j} \overline{Y}_j , \\ 0 + b_1 \sum_{j=1}^{N} X_j^2 = \sum_{j=1}^{N} X_j \overline{Y}_j , \end{cases} \tag{69}$$

позволяющий рассчитать коэффициенты b_0 и b_1 по простым формулам:

$$b_0 = \frac{\sum_{j=1}^{N} X_{0j} \overline{Y}_j}{\sum_{j=1}^{N} X_{0j}^2}, \tag{70}$$

$$b_1 = \frac{\sum_{j=1}^{N} X_j \overline{Y}_j}{\sum_{j=1}^{N} X_j^2} . \tag{71}$$

Свойство $\sum_{j=1}^{N} X_{0j} X_j = 0$ для двух факторов X_0 и X называется ортогональностью этих факторов, и оно позволяет достаточно просто рассчитать регрессионные коэффициенты b_0 и b_1 по уравнениям (70), (71). Свойство ортогональности обеспечивает независимость оценок коэффициентов уравнения регрессии и значительно упрощает его интерпретацию. В связи с этим отметим, что фактор X_0 необходим для проверки всех факторов уравнения регрессии на ортогональность.

В дальнейшем для построения уравнений регрессии будем рассматривать только ортогональные планы, для которых выполняется условие (68) и которые приводят систему уравнений (67) к диагональному виду (уравнения (69)).

Добиться выполнения условия ортогональности, задаваемого уравнением (68), для факторов X_0 и X в этом случае можно выбором соответствующих уровней главного управляющего фактора. В общем случае планирование эксперимента – один из важнейших элементов математической теории эксперимента, под которым подразумевается выбор факторов и их уровней для стохастического моделирования технических систем.

Для построения уравнения (50) можно предложить большое количество различных планов эксперимента. В табл. 6 приведены 4 плана различного типа.

Таблица 6

Примеры планов однофакторного эксперимента

N	X_{0j}	1-й план			2-й план			3-й план			4-й план		
		x_j	X_j	$X_{0j}X_j$	x_j	X_j	$X_{0j}X_j$	x_{1j}	X_j	$X_{0j}X_j$	x_{1j}	X_j	$X_{0j}X_j$
1	1	60	–1	–1	60	–1	–1	60	–1	–1	60	–1	–1
2	1	84	–0,2	–0,2	63	–0,9	–0,8	66	–0,8	–0,8	72	–0,6	–0,6
3	1	90	0	0	90	0	0	72	–0,6	–0,6	84	–0,2	–0,2
4	1	108	0,6	0,6	99	0,3	0,2	108	0,6	0,6	96	0,2	0,2
5	1	114	0,8	0,8	108	0,6	0,6	114	0,8	0,8	108	0,6	0,6
6	1	120	1	1	120	1	1	120	1	1	120	1	1
$\sum X$	**6**	–	**1,2**	**1,2**	–	**0**	**0**	–	**0**	**0**	–	**0**	**0**

В 1-м плане уровни нормированного фактора X_j неравномерны, несимметричны относительно центра плана и $\sum_{j=1}^{N} X_j \neq 0$. Факторы X_0 и X_j не ортогональны, так как $\sum_{j=1}^{N} X_{0j} X_j = 1,20 \neq 0$. Система уравнений (67) для такого плана не сводится к диагональному виду. В этом случае для нахождения коэффициентов регрессии b_0 и b_1 необходимо решить систему двух линейных уравнений с двумя неизвестными.

Во 2-м плане уровни фактора X_j подобраны так, что $\sum_{j=1}^{N} X_j = 0$, хотя уровни этого фактора неравномерны и несимметричны относительно центра плана. В этом случае факторы X_0 и X_j ортогональны, так как $\sum_{j=1}^{N} X_{0j} X_j = 0$, и поэтому система линейных уравнений (67) будет иметь диагональный вид.

В 3-м плане уровни нормированного фактора X_j симметричны относительно центра плана, вследствие чего условие $\sum_{j=1}^{N} X_j = 0$ выполняется автоматически, хотя уровни нормированного фактора X_j неравномерны. В этом случае факторы X_0 и X_j также ортогональны, так как $\sum_{j=1}^{N} X_{0j} X_j = 0$, и потому система уравнений (67) будет также иметь диагональный вид.

В 4-м равномерном симметричном плане (РСП) уровни нормированного фактора X_{1j} и равномерны, и симметричны относительно центра плана. Как и в 3-м плане, условие $\sum_{j=1}^{N} X_j = 0$ выполняется автоматически. В этом случае ортогональность факторов X_0 и X_{1j} очевидна $\sum_{j=1}^{N} X_{0j} X_j = 0$. Система уравнений (67) для РСП будет иметь диагональный вид; коэффициенты регрессии b_0 и b_1 рассчитываются по уравнениям (70), (71).

С точки зрения требования ортогональности планы 2-го, 3-го и 4-го типов равноценны, однако 4-й план (РСП) является предпочтительным, если отсутствует априорная информация об изучаемой технической системе. Предпочтительность плана 4-го типа обусловлена тем, что он имеет однозначный алгоритм построения. К тому же равномерное «покрытие» изучаемого интервала варьирования фактора повышает вероятность реализации опытов вблизи его оптимального значения.

В дальнейшем для построения матрицы планирования однофакторных уравнений регрессии любого порядка будет использована только РСП, натуральные значения которого рассчитываются по формуле (51).

Так как уравнение регрессии строится всегда в нормированных координатах, то натуральные значения фактора x_j, задаваемые для РСП уравнением (51), следует преобразовать в нормированные, используя уравнение (60). В результате чего для любого РСП справедлива следующая формула:

$$X_j = -1 + \frac{2(j-1)}{N-1}, \quad j = 1, \dots, N. \tag{72}$$

Обратите внимание! X_j в РСП не зависит ни от $x_{\max}$, ни от $x_{\min}$, а только от числа опытов N (вывод уравнения (72) приведен в прилож. 6 (П6.22)).

$$\sum_{j=1}^{N} X_{0j} X_j = \sum_{j=1}^{N} X_j = 0. \tag{73}$$

Ортогональность факторов X_{0j} и X_j позволяет рассчитать коэффициентов уравнения регрессии 1-го порядка для РСП по уравнениям (70), (71).

Проверка коэффициентов уравнения регрессии на значимость по критерию Стьюдента. Коэффициенты уравнения регрессии, рассчитываемые на основе экспериментальных данных, имеют случайный характер. Поэтому для оценки их достоверности важны не только значения коэффициентов, но и доверительный интервал, задаваемый абсолютной погрешностью. Эта процедура реализуется при проверке коэффициентов регрессии на значимость.

Дисперсии значимости коэффициентов уравнения регрессии 1-го порядка, но только при условии ортогональности факторов, определяют по следующему алгоритму:

– расчет дисперсии значимости коэффициентов уравнения регрессии:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} X^2_{0j}}, \tag{74}$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n\sum\limits_{j=1}^{N} X_j^2}; \quad (75)$$

– расчет доверительного интервала для каждого коэффициента уравнения регрессии:

$$\Delta b_0 = t_{f_{\text{воспр}};\, p} S(b_0), \quad (76)$$

$$\Delta b_1 = t_{N(n-1);\, p} S(b_1), \quad (77)$$

где $t_{f_{\text{воспр}};\, p}$ – критическое значение критерия Стьюдента с числом степеней свободы $f_{\text{воспр}} = N(n-1)$ при доверительной вероятности p = 0,95 (прилож. 2);

если

$$\Delta b_0 < |b_0|, \quad (78)$$

$$\Delta b_1 < |b_1|, \quad (79)$$

то коэффициенты регрессии b_0 и b_1 значимы,

если условия (78) и (79) не выполняются, то соответствующие коэффициенты незначимы.

Незначимые коэффициенты следует исключить из уравнения регрессии. Количество оставшихся значимых коэффициентов регрессии равно B, причем $B \leq k$.

Проверка уравнения регрессии на адекватность по критерию Фишера. Процедура проверки адекватности направлена на установление соответствия расчетных значений параметра оптимизации по уравнению регрессии и экспериментальных данных. Уравнение регрессии со значимыми коэффициентами позволяет рассчитать значение параметра оптимизации Y_j^p в каждом опыте. Проверка уравнения регрессии на адекватность заключается в проверке степени совпадения экспериментальных и расчетных значений параметра оптимизации. Степень рассеяния экспериментальных значений параметра оптимизации характеризуется дисперсией воспроизводимости $S^2_{\text{воспр}}$

с числом степеней свободы $f_{\text{воспр}} = N(n-1)$. Степень несовпадения экспериментальных и расчетных значений параметра оптимизации характеризуется дисперсией адекватности $S^2_{\text{ад}}$ с числом степеней свободы $f_{\text{ад}} = N - B$. Полученное уравнение регрессии является адекватным, если дисперсии $S^2_{\text{воспр}}$ и $S^2_{\text{ад}}$ примерно одинаковы, что математически означает однородность по критерию Фишера.

Алгоритм проверки полученного уравнения регрессии 1-го порядка на адекватность включает следующие этапы:

– расчет значения параметра оптимизации Y_j^p по уравнению регрессии 1-го порядка в каждом опыте: $Y_j^p = b_0 + b_1 X_j$, $j = 1, \ldots, N$;

– расчет квадрата разности между средним значением $\overline{Y}_j$ и расчетным значением Y_j^p в каждом опыте $\left(\overline{Y}_j - Y_j^p\right)^2$, $j = 1, \ldots, N$;

– расчет остаточной суммы квадратов:

$$\varphi = \sum_{j=1}^{N} \left(\overline{Y}_j - Y_j^p\right)^2; \tag{80}$$

– расчет дисперсии адекватности $S^2_{\text{ад}}$ и ее число степеней свободы:

$$S^2_{\text{ад}} = \frac{n\varphi}{N - B}, \tag{81}$$

$$f_{\text{ад}} = N - B, \tag{82}$$

где B – число значимых коэффициентов уравнения регрессии;

– расчет экспериментального значения критерия Фишера:

$$F_{\text{э}} = \frac{\max\left(S^2_{\text{ад}}, S^2_{\text{воспр}}\right)}{\min\left(S^2_{\text{ад}}, S^2_{\text{воспр}}\right)}; \tag{83}$$

– определение критического значения критерия Фишера $F_{f_1;\, f_2;\, p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии f_1, на втором – число степеней свободы *меньшей* дисперсии f_2 при доверительной вероятности p = 0,95 (прилож. 4);

если

$$F_э < F_{f_1;\, f_2;\, p}\,, \tag{84}$$

то выборочные дисперсии $S^2_{\text{воспр}}$ и $S^2_{\text{ад}}$ однородны, следовательно, полученное уравнение регрессии 1-го порядка со значимыми коэффициентами адекватно;

если

$$F_э \geq F_{f_1;\, f_2;\, p}\,, \tag{85}$$

то выборочные дисперсии $S^2_{\text{воспр}}$ и $S^2_{\text{ад}}$ неоднородны, следовательно, полученное уравнение регрессии со значимыми коэффициентами не адекватно. В этом случае необходимо перейти к уравнению регрессии 2-го порядка.

Результаты расчетов для построения уравнения регрессии 1-го порядка представлены в виде матрицы моделирования (табл. 7).

Таблица 7

Матрица моделирования и результаты обработки данных
для построения уравнения регрессии 1-го порядка

j	X_{0j}	X_j	$\bar{Y}_j$	$X_{0j}\bar{Y}_j$	$X_j\bar{Y}_j$	Y_j^p	$\left(\bar{Y}_j - Y_j^p\right)^2$
1	1	–1	$\bar{Y}_1$	$X_{01}\bar{Y}_1$	$X_1\bar{Y}_1$	Y_1^p	$\left(\bar{Y}_1 - Y_1^p\right)^2$
…	…	…	…	…	…	…	…
N	1	1	$\bar{Y}_N$	$X_{0j}\bar{Y}_N$	$X_N\bar{Y}_N$	Y_N^p	$\left(\bar{Y}_N - Y_N^p\right)^2$
N =				$\sum\limits_{j=1}^{N} X_{0j}^2 =$	$\sum\limits_{j=1}^{N} X_{0j}\bar{Y}_j =$	φ =	
n =				$\sum\limits_{j=1}^{N} X_{1j}^2 =$	$\sum\limits_{j=1}^{N} X_{1j}\bar{Y}_j =$	$S^2_{\text{ад}} =$ $f_{\text{ад}} =$	
				b_0 =	b_1 =	$F_э$ =	

Окончание таблицы 7

j	X_{0j}	X_j	$\bar{Y}_j$	$X_{0j}\bar{Y}_j$	$X_j\bar{Y}_j$	Y_j^p	$\left(\bar{Y}_j - Y_j^p\right)^2$
$S^2_{\text{воспр}} =$				$S^2(b_0) =$	$S^2(b_1) =$	$F_{f_1;\, f_2,\, 0{,}95} =$	
$f_{\text{воспр}} =$				$S(b_0) =$	$S(b_1) =$	Модель (не) адекватна	
$t_{N(n-1);\, 0{,}95} =$				$\Delta b_0 =$	$\Delta b_1 =$	$Y = b_0 + b_1X$	

Анализ уравнения (81) показывает, что минимальное количество опытов N, необходимое для проверки однофакторного уравнения регрессии на адекватность, должно быть больше числа значимых коэффициентов в уравнении регрессии. С целью повышения надежности проверки полученного уравнения регрессии на адекватность рекомендуем придерживаться следующего правила: $N > 2k$.

Поэтому при моделировании исследуемой технической системы однофакторным уравнением регрессии 1-го порядка количество опытов целесообразно брать не менее пяти ($k = 2$), при моделировании однофакторным уравнением 2-го порядка – не менее семи ($k = 3$).

3.2.4. Оптимизация однофакторной технической системы по уравнению регрессии 1-го порядка

Уравнение регрессии 1-го порядка позволяет рассчитать значение параметра оптимизации и его абсолютной погрешности в любой точке факторного пространства (компьютерный эксперимент). Оптимизация технической системы осуществима только при следующих условиях:

1) выборочные дисперсии всех опытов однородны по критерию Кохрена;

2) все коэффициенты, включенные в уравнение регрессии, статистически значимы по критерию Стьюдента;

3) полученное уравнение регрессии адекватно по критерию Фишера;

4) расчет осуществляется только в экспериментально исследованном интервале варьирования изученного факторного пространства, в котором уравнение регрессии адекватно. Оптимизация функционирования изучаемой технической системы в некоторой точке факторного отрезка выполняется по следующему алгоритму:

5) перевести заданное натуральное значение фактора x в нормированное X по уравнению (60):

$$X = \frac{x - x_0}{\Delta x};$$

– рассчитать значение параметра оптимизации по уравнению (50):

$$Y = b_0 + b_1 X \ ;$$

– рассчитать абсолютную погрешность параметра оптимизации для уравнения регрессии 1-го порядка:

$$\Delta Y(X) = t_{N(n-1);\, p} \sqrt{S^2(b_0) + X^2 S^2(b_1)} \ . \qquad (86)$$

Анализ уравнения (86) позволяет для РСП сделать несколько выводов:

а) минимальная абсолютная погрешность достигается в центре факторного пространства $X = 0$:

$$\Delta Y(0) = t_{N(n-1);\, p} \sqrt{S^2(b_0)} = t_{N(n-1);\, p} S(b_0) = \Delta b_0 \ ; \qquad (87)$$

б) максимальная абсолютная погрешность достигается на границах интервала варьирования фактора при $X = \pm 1$:

$$\begin{aligned} \Delta Y(\pm 1) &= t_{N(n-1);\, p} \sqrt{S^2(b_0) + S^2(b_1)} = \\ &= \sqrt{t^2_{N(n-1);\, p} \left[S^2(b_0) + S^2(b_1) \right]} = \sqrt{\Delta b_0^2 + \Delta b_1^2} \ . \end{aligned} \qquad (88)$$

Абсолютная погрешность параметра оптимизации Y, рассчитанного по адекватному уравнению регрессии 1-го порядка, значительно меньше абсолютной погрешности, найденной по экспериментальным данным в каждом опыте, что возможно обосновать количественно.

Так как выборочные дисперсии в каждом опыте однородны по критерию Кохрена, то их среднее значение равно $S^2_{\text{воспр}}$.

Отношение экспериментальных и расчетных абсолютных погрешностей в центре плана ($X = 0$), согласно формуле (87), равно:

$$\frac{\Delta \overline{Y}_j}{\Delta Y(0)} = \frac{t_{n-1,\,p} S_j}{t_{N(n-1),\,p} S(b_0)\sqrt{n}} \approx \frac{t_{n-1,\,p} S_j \sqrt{nN}}{t_{N(n-1),\,p} S_{\text{воспр}} \sqrt{n}}\,. \tag{89}$$

Например, для эксперимента $N = 10$, $n = 4$, $t_{30,\,0,95} = 2{,}042$, $t_{3;\,0,95} =$ $= 3{,}182$ (прилож. 2), из уравнения (89) и условия $\overline{S_j} \approx S_{\text{воспр}}$ следует, что в центре плана

$$\frac{\Delta \overline{Y}_j}{\Delta Y(0)} \approx \frac{3{,}182\sqrt{10}}{2{,}042} = 4{,}9 \approx 5\,. \tag{90}$$

В результате предельная абсолютная погрешность рассчитанного параметра Y в центре РСП примерно в 5 раз меньше предельной абсолютной погрешности средних значений параметров, найденных экспериментально в каждом опыте.

Таким образом, математическая модель технической системы, построенная по экспериментальным данным (N опытов, n дублей) позволяет рассчитать значение параметра Y по уравнению регрессии точнее, чем его экспериментальное значение в каждом опыте, на базе которых построено само уравнение регрессии. Это одно из достоинств, характеризующих всю мощь и красоту регрессионного анализа.

