

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
КУРГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА «ИНФОРМАТИКА»

***РЕШЕНИЕ ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ
В СРЕДЕ MATHCAD***

Методические указания
к проведению лабораторных занятий
по курсам «Спец. главы информатики», «Прикладное программное
обеспечение»
для студентов специальностей 120100,120200,120500,150100,
150200,150300,330100,030500

Кафедра: «Информатика»

Дисциплины: «Спец. главы информатики», «Прикладное программное обеспечение»
(специальности 120100, 120200, 120500, 150100, 150200, 150300, 330100, 030500)

Составила:

старший преподаватель Сысолятина Л. Г.

Утверждены на заседании кафедры « » _____ 2004 г.

Рекомендованы редакционно-издательским советом университета

« » _____ 2004 г.

РЕШЕНИЕ ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ В СРЕДЕ MATHCAD

1. ИСПОЛЬЗУЕМЫЕ ИНСТРУМЕНТЫ MATHCAD

Познакомимся с основными функциями Mathcad, предназначенными для решения задач математической статистики, а также с методами ввода данных для последующей статистической обработки. Попутно будут рассматриваться основные понятия математической статистики, постановка задач, алгоритмы и методы их решения.

Ввод и вывод файлов данных

При решении практических задач статистического анализа данных чаще всего приходится иметь дело с большими объемами исходной информации. Зачастую такие данные представляют собой заранее введенные в файл аппаратными средствами экспериментальные данные, подготовленные специальными приложениями и сохраненные в файле таблицы чисел. Ниже будет рассказано о том, как в Mathcad можно генерировать последовательности случайных чисел. Такие последовательности позволяют имитировать результаты реальных измерений той или иной случайной величины. При многократном использовании больших массивов данных их удобно хранить в файлах на диске.

Mathcad предоставляет пользователю специальные функции ввода данных из файла на диске и вывода данных в файл, т. е. функции доступа к файлам – READ, WRITE, APPEND, READPRN, WRITEPRN, APPENDPRN. Подробное описание этих функций и правила работы с ними можно найти в литературе по пакету, во встроенном в систему справочнике, в руководстве пользователя.

Познакомимся подробнее с функциями READ (file) и WRITE (file), предназначенными соответственно для чтения и записи числового значения. Файл данных для Mathcad – это файл чисел, записанных в формате ASCII, разделенных пробелом, запятой или символом конца строки. Числа могут быть целыми или с плавающей запятой, записанными с десятичной точкой или в экспоненциальной форме. При обращении к файлу Mathcad по умолчанию обращается в ту папку (каталог, директорию), из которой загружался рабочий документ или в которую документ последний раз загружался. Однако можно работать с файлами из любых папок, указывая полное имя файла. В приведенных ниже примерах всегда будет указываться полное имя файла.

Функция READ (file) считывает значение из файла и присваивает его переменной. Поскольку чаще всего читаются массивы чисел, обращение к функции записывается следующим образом: $x_i := \text{READ}(\text{file})$.

Предположим, что на диске **c:** в папке **tmp** в файле с именем **data.txt** записаны 20 различных чисел, подготовленных текстовым процессором и сохраненных в указанном файле. Ниже представлен фрагмент рабочего

документа Mathcad, в котором этот файл прочитан (в приведенном фрагменте отображены первые восемь чисел.).

$i := 0..19$ $x_i := \text{READ} ("c:\tmp\data.txt")$

$xT =$		0	1	2	3	4	5	6	7	8
	0	4.591	4.326	4.868	6.018	5.181	5.513	5.74	4.984	5.279

Указание. Поскольку в Mathcad массив – это вектор-столбец, запишите в рабочем документе xT , чтобы вывести массив x в виде строки. Для того чтобы просмотреть все данные, щелкните по полю вывода элементов массива и просмотрите содержимое массива с помощью линеек прокрутки.

Функция **WRITE(file)** записывает в файл на диске числовое значение переменной. Поскольку, как правило, записываются массивы чисел, то чаще всего она указывается следующим образом: **WRITE(file) := x_i** . Если файла с указанным именем не существует, то он будет создан; если такой файл есть, то при записи предыдущая информация будет потеряна.

Ниже приведен фрагмент рабочего документа Mathcad, в котором массив, сформированный в предыдущем фрагменте, записан в файл с именем **data1.txt** в папке **tmp** на диске **c:**, а затем (для проверки) прочитан и выведен в рабочий документ.

$i := 0..19$

$\text{WRITE} ("c:\tmp\data1.txt") := x_i$ $y_i := \text{READ} ("c:\tmp\data1.txt")$

$yT =$		0	1	2	3	4	5	6	7	8
	0	4.591	4.326	4.868	6.018	5.181	5.513	5.74	4.984	5.279

Указание: Здесь продемонстрирована работа двух функций доступа к файлам, позволяющих простейшим способом читать и записывать файлы. Для того чтобы аккуратно работать с файлами, необходимо обязательно ознакомиться с полным описанием того, как с помощью этих функций происходит обращение к файлу.

Функция вычисления выборочных характеристик

Первичная обработка данных состоит обычно в отыскании максимального $x_{\max}$ и минимального $x_{\min}$ значений выборки, а также в построении вариационного ряда – массива выборочных значений, занумерованных (записанных) в порядке возрастания. Для выполнения этих вычислений в Mathcad предназначены соответственно функции **$\max(A)$** , **$\min(A)$** и **$\text{sort}(A)$** .

Кроме того, Mathcad имеет шесть функций, вычисляющих точечные оценки параметров распределения случайной величины. В последующих разделах главы даны все необходимые определения и описаны методы получения оценок. Здесь приведем только определения функций и правила обращения к ним.

Следующие четыре функции вычисляют числовые характеристики выборки, содержащиеся в массиве A размерности $m \times n$.

Функция **$\text{mean}(A)$** вычисляет значение выборочного среднего:

$$mean(A) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} A_{ij}.$$

Функция $var(A)$ вычисляет смещенную точечную оценку дисперсии, называемую *выборочной дисперсией* (для комплексных выборочных значений выборочная дисперсия вычисляется по формуле $var(A) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |A_{ij} - mean(A)|^2$):

$$var(A) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A_{ij} - mean(A))^2.$$

Функция $stdev(A)$ определяет среднеквадратичное отклонение, т. е. $stdev(A) = \sqrt{var(A)}$.

Функция $median(A)$ вычисляют медиану – величину, меньше и больше которой в выборке содержится одинаковое количество элементов.

Еще две функции предназначены для вычисления числовых характеристик двумерного случайного вектора, выборочные значения двух, компонент которого расположены соответственно в массивах A и B размерности $m \times n$.

Функция $cvar(A, B)$ вычисляет значение выборочной ковариации: (для комплексных выборочных значений выборочная ковариация вычисляется по формуле

$$cvar(A, B) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A_{ij} - mean(A))(\overline{B_{ij} - mean(B)}),$$

в которой горизонтальная черта – знак комплексно сопряженного числа)

$$cvar(A, B) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (A_{ij} - mean(A))(B_{ij} - mean(B)).$$

Функция $corr(A, B)$ определяет коэффициент корреляции:

$$corr(A, B) = \frac{cvar(A, B)}{\sqrt{var(A)} \sqrt{var(B)}}.$$

Ниже представлен фрагмент рабочего документа Mathcad, в котором выполнены следующие действия: прочитаны из файла на диске некоторые экспериментальные данные, а затем с помощью описанных выше функций вычислены числовые характеристики этих выборочных данных.

i := 0..19

```

xi := READ("c:\tmp\datax.txt")   yi := READ("c:\tmp\datay.txt")
Xmin := min(x)   Xmin=3,72       Xmax := max(x)   Xmax = 8.048
Ymin := (y)   Ymin=3.72         Ymax := max(y)   Ymax = 8.048
X := sort(x)           y:= sort(y)
Xmean := mean(x)       Xvar := var(x)           Xstdev := stdev(x)
Xmean := 5.026         Xvar := 0.888           Xstdev = 0.943
Ymean := mean(y)       Yvar := var(y)           Ystdev : stdev(y)
Ymean := 5.026         Yvar := 0.888           Ystdev =0.943
covxy := cvar (x, y)   corrxy := corr (x, y)

```

covxy = 0.888

corxy = 1

Указание. Выборочные данные прочитаны из файлов **datax.txt** и **datay.txt**, расположенных на диске **c:** в папке **tmp**, в массивы (переменные) соответственно **x** и **y**. Для экономии места прочитанные выборочные значения выведены в рабочий документ в виде строки.

Построение эмпирических распределений

Наиболее наглядной формой графического представления выборок является *гистограмма*. В Mathcad для построения гистограмм предназначена функция *hist* (Δ , A). Для того чтобы построить *гистограмму*, нужно сначала сгруппировать выборочные данные, записанные в массиве A , и сохранить граничные точки интервалов группировки в векторе Δ , размерность которого равна числу интервалов. Результат вычислений функций *hist* (Δ , A) – вектор, каждый элемент которого равен количеству выборочных значений, попадающих в соответствующий интервал группировки.

Размерность вектора *hist* (Δ , A) совпадает с размерностью вектора Δ и равна числу интервалов группировки. Если $\delta_1, \delta_2, \dots, \delta_m$ – длины интервалов группировки, а $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ – их середины и $h_j = \frac{n_j}{n}$ – относительные частоты попадания наблюдений в j -й интервал группировки, то можно построить изображение ступенчатой функции $f(x) = \frac{h_j}{\delta_j}, x \in \Delta_j, j = 1, 2, \dots, m$, в виде столбчатой диаграммы, которая называется *гистограммой*.

Используя функцию *hist* (Δ , A), можно построить полигон частот – ломаную линию, соединяющую точки с абсциссами, равными серединам интервалов группировки, и ординатами, равными соответствующим частотам.

Другие способы графического представления выборок описаны в разд. 2.

