**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И КУЛЬТУРЫ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ**

**КЫРГЫЗСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ им. И.РАЗЗАКОВА**

КАФЕДРА «АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ»

СТАТИСТИЧЕСКИЕ МЕТОДЫ

**Методическое руководство к выполнению лабораторных
и практических занятий для студентов направления
550201.01 «Управление и информатика в технических
системах», и других родственных направлений
всех форм обучения**

БИШКЕК – 2011

«РАССМОТРЕНО»
на заседании кафедры
«Автоматическое управление»
Протокол № 4 от 04.01.2011г.

«УТВЕРЖДЕНО»
учебно-методической комиссией
ФИТ
Протокол №5 от 20.01.2011г.

Составители: к.т.н., доц. МИХЕЕВА Н.И., ст. преп. АЙДАРАЛИЕВА Н.Ш.

УДК 519.2(075.8).681.3

Статистические методы. Методическое руководство к выполнению лабораторных и практических занятий для студентов направления 550201.01 «Управление и информатика в технических системах», и других родственных направлений всех форм обучения./ КГТУ им. И. Раззакова, сост.: Н.И.Михеева, Н.Ш.Айдаралиева. – Б.: «Текник», 2011. – 47 с.

Приведены задания к лабораторным и практическим работам и указания к их выполнению.

Предназначены для студентов направления 550201.01 «Управление и информатика в технических системах» всех форм обучения.

Ил. 1. Табл. 12. Библиогр.: 19 назв.

Рецензент к.т.н., доц. Молдобеков К.М.

Статистические методы
Методическое руководство к выполнению лабораторных и практических занятий
для студентов направления 550201.01 «Управление и информатика в технических системах»,
и других родственных направлений всех форм обучения

Составители: ***Михеева Н.И., Айдаралиева Н.Ш.***

Тех. редактор ***Субанбердиева Н.Е.***

Подписано к печати 14.03.2011 г. Формат бумаги 60х84 $^{1}/_{16}$.
Бумага офс. Печать офс. Объем 3 п.л. Тираж 50 экз. Заказ 87. Цена 53 сом.

Бишкек, ул. Сухомлинова, 20. ИЦ "Текник" КГТУ им. И.Раззакова, т.: 54-29-43
e-mail: beknur@mail.ru

Содержание

Введение…………………………………………………………….………… **4**

Лабораторная работа 1. МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН, РАСПРЕДЕЛЕННЫХ ПО НОРМАЛЬНОМУ ЗАКОНУ………... 6

Лабораторная работа 2. СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ….…………………………….. 10

Лабораторная работа 3. ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ДВУМЯ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ…...………………………….. 20

Лабораторная работа 4. ОДНОМЕРЕНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ …………………………………………………………..……… 26

Лабораторная работа 5. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ ……………………………………………………………... 32

Приложение 1. Критические точки распределения *Стьютенда*………... 40

Приложение 2. Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ ……….. 41

Приложение 3. Критические точки распределения *Кочрена*.................... 43

Приложение 4. Критические точки распределения **F** *Фишера* – Снедекора……………………………………………………….. 44

Приложение 5. Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} l^{\frac{-x^2}{2}}$ ………… 46

Библиографический список……………………………………………….. 47

Введение

В современной инженерной практике создания систем автоматического и автоматизированного управления статистические методы занимают весьма важное место. Так, идентификация объектов и систем, играющая особую роль при создании и использовании автоматических систем управления, основывается на методах планирования эксперимента и обработки его результатов. Это позволяет реализовать системный подход в изучении сложных явлений и получить надежные оценки при минимальных затратах времени и средств. Учет стохастической природы внешних воздействий на систему автоматического управления и случайного разброса ее параметров дает возможность оценивать свойства реальных систем и оптимизировать их в процессе проектирования. Соответственно, дисциплина «Статистические методы» входит в число основных профилирующих предметов при подготовке специалистов направления 550201.01 «Управление и информатика в технических системах». При этом важная роль отводится лабораторным работам, выполнение которых, позволят глубже освоить и закрепить положения, излагаемые в лекционном курсе.

Реализация лабораторных работ на основе результатов натурного эксперимента весьма затруднительна. Поэтому в данном пособии предлагается осваивать статистические процедуры с использованием массивов случайных чисел, смоделированных на ЭВМ по заданным законам распределения. Это оправдано с методологической точки зрения, так как облегчает контроль правильности осуществления студентами статистической обработки полученного массива. Таким образом, основным техническим оснащением лабораторных работ является персональный компьютер. В качестве программного средства выбран табличный процессор Microsoft Excel. Это обусловлено простотой освоения и использования данного средства, гибкостью. К достоинствам Microsoft Excel можно отнести необходимость активного творческого участия студентов в реализации алгоритмов выполнения лабораторных заданий. Следует отметить, что при выполнении лабораторных работ могут быть использованы другие программные средства.

Методическое пособие содержит 5 лабораторных работ. Первая работа посвящена ознакомлению с табличным процессором Microsoft Excel и моделированию массивов случайных чисел с требуемым законом распределения. Работы 2-5 выполняются с использованием результатов 1- работы.

Описание каждой работы содержит задание, краткие теоретические сведения, методические указания к выполнению задания, контрольные вопросы и задачи, что позволяет варьировать объем заданий и индивидуализировать процесс обучения.

Общие указания к выполнению лабораторных работ:

1. В каждой работе предлагается несколько вариантов задания. Выберите свой вариант, номер которого совпадает с номером по списку студентов в журнале преподавателя.

   Примечание: Задание может быть подготовлено преподавателем для студента индивидуально и отличаться от типового.
2. Тщательно изучите теоретические основы задания, используя указанную литературу и настоящее руководство.
3. Составьте план проведения работы: обоснуйте и изложите основные расчетные соотношения и исходные данные с указанием источников.
4. Согласуйте план работы с преподавателем и выполняйте ее.
5. Оформите отчет о проделанной работе в следующем составе: цель работы, задание, теоретические сведения, результаты в числовой и графической форме, выводы.
6. Защитите результаты проделанной работы.

Лабораторная работа 1

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН, РАСПРЕДЕЛЕННЫХ ПО НОРМАЛЬНОМУ ЗАКОНУ

Цель работы

Ознакомление с методами математического моделирования на ЭВМ случайных величин, распределенных по нормальному закону.

Задание к работе

Получить массивы случайных величин Y и Z, распределенных по нормальному закону с различными параметрами. Параметры и размеры массивов m по вариантам указаны в таблице 1.1.

Таблица 1.1

№ Вариантов	Параметры нормального закона распределения				Размер
	Y		Z		
	M	σ	C_1	C_2	m
1	15	3	6	3	45
2	25	5	5	2	50
3	6	1	7	3	55
4	50	10	10	2	60
5	90	18	9	4	45
6	40	8	10	5	50
7	20	4	6	2	55
8	12	2	5	1	60
9	45	9	10	4	45
10	60	11	9	3	50
11	90	9	8	2	55
12	80	15	7	2	60
13	120	20	8	3	45
14	50	7	7	4	50
15	15	2	4	1	55
16	40	6	10	3	60
17	90	16	11	4	45
18	20	3	8	4	50
19	30	5	9	2	55
20	150	30	11	3	60

Примечание. Смысл параметров M, σ, C_1, C_2 будет пояснен ниже.

Краткие теоретические сведения

Случайная величина.

Случайной называют величину, которая в результате испытаний принимает значение, наперед неизвестное и зависящее от случайных причин, которые заранее не могут быть учтены.

Дискретной (прерывной) называют случайную величину (СВ), которая принимает отдельные, изолированные возможные значения с определенными вероятностями . Число возможных значений дискретной СВ может быть конечным или бесконечным.

Непрерывной называют СВ, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Число возможных значений непрерывной СВ бесконечно.

Законом распределения СВ называется соотношение между значениями СВ и их вероятностными характеристиками. Закон распределения непрерывной СВ может быть задан в графической и аналитической формах, дискретной СВ - в графической, аналитической и табличной формах.

Если рассматривать появление конкретного значения СВ как событие, то вероятностью события называется отношение числа исходов, благоприятствующих этому событию, к общему числу исходов.

Относительной частотой события называют отношение числа испытаний, в которых событие появилось, к общему числу фактически произведенных испытаний. Таким образом, вероятность вычисляют до опыта, относительную частоту - после опыта. Значение частоты стремится к вероятности при количестве опытов, стремящемся к бесконечности.

Моделирование случайной величины.

Задача получения случайных величин, распределенных по заданному закону, решается либо с помощью специальных физических датчиков случайных чисел, либо с помощью специальных прикладных программ. Применение датчиков случайных чисел связано с серьезными техническими трудностями. Наибольшее распространение в инженерной практике получили специальные подпрограммы, в основу которых могут быть положены различные принципы. Один из наиболее простых, основан на формировании случайной величины по закону равной вероятности и последующем преобразовании этого закона. Многие программные средства содержат генератор случайных чисел. Так, функция СЛЧИС в MS Excel формирует равномерную случайную величину ξ, в диапазоне [*0, -1*].

Примечание. Значения ξ рассчитываются по определенному алгоритму, поэтому ξ является псевдослучайной величиной.

Чтобы получить нормальную величину Y с параметрами M (математическое ожидание) и σ (среднее квадратическое отклонение), необходимо использовать равномерные случайные величины x_i в диапазоне [a, $-b$], причем a и b можно найти из соотношений:

$$M = \frac{a+b}{2}; \quad (1.1)$$

$$\sigma^2 = \frac{(b-a)^2}{12}. \quad (1.2)$$

Равномерная случайная величина X в диапазоне [*a, -b*] может быть получена на основе ξ (СЛЧИС), при этом значения x_i могут быть рассчитаны по формуле:

$$x_i = a + (b-a)\xi \quad . \quad (1.3)$$

Нормальная случайная величина *Y* с математическим ожиданием *M* и средним квадратическим отклонением σ может быть получена как композиция *n* равномерных распределений *X*, причем *n*=$12 \div 15$.

Примечание. В данном случае под композицией понимается СВ *Y*, каждое значение которой получается по следующему правилу. Пусть имеется m реализаций СВ *X* одинаковой размерности *n*:

$$\begin{matrix} X_1 & x_{11} & x_{12}\ldots & x_{1j}\ldots & x_{1n} \\ X_2 & x_{21} & x_{22}\ldots & x_{2i}\ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_i & x_{i1} & x_{i2}\ldots & x_{ij}\ldots & x_{in} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_m & x_{m1} & x_{m2}\ldots & x_{mj}\ldots & x_{mn} \end{matrix}.$$

Тогда *m* значений СВ *Y*

$$Y \; y_1 \quad y_2 \ldots . \; y_j \ldots y_m$$

определяются соотношением:

$$y_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \quad . \quad (1.4)$$

Полученная совокупность значений *Y* является реализацией одномерной случайной величины.

Двумерная нормальная случайная величина - совокупность двух нормальных величин, между которыми может существовать взаимосвязь. В качестве одной компоненты можно использовать одномерную нормальную величину *Y*, вторая компонента *Z* может быть получена путем линейного преобразования *Y* по следующему правилу:

$$Z = C_1\vec{Y} + C_2 Y, \quad (1.5)$$

где C_1, C_2 - постоянные; $\vec{Y}$ упорядоченные по возрастанию значения СВ *Y*.

Соотношение между величинами C_1 и C_2 определяет степень взаимосвязи между *Y* и *Z*. При этом, чем больше C_2, тем слабее взаимосвязь.

Очевидно, СВ *Z*, являющаяся линейной комбинацией *Y*, распределена по нормальному закону. Таким образом, двумерная совокупность (*Y*, *Z*) характеризует нормальный закон распределения на плоскости.

Методические указания к выполнению работы

1. Вычислите параметры a и b по заданным значения M и σ из табл. 1.1, используя формулы (1.1), (1.2).
2. Получите двумерный массив X_{mn}:

- используя формулу (1.3), получите случайную величину x_{11} в ячейку A1 в новом файле MS Excel.
- значение x_{11} растяните в m строк (по варианту), затем n столбцов ($n=12 \div 15$);
- скопируйте полученный массив X_{mn} (<u>только значения</u>) в лист 2.

