МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский национальный технический университет

Кафедра «Технология машиностроения»

М. М. Кане

# ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ. ОЦЕНКА ПРАВОМЕРНОСТИ ПРИМЕНЕНИЯ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА В ЗАДАННЫХ УСЛОВИЯХ

Пособие
для студентов специальностей
1-36 80 02 «Инновационные технологии в машиностроении»
и 1-53 80 01 «Автоматизация»

*Рекомендовано учебно-методическим объединением по образованию в области машиностроительного оборудования и технологий, в области автоматизации технологических процессов, производств и управления*

Минск
БНТУ
2020

УДК 621.75.002 (076.5)
ББК 34.5я7
К19

Р е ц е н з е н т ы:
кафедра «Технология металлов» Белорусского государственного аграрного технического университета;
профессор кафедры «Технология металлов» БГАТУ д-р техн. наук, профессор *Л. М. Акулович*;
начальник отделения технологии машиностроения и металлургии Объединённого института машиностроения Национальной Академии наук Беларуси, д-р техн. наук, доцент *В. И. Жорник*

**Кане, М. М.**

К19 Предварительный анализ экспериментальных данных. Оценка правомерности применения корреляционно-регрессионного анализа в заданных условиях : пособие для студентов специальностей 1-36 80 02 «Инновационные технологии в машиностроении» и 1-53 80 01 «Автоматизация» / М. М. Кане. – Минск : БНТУ, 2020. – 49 с.
ISBN 978-985-583-538-8.

Пособие предназначено для изучения дисциплин «Анализ и упорядочение исходных данных при статистической обработке результатов научных исследований» (для специальности 1-36 80 02 «Инновационные технологии в машиностроении») и «Системный анализ в научных исследованиях» (для специальности 1-53 80 01 «Автоматизация»). В издании рассмотрены основные понятия и задачи корреляционно-регрессионного анализа (КРА), оценка достоверности его результатов. Сформулированы условия эффективного применения КРА для моделирования и оптимизации процессов, описаны методики проверки соблюдения этих условий для заданных параметров исследования. Рассмотрен пример использования указанных методик, даны варианты заданий и справочный материал, необходимый для решения поставленных задач. Данное пособие позволит углубить теоретическую подготовку и приобрести практические навыки для предварительного анализа экспериментальных данных и повышения эффективности научных исследований.

**УДК 621.75.002 (076.5)**
**ББК 34.5я7**

**ISBN 978-985-583-538-8**

# 1. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

## 1.1. Основные понятия и задачи корреляционно-регрессионного анализа (КРА)

Для технологических экспериментов характерны статистические объекты исследований, в которых имеют место стохастические или корреляционные взаимосвязи между зависимыми и независимыми переменными. Найти математическое описание этих взаимосвязей – значит получить математическую модель изучаемого процесса или объекта. В задачу корреляционно-регрессионного анализа (КРА) входит получение на основании экспериментальных данных о работе объекта (протекании процесса) математической модели объекта (процесса) и её анализ.

Ценность математической модели заключается в том, что с её помощью можно решать задачи оптимизации процесса (объекта) и предсказанию его результатов (характеристик) при изменении условий его протекания (эксплуатации). Вторую задачу называют интерполяционной.

Методы КРА могут быть применимы только для таких параметров, которые при качественном их рассмотрении, т. е. при изучении физической природы процесса (объекта), являются взаимосвязанными. На первом этапе применения этих методов обычно оценивают степень тесноты взаимосвязи значений функции отклика ($y$) с одним или несколькими независимыми переменными ($x$). Первая задача решается с помощью коэффициента парной корреляции $r_{xy}$, вторая – с помощью коэффициента множественной корреляции $R_{y.x_1,x_2 \ldots x_m}$

$$r_{yx} = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{nS_xS_y}, \tag{1.1}$$

где $\overline{y}$, $\overline{x}$, $S_y$, $S_x$ – средние арифметические значения и средние квадратические отклонения значений соответственно $y_i$ и $x_i$ в рассматриваемой выборке;

$n$ – объём выборки (обычно $n$ = 50–200 шт.).

Обозначим зависимую переменную цифрой 1, а независимые переменные цифрами 2, 3 … $m$, тогда коэффициенты парной корреляции запишутся как $r_{12}$, $r_{13}$ … $r_{1m}$, $r_{23}$, $r_{24}$ и т. д., а коэффициент множественной корреляции между $y$ и $x_1$, $x_2$ … $x_m$ – $R_{1.23\ldots m}$

Коэффициент множественной корреляции с использованием метода определителей находится по формуле

$$R_{1.23\ldots m} = \sqrt{1 - \frac{D}{D_{11}}}, \tag{1.2}$$

где $D$ – определитель, составленный из всех коэффициентов парной корреляции:

$$D = \begin{vmatrix} 1 & r_{12} & r_{13} & \ldots & r_{1m} \\ r_{21} & 1 & r_{23} & \ldots & r_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & r_{m3} & \ldots & 1 \end{vmatrix}. \tag{1.3}$$

$D_{11}$ – определитель, получающийся из определителя $D$ исключением нулевого (первого слева) столбца и нулевой (верхней) строки:

$$D_{11} = \begin{vmatrix} 1 & r_{23} & \ldots & r_{2m} \\ r_{32} & 1 & \ldots & r_{3m} \\ \ldots & \ldots & \ldots & \ldots \\ r_{m2} & r_{m3} & \ldots & 1 \end{vmatrix}. \tag{1.4}$$

Для трёх переменных

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}. \tag{1.5}$$

Значения $r_{yx}$ и $R_{y.x_1,x_2 \ldots x_m}$ находятся в пределах ±1. Если они достоверны, т. е. существенно отличаются от нуля, значит между исследуемыми факторами имеется линейная корреляционная зависимость. В противном случае эта зависимость отсутствует либо является существенно нелинейной. Если $r_{yx}$ или $R_{y.x_1,x_2 \ldots x_m}$ равны +1 или –1, что встречается крайне редко, значит между исследуемыми факторами существует функциональная взаимосвязь. Знак величин $r_{yx}$ и $R_{y.x_1,x_2 \ldots x_m}$ говорит о прямом (+) или обратном (–) характере взаимосвязей между исследуемыми факторами.

Для оценки степени тесноты нелинейных одно- и многофакторных взаимосвязей используются корреляционные отношения η.

Если корреляционный анализ подтвердил наличие взаимосвязей между исследуемыми факторами, то на следующем этапе обработки экспериментальных данных выбирают математическую модель с помощью регрессионного анализа, которая наилучшим образом описывает указанные взаимосвязи. Уравнение, с помощью которого могут быть найдены числовые значения выборочных средних функции отклика в зависимости от соответствующих значений независимых переменных, называется уравнением регрессии. В общем случае оно может быть записано в следующем виде:

$$\overline{y} = f\left(\overline{x}_1,\ \overline{x}_2,\ \overline{x}_m\right). \tag{1.6}$$

Операция замены одной функции другой, в каком-то смысле эквивалентной функцией, называется аппроксимацией. При аппроксимации неизвестных функций отклика в математической статистике наиболее часто используют полиномиальные модели. Степень полинома определяется максимальной степенью входящих в них переменных.

Выпишем полиномы различных степеней для случая двух факторов. Полином нулевой степени:

$$y = b_0. \tag{1.7}$$

Полином первой степени:

$$y = b_0 + b_1 x_1 + b_2 x_2. \tag{1.8}$$

Полином второй степени:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2. \tag{1.9}$$

Полином третьей степени:

$$\begin{gathered} y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + \\ + b_{112} x_1^2 x_2 + b_{122} x_1 x_2^2 + b_{111} x_1^3 + b_{222} x_2^3. \end{gathered} \tag{1.10}$$

Уравнения регрессии вида (1.7) и (1.8) называются линейными, а вида (1.9) и (1.10) – нелинейными. Линейные уравнения регрессии можно использовать на первом этапе исследований нелинейных корреляционных связей с тем, чтобы в дальнейшем ввести в них необходимые поправки. Расчёт коэффициентов уравнений регрессии выполняют чаще всего методом наименьших квадратов.

С позиций статистики полиномиальная модель удобна, так как позволяет увеличить степень точности аппроксимации за счет повышения порядка полинома. При этом аппроксимирующая функция остается линейной по параметрам, что облегчает все статистические операции (применение метода наименьших квадратов для оценки параметров, выбор оптимального расположения уровней факторов в области их изменения и т. д.).

При определении параметров уравнения регрессии все переменные и соотношения между ними иногда удобно выражать в стандартизованном масштабе, где за начало отсчета для каждой переменной принимается ее среднее значение, а за единицу масштаба – ее же среднее квадратическое отклонение. В стандартизованном масштабе упрощаются соотношения между переменными, что удобно при анализе многомерных связей. Значения переменных $t_{x_i}$ в стандартизованном масштабе определяются по формуле

$$t_{x_i} = \frac{x_i - \overline{x}}{s_x}, \tag{1.11}$$

где $x_i$ – значения переменных (зависимой или независимой) в натуральном масштабе.

Если в качестве модели процесса используется полином первой степени, то после перевода переменных в стандартизованный масштаб значительно упрощается формула для расчета коэффициента множественной корреляции:

$$R_{1.23\,\ldots\,m} = \sqrt{\beta_1^2 r_{12}^2 + \beta_2^2 r_{13}^2 + \ldots + \beta_m^2 r_{1m}^2}. \tag{1.12}$$

Здесь $\beta_1$, $\beta_2$, …, $\beta_m$ – коэффициенты уравнения регрессии в стандартизованном масштабе.

Соответствие математической модели процесса экспериментальным данным называют адекватностью. Уравнение адекватно описывает результаты опытов, если квадратическое отклонение значений зависимой переменной $y_{p_i}$, рассчитанных по уравнению регрессии, от экспериментальных данных

$y_i$ обусловлено только ошибкой воспроизведения, т. е. случайным характером этого параметра.

Применение КРА правомерно и эффективно при соблюдении ряда требований к экспериментальным данным. Содержание этих требований и методы их проверки рассмотрены ниже в п. 2 данного пособия.

## 1.2. Оценка достоверности результатов КРА

Поскольку результаты корреляционно-регрессионного анализа, полученные на базе ограниченного числа экспериментальных данных, являются случайными величинами, необходимо оценить их достоверность, определить доверительные интервалы, в которых находятся их истинные значения. Для этого последовательно производятся следующие операции: 1) оценка достоверности коэффициентов корреляции; 2) оценка значимости коэффициентов регрессии; 3) оценка адекватности уравнения регрессии.

1. При любом объеме выборки и нормальном многомерном распределении рассматриваемых факторов вычисляется статистика:

$$t_{\text{н}} = r\sqrt{(n-2)/\left(1-r^2\right)}, \tag{1.13}$$

имеющая распределение Стьюдента с $f = n-2$ степенями свободы. Для проверки нулевой гипотезы $H_0$: $\rho = 0$ ($\rho$ – коэффициент корреляции генеральной совокупности) находят по соответствующим таблицам прил. 5 при фиксированном уровне значимости $\alpha$ и числе степеней свободы $f = n-2$, критическое значение $t_{\alpha/2,\, n-2}$, удовлетворяющее условию $P\left(|t| \geq t_{\alpha/2,\, n-2}\right) = \alpha$. Если наблюдаемое значение $t_{\text{н}} \geq t_{\alpha/2,\, n-2}$, то нулевую гипотезу об отсутствии линейной зависимости между переменными $x$ и $y$ следует отвергнуть. Этот метод часто применяют при малых объемах выборок.