Оптимизация технической системы по уравнению регрессии. Задача оптимизации представляет особый класс математических задач. Целью оптимизации является поиск комплекса факторов, который обеспечит оптимальное (наилучшее) функционирование технической системы (параметр оптимизации должен иметь минимальное или максимальное значение). В этом случае критерий оптимизации математически задается целевой функцией следующего вида:

$$Y(X) \to \max \wedge \min\,. \tag{91}$$

В зависимости от вида целевой функции, задаваемой уравнением регрессии, возможны 2 вида экстремумов: локальный и граничный. Локальный максимум находится из условий: $Y'(X) = 0$ (необходимое условие) и $Y''(X) < 0$ (достаточное условие). Локальный минимум находится из условий: $Y'(X) = 0$ (необходимое условие) и $Y''(X) > 0$

(достаточное условие). Причем локальный экстремум, найденный по уравнению регрессии, должен находиться внутри изученного факторного пространства, в котором уравнение регрессии адекватно. Локальные экстремумы вне факторного изученного пространства не рассматриваются. Граничные экстремумы всегда находятся на границе изученного факторного пространства.

Для уравнения регрессии 1-го порядка (это всегда прямая линия) экстремальные значения параметра оптимизации всегда являются граничными, то есть достигаются на границах изученного факторного отрезка: в нормированных координатах $X_{\text{опт}} = -1$ или $X_{\text{опт}} = +1$, что соответствует $x_{\min}$, или $x_{\max}$ в натуральных координатах (см. табл. 7). В этом случае оптимальное значение параметра оптимизации

$$Y_{\text{опт}} = b_0 + b_1 X_{\text{опт}} . \tag{92}$$

Абсолютная погрешность определения оптимального параметра рассчитывается по уравнению (86):

$$\Delta Y_{\text{опт}} = t_{N(n-1),\, p} \sqrt{S^2(b_0) + X^2_{\text{опт}} S^2(b_1)} = \sqrt{\Delta b_0^2 + X^2_{\text{опт}} \Delta b_1^2} . \tag{93}$$

3.2.5. Алгоритм решения прикладных задач с помощью однофакторных уравнений регрессии 1-го порядка

1. Выбрать объект моделирования – техническую систему – и на основе литературных и патентных источников собрать необходимую предварительную (априорную) информацию о параметрах и факторах, характеризующих функционирование моделируемой технической системы.

1.1. Выбрать параметр оптимизации *Y*, эффективно характеризующий функционирование технической системы.

1.2. Выбрать управляющий фактор *x*, существенно влияющий на параметр оптимизации, а также интервал его варьирования [$x_{\min}$, $x_{\max}$].

1.3. Установить математическую взаимосвязь между натуральными и нормированными значениями управляющего фактора.

1.4. Выбрать и обосновать степень алгебраического ортогонализированного полинома 1-го порядка в качестве уравнения регрессии:

$$Y = b_0 X_0 + b_1 X .$$

1.5. Выбрать число опытов N и число дублей n в каждом опыте. Рекомендуемое число опытов для полинома 1-го порядка $N > 2k$, где $k = 2$. Рекомендуемое число дублей $n \geq 4$.

1.6. Построить матрицу планирования эксперимента, включающую столбцы j, $x_j, Y_{j1}, ..., Y_{jn}, \overline{Y}_j, S_j^2$, для РСП, в котором натуральные значения фактора x_j для каждого опыта рассчитать по формуле

$$x_j = x_{\min} + \frac{j-1}{N-1}(x_{\max} - x_{\min}),\ j = 1, \ldots, N.$$

1.7. Обеспечить приборное и технологическое оснащение эксперимента, позволяющее измерять исследуемый параметр Y и варьируемый управляющий фактор x с требуемой точностью и провести эксперимент.

2. Провести предварительную обработку экспериментальных данных.

2.1. Расчет выборочного среднего в каждом опыте:

$$\overline{Y}_j = \frac{\sum_{i=1}^{n} Y_{ji}}{n},\ j = 1, \ldots, N.$$

2.2. Расчет выборочной дисперсии и ее числа степеней свободы в каждом опыте:

$$S_j^2 = \frac{\sum_{i=1}^{n}\left(Y_{ji} - \overline{Y}_j\right)^2}{n-1},\ j = 1, \ldots, N.$$

$$f_S = n - 1.$$

2.3. Проверка выборочных дисперсий на однородность по критерию Кохрена:

– расчет экспериментального значения критерия Кохрена $G_э$

$$G_э = \frac{\max S_j^2}{\sum_{j=1}^{N} S_j^2};$$

– определение критического значения критерия Кохрена $G_{f1;\,f2;\,p}$, в котором на первом месте стоит число степеней свободы максимальной дисперсии $f_1 = n - 1$, на втором – число степеней свободы для суммы всех выборочных дисперсий $f_2 = N$ при доверительной вероятности $p = 0{,}95$ (прилож. 5):

если

$$G_э < G_{n-1;\,N;\,0,95},$$

то все выборочные дисперсии однородны,

если

$$G_э \geq G_{n-1;\,N;\,0,95},$$

то все выборочные дисперсии неоднородны.

Внимание! Дальнейшая обработка экспериментальных данных возможна только в том случае, если дисперсии всех опытов однородны.

2.4. Расчет дисперсии воспроизводимости и ее числа степеней свободы:

$$S^2_{\text{воспр}} = \frac{\sum_{j=1}^{n} S_j^2}{N},$$

$$f_{\text{воспр}} = N(n-1).$$

3. Построение уравнения регрессии и проверка его на статистическое качество:

3.1. Построение матрицы моделирования, включающей столбцы j, X_{0j}, X_j, $\overline{Y_j}$, $X_{0j}\overline{Y_j}$, $X_j\overline{Y_j}$, Y_j^p, $(\overline{Y_j} - Y_j^p)^2$, где:

– фиктивный фактор $X_{0j} = 1$, $j = 1, \ldots, N$;

– нормированные значения фактора X_j для каждого опыта:

$$X_j = -1 + \frac{2(j-1)}{N-1}, \quad j = 1, \ldots, N.$$

3.2. Проверка факторов на ортогональность:

$$\sum_{j=1}^{N} X_{0j} X_j = 0.$$

3.3. Если выполнено условие ортогональности факторов, рассчитать коэффициенты уравнения регрессии 1-го порядка b_0, b_1:

$$b_0 = \frac{\sum_{j=1}^{N} X_{0j} \bar{Y}_j}{\sum_{j=1}^{N} X_{0j}^2},$$

$$b_1 = \frac{\sum_{j=1}^{N} X_j \bar{Y}_j}{\sum_{j=1}^{N} X_j^2}.$$

3.4. Проверка коэффициентов уравнения регрессии b_0, b_1 1-го порядка $Y = b_0 X_0 + b_1 X$ на значимость по критерию Стьюдента:

– расчет дисперсии значимости коэффициентов b_0, b_1:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{N} X_{0j}^2},$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{N} X_j^2};$$

– расчет доверительных интервалов для коэффициентов b_0, b_1:

$$\Delta b_0 = t_{N(n-1);\,0,95} \sqrt{S^2(b_0)} = t_{N(n-1);\,0,95} S(b_0),$$

$$\Delta b_1 = t_{N(n-1);\,0,95} \sqrt{S^2(b_1)} = t_{N(n-1);\,0,95} S(b_1),$$

где $t_{f\text{воспр};\, p}$ – критическое значение критерия Стьюдента с числом степеней свободы $f_{\text{воспр}} = N(n-1)$ при доверительной вероятности p = 0,95 (прилож. 2):

если

$$\Delta b_0 < |b_0|,$$

$$\Delta b_1 < |b_1|,$$

то коэффициенты регрессии b_0 и b_1 значимы,

если условия неравенства не выполняются, то соответствующие коэффициенты незначимы.

Незначимые коэффициенты уравнения регрессии следует исключить из уравнения. Количество оставшихся значимых коэффициентов уравнения регрессии равно B, причем $B \le k$.

3.5. Проверка полученного уравнения регрессии 1-го порядка $Y = b_0 X_0 + b_1 X$ на адекватность по критерию Фишера:

если

$$\Delta b_0 < |b_0|, \tag{94}$$

$$\Delta b_1 < |b_1|, \tag{95}$$

то коэффициенты регрессии b_0 и b_1 значимы,

если условия (94) и (95) не выполняются, то соответствующие коэффициенты незначимы;

– расчет в каждом опыте квадрата разности экспериментального и расчетного значения параметра оптимизации $(\overline{Y}_j - Y_j^p)^2$, $j = 1, \ldots, N$;

– расчет остаточной суммы квадратов:

$$\varphi = \sum_{j=1}^{N} \left(\overline{Y}_j - Y_j^p\right)^2;$$

– расчет дисперсии адекватности и ее число степеней свободы:

$$S_{\text{ад}}^2 = \frac{n\varphi}{N - B},$$

$$f_{\text{ад}} = N - B;$$

– расчет экспериментального значения критерия Фишера:

$$F_{\text{э}} = \frac{\max\left(S_{\text{ад}}^2, S_{\text{воспр}}^2\right)}{\min\left(S_{\text{ад}}^2, S_{\text{воспр}}^2\right)};$$

– определение критического значения критерия Фишера $F_{f_1;\,f_2;\,p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии f_1, на втором – число степеней свободы *меньшей* дисперсии f_2 при доверительной вероятности p = 0,95 (прилож. 4):

если

$$F_э < F_{f_1;\,f_2;\,p},$$

то выборочные дисперсии $S^2_{воспр}$ и $S^2_{ад}$ однородны, следовательно, полученное уравнение регрессии 1-го порядка со значимыми коэффициентами адекватно,

если

$$F_э \geq F_{f_1;\,f_2;\,p},$$

то выборочные дисперсии $S^2_{воспр}$ и $S^2_{ад}$ неоднородны, следовательно, полученное уравнение регрессии со значимыми коэффициентами не адекватно. В случае неадекватной модели следует перейти к построению уравнения регрессии 2-го порядка.

4. Оптимизация технической системы по адекватному уравнению регрессии, в котором все коэффициенты значимы:

– для нахождения максимального (минимального) значения параметра оптимизации рассчитать оптимальное значение управляющего фактора в нормированных и натуральных координатах:

– при $b_1 > 0$:

$$X_{опт} = +1, \quad (X_{опт} = -1),$$

$$x_{опт} = x_{max}, \quad (x_{опт} = x_{min});$$

– при $b_1 < 0$:

$$X_{опт} = -1, \quad (X_{опт} = +1),$$

$$x_{опт} = x_{min}, \quad (x_{опт} = x_{max});$$

– расчет оптимального значения параметра оптимизации:

$$Y_{опт} = b_0 + b_1 X_{опт} = b_0 \pm b_1;$$

– расчет абсолютной погрешности оптимального значения параметра оптимизации:

$$\Delta Y_{\text{опт}} = t_{N(n-1);\,0,95}\sqrt{S^2(b_0) + X_{\text{опт}}^2 S^2(b_1)} = \sqrt{\Delta b_0^2 + X_{\text{опт}}^2 \Delta b_1^2}\ .$$

3.2.6. Типовая задача: стохастическое моделирование технических систем уравнением регрессии 1-го порядка

Цель – исследование функционирования однофакторных технических систем с помощью уравнения регрессии 1-го порядка.

Проиллюстрируем моделирование и оптимизацию однофакторных технических систем уравнениями регрессии 1-го порядка на примере разработки процесса кормления цыплят с учетом использования витаминизированной добавки.

Математическая постановка задачи – построить математическую модель процесса кормления, проверить модель на статистическое качество и на ее базе выполнить оптимизацию режима кормления цыплят 2-недельного возраста.

1. Планирование и проведение эксперимента. В качестве параметра оптимизации выбрана масса Y (г) цыплят 2-недельного возраста.

В качестве управляющего фактора, влияющего на массу цыплят, выбрано содержание (г/сут) витаминизированной добавки в корме. По условиям эксперимента, интервал варьирования выбран в диапазоне $x \in [0, 12]$ г/сут, то есть $x_{\min} = 0$ г/сут, $x_{\max} = 12$ г/сут.

Моделирование процесса вскармливания начнем с уравнения регрессии 1-го порядка: $Y = b_0 X_0 + b_1 X$.

Для большей информативности при построении уравнения регрессии 1-го порядка ($k = 2$) выберем количество опытов ($N = 7$), число дублей ($n = 4$) в каждом опыте. При выборе количества опытов руководствуемся соотношением: $N > 2k$.

Для проведения опытов было организовано 7 контрольных вольеров (по числу опытов). В каждый вольер поместили по 200 цыплят суточного возраста.

В связи с тем, что конечная масса цыплят зависит от их стартовой массы, до проведения основного эксперимента была проверена гипотеза: различие средней массы цыплят в каждом вольере не существенно. Для этого из каждого вольера случайным образом взяли по 4 цыпленка суточного возраста и взвешиванием определили массу каждого. Результаты предварительной статистической обработки приведены в табл. 8.

Таблица 8

Результаты предварительной статистической обработки измеренной массы цыплят суточного возраста в семи вольерах

№ вольера	Y_{j1}, г	Y_{j2}, г	Y_{j3}, г	Y_{j4}, г	$\overline{Y_j}$, г	S_j^2
1	48,6	49,3	49,4	51,3	49,65	1,337
2	49,1	50,3	49,8	51,0	50,05	0,643
3	50,3	51,7	48,5	51,9	50,60	2,467
4	47,2	51,2	48,9	48,8	49,03	2,709
5	48,4	51,3	51,2	49,9	50,20	1,847
6	46,9	49,7	49,6	48,9	48,78	1,689
7	48,2	52,0	45,3	49,0	48,63	7,589
$G_э = 0{,}415$				$\sum_{j=1}^{7} S_j^2 = 18{,}28$		
$G_{3;\,7;\,0{,}95} = 0{,}480$				$S^2_{\text{воспр}} = 2{,}611$; $f_S = 21$		

Установлено, что выборочные дисперсии массы цыплят в каждом вольере однородны по критерию Кохрена (см. табл. 8). В связи с тем, что выборочные дисперсии однородны, можно рассчитать дисперсию воспроизводимости опытов, как среднюю всех выборочных дисперсий, и ее число степеней свободы

$$S^2_{\text{воспр}} = \frac{\sum_{j=1}^{N} S_j^2}{N} = \frac{18{,}28}{7} = 2{,}611;$$

$$f_{\text{воспр}} = N(n-1) = 7 \cdot (4-1) = 21.$$

Чтобы проверить, есть ли существенное различие средней массы цыплят в разных вольерах, выберем для сравнения вольеры № 3 и № 7, для которых различие средних масс цыплят наибольшее. Проверка оценки качества однотипной продукции по алгоритму (см. раздел 2.3.1) показывает, что по критерию Стьюдента разница средних значений средней массы цыплят в вольерах № 3 и № 7 несущественная (решите эту задачу самостоятельно).

Так как между наибольшей и наименьшей средней массой цыплят в вольерах нет существенного различия, а дисперсии их массы однородны, то можно утверждать, что с доверительной вероятностью p = 0,95 все 7 выборочных совокупностей принадлежат к одной генеральной совокупности.

$$\bar{Y} = \frac{\sum_{j=1}^{N} \bar{Y}_j}{N} = \frac{49,65 + 50,05 + 50,6 + 49,03 + 50,2 + 48,78 + 48,63}{7} = 49,56;$$

$$\Delta\bar{Y} = t_{N(n-1)}\sqrt{S^2(b_0)} = t_{N(n-1)}\sqrt{\frac{S^2_{\text{воспр}}}{nN}} = 2,080\sqrt{\frac{2,612}{28}} = 0,6353 \approx 0,6.$$

Таким образом, средняя стартовая масса цыплят равна $\bar{Y}$ = (49,6 ± 0,6) г (записано по алгоритму корректного оформления результата расчетов).

В качестве плана эксперимента для разработки процесса кормления цыплят с учетом влияния витаминизированной добавки выбран РСП, в котором значение управляющего фактора в каждом опыте (количество витаминизированной добавки в вольерах) рассчитывалось по уравнению

$$x_j = x_{\min} + \frac{j-1}{N-1}(x_{\max} - x_{\min}), \; j = 1, \ldots, 7.$$

Цыплята во всех вольерах с 2-дневного возраста получали одинаковый по количеству и качеству корм, за исключением витаминизированной добавки. В течение 14 дней в корм цыплят добавляли витаминизированную добавку в соответствии с планом эксперимента: в вольер № 1 – 0 г/сут, № 2 – 2 г/сут, …, № 7 – 12 г/сут (табл. 9).

В конце периода вскармливания (через 14 дней) из каждого вольера случайным образом выбрали по 4 цыпленка, определили их массу за вычетом средней стартовой массы, зафиксированной в начале эксперимента. Полученные результаты внесены в матрицу планирования (см. табл. 9).