Ниже приведен фрагмент рабочего документа Mathcad, содержащий гистограммы и полигоны частот, построенные для выборок, описанных в приведенном выше фрагменте. Здесь же представлены точечные графики каждой выборки и точечный график выборки из двумерного случайного вектора, выборочные значения, компоненты которого находятся соответственно в массивах x и y .

i := 0..19

$x_i := \text{READ}("c:\text{tmp}\backslash\text{datax.txt}")$ $y_i := \text{READ}("c:\text{tmp}\backslash\text{datay.txt}")$

$X_{\min} := \min(x)$ $X_{\min} = 3,72$ $X_{\max} := \max(x)$ $X_{\max} = 8.048$

$Y_{\min} := \min(y)$ $Y_{\min} = 3.72$ $Y_{\max} := \max(y)$ $Y_{\max} = 8.048$

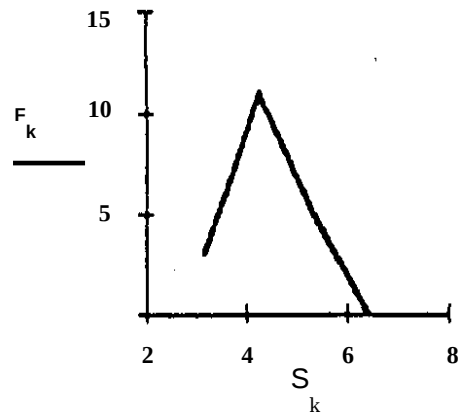
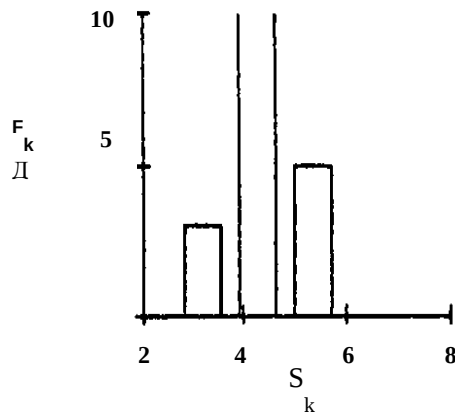
$x := \text{sort}(x)$ $y := \text{sort}(y)$

$R_x := X_{\max} - X_{\min}$ $R_y := Y_{\max} - Y_{\min}$ **$n := 20$**

$R_x = 4.328$ $R_y = 4.328$

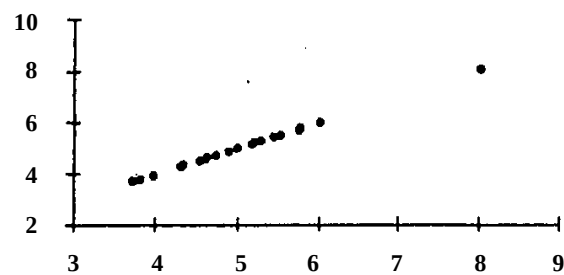
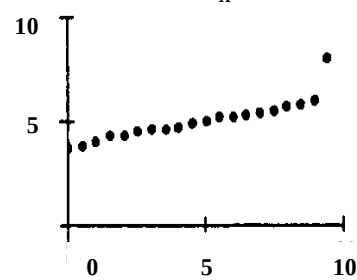
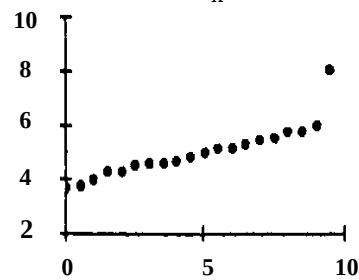
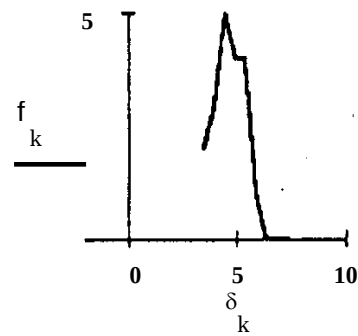
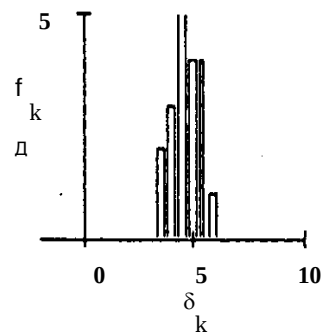
$$mx := 4 \quad \Delta := \frac{Rx}{mx} \quad j := 0..mx \quad k := 0..mx - 1$$

$$\delta_j := Xmin + \frac{\Delta}{2} (2j - 1) \quad f := \text{hist}(\delta, x) \quad \Delta = 1.082$$



$$my := 9 \quad \Delta := \frac{Ry}{my} \quad j = 0..my \quad k := 0..my - 1$$

$$\delta_j := Xmin + \frac{\Delta}{2} \cdot (2j - 1) \quad f := \text{hist}(\delta, y) \quad \Delta = 0.481$$



Указание. Прежде чем приступать к группировке выборок, нужно их упорядочить с помощью функции *sort*. Перед обращением к функции *hist* следует вычислить середины

интервалов группировки и присвоить их значения элементам массива x . Используйте окна настройки параметров изображения гистограмм и графиков Formatting Currently Selected X-Y Plot.

Моделирование выборок из стандартных распределений

Mathcad обладает богатой библиотекой встроенных функций, предназначенных для генерирования выборок из генеральных совокупностей с наиболее распространенными стандартными распределениями. Например, для генерации нормального распределения предназначена функция $rnorm(k, \mu, \sigma)$, значением которой является вектор, содержащий k выборочных значений нормально распределенной случайной величины с математическим ожиданием $M\xi = \mu$ и дисперсией $D\xi = \sigma^2$.

Ниже приведен список функций Mathcad, генерирующих выборки:

- ▣ Бета-распределение: $rbeta(k, s_1, s_2)$.
- ▣ Биноминальное распределение: $rbinom(k, n, p)$.
- ▣ Распределение Коши: $rcauchy(k, l, s)$.
- ▣ χ^2 -распределение: $rchisq(k, d)$.
- ▣ Экспоненциальное распределение: $rexp(k, r)$.
- ▣ Распределение Фишера (F -распределение): $rF(k, m, n)$.
- ▣ Гамма-распределение: $rgamma(k, s)$.
- ▣ Геометрическое распределение: $rgeom(k, p)$.
- ▣ Логнормальное распределение: $rlnorm(k, \mu, \sigma)$.
- ▣ Логистическое распределение: $rlogis(k, l, s)$.
- ▣ Отрицательное биномиальное распределение: $rnbinom(k, n, p)$.
- ▣ Нормальное распределение: $rnorm(k, \mu, \sigma)$.
- ▣ Распределение Пуассона: $rpois(k, \lambda)$.
- ▣ Распределение Стьюдента: $rt(k, d)$.
- ▣ Равномерное распределение: $runif(k, a, b)$.
- ▣ Распределение Вейбулла: $rweibull(k, s)$.

Выборка x , которая исследовалась в приведенных выше фрагментах, - выборка 20 значений из стандартного нормального распределения. Ниже приведен фрагмент рабочего документа Mathcad, содержащий функции построения этой выборки и сохранения ее в файле **datx.txt** на диске **c:** в папке **tmp**.

$x := rnorm(20, 0, 1) \quad i := 0..19 \quad \text{WRITE}("c:\tmp\datx.txt") := x_i$

2. ОСНОВНЫЕ ЗАДАЧИ СТАТИСТИКИ. ВЫБОРКИ. ГИСТОГРАММЫ. ПОЛИГОНЫ ЧАСТОТ

Математическая статистика в основном занимается изучением случайных величин и случайных событий по результатам наблюдений. Ее главная задача – извлечь максимум информации из эмпирических данных. Важнейшими понятиями математической статистики являются

генеральная совокупность и выборка.

Важнейшими понятиями математической статистики являются генеральная совокупность и выборка.

Генеральная совокупность – это вероятностное пространство с определенной на нем случайной величиной ξ . Функцию распределения этой случайной величины $F_{\xi}(x)$ часто называют *теоретической функцией распределения*, хотя более правильным представляется другой термин – *истинная функция распределения*, в отличие от эмпирической (экспериментальной, приближенной) функции распределения, которая будет определена ниже.

В результате проведения n экспериментов со случайной величиной ξ получаем n выборочных значений $x_i, i=1, 2, \dots, n$. Вся совокупность этих значений называется *выборкой*.

Выборка – это, вообще говоря, случайный вектор: если в одной серии из n испытаний получена выборка $(x_1, x_2, \dots, x_n)$, то в другой серии будет получена, скорее всего, другая выборка $(x'_1, x'_2, \dots, x'_n)$.

Эмпирические распределения и числовые характеристики

Выборка из генеральной совокупности является основным источником информации о случайной величине. По выборке оценивается класс распределений, к которому принадлежит распределение исследуемой случайной величины, устанавливаются интервалы, в которых лежат истинные значения параметров распределения, проверяются гипотезы об этой случайной величине и формулируются выводы о других ее свойствах.

Чтобы использовать аппарат математической статистики, нужно, прежде всего, уметь находить некоторые числовые характеристики выборок и строить эмпирические распределения, с помощью которых в дальнейшем можно делать соответствующие выводы.

Рассмотрим некоторые правила предварительной обработки выборочных данных. Представленная ниже таблица выборки объема (*Объемом выборки* называют количество наблюдений или количество значений случайной величины) $n = 250$ будет использоваться далее во всех вычислениях, а также станет источником построения выборок для индивидуальных вариантов заданий.

145.61	143.206	145.267	140.485	133.143	150.435	148.794	155.564	171.918
158.087	159.851	158.622	159.156	156.73	139.557	150.691	142.444	156.967
148.181	143.556	142.769	144.834	155.58	147.552	150.895	162.618	142.945
150.019	161.076	158.926	120.991	128.429	152.06	143.842	138.023	150.99
157.708	153.059	150.113	142.355	145.909	143.262	148.678	160.181	151.805
155.133	157.398	149.837	152.788	151.622	154.285	145.248	143.045	180.482
147.135	137.201	157.594	146.073	137.964	139.631	149.807	150.32	152.649
154.915	152.383	143.155	133.852	164.113	159.715	138.44	151.437	166.972
146.797	129.688	135.888	136.747	144.829	150.621	144.042	146.693	155.391
152.186	154.05	138.441	138.949	138.966	145.927	136.867	121.596	162.762

157.911	151.429	139.937	140.73	141.22	152.777	145.978	163.02	136.219
153.803	154.377	167.608	143.527	155.51	165.465	131.784	163.079	139.511
154.591	139.478	137.579	154.241	130.834	148.761	154.132	164.656	137.711
146.154	154.763	151.862	151.96	155.206	158.229	159.314	158.972	152.601
143.066	154.656	148.493	141.368	171.144	137.64	133.062	153.865	135.711
145.891	158.742	144.311	140.903	141.323	160.971	139.771	137.484	156.247
142.623	155.409	156.641	155.196	151.459	149.488	153.16	152.488	148.294
145.475	152.937	151.507	140.659	157.925	157.163	160.438	158.11	156.17
147.549	149.142	156.848	157.911	153.578	147.887	148.445	151.36	158.639
169.584	150.688	155.646	155.572	168.911	164.788	127.059	156.623	145.593
145.263	150.889	143.012	153.472	141.25	169.001	122.741	158.702	171.791
160.849	161.757	140.268	134.241	154.64	164.744	161.654	142.365	155.094
154.96	141.977	143.729	144.466	146.54	145.355	152.509	146.266	147.269
162.895	151.941	170.865	134.377	150.79	154.205	166.274	156.198	132.828
136.274	173.96	157.332	149.975	141.54	139.826	133.692	139.462	161.159
159.455	157.597	139.385	145.867	166.069	150.237	146.685	145.436	153.969
154.961	149.211	150.83	154.224	142.28	148.655	135.371	152.018	166.807
140.923	157.864	148.745	138.823	157.239	152.912	141.182		

Первичная обработка данных состоит обычно в отыскании максимального $x_{\max}$ и минимального $x_{\min}$ значений выборки (в Mathcad они вычисляются соответственно функциями $\max(\xi)$ и $\min(\xi)$, а также размаха варьирования $R = x_{\max} - x_{\min}$. Для приведенной выше выборки эти величины равны: $x_{\max}=180.482$, $x_{\min} = 120.991$, $R=59.49$.

