

Министерство образования Республики Беларусь

**Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»**

В. М. ЕФИМЕНКО

ЛЕСНАЯ БИОМЕТРИЯ

**ПРАКТИЧЕСКОЕ ПОСОБИЕ
по выполнению лабораторных работ
для студентов специальности
1-75 01 01 «Лесное хозяйство»**

Гомель 2007

УДК 630: 57.087.1(075.8)
ББК 43.4В631.8 Я73
Е-911

Рецензент:
кафедра лесохозяйственных дисциплин учреждения
образования «Гомельский государственный университет имени
Франциска Скорины»

Рекомендовано к изданию на заседании научно-методического
совета учреждения образования «Гомельский государственный
университет имени Франциска Скорины»

Ефименко, В. М.
Е-911 Лесная биометрия : Практическое пособие по
выполнению лабораторных работ для студентов
специальности 1-750101 «Лесное хозяйство» /
В. М. Ефименко; М-во обр. РБ,
Гомельский государственный университет им Ф. Скори-
ны.—
Гомель: ГГУ им. Ф. Скорины, 2007 – 70 с.

Целью подготовки практического пособия является оказание по-
мощи студентам в овладении теоретическими основами курса
«Лесная биометрия» и применении их на лабораторных занятиях и
самостоятельной работе.

Практическое пособие адресовано студентам специальности 1-75
01 01 «Лесное хозяйство».

УДК 630: 57.087.1(075.8)
ББК 43.4В631.8 Я73

© Ефименко В. М., 2007
© УО «ГГУ им.Ф.Скорины», 2007

СОДЕРЖАНИЕ

Введение	4
Тема 1 Составление вариационных рядов.....	5
Тема 2 Статистическая характеристика совокупностей наблюдений.....	12
Тема 3 Функции распределений статистической величины.....	21
Тема 4 Статистическое оценивание.....	30
Тема 5 Статистическая проверка гипотез.....	35
Тема 6 Корреляционный анализ.....	38
Тема 7 Регрессионный анализ.....	42
Тема 8 Дисперсионный анализ.....	46
Тема 9 Статистические функции анализа данных на ПЭВМ.....	52
Литература.....	58
Приложение А.....	59
Приложение Б.....	60
Приложение В.....	61
Приложение Г.....	62
Приложение Д.....	64
Приложение Е.....	67
Приложение Ж.....	68

ВВЕДЕНИЕ

Переход лесного хозяйства Республики Беларусь на самоокупаемость и самофинансирование требует решения задач связанных с организацией систем его ведения, успешное решение которых возможно при условии высококачественной подготовки специалистов.

Специалист лесного хозяйства должен оценивать деревья в лесу не только в качестве источника ресурсов, но и уметь выявлять их роль в средообразующих процессах, формализуя их в статистических и математических взаимосвязях и моделях.

Лабораторные занятия являются составной частью учебного процесса при подготовке специалистов, проводятся согласно учебным планам в сроки, соответствующие графику учебного процесса.

В пособии приводится краткое содержание понятий основных тем курса «Лесная биометрия», цели и хода выполнения лабораторных работ, а также вопросы для самоконтроля.

ТЕМА 1 СОСТАВЛЕНИЕ ВАРИАЦИОННЫХ РЯДОВ

- 1 Образование выборочной совокупности.
- 2 Образование вариационных рядов.
- 3 Графическое отображение вариационных рядов.

Основные понятия по теме

Образование выборочной совокупности. Статистической совокупностью называют некоторое множество относительно однородных предметов или объектов, объединяемых по выбранному признаку.

Теоретически бесконечно большая или приближающаяся к бесконечности совокупность всех единиц или членов вариационного ряда называют *генеральной*.

Генеральная совокупность может состоять из такого большого количества единиц, что изучить их все не представляется возможным. Поэтому приходится иметь дело со сравнительно небольшими, *выборочными* совокупностями.

Выборочная совокупность наиболее полно отражающая свойства генеральной называется репрезентативной, например, при изучении роста деревьев в высоту исключаются деревья, сломанные бурей, поврежденные огнем и т.д.

При образовании выборки используется метод случайного отбора, то есть выдерживается принцип объективности.

Например, если необходимо определить среднюю высоту 20 деревьев, то нельзя их выбирать по своему вкусу. Следует измерять высоту у каждого 10 или 20 и т.д. дерева или предварительно пронумеровав отобрать по таблице случайных чисел.

Образование вариационных рядов. Статистическая совокупность подвергается упорядочиванию, которое заключается в следующем:

а) находится минимальная и максимальная варианты

$$x_{\min} = \min_{1 \leq i \leq n} x(i), x_{\max} = \max_{1 \leq i \leq n} x(i)$$

б) весь диапазон значений признака [от X_{\min} , до X_{\max}] разбивается на «к» интервалов одинаковой протяженности $h = (x_{\max} - x_{\min}) / k$.

Величину интервала обычно определяют по формуле:

$$k = \frac{X_{\max} - X_{\min}}{\text{от } 8 \text{ до } 12}, \text{ где}$$

X_{\max} и X_{\min} - соответственно наибольшее и наименьшее значение признака или вариант. Числа от 8 до 12 означают предварительно принимаемое количество классов.

В качестве k принимают круглое число, ближайшее к полученному частному. При этом действительное число классов определится как частное от деления размаха вариант ($X_{\max} - X_{\min}$) на округленное значение интервала. В качестве полученного частного принимают целочисленную величину. Округление делается всегда в большую сторону. Хороший результат вычисления величины интервала получается при применении формулы:

$$k = \frac{X_{\max} - X_{\min}}{1 + 3,322 \lg N}.$$

в) Находятся граничные точки каждого из интервалов. Границы и срединные значения классов устанавливаются следующим образом.

В качестве среднего значения первого класса принимают число k , ближайшее к наименьшей (в возрастающем ряду) или наибольшей (в убывающем ряду) вариантам ряда распределения.

Например, для диаметров, при составлении возрастающего ряда (14,0-17,9; 18,0-21,9 и т.д.), таким числом оказывается 16,0. Нижние границы классов определяют путем вычитания половины величины интервала из срединных значений каждого класса, а верхние границы - путем прибавления этой половины. Для первого класса получим соответственно числа 14,0 и 18,0; для второго - 18,0 и 22,0.

Срединные значения последующих классов получают путем последовательного прибавления величины интервала. Для ряда диаметров они окажутся равными: 20,24, 28 и т.д. Эти значения называют классовыми вариантами, так как они представляют собою классы.

В целях исключения перекрытия верхней границы предыдущего класса с нижней границей последующего класса, (например, чисел 18,0 и 18,0, входящих в первый и второй классы), нижние границы классов увеличивают на величину, равную точности измерения признака (можно верхние границы классов уменьшить на ту же величину и таким образом получить значения границ классов: для диаметров она равна 0,1 см, для высот 0,1 м.

г) Затем подсчитывается число вариант N_i , попавших в интервал, причем варианты, попавшие на границы интервалов, относят только к одному из интервалов, результат заносится в таблицу:

Классы	Частоты	Классы	Частоты

Пример:

Границы классов, см	Срединные значения классов, см	Частоты (число стволов)	
		в рабочей записи	в цифрах
14,1-18,0	16	: :	4
18,1-22,0	20	□	7
22,1-26,0	24	□	8
26,1-30,0	28	⊠ □	28
30,1-34,0	32	⊠ ⊠	20
34,1-38,0	36	⊠ □	18
38,1-42,0	40	⊠	9
	Σ		94

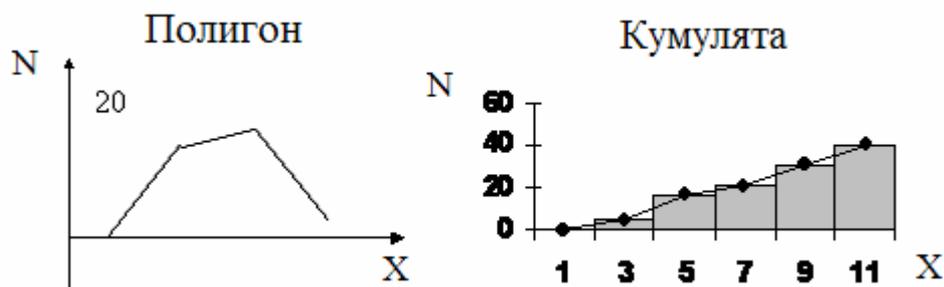
Графическое изображение вариационных рядов. Обычно табличное распределение частот дополняют его графическим представлением. Схематически все множество графических представлений статистических данных разделяют на два класса: диаграммы и линейные изображения. К классу линейных графиков относятся полигон, кумулятивная кривая, огива.

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1, N_1), \dots, (x_k, N_k)$.

Кумулятивная кривая (кривая сумм) - ломаная, составленная по последовательно суммированным, т.е. накопленным частотам или относительным частотам. При построении кумулятивной кривой дискретного признака на ось абсцисс наносятся значения признака, а ординатами служат нарастающие итоги частот. Соединением вершин ординат прямыми линиями получают **кумуляту**. Кумулятивную кривую называют полигоном накопленных частот.

Если на ось ординат нанести значение признака, а накопленные частоты - на ось абсцисс, то получим кривую, называемую **огивой**.

Пример графического отображения статистических данных



Диаграммы. Диаграмма (от греческого *diagramma* - изображение, чертеж, рисунок) - это графическое изображение, наглядно показывающее соотношение между сравниваемыми величинами. Диаграммы бывают различных видов: полосовые (ленточные), столбиковые, круговые и т. д.

Полосовые - особенно наглядны при сравнении величин, связанных между собой в единое целое. Ширина полос должна быть одинаковой. По длине полосы разбиваются на части, пропорциональные изображаемым величинам.

Гистограмма частот является видом столбиковых диаграмм представляющая ступенчатую фигуру, состоящую из прямоугольников, основанием которых служат частотные интервалы длины h , а высоты равны отношению N_i/h - плотность частоты. Гистограмма относительных частот - аналог плотности распределения непрерывной случайной величины.

Для построения гистограммы частот на оси абсцисс откладывают классовые интервалы, а над ними на расстоянии N_i/h проводят отрезки параллельные основанию, см. рисунок.

Масштабированной называют диаграмму, в которой значения частот обозначены над столбиками и над осью ординат указывается размерность масштабирования (например, разделить на h).

Каждое значение изучаемого показателя изображается в виде вертикального столбика. Количество столбиков определяется числом классов изучаемых показаний (данных). Расстояние между столбиками должно быть одинаковым. У основания столбиков приводится название изучаемого показателя.

Пример построения гистограммы

Срединное значение классового интервала	Частоты			Графическое изображение ряда
	абсо- лютные	относи- тельные	накоп- ленные	
14	12	6	12	
16	24	12	36	
18	45	22,5	81	
20	35	17,5	116	
22	26	13	142	
24	21	10,5	163	
26	10	5	173	

Структурная диаграмма - позволяет сопоставить статистические совокупности по составу.

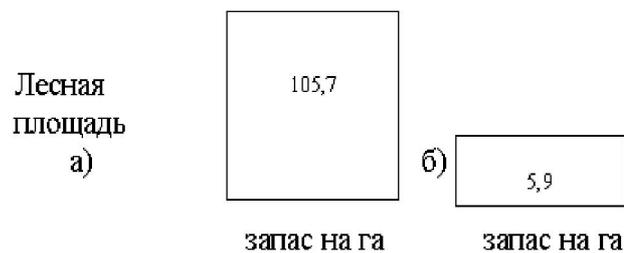
При построении квадратных и круговых диаграмм площади квадратов или кругов выражают изображаемые величины.

Секторная диаграмма строится таким образом, чтобы каждый сектор занимал площадь круга пропорционально удельному весу отображаемых частей целого, исходя из соотношения - $1\% = 3,6$ градуса.

Знак Варзара. - Известный русский статистик В. Е. Варзар предложил использовать прямоугольные фигуры для графического изображения трех показателей, один из которых является произведением двух других. В каждом таком прямоугольнике основание пропорционально одному из показателей — сомножителей, а высота его соответствует второму показателю — сомножителю. Площадь прямоугольника равна величине третьего показателя, являющегося произведением двух первых. Располагая рядом несколько прямоугольников, относящихся к разным показателям, можно сравнивать не только размеры показателя — произведения, но и значения показателей — сомножителей.

Пример. Запас лесного фонда равен произведению запаса (m^3) на единице площади (га) и общей занимаемой площади (тыс. га).

Общий запас лесного фонда, тыс. m^3	Запас древесины на единице площади, m^3 на га	Занимаемая насаждениями площадь, га
105,7	100	1057
5,9	65	91



а - молодняки,
б - средневозрастные.

На этом графике можно сравнить между собой:

- * запас древостоя на единице площади (по длине основания);
- * занимаемая насаждениями площадь (по длине боковой стороны);
- * общий запас лесного фонда (по площади прямоугольника).

При использовании ПЭВМ описанные диаграммы и иной структуры строятся при помощи «мастер диаграмм» в приложении EXCEL.

Вопросы для самоконтроля

- 1 Что такое выборочная и сгруппированная статистическая совокупность?
- 2 Как образуется классовый интервал?
- 3 Что такое кумулята, огива и как они отображаются графически?
- 4 Как построить полигон распределения статистических частот, гистограмму, диаграмму?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин; миллиметровка; калькуляторы.

Цель. Образование выборочных, сгруппированных статистических совокупностей и их графическое отображение.

Ход работы

1 Образование выборочной совокупности. На основании совокупности замеров диаметров и высот сосны (полученной от преподавателя) и таблицы случайных чисел (приложение А) образовать выборочную частичную совокупность, которая приводится в следующей форме:

Порядковый номер в вариационном ряду 1.....50	Порядковый номер отобранных деревьев	Значение признака	
		Замер диаметра, (Д) см	Замер высоты, (Н) м

Просматривается с любого места таблица случайных чисел (Приложение А) в пределе до числового значения - 150, выбираются и заносятся в таблицу значения, соответствующие данному числовому интервалу в количестве 50 штук.

2 Образование вариационных рядов. На основании совокупности с 150 значениями диаметров и высот и основных положений по теме определяется размерность классов и проводится разноска частот вариантов по классам диаметров и высот.

3 Графическое изображение вариационных рядов. На основании сгруппированных вариационных рядов на миллиметровой бумаге вычерчивается (либо строится при помощи ПЭВМ) полигон частот диаметров и высот, полигон накопленных частот диаметров и высот, столбиковая диаграмма, кумулята, огива.

ТЕМА 2 СТАТИСТИЧЕСКАЯ ХАРАКТЕРИСТИКА СОВОКУПНОСТЕЙ НАБЛЮДЕНИЙ

- 1 Статистические показатели малой выборочной совокупности
- 2 Статистические показатели сгруппированных данных

Основные понятия по теме

Статистические показатели малой выборочной совокупности. Наиболее распространенным и используемым показателем является ***среднеарифметическая величина***. Она рассчитывается по формуле:

$$\bar{X} = (x_1 + x_2 + \dots + x_n) / N = (\sum x_i) / N,$$

где N - общее число вариантов; Σ - знак суммирования;
 x_i - значения вариантов.

Средняя квадратическая величина используется при вычислении средней из величин объема, запаса, площади. Рассчитывается по формуле:

$$\bar{X}_q = \sqrt{(\sum x_i^2) / N}$$

где x_i^2 - квадраты измеряемых величин – объем, площадь и т. п.; N - общее число деревьев в выборке; x_i - значения вариантов.

Средняя геометрическая M_g (или \bar{X}_g) используется для расчета среднего темпа роста изучаемого признака. Она известна также как средняя логарифмическая, так как ее логарифм есть арифметическая средняя логарифмов составляющих величин.

Вычисляется по формуле:

$$\bar{X}_g = \sqrt{x_1 \cdot x_2 \cdot x_3 \dots x_n},$$

где x_1, x_2, \dots, x_n - темпы роста (величины, показывающие, во сколько раз увеличивался признак от периода к периоду); n - число периодов.