Примечание: Необходимость этого шага вызвана тем, что при выполнение действий над случайной величиной каждый раз запускается оператор СЛЧИС, что не позволяет получить однозначные результаты;
- полученные данные вам необходимы для следующих лабораторных работ, поэтому необходимо сохранить файл значений X_{mn} на листе 2.

3. Получите одномерный массив Y:

- рассчитайте значение y_i СВ Y по выражению (1.4), где n- это количество столбцов в X_{mn};
- растяните вычисленное значение Y в m строк;
- скопируйте полученный столбец Y_m и поместите столбец рядом по возрастанию, используя оператора сортировки по возрастанию $\frac{A}{Z}\downarrow$. Полученный столбец чисел будет обозначаться $\vec{Y}$.

4. Получите одномерный массив Z:

- рассчитайте СВ Z, используя формулу (1.5), и получите столбец чисел Z_m;
- постройте диаграмму СВ ($\vec{Y}$,Z_m) с помощью мастера Диаграмм, тип диаграммы – точечная, без соединений между точками.

Примечание. Столбцы со значениями $\vec{Y}$ и Z_m должны находиться рядом, причем столбец $\vec{Y}$ должен находиться перед столбцом Z_m.

5. Сохраните полученные массивы X_{mn}, Y_m, $\vec{Y}$ и Z_m для использования в последующих работах (<u>обязательно</u>!).

Контрольные вопросы

1. Дайте определение случайных величин и их классификацию.
2. Что такое вероятность и каковы ее свойства?
3. Что такое относительная частота и как она связана с вероятностью?
4. Что характеризует закон распределения СВ и как он задается?
5. Какие существуют законы распределения СВ и от чего они зависят?
6. Зачем нужно моделирование СВ и как оно осуществляется?

Задачи

1. Брошена игральная кость. Найти вероятность того, что выпадет четное число очков.
2. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100.Найти вероятность того, что номер первого, наудачу извлеченного жетона, не содержит цифры 5.
3. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера, которые затем тщательно перемешаны. Найти вероятность того, что наудачу извлеченный кубик имеет окрашенных граней : а) одну; б) две; в) три.
4. Отдел технического контроля обнаружил пять бракованных книг в партии из случайно отобранных 100 книг. Найти относительную частоту появления бракованных книг.
5. По цели произведено 20 выстрелов, причем зарегистрировано 18 попаданий. Найти относительную частоту попадания в цель.
6. При испытании партии приборов относительная частота годных приборов оказалась равной 0,9. Найти число годных приборов, если всего было проверено 200 приборов.

Лабораторная работа 2

СТАТИСТИЧЕСКИЙ АНАЛИЗ ОДНОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Цель работы

Освоение методов статистического анализа одномерных случайных величин.

Задание к работе

Используя одномерную случайную величину (СВ) *Y,* полученную в работе 1, определите интегральную и дифференциальную функции распределения, рассчитайте точечные оценки математического ожидания, дисперсии, среднего квадратического отклонения, моды, коэффициента вариации, медианы, асимметрии, эксцесса; определите границы доверительного интервала при значении доверительной вероятности *P*, указанной в таблице 2.1 по вариантам:

Таблица 2.1

№ вар.	1	2	3	4	5	6	7	8	9	10
P	0.95	0.99	0.999	0.95	0.99	0.95	0.80	0.79	0.75	0.8
№ вар.	11	12	13	14	15	16	17	18	19	20
P	0,85	0,999	0,95	0,99	0,999	0,95	0,99	0,999	0,95	0,95

Краткие теоретические сведения

Закон распределения.

Исчерпывающей характеристикой СВ является закон распределения, который может быть представлен в формах интегральной функции распределения ИФР и дифференциальной функции распределения ДФР вероятностей СВ.

Интегральная функция распределения (ИФР) *F(y)* СВ *Y* показывает вероятность того, что СВ не превышает некоторого заданного или текущего значения *y*:

$$F(y) = P\{Y \le y\},$$

где *P*- оператор вероятности (конкретную величину вероятности принято обозначать *p*).

Основные свойства *F(y)*:

1) $F(y) \ge 0$ для всех y; 2) $F(y_2) > F(y_1)$, если $y_2 > y_1$,

3) $\lim\limits_{y \to -\infty} F(y) = 0$; 4) $\lim\limits_{y \to \infty} F(y) = 1$.

Вероятности:

$$\begin{aligned} &P\{y_1 < Y \le y_2\} = F(y_2) - F(y_1) \\ &P\{Y > y\} = 1 - F(y) \end{aligned}.$$

Если функция *F(y)* дифференцируема для всех значений СВ *Y*, то закон распределения вероятностей можно выразить в аналитической форме с помощью дифференциальной функции распределения (ДФР) вероятностей:

$$f(y) = \frac{dF(y)}{dy} = \lim_{\Delta y \to 0} \frac{P\{y < Y > y + \Delta y\}}{\Delta y}.$$

Функцию *f(y)* называют также плотностью распределения вероятностей (плотностью вероятностей).

Основные свойства *f(y)*:

1) $f(y) \ge 0$; 2) $\int\limits_{-\infty}^{\infty} f(y)dy = 1$;

3) $\int\limits_{-\infty}^{y} f(x)dx = F(y)$; 4) $\lim\limits_{|y| \to \infty} f(y) = 0$,

где *x* - переменная интегрирования.

С помощью *f(y)* можно вычислить вероятность нахождения СВ в любой области из множества ее возможных значений:

$$P\{Y \le a_1\} = \int_{-\infty}^{a} f(y)dy;$$

$$P\{Y \ge a_2\} = \int_{a_2}^{\infty} f(y)dy;$$

$$P\{a_1 < Y \le a_2\} = \int_{a_1}^{a_2} f(y)dy.$$

Числовые характеристики СВ.

Статистические свойства СВ часто представляют с помощью числовых характеристик, так как:

1) получить ИФР $F(y)$, ДФР $f(y)$ затруднительно, а иногда невозможно;
2) расчеты с использованием ИФР и ДФР оказываются громоздкими и сложными.

Различают 3 группы числовых характеристик:

1) характеристики, определяющие центр закона распределения (математическое ожидание, медиана, мода) или конкретное значение Y при заданной вероятности (квантиль);
2) характеристики, определяющие масштаб распределения (дисперсия, среднеквадратическое отклонение (СКО), коэффициент вариации);
3) характеристики, определяющие форму распределения (асимметрия, эксцесс).

Математическое ожидание m_y СВ Y характеризует центр для симметричных законов распределения:

$$m_y = M(Y) = \int_{-\infty}^{\infty} yf(y)dy.$$

Медиана M_e СВ Y есть такое ее значение y, когда

$$F(y) = \int_{-\infty}^{y} f(x)dx = 0{,}5,$$

где x - переменная интегрирования, т.е. медиана делит площадь под кривой $f(y)$ пополам и является удобной характеристикой центра распределения для СВ, у которых ДФР $f(y)$ не является симметричной.

Мода M_O есть экстремальное (максимальное) значение ДФР $f(y)$, т.е. соответствует такому значению СВ Y=y, когда

$$F(y) = \max_{y \in Y} f(y).$$

Квантиль y_α уровня вероятности α есть такое значение СВ Y= y_α, когда вероятность попадания значения СВ в интервал [-∞, y_a] равна α, т.е.

$$F(y_\alpha) = P\{Y \le y_\alpha\} = \alpha.$$

Для симметричных распределений квантиль $y_{0,5}$ совпадает с математическим ожиданием.

Частными случаями квантиля y_α являются квартиль, дециль, процентиль, которые определяют такие значения *y*, когда диапазон изменения СВ *Y* делится соответственно на 4, 10, 100 интервалов, попадание в которые имеет равные вероятности.

Степень рассеяния СВ *Y* относительно m_y можно описать с помощью дисперсии:

$$D_y = \int_{-\infty}^{\infty} (y - m_y)^2 f(y) dy = \int_{-\infty}^{\infty} y^2 f(y) dy - m_y^2 = M\left[Y^2\right] - m_y^2,$$

а также среднего квадратического отклонения, которое представляет собой положительный квадратный корень из дисперсии:

$$\sigma_y = \sqrt{D_y}.$$

Иногда масштаб распределения удобно представлять безразмерной величиной, называемой коэффициентом вариации (относительным СКО) *V*;

$$V = \frac{\sigma_y}{m_y},$$

который может быть выражен в процентах.

При нормальном законе распределения функции *f(y)* и *F(y)* имеют вид:

$$f(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp\left[-\frac{(y - m_y)^2}{2\sigma^2}\right];$$

$$F(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \int_{-\infty}^{y} \exp\left[-\frac{(x - m_y)^2}{2\sigma^2 y}\right] dx,$$

где *x* - переменная интегрирования.

Если закон распределения СВ отличается от нормального, то степень этого различия может быть оценена с помощью параметров: асимметрии и эксцесса.

Асимметрия характеризует несимметричность функции *f(y)* относительно математического ожидания и определяется соотношением:

$$A_S = \frac{\mu_3}{\sigma^3},$$

где $\mu_3 = \int_{-\infty}^{\infty} (y - m_y)^3 \cdot f(y) dy$ - центральный момент третьего порядка.

Для симметричного распределения $\mu_1 = \mu_2 = \mu_3 = \ldots = 0$, следовательно, A_S=0.

Если $A_S > 0$, то «длинная часть кривой» *f(y)* расположена справа от m_y; если $A_S < 0$, то слева.

Величина эксцесса характеризует степень островершинности исследуемого закона распределения по сравнению с нормальным и определяется выражением:

$$E_k = \frac{\mu_4}{\sigma^4} - 3,$$

где $\mu_4 = \int\limits_{-\infty}^{\infty} (y - m_y)^4 \cdot f(y)dy$ - центральный момент четвёртого порядка.

Для СВ с нормальным законом распределения A_S=0, E_k=0.

Если $E_k > 0$, то кривая *f(y)* исследуемого распределения имеет более высокую и «острую» вершину по сравнению с нормальной; если $E_k < 0$, то вершина исследуемого распределения имеет более низкую и «плоскую» вершину, чем нормальная кривая.

Примечание: Предполагается, что исследуемое распределение и нормальное имеют одинаковые математические ожидания и дисперсии.

Вероятность того, что нормальная СВ Y примет значение, принадлежащее интервалу $[\alpha, \beta]$, определяется выражением:

$$P(\alpha < Y < \beta) = \varPhi\left(\frac{\beta - m_y}{\sigma_y}\right) - \varPhi\left(\frac{\alpha - m_y}{\sigma_y}\right),$$

где $\varPhi(y) = \frac{1}{\sqrt{2\pi}} \int\limits_0^y e^{\frac{-y^2}{2}}$ - функция Лапласа (табулированная).

Моменты распределения СВ. Приведенные выше числовые параметры СВ тесно связаны с понятием моментов распределения СВ. Различают начальные и центральные моменты.

Начальный момент *k*-го порядка вычисляются из условия:

$$\nu_k = MY^k = \int\limits_{-\infty}^{\infty} y^k f(y)dy.$$

Следует заметить, что $\nu_0 = 1$, а $\nu_1 = m_y$.

Случайная величина $\overset{\circ}{Y} = Y - MY$ или $\overset{\circ}{Y} = Y - m_y$ называется центрированной. Центральный момент *k*-го порядка μ_k представляет собой математическое ожидание от центрированной СВ степени *k*, т.е.:

$$\mu_k = M(Y - m_y)^k = \int\limits_{-\infty}^{\infty} (y - m_y)^k f(y)dy.$$

Заметим, что $\mu_0 = 1$, $\mu_1 = 0$, и $\mu_2 = D_y$.

Таким образом, начальный момент 1-порядка есть математическое ожидание, а центральный момент 2-го порядка - дисперсия.

Центральный момент 2-го порядка μ_2 (дисперсию) можно выразить через начальные моменты:

$$\mu_2 = \nu_2 - \nu_1^2.$$

Понятия и процедуры_статистического анализа СВ.

Статистический анализ случайной величины направлен на определение её характеристик на основе экспериментальных данных. В основе статистического анализа лежат понятия генеральной и выборочной совокупностей. Генеральная совокупность - это множество всех мыслимых результатов наблюдений над СВ, которые в принципе могут быть проведены при заданных условиях.