Следующий этап первичной обработки – группировка и ее графическое представление. Группировка выборки объема n состоит в следующем. Промежуток $[x_{\min}, x_{\max}]$ разбивают на m интервалов группировки (чаще всего одинаковой длины) и подсчитывают число n_j выборочных значений, которые попали в j -й интервал. Обычно выбирают $m = 7 - 20$. Теперь каждый интервал группировки $\Delta_j = (a_j, b_j)$ представлен своими левой a_j и правой b_j границами и числом n_j элементов выборки, принадлежащих ему. Каждый интервал удобно представлять не двумя границами, а одним числом – срединным значением.

Наиболее наглядная форма графического представления группировки – гистограмма.

Если $\delta_1, \delta_2, \dots, \delta_m$ - длины интервалов группировки, а $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ – их середины и $h_j = \frac{n_j}{n}$ - относительные частоты попадания наблюдений в j -й интервал группировки, то можно построить график ступенчатой функции: $f(x) = \frac{h_j}{\delta_j}, x \in \Delta_j, j = 1, 2, \dots, m$.

Этот график называется гистограммой. В Mathcad для построения гистограмм предназначена функция $hist(\Delta, \xi)$.

Очевидно, что величина интервала группировки существенно влияет на вид гистограммы. При малой их ширине в каждый интервал попадает незначительное число наблюдений или даже не попадает ни одного, в

результате гистограмма становится сильно «изрезанной» и плохо передает основные особенности изучаемого распределения. Другая крайность – большие интервалы группировки; в этом случае скрадываются характерные черты распределения.

Иная форма графического представления группированных данных – полигон частот. *Полигон частот* – это ломаная линия, соединяющая точки с координатами $(\bar{x}_i, h_i)$, т.е. с абсциссами, равными серединам интервалов группировки, и ординатами, равными соответствующим частотам.

Можно также построить *полигон накопленных частот* – график ломаной, соединяющей точки с координатами $(b_j, \sum_{k=1}^j n_k)$ или $(b_j, \sum_{k=1}^j \frac{n_k}{n})$, т.е. с абсциссами, равными правым границам интервалов группировки, и ординатами, равными соответствующим накопленным частотам или относительным накопленным частотам.

Ниже приведен фрагмент рабочего документа Mathcad с вычислением $x_{\max}$, $x_{\min}$ $R = x_{\max} - x_{\min}$ для исследуемой выборки, а также с гистограммами и полигонами частот для различных интервалов группировки.

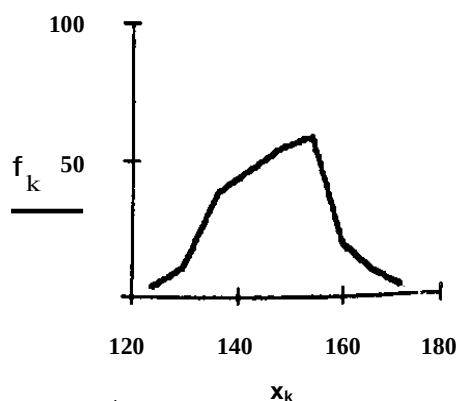
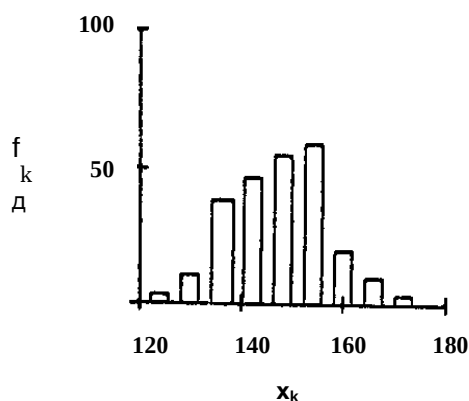
ORIGIN := 1

$x_{\max} := \max(\xi)$ $x_{\min} := \min(\xi)$ $R := x_{\max} - x_{\min}$

$x_{\max} = 180.482$ $x_{\min} = 120.991$ $R = 59.491$ $n := 250$

$m := 10$ $\Delta := \frac{R}{m}$ $j := 1..m$ $k := 1..m - 1$

$\delta_j := x_{\min} + \frac{\Delta}{2} \cdot (2 \cdot j - 1)$ $f := \text{hist}(x, \xi)$ $\Delta := 5.949$



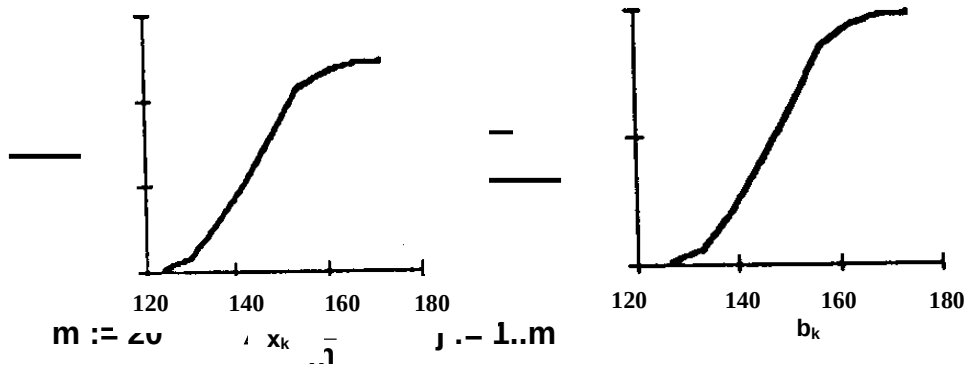
$a_j := x_{\min} + \Delta \cdot (j - 1)$ $b_j := a_j + \Delta$ $F_k := \sum_{j=1}^k f_j$

300

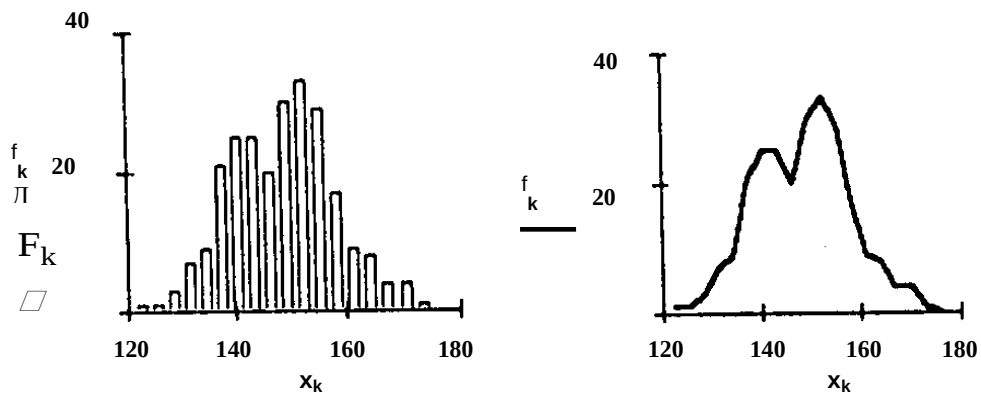
1

F_k
200
100

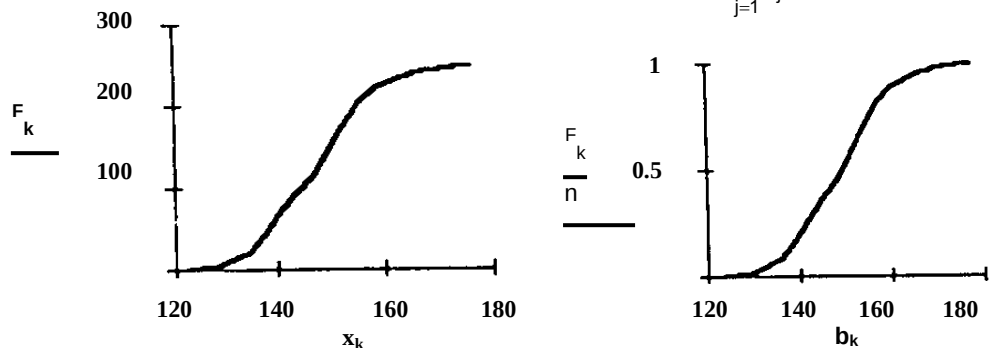
F_k
0.5
n



$$\delta_j := \text{xmin} + \frac{\Delta}{2} \cdot (2 \cdot j - 1) \quad f := \text{hist}(x, \xi) \quad \Delta = 2.975$$

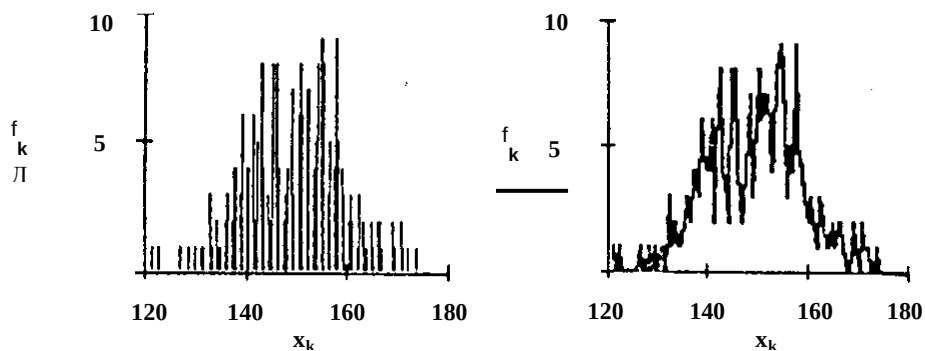


$$a_j := \text{xmin} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta \quad F_k := \sum_{j=1}^k f_j$$

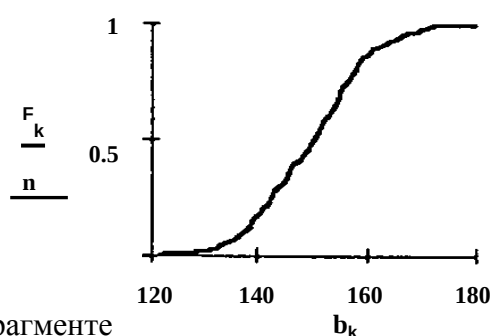
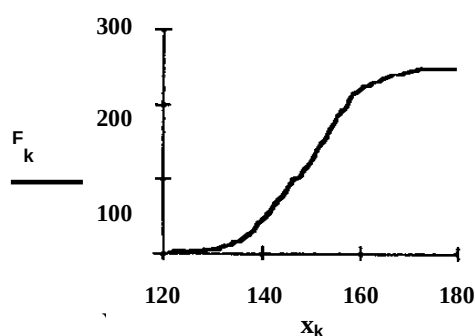


$$m := 100 \quad \Delta := \frac{R}{m} \quad j := 1..m \quad k := 1..m - 1$$

$$x_j := \text{xmin} + \frac{\Delta}{2} \cdot (2 \cdot j - 1) \quad f := \text{hist}(x, \xi) \quad \Delta = 0.595$$



$$a_j := \text{xmin} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta \quad F_k := \sum_{j=1}^k f_j$$



сохранены в массиве x_k . Приступить к группировке выборки, необходимо упорядочить выборочные значения в порядке их возрастания. Эту операцию выполняет функция $\text{sort}(\xi)$. Группировка производится с помощью функции $\text{hist}(x, \xi)$, где x – массив, содержащий значения середин интервалов группировки. Прежде чем обратиться к функции $\text{hist}(x, \xi)$, необходимо вычислить середины интервалов группировки и присвоить их значения элементам массива x . Значения функции $\text{hist}(x, \xi)$ – вектор, компоненты которого равны количеству элементов массива ξ , которые попадают в интервал группировки, середина которого равна соответствующей компоненте массива x . Используйте окна настройки параметров изображения гистограмм.