При $n > 2$ формулу удобнее применять в логарифмическом виде:

$$\ln \bar{X}_g = \frac{1}{n} (\ln x_1 + \ln x_2 + \dots + \ln x_n) = (1/n) \cdot (\sum \ln x_i).$$

откуда $\bar{X}_g = e^{\ln}$, где e – основание натуральных логарифмов, равно 2,72.

Средняя гармоническая используется для вычисления средней величины отношений двух варьирующих величин. Она определяется по формуле:

$$\bar{X}_h = N / (\sum 1/x_i),$$

где N - число значений; x_i – значения соотношений величин
 Рассеяние вариант выборки относительно средней характеризуется:

линейным отклонением;

среднеквадратическим отклонением - дисперсией;

основным отклонением;

коэффициентом изменчивости.

Их вычисляют по формулам:

линейное отклонение $\alpha = x_i - \bar{x}$,

среднеквадратическое отклонение (дисперсия)

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^k (x_i - \bar{x})^2$$

основное отклонение $D = \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^k (x_i - \bar{x})^2}$

коэффициент изменчивости $C_x = \frac{\sigma_x \cdot 100}{\bar{x}}$

При $C_x \leq 30\%$ - выборка имеет большую степень концентрации вариант возле величины. При $30\% \leq C_x \leq 100\%$ - степень концентрации допустимая. При $C_x > 100\%$ - делается вывод о неоднородности выборки.

Статистические показатели сгруппированных данных. Для упрощения расчетов показателей сгруппированного статистического ряда применяется система так называемых моментов.

Моменты случайной величины являются обобщенным понятием числовой характеристики случайных величин. Оно широко применяется в механике для описания распределения масс (статистические моменты, моменты инерции и т.д.). Совершенно теми же приемами пользуются в теории вероятностей для описания

основных свойств распределения случайной величины. Моментом статистической величины называется среднее из отклонений индивидуальных значений признака от определенной величины.

Различают моменты двух видов: начальные и центральные.

В общем случае моменты распределения вариационного ряда определяются выражением
$$\mu_p = \sum_{i=1}^k n_i \cdot (x_i - a)^p$$

В зависимости от значения «а» общая схема моментов порядка «р» разбивается на подсистемы.

а=0, получаем систему начальных моментов

$$\mu_{p,0} = \frac{1}{N} \sum_{i=1}^k n_i x_i^p$$

а=x, получаем систему центральных (основных) моментов

$$\mu_{p,0} = \sum_{i=1}^k n_i (x_i - \bar{x})^p, \text{ где } n_i - \text{ веса отдельных значений; } N$$

– общее число значений.

а=const, получаем систему условных моментов.

Показатель степени определяет порядок момента - первая степень показывает, что рассчитывается момент первого порядка, вторая степень – рассчитывается момент второго порядка и т. д. В большинстве случаев используются моменты четырех порядков.

Расчет начальных моментов по способу произведений. По данным разности численности вариационного ряда диаметров и высот по классам проводим поэтапные **расчеты моментов по способу произведений**, которые приводятся в таблице:

Пример расчета моментов по способу произведений

X	n	A _x	a	a ²	a·n	a ² ·n	a ³ ·n	a ⁴ ·n
77	4		-3	9	-12	36	-107	324
91	6		-2	4	-12	24	-48	96
105	15		-1	1	-15	15	-15	15
119	27	119	0	0	0	0	0	0
133	16		1	1	16	16	16	16
147	10		2	4	20	40	80	160
161	5		3	9	15	45	135	405
175	0		4	16	0	0	0	0
189	2		5	25	10	50	250	1250
	Сумма	Сумма верхней части			-39	75	-171	435
	N=85	Сумма нижней части			61	151	481	1831
		Сумма алгебр. d =			22	226	310	2266

В этой таблице A_x – условное среднее значение, за которое принимается среднее значение класса с наибольшим числом вариант, а – начальное отклонение – рассчитывается по формуле:

$a = \frac{X - Ax}{k}$, где k – величина классового интервала (размер класса); X – среднее значение класса; Ax – условное среднее.

Начальный момент первой степени равен:

$$m_1 = \frac{\sum a \cdot n}{N} = 22/85 = 0.2588$$

Начальный момент второй степени равен:

$$m_2 = \frac{\sum a^2 \cdot n}{N} = 226/85 = 2,659$$

Начальный момент третьей степени равен:

$$m_3 = \frac{\sum a^3 \cdot n}{N} = 310/85 = 3,647$$

Начальный момент четвертой степени равен:

$$m_4 = \frac{\sum a^4 \cdot n}{N} = 2266/85 = 26,659$$

В представленной схеме первые две колонки – вариационный ряд. Третья колонка содержит условную среднюю величину, за которую обычно принимается одно из средних значений классов, ближайшее по значению к средней.

Расчет начальных моментов по способу сумм. Первые три колонки те же, что и в предыдущей таблице.

Пример расчета моментов по способу сумм

x	n	A_x	S_1	S_2	S_3	S_4
77	4		4	4	4	-
91	6		10	14	-	-
105	15		25	-	-	-
119	27	119	-	-	-	-
133	16		33	-	-	-
147	10		17	28	-	-
161	5		7	11	17	-
175	0		2	4	6	8
189	2		2	2	2	2
	Сумма верхней части		-39	-18	-4	0
	Сумма нижней части		61	45	25	10
	Сумма алгебр. d		22	27	21	10
	Сумма арифм. s		100	63	29	10

При этом важно, чтобы условное среднее было записано в строке избранного класса. (в нашем примере оно равно 119 и записано против четвертого класса). Затем в схеме наносятся прочерки в клетки, суммы которых исключаются.

Последовательное суммирование проводится вначале сверху, а затем снизу. Частоту первого класса переносят в колонку S_1 . Далее эту частоту суммируют с частотой второго класса и т. д.

Все суммы в таблице выше условной середины – отрицательные, а ниже – положительные.

Расчеты моментов производятся в соответствии с приведенными формулами, где d – алгебраические суммы, а S – суммы арифметические.

Начальный момент первой степени равен:

$$m_1 = \frac{d_1}{N} = 22/85 = 0,2588.$$

Начальный момент второй степени равен:

$$m_2 = \frac{S_1 + 2 \cdot S_2}{N} = 100 + 2 \cdot 63/85 = 2,659$$

Начальный момент третьей степени равен:

$$m_3 = \frac{d_1 + 6d_2 + 6d_3}{N} = (22 + 6 \times 27 + 6 \times 21)/85 = 3,647$$

Начальный момент четвертой степени равен:

$$m_4 = (S_1 + 14S_2 + 36S_3 + 24 S_4)/N = (100 + 14 \cdot 63 + 36 \cdot 29 + 24 \cdot 10)/85 = 26,659.$$

Для характеристики формы распределения статистических величин используют **центральные моменты**, которые рассчитываются по начальным моментам.

Центральный момент первой степени всегда равен нулю, поскольку вычисляется как $\mu_1 = m_1 - m_1$.

Центральный момент второй степени рассчитывается через начальные моменты по формуле: $\mu_2 = m_2 - m_1^2$.

Центральный момент третьей степени рассчитывается через начальные моменты по формуле: $\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$.

Центральный момент четвертой степени рассчитывается через начальные моменты по формуле: $\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$.

Отклонение распределений фактических данных от нормального типа характеризуется **основными моментами** - r_3, r_4 , которые показывают асимметричность **коэффициент асимметрии** - A и крутость **коэффициент эксцессов** распределений - E :

$$A = r_3 = \mu_3 / (\sqrt{\mu_2})^3; \quad E = r_4 - 3 = \mu_4 / (\mu_2)^2 - 3$$

При помощи моментов статистических величин рассчитываются также остальные характеристики вариационного ряда.

Средняя рассчитывается по формуле: $\bar{X} = A_x + k \cdot m_1$

Дисперсия учитывает отклонение каждого из значений выборки от среднеарифметической величины: $S = k \mu_2$.

Основное отклонение представлено в тех же единицах измерения, что и средняя величина и рассчитывается в двух вариантах – неполное без учета размера класса

$$\sigma' = \sqrt{\mu_2} \text{ и полное с учетом размера класса } \sigma = k \cdot \sqrt{\mu_2}.$$

Величина **коэффициента вариации** является относительным выражением основного отклонения, которая позволяет судить о степени изменчивости вне зависимости от абсолютных значений выборки:

$$C_x = \frac{\sigma_x \cdot 100}{\bar{X}}.$$

Вычисление смешанных моментов позволяет характеризовать степень взаимообусловленных изменений наблюдений относительно друг друга.

Исходными данными для вычисления смешанного начального момента первого порядка m_{1x} служит таблица распределения диаметров и высот, которая составляется параллельно с рядами распределений.

Разноску наблюдаемых значений признаков производят сразу по обоим признакам. Каждую варианту в клетке, образуемой пересечением строки и столбца, соответствующих значениям варианты по обоим признакам (Приложение Б), обозначают постановкой точки или черточки, как при составлении ряда распределения. Если диаметр ствола сосны равен 20,8, а высота 28,0 м, варианта должна быть занесена в клетку, образуемую 3- строкой и 2-м столбцом таблицы. После разnosки всех вариантов и обозначения их в цифровой записи подводят итоги частот по каждой строке и столбцу. В таблице процесс разnosки вариантов и обозначения их точками не показаны. Приведены итоги распределения, показанные в цифрах.

Оформление таблицы распределения заканчивается вычислением эмпирических средних значений зависимого признака (высоты) по классам независимого (диаметра). Для этого сначала в пределах каждого класса независимого признака находят суммы произведений срединных значений классов зависимого признака на соответствующие им частоты, помещенные в корреляционной таблице. Найденные по каждому разряду суммы произведений (они вписаны во 2-й строке снизу) делят на общее количество частот класса.

Полученные в результате этого деления средние значения зависимого признака по классам независимого, то есть средние эмпирические высоты, вписаны в 3-й строке снизу.

Смешанный начальный момент вычисляется непосредственно на основе таблицы распределения соотношений диаметров и высот. Для нахождения моментов в таблицу с исходной информацией вводят 5 дополнительных столбцов и 7 10 дополнительных строк.

В первую дополнительную строку вписывают отклонения классовых вариантов ряда независимого признака X от условной средней

величины этого ряда M' , выраженные в рабочих единицах, то есть поделенные на величину классового промежутка (k_x).

$$X_k = (X - M'_x) / k_x$$

Во вторую дополнительную строку вписывают произведения частот на условные отклонения ($n_x x_k$) и в третью – произведения частот на квадраты условных отклонений ($n_x x_k^2$). В следующую строку по каждому частному ряду распределения, то есть по каждому частному ряду распределения, то есть по каждому столбцу, вписывают произведения отклонений классовых вариантов ряда зависимой переменной от их условной средней величины на частоты каждого частного ряда зависимой.

Для класса $X=16$, например значения этих произведений получают так: условное отклонение y_k (в первом дополнительном столбце и восьмой строке таблицы), равное -3 , умножают на численность $n_{yx}=1$; условное отклонение (в 9-й строке таблицы), равное -4 , умножают на численность $n_{yx}=1$; условные отклонения 10-й и 12-й строк таблицы, соответственно равные -5 и -7 , умножают на численности, равные 1. Полученные произведения -3 , -4 , -5 , -7 суммируют. Сумма произведений -19 .

В пятую дополнительную строку вписывают значения произведений условных отклонений ряда независимой переменной x_k на алгебраические суммы произведений ($n_{yx} y_k$), вписанные в предыдущей строке.

В шестую строку таблицы вписывают значения средних квадратов отклонений частных рядов зависимого признака по классам независимого. Для первого частного ряда (1-й столбец) средний квадрат

$$[(n_{yx} y_k)^2] / n_x = (-19)^2 / 4 = 90,2.$$

В седьмую (самую нижнюю строку) вписывают средние значения зависимого признака \bar{y}_i по каждому классу независимого. Эти значения впервые вычислены в процессе составления рабочей таблицы.

Содержание дополнительных пяти столбцов таблицы аналогично содержанию рассмотренных строк без двух последних. Оно ясно из символов, указанных в заголовках столбцов. Столбцы эти служат для проверки расчетов, произведенных в строках. Суммы третьей строки снизу и последнего столбца должны совпадать (см. число 225).

После проверки этих сумм и нахождения сумм других строк и столбцов вычисляют начальные моменты и основные отклонения. Формулы и последовательность расчетов приведены в нижней части таблицы справа.

В этих формулах:

m_{1x} и m_{2x} - первый и второй начальные моменты ряда распределения независимого признака;

m_{1y} m_{2y} - 1 и 2-й начальные моменты ряда зависимого признака;

m_{1xy} - момент произведения отклонений;

$m_{2y/x}$ - средний квадрат условных произвольных отклонений частных средних ряда y по классам x (второй момент);

σ'_x , σ'_y - основные отклонения вариантов, соответственно ряда x и y от их средних \bar{x} и \bar{y} , выраженные в долях интервалов.

При помощи смешанных моментов вычисляются следующие статистические показатели.

Коэффициент корреляции является мерой тесноты связи между двумя величинами (например, D и H). Он может вычисляться при малом и большом числе наблюдений, изменяется в пределах от -1 до $+1$. Вычисляется по формуле:

$$r = (m_{1xy} - m_{1x} m_{1y}) / \sigma'_x \sigma'_y$$

Ошибка коэффициента вычисляется по формуле:

$$m_r = 1 - r^2 / \sqrt{N}.$$

Корреляционное отношение. В тех случаях, когда зависимость между величинами криволинейная, вычисляется корреляционное отношение η . Оно изменяется в пределах от 0 до $+1$ и отрицательным не может быть. Если корреляционное отношение стремится к $+1$, то это означает, что криволинейная корреляция стремится к функциональной зависимости. Корреляционное отношение всегда по своей абсолютной величине больше коэффициента корреляции.

Вычисляется по формуле: $\eta = \sqrt{\frac{(m_{2y/x} - m_{1y}^2)}{\sigma'_y}}$

Вопросы для самоконтроля

- 1 Что такое статистические моменты и как они вычисляются?
- 2 Что такое основное отклонение, дисперсия, коэффициент вариации и как они вычисляются?
- 3 Что такое средняя арифметическая, средняя квадратическая, средняя гармоническая, средняя геометрическая и как они вычисляются?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 1; калькуляторы.

Цель: Определение статистических показателей выборочной и сгруппированной совокупности

Ход работы

1 На основе основных положений темы определить: среднюю арифметическую, среднюю квадратическую, среднюю геометрическую, среднюю гармоническую. В качестве отдельных вариантов используются значения диаметров или высот. Промежуточные показатели расчетов располагают в виде таблицы.

x_i (высота или диаметр)	x_i^2	$\ln x_i$	$1/x_i$	$\alpha=(x_i-\bar{x})$	α^2
ΣX_i	ΣX_i^2	$\Sigma(\ln X_i)$	Σ	Σ	Σ

2 На основе основных положений темы и данных выборочной совокупности определить: среднеквадратическое отклонение, дисперсию, коэффициент вариации.

3 На основе основных положений темы и данных сгруппированной совокупности найти начальные статистические моменты по способу произведений, по способу сумм.

4 На основе основных положений темы и данных сгруппированной совокупности найти центральные (основные) моменты рядов распределений по диаметру и высоте.

5 На основе основных положений темы и данных сгруппированной совокупности найти значение средней арифметической, дисперсии, основного отклонения, асимметрии и эксцесса.

6 На основе основных положений темы и данных сгруппированной совокупности по высоте и диаметру найти смешанные моменты двух случайных величин и степень их взаимозависимости – коэффициента корреляции, корреляционного отношения.

ТЕМА 3 ФУНКЦИИ РАСПРЕДЕЛЕНИЙ СТАТИСТИЧЕСКОЙ ВЕЛИЧИНЫ

- 1 Как осуществить расчет биномиального распределения?
- 2 Как осуществить расчет распределения Пуассона?