Выборочной совокупностью или просто выборкой называют случайный отбор объектов из генеральной совокупности. Результатом случайной выборки объема *n* является множество $(y_1, \ldots, y_n)$ значений исследуемого признака,

Пусть имеется генеральная совокупность, в которой признак Y распределён по закону $F(y)$. Тогда *n*-мерный случайный вектор $(Y_1, \ldots, Y_n)$, в котором Y_i независимы друг от друга и все имеют распределение $F(y)$, называется математической выборкой объема *n*. Каждая реализация $(y_1, \ldots, y_n)$ есть выборка. Методы математической статистики позволяют на основе анализа выборки делать заключение о параметрах генеральной совокупности.

Исходными данными для анализа является набор (выборка) *n* экспериментальных значений $(y_1, y_2, \ldots, y_i, \ldots, y_n)$ СВ *Y*, называемых вариантами.

Вариационным рядом называется упорядоченная по возрастанию последовательность значений вариант.

Размахом вариационного ряда называется разность:

$$R = y_{\max} - y_{\min},$$

где $y_{\max}$ и $y_{\min}$ - максимальное и минимальное значение СВ *Y* в вариационном ряду.

Теоретические функции распределения генеральной совокупности характеризуются эмпирическими функциями распределения выборочной совокупности. Для наглядности часто используют графическое представление эмпирических распределений СВ. Дифференциальный закон распределения СВ иллюстрируют гистограмма и полигон частот, интегральный закон распределения - диаграмма накопленных частот и кумулята, соответственно.

Построение гистограммы и полигона частот осуществляется в такой последовательности:

1.Находится количество интервалов *К*, на которое разбивается размах вариационного ряда. Величина *К* оценивается по выражению:

$$K = 1 + 3{,}2 \cdot \lg n \quad ,$$

причем найденное значение округляют до ближайшего большего целого числа.

2. Определяется длина интервала: $\Delta y = \dfrac{R}{K}$,

причем величину Δy можно несколько округлить в большую сторону для удобства вычислений.

3.Середина области изменения выборки $(y_{\max} + y_{\min})/2$ принимается за центр графика, после чего уточняются границы и окончательное количество интервалов так, чтобы они в совокупности перекрывали всю область от $y_{\min}$ до $y_{\max}$.

4.Подсчитывается количество наблюдений $n_к$, попавших в каждый интервал, $n_к$ равно числу членов вариационного ряда, для которых справедливо неравенство:

$$y_k \le y_i < y_k + \Delta y,$$

где y_k и $y_k + \Delta y$ - границы *k*- го интервала. Значение y_i, попавшее на границу между $(k-1)$-м и k-м интервалами, относится к k-му интервалу.

5.Подсчитывается относительная частота наблюдений n_k / n, попавших в каждый интервал.

6.Строится гистограмма - ступенчатая фигура, состоящая из прямоугольников, высота каждого из которых на k-м интервале $(y_k,\ y_k + \Delta y)$, $(k = 1,2,...,K)$ постоянна и равна n_k / n.

Примечания:

1 . При построении гистограмм предполагается, что все значения СВ в пределах одного интервала имеют одинаковую частоту, что является приближением. Поэтому гистограмма служит для наглядного представления распределения частот, но не используется для оценок вероятностей и сравнения различных распределений.

2.Гистограмму можно построить также, используя абсолютные частоты n_k.

Для построения полигона частот подсчитывается плотность частоты на каждом интервале φ_k - эмпирический аналог плотности вероятности $f(y)$;

$$\varphi_k = \frac{n_k}{n \cdot \Delta y}.$$

Затем рассчитываются координаты середин каждого интервала $\overline{y_k} = y_{k-1} + \frac{\Delta y}{2}$ и строится полигон частот - ломаная линия, соединяющая точки $(\overline{y_k}, \varphi_k)$.

Примечание: Описанный полигон частот является эмпирическим аналогом дифференциальной функции распределения и может использоваться для вычисления вероятностей и для сравнения различных распределений.

Построение диаграммы накопленных частот производится в соответствии с формулой:

$$F_k(y) = \frac{1}{n}\sum_{j=1}^{k} n_j, k = \overline{1,K},$$

где k-номер интервала, для которого рассчитывается значение $F_k(y)$; $\mathrm{j} = \overline{1,\mathrm{k}}$ - номера интервалов, предшествующих k-му, включая k-й; n_j - количество наблюдений, попавших в j-й интервал.

Значение $F_k(y)$=0 при $y<y_{min}$, $F_k(y)$=1 при $y>y_{max}$. Диаграмма накопленных частот является эмпирическим аналогом ИФР, поэтому $F_k(y) \to F(y)$ при $K \to \infty$

Построение $F_k(y)$ можно осуществить аналогично построению гистограммы или полигона частот.

Рассмотрим способы нахождения точечных оценок некоторых числовых параметров на основе заданных вариант.

Мода M_o находится как варианта, имеющая наибольшую частоту.

Медиана M_e находится как варианта, которая делит вариационный ряда на две части, равные по числу вариант. Если число вариант нечетно, т.е. $n = 2m + 1$, то $M_e = y_{m+1}$; при четном $n=2 \cdot m$:

$$M_e = \frac{y_m + y_{m+1}}{2}.$$

Выборочная средняя служит оценкой математического ожидания и определяется по формуле:

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

Выборочная дисперсия служит оценкой дисперсии теоретического распределения и определяется формулой:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 .$$

Выборочным средним квадратическим отклонением (стандартом) S называется корень квадратный из выборочной дисперсии.

Коэффициент вариации V характеризует степень рассеяния вариационного ряда и определяется следующей формулой:

$$V = \frac{S}{\overline{y}} \cdot 100\% .$$

Выборочные значения ассиметрии A_{SB} и эксцесса E_{KB} рассчитываются по формулам:

$$A_{SB} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^3}{n \cdot S^3};$$

$$E_{KB} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^4}{n \cdot S^4} - 3.$$

Для оценки истинного значения θ СВ, исследуемой экспериментально, используется доверительный интервал, границы которого y_{r1} и y_{r2} являются функциями выборочных значений $y_1, y_2, \ldots, y_n$ и который с заданной вероятностью p накрывает оцениваемый параметр θ:

$$P\{y_{r1} < \theta \le y_{r2}\} = p .$$

Интервал $[y_{r1}, y_{r2}]$ называется доверительным, его границы y_{r1} и y_{r2} – соответственно нижним и верхним доверительными пределами, вероятность p – доверительной вероятностью или надежностью, а величина $\alpha = 1 - p$ - уровнем зависимости, используемым при построении доверительного интервала.

На практике используют значение p =0,95; реже p =0,9 и p= =0,99; совсем редко p=0,8 и p =0,999.

При заданном значении p порядок определения границ доверительного интервала зависит от имеющейся информации о законе распределения исследуемой СВ. Так, если известно, что СВ Y распределена по нормальному закону с неизвестными параметрами, то значения СВ в выборке объемом n используется для определения $\overline{y}$ и S, после чего границы доверительного интервала y_{r1} и y_{r2} находят из соотношения:

$$y_{r1} = \overline{y} - t_{p,n} * \frac{S}{\sqrt{n}};$$
$$y_{r2} = \overline{y} + t_{p,n} * \frac{S}{\sqrt{n}}, \qquad (2.1)$$

где $t_{p.n}$ - число, определяемое из таблиц t - распределения (Стьюдента) при доверительной вероятности p и размере выборки n (см. Прил. 1) для двусторонней критической области с учетом того, что α =1-p.

Примечания:

1. В других литературных источниках формы представления таблиц t-распределения и методика определения числа t несколько отличны, на что следует обращать внимание.

2. При возрастании объёма выборки n распределение Стьюдента стремится к нормальному, поэтому при n>30 можно вместо распределения Стьюдента пользоваться нормальным (Прил.2). При этом в формулах (2.1) используют число t_p, определяемое из равенства $2\Phi(t_p)=p$ или $\Phi(t_p)=p/2$.

Нормальным распределением пользуются также, если среднее квадратическое отклонение исследуемой СВ известно до эксперимента.

Методические указания к выполнению работы

1 . В качестве совокупности экспериментальных данных, подлежащих статистическому анализу, используйте одномерную нормальную СВ $\vec{Y}$, полученную ранее.

2.Получите вариационный ряд и оцените его размах R.

3.Постройте гистограмму, полигон частот и диаграмму накопленных частот.

4.Найдите числовые характеристики СВ Y: $M_0, M_e, \bar{y}, S, V, A_{SB}, E_{kb}$, а также границы доверительного интервала y_{r1} и y_{r2}. Нанесите числовые характеристики на эмпирически найденные графики закона распределения (гистограмму, полигон частот, диаграмму накопленных частот).

Контрольные вопросы

1. Дайте определение одномерных интегрального и дифференциального законов распределения СВ, опишите их свойства.
2. Приведите аналитическое выражение интегрального и дифференциального законов распределения одномерной нормальной СВ, охарактеризуйте входящие в них параметры.
3. Охарактеризуйте числовые параметры одномерной СВ.
4. Что такое генеральная и выборочная совокупности?
5. Что такое вариационный ряд?
6. Что такое гистограмма, полигон частот и диаграмма накопленных частот и как производится их построение?
7. Перечислите числовые характеристики вариационного ряда и способы их оценки.
8. Дайте определение доверительному интервалу, поясните его смысл и опишите способ его построения.

Задачи

1. Случайная величина X задана функцией распределения:

$$F(x)=\begin{cases}0, & при \quad x\le 2\\ 0{,}5, & при \quad 2<x\le 4\\ 1, & при \quad x>4\end{cases}.$$

Найти вероятность того, что в результате испытания X примет значения:

а) меньше 0,2;
б) меньше 3;
в) не меньше 3;
г) не меньше 5.

2. Дискретная случайная величина задана законом распределения:

X	3	4	7	10
p	0,2	0,1	0,4	0,3

Найти функцию распределения *F(x)* и построить ее график.

3. Задана плотность распределения непрерывной случайной величины X:

$$f(x)=\begin{cases}0, & при \quad x\le 0\\ \sin x, & при \quad 0<x\le \dfrac{\pi}{2}\\ 0, & при \quad x>\dfrac{\pi}{2}\end{cases}.$$

Найти функцию распределения *F(x)*.

4. Случайная величина X задана плотностью распределения *f(x)=(1/2)x* в интервале [0, 2]; вне этого интервала *f(x)*=0. Найти математическое ожидание величины X.
5. Найти дисперсию случайной величины X, заданной функцией распределения:

$$F(x)=\begin{cases}0, & при \quad x\le -2\\ \dfrac{x}{4}+\dfrac{1}{2}, & при \quad -2<x\le 2\\ 1, & при \quad x>2\end{cases}.$$

6. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X соответственно равно 10 и 2. Найти вероятность того, что в результате испытания X примет значение, заключенное в интервал (12, 14).
7. Случайная величина X распределена нормально с математическим ожиданием a=25. Вероятность попадания X в интервал [10; 15] равна 0,2. Чему равна вероятность попадания X в интервал [35; 40]?
8. Случайная величина X распределена нормально со средним квадратиче-

ским отклонением $\sigma = 5$ мм. Найти длину интервала, симметричного относительно математического ожидания, в котором с вероятностью 0,9973 попадает X в результате испытания.

9. Найти доверительный интервал для оценки с надежностью 0,95 неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если генеральное среднее квадратическое отклонение $\sigma = 5$, выборочная средняя X_B =14 и объем выборки n=25.
10. Найти минимальный объем выборки, при котором с надежностью 0,975 точность оценки математического ожидания a генеральной совокупности по выборочной средней равна $\delta = 0{,}3$, если известно среднее квадратическое отклонение $\sigma = 1{,}2$ нормально распределенной генеральной совокупности.
11. По данным девяти независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений X_B = 30,1 и выборочное среднее квадратическое отклонение s=6. Оценить истинное значение измеряемой величины с помощью доверительного интервала с надежностью p=0,99, предполагая, что результаты измерений распределены нормально.
12. По данным выборки объема n=16 из генеральной совокупности найдено «выборочное» среднее квадратическое отклонение s=1 нормально распределенного количественного признака. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95.

Лабораторная работа 3

ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ДВУМЯ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ

Цель работы

Освоение методов установления формы и тесноты взаимосвязи между двумя случайными величинами.