При первичной обработке выборочных данных можно рекомендовать несколько общих правил:

1. Перед началом группировки следует упорядочить выборочные значения в порядке возрастания. Такая упорядоченная в порядке возрастания выборка называется *вариационным рядом*.
2. При выборе числа интервалов группировки следует ориентироваться на 10-20 интервалов.
3. Предпочтительнее использовать интервалы одинаковой длины.
4. При анализе охватывайте всю область данных.
5. Избегайте полуоткрытых промежутков.
6. Интервалы группировки не должны перекрываться.

Задание 1

Вычислите максимальное, минимальное значение и размах для заданной части приведенной выше выборки. Выполните группировку для заданных значений m , постройте соответствующие гистограммы, полигоны частот и полигоны накопленных частот.

Порядок выполнения заданий

1. Определите и введите вектор-столбец выборочных значений.
2. Упорядочите выборку в порядке возрастания выборочных значений.
3. Вычислите минимальное значение и размах для полученной выборки.
4. Определите число интервалов группировки и их длину.
5. Определите вектор-столбец, содержащий середины интервалов группировки.
6. Определите с помощью функции $\text{hist}(x, \xi)$ вектор-столбец частот для полученных интервалов группировки.
7. Определите вектор-столбец накопленных частот.

8. Постройте гистограмму, полигон частот.
9. Постройте полигон накопленных частот и полигон относительных накопленных частот.
10. Выполните вычисления пп. 6-9 для всех заданных значений m .
11. Сохраните рабочий документ в файле на диске.

Пример выполнения задания

Примерный вариант выполнения задания для всей выборки для $m=10, 20, 100$ приведен выше.

Числовые характеристики выборки

Показатели положения. Среднее значение выборки вычисляется по формуле

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

В Mathcad для вычисления выборочного среднего значения выборки, сохраненной в матрице A , предназначена функция $\text{mean}(A)$.

Выборочной квантилью уровня p называется решение уравнения

$$F_n(x) = p,$$

где $F_n(x)$ – выборочная функция распределения.

В частотности, *выборочная медиана* есть решение уравнения $F_n(x) = 0.5$, т. е. выборочная медиана – это выборочная квантиль уровня 0.5. Выборочная медиана разбивает выборку пополам: слева и справа от нее оказывается одинаковое число элементов выборки. Если число элементов выборки четно, $n=2k$, то выборочную медиану определяют по формуле

$$\frac{x_k + x_{k+1}}{2}, \text{ где } x_k \text{ и } x_{k+1} - k\text{-е и } (k+1)\text{-е выборочные значения из}$$

вариационного ряда. При нечетном объеме выборки ($n=2k+1$) в качестве значения медианы принимают величину x_{k+1} .

В Mathcad для вычисления выборочной медианы выборки, сохраненной в матрице A , предназначена функция $\text{median}(A)$.

К показателям положения относятся минимальный и максимальный элементы выборки, а также верхняя и нижняя квартили (они ограничивают зону, в которой сосредоточены 50% элементов выборки).

Для вычисления минимального и максимального элементов выборки, размещенной в матрице A , в Mathcad предназначены соответственно функции $\text{min}(A)$ $\text{max}(A)$.

Показатели разброса. К показателям разброса относятся дисперсия выборки (выборочная дисперсия), стандартное отклонение, размах выборки, межквартильный размах, коэффициент эксцесса (выборочный эксцесс).

Выборочной дисперсией называется величина

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Однако в статистике чаще в качестве выборочной дисперсии

используется величина

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Причина такого, на первый взгляд неожиданного, способа вычисления дисперсии в том, что оценка дисперсии в этом случае является состоятельной и несмещенной.

В Mathcad для определения дисперсии выборки, сохраненной в матрице A, предназначена функция $\text{var}(A)$, а величину s^2 можно вычислить по формуле $s^2 = \frac{n}{n-1} \text{var}(A)$.

Стандартное отклонение рассчитывается по формуле $\hat{\sigma} = \sqrt{s^2}$.

Размах выборки вычисляется по формуле $R = x_{\max} - x_{\min}$.

Межквартильный размах равен $x_{0.75} - x_{0.25}$, где $x_{0.75}$ – 75%-ая квартиль, решение уравнения $F_n(x_{0.75}) = 0.75$, $x_{0.25}$ – 25%-ная квартиль, решение уравнения $F_n(x_{0.25}) = 0.25$.

Выборочный эксцесс определяется следующим образом. Сначала отыскивается величина выборочного центрального момента 4-ого порядка

$$\hat{\mu}_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4.$$

А затем по формуле $\hat{E} = \hat{\mu}_4 (s^2)^{-2} - 3$ вычисляется выборочный эксцесс.

Показатели асимметрии. На основании этих показателей изучают информацию о симметрии распределения выборочных данных около центра выборки. Сюда в первую очередь относится коэффициент асимметрии, который вычисляется по формуле

$$\hat{a} = \frac{\hat{\mu}_3}{\hat{\sigma}^3},$$

где $\hat{\mu}_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$ – выборочный центральный момент 3-его порядка, а $\hat{\sigma}$ – стандартное отклонение, формула для вычисления которого приведена выше.

Задание 2

Для выборки, сформированной в предыдущем задании, вычислите все описанные выше выборочные характеристики.

Порядок выполнения задания

1. Прочитайте сохраненный ранее файл, содержащий выборку.
2. Вычислите максимальный и минимальный элементы и размах выборки.
3. Рассчитайте выборочное среднее.
4. Найдите медиану.
5. Вычислите выборочную дисперсию и стандартное отклонение.
6. Найдите выборочные моменты 3-его и 4-его порядков.
7. Вычислите выборочный эксцесс.
8. Определите коэффициент асимметрии.

Пример выполнения задания

Ниже представлен фрагмент рабочего документа Mathcad, содержащий вычисление характеристик выборочных данных, приведенных в начале раздела.

$$n := 250$$

$$x_{\max} := \max(\xi) \quad x_{\min} := \min(\xi) \quad R := x_{\max} - x_{\min}$$

$$x_{\max} = 180.482 \quad x_{\min} = 120.991 \quad R = 59.49$$

$$\text{mean} := \text{mean}(\xi) \quad s2 := \frac{n}{n-1} \text{var}(\xi) \quad \sigma := \sqrt{s2}$$

$$\text{mean} = 149.849 \quad s2 = 98.174 \quad \sigma = 9.908$$

$$\mu3 := \frac{1}{n} \cdot \sum_{i=1}^n (\xi_i - \text{mean})^3 \quad \mu4 := \frac{1}{n} \cdot \sum_{i=1}^n (\xi_i - \text{mean})^4$$

$$\text{median} := \text{median}(\xi) \quad E := \frac{\mu4}{s2^2} - 3 \quad \alpha := \frac{\mu3}{\sigma^3}$$

$$\text{median} = 150.69 \quad E = 0.136 \quad \alpha = -0.055$$

Указание: В Mathcad нет встроенных функций для вычисления выборочных моментов. Для определения среднеквадратичного отклонения в Mathcad предназначена функция $\text{stdev}(A) = \sqrt{\text{var}(A)}$. Рассчитываемое с ее помощью значение среднеквадратичного отклонения отлично от определенного выше, поэтому среднеквадратичное отклонение следует вычислять как $\sqrt{s^2}$.

Оценка функций распределения

Как уже упоминалось ранее, распределение случайной величины является ее «паспортом», содержащим всю информацию о случайной величине.

Рассмотрим методы оценивания функции распределения $F_{\xi}(x)$ случайной величины, о которой известно, что она непрерывна.

Пусть $\hat{x} = \{x_1, x_2, \dots, x_n\}$ – совокупность выборочных значений случайной величины ξ , т.е. выборка из случайной величины ξ . Расположим наблюдения $x_1, x_2, \dots, x_n$ в порядке их возрастания. Обозначим новую упорядоченную последовательность – вариационный ряд* (напомним, что *вариационным рядом* называется упорядоченная совокупность выборочных значений.) – $x'_1, x'_2, \dots, x'_n$, где $x'_1 < x'_2 < \dots < x'_n$. По этому вариационному ряду построим следующую неубывающую ступенчатую функцию:

$$\hat{F}_n(x) = \begin{cases} 0, & x \leq x'_1, \\ \frac{k-1}{n}, & x'_{k-1} < x \leq x'_k, \quad k=1, 2, \dots, n, \\ 1, & x > x'_n \end{cases}$$

Из приведенной выше формулы видно, что функция $\hat{F}_n(x)$ претерпевает в каждой точке вариационного ряда скачок, равный по величине $1/n$. Если какая-нибудь точка вариационного ряда повторяется m раз (m точек вариационного ряда совпадают), то скачок функции $\hat{F}_n(x)$ в этой точке равен m/n .

Функция $\hat{F}_n(x)$ называется *эмпирической функцией распределения*.

Замечание: Эмпирическая функция распределения $\hat{F}_n(x)$ зависит не только от x , но и от всей выборки $\hat{x}$. Чтобы обратить внимание на этот факт, будем обозначать эмпирическую функцию распределения через $F_n(x, \hat{x})$. Именно $F_n(x, \hat{x})$ принимают за оценку теоретической функции распределения $F(x)$.

Остается выяснить, насколько хорошо эмпирическая функция распределения аппроксимирует теоретическую функцию распределения.

