

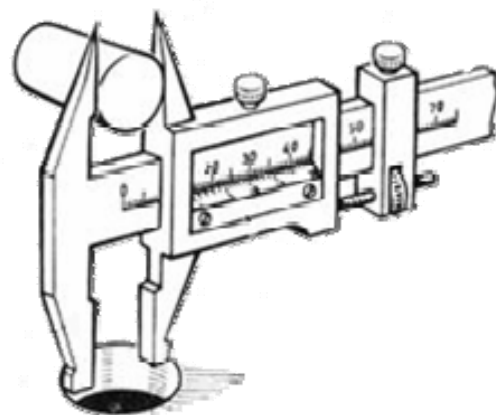
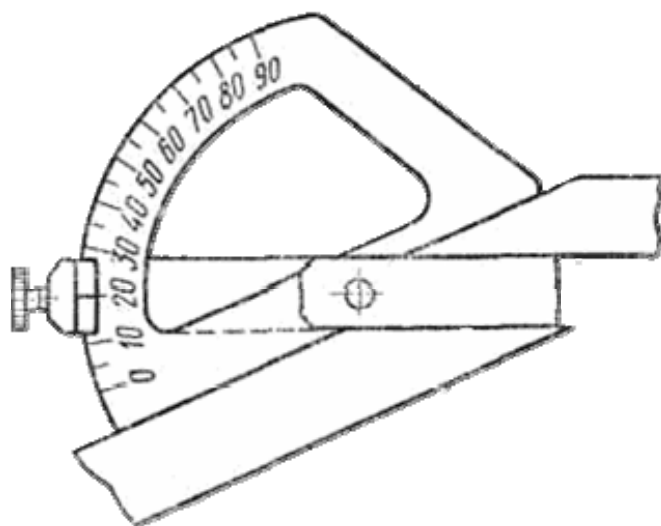
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ,
МЕХАНИКИ И ОПТИКИ

Ю.Т.Нагибин

**МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В
ОПТОЭЛЕКТРОНИКЕ.
РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ
АНАЛИЗ.**

Учебное пособие



Санкт-Петербург

2011

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**

Ю.Т. Нагибин

**МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ
В ОПТОЭЛЕКТРОНИКЕ.
РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ
АНАЛИЗ**

Учебное пособие



Санкт-Петербург

2011

Ю.Т. Нагибин. Методы статистической обработки экспериментальных данных в оптоэлектронике. Регрессионный и корреляционный анализ./ Учебное пособие. – СПб: СПбГУ ИТМО, 2011 - 53с.

В учебном пособии изложена сущность, задачи и методика проведения регрессионного и корреляционного анализа при статистической обработке результатов экспериментальных измерений в оптоэлектронике. Описана методика проверки однородности ряда дисперсий. Рассмотрены линейная и нелинейная модели с одной независимой переменной в случае равенства и неравенства дисперсий, а также методика построения точечных и интервальных оценок коэффициентов регрессии и дисперсии. Изложена методика ортогонализации системы функций на множестве значений аргумента. Описана методика проведения корреляционного анализа для случаев нормального двумерного распределения и его отсутствия. Студентам предлагаются 20 вариантов заданий для выполнения расчетной работы.

Учебное пособие предназначено для студентов 3-х – 4-х курсов инженерно-физического факультета, обучающихся по специальностям: 20020104 «Лазерная техника и лазерные технологии», 200201101 «Биомедицинская оптика», 20020107 «Оптика светового дизайна», 140400 «Техническая физика».

Рекомендовано к печати Ученым Советом ИФФ (протокол № 3 от 15 марта 2011 г.).



В 2009 году Университет стал победителем многоэтапного конкурса, в результате которого определены 12 ведущих университетов России, которым присвоена категория «Национальный исследовательский университет». Министерством образования и науки Российской Федерации была утверждена Программа развития государственного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет информационных технологий, механики и оптики» на 2009–2018 годы.

© Санкт-Петербургский государственный университет информационных технологий, механики и оптики, 2011

© Ю.Т.Нагибин, 2011

Содержание	стр.
1. Регрессионный анализ.....	4
Введение.....	4
1.1 Проверка однородности ряда дисперсий.....	6
1.2 Нелинейная модель с одной независимой переменной (случай равенства дисперсий).....	8
1.3 Линейная модель с одной независимой переменной (случай равенства дисперсий).....	12
1.4 Нелинейная модель с одной независимой переменной (случай неравенства дисперсий).....	16
1.5 Линейная модель с одной независимой переменной (случай неравенства дисперсий).....	18
2. Корреляционный анализ.....	21
2.1 Уравнения прямых регрессии.....	21
2.2 Корреляционный анализ при нормальном распределении.....	26
2.3 Корреляционный анализ при отсутствии нормальности распределения.....	31
Заключение.....	34
3. Задание на расчетную работу.....	35
Приложение 1. Ортогонализация системы функций на множестве Значений аргумента.....	36
Приложение 2. Выбор функции $f(x)$ для проведения регрессионного анализа в случае неравенства дисперсий.....	40
Приложение 3.....	43
Варианты заданий.....	47
Литература.....	50

1. Регрессионный анализ

Введение

Регрессионный анализ представляет собой метод статистического исследования, позволяющий строить эмпирические модели, основанные на функциональных статистических зависимостях. Основной задачей регрессионного анализа является построение математической модели (кривой регрессии), описывающей связь между переменными величинами: зависимой, случайной и одним либо несколькими независимыми неслучайными параметрами. Наиболее полно и доступно регрессионный анализ изложен в работах [12, 13].

Пусть требуется определить зависимость между некоторыми величинами Q и x по результатам L групп наблюдений объемом n_j каждая - $\{y_{ij}\}_{i=1}^{n_j}, j = 1, 2, \dots, L$, (см. рис.1), т.е. требуется построить зависимость

$$Q = \psi(x). \quad (1)$$

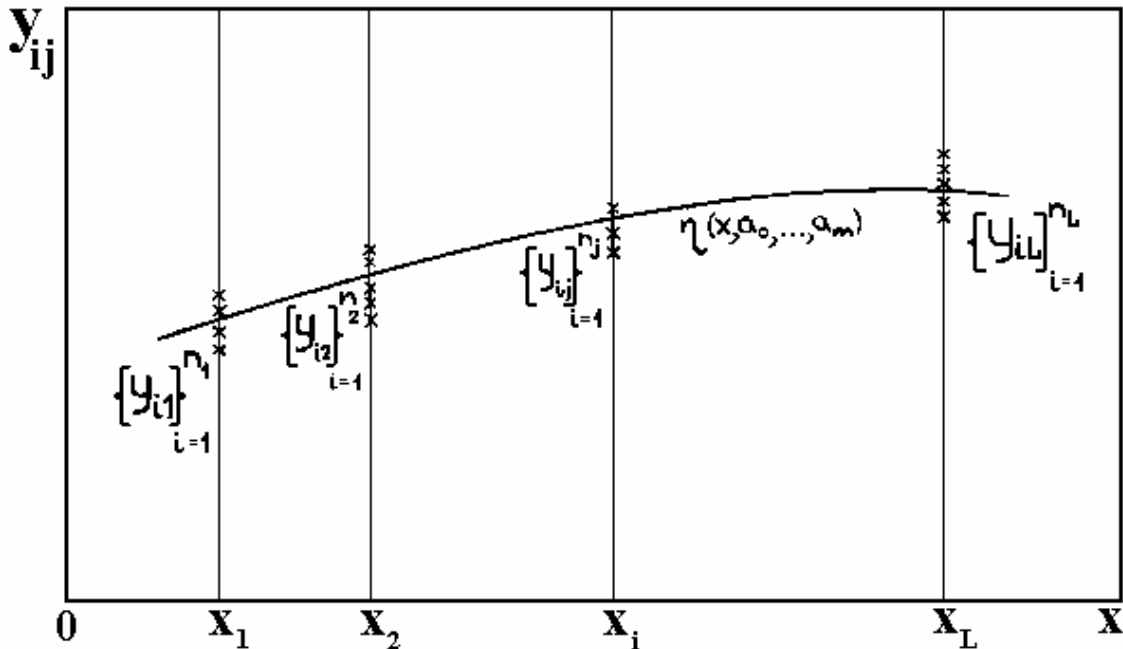


Рис.1

Зависимость между результатами групп наблюдений

$\{y_{ij}\}_{i=1}^{n_j}, j = 1, 2, \dots, L$ от аргумента x .

При отсутствии какой-либо информации относительно вида функции $\psi(x)$ сформулированная задача не может быть решена. Поэтому считают, что зависимость Q от x при $x_1 \leq x \leq x_L$ представляется функциональной зависимостью известного вида, но с неизвестными параметрами, т.е.

$$Q = \psi(x) = \eta(x, a_0, a_1, \dots, a_m), \quad (2)$$

где a_0, a_1, \dots, a_m - неизвестные параметры, η - функция известного вида (кривая регрессии), вид которой устанавливается по каким-либо теоретическим или практическим соображениям.

Тогда задача определения $\psi(x)$ сводится к построению статистических оценок коэффициентов регрессии a_0, a_1, \dots, a_m по результатам эксперимента.

При решении этой задачи будем полагать, что результаты наблюдений (см. рис.1) представляют собой независимые выборки $\{y_{i1}\}_{i=1}^{n_1}, \dots, \{y_{iL}\}_{i=1}^{n_L}$ из нормальных распределений с параметрами

$$N[\eta(x_1, a_0, \dots, a_m), \sigma_1^2], \dots, N[\eta(x_L, a_0, \dots, a_m), \sigma_L^2].$$

Тогда оценки коэффициентов регрессии, определяющие положение теоретической кривой $\eta(x_1, a_0, \dots, a_m)$ на рис.1 можно строить как оценки неизвестных параметров, используя метод наименьших квадратов.

Согласно этому методу оценки $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_m$ нужно выбрать так, чтобы сумма квадратов отклонений теоретического значения

$$Q_j = \eta(x_j, \tilde{a}_0, \dots, \tilde{a}_m) \text{ от экспериментального } Q_j = \bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \text{ по всем}$$

точкам $\{x_j\}_{j=1}^L$ была бы минимальной. Тогда получим систему уравнений относительно точечных оценок коэффициентов регрессии:

$$\frac{\partial M}{\partial \tilde{a}_g} = 0, \quad g = 0, 1, \dots, m$$

$$M = \sum_{j=1}^L n_j (\bar{y}_j - \eta(x_j, \tilde{a}_0, \dots, \tilde{a}_m))^2 \quad (3)$$

Здесь величина M характеризует суммарное рассогласование между экспериментальным значением Q_j и теоретическим расчетом Q_j с помощью модели (2) по всем экспериментальным условиям $\{x_j\}_{j=1}^L$ с учетом количества наблюдений n_j в каждой точке.

Система уравнений (3) будет непротиворечива только при условии

$$L \geq m + 1, \quad (4)$$

т.е. число экспериментальных условий L должно быть больше или равно числу оцениваемых параметров модели (2).

Рассмотрим регрессионный анализ для случаев равенства дисперсий $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2 = \sigma^2$ и их неравенства $\sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_L^2$.

В соответствии с этим, перед проведением регрессионного анализа необходимо проверить гипотезу об однородности дисперсий выборок $\{y_{ij}\}_{i=1}^{n_j}, j = 1, 2, \dots, L$.

