Лекция. БАЗОВЫЕ ПОНЯТИЯ И ОПЕРАЦИИ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Учебные вопросы

2.1. Эмпирическая функция распределения

2.2. Оценки параметров распределения и их свойства

2.3. Оценки моментов и квантилей распределения

2.1. Эмпирическая функция распределения

Случайность – это старейшая знать мира, которая избавлена … от рабства под игом цели.

Ф. Ницше. "Так говорил Заратустра".

Методы обработки ЭД опираются на базовые понятия теории вероятностей и математической статистики. К их числу относятся понятия генеральной совокупности, выборки, эмпирической функции распределения [3, 5].

Под *генеральной совокупностью* понимают все возможные значения параметра, которые могут быть зарегистрированы в ходе неограниченного по времени наблюдения за объектом. Такая совокупность состоит из бесконечного множества элементов. В результате наблюдения за объектом формируется ограниченная по объему совокупность значений параметра $x_1, x_2, \ldots, x_n$. С формальной точки зрения такие данные представляют собой ***выборку* из генеральной совокупности**.

Будем считать, что выборка содержит полные наработки до системных событий (цензурирование отсутствует). Наблюдаемые значения x_i **называют *вариантами***, а их количество – **объемом выборки *n*.** Для того чтобы по результатам наблюдения можно было делать какие-либо выводы, выборка должна быть *репрезентативной* (представительной), т. е. правильно представлять пропорции генеральной совокупности. Это требование выполняется, если объем выборки достаточно велик, а каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в выборку.

Пусть в полученной выборке значение x_1 параметра наблюдалось n_1 раз, значение x_2 – n_2 раз, значение x_k – n_k раз, $\boldsymbol{n_1+n_2+\ldots+n_k=n}$.

Совокупность значений, записанных в порядке их возрастания, называют *вариационным рядом,* величины $\boldsymbol{n_i}$ **– частотами**, а их отношения к объему выборки $\boldsymbol{w_i=n_i/n}$ – ***относительными частотами*** **(частостями).** Очевидно, что сумма относительных частот равна единице.

Под распределением понимают соответствие между наблюдаемыми вариантами и их частотами или частостями. Пусть n_x – количество наблюдений, при которых случайные значения параметра X меньше x. Частость события $X<x$ равна n_x/n . Это отношение является функцией от x и от объема выборки: $\boldsymbol{F_n(x)=n_x/n}$. Величина $F_n(x)$ обладает всеми свойствами функции:

распределения: $F_n(x)$ неубывающая функция, ее значения принадлежат отрезку [0 – 1];

если x_1 – наименьшее значение параметра, а x_k – наибольшее, то $F_n(x)=0$, когда $x<x_1$, и $F_n(x_k)=1$, когда $x>=x_k$.

Функция $\boldsymbol{F_n(x)}$ определяется по ЭД, поэтому ее называют *эмпирической функцией распределения*. В отличие от эмпирической функции $F_n(x)$ функцию распределения $\boldsymbol{F(x)}$ **генеральной совокупности называют теоретической функцией распределения,** она характеризует не частость, а вероятность события $X<x$. Из теоремы Бернулли вытекает, что частость $F_n(x)$ стремится по вероятности к вероятности $F(x)$ при

неограниченном увеличении n. Следовательно, **при большом объеме наблюдений теоретическую функцию распределения $F(x)$ можно заменить эмпирической функцией $F_n(x)$.**

График эмпирической функции $F_n(x)$ представляет собой ломаную линию. В промежутках между соседними членами вариационного ряда $F_n(x)$ сохраняет постоянное значение. При переходе через точки оси x, равные членам выборки, $F_n(x)$ претерпевает разрыв, скачком возрастая на величину $1/n$, а при совпадении m наблюдений – на m/n.

Пример 2.1. Построить вариационный ряд и график эмпирической функции распределения по результатам наблюдений, табл. 2.1.