Если $F_\xi(x)$ – теоретическая функция распределения, а $F_n(x)$ – эмпирическая функция распределения, построенная по заданной выборке $\hat{x}$ значений случайной величины ξ , то в качестве меры расхождения теоретической и эмпирической функций распределения возьмем величину

$$D_n(\hat{x}) = \sup_x |F_n(x) - F_\xi(x)|.$$

Эта функция от выборочных значений x называется *статистикой Колмогорова*. Следует помнить, что $D_n(\hat{x})$ – случайная величина и что ее распределение не зависит от неизвестной теоретической функции распределения $F_\xi(x)$, если она непрерывна. Более того, справедлива теорема Колмогорова: если функция распределения $F_\xi(x)$ случайной величины ξ непрерывна, а $\hat{F}_n(x)$ – ее выборочная функция распределения, то при $n \rightarrow \infty$

$$P\left(\sup_x |\hat{F}_n(x) - F_\xi(x)| < \frac{z}{\sqrt{n}}\right) \rightarrow K(z) = \begin{cases} 0, & z \leq 0, \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2}, & z > 0. \end{cases}$$

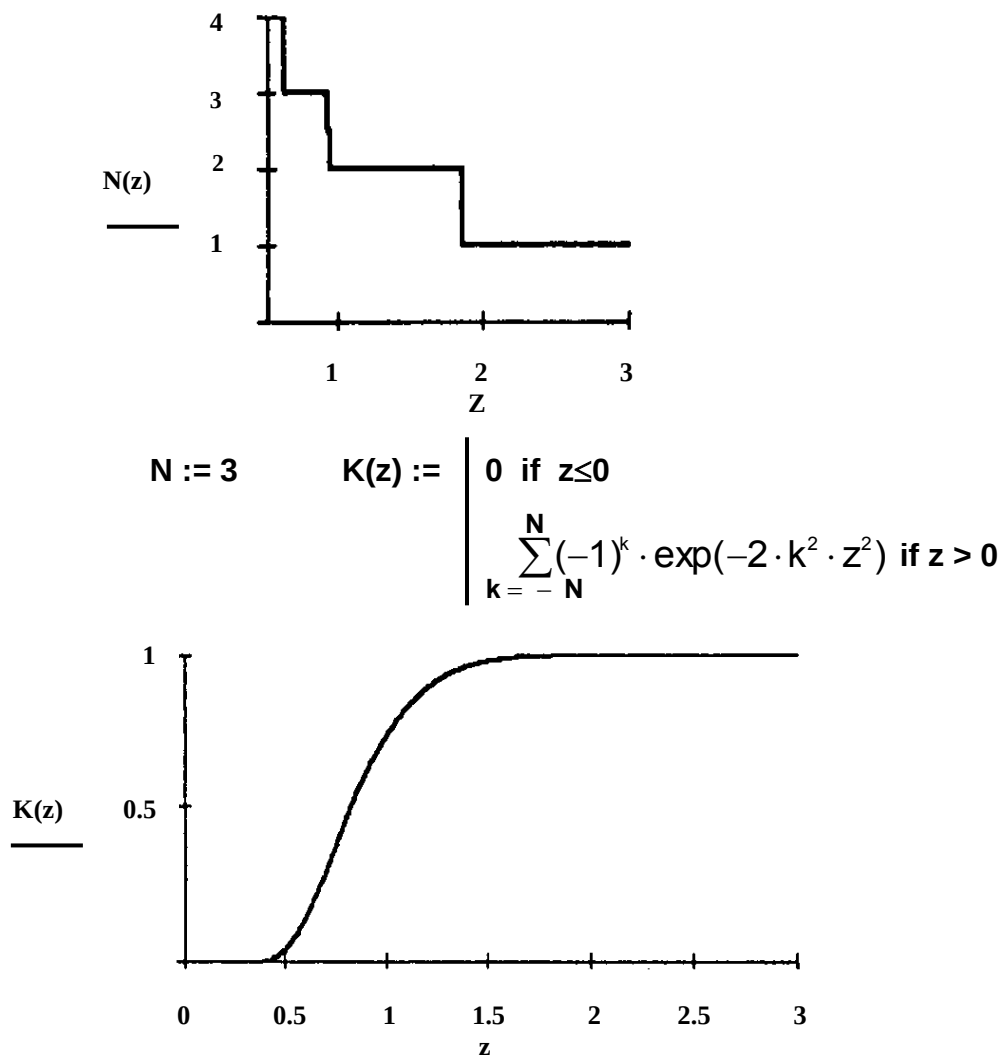
Функция $K(z)$ представляет собой функциональный ряд, который следует протабулировать. Сразу обратим внимание на то, что этот ряд сходится абсолютно для всех $z > 0$, но неравномерно на промежутке $[0; +\infty)$. Это означает, что для достижения заданной точности при вычислении $K(z)$ число N членов в соответствующей частичной сумме зависит от z . Если ε – требуемая точность вычисления $K(z)$, то число N вычисляется по формуле

$$N = \left\lceil \frac{1}{z} \sqrt{\frac{1}{2} \ln \frac{1}{\varepsilon}} \right\rceil + 1,$$

где символом $\lceil \rceil$ обозначена целая часть числа.

Ниже приведен фрагмент рабочего документа Mathcad, содержащий приближенное определение функции $K(z)$ для $\varepsilon = 0.001$, $N = 3$, и соответствующие графики.

$$\varepsilon := 0.001 \quad N(z) := \text{floor} \left(\frac{1}{z} \cdot \sqrt{\frac{1}{2} \cdot \ln \left(\frac{1}{\varepsilon} \right)} \right) + 1$$



Из приведенных в документе графиков видно, что для малых z величину $K(z)$ можно положить равной нулю, а для $z > 2$ можно считать $K(z)$ равной единице.

Зададимся вероятностью α такой, что событие, происходящее с вероятностью $1-\alpha$, представляется практически достоверным. Вычислим корень z_α уравнения $1-K(z) = \alpha$, тогда неравенство

$$\widehat{F}_n(x) - \frac{z_\alpha}{\sqrt{n}} < F_\xi(x) < \widehat{F}_n(x) + \frac{z_\alpha}{\sqrt{n}}$$

выполняется для всех действительных x с вероятностью, близкой к $1-\alpha$.

Таким образом, в окрестности эмпирической функции распределения построен «коридор», в котором лежит истинная, теоретическая функция распределения $F_\xi(x)$. С ростом n «ширина» этого коридора стремится к нулю.

Вместо эмпирической функции распределения будем использовать функцию накопленных относительных частот, поскольку $\widehat{F}_n(x) = F_k$ для $x \in (\xi_{k-1}, \xi_k]$ и значения функций совпадают вне промежутка $[x_{\min}, x_{\max}]$.

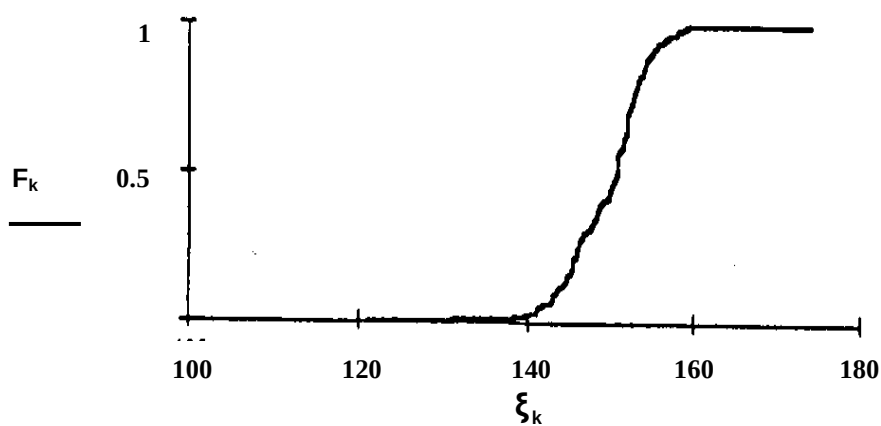
Ниже приведен фрагмент рабочего документа Mathcad с построением 95%-ного «коридора» для функции распределения случайной

величины по приведенной выборке.

$$M := 250 \quad \Delta := \frac{R}{m} \quad j := 1..m \quad k := 1..m - 1$$

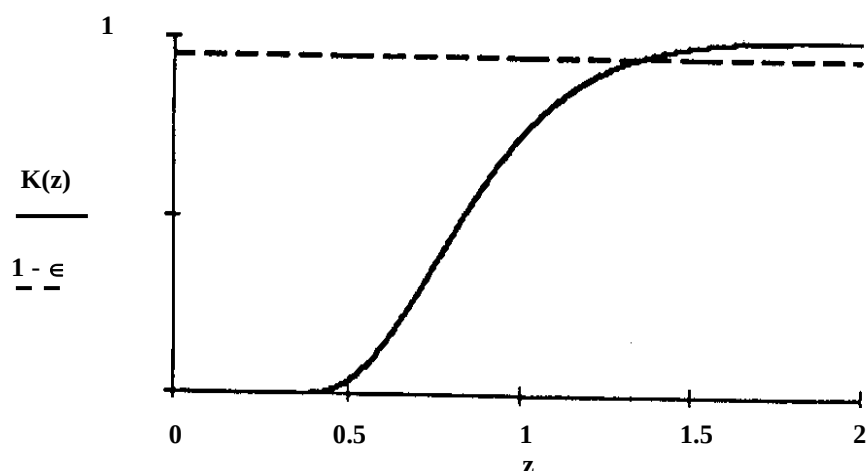
$$X_j := x_{\min} + \Delta \cdot (2 \cdot j - 1) \quad f := \text{hist}(x, \xi) \quad \Delta = 0.238$$

$$a_j := x_{\min} + \Delta \cdot (j - 1) \quad b_j := a_j + \Delta \quad F_k := \sum_{j=1}^k \frac{f_j}{n}$$



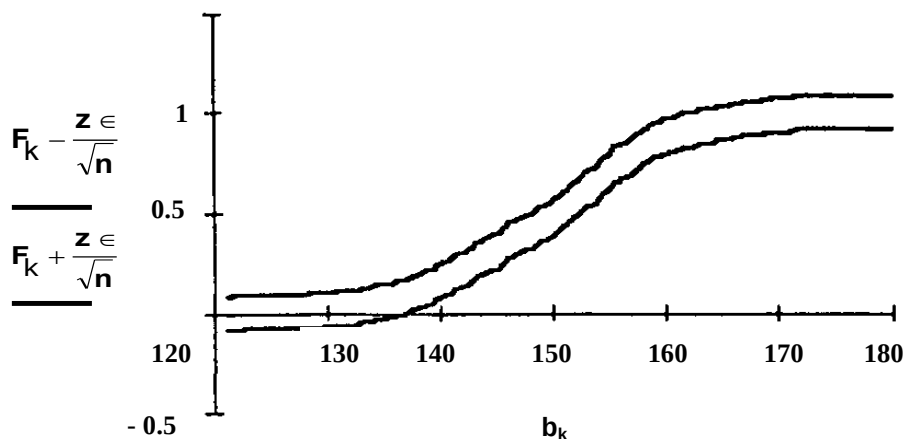
$$\alpha := 0.05 \quad K(z) := \begin{cases} 0 & \text{if } z \leq 0 \\ \sum_{k=3}^3 (-1)^k \cdot \exp(-2 \cdot k^2 \cdot z^2) & \text{if } z > 0 \end{cases}$$

$$p := 1 - \alpha$$



$$z\alpha := 1,358$$

$$1.5$$



Указание: Как уже отмечалось выше, в качестве эмпирической функции распределения использована эмпирическая функция накопленных частот. Заметим, что Mathcad вместо графика ступенчатой функции строит ломаную линию, соединяя «ступеньки» вертикальными отрезками прямых. Корень уравнения $1 - K(z) = \alpha$ проще всего найти графически, используя операцию Трассе пункта Graph меню Format как точку пересечения графика $K(z)$ и прямой $y = 1 - \alpha$. Ниже приведен фрагмент окна



Для оценки плотности распределения случайной величины можно воспользоваться полигоном частот, который представлен выше. При не очень обременительных ограничениях доказано, что выборочная плотность вероятностей, т. е. полигон частот, с ростом объема выборки до бесконечности стремится к истинной, теоретической, плотности распределения исследуемой случайной величины.

Задание 3

Постройте для выборки, сформированной в задании 1, 95%-ный «коридор» для функции распределения исследуемой случайной величины.