При этом различают два случая: число выборок $L=2$ и $L>2$.

1.1 Проверка однородности ряда дисперсий

1) $L=2$, т.е. имеется две выборки $\{y_{i1}\}_{i=1}^{n_1}$ и $\{y_{i2}\}_{i=1}^{n_2}$. Требуется проверить гипотезу $\sigma_1^2 = \sigma_2^2$ против альтернативы $\sigma_1^2 \neq \sigma_2^2$. Приведем критерий для такой проверки:

$$\text{если } \Pi_1 \leq U \leq \Pi_2, \text{ то } \sigma_1^2 = \sigma_2^2, \quad (5)$$

$$\text{если } U < \Pi_1 \text{ или } U > \Pi_2, \text{ то } \sigma_1^2 \neq \sigma_2^2, \quad (6)$$

где

$$U = \frac{S_1^2}{S_2^2}, \quad S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}, \quad j = 1, 2$$

Здесь S_j^2 и \bar{Y}_j - выборочная дисперсия и выборочное среднее соответственно. В качестве порогов Π_1 и Π_2 выбираются квантили распределения Фишера (см. табл.1,2 Приложения 3).

$$\Pi_1 = f_{n_1-1; n_2-1; q/2}; \quad \Pi_2 = f_{n_1-1; n_2-1; 1-q/2}, \quad (7)$$

где q-уровень значимости.

2) $L > 2$. Проверка гипотезы $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2 = \sigma^2$ против альтернативы $\sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma^2$.

Имеется L выборок $\{y_{ij}\}_{i=1}^{n_j}, j = 1, 2, \dots, L$ из нормальных распределений $N(a_j, \sigma_j^2), j = 1, 2, \dots, L$. Требуется проверить гипотезу об однородности ряда дисперсий. Для этого используется критерий Бартлетта (условие его использования $n_j \geq 3$):

если $U < X_{L-1, 1-q}^2$, то дисперсии равны,

если $U > X_{L-1, 1-q}^2$, то дисперсии не равны, (8)

где $q=0,02 \div 0,1$

$$U = -\frac{1}{c} \sum_{j=1}^L (n_j - 1) \ln \frac{S_j^2}{S^2}, \quad (9)$$

$$S^2 = \frac{1}{N - L} \sum_{j=1}^L S_j^2 (n_j - 1), \quad (10)$$

$$N = \sum_{j=1}^L n_j, \quad (11)$$

$$c = 1 + \frac{1}{3(L-1)} \left(\sum_{j=1}^L \frac{1}{n_j - 1} - \frac{1}{N - L} \right), \quad (12)$$

$$S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2, \quad j = 1, 2, \dots, L \quad (13)$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}. \quad (14)$$

$X_{L-1,1-q}^2$ - соответствующий квантиль распределения X^2 (таблица 3 Приложения 3).

1.2 Нелинейная модель с одной независимой переменной (случай $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2 = \sigma^2$)

Рассмотрим проведение регрессионного анализа для модели следующего вида:

$$\eta(x, a_0, \dots, a_m) = \sum_{g=0}^m a_g \varphi_g(x), \quad (15)$$

где $\{\varphi_g(x)\}_{g=0}^m$ - некоторая система нелинейных функций от x , которая выбирается в зависимости от характера объекта. Например, если взять $\varphi_0(x)=1$, $\varphi_1(x)=x$, ..., $\varphi_m(x)=x^m$, то получится модель в виде полинома m -ой степени. При $\varphi_g(x) = \sin(\omega_g x + d_g)$ имеем модель, образованную суммой тригонометрических функций. Эти два типа моделей получили наиболее широкое распространение.

Построим точечные оценки коэффициентов регрессии \tilde{a}_g и дисперсии S^2 . В рассматриваемом случае величина M из системы уравнений (3) имеет вид:

$$M = \sum_{j=1}^L n_j (\bar{y}_j - \sum_{g=0}^m \tilde{a}_g \varphi_g(x_j))^2. \quad (16)$$

Дифференцируя, получаем:

Тогда легко находим точечные оценки коэффициентов \tilde{a}_g

$$\tilde{a}_g = \frac{(\varphi_g, \bar{y})}{(\varphi_g, \varphi_g)} = \frac{\sum_{j=1}^L n_j \varphi_g(x_j) \bar{y}_j}{\sum_{j=1}^L n_j \varphi_g^2(x_j)}, \quad (22)$$

$g=0,1,\dots,m$.

Точечная оценка S^2 дисперсии σ^2 имеет вид:

$$S^2 = \frac{1}{N - m - 1} \sum_{j=1}^L \sum_{i=1}^{n_j} (y_{ij} - \sum_{g=0}^m \tilde{a}_g \varphi_g(x_j))^2 \quad (23)$$

$$N = \sum_{j=1}^L n_j$$

Построим интервальные оценки для \tilde{a}_g и σ^2 при заданной доверительной вероятности α в виде:

$$[\tilde{a}_g \pm t(N-m-1, 1-\varepsilon/2) \cdot S_{\tilde{a}_g}], \quad (24)$$

где

$$S_{\tilde{a}_g} = \frac{S}{\sqrt{\sum_{j=1}^L n_j \varphi_g^2(x_j)}} \quad (25)$$

$$g = 0, 1, \dots, m$$

Интервальная оценка для σ^2 имеет вид:

$$\left[S \cdot \sqrt{\frac{N-m-1}{X_{N-m-1, 1-\varepsilon/2}^2}}, S \cdot \sqrt{\frac{N-m-1}{X_{N-m-1, \varepsilon/2}^2}} \right] \quad (26)$$

$$\varepsilon = 1 - \alpha$$

Здесь используются соответствующие квантили распределений Стьюдента $t(N-m-1, 1-\varepsilon/2)$ и X^2 распределения ($X^2_{N-m-1, \varepsilon/2}$), определяемые по таблицам 3 и 4 Приложения 3.

Теперь, зная оценки коэффициентов, легко найти точечную и интервальную оценки для кривой регрессии при любом значении x :

$$\tilde{Q} = \sum_{g=0}^m \tilde{a}_g \varphi_g(x) \quad (27)$$

$$Q \pm t(N-m-1, 1-\varepsilon/2) \cdot S_{\tilde{Q}} \quad (28)$$

и

$$S_{\tilde{Q}} = \sqrt{\sum_{g=0}^m S_{\tilde{a}_g}^2 \varphi_g^2(x)} \quad (29)$$

После построения модели необходимо проверить степень ее адекватности (или соответствия) описываемому экспериментальному материалу, что достигается проверкой гипотезы адекватности.

Для этого вычисляется величина F .

$$F = \frac{S_{AD}^2}{S_{BM}^2} \quad (30)$$

Оценка дисперсии, характеризующая рассогласование теоретических и экспериментальных данных S_{AD}^2 , т.е. адекватность модели, рассчитывается по формуле:

$$S_{AD}^2 = \frac{1}{L-m-1} \cdot \sum_{j=1}^L n_j (\bar{y}_j - \sum_{g=0}^m \tilde{a}_g \varphi_g(x_j))^2, \quad (31)$$

а характеристика случайной погрешности наблюдений (оценка дисперсии без модели) определяется по формуле:

$$S_{3M}^2 = \frac{1}{N-L} \cdot \sum_{j=1}^L (n_j - 1) S_j^2, \quad (32)$$

где

$$S_j^2 = \frac{1}{n_j - 1} \cdot \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

Для проверки адекватности модели сравнивают величину F с порогом Π :

$$\Pi = f_{L-m-1, N-L, 1-q} \quad (33)$$

За порог Π принимают квантиль распределения Фишера $f_{L-m-1, N-L, 1-q}$, определяемый по таблицам 1,2 Приложения 3 при заданном уровне значимости q (обычно q принимается равным 0,05). Если $F < \Pi$, то модель адекватна и на этом регрессионный анализ заканчивается. Если $F > \Pi$, то модель неадекватна. Тогда необходимо изменить исходную модель и вновь провести все вычисления по формулам (22)-(33). Заметим, что проверка адекватности возможна только при условии $L > m+1$, $n_j \geq 2$.

1.3 Линейная модель с одной независимой переменной

(случай $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2 = \sigma^2$).

В этом случае модель имеет вид:

$$Q = a_0 + a_1 x \quad (34)$$

В соответствии с неравенством (4) для построения этой модели необходимо использовать, как минимум, два экспериментальных условия, т.е. $L \geq 2$.

Регрессионный анализ проводится по рассмотренной выше схеме для нелинейной модели. Поэтому приведем соответствующие выражения, получающиеся из формулы (22) при $m=1$, $\varphi_0(x)=1$, $\varphi_1(x)=x$.

Точечные оценки \tilde{a}_1 и \tilde{a}_0 для коэффициентов регрессии равны:

$$\tilde{a}_1 = \frac{\sum_{j=1}^L n_j (x_j - \bar{x}) \bar{y}_j}{\sum_{j=1}^L (x_j - \bar{x})^2} \quad (35)$$

$$\tilde{a}_0 = \left(\frac{1}{N} \sum_{j=1}^L n_j \bar{y}_j \right) - \tilde{a}_1 \bar{x} \quad (36)$$

где

$$\bar{x} = \frac{1}{N} \sum_{j=1}^L n_j x_j \quad (37)$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij} \quad (38)$$

$$N = \sum_{j=1}^L n_j \quad (39)$$

Точечная оценка дисперсии в соответствии с выражением (23) имеет вид:

$$S^2 = \frac{1}{N-2} \sum_{j=1}^L \sum_{i=1}^{n_j} (y_{ij} - \tilde{a}_0 - \tilde{a}_1 x_j)^2 \quad (40)$$

Заметим, что запись модели (34) в виде

$$Q = a_0 + a_1 (x - \bar{x}) \quad (41)$$

дает более простое выражение для \tilde{a}_0 :

$$\tilde{a}_0 = \frac{1}{N} \sum_{j=1}^L n_j \bar{y}_j \quad (42)$$

при сохраняющихся соотношениях (35) и (40) для \tilde{a}_1 и S^2 соответственно.

Интервальные оценки a_0 и a_1 при доверительной вероятности α имеют вид для a_0 :

$$[\tilde{a}_0 \pm t \cdot (N-2, 1-\varepsilon/2) \cdot S_{\tilde{a}_0}] \quad , \quad (43)$$

а для a_1 :

$$[\tilde{a}_1 \pm t \cdot (N-2, 1-\varepsilon/2) \cdot S_{\tilde{a}_1}], \quad (44)$$

где

$$S_{\tilde{a}_0} = \frac{S}{\sqrt{N}} \quad (45)$$

$$S_{\tilde{a}_1} = \frac{S}{\sqrt{\sum_{j=1}^L n_j (x_j - \bar{x})^2}} \quad (46).$$

$t(N-2, 1-\varepsilon/2)$ - соответствующие квантили распределения Стьюдента (Таблица 4 Приложения 3).

Интервальная оценка для σ имеет вид:

$$\left[S \cdot \sqrt{\frac{N-2}{X_{N-2, 1-\varepsilon/2}^2}}, S \cdot \sqrt{\frac{N-2}{X_{N-2, \varepsilon/2}^2}} \right], \quad (47)$$

где $X_{k, \gamma}^2$ - квантили распределения X^2 (Таблица 3 Приложения 3).