Основные понятия по теме

Определение типа распределения частот вариационного ряда. Показатели частот вариационных рядов по диаметру либо по высоте являются отражением частного случая распределения – для конкретного насаждения, в котором производился обмер деревьев. Если взять близкое по характеристикам насаждение, то его распределение деревьев обязательно совпадает с полученными аналитическими данными.

При решении задач связанных с обработкой лесоводственной информации часто необходимо выявлять принадлежность кривой отражающей распределение частот наблюдаемых экспериментальных значений к определенному типу.

Наиболее распространенным типом распределения является нормальное – которое характеризуется колокообразной формой с осью симметрии совпадающей со среднеарифметической величиной. При отклонении от оси симметрии вправо или влево используются другие типы распределений.

Тип распределения выбирается в зависимости от величины асимметрии и эксцесса. При нормальном распределении величина асимметрии стремится к нулю, а эксцесса к трем. Если значение асимметрии лежит в пределах 0,4-0,8, то расчет распределения ведут по типу А, при больших абсолютных значениях асимметрии рассчитывают распределение типа В (кривая Грама-Шарлье), при наличии максимальной частоты в выборке в одном из крайних классов (первый или последний) рассчитывают распределение Пуассона, при симметричном расположении частот – биномиальное распределение.

Определение проводится при помощи критерия Пирсона. Конкретный тип кривой устанавливают в зависимости от величины критерия - χ (хи) , который называют критерием Пирсона и определяют через величины r_3 и r_4 (основные моменты):

$$\chi = \frac{r_3^2(r_4 - 3)}{4 \cdot (4r_4 - 3r_3^2) \cdot (2r_4 - 3r_3^2 - 6)}$$

где r_3 и r_4 – третий и четвертый основные моменты,
 Тип кривых определяемых критерием Пирсона

Тип кривой	Характерные признаки кривой	χ	Теоретический аналог распределения
I	ограничена с обеих сторон, асимметрична	$\chi < 0$	бета-распределение
II	ограничена с обеих сторон, симметрична	$\chi = 0$	
III	ограничена слева, асимметрична	$\chi = \infty$	при $ \chi > 4$, – гамма-распределение
IV	не ограничена по обоим сторонам, асимметрична	$0 < \chi < 1$	
V	ограничена с одной стороны, асимметрична	$\chi = 1$	
VI	ограничена с одной стороны, асимметрична	$1 < \chi < \infty$	
VII	не ограничена по обоим сторонам, симметрична	$\chi = 0$	
VIII	не ограничена по обоим сторонам, симметрична	$\chi = 0$	нормальное распределение

Расчет биномиального распределения. Соответствие эмпирического распределения биномиальному определяют разными способами в зависимости от величины частоты вариантов. Если величина частоты больше 10, то теоретические частоты биномиального распределения получают с достаточной точностью при помощи формулы частот нормального распределения. Если частоты меньше 10, то лучше воспользоваться формулой разложения бинома, отдельные члены которого представляют собой теоретические частоты биномиального распределения.

Теоретические частоты биномиального распределения по схеме нормального распределения вычисляются по формуле:

$$f' = \frac{(N_n \cdot k \cdot \varphi(t))}{\sigma}, \text{ где}$$

f' - теоретические частоты биномиального распределения;
 N_n - сумма частот;

$\varphi(t)$ - функция от t равна $0,39894/e^{0.5 \cdot t^2}$ (см. приложение В);

t – нормированное отклонение $(x - \bar{X})/\sigma$;

σ – основное отклонение

x – номера градаций классов статистического признака;

\bar{X} – средняя арифметическая;

f – эмпирические частоты;

k – классовый интервал.

Теоретические частоты вычисляются согласно действиям по приведенному примеру в таблице в следующем порядке.

Перемножаем номера классов из столбца 1 на соответствующие им частоты из столбца 2, результаты записываем в столбце 3.

Числа столбца 3 умножаем на числа столбца 1, результаты записываем в столбце 4.

Суммируем числа столбцов 2-4, получаем итоги

$$N_n = 120, \quad \sum(x \cdot f) = 585, \quad \sum(x^2 \cdot f) = 3159.$$

Вычисляем среднюю арифметическую $\bar{X} = 585 / 120 = 4,88$.

Вычисляем среднее квадратическое и основное отклонение

$$\sigma^2 = \sqrt{\frac{((\sum(x^2 \cdot f) - (\sum(f \cdot x))^2 / N_n))}{N - 1}}$$

$$\sigma^2 = (3159 - (585^2 / 120)) / 120 - 1 = 2,581$$

$$\sigma = \sqrt{2,581} = 1,607$$

Дисперсия $\sigma^2 = 2,581$ меньше чем средняя 4,88, поэтому в данном случае можно применить биномиальное распределение.

Разности средней арифметической с номерами классов (без учета знаков) делим на $\sigma = 1,607$, получаем значения нормированных отклонений в столбце 5.

По значению t (без учета знака) в таблице приложения В находим значение $\varphi(t)$ и выписываем их в столбец 6.

Находим значение выражения $(N_n \cdot k) / \sigma = (120 \cdot 1) / 1,607 = 74,67$

Перемножаем числа из столбца 6 на полученное значение и, округляя до целых чисел, получаем в столбце 7 теоретические частоты кривой биномиального распределения, уравнение которой имеет вид:

$$f' = 74,67 \varphi(t), \quad \text{где } t = (x - 4,88) / 1,607.$$

Пример расчета биномиального распределения

х – номер классов	f – эмпирические частоты	х · f	х ² · f	t = (х - \bar{X}) / σ	$\varphi(t)$	$f' = N_n \cdot k \cdot \varphi(t) / \sigma$
1	4	4	4	2,41	0,02186	2
2	7	14	28	1,79	0,08038	6
3	8	24	72	1,17	0,20121	15
4	23	92	368	0,55	0,34294	26
5	40	200	1000	0,07	0,39797	30
6	23	138	828	0,7	0,31225	23
7	9	63	441	1,32	0,16694	12
8	4	32	256	1,94	0,06077	5
9	2	18	162	2,56	0,01506	1
Σ	120	585	3159			119

Расчет распределения Пуассона. Распределение Пуассона характерно для событий, вероятность которых очень мала. Также как и при биномиальном распределении, эмпирические частоты распределения Пуассона являются числом наблюдений по градациям наблюдаемого признака. В распределении средняя арифметическая примерно равна дисперсии: $\bar{X} \approx \sigma^2$, что является основным его признаком.

Теоретические частоты распределения Пуассона вычисляют по формуле:

$$f' = \left(\frac{\bar{X}^x}{x!} \right) \cdot N_n \cdot e^{-\bar{X}}, \text{ где}$$

f' - теоретические частоты распределения Пуассона;

x – градации значений наблюдаемого признака;

$x!$ – (икс-факториал) обозначает произведение ряда натуральных чисел, например: $3! = 1 \cdot 2 \cdot 3 = 6$;

\bar{X} – средняя арифметическая данного ряда;

N_n - общее число наблюдений;

$e^{-\bar{X}}$ - основание натурального логарифма (2,7182818) в степени \bar{X} .

Пример расчета частот проводим в таблице. Средняя арифметическая находится по формуле: $\bar{X} = \frac{\sum x \cdot f}{N_n}$. Для приведенного примера она равна 0,672. Возводим последовательно \bar{X} в степень равную величинам градаций классов. Затем вычислим факториалы от x . Находим последовательно частное $\frac{\bar{X}^x}{x!}$. Вычисляем выражение $N_n \cdot e^{-\bar{X}}$ равное для данного примера 62,43. Полученное частное последовательно перемножаем на 62,43 и записываем с округлением до целого в качестве найденных частот. Оценка совпадения эмпирических частот с теоретическими проводится по критерию χ^2 хи-квадрат (критерий Пирсона).

Пример расчета распределения Пуассона

Средние градации классов	Частоты	$\bar{X}^x = 0,672^x$	$x!$	$\frac{\bar{X}^x}{x!}$	$\frac{\bar{X}^x}{x!} \cdot N_n \cdot e^{-\bar{X}}$	f'
0	70	1,00	0! = 1	1,0	62,43	62
1	32	0,672	1	0,672	41,95	42
2	14	0,4516	2	0,2258	14,1	14
3	2	0,3035	6	0,0506	3,16	3
4	4	0,2039	24	0,0085	0,53	1
	$\Sigma=122$					122

Расчет нормального распределения. Основная формула, по которой строится эта кривая, - формула, выражающая зависимость между переменными величинами Y и X :

$$y = \frac{N}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x_i - \bar{X}}{\sigma} \right)^2}$$

При использовании ПЭВМ расчет проводят по соотношению:

$$y = N \cdot \exp(\ln k + \ln 0,39894 - \ln \sigma - 0,5 \cdot t^2), \text{ где}$$

y – расчетная, теоретическая частота;
 N – общее число вариант данной совокупности;
 σ – основное отклонение;
 $(x_i - \bar{X})$ – отклонение варианты (середины класса) от средней арифметической;
 $t = (x - \bar{X})/\sigma$ – нормированное отклонение;
 $\pi = 3,1416$,
 k – классовый интервал,
 e – основание натуральных логарифмов, равное 2,7182818;

При отклонении $(x_i - \bar{X})$ равном нулю $e^{-\frac{1}{2}\left(\frac{x_i - \bar{X}}{\sigma}\right)^2}$ превращается в единицу. В этом случае функция нормального распределения дает максимальную частоту. Используя эту особенность и нормируя отклонение, расчет теоретических частот проводится следующим образом.

Вначале определяется максимальная частота:

$$N_{\max} = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \quad \text{или}$$

$$n_{\max} = \frac{N}{\sigma' \cdot \sqrt{2 \cdot \pi}} = 85 / (1,61 \cdot \sqrt{6,2832}) = 21,064, \text{ где}$$

σ – основное отклонение полное;
 σ' – основное отклонение неполное;
 N – сумма частот.

Затем расчет ведется по схеме, приведенной в таблице.

В первой колонке записываются средние значения классов с добавкой вверх и вниз по одному-два класса (если число классов меньше 10). За условное значение классов, помещенные во второй колонке, берутся начальные отклонения (см. вычисление моментов по способу произведений).

В третьей записываются частоты.

В четвертой колонке записывается алгебраическая разница между условным значением класса и первым начальным моментом. Эта же разница, деленная на основное отклонение (в условных единицах), представляет собой нормированное отклонение, которое используется для нахождения по приложению - Г функции $e^{-\frac{1}{2}(x_i - \bar{X})^2}$. При этом для нахождения искомых значений используется линейная

интерполяция между табличными данными. Умножением расчетной максимальной частоты построчно на найденные значения функции находятся теоретические частоты, которые записываются в последней колонке таблицы.

Пример расчета нормального распределения

Х - среднее значе- ние классов	Х' - ус- лов- ное значе- ние классов	n - час- то- та,	Х' - m _l	Х'' = (Х' - m _l)/σ' нормиро- ванное отклоне- ние	$e^{-\frac{1}{2}(x_i - \bar{X})^2}$	$n' = n'_{\max}$ · $e^{-\frac{1}{2}(x_i - \bar{X})^2}$	n' окргл
63	-4	0	-4,2588	-2,6454	0,03023	0,637	0
77	-3	4	-3,2588	-2,0242	0,12891	2,715	3
91	-2	6	-2,2588	-1,4031	0,37369	7,871	8
105	-1	15	-1,2588	-0,7819	0,73661	15,516	15
119	0	27	-0,2588	-0,1608	0,98598	20,769	21
133	1	16	0,7412	0,4604	0,89793	18,914	19
147	2	10	1,7412	1,0816	0,55715	11,736	12
161	3	5	2,7412	1,7027	0,23467	4,943	5
175	4	0	3,7412	2,3239	0,06719	1,415	1
189	5	2	4,7412	2,9450	0,01309	0,276	1
203	6	0	5,7412	3,5662	0,00173	0,037	0
Σ		85			4,02718	84,829	85

Расчет кривой нормального распределения типа А. Распределение этого типа является разложением в ряд уравнения нормального распределения. Оно полнее учитывает асимметрию, эксцесс и дает лучшую аппроксимацию экспериментального ряда, чем уравнение Лапласа-Гаусса (нормальное распределение).

Расчет теоретических частот ведется по формуле: $n' = \frac{N}{\sigma^4} f_A(x)$,

где функция $f_A(x) = f(x) - \frac{r_3}{6} f'''(x) + \frac{r_4 - 3}{24} f^{IV}(x)$, где

$f(x)$ – функция кривой нормального распределения;

$f'''(x)$, $f^{IV}(x)$ - производные функции кривой нормального распределения;

r_3 , r_4 – основные моменты;

Форма записи результатов расчета приведена в таблице.

В ней первые пять колонок заполняются так же, как при вычислении частот при нормальном распределении вариант.

Значения для шестой, седьмой и восьмой колонок берутся из таблицы приложения Д на основе нормированного отклонения (пятая колонка). При этом для нахождения искоемых значений используется линейная интерполяция между табличными данными.

В таблице величина $f'''(x)$ приведена для положительных значений нормированного отклонения. Для отрицательных значений X'' значность (плюс или минус) выбранных табличных значений меняется на противоположную. Значность табличных значений $f(x)$ и $f^{IV}(x)$ не зависит от знака X'' поэтому при выборке их из таблицы знаки не меняются.

В девятой колонке построчно записывается произведение $f'''(x)$ и коэффициента r_3 / σ . При этом следует внимательно учитывать знаки используемых при расчетах чисел.

При расчетах по десятой колонке следует иметь ввиду, что иногда выражение $r_4 - 3$ может иметь отрицательное значение, а значение четвертой производной $f^{IV}(x)$ может менять значность при различных отклонениях.

В одиннадцатой колонке построчно приводятся алгебраические суммы значений шестой, девятой и десятой колонок.

В двенадцатой колонке приведены расчетные частоты. Их сумма должна быть близка к общей численности ряда (N).

Для контроля расчетов сопоставляется сумма значений одиннадцатой колонки и сумма значений из шестой, девятой и десятой колонок.

Вопросы для самоконтроля

- 1 Как определяется тип кривой при помощи критерия Пирсона?
- 2 Как осуществить расчет нормального распределения?
- 3 Как осуществить расчет нормального распределения типа А?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 2; калькуляторы.

Цель. Определение типа распределения и расчет теоретических частот биномиального, Пуассона, нормального распределения.

Схема расчета частот распределения типа А

Среднее значение классов X	Условное значение классов X'	Частота, n	Нормированное отклонение $X''=(X'-m)/\sigma'$	$f(x)$	$f^{III} (x)$	$f^{IV} (x)$	$-\frac{r_3}{6} f^{III}(x)$	$\frac{r_4-3}{24} f^{IV}(x)$	$f_A(x)$	$n' = \frac{N}{f} f_A(x)$
63	-4	0	-2.6454	0,01206	+0,12753	+0,12039	-0,00824	+0,00283	0,00665	0,351
77	-3	4	-2.0242	0,05142	+0,11422	-0,24663	-0,00738	-0,00579	0,03825	2,020
91	-2	6	-1.4031	0,14908	-0,21571	-0,73587	+0,01394	-0,01728	0,14574	7,695
105	-1	15	-0.7819	0,29386	-0,54883	-0,08651	+0,03546	-0,00203	0,32729	17,28
119	0	27	-0.1608	0,39382	-0,18833	+1,12061	+0,01217	+0,02631	0,43230	22,825
133	1	16	0.4604	0,35882	-0,46059	+0,63623	-0,02976	+0,01494	0,34400	18,163
147	2	10	1.0816	0,22227	-0,43997	-0,58911	-0,02843	-0,01383	0,18001	9,504
161	3	5	1.7027	0,09362	+0,01608	-0,56075	-0,00104	-0,01317	0,07941	4,093
175	4	0	2.3239	0,02681	-0,14952	-0,00638	+0,00966	-0,00015	0,03632	1,918
189	5	2	2.9450	0,00522	-0,08719	+0,13664	+0,00563	+0,00321	0,01406	0,742
203	6	0	3.5662	0,00069	-0,02394	+0,06108	+0,00154	+0,00143	0,00366	0,193
Σ		85		1,60767			+0,00355	-0,00353	1,60769	84,783

Ход работы

1 На основе основных положений темы и данных сгруппированной совокупности по диаметру либо высоте определить тип кривой отражающей распределение частот вариационного ряда по критерию Пирсона.