Таблица 9

Матрица планирования для построения однофакторного ортогонализированного уравнения регрессии 1-го порядка на базе РСП с числом опытов N = 7

j	x_j, г/сут	Y_{j1}, г	Y_{j2} г	Y_{j3}, г	Y_{j4}, г	$\overline{Y}_j$, г	S_j^2
1	0	408	406	415	414	410,8	19,58
2	2	430	419	427	428	426,0	23,33
3	4	424	423	0427	419	423,3	10,92
4	6	443	437	432	439	437,8	20,92
5	8	443	452	451	453	449,8	20,92
6	10	445	440	448	447	445,0	12,67
7	12	464	454	456	458	458,0	18,67
$G_э$ = 0,184				$\sum_{j=1}^{7} S_j^2 = 127,0$			
$G_{3;\,7;\,0,95}$ = 0,480				$S^2_{\text{воспр}} = 18,14$; $f_{\text{воспр}}$ = 21			

2. Предварительная обработка экспериментальных данных. Результаты расчетов по мере их получения вносятся в матрицу планирования (см. табл. 9):

– выборочное среднее в каждом опыте:

$$\overline{Y}_j = \frac{\sum_{i=1}^{n} Y_{ji}}{n} = \frac{\sum_{i=1}^{4} Y_{ji}}{4}, \quad j = 1, \ldots, 7,$$

например: $\overline{Y}_1$ = (408 + 406 + 415 + 414) / 4 = 410,8;

– выборочная дисперсия и ее число степеней свободы в каждом опыте:

$$S_j^2 = \frac{\sum_{i=1}^{n} (Y_{ji} - \overline{Y}_j)^2}{n-1} = \frac{\sum_{i=1}^{4} (Y_{ji} - \overline{Y}_j)^2}{4-1}, \quad j = 1, \ldots, 7,$$

например: $S_1^2 = [(408 - 410,8)^2 + (406 - 410,8)^2 + (415 - 410,8)^2 + (414 - 410,8)^2] / (4 - 1) = 19,58$;

$$f_{S_j} = n - 1 = 4 - 1 = 3, \quad j = 1, \ldots, 7;$$

– проверка всех выборочных дисперсий на однородность по критерию Кохрена:

$$G_э = \frac{\max S_j^2}{\sum\limits_{j=1}^{7} S_j^2} = \frac{23,33}{11,7} = 0,209\,,$$

$$G_{кр} = G_{n-1,N,p} = G_{3;\,7;\,0,95} = 0,480 \text{ (прилож. 5)};$$

– так как $G_э < G_{кр}$, то выборочные дисперсии однородны, что дает право рассчитать дисперсию воспроизводимости и ее число степеней свободы:

$$S^2_{воспр} = \frac{\sum\limits_{j=1}^{n} S_j^2}{N} = \frac{111,7}{7} = 15,95\,;$$

$$f_{воспр} = N(n-1) = 7(4-1) = 21.$$

3. Построение уравнения регрессии 1-го порядка и проверка его на статистическое качество. Для расчета коэффициентов уравнения регрессии создадим матрицу моделирования для построения однофакторного ортогонализированного уравнения регрессии 1-го порядка $Y = b_0 X_0 + b_1 X$ с числом опытов N = 7. Результаты расчета по мере их получения вносятся в табл. 10.

Таблица 10

Матрица моделирования и результаты обработки данных по вскармливанию цыплят

j	X_{0j}	X_j	$\bar{Y}_j$	$X_{0j}\bar{Y}_j$	$X_j\bar{Y}_j$	Y_j^p	$\left(\bar{Y}_j - Y_j^p\right)^2$
1	1	–1,000	410,8	410,8	–410,8	420,5	94,54
2	1	–0,6667	426,0	426,0	–284,0	425,8	0,054
3	1	–0,3333	433,3	433,3	–144,4	431,1	4,785
4	1	0	447,8	447,8	0	436,4	129,8
5	1	0,3333	449,8	449,8	149,9	441,7	65,58
6	1	0,6667	445,0	445,0	296,7	446,9	3,789
7	1	1,000	442,0	442,0	442,0	452,2	104,8

j	X_{0j}	X_j	$\overline{Y}_j$	$X_{0j}\overline{Y}_j$	$X_j\overline{Y}_j$	Y_j^p	$\left(\overline{Y}_j - Y_j^p\right)^2$
$N = 7$				$\sum\limits_{j=1}^{7} X_{0j}^2 = 7$	$\sum\limits_{j=1}^{7} X_{0j}\overline{Y}_j = 3054{,}5$	$\varphi = 403{,}4$	
$n = 4$				$\sum\limits_{j=1}^{7} X_{1j}^2 = 3{,}111$	$\sum\limits_{1}^{7} X_{1j}\overline{Y}_j = 49{,}42$	$S_{\text{ад}}^2 = 332{,}7$, $f = 5$	
				$b_0 = 436{,}4$	$b_1 = 15{,}9$	$F_{\text{э}} = 20{,}23$	
$S_{\text{воспр}}^2 = 15{,}95$				$S^2(b_0) = 0{,}5697$	$S^2(b_1) = 1{,}282$	$F_{5;\,21;\,0{,}95} = 2{,}685$	
$f_{\text{воспр}} = 21$				$S(b_0) = 0{,}7548$	$S(b_1) = 1{,}132$	Модель не адекватна	
$t_{21;\,0{,}95} = 2{,}080$				$\Delta b_0 = 1{,}57$	$\Delta b_1 = 2{,}4$	$Y = 436{,}4 + 15{,}9X$	

В матрицу моделирования включены столбцы j, X_{0j}, X_j, $\overline{Y}_j$, X_{0j}, $\overline{Y}_j$, X_j, $\overline{Y}_j$, Y_j^p, $\left(\overline{Y}_j - Y_j^p\right)^2$, элементы которых рассчитывают следующим образом:

– фиктивный фактор X_0, значения которого во всех опытах равны единице ($X_{0j} = 1$), при этом $\sum\limits_{j=1}^{7} X_{0j}^2 = 7$;

– нормированные значения управляющего фактора X_j для РСП:

$$X_j = -1 + 2 \cdot \frac{j-1}{N-1}, \quad j = 1;\ldots, 7,$$

например: $X_5 = -1 + \dfrac{2\,(5-1)}{7-1} = 0{,}3333$, при этом

$$\sum_{j=1}^{7} X_j^2 = (-1{,}000)^2 + (-0{,}6667)^2 + (-0{,}3333)^2 + 0^2 +$$
$$+ (0{,}6667)^2 + (0{,}3333)^2 + 1^2 = 3{,}111;$$

– суммы столбцов:

$$\sum_{j=1}^{7} X_{0j}\,\overline{Y}_j = 1\cdot 410{,}8 + 1\cdot 426{,}0 + 1\cdot 433{,}3 + 1\cdot 447{,}8 +$$
$$+\ 1\cdot 449{,}8 + 1\cdot 445{,}0 + 1\cdot 442{,}0 = 3054{,}5\,;$$
$$\sum_{j=1}^{7} X_{1j}\,\overline{Y}_j = (-1)\,410{,}8 + (-0{,}6667)\,426{,}0 + (-0{,}3333)\,433{,}3 +$$
$$+\ 0\cdot 447{,}8 + 0{,}3333\cdot 449{,}8 + 0{,}6667\cdot 445{,}0 + 1\cdot 442{,}0 = 49{,}42\,;$$

– факторы ортогональны, так как

$$\sum_{j=1}^{7} X_{0j}\,X_j = 1\,(-1) - 1(0{,}667) - 1\,(0{,}333) + 1\cdot 0 + 1\cdot 0{,}333 + 1\cdot 0{,}667 + 1\cdot 1 = 0\,;$$

– коэффициенты уравнения регрессии для ортогональных факторов:

$$b_0 = \frac{\sum_{j=1}^{N} X_{0j}\overline{Y}_j}{\sum_{j=1}^{7} X_{0j}^2} = \frac{3054{,}5}{7} = 436{,}4\,;$$

$$b_1 = \frac{\sum_{j=1}^{N} X_{1j}\overline{Y}_j}{\sum_{j=1}^{N} X_{1j}^2} = \frac{49{,}42}{3{,}111} = 15{,}88\,;$$

– дисперсии значимости и стандартные отклонения коэффициентов уравнения регрессии:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{7} X_{0j}^2} = \frac{15{,}95}{4\cdot 7} = 0{,}5697\,;$$

$$S(b_0) = \sqrt{S^2(b_0)} = \sqrt{0{,}5697} = 0{,}7548\,;$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{7} X_j^2} = \frac{15{,}95}{4\cdot 3{,}111} = 1{,}281;$$

$$S(b_1) = \sqrt{S^2(b_1)} = \sqrt{1,281} = 1,132\,;$$

– доверительные интервалы коэффициентов уравнения регрессии:

$$\Delta b_0 = t_{N(n-1),\,p}\, S(b_0) = 2,080 \cdot 0,7548 = 1,570 \approx 1,6\,;$$

$$\Delta b_1 = t_{N(n-1),\,p}\, S(b_1) = 2,080 \cdot 1,1322 = 2,355 \approx 2,4\,,$$

где $t_{N(n-1);\,p}$ = 2,080 – критическое значение критерия Стьюдента при числе степеней свободы $f = N(n - 1)$ и доверительной вероятности p = 0,95 (прилож. 2);

Обратите внимание! Абсолютные погрешности рассчитанных коэффициентов регрессии следует округлить по алгоритму корректного оформления результатов расчета: «легкие» – до двух значащих цифр, «тяжелые» – до одной значащей цифры.

– так как $\Delta b_0 < |b_0|$ (1,6 < 436,4) и $\Delta b_1 < |b_1|$ (2,4 < 15,88), то оба коэффициента уравнения регрессии значимы, B = 2.

Проверка уравнения регрессии 1-го порядка на адекватность по критерию Фишера включает расчет следующих параметров:

– параметр оптимизации Y_j^p в каждом опыте, например для j = 1:

$$Y_1^p = 436,4 + 15,9\cdot(-1) = 420,5;$$

– остаточная сумма квадратов:

$$\varphi = \sum_{j=1}^{7}\left(\overline{Y}_j - Y_j^p\right)^2 = (410,8 - 420,5)^2 + (426,0 - 425,8)^2 +$$
$$+ (433,3 - 431,1)^2 + (447,8 - 436,4)^2 + (449,8 - 441,7)^2 +$$
$$+ (445,0 - 446,9)^2 + (458,0 - 452,2)^2 = 403,4;$$

– дисперсия адекватности и ее числа степеней свободы:

$$S^2_{\text{ад}} = \frac{n\varphi}{N - B} = \frac{4 \cdot 212,9}{7 - 2} = 170,28\,;$$

$$f = N - B = 7 - 2 = 5;$$

– экспериментальное значение критерия Фишера:

$$F_{э} = \frac{\max\,(S^2_{ад}, S^2_{воспр})}{\min\,(S^2_{ад}, S^2_{воспр})} = \frac{S^2_{ад}}{S^2_{воспр}} = \frac{170,3}{18,4} = 9,386\text{, так как } S^2_{ад} > S^2_{воспр};$$

– критическое значение критерия Фишера:

$$F_{кр} = F_{f_{числ};\, f_{знам};\, p} = F_{5;\,21;\,0,95} = 2,685 \text{ (прилож. 4)},$$

где на первом месте стоит $f_{числ} = N - B = 7 - 2 = 5$ число степеней свободы большей дисперсии $S^2_{ад}$, на втором – число степеней свободы $f_{знам} = N(n - 1) = 7 \cdot 3 = 21$ меньшей дисперсии $S^2_{воспр}$;

– так как $F_{э} > F_{кр}$, то полученное уравнение регрессии 1-го порядка не адекватно.

Таким образом, полученное уравнение регрессии:

$$Y = 436,4 + 15,9X,$$

в котором все коэффициенты значимы, не адекватно и не может быть использовано для оптимизации технической системы.

Для построения адекватной математической модели, описывающей зависимость привеса цыплят от количества витаминизированной добавки, необходимо перейти к уравнению регрессии 2-го порядка.

3.3. Моделирование технических систем однофакторными регрессионными уравнениями 2-го порядка

Для описания функциональной зависимости параметра оптимизации от управляющего фактора при стохастическом моделировании используем алгебраический полином (64), последовательно повышая его степень. Поэтому очередной этап моделирования технической системы заключается в построении математической модели на базе однофакторного уравнения регрессии 2-го порядка:

$$Y = b_0 X_0 + b_1 X + b_2 X^2. \tag{96}$$

3.3.1. Матрица планирования и проведение эксперимента

Планирование эксперимента для построения однофакторных регрессионных моделей 2-го порядка практически ничем не отличается от алгоритма построения моделей 1-го порядка.

Выбор параметра оптимизации выполняется на тех же принципах, что и для модели 1-го порядка.

Параметр оптимизации должен быть:

– количественным;

– экспериментально определяемым при всех уровнях управляющего фактора;

– непрерывным в исследуемом интервале варьирования фактора;

– имеющим наглядный физический смысл;

– функциональным, то есть характеризовать важное свойство технической системы.

Выбор главного фактора, который должен быть:

– существенным, то есть существенно влияющим на параметр оптимизации;

– количественным;

– детерминированным, то есть его значения на всех уровнях варьирования должны фиксироваться с относительно незначимой приборной погрешностью;

– задаваемым однозначно исследователем.

Построение матрицы планирования начинается с выбора количества опытов N и количества дублей в каждом опыте n (неформализуемый этап). Для надежной оценки адекватности и точности прогнозирования целесообразно провести как можно больше опытов. Как и для уравнений регрессии 1-го порядка, рекомендуемое количество опытов $N > 2k$, где k – число факторов в однофакторном уравнении регрессии 2-го порядка ($k = 3$).

РСП, выбранный в качестве предпочтительного плана при построении уравнения регрессии 1-го порядка, используется и при построении уравнения регрессии 2-го порядка, сохраняя все свои преимущества.

3.3.2. Предварительная обработка экспериментальных данных

Предварительная обработка экспериментальных данных такая же, как при построении уравнения регрессии 1-го порядка, с использованием тех же формул:

– расчет выборочных средних, дисперсий и числа степеней свободы в каждом опыте:

$$\overline{Y}_j = \frac{Y_{j1} + \ldots + Y_{jn}}{n} = \frac{\sum_{i=1}^{n} Y_{ji}}{n}, \quad j = 1, \ldots, N; \tag{97}$$

$$S_j^2 = \frac{(Y_{j1} - \overline{Y}_j)^2 + \ldots + (Y_{jn} - \overline{Y}_j)^2}{n-1} = \frac{\sum_{i=1}^{n} (Y_{ji} - \overline{Y}_j)^2}{n-1}, \quad j = 1, \ldots, N; \tag{98}$$

$$f_{S_j} = n - 1, \quad j = 1, \ldots, N; \tag{99}$$

– проверка выборочных дисперсий всех опытов на однородность по критерию Кохрена, который применяется при условии равенства числа дублей n в каждом опыте. Проверка N выборочных дисперсий с n дублями в каждой выборке на однородность осуществляется по следующему алгоритму:

– расчет экспериментального значения критерия Кохрена:

$$G_э = \frac{\max S_j^2}{\sum_{j=1}^{N} S_j^2}; \tag{100}$$

– определение критического значения критерия Кохрена $G_{n-1;\,N;\,p}$ с числом степеней свободы в числителе $f_{числ} = n - 1$, в знаменателе $f_{зн} = N$ (прилож. 5):

если

$$G_э < G_{n-1;\,N;\,p},$$

то все выборочные дисперсии однородны,

если

$$G_э \geq G_{n-1;\,N;\,p},$$

то все выборочные дисперсии неоднородны. В этом случае для дальнейшей корректной обработки экспериментальных данных необходимо повторить опыт, в котором обнаружено максимальное значение дисперсии. Если и после этого однородность дисперсий не достигается, то дальнейшее моделирование по предложенной методике некорректно;

если все выборочные дисперсии однородны, то они принадлежат генеральным совокупностям с одинаковыми генеральными дисперсиями $\sigma_1^2 = \sigma_2^2 = ... = \sigma_j^2 = ... = \sigma_N^2$. Только в этом случае можно определить среднее значение выборочных дисперсий, которое называется дисперсией воспроизводимости $S^2_{\text{воспр}}$ с числом степеней свободы $f_{\text{воспр}}$, характеризующей точность эксперимента в целом.

$$S^2_{\text{воспр}} = \frac{\sum_{j=1}^{n} S_j^2}{N}, \tag{101}$$

$$f_{\text{воспр}} = N(n-1). \tag{102}$$

Следует отметить, что статистическая значимость любой выборочной дисперсии определяется числом ее степеней свободы. Поэтому статистическая значимость дисперсии воспроизводимости $S^2_{\text{воспр}}$ существенно больше статистической значимости выборочной дисперсии в каждом опыте S_j^2, так как $N(n - 1) > (n - 1)$.

3.3.3. Построение ортогонализированного уравнения регрессии 2-го порядка и проверка его на статистическое качество

Для построения модели технической системы в виде уравнения регрессии 2-го порядка $Y = b_0X_0 + b_1X + b_2X^2$ необходимо определить три коэффициента регрессии b_0, b_1, b_2 (k = 3). Для этого с помощью МНК составляется система уравнений. В данном случае - три уравнения с тремя неизвестными. Однако такая система уравнений для определения b_0, b_1, b_2 не имеет диагональный вид, так как добавленный к уравнению регрессии 1-го порядка $Y = b_0X_0 + b_1X$ квадратичный фактор X^2 в уравнении $Y = b_0X_0 + b_1X + b_2X^2$ не ортогонален фактору X_0:

$$\sum_{j=1}^{N} X_{0j} X_j^2 = \sum_{j=1}^{N} X_j^2 > 0. \tag{103}$$

Поскольку для РСП факторы X_0 и X ортогональны (уравнение (68)), то для упрощения расчета коэффициентов регрессии квадратичный фактор X^2 должен быть ортогонален с X_0 и X. Для ортогонализации фактора X^2 представим его в виде:

$$X^2 \to X^2 - \lambda , \tag{104}$$

где λ – константа.

По условию ортогональности факторов X_0 и $X^2 - \lambda$:

$$\sum_{j=1}^{N} X_{0j}(X_j^2 - \lambda) = \sum_{j=1}^{N} X_j^2 + \sum_{j=1}^{N} \lambda = \sum_{j=1}^{N} X_j^2 + \lambda N = 0 , \tag{105}$$

откуда следует, что ортогонализирующий фактор

$$\lambda = \frac{\sum_{j=1}^{N} X_j^2}{N}. \tag{106}$$

Ортогональность факторов X и $X^2 - \lambda$ следует из условия:

$$\sum_{j=1}^{N} X_j (X_j^2 - \lambda) = 0 .$$

Так как все три фактора X_0, X и $X^2 - \lambda$ ортогональны, то уравнение для расчета коэффициента регрессии b_2 аналогично уравнениям для расчета коэффициентов регрессии b_0 и b_1 (уравнения (70), (71)):

$$b_0 = \frac{\sum_{j=1}^{N} X_{0j} \overline{Y}_j}{\sum_{j=1}^{N} X_{0j}^2} , \tag{107}$$

$$b_1 = \frac{\sum_{j=1}^{N} X_j \overline{Y}_j}{\sum_{j=1}^{N} X_j^2} , \tag{108}$$

$$b_2 = \frac{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right) \overline{Y}_j}{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right)^2} . \qquad (109)$$

Проверка полученного уравнения регрессии 2-го порядка на статистическое качество (проверка коэффициентов регрессии на значимость, а самого уравнения – на адекватность) также полностью повторяет аналогичную процедуру для уравнения регрессии 1-го порядка.