Задание к работе

Пользуясь выборочными значениями двумерной СВ (Y,Z), рассчитать коэффициент корреляции и корреляционное отношение, оценить значимость коэффициента корреляции, то есть проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции $H_0 : r_r = 0$ при конкурирующей гипотезе $H_1 : r_r \neq 0$; уровни значимости a по вариантам приведены в таблице 3.1.

Таблица 3.1

N вар.	1	2	3	4	5	6	7	8	9	10	11
α	0,1	0,05	0,02	0,01	0,002	0,001	0,1	0,05	0,02	0,1	0,002
N вар.	12	13	14	15	16	17	18	19	20	21	22
α	0,001	0,1	0,05	0,02	0,01	0,002	0,001	0,1	0,05	0,02	0,01

Краткие теоретические сведения

Часто явление характеризуется несколькими параметрами, носящими случайный характер. Если таких параметров 2, 3, . . . , n, то говорят о двухмерной, трехмерной, ..., n-мерной случайной величине. Многомерность случайных величин принято обозначать как (X, Y, …). Геометрически двухмерную СВ можно представить либо как случайную точку $M(X,Y)$ на плоскости (т.е. как точку со случайными координатами), либо как случайный вектор OM, соединяющий начало координат О и точку М. Трехмерную СВ геометрически можно истолковать как точку $M(x,y,z)$ в трехмерном пространстве, n-мерную СВ- как точку в n-мерном пространстве или как соответствующий вектор OM.

Компоненты многомерной СВ могут быть связаны либо функциональной зависимостью, либо статистической, либо быть независимыми. Например, если компоненты двухмерной СВ (Y, Z) связаны функциональной зависимостью, то каждому возможному значению СВ Y соответствует одно возможное значение Z:

$$z = \varphi(y). \tag{3.1}$$

Строгая зависимость (3.1) реализуется редко, так как Y и Z подвержены действию случайных факторов; при этом возникает статистическая зависимость, когда изменение одной из СВ влечет изменение закона распределения другой. В частности, если при изменении одной из СВ меняется среднее значение другой, то статистическая зависимость называется корреляционной.

Корреляция может быть линейной и нелинейной. Для оценки тесноты линейной корреляционной связи между Y и Z служит коэффициент корреляции:

$$\rho_{yz} = \frac{\mu_{yz}}{\sigma_y \sigma_z},$$

где σ_y и σ_z -средние квадратические отклонения СВ Y и СВ Z, соответственно; μ_{yz} - корреляционный момент системы (Y, Z), определяемый соотношением:

$$\mu_{yz} = M\{[Y - m_y] \cdot [Z - m_z]\},$$

где m_y и m_z - математические ожидания Y и Z, соответственно.

Если Y и Z непрерывны, то

$$\mu_{yz} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - m_y)\cdot(z - m_z) f(y,z)\,dydz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yzf(y,z)\,dydz - m_y m_z.$$

Как известно, $|\rho_{yz}| \le 1$, причем $\rho_{yz} = 1$ соответствует функциональной линейной связи. При ρ_{yz}=0 линейная корреляционная связь отсутствует.

Если случайный вектор (*Y, Z*) представлен выборкой объёма *n*, то оценку ρ_{yz} можно произвести с помощью выборочного коэффициента корреляции r_{yz} по формуле:

$$r_{yz} = \frac{1}{(n-1)S_y S_z} \sum_{i=1}^{n} \left(y_i - \overline{y}\right)\left(z_i - \overline{z}\right),$$

где y_i , z_i - пары значений *Y* и *Z*, $\overline{y}, \overline{z}$ - средние значения *Y* и *Z*, S_y, S_z - выборочные средние квадратические отклонения *Y* и *Z*.

Для обоснованного суждения о наличии линейной связи между СВ *Y* и *Z* следует проверить, значимо ли выборочный коэффициент корреляции r_{yz} отличается от 0. Это достаточно легко сделать, если исследуемая генеральная совокупность *(Y, Z)* распределена по нормальному закону.

Пусть из нормальной СВ *(Y, Z*) извлечена выборка объема *n* и по ней найден выборочный коэффициент корреляции $r_{yz} \neq 0$. Требуется проверить нулевую гипотезу $H_0 : \rho_{yz}$=0 о равенстве нулю генерального коэффициента корреляции, при этом конкурирующая гипотеза формируется в виде $H_1 : \rho_{yz} \neq 0$. Тогда для проверки нулевой гипотезы при уровне значимости α надо вычислить наблюдаемое значение критерия:

$$T_{набл} = \frac{r_{yz}\sqrt{n-2}}{\sqrt{1-(r_{yz})^2}},$$

и по таблице критических точек распределения Стьюдента (Прил.1) при заданном уровне значимости α и числу степеней свободы *k=n-2* найти критическую точку $t_{кр}$ (α, k) двусторонней критической области.

Если установлено, что | $T_{набл}$ | > $t_{кр}$, то нулевую гипотезу отвергают, т.е. r_{yz} значимо отличается от нуля и величину r_{yz} можно использовать для оценки параметра p_{yz}, характеризующего тесноту линейной корреляционной связи. Для этого при $n \geq 50$ можно воспользоваться формулой:

$$r_{yz} - 3\frac{1-r^2{}_{yz}}{\sqrt{n}} \leq \rho_{yz} \leq r_{yz} + 3\frac{1+r^2{}_{yz}}{\sqrt{n}},$$

выражающей доверительный интервал для ρ_{yz}.

Если $|T_{набл}| < t_{кр}$ -нет основании отвергнуть нулевую гипотезу, т.е. r_{yz} отличается от 0 незначимо. Из этого следует, что генеральный коэффициент корреляции ρ_{yz}=0, т.е. линейная корреляционная взаимосвязь отсутствует.

Следует отметить, что при ρ_{yz} =0 признаки *Y* и *Z* могут быть связаны нелинейной корреляционной и даже функциональной связью. Поэтому, если нулевая гипотеза не отвергнута и r_{yz} отличается от нуля незначимо, следует проверить совокупность (*Y, Z*) на наличие нелинейной взаимосвязи.

Для оценки степени нелинейной корреляционной взаимосвязи между *Y* к *Z* служит выборочное корреляционное отношение. Его величина может быть рассчитана, если совокупность (*Y, Z*) разбита на группы. Каждая группа характеризуется конкретными значениями (y_i, z_j), причём хотя бы некоторые сочетания встречаются в совокупности (Y, Z) несколько раз, как, например, представлено в корреляционной табл. 3.2.

В таблице жирной линией обведена совокупность исходных экспериментальных данных, показывающих количество n_{ij} сочетаний (y_i, z_j); остальные параметры рассчитаны и означают: n_y - частота признака y, её i-е значение:

$$n_{yi} = \sum_{j=1}^{k} n_{ij};$$

n_z - частота признака z, её j-е значение: $n_{zj} = \sum_{i=1}^{m} n_{ij}$;

n - объем выборки: $$n = \sum_{i=1}^{m} n_{yi} = \sum_{j=1}^{k} n_{zj} = \sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij};$$

$\bar{y}_{zj}$ и $\bar{z}_{yi}$ - условные средние: $$\bar{y}_{zj} = \frac{1}{n_{zj}} \sum_{i=1}^{m} y_i \cdot n_{ij};$$

$$\bar{z}_{yi} = \frac{1}{n_{yi}} \sum_{j=1}^{k} z_j \cdot n_{ij}.$$

Таблица 3.2

Z / Y	z_1	…	z_j	…	z_k	n_y	$\bar{z}_y$
y_1	n_{11}	…	n_{1j}	…	n_{1k}	n_{y1}	$\bar{z}_{y1}$
…							
y_i	n_{i1}	…	n_{ij}	…	n_{ik}	n_{yj}	$\bar{z}_{yi}$
…							
y_m	n_{m1}	…	n_{mj}	…	n_{mk}	n_{ym}	$\bar{z}_{ym}$
n_z	n_{z1}	…	n_{zj}	…	n_{zk}	n	
$\bar{y}_z$	$\bar{y}_{z1}$	…	$\bar{y}_{zj}$	…	$\bar{y}_{zk}$		

Приведенные данные позволяют оценить выборочное корреляционное отношение Z к Y в форме:

$$\eta_{ZY} = \frac{\sigma_{\bar{z}y}}{\sigma_{\bar{z}}},$$

где $\sigma_{\bar{z}y}$ - среднее квадратичное отклонение условных средних $\bar{z}_y$ от общей средней $\bar{z}$:

$$\sigma_{\bar{z}y} = \sqrt{\frac{1}{n}\sum_{i=1}^{m} n_{yi}(\bar{z}_{yi} - \bar{z})^2},$$

$$\bar{z}=\frac{1}{n}\sum_{j=1}^{k} n_{zj}\cdot z_j\ ; \qquad (3.2)$$

$\sigma_{\bar{z}}$ - среднее квадратическое отклонение значений признака z всей совокупности относительно общей средней:

$$\sigma_{\bar{z}}=\sqrt{\frac{1}{n}\sum_{j=1}^{k} n_{zj}(\bar{z}_j-\bar{z})^2}\ .$$

Примечания:

1 . Формула выборочного корреляционного отношения Y к Z η_{YZ} может быть получена аналогично.

2.Корреляционная таблица может быть использована для расчета выборочного коэффициента корреляции по формуле:

$$r_{yz}=\frac{1}{(n-1)S_y\cdot S_z}\sum_{i=1}^{m}\ \sum_{j=1}^{k} n_{ij}y_i z_j - n\overline{y}\overline{z};$$

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{m}\ n_{yi}y_i\,,$$

где $\overline{y}$ - общая средняя y, $\bar{z}$ - определяется формулой (3.2).

Возможные значения η лежат в пределах $0\le\eta\le1$.

Если η=0, то исследуемые признаки Y и Z корреляционной зависимостью не связаны, при η=1 между признаками Y и Z функциональная связь. Между величинами η и r существует соотношение $\eta\ge|r|$, , причем, если $\eta=|r|$, то между признаками Y и Z имеет место линейная корреляционная зависимость.

Методические указания к выполнению работы

1 . Для выполнения работы воспользуйтесь выборочными значениями нормальной СВ ($\overline{Y}$, Z), полученными в лабораторной работе 1. Поле рассеяния ($\overline{Y}$, Z) приведите в отчете.
2. Рассчитайте величину выборочного коэффициента корреляции, проверьте его значимость при уровне значимости, заданном по варианту.

Контрольные вопросы

1. Дайте определение многомерной СВ, приведите примеры.
2. Охарактеризуйте виды взаимосвязи между компонентами многомерной СВ, приведите примеры.
3. Что такое корреляционная связь и каковы ее разновидности?
4. Как определяется теснота линейной корреляционной связи?
5. Как оценивается значимость выборочного коэффициента корреляции?
6. Охарактеризуйте свойства коэффициента корреляции.
7. Как определяется теснота нелинейной корреляционной связи?

8. Опишите свойства выборочного корреляционного отношения.
9. Произведите сопоставление коэффициента корреляции и корреляционного отношения.

Задачи

1. Даны корреляционные таблицы 3.3 и 3.4. Найти:
   а) выборочный коэффициент корреляции r_{yx};
   б) выборочное корреляционные отношение η_{yx};
   в) выборочное корреляционные отношение η_{xy};
   г) определить значимость коэффициента корреляции при уровне значимости $\alpha = 0{,}01$.

Таблица 3.3

Y \ X	20	25	30	35	40
10	4	6	-	-	-
20	-	8	10	-	-
30	-	-	32	3	9
40	-	-	4	12	6
50	-	-	-	1	5

Таблица 3.4

X / Y	5	10	15	20
10	2	-	-	-
20	5	4	-	-
30	3	8	6	3
40	-	3	6	6
50	-	-	2	1

3.По выборке объема n=62, извлеченной из нормальной двухмерной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции $r_в$=0,3. Требуется при уровне значимости 0,01 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1\colon \rho \neq 0$.

4. По выборке объема n=120, извлеченной из нормальной двухмерно генеральной совокупности (X, Y), найден выборочный коэффициент корреляции $r_в$=0,4. Требуется при уровне значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1\colon \rho \neq 0$.

Лабораторная работа 4

ОДНОМЕРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Цель работы

Ознакомление с идеей и принципами применения регрессионного анализа для построения математической модели объекта.