После нахождения оценок коэффициентов регрессии строят точечную и интервальную оценки для искомой модели (34) при любом значении x (рис.2):

$$\tilde{Q} = \tilde{a}_0 + \tilde{a}_1 x \quad (48)$$

$$[\tilde{Q} \pm t \cdot (N-2, 1-\varepsilon/2) \cdot S_{\tilde{Q}}] \quad (49)$$

где

$$S_{\tilde{Q}} = \sqrt{S_{\tilde{a}_0}^2 + S_{\tilde{a}_1}^2 (x - \bar{x})^2} \quad (50)$$

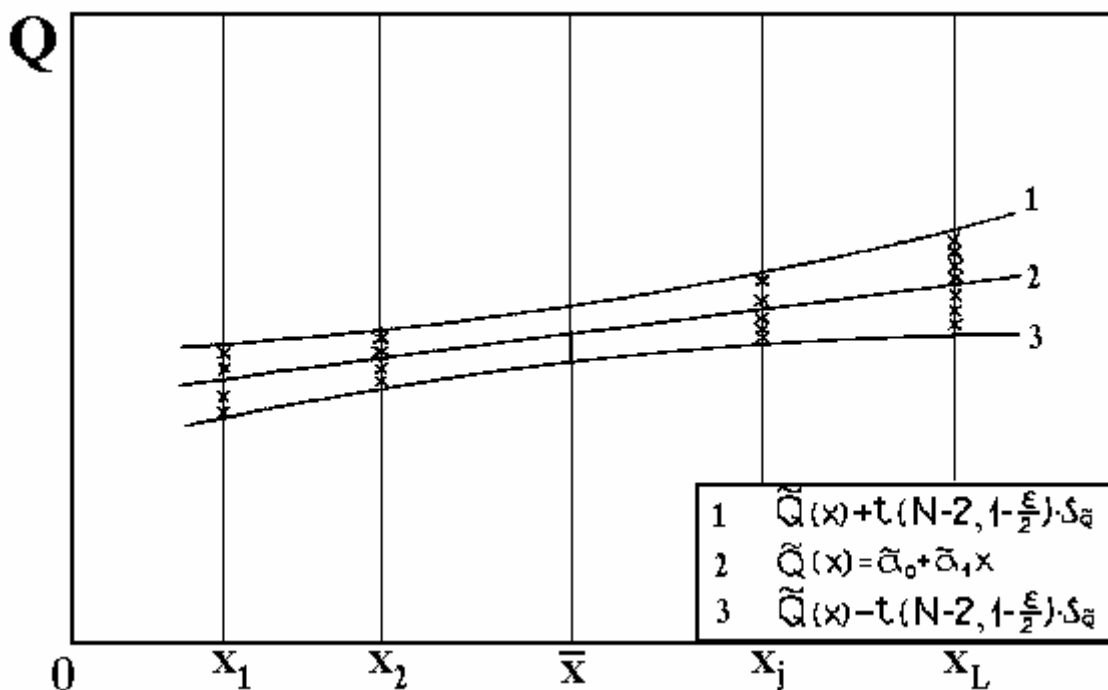


Рис.2

Построение точечной и интервальной оценок для искомой модели Q при различных значениях аргумента x .

Проверка гипотезы адекватности возможна только при условии $L > 2$, $n_j \geq 2$. Для этого рассчитывается оценка дисперсии адекватности

$$S_{AD}^2 = \frac{1}{L - 2} \sum_{j=1}^L n_j (\bar{y}_j - \tilde{\alpha}_0 - \tilde{\alpha}_1 x_j)^2 \quad (51)$$

и оценка дисперсии без модели

$$S_{BM}^2 = \frac{1}{N - L} \sum_{j=1}^L (n_j - 1) S_j^2, \quad (52)$$

где S_j^2 определяется по формуле (6). По формуле (30) рассчитывается величина F .

Если $F < \Pi$, то модель (34) адекватна, если $F > \Pi$, то модель (34) неадекватна, где $\Pi = f_{L-2, N-L, 1-q}$ - соответствующий квантиль распределения Фишера (Таблица 1,2 Приложения 3).

1.4 Нелинейная модель с одной независимой переменной

(случай $\sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_L^2$)

В этом случае регрессионный анализ выполняется в предположении, что известна зависимость $\sigma^2(x)$:

$$\sigma^2(x) = \sigma^2 f(x), \quad (53)$$

где σ^2 неизвестна, а функция $f(x)$ известна. В этом случае

$$\sigma_j^2 = \sigma^2 f(x_j). \quad (54)$$

В качестве функции $f(x)$ обычно выбирают аналитическую аппроксимацию зависимости значений выборочных дисперсий S_j^2 от аргумента x_j . Аппроксимация производится с помощью метода наименьших квадратов (Приложение 2).

Регрессионный анализ в этом случае проводится в той же последовательности, что и рассмотренный ранее. Приведем все необходимые формулы, обозначив

$$\omega_j = \frac{1}{f(x_j)}, \quad (55)$$

где $f(x_j)$ определяется по методике, изложенной в Приложении 2, для всех значений аргумента x_j ($j=1, 2, \dots, L$). Перед проведением регрессионного анализа для построения системы ортогональных функций необходимо использовать веса $\{\omega_j n_j\}_{j=1}^L$, т.е. ортогональные функции $\{\varphi_g(x)\}_{g=0}^m$ должны обладать свойством

$$\sum_{j=1}^L \omega_j n_j \varphi_{g_1}(x_j) \varphi_{g_2}(x_j) = 0 \quad g_1 \neq g_2. \quad (56)$$

Перейти от неортогональной системы к ортогональной можно по формулам (82)-(90) (Приложение 1), где в качестве $(\varphi_{g_1}, \varphi_{g_2})$ теперь выступает

$$(\varphi_{g_1}, \varphi_{g_2}) = \sum_{j=1}^L \omega_j n_j \varphi_{g_1}(x_j) \varphi_{g_2}(x_j) \quad (57)$$

(сравните с выражением (18)).

При использовании в качестве $\{\varphi_g(x)\}_{g=0}^m$ системы функций, ортогональных на множестве значений аргумента $\{x_j\}_{j=1}^L$ с весами $\{\omega_j n_j\}_{j=1}^L$, выражения для точечных оценок коэффициентов a_ν принимают вид:

$$\tilde{a}_g = \frac{\sum_{j=1}^L \omega_j n_j \varphi_g(x_j) \bar{y}_j}{\sum_{j=1}^L \omega_j n_j \varphi_g^2(x_j)}, \quad g = 0, 1, \dots, m \quad (58)$$

Точечная оценка для σ^2 имеет вид:

$$S^2 = \frac{1}{N - m - 1} \sum_{j=1}^L \omega_j \sum_{i=1}^{n_j} \left[y_{ij} - \sum_{g=0}^m \tilde{a}_g \varphi_g(x_j) \right]^2 \quad (59)$$

$$N = \sum_{j=1}^L n_j$$

Интервальные оценки для a_ν при вероятности α имеют вид:

$$\left[\tilde{a}_g \pm t(N - m - 1, 1 - \varepsilon / 2) \cdot S_{\tilde{a}_g} \right] \quad \varepsilon = 1 - \alpha \quad (60)$$

$$S_{\tilde{a}_g} = \frac{S}{\sqrt{\sum_{j=1}^L \omega_j n_j \varphi_g^2(x_j)}} \quad (61)$$

Точечная оценка значений Q при любом X вычисляются по формуле

$$\tilde{Q} = \sum_{g=0}^m \tilde{a}_g \varphi_g(x), \quad (62)$$

а интервальная оценка для этого значения вычисляется по формуле

$$\left[\tilde{Q} \pm t(N - m - 1, 1 - \varepsilon / 2) \cdot S_{\tilde{Q}} \right], \quad \varepsilon = 1 - \alpha, \quad (63)$$

где

$$S_{\tilde{Q}} = \sqrt{\sum_{g=0}^m S_{\tilde{a}_g}^2 \varphi_g^2(x)}. \quad (64)$$

Гипотеза адекватности проверяется сравнением величины

$$F = \frac{S_{AD}^2}{S_{BM}^2} \quad (65)$$

с порогом $\Pi = f_{L-m-1, N-L, 1-\alpha}$ по правилу, указанному ранее. Здесь

$$S_{AD}^2 = \frac{1}{L - m - 1} \sum_{j=1}^L \omega_j n_j \left[\bar{y}_j - \sum_{g=0}^m \tilde{a}_g \varphi_g(x_j) \right]^2 \quad (66)$$

$$S_{\zeta l}^2 = \frac{1}{N - L} \sum_{j=1}^L \omega_j (n_j - 1) S_j^2 \quad (67)$$

1.5 Линейная модель с одной независимой переменной

(случай $\sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma^2$)

Для линейной модели вида (34), так же, как и в предыдущем случае будем считать, что

$$\sigma_j^2 = \sigma^2 f(x_j) = \frac{\sigma^2}{\omega_j} \quad (68)$$

Тогда точечные оценки для a_0 и a_1 имеют вид:

$$\tilde{a}_0 = \frac{1}{P} \sum_{j=1}^L \omega_j n_j \bar{y}_j - \tilde{a}_1 \bar{x} \quad (69)$$

$$\tilde{a}_1 = \frac{\sum_{j=1}^L \omega_j n_j (x_j - \bar{x}) \bar{y}_j}{\sum_{j=1}^L \omega_j n_j (x_j - \bar{x})^2} \quad (70)$$

где

$$\bar{x} = \frac{1}{P} \sum_{j=1}^L \omega_j n_j x_j, \quad P = \sum_{j=1}^L \omega_j n_j \quad (71)$$

Точечная оценка для σ^2 имеет вид:

$$S^2 = \frac{1}{N-2} \sum_{j=1}^L \omega_j \sum_{i=1}^{n_j} (y_{ij} - \tilde{a}_0 - \tilde{a}_1 x_j)$$

$$N = \sum_{j=1}^L n_j \quad (72)$$

Интервальные оценки для a_0 и a_1 при вероятности α :

$$[\tilde{a}_0 \pm t(N-2, 1-\varepsilon/2) \cdot S_{\tilde{a}_0}] , \quad (73)$$

$$[\tilde{a}_1 \pm t(N-2, 1-\varepsilon/2) \cdot S_{\tilde{a}_1}] , \quad (74)$$

$$S_{\tilde{a}_0} = \frac{S}{\sqrt{P}} \quad ; \quad S_{\tilde{a}_1} = \frac{S}{\sqrt{\sum_{j=1}^L \omega_j n_j (x_j - \bar{x})^2}} . \quad (75)$$

Точечная оценка Q при любом X определяется по формуле:

$$\tilde{Q} = \tilde{a}_0 + \tilde{a}_1 x \quad (76)$$

а интервальная оценка для этого значения имеет вид:

$$[\tilde{Q} \pm t(N-2, 1-\varepsilon/2) \cdot S_{\tilde{Q}}], \quad (77)$$

где

$$S_{\tilde{Q}} = \sqrt{S_{\tilde{a}_0}^2 + S_{\tilde{a}_1}^2 (x - \bar{x})^2}. \quad (78)$$

Гипотеза адекватности проверяется путем сравнения величины F с порогом $\Pi = f_{L-2, N-L, 1-\alpha}$.

$$F = \frac{S_{AD}^2}{S_{BM}^2}, \quad (79)$$

$$S_{AD}^2 = \frac{1}{L-2} \sum_{j=1}^L \omega_j n_j (\bar{y}_j - \tilde{a}_0 - \tilde{a}_1 x_j)^2, \quad (80)$$

$$S_{BM}^2 = \frac{1}{N-L} \sum_{j=1}^L (n_j - 1) \omega_j S_j^2. \quad (81)$$

2. Корреляционный анализ.