Таблица 2.1

i	1	2	3	4	5	6
x_i	51	43	56	60	64	56

Решение. Построим вариационный ряд, упорядочив по возрастанию значения варианты, табл. 2.2.

Таблица 2.2

i	1	2	3	4	5	6
x_i	43	51	56	56	60	64

Искомая эмпирическая функция, рис. 2.1:

$$F_6(x) = \begin{cases} 0, & \text{при } x < 43, \\ 0{,}16, & \text{при } 43 \le x < 51, \\ 0{,}33, & \text{при } 51 \le x < 56, \\ 0{,}67, & \text{при } 56 \le x < 60, \\ 0{,}84, & \text{при } 60 \le x < 64, \\ 1, & \text{при } x \ge 64. \end{cases}$$

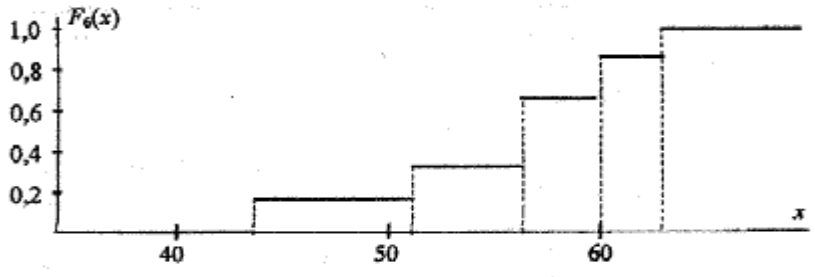

Рис. 2.1. Эмпирическая функция распределения

При большом объеме выборки (понятие «большой объем» зависит от целей и методов обработки, в данном случае будем считать n большим, если n>40) в целях удобства обработки и хранения сведений **прибегают к группированию ЭД в интервалы.** Количество интервалов следует выбрать так, чтобы в необходимой мере отразилось разнообразие значений параметра в совокупности и в то же время закономерность распределения не искажалась случайными колебаниями частот по отдельным разрядам.

Существуют нестрогие рекомендации по выбору **количества у** и **размера *h*** таких интервалов, в частности:

в каждом интервале должно находиться не менее 5 – 7 элементов. В крайних разрядах допустимо всего два элемента;

количество интервалов не должно быть очень большим или очень маленьким. Минимальное **значение у должно быть не менее 6 – 7.** При объеме выборки, не превышающем несколько сотен элементов, величину **у задают в пределах от 10 до 20.** Для очень большого объема выборки (n>1000) количество интервалов может превышать указанные значения. Некоторые исследователи рекомендуют пользоваться соотношением **$y=1{,}441*\ln(n)+1$;**

при относительно небольшой неравномерности длины интервалов удобно выбирать одинаковыми и равными величине

$$h=(x_{max}-x_{min})/y,$$

где x_{max} – максимальное и x_{min} – минимальное значение параметра. При существенной неравномерности закона распределения длины интервалов можно задавать меньшего размера в области быстрого изменения плотности распределения;

при значительной неравномерности лучше в каждый разряд назначать примерно одинаковое количество элементов выборки. Тогда длина конкретного интервала будет определять крайними значениями элементов выборки, сгруппированными в этот интервал, т.е. будет различна для разных интервалов (в этом случае при построении гистограммы нормировка по длине интервала обязательна - в противном случае высота каждого элемента гистограммы будет одинакова).

Группирование результатов наблюдений по интервалам предусматривает: определение размаха изменений параметра x; выбор количества интервалов и их величины; подсчет для каждого i-го интервала $[x_i–x_{i+1}]$ частоты n_i или относительной частоты (частости n_i) попадания варианты в интервал. **В результате формируется представление ЭД в виде *интервального или статистического ряда*.**

Графически статистический ряд отображают в виде гистограммы, полигона и ступенчатой линии. **Часто *гистограмму* представляют как фигуру, состоящую из прямоугольников, основаниями которых служат интервалы длиною *h*, а высоты равны соответствующей частости. Однако такой подход неточен. Высоту *i*-го прямоугольника z_i следует выбрать равной $n_i/(nh)$.** Такую гистограмму можно интерпретировать как графическое представление эмпирической функции плотности распределения $f_n(x)$, в ней суммарная площадь всех прямоугольников составит единицу. Гистограмма помогает подобрать вид теоретической функции распределения для аппроксимации ЭД.