Задание к работе

На основании выборки объемом n двухмерной нормальной СВ (Z, Y) получить выборочное уравнение регрессии Z на Y; оценить адекватность описания выборочной совокупности с помощью уравнения регрессии при уровне значимости, указанном в задании к лабораторной работе 3.

Краткие теоретические сведения

Задача построения математической модели, связывающей параметры случайного характера, широко распространена в практике; например, при идентификации объектов. В простейшем случае таких параметров два.

Пусть имеются количественные признаки Y и Z, связанные корреляционной зависимостью. Взаимосвязь между Y и Z характеризуется выражениями:

$$\bar{z}(y) = M(Z / Y = y) \tag{4.1}$$

$$\bar{y}(z) = M(Y / Z = z); \tag{4.2}$$

где $M(Z/Y=y)$- условное математическое ожидание Z при условии, что Y имеет значение y, а $M(Y/Z=z)$- условное математическое ожидание Y при условии $Z=z$, $\bar{z}(y)$ и $\bar{y}(z)$- соответствующие средние.

В случае непрерывных Y и Z:

$$M(Z / Y = y) = \int_{-\infty}^{\infty} zf(x / y)dz;$$

$$M(Y / Z = z) = \int_{-\infty}^{\infty} yf(y / z)dy,$$

где $f(z/y)$ и $f(y/z)$- условные плотности распределения вероятности.

Уравнение (4.1) имеет смысл: наилучшее предсказание Z при условии, что $Y=y$, есть $\bar{z}(y)$. При этом наилучшее означает, что для произвольной функции $u(y)$ справедливо соотношение:

$$M(Z - u(y))^2 \geq M(Z - \bar{z}(y))^2,$$

то есть $\bar{z}(y)$- это функция регрессии, минимизирующая среднюю квадратическую ошибку величины предсказания Z на основании значений Y. Соответствующим образом можно интерпретировать $\bar{y}(z)$ (4.2).

Уравнения (4.1) и (4.2) называются уравнениями регрессии *Z* относительно *Y* и *Y* относительно *Z*, соответственно. Графики, соответствующие уравнениям (4.1), (4.2), называются линиями регрессии.

Если случайные величины *Y* и *Z* линейно коррелированны, то линии регрессии являются прямыми и задаются следующими уравнениями:

- регрессия *Z* относительно *Y:* $z = M(Z) + \beta_{zy}(y - M(Y));$ (4.3)
- регрессия *Y* относительно *Z:* $y = M(Y) + \beta_{yz}(z - M(Z)),$ (4.4)

где *M(Z)* и *M(Y)* – математические ожидания *Z* и *Y*; β_{zy} и β_{yz} называются теоретическими коэффициентами регрессии:

$$\beta_{zy} = \frac{\sigma_z}{\sigma_y}\rho;$$

$$\beta_{yz} = \frac{\sigma_y}{\sigma_z}\rho,$$

где ρ - коэффициент корреляции; σ_z и σ_y - средние квадратические отклонения *z* и *y*, соответственно.

Параметры уравнений (4.3) и (4.4) могут быть получены из результатов эксперимента. Пусть в процессе измерений характеристик объекта (*Y,Z*) получена связанная выборка объемом *n*, т.е. *n* пар значений (y_1, z_1), (y_2, z_2), … , (y_n, z_n), причем *Y* – независимый управляемый фактор, *Z* – отклик, на взаимосвязь между которыми влияет случайная помеха ε (рис. 4.1).

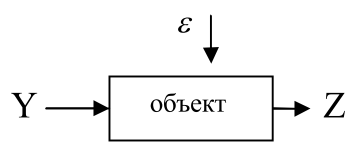

Рисунок 4.1

Эти значения (*Y,Z*) можно использовать для нахождения оценок параметров уравнений регрессии, если выполняются следующие предпосылки:

1. Результаты наблюдений Z_1, …, Z_n отклика представляют собой независимые нормально распределенные случайные величины, т.е. на них воздействуют нормально распределенные случайные помехи ε с нулевым математическим ожиданием $M(\varepsilon) = 0$.

2. Дисперсии $\sigma^2(Z_i)$ (i=1, 2 , 3,… n) равны. Это значит, что при проведении многократных повторных наблюдений над величиной Z_i в точках y_i, выборочные оценки $s^2(Z_i)$ однородны, а дисперсия $\sigma^2(Z_i)$ не зависит от математического ожидания $M(Z_i)$, то есть не отличается от дисперсии $\sigma^2(Z_q)$, полученной при повторных наблюдениях в любой другой точке y_q (воспроизводимость с равной точностью).

3. Независимый управляемый фактор *Y* измеряется с пренебрежимо малы-

ми ошибками по сравнению с ошибкой в определении Z (имеется в виду незначительность влияния ошибок Y на величину Z по сравнению с влиянием неуправляемых и неконтролируемых факторов ε).

При выполнении перечисленных условий экспериментальные данные можно использовать для получения оценок параметров уравнений регрессии.

$$\left.\begin{array}{ll} \text{Оценка } M(Z) & \bar{z} = \dfrac{1}{n}\sum_{i=1}^{n} z_i; \\ \text{оценка } M(Y) & \bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i . \end{array}\right\} \quad (4.5)$$

$$\left.\begin{array}{ll} \text{оценка } \sigma^2{}_z & s_z{}^2 = \dfrac{1}{n-1}\sum_{i-1}^{n} (z_i - \bar{z})^2; \\ \text{оценка } \sigma^2{}_y & s_y{}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2 . \end{array}\right\} \quad (4.6)$$

Эмпирический коэффициент корреляции:

$$r = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 * \sum_{i=1}^{n} (z_i - \bar{z})^2}}. \quad (4.7)$$

Примечание. Если некоторые сочетания значений встречаются в совокупности (Y,Z) несколько раз, то (y, z) удобно представить в корреляционной таблице, а оценки (4.5)-(4.6) можно сделать по формулам, приведенным в лабораторной работе 3.

Тогда оценками β_{zy} и β_{yz} служат соответственно:

$$b_{zy} = \frac{s_z}{s_y} r \quad \text{и} \quad b_{yz} = \frac{s_y}{s_z} r. \quad (4.8)$$

Правомочность использования найденных оценок (4.5), (4.8) в уравнениях регрессии (4.3), (4.4) необходимо подтвердить, для чего используется специальные статистические процедуры, осуществление которых возможно, если при каждом уровне управляемого параметра Y производится измерение нескольких значений отклика Z. Таких процедур три.

1. Прежде всего, перед расчетом оценок (4.5), (4.8) необходимо проверить выполнение второй предпосылки регрессионного анализа об однородности выборочных дисперсий измерения Z. Задача формулируется как проверка гипотезы о равенстве $\sigma^2(Z_1) = \sigma^2(Z_2) = \ldots = \sigma^2(Z_n)$ генеральных дисперсий при значениях управляемого фактора y_1 y_2, ..., y_n соответственно. Оценки этих дисперсий можно определить по формуле:

$$s_i^2 = \frac{1}{m_i - 1}\sum_{k=1}^{m_i} (z_{ik} - \bar{z}_i)^2;$$

где m_i-объем выборки значений Z при i-том уровне y_i фактора Y; z_{ik}, и $\bar{z}_i$ выборочные и среднее значения Z, соответственно, при i-том уровне фактора Y, причем $\bar{z}_i$ определяется формулой:

$$\overline{z_i} = \frac{1}{m_i}\sum_{k=1}^{m_i} z_{ik} .$$

Если при всех y_i объемы выборок Z одинаковы, т.е. $m_1 = ...m_i = ...m_n$, то для проверки гипотезы об однородности оценок дисперсий следует пользоваться критерием Кочрена, наблюдаемое значение которого определяется формулой:

$$G = \frac{\max(s_i^2)}{\sum_{i=1}^{n} s_i^2}.$$

Критическое значение $G_{кр}$ находят по таблице приложения 3 для k_l=m-1 и k_2=n и выбранного уровня значимости α (обычно 5%).

Если G>$G_{кр}$, то эксперимент признают не воспроизводимым относительно управляемых факторов вследствие наличия значительных флуктуаций неуправляемых и неконтролируемых факторов, влияющих на Z. При этом следует увеличить число параллельных опытов m для вариантов варьирования с большими значениями выборочных дисперсий s_i^2 либо принять меры по стабилизации дисперсии, либо использовать модификацию метода наименьших квадратов, пригодную при невыполнении предпосылки о воспроизводимости эксперимента.

Если G<$G_{кр}$, то гипотеза об однородности выборочных дисперсий не противоречит результатам наблюдений и эксперимент можно считать воспроизводимым. При этом всю группу выборочных дисперсий s_i^2 можно считать оценками одной и той же генеральной дисперсии σ_z^2 воспроизводимости эксперимента, откуда ее наилучшая оценка имеет вид:

$$s_{ВОС}^2(z) = \frac{1}{n}\sum_{i=1}^{n} s_i^2 \;; \qquad (4.9)$$

с числом степеней свободы

$$k_{зн}=n(m-1). \qquad (4.10)$$

При установлении воспроизводимости эксперимента можно рассчитать коэффициенты уравнения регрессии по формклам (4.5), (4.8).

Примечание. Если объемы выборок m_i при различных y_i неодинаковые, то для проверки воспроизводимости эксперимента следует использовать критерий Бартлета /1-4/.

3.После определения оценок коэффициентов уравнения регрессии (4.5), (4.8) следует проверить гипотезы об их значимости, т.е. проверить соответствующие нуль-гипотезы об их равенстве нулю. Так, например, для уравнения (4.3) нулевая гипотеза может быть сформулирована в виде:

$$H_0{:}M(Z)=0,\;\; \beta_{zy}=0; \qquad \text{при } H_1{:}M(Z)\neq 0,\;\; \beta_{zy}\neq 0.$$

Проверку такой гипотезе производят по критерию Стьюдента, эмпирические значения которого:

$$t_{\bar{z}} = \frac{|\bar{z}|}{s}, \qquad t_{в} = \frac{|b_{zy}|}{s}; \tag{4.11}$$

где $s^2 = \frac{1}{nm} s^2_{вос}(z)$ -дисперсия оценки коэффициентов уравнения (4.3), $s^2_{вос}(z)$ описывается уравнением (4.9).

Если $t_{\bar{z}}$ и $t_в$ превышают значение $t_{кр}$ из Прил.1 для числа степеней свободы $k_{зн}$ (4.10) при заданном уровне значимости α (обычно 5%) (двусторонняя критическая область), то проверяемую гипотезу H_0 отвергают и соответствующие оценки (4.5) и (4.8) принимают значимыми.

В противном случае, т.е. при $t_{\bar{z}} < t_{кр}$ или $t_в < t_{кр}$, нуль-гипотезу не отвергают и оценки, для которых эти неравенства выполняются, считают незначимыми. Соответствующие коэффициенты уравнения регрессии полагают равными "0" и в математическую модель связи Z и Y включают только члены, содержащие значимые оценки коэффициентов.

3. Полученное математическое описание объекта следует проверить на адекватность опытным данным. Для этого нужно оценить отклонение предсказанной по полученному уравнению регрессии величины отклика $\hat{z}_i$ от результатов наблюдений $\bar{z}_i$ с помощью дисперсии адекватности:

$$s^2_{АД} = \frac{m}{n-d}\sum_{i=1}^{n}(\bar{z}_i - \hat{z}_i)^2, \tag{4.12}$$

где d- число членов уравнения регрессии, найденного по экспериментальным данным; $\hat{z}_i$ определяется по формуле:

$$\hat{z}_i = \bar{z}_i + b_{zy}(y_i - \bar{y}). \tag{4.13}$$

Число степеней свободы дисперсии адекватности составляет:

$$k_{АД} = n - d.$$

Примечание. Если каждому значению управляемого фактора y_i соответствует одно значение отклика z_i, то дисперсия адекватности оценивается по формуле:

$$s^2_{АД} = \frac{1}{n-d}\sum_{i=1}^{n}(z_i - \hat{z}_i)^2, \tag{4.14}$$

где $\hat{z}$, определяется формулой (4.13), в которой $\bar{z}_i$ заменяется на z_i.

