1. Основные задачи первичной обработки данных

Современная практика прикладных статистических исследований свидетельствует, что для достижения успеха в статистических приложениях исследователь должен хорошо ориентироваться в прикладных методах математической статистики.

В настоящей работе предпринята попытка систематического и взаимосвязанного изложения всех аспектов проблемы статистического анализа выборочных данных, которые приходится решать исследователю на первом этапе обработки результатов эксперимента или наблюдения, представленных выборочными данными. Этот этап назван этапом первичной статистической обработки и включает в себя следующие задачи:

1. Составить дискретный вариационный ряд для статистической обработки результатов наблюдений.
2. Для полученного вариационного ряда определить следующие точечные оценки параметров распределения:
   (a) Меры центральной тенденции.
   (b) Абсолютные и средние показатели вариации данных.
   (c) Показатели относительного рассеивания.
3. Построить доверительные интервалы для математического ожидания и дисперсии генеральной совокупности с надежностью $p = 0.95$.
4. Построить гистограмму (эмпирическую плотность распределения) и кумуляту (эмпирическую функцию распределения).
5. Проверить гипотезу о том, что случайная величина распределена по нормальному закону, используя точечные оценки.
6. Проверить гипотезу о том, что случайная величина распределена по нормальному закону, используя критерий согласия χ^2 Пирсона при уровне значимости $\alpha = 0.05$. Построить на одном чертеже гистограмму и соответствующую нормальную кривую. Построить на одном чертеже кумуляту и соответствующую теоретическую кривую.

2. Точечные оценки параметров распределения

Для данного статистического распределения выборки построим вариационный ряд x_i, $i = 1, \dots, n$, для чего выборочные значения расположим в порядке возрастания. Для данного вариационного ряда определяем следующие точечные оценки.

2.1. Меры центральной тенденции

Меры центральной тенденции отражают тенденцию развития, т. е. действие главных причин (факторов) на распределение изучаемого признака.

1. Выборочная средняя арифметическая (оценка математического ожидания теоретического распределения)

$$\bar{X} = \frac{1}{n} \sum x_i.$$

2. Выборочная медиана. Вычисление медианы производится по формуле

$$M_{\mathrm{e}} = x_{(n-1)/2+1} \text{ при нечетном } n,$$

$$M_{\mathrm{e}} = \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right) \text{ при четном } n.$$

3. Выборочная мода. Вычисление моды производится по формуле

$$M_{\mathrm{o}} = \bar{X} + 3\left(M_{\mathrm{e}} - \bar{X}\right).$$

Выборочная медиана и мода относятся к классу порядковых статистик.

2.2. Абсолютные и средние показатели вариации данных

Различие индивидуальных значений признака внутри изучаемой совокупности в статистике называется вариацией признака. Она возникает в результате того, что его индивидуальные значения складываются под совокупным влиянием разнообразных факторов (условий), которые поразному сочетаются в каждом отдельном случае.

Колеблемость отдельных значений характеризует показатели вариации. Термин 'вариация' произошел от латинского variatio – изменение, колеблемость, различие. Однако не всякие различия принято называть вариацией. Под вариацией в статистике понимают такие количественные изменения величины исследуемого признака в пределах однородной совокупности, которые обусловлены перекрещивающимся влиянием действия различных факторов.

Для характеристики колеблемости признака используется ряд показателей.