Полигоном называют ломаную линию, отрезки которой соединяют точки с координатами по оси абсцисс, равными серединам интервалов, а по оси ординат – соответствующим частостям. Эмпирическая функция распределения отображается ступенчатой ломаной линией: над каждым интервалом проводится отрезок горизонтальной линии на высоте, пропорциональной накопленной частости в текущем интервале. Накопленная частость равна сумме всех частостей, начиная с первого и до данного интервала включительно.

Пример 2.2. Имеются результаты регистрации значений затухания сигнала x_i на частоте 1000 Гц коммутируемого канала телефонной сети. Эти значения, измеренные в дБ, в виде вариационного ряда представлены в табл. 2.3. Необходимо построить статистический ряд.

Таблица 2.3

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
x_i	25,79	25,98	25,98	26,12	26,13	26,49	26,52	26,60	26,66	26,69	26,74
i	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
x_i	26,85	26,90	26,91	26,96	27,02	27,11	27,19	27,21	27,28	27,30	27,38
i	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
x_i	27,40	27,49	27,64	27,66	27,71	27,78	27,89	27,89	28,01	28,10	28,11
i	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**
x_i	28,37	28,38	28,50	28,63	28,67	28,90	28,99	28,99	29,03	29,12	29,28

Решение . Количество разрядов статистического ряда следует выбрать минимальным, чтобы обеспечить достаточное количество попаданий в каждый из них, возьмем y = 6. Определим размер разряда

$$h = (x_{max} - x_{min})/y = (29,28 - 25,79)/6 = 0,58.$$

Сгруппируем наблюдения по разрядам, табл. 2.4.

Таблица 2.4

i	**1**	**2**	**3**	**4**	**5**	**6**
x_i	**25,79**	**26,37**	**26,95**	**27,5 3**	**28,12**	**28,70**
n_i	**5**	**9**	**10**	**9**	**5**	**6**
$w_i = n_i/n$	**0,114**	**0,205**	**0,227**	**0,205**	**0,11 4**	**0,136**
$F_n(x)$	**0,114**	**0,319**	**0,546**	**0,751**	**0,864**	**1**
$z_i = w_i/h$	**0,196**	**0,353**	**0,392**	**0,353**	**0,196**	**0,235**

На основе статистического ряда построим гистограмму, рис. 2.2, и график эмпирической функции распределения, рис. 2.3.

График эмпирической функции распределения, рис. 2.3, отличается от графика, представленного на рис. 2.1 равенством шага изменения варианты и величиной шага приращения функции (при построении по вариационному ряду шаг приращения кратен

$1/n$, а по статистическому ряду – зависит от частости в конкретном разряде).

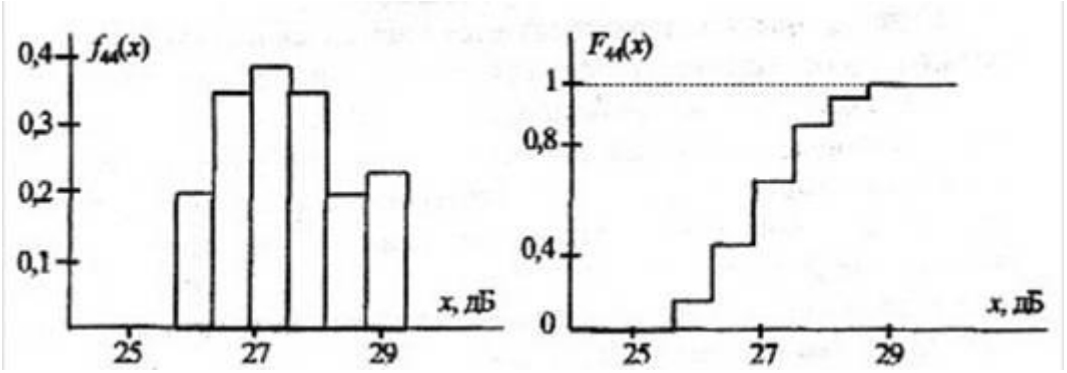

Рис.2..2. Гистограмма распределения

Рис. 2.3. Эмпирическая функция распределения

Рассмотренные представления ЭД являются исходными для последующей обработки и вычисления различных параметров.