2 На основе основных положений темы и данных сгруппированной совокупности по диаметру либо высоте рассчитать теоретические частоты биномиального и распределения Пуассона.

3 На основе основных положений темы и данных сгруппированной совокупности по диаметру либо высоте рассчитать теоретические частоты нормального распределения и типа А.

ТЕМА 4 СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

- 1 Расчет ошибок основных статистических показателей.
- 2 Расчет доверительных интервалов статистических показателей.

Основные понятия по теме

Расчет ошибок основных статистических показателей. Выборочная совокупность довольно точно воспроизводит свойства и соотношения в генеральной совокупности, но не абсолютно точно вследствие колеблемости изучаемых признаков. Поэтому между статистическими показателями выборочной совокупности и действительными значениями этих показателей генеральной совокупности всегда будут некоторые расхождения, которые являются случайными ошибками выборки (иначе - случайными ошибками репрезентативности) и называются **основными ошибками** того или иного статистического показателя.

На основании величины этой основной ошибки и значения соответствующего показателя выборки можно судить о действительном значении данного показателя в генеральной совокупности. Так, с вероятностью равной 0,68 (в 68% случаев из ста), можно утверждать, что расхождение между действительным значением данного показателя в генеральной совокупности и вычисленным его значением для выборки не превышает однократного значения основной ошибки этого показателя (со знаком плюс или минус); предельное же расхождение не превышает трехкратного значения основной ошибки (о чем можно утверждать с вероятностью 0,997 или 99,7% случаев из ста).

Основные ошибки статистических показателей вычисляются по формулам:

Ошибка среднего значения (или ошибка среднеквадратического отклонения) $m_x = \frac{\sigma}{\sqrt{n}}$;

Если распределение частот сильно отличается от нормального, то ошибку сигмы вычисляют по формуле: $m_\sigma = \sqrt{\frac{(\mu_4 - \mu_2^2)}{4 \cdot \mu_2 \cdot N}}$

Ошибка коэффициента изменчивости $m_c = \frac{C}{\sqrt{n}} \left[0.5 + \left(\frac{C}{100} \right)^2 \right]$

Точность опыта, или процент ошибки наблюдения – это процент расхождения между генеральной и выборочной средней, который вычисляется по формуле: $p = \frac{100 \cdot m_x}{X}$ или же по формуле

$p = \frac{C}{\sqrt{N}}$. В 68% случаях расхождение между генеральной и выборочной средней не превышает однократного значения точности опыта (в одну или другую сторону), а предельное расхождение – не превосходит трехкратного значения.

Точность опыта показывает, насколько процентов можно ошибиться, если утверждать, что генеральная средняя равна полученной выборочной средней.

Полученный процент ошибки сопоставляется с заданным: если он не больше заданного, точность достаточная, а если больше, то точность результата является неудовлетворительной; значит, следует увеличить число наблюдений.

Ошибка точности опыта может быть вычислена через ошибку коэффициента изменчивости: $m_p = \frac{m_c}{\sqrt{N}}$.

Ошибка показателя асимметрии проводится по формуле: $m_A = \sqrt{\frac{6}{N}}$. Ошибка показателя эксцесса равна удвоенной ошибке показателя асимметрии.

Более точное значение ошибок получают по формуле

$$s_A = \sqrt{\frac{6n \cdot (n-1)}{(n-2) \cdot (n+1) \cdot (n+3)}}$$

$$s_E = \sqrt{\frac{24n \cdot (n-1)^2}{(n-3) \cdot (n-2) \cdot (n+2) \cdot (n+5)}}$$

После вычисления того или иного статистического показателя необходимо проверить степень его надежности или достоверности

путем деления величины данного показателя на величину его

основной ошибки: $t = \frac{X}{m_x}$, где

X - величина любого статистического показателя;

m_x - величина ошибки любого статистического показателя.

Если частное t получится равным или больше трех, то значение показателя является надежным, достоверным, и им можно пользоваться для разных сопоставлений и выводов. Если же это отношение будет меньше трех, то данный показатель оказывается ненадежным, величина его не достоверна и является лишь в той или иной мере вероятной. Такие показатели нельзя сопоставлять между собой или производить на основе их заключения.

Нередко приходится решать вопрос, насколько существенно различие в значениях показателей какого-либо признака, вычисленных для разных совокупностей. С этой целью находится основная ошибка разницы чисел и доказываемая ее достоверность по выше описанному принципу.

Ошибка разности вычисляется как корень квадратный из суммы квадратов основных ошибок исследуемого показателя, то есть

$$m_s = \sqrt{m_1^2 + m_2^2}.$$

Полученную разность показателей делят на его ошибку. Находится показатель существенности различия средних значений:

$$t = \frac{X_1 - X_2}{m_s} = \frac{X_1 - X_2}{\sqrt{m_1^2 + m_2^2}}.$$

Если этот показатель получится больше трех, то различие существенно, доказано, и данное мероприятие вызвало существенное изменение; обе сравниваемые выборочные совокупности являются представителями качественно разных генеральных совокупностей. Если же он получится меньше трех, то можно утверждать, что расхождение оказалось случайным, недостоверным и во всяком случае целесообразность данного мероприятия осталась недоказанной.

Расчет доверительных интервалов статистических показателей. Среднее значение, основное (квадратическое) отклонение, коэффициент изменчивости, асимметрия, эксцесс дают представление о величине и форме распределений наблюдений. Однако они не дают представления о возможных значениях случайной величины. Оно заключается в вычислении вероятности того, что значение величины будет заключаться в определенных границах. Считается, что границы достоверно определены если вероятность близка к

единице, например, 0,99 или 0,999 (99,0% или 99,9%). Соответствующие границы называются доверительными.

В зависимости от типа распределения данных доверительный интервал рассчитывается двумя способами.

В симметричных распределениях, близких к нормальному, размах отклонений данных от средней арифметической обычно равен приблизительно 3σ в обе стороны от значения средней. (так называемый закон трех сигм).

При расчете доверительного интервала для трех стандартных доверительных уровней: 95%, 99%, 99,9% t выбирается по числу степеней свободы из таблицы.

Значения показателя t (критерия Стьюдента)

Число степеней свободы	Доверительные уровни		
	95%	99%	99,9%
9	2,3	3,2	4,8
10	2,2	3,2	4,6
11-14	2,2	3,0	4,3
15-20	2,1	2,9	3,9
21-30	2,1	2,8	3,7
31-60	2,0	2,7	3,5
61-120	2,0	2,6	3,4
∞	1,96	2,58	3,29

Доверительный интервал статистического показателя, например, средней арифметической строится по формуле:

$$\bar{X} - t \cdot m_{\bar{x}} < \text{истинное значение} < \bar{X} + t \cdot m_{\bar{x}}$$

Подставляя в формулу величины среднего арифметического, коэффициента вариации, коэффициента асимметрии, эксцесса и их ошибок определяются доверительные интервалы для этих показателей.

Когда доверительный интервал необходимо определить при неизвестном типе распределения вариант применяется неравенство Чебышева. В этом случае подразумевается, что размах отклонений от средней арифметической в совокупности может быть больше, чем в нормальном распределении, то есть больше, чем 3σ .

Согласно неравенству Чебышева истинная средняя арифметическая находится в интервале:

$$\bar{X} - \frac{m_{\bar{x}}}{\sqrt{1-P'}} < \text{истинное значение} < \bar{X} + \frac{m_{\bar{x}}}{\sqrt{1-P'}}, \text{ где}$$

\bar{x} - средняя арифметическая, рассчитанная по выборке; m_M - ее ошибка; P' - доверительный уровень в долях единицы или $1 - P' = W$ - уровень значимости, например, доверительный уровень вероятности равен 0,95 95%, а уровень значимости $W = 1 - 0,95 = 0,05$).

Подставляя в формулу величины среднего арифметического, коэффициента вариации, коэффициента асимметрии, эксцесса и их ошибок строятся доверительные интервалы для этих показателей.

Вопросы для самоконтроля

- 1 Как определяются основные ошибки статистических показателей?
- 2 Как определяется доверительный интервал при нормальном распределении статистических показателей?
- 3 Как определяется доверительный интервал при неизвестном типе распределений статистических вариантов?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 2; калькуляторы.

Цель. Определение основных ошибок статистических показателей и их доверительных интервалов.

Ход работы

1 На основе основных положений темы и расчетов по лабораторной работе темы 2 определить основные ошибки статистических показателей;

2 На основе основных положений темы и расчетов по лабораторной работе темы 2 определить доверительный интервал для основных ошибок статистических показателей.

ТЕМА 5 СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

- 1 Критерий согласия Пирсона.
- 2 Критерий согласия Колмогорова-Смирнова.

Основные понятия по теме

Критерий согласия Пирсона. Критерии различия, при помощи которых могут быть сравнены статистические совокупности, разделяются на две группы: параметрические и непараметрические. К первой группе относятся критерии, для применения которых необходимо вычислить среднюю арифметическую, сигму или ошибки параметров (критерий Стьюдента, Фишера).

Непараметрические критерии не требуют для своего применения вычисления названных показателей, что упрощает процесс сравнения совокупностей. Критерий согласия Пирсона (или χ^2), критерий Колмогорова (или лямбда λ) относятся к непараметрическим критериям.

Оценка близости, согласованности в распределении частот, вычисленных для любого типа распределений и полученных по фактическим данным производится при помощи критерия Пирсона. Он может быть также применен как для сравнения двух вариационных рядов, так и для установления правильности выбора теоретического распределения. Он рассчитывается по формуле:

$$\chi^2 = \frac{1}{K-1} \sum \frac{(n - n^1)^2}{n^1}, \text{ где}$$

χ^2 - критерий Пирсона;

K – количество классов включая добавленные при проведении расчетов;

n – частота фактическая;

n^1 - частота расчетная.

Если критерий согласия равен или больше 2, то расхождение сравниваемых рядов считается существенным, и если меньше 2, - расхождение несущественное. Более точная оценка значимости коэффициента проводится по специальным таблицам.

Расчет критерия согласия Пирсона оформляется в виде таблицы.

Пример расчета критерия согласия Пирсона

X	n	n'	$n - n'$	$(n - n')^2$	$(n - n')^2 / n'$
63	0	1	-1	1	1
77	4	3	1	1	0,33
91	6	8	-2	4	0,50
105	15	16	-1	1	0,06
119	27	21	6	36	1,71
133	16	19	-3	9	0,47
147	10	12	-2	4	0,33
161	5	4	1	1	0,25
175	0	1	-1	1	1
189	2	1	1	1	1
203	0	1	-1	1	1
	Σ 85	Σ 87			Σ 4,65
$\chi^2 = (1/ 11-1) \times 4,65 = 0,465$					

Критерий согласия Колмогорова-Смирнова. Один из наиболее простых и удобных при сопоставлении эмпирических совокупностей большого объема – критерий, предложенный А.Н. Колмогоровым и Н.В. Смирновым. Этот непараметрический показатель, обозначаемый греческой буквой λ (лямбда), представляет собой максимальную разность (d_{\max}) между значениями накопленных частот эмпирического и вычисленного рядов (без учета знаков d), отнесенную к корню квадратному из суммы всех вариант совокупности:

$$\lambda = \frac{d_{\max}}{\sqrt{n}} .$$

Условием применения критерия «лямбда» служит достаточное число (не менее 100) наблюдений.

Предельные значения критерия лямбда, соответствующие трем уровням доверительной вероятности – $P_1 = 0,95$, $P_2 = 0,99$ и $P_3 = 0,999$ – соответственно равны 1,36, 1,63 и 1,95.

Расчет критерия «лямбда» показан на примере распределения высот в 40-летнем сосняке.

Пример расчета критерия Колмогорова-Смирнова

Срединные значения классов (x)	Эмпирические частоты (p)	Теорет. вычисл. частоты (окргл.) (p')	Накопленные Частоты		p-p'=d
			p	p'	
8,8	1	1	1	1	0
9,3	2	2	3	3	0
9,8	3	4	6	7	1
10,3	7	8	13	15	2
10,8	10	12	23	27	4
11,3	16	15	39	42	3
11,8	16	16	55	58	3
12,3	17	15	72	73	1
12,8	12	12	84	85	1
13,3	7	8	91	93	2
13,8	4	4	95	97	2
14,3	3	2	98	99	1
14,8	2	1	100	100	0
Сумма	100	100	-	-	-

Расчет необходимых значений показан в таблице. Максимальное значение разности $p - p' = 4$, откуда $\lambda = \frac{4}{\sqrt{100}} = 0,4$. Полученная величина значительно меньше предельного значения лямбда (1,36) для $P = 0,05$. следовательно, расхождения между эмпирическими и вычисленными частотами симметричного распределения лежат в пределах случайных колебаний, они не достоверны. На этом основании распределение высот в исследованном древостое можно считать нормальным.

Вопросы для самоконтроля

- 1 Что такое параметрические и непараметрические критерии и как они вычисляются?
- 2 Для каких целей применяются статистические критерии?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 3; калькуляторы.

Цель. Определение соответствия расчетных и фактических кривых распределений диаметров и высот по критериям согласия Пирсона, Колмогорова – Смирнова.

Ход работы

1 На основе основных положений темы и расчетов по лабораторной работе темы 3 определить соответствие расчетных и фактических кривых распределений диаметров и высот по критерию согласия Пирсона;

2 На основе основных положений темы и расчетов по лабораторной работе темы 3 определить соответствие расчетных и фактических кривых распределений диаметров и высот по критерию Колмогорова – Смирнова.

ТЕМА 6 КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

- 1 Коэффициент корреляции.
- 2 Коэффициент множественной корреляции.

Основные понятия по теме

Коэффициент корреляции. Отличительной особенностью биологических объектов является многообразие признаков, характеризующих каждый из них. Часто наблюдается связь между вариациями по различным признакам. В простейшем случае связь между двумя переменными величинами строго однозначна. Например, вес образцов, сделанных из одного и того же материала, определяется их объемом. Такого рода зависимость принято называть **функциональной**. Для биологических объектов связь обычно бывает менее жесткой: объекты с одинаковым значением одного признака имеют, как правило, разные значения по другим признакам. Такую связь между вариациями разных признаков называют корреляцией (дословный перевод: соотношение) между признаками.

Если устанавливается наличие и характер корреляции между признаками (например, диаметром и высотой), то производится

разноска полученных численных значений в одну общую таблицу называемую *корреляционной решеткой* (см. вычисление смешанных моментов).

Для расчета коэффициента корреляции формируется выборка. Для этого используются значения частичной совокупности в объеме 50 пар значений диаметров и высот из первой лабораторной работы (задание 1). Данные группируются согласно следующей таблице. В каждой колонке должно быть 25 значений.

Диаметр X	Высота Y	Диаметр	Высота Z
Выписываются значения с номерами: 2, 4, 6, и т. д. до 50	Выписываются значения с номерами: 2, 4, 6, и т. д. до 50	Выписываются значения с номерами: 1, 3, 5, и т. д. до 49	Выписываются значения с номерами: 1, 3, 5, и т. д. до 49

Значения с четными или нечетными номерами преобразуются согласно следующей таблице.

Диаметр, см X	Высота, м Y	X ²	Y ²	X · Y
Значения из частичной совокупности 25 шт.				
Σ	Σ	Σ	Σ	Σ

Затем вычисляется коэффициент корреляции, его ошибка и достоверность.

Коэффициент корреляции вычисляется в тех случаях, когда требуется выразить количественно силу связи между двумя сопряженными признаками, если известно, что зависимость одного из признаков от другого близка к прямолинейной.