**Пример** [4]. Пусть по выборке объема $n$ = 11, извлеченной из двумерной нормальной совокупности, вычислен выборочный (эмпирический) коэффициент корреляции $r = 0,76$. Требуется проверить нулевую гипотезу $H_0$: $\rho = 0$ против альтернативной гипотезы $H_\alpha$: $\rho \neq 0$ при уровне значимости $\alpha = 0,01$.

**Решение.** Вычисляем статистику

$$t_{\text{н}} = 0,76\sqrt{\frac{11-2}{1-(0,76)^2}} = 3,5.$$

По таблице распределения Стьюдента при $\alpha = 0,01$ и числе степеней свободы $f = n - 2 = 9$ находим $t_{\alpha/2,\, n-2} = t_{0,005;\, 9} = 3,25.$ Так как 3,5 > 3,25, то $r$ значимо отличается от нуля, т. е. переменные $x$ и $y$ являются коррелированными.

При числе наблюдений $n > 50$ надежность коэффициента корреляции можно оценить по его среднему квадратическому отклонению:

$$s_r = \left(1 - r^2\right) / \sqrt{n} \tag{1.14}$$

и нормированному отклонению:

$$t_r = |r| / s_r. \tag{1.15}$$

Достоверность коэффициента корреляции считается доказанной с вероятностью 0,997, если $t_r \geq 3$; с вероятностью 0,990 при $t_r \geq 2,58$. Если $n$ достаточно велико, а $r$ близко к 0,5, границы доверительного интервала для коэффициента корреляции $\rho$ той генеральной совокупности, из которой взята выборка, можно определить обычным способом:

$$r - t_{\text{кр}} s_r \le \rho \le r + t_{\text{кр}} s_r. \tag{1.16}$$

Значение $t_{\text{кр}}$ устанавливается по таблице функции Лапласа для выбранной вероятности. Если левая и правая части неравенства (1.16) имеют одинаковый знак, т. е. $\rho$ не принимает нулевого значения, то $r$ имеет достоверный знак и является значимым.

В случае выборок малых объемов доверительный интервал для $r$ можно определить по номограмме.

Формулы (1.13–1.16) справедливы и при оценках достоверности коэффициента множественной корреляции и корреляционных отношений.

2. Проверку значимости коэффициентов регрессии можно производить двумя способами: сравнением абсолютного значения коэффициента с доверительным интервалом; с помощью t-критерия Стьюдента. В первом случае доверительный интервал для коэффициента $b_i$ вычисляют по формуле

$$\Delta b_i = \pm t_{\text{т}} s_{b_i}, \tag{1.17}$$

где $t_{\text{т}}$ – табличное значение критерия Стьюдента при уровне значимости и числе степеней свободы, для которых определялось $s_{b_i}$; $s_{b_i}$ – среднее квадратическое отклонение $b_i$.

Коэффициент значим, если его абсолютное значение больше доверительного интервала.

При проверке значимости коэффициентов вторым способом вычисляют

$$t_p = |b_i| / s_{b_i} \tag{1.18}$$

и сравнивают его с критическим значением этого критерия $t_{\text{кр}}$. Коэффициент значим, если $t_p > t_{\text{кр}}$ для принятого уровня значимости и числа степеней свободы, при которых определялось $s_{b_i}$.

Статистически незначимые коэффициенты могут быть исключены из уравнения регрессии. Причем, если уравнение получено с помощью методов планирования эксперимента, остальные коэффициенты пересчитывать не надо.

Методика определения $s_{b_i}$ зависит от способа получения уравнения регрессии. В случае применения планирования эксперимента

$$s_{b_i}^2 = \frac{s_{(y)}^2}{nN}, \tag{1.19}$$

где $s_{(y)}^2$ – дисперсия воспроизводимости эксперимента;

$n$ – число параллельных опытов в каждой точке матрицы при равномерном дублировании опытов (при отсутствии дублирования опытов $n = 1$); $N$ – общее число опытов в матрице плана.

При равномерном дублировании опытов во всех строках матрицы плана число параллельных опытов одинаково. Для каждой строки этой матрицы вычисляют дисперсию $s_j^2$ результатов по данным $n$ параллельных опытов:

$$s_j^2 = \frac{1}{n-1}\sum\nolimits_{u=1}^{n}(y_{j_u} - \overline{y}_j)^2, \tag{1.20}$$

где $y_{j_u}$ – значение функции отклика в $j$-й строке для $u$-го опыта.

Если $s_j^2$ результатов опытов однородны, то дисперсия $s_{(y)}^2$ воспроизводимости эксперимента

$$s_{(y)}^2 = \frac{1}{N}\sum_{j=1}^{N} s_j^2, \tag{1.21}$$

где $N$ – число опытов или число строк матрицы плана.

При отсутствии дублирования опытов для определения дисперсии воспроизводимости эксперимента выполняют $n_0$ параллельных опытов при средних уровнях всех независимых факторов (в нулевой точке плана). По результатам этих опытов вычисляют

$$s_{(y)}^2 = \frac{1}{n_0 - 1}\sum_{u=1}^{n_0} \left(y_u - \overline{y}\right)^2, \tag{1.22}$$

где $y_u$ – значение функции отклика в $u$-м параллельном опыте.

При равномерном дублировании опытов число степеней свободы для расчета $s_{(y)}^2$ и, следовательно, $s_{b_i}$ находится как $f = N\left(n-1\right)$ при отсутствии дублирования опытов $f = n_0 - 1$.

Если коэффициенты уравнения регрессии получены без планирования эксперимента (по результатам пассивного эксперимента), средние квадратические отклонения коэффициентов регрессии определяются следующим образом.

Для парной регрессии $\overline{y} = a + bx$:

$$s_b = \frac{s_y\sqrt{1 - r_{xy}^2}}{s_x\sqrt{n}}; \tag{1.23}$$

для регрессии трех переменных $\overline{y} = a + bx + cz$:

$$s_b = \frac{s_y \sqrt{1 - R_{y.xz}^2}}{s_x \sqrt{1 - r_{xz}^2}\,\sqrt{n}}; \tag{1.24}$$

$$s_c = \frac{s_y \sqrt{1 - R_{y.xz}^2}}{s_z \sqrt{1 - r_{xz}^2}\,\sqrt{n}}; \tag{1.25}$$

для регрессии многих переменных $\overline{y} = b_0 + b_1 x_1 + b_2 x_2 + {} + \ldots + b_k x_k$:

$$s_{b_1} = \frac{s_y \sqrt{1 - R_{y.x_1 x_2 \ldots x_k}^2}}{s_{x_1} \sqrt{1 - R_{x_1.x_2 x_3 \ldots x_k}^2} \cdot \sqrt{n}}. \tag{1.26}$$

Аналогичные формулы могут быть записаны для расчета $s_{b_2}$, $s_{b_3}$, …, $s_{b_r}$. Дальнейшая проверка значимости коэффициентов регрессии выполняется по формулам (1.17) и (1.18).

3. В зависимости от наличия сведений о дисперсии воспроизводимости эксперимента $s_{(y)}^2$ проверку адекватности уравнения регрессии можно производить по двум схемам. Первая из них применяется при отсутствии оценки дисперсии воспроизводимости, что характерно для пассивного эксперимента и состоит из следующих этапов:

а) вычисление дисперсии относительно среднего значения параметра оптимизации (остаточной дисперсии для уравнения нулевого порядка):

$$s_{y_0}^2 = \frac{\sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2}{N - 1}; \tag{1.27}$$

б) расчет дисперсии, характеризующей отклонение экспериментальных значений величин от найденных по уравнению регрессии. Если порядок уравнения заранее неизвестен, то в случае многофакторного пространства имеет смысл начинать с уравнения первого порядка

$$s_{y_1}^2 = \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{f}, \tag{1.28}$$

где $\hat{y}_i$ – значение параметра оптимизации, вычисленное по уравнению регрессии для условий $i$-го опыта; $f = N - g$ – число степеней свободы; $g$ – число коэффициентов регрессии; для линейного уравнения $g = k + 1$; для неполного квадратного уравнения, включающего члены типа $b_j x_j$ и $b_{ij} x_i x_j$, $g = k\left(k+1\right)/2+1$; для полного квадратного уравнения $g = k\left(k+3\right)/2+1$; $k$ – число факторов;

в) вычисление опытного значения отношения дисперсий

$$F_0 = \frac{s_{y_0}^2}{s_{y_1}^2}, \tag{1.29}$$

которое затем сравнивают с критическим $F_{\text{кр}(f_1, f_2)}$. Числа степеней свободы $f_1$ и $f_2$ принимают равными соответственно знаменателям в формулах (1.27) и (1.28).

Если $F_0 \le F_{\text{кр}}$, пользоваться уравнением регрессии первого порядка не имеет смысла, так как в изученном интервале изменения уровней факторов оно описывает исследуемую систему (процесс) не лучше, чем уравнение нулевого порядка. Затем составляют уравнение второго порядка, рассчитывают

$s_{y_2}^2$ и $F_1 = s_{y_2}^2 / s_{y_1}^2$. Далее проверяют значимость этого отношения по $F$-критерию. Процедуру повторяют до тех пор, пока не будет выполнено условие $F_r \geq F_{кр(f_1, f_2)}$. Индекс $r$ соответствует степени предпоследнего полинома.

Если известна дисперсия воспроизводимости $s_{(y)}^2$ эксперимента для оценки адекватности модели вначале рассчитывают дисперсию адекватности по формуле (1.28), а затем вычисляют опытное значение $F$-критерия (вторая схема):

$$F_p = s_{ад}^2 / s_{(y)}^2. \tag{1.30}$$

Если $F_r < F_{кр(f_1, f_2)}$, модель считают адекватной. Значения $s_{(y)}^2$ в зависимости от характера дублирования опытов определяют по формулам (1.21) или (1.22), значения $f_1$ и $f_2$ равны соответственно знаменателям в формулах для расчета $s_{ад}^2$ и $s_{(y)}^2$. $F_{кр}(t_1, t_2)$ приведены в прил. 3.

# 2. ПРАКТИЧЕСКАЯ РАБОТА.
# ОЦЕНКА ПРАВОМЕРНОСТИ ПРИМЕНЕНИЯ КРА В ЗАДАННЫХ УСЛОВИЯХ

## 2.1. Основные положения предварительного анализа экспериментальных данных

**Цель работы**: изучение методики предварительного анализа экспериментальных данных при статистическом исследовании зависимостей и приобретение практических навыков.

Работа рассчитана на 5 академических часов.

Для моделирования различных случайных процессов, в том числе процессов механической обработки деталей машин, наиболее часто используется такой статистический метод как корреляционно-регрессионный анализ (КРА). При этом качество модели во многом зависит от качества экспериментальных данных, положенных в её основу. Поэтому предварительный анализ экспериментальных данных является обязательным этапом статистического исследования зависимостей [1].

При выборе КРА в качестве метода исследования необходимо выполнить анализ физической природы изучаемого процесса или объекта и убедиться в том, что между рассматриваемыми зависимыми и независимыми переменными имеются взаимосвязи. Только в этом случае для их анализа могут использоваться статистические методы.