Порядок выполнения задания

1. Прочитайте файл, сохраненный при выполнении задания 1.
2. Определите статистику Колмогорова – функцию $K(z)$ и постройте ее график.
3. Определите значение величины α .
4. Решите графически уравнение $1 - K(z) = \alpha$.
5. Постройте «коридор» для теоретической функции распределения.

Пример выполнения заданий

Пример построения 95%-ного «коридора» функции распределения для исследуемой во всех примерах этого раздела выборки 250 значений случайной величины приведен выше.

При анализе статистических данных большую роль играет опыт и интуиция исследователя. В этой связи чрезвычайно полезными представляются следующие упражнения. Пользователь генерирует достаточно большую выборку значений случайной величины, имеющей известное непрерывное распределение с известными параметрами. А затем производит описанные выше вычисления, изменяя параметры задачи – объем выборки, количество интервалов группировки, доверительные вероятности и др., и сравнивает полученные оценки с известными теоретическими значениями. Здесь, прежде всего, полезно изучить равномерное и нормальное распределения. Приведенное ниже задание 4 заключается в решении именно такой задачи – исследование выборки значений случайной величины с заданным распределением.

Напомним, что исследованная во всех примерах раздела выборка представляет собой сгенерированную функцией Mathcad *rnorm* выборку 250 значений случайной величины, имеющей нормальное распределение $N(150, 10)$. Следовательно, внимательный читатель может не затрудняться ручным вводом выборки для индивидуального варианта задания, а просто аккуратно сгенерировать ее.

Задание 4

Сгенерируйте выборку значений случайной величины с заданным непрерывным распределением и выполните ее полный предварительный анализ для указанных значений объема выборки, числа интервалов группировки и доверительной вероятности. Постройте графики плотности вероятностей и функции распределения и сравните их с полученными графиками соответствующих выборочных функций.

Порядок выполнения задания

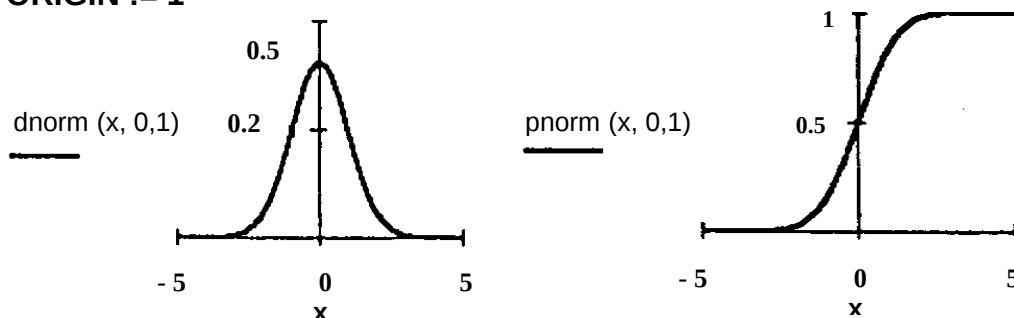
1. Установите в меню Math режим **Optimization**.
2. Присвойте переменной n значение, равное 100.
3. Постройте для заданного распределения графики плотности вероятностей и функции распределения.
4. Найдите математическое ожидание, дисперсию, среднеквадратичное отклонение, медиану, моменты 3- и 4-го порядка, асимметрию и эксцесс заданного распределения.
5. Сгенерируйте выборку объема n значений случайной величины, имеющей заданное распределение.
6. Определите как функции переменной n и найдите выборочные значения среднего, среднеквадратичного отклонения, моментов 3- и 4-го порядка, асимметрии и эксцесса.

7. Постройте гистограмму, полигон частот, график накопительных относительных частот.
8. Постройте 95%-ный «коридор» для теоретической функции распределения и изобразите на этом же графике функцию заданного в условии распределения вероятностей.
9. Сравните вычисленные теоретические и выборочные значения параметров.
10. Выполните вычисления пп. 4-7 для $n = 150, 200, 300, 500$.

Пример выполнения задания

Ниже приведен пример выполнения задания для стандартного нормального распределения $N(0, 1)$.

ORIGIN := 1



$\mu_{\xi} := 0$ $D_{\xi} := 1$ $\sigma_{\xi} := 1$ $\mu_{2\xi} := 1$ $\mu_{3\xi} := 0$

$$\mu_{4\xi} := 2 \cdot \int_0^{\infty} x^4 \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{x^2}{2}\right) dx \quad \mu_{4\xi} = 3$$

$$\beta_{\xi} := \frac{\mu_{3\xi}}{\mu_{2\xi}^2} \quad \beta_{\xi} = 0 \quad y_{\xi} := \frac{\mu_{4\xi}}{\mu_{2\xi}^2} - 3 \quad y_{\xi} = 0$$

median := qnorm(0.5, 0, 1) median = 0

Варианты заданий

Задание 1. Вычислите максимальные, минимальные значения и размах для заданной выборки. Выполните группировку для значений $m = 10, 20$, построьте соответствующие гистограммы, полигоны частот и полигоны накопленных частот. Выполните вычисления для 100 чисел из приведенной в разд. 2 выборки, начиная с числа n , номер которого указан в таблице.

N – номер варианта

N	n	N	n	N	n	N	n	N	n
1	10	5	50	9	90	13	95	17	135
2	20	6	60	10	270	14	105	18	145
3	30	7	70	11	75	15	115	19	155

4	40	8	80	12	85	16	125	20	165
---	----	---	----	----	----	----	-----	----	-----

Задание 2. Для выборки, сформированной в задании 1, вычислите все описанные в разд. 2 выборочные характеристики.

Задание 3. Постройте для выборки, сформированной в задании 1, 95%-ный коридор для функции распределения исследуемой случайной величины.

Задание 4. Сгенерируйте выборку объема n значений случайной величины с заданным непрерывным распределением и выполните полный предварительный ее анализ для числа интервалов группировки, равного целой части размаха и доверительной вероятности α . Постройте графики плотности вероятностей и функции распределения и сравните их с полученными графиками соответствующих выборочных функций.

N – номер варианта

N	Распределение	Параметры	n	α
1	Биноминальное	$p = 0.1$	50	0.95
2	Геометрическое	$p = 0.2$	50	0.90
3	Распределение Пуассона	$\lambda = 3$	50	0.95
4	Равномерное	$a = 0, b = 3$	50	0.90
5	Нормальное	$a = 1, \sigma = 3$	50	0.95
6	Экспоненциальное	$\lambda = 3$	50	0.90
7	χ^2 -распределение	$n = 5$	50	0.95
8	Распределение Стьюдента	$n = 7$	50	0.90
9	Распределение Фишера	$n = 5, m = 7$	50	0.95
10	Логистическое	$a = 0.3, \beta = 2$	50	0.90
11	Биноминальное	$p = 0.3$	60	0.95
12	Геометрическое	$p = 0.4$	70	0.90
13	Распределение Пуассона	$\lambda = 2$	80	0.95
14	Равномерное	$a = 1, b = 5$	90	0.90
15	Нормальное	$a = -1, \sigma = 2$	100	0.95
16	Экспоненциальное	$\lambda = 5$	60	0.90
17	χ^2 -распределение	$n = 3$	70	0.95
18	Распределение Стьюдента	$n = 5$	80	0.90
19	Распределение Фишера	$n = 3, m = 5$	90	0.95
20	Логистическое	$a = 2, \beta = 3$	100	0.90

3. ЛИНЕЙНАЯ РЕГРЕССИЯ. КРАТКИЕ СВЕДЕНИЯ

Пусть требуется исследовать зависимость $y(x)$, причем величины y и x измеряются в одних и тех же экспериментах. Без ограничения общности можно считать, что величина x измеряется точно, в то время как измерение

величины y содержит случайные погрешности. Это означает, что погрешность измерения величины x пренебрежимо мала по сравнению с погрешностью измерения величины y . Таким образом, результаты эксперимента можно рассматривать как выборочные значения случайной величины $\eta(x)$, зависящей от x , как от параметра.

Регрессией называют зависимость $y(x)$ условного математического ожидания величины $\eta(x)$ от переменной x , т. е. $y(x) \equiv M(\eta/x)$.

Задача регрессионного анализа состоит в восстановлении функциональной зависимости $y(x)$ по результатам измерений (x_i, y_i) , $i = 1, 2, \dots, n$.

Аппроксимируем неизвестную зависимость $y(x)$ заданной функцией $f(x, a_0, a_1, \dots, a_k)$. Это означает, что результаты измерений можно представить в виде

$$y_i = f(x, a_0, a_1, \dots, a_k) + \xi_i,$$

где $a_0, a_1, \dots, a_k$ – неизвестные параметры регрессии, а ξ_i – случайные величины, характеризующие погрешности эксперимента.

Обычно предполагается, что ξ_i – независимые нормально распределенные случайные величины с математическим ожиданием $M\xi_i = 0$ и одинаковыми дисперсиями $D\xi_i = \sigma^2$.

Параметры $a_0, a_1, \dots, a_k$ следует выбирать таким образом, чтобы отклонение значений предложенной функции от результатов эксперимента было минимальным. Часто в качестве меры отклонения выбирают величину

$$\Phi(a_0, a_1, \dots, a_k) = \sum_{i=1}^n (f(x_i, a_1, \dots, a_k) - y_i)^2,$$

и, следовательно, параметры $a_0, a_1, \dots, a_k$ определяют методом наименьших квадратов.

Рассмотрим простейший случай линейной регрессии. Пусть выдвинута гипотеза о том, что функция $f(x, a_0, a_1, \dots, a_k)$ имеет вид $f(x, a_0, a_1) = a_0 + a_1 x$. Найдем оценку параметров a_0 и a_1 методом наименьших квадратов. Для этого минимизируем функцию

$$\Phi(a_0, a_1) = \sum_{i=1}^n (a_0 + a_1 x_i - y_i)^2,$$

приравнивая нулю частные производные $\frac{\partial \Phi}{\partial a_0}$ и $\frac{\partial \Phi}{\partial a_1}$, откуда

$$a_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad a_1 = \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}.$$

В Mathcad для вычисления параметров a_0 и a_1 предназначены соответственно функции `intersept(x, y)` и `slope(x, y)`.

Ниже приведен фрагмент рабочего документа Mathcad, содержащий вычисление коэффициентов линейной регрессии a_0 и a_1 и

соответствующие графики для представленных ниже экспериментальных данных.