Проверка коэффициентов регрессии на значимость по критерию Стьюдента. Дисперсии значимости коэффициентов ортогонализированного уравнения регрессии b_0, b_1, b_2 в случае ортогональности факторов (и только в этом случае) рассчитываются по уравнениям, аналогичным уравнениям (74), (75):

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} X_{0j}^2} , \qquad (110)$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} X_j^2} , \qquad (111)$$

$$S^2(b_2) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} \left(X_j^2 - \lambda \right)^2} . \qquad (112)$$

Процедура построения доверительных интервалов для каждого коэффициента уравнения регрессии 2-го порядка также совпадает с процедурой построения доверительных интервалов для коэффициентов уравнения регрессии 1-го порядка.

$$\Delta b_0 = t_{f_{\text{воспр}};\, p} S(b_0) , \qquad (113)$$

$$\Delta b_1 = t_{N(n-1);\, p} S(b_1) , \qquad (114)$$

$$\Delta b_2 = t_{N(n-1);\, p} S(b_2) , \qquad (115)$$

где $t_{N(n-1);\,p}$ – критическое значение критерия Стьюдента с числом степеней свободы $f_{\text{воспр}} = N\left(n-1\right)$ при доверительной вероятности p = 0,95 (прилож. 2),

если

$$\Delta b_0 < |b_0|, \tag{116}$$

$$\Delta b_1 < |b_1|, \tag{117}$$

$$\Delta b_2 < |b_2|, \tag{118}$$

то все коэффициенты уравнения регрессии статистически значимы,

если условия (116)–(118) не выполняются, то соответствующие коэффициенты незначимы и должны быть исключены из уравнения регрессии.

Проверка уравнения регрессии 2-ого порядка на адекватность по критерию Фишера также полностью совпадает с аналогичной процедурой для однофакторного уравнения регрессии 1-го порядка:

– расчет экспериментального значения критерия Фишера:

$$F_{\text{э}} = \frac{\max\left(S^2_{\text{ад}}, S^2_{\text{воспр}}\right)}{\min\left(S^2_{\text{ад}}, S^2_{\text{воспр}}\right)}\ ; \tag{119}$$

– определение критического значения критерия Фишера $F_{f_1;\,f_2;\,p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии f_1, на втором – число степеней свободы *меньшей* дисперсии f_2 при доверительной вероятности p = 0,95 (прилож. 4):

если

$$F_{\text{э}} < F_{f_1;\,f_2;\,p}, \tag{120}$$

то выборочные дисперсии $S^2_{\text{воспр}}$ и $S^2_{\text{ад}}$ однородны, следовательно, полученное уравнение регрессии 1-го порядка со значимыми коэффициентами адекватно,

если

$$F_{\text{э}} \geq F_{f_1;\,f_2;\,p}, \tag{121}$$

то выборочные дисперсии $S^2_{\text{воспр}}$ и $S^2_{\text{ад}}$ неоднородны, следовательно, полученное уравнение регрессии со значимыми коэффициентами не адекватно. В этом случае необходимо перейти к уравнению регрессии 2-го порядка.

В случае неадекватности однофакторного уравнения регрессии 2-го порядка целесообразно перейти к построению уравнения регрессии 3-го порядка.

3.3.4. Оптимизация технической системы по уравнению регрессии 2-го порядка

Уравнение регрессии 2-го порядка позволяет решить задачу оптимизации технической системы, которая осуществима только при следующих условиях:

1) выборочные дисперсии всех опытов однородны по критерию Кохрена;

2) все коэффициенты, включенные в уравнение регрессии, статистически значимы по критерию Стьюдента;

3) полученное уравнение регрессии адекватно по критерию Фишера;

4) оптимизация осуществляется только в факторном пространстве, в котором уравнение регрессии адекватно.

Если уравнение регрессии 2-го порядка адекватно, то для него всегда существует глобальный экстремум – максимум или минимум исследуемого параметра, но он может лежать и вне интервала варьирования управляющего фактора:

– необходимым условием глобального экстремума является равенство нулю первой производной уравнения регрессии 2-го порядка $Y(X)=b_0+b_1X+b_2\left(X^2-\lambda\right)$:

$$Y'(X)=\left[b_0+b_1X+b_2\left(X^2-\lambda\right)\right]'=b_1+2b_2X=0; \qquad (122)$$

– достаточное условие максимума:

$$Y''<0; \qquad (123)$$

– достаточное условие минимума:

$$Y''>0. \qquad (124)$$

Для уравнения регрессии 2-го порядка $Y(X) = b_0 + b_1 X + b_2 \left(X^2 - \lambda \right)$ вторая производная равна

$$Y''(X) = \left(Y'(X) \right)' = \left(b_1 + 2b_2 X \right)' = 2b_2 . \qquad (125)$$

Поэтому уравнения (123) и (124) с учетом уравнения (125) имеют следующий вид:

– достаточное условие максимума:

$$Y''\left(X \right) = 2b_2 < 0 \quad \to \quad b_2 < 0 ; \qquad (126)$$

– достаточное условие минимума:

$$Y''\left(X \right) = 2b_2 > 0 \quad \to \quad b_2 > 0 . \qquad (127)$$

Из уравнения (122) следует, что значение управляющего фактора, при котором уравнение регрессии имеет глобальный максимум (минимум), равно:

$$X_{\text{опт}} = -\frac{b_1}{2b_2} . \qquad (128)$$

Если значение управляющего фактора $X_{\text{опт}}$, при котором исследуемый параметр имеет оптимальное значение, лежит в интервале варьирования управляющего фактора $X_{\text{опт}} \in \left[-1; +1 \right]$, то оптимальное значение параметра оптимизации (максимальное или минимальное) определяется следующим уравнением:

$$Y_{\text{опт}} = b_0 + b_1 \left(-\frac{b_1}{2b_2} \right) + b_2 \left[\left(-\frac{b_1}{2b_2} \right)^2 - \lambda \right], \qquad (129)$$

где $Y_{\text{опт}} = Y_{\min}$, если $b_2 > 0$; $Y_{\text{опт}} = Y_{\max}$, если $b_2 < 0$.

Если значение управляющего фактора X лежит вне интервала варьирования, то анализ полученной модели не может быть выполнен, так как отсутствуют основания считать уравнение регрессии адекватным вне области [–1;+1].

В этом случае оптимальное значение управляющего фактора следует выбрать на границе интервала варьирования $X_{опт}$ = –1 или $X_{опт}$ = +1, которое определяется следующими уравнениями:

$$Y_{\min} = \min\left[Y(-1), Y(+1)\right], \tag{130}$$

$$Y_{\max} = \max\left[Y(-1), Y(+1)\right]. \tag{131}$$

Приведем без доказательства несколько формул, полезных для РСП.

$$\sum_{j=1}^{N} X_{0j}^{2} = N\,. \tag{132}$$

$$\sum_{j=1}^{N} X_{j}^{2} = \frac{N\left(N+1\right)}{3\left(N-1\right)}. \tag{133}$$

$$\lambda = \frac{N+1}{3\left(N-1\right)}. \tag{134}$$

$$\sum_{j=1}^{N}\left(X_{j}^{2}-\lambda\right)^{2} = \frac{4N\left(N+1\right)\left(N^{2}-4\right)}{45\left(N-1\right)^{3}}. \tag{135}$$

Подставляя в уравнения (107)–(109) уравнения (132)–(135), получим:

$$b_0 = \frac{\sum_{j=1}^{N} X_{0\,j}\overline{Y}_j}{N}; \tag{136}$$

$$b_1 = \frac{3(N-1)}{N(N+1)}\sum_{j=1}^{N} X_{\,j}\overline{Y}_j\;; \tag{137}$$

$$b_2 = \frac{45(N-1)^3}{4N(N+1)(N^2-4)}\sum_{j=1}^{N}\left(X_{j}^{2}-\frac{N+1}{3\left(N-1\right)}\right)\overline{Y}_j\;. \tag{138}$$

Подставляя в уравнения (110)–(112) уравнения (132)–(135), получим:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{Nn}; \tag{139}$$

$$S^2(b_1) = \frac{3(N-1)}{N+1} \cdot \frac{S^2_{\text{воспр}}}{Nn}; \tag{140}$$

$$S^2(b_2) = \frac{45(N-1)^3}{4(N+1)(N^2-4)} \cdot \frac{S^2_{\text{воспр}}}{Nn}. \tag{141}$$

3.3.5. Алгоритм решения прикладных задач с помощью однофакторных уравнений регрессии 2-го порядка

1. Выбрать объект моделирования – техническую систему – и на основе литературных и патентных источников собрать необходимую предварительную (априорную) информацию о факторах и параметрах, характеризующих функционирование моделируемой технической системы.

1.1. Выбрать параметр оптимизации Y, эффективно характеризующий функционирование технической системы.

1.2. Выбрать управляющий фактор x, значимо влияющий на параметр оптимизации, а также интервал его варьирования [$x_{\min}$, $x_{\max}$].

1.3. Выбрать в качестве уравнения регрессии алгебраический ортогонализированный полином 2-го порядка:

$$Y = b_0 X_0 + b_1 X + b_2\left(X^2 - \lambda\right).$$

1.4. Выбрать число опытов N и число дублей n в каждом опыте. Рекомендуемое число опытов для полинома 2-го порядка $N > 2k$, где $k = 3$. Рекомендуемое число дублей $n > 3$.

1.5. Построить матрицу планирования эксперимента для РСП, включающую столбцы $j,\ x_j, Y_{j1}, ..., Y_{jn}, \overline{Y_j}\ , S_j^2$, в которой натуральные значения фактора x_j для каждого опыта рассчитать по формуле:

$$x_j = x_{\min} + \frac{j-1}{N-1}(x_{\max} - x_{\min}),\ j = 1, \dots, N.$$

1.6. Обеспечить приборное и технологическое оснащение эксперимента, позволяющее измерять управляющий фактор x с требуемой точностью, и провести эксперимент.

2. Провести предварительную обработку экспериментальных данных.

2.1. Расчет выборочного среднего в каждом опыте:

$$\overline{Y_j} = \frac{\sum_{i=1}^{n} Y_{ji}}{n}, \quad j = 1, \ldots, N.$$

2.2. Расчет выборочной дисперсии и ее числа степеней свободы в каждом опыте:

$$S_j^2 = \frac{\sum_{i=1}^{n} \left(Y_{ji} - \overline{Y}_j \right)^2}{n-1}, \quad j = 1, \ldots, N.$$

$$f_S = n - 1.$$

2.3. Проверка выборочных дисперсии на однородность по критерию Кохрена:

– определение критического значения критерия Кохрена $G_{n-1;\,N;\,p}$, в котором на первом месте – число степеней свободы максимальной дисперсии $f_1 = n - 1$, а на втором – число степеней свободы для суммы всех выборочных дисперсий $f_2 = N$ при доверительной вероятности $p = 0{,}95$ (прилож. 5):

если

$$G_э < G_{n-1;\,N;\,p},$$

то все выборочные дисперсии однородны,

если

$$G_э \geq G_{n-1;\,N;\,p},$$

то все выборочные дисперсии неоднородны.

Внимание! Дальнейшая обработка экспериментальных данных возможна только в том случае, если выборочные дисперсии всех опытов однородны.

2.4. Расчет дисперсии воспроизводимости и ее числа степеней свободы:

$$S_{\text{воспр}}^2 = \frac{\sum_{j=1}^{N} S_j^2}{N};$$

$$f_{\text{воспр}} = N(n-1).$$

3. Построить уравнение регрессии и проверить его на статистическое качество.

3.1. Построить матрицу моделирования, включающую следующие столбцы $j, X_{0j}, X_j, (X_j^2-\lambda), \overline{Y_j}, X_{0j}\overline{Y_j}, X_j\overline{Y_j}, (X_j^2-\lambda)\overline{Y_j}$, $Y_j^p, (\overline{Y_j}-Y_j^p)^2$:

– расчет нормированных значений фактора X_j для каждого опыта:

$$X_j = -1 + \frac{2(j-1)}{N-1}, \quad j = 1, \ldots, N;$$

– расчет ортогонализирующего коэффициента:

$$\lambda = \frac{\sum_{j=1}^{N} X_j^2}{N} = \frac{N+1}{3(N-1)};$$

– расчет нормированного значения ортогонализированного фактора $\left(X_j^2-\lambda\right)$ для каждого опыта.

3.2. Расчет коэффициентов b_0, b_1, b_2 уравнения регрессии 2-го порядка:

$$b_0 = \frac{\sum_{j=1}^{N} X_{0j}\overline{Y_j}}{\sum_{j=1}^{N} X_{0j}^2};$$

$$b_1 = \frac{\sum_{j=1}^{N} X_j\overline{Y_j}}{\sum_{j=1}^{N} X_j^2};$$

$$b_2 = \frac{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right) \overline{Y}_j}{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right)^2} .$$

3.3. Проверка коэффициентов b_0, b_1, b_2 уравнения регрессии 2-го порядка $Y = b_0 X_0 + b_1 X + b_2 \left(X^2 - \lambda \right)$ на значимость по критерию Стьюдента:

– расчет дисперсии значимости коэффициентов b_0, b_1, b_2:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} X^2_{0j}};$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} X^2_{j}};$$

$$S^2(b_2) = \frac{S^2_{\text{воспр}}}{n \sum_{j=1}^{N} \left(X_j^2 - \lambda \right)^2};$$

– расчет доверительных интервалов для коэффициентов b_0, b_1, b_2:

$$\Delta b_0 = t_{N(n-1);\, p} S(b_0);$$

$$\Delta b_1 = t_{N(n-1);\, p} S(b_1);$$

$$\Delta b_2 = t_{N(n-1);\, p} S(b_2),$$

где $t_{N(n-1);\, p}$ – критическое значение критерия Стьюдента с числом степеней свободы $f_{\text{воспр}} = N(n - 1)$ при доверительной вероятности $p = 0{,}95$ (прилож. 2):

если

$$\Delta b_0 < |b_0|,$$

$$\Delta b_1 < |b_1|,$$

$$\Delta b_2 < |b_2|,$$

то коэффициенты регрессии b_0, b_1 и b_2 статистически значимы,

если для некоторых коэффициентов уравнения регрессии эти неравенства не выполняются, то соответствующие коэффициенты незначимы. Незначимые коэффициенты следует исключить из уравнения регрессии. Количество оставшихся значимых коэффициентов регрессии равно B.

3.4. Проверка однофакторного уравнения регрессии 2-го порядка $Y = b_0 X_0 + b_1 X + b_2 \left(X^2 - \lambda \right)$ на адекватность по критерию Фишера:

– расчет значения параметра оптимизации в каждом опыте по уравнению регрессии 2-го порядка:

$$Y_j^p = b_0 X_{0j} + b_1 X_j + b_2 \left(X_j^2 - \lambda \right), \quad j = 1, \ldots, N;$$

– расчет квадрата разности экспериментальных и расчетных значений параметра оптимизации:

$$(\overline{Y}_j - Y_j^p)^2, \quad j = 1, \ldots, N;$$

– расчет остаточной суммы квадратов:

$$\varphi = \sum_{j=1}^{N} \left(\overline{Y}_j - Y_j^p \right)^2;$$

– расчет дисперсии адекватности и ее числа степеней свободы:

$$S_{\text{ад}}^2 = \frac{n\,\varphi}{N - B};$$

$$f_{\text{ад}} = N - B;$$

– расчет экспериментального значение критерия Фишера:

$$F_э = \frac{\max\left(S_{ад}^2, S_{воспр}^2\right)}{\min\left(S_{ад}^2, S_{воспр}^2\right)};$$

– определение критического значения критерия Фишера $F_{f_1; f_2; p}$, в котором на первом месте стоит число степеней свободы *большей* дисперсии f_1, на втором – число степеней свободы *меньшей* дисперсии f_2, при доверительной вероятности p = 0,95 (прилож. 4):

если

$$F_э < F_{f_1; f_2; p},$$

то выборочные дисперсии $S_{воспр}^2$ и $S_{ад}^2$ однородны, следовательно, полученное уравнение регрессии 2-го порядка со значимыми коэффициентами адекватно,

если

$$F_э \geq F_{f_1; f_2; p},$$

то выборочные дисперсии $S_{воспр}^2$ и $S_{ад}^2$ неоднородны, и, следовательно, полученное уравнение регрессии со значимыми коэффициентами не адекватно.

В случае неадекватной модели следует перейти к построению уравнения регрессии 3-го порядка.

4. Оптимизация технической системы по адекватному уравнению регрессии 2-го порядка, в котором все коэффициентами значимы:

– расчет оптимального нормированного значения управляющего фактора, в котором параметр оптимизации равен минимуму или максимуму:

$$X_{опт} = -\frac{b_1}{2b_2}.$$

Если значение управляющего фактора $X_{опт}$, при котором исследуемый параметр равен оптимальному значению, лежит в интервале варьирования управляющего фактора $X_{опт} \in [-1;+1]$, то оптимальное значение параметра оптимизации (максимальное или минимальное) рассчитать по уравнению

$$Y_{\text{опт}} = b_0 + b_1\left(-\frac{b_1}{2b_2}\right) + b_2\left[\left(-\frac{b_1}{2b_2}\right)^2 - \lambda\right].$$

Если значение управляющего фактора X лежит вне интервала варьирования, то анализ полученной модели не может быть выполнен, так как отсутствуют основания считать уравнение регрессии адекватным вне области [–1;+1]. В этом случае оптимальное значение управляющего фактора следует выбрать на границе интервала варьирования $X_{\text{опт}}$ = –1 или $X_{\text{опт}}$ = +1. Оптимальное значение параметра оптимизации (максимальное или минимальное) определяется следующими уравнениями:

$$Y_{\min} = \min\left[Y(-1), Y(+1)\right],$$

$$Y_{\max} = \max\left[Y(-1), Y(+1)\right];$$

– расчет оптимального натурального значения управляющего фактора, в котором параметр оптимизации достигает минимума или максимума:

$$x_{\text{опт}} = x_0 + X_{\text{опт}}\Delta x;$$

$$x_0 = \frac{x_{\max} + x_{\min}}{2};$$

$$\Delta x = \frac{x_{\max} - x_{\min}}{2};$$

– расчет оптимального значение параметра оптимизации:

$$Y_{\text{опт}} = b_0 + b_1 X_{\text{опт}} + b_2\left(X_{\text{опт}}^2 - \lambda\right);$$

– расчет абсолютной погрешности оптимального параметра оптимизации:

$$\Delta Y_{\text{опт}} = t_{N(n-1),\,p}\sqrt{S^2(b_0) + X_{\text{опт}}^2 S^2(b_1) + (X_{\text{опт}}^2 - \lambda)^2 S^2(b_2)}\,.$$

3.3.6. Типовая задача: стохастическое моделирование технических систем уравнением регрессии 2-го порядка

Цель – исследование функционирования однофакторных технических систем с помощью уравнения регрессии 2-го порядка.