Проверка гипотезы адекватности состоит в сравнении дисперсий адекватности $s^2_{АД}$ (4.12) или (4.14) и воспроизводимости (4.9) с использованием F-критерия Фишера, который при $s^2_{АД} > s^2_{ВОС}(z)$ имеет вид:

$$F = \frac{s^2{}_{АД}}{s^2{}_{ВОС}(z)}.$$

Если $F < F_{кр}$, где $F_{кр}$ находится из Прил. 4 для соответствующих, степеней свободы:

$$k_1 = k_{АД} = n - d;$$

$$k_2=k_{зн}=n(m-1),$$

при заданном уровне значимости α , то гипотезу об адекватности не отвергают. В противном случае ($F>F_{кр}$) полученное математическое описание экспериментальных данных признается неадекватным.

В случае, если гипотеза об адекватности отвергается, следует переходить к более сложной форме математического описания или, если это возможно, уменьшить интервал изменения управляемого фактора Y. Однако в последнем случае растет отношение случайной помехи к полезному сигналу, что может привести к незначимости коэффициентов уравнения регрессии (4.8) . В этом случае необходимо увеличивать число измерений отклика Z при фиксированном y_i для выделения полезного сигнала на фоне шума.

Уравнение регрессии можно определять также методом наименьших квадратов, причем в этом случае его целесообразно искать в виде

$$\hat{z}_i = b_0 + b_1 \cdot y_i \tag{4.15}$$

где $\hat{z}_i$ - вычисленное по уравнению (4.15) значение z, соответствующее наблюдаемой величине y_i.

Сущность метода наименьших квадратов состоит в том, что параметры b_0 и b_i подбирают из условия, чтобы обеспечить минимальное значение функции:

$$F(b_0, b_1) = \sum_{i=1}^{n} (\hat{z}_i - z_i)^2$$

или

$$F(b_0, b_1) = \sum_{i=1}^{n} (b_0 + b_1 y_i - z_i)^2 \ .$$

Для этого решают уравнения

$$\frac{\partial F}{\partial b_0} = 0 \quad \text{и} \quad \frac{\partial F}{\partial b_1} = 0,$$

откуда

$$b_0 = \frac{\sum_{i=1}^{n} y_i^2 \sum_{i=1}^{n} z_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i z_i}{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}; \tag{4.16}$$

$$b_1 = \frac{n \sum_{i=1}^{n} y_i z_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} z_i}{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}. \tag{4.17}$$

Адекватность математического описания экспериментальных данных с помощью уравнения (4.15) можно проверить вышеописанным способом.

Методические указания к выполнению работы

1. Для выполнения работы воспользуйтесь выборочными значениями нормальной СВ ($\vec{Y}$,Z), полученными в лабораторной работе 1.

Примечание. Если ставятся задачи проверки воспроизводимости эксперимента, значимости коэффициентов уравнения регрессии и адекватности математического описания, то в совокупности ($\vec{Y},Z$) каждому значению y_i должно соответствовать несколько значений z_i. Такую совокупность можно получить, например, осуществив моделирование ($\vec{Y},Z$) несколько раз при одних и тех же параметрах (см. лабораторную работу 1), а затем использовать одну из совокупностей Y совместно со всеми полученными совокупностями Z.

2.Опишите совокупность ($\vec{Y},Z$) моделями (4.3) и (4.15), найдя их коэффициенты по уравнениям (4.5)-(4.8) и (4.16),(4.17) соответственно.

3.Изобразите линии регрессии (4.3) и (4.15) на рисунке, где представлена совокупность ($\vec{Y},Z$).

4.Рассчитайте $S^2_{АД}$ (4.14) для моделей (4.3) и (4.15) и сравните их с использованием критерия Фишера при заданном уровне значимости.

5.Сделайте выводы об эквивалентности использования моделей (4.3) и (4.15) для описания экспериментальных данных.

Контрольные вопросы

1. В чем состоит цель регрессионного анализа?
2. Каковы предпосылки использования одномерного регрессионного анализа?
3. Как проверить предпосылку об однородности выборочных дисперсий измерения отклика (выходного сигнала) объекта?
4. Как проверять полученное уравнение регрессии на адекватность опытным данным?
5. В чем состоит сущность метода наименьших квадратов?

Задача

По данным корреляционных таблиц 3.3 и 3.4 (см. задачу 1 к лабораторной работе 3) найдите выборочные уравнения прямых регрессии.

Лабораторная работа 5

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Цель работы

Ознакомление с назначением, предпосылками применения метода дисперсионного анализа и его алгоритмом.

Задание к работе

Выбрав массив экспериментальных данных Y из табл. 5.2 в соответствии с вариантом согласно табл. 5.1, оценить значимость влияния фактора ***Ф*** на Y при уровне значимости из табл.5.1.

Примечания.

1. В табл. 5.2 представлена совокупность значений параметра Y при 30 величинах фактора *Ф,* причем при каждом уровне *Ф* приводятся 5 значений Y.

2. Каждому из 27 вариантов (табл.5.1) соответствует определенный набор значений Y при 4 номерах уровней фактора Ф и величина уровня значимости.

Таблица 5.1

№ вар.	№ уровня фактора				Уровень значимости α
1	1	3	5	7	0,01
2	2	4	6	8	0,05
3	3	5	7	9	0,01
4	4	6	8	10	0,05
5	5	7	9	11	0,01
6	6	8	10	12	0,05
7	7	9	11	13	0,01
8	8	10	12	14	0,05
9	9	11	13	15	0,01
10	10	12	14	16	0,05
11	11	13	15	17	0,01
12	12	14	16	18	0,05
13	13	15	17	19	0,01
14	14	16	18	20	0.05
15	1	2	3	4	0,01
16	5	6	7	8	0,05
17	9	10	11	12	0,01
18	13	14	15	16	0,05
19	17	18	19	20	0,01
20	1	4	7	10	0,05
21	4	7	10	13	0,01
22	7	10	13	16	0,05
23	10	13	16	19	0,01
24	2	5	8	11	0,05
25	5	8	11	14	0,01
26	8	11	14	17	0,05
27	11	14	17	20	0.01

Таблица 5.2

№ уровня фактора	Значение параметра Y				
1.	17	18	16	20	19
2.	21	18	19	20	22
3.	22	20	23	22	21
4.	22	19	19	15	21
5.	20	15	19	19	20
6.	25	23	26	26	24
7.	27	28	27	26	30

8.	28	27	26	26	24
9.	26	29	30	24	25
10.	29	56	26	27	25
11.	31	58	32	30	30
12.	35	37	36	34	36
13.	42	43	41	40	43
14.	54	53	49	50	49
15.	57	59	58	56	55
16.	63	64	62	67	61
17.	73	70	67	73	68
18.	74	80	76	77	78
19.	81	82	78	84	83
20.	91	85	89	90	92
21.	95	90	99	94	98
22.	103	97	105	106	102
23.	111	105	113	114	112
24.	118	119	120	121	122
25.	124	125	138	123	128
26.	135	130	133	140	137
27.	141	135	139	142	150
28.	146	147	151	152	161
29.	158	161	163	155	171
30.	170	177	160	182	179

Краткие теоретические сведения

Во многих областях практической деятельности встречаются объекты исследования, состояние которых определяется входными переменными (факторами), не имеющими количественного описания. Такими факторами могут быть управляемые и неуправляемые переменные, которые невозможно измерить в данном эксперименте. Для изучения влияния таких факторов на выходной параметр (отклик) объекта и выделения среди них существенных могут использоваться методы дисперсионного анализа.

В качестве примеров применения дисперсионного анализа можно привести следующие.

1 . Рассмотрим процесс измерения рядом операторов какой-либо физической величины несколькими приборами (несколькими методами), причем каждый оператор с помощью каждого прибора производит некоторое число параллельных измерений. Осреднение наблюдений, относящихся к каждому из всех возможных сочетаний прибор-оператор, дает множество средних арифметических. Их рассеивание может быть связано со случайной погрешностью, систематической приборной или методической ошибкой и влиянием оператора. Требуется определить, насколько существенно влияние на результат измерения двух факторов: прибора (метода) и оператора.

2.Аналогичная задача возникает при обработке деталей параллельно на нескольких станках автоматической линии. Требуется установить, однотипны ли средние размеры деталей, изготовляемых на разных станках, т.е. оценить существенно ли воздействует фактор индивидуальности станка на процесс обработки.

3. При использовании радиодеталей из нескольких партий надо определить, существенно ли отличаются параметры деталей различных партий.

4. При нестабильности величины выходного показателя во времени необходимо оценить, насколько существенно влияние медленного временного дрейфа на фоне случайных погрешностей наблюдений.

5. При настройке радио-, теле- и радиолокационной аппаратуры визуальным и акустическим методами несколькими операторами надо оценить влияние на результат настройки трех факторов: аппаратуры, метода и оператора.

Обобщая особенности перечисленных ситуаций, постановку задачи применения дисперсионного анализа в общем виде можно сформулировать так.

Дано: 1) Отклик Y может зависеть (по физическим причинам) от n независимых управляемых факторов $Ф_1, Ф_2,...,Ф_l,..., Ф_n$, не имеющих количественного описаниями их парных взаимодействий;

2) каждый фактор $Ф_l$ может варьироваться на p_l уровнях;

3) полный факторный эксперимент состоит из $N=p_1*p_2*...*p_l *...*p_n$ серий независимых наблюдений по числу всех возможных неповторяющихся сочетаний уровней и факторов;

4)каждая k-тая серия содержит q_k наблюдений $y_{k1}, y_{k2},..., y_{ki},...y_{kq}$ параллельных опытов.

Требуется: Определить, в какой мере существенно на фоне случайных погрешностей влияние того или иного фактора $Ф_e$ или взаимодействия факторов на отклик Y; провести сравнение с другими факторами и выделить наиболее существенные.

Допущения: 1) Результат наблюдения отклика Y-нормально распределенная СВ с центром распределения

$$M(Y)=f(Ф1, Ф2,.....,Фl,Фn),$$

то есть факторы определяют величину Y в среднем, а разброс экспериментальных данных обусловлен случайными причинами;

2) дисперсия единичного наблюдения $\sigma_\varepsilon^2(Y)$, обусловленная случайными причинами ε, одинакова во всех опытах, и не зависит от $Ф_1Ф_2,...,Ф_l,...,Ф_n$ т.е. дисперсии $\sigma_k^2(Y)$ (Y) равны ($k=l,2,..., N$), а их выборочные оценки $s_k^2(Y)$ однородны, что характеризует условие воспроизводимости опытов с равной точностью.

Примечание. Каждое из этих допущений необходимо проверить по результатам наблюдений анализируемого эксперимента.

Идея дисперсионного анализа

Дисперсионный анализ (ДА) предложен в 20-х годах 20-го столетия Р.А.Фишером. Идею ДА удобно проиллюстрировать на примере изучения влияния одного фактора $Ф$ на фоне случайных погрешностей, воздействующих на результат эксперимента Y, причем степень этого воздействия оценивается дисперсией воспроизводимости σ_{δ}^{2}, которая известна. Пусть фактор $Ф$ варьируется на p уровнях, причем при каждом значении уровня $Фj$ производится q измерений, т.е. серии наблюдений равночисленны. Тогда будет получено $q \times p$-результатов наблюдений, которые можно свести в табл. 5.3 (обведено жирной – чертой).

Таблица 5.3

Номер испыта	*Уровни фактора*					
	$Ф_1$	$Ф_2$	...	$Фj$	...	$Ф_p$
1	y_{11}	y_{12}		y_{1i}		y_{1p}
2	y_{21}	y_{22}		y_{2i}		y_{2p}
…						
i	y_{i1}	y_{i2}		y_{ii}		y_{ip}
…						
q	y_{q1}	y_{q2}		y_{qj}		y_{qp}
Групповая средняя	$\bar{y}_1$	$\bar{y}_2$	...	$\bar{y}_I$		$\bar{y}_P$

В табл.5.3 помещаются также средние арифметические из данных повторных наблюдений для каждого j-го уровня фактора:

$$\bar{y}_j = \frac{1}{q}\sum_{i=1}^{q} y_{ij} .$$

Общее среднее арифметическое $\bar{y}$ всех $q \times p$ наблюдений:

$$\bar{y} = \frac{1}{qp}\sum_{j=1}^{p} \sum_{i=1}^{q} y_{ij} = \frac{1}{p}\sum_{j=1}^{p} \bar{y}_j .$$

Рассеяние отдельных наблюдений y_{ij} относительно общего среднего $\bar{y}$ обусловлено действием, как случайных причин, так и влиянием фактора $ф$. Действие фактора случайности проявляется в рассеянии наблюдений серий параллельных опытов y_{ij} при каждом уровне $\boldsymbol{Ф}_j$ вокруг среднего арифметического $\bar{y}_j$ своей серии. Влияние же фактора $\boldsymbol{Ф}$ вызывает повышенное рассеяние средних арифметических $\bar{y}_j$ серий относительно общего среднего $\bar{y}$. Каждое из этих трех рассеяний можно охарактеризовать соответствующей суммой квадратов отклонений.