2.2. Оценки параметров распределения и их свойства

Для понимающего достаточно и немногого.

Стендаль. "Пармская обитель".

Значение параметра, вычисленное по ограниченному объему ЭД, является случайной величиной, т. е. значение такой величины от выборки к выборке может меняться заранее не предвиденным образом. Следовательно, в результате обработки ЭД определяется не значение **параметра *T*,** а только лишь его приближенное значение – **статистическая *оценка* параметра q.** Получить статистическую оценку параметра теоретического распределения означает найти функцию от имеющихся результатов наблюдения, которая и даст приближенное значение искомого параметра. Различают два вида оценок – точечные и интервальные. ***Точечными*** **называют такие оценки, которые характеризуются одним числом.** При малых объемах выборки точечные оценки могут значительно отличаться от истинных значений параметров, поэтому их применяют при большом объеме выборки. ***Интервальные*** **оценки задаются двумя числами, определяющими вероятный диапазон возможного значения параметра**. Эти оценки применяются для малых и для больших выборок. Рассмотрим вначале точечные оценки.

Применительно к каждому оцениваемому параметру закона распределения генеральной совокупности существует множество функций, позволяющих вычислить искомые значения. Например, **оценку математического ожидания** можно вычислить, взяв среднее арифметическое выборочных значений, половину суммы крайних членов вариационного ряда, средний член выборки и т.д. Указанные функции отличаются качеством оценок и трудоемкостью реализации.

Качество оценок **характеризуется такими свойствами, как состоятельность, несмещенность, эффективность и достаточность** [3, 5, 9].

Состоятельность характеризует сходимость по вероятности оценки q к истинному значению параметра *T* при неограниченном увеличении объема выборки *n*. Для состоятельности оценки достаточно, но не обязательно, чтобы математическое ожидание квадрата отклонения оценки от параметра $M(T - q)^2$ стремилось к нулю с увеличением объема выборки (здесь и далее символ *M* означает математическое ожидание). Свойство состоятельности проявляется при неограниченном увеличении *n*, а при небольших объемах ЭД наличие этого свойства еще недостаточно для применения оценки.

Несмещенность характеризует отсутствие систематических (в среднем) отклонений оценки от параметра при любом конечном, в том числе и малом, объеме выборки, т. е. $M(\mathrm{q}) = T$. Использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, приводит к систематическим ошибкам. Не всегда наличие смещения плохо. Оно может быть существенно меньше погрешности регистрации значений параметра или давать дополнительную гарантию выполнения требований к значению параметра (если даже при положительном смещении оценка q меньше предельно допустимого значения, то несмещенное значение тем более будет отвечать этому условию). В таких ситуациях допустимо применение смещенных оценок, если они вычисляются проще, чем несмещенные. Но даже несмещенная оценка может быть удалена от истинного значения.

Эффективность характеризует разброс случайных значений оценки около истинного значения параметра. Среди всех оценок следует выбрать ту, значения которой теснее сконцентрированы около оцениваемого параметра. Для многих применяемых способов оценивания выборочные распределения параметров асимптотически нормальны, поэтому часто мерой эффективности служит дисперсия оценки. В таком понимании эффективная оценка – это оценка с минимальной дисперсией. При неограниченном увеличении n эффективная оценка является и состоятельной. В случае оценивания одного параметра дисперсия несмещенной оценки отвечает условию Рао – Крамера,

$$D(fAE)\left[-nM\left(\frac{\partial^2 \ln f(x,T)}{\partial T^2}\right)\right]^{-1}$$

где $f(x, T)$ – плотность распределения варианты; n – количество наблюдений.