Анализ статистических методов моделирования и распространенных методов оценки адекватности и достоверности полученных результатов показывает, что их эффективное применение возможно при соблюдении следующих требований к экспериментальным данным:

1. зависимые и независимые переменные являются случайными величинами с нормальным законом распределения*;

---

* независимые переменные могут быть и неслучайными величинами, в частности, в активных экспериментах.

2. дисперсия зависимой переменной $y$ не зависит от абсолютных значений $y$ – остается постоянной или однородной при различных наблюдениях $y$;

3. значения независимых переменных $x_1$, $x_2$ … $x_m$ измеряются с пренебрежимо малыми ошибками по сравнению с ошибкой измерения $y$;

4. переменные $x_1$, $x_2$ … $x_m$ линейно независимы;

5. процесс формирования $y$ является стационарным и эргодическим;

6. экспериментальные данные получены из ряда независимых испытаний, наблюдений и образуют случайную выборку из данной генеральной совокупности;

7. результаты наблюдений не содержат резко выделяющихся значений, не принадлежащих к данной генеральной совокупности.

Только экспериментальные данные, отвечающие перечисленным требованиям, позволяют получить надёжные и достоверные результаты статистического исследования взаимосвязей между зависимыми ($y$) и независимыми ($x$) переменными. Рассмотрим методы проверки соблюдения указанных требований.

***Соответствие распределения у и х нормальному закону распределения*** устанавливается либо по большим выборкам с помощью критериев Пирсона или Колмогорова [2], либо на основании анализа природы величин $y$ или $x$.

Критерий Колмогорова λ даёт достаточно точные результаты даже при сравнительно небольших (около 30 шт.) объёмах выборок и прост для вычислений. Для расчёта значения λ необходимо предварительно определить эмпирические $F_n(x)$ и теоретические $F(x)$ функции предполагаемого закона распределения для каждого наблюденного значения случайной величины $x$ (в качестве случайной величины могут рассматриваться как зависимые, так и независимые случайные величины). Методики определения $F_n(x)$ и $F(x)$ для нормального закона распределения рассмотрены в [3, 4]. Затем по максимальной разности этих функций определяется λ по следующей формуле:

$$\lambda = | F_n(x) - F(x) |_{\max} \sqrt{n} = D\sqrt{n} \qquad (2.1)$$

здесь $n$ – объём выборки, по которой производится оценка закона распределения.

По вычисленному значению λ по прил. 1 определяют $P(\lambda)$. Если $P(\lambda) > 0{,}15$, то гипотеза о соответствии эмпирического распределения нормальному распределению принимается.

Критерий $\chi^2$ применим для любых сгруппированных совокупностей, но при достаточно большом их объёме ($n > 50$). Для вычисления $\chi^2$ необходимо предварительно вычислить теоретические частоты для наблюденных значений эмпирического распределения, т. е. произвести сопоставление этого распределения с предполагаемым теоретическим (в нашем случае – с нормальным распределением). При этом необходимо, чтобы число эмпирических частот $f_i$ в каждом интервале значений $x$ было не менее 5. Если в каком-либо интервале $f_i < 5$, то такой интервал следует объединить с соседним.

Критерий $\chi^2$ вычисляется по следующей формуле:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(f_i - f_i^{'}\right)^2}{f_i^{'}}, \qquad (2.2)$$

где $m$ – число интервалов значений случайной величины $x$;

$f_i$, $f_i^{'}$ – эмпирическая и теоретическая частоты $i$-го интервала значений случайной величины $x$.

Методы расчёта $f_i$ и $f_i^{'}$ показаны в [3, 4].

Далее необходимо определить число степеней свободы $k$ по формуле:

$$k = m - p - 1, \qquad (2.3)$$

где $p$ – число параметров теоретического распределения.

Для нормального распределения $k$ = 2 (математическое ожидание и дисперсия).

Для величины $k$ найден закон распределения, по которому вычислены вероятности $P(\chi^2)$ для различных значений $\chi^2$ и $k$. Значения $P(\chi^2)$ приведены в прил. 2. Если $P(\chi^2) > 0{,}05$, то гипотеза о соответствии рассматриваемых эмпирического и теоретического распределений принимается. Здесь принят уровень значимости $\alpha = 0{,}05$.

Необходимо отметить, что критерии согласия Пирсона и Колмогорова не позволяют однозначно установить теоретический закон распределения рассматриваемой случайной величины $x$. С помощью этих критериев нулевая гипотеза о виде функции распределения либо отклоняется, либо считается, что имеющаяся информация не даёт повода для отклонения выдвинутой гипотезы о виде функции распределения. Если объём выборок $n$ невелик ($n \le 50$) или результаты измерений изменяются в узком диапазоне, то экспериментальные данные могут согласовываться с рядом законов распределения. На практике критерии согласия часто применяют для подбора закона распределения, который с наилучшей точностью описывает эмпирическое распределение из ряда рассматриваемых законов. Если критерии согласия Пирсона и Колмогорова дают различные результаты, то предпочтение следует отдать результату, полученному с помощью критерия Пирсона, который считается более строгим.

Выбор подходящего закона распределения должен базироваться прежде всего на понимании механизма изучаемого явления. Если механизм изучаемого явления неизвестен, то предварительный выбор закона распределения может быть выполнен с помощью следующих методов: по виду гистограммы частостей эмпирического распределения; с помощью графического представления эмпирической функции распределения на вероятностных бумагах; по величине эмпирического коэффициента вариации $v = s / \bar{x}$ (здесь $s$ – среднее квадратическое отклонение случайной величины $x$; $\overline{x}$ – её среднее арифметическое значение); с помощью выборочных коэффициентов асимметрии и эксцесса [4, 5].

Для оценки ***однородности дисперсий у*** в условиях планирования эксперимента проводят параллельные опыты в различных точках матрицы плана, т. е. при различных значениях $x_1, x_2, \ldots x_m$. При пассивном эксперименте $s_{y_j}^2$ определяют по результатам отдельных выборок, взятых в примерно одинаковых условиях работы объекта. Если сравниваются **два** значения $s_{y_1}^2$ и $s_{y_2}^2$ при разных числах степеней их свободы $f_j$ ($f = N - 1$, $N$ – число параллельных опытов или объем выборки), то используется критерий Фишера, рассчитываемый как отношение большей дисперсии к меньшей:

$$F_p = \frac{s_{y_1}^2}{s_{y_2}^2}; \quad \left(s_{y_1}^2 > s_{y_2}^2\right). \tag{2.4}$$

Если наблюдаемое значение $F_p$ меньше критического $F_{\text{кр}}$ для соответствующих чисел степеней свободы и принятого уровня значимости, то опыты считаются воспроизводимыми, а дисперсии однородными [5]. В Прил. 3 приведены значения $F_{\text{кр}}$ для уровня значимости α = 0,05.

Однородность **ряда** дисперсий при одинаковом числе опытов для определения каждой из них оценивают с помощью критерия Кохрена – отношения максимальной дисперсии к сумме всех дисперсий ряда:

$$G_p = \frac{s_{y_{\max}}^2}{s_{y_1}^2 + s_{y_2}^2 + \ldots + s_{y_N}^2} = \frac{s_{y_{\max}}^2}{\sum_{j=1}^{N} s_{y_j}^2}, \tag{2.5}$$

где $N$ – число параллельных опытов или различных выборок (обычно $N$ = 2–5).

Дисперсии однородны, если расчетное значение $G_p$ не превышает критического $G_{кр}$ [2]. В Прил. 4 приведены значения $G_{кр}$ для уровня значимости $\alpha = 0{,}05$.

Если дисперсии неоднородны, необходимо выполнить такое преобразование *y*, чтобы они стали однородными. Довольно часто помогает замена *y* на *lny*.

Воспроизводимость опытов и однородность дисперсий достигается, когда выявлены и устранены источники нестабильности эксперимента, а также с помощью более точных методов и средств измерений.

Проверку ***достаточной точности измерения*** значений независимых переменных можно произвести, сопоставив ее с диапазоном изменения последних. Считается, что ошибки определения независимых переменных не должны превышать 5–7 % интервала их варьирования. Ошибки в определении значения зависимой переменной не влияют столь значительно на точность регрессионного анализа (они могут составлять до 30 % интервала варьирования).

Методы определения ошибок прямых и косвенных измерений описаны в [4, 6]. В данной работе этот вопрос не рассматривается.

***Отсутствие коррелированности независимых переменных*** проверяется расчетом парных коэффициентов корреляции между ними [4, 5]. Способ оценки достоверности значений коэффициентов *r* парной корреляции зависит от объёма *n* выборки значений независимых переменных, по которым этот коэффициент был рассчитан.

Для выборки $n \le 30$ вычисляется критерий Стьюдента *t* по формуле:

$$t = \sqrt{\frac{n-2}{1-r^2}} \qquad (2.6)$$

имеющий распределение Стьюдента с $\upsilon = n - 2$ степенями свободы. Для проверки нулевой гипотезы $H_0$: $\rho = 0$ (здесь $\rho$ – коэффициент корреляции генеральной совокупности) находят по таблицам распределения Стьюдента (Прил. 5) по принятому уровню значимости $\alpha$ и числу степеней свободы $\upsilon$ критическое значение $t_{\alpha/2;n-2}$, удовлетворяющее условию $P(|t| \geq t_{\alpha;n-2}) = \alpha$. Если $|t_{набл}| \geq t_{\alpha;n-2}$, то нулевую гипотезу об отсутствии линейной зависимости между рассмотренными независимыми переменными следует отвергнуть. Если же $|t_{набл}| < t_{\alpha;n-2}$, то нет оснований отвергать нулевую гипотезу о некоррелированности рассматриваемых независимых переменных.

При достаточно большом объёме выборки ($n \geq 50$) можно воспользоваться $r$-распределением, соответствующем $\rho = 0$ [7]. В Прил. 6 приведены квантили $r_{1-\alpha/2}$ этого распределения для некоторых уровней значимости $\alpha$ и числа степеней свободы $\upsilon = n - 2$. Если окажется, что найденный по выборке коэффициент корреляции $r$ удовлетворяет неравенству

$$|r| > r_{1-\alpha/2}, \upsilon, \tag{2.7}$$

то его нужно признать значимым, т. е. нужно считать, что нулевая гипотеза неверна. А это значит, что $\rho \neq 0$ и между наблюдаемыми величинами есть корреляция. При несоблюдении условия (2.7) можно считать, что между наблюдаемыми величинами корреляция отсутствует.

***Стационарным*** называют такой случайный процесс, основные характеристики которого $\left(M\left[x\right], \sigma^2\right)$ постоянны или однородны во времени. Поскольку при пассивном эксперименте свойства процесса определяются по одной представительной выборке, распространить полученные результаты на весь процесс в целом можно лишь при условии его стационарности.

Проверка стационарности процесса производится в следующем порядке.

По результатам измерений параметра строится случайная последовательность значений этого параметра, соответствующая порядку проведения измерений. Полученную реализацию разбивают на несколько равных отрезков (5–10), для каждого отрезка устанавливают дисперсию $s_{y_j}^2$ и с помощью критерия Кохрена по формуле (2.5) определяют, являются ли значения $s_{y_j}^2$ на каждом из отрезков оценками одной и той же генеральной дисперсии. Критическое значение критерия Кохрена можно для данных условий выбрать по прил. 4.