Общая сумма квадратов отклонений S_0 наблюдений y_{ij} от общего среднего $\bar{y}$:

$$s_0 = \sum_{j=1}^{p} \sum_{i=1}^{q} (y_{ij} - \bar{y})^2 = s_{\delta} + s_{ф},$$

где

$$S_{\delta} = \sum_{j=1}^{p} \sum_{i=1}^{q} (y_{ij} - \bar{y}_j)^2$$

- сумма квадратов отклонений внутри серий и характеризует влияние случайных причин на результат;

$$S_{\phi} = q\sum_{j=1}^{p} (\bar{y}_j - \bar{y})^2$$

- факторная сумма квадратов, характеризующая рассеяние между группами вследствие влияния фактора ***Ф***, т. е. взвешенная с учетом числа q параллельных наблюдений в каждой серии сумма квадратов разностей между средними $\bar{y}_j$ отдельных серий и общим средним по всей совокупности наблюдений $\bar{y}$.

Разделив суммы квадратов отклонений на соответствующее число степеней свободы, можно получить общую, факторную и остаточную дисперсии, соответственно:

$$s_0^2 = \frac{S_0}{pq-1};$$

$$s_{\Phi}^2 = \frac{S_{\Phi}}{p-1};$$

$$s_{\delta}^2 = \frac{S_{\delta}}{p(q-1)}.$$

Примечание. Число степеней свободы *p(q-l)* остаточной дисперсии равно разности между числами степеней свободы общей и факторной дисперсией.

Полученные значения дисперсий используются для оценки значимости влияния фактора ***Ф***. Для того, чтобы влияние ***Ф*** было признано существенным, необходимо и достаточно, чтобы оценка дисперсии s_{Φ}^2 значимо превышала s_{δ}^2. Проверку нуль-гипотезы об однородности этих выборочных дисперсий можно осуществить с помощью критерия Фишера, если $s_{\Phi}^2 > s_{\delta}^2$:

$$F = \frac{s_{\Phi}^2}{s_{\delta}^2}.$$

Если вычисленное по результатам наблюдений дисперсионное отношение $F_{набл}$ превосходит критическое $F_{кр}$ найденное по таблицам распределений Фишера для выбранного уровня: значимости α при соответствующих степенях свободы k_{Φ} и k_{δ}:

$$k_1 = k_{\Phi} = p-1;$$

$$k_2 = k_{\delta} = p(q-1),$$

то влияние фактора ***Ф*** признается существенным. Если же $F_{набл} \le F_{кр}(\alpha, k_{\Phi}, k_{\delta})$, то влияние ***Ф*** несущественно.

Примечания.

1. В дисперсионном анализе проверяют нуль-гипотезу H_0: $M\{s_{\Phi}^2\} = M\{s_{\delta}^2\}$ при альтернативе H_1: $M\{s_{\Phi}^2\} > M\{s_{\delta}^2\}$, поэтому пользуются односторонним F-критерием (см. Прил.4).

2. Дисперсионный анализ позволяет оценивать влияние фактора Ф лишь в

целом; полученные выводы применимы только к данному экспериментальному материалу при данной его систематизации. При изменении условий эксперимента оценка влияния фактора может измениться.

Следует отметить, что идея "ДА" при двух и более факторах аналогична, но реализация более трудоемка и включает процедуры планирования эксперимента и обработки результатов с использованием более сложных таблиц наблюдений и расчетных формул по сравнению с рассмотренными выше.

Методические указания к выполнению работы

1 . Используя исходные данные, соответствующие Вашему варианту, произведите расчет оценок факторной s_{Φ}^2 и остаточной s_{δ}^2 дисперсий по вышеприведенным соотношениям.

2. Проверьте гипотезу H_0 : $M\left\{s_{\Phi}^2\right\}=M\left\{s_{\delta}^2\right\}$ при альтернативе H_1: $M\left\{s_{\Phi}^2\right\}>M\left\{s_{\delta}^2\right\}$, при заданном уровне значимости α. Сделайте выводы о значимости влияния фактора *Ф*.

Контрольные вопросы

1 . Охарактеризуйте ситуации и цели применения ДА.
2. Приведите разновидности ДА.
3. Каковы исходные данные при ДА?
4. Каковы допущения при ДА?
5. Опишите идею ДА.
6. Докажите, что для сумм квадратов справедливо соотношение $S^2{}_0 = S^2{}_{ф} + S^2{}_{б}$.
7. Докажите, что для количеств степеней свободы справедливо соотношение: количество степеней свободы обшей дисперсии $k_0 = k_{\Phi} + k_{\delta}$.
8. Проиллюстрируйте графически найденное значение $F_{набл}$ и критическую область.

Задачи

1. Произведено по четыре испытания на каждом из трех уровней фактора ***Ф***. Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний приведены в табл. 5.4

Таблица 5.4

Номер испытания	Уровни фактора		
i	$Ф_1$	$Ф_2$	$Ф_3$
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34
$\overline{y_{грj}}$	35	25	27

2. Произведено по четыре испытания на каждом из трех уровней фактора ***Ф***. Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний в табл.5.5.

Таблица 5.5

Номер испытания	Уровни фактора		
i	$Ф_1$	$Ф_2$	$Ф_3$
1	35	30	21
2	32	24	22
3	31	26	34
4	30	20	31
$\overline{y}_{rpj}$	32	25	27

3. Произведено по четыре испытания на каждом из трех уровней фактора *Ф*. Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний приведены в табл.5.6.

Таблица 5.6

Номер испытания	Уровни фактора		
i	$Ф_1$	$Ф_2$	$Ф_3$
1	27	24	22
2	23	20	21
3	29	26	36
4	29	30	37
$\overline{y}_{rpj}$	27	25	27

Приложение 1

Критические точки распределения Стьютенда

Число степеней свободы	Уровень значимости α (двусторонняя критическая область)					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.31	12.7	31.82	63.7	318,3	637.0
2	2.92	4.30	6.97	9.92	22.33	31.6
3	2.35	3.18	4.54	5.84	10.22	12.9
4	2.13	2.78	3.75	4.60	7.17	8.61
5	2.01	2.57	3.37	4.03	5.89	6.86
6	1.94	2.45	3.14	3.71	5.21	5.96
7	1.89	2.36	3.00	3.50	4.79	5.40
8	1.86	231	2.90	3.36	4.50	5.04
9	1.83	2.26	2.82	3,25	4.30	4.78
10	1.81	2.23	2.76	3.17	4.14	4.59
11	1.80	2.20	2.72	3.11	4.03	4.44
12	1.78	2.18	2.68	3.05	3.93	4.32
13	1.77	2.16	2.65	3.01	3.85	4.22
14	1.76	2.14	2.62	2.98	3.79	4.14
15	1.75	2.13	2.60	2.95	'3.73	4.07
16	1.75	2.12	2.58	2,92	3.69	4.01
17	1.74	2.11	2.57	2.90	3.65	3.96
18	1.73	2.10	2.55	2.88	3.61	3.92
19	1.73	2.09	2.54	2.86	3.58	3.88
20	1.73	2.09	2.53	185	3.55	3.85
21	1.72	2.08	2,52	2.83	3.53	3.82
22	1-72	2.07	2.51	2.82	3.51	3.79
23	1.71	2,07	2.50	2.81	3.49	3.77
24	1.71	2.06	2.49	2.80	3.47	3.74
25	1.71	2.06	2.49	1.79	3.45	3.72
26	1.71	2.06	2.48	2.78	3.44	3,71
27	1.71	2.05	2.47	2.77	3.42	3.69
28	1.70	2.05	2.46	2.76	3.40	3.66
29	1.70	2.05	2.46	2.76	3.40	3.66
30	1.70	2.04	2.46	2.75	3.39	3.65
	0.05	0.025	0.01	0.005	0.001	0.0005
	Уровень значимости α (односторонняя критическая область)					

Число степеней свободы	Уровень значимости α (двусторонняя критическая область)					
	0.10	0.05	0.02	0.01	0.002	0.001
40	1.68	2.02	2.42	2.70	3.31	3.55
60	1.67	2.00	2.39	2.66	3.23	3.46
120	1.66	1.98	2.36	2.62	3.17	3.37
∞	1.64	1.96	2.33	2.58	3.09	2.29
	0.05	0.025	0.01	0.005	0.001	0.0005
	Уровень значимости α (односторонняя критическая область)					

Приложение 2

Таблица значений функции $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$

x	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*
0,00	0,0000	0,35	0.1368	0,70	0,2580	1,05	03531
0,01	0,0040	0,36	0,1406	0,71	0,2611	1,06	0,3554
0,02	0,0080	0,37	0,1443	0,72	0,2642	1,07	0.3577
0,03	0,0120	0,38	0,1480	0,73	0,2673	1,08	0,3599
0,04	0,0160	0,39	0,1517	0,74	0,2703	1,09	0,3621
0,05	0,0199	0,40	0,1554	0,75	0,2734	1,10	0,3643
0,06	0,0239	0,41	0.1591	0,76	0,2764	1.11	0,3665
0,07	0,0279	0,42	0,1628	0,77	0,2794	1,12	0,3686
0,08	0,0319	0,43	0,1664	0,78	0,2823	1,13	0,3708
0,09	0,0359	0,44	0,1700	0,79	0,2852	1Л4	0,3729
0,10	0,0398	0,45	0,1736	0,80	0,2881	1,15	03749
0,11	0,0438	0,46	0,1772	0,81	0,2910	1,16	03770
0,12	0,0478	0,47	0,1808	0,82	0,2939	1,17	0,3790
0,13	0,0517	0,48	0,1844	0,83	0,2967	1,18	0,3810
0,14	0,0557	0,49	0,18.79	0,84	0,2995	1,19	0,3830
0,15	0,0596	0,50	0,1915	0,85	0,3023	1,20	0,3849
0,16	0,0636	0,51	0,1950	0,86	0,3051	1.21	0,3869
0,17	0,0675	0,52	0,1985	0,87	0,3078	1,22	0,3883
0,18	0,0714	0,53	0,2019	0,88	0,3106	1,23	0,3907
0.19	0,0757	0,54	0,2054	0,89	0,3133	1,24	0,3925
0.20	0,0793	0,55	0,2088	0.90	0,3159	1,25	0,2944
0,21	0,0832	0,56	0,2123	0,91	0,3186	1,26	0,3962
0,22	0,0871	0,57	0,2157	0,92	0,3212	1.27	0,3980
0,23	0,0910	0,58	0,2190	0,93	0,3238	1,28	0,3997
0,24	0,0948	0,59	0.2224	0,94	0,3264	1.29	0,4015
0.25	0,0987	0.60	0.2257	0,95	0,3289	1,30	0,4032
0,26	0,1026	0,61	0,2291	0,96	0,3315	1,31	0,4049
0,27	0,1064	0,62	0,2324	0,97	0,3340	1,32	0,4066
0.28	0,1103	0,63	0,2357	0,98	0,3365	1.33	0,4082
0.29	0,1141	0,64	0,2389	0,99	0.3389	1,34	0,4099