1. Размах вариации – простейший показатель, определяемый как разность между наибольшим (x_{max}) и наименьшим (x_{min}) значениями вариантов,
$$X_R = x_{max} - x_{min}.$$
2. Среднее линейное отклонение, которое учитывает различие всех единиц изучаемой совокупности,
$$\bar{D} = \frac{1}{n} \sum |x_i - \bar{X}|.$$
3. Выборочная дисперсия (оценка дисперсии теоретического распределения),
$$S^2 = m_2 = \frac{1}{n} \sum (x_i)^2 - \bar{X}^2.$$
Здесь m_k – центральный момент k-го порядка,
$$m_k = \frac{1}{n} \sum (x_i - \bar{X})^k.$$
4. Несмещенная оценка дисперсии (исправленная выборочная дисперсия) определяется соотношением
$$\bar{S}^2 = \frac{1}{n-1} \sum (x_i - \bar{X})^2 = \frac{n}{n-1} S^2.$$
5. Выборочные среднеквадратичные отклонения соответственно могут быть найдены по формулам
$$S = \sqrt{S^2}, \qquad \bar{S} = \sqrt{\bar{S}^2}.$$
6. Эти показатели на практике более объективно отражают меру вариации и являются мерилом надежности оценки среднего. Эта надежность определяется ошибкой выборочного среднего
$$S_{\bar{X}} = \frac{\bar{S}}{\sqrt{n}}.$$

Чем меньше $\bar{S}$, тем лучше средняя арифметическая отражает собой всю представляемую совокупность.

2.3. Показатели относительного рассеивания

Эти показатели характеризуют меру колеблемости изучаемого признака в относительных величинах:

1. Коэффициент вариации, являющийся мерой относительной изменчивости наблюдаемой случайной величины. Выборочный коэффициент вариации V и исправленный коэффициент вариации $\bar{V}$ вычисляются по формулам

$$V = \frac{S}{\bar{X}} \cdot 100(\%), \quad \bar{V} = \frac{\bar{S}}{\bar{X}} \cdot 100(\%).$$

2. Относительное линейное отклонение характеризует долю усредненного значения абсолютных отклонений от средней величины

$$K_D = \frac{\bar{D}}{\bar{X}} \cdot 100(\%).$$

3. Коэффициент осцилляции, который отражает относительную колеблемость крайних значений признака вокруг средней

$$K_O = \frac{X_R}{\bar{X}} \cdot 100(\%).$$

Близость выборочных оценок мер центральной тенденции и показателей относительного рассеивания свидетельствует о высокой вероятности отсутствия грубых погрешностей опытных данных.

2.4. Показатели асимметрии и эксцесса

1. Точечная оценка асимметрии и соответствующая средняя квадратичная ошибка определения неисправленной асимметрии:

$$A = \frac{m_3}{(m_2)^{3/2}} = \frac{\Sigma(x_i - \bar{X})^3}{nS^3}, \quad S_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

2. Несмещенная (исправленная) оценка асимметрии и соответствующая средняя квадратичная ошибка определения исправленной асимметрии

$$\bar{A} = \frac{\sqrt{n(n-1)}}{n-2}A, \quad S_{\bar{A}} = \sqrt{\frac{6n(n-1)}{(n-1)(n+1)(n+3)}}.$$

3. Точечная оценка эксцесса и соответствующая средняя квадратичная ошибка определения неисправленного эксцесса

$$E = \frac{m_4}{m_2^4} = \frac{\Sigma(x_i - \bar{X})^4}{nS^4} - 3, \quad S_E = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}.$$

4. Несмещенная (исправленная) оценка эксцесса и соответствующая средняя квадратичная ошибка определения исправленного эксцесса

$$\bar{E} = \frac{n-1}{(n-2)(n-3)}[(n+1)E+6], \quad S_{\bar{E}} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}}.$$

3. Интервальные оценки математического ожидания и дисперсии генеральной совокупности

Оценка математического ожидания является несмещенной. Поэтому рассматривается симметричный доверительный интервал. При неизвестной дисперсии статистика

$$g = \frac{\bar{X} - M(X)}{\bar{S}}\sqrt{n}$$

имеет t – распределение с $\nu = n - 1$ степенями свободы. Отсюда получаются следующие границы доверительного интервала, который покрывает математическое ожидание с надежностью p:

$$\bar{X} - S_{\bar{X}}t((1+p)/2, n-1) < M(X) < \bar{X} + S_{\bar{X}}t((1+p)/2, n-1).$$

Здесь $t(q,\nu)$ – квантиль распределения Стьюдента на уровне q с ν степенями свободы.