2.1. Уравнения прямых регрессии.

В разделе 1 данного учебного пособия был рассмотрен один из методов статистического исследования результатов наблюдений в оптоэлектронике, позволяющий строить эмпирические модели, основанные на функциональных статистических зависимостях. Основой задачей регрессионного анализа являлось построение математической модели (кривой регрессии), описывающей связь между переменными величинами: зависимой, случайной и одним либо несколькими независимыми неслучайными пар.

В этом разделе учебного пособия мы рассмотрим основные понятия, связанные с корреляционным анализом, которые, по нашему мнению, наиболее полно и доступно изложены в [12].

Итак, изучаемый объект имеет два параметра, описываемых совокупностью из двух случайных величин (x, y) с двумерной совместной функцией плотности распределения $f(x, y)$. Искомыми моделями являются две функции условных математических ожиданий:

$$E\left\{\frac{x}{y=y}\right\} = a_{x/y} = \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_2(y)} dx = \int_{-\infty}^{\infty} x f(x/y) dx, \quad (82)$$

$$E\left\{\frac{y}{x=x}\right\} = a_{y/x} = \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_1(x)} dy = \int_{-\infty}^{\infty} y f(y/x) dy, \quad (83)$$

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y) dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x,y) dx, \quad (84)$$

где $f_1(x)$ - функция плотности распределения случайной величины X , $f_2(y)$ - функция плотности распределения случайной величины Y , $f_1(x)$ и $f_2(y)$ связаны с функцией плотности совместного распределения соотношениями (84).

$f\left(\frac{x}{y}\right) = x \frac{f(x,y)}{f_2(y)}$ называется условной плотностью распределения X при условии, что Y приняла заданное значение y . Аналогичный смысл имеет и $f\left(\frac{y}{x}\right)$ из (83). $a_{x/y}$ называется линией регрессии случайной величины X по Y , а $a_{y/x}$ - линией регрессии Y по X .

Теперь рассмотрим важный для практики случай, когда функции $a_{x/y}$ и $a_{y/x}$ линейны, т.е.

$$a_{x/y} = c + by, \quad (85)$$

$$a_{y/x} = l + hx. \quad (86)$$

Оказывается, что коэффициенты c , b , l и h из (85) и (86) можно выразить через числовые характеристики системы случайных величин (X, Y) – их математические ожидания, средние квадратические отклонения и коэффициент корреляции.

Рассмотрим, например, $a_{x/y}$. Перепишем (85) с учетом (82)

$$a_{x/y} = \int_{-\infty}^{\infty} x \frac{f(x,y)}{f_2(y)} dx = c + by. \quad (87)$$

Умножим обе части (87) на $f_2(y)$ и проинтегрируем от $-\infty$ до $+\infty$ по y :

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f(x,y)}{f_2(y)} dx \right) f_2(y) dy = c \int_{-\infty}^{\infty} f_2(y) dy + b \int_{-\infty}^{\infty} y f_2(y) dy. \quad (88)$$

Так как

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f(x,y)}{f_2(y)} dx \right) f_2(y) dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} \frac{f(x,y)}{f_2(y)} f_2(y) dy \right) dx =$$

$$\int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y) dy \right) dx = \int_{-\infty}^{\infty} x f_1(x) dx = E\{X\} = a_x,$$

$$\int_{-\infty}^{\infty} f_2(y) dy = 1, \quad \int_{-\infty}^{\infty} y f_2(y) dy = E\{Y\} = a_y, \quad (89)$$

где a_x – математическое ожидание X , a_y – математическое ожидание Y (не путать с условными математическими ожиданиями $a_{x/y}$ и $a_{y/x}$), то

$$a_{x/y} = c + b a_y. \quad (90)$$

Вычитая из (85) равенство (90), получаем:

$$a_{x/y} - a_x = b(y - a_y). \quad (91)$$

Снова подставим в (91) $a_{x/y}$ из (82):

$$\frac{\int_{-\infty}^{\infty} f(x,y) dx}{f_2(y)} - a_x = b(y - a_y). \quad (92)$$

Отсюда

$$\int_{-\infty}^{\infty} x f(x,y) dx - a_x \int_{-\infty}^{\infty} f(x,y) dx = b(y - a_y) f_2(y) \quad (93)$$

и

$$(y - a_y) \int_{-\infty}^{\infty} (x - a_x) f(x,y) dx = b(y - a_y)^2 f_2(y). \quad (94)$$

Интегрируя (94) по y от $-\infty$ до ∞ , имеем:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - a_x)(y - a_y) f(x,y) dx dy = b \int_{-\infty}^{\infty} (y - a_y)^2 f_2(y) dy, \quad (95)$$

или

$$E\{(X - a_x)(Y - a_y)\} = b\sigma_y^2. \quad (96)$$

И тогда

$$b = \frac{E\{(X - a_x)(Y - a_y)\}}{\sigma_y^2}, \quad (97)$$

где σ_y - среднее квадратическое отклонение Y .

Вспоминая, что

$$E\{(X - a_x)(Y - a_y)\} = \rho_{xy} \sigma_x \sigma_y, \quad (98)$$

где σ_x - среднее квадратическое отклонение X , ρ_{xy} - коэффициент корреляции случайных величин X и Y (напомним, что $-1 \leq \rho_{xy} \leq 1$). Окончательно получаем:

$$b = \frac{\rho_{XY}\sigma_X}{\sigma_Y} . \quad (99)$$

Из (91) и (99) следует, что уравнение прямой регрессии X по Y имеет вид

$$a_{\frac{X}{Y}} = a_X + \frac{\rho_{XY}\sigma_X}{\sigma_Y} (y - a_Y). \quad (100)$$

Аналогично для $a_{Y/X}$ находим

$$a_{\frac{Y}{X}} = a_Y + \frac{\rho_{XY}\sigma_Y}{\sigma_X} (x - a_X). \quad (101)$$

Остановимся несколько подробнее на свойствах ρ_{XY} . Из (100) и (101) вытекает, что если $\rho_{XY} = 0$, то

$$a_{\frac{X}{Y}} = a_X, \quad a_{\frac{Y}{X}} = a_Y, \quad (102)$$

что соответствует отсутствию связи между X и Y в рамках моделей (82), (83). Такие случайные величины называются некоррелированными. Кроме этого, можно доказать два важных свойства ρ_{XY} :

а) если $|\rho_{XY}| = 1$, то X и Y связаны строгой линейной зависимостью, т.е. справедливы не только модели

$$a_{\frac{X}{Y}} = a_X \pm \frac{\sigma_X}{\sigma_Y} (y - a_Y), \quad a_{\frac{Y}{X}} = a_Y \pm \frac{\sigma_Y}{\sigma_X} (x - a_X), \quad \rho_{XY} = \pm 1 \quad (103)$$

для условных математических ожиданий $a_{\frac{X}{Y}}$ и $a_{\frac{Y}{X}}$, но и соотношения

$$X = a_X \pm \frac{\sigma_X}{\sigma_Y} (Y - a_Y), \quad Y = a_Y \pm \frac{\sigma_Y}{\sigma_X} (X - a_X), \quad (104)$$

для самих случайных величин X и Y;

б) если X и Y связаны нелинейной функциональной зависимостью, то модуль ρ_{XY} не равен 1.

Все это позволяет интерпретировать ρ_{XY} , как меру (характеристику) степени связи между X и Y в рамках моделей (82), (83) при условии их линейности.

Прокомментируем ограничение «в рамках моделей (82), (83)», при котором ρ_{XY} можно рассматривать как меру степени связи. Дело заключается в том, что вовсе не обязательно в качестве моделей, описывающих связь X и Y , брать функции условных математических ожиданий. С не более неудачными основаниями можно было бы, например, использовать функции условных дисперсий:

$$\sigma_{\frac{x}{y}}^2 = \int_{-\infty}^{\infty} \left(x - a_{\frac{x}{y}}\right)^2 f\left(\frac{x}{y}\right) dx,$$

$$\sigma_{y/x}^2 = \int_{-\infty}^{\infty} \left(y - a_{\frac{y}{x}}\right)^2 f\left(\frac{y}{x}\right) dy. \quad (105)$$

При этом, если функций условных математических ожиданий постоянны, то функции условных дисперсий могут быть, вообще говоря, непостоянными, т.е. в рамках моделей (105) связь X и Y существует, если они не являются статистически независимыми. Иными словами при

$$f(x, y) \neq f_1(x)f_2(y). \quad (106)$$

Когда же мы исследуем связь между X и Y в рамках моделей (82), (83), то неизбежно рассматриваем только какую-то одну сторону стохастической связи, какое-то одно ее проявление. И отсутствие какого-то одного проявления стохастической связи еще не говорит о ее полном отсутствии, заключающемся в выполнении равенства

$$f(x, y) = f_1(x)f_2(y). \quad (107)$$

В разобранном случае стохастическая связь не проявляется в условных математических ожиданиях, но может проявиться в условных дисперсиях. Исследование связи между X и Y с помощью условных математических ожиданий наглядно и удобно, но не полно отражает стохастическую связь между X и Y . Это необходимо учитывать при выполнении корреляционного анализа.

Остановимся на втором ограничении – условии линейности моделей (82), (83). Оно очень важно. Можно привести примеры, когда не только функции условных математических ожиданий непостоянны (и нелинейные), но даже X и Y непосредственно связаны функциональной нелинейной зависимостью (т.е. в рамках условных математических ожиданий стохастическая связь проявляется), а $\rho_{XY} = 0$.

Теперь рассмотрим наиболее распространенную функцию плотности двумерного нормального распределения

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}}} \exp\left(-\frac{1}{2(1-\rho_{XY})^2} \left[\frac{(x-a_X)^2}{\sigma_X^2} + \frac{(y-a_Y)^2}{\sigma_Y^2} - 2\rho_{XY} \frac{(x-a_X)(y-a_Y)}{\sigma_X\sigma_Y} \right]\right). \quad (108)$$

Для нее функции условных математических ожиданий линейны.

Этот факт доказывается путем непосредственного расчета $\frac{a_X}{Y}$ и $\frac{a_Y}{X}$ по формулам (82) – (84). Для случайных величин X и Y , имеющих плотность совместного распределения (108), ρ_{XY} является полной характеристикой степени связи, так как вместе с математическими ожиданиями и дисперсиями X и Y он дает исчерпывающие сведения о двумерной плотности распределения. В частности, если $\rho_{XY} = 0$, то X и Y статистически независимы.

Таким образом, можно сделать вывод о том, что для совместно нормально распределенных случайных величин, при условии их некоррелированности, вытекает их статистическая независимость. В общем случае это свойство отсутствует.

Как видно, при нормальном распределении (108) дело обстоит наиболее просто, и поэтому изложение корреляционного анализа мы начнем с этого случая.

2.2. Корреляционный анализ при нормальном распределении.