0,30	0,1179	0,65	0,2422	1,00	0,3413	1,35	0,4115
0,31	0,1217	0,66	0.2454	1.,01	0,3438	1.36	0,4131
0.32	0,1255	0,67	0,2486	1,02	0,3461	1,37	0,4147
0,33	0,1293	0,68	0,2517	1,03	0.3485	1,38	0,4162
0.34	0,1331	0,69	0,2549	1,04	0,3508	1,39	0.4177
1,40	0,4192	1,70	0,4554	2,00	0,4772	2,60	0,4953
1,41	0,4207	1,71	0,4564	2,02	0,4783	2,62	0,4956
1,42	0,4222	1,72	0,4573	2,04	0,4793	2,64	0,4959
1,43	0,4236	1,73	0,4582	2,06	0,4803	2,66	0,4961
1,44	0,4251	1,74	0,4591	2,08	0,4812	2,68	0,4963
1,45	0,4265.	1,75	0,4599	2,10	0,4821	2,70	0,4965
1,46	0,4279	1,76	0,4608	2,12	0,4730	2,72	0,4967
1,47	0,4292	1,77	0,4616	2,14	0,4838	2,74	0,4969
1,48	0,4306	1,78	0,4625	2,16	0,4846	2,76	0,4971
1,49	0,4319	1,79	0,4633	2,18	0,4854	2,78	0,4973
1,50	0,4332	1,80	0,4641	2,20	0,4861	2,80	0,4974
1,51	0,4345	1,81	0,4649	2,22	0,4868	2,82	0,4976
1,52	0,4357	1,82	0,4656	2,24	0,4875	2,84	0,4977
1,53	0,4370	1,83	0,4664	2,26	0,4881	2,86	0,4979
1,54	0,4382	1,84	0,4671	2,28	0,4887	2,88	0,4980
1,55	0,4394	1,85	0,4678	2,30	0,4893	2,90	0,4981
1,56	0,4406	1,86	0,4686	2,32	0,4898	2,92	0,4982
1,57	0,4418	1,87	0,4693	2,34	0,4904	2,94	0,4984
1,58	0,4429	1,88	0,4699	2,36	0,4909	2,96	0,4915
1,59	0,4441	1,89	0,4706	2,38	0,4913	2,98	0,4986
1,60	0,4452	1,90	0,4713	2,40	0,491S	3,00	0,49865
1,61	0,4463	1,91	0,4719	2,42	0,4922	3,20	0,49931
1,62	0,4474	1,92	0,4726	2,44	0,4927	3,40	0,49966
1,63	0,4484	1,93	0,4732	2,46	0,4931	3,60	0,49984
1,64	0,4495	1,94	0,4738	2,48	0,4934	3,80	0,49992
1,65	0,4505	1,95	0,4744	2,50	0,4938	4,00	0,49996
1,66	0,4515	1,96	0,4750	2,52	0,4941	4,50	0,49999
1,67	0,4525	1,97	0,4756	2,54	0,4945	5,00	0,49999
1,68	0,4535	1,98	0;4761	2,56	0,4948 .		
1,69	0,4545	1,99	0,4767	2,58	0,4951		

Приложение 3

Критические точки распределения Кочрена

(k_1 -число степеней свободы, k_2 -количество выборок)

Уровень значимости α =0.05							
k_2 \ k_1	1	2	3	4	5	6	7
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332
3	9969	8709	7977	7457	7071	6771	6530
4	9065	7679	6841	6287	5895	5598	5395
5	0.8412	0.6338	0.5981	0.5440	0.5063	0.4783	0.4564
6	7808	6161	5321	4803	4447	4184	3980
7	7271	5612	4800	4307	3974	3726	3535
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185
9	6385	4775	4027	3584	3286	3067	2901
10	6020	4450	3733	3311	3029	2823	2666
12	0.5410	0.3924	03624	0.2880	0.2624	0.2439	0.2299
15	4709	3346	2758	2419	2195	2034	1911
20	3894	2705	2205	1921	1735	1606	1501
24	0.3434	O.2354	0.1907	0.1656	0.1493	0.1374	0.1286
30	2929	1980	1593	1377	1237	1137	1061
40	2370	1576	1259	1082	0968	0887	0827
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583
120	0998	0632	0.0495	0419	0371	0337	0312
∞	0000	0000	0000	0000	0000	0000	0000
Уровень значимости α =0.05							
k_2 \ k_1	8	9	10	16	36	144	∞
2	0,8159	0,8010	0,7880	0,7341	0,6602	0,5813	0,5000
3	6333	6167	6025	5466	4748	4031	3333
4	5175	5017	4884	4366	3720	3093	2500
5	0,4387	0,4241	0,4118	0,3645	0,3066	0,2013	0,2000
6	3817	3682	3568	3135	2612	2119	1667
7	3384	3259	3154	2756	2278	1833	1429
8	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0.1250
9	2768	2659	2568	2226	1820	1446	1111
10	2541	2439	2353	2032	1655	1308	1000
12	0,2187	0,2098	0,2020	0.1737	0,1403	0,1100	0,0833
Уровень значимости α =0.05							
k_2 \ k_1	8	9	10	16	36	144	∞
15	1815	1736	1671	1429	1144	0889	0667
20	1422	1357	1303	1108	0879	0675	0500
24	0,1216	0.1160	0.1113	0,0942	0,0743	0,0567	0,0417
30	1002	0958	0921	0771	0604	0457	0333
40	0780	0745	0713	0595	0462	0347	0250
60	0,0552	0,0520	0,0497	0,0411	0,0316	0,0234	0,0167
120	0292	0279	0266	0218	0165	0120	0083
∞	0000	0000	0000	0000	0000	0000	0000

Приложение 4

Критические точки распределения F Фишера-Снедекора

(k_1 - число степеней свободы большей дисперсии,
k_2 - число степеней свободы меньшей дисперсии)

Уровень значимости α =0.01								
k_1 \ k_2	1	2	3	4	5	6	7	8
1	4052	98,49	34,12	21,20	16,26	13,74	12,25	11,26
2	4999	99,01	30,81	18,00	13,27	10,92	9,55	8,65
3	5403	90,17	29,46	16,69	12,06	9,78	8,45	7,59
4	5625	99,25	28,71	15,98	11,39	9,15	7,85	7,01
5	5764	99,33	28,24	15,52	10,97	8,75	7,46	6,63
6	5889	99,30	27,91	15,21	10,67	8,47	7.19	6,37
т	5928	99,34	27,67	14,98	10,45	8,26	7,00	6,19
8	5981	99,36	27,49	14,80	10,27	8,10	6,84	6,03
9	6022	99,36	27,34	14,66	10,15	7,98	6,71	5,91
10	6056	99,40	27,23	14,54	10,05	7,87	6,62	5,82
11	6082	99,41	27,13	14,45	9,96	7,79	6,64	5,74
12	6106	99,42	27,05	14,37	9,89	7,72	6,47	5,67

Уровень значимости α =0.01									
k_1 \ k_2	9	10	11	12	13	14	15	16	17
1	10,56	10,04	9,86	9,33	9,07	8,86	8,68	8,53	8,40
2	8,02	7,56	7,20	6,93	6,70	6,51	6,36	6,23	6,11
3	6,99	6,55	6,22	5,95	5,74	5,56	5,42	5,29	5,18
4	6,42	5,99	5,67	5,41	5,20	5,03	4,89	4,77	4,67
5	6,06	5,64	5,32	5,06	4,86	4,69	4,56	4,44	4,34
6	5,80	5,39	5,07	4,82	4,62	4,46	4,32	4,40	4,10
7	5,62	5,21	4,88	4,65	4,44	4,28	4,14	4,03	3,93
8	5,47	5.06	4,74	4,50	4,30	4,14	4.00	3,89	3,79
9	5,35	4,95	4,63	4,39	4,19	4,03	3,89	3,78	3,68
10	5,26	4,85	4,54	4,30	4,10	3,94	3,80	3,69	3,59
11	5,18	4.78	4,46	4,22	4,02	3,86	3,73	3,61	3,52
12	5,11	4,71	4,40	4,16	3,96	3,80	3,67	3,55	3,45

(k_1 число степеней свободы большей дисперсии,
k_2- число степеней свободы меньшей дисперсии)

Уровень значимости α =0.05								
k_1 \ k_2	1	2	3	4	5	6	7	8
1	161	18,51	10,13	7,71	6,61	5,99	5.59	5,32
2	200	19,00	9,55	6,94	5,79	5,14	4,74	4,46
3	216	19,16	9,28	6,59	5,41	4,76	4,35	4,07
4	225	19,25	9,12	6,39	5,19	4,53	4,12	3,84
5	230	19,30	9,01	6,26	5,05	4,39	3,97	3,69
6	234	19,33	8,94	6,16	4,95	4,28	3,87	3,58
7	237	19,36	8,88	6,09	4,88	4,21	3,79	3,50
8	239	19,37	8,84	6,04	4,82	4,15	3,73	3,44
9	241	19,38	8,81	6,00	4,78	4,10	3,68	3,39
10	242	19,39	8,78	5,96	4,74	4,06	4,63	3,34
11	243	19,40	8,76	5,93	4,70	4,03	4,60	3,31
12	244	19,41	8,74	5,91	4,68	4,00	4,57	3,28

Уровень значимости α =0.05									
k_1 \ k_2	9	10	11	12	13	14.	15	16	17
1	5,12	4,96	4,84	4,75	4,67	4,60	4,54	4,49	4,45
2	4,26	4,10	3,98	3,88	3,80	3,74	3,68	3,63	3,59
3	3,86	3,71	3,59	3.49	3.41	3,34	3,29	3.24	3,20
4	3,63	3,48	3,36	3,26	3,18	3,11	3,06	3,01	2,96
5	3,48	3,33	3,20	3,11	3,02	2,96	2,90	2,85	2,81
6	3,37	3,22	3,09	3,00	2,92	2,85	2,79	2,74	2,70
7	3,29	3,14	3,01.	2,92	2,84	2,77	2,77	2,66	2,62
8	3,23	3,07	2,95	2,85	2.77	2,70	2,64	2,59	2,55
9	3,18	3,02	2,90	2,80	2,72	2,65	2,59	2,54	2,50
10	3,13	2,97	2,86	2,76	2,67	2,60	2,55	2,49	2,45
11	3,10	2,94	2,82	2,72	2,63	2,56	2,51	2,45	2,41
12	3,07	2,91	2,79	2,69	2,60	2,53	2,48	2,42	2,38

Приложение 5

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} l^{\frac{-x^2}{2}}$

X	Φ(x)									
	0	1	2	3	4	5	6	7	8	9
0.0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0.4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2079	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2.0	0,540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	O084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3.5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3.6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3.8	0003	0003	0005	0003	0003	0002	0002	0002	0002	0002
3.9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Библиографический список

1. Ван дер Варден БЛ. Математическая статистика. – М.: ИЛ, 1960. – 434 с., ил.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: ВШ, 1997. – 479 с., ил.
3. Крамер Г. Математические методы статистики. – М.: ИЛ, 1975. – 648 с., ил.
4. Пугачев В.С. Теория вероятностей и математическая статистика. – М.: Наука, 1979. – 496 с., ил.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1965.
6. Айвазян С.А. Прикладная статистика: основы моделирования и первичной обработки данных. Справ. издание. – М.: Финансы и статистика, 1983. – 417 с., ил.
7. Бикел П.Д. Математическая статистика. – М.: Финансы и статистика, 1983. – 254 с., ил.
8. Дюге Д. Теоретическая и прикладная статистика. – М.: Наука, 1972. – 383 с., ил.
9. Львовский Е.Н. Статистические методы построения эмпирических формул. – М.: ВШ, 1982. – 244 с., ил.
10. Мелник М. Основы прикладной статистики. – М.: Энергоатомиздат, 1983. – 416 с., ил.
11. Пустыльник Б.Н. Статистические методы анализа и обработки наблюдении. – М.: Наука, 1968. – 288 с., ил.
12. Пытьев Ю.П.. Курс теории вероятностей для физиков. – М.: МГУ. – 252 с., ил.
13. Румшиский Л.З. Математическая обработка результатов эксперимента. Справочник. – М.: Наука, 1971. – 192 с., ил.
14. Статистические методы в инженерных исследованиях. – М.: ВШ, 1983. – 216 с., ил. .
15. Статистические методы в экспериментальной физике. – М.: Атомиздат, 1976. – 334 с., ил.
16. Тарасенко Ф.П. Непараметрическая статистика. – Томск: Том. ун-т, 1976.- 2952 с., ил.
17. Тескин О.И. Учебное пособие по применению вероятностных методов и статистических моделей. М., 1983.
18. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. – М.: Статистика, 1980. – 95 с., ил.
19. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: Финансы и статистика, 1995. – 384 с., ил.