При малых объемах выборок (до 30-50) коэффициент корреляции вычисляется по формуле:

$$r = \frac{\sum x \cdot y - (\sum x \cdot \sum y) / N}{\sqrt{[\sum x^2 - (\sum x)^2 / N] \cdot [\sum y^2 - (\sum y)^2 / N]}}$$

При расчетах коэффициента корреляции на ПЭВМ как для малых, так и для больших рядов (без группирования в классы) удобна формула:

$$r = [\sum (X - \bar{X}) (Y - \bar{Y})] / \sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}$$

Достоверность коэффициента корреляции можно оценить по формуле $t = (0,5 \cdot \ln(1 + r / (1 - r))) \cdot \sqrt{N - 3}$, где t – критерий Стьюдента при числе свободы $v = N - 2$.

По результатам расчетов делается вывод о характере связи:

- связь между признаками прямая ($r > 0$) или обратная ($r < 0$);
- теснота связи близка к функциональной $r = 1.0$;
- $r = 0.901 - 0.999$ связь очень высокая;
- $r = 0.701 - 0.900$ высокая;
- $r = 0.501 - 0.700$ значительная;
- $r = 0.301 - 0.500$ слабая;
- $r = 0 - 0.300$ отсутствует.

Затем определяется достоверность вычисленных показателей и сравниваются их значения, полученные по способу смешанных моментов и по приведенной формуле для малой выборки.

Коэффициент множественной корреляции. При анализе множественных взаимосвязей **выявляются структуры** связей в определенном наборе признаков, и **проводится отбор** наиболее тесно связанных между собой признаков.

Силу совокупной взаимосвязи трех признаков, когда измеряется совместное взаимодействие двух признаков с третьим можно определить при помощи множественных коэффициентов корреляции по формулам:

$$R^2_{y \cdot xz} = (r^2_{xy} + r^2_{yz} - A) / (1 - r^2_{xz})$$

$$R^2_{x \cdot yz} = (r^2_{xy} + r^2_{xz} - A) / (1 - r^2_{yz})$$

$$R^2_{z \cdot xy} = (r^2_{zx} + r^2_{zy} - A) / (1 - r^2_{xy}), \text{ где}$$

$R_{y \cdot xz}$ - множественный коэффициент корреляции; точка например, после y означает, что изучается совместное или объединенное влияние аргументов x, z на функцию y;

$A = 2 r_{xy} r_{xz} r_{yz}$, r_{yz} , r_{yx} , r_{xz} - полные (обычные) коэффициенты корреляции попарно между признаками x, y, z;

$r_{yx \cdot x}, r_{yx \cdot z}, r_{xz \cdot y}$ - частные коэффициенты корреляции попарно при исключении влияния третьего признака, индекс последнего ставится справа от точки.

Критерий достоверности (t) частного коэффициента корреляции при любом объеме совокупности вычисляют по формуле:

$$t = R_{xy \cdot z} \sqrt{(N - n - 2) / (1 - R_{xy \cdot z}^2)}$$

В формулу вместо $R_{y \cdot xz}$ можно подставлять величину любого частного коэффициента корреляции. При оценке достоверности частного коэффициента корреляции по таблице значений критерия Стьюдента число степеней свободы принимается $v = N - 2 - n$, где N – объем выборки; n – число факторов, действие которых устраняется (в данном случае $n = 1$).

Вычисление множественного и частных коэффициентов корреляции производится в следующем порядке. Вычисляем полные коэффициенты корреляции между значениями высот и диаметров, для чего произведем действия, указанные в таблице. Подставляя суммы столбцов этой таблицы в формулу вычисления коэффициента корреляции, получим значения r_{yx}, r_{xz}, r_{yz} и оцениваем их величину и достоверность.

Диаметр X	Высота Y	Высота Z	X^2	Y^2	Z^2	XY	XZ	YZ
Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ

Вопросы для самоконтроля

- 1 Что такое коэффициент корреляции и как он вычисляется?
- 2 Что такое множественный коэффициент корреляции и как он вычисляется?
- 4 Какова сфера применения корреляционного анализа в лесном хозяйстве?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 2; калькуляторы.

Цель. Определение простого и множественного коэффициентов корреляции.

Ход работы

1 На основе положений темы 6 и выборочной совокупности из лабораторной работы по теме 2 отразить взаимообусловленность диаметров и высот деревьев аналитическими показателями.

ТЕМА 7 РЕГРЕССИОННЫЙ АНАЛИЗ

- 1 Расчет линии регрессии.
- 2 Оценка регрессионных уравнений.

Основные понятия по теме

Расчет линии регрессии. Регрессионный анализ предполагает аналитическое выражение вероятностной связи между признаками уравнениями различного вида.

Регрессионные модели обычно используют для выражения разного рода связей в лесной таксации, лесоводстве и в других лесных дисциплинах. Чаще всего они применяются для нахождения общей зависимости по экспериментальным данным. Введенное уравнение сглаживает (выравнивает) полученные (экспериментальные) данные. В этом случае сохраняется главная тенденция изменения функции в зависимости от изменения аргументов, и устраняются случайные отклонения.

По форме различают линейную регрессию и не линейную. По направлению связи различают прямую т.е. с увеличением признака x увеличивается признак y и обратную т.е. с увеличением x уменьшается y . Наиболее точная оценка принадлежности к виду связи производится с помощью метода наименьших квадратов (МНК). При МНК \min сумма квадратов отклонений эмпирических значений y от теоретических полученных по выбранному уравнению регрессии стремится к минимуму.

Простейшей теоретической линией регрессии является прямая линия, или парабола первого порядка, которая имеет вид:

$$Y' = a_0 + a_1 \cdot x,$$

где Y' - теоретические значения функции или зависимой переменной; x - аргумент или независимая переменная; a_0, a_1 - коэффициенты уравнения, имеющие различное значение в зависи-

мости от специфики изучаемого явления; \bar{Y} – эмпирические значения зависимой переменной.

Коэффициенты определяются по формулам:

$$a_0 = \bar{Y} - a_1 \cdot \bar{X} \quad a_1 = \frac{\sum(x - \bar{X}) \cdot (y - \bar{Y})}{\sum(x - \bar{X})^2},$$

где \bar{X} и \bar{Y} средние арифметические рядов аргументов (X) и функции (Y).

Ход вычислений приведен в таблице.

x	y	$x - \bar{X}$	$(x - \bar{X})^2$	$(y - \bar{Y})$	$(x - \bar{X}) \cdot (y - \bar{Y})$	$(y - \bar{Y})^2$	y'
Σ	Σ	Σ	Σ		Σ	Σ	

В качестве исходных данных используются вариационные ряды частичной совокупности или средние значения высот и диаметров при вычислении смешанных моментов.

Если известны среднеквадратические отклонения для рядов x и y и найден коэффициент корреляции между ними (см. вычисление смешанных моментов), то величина a_1 вычисляется по формуле:

$a_1 = r_{xy} \cdot \left(\frac{\sigma_y}{\sigma_x} \right)$, где σ_y , σ_x - дисперсии выборочных (или усредненных) рядов; r_{xy} - мера корреляционной взаимосвязи.

Точки пересечения с осями ординат и абсцисс равны соответственно:

$$y = a_0, \quad x = -(a_0/a_1).$$

При исследованиях нередко применяется аппроксимация криволинейных эмпирических зависимостей параболой второго порядка и другими видами уравнений, нахождение аналитического выражения их проводится на ПЭВМ при помощи встроенных программ в EXCEL(см. лаб. работу темы 9).

Оценка регрессионных уравнений. Поскольку в определении линий регрессии участвуют несколько параметров, то необходимо оценить пределы изменчивости каждого из них.

Наиболее вероятная область расположения линии прямой регрессии по отношению к оси абсцисс определяется величиной коэффициента a_1 и тангенсом угла, геометрическим смыслом которого является коэффициент корреляции. При отсутствии регрессии $r=0$, и тогда линия регрессии y по x располагается горизон-

тально по отношению к оси абсцисс, а линия регрессии x по y - вертикально. Место их пересечения соответствует средним значениям обоих признаков.

Второй коэффициент определяет величину отрезка, отсекаемого на оси y линией регрессии. Величина его определяет границы колеблемости регрессии по ординате, которая расширяется в обе стороны от средней точки (\bar{x}, \bar{y}) .

Поскольку опытные данные всегда имеют определенную величину изменчивости, то и все показатели в том числе и уравнения регрессии определяются с некоторой степенью достоверности.

Определение величины ошибки найденных уравнений и оценка достоверности полученных коэффициентов уравнения прямой проводится по формулам:

$$m_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{N - n}}, \text{ или } \sigma_{y \cdot x} = \sqrt{\frac{\sum (y_i - y'_i)^2}{n - 2}} \quad \text{где } \sigma_{y \cdot x}, m_{y \cdot x} -$$

ошибка уравнения; y - эмпирические значения функции; y' , - теоретические значения функции; N - число точек эмпирической линии регрессии, по которым вычислялось уравнение регрессии; n - число коэффициентов уравнения, включая свободный член.

Здесь величина $\sigma_{y \cdot x}$ имеет такое же значение как и σ в вариационном ряду. В пределах одной $\sigma_{y \cdot x}$ отклонения распределяются вверх и вниз от линии регрессии в 68% случаев. В 95% они лежат в пределах $2\sigma_{y \cdot x}$, а в 99,7% случаев отклонения от теоретической линии регрессии составляют величину $3\sigma_{y \cdot x}$. Ошибку уравнения регрессии можно определить и по формуле: $m_{y \cdot x} = \sigma_y \sqrt{1 - r^2}$, где $m_{y \cdot x}$ - ошибка теоретических значений функции; σ_y - среднее квадратическое отклонение ряда y ; r - коэффициент корреляции между x и y (можно использовать и корреляционное отношение при наличии криволинейности связи между признаками). Эта формула представляет упрощенный вариант вычислений и применяется для больших выборок.

В таблицах по вычислению коэффициентов уравнений в последней колонке рассчитываются теоретические значения функции. Получаем попарно разности $(y - y')$, возводим все разности в квадрат и получаем их сумму: $\sum (y - y')^2$. Применив формулу:

$m_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{N - n}}$, к уравнению прямой, параболы определим их ошибку.

Достоверность найденного коэффициента a_1 определяется по формуле: $t = a_1 \cdot \sigma_x \cdot \sqrt{N-1} / m_{yx}$, где t – величина критерия Стьюдента, сравниваемая с критической при числе степеней свободы $\nu = N-2$; σ_x – среднее квадратическое отклонение ряда аргументов; m_{yx} – ошибка уравнения; N – объем выборки.

Если вычисленная величина меньше табличной, то связь между x , y и значение a_1 достоверны, а если вычисленная будет больше табличной величины, то связь данных признаков и значение первого коэффициента недостоверны.

Достоверность отличия от нуля коэффициента a_0 можно оценить по формуле:

$$t = \frac{a_0}{m_{y \cdot x} \sqrt{\frac{1}{N} + \frac{1}{N-1} \cdot \left(\frac{\bar{X}}{\sigma_x}\right)^2}}, \text{ где } t \text{ – величина критерия Стьюдента}$$

та, сравниваемая с критической при числе степеней свободы $\nu = N-2$; σ_x – среднее квадратическое отклонение ряда аргументов; m_{yx} – ошибка уравнения; N – объем выборки.

Вопросы для самоконтроля

- 1 Что такое метод МНК?
- 2 В чем состоит геометрический смысл коэффициентов уравнения прямой и как они определяются?
- 3 Что лежит в основе статистической оценки регрессионных уравнений и как ее определить?

Лабораторная работа

Материалы и оборудование: данные замеров статистических величин и расчетов по лабораторной работе темы 2; калькуляторы.

Цель. Определение величины коэффициентов уравнений регрессии методом наименьших квадратов и их статистической оценки.

Ход работы

- 1 На основе основных положений темы и расчетов по лабораторной работе темы 2 определить взаимобусловленность диаметров и высот деревьев аналитическими уравнениями.

- 2 Определить ошибки регрессионных уравнений и достоверность коэффициентов линейных уравнений.

ТЕМА 8 ДИСПЕРСИОННЫЙ АНАЛИЗ

- 1 Схема расчетов при дисперсионном анализе.
- 2 Подготовка статистических массивов для дисперсионного анализа.

Основные понятия по теме

Схема расчетов при дисперсионном анализе. Дисперсионный или вариантный анализ (analysis of variance) предполагает установление роли отдельных факторов в изменчивости того или иного признака, при котором общая дисперсия как количественных, так и качественных признаков раскладывается на отдельные составляющие.

У изучаемых признаков в эксперименте имеется не одно, а несколько значений, которые называют *градациями или уровнями фактора А*.

Число наблюдений (вариант) в каждой группе обозначается как « n », но равное число наблюдений в группах не обязательно. При неравном числе можно исходить из среднего числа n_i . $N = k \cdot n$ ($=k \cdot n_i$)

Обычно разные уровни принято обозначать буквой i , а отдельные варианты (наблюдения) — буквой j . Поэтому каждую варианту, независимо от того, где она находится, обозначают в общем виде как x_{ij} . В пределах каждого уровня (группы) отдельные варианты принимают случайные значения:

$x_{.1}, x_{.2}, x_{.3}, \dots, x_{.j}, \dots, x_{.n}$

Средние по группам: $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i, \dots, \bar{x}_a$. В общем виде групповые средние обозначают через \bar{x}_i . Общую же среднюю для всех вариантов всех групп — как \bar{x} .

Схема обозначения членов вариационного ряда при однофакторном дисперсионном анализе

Группы фактора	Отдельные варианты (наблюдения) x_{ij}						Суммы по группам	Средние по группам \bar{x}_i	
	1	2	3	..	j	..			n
1	x_{11}	x_{12}	x_{13}		x_{1j}		x_{1n}	$\sum x_{1i}$	\bar{x}_1
2	x_{21}	x_{22}	x_{23}		x_{2j}		x_{2n}	$\sum x_{2i}$	\bar{x}_2
i	x_{i1}	x_{i2}	x_{i3}		x_{ij}		x_{in}	$\sum x_{i1}$	\bar{x}_i
k	x_{k1}	x_{k2}	x_{k3}		x_{kj}		x_{kn}	$\sum x_{ki}$	\bar{x}_k
								$\sum x_{ij}$	\bar{x}

Различают 3 типа варьирования:

а) S_0^2 - общее варьирование вариант (x_{ij}), независимо от того, в какой группе они находятся, вокруг общей средней \bar{x} ;

б) S_M^2 - варьирование групповых средних \bar{x}_i , или, иначе, средних каждого уровня данного изучаемого фактора, вокруг общей средней \bar{x} ;

в) S_b^2 - варьирование вариант x_{ij} внутри каждой группы вокруг каждой групповой средней \bar{x}_i (так называемая остаточная).

Между ними существует соотношение:

$$S_0^2 = S_M^2 + S_b^2$$

Для каждого типа варьирования вычисляются суммы квадратов отклонений по следующим формулам:

Схема дисперсионного анализа (анализа варiances)
при влиянии одного фактора

Источник варьирования	Сумма квадратов SS	Число степеней свободы df	Средний квадрат ms
Общее (все варианты)	$S_0^2 = \sum_{ij} (x_{ij} - \bar{x})^2$	N-1	$\frac{1}{N-1} \sum_{ij} (x_{ij} - \bar{x})^2$
Групповые средние (фактор A)	$S_M^2 = \sum_i n_i (\bar{x}_i - \bar{x})^2$	k-1	$\frac{1}{k-1} \sum_i n_i (\bar{x}_i - \bar{x})^2$
Варианты внутри групп (случайные отклонения)	$S_b^2 = \sum_i \left[\sum_j (x_{ij} - \bar{x}_i)^2 \right]$	N-k	$\frac{1}{N-k} \sum_i \left[\sum_j (x_{ij} - \bar{x}_i)^2 \right]$

Где $\sum x_i$ для каждой группы (уровня фактора A);

n_i - число наблюдений в каждой группе;

N - общее число вариантов.

k - число уровней (градаций) фактора.