Итак, имеется совокупность результатов наблюдений, представляющая набор $\{x_i, y_i\}_{i=1}^n$ реализаций случайных величин (X, Y), распределенных предположительно по двумерному нормальному распределению (рис.3).

Необходимо определить эмпирическую модель $\frac{a_X}{Y}$ и $\frac{a_Y}{X}$ (или обе модели). В дальнейшем для определенности мы будем рассматривать $\frac{a_Y}{X}$.

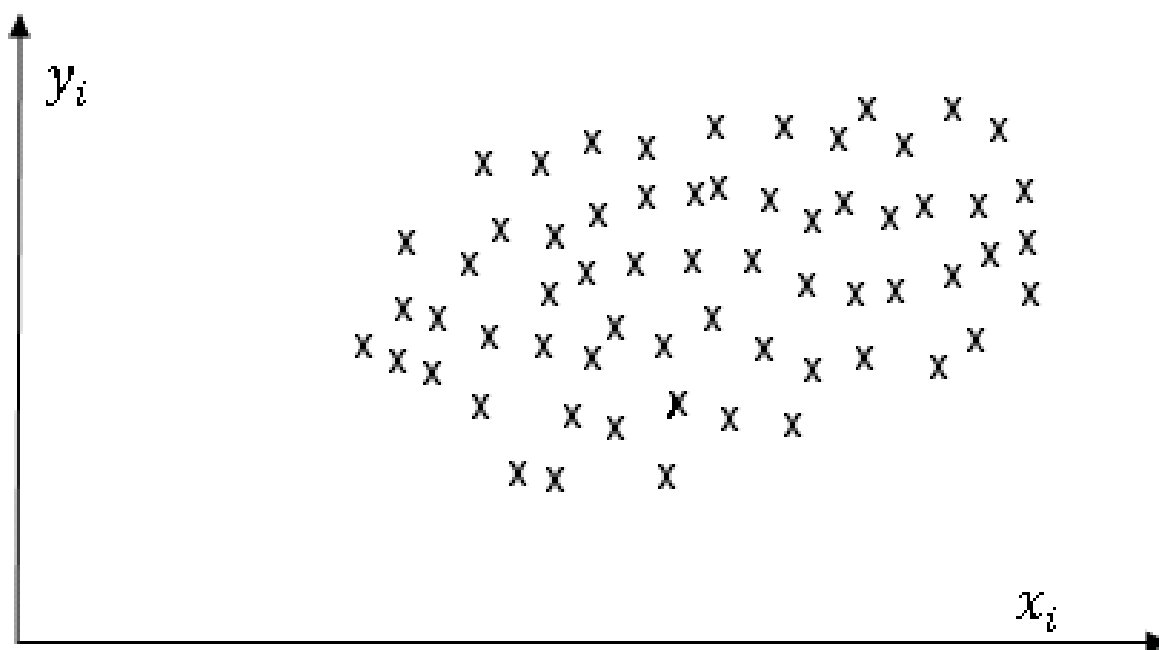


Рис. 3.

Двумерное нормальное распределение совокупности результатов наблюдений случайных величин (X, Y) .

Статистический анализ состоит из трех этапов:

- 1) проверяется гипотеза о нормальности совместного распределения случайных величин X и Y ;
- 2) проверяется гипотеза $\rho_{XY} = 0$ – статистическая связь между X и Y отсутствует;
- 3) рассчитываются оценки коэффициентов модели (86).

Разберем эти этапы более подробно. При этом ограничимся изложением основных алгоритмов статистического анализа.

Проверка гипотезы нормальности.

Сначала проводится группировка результатов наблюдений. Для этого интервал результатов наблюдений первого параметра X (так как рассматривается $\frac{\sigma_Y}{X}$) $[x_{min}, x_{max}]$ разбивается на L интервалов шириной $\Delta x_j, j = 1, \dots, L$. Обозначим через $\{y_{ij}\}_{i=1}^{n_j}$ результаты наблюдений второго параметра, для которых соответствующие результаты наблюдений первого параметра попали в j -ый интервал разбиения шириной Δx_j (см. рис. 4).

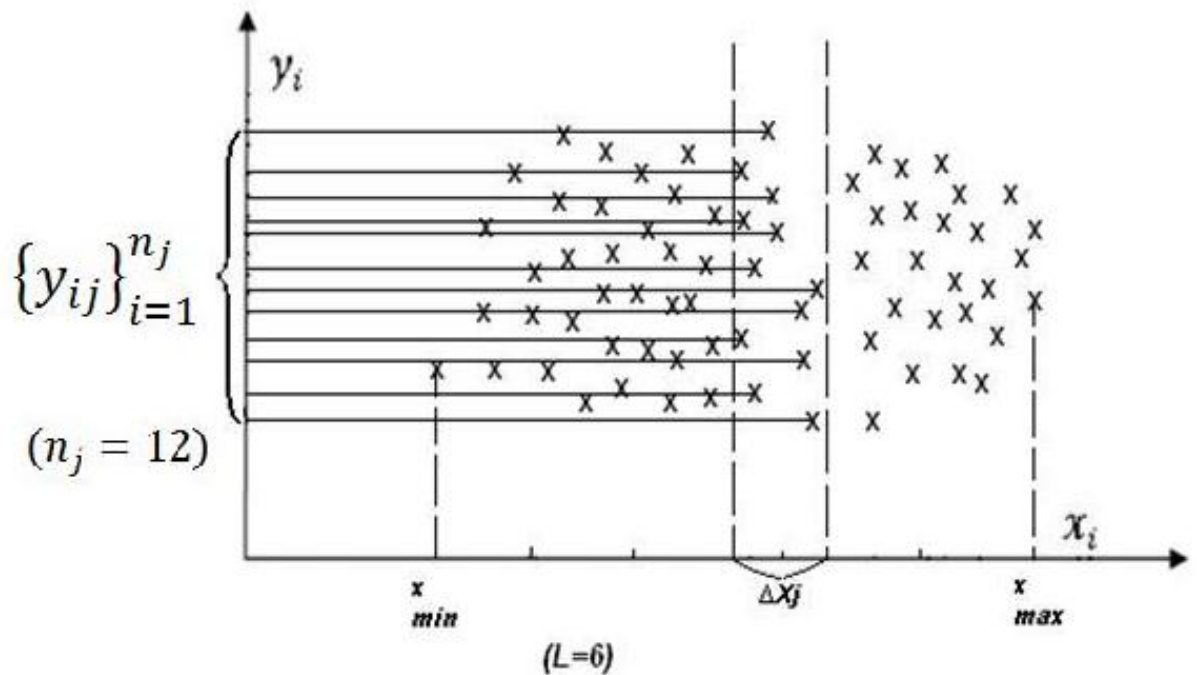


Рис. 4.

Распределение результатов наблюдений первого параметра X относительно второго параметра Y.

Введем $\{x_j^c\}_{j=1}^L$ - средние точки интервалов разбиения диапазона $[x_{\min}, x_{\max}]$. Результаты наблюдений $\{y_{ij}\}_{i=1}^{n_j}$ считаются соответствующими x_j^c , как в регрессионном анализе (рис.5). Далее проводится следующий статистический анализ:

- a) проверяется гипотеза о нормальности каждой выборки

$$\{y_{ij}\}_{i=1}^{n_j}, j = 1, \dots, L,$$

- b) проверяется гипотеза о равенстве дисперсий выборок

$$\{y_{ij}\}_{i=1}^{n_j}, j = 1, \dots, L, \text{ (см. раздел 1.1).}$$

с) результаты наблюдений $\{y_{ij}\}_{i=1}^m, j = 1, \dots, L$, при $\{x_j^c\}_{j=1}^L$ описываются линейной моделью по методике регрессионного анализа, и проверяется гипотеза об адекватности линейной модели (подраздел 1.3).

д) проверяется гипотеза о нормальности для результатов наблюдений первого параметра $\{x_i\}_{i=1}^m$.

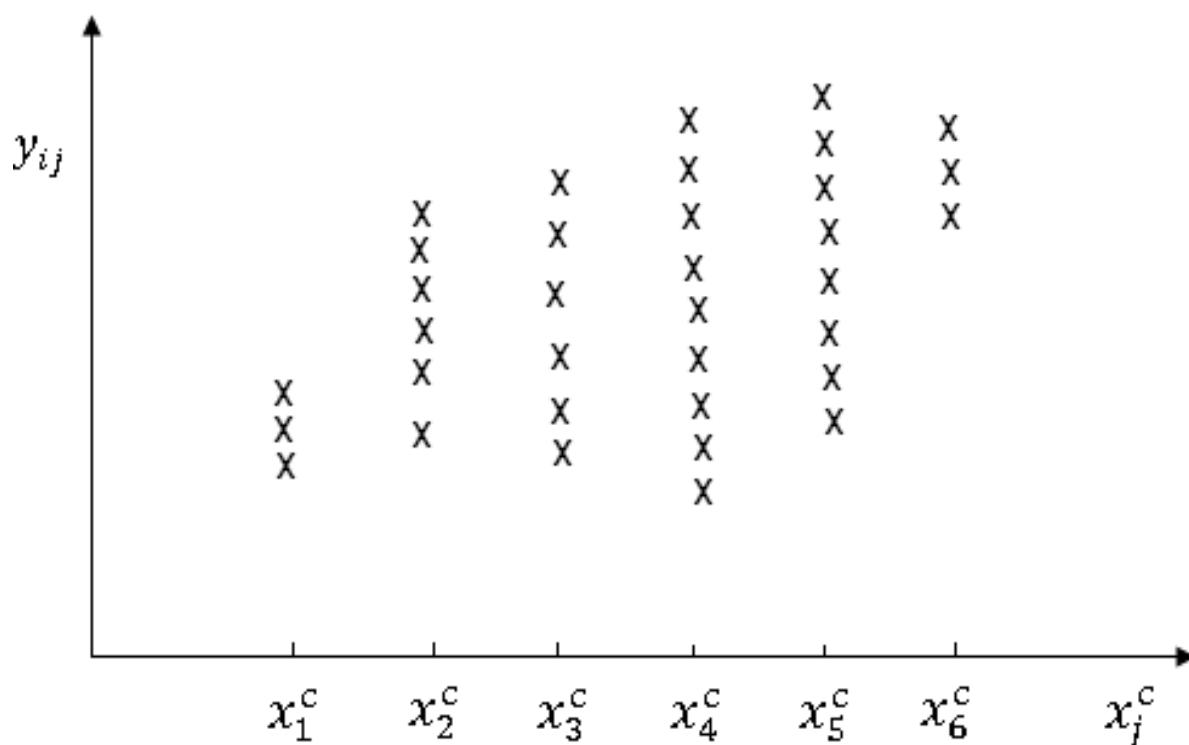


Рис.5

Распределение результатов наблюдений средних точек интервалов разбиения x_j^c относительно Y .

Только при положительном исходе всех проверок принимается гипотеза о нормальности двумерного распределения. Очевидно, что если при какой-то проверке будет получен отрицательный ответ, то следующие этапы проводить не имеет смысла.

Проверка гипотезы $\rho_{xy} = 0$.

Эта проверка проводится только при принятии гипотезы о нормальности двумерного распределения.