Затем на каждом из отрезков производится сравнение средних арифметических $\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_N$, соответствующих выборочным дисперсиям $s_1^2, s_2^2, \ldots, s_N^2$, числам степеней свободы $f_1, f_2, \ldots, f_N$ (обычно $f_j = n_j - 1$, где $n_j$ – объем соответствующей выборки, отрезка). Всем выборкам соответствует единая генеральная дисперсия $\sigma^2$. В качестве ее оценки можно взять средневзвешенную дисперсию $s_y^2$, которая рассчитывается по формуле (2.8)

$$\overline{s_y^2} = \frac{\sum_{j=1}^{N} f_j s_{y_j}^2}{f}, \quad f = \sum_{j}^{N} f_j. \tag{2.8}$$

Если справедлива нулевая гипотеза о равенстве всех генеральных средних, то в качестве оценки единого генерального среднего можно взять общее среднее всех элементов, отрезков, как бы объединенных в одну выборку.

Обозначим это среднее через $\overline{\overline{y}}$. Для дисперсии $\sigma^2$ можно теперь дать другую оценку:

$$\overline{s^2} = \frac{1}{N-1}\sum_{j=1}^{N} n_j \left(\overline{y_j} - \overline{\overline{y}}\right)^2, \qquad (2.9)$$

которой соответствует $N - 1$ степеней свободы. Чтобы нулевая гипотеза была справедлива, должно соблюдаться условие

$$\overline{s_y^2} / \overline{s^2} \le F_{1-p(N-1,f)}, \qquad (2.10)$$

где значение критерия Фишера $F_{1-p(N-1,f)}$ берется по Прил. 3.

Стационарная случайная функция обладает **эргодическим свойством**, если ее характеристики $M[x]$, $\sigma^2$ могут быть рассчитаны как соответствующие средние по времени для одной реализации большой продолжительности. Иными словами, эргодичность определяет способность процесса к воспроизведению своих характеристик в различных реализациях. Следовательно, условием эргодичности процесса является соблюдение описанных выше условий стационарности процесса, но рассчитанных не для различных отрезков одной реализации процесса, а для различных реализаций процесса (3–5 реализаций объёмом 50–100 шт.). Эргодичными могут быть только стационарные процессы. Если стационарный процесс обладает эргодическим свойством, его основные характеристики могут быть установлены по одной реализации достаточно большой продолжительности (объем выборки $n$ = 50–100 шт.) Если анализ выборки не подтвердил стационарность процесса, необходимо либо учесть фактор времени в модели процесса как независимую переменную, либо найти объем выборки (период времени для ее отбора), для которой процесс будет стационарным.

**Случайность и независимость результатов измерений** в выборке свидетельствует об отсутствии монотонного или циклического смещения результатов измерений, вызванного влиянием неучтённого фактора, систематической погрешностью процесса. Подобный случай может иметь место при анализе размеров деталей, обрабатываемых на настроенном станке, когда вследствие изнашивания инструмента или нагрева станка, центр группирования размеров постепенно смещается при сохранении разброса размеров (значения дисперсии).

Проверку ***гипотезы о случайности выборки***, необходимую при пассивном эксперименте, можно произвести методом последовательных разностей. Этот метод заключается в следующем: по значениям $x_i$ выборки, расположенным в последовательности их наблюдения $x_1$, $x_2$, $x_3 \ldots x_n$ образуется $n - 1$ разностей между соседними членами:

$$a_1 = x_2 - x_1,\ a_2 = x_3 - x_2,\ \ldots,\ a_{n-1} = x_n - x_{n-1}.$$

Затем определяются

$$C^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} a_i^2, \qquad (2.11)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2. \qquad (2.12)$$

Для оценки случайности выборки используется критерий $\tau = C^2 / s^2$. Если $\tau > \tau_q$ где $\tau_q$ – критическое значение критерия при уровне значимости $q$, то гипотеза «случайности» верна [2, 7]. Значения $\tau_q$ приведены в прил. 7.

Одним из важных условий правильного применения статистических оценок является **отсутствие грубых ошибок при наблюдениях**. Поэтому все грубые ошибки должны быть выявлены и исключены из рассмотрения в самом начале обработки результатов наблюдений. При решении данной задачи предполагается, что результаты наблюдений подчиняются нормальному закону распределения. Грубые погрешности возникают при случайном резком изменении условий измерения.

Для анализа резко выделяющихся значений случайных величин в выборке используют графический и аналитические методы [1]. Наиболее распространенным критерием для исключения одного экстремального значения является критерий К. Пирсона–Н. В. Смирнова–Ф. Граббса

$$\upsilon = \frac{\left|\bar{x} - x_i\right|}{s}, \tag{2.13}$$

где $x_i$ – крайний (max или min) элемент в выборке;

$\bar{x}$, $s$ – среднее арифметическое значение и среднее квадратическое отклонение в полной выборке (до исключения крайнего значения).

Критические значения $\upsilon_{1-\alpha}$ для данного уровня значимости $\alpha$ зависят только от объема выборки $n$. При $\upsilon > \upsilon_{1-\alpha}$ крайнее значение $x$ отбрасывается как ошибочное. Процесс исключения подозрительных значений рекомендуется начинать с максимального значения и проводить последовательно для крайних значений вариационного ряда. После исключения ошибочного результата значения $\bar{x}$ и $s$ необходимо пересчитать. Критерий $\upsilon$ справедлив для нормально распределенных случайных величин и чувствителен к отклонениям от этого закона. Для уменьшения вероятности ошибочного решения необходимо предварительно убедиться в том, что результаты получены при соблюдении требований к точности измерений в постоянных условиях реализации процесса. Значения $\upsilon_{1-\alpha}$ приведены в прил. 8.

## 2.2. Методические указания

Для приобретения практических навыков оценки правомерности применения корреляционно-регрессионного анализа в заданных условиях исследования группе студентов (2–4 человека) выдаётся вариант задания с описанием промежуточных результатов исследования. Необходимо с помощью приведенных выше расчётных зависимостей и прил. 1–8 оценить соблюдение приведенных выше семи требований к экспериментальным данным.

Порядок выполнения работы и форму представления её результатов рассмотрим на примере.

## Пример

**1. Проверка соответствия эмпирического распределения случайной величины нормальному закону.**

1.1. По критерию Колмогорова.

Исходные данные:

Объём выборки $n = 50$; $D = 0{,}06$ (по формуле (2.1)).

**Решение:**

По формуле (2.1) $\lambda_{набл} = 0{,}06\sqrt{50} = 0{,}4$. По прил. 1 этому значению $\lambda_{набл}$ соответствует $P(\lambda) = 0{,}9972$.

Т. к. $P(\lambda) > 0{,}15$, гипотеза о соответствии эмпирического распределения нормальному распределению принимается.

1.2. По критерию Пирсона.

Исходные данные:

Число интервалов значений случайной величины $m = 7$, критерий Пирсона $\chi^2 = 6{,}5$ (по формуле (2.2)).

**Решение:**

По формуле (2.3) число степеней свободы $k = 7 - 2 - 1 = 4$, по прил. 2 $P(\chi^2) = P(6{,}5;4) = 0{,}23$.

Т. к. $P(\chi^2) > 0{,}05$, гипотеза о соответствии эмпирического распределения нормальному распределению принимается.

**2. Проверка однородности дисперсий *y*.**

2.1. Сравнение двух дисперсий.

Исходные данные:

Объём выборки $N = 30$; дисперсии $s_{y_1}^2 = 0{,}031$; $s_{y_2}^2 = 0{,}024$; уровень значимости $\alpha = 0{,}05$.

**Решение:**

По формуле (2.4) находим наблюдаемое значение критерия Фишера $F_р$ = 0,031 / 0,024 = 1,29; числа степеней свободы $f_1 = f_2$ = 30 – 1 = 29; по прил. 3 $F_{кр} = F_{29;29;0.05}$ = 1,6; т. к. $F_р < F_{кр}$, опыты считаются воспроизводимыми, а сравниваемые дисперсии однородными.

2.2. Сравнение ряда дисперсий.

Исходные данные:

$$N = 30;\ s_{y_1}^2 = 0{,}041;\ s_{y_2}^2 = 0{,}037;\ s_{y_3}^2 = 0{,}026;\ \alpha = 0{,}05.$$

**Решение:**

По формуле (2.5) находим наблюдаемое значение критерия Кохрена $G_p = \dfrac{0{,}041}{0{,}041 + 0{,}037 + 0{,}26} = \dfrac{0{,}041}{0{,}104} = 0{,}39$; по прил. 4 $G_{кр} = G_{3;30;0{,}05} = 0{,}45$.

Т. к. $G_р < G_{кр}$, опыты считаются воспроизводимыми, а сравниваемые дисперсии однородными.

**3. Проверка достаточной точности измерений независимых переменных.**

Исходные данные:

Погрешность измерений независимой переменной $\Delta_\Sigma$ составляет $\Delta_\Sigma$ = 0,025 мм, диапазон $z$ изменения независимой переменной –$z$ = 1,5 мм.

**Решение:**

Ошибка определения независимой переменной составляет $\Delta_\Sigma / z$ = 0,025 / 1,5 = 0,017 или 1,7 % от интервала её варьирования. Т. к. это значение не превышает 5–7 %, точность измерения независимой переменной является достаточной.

**4. Проверка отсутствия коррелированности независимых переменных.**

4.1. Малый объём выборки ($n \le 30$).

Исходные данные:

Коэффициент парной корреляции $r$ = 0,22; объём выборки $n$ = 20; уровень значимости $\alpha$ = 0,05.

**Решение:**

По формуле (2.6) определяем наблюдаемое значения критерия Стьюдента $t = \sqrt{\frac{20-2}{1-0,22^2}} = 4,35.$ Определяем число степеней свободы $\upsilon = n - 2 = 20 - 2 = 18$. По прил. 5 находим критическое значение этого критерия $t_{кр} = t_{0,05;18} = 1,734$. Т. к. $t_{набл} > t_{кр}$ (4,35 > 1,734), между независимыми переменными имеется корреляционная взаимосвязь.

4.2. Большой объём выборки ($n \geq 50$).

Исходные данные:

$r = 0,15$; $n = 60$; $\alpha = 0,05$.

**Решение:**

По прил. 6 находим критическое значение коэффициента корреляции $r_{кр} = r_{1-0,05/2;58} = 0,255$. Т. к. $|r| < r_{кр}$, можно считать, что между независимыми переменными корреляционная зависимость отсутствует.

**5. Проверка стационарности процесса.**

Исходные данные:

Характеристики отрезков выборки: $\overline{y_1} = 1,2;$ $\overline{y_2} = 1,5;$ $\overline{y_3} = 1,1;$ $\overline{y_4} = 1,4;$ $\overline{y_5} = 1,6;$ $s_{y_1}^2 = 0,04;$ $s_{y_2}^2 = 0,06;$ $s_{y_3}^2 = 0,03;$ $s_{y_4}^2 = 0,05;$ $s_{y_5}^2 = 0,07.$

Объём каждого из отрезков $n_j = 5$ шт. Число отрезков $N = 5$. Доверительная вероятность $p = 0,95$; уровень значимости $\alpha = 1 - p = 1 - 0,95 = 0,05$.

**Решение:**

По формуле (2.5) определяем наблюдаемое значение критерия Кохрена

$$G_p = \frac{0,07}{0,04 + 0,06 + 0,03 + 0,05 + 0,07} = 0,28.$$

По прил. 4 $G_{кр}$ = 0,544. Т. к. $G_р < G_{кр}$, дисперсии отрезков однородны.