При делении сумм квадратов, обозначаемых SS, на число степеней свободы получают **средние квадраты (варiances)** — ms, непосредственно измеряющие суммарную вариацию.

Оценка дисперсии каждой из групп связана с $n_i - 1$ степенью свободы, общее число степеней свободы равно $k \cdot \sum (n_i - 1) = n - k$, где n — число наблюдений.

Далее проводится проверка гипотезы $H_0 : \bar{X}_1 = \bar{X}_2 = \dots = \bar{X}_k$, т.е. утверждения, что все групповые средние не зависят от влияния

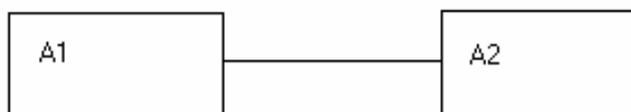
фактора A по величине соотношения: $F(k, n-k) = \frac{s_M^2 / (k - 1)}{s_b^2 / (n - k)}$ при

условии, что x_{ij} - независимые наблюдения над случайной величиной \bar{X} , распределенной нормально со средним \bar{X} и дисперсией σ^2 .

Если верна H_0 , то межгрупповая дисперсия (в генеральной совокупности) должна быть равна внутригрупповой, т. е. $H_0 : \sigma_M^2 = \sigma_b^2$. При этом вычисленное значение F меньше табличного при уровне значимости α . Следовательно гипотезу об отсутствии влияния фактора A не отклоняют.

Подготовка статистических массивов для дисперсионного анализа. Для дисперсионного анализа используем данные частичной совокупности, модифицированные при выполнении лабораторной работы по теме 6. Затем, используя схему однофакторного дисперсионного анализа, образуем

Схема однофакторного анализа



A1, A2 - градации диаметров или
высот

вторую градацию фактора высоты. Для этого следует значения высоты первой градации изменить на величину плюс минус сигма или средней ошибки. Затем проводим расчеты по приведенному примеру. На основе сравнения фактического критерия и табличного (приложение Е) делаем вывод о принятии или отклонении нулевой гипотезы.

Высота У	Высота	Высота Z	Высота
	A1		A2
Выписываются значения с номерами: 2, 4, 6, и т. д. до 50	К значениям высот с номерами 2,4,6,8.. прибавляется или отнимается средняя ошибка	Выписываются значения с номерами: 1, 3, 5, и т. д. до 49	К значениям высот с номерами 1, 3, 5.. прибавляется или отнимается средняя ошибка

Пример выходных данных в **Excel** после просчета однофакторного дисперсионного анализа.

ИТОГИ						
<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
в26 50	22	97,99	4,454091	1,719778		
д26 50	22	108,248	4,920364	1,045343		
Дисперсионный анализ						
<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
Между группами	2,391513	1	2,391513	1,729771	0,19557565	4,072660431
Внутри групп	58,06754	42	1,382561			
Итого	60,45906	43				

ВЫЧИСЛЕНИЕ ДИСПЕРСИЙ ПО ФОРМУЛАМ в Excel
Вычисление суммы квадратов отклонений

x_{i1}	x_{i2}	$(x_{i1} - X)^2$ общей	$(x_{i2} - X)^2$	$(x_{i1} - X_1)^2$ внутригрупповых	$(x_{i2} - X_2)^2$
0,4	0,5	0,2061	0,1384	0,1901	0,1384
0,5	0,6	0,1253	0,0740	0,1129	0,0740
0,6	0,6	0,0645	0,0740	0,0557	0,0740
0,6	0,7	0,0645	0,0296	0,0557	0,0296
0,6	0,7	0,0645	0,0296	0,0557	0,0296
0,6	0,7	0,0645	0,0296	0,0557	0,0296
0,6	0,7	0,0645	0,0296	0,0557	0,0296
0,6	0,7	0,0645	0,0296	0,0557	0,0296
0,7	0,7	0,0237	0,0296	0,0185	0,0296
0,7	0,7	0,0237	0,0296	0,0185	0,0296
0,7	0,8	0,0237	0,0052	0,0185	0,0052
0,8	0,8	0,0029	0,0052	0,0013	0,0052
0,9	0,8	0,0021	0,0052	0,0041	0,0052
0,9	0,8	0,0021	0,0052	0,0041	0,0052
0,9	0,8	0,0021	0,0052	0,0041	0,0052
0,9	0,9	0,0021	0,0008	0,0041	0,0008
0,9	0,9	0,0021	0,0008	0,0041	0,0008
0,9	0,9	0,0021	0,0008	0,0041	0,0008
1,1	1	0,0605	0,0164	0,0697	0,0164
1,1	1	0,0605	0,0164	0,0697	0,0164
1,1	1,1	0,0605	0,0520	0,0697	0,0520
1,1	1,3	0,0605	0,1832	0,0697	0,1832
1,2	1,3	0,1197	0,1832	0,1325	0,1832
1,2	1,3	0,1197	0,1832	0,1325	0,1832
1,3	1,5	0,1989	0,3944	0,2153	0,3944

Дисперсии

Средние по градациям фактора	X_1 0,836	X_2 0,872	Общая	Факториальная
	X 0,854		$(S_0)\Sigma$ 3,0361	$(S_M)\Sigma$ 3,0280
Средняя общая			в том числе обусловленная градацией фактора (от общей средней)	
			$S=(x_1 - X)^2+(x_2 - X)^2$ n	0,0162
			Неучтенная	S 0,0081

Сводная таблица

	Сумма квадрат	Число свободы	Дисперс	Оценка	Стандарт
Общее	3,0361	49	0,062457	1,010026	1,6
внутригрупповое	3,0280	48	0,063083	3,894033	4,03
По градациям опыта	0,0162	1	0,0162	96	251,8
Неучтенное(остаточное)	0,0081	48	0,000169		

Вопросы для самоконтроля

- 1 В чем состоит сущность метода дисперсионного анализа?
- 2 Как проводится оценка варьирования при дисперсионном анализе?

Лабораторная работа

Материалы и оборудование: данные замеров малой статистической совокупности, калькуляторы, ПЭВМ.

Цель. Определение значимости различия высот деревьев методом однофакторного дисперсионного анализа.

Ход работы

- 1 На основе основных положений темы подготовить к анализу градации замеров высот деревьев.
- 2 На основе основных положений темы провести расчеты соответствующие однофакторному дисперсионному комплексу.
- 3 На основе основных положений темы и при помощи «Анализ данных» провести расчеты по дисперсионному анализу, сохранить в Excel и распечатать.

ТЕМА 9 СТАТИСТИЧЕСКИЕ ФУНКЦИИ АНАЛИЗА ДАННЫХ НА ПЭВМ

- 1 Расчет основных статистических показателей для диапазона данных в приложении Excel.

Основные понятия по теме

Расчет основных статистических показателей для диапазона данных в приложении Excel. Ввести данные диаметров и соответствующих им высот из образованной частичной совокупности (см лаб. темы 1) в ячейки Excel. При помощи пакета «Анализ данных» в Excel (см табл.), «Описательная статистика» провести оценку среднего, его дисперсии, величины корреляции между частями совокупности.

Функции позволяющие выполнить расчет основных статистических показателей для диапазона данных

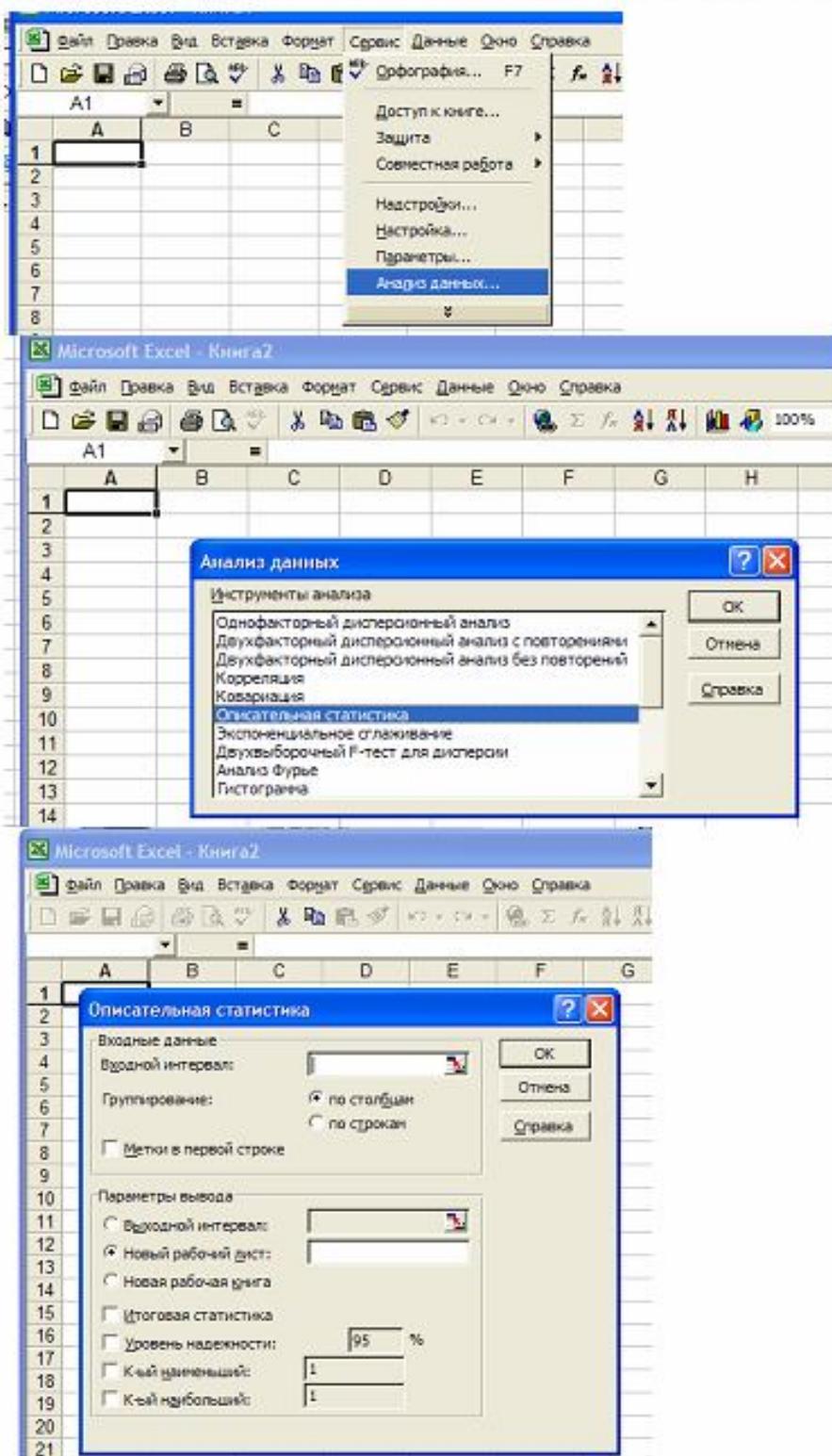
БЕТАОБР	ФРАСП	ПУАССОН
БЕТАРАСП	ФРАСПОБР	РАНГ
БИНОМРАСП	ZТЕСТ	РОСТ
ВЕЙБУЛЛ	ХИ2ОБР	СКОС
ВЕРОЯТНОСТЬ	ХИ2РАСП	СРГАРМ
ГАММАНЛОГ	ХИ2ТЕСТ	СРГЕОМ
ГАММАОБР	ЧАСТОТА	СРЗНАЧ
ГАММАРАСП	ЭКСПРАСП	СРЗНАЧА
ГИПЕРГЕОМЕТ	ЭКСЦЕСС	СРОТКЛ
ДИСП	ТЕНДЕНЦИЯ	СТАНДОТКЛОН
ДИСПА	ТТЕСТ	СТАНДОТКЛОНА
ДИСПР	УРЕЗСРЕДНЕЕ	СТАНДОТКЛОНП
ДИСПРА	ФИШЕР	СТАНДОТКЛОНПА
ДОВЕРИТ	ФИШЕРОБР	СТОШУХ
КВАДРОТКЛ	ФТЕСТ	СТЮДРАСП
КВАРТИЛЬ	ПЕРСЕНТИЛЬ	СТЮДРАСПОБР
КВПИРСОН	ПИРСОН	СЧЁТ
КОВАР	ПРЕДСКАЗ	СЧЁТЗ
КОРРЕЛ	ПРОЦЕНТРАНГ	ОТРЕЗОК
КРИТБИНОМ	НОМРАСП	ПЕРЕСТ
ЛГРФПРИБЛ	НОРМСТОБР	НАИМЕНЬШИЙ
ЛИНЕЙН	НОРМСТРАСП	НАКЛОН
ЛОГНОРМОБР	ОТРБИНОМРАСП	НОРМАЛИЗАЦИЯ
ЛОГНОМРАСП	МИНА	НОРМОБР
МАКС	МОДА	МЕДИАНА
МАКСА	НАИБОЛЬШИЙ	МИН

Для расширенного статистического анализа данных необходимо использовать пакет программ STATISTICA, которая представляет интегрированную систему анализа и обработки данных под Windows. пакета фирма Stat Soft. Inc.

Для анализа и обработки данных в пакете используются статистические модули, в которых объединены группы логически связанных статистических процедур.

Перечень основных модулей пакета и их вид на экране монитора представлен в приложении Ж.

Окна приложения Excel для настройки анализа данных



<i>Вариант 1</i>	
Среднее	4,920364
Стандартная ошибка	0,217981
Медиана	4,95
Мода	3,66
Стандартное отклонение	1,02242
Дисперсия выборки	1,045343
Экссесс	-0,22556
Асимметричность	0,438238
Интервал	3,846
Минимум	3,48
Максимум	7,326
Сумма	108,248
Счет	22

Функции для выполнения статистического анализа

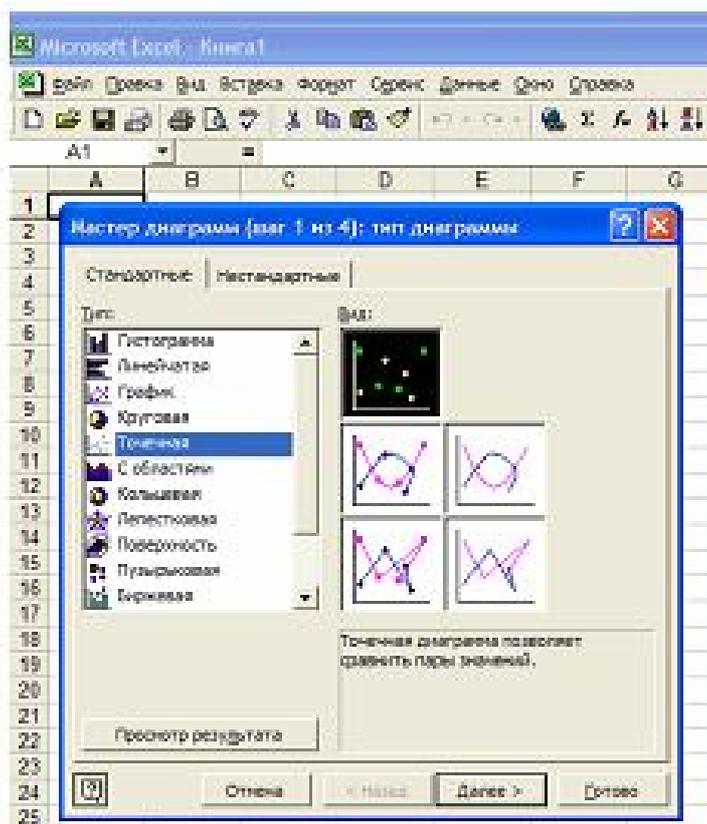
БЕТАОБР	ФРАСП	ПУАССОН
БЕТАРАСП	ФРАСПОБР	РАНГ
БИНОМРАСП	ZТЕСТ	РОСТ
ВЕЙБУЛЛ	ХИ2ОБР	СКОС
ВЕРОЯТНОСТЬ	ХИ2РАСП	СРГАРМ
ГАММАНЛОГ	ХИ2ТЕСТ	СРГЕОМ
ГАММАОБР	ЧАСТОТА	СРЗНАЧ
ГАММАРАСП	ЭКСПРАСП	СРЗНАЧА
ГИПЕРГЕОМЕТ	ЭКССЕСС	СРОТКЛ
ДИСП	ТЕНДЕНЦИЯ	СТАНДОТКЛОН
ДИСПА	ТТЕСТ	СТАНДОТКЛОНА
ДИСПР	УРЕЗСРЕДНЕЕ	СТАНДОТКЛОНП
ДИСПРА	ФИШЕР	СТАНДОТКЛОНПА
ДОВЕРИТ	ФИШЕРОБР	СТОШУХ
КВАДРОТКЛ	ФТЕСТ	СТЬЮДРАСП
КВАРТИЛЬ	ПЕРСЕНТИЛЬ	СТЬЮДРАСПОБР
КВПИРСОН	ПИРСОН	СЧЁТ
КОВАР	ПРЕДСКАЗ	СЧЁТЗ
КОРРЕЛ	ПРОЦЕНТРАНГ	ОТРЕЗОК
КРИТБИНОМ	НОМРАСП	ПЕРЕСТ
ЛГРФПРИБЛ	НОРМСТОБР	НАИМЕНЬШИЙ
ЛИНЕЙН	НОРМСТРАСП	НАКЛОН
ЛОГНОРМОБР	ОТРБИНОМРАСП	НОРМАЛИЗАЦИЯ
ЛОГНОМРАСП	МИНА	НОРМОБР
МАКС	МОДА	МЕДИАНА
МАКСА	НАИБОЛЬШИЙ	МИН