Сравнительная эффективность оценки с дисперсией $Dk(\mathrm{q})$ измеряется коэффициентом эффективности

$$\mathrm{e} = D(\mathrm{q})/Dk(\mathrm{q}),$$

который не превышает единицы. Чем ближе коэффициент e к единице, тем эффективнее оценка. Отмеченное ограничение применимо и к дискретным распределениям, если вместо плотности распределения подставить в него функцию вероятности.

Достаточность характеризует полноту использования информации, содержащейся в выборке. Другими словами, оценка q будет достаточной, если все другие независимые оценки на основе данной выборки не дают дополнительной информации об оцениваемом параметре. Эффективная оценка обязательно является и достаточной.

Рассмотренные свойства применимы также и к ЭД, которые характеризуются многомерными распределениями вероятностей.

Подходы к формированию оценок разработаны в теории несмещенных оценок , предложенной А. Н. Колмогоровым и С. Рао. В данной теории предполагается известным с точностью до параметра T вид функции плотности распределения наблюдаемой величины $f(x, T)$. Вид распределения устанавливается исходя из априорных соображений, например, на основе общепринятых суждений о характере безотказной работы технических средств. Тогда задача сводится к нахождению такой функции от результатов наблюдений, которая дает несмещенную и эффективную оценку.

2.3. Оценки моментов и квантилей распределения

> Первые понятия, с которых начинается какая-нибудь наука, должны быть ясны и приведены к самому меньшему числу.
>
> *Лобачевский Н.И. "О началах геометрии".*

Для характеристики эмпирического распределения можно использовать оценки **центральных и начальных моментов.** Применение находят моменты до четвертого порядка включительно, так как точность выборочных моментов резко падает с увеличением их порядка, в частности, дисперсия начальных моментов порядка r зависит от моментов порядка $2r$. Она становится значительной для моментов высокого порядка даже при больших объемах выборки. Выборочные значения моментов определяют непосредственно по выборке или по сгруппированным данным [3, 5, 9].

Выборочные значения центральных моментов случайной величины X вычисляются по выборке с применением с формул

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \tilde{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_1)^k, \quad k = 2,\ 3,\ 4. \tag{2.1}$$

Эти величины являются оценками соответствующих теоретических моментов

$m_1 - m_4$ и должны рассматриваться как случайные. Вычисления по формулам (2.1) дают состоятельные, но смещенные оценки моментов старше первого. Смещение удается устранить введением поправочных коэффициентов, зависящих от объема выборки. Несмещенными и состоятельными будут оценки

$$\mu_2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_1)^2, \tag{2.2}$$

$$\mu_3 = \frac{n^2}{(n-1)(n-2)}\mu_3,$$

$$\mu_4 = \frac{n(n^2 - 2n + 3)\mu_4 - 3n(2n-3)\mu_2^2}{(n-1)(n-2)(n-3)}.$$

Оценки моментов по сгруппированным ЭД

$$\tilde{\mu}_{1,g} = \frac{1}{n}\sum_{j=1}^{w} n_j X_{ц,j}, \tag{2.3}$$

$$\tilde{\mu}_{k,g} = \frac{1}{n}\sum_{j=1}^{w} n_j (X_{ц,j} - \mu_{1,g})^k, \quad k = 2,\ 3,\ 4,\ \ldots,$$

где $X_{ц\,j}$ – центр j-го интервала; y – количество интервалов.