Определяем число степеней свободы отрезков выборки $f_j = n_j - 1 = 5 - 1 = 4$ и выборки в целом $f = \sum_j^N f_j = 4 \cdot 5 = 20.$

По формуле (2.8) определяем $s_y^2$

$$s_y^2 = \frac{4 \cdot 0,04 + 4 \cdot 0,06 + 4 \cdot 0,03 + 4 \cdot 0,05 + 4 \cdot 0,07}{20} = 0,05;$$

$$\overline{\overline{y}} = \frac{1,2 + 1,5 + 1,1 + 1,4 + 1,6}{5} = 1,36.$$

По формуле (2.9) определяем $\overline{s^2}$

$$\overline{s^2} = \frac{5(1,2-1,36)^2 + 5(1,5-1,36)^2 + 5(1,1-1,36)^2 + 5(1,4-1,36)^2 + 5(1,6-1,36)^2}{5} =$$
$$= 0,172.$$

По формуле (2.10) находим наблюдаемое значение критерия Фишера

$$F_н = \frac{0,172}{0,05} = 3,44.$$

По прил. 3 находим критическое значение критерия Фишера

$$F_{кр} = F_{(0,05;4;20)} = 5,8.$$

Т. к. $F_{н-} < F_{кр}$, гипотеза о равенстве всех средних арифметических значений в отрезках выборки справедлива. Таким образом, подтверждена стабильность значений $s_{y_j}^2$ и $\overline{y}_j$. В общей выборке, что указывает на стационарность рассматриваемого процесса в пределах объёма данной выборки.

**6. Проверка гипотезы о случайности выборки.**

Исходные данные:

Характеристики выборки (см. формулы (2.11), (2.12)) $s^2 = 0{,}122$ мм$^2$, $c^2 = 0{,}118$, объём выборки $m = 30$, доверительная вероятность $p = 0{,}95$.

**Решение:**

Определяем наблюдаемое значение критерия $\tau$.

$$\tau_{\text{н}} = c^2 / s^2 = 0{,}118 / 0{,}122 = 0{,}967.$$

По прил. 7 находим критическое значение критерия $\tau$.

Для $m = 30$, $p = 0{,}95$, $\tau_q = 0{,}704$.

Т. к. $\tau_{\text{н}} > \tau_q$, то гипотеза независимости и случайности измерений в выборке принимается.

**7. Проверка отсутствия грубых ошибок наблюдений.**

Исходные данные:

Объём выборки $m = 12$, среднее арифметическое значение выборки $\bar{x} = 4{,}3$, среднее квадратическое отклонение в выборке $s = 0{,}31$, доверительная вероятность $p = 0{,}95$, уровень значимости $\alpha = 1 - p = 1 - 0{,}95 = 0{,}05$, минимальное значение в выборке $x_{\min} = 3{,}7$.

**Решение:**

По формуле (2.13) находим наблюдаемое значение критерия $\upsilon_{\text{н}}$:

$$\upsilon_{\text{н}} = \frac{4{,}3 - 3{,}7}{0{,}31} = 1{,}93.$$

По прил. 8 находим критическое значение критерия $\upsilon$: для $m = 12$, $p = 0{,}95$ $\upsilon_{\text{к}} = 2{,}387$. Т. к. $\upsilon_{\text{н}} < \upsilon_{\text{к}}$, то размер $x_{\min} = 3{,}7$ не является грубой ошибкой и должен быть сохранён в выборке.

## 2.3. Порядок выполнения работы

1. Ознакомиться с работой, изучить её задачи и методы их решения.
2. Образовать подгруппу (2–4 студента) для выполнения работы. Получить у преподавателя вариант задания.
3. Выполнить задание.
4. Индивидуально оформить отчёт и представить преподавателю.

Варианты заданий приведены в прил. 9.

## 2.4. Содержание отчёта

1. Название работы.
2. В соответствии с рекомендациями раздела 2.2 данной работы (Методические указания) привести перечень задач, для каждой задачи выполнить необходимые расчёты и привести их результаты.
3. Для каждой задачи дать обоснованные выводы.

Отчёт может быть оформлен в виде таблицы. Форма таблицы показана ниже.

Отчёт по практической работе «Оценка правомерности применения КРА в заданных условиях» магистранта специальности (№ специальности) Ф.И.О.

| № п/п | Название задачи | Основные условия задачи | Выполненные расчёты | Выводы |
|---|---|---|---|---|
| | | | | |

## 2.5. Контрольные вопросы

1. Что является основным условием применения статистических методов для изучения взаимосвязей между зависимыми и независимыми факторами рассматриваемого процесса или объекта?

2. Перечислите 7 основных требований к экспериментальным данным, соблюдение которых является предпосылкой успешного использования КРА для моделирования процесса или объекта.

3. Каким образом можно установить, что фактическое распределение случайной величины не противоречит закону нормального распределения?

4. Как можно предварительно установить закон распределения случайной величины?

5. Как можно оценить однородность двух или ряда дисперсий?

6. Как можно оценить достаточность точности измерений зависимых и независимых переменных?

7. Как проверить отсутствие коррелированности независимых переменных?

8. Какой процесс является стационарным и как это можно определить?

9. Какой процесс обладает эргодическим свойством и как это можно проверить?

10. Как проверить гипотезу случайности выборки? О чём свидетельствует справедливость этой гипотезы?

11. Как проверить отсутствие грубых ошибок в результатах наблюдений?

## СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Айвазян, С. А. Прикладная статистика : Основы моделирования и первичная обработка данных : справочное изд. / С. А. Айвазян, И. С. Евсюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983.

2. Солонин, И. С. Математическая статистика в технологии машиностроения / И. С. Солонин. – М. : Машиностроение, 1972.

3. Бабук, В. В. Лабораторный практикум по технологии машиностроения / В. В. Бабук [и др.] ; под ред. В. В. Бабука. – Минск : Вышэйшая школа, 1983.

4. Кане, М. М. Основы исследований, изобретательства и инновационной деятельности в машиностроении : учебник / М. М. Кане. – Минск : Вышэйшая школа, 2018.

5. Герасимович, А. И. Математическая статистика / А. И. Герасимович, Я. И. Матвеева. – Минск : Вышэйшая школа, 1978.

6. Кане, М. М. Исследования и изобретательство в машиностроении. Практикум / М. М. Кане [и др.] ; под общ. ред. М. М. Кане. – Минск : Вышэйшая школа, 2020.

7. Пустыльник, Е. И. Статистические методы анализа и обработки наблюдений / Е. И. Пустыльник. – М. : Изд. Наука, Глав. редакция физико-математической литературы, 1968.

# ПРИЛОЖЕНИЕ 1

Таблица П1

Значения вероятностей $P(\lambda)$ для различных $\lambda$

| $\lambda$ | $P(\lambda)$ | $\lambda$ | $P(\lambda)$ | $\lambda$ | $P(\lambda)$ | $\lambda$ | $P(\lambda)$ |
|---|---|---|---|---|---|---|---|
| 0,30 | 1,000 | 0,70 | 0,7112 | 1,20 | 0,1122 | 1,90 | 0,0015 |
| 0,35 | 0,9997 | 0,75 | 0,6272 | 1,30 | 0,0681 | 2,00 | 0,0007 |
| 0,40 | 0,9972 | 0,80 | 0,5441 | 1,40 | 0,0397 | 2,10 | 0,0003 |
| 0,45 | 0,9874 | 0,85 | 0,4653 | 1,50 | 0,0222 | 2,20 | 0,0001 |
| 0,50 | 0,9639 | 0,90 | 0,3927 | 1,60 | 0,0120 | 2,30 | 0,0001 |
| 0,55 | 0,9228 | 0,95 | 0,3275 | 1,70 | 0,0062 | 2,40 | 0,0000 |
| 0,60 | 0,8643 | 1,00 | 0,2700 | 1,80 | 0,0032 | 2,50 | 0,0000 |
| 0,65 | 0,7920 | 1,10 | 0,1777 | | | | |

# ПРИЛОЖЕНИЕ 2

Таблица П2

Таблица вероятностей *P* для критерия $\chi^2$

| $\chi^2$ | *k* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 0,317 | 0,6055 | 0,8013 | 0,9098 | 0,9626 | 0,9856 | 0,9948 | 0,998 |
| 2 | 1574 | 3679 | 5724 | 7358 | 8491 | 9197 | 9598 | 9810 |
| 3 | 0833 | 2231 | 3916 | 5578 | 7000 | 8088 | 8850 | 9344 |
| 4 | 0455 | 1353 | 2615 | 4060 | 5494 | 6767 | 7798 | 8571 |
| 5 | 0254 | 0821 | 1718 | 2873 | 4159 | 5438 | 6600 | 7576 |
| 6 | 0143 | 0498 | 1116 | 1991 | 3062 | 4232 | 5398 | 6472 |
| 7 | 0081 | 0302 | 0719 | 1359 | 2206 | 3208 | 4289 | 5366 |
| 8 | 0047 | 0183 | 0460 | 0916 | 1562 | 2381 | 3326 | 4335 |
| 9 | 0027 | 0111 | 0293 | 0611 | 1091 | 1736 | 2527 | 3423 |
| 10 | 0016 | 0067 | 0186 | 0404 | 0752 | 1247 | 1886 | 2650 |
| 11 | 0009 | 0041 | 0117 | 0266 | 0514 | 0884 | 1386 | 2017 |
| 12 | 0005 | 0025 | 0074 | 0174 | 0348 | 0620 | 1006 | 1512 |
| 13 | 0003 | 0015 | 0046 | 0113 | 0234 | 0430 | 0721 | 1119 |
| 14 | 0002 | 0009 | 0029 | 0073 | 0156 | 0296 | 0512 | 0818 |
| 15 | 0001 | 0006 | 0018 | 0047 | 0104 | 0203 | 0360 | 0591 |
| 16 | 0001 | 0003 | 0011 | 0030 | 0068 | 0138 | 0251 | 0424 |
| 17 | 0000 | 0002 | 0007 | 0019 | 0045 | 0093 | 0174 | 0301 |
| 18 | | 0001 | 0004 | 0012 | 0029 | 0062 | 0120 | 0212 |
| 19 | | 0001 | 0003 | 0008 | 0019 | 0042 | 0082 | 0149 |
| 20 | | 0000 | 0002 | 0005 | 0013 | 0028 | 0056 | 0103 |
| 21 | | 0000 | 0001 | 0003 | 0008 | 0018 | 0038 | 0071 |
| 22 | | 0000 | 0001 | 0002 | 0005 | 0012 | 0025 | 0049 |
| 23 | | 0000 | 0000 | 0001 | 0003 | 0008 | 0017 | 0034 |
| 24 | | 0000 | 0000 | 0001 | 0002 | 0005 | 0011 | 0023 |
| 25 | | 0000 | 0000 | 0001 | 0001 | 0003 | 0008 | 0016 |
| 26 | | 0000 | 0000 | 0000 | 0001 | 0002 | 0005 | 0010 |
| 27 | | 0000 | 0000 | 0000 | 0001 | 0001 | 0003 | 0007 |
| 28 | | 0000 | 0000 | 0000 | 0000 | 0001 | 0002 | 0005 |
| 29 | | 0000 | 0000 | 0000 | 0000 | 0001 | 0001 | 0003 |
| 30 | | 0000 | 0000 | 0000 | 0000 | 0000 | 0001 | 0002 |