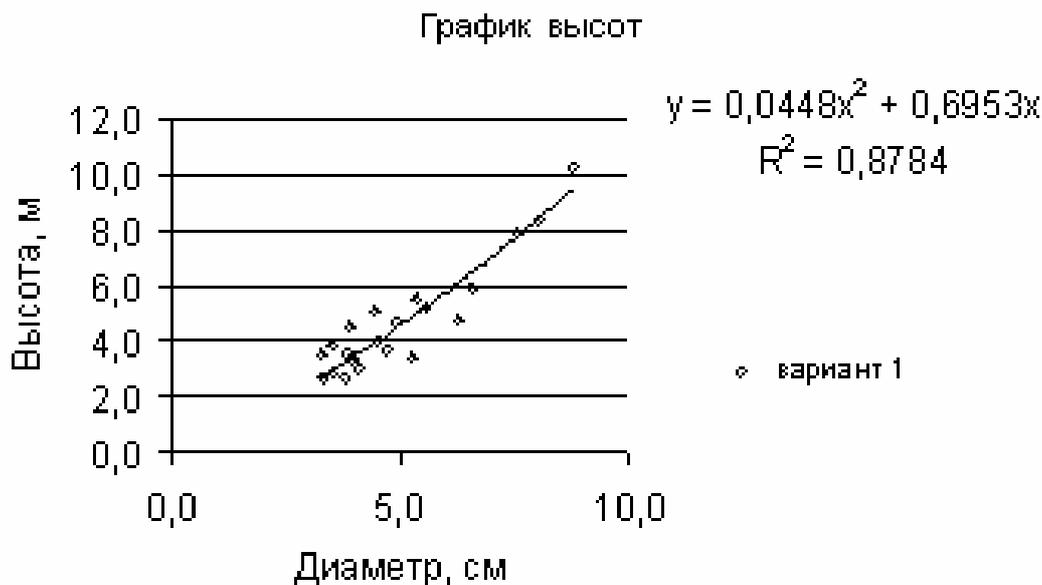
Пример выходных данных в Excel после просчета коэффициента корреляции

	<i>Д</i>	<i>Н</i>
<i>Д</i>	1	
<i>Н</i>	0,504430	1

При помощи «**Мастера диаграмм**» в Excel. выбрать вид диаграммы «**точечная**», ввести данные полученные в лабораторной работе темы 7 построить графики, отобразить на нем вид функции и коэффициент детерминации.

Пример вывода окна приложения Excel для построения графиков и выходных данных после просчета уравнения .





Вопросы для самоконтроля

- 1 Какие существуют встроенные статистические функции в приложении Excel?
- 2 Какова последовательность действий при обработке количественной и качественной информации на ПЭВМ?

Лабораторная работа

Материалы и оборудование: данные замеров малой статистической совокупности, калькуляторы, ПЭВМ.

Цель. Освоение методов обработки экспериментальных данных при помощи статистических программ и модулей на ПЭВМ.

Ход работы

- 1 На основе основных положений темы **получить** статистические показатели выборочной совокупности в приложении *Excel*.
- 2 На основе основных положений темы **получить** коэффициенты корреляции в приложении *Excel*.
- 3 На основе основных положений темы **получить** графики функций прямой, параболы и других уравнений в приложении *Excel*

Литература

- 1 Андронов, А. М. Теория вероятностей и математическая статистика: учебник для вузов / А. М. Андронов, Е. А. Копытов, Л. Я. Гринглаз. – СПб.: Питер, 2004. – 461 с.
- 2 Лакин, Г.Ф. Биометрия / Г.Ф., Лакин. — М., 1990. — 352 с.
- 3 Зайцев, Г. Н. Математическая статистика в экспериментальной ботанике / Г. Н., Зайцев. – М.: Наука, 1984.- 424 с.

Таблица случайных чисел

	00 04	05 09	10 14	15 19	20 24	25 29	30 34	35 39	40 44	45 49
0	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
1	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
2	85941	40756	82414	20150	13858	78030	16269	65978	01385	15345
3	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
4	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
5	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
6	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
7	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
8	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
9	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	72443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067
20	75884	12952	84318	95108	72305	64620	91318	89872	45375	85436
21	16777	37116	58550	42958	21460	43910	01175	87894	81378	10620
22	46230	43877	80207	88877	89380	32992	91380	03164	98656	59337
23	42902	66892	46134	01432	94710	23474	20423	60137	60609	13119
24	81007	00333	39693	28039	10154	95425	39220	19774	31782	49037
25	68089	01121	51111	72373	06902	74373	96199	97017	41273	21546
26	20411	67081	89950	16944	93054	87687	96693	87236	77054	33848
27	58212	13160	06468	15718	82627	76999	05999	58680	96739	63700
28	70577	42866	24969	61210	76046	67699	42054	12696	93758	03283
29	94522	74358	71659	62038	79643	79169	44741	05437	39038	13162
30	42626	86819	85651	88678	17401	03252	99547	32404	17918	62880
31	16051	33763	57194	16752	54450	19031	58580	47629	54132	60631
32	08244	27647	33851	44705	94211	46716	11738	55784	95374	72655
33	59497	04392	09419	89964	51211	04894	72882	17805	21896	83864
34	97155	13428	40293	09985	58434	01412	69124	82171	59058	82859
35	98409	66162	95763	47420	20792	61527	20441	39435	11859	41567
36	45476	84882	65109	96597	25930	66790	65706	61203	53634	22557
37	89300	69700	50741	30329	11658	23166	05400	66669	48708	03887
38	50051	95137	91631	66315	91428	12275	24816	68091	71710	33258
39	31753	85178	31310	89642	98364	02306	24617	09609	83942	22716
40	79152	53829	77250	20190	56535	18760	69942	77448	33278	48805
41	44560	38750	83635	56540	64900	42912	13953	79149	18710	68618
42	68328	83378	63369	71381	39564	05615	42451	64559	97501	65747
43	46939	38689	58625	08342	30459	85863	20781	09284	26333	91777
44	83544	86141	15707	86256	23068	13782	08467	89469	93842	55349
45	91621	00881	04900	54224	46177	55309	17852	27491	89415	23466
46	91896	67126	04151	03795	59077	11848	12630	98375	52068	60142
47	55751	62515	21108	80830	02263	29303	37204	96926	30506	09808
48	85156	87689	95493	88842	00664	55017	55539	17771	69448	87530
49	07521	56898	12236	60277	39102	62315	12239	07105	11844	01117

Схема вычисления смешанных моментов

Высо- та, м	Диаметр, см						Итого n_y	y_k	$n_y y_k$	$n_y y_k^2$	$n_{xy} x_k y_k$	$n_{xy} x_k^2 y_k$
	16	20	24	28	$32 M_x^1$	36						
30					2	6	2	10	30	90	0+ 6+4	10
29				3	6	5	4	18	36	72	-3+ 0+5+8	10
28		1		10	7	5	2	25	25	25	-3-10+0+5+4	4
$M_y^1=27$		1	2	5	3	2	1	14	0	0	-3-4-5+0+2+2	-8
26			4	7	2			13	-1	13	-8-7+0	-15
25			1	1				2	-2	8	-2-1	-3
24	1	2		1				4	-3	36	-4-6-1	-11
23	1	1	1					3	4	48	-4-3-2	-9
22	1			1				2	-5	50	-4-1	-5
21		2						2	-6	72	-6	-6
20	1							1	-7	49	-4	-4
Итого n_x	4	7	8	28	20	18	9	94	21	463		225
x_k	-4	-3	-2	-1	0	+1	+2				МОМЕНТЫ	
$n_x x_k$	-16	-21	-16	-28	0	+18	+18	-45			$m_{1x} = (\sum n_x x_k) / N = 45 / 94 = -0,479;$	
$n_x x_k^2$	64	63	32	28	0	18	36	241			$m_{2x} = \sum n_x x_k^2 / N = 21 / 94 = 0,223;$	
	-3	+1	0	+6	+6	+18	+6				$m_{3x} = \sum n_x x_k^3 / N = 241 / 94 = 2,564;$	
	-4	0	-4	+10	+12	+10	+8				$m_{4x} = \sum n_x x_k^4 / N = 463 / 94 = 4,926;$	
	-5	-6	-2	0	+7	+5	+2				$m_{5x} = \sum n_x x_k^5 / N = 225 / 94 = 2,394;$	
$n_{xy} x_k y_k$	-7	-4	-4	-7	0	0	0				$m_{xy} = \sum n_{xy} x_k y_k / N = 225 / 94 = 2,394;$	
		-12		-2	-2						$m_{2xy} = \sum n_{xy} x_k^2 y_k / N = 225 / 94 = 2,394;$	
				-3							$m_{3xy} = \sum n_{xy} x_k^3 y_k / N = 225 / 94 = 2,394;$	
				-5							$m_{4xy} = \sum n_{xy} x_k^4 y_k / N = 225 / 94 = 2,394;$	
$n_{xy} x_k^2 y_k$	-19	-21	-10	-1	23	33	16	225			$m_{xy^2} = (1/N) \sum [(n_{xy} x_k)^2 n_{yk}] = 1/94 \cdot 281,0 = 2,989;$	
$(n_{xy} x_k)^2 / n_x$	76	63	20	1	0	33	32	281,0				
	90,2	63,0	12,5	0	26,4	60,5	28,4	281,0				
\bar{y}_i	22,2	24,0	25,8	27,0	28,2	28,8	28,8					

Приложение В

Значение функции $\Psi(t) = 0,39894 / e^{0.5 \cdot t^2}$
 (перед числами в таблице необходимо ставить 0,)

t	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29430	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08938	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
4,0	00013	00009	00006	00004	00002	00002	00001	00001	00000	00000

Приложение Г

Значения функции $\varphi(x) = e^{-\frac{x^2}{2}} = e^{-(x_i - \bar{X})^2 / 2}$

x"	0	1	2	3	4	5	6	7	8	9
0,0	1,00000	99995	99980	99955	99920	99875	99820	99755	99680	99596
0,1	99501	99397	99283	99159	99025	98881	98728	98565	98393	98211
0,2	98020	97819	97609	97390	97161	96923	96676	96421	96156	95882
0,3	95600	95309	95009	94701	94384	94059	93725	93384	93034	92677
0,4	92312	91939	91558	91169	90774	90371	89960	89543	89119	88688
0,5	88250	87805	87354	86897	86433	85963	85487	85006	84518	84025
0,6	83527	83023	82514	82000	81481	80957	80429	79896	79358	78816
0,7	78270	77721	77167	76609	76048	75484	74916	74345	73771	73194
0,8	72615	72033	71448	70861	70272	69680	69087	68492	67896	67297
0,9	66698	66097	65495	64892	64288	63683	63078	62472	61866	61260
1,0	60653	60047	59440	58834	58228	57623	57018	56414	55811	55209
1,1	54607	54007	53408	52811	52215	51621	51028	50437	49848	49260
1,2	48675	48092	47511	46933	46357	45783	45212	44644	44078	43516
1,3	42956	42399	41845	41294	40748	40202	39661	39123	38589	38058
1,4	37531	37007	36488	35971	35459	34950	34445	33944	33447	32954
1,5	32465	31980	31499	31023	30550	30082	29618	29158	28702	28251
1,6	27804	27361	26923	26489	26059	25634	25213	24797	24385	23978
1,7	23575	23176	22782	22392	22007	21627	21250	20879	20511	20148
1,8	19790	19436	19086	18741	18400	18064	17732	17404	17081	16762
1,9	16447	16137	15831	15529	15232	14938	14649	14364	14083	13806
2,0	13534	13265	13000	12740	12483	12230	11982	11737	11496	11258
2,1	11025	10795	10569	10347	10129	09914	09702	09495	09290	09090
2,2	08892	08698	08508	08320	08137	07956	07779	07604	07433	07265
2,3	07101	06939	06780	06624	06471	06321	06174	06030	05888	05750
2,4	05613	05480	05349	05221	05096	04972	04852	04734	04618	04505

x"	0	1	2	3	4	5	6	7	8	9
2,5	04394	04285	04179	04079	03972	03873	03775	03679	03586	03494
2,6	03405	03317	03232	03148	03066	02986	02908	02831	02757	02683
2,7	02612	02542	02474	02408	02343	02279	02217	02157	02098	02040
2,8	01984	01929	01876	01823	01772	01723	01674	01627	01581	01536
2,9	01492	01449	01408	01367	01328	01289	01252	01215	01179	01145
3,0	01111	01078	01046	01015	00985	00955	00926	00898	00871	00845
3,1	00819	00794	00769	00746	00723	00700	00679	00658	00637	00617
3,2	00598	00579	00560	00543	00525	00509	00492	00477	00461	00446
3,3	00432	00418	00404	00391	00378	00366	00354	00342	00331	00320
3,4	00309	00299	00289	00279	00269	00260	00251	00243	00235	00227
3,5	00219	00211	00204	00197	00190	00183	00177	00171	00165	00159
3,6	00153	00148	00143	00138	00133	00128	00123	00119	00115	00110
3,7	00106	00103	00099	00095	00092	00088	00085	00082	00079	00076
3,8	00073	00070	00068	00065	00063	00060	00058	00056	00054	00052
3,9	00050	00048	00046	00044	00043	00041	00039	00038	00036	00035
4,0	00034	00022	00015	00010	00006	00004	00003	00002	00001	00001