Проиллюстрируем моделирование и оптимизацию однофакторных технических систем уравнениями регрессии 2-го порядка на примере разработки процесса кормления цыплят с учетом влияния витаминизированной добавки.

Математическая постановка задачи – построение математической модели процесса кормления на базе однофакторного уравнения регрессии 2-го порядка, проверка модели на статистическое качество и оптимизация режима кормления цыплят до 2-недельного возраста.

1. Планирование и проведение эксперимента. В качестве параметра оптимизации выбрана масса Y (г) цыплят, достигнутая в результате 2-недельного вскармливания.

В качестве управляющего фактора, влияющего на массу цыплят, выбрано содержание витаминизированной добавки в корме. Принято решение интервал варьирования управляющего фактора принять равным $x \in [0, 20]$ г/сут.

Для уравнения регрессии 2-го порядка ($m = 2$) построим матрицу планирования на базе РСП с количеством опытов $N = 11$. При этом выполнено соотношение $N > 2k$. Принимаем число дублей $n = 4$.

Для реализации РСП были организованы еще 11 контрольных вольеров, в каждом из которых находилось по 200 цыплят суточного возраста. Цыплята получали одинаковый корм по количеству и составу, за исключением количества витаминизированной добавки.

Для обеспечения корректности начальных условий эксперимента необходимо, чтобы различие стартовых масс цыплят во всех опытах было несущественным. Средняя стартовая масса цыплят составила $m_0 = (50{,}2 \pm 0{,}7)$ г. Расчеты показали, что различие между выборочными средними по критерию Стьюдента в этом случае несущественно (см. табл. 7).

В качестве плана эксперимента выбран РСП. Количество добавки в вольерах рассчитывалось по уравнению

$$x_j = x_{\min} + \frac{j-1}{N-1}(x_{\max} - x_{\min}), \; j = 1, \ldots, 11. \tag{142}$$

Цыплята во всех вольерах получали одинаковый корм по количеству и качеству, за исключением витаминизированной добавки. Затем в течение 14 дней при каждом кормлении в каждом вольере добавлялась витаминизированная добавка в соответствии с планом эксперимента (табл. 11, столбец x_j).

Таблица 11

Матрица планирования для построения однофакторного ортогонализированного уравнения регрессии 2-го порядка на базе РСП с числом опытов N = 11

j	x_j, г/сут	Y_{j1}, г	Y_{j2} г	Y_{j3}, г	Y_{j4}, г	$\overline{Y_j}$, г	S_j^2
1	0	408	406	415	414	410,8	19,58
2	2	430	419	427	428	426,0	23,33
3	4	424	423	427	419	423,3	10,92
4	6	443	437	432	439	437,8	20,92
5	8	443	452	451	453	449,8	20,92
6	10	445	440	448	447	445,0	12,67
7	12	464	454	456	458	458,0	18,67
8	14	457	465	470	461	463,3	30,92
9	16	465	459	466	455	461,3	26,92
10	18	453	458	465	459	458,8	24,25
11	20	448	453	460	463	456,0	46,00
$G_э = 0{,}180$				$\sum_{j=1}^{11} S_j^2 = 255{,}1$			
$G_{3;\,20;\,0{,}95} = 0{,}348$				$S^2_{воспр} = 23{,}19$ $f_{воспр} = 33$			

2. Предварительная обработка экспериментальных данных:

– выборочное среднее в каждом опыте:

$$\overline{Y_j} = \frac{\sum_{i=1}^{4} Y_{ji}}{4}, \quad j = 1, \ldots, 11,$$

например: $\overline{Y}_{10} = (453 + 458 + 465 + 459)/4 = 458{,}8$;

– выборочная дисперсия и ее число степеней свободы в каждом опыте:

$$S_j^2 = \sum_{i=1}^{4}(Y_{ji} - \overline{Y}_j)^2 \,/\, (4-1),$$

например:

$$S_{10}^2 = \frac{(453-458{,}8)^2 + (458-458{,}8)^2 + (465-458{,}8)^2 + (459-458{,}8)^2}{11-1} = 24{,}25,$$

$$f_{S_j} = n-1, \quad j = 1, \ldots, 11;$$

– проверка всех выборочных дисперсий на однородность по критерию Кохрена:

$$G_э = \frac{\max(S_j^2)}{\sum_{j=1}^{11} S_j^2} = \frac{46{,}00}{255{,}1} = 0{,}180;$$

$$G_{кр} = G_{n-1,N,p} = G_{3;\,11;\,0{,}95} = 0{,}348 \text{ (прилож. 5)};$$

– так как $G_э < G_{кр}$, то выборочные дисперсии однородны;

– дисперсия воспроизводимости и ее число степеней свободы:

$$S_{воспр}^2 = \frac{\sum_{j=1}^{11} S_j^2}{11} = \frac{255{,}1}{11} = 23{,}19,$$

$$f_{воспр} = N(n-1) = 11\,(4-1) = 33.$$

Результаты предварительной обработки экспериментальных данных внесены в матрицу планирования (см. табл. 11).

3. Построение уравнения регрессии 2-го порядка и проверка его на статистическое качество. Для расчета коэффициентов ортогонализированного уравнения регрессии 2-го порядка $Y = b_0 X_0 + {}$ ${} + b_1 X + b_2(X^2 - \lambda)$ построим матрицу моделирования с числом опытов $N = 11$ (табл. 12), включающую столбцы $j,\ X_{0j},\ X_j, (X_j^2 - \lambda), \overline{Y_j},$ $Y_j^p, (\overline{Y_j} - Y_j^p)^2 \quad X_{0j}\overline{Y_j},\ X_j\overline{Y_j}, (X_j^2 - \lambda)\overline{Y_j}$, элементы которых рассчитываются следующим образом:

Таблица 12

Матрица моделирования и результаты обработки данных по вскармливанию цыплят

j	X_{0j}	X_j	$X_j^2-\lambda$	$\bar{Y}_j$	$X_{0j}\bar{Y}_j$	$X_j\bar{Y}_j$	$\left(X_j^2-\lambda\right)\bar{Y}_j$	Y_j^p	$\left(\bar{Y}_j-Y_j^p\right)^2$
1	1	–1,00	0,600	410,8	410,8	–410,8	246,5	409,9	0,704
2	1	–0,80	0,240	426,0	426,0	–40,8	102,2	421,0	24,676
3	1	–0,60	–0,040	423,3	423,3	–254,0	–16,9	430,8	56,316
4	1	–0,40	–0,240	437,8	437,8	–175,1	–105,1	439,1	1,760
5	1	–0,20	–0,360	449,8	449,8	–90,0	–161,9	446,0	14,070
6	1	0	–0,400	445,0	445,0	0	–178,0	451,5	42,531
7	1	0,20	–0,360	458,0	458,0	91,6	–164,9	455,6	5,549
8	1	0,40	–0,240	463,3	463,3	185,3	–111,2	458,4	23,839
9	1	0,60	–0,040	461,3	461,3	276,8	–18,5	459,7	2,431
10	1	0,80	0,240	458,8	458,8	367,0	110,1	459,6	0,747
11	1	1,00	0,600	456,0	456,0	456,0	273,6	458,1	4,572

$N = 11$	$\sum_{j=1}^{11} X_{0j}^2 = 11$	$\sum_{j=1}^{11} X_{0j}\bar{Y}_j = 4890$	$\sum_{j=1}^{11}\left(X_j^2-\lambda\right) = 1,373$	$\varphi = 177,2$
$n = 4$	$\sum_{j=1}^{11} X_j^2 = 4,40$	$\sum_{1}^{11} X_j\bar{Y}_j = 106,1$	$\sum_{j=1}^{11}\left(X_j^2-\lambda\right)\bar{Y}_i = -24,02$	$S^2_{\text{ад}} = 44,30$, $f = 8$
$\lambda = 0,400$	$b_0 = 444,5$	$b_1 = 24,1$	$b_2 = 17,5$	$F_э = 1,910$
$S^2_{\text{воспр}} = 23,19$	$S^2(b_0) = 0,5270$	$S^2(b_1) = 1,318$	$S^2(b_2) = 4,223$	$F_{8,\ 33,\ 0,95} = 2,235$
$f_{\text{воспр}} = 33$	$S(b_0) = 0,7260$	$S(b_1) = 1,1479$	$S(b_2) = 2,055$	Модель адекватна
$t_{33;\ 0,95} = 2,035$	$\Delta b_0 = 1,5$	$\Delta b_1 = 2,3$	$\Delta b_2 = 4,2$	$Y = 444,5 + 24,1X - 17,5\left(X^2 - 0,4\right)$

– фиктивный фактор X_0, значения которого во всех опытах равны единице X_0 = 1, причем $\sum_{j=1}^{11} X_{0j}^2 = 11$;

– значения нормированного главного фактора:

$$X_j = 2\,\frac{j-1}{N-1} - 1, \quad j = 1, \ldots, 11,$$

например: $X_{10} = -1 + \dfrac{2(10-1)}{11-1} = 0{,}8$, при этом $\sum_{j=1}^{11} X_j^2 = 4{,}400$;

– ортогонализирующий коэффициент

$$\lambda = \frac{\sum_{i=1}^{11} X_j^2}{11} = \frac{4{,}400}{11} = 0{,}4000;$$

– ортогонализированный фактор $(X_j^2 - \lambda)$, при этом

$$\sum_{j=1}^{11} (X_j^2 - \lambda)^2 = 1{,}3728;$$

– суммы столбцов:

$$\sum_{j=1}^{11} X_{0j}\,\overline{Y}_j = 1 \cdot 410{,}8 + 1 \cdot 426{,}0 + 1 \cdot 423{,}3 + 1 \cdot 437{,}8 +$$
$$+1 \cdot 449{,}8 + 1 \cdot 445{,}0 + 1 \cdot 458{,}0 = 3050{,}5;$$

$$\sum_{j=1}^{11} X_j \overline{Y}_j = (-1)\,410{,}8 + (-0{,}6667)\,426{,}0 + (-0{,}3333)\,423{,}3 +$$
$$+0 \cdot 437{,}8 + 0{,}3333 \cdot 449{,}8 + 0{,}6667 \cdot 445{,}0 + 1 \cdot 458{,}0 = 68{,}75;$$

$$\sum_{j=1}^{11} (X_j^2 - \lambda)\overline{Y}_j = -24{,}02;$$

– коэффициенты уравнения регрессии:

$$b_0 = \frac{\sum_{j=1}^{11} X_{0j}\overline{Y}_j}{\sum_{j=1}^{11} X_{0j}^2} = \frac{3050,5}{7} = 435,8\,;$$

$$b_1 = \frac{\sum_{j=1}^{11} X_j\overline{Y}_j}{\sum_{j=1}^{11} X_j^2} = \frac{68,75}{3,111} = 22,10\,;$$

$$b_2 = \frac{\sum_{j=1}^{11} X_j\overline{Y}_j}{\sum_{j-1}^{11} X_j^2} = \frac{68,75}{3,111} = 22,10\,;$$

– дисперсии и стандартные отклонения значимости коэффициентов уравнения регрессии:

$$S^2(b_0) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{11} X_{0j}^2} = \frac{23,19}{4\cdot 11} = 0,5270;$$

$$S(b_0) = \sqrt{S^2(b_0)} = \sqrt{0,521\,70} = 0,7260;$$

$$S^2(b_1) = \frac{S^2_{\text{воспр}}}{n\sum_{j=1}^{11} X_j^2} = \frac{23,19}{4\cdot 4,40} = 1,318;$$

$$S(b_1) = \sqrt{S^2(b_1)} = \sqrt{1,318} = 1,1479;$$

$$S^2(b_2) = \frac{S^2_{\text{воспр}}}{n\sum\limits_{j=1}^{11}\left(X_j^2 - \lambda\right)} = \frac{23{,}19}{4\cdot 1{,}373} = 4{,}223;$$

$$S(b_2) = \sqrt{S^2(b_2)} = \sqrt{4{,}223} = 2{,}055;$$

– доверительные интервалы коэффициентов уравнения регрессии:

$$\Delta b_0 = t_{N(n-1),\,p}\cdot S(b_0) = 2{,}035\cdot 0{,}7260 = 1{,}477 \approx 1{,}5\,;$$

$$\Delta b_1 = t_{N(n-1),\,p}\cdot S(b_1) = 2{,}035\cdot 1{,}1479 = 2{,}335 \approx 2{,}3\,;$$

$$\Delta b_2 = t_{N(n-1),\,p}\cdot S(b_2) = 2{,}035\cdot 2{,}055 = 4{,}181 \approx 4{,}2\,,$$

где $t_{N(n-1),\,p}$ = 2,080 – критическое значение критерия Стьюдента при числе степеней свободы $f = N(n-1)$ и доверительной вероятности p = 0,95 (прилож. 2);

Обратите внимание! Абсолютные погрешности рассчитанных коэффициентов регрессии следует округлить по алгоритму корректной записи результата: «легкие» – до двух значащих цифр, «тяжелые» – до одной значащей цифры.

– так как $\Delta b_0 < |b_0|$ (1,5 < 444,5), $\Delta b_1 < |b_1|$ (2,3 < 24,1) и $\Delta b_2 < |b_2|$ (4,2 < 17,5), то все коэффициенты уравнения регрессии значимы.

Проверка уравнения регрессии 2-го порядка на адекватность по критерию Фишера:

– расчетные значения параметра Y_j^p, например, для j = 10:

$$Y_{10}^p = b_0 + b_1 X_{10} + b_2\left(X_{10}^2 - \lambda\right) =$$
$$= 444{,}5 + 24{,}1\cdot 0{,}8 - 17{,}5\left(0{,}8^2 - 0{,}40\right) = 459{,}6;$$

– остаточная сумма квадратов:

$$\varphi = \sum_{j=1}^{11}(\overline{Y}_j - Y_j^p)^2 = (410{,}8 - 409{,}9)^2 + (426{,}0 - 421{,}0)^2 +$$
$$+ (423{,}3 - 430{,}8)^2 + (437{,}8 - 439{,}1)^2 + (449{,}8 - 446{,}0)^2 +$$

$$+(445,0-451,5)^2+(458,0-455,6)^2+(463,3-458,4)^2+$$
$$+(461,3-459,7)^2+(458,8-459,6)^2+(456,0-458,1)^2=177,2\,;$$

– дисперсия адекватности и ее число степеней свободы:

$$S^2_{\text{ад}}=\frac{n\varphi}{N-B}=\frac{4\cdot 177,2}{11-2}=44,30\,;$$

$$f=N-B=11-2=9\,;$$

– экспериментальное значение критерия Фишера:

$$F_{\text{э}}=\frac{S^2_{\text{ад}}}{S^2_{\text{воспр}}}=\frac{44,3}{23,19}=1,910\text{, так как } S^2_{\text{ад}}>S^2_{\text{воспр}}\,;$$

– критическое значение критерия Фишера:

$$F_{(N-B),\,N(n-1),\,p}=F_{9,\,33,\,0{,}95}=2{,}235\ \text{(прилож. 4)};$$

– так как $F_{\text{э}} < F_{\text{кр}}$, то полученное уравнение регрессии 2-го порядка адекватно.

Таким образом, полученное уравнение регрессии:

$$Y^p_{10}=444,5+24,1X-17,5\left(X^2-0,40\right),$$

является адекватным, в котором все коэффициенты уравнения регрессии значимы, и может быть использовано для прогнозирования и оптимизации процесса вскармливания цыплят.

Оптимизация однофакторной технической системы. Полученное уравнение является уравнением регрессии 2-го порядка:

$$Y^p_{10}=444,5+24,1X-17,5\left(X^2-0,40\right),$$

имеет глобальный максимум, так как $b_2 = -17,5 < 0$. Для решаемой задачи представляет интерес найти оптимальное значение управляющего фактора (содержание витаминизированной добавки), который обеспечит максимум параметра оптимизации (масса цыплят, достигнутая в результате 2-недельного вскармливания):

– расчет оптимального значения управляющего фактора в нормированных координатах:

$$X_{\text{опт}} = -\frac{b_1}{2b_2} = -\frac{24,1}{2\left(-17,5\right)} = 0,6886;$$

– расчет оптимального значение управляющего фактора в натуральных координатах:

$$x_{1\text{опт}} = x_{1\text{о}} + X_{1\text{опт}} \cdot \Delta x_1 = 10 + 0,6886 \cdot 10 = 16,9 \text{ г/сут};$$

– расчет максимального значение параметра оптимизации:

$$Y_{\max} = 444,5 + 24,1 \cdot X_{\text{опт}} - 17,5\left(X_{\text{опт}}^2 - 0,40\right) =$$

$$= 444,5 + 24,1 \cdot 0,6886 + \left(0,6886^2 - 0,40\right) = 459,7 \text{ г};$$

– расчет предельной абсолютной погрешности параметра оптимизации $Y_{\max}$:

$$\Delta Y_{\text{опт}} = t_{N(n-1),\,p}\sqrt{S^2(b_0) + X_{\text{опт}}^2 S^2(b_1) + (X_{\text{опт}}^2 - \lambda)^2 S^2(b_2)} =$$

$$= 2,035\sqrt{0,5270 + 0,6886^2 \cdot 1,318 + \left(0,6886^2 - 0,40\right)^2 4,223} = 1,615 \approx 1,6.$$

Вывод. Максимальная масса цыпленка 2-недельного возраста при использовании витаминизированной добавки в количестве 16,9 г/сут составит (459,7 ± 1,6) г, а при кормлении без внесения витаминизированной добавки – 409,9 г. Таким образом, использование витаминизированной добавки увеличивает средний привес цыплят 2-недельного возраста на 459,7 – 409,9 = 49,8 г (≈ 12 %).