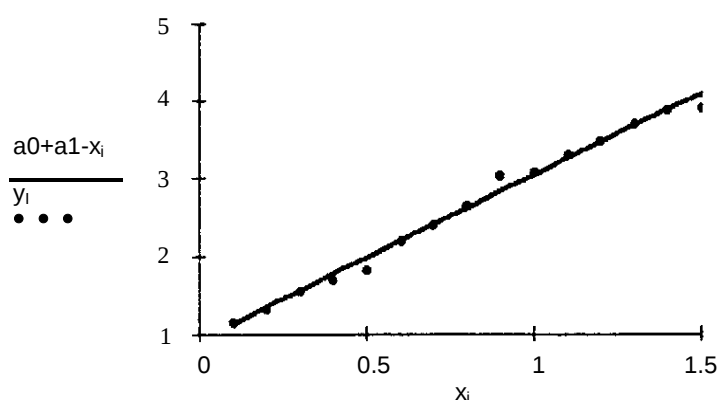
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
y	1.156	1.332	1.553	1.705	1.831	2.204	2.338	2.656
x	0.9	1.0	1.1	1.2	1.3	1.4	1.5	
y	3.019	3.081	3.299	3.486	3.692	3.867	3.896	

Значения x и y записаны на диске **c:** в папке **tmp** в файле **data5.txt**.

```

ORIGIN := 1      N := 15
i := 1..N      xi := i·0.1      yi := READ("c:\tmp\data5.txt")
a0 := intercept(x,y)      a0 = 0.925
a1 := slope(x,y)          a1 = 2.107

```



Эти же оценки дает метод максимального правдоподобия. В практических задачах дисперсия $D\xi_i = \sigma^2$ обычно неизвестна, но с помощью метода максимального правдоподобия можно получить ее оценку $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2.$$

Следует помнить, что $\hat{\sigma}^2, \hat{a}_0, \hat{a}_1$ – случайные величины. При сделанных выше предположениях, $\hat{a}_0$ и $\hat{a}_1$ распределены нормально, $M\hat{a}_0 = a_0$, $M\hat{a}_1 = a_1$, т.е. эти оценки несмещенные, а дисперсии этих оценок вычисляются по формулам

$$D\hat{a}_0 = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \hat{\sigma}^2, \quad D\hat{a}_1 = \frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Поскольку приведенная выше оценка дисперсии $\hat{\sigma}^2$ смещена (ее математическое ожидание равно $M\hat{\sigma}^2 = \frac{n-2}{n}\sigma^2$), будем использовать для оценки дисперсии другую, несмещенную оценку:

$$s^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2 = \frac{1}{n-2} \left(\sum_{i=1}^n y_i^2 - \hat{a}_0 \sum_{i=1}^n y_i - \hat{a}_1 \sum_{i=1}^n x_i y_i \right).$$

Величина $\frac{(n-2)s^2}{\sigma^2}$ имеет χ^2 -распределение с $n - 2$ степенями свободы.

Используя информацию о свойствах случайных величин s^2 , $\hat{a}_0, \hat{a}_1$ можно построить доверительные интервалы для оцениваемых параметров a_0, a_1 и σ^2 . Начнем с оценки $\hat{a}_0$. Если дисперсия σ^2 известна, то случайная величина

$$\frac{\hat{a}_0 - a_0}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

имеет стандартное нормальное распределение. Если α - доверительная вероятность и x_α - решение уравнения $\Phi(x_\alpha) = 1 - 0.5\alpha$, где $\Phi(x)$ - функция Лапласа, доверительный интервал

$$\hat{a}_0 - x_\alpha \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < a_0 < \hat{a}_0 + x_\alpha \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

накрывает неизвестный параметр с вероятностью $1 - \alpha$.

Если же дисперсия неизвестна, то в качестве критерия можно взять величину

$$\frac{\hat{a}_0 - a_0}{s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}.$$

Здесь

$$s^2 = \frac{1}{n-2} \left(\sum_{i=1}^n y_i^2 - \hat{a}_0 \sum_{i=1}^n y_i - \hat{a}_1 \sum_{i=1}^n x_i y_i \right)$$

имеет распределение Стьюдента с $n - 2$ степенями свободы.

По заданному значению α найдем корень t_α уравнения $F_{n-2}(t_\alpha) = 1 - 0.5\alpha$, где $F_{n-2}(x)$ - распределение Стьюдента с $n - 2$ степенями свободы.

Теперь доверительный интервал имеет вид

$$\hat{a}_0 t_{\alpha} s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} < a_0 < \hat{a}_0 + t_{\alpha} s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Ниже приведен фрагмент рабочего документа Mathcad, содержащий вычисление доверительного интервала для a_0 по выборке, которая приведена выше.

ORIGIN := 1 N := 15

i := 1..N x_i := i·0.1 y_i := READ("c:\tmp\data5.txt")

a0 := intercept(x,y) a0 = 0.925 a1 := slope(x,y) a1 = 2.107

y_r_i := a0+a·x_i

Xmean := mean(x) Xmean = 0.8

Ymean := mean(y) Ymean = 2.611

$$s2 = \frac{1}{N-2} \cdot \sum_{k=1}^N (y_k - y_{r_k})^2$$

Построение доверительного интервала для a₀

α := 0.1 t := qt(1 - $\frac{1}{\alpha}$, N - 2) t = 1.771

$$a0left := a0 - t \cdot \sqrt{s2} \cdot \sqrt{\frac{1}{N} + \frac{Xmean^2}{\sum_{k=1}^N (x_k - Xmean)^2}} \quad a0left = 0.836$$

$$a0right := a0 + t \cdot \sqrt{s2} \cdot \sqrt{\frac{1}{N} + \frac{Xmean^2}{\sum_{k=1}^N (x_k - Xmean)^2}} \quad a0right = 1.014$$

Доверительный интервал для a0 (0.836, 1.014)

Аналогично строится доверительный интервал для параметра $\hat{a}_1$. Если дисперсия σ^2 известна, то случайная величина

$$\frac{\hat{a}_1 - a_1}{\frac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

имеет стандартное нормальное распределение и, значит, с вероятностью

$1 - \alpha$ доверительный интервал

$$\left(\hat{a}_1 - \frac{x_{\alpha} \sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{a}_1 + \frac{x_{\alpha} \sigma}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$

накрывает оцениваемый параметр a_1 . Здесь, как и выше, x_{α} – решение уравнения $\Phi(x_{\alpha}) = 1 - 0.5\alpha$.

Если же дисперсия неизвестна, то в качестве критерия можно взять величину

$$\frac{\hat{a}_1 - a_1}{\frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}},$$

которая имеет распределение Стьюдента с $n - 2$ степенями свободы, и поэтому интервал

$$\left(\hat{a}_1 - \frac{t_\alpha s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{a}_1 + \frac{t_\alpha s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$

накрывает оцениваемый параметр a_1 с доверительной вероятностью $1 - \alpha$. Здесь t_α – корень уравнения $F_{n-2}(t_\alpha) = 1 - 0.5\alpha$, где $F_{n-2}(x)$ – функция распределения Стьюдента с $n - 2$ степенями свободы.

Ниже приведен фрагмент рабочего документа Mathcad, содержащий вычисление доверительного интервала для a_1 по той же выборке.

```

ORIGIN := 1           N := 15
i := 1..N    xi := i·0.1  yi := READ("c:\tmp\data5.txt")
a0 := intercept(x,y)  a0 = 0.925    a1 := slope(x,y)    a1 = 2.107
yri := a0+a1·xi
Xmean := mean(x)      Xmean = 0.8
Ymean := mean(y)      Ymean = 2.611

```

$$s2 = \frac{1}{N-2} \cdot \sum_{k=1}^n (y_k - yr_k)^2$$

$$\alpha := 0.1 \quad t := qt\left(1 - \frac{1}{\alpha}, N-2\right) \quad t = 1.771$$

Построение доверительного интервала для a_1

$$a1left := a1 - \frac{t \cdot \sqrt{s2}}{\sqrt{\sum_{k=1}^N (x_k - Xmean)^2}} \quad a1left = 2.009$$

$$a1right := a1 + \frac{t \cdot \sqrt{s2}}{\sqrt{\sum_{k=1}^N (x_k - Xmean)^2}} \quad a1right = 2.205$$

Доверительный интервал для a_1 (2.009, 2.205)

И, наконец, построим доверительный интервал для дисперсии. Как уже отмечалось выше, случайная величина $\frac{(n-2)s^2}{\sigma^2}$ имеет χ^2 -распределение с $n - 2$ степенями свободы. Задавшись малой вероятностью α , решим два уравнения: $F_{n-2}(\chi_l, \alpha) = 0.5\alpha$ и $F_{n-2}(\chi_r, \alpha) = 1 - 0.5\alpha$, где $F_{n-2}(x)$ – функция χ^2 -распределения с $n - 2$ степенями свободы. Случайная

величина $\frac{(n-2)s^2}{\sigma^2}$ попадает в интервал $(\chi_{l,\alpha}, \chi_{r,\alpha})$ с вероятностью $1 - \alpha$.
 Отсюда получаем доверительный интервал для дисперсии:

$$\left(\frac{(n-2)s^2}{\chi_{r,\alpha}}, \frac{(n-2)s^2}{\chi_{l,\alpha}} \right).$$

Ниже приведен фрагмент рабочего документа Mathcad, содержащий вычисление доверительного интервала для дисперсии по той же выборке.

```

ORIGIN := 1      N := 15      α := 0.1
i := 1..N      xi := 0.1·i      yi := READ("c:\tmp\data5.txt")
a0 := intercept(x,y)      a0 = 0.925      a1 := slope(x,y)      a1 = 2.107
yri := a0+a1·xi
Xmean := mean(x)      Ymean := mean(y)
Xmean = 0.8      Ymean = 2.611

s2 :=  $\frac{1}{N-2} \cdot \sum_{k=1}^N (y_k - yr_k)^2$       s2 = 8.556·10-3

Построение доверительного интервала для дисперсии
χleft := qchisq( $\frac{\alpha}{2}, N-2$ )      χleft = 5.892
χright := qchisq( $1-\frac{\alpha}{2}, N-2$ )      χright = 22.362

σ left :=  $\frac{(N-2) \cdot s2}{\chi_{right}}$       σ left = 4.974
σ right :=  $\frac{(N-2) \cdot s2}{\chi_{left}}$       σ right = 0.019

Доверительный интервал для дисперсии (0.005, 0.019)
    
```

Теперь задача о построении доверительных интервалов для параметров линейной регрессии решена полностью.

Рассмотрим еще некоторые свойства линейной регрессии.

Пусть линейная регрессия построена: $f(x, a_0, a_1) = \hat{a}_0 + \hat{a}_1 x$. Возьмем в области изменения аргумента некоторую точку x_0 и вычислим $\hat{y}_0 = \hat{a}_0 + \hat{a}_1 \cdot x_0$. Величина $\hat{y}_0$ случайная и меняется от выборки к выборке. Ее математическое ожидание равно истинному значению функции $f(x)$ в точке x_0 , т.е. $y_0 = a_0 + a_1 x_0$. Найдем доверительный интервал для величины y_0 . Для этого рассмотрим статистику

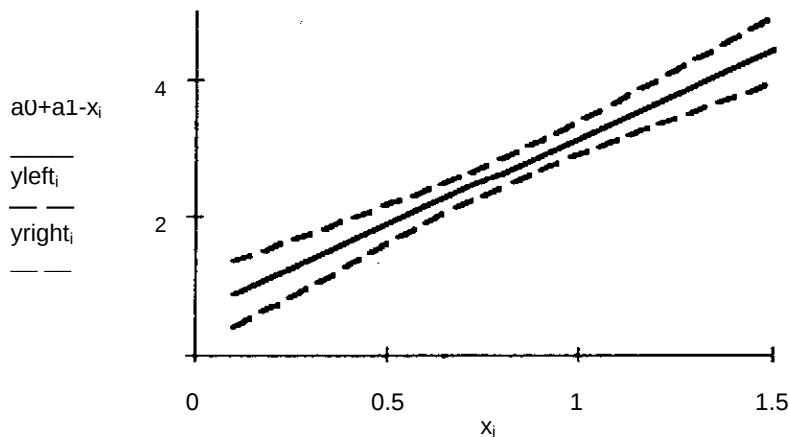
$$\frac{(\hat{a}_0 + \hat{a}_1 x_0) - (a_0 + a_1 x_0)}{s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}.$$

Доказано, что она имеет распределения Стьюдента с $n - 2$ степенями свободы, и поэтому доверительный интервал

$$\left(\hat{a}_0 + \hat{a}_1 x_0 - t_\alpha s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^{-2}}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{a}_0 + \hat{a}_1 x_0 + t_\alpha s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^{-2}}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right)$$

накрывает истинное значение $y_0 = a_0 + a_1 x_0$ с вероятностью $1 - \alpha$. Величина t_α определена выше.