36	15822	24755	73995	69	09566	02326	57202	02	05186	11319	25064	35	02522	14949	00915
37	15608	24015	73961	70	09405	01759	56316	03	05082	11565	24109	36	02463	14937	01485
38	15395	23276	73890	71	09246	01200	55422	04	04980	11801	23160	37	02406	14919	02040
39	15183	22537	73784	72	09089	00650	54516	05	04879	12028	22220	38	02349	14896	02582
40	14973	21800	73642	73	08933	+0,00110	53599	06	04780	12245	21287	39	02294	14868	03109
41	14764	21065	73466	74	08780	-0,00422	52671	07	04682	12454	20363	40	02239	14834	03623
42	14556	20331	73256	75	08628	00944	51733	08	04586	12653	19448	41	02186	14795	04123
43	14350	19600	73012	76	08478	01456	50785	09	04491	12844	18542	42	02134	14752	04608
44	14146	18871	72736	77	08329	01959	49829	10	04398	13024	17646	43	02083	14703	05079
45	13943	18145	72427	78	08183	02453	48865	11	04307	13196	16759	44	02033	14650	05537
46	13742	17423	72067	79	08038	02937	47893	12	04217	13359	15883	45	01984	14593	05981
47	13542	16704	71716	80	07895	03411	46915	13	04128	13513	15017	46	01936	14531	06411
48	13344	15988	71315	81	07754	03875	45932	14	04041	13659	14162	47	01888	14464	06828
49	13147	15277	70885	82	07614	04329	44943	15	03955	13797	13318	48	01842	14394	07231
50	12952	+0,14571	-0,70425	83	07477	04774	43950	16	03871	13926	12486	49	01797	14320	07621
51	12758	13869	69937	84	07341	05208	42953	17	03788	14046	11665	2,50	0,01753	-0,14242	+0,07997
52	12566	13172	69423	85	07206	05633	41953	18	03706	14159	10858	51	01709	14160	08360
53	12376	12481	68881	86	07074	06047	40950	19	03626	14263	10059	52	01667	14075	08710
54	12188	11795	68314	87	06943	06452	39946	20	03547	14360	09274	53	01625	13986	09047
55	12001	11114	67721	88	06814	06846	38940	21	03470	14449	08502	54	01585	13894	09372
56	11816	10440	67104	89	06687	07231	37934	22	03394	14530	07743	55	01545	13798	09683
57	11632	09772	66463	90	06562	07605	36928	23	03319	14604	06996	56	01506	13700	09982
58	11450	09111	65799	91	06438	07969	35923	24	03246	14670	06263	57	01468	13599	10268
59	11270	08456	65113	92	06316	08323	34918	25	03174	14729	05542	58	01431	13495	10542
60	11092	07809	64405	93	06195	08667	33916	26	03103	14781	04835	59	01394	13388	10804
61	10915	07168	63677	94	06077	09002	32916	27	03034	14826	04141	60	01358	13279	11053
62	10741	06535	62928	95	05959	09326	31919	28	02965	14864	03461	61	01323	13167	11291
63	10567	05910	62161	96	05844	09640	30925	29	02898	14895	02794	62	01289	13053	11517
64	10396	05292	61375	97	05730	09944	29936	30	02833	14920	02141	63	01256	12937	11732
65	10226	04682	60571	98	05618	10239	28950	31	02768	14938	01502	64	01223	12818	11935
66	10059	04081	59751	99	05508	10523	27970	32	02705	14950	00877	65	01191	12698	12127
67	09893	03487	58914	2,00	0,05399	-0,10798	-0,26996	33	02643	14956	-0,00265	66	01160	12576	12308
68	09728	02903	58063	01	05292	11063	26027	34	02582	14955	+0,00332	67	01130	12452	12479

X ^{II}	f ^(x)	f ^{III} (x)	f ^V (x)	X ^{II}	f ^(x)	f ^{III} (x)	f ^V (x)	X ^{II}	f ^(x)	f ^{III} (x)	f ^V (x)	X ^{II}	f ^(x)	f ^{III} (x)	f ^V (x)	X ^{II}	f ^(x)	f ^{III} (x)	f ^V (x)
68	01100	12326	12638	01	00430	07845	13214	34	00151	04109	09128	67	00047	01823	04915	71	00041	01636	04497
69	01071	12189	12787	02	00417	07713	13128	35	00146	04018	08987	68	00046	01774	04808	72	00039	01590	04396
70	01042	12071	12926	03	00405	07582	13038	36	00141	03929	08847	69	00044	01727	04703	73	00038	01547	04297
71	01014	11941	13055	04	00393	07452	12944	37	00136	03841	08707	70	00042	01680	04599	74	00037	01504	04200
72	00987	11810	13174	05	00381	07323	12847	38	00132	03755	08567	71	00041	01636	04497	75	00036	01463	04103
73	00961	11677	13283	06	00370	07195	12747	39	00127	03670	08428	72	00039	01590	04396	76	00035	01422	04009
74	00935	11544	13383	07	00358	07068	12643	40	00123	03586	08290	73	00038	01547	04297	77	00033	01383	03916
75	00909	11410	13473	08	00348	06943	12536	41	00119	03504	08151	74	00037	01504	04200	78	00031	01344	03824
76	00885	11274	13555	09	00337	06818	12426	42	00115	03423	08014	75	00036	01463	04103	79	00030	01306	03734
77	00861	11139	13627	10	00327	06694	12313	43	00111	03344	07877	76	00034	01422	04009	80	00029	01269	03646
78	00837	11002	13691	11	00317	06571	12198	44	00107	03266	07741	77	00033	01383	03916	81	00028	01233	03559
79	00814	10865	13746	12	00307	06450	12080	45	00104	03189	07606	78	00031	01344	03824	82	00027	01198	03473
80	00792	10727	13793	13	00298	06330	11960	46	00100	03114	07471	79	00030	01306	03734	83	00026	01164	03389
81	00770	10589	13832	14	00288	06211	11838	47	00097	03040	07338	80	00029	01269	03646	84	00025	01130	03307
82	00748	10450	13863	15	00279	06093	11714	48	00094	02967	07205	81	00028	01233	03559	85	00024	01098	03226
83	00727	10312	13886	16	00271	05977	11588	49	00090	02895	07074	82	00027	01198	03473	86	00023	01066	03146
84	00707	10173	13902	17	00262	05861	11460	3,50	0,00087	-0,02825	+0,06943	83	00026	01164	03389	87	00022	01035	03068
85	00687	10034	13910	18	00254	05747	11330	51	00084	02757	06814	84	00025	01130	03307	88	00021	01004	02991
86	00668	99895	13912	19	00246	05635	11199	52	00081	02689	06685	85	00024	01098	03226	89	00020	00975	02916
87	00649	99755	13906	20	00238	05523	11066	53	00079	02623	06558	86	00023	01066	03146	90	00020	00946	02842
88	00631	99616	13894	21	00231	05413	10933	54	00076	02558	06432	87	00022	01035	03068	91	00019	00918	02770
89	00613	99478	13875	22	00224	05305	10798	55	00073	02494	06308	88	00021	01004	02991	92	00018	00891	02699
90	00595	99339	13850	23	00216	05198	10662	56	00071	02432	06184	89	00021	00975	02916	93	00018	00864	02630
91	00578	99201	13819	24	00210	05092	10525	57	00068	02370	06062	90	00020	00946	02842	94	00017	00838	02562
92	00562	99063	13782	25	00203	04987	10387	58	00066	02310	05941	91	00019	00918	02770	95	00016	00813	02495
93	00545	98925	13739	26	00196	04884	10249	59	00063	02252	05821	92	00018	00891	02699	96	00016	00788	02430
94	00530	98788	13691	27	00190	04782	10110	60	00061	02194	05703	93	00018	00864	02630	97	00015	00764	02366
95	00514	98651	13638	28	00184	04682	99970	61	00059	02138	05586	94	00017	00838	02562	98	00014	00741	02303
96	00499	98515	13579	29	00178	04583	99830	62	00057	02082	05471	95	00016	00813	02495	99	00014	00718	02242
97	00485	98380	13515	30	00172	04485	99690	63	00055	02028	05357	96	00016	00788	02430	4,00	0,00013	-0,00696	+0,02181
98	00470	98245	13446	31	00167	04389	99549	64	00053	01975	05244	97	00015	00764	02366				
99	00457	98111	13373	32	00161	04294	99409	65	00051	01923	05133	98	00014	00741	02303				
3,00	0,00443	-0,07977	+0,13296	33	00156	04201	99268	66	00049	01873	05023	99	00014	00718	02242				

Приложение Е

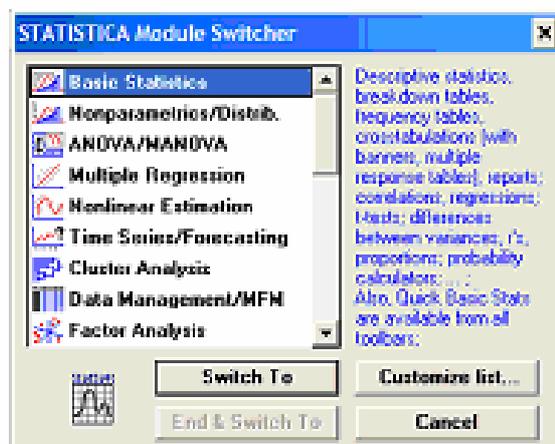
Стандартные значения критерия Фишера

число шагов дисперсии	доверительный уровень 95%											
	степени свободы большей дисперсии											
	1	2	3	4	5	6	8	12	16	24	50	∞
1	161,4	199,5	215,7	224,6	230,2	234	238,9	243,9	246,5	249	251,8	254,3
2	18,51	19	19,16	19,25	19,3	19	19,37	19,41	19,43	19,45	19,47	19,5
3	10,13	9,55	9,28	9,12	9,01	9	8,84	8,74	8,69	8,64	8,58	8,53
4	7,71	6,94	6,59	6,39	6,26	6	6,04	5,91	5,84	5,77	5,7	5,63
5	6,61	5,79	5,41	5,19	5,05	5	4,82	4,68	4,6	4,53	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4	4,15	4	3,92	3,84	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	4	3,73	3,57	3,49	3,41	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	4	3,44	3,28	3,2	3,12	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3	3,23	3,07	2,98	2,9	2,8	2,71
10	4,96	4,1	3,71	3,48	3,33	3	3,07	2,91	2,82	2,74	2,64	2,54
11	4,84	3,98	3,59	3,36	3,2	3	2,95	2,79	2,7	2,61	2,5	2,4
12	4,75	3,88	3,49	3,26	3,11	3	2,85	2,69	2,6	2,5	2,4	2,3
13	4,67	3,8	3,41	3,18	3,02	3	2,77	2,6	2,51	2,42	2,32	2,21
14	4,6	3,74	3,34	3,11	2,96	3	2,7	2,53	2,44	2,35	2,24	2,13
15	4,54	3,68	3,29	3,06	2,9	3	2,64	2,48	2,39	2,29	2,18	2,07
16	4,49	3,63	3,24	3,01	2,85	3	2,59	2,42	2,33	2,24	2,13	2,01
17	4,45	3,59	3,2	2,96	2,81	3	2,55	2,38	2,29	2,19	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	3	2,51	2,34	2,25	2,15	2,04	1,92
19	4,38	3,52	3,13	2,9	2,74	3	2,48	2,31	2,21	2,11	2	1,88
20	4,35	3,49	3,1	2,87	2,71	3	2,45	2,28	2,18	2,08	1,96	1,84
21	4,32	3,47	3,07	2,84	2,68	3	2,42	2,25	2,15	2,05	1,93	1,81
22	4,3	3,44	3,05	2,82	2,66	3	2,4	2,23	2,13	2,03	1,91	1,78
23	4,28	3,42	3,03	2,8	2,64	3	2,38	2,2	2,11	2	1,88	1,76
24	4,26	3,4	3,01	2,78	2,62	3	2,36	2,18	2,09	1,98	1,86	1,73
25	4,24	3,38	2,99	2,76	2,6	2	2,34	2,16	2,07	1,96	1,84	1,71
30	4,17	3,32	2,92	2,69	2,53	2	2,27	2,09	1,99	1,89	1,76	1,62
40	4,08	3,23	2,84	2,61	2,45	2	2,18	2	1,9	1,79	1,66	1,51
50	4,03	3,18	2,79	2,56	2,4	2	2,13	1,95	1,85	1,74	1,6	1,44
60	4	3,15	2,76	2,52	2,37	2	2,1	1,92	1,81	1,7	1,56	1,39
70	3,98	3,13	2,74	2,5	2,35	2	2,07	1,89	1,79	1,67	1,53	1,35
80	3,96	3,11	2,72	2,49	2,33	2	2,06	1,88	1,77	1,65	1,51	1,32
90	3,95	3,1	2,71	2,47	2,32	2	2,04	1,86	1,76	1,64	1,49	1,3
100	3,94	3,09	2,7	2,46	2,3	2	2,03	1,85	1,75	1,63	1,48	1,28
150	3,9	3,06	2,66	2,43	2,27	2	2	1,82	1,71	1,59	1,44	1,22
200	3,89	3,04	2,65	2,42	2,26	2	1,98	1,8	1,69	1,57	1,42	1,19
300	3,87	3,03	2,64	2,41	2,25	2	1,97	1,79	1,68	1,55	1,39	1,15
400	3,86	3,02	2,63	2,4	2,24	2	1,96	1,78	1,67	1,54	1,38	1,13
500	3,86	3,01	2,62	2,39	2,23	2	1,96	1,77	1,66	1,54	1,38	1,11
1000	3,85	3	2,61	2,38	2,22	2	1,95	1,76	1,65	1,53	1,36	1,08
∞	3,84	2,99	2,6	2,37	2,21	2	1,94	1,75	1,64	1,52	1,35	1

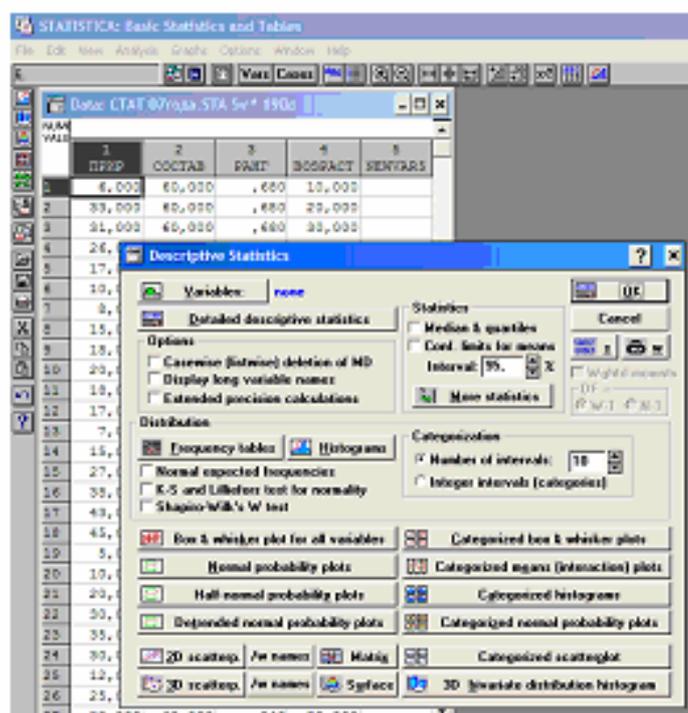
Основные статистические модули пакета STATISTICA

Наименование модуля	Назначение
ANOVA/MANOVA	Одномерный и многомерный, дисперсионный и ковариационный анализ
Basic Statistic and Tables	Описательные статистики, группировка, вычисление корреляции, t - критерий
Cluster Analysis	Кластерный анализ
Factor Analysis	Факторный анализ
Multiple Regression	Линейный регрессионный анализ
Nonparametric distribution	Непараметрическая статистика
Times Series / Forecasting	Анализ и прогнозирование временных рядов

Окно переключений модулей STATISTICA



Окно пакета описательные статистики в Statistica 5



Учебное издание

Ефименко Владимир Макарович

ЛЕСНАЯ БИОМЕТРИЯ

**ПРАКТИЧЕСКОЕ ПОСОБИЕ
по выполнению лабораторных работ
для студентов специальности
1-75 01 01 «Лесное хозяйство»**

В авторской редакции

Подписано в печать 10.06.2007г. (43) Формат 60x84 1/16
Бумага писчая № 1. Гарнитура «Таймс». Усл. п. л. 4,0
Уч.-изд. л. 3,2. Тираж 25 экз.

Отпечатано с оригинал-макета на ризографе
в учреждении образования
«Гомельский государственный университет
Имени Франциска Скорины»
Лицензия № 02330/0056611 от 16.02.04.
246019, г.Гомель, ул. Советская, 104.