Рассчитывается точечная оценка ρ_{XY} - выборочный коэффициент корреляции r_{XY} :

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_X \cdot S_Y}, \quad (109),$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad S_X = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \\ S_Y = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

Далее применяется критерий

$$\begin{cases} \text{если } \frac{|r_{XY}| \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} < t_{n-2, 1-q/2}, \text{ то } \rho_{XY} = 0, \\ \text{если } \frac{|r_{XY}| \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} > t_{n-2, 1-q/2}, \text{ то } \rho_{XY} \neq 0, \end{cases} \quad (110)$$

где q – уровень значимости.

Иногда бывает необходимо построить доверительный интервал для ρ_{XY} . Он имеет вид:

$$[th z_1, th z_2], \quad (111)$$

где

$$z_{1,2} = \frac{1}{2} \ln \frac{1+r_{XY}}{1-r_{XY}} \pm \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{r_{XY}}{2(n-1)}, \quad \varepsilon = 1 - \alpha, \quad (112)$$

а $th z = \frac{t^z - t^{-z}}{t^z + t^{-z}}$ - гиперболический тангенс от аргумента z .

Если по критерию (110) гипотеза $\rho_{XY} = 0$ принимается, то величины X и Y считаются статистически независимыми. В этом случае нет смысла рассчитывать эмпирические модели. Поэтому их определение производится только в том случае, когда гипотеза $\rho_{XY} = 0$ отвергается.

Определение эмпирической модели.

Результаты наблюдений, $\{y_{ij}\}_{i=1}^{n_j}, j = 1, \dots, L$, полученные после группировки, считаются соответствующими $x_j^c, j = 1, \dots, L$, как в регрессионном анализе. Далее по формулам раздела 1.3 рассчитываются оценки коэффициентов линейной модели (86).

2.3. Корреляционный анализ при отсутствии нормальности распределения.

Этот случай соответствует тому, когда гипотеза о нормальности совместного распределения X и Y отвергнута. Как правило, при этом не приходится надеяться на линейность линий регрессии, и, следовательно, целесообразно использовать коэффициент корреляции в качестве характеристики степени связи между X и Y в рамках моделей условных математических ожиданий. В качестве такой характеристики надо использовать корреляционные отношения $\eta_{\frac{Y}{X}}$ и $\eta_{\frac{X}{Y}}$:

$$\eta_{\frac{Y}{X}} = \left(1 - \frac{\sigma_{\frac{Y}{X}}^2}{\sigma_Y^2}\right)^{\frac{1}{2}}, \quad \eta_{\frac{X}{Y}} = \left(1 - \frac{\sigma_{\frac{X}{Y}}^2}{\sigma_X^2}\right)^{\frac{1}{2}}, \quad (113)$$

$$0 \leq \eta_{\frac{Y}{X}} \leq 1, \quad 0 \leq \eta_{\frac{X}{Y}} \leq 1, \quad (114)$$

$$\sigma_{\frac{Y}{X}}^2 = \int_{-\infty}^{\infty} \sigma_{\frac{Y}{X}}^2 f_1(x) dx, \quad \sigma_{\frac{X}{Y}}^2 = \int_{-\infty}^{\infty} \sigma_{\frac{X}{Y}}^2 f_2(y) dy, \quad (115)$$

где $\sigma_{\frac{Y}{X}}^2$ и $\sigma_{\frac{X}{Y}}^2$ - функции условных дисперсий (см. формулы 105), а σ_Y^2 и σ_X^2 - дисперсии случайных величин X и Y . $\eta_{\frac{Y}{X}}$ называется корреляционным отношением Y к X , а $\eta_{\frac{X}{Y}}$ - называется корреляционным отношением X к Y .

Корреляционные отношения обладают следующими свойствами:

1) если $\eta_{\underline{y}}^{\underline{x}} = 0$, то функция $\alpha_{\underline{x}}^{\underline{y}}$ является постоянной, т.е. условное математическое ожидание Y не зависит от X . Если $\eta_{\underline{x}}^{\underline{y}} = 0$ то функция $\alpha_{\underline{y}}^{\underline{x}}$ является постоянной, т.е. условное математическое ожидание X не зависит от Y .

2) если $\eta_{\underline{x}}^{\underline{y}} = 1$, то Y является однозначной функцией X .

Справедливы и обратные утверждения:

3) если $\alpha_{\underline{x}}^{\underline{y}}$ не зависит от x , то $\eta_{\underline{x}}^{\underline{y}} = 0$. Если $\alpha_{\underline{y}}^{\underline{x}}$ не зависит от y , то $\eta_{\underline{y}}^{\underline{x}} = 0$;

4) если Y – однозначная функция X , то $\eta_{\underline{x}}^{\underline{y}} = 1$. Если X – однозначная функция Y , то $\eta_{\underline{y}}^{\underline{x}} = 1$.

Перечисленные свойства объясняют тот факт, почему корреляционные отношения целесообразно использовать в качестве характеристик связи X и Y в рамках условных математических ожиданий.

Отметим, что между $\eta_{\underline{x}}^{\underline{y}}$ и $\eta_{\underline{y}}^{\underline{x}}$ не существует какой-либо простой зависимости. Например, когда $Y = X^2$ и $\eta_{\underline{x}}^{\underline{y}} = 1$, а $\eta_{\underline{y}}^{\underline{x}} = 0$. Подчеркнем, что равенство $\eta_{\underline{y}}^{\underline{x}} = 0$ не противоречит четвертому свойству, так как X не является однозначной функцией Y ($X = \pm\sqrt{Y}$). Ясно, что если $\eta_{\underline{x}}^{\underline{y}} = \eta_{\underline{y}}^{\underline{x}} = 1$, то X является однозначной монотонной функцией Y . Наконец, подчеркнем, что равенства $\eta_{\underline{x}}^{\underline{y}} = \eta_{\underline{y}}^{\underline{x}} = 0$ означают отсутствие связи между X и Y только в рамках условных математических ожиданий.

Теперь рассмотрим процедуру статистического анализа для построения по результатам наблюдений $\{x_i, y_i\}_{i=1}^n$ эмпирических линий регрессии. Для определенности возьмем линию $\alpha_{\underline{x}}^{\underline{y}}$ и ограничимся только изложением алгоритма анализа.

Сначала производят группировку результатов наблюдений и получают группы $\{y_{ij}\}_{i=1}^{m_j}$, $j = 1, \dots, L$, и средние точки интервалов $\{x_j^c\}_{j=1}^L$. Далее, если есть, какая-то надежда на благоприятный исход, то можно проверить гипотезу о нормальности двумерного распределения.

Если такой надежды нет или гипотеза отвергнута, то необходимо проверить предположение о независимости \underline{a}_x от X . Для этого проверяют гипотез о том, что $\underline{\eta}_x = \mathbf{0}$. Процедура проверки следующая. Сначала проверяется гипотеза о равенстве дисперсий выборок $\{y_{ij}\}_{i=1}^{n_j}, j = 1, \dots, L$.

Если она принимается, то рассчитывается выборочное корреляционное отношение $\underline{\eta}_x$ по формуле:

$$\underline{\eta}_x = \left(\frac{\sum_{j=1}^L n_j (\bar{y}_j - \bar{y})^2}{\sum_{j=1}^L \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2} \right)^{\frac{1}{2}}, \quad (116)$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij},$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Если гипотеза о равенстве дисперсий отвергается, то определяются весовые коэффициенты $\{w_j\}_{j=1}^L$ (путем аппроксимации зависимости выборочных дисперсий выборок $\{y_{ij}\}_{i=1}^{n_j}$ от x_j^2), и рассчитывается выборочное корреляционное отношение $\underline{\eta}_x$ по формуле:

$$\underline{\eta}_x = \left(\frac{\sum_{j=1}^L n_j w_j (\bar{y}_j - \bar{y})^2}{\sum_{j=1}^L w_j \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2} \right)^{\frac{1}{2}}, \quad (117)$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij},$$

$$\bar{y} = \frac{\sum_{j=1}^L w_j \bar{y}_j}{\sum_{j=1}^L w_j}.$$

Далее при достаточно большом n применяется критерий:

$$\left\{ \begin{array}{l} \text{если } \frac{\hat{\eta}_Y \sqrt{n-2}}{\frac{\bar{x}}{1-(\hat{\eta}_Y)^2}} < t_{n-2, 1-q}, \text{ то } \eta_{\frac{Y}{X}} = 0, \\ \text{если } \frac{\hat{\eta}_Y \sqrt{n-2}}{\frac{\bar{x}}{1-(\hat{\eta}_Y)^2}} > t_{n-2, 1-q}, \text{ то } \eta_{\frac{Y}{X}} \neq 0, \end{array} \right. \quad (118)$$

где q – уровень значимости.

Если по критерию (118) гипотеза $\eta_{\frac{Y}{X}} = 0$ отвергается, то можно строить эмпирическую линию регрессии. Для этого проверяется гипотеза о нормальности для каждой выборки $\{y_{ij}\}_{i=1}^{m_j}, j = 1, \dots, L$, и в случае положительного исхода всех проверок результаты наблюдений $\{y_{ij}\}_{i=1}^{m_j}, j = 1, \dots, L$, при $\{x_j^2\}_{j=1}^L$ описываются подходящей моделью регрессионного анализа.

Заключение

В данном учебном пособии изложена сущность, задачи и методика проведения регрессионного и корреляционного анализа при статистической обработке результатов экспериментальных измерений в оптоэлектронике. Описана методика проверки однородности ряда дисперсий. Рассмотрены линейная и нелинейная модели с одной независимой переменной в случае равенства и неравенства дисперсий, а также методика построения точечных и интервальных оценок коэффициентов регрессии и дисперсии. Изложена методика ортогонализации системы функций на множестве значений аргумента. Описана методика проведения корреляционного анализа для случаев нормального двумерного распределения и его отсутствия.

Ниже приводятся методические указания и даются исходные данные для выполнения расчетной работы по теме «Проведение регрессионного анализа по результатам измерений распределения показателя преломления в сечении графана».

3. Задание на расчетную работу.

1. Получить у преподавателя номер варианта с экспериментальными данными по измерению распределения показателя преломления в сечении графана, по которым необходимо провести регрессионный анализ (см. стр. 45, 46).
2. Проверить гипотезу об однородности дисперсий выборок по формулам (8)-(14) (q принять равным 0.02).
3. В качестве модели, описывающей экспериментальные данные выбрать полином 2-й степени.
4. Провести ортогонализацию системы функций на заданном наборе значений аргумента по методике, изложенной в Приложении 1. (см. формулы (128)-(135)).
5. Построить точечные оценки коэффициентов регрессии a_v ($v=0,1,\dots,m$) и дисперсии δ^2 по формулам (22) и (23) соответственно.
6. Построить интервальные оценки коэффициентов регрессии a_v ($v=0,1,\dots,m$) и дисперсии δ^2 при заданной доверительной вероятности $\alpha = 0.9$ по формулам (24)-(26).
7. Найти точечную и интервальную оценки для кривой регрессии при значении аргумента X_4 (см. варианты заданий стр.47, 48) по формулам (27)-(29).
8. Результаты расчетов представить в виде пояснительной записки к расчетной работе.

Ортогонализация системы функций на множестве значений аргумента.