3.4. Вопросы и задания для самопроверки

1. Напишите уравнение регрессии 2-го порядка в ортогонализированном виде.
2. Напишите формулу для расчета ортогонализирующего коэффициента квадратичного члена уравнения регрессии 2-го порядка.

3. Напишите формулы для расчета коэффициентов уравнения регрессии 2-го порядка в ортогонализированном виде (N опытов, n дублей).

4. Напишите формулу для расчета дисперсии воспроизводимости (N опытов, n дублей). Чему равно число степеней свободы для этого параметра?

5. Напишите формулы для расчета дисперсий значимости коэффициентов уравнения регрессии 2-го порядка в общем и ортогонализированном видах (N опытов, n дублей).

6. Напишите формулы для расчета доверительных интервалов коэффициентов уравнения регрессии 2-го порядка.

7. Сформулируйте критерий значимости коэффициентов регрессии в общем виде.

8. Напишите формулу для остаточной суммы квадратов.

9. Напишите формулу для расчета дисперсии адекватности (N опытов, n дублей). Чему равно число степеней свободы для этого параметра?

10. Сформулируйте критерий адекватности уравнения регрессии любого порядка.

11. Сформулируйте необходимое и достаточное условие экстремума для однофакторного уравнения регрессии 2-го порядка. Напишите выражение для экстремума однофакторного уравнения регрессии 2-го порядка? Всегда ли оптимальное значение фактора находится внутри исследованного факторного пространства?

12. Напишите формулу для предельной абсолютной погрешности параметра $Y(X)$, рассчитанного по однофакторному уравнению регрессии 2-го порядка.

13. Однофакторное уравнение регрессии 2-го порядка строится в ортогональной форме $Y = b_0 X_0 + b_1 X + b_2 (X^2 - \lambda)$, где:

а) $\lambda = \dfrac{\sum_{j=1}^{N} X_j^2}{N}$; б) $\lambda = \sum_{j=1}^{N} X_j^2$; в) $\lambda = \dfrac{\sum_{j=1}^{N} X_j}{N}$.

14. Коэффициенты регрессии 2-го порядка однофакторного эксперимента рассчитываются по формулам:

а) $b_0 = \dfrac{\sum_{j=1}^{N} X_{0j} \overline{Y}_j}{\sum_{j=1}^{N} X_{0j}}$, $b_1 = \dfrac{\sum_{j=1}^{N} X_j \overline{Y}_j}{\sum_{j=1}^{N} X_j}$, $b_2 = \dfrac{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right) \overline{Y}_j}{\sum_{j=1}^{N} \left(X_j^2 - \lambda \right)}$;

б) $b_0 = \dfrac{\sum\limits_{j=1}^{N} X_{0j}\overline{Y}_j}{\sum\limits_{j=1}^{N} X_j^2}$, $b_1 = \dfrac{\sum\limits_{j=1}^{N} \overline{Y}_j}{\sum\limits_{j=1}^{N} X_j^2}$, $b_2 = \dfrac{\sum\limits_{j=1}^{N} (X_j^2 - \lambda)\overline{Y}_j}{\sum\limits_{j=1}^{N} X_j^2}$;

в) $b_0 = \dfrac{\sum\limits_{j=1}^{N} X_{0j}\overline{Y}_j}{N\sum\limits_{j=1}^{N} X_{0j}^2}$, $b_1 = \dfrac{\sum\limits_{j=1}^{N} X_j\overline{Y}_j}{N\sum\limits_{j=1}^{N} X_j^2}$, $b_2 = \dfrac{\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)\overline{Y}_j}{N\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)^2}$;

г) $b_0 = \dfrac{\sum\limits_{j=1}^{N} X_{0j}\overline{Y}_j}{\sum\limits_{j=1}^{N} X_{0j}^2}$, $b_1 = \dfrac{\sum\limits_{j=1}^{N} X_j^2\overline{Y}_j}{\sum\limits_{j=1}^{N} X_j^2}$, $b_2 = \dfrac{\sum\limits_{j=1}^{N} \left(X_{1j}^2 - \lambda\right)\overline{Y}_j}{\sum\limits_{j=1}^{N} \left(X_{1j}^2 - \lambda\right)^2}$.

15. В однофакторном эксперименте для РСП имеем:

а) $\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)^2 = \dfrac{4(N+1)(N^2-4)}{45(N-1)^3}$;

б) $\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)^2 = \dfrac{4N(N^2-4)}{45(N-1)^3}$;

в) $\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)^2 = \dfrac{4N(N+1)(N^2-4)}{45(N-1)^3}$.

16. Дисперсии значимости коэффициентов однофакторного уравнения регрессии 2-го порядка в обычной форме:

а) $S^2(b_1) = \dfrac{S^2_{\text{воспр}}}{n\sum\limits_{j=1}^{N} X_j^2}$, $S^2(b_2) = \dfrac{S^2_{\text{воспр}}}{n\sum\limits_{j=1}^{N} \left(X_j^2 - \lambda\right)^2}$,

б) $S^2(b_1) = \dfrac{S^2_{\text{воспр}}}{N\sum\limits_{j=1}^{N} X_j^2}$, $S^2(b_2) = \dfrac{S^2_{\text{воспр}}}{N\sum\limits_{j=1}^{N} \left(X_{1j}^2 - \lambda\right)^2}$.

17. Пусть температура сушильного агента в аппарате для сушки зерна является фактором x. Представляет интерес следующий диапазон изменения фактора x: x_{min} = 45 °С, x_{max} = 93 °С. Построить РСП для 9 опытов.

Ответ: РСП для 9 опытов

N	$x_j = T_j$, °С	X_j
1	45	–1,00
2	51	–0,75
3	57	–0,50
4	63	–0,25
5	69	0
6	75	0,25
7	81	0,50
8	87	0,75
9	93	1,00

ЗАКЛЮЧЕНИЕ

Моделирование стохастических систем, описываемых непрерывными количественными параметрами, подчиняющимися нормальному закону распределения, при воздействии на них количественных детерминированных факторов базируется на аппарате регрессионного анализа, основы которого подробно изложены в данном издании применительно к однофакторным и многофакторным моделям первого и второго порядка. Регрессионные модели более высоких порядков не содержат принципиальных отличий в методах расчета и статистического анализа параметров, поэтому при необходимости вполне могут быть освоены самостоятельно на базе данного курса.

Существует большое количество задач других классов, анализ которых базируется на широком спектре различных статистических методов. В их числе: дисперсионный анализ (количественный параметр – качественные факторы), ковариационный анализ (количественный параметр – количественные и качественные факторы), дискриминантный анализ (качественный параметр – количественные факторы), модели распознавания образов (качественные параметры – качественные факторы) и др. Рассмотрение этих методов лежит за рамками начального курса основ моделирования. В литературных источниках можно найти работы, рассматривающие подобные задачи. Необходимо принять во внимание и то, что широкое использование средств вычислительной техники в современных научных исследованиях привело к разработке большого числа специализированных программных пакетов математической статистики, начиная от пакета простейшего статистического анализа в составе MS Excel (Analysis ToolPak) до таких мощных программных комплексов, как STATISTICA (TIBCO Software Inc.) и SPSS Statistics (SPSS Inc.). Применение подобных программных пакетов избавляет исследователя от рутинных вычислительных процедур, но требует знания теоретических и методологических основ применения тех или иных методов статистического анализа.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Основной

1. Леонов, А. Н. Основы научных исследований в примерах и задачах : учебно-методическое пособие / А. Н. Леонов, М. М. Дечко, В. Б. Ловкис ; под ред. А. Н. Леонова. – Минск : БГАТУ, 2013. – 136 с.

2. Леонов, А. Н. Основы научных исследований и моделирование : тестовые задания / А. Н. Леонов, М. М. Дечко, В. Б. Ловкис ; под ред. А. Н. Леонова. – Минск : БГАТУ, 2010. – 80 с.

3. Леонов, А. Н. Основы научных исследований и моделирования : учебно-методический комплекс / А. Н. Леонов, М. М. Дечко, В. Б. Ловкис. – Минск : БГАТУ, 2010. – 276 с.

Дополнительный

4. Воробьева, Ф. И. Информатика. MS Excel 2010 : учебное пособие / Ф. И. Воробьева, Е. С. Воробьев. – Казань : Изд-во КНИТУ, 2014. – 96 с.

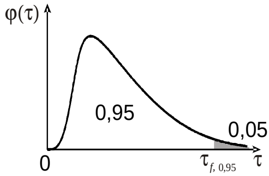
5. Кирьянов, Д. В. Mathcad 15/Mathcad Prime 1.0. / Д. В. Кирьянов. – СПб. : БХВ-Петербург, 2012. – 432 с.

6. Нагорский, И. С. Основы научных исследований : пособие по изучению дисциплины : в 4 ч. / И. С. Нагорский. – Минск : БГАТУ, 2006. – Ч. 1, 2. – 132 с.

7. Нагорский, И. С. Основы научных исследований : пособие по изучению дисциплины : в 4 ч. /И. С. Нагорский, В. Б. Ловкис, Ю. Т. Антонишин. – Минск : БГАТУ, 2008. – Ч. 3, 4. – 108 с.

Приложение 1

Критерий Смирнова–Граббса



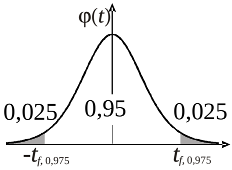
$$\tau = \max \left| \frac{Y_i - \overline{Y}}{S} \right|$$

Табличные значения критерия Смирнова–Граббса $\tau_{f;\ 0,95}$
при числе степеней свободы f и доверительной вероятности $p = 0,95$

f	$\tau_{f;\ 0,95}$	f	$\tau_{f;\ 0,95}$
1	1,412	26	2,764
2	1,689	27	2,778
3	1,869	28	2,792
4	1,996	29	2,805
5	2,093	30	2,818
6	2,172	31	2,830
7	2,238	32	2,842
8	2,294	33	2,853
9	2,343	34	2,864
10	2,387	35	2,874
11	2,426	36	2,885
12	2,461	37	2,894
13	2,494	38	2,904
14	2,523	39	2,913
15	2,551	40	2,922
16	2,577	41	2,931
17	2,601	42	2,940
18	2,623	43	2,948
19	2,644	44	2,956
20	2,664	45	2,964
21	2,683	46	2,972
22	2,701	47	2,980
23	2,718	48	2,987
24	2,734	49	2,994
25	2,749	50	3,001

Приложение 2

Критерий Стьюдента



$$t = \frac{(\overline{Y} - \mu)\sqrt{n}}{S}$$

Табличные значения критерия Стьюдента $t_{f;\ 0,95}$

при числе степеней свободы f и доверительной вероятности $p = 0,95$

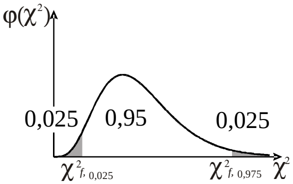
f	$t_{f;\ 0,95}$	f	$t_{f;\ 0,95}$	f	$t_{f;\ 0,95}$
1	12,706	36	2,028	71	1,994
2	4,303	37	2,026	72	1,993
3	3,182	38	2,024	73	1,993
4	2,776	39	2,023	74	1,993
5	2,571	40	2,021	75	1,992
6	2,447	41	2,020	76	1,992
7	2,365	42	2,018	77	1,991
8	2,306	43	2,017	78	1,991
9	2,262	44	2,015	79	1,990
10	2,228	45	2,014	80	1,990
11	2,201	46	2,013	81	1,990
12	2,179	47	2,012	82	1,989
13	2,160	48	2,011	83	1,989
14	2,145	49	2,010	84	1,989
15	2,131	50	2,009	85	1,988
16	2,120	51	2,008	86	1,988
17	2,110	52	2,007	87	1,988
18	2,101	53	2,006	88	1,987
19	2,093	54	2,005	89	1,987
20	2,086	55	2,004	90	1,987
21	2,080	56	2,003	91	1,986
22	2,074	57	2,002	92	1,986
23	2,069	58	2,002	93	1,986
24	2,064	59	2,001	94	1,986
25	2,060	60	2,000	95	1,985

Окончание приложения 2

f	$t_{f;\ 0,95}$	f	$t_{f;\ 0,95}$	f	$t_{f;\ 0,95}$
26	2,056	61	2,000	96	1,985
27	2,052	62	1,999	97	1,985
28	2,048	63	1,998	98	1,984
29	2,045	64	1,998	99	1,984
30	2,042	65	1,997	100	1,984
31	2,040	66	1,997	101	1,984
32	2,037	67	1,996	102	1,983
33	2,035	68	1,995	103	1,983
34	2,032	69	1,995	104	1,983
35	2,030	70	1,960	105	1,983

Приложение 3

Критерий Пирсона



$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

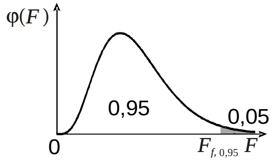
Табличные значения критерия Пирсона $\chi^2_{f;\,0,025}$ и $\chi^2_{f;\,0,975}$
при числе степеней свободы *f* и доверительной вероятности $p = 0,95$

f	$\chi^2_{f;\,0,025}$	$\chi^2_{f;\,0,975}$	f	$\chi^2_{f;\,0,025}$	$\chi^2_{f;\,0,975}$
1	0,000982	5,024	31	17,539	48,232
2	0,0506	7,378	32	18,291	49,480
3	0,216	9,348	33	19,047	50,725
4	0,484	11,143	34	19,806	51,966
5	0,831	12,833	35	20,569	53,203
6	1,237	14,449	36	21,336	54,437
7	1,690	16,013	37	22,106	55,668
8	2,180	17,535	38	22,878	56,896
9	2,700	19,023	39	23,654	58,120
10	3,247	20,483	40	24,433	59,342
11	3,816	21,920	41	25,215	60,561
12	4,404	23,337	42	25,999	61,777
13	5,009	24,736	43	26,785	62,990
14	5,629	26,119	44	27,575	64,201
15	6,262	27,488	45	28,366	65,410
16	6,908	28,845	46	29,160	66,617
17	7,564	30,191	47	29,956	67,821
18	8,231	31,526	48	30,755	69,023
19	8,907	32,852	49	31,555	70,222
20	9,591	34,170	50	32,357	71,420
21	10,283	35,479	51	33,162	72,616
22	10,982	36,781	52	33,968	73,810

f	$\chi^2_{f;\,0,025}$	$\chi^2_{f;\,0,975}$	f	$\chi^2_{f;\,0,025}$	$\chi^2_{f;\,0,975}$
23	11,689	38,076	53	34,776	75,002
24	12,401	39,364	54	35,586	76,192
25	13,120	40,646	55	36,398	77,380
26	13,844	41,923	56	37,212	78,567
27	14,573	43,195	57	38,027	79,752
28	15,308	44,461	58	38,844	80,936
29	16,047	45,722	59	39,662	82,117
30	16,791	46,979	60	40,482	83,298

Приложение 4

Критерий Фишера



$$F = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)}$$

Табличные значения критерия Фишера $F_{f_{числ};\, f_{знам};\, 0,95}$

при числах степеней свободы $f_{числ}$ и $f_{знам}$ и доверительной вероятности $p = 0,95$

$f_{знам}$	$f_{числ}$																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9	246,5	246,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,44
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763	8,745	8,729	8,715	8,703	8,692	8,683
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,891	5,873	5,858	5,844	5,832
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,655	4,636	4,619	4,604	4,590
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,976	3,956	3,938	3,922	3,908
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,550	3,529	3,511	3,494	3,480
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,259	3,237	3,218	3,202	3,187
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,073	3,048	3,025	3,006	2,989	2,974
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,913	2,887	2,865	2,845	2,828	2,812
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818	2,788	2,761	2,739	2,719	2,701	2,685
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717	2,687	2,660	2,637	2,617	2,599	2,583

Продолжение приложения 4

$f_{знам}$	$f_{числ}$																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,635	2,604	2,577	2,554	2,533	2,515	2,499
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,565	2,534	2,507	2,484	2,463	2,445	2,428
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,507	2,475	2,448	2,424	2,403	2,385	2,368
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,456	2,425	2,397	2,373	2,352	2,333	2,317
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,413	2,381	2,353	2,329	2,308	2,289	2,272
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,374	2,342	2,314	2,290	2,269	2,250	2,233
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,340	2,308	2,280	2,256	2,234	2,215	2,198
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,310	2,278	2,250	2,225	2,203	2,184	2,167
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321	2,283	2,250	2,222	2,197	2,176	2,156	2,139
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,259	2,226	2,198	2,173	2,151	2,131	2,114
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,236	2,204	2,175	2,150	2,128	2,109	2,091
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,216	2,183	2,155	2,130	2,108	2,088	2,070
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,198	2,165	2,136	2,111	2,089	2,069	2,051
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,181	2,148	2,119	2,094	2,072	2,052	2,034
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,166	2,132	2,103	2,078	2,056	2,036	2,018
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,151	2,118	2,089	2,064	2,041	2,021	2,003
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,138	2,104	2,075	2,050	2,027	2,007	1,989
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,126	2,092	2,063	2,037	2,015	1,995	1,976
31	4,160	3,305	2,911	2,679	2,523	2,409	2,323	2,255	2,199	2,153	2,114	2,080	2,051	2,026	2,003	1,983	1,965
32	4,149	3,295	2,901	2,668	2,512	2,399	2,313	2,244	2,189	2,142	2,103	2,070	2,040	2,015	1,992	1,972	1,953
33	4,139	3,285	2,892	2,659	2,503	2,389	2,303	2,235	2,179	2,133	2,093	2,060	2,030	2,004	1,982	1,961	1,943
34	4,130	3,276	2,883	2,650	2,494	2,380	2,294	2,225	2,170	2,123	2,084	2,050	2,021	1,995	1,972	1,952	1,933
35	4,121	3,267	2,874	2,641	2,485	2,372	2,285	2,217	2,161	2,114	2,075	2,041	2,012	1,986	1,963	1,942	1,924
36	4,113	3,259	2,866	2,634	2,477	2,364	2,277	2,209	2,153	2,106	2,067	2,033	2,003	1,977	1,954	1,934	1,915