Границы доверительных интервалов в каждой точке x_0 образуют доверительную полосу, или доверительный коридор (см. ниже). Эта полоса, однако, не является доверительной областью для всей линии регрессии. Она определяет только концы доверительных интервалов для y при каждом значении x . С помощью коридора регрессии нельзя, например, построить одновременно два доверительных интервала в различных точках



x_0 и x_1 .

Ниже приведен фрагмент рабочего документа Mathcad, содержащий построение коридора регрессии.

Построение доверительного коридора $i := 1..N$

$$\begin{aligned} \text{yleft}_i &:= a_0 + a_1 \cdot x_i - t \cdot \sqrt{s^2} \cdot \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^N (x_k - \bar{x})^2}} \\ \text{yright}_i &:= (a_0 + a_1 \cdot x_i) + t \cdot \sqrt{s^2} \cdot \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^N (x_k - \bar{x})^2}} \end{aligned}$$

Доверительная область для всей *линии регрессии* определяется с помощью следующих уравнений соответственно нижней и верхней границ полосы:

$$y = \hat{a}_0 + \hat{a}_1 x - 2 f_\alpha s \sqrt{\frac{1}{n} + \frac{\bar{x}^{-2}}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

$$y = \hat{a}_0 + \hat{a}_1 x + 2 f_\alpha s \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^{-2}}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

где f_α - корень уравнения $F_{2,n-2}(f_\alpha) = 1 - \alpha$; $F_{2,n-2}(x)$ - функция распределения Фишера с 2 и $n - 2$ степенями свободы.

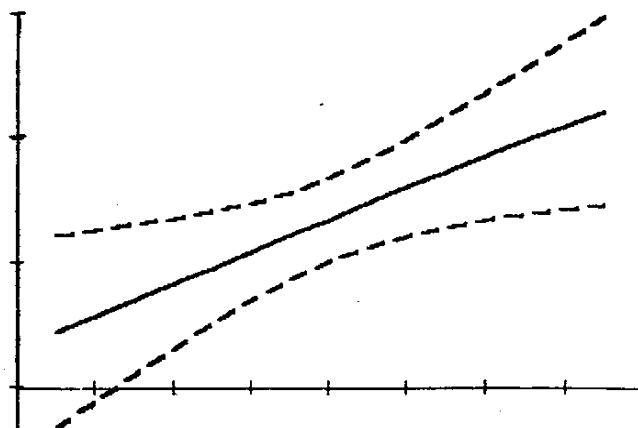
Ниже приведен фрагмент рабочего документа Mathcad, содержащий вычисление доверительной области регрессии для выборки, анализируемой во всех предыдущих примерах.

Построение доверительной области

$$f := qF(1 - \alpha, 2, N - 2)$$

$$i := 1..N$$

$$yleft_i := a_0 + a_1 \cdot x_i - 2 \cdot f \cdot \sqrt{s^2 \cdot \left(\frac{1}{N} + \frac{(x_i - X_{mean})^2}{\sum_{k=1}^N (x_k - X_{mean})^2} \right)}$$



$$yright_i := (a_0 + a_1 \cdot x_i) + 2 \cdot f \cdot \sqrt{s^2 \cdot \left(\frac{1}{N} + \frac{(x_i - X_{mean})^2}{\sum_{k=1}^N (x_k - X_{mean})^2} \right)}$$

6

$a_0 + a_1 \cdot x_i$

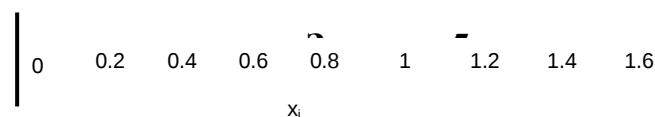
4

$yleft_i$

$yright_i$

2

--



Для заданной в условии выборки вычислите регрессию и найдите доверительные интервалы коэффициентов регрессии и дисперсии для заданной доверительной вероятности. Вычислите коридор и доверительную область регрессии. Изобразите выборку графически на одном графике с линией регрессии. Изобразите графически коридор и доверительную область регрессии.

Порядок выполнения задания

1. Определите и введите заданную выборку.
2. Найдите точечные оценки математического ожидания обеих переменных.
3. Вычислите точечную несмещенную оценку неизвестной дисперсии.
4. Найдите коэффициенты регрессии.
5. Постройте график линии регрессии и изобразите на нем экспериментальные точки.
6. Вычислите значение критерия для оценки коэффициента регрессии a_0 .
7. Найдите доверительный интервал для a_0 .
8. Вычислите значение критерия для оценки коэффициента регрессии a_1 .
9. Найдите доверительный интервал для a_1 .
10. Вычислите значение критерия для оценки дисперсии.
11. Найдите доверительный интервал для дисперсии.
12. Вычислите коридор регрессии.
13. Изобразите на графике линию регрессии и границы коридора для нее.
14. Вычислите доверительную область для всей регрессии.
15. Изобразите на графике линию регрессии и доверительную область для нее.

Пример выполнения задания

Все приведенные выше фрагменты рабочих документов Mathcad содержат последовательные этапы выполнения задания для выборки, описанной в тексте.

Варианты заданий

1

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	- 1.45	- 1.829	- 1.247	- 1.051	- 1.241	- 0.988	- 0.766	- 0.504
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	- 0.399	0.075	0.088	0.318	0.987	0.858	1.626	

2

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
-----	-------	-------	-------	-------	-------	-------	-------	-------

y	- 2.169	- 1.376	- 0.974	- 0.312	- 0.314	- 0.715	- 0.312	1.119
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	0.92	0.999	1.046	1.295	1.411	1.884	2.835	

3

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	0.484	0.628	0.282	0.676	1.482	1.207	1.301	1.463
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	1.919	2.149	2.176	2.425	2.727	2.568	2.96	

4

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	-0.139	0.661	1.404	0.928	1.736	1.762	1.765	2.617
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	2.787	2.735	2.72	3.312	3.502	4.082	4.197	

5

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	2.016	2.073	2.442	2.708	2.956	2.907	3.315	3.493
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	3.457	3.971	4.12	3.939	4.681	4.924	4.221	

6

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	2.318	2.451	2.917	2.954	3.486	3.725	4.106	4.936
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	4.678	4.859	5.611	6.017	5.46	6.586	6.15	

7

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	3.911	3.893	4.704	4.993	4.935	5.477	5.384	5.489
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	5.202	5.714	6.524	6.348	6.516	7.136	7.069	

8

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	3.972	4.811	4.932	5.355	5.821	5.789	6.266	6.857
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	6.206	6.857	7.366	7.527	7.962	8.402	8.569	

9

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	2.258	0.738	1.479	1.094	1.177	1.126	0.523	0.741

x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	-0.364	0.673	0.259	-0.378	-0.568	-1.266	-1.376	

10

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	2.235	2.849	2.237	2.63	1.761	2.163	1.813	1.707
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	1.137	1.348	0.799	0.997	0.273	0.057	-0.321	

11

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	4.005	3.637	2.987	3.19	3.102	3.236	2.68	2.241
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	2.258	2.249	1.958	1.188	1.501	1.035	0.911	

12

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	4.592	4.775	4.487	4.204	3.147	3.943	3.543	3.622
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	2.608	2.776	2.767	2.509	2.732	1.804	2.016	

13

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	5.892	5.103	5.624	5.197	4.749	4.653	4.253	4.249
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	4.555	3.955	4.076	3.869	3.241	2.782	2.667	

14

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	7.689	7.513	7.314	6.951	6.632	6.515	5.653	5.61
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	5.076	4.768	4.503	4.224	3.35	3.869	3.405	

15

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	4.874	4.996	5.073	5.438	5.356	5.214	5.502	5.64
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	6.088	5.798	5.916	6.357	6.077	6.859	6.416	

16

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	10.217	9.632	8.604	9.345	8.472	8.207	7.739	7.278
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	7.208	6.73	6.933	6.434	6.15	5.736	6.092	

17

X	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	-2.83	-2.633	-2.227	-1.281	-2.609	-1.574	-1.986	-1.48
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	-0.829	-1.305	-1.065	-0.552	-0.941	-0.55	0.168	

18

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	-3.388	-3.367	-3.459	-3.025	-2.191	-1.763	-1.605	-1.658
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	-1.433	-1.135	-0.824	-0.663	0.591	0.038	0.321	

19

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	-3.831	-3.335	-2.974	-3.287	-2.876	-2.821	-2.409	-2.421
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	-2.257	-2.209	-1.634	-1.222	-1.333	-1.676	-1.006	

20

x	- 0.9	- 0.8	- 0.7	- 0.6	- 0.5	- 0.4	- 0.3	- 0.2
y	-5.315	-5.622	-5.509	-4.718	-4.679	-4.235	-3.742	-3.459
x	- 0.1	0	0.1	0.2	0.3	0.4	0.5	
y	-2.848	-3.381	-3.083	-2.167	-1.688	-1.325	-1.641	

СОДЕРЖАНИЕ

	<i>стр.</i>
1. Используемые инструменты Mathcad	3
☞ Ввод и вывод файлов данных	3
☞ Функции вычисления выборочных характеристик	4
☞ Построение эмпирических распределений	6
☞ Моделирование выборок из стандартных распределений	8
2. Основные задачи статистики. Выборки. Гистограммы.	
Полигоны частот.....	8
☞ Эмпирические распределения и числовые характеристики	9
☞ Задание 1	13
☞ Числовые характеристики выборки	14
☞ Задание 2	15
☞ Оценка функций распределения	16
☞ Задание 3	20
☞ Задание 4	21
☞ Варианты заданий	22
3. Линейная регрессия. Краткие сведения	23
☞ Задание 5	31
☞ Варианты заданий.....	32

СПИСОК ЛИТЕРАТУРЫ

1. Дьяконов В. Mathcad 8/2000: Специальный справочник, - СПб.: Питер, 2000. – 592 с.: ил.
2. Кудрявцев Е. М. Mathcad 8. – М.: ДМК, 2000. – 320 с.: ил.
3. Плис А. И., Сливина Н. А. Mathcad: Математический практикум для экономистов и инженеров: Учебное пособие. – М.: Финансы и статистика. 1999. – 656 с.: ил.