Система функций $\{\varphi_{\nu}(X)\}_{\nu=1}^m$ называется ортогональной на множестве значений аргумента $\{x_j\}_{j=1}^m$ с весами $\{n_j\}_{j=1}^m$, если

$$\sum_{j=1}^L n_j \varphi_{\nu_1}(x_j) \varphi_{\nu_2}(x_j) = (\varphi_{\nu_1}, \varphi_{\nu_2}) = 0 \quad (119)$$

при $\nu_1 \neq \nu_2$.

Рассмотрим процедуру ортогонализации для произвольной системы функций $\varphi_0, \varphi_1, \dots, \varphi_m$. Необходимо построить эквивалентную ей систему функций, $\psi_0, \psi_1, \dots, \psi_m$, ортогональную на множестве значений аргумента. Процедура ортогонализации заключается в последовательном вычислении функций $\psi_0, \psi_1, \dots, \psi_m$ на основании выполнения требования в соответствии с выражением (119), чтобы каждая вновь получаемая функция была бы ортогональна всем предыдущим. Перечислим этапы этой процедуры:

ψ_0 берется равной φ_0

$$\psi_0 = \varphi_0, \quad (120)$$

ψ_1 ищется в виде

$$\psi_1 = b\psi_0 + \varphi_1, \quad (121)$$

где b определяется из условия

$$(\psi_1, \psi_0) = 0 \quad (122)$$

ψ_2 ищется в виде

$$\psi_2 = c\psi_0 + d\psi_1 + \varphi_2, \quad (123)$$

где c и d рассчитываются из уравнений

$$(\psi_2, \psi_0)=0 ; (\psi_2, \psi_1)=0 \quad (124)$$

В общем виде ψ_q разыскивается в виде:

$$\psi_q = \sum_{v=0}^{q-1} l_v \psi_v + \varphi_q, \quad (125)$$

где l_v находится из соотношений:

$$(\psi_q, \psi_0) = 0, (\psi_q, \psi_1) = 0, \dots, (\psi_q, \psi_{q-1}) = 0 \quad (126)$$

по формуле:

$$l_q = - \frac{(\varphi_q, \psi_v)}{(\psi_v, \psi_v)} \quad (127)$$

Поясним на конкретном примере, как любую модель вида (15) с неортогональными функциями можно преобразовать в эквивалентную ей модель вида (15) с ортогональными функциями.

Предположим, что в качестве модели используется полином второй степени, а в качестве множества значений независимой переменной совокупность

$$x_1=0, x_2=1, x_3=2 \quad (L=3 \geq m+1=3) \quad (128)$$

Будем считать, что в каждом экспериментальном условии проводится одно и то же число опытов $n_1=n_2=n_3=n$. Тогда условие ортогональности (119) описывается соотношениями:

$$(\varphi_{v_1}, \varphi_{v_2}) = n \sum_{j=1}^3 \varphi_{v_1}(x_j) \varphi_{v_2}(x_j) = 0 \quad (129)$$

при $v_1 \neq v_2$

Выбранную модель в виде полинома второй степени можно представить в виде:

$$Q = a_0 + a_1 x + a_2 x^2. \quad (130)$$

$$\text{т.е. } \varphi_0(x)=1, \varphi_1(x)=x, \varphi_2(x)=x^2 \quad (131)$$

Система (131) не является ортогональной на множестве (10), т.к. $(\varphi_0, \varphi_1)=3n$, $(\varphi_0, \varphi_2)=5n$, $(\varphi_1, \varphi_2)=9n$. Построим систему $\{\Psi_v\}_{v=0}^2$, эквивалентную системе $\{\varphi_v\}_{v=0}^2$, но обладающую свойством ортогональности. В соответствии с (83),(84) функция ψ_0 полагается равной φ_0 , $\psi_0=\varphi_0=1$, а ψ_1 ищется в виде $\psi_1=b\psi_0+\varphi_1$, где коэффициент b находится из требования ортогональности ψ_0 и ψ_1

$$(\psi_0, \psi_1)=0 \quad (132)$$

$$\text{Отсюда } b = -\frac{(\varphi_1, \psi_0)}{(\psi_0, \psi_0)} = -\frac{3n}{3n} = -1$$

$$\psi_1 = -1 + x$$

В соответствии с (123) и (124) ψ_2 разыскивается в виде:

$$\psi_2 = c\psi_0 + d\psi_1 + \varphi_2, \quad (133)$$

где коэффициенты c и d определяются из условия ортогональности ψ_0 , ψ_1 и ψ_2 :

$$(\psi_2, \psi_0)=0; (\psi_2, \psi_1)=0. \quad (134)$$

Из (133) и (134) получаем:

$$c = -\frac{(\varphi_2, \psi_0)}{(\psi_0, \psi_0)} = -\frac{5}{3}, \quad d = -\frac{(\varphi_2, \psi_1)}{(\psi_1, \psi_1)} = -\frac{4n}{2n} = -2$$

$$\psi_2 = -\frac{5}{3} - 2(-1 + x) + x^2 = \frac{1}{3} - 2x + x^2$$

Требуемая система $\{\Psi_v\}_{v=0}^2$ построена. Для этой системы модель в виде полинома второй степени отличается от (130):

$$Q = a'_0 + a'_1(x - 1) + a'_2(x^2 - 2x + \frac{1}{3}) \quad (135)$$

Но модель (135) эквивалентна модели (130), т.к. переходит в нее при

$$a'_2 = a_2, \quad a'_1 = a_1 + 2a_2, \quad a'_0 = a_0 + a_1 + \frac{5}{3}a_2.$$

Таким образом, модель в виде полинома второй степени можно записать в двух вариантах (130) и (135). Однако, второй вариант имеет большее преимущество, т.к. для него значительно упрощается процедура регрессионного анализа.

**Выбор функции $f(x)$ для проведения регрессионного анализа
в случае неравенства дисперсий ($\sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_L^2$)**

В случае, если гипотеза об однородности дисперсий отклоняется по критерию (5) или (8), для проведения регрессионного анализа необходимо определить аналитическую зависимость $f(x)$, входящую в уравнение (53). Для этого строят график зависимости значений выборочных средних S_j^2 от аргумента x_j . Затем выбирают вид функции $f(x)$ в виде линейной или нелинейной модели. Рассмотрим определение функции $f(x)$ в случае аппроксимации ее линейной зависимостью (полиномом первой степени) и параболической зависимостью (полиномом второй степени).

Линейная зависимость

Будем искать функцию $f(x)$ в виде

$$f(x) = b_0 + b_1 x \quad (136)$$

Найдем приближенные значения коэффициентов b_0 и b_1 , которые обозначим через b'_0 и b'_1 , из условия аппроксимации по методу наименьших квадратов зависимости S_j^2 от x_j , т.е. из условия минимума величины

$$M = \sum_{j=1}^L (S_j^2 - b'_0 - b'_1 x_j)^2 \quad (137)$$

Коэффициенты b'_0 и b'_1 в этом случае можно определить по формулам:

$$b'_1 = \frac{\sum_{j=1}^L (x_j - \frac{1}{L} \sum_{j=1}^L x_j) S_j^2}{\sum_{j=1}^L (x_j - \frac{1}{L} \sum_{j=1}^L x_j)^2}, \quad (138)$$

$$b'_0 = \frac{1}{L} \sum_{j=1}^L S_j^2 - b'_1 \frac{1}{L} \sum_{j=1}^L x_j . \quad (139)$$

Таким образом, в качестве функции $f(x)$ можно взять зависимость вида:

$$f(x) = b'_0 + b'_1 x \quad (140)$$

Параболическая зависимость

Будем искать функцию $f(x)$ в виде:

$$f(x) = b_0 + b_1 x + b_2 x^2 \quad (141)$$

Найдем приближенные значения коэффициентов b_0, b_1, b_2 (обозначим их через b'_0, b'_1, b'_2) из условия минимума величины

$$M = \sum_{j=1}^L (S_j^2 - b'_0 - b'_1 x_j - b'_2 x_j^2)^2 \quad (142)$$

Для упрощения вычислений целесообразно переписать зависимость (141) в виде:

$$f(x) = b_0 + b_1 (x - \bar{x}) + b_2 (x - \bar{x})^2 , \quad (143)$$

$$\bar{x} = \frac{1}{L} \sum_{j=1}^L x_j . \quad (144)$$

Коэффициенты b'_0, b'_1, b'_2 можно определить по формулам:

$$b'_2 = \frac{1}{c} \left[\sum_{j=1}^L S_j^2 (x_j - \bar{x})^2 - \frac{1}{L} \left(\sum_{j=1}^L S_j^2 \right) \sum_{j=1}^L (x_j - \bar{x})^2 \right] , \quad (145)$$

$$c = \sum_{j=1}^L (x_j - \bar{x})^4 - \frac{1}{L} \left[\sum_{j=1}^L (x_j - \bar{x})^2 \right]^2 , \quad (146)$$

$$b_1' = \frac{\sum_{j=1}^L S_j^2 (x_j - \bar{x})}{\sum_{j=1}^L (x_j - \bar{x})^2} , \quad (147)$$

$$b_0' = \frac{1}{L} \sum_{j=1}^L S_j^2 - b_2' \frac{1}{L} \sum_{j=1}^L (x_j - \bar{x})^2 . \quad (148)$$

Таким образом, искомая функция $f(x)$ определяется в виде (143).

Таблица 1

Квантили $f_{k_1, k_2; 0.95}$ плотности распределения Фишера.

K ₂	Число степеней свободы K ₁											
	1	2	3	4	5	6	8	12	16	24	50	∞
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.43	19.45	19.47	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.69	8.64	8.58	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.84	5.77	5.70	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.60	4.53	4.44	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.92	3.84	3.75	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.49	3.41	3.32	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.20	3.12	3.03	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.98	2.90	2.80	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.82	2.74	2.64	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.70	2.61	2.50	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.60	2.50	2.40	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.51	2.42	2.32	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.44	2.35	2.24	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.39	2.29	2.18	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.33	2.24	2.13	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.29	2.19	2.08	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.25	2.15	2.04	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.21	2.11	2.00	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.29	2.18	2.08	1.96	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.15	2.05	1.93	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.13	2.03	1.91	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.11	2.00	1.88	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	2.09	1.98	1.86	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	2.07	1.96	1.84	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	2.05	1.95	1.82	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	2.03	1.93	1.80	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	2.02	1.91	1.78	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	2.00	1.90	1.77	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.99	1.88	1.76	1.62
35	4.12	3.26	2.87	2.64	2.48	2.37	2.22	2.04	1.94	1.83	1.70	1.57

Таблица 2

Квантили $f_{k_1, k_2; 0.99}$ плотности распределения Фишера.