$f_{знам}$	$f_{числ}$																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
37	4,105	3,252	2,859	2,626	2,470	2,356	2,270	2,201	2,145	2,098	2,059	2,025	1,995	1,969	1,946	1,926	1,907
38	4,098	3,245	2,852	2,619	2,463	2,349	2,262	2,194	2,138	2,091	2,051	2,017	1,988	1,962	1,939	1,918	1,899
39	4,091	3,238	2,845	2,612	2,456	2,342	2,255	2,187	2,131	2,084	2,044	2,010	1,981	1,954	1,931	1,911	1,892
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077	2,038	2,003	1,974	1,948	1,924	1,904	1,885
41	4,079	3,226	2,833	2,600	2,443	2,330	2,243	2,174	2,118	2,071	2,031	1,997	1,967	1,941	1,918	1,897	1,879
42	4,073	3,220	2,827	2,594	2,438	2,324	2,237	2,168	2,112	2,065	2,025	1,991	1,961	1,935	1,912	1,891	1,872
43	4,067	3,214	2,822	2,589	2,432	2,318	2,232	2,163	2,106	2,059	2,020	1,985	1,955	1,929	1,906	1,885	1,866
44	4,062	3,209	2,816	2,584	2,427	2,313	2,226	2,157	2,101	2,054	2,014	1,980	1,950	1,924	1,900	1,879	1,861
45	4,057	3,204	2,812	2,579	2,422	2,308	2,221	2,152	2,096	2,049	2,009	1,974	1,945	1,918	1,895	1,874	1,855
46	4,052	3,200	2,807	2,574	2,417	2,304	2,216	2,147	2,091	2,044	2,004	1,969	1,940	1,913	1,890	1,869	1,850
47	4,047	3,195	2,802	2,570	2,413	2,299	2,212	2,143	2,086	2,039	1,999	1,965	1,935	1,908	1,885	1,864	1,845
48	4,043	3,191	2,798	2,565	2,409	2,295	2,207	2,138	2,082	2,035	1,995	1,960	1,930	1,904	1,880	1,859	1,840
49	4,038	3,187	2,794	2,561	2,404	2,290	2,203	2,134	2,077	2,030	1,990	1,956	1,926	1,899	1,876	1,855	1,836
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026	1,986	1,952	1,921	1,895	1,871	1,850	1,831
51	4,030	3,179	2,786	2,553	2,397	2,283	2,195	2,126	2,069	2,022	1,982	1,947	1,917	1,891	1,867	1,846	1,827
52	4,027	3,175	2,783	2,550	2,393	2,279	2,192	2,122	2,066	2,018	1,978	1,944	1,913	1,887	1,863	1,842	1,823
53	4,023	3,172	2,779	2,546	2,389	2,275	2,188	2,119	2,062	2,015	1,975	1,940	1,910	1,883	1,859	1,838	1,819
54	4,020	3,168	2,776	2,543	2,386	2,272	2,185	2,115	2,059	2,011	1,971	1,936	1,906	1,879	1,856	1,835	1,816
55	4,016	3,165	2,773	2,540	2,383	2,269	2,181	2,112	2,055	2,008	1,968	1,933	1,903	1,876	1,852	1,831	1,812
56	4,013	3,162	2,769	2,537	2,380	2,266	2,178	2,109	2,052	2,005	1,964	1,930	1,899	1,873	1,849	1,828	1,809
57	4,010	3,159	2,766	2,534	2,377	2,263	2,175	2,106	2,049	2,001	1,961	1,926	1,896	1,869	1,846	1,824	1,805
58	4,007	3,156	2,764	2,531	2,374	2,260	2,172	2,103	2,046	1,998	1,958	1,923	1,893	1,866	1,842	1,821	1,802
59	4,004	3,153	2,761	2,528	2,371	2,257	2,169	2,100	2,043	1,995	1,955	1,920	1,890	1,863	1,839	1,818	1,799
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,952	1,917	1,887	1,860	1,836	1,815	1,796

Продолжение приложения 4

$f_{знам}$	$f_{числ}$																
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	247,3	247,7	248,0	248,3	248,6	248,8	249,1	249,3	249,5	249,6	249,8	250,0	250,1	250,2	250,4	250,5	250,6
2	19,44	19,44	19,45	19,45	19,45	19,45	19,45	19,46	19,46	19,46	19,46	19,46	19,46	19,46	19,46	19,47	19,47
3	8,675	8,667	8,660	8,654	8,648	8,643	8,639	8,634	8,630	8,626	8,623	8,620	8,617	8,614	8,611	8,609	8,606
4	5,821	5,811	5,803	5,795	5,787	5,781	5,774	5,769	5,763	5,759	5,754	5,750	5,746	5,742	5,739	5,735	5,732
5	4,579	4,568	4,558	4,549	4,541	4,534	4,527	4,521	4,515	4,510	4,505	4,500	4,496	4,492	4,488	4,484	4,481
6	3,896	3,884	3,874	3,865	3,856	3,849	3,841	3,835	3,829	3,823	3,818	3,813	3,808	3,804	3,800	3,796	3,792
7	3,467	3,455	3,445	3,435	3,426	3,418	3,410	3,404	3,397	3,391	3,386	3,381	3,376	3,371	3,367	3,363	3,359
8	3,173	3,161	3,150	3,140	3,131	3,123	3,115	3,108	3,102	3,095	3,090	3,084	3,079	3,075	3,070	3,066	3,062
9	2,960	2,948	2,936	2,926	2,917	2,908	2,900	2,893	2,886	2,880	2,874	2,869	2,864	2,859	2,854	2,850	2,846
10	2,798	2,785	2,774	2,764	2,754	2,745	2,737	2,730	2,723	2,716	2,710	2,705	2,700	2,695	2,690	2,686	2,681
11	2,671	2,658	2,646	2,636	2,626	2,617	2,609	2,601	2,594	2,588	2,582	2,576	2,570	2,565	2,561	2,556	2,552
12	2,568	2,555	2,544	2,533	2,523	2,514	2,505	2,498	2,491	2,484	2,478	2,472	2,466	2,461	2,456	2,452	2,447
13	2,484	2,471	2,459	2,448	2,438	2,429	2,420	2,412	2,405	2,398	2,392	2,386	2,380	2,375	2,370	2,366	2,361
14	2,413	2,400	2,388	2,377	2,367	2,357	2,349	2,341	2,333	2,326	2,320	2,314	2,308	2,303	2,298	2,293	2,289
15	2,353	2,340	2,328	2,316	2,306	2,297	2,288	2,280	2,272	2,265	2,259	2,253	2,247	2,241	2,236	2,232	2,227
16	2,302	2,288	2,276	2,264	2,254	2,244	2,235	2,227	2,220	2,212	2,206	2,200	2,194	2,188	2,183	2,178	2,174
17	2,257	2,243	2,230	2,219	2,208	2,199	2,190	2,181	2,174	2,167	2,160	2,154	2,148	2,142	2,137	2,132	2,127
18	2,217	2,203	2,191	2,179	2,168	2,159	2,150	2,141	2,134	2,126	2,119	2,113	2,107	2,102	2,096	2,091	2,087
19	2,182	2,168	2,155	2,144	2,133	2,123	2,114	2,106	2,098	2,090	2,084	2,077	2,071	2,066	2,060	2,055	2,050
20	2,151	2,137	2,124	2,112	2,102	2,092	2,082	2,074	2,066	2,059	2,052	2,045	2,039	2,033	2,028	2,023	2,018
21	2,123	2,109	2,096	2,084	2,073	2,063	2,054	2,045	2,037	2,030	2,023	2,016	2,010	2,004	1,999	1,994	1,989
22	2,098	2,084	2,071	2,059	2,048	2,038	2,028	2,020	2,012	2,004	1,997	1,990	1,984	1,978	1,973	1,968	1,963
23	2,075	2,061	2,048	2,036	2,025	2,014	2,005	1,996	1,988	1,981	1,973	1,967	1,961	1,955	1,949	1,944	1,939
24	2,054	2,040	2,027	2,015	2,003	1,993	1,984	1,975	1,967	1,959	1,952	1,945	1,939	1,933	1,927	1,922	1,917

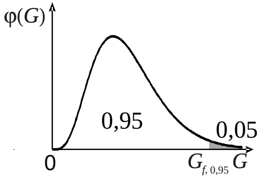
$f_{знам}$	$f_{числ}$																
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
25	2,035	2,021	2,007	1,995	1,984	1,974	1,964	1,955	1,947	1,939	1,932	1,926	1,919	1,913	1,908	1,902	1,897
26	2,018	2,003	1,990	1,978	1,966	1,956	1,946	1,938	1,929	1,921	1,914	1,907	1,901	1,895	1,889	1,884	1,879
27	2,002	1,987	1,974	1,961	1,950	1,940	1,930	1,921	1,913	1,905	1,898	1,891	1,884	1,878	1,872	1,867	1,862
28	1,987	1,972	1,959	1,946	1,935	1,924	1,915	1,906	1,897	1,889	1,882	1,875	1,869	1,863	1,857	1,851	1,846
29	1,973	1,958	1,945	1,932	1,921	1,910	1,901	1,891	1,883	1,875	1,868	1,861	1,854	1,848	1,842	1,837	1,832
30	1,960	1,945	1,932	1,919	1,908	1,897	1,887	1,878	1,870	1,862	1,854	1,847	1,841	1,835	1,829	1,823	1,818
31	1,948	1,933	1,920	1,907	1,896	1,885	1,875	1,866	1,857	1,849	1,842	1,835	1,828	1,822	1,816	1,811	1,805
32	1,937	1,922	1,908	1,896	1,884	1,873	1,864	1,854	1,846	1,838	1,830	1,823	1,817	1,810	1,804	1,799	1,794
33	1,926	1,911	1,898	1,885	1,873	1,863	1,853	1,844	1,835	1,827	1,819	1,812	1,806	1,799	1,793	1,788	1,783
34	1,917	1,902	1,888	1,875	1,863	1,853	1,843	1,833	1,825	1,817	1,809	1,802	1,795	1,789	1,783	1,777	1,772
35	1,907	1,892	1,878	1,866	1,854	1,843	1,833	1,824	1,815	1,807	1,799	1,792	1,786	1,779	1,773	1,768	1,762
36	1,899	1,883	1,870	1,857	1,845	1,834	1,824	1,815	1,806	1,798	1,790	1,783	1,776	1,770	1,764	1,758	1,753
37	1,890	1,875	1,861	1,848	1,837	1,826	1,816	1,806	1,798	1,789	1,782	1,775	1,768	1,761	1,755	1,750	1,744
38	1,883	1,867	1,853	1,841	1,829	1,818	1,808	1,798	1,790	1,781	1,774	1,766	1,760	1,753	1,747	1,741	1,736
39	1,875	1,860	1,846	1,833	1,821	1,810	1,800	1,791	1,782	1,774	1,766	1,759	1,752	1,745	1,739	1,733	1,728
40	1,868	1,853	1,839	1,826	1,814	1,803	1,793	1,783	1,775	1,766	1,759	1,751	1,744	1,738	1,732	1,726	1,721
41	1,862	1,846	1,832	1,819	1,807	1,796	1,786	1,777	1,768	1,759	1,752	1,744	1,737	1,731	1,725	1,719	1,713
42	1,855	1,840	1,826	1,813	1,801	1,790	1,780	1,770	1,761	1,753	1,745	1,738	1,731	1,724	1,718	1,712	1,707
43	1,849	1,834	1,820	1,807	1,795	1,784	1,773	1,764	1,755	1,747	1,739	1,731	1,724	1,718	1,712	1,706	1,700
44	1,844	1,828	1,814	1,801	1,789	1,778	1,767	1,758	1,749	1,741	1,733	1,725	1,718	1,712	1,706	1,700	1,694
45	1,838	1,823	1,808	1,795	1,783	1,772	1,762	1,752	1,743	1,735	1,727	1,720	1,713	1,706	1,700	1,694	1,688
46	1,833	1,817	1,803	1,790	1,778	1,767	1,756	1,747	1,738	1,729	1,721	1,714	1,707	1,700	1,694	1,688	1,683
47	1,828	1,812	1,798	1,785	1,773	1,762	1,751	1,742	1,733	1,724	1,716	1,709	1,702	1,695	1,689	1,683	1,677
48	1,823	1,807	1,793	1,780	1,768	1,757	1,746	1,737	1,728	1,719	1,711	1,704	1,697	1,690	1,684	1,678	1,672

Окончание приложения 4

$f_{знам}$	$f_{числ}$																
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
49	1,819	1,803	1,789	1,775	1,763	1,752	1,742	1,732	1,723	1,714	1,706	1,699	1,692	1,685	1,679	1,673	1,667
50	1,814	1,798	1,784	1,771	1,759	1,748	1,737	1,727	1,718	1,710	1,702	1,694	1,687	1,680	1,674	1,668	1,662
51	1,810	1,794	1,780	1,767	1,754	1,743	1,733	1,723	1,714	1,705	1,697	1,690	1,683	1,676	1,670	1,664	1,658
52	1,806	1,790	1,776	1,763	1,750	1,739	1,729	1,719	1,710	1,701	1,693	1,685	1,678	1,672	1,665	1,659	1,654
53	1,802	1,786	1,772	1,759	1,746	1,735	1,725	1,715	1,706	1,697	1,689	1,681	1,674	1,667	1,661	1,655	1,649
54	1,798	1,782	1,768	1,755	1,743	1,731	1,721	1,711	1,702	1,693	1,685	1,677	1,670	1,663	1,657	1,651	1,645
55	1,795	1,779	1,764	1,751	1,739	1,727	1,717	1,707	1,698	1,689	1,681	1,674	1,666	1,660	1,653	1,647	1,641
56	1,791	1,775	1,761	1,748	1,735	1,724	1,713	1,703	1,694	1,686	1,678	1,670	1,663	1,656	1,650	1,643	1,638
57	1,788	1,772	1,757	1,744	1,732	1,720	1,710	1,700	1,691	1,682	1,674	1,666	1,659	1,652	1,646	1,640	1,634
58	1,785	1,769	1,754	1,741	1,729	1,717	1,706	1,697	1,687	1,679	1,671	1,663	1,656	1,649	1,642	1,636	1,631
59	1,781	1,766	1,751	1,738	1,725	1,714	1,703	1,693	1,684	1,675	1,667	1,660	1,652	1,646	1,639	1,633	1,627
60	1,778	1,763	1,748	1,735	1,722	1,711	1,700	1,690	1,681	1,672	1,664	1,656	1,649	1,642	1,636	1,630	1,624

Приложение 5

Критерий Кохрена



$$G = \frac{\max S_j^2}{\sum\limits_{j=1}^{N} S_j^2}$$

Табличные значения критерия Кохрена $G_{f_{\text{числ}};\,f_{\text{знам}};\,0{,}95}$

при числах степеней свободы $f_{\text{числ}}$ и $f_{\text{знам}}$ и доверительной вероятности $p = 0{,}95$

$f_{\text{знам}}$	$f_{\text{числ}}$							$f_{\text{знам}}$	$f_{\text{знам}}$						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
2	0,998	0,975	0,939	0,906	0,877	0,853	0,833	16	0,452	0,319	0,262	0,230	0,208	0,193	0,181
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	17	0,434	0,305	0,250	0,219	0,198	0,183	0,172
4	0,906	0,768	0,684	0,629	0,589	0,560	0,536	18	0,418	0,293	0,240	0,209	0,189	0,175	0,164
5	0,841	0,684	0,598	0,544	0,506	0,478	0,456	19	0,403	0,281	0,230	0,200	0,181	0,167	0,157
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	20	0,389	0,270	0,221	0,192	0,174	0,160	0,150
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	21	0,377	0,261	0,212	0,185	0,167	0,154	0,144
8	0,680	0,516	0,438	0,391	0,359	0,336	0,318	22	0,365	0,252	0,204	0,178	0,160	0,148	0,138
9	0,638	0,477	0,403	0,358	0,328	0,307	0,290	23	0,354	0,243	0,197	0,171	0,155	0,142	0,133
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	24	0,343	0,235	0,191	0,166	0,149	0,137	0,129
11	0,570	0,417	0,348	0,308	0,281	0,262	0,247	25	0,334	0,228	0,185	0,160	0,144	0,133	0,124
12	0,541	0,392	0,326	0,288	0,262	0,244	0,230	26	0,325	0,221	0,179	0,155	0,139	0,128	0,120
13	0,515	0,371	0,307	0,271	0,246	0,229	0,215	28	0,308	0,209	0,168	0,146	0,131	0,121	0,113
14	0,492	0,352	0,291	0,255	0,232	0,215	0,202	30	0,293	0,198	0,159	0,138	0,124	0,114	0,106
15	0,471	0,335	0,276	0,242	0,220	0,203	0,191	32	0,279	0,188	0,151	0,130	0,117	0,108	0,100

Приложение 6

Вывод математических формул

1. Определения (Y и Z – случайные величины).
Выборочное среднее:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \qquad (П6.1)$$

$$\overline{Z} = \frac{\sum_{i=1}^{n} Z_i}{n}. \qquad (П6.2)$$

Выборочная дисперсия:

$$S^2(Y) = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n}, \qquad (П6.3)$$

$$S^2(Z) = \frac{\sum_{i=1}^{n} (Z_i - \overline{Z})^2}{n}. \qquad (П6.4)$$

Выборочное стандартное отклонение:

$$S(Y) = \sqrt{S^2(Y)}, \qquad (П6.5)$$

$$S(Z) = \sqrt{S^2(Z)}. \qquad (П6.6)$$

Выборочный коэффициент корреляции между Y и Z:

$$r(YZ) = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})(Z_i - \overline{Z})}{S(Y) \cdot S(Z)}. \qquad (П6.7)$$

Для независимых величин Y и Z коэффициент корреляции $r(Y, Z) = 0$.