# ПРИЛОЖЕНИЕ 3

Таблица П3

Квантили распределения Фишера $F_{1-p}$

| $f_1$ / $f_2$ | Уровень значимости 0,05 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 12 | 24 | ∞ |
| 1 | 164,4 | 199,5 | 215,7 | 224,6 | 230,2 | 234,0 | 244,9 | 249,9 | 254,3 |
| 2 | 18,5 | 19,2 | 19,2 | 19,3 | 19,3 | 19,3 | 19,4 | 19,5 | 19,5 |
| 3 | 10,1 | 9,6 | 9,3 | 9,1 | 9,0 | 8,9 | 8,7 | 8,6 | 8,5 |
| 4 | 7,7 | 6,9 | 6,6 | 6,4 | 6,3 | 6,2 | 5,9 | 5,8 | 5,6 |
| 5 | 6,6 | 5,8 | 5,4 | 5,2 | 5,1 | 5,0 | 4,7 | 4,5 | 4,4 |
| 6 | 6,0 | 5,1 | 4,8 | 4,5 | 4,4 | 4,3 | 4,0 | 3,8 | 3,7 |
| 7 | 5,6 | 4,7 | 4,4 | 4,1 | 4,0 | 3,9 | 3,6 | 3,4 | 3,2 |
| 8 | 5,3 | 4,5 | 4,1 | 3,8 | 3,7 | 3,6 | 3,3 | 3,1 | 2,9 |
| 9 | 5,1 | 4,3 | 3,9 | 3,6 | 3,5 | 3,4 | 3,1 | 2,9 | 2,7 |
| 10 | 5,0 | 4,1 | 4,7 | 3,5 | 3,3 | 3,2 | 2,9 | 2,7 | 2,5 |
| 11 | 4,8 | 4,0 | 3,6 | 3,4 | 3,2 | 3,1 | 2,8 | 2,6 | 2,4 |
| 12 | 4,8 | 3,9 | 3,5 | 3,3 | 3,1 | 3,0 | 2,7 | 2,5 | 2,3 |
| 13 | 4,7 | 3,8 | 3,4 | 3,2 | 3,0 | 2,9 | 2,6 | 2,4 | 2,2 |
| 14 | 4,6 | 3,7 | 3,3 | 3,1 | 3,0 | 2,9 | 2,5 | 2,3 | 2,1 |
| 15 | 4,5 | 3,7 | 3,3 | 3,0 | 2,9 | 2,7 | 2,4 | 2,2 | 2,0 |
| 16 | 4,5 | 3,6 | 3,2 | 3,0 | 2,9 | 2,7 | 2,4 | 2,2 | 2,0 |
| 17 | 4,5 | 3,6 | 3,2 | 3,0 | 2,8 | 2,7 | 2,4 | 2,2 | 2,0 |
| 18 | 4,4 | 3,6 | 3,2 | 2,9 | 2,8 | 2,7 | 2,3 | 2,1 | 1,9 |
| 19 | 4,4 | 3,5 | 3,1 | 2,9 | 2,7 | 2,6 | 2,3 | 2,1 | 1,8 |
| 20 | 4,4 | 3,5 | 3,1 | 2,9 | 2,7 | 2,6 | 2,3 | 2,1 | 1,8 |
| 22 | 4,3 | 3,4 | 3,1 | 2,8 | 2,7 | 2,6 | 2,2 | 2,0 | 1,8 |
| 24 | 4,3 | 3,4 | 3,0 | 2,8 | 2,6 | 2,5 | 2,2 | 2,0 | 1,7 |
| 26 | 4,2 | 3,4 | 3,0 | 2,7 | 2,6 | 2,4 | 2,1 | 1,9 | 1,7 |
| 28 | 4,2 | 3,3 | 2,9 | 2,7 | 2,6 | 2,4 | 2,1 | 1,9 | 1,6 |
| 30 | 4,2 | 3,3 | 2,9 | 2,7 | 2,5 | 2,3 | 2,1 | 1,9 | 1,6 |
| 40 | 4,1 | 3,2 | 2,9 | 2,6 | 2,5 | 2,3 | 2,0 | 1,8 | 1,5 |
| 60 | 4,0 | 3,2 | 2,8 | 2,5 | 2,4 | 2,3 | 1,9 | 1,7 | 1,4 |
| 120 | 3,9 | 3,1 | 2,7 | 2,5 | 2,3 | 2,2 | 1,8 | 1,6 | 1,3 |
| ∞ | 3,8 | 3,0 | 2,6 | 2,4 | 2,2 | 2,1 | 1,8 | 1,5 | 1,0 |

## ПРИЛОЖЕНИЕ 4

Таблица П4

Критические значения критерия Кохрена $G_к$

| $n$ (число выборок) | $m$ (объем выборки) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 17 |
| $P = 0{,}95$ | | | | | | | | |
| 3 | 0,798 | 0,746 | 0,707 | 0,677 | 0,653 | 0,633 | 0,617 | 0,547 |
| 4 | 0,684 | 0,629 | 0,589 | 0,560 | 0,537 | 0,518 | 0,502 | 0,437 |
| 5 | 0,589 | 0,544 | 0,507 | 0,478 | 0,456 | 0,439 | 0,424 | 0,365 |
| 6 | 0,532 | 0,480 | 0,445 | 0,418 | 0,398 | 0,382 | 0,368 | 0,314 |
| 7 | 0,480 | 0,431 | 0,397 | 0,373 | 0,354 | 0,338 | 0,326 | 0,276 |
| 8 | 0,438 | 0,391 | 0,359 | 0,336 | 0,319 | 0,304 | 0,293 | 0,246 |
| 10 | 0,373 | 0,331 | 0,303 | 0,282 | 0,267 | 0,254 | 0,244 | 0,203 |
| 12 | 0,326 | 0,288 | 0,262 | 0,244 | 0,230 | 0,219 | 0,210 | 0,174 |
| 15 | 0,276 | 0,242 | 0,219 | 0,203 | 0,191 | 0,182 | 0,144 | 0,143 |
| 20 | 0,220 | 0,192 | 0,174 | 0,160 | 0,150 | 0,142 | 0,136 | 0,111 |
| $P = 0{,}99$ | | | | | | | | |
| 3 | 0,883 | 0,834 | 0,793 | 0,761 | 0,734 | 0,711 | 0,691 | 0,606 |
| 4 | 0,781 | 0,721 | 0,676 | 0,641 | 0,613 | 0,590 | 0,570 | 0,488 |
| 5 | 0,696 | 0,633 | 0,588 | 0,553 | 0,526 | 0,504 | 0,485 | 0,409 |
| 6 | 0,626 | 0,564 | 0,520 | 0,487 | 0,461 | 0,440 | 0,423 | 0,353 |
| 7 | 0,569 | 0,508 | 0,466 | 0,435 | 0,411 | 0,391 | 0,375 | 0,310 |
| 8 | 0,521 | 0,463 | 0,423 | 0,393 | 0,370 | 0,352 | 0,337 | 0,278 |
| 10 | 0,447 | 0,393 | 0,357 | 0,331 | 0,311 | 0,295 | 0,281 | 0,230 |
| 12 | 0,392 | 0,343 | 0,310 | 0,286 | 0,268 | 0,254 | 0,242 | 0,196 |
| 15 | 0,332 | 0,288 | 0,259 | 0,239 | 0,223 | 0,210 | 0,200 | 0,161 |
| 20 | 0,265 | 0,229 | 0,205 | 0,188 | 0,175 | 0,165 | 0,157 | 0,125 |

## ПРИЛОЖЕНИЕ 5

Распределение Стьюдента

Значения $t_{\alpha,\nu}$ удовлетворяют условию

$$P(t \ge t_{\alpha,\nu}) = \int_{t_{a.v}}^{\infty} S\left(t,\nu\right) dt = \alpha$$

В таблице приведены значения квантилей $t_{\alpha,\nu}$ в зависимости от числа степеней свободы ν и вероятности α.

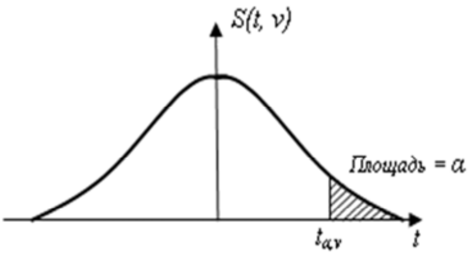


Таблица П5

Квантили $t_{\alpha,\nu}$ в зависимости от числа степеней свободы ν и вероятности α

| ν \ α | 0,40 | 0,30 | 0,20 | 0,10 | 0,050 | 0,025 | 0,010 | 0,005 | 0,001 | 0,0005 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,325 | 0,727 | 1,376 | 3,078 | 6,314 | 12,71 | 31,82 | 63,66 | 318,3 | 636,6 |
| 2 | 0,289 | 0,617 | 1,061 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 | 22,33 | 31,60 |
| 3 | 0,277 | 0,584 | 0,978 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 | 10,22 | 12,94 |
| 4 | 0,271 | 0,569 | 0,941 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 | 7,173 | 8,610 |
| 5 | 0,267 | 0,559 | 0,920 | 1,476 | 2,015 | 2,571 | 3,365 | 5,032 | 5,893 | 6,859 |
| 6 | 0,265 | 0,553 | 0,906 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 | 5,208 | 5,959 |
| 7 | 0,263 | 0,549 | 0,896 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 | 4,785 | 5,405 |
| 8 | 0,262 | 0,546 | 0,889 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 | 4,501 | 5,041 |
| 9 | 0,261 | 0,543 | 0,883 | 1,383 | 1,833 | 2,262 | 2,821 | 3,250 | 4,297 | 4,781 |