Группирование и приписывание соответствующей частости значения варианты в середине интервала группирования вносят некоторые искажения. Если распределение

непрерывно и имеет достаточно высокий порядок соприкосновения с осью абсцисс (значения функции плотности распределения быстро убывают при удалении от центра распределения), то для снижения ошибок группирования используют **поправки Шеппарда**. Уточненные значения выборочных моментов для случая равной длины всех интервалов определяются через оценки моментов, вычисленные по сгруппированным данным:

$$\mu_1 = \mu_{1,g},\ \mu_2 = \mu_{2,g} - h^2/12, \qquad (2.4)$$
$$\mu_3 = \mu_{3,g},\ \mu_4 = \mu_{4,g} - \mu_{2,g}h^2/2 + 7h^4/240,$$

где h – длина интервала группирования. Указанные поправки ведут к уточнению только при соблюдении указанного условия, в противном случае они могут привести к еще большей ошибке.

Начальный эмпирический момент порядка r по несгруппированным данным определяется соотношением

$$\eta_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r, \quad r = 1, 2, 3, \ldots \qquad (2.5)$$

.

Центральные и начальные оценки моментов связаны между собой следующими зависимостями:

$$\mu_1 = \eta_1; \qquad \tilde{\mu}_2 = \eta_2 - \eta_1^2; \qquad (2.6)$$
$$\mu_3 = \eta_3 - 3\eta_1\eta_2 + 2\eta_1^3; \quad \mu_4 = \eta_4 - 4\eta_1\eta_3 + 6\eta_1^2\eta_2 - 3\eta_1^4.$$

В процессе обработки ЭД проще сначала определить оценки начальных моментов, потом перейти к смещенным оценкам центральных моментов и затем вычислить несмещенные оценки.

Квантилью, отвечающей уровню вероятности g, называют такое значение варианты x_g, при котором функция распределения случайной величины принимает значение g, т. е. квантиль – это значение аргумента x_g функции распределения, при котором $F(x_g)=g$. Эмпирическую квантиль находят по заданному значению вероятности g, используя вариационный ряд или ступенчатую ломаную линию.

Наряду с указанными параметрами для описания распределений применяются и другие характеристики:

среднеквадратическое отклонение $\sigma = \sqrt{\mu_2}$;

коэффициенты асимметрии $\beta_1 = \mu_3/\mu^3$ и эксцесса $\beta_2 = \mu_4/\mu_2^2$;

стандартизованные переменные $\boldsymbol{u=(x-m_1)/s}$.

Коэффициент асимметрии характеризует "скошенность" распределения относительно симметричного нормального распределения (у любого симметричного распределения $b_1=0$), рис. 2.4. Этот показатель в основном зависит от крайних значений выборки.

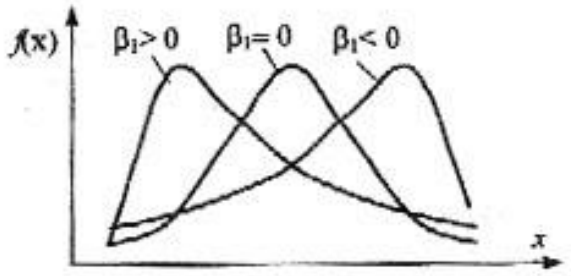

Рис. 2.4. Асимметрия распределения

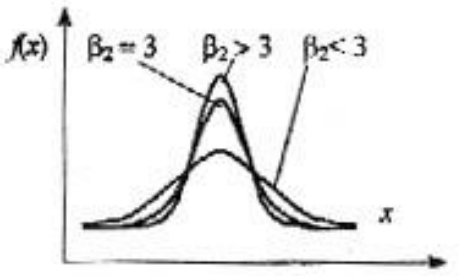

Рис. 2.5. Эксцесс распределения

Коэффициент эксцесса характеризует островершинность распределения относительно нормального распределения (этот коэффициент у нормального распределения равен трем), рис. 2.5. Термин "эксцесс" (превышение) целесообразно применять не к величине b_2, а к сравнению этой величины изучаемого распределения с величиной данного коэффициента нормального распределения, т. е. с величиной, равной трем. Исходя из этого, часто вместо b_2 используют величину $b_2–3$.