К ₂	Число степеней свободы К ₁											
	1	2	3	4	5	6	8	12	16	24	50	∞
2	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.42	99.44	99.46	99.48	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.83	26.60	26.35	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	14.15	13.93	13.69	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	9.88	9.68	9.47	9.24	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.52	7.31	7.09	6.88
7	12.35	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.27	6.07	5.85	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.48	5.28	5.06	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.92	4.73	4.51	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.52	4.33	4.12	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.21	4.02	3.80	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.98	3.78	3.56	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.78	3.59	3.37	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.62	3.43	3.21	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.48	3.29	3.01	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.37	3.18	2.96	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.27	3.08	2.86	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.20	3.00	2.79	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	3.12	2.92	2.70	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	3.05	2.86	2.63	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.99	2.80	2.58	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.94	2.75	2.53	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.89	2.70	2.48	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.85	2.66	2.44	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.81	2.62	2.40	2.17
26	7.62	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.78	2.58	2.36	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.74	2.55	2.33	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.71	2.52	2.30	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.68	2.49	2.27	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.66	2.47	2.24	2.01
35	7.42	5.27	4.40	3.91	3.59	3.37	3.07	2.74	2.56	2.37	2.13	1.90
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.48	2.29	2.05	1.80
45	7.23	5.11	4.25	3.77	3.45	3.23	2.94	2.61	2.43	2.23	1.99	1.75
50	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.56	2.38	2.18	1.94	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.32	2.12	1.87	1.60
70	7.01	4.92	4.07	3.60	3.29	3.07	2.78	2.45	2.28	2.07	1.82	1.53
80	6.96	4.88	4.04	3.53	3.23	3.01	2.72	2.39	2.21	2.00	1.75	1.45
100	6.90	4.82	3.96	3.51	3.21	2.99	2.69	2.37	2.19	1.98	1.73	1.43
125	6.84	4.78	3.94	3.47	3.17	2.95	2.66	2.33	2.15	1.94	1.69	1.37
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.61	2.18	2.13	1.79	1.52	1.00

Таблица 3

Квантили $X^2_{k,\gamma}$ плотности X^2 -распределения.

K_2	вероятность γ				вероятность γ			
	0.01	0.02	0.05	0.1	0.9	0.95	0.98	0.99
1	0.000	0.001	0.004	0.016	2.706	3.841	5.412	6.635
2	0.020	0.040	0.103	0.211	4.605	5.991	7.824	9.210
3	0.115	0.185	1.352	0.584	6.251	7.815	9.837	11.345
4	0.297	0.429	0.711	0.064	7.779	9.488	11.668	13.277
5	0.554	0.752	1.145	1.610	9.236	11.070	13.388	15.086
6	0.872	1.134	1.635	2.204	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	23.296	26.296	29.633	32.000
17	6.408	7.225	8.672	10.085	24.769	27.587	30.992	33.409
18	7.015	7.906	9.390	10.865	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	28.412	31.410	35.020	37.556
21	8.897	9.915	11.591	13.240	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	32.172	38.968	41.638	41.638
24	10.856	11.992	13.848	15.659	33.196	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	35.563	42.886	42.856	45.642
27	12.879	14.125	16.151	18.114	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	40.256	43.773	47.962	50.982

Таблица 4

Квантили $t_{k,\gamma}$ плотности распределения Стьюдента.

k	γ		
	0.95	0.975	0.995
1	6.31	12.71	63.66
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.90	2.36	3.50
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17
12	1.78	2.18	3.06
14	1.76	2.14	2.98
16	1.75	2.12	2.92
18	1.73	2.10	2.88
20	1.72	2.09	2.84
22	1.72	2.07	2.82
24	1.71	2.06	2.80
26	1.71	2.06	2.78
28	1.70	2.05	2.76
30	1.70	2.04	2.75
∞	1.64	1.96	2.58

Варианты заданий

Вариант №	X ₁ ,М М	n ₁₁ ÷n ₁₄	X ₂ ,ММ	n ₂₁ ÷n ₂₄	X ₃ ,ММ	n ₃₁ ÷n ₃₄	X ₄ ,ММ
1	2	1.5105 1.5107 1.5108 1.5107	8	1.5198 1.5195 1.5193 1.5196	16	1.5115 1.5113 1.5116 1.5117	18
2	4	1.5119 1.5115 1.5121 1.5117	10	1.5199 1.5202 1.5197 1.5200	16	1.5116 1.5114 1.5117 1.5118	18
3	2	1.5104 1.5107 1.5106 1.5108	10	1.5198 1.5201 1.5197 1.5200	14	1.5158 1.5155 1.5160 1.5156	18
4	4	1.5120 1.5119 1.5116 1.5120	12	1.5162 1.5160 1.5158 1.5162	18	1.5108 1.5110 1.5107 1.5111	16
5	2	1.5109 1.5110 1.5107 1.5105	12	1.5163 1.5165 1.5159 1.5162	18	1.5109 1.5111 1.5108 1.5112	16
6	6	1.5160 1.5162 1.5157 1.5159	12	1.5164 1.5165 1.5160 1.5162	18	1.5109 1.5110 1.5107 1.5113	16
7	6	1.5158 1.5162 1.5160 1.5159	10	1.5197 1.5202 1.5199 1.5200	18	1.5108 1.5111 1.5107 1.5113	16
8	2	1.5107 1.5103 1.5109 1.5106	12	1.5160 1.5155 1.5157 1.5163	16	1.5115 1.5112 1.5117 1.5114	18
10	4	1.5123 1.5117 1.5118 1.5120	8	1.5198 1.5200 1.5195 1.5197	18	1.5111 1.5107 1.5105 1.5108	16

11	2	1.5105 1.5109 1.5111 1.5104	8	1.5199 1.5196 1.5202 1.5200	18	1.5110 1.5106 1.5112 1.5109	16
12	2	1.5104 1.5110 1.5107 1.5106	10	1.5190 1.5192 1.5196 1.5198	18	1.5108 1.5112 1.5110 1.5106	16
13	3	1.5193 1.5190 1.5194 1.5197	9	1.5217 1.5215 1.5220 1.5216	18	1.5200 1.5205 1.5203 1.5199	15
14	3	1.5192 1.5191 1.5193 1.5198	9	1.5216 1.5219 1.5215 1.5218	15	1.5208 1.5210 1.5205 1.5208	18
15	6	1.5207 1.5210 1.5206 1.5209	12	1.5210 1.5216 1.5212 1.5215	18	1.5202 1.5200 1.5204 1.5205	15
16	3	1.5190 1.5189 1.5195 1.5193	6	1.5207 1.5208 1.5211 1.5205	15	1.5210 1.5206 1.5207 1.5209	18
17	6	1.5204 1.5210 1.5206 1.5209	12	1.5214 1.5210 1.5218 1.5216	18	1.5199 1.5203 1.5207 1.5200	15
18	3	1.5185 1.5188 1.5182 1.5183	9	1.5209 1.5210 1.5212 1.5208	18	1.5201 1.5208 1.5205 1.5206	15
19	3	1.5185 1.5184 1.5187 1.5185	12	1.5219 1.5217 1.5222 1.5221	18	1.5206 1.5204 1.5209 1.5207	15
20	6	1.5213 1.5215 1.5210 1.5212	12	1.5221 1.5223 1.5217 1.5221	18	1.5205 1.5203 1.5208 1.5207	15

Примечание

1) X_1, X_2, X_3 - значения аргумента, соответствующие экспериментальным условиям X_1, \dots, X_L $L=3$ (см. рис.1).

2) $n_{11} \div n_{14}, n_{21} \div n_{24}, n_{31} \div n_{34}$ - измеренные значения показателей преломления при аргументах X_1, X_2, X_3 (соответствуют результатам $L=3$ групп наблюдений

объемом $n_j=4$ каждая $\{y_{ij}\}_{i=1}^{n_j}, j=1, \dots, L$ см. стр.3, 4)

Литература

1. Сергеев А.Г. Метрология. – М.; Логос, 2005. – 272с.
2. Зайцев С.А., Грибанов Д.Д., Куранов А.Д. и др. Метрология. – М.: ФОРУМ, ИНФРА – М, 2011. – 464с.
3. Сергеев А.Г., Терегея В.В. Метрология, стандартизация и сертификация. – М.: Юрайт-издат, 2011. – 820с.
4. Колчков В.И. Метрология, стандартизация и сертификация. Учебник. – М.: Владос. – 2010. – 401с.
5. Герасимова Е.Б., Герасимов Б.И. Метрология, стандартизация и сертификация. – М.: ФОРУМ, ИНФРА – М, 2010. – 224с.
6. Шишкин И.Ф. Теоретическая метрология. Учебник для вузов. – СПб.: Питер, 2009 – 192с.
7. Дегтярев А.А. и др. Метрология. Учебное пособие для вузов.- М.: Академ. Проект, 2006. – 254с.
8. Димов Ю.В. Метрология, стандартизация и сертификация. Доп. М-вом образов. РФ в кач. Учебника для студентов вузов. – СПб.: Питер, 2006. – 432с.
9. Лисиц И.М. Стандартизация, метрология и сертификация. Учебник. – М.: Юрайт-издат, 2007. – 399с.
10. Кошечкина И.П. Метрология, стандартизация, сертификация. – М.: ФОРУМ, ИНФРА – М, 2007. – 414с.
11. Ким К.К. Метрология, стандартизация и сертификация. – СПб.: Питер, 2006. – 344с.
12. Парфенов В.Г. Регрессионный и корреляционный анализ. Обработка результатов наблюдений при измерениях. – Л., ЛИТМО, 1983. – 78 с.
13. Гордов А.Н., Парфенов В.Г., Потягайло А.Ю., Шарков А.В. Статистические методы обработки результатов теплофизического эксперимента. – Л., ЛИТМО, 1981. – 72 с.



В 2009 году Университет стал победителем многоэтапного конкурса, в результате которого определены 12 ведущих университетов России, которым присвоена категория «Национальный исследовательский университет». Министерством образования и науки Российской Федерации была утверждена Программа развития государственного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет информационных технологий, механики и оптики» на 2009–2018 годы.

КАФЕДРА ТВЕРДОТЕЛЬНОЙ ОПТОЭЛЕКТРОНИКИ

Кафедра была организована в 1983 году в период выделения оптоэлектроники в самостоятельную область науки и производства. Кафедра укомплектована высококвалифицированными специалистами, имеет оснащенные учебные и научные лаборатории. Выпускники кафедры получают усиленную подготовку по физической оптике, физике твердого тела, оптическим методам передачи и обработки информации, оптико-физическим измерениям, лазерной технике, волоконной и интегральной оптике, электронике, компьютерной технике и технологиям, оптико-электронным средствам массовой информации, живописи, экологии, иностранному языку, экономике, менеджменту и другим дисциплинам.

Широкая подготовка по ряду дисциплин в области оптоэлектроники дает возможность выпускникам легко ориентироваться на рынке труда. Выпускники кафедры работают в исследовательских институтах и лабораториях, производстве, бизнесе. Многие из них занимают ведущие административные должности.

Область научных интересов сотрудников кафедры лежит в направлении оптического материаловедения, спектроскопии твердого тела, информатики, оптико-физических измерений и их автоматизации, взаимодействия оптического излучения с различными веществами.

Юрий Тихонович Нагибин

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В ОПТОЭЛЕКТРОНИКЕ.
РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ АНАЛИЗ.

Учебное пособие

В авторской редакции

Дизайн

Верстка

А.А. Горбенко

М.А. Смуров,

К.А. Шишова

Редакционно-издательский отдел Санкт-Петербургского
государственного университета информационных технологий,
механики и оптики

Зав. РИО

Н.Ф. Гусарова

Лицензия ИД № 00408 от 05.11.99

Подписано к печати

2011г.

Заказ №

Тираж 100 экз.

Отпечатано на ризографе

Редакционно-издательский отдел
Санкт-Петербургского государственного уни-
верситета информационных технологий, механики и
197101, Санкт-Петербург, Кронверкский пр., 49



верси-
оптики