Окончиние табл. П5

| ν \ α | 0,40 | 0,30 | 0,20 | 0,10 | 0,050 | 0,025 | 0,010 | 0,005 | 0,001 | 0,0005 |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0,260 | 0,542 | 0,879 | 1,372 | 1,812 | 2,228 | 2,764 | 3,169 | 4,144 | 4,587 |
| 11 | 0,260 | 0,540 | 0,876 | 1,363 | 1,796 | 2,201 | 2,718 | 3,106 | 4,025 | 4,487 |
| 12 | 0,259 | 0,539 | 0,873 | 1,356 | 1,782 | 2,179 | 2,681 | 3,055 | 3,930 | 4,318 |
| 13 | 0,259 | 0,538 | 0,870 | 1,350 | 1,771 | 2,160 | 2,650 | 3,012 | 3,852 | 4,221 |
| 14 | 0,258 | 0,537 | 0,868 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 | 3,787 | 4,140 |
| 15 | 0,258 | 0,536 | 0,866 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 | 3,733 | 4,073 |
| 16 | 0,258 | 0,535 | 0,865 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 | 3,686 | 4,015 |
| 17 | 0,257 | 0,534 | 0,863 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 | 3,646 | 3,965 |
| 18 | 0,257 | 0,534 | 0,862 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 | 3,611 | 3,922 |
| 19 | 0,257 | 0,533 | 0,861 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 | 3,579 | 3,883 |
| 20 | 0,257 | 0,533 | 0,860 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 | 3,552 | 3,850 |
| 21 | 0,257 | 0,532 | 0,859 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 | 3,527 | 3,819 |
| 22 | 0,256 | 0,532 | 0,858 | 1,321 | 1,717 | 2,074 | 2,508 | 2,819 | 3,505 | 3,792 |
| 23 | 0,256 | 0,532 | 0,858 | 1,319 | 1,714 | 2,069 | 2,500 | 2,807 | 3,485 | 3,767 |
| 24 | 0,256 | 0,531 | 0,857 | 1,318 | 1,711 | 2,064 | 2,492 | 2,797 | 3,467 | 3,745 |
| 25 | 0,256 | 0,531 | 0,856 | 1,316 | 1,708 | 2,060 | 2,485 | 2,787 | 3,450 | 3,725 |
| 26 | 0,256 | 0,531 | 0,856 | 1,315 | 1,706 | 2,056 | 2,479 | 2,779 | 3,435 | 3,707 |
| 27 | 0,256 | 0,531 | 0,855 | 1,314 | 1,703 | 2,052 | 2,473 | 2,771 | 3,421 | 3,690 |
| 28 | 0,256 | 0,530 | 0,855 | 1,313 | 1,701 | 2,048 | 2,467 | 2,763 | 3,408 | 3,674 |
| 29 | 0,256 | 0,530 | 0,854 | 1,311 | 1,699 | 2,045 | 2,462 | 2,756 | 3,396 | 3,659 |
| 30 | 0,256 | 0,530 | 0,854 | 1,310 | 1,697 | 2,042 | 2,457 | 2,750 | 3,385 | 3,646 |
| 40 | 0,255 | 0,529 | 0,851 | 1,303 | 1,684 | 2,021 | 2,423 | 2,704 | 3,307 | 3,551 |
| 50 | 0,255 | 0,528 | 0,849 | 1,298 | 1,676 | 2,009 | 2,403 | 2,678 | 3,262 | 3,495 |
| 60 | 0,254 | 0,527 | 0,848 | 1,296 | 1,671 | 2,000 | 2,390 | 2,660 | 3,232 | 3,460 |
| 80 | 0,254 | 0,527 | 0,846 | 1,292 | 1,664 | 1,990 | 2,374 | 2,639 | 3,195 | 3,415 |
| 100 | 0,254 | 0,526 | 0,845 | 1,290 | 1,660 | 1,984 | 2,365 | 2,626 | 3,174 | 3,389 |
| 200 | 0,254 | 0,525 | 0,843 | 1,286 | 1,653 | 1,972 | 2,345 | 2,601 | 3,131 | 3,339 |
| 500 | 0,253 | 0,525 | 0,842 | 1,283 | 1,648 | 1,965 | 2,334 | 2,586 | 3,106 | 3,310 |
| ∞ | 0,253 | 0,524 | 0,842 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,090 | 3,291 |

**ПРИЛОЖЕНИЕ 6**

Таблица П6

Квантили распределения выборочного
коэффициента корреляции $r_{1-p/2}$

| Число степеней свободы $f$ | Уровни значимости $p$ | | | | |
|---|---|---|---|---|---|
| | 0,10 | 0,05 | 0,02 | 0,01 | 0,001 |
| 1 | 0,988 | 0,997 | 0,999 | 1,000 | 1,000 |
| 2 | 0,900 | 0,950 | 0,980 | 0,990 | 0,999 |
| 3 | 0,805 | 0,878 | 0,934 | 0,959 | 0,992 |
| 4 | 0,729 | 0,811 | 0,882 | 0,917 | 0,974 |
| 5 | 0,669 | 0,754 | 0,833 | 0,874 | 0,951 |
| 6 | 0,621 | 0,707 | 0,789 | 0,834 | 0,925 |
| 7 | 0,582 | 0,666 | 0,750 | 0,798 | 0,898 |
| 8 | 0,549 | 0,632 | 0,716 | 0,765 | 0,872 |
| 9 | 0,521 | 0,602 | 0,685 | 0,735 | 0,847 |
| 10 | 0,497 | 0,576 | 0,658 | 0,708 | 0,823 |
| 11 | 0,476 | 0,553 | 0,634 | 0,684 | 0,801 |
| 12 | 0,457 | 0,532 | 0,612 | 0,661 | 0,780 |
| 13 | 0,441 | 0,514 | 0,592 | 0,641 | 0,760 |
| 14 | 0,426 | 0,497 | 0,574 | 0,623 | 0,742 |
| 15 | 0,412 | 0,482 | 0,558 | 0,606 | 0,725 |
| 16 | 0,400 | 0,468 | 0,543 | 0,590 | 0,708 |
| 17 | 0,389 | 0, 456 | 0,528 | 0,575 | 0,693 |
| 18 | 0,378 | 0,444 | 0,516 | 0,561 | 0,679. |
| 19 | 0,369 | 0,433 | 0,503 | 0,549 | 0,665 |
| 20 | 0,360 | 0,423 | 0,492 | 0,537 | 0,652 |
| 25 | 0,323 | 0,381 | 0,445 | 0,487 | 0,597 |
| 30 | 0,296 | 0, 349 | 0,409 | 0,449 | 0,554 |
| 35 | 0,275 | 0,325 | 0,381 | 0,418 | 0,519 |
| 40 | 0,257 | 0,304 | 0,358 | 0,393 | 0,490 |
| 45 | 0,243 | 0,287 | 0,338 | 0,372 | 0,465 |
| 50 | 0,231 | 0,273 | 0,322 | 0,354 | 0,443 |
| 60 | 0,211 | 0,250 | 0,295 | 0,325 | 0,408 |
| 70 | 0,195 | 0,232 | 0,274 | 0,302 | 0,380 |
| 80 | 0,183 | 0,217 | 0,256 | 0,283 | 0,357 |
| 90 | 0,173 | 0,205 | 0,242 | 0,267 | 0,337 |
| 100 | 0,164 | 0,195 | 0,230 | 0,254 | 0,321 |

## ПРИЛОЖЕНИЕ 7

Таблица П7

Значения критерия $\tau_к$

| *m* | *P* | | *m* | *P* | |
|---|---|---|---|---|---|
| | 0,95 | 0,99 | | 0,95 | 0,99 |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 4 | 0,390 | 0,256 | 19 | 0,642 | 0,510 |
| 5 | 0,410 | 0,269 | 20 | 0,650 | 0,520 |
| 6 | 0,445 | 0,281 | 25 | 0,676 | 0,542 |
| 7 | 0,468 | 0,307 | 30 | 0,704 | 0,508 |
| 8 | 0,491 | 0,331 | 35 | 0,725 | 0,611 |
| 9 | 0,514 | 0,354 | 40 | 0,742 | 0,636 |
| 10 | 0,531 | 0,376 | 45 | 0,757 | 0,658 |
| 11 | 0,548 | 0,397 | 50 | 0,769 | 0,674 |
| 12 | 0,564 | 0,414 | 60 | 0,789 | 0,702 |
| 13 | 0,578 | 0,431 | 70 | 0,804 | 0,724 |
| 14 | 0,591 | 0,447 | 80 | 0,817 | 0,741 |
| 15 | 0,603 | 0,461 | 90 | 0,827 | 0,756 |
| 16 | 0,614 | 0,475 | 100 | 0,836 | 0,767 |
| 17 | 0,624 | 0,487 | 110 | 0,843 | 0,778 |
| 18 | 0,633 | 0,499 | 120 | 0,850 | 0,788 |

# ПРИЛОЖЕНИЕ 8

Таблица П8

Критические значения критерия $\upsilon(P, m)$

| $m$ | Доверительная вероятность $P$ | | | |
|---|---|---|---|---|
| | 0,9 | 0,95 | 0,975 | 0,99 |
| 3 | 1,406 | 1.412 | 1,414 | 1,414 |
| 4 | 1,645 | 1.689 | 1,710 | 1,723 |
| 5 | 1,791 | 1.869 | 1,917 | 1,955 |
| 6 | 1,894 | 1,996 | 2,067 | 2,130 |
| 7 | 1,947 | 2,093 | 2,182 | 2,265 |
| 8 | 2,041 | 2,172 | 2,273 | 2,374 |
| 9 | 2,097 | 2,238 | 2,349 | 2,464 |
| 10 | 2,146 | 2,294 | 2,414 | 2,540 |
| 11 | 2,190 | 2,343 | 2,470 | 2,606 |
| 12 | 2,229 | 2,387 | 2,519 | 2,663 |
| 13 | 2,264 | 2,426 | 2,563 | 2,713 |
| 14 | 2,297 | 2,461 | 2,602 | 2,759 |
| 16 | 2,354 | 2,523 | 2,670 | 2,837 |
| 18 | 2,404 | 2,577 | 2,728 | 2,903 |
| 20 | 2,447 | 2,623 | 2,779 | 2,959 |
| 22 | 2,486 | 2,664 | 2,823 | 3,008 |
| 24 | 2,521 | 2,701 | 2,862 | 3,051 |
| 26 | 2,553 | 2,734 | 2,897 | 3,089 |
| 28 | 2,582 | 2,764 | 2,929 | 3,124 |
| 30 | 2,609 | 2,792 | 2,958 | 3,156 |
| 35 | 2,668 | 2,853 | 3,022 | 3,224 |
| 40 | 2,718 | 2,904 | 3,075 | 3,281 |
| 45 | 2,762 | 2,948 | 3,120 | 3,329 |
| 50 | 2,800 | 2,987 | 3,160 | 3,370 |

## ПРИЛОЖЕНИЕ 9

Варианты заданий

Таблица П9

| № п/п | Наименование задачи | Исходные данные |
|---|---|---|
| ВАРИАНТ 1 | | |
| 1. | Проверка соответствия распределения случайной величины нормальному закону | а. По критерию Колмогорова: $n = 100$, $D = 0{,}1$<br>б. По критерию Пирсона: $m = 8$, $\chi^2 = 12$ |
| 2. | Проверка однородности дисперсий $y$ | а. Сравнение 2-х дисперсий: $N = 40$; $s_{y_1}^2 = 0{,}43$; $s_{y_2}^2 = 0{,}34$; $\alpha = 0{,}05$<br>б. Сравнение 4-х дисперсий: $N = 40$; $s_{y_1}^2 = 0{,}42$; $s_{y_2}^2 = 0{,}0{,}37$; $s_{y_3}^2 = 0{,}25$; $s_{y_4}^2 = 0{,}31$; $\alpha = 0{,}05$ |
| 3. | Проверка достаточной точности измерений независимых переменных | $\Delta_\Sigma = 0{,}1$ мм; $z = 1{,}0$ мм |
| 4. | Проверка отсутствия коррелированности независимых переменных | а. Малый объём выборки $n = 25$, $r = 0{,}45$, $\alpha = 0{,}05$<br>б. Большой объём выборки $n = 70$, $r = 0{,}52$, $\alpha = 0{,}05$ |
| 5. | Проверка стационарности процесса | $\overline{y_1} = 12$, $\overline{y_2} = 14$, $\overline{y_3} = 10$, $\overline{y_4} = 11$, $\overline{y_5} = 13$, $\overline{y_6} = 15$, $s_{y_1}^2 = 3$, $s_{y_2}^2 = 4$, $s_{y_3}^2 = 2$, $s_{y_4}^2 = 2{,}5$, $s_{y5}^2 = 3{,}5$, $s_{y6}^2 = 4{,}5$, $n_j = 5$, $N = 6$, $\alpha = 0{,}05$, $p = 0{,}95$ |
| 6. | Проверка гипотезы о случайности выборки | $S^2 = 2{,}1$, $c^2 = 1{,}8$, $m = 40$, $p = 0{,}95$ |
| 7. | Проверка отсутствия грубых ошибок наблюдений | $m = 18$, $\overline{x} = 24{,}3$, $S = 7{,}5$, $x_{\max} = 30{,}6$, $p = 0{,}95$, $\alpha = 0{,}05$ |