Стандартизация переменной позволяет упростить расчеты, кроме того, в литературе многие справочные статистические таблицы приводятся именно для стандартизованных переменных. Нетрудно показать, что математическое ожидание стандартизованной переменной равно нулю, а дисперсия равна единице, т.е. после такого преобразования ЭД справедливы следующие соотношения:

$$M(u)=\frac{1}{n}\sum_{i=1}^{n}u_i=0\,;\quad D(u)=\frac{1}{n}\sum_{i=1}^{n}(u_i-0)^2=1\,.$$

Величина *u* называется *центрированной и нормированной*. Переход от центрированной и нормированной величины к исходной осуществляется простым преобразованием $x=us+m_1$. Потери информации при стандартизации и обратном преобразовании не происходит.

Пример 2.3. Необходимо определить числовые характеристики распределения по данным, представленным в виде вариационного и статистического ряда, табл. 2.3 и 2.4 соответственно.

Решение. Вычислим значения центральных моментов по вариационному ряду, пользуясь формулами (2.1)

$$m_1 = 27{,}508;\ \mu_2 = 0{,}893;\ \mu_3 = 0{,}123;\ \mu_4 = 1{,}656.$$

Эти оценки, кроме математического ожидания, являются смещенными. Несмещенные оценки получим на основе (2.2)

$$m_2 = 0{,}913;\ m_3 = 0{,}132;\ m_4 = 1{,}819;\ s = 0{,}956.$$

Вычисление оценок моментов на основе статистического ряда по (2.3) дает следующие результаты:

$$m_1 = 27{,}482;\ \mu_2 = 0{,}805;\ \mu_3 = 0{,}137;\ \mu_4 = 1{,}320.$$

Судя по гистограмме, по крайней мере левый край распределения не имеет гладкого соприкосновения с осью *x*, поэтому поправки Шеппарда нецелесообразны.

Значения оценок моментов различаются при их вычислении по вариационному ряду и по сгруппированным данным. Можно предполагать, что оценки, вычисленные по вариационному ряду, будут точнее оценок, рассчитанных по статистическому ряду.

Оценка коэффициента асимметрии $b_1=0{,}132/0{,}9131{,}5=0{,}15$ говорит о небольшой положительной асимметрии распределения (мода функции плотности распределения находится левее математического ожидания), а оценка коэффициента эксцесса

b_2=1,819/0,9132=2,18 – о пологости распределения ("островершинность" выражена слабее, чем у нормального распределения).

Анализируя назначение рассмотренных параметров, необходимо отметить следующее. Одни параметры характеризует средние величины, а другие – вариацию. ***Главное назначение средних величин (оценок начальных моментов и в первую очередь первого момента распределения) состоит в их обобщающей функции.*** Это обобщение позволяет заменить множество различных индивидуальных значений показателя средней величиной, характеризующей всю однородную совокупность. Иначе говоря, средняя величина является типической характеристикой варианты в конкретной выборке. Иногда средняя величина обобщает и неоднородные совокупности данных. Например, может применяться такой показатель как среднее количество обработанных запросов на сервере в течение суток, хотя очевидно, что дневная загрузка сервера сильно отличается от загрузки в ночное время. Указанный показатель имеет смысл для оценки ресурса накопителей на жестких дисках. **Наряду с оценками математического ожидания (средними величинами в формулах 2.1 и 2.5) находят применение и другие оценки – среднее геометрическое, среднее гармоническое значение [5].**

Каждый элемент ЭД формируется под влиянием как общих закономерностей, так и особых условий и случайных событий. Следовательно, **в обработке ЭД большой интерес представляют вопросы оценки величин, характеризующих вариацию значений параметра у разных объектов или у одного и того же объекта в разные моменты времени.** ***Вариацией какого-либо параметра (показателя) в совокупности наблюдений называется различие его значений у разных элементов этой совокупности.*** **Вариация является характерным свойством большинства информационных параметров АСОИУ.** Именно это свойство является объектом исследования большинства методов обработки ЭД. Для характеристики вариации нет единого показателя, в этих целях применяются моменты распределения выше первого, производные от них величины, размах выборки, квантили и др.