2. Доказать, что

$$\overline{CY} = C\overline{Y}. \qquad (П6.8)$$

С учетом уравнения (П6.1) получаем:

$$\overline{CY} = \frac{\sum_{i=1}^{n}(CY)_i}{n} = \frac{\sum_{i=1}^{n}CY_i}{n} = \frac{C\sum_{i=1}^{n}Y_i}{n} = C\overline{Y}.$$

3. Доказать, что

$$\overline{Y+Z} = \overline{Y} + \overline{Z}. \qquad (П6.9)$$

С учетом уравнений (П6.1), (П6.2) получаем:

$$\overline{Y+Z} = \frac{\sum_{i=1}^{n}(Y+Z)_i}{n} = \frac{\sum_{i=1}^{n}Y_i + Z_i}{n} = \frac{\sum_{i=1}^{n}Y_i}{n} + \frac{\sum_{i=1}^{n}Z_i}{n} = \overline{Y} + \overline{Z}.$$

4. Доказать, что

$$S^2(C) = 0. \qquad (П6.10)$$

С учетом уравнения (П6.3) получаем:

$$S^2(C) = \frac{\sum_{i=1}^{n}\left(C_i - \overline{C}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}(C - C)^2}{n-1} = 0.$$

5. Доказать, что

$$S^2(CY) = C^2 S^2(Y). \qquad (П6.11)$$

С учетом уравнений (П6.3), (П6.8) получаем:

$$S^2(CY) = \frac{\sum_{i=1}^{n}\left((CY)_i - \overline{CY}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}\left(CY_i - C\overline{Y}\right)^2}{n-1} =$$

$$= \frac{C^2 \sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1} = C^2 S^2(Y).$$

Аналогично можно показать, что в случае независимости случайных величин Y и Z выполняется равенство $S^2(Y - Z) = S^2(Y) + S^2(Z)$.

6. Доказать, что

$$S^2(\overline{Y}) = \frac{S^2(Y)}{n} \text{ и } S(\overline{Y}) = \frac{S(Y)}{\sqrt{n}}. \quad \text{(П6.12)}$$

Параметр $\overline{Y}$ рассматривается как случайная величина, а не константа. *Пояснение*: пусть случайная величина Y принадлежит некоторой генеральной совокупности с параметрами μ и σ. Выборка $Y_1', Y_2', ..., Y_n'$ характеризуется выборочным средним $\overline{Y'}$. Если на базе той же генеральной совокупности провести еще несколько серий наблюдений такого же объема, то получим: $Y_1'', Y_2'', ..., Y_n''$; $Y_1''', Y_2''', ..., Y_n'''$ и т. п. Все они будут характеризоваться выборочными средними $\overline{Y''}$, $\overline{Y'''}$ и т. п. Поэтому случайные величины Y_1, Y_2, …, Y_n в серии из n наблюдений можно рассматривать как случайные величины с дисперсиями $S^2(Y_1), S^2(Y_2), ..., S^2(Y_n)$. Поскольку эти случайные величины возникают при измерении одной и той же случайной величины Y, то дисперсии их естественно считать одинаковыми (то есть случайные величины Y_2, …, Y_n можно считать различными значениями Y_1):

$$S^2(Y_1) = S^2(Y_2) = ... = S^2(Y_n). \quad \text{(П6.13)}$$

Применив уравнения (П6.1), (П6.2) к $S^2(\overline{Y})$, получим:

$$S^2(\overline{Y}) = S^2\left(\frac{Y_1 + Y_2 + ... + Y_n}{n}\right) = \frac{1}{n^2}\left[S^2(Y_1) + S^2(Y_2) + ... + S^2(Y_1)\right] =$$

$$= \frac{1}{n^2} n S^2(Y) = \frac{S^2(Y)}{n}, \text{ или } S(\overline{Y}) = \frac{S(Y)}{\sqrt{n}}.$$

$$\sum_{j=1}^{n} j = 1 + \ldots + N = \frac{N(N+1)}{2}; \qquad (П6.14)$$

$$\sum_{j=1}^{N} j^2 = 1 + \ldots + N^2 = \frac{N(N+1)(2N+1)}{6}; \qquad (П6.15)$$

$$\sum_{j=1}^{N} j^3 = 1 + \ldots + N^3 = \frac{N^2(N+1)^2}{4}; \qquad (П6.16)$$

$$\sum_{j=1}^{N} j^4 = 1 + \ldots + N^4 = \frac{N(N+1)(2N+1)(3N^2+3N-1)}{30}. \qquad (П6.17)$$

Уравнения (П6.14)–(П6.17) доказываются методом математической индукции:

1) проверить справедливость уравнений (П6.14)–(П6.17) для малого числа N, например, для $N = 2$;

2) предположить, что уравнения (П6.14)– (П6.17) справедливы для $N = k - 1$;

3) доказать, исходя из этого предположения, справедливость уравнений (П6.14)–(П6.17) для $N = k$.

Адаптированные данные

$$\sum_{j=1}^{N} (j-1) = \frac{(N-1)N}{2}; \qquad (П6.18)$$

$$\sum_{j=1}^{N} (j-1)^2 = \frac{(N-1)N(2N-1)}{6}; \qquad (П6.19)$$

$$\sum_{j=1}^{N} (j-1)^3 = \frac{(N-1)^2 N^2}{4}; \qquad (П6.20)$$

$$\sum_{j=1}^{N} (j-1)^4 = \frac{(N-1)N(2N-1)(3N^2-3N-1)}{30}. \qquad (П6.21)$$

Уравнения (П6.18)–(П6.21) получаются из уравнений (П6.14)–(П6.17) путем замены $N \to (N-1)$.

7. Дано: а) план РСП в натуральных значениях фактора: $x_j = x_{\min} + \dfrac{j-1}{N-1}(x_{\max} - x_{\min})$ (см. уравнение (45)), б) взаимосвязь нормированных и натуральных значений факторов X и x: $X_j = \dfrac{2x_j - (x_{\max} + x_{\min})}{x_{\max} - x_{\min}}$ (см. уравнение (54)–(57)). Доказать, что

$$X_j = 2 \cdot \frac{j-1}{N-1} - 1. \qquad \text{(П6.22)}$$

Используя уравнения (54)–(57), получаем:

$$X_j = \frac{2x_j - (x_{\max} + x_{\min})}{x_{\max} - x_{\min}} = \frac{2\left(x_{\min} + \dfrac{j-1}{N-1}(x_{\max} - x_{\min})\right) - (x_{\max} + x_{\min})}{x_{\max} - x_{\min}} =$$

$$= \frac{2x_{\min} + 2\dfrac{j-1}{N-1}(x_{\max} - x_{\min}) - (x_{\max} + x_{\min})}{x_{\max} - x_{\min}} =$$

$$= \frac{2\dfrac{j-1}{N-1}(x_{\max} - x_{\min}) - (x_{\max} - x_{\min})}{x_{\max} - x_{\min}} = 2\frac{j-1}{N-1} - 1.$$

8. Доказать, что для РСП факторы X_0 и X ортогональны:

$$\sum_{j=1}^{N} X_{0j} X_j = \sum_{j=1}^{N} X_j = \sum_{j=1}^{N}\left(\frac{2(j-1)}{N-1} - 1\right) = 0. \qquad \text{(П6.23)}$$

С учетом того, что $X_{0j} = 1$, используя уравнения (П6.18) (П6.22), получаем:

$$\sum_{j=1}^{N}\left(\frac{2(j-1)}{N-1} - 1\right) = \frac{2}{N-1}\sum_{j=1}^{N}(j-1) - \sum_{j=1}^{N} 1 = \frac{2}{N-1} \cdot \frac{(N-1)N}{2} - N = 0.$$

Из уравнения (П.23) следует, что

$$\sum_{j=1}^{N} X_j = 0\,, \tag{П6.24}$$

$$\overline{X} = \frac{\sum_{j=1}^{N} X_j}{N} = 0\,. \tag{П6.25}$$

9. Доказать, что

$$\sum_{j=1}^{N} X_j^2 = \frac{N(N+1)}{3(N-1)}\,. \tag{П6.26}$$

Используя уравнения (П6.18), (П6.19), (П6.22), получаем:

$$\sum_{j=1}^{N} X_j^2 = \sum_{j=1}^{N}\left(\frac{2(j-1)}{N-1}-1\right)^2 = \sum_{j=1}^{N}\frac{4(j-1)^2}{(N-1)^2} - \sum_{j=1}^{N} 2\frac{2(j-1)}{N-1} + N\sum_{j=1}^{N} 1 =$$

$$= \frac{4}{(N-1)^2}\sum_{j=1}^{N}(j-1)^2 - \frac{4}{N-1}\sum_{j=1}^{N}(j-1) + N\sum_{j=1}^{N} 1 =$$

$$= \frac{4}{(N-1)^2}\cdot\frac{(N-1)N(2N-1)}{6} - \frac{4}{N-1}\cdot\frac{(N-1)N}{2} + N =$$

$$= \frac{2N(2N-1)}{3(N-1)} - 2N + N = N\left[\frac{2(2N-1)}{3(N-1)} - 1\right] =$$

$$= N\frac{4N-2-3N+3}{3(N-1)} = \frac{N(N+1)}{3(N-1)}\,.$$

Из уравнения (П6.26) следует, что

$$\overline{X^2} = \frac{\sum_{j=1}^{N} X_j^2}{N} = \frac{N+1}{3(N-1)}\,. \tag{П6.27}$$

10. Доказать, что решение системы уравнений:

$$\begin{cases} \sum_{j=1}^{N} X_{0j}(X_j^2 - \gamma_1 X_j - \lambda_1) = 0, \\ \sum_{j=1}^{N} X_j (X_j^2 - \gamma_1 X_j - \lambda_1) = 0, \end{cases}$$

имеет следующий вид:

$$\begin{cases} \gamma_1 = 0; \\ \lambda_1 = \dfrac{\sum_{j=1}^{N} X_j^2}{N} = \overline{X^2} = \dfrac{N+1}{3(N-1)}. \end{cases} \qquad (П6.28)$$

Докажем сначала справедливость для РСП вспомогательного равенства:

$$\sum_{j=1}^{N} X_j^3 = 0. \qquad (П6.29)$$

Используя для РСП уравнения (П6.18)–(П6.21), (П6.22), получаем:

$$\sum_{j=1}^{N} X_j^3 = \sum_{j=1}^{N}\left(\frac{2(j-1)}{N-1} - 1\right)^3 = \frac{8}{(N-1)^3}\sum_{j=1}^{N}(j-1)^3 - 3\cdot\frac{4}{(N-1)^2}\sum_{j=1}^{N}(j-1)^2 +$$

$$+3\cdot\frac{2}{N-1}\sum_{j=1}^{N}(j-1) - \sum_{j=1}^{N}1 = \frac{8}{(N-1)^3}\cdot\frac{(N-1)^2N^2}{4} - \frac{12}{(N-1)^2}\cdot\frac{(N-1)N(2N-1)}{6} +$$

$$+\frac{6}{N-1}\cdot\frac{(N-1)N}{2} - N = \frac{2N^2}{N-1} - \frac{2N(2N-1)}{N-1} + 3N - N =$$

$$= \frac{2N^2 - 4N^2 + 2N}{N-1} + 2N = \frac{-2N^2 + 2N}{N-1} + 2N = -2N + 2N = 0.$$

Исходная система уравнений, так как $X_{0j} = 1$, преобразуется следующим образом:

$$\begin{cases} \sum_{j=1}^{N} X_j^2 - \gamma_1 \sum_{j=1}^{N} X_j \ - \lambda_1 N = 0; \\ \sum_{j=1}^{N} X_j^3 - \gamma_1 \sum_{j=1}^{N} X_j^2 - \lambda_1 \sum_{j=1}^{N} X_j \ = 0. \end{cases} \quad (\text{П}6.30)$$

Используя уравнения (П6.24), (П6.26), (П6.28), (П6.29), преобразуем систему уравнений (П6.30) следующим образом:

$$\begin{cases} \frac{N(N+1)}{3(N-1)} + 0 - \lambda_1 N = 0; \\ 0 - \gamma_1 \frac{N(N+1)}{3(N-1)} - 0 \ = 0. \end{cases} \quad (\text{П}6.31)$$

Из системы уравнений (П6.30) и (П6.31) следует, что

$$\gamma_1 = 0, \quad (\text{П}6.32)$$

$$\lambda_1 = \frac{\sum_{j=1}^{N} X_j^2}{N} = \overline{X^2} = \frac{N+1}{3(N-1)}. \quad (\text{П}.33)$$

11. Доказать, что

$$\sum_{j=1}^{N} \left(X_j^2 - \lambda_1 \right)^2 = \frac{4N(N+1)(N^2-4)}{45(N-1)^3}. \quad (\text{П}.34)$$

Докажем сначала справедливость для РСП вспомогательного равенства:

$$\sum_{j=1}^{N} X_j^4 = \frac{N(N+1)(3N^2-7)}{15(N-1)^3}. \quad (\text{П}6.35)$$

Используя для РСП уравнения (П6.18) и (П6.21), получаем:

$$\sum_{j=1}^{N} X_j^4 = \sum_{j=1}^{N}\left[\frac{2(j-1)}{N-1}-1\right]^2 = \frac{N(N+1)(3N^2-7)}{15(N-1)^3}. \qquad (П6.36)$$

$$\sum_{j=1}^{N} X_j^4 = \sum_{j=1}^{N}\left[\frac{16(j-1)^4}{(N-1)^4} - 4\cdot\frac{8(j-1)^3}{(N-1)^3} + 6\cdot\frac{4(j-1)^2}{(N-1)^2} - 4\cdot\frac{2(j-1)}{(N-1)} + 1\right] =$$

$$= \frac{16}{(N-1)^4}\sum_{j=1}^{N}(j-1)^4 - \frac{32}{(N-1)^3}\sum_{j=1}^{N}(j-1)^3 + \frac{24}{(N-1)^2}\sum_{j=1}^{N}(j-1)^2 -$$

$$-\frac{8}{(N-1)}\sum_{j=1}^{N}(j-1) + N = \frac{16}{(N-1)^4}\cdot\frac{(N-1)N(2N-1)(3N^2-3N-1)}{30} -$$

$$-\frac{32}{(N-1)^3}\cdot\frac{(N^2-1)N^2}{4} + \frac{24}{(N-1)^2}\cdot\frac{(N-1)N(2N-1)}{6} - \frac{8}{N-1}\cdot\frac{(N-1)N}{2} + N =$$

$$= \frac{8N(2N-1)(3N^2-3N-1)}{15(N-1)^3} - \frac{8N^2}{N-1} + \frac{4N(2N-1)}{N-1} - 4N + N =$$

$$= \frac{8N(2N-1)(3N^2-3N-1)}{15(N-1)^3} - \frac{4N}{N-1} - 3N =$$

$$= \frac{N}{N-1}\left[\frac{8(2N-1)(3N^2-3N-1)}{15(N-1)^2} - 4 - 3(N-1)\right] =$$

$$= \frac{N}{N-1}\left[\frac{8(2N-1)(3N^2-3N-1)}{15(N-1)^2} - (3N+1)\right] =$$

$$= \frac{N}{N-1}\left[\frac{8(2N-1)(3N^2-3N-1) - 15(N-1)^2(3N+1)}{15(N-1)^2}\right] =$$

$$= \frac{N}{N-1}\left[\frac{48N^3 - 72N^2 + 8N + 8 - 45N^3 + 75N^2 - 15N - 15}{15(N-1)^2}\right] =$$

$$= \frac{N(N+1)(3N^2-7)}{15(N-1)^3}.$$

Используя уравнения (П6.26), (П6.33), (П6.36), получаем:

$$\sum_{j=1}^{N}\left(X_j^2-\lambda_1\right)^2=\sum_{j=1}^{N}(X_j^4-2\lambda_1 X_j^2+\lambda_1^2)=$$

$$=\frac{N(N+1)(3N^2-7)}{15(N-1)^3}-2\cdot\frac{N+1}{3(N-1)}\cdot\frac{N(N+1)}{3(N-1)}+N\cdot\frac{(N+1)^2}{9(N-1)^2}=$$

$$=\frac{N(N+1)(3N^2-7)}{15(N-1)^3}-\frac{N(N+1)^2}{9(N-1)^2}=\frac{N(N+1)\left[3(3N^2-7)-5(N-1)(N+1)\right]}{9\cdot 15(N-1)^3}=$$

$$=\frac{N(N+1)(4N^2-16)}{9\cdot 15(N-1)^3}=\frac{4N(N+1)(N^2-4)}{45(N-1)^3}.$$

Учебное издание

Леонов Андрей Николаевич, **Дечко** Михаил Михайлович,
Ловкис Виктор Болеславович

ОСНОВЫ МОДЕЛИРОВАНИЯ

Учебно-методическое пособие

Ответственный за выпуск *В. Б. Ловкис*
Редактор *Т. В. Каркоцкая*
Компьютерная верстка *Д. А. Значёнок*
Дизайн обложки *Д. О. Бабаковой*

Подписано в печать 26.03.2020. Формат $60\times84^{1}/_{16}$.
Бумага офсетная. Ризография.
Усл. печ. л. 9,30. Уч.-изд. л. 7,27. Тираж 99 экз. Заказ 76.

Издатель и полиграфическое исполнение:
Учреждение образования
«Белорусский государственный аграрный технический университет».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий
№ 1/359 от 09.06.2014.
№ 2/151 от 11.06.2014.
Пр-т Независимости, 99–2, 220023, Минск.