Продолжение табл. П9

| | | |
|---|---|---|
| ВАРИАНТ 2 | | |
| 1. | Проверка соответствия распределения случайной величины нормальному закону | а. По критерию Колмогорова: $n = 80$, $D = 0{,}21$<br>б. По критерию Пирсона: $m = 7$, $\chi^2 = 14$ |
| 2. | Проверка однородности дисперсий $y$ | а. Сравнение 2-х дисперсий: $N = 50$; $s_{y_1}^2 = 3{,}2$; $s_{y_2}^2 = 2{,}6$; $\alpha = 0{,}05$<br>б. Сравнение 4-х дисперсий: $N = 60$; $s_{y1}^2 = 1{,}5$; $s_{y2}^2 = 1{,}9$; $s_{y_3}^2 = 2{,}3$; $s_{y_4}^2 = 2{,}1$; $\alpha = 0{,}05$ |
| 3. | Проверка достаточной точности измерений независимых переменных | $\Delta_\Sigma = 0{,}01$ мм; $z = 2{,}0$ мм |
| 4. | Проверка отсутствия коррелированности независимых переменных | а. Малый объём выборки $n = 15$, $r = 0{,}35$, $\alpha = 0{,}05$<br>б. Большой объём выборки $n = 60$, $r = 0{,}47$, $\alpha = 0{,}05$ |
| 5. | Проверка стационарности процесса | $\overline{y_1} = 2{,}2$, $\overline{y_2} = 2{,}5$, $\overline{y_3} = 3{,}1$, $\overline{y_4} = 2{,}7$, $\overline{y_5} = 2{,}9$, $\overline{y_6} = 3{,}4$, $\overline{y_7} = 3{,}2$, $s_{y_1}^2 = 0{,}42$, $s_{y_2}^2 = 0{,}47$, $s_{y_3}^2 = 0{,}53$, $s_{y_4}^2 = 0{,}46$, $s_{y_5}^2 = 0{,}50$, $s_{y_6}^2 = 0{,}64$, $s_{y_7}^2 = 0{,}62$, $n_j = 6$, $N = 7$, $\alpha = 0{,}05$, $p = 0{,}95$ |
| 6. | Проверка гипотезы о случайности выборки | $S^2 = 0{,}62$, $c^2 = 0{,}53$, $m = 30$, $p = 0{,}95$ |
| 7. | Проверка отсутствия грубых ошибок наблюдений | $m = 22$, $\overline{x} = 14.6$, $S = 2{,}7$, $x_{\max} = 16{,}1$, $p = 0{,}95$, $\alpha = 0{,}05$ |

Продолжение табл. П9

| ВАРИАНТ 3 | | |
|---|---|---|
| 1. | Проверка соответствия распределения случайной величины нормальному закону | а. По критерию Колмогорова: $n = 90$, $D = 0{,}23$<br>б. По критерию Пирсона: $m = 5$, $\chi^2 = 10$ |
| 2. | Проверка однородности дисперсий $y$ | а. Сравнение 2-х дисперсий: $N = 60$; $s_{y_1}^2 = 1{,}43$; $s_{y_2}^2 = 1{,}24$; $\alpha = 0{,}05$<br>б. Сравнение 4-х дисперсий: $N = 50$; $s_{y_1}^2 = 2{,}42$; $s_{y_2}^2 = 2{,}37$; $s_{y_3}^2 = 2{,}25$; $s_{y_4}^2 = 2{,}36$; $\alpha = 0{,}05$ |
| 3. | Проверка достаточной точности измерений независимых переменных | $\Delta_\Sigma = 0{,}05$ мм; $z = 1{,}8$ мм |
| 4. | Проверка отсутствия коррелированности независимых переменных | а. Малый объём выборки $n = 18$, $r = 0{,}25$, $\alpha = 0{,}05$<br>б. Большой объём выборки $n = 55$, $r = 0{,}52$, $\alpha = 0{,}05$ |
| 5. | Проверка стационарности процесса | $\overline{y_1} = 2{,}6$, $\overline{y_2} = 2{,}7$, $\overline{y_3} = 2{,}4$, $\overline{y_4} = 3{,}0$, $\overline{y_5} = 2{,}9$, $\overline{y_6} = 3{,}3$, $\overline{y_7} = 3{,}5$, $s_{y_1}^2 = 0{,}47$, $s_{y_2}^2 = 0{,}52$, $s_{y_3}^2 = 0{,}45$, $s_{y_4}^2 = 0{,}61$, $s_{y_5}^2 = 0{,}58$, $s_{y_6}^2 = 0{,}61$, $s_{y_7}^2 = 0{,}65$, $n_j = 5$, $N = 7$, $\alpha = 0{,}05$, $p = 0{,}95$ |
| 6. | Проверка гипотезы о случайности выборки | $S^2 = 1{,}2$, $c^2 = 0{,}83$, $m = 30$, $p = 0{,}95$ |
| 7. | Проверка отсутствия грубых ошибок наблюдений | $m = 35$, $\bar{x} = 12{,}6$, $S = 2{,}5$, $x_{\max} = 14{,}1$, $p = 0{,}95$, $\alpha = 0{,}05$ |

Продолжение табл. П9

| ВАРИАНТ 4 | | |
|---|---|---|
| 1. | Проверка соответствия распределения случайной величины нормальному закону | а. По критерию Колмогорова: $n = 70$, $D = 0{,}16$<br>б. По критерию Пирсона: $m = 6$, $\chi^2 = 12$ |
| 2. | Проверка однородности дисперсий $y$ | а. Сравнение 2-х дисперсий: $N = 60$; $s_{y_1}^2 = 3{,}45$; $s_{y_2}^2 = 3{,}14$; $\alpha = 0{,}05$<br>б. Сравнение 5-и дисперсий: $N = 50$; $s_{y_1}^2 = 3{,}46$; $s_{y_2}^2 = 3{,}39$; $s_{y_3}^2 = 3{,}24$; $s_{y_4}^2 = 3{,}33$; $s_{y5}^2 = 3{,}18$; $\alpha = 0{,}05$ |
| 3. | Проверка достаточной точности измерений независимых переменных | $\Delta_\Sigma = 0{,}03$ мм; $z = 1{,}5$ мм |
| 4. | Проверка отсутствия коррелированности независимых переменных | а. Малый объём выборки $n = 15$, $r = 0{,}28$, $\alpha = 0{,}05$<br>б. Большой объём выборки $n = 50$, $r = 0{,}35$, $\alpha = 0{,}05$ |
| 5. | Проверка стационарности процесса | $\overline{y_1} = 1{,}7$, $\overline{y_2} = 1{,}9$, $\overline{y_3} = 1{,}4$, $\overline{y_4} = 2{,}0$, $\overline{y_5} = 2{,}2$, $\overline{y_6} = 2{,}1$, $\overline{y_7} = 2{,}3$, $s_{y_1}^2 = 0{,}09$, $s_{y_2}^2 = 0{,}11$, $s_{y_3}^2 = 0{,}07$, $s_{y_4}^2 = 0{,}12$, $s_{y_5}^2 = 0{,}14$, $s_{y_6}^2 = 0{,}13$, $s_{y_7}^2 = 0{,}16$, $n_j = 5$, $N = 7$, $\alpha = 0{,}05$, $p = 0{,}95$ |
| 6. | Проверка гипотезы о случайности выборки | $S^2 = 3{,}2$, $c^2 = 1{,}85$, $m = 20$, $p = 0{,}95$ |
| 7. | Проверка отсутствия грубых ошибок наблюдений | $m = 40$, $\overline{x} = 22{,}3$, $S = 3{,}8$, $x_{\max} = 24{,}7$, $p = 0{,}95$, $\alpha = 0{,}05$ |

Окончание табл. П9

| ВАРИАНТ 5 | | |
|---|---|---|
| 1. | Проверка соответствия распределения случайной величины нормальному закону | а. По критерию Колмогорова: $n = 60$, $D = 0{,}26$<br>б. По критерию Пирсона : $m = 7$, $\chi^2 = 18$ |
| 2. | Проверка однородности дисперсий $y$ | а. Сравнение 2-х дисперсий: $N = 80$; $s_{y_1}^2 = 2{,}4$; $s_{y_2}^2 = 2{,}1$; $\alpha = 0{,}05$<br>б. Сравнение 5-и дисперсий: $N = 70$; $s_{y_1}^2 = 2{,}46$; $s_{y_2}^2 = 2{,}39$; $s_{y_3} = 2{,}27$; $s_{y_4}^2 = 2{,}34$; $s_{y_5}^2 = 2{,}15$; $\alpha = 0{,}05$ |
| 3. | Проверка достаточной точности измерений независимых переменных | $\Delta_\Sigma = 0{,}04$ мм; $z = 2{,}5$ мм |
| 4. | Проверка отсутствия коррелированности независимых переменных | а. Малый объём выборки $n = 25$, $r = 0{,}38$, $\alpha = 0{,}05$<br>б. Большой объём выборки $n = 60$, $r = 0{,}30$, $\alpha = 0{,}05$ |
| 5. | Проверка стационарности процесса | $\overline{y_1} = 3{,}4$, $\overline{y_2} = 3{,}9$, $\overline{y_3} = 3{,}6$, $\overline{y_4} = 3{,}0$, $\overline{y_5} = 3{,}2$, $\overline{y_6} = 3{,}1$, $\overline{y_7} = 3{,}5$, $s_{y_1}^2 = 0{,}04$, $s_{y_2}^2 = 0{,}08$, $s_{y_3}^2 = 0{,}07$, $s_{y_4}^2 = 0{,}03$, $s_{y_5}^2 = 0{,}04$, $s_{y_6}^2 = 0{,}03$, $s_{y_7}^2 = 0{,}06$, $n_j = 5$, $N = 7$, $\alpha = 0{,}05$, $p = 0{,}95$ |
| 6. | Проверка гипотезы о случайности выборки | $S^2 = 2{,}2$, $c^2 = 1{,}35$, $m = 20$, $p = 0{,}95$ |
| 7. | Проверка отсутствия грубых ошибок наблюдений | $m = 50$, $\bar{x} = 20{,}7$, $S = 3{,}6$, $x_{\min} = 18{,}7$, $p = 0{,}95$, $\alpha = 0{,}05$ |

## ОГЛАВЛЕНИЕ

Учебное издание

**КАНЕ** Марк Моисеевич

# ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ. ОЦЕНКА ПРАВОМЕРНОСТИ ПРИМЕНЕНИЯ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА В ЗАДАННЫХ УСЛОВИЯХ

Пособие
для студентов специальностей
1-36 80 02 «Инновационные технологии в машиностроении»
и 1-53 80 01 «Автоматизация»

Редактор *А. С. Кириллова*
Компьютерная верстка *Е. А. Беспанской*

Подписано в печать 15.09.2020. Формат 60×84 $^1/_{16}$. Бумага офсетная. Ризография.
Усл. печ. л. 2,91. Уч.-изд. л. 2,27. Тираж 100. Заказ 386.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет.
Свидетельство о государственной регистрации издателя, изготовителя, распространителя
печатных изданий № 1/173 от 12.02.2014. Пр. Независимости, 65. 220013, г. Минск.