При неизвестном математическом ожидании статистика

$$g = (n-1)\frac{\bar{S}^2}{D(X)}$$

имеет χ^2 – распределение с $\nu = n-1$ степенями свободы. Отсюда получаются следующие границы доверительного интервала, который покрывает дисперсию с надежностью p:

$$\frac{(n-1)\bar{S}^2}{\chi^2((1+p)/2; n-1)} < D(X) < \frac{(n-1)\bar{S}^2}{\chi^2((1-p)/2; n-1)}.$$

Здесь $\chi^2(q,\nu)$ – квантиль распределения Пирсона на уровне q с ν степенями свободы.

4. Пример выполнения индивидуального задания

4.1. Задание

Типичное задание для самостоятельной работы состоит в следующем. В табл. 1 представлены выборочные данные, объем выборки равен 30.

Табл. 1. Выборочные данные

151	168	170	188	158	170	146	162	166	180	176	176	182	162	173	166	173
166	162	169	181	164	166	183	172	174	166	154	167	158				

Провести первичную обработку данных согласно описанному выше алгоритму.

4.2. Вариационный ряд

Для данной выборки построим вариационный ряд x_i, $i = 1, \ldots, n$, для чего выборочные значения расположим в порядке возрастания. Получим вариационный ряд, представленный в табл. 2.

Табл. 2. Вариационный ряд

146	151	154	158	158	162	162	162	164	166	166	166	166	166	167	168	169
170	170	172	173	173	174	176	176	180	181	182	183	188				

4.3. Вычисление точечных оценок

Для вычисления точечных оценок используем вариационный ряд и составим вспомогательную табл. 3.

Табл. 3. Таблица для расчета точечных оценок

i	x_i	x_i^2	$x_i - \bar{X}$	$\lvert x_i - \bar{X}\rvert$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^3$	$(x_i - \bar{X})^4$
1	146	21316	-22.3	22.3	497.29	-11089.6	247297
2	151	22801	-17.3	17.3	299.29	-5177.72	89574.5
3	154	23716	-14.3	14.3	204.49	-2924.21	41816.2
4	158	24964	-10.3	10.3	106.09	-1092.73	11255.1
5	158	24964	-10.3	10.3	106.09	-1092.73	11255.1
6	162	26244	-6.3	6.3	39.69	-250.047	1575.3
7	162	26244	-6.3	6.3	39.69	-250.047	1575.3
8	162	26244	-6.3	6.3	39.69	-250.047	1575.3
9	164	26896	-4.3	4.3	18.49	-79.507	341.88
10	166	27556	-2.3	2.3	5.29	-12.167	27.9841
11	166	27556	-2.3	2.3	5.29	-12.167	27.9841
12	166	27556	-2.3	2.3	5.29	-12.167	27.9841
13	166	27556	-2.3	2.3	5.29	-12.167	27.9841
14	166	27556	-2.3	2.3	5.29	-12.167	27.9841
15	167	27889	-1.3	1.3	1.69	-2.197	2.8561
16	168	28224	-0.3	0.3	0.09	-0.027	0.0081
17	169	28561	0.7	0.7	0.49	0.343	0.2401
18	170	28900	1.7	1.7	2.89	4.913	8.3521
19	170	28900	1.7	1.7	2.89	4.913	8.3521
20	172	29584	3.7	3.7	13.69	50.653	187.416
21	173	29929	4.7	4.7	22.09	103.823	487.968
22	173	29929	4.7	4.7	22.09	103.823	487.968
23	174	30276	5.7	5.7	32.49	185.193	1055.6
24	176	30976	7.7	7.7	59.29	456.533	3515.3
25	176	30976	7.7	7.7	59.29	456.533	3515.3
26	180	32400	11.7	11.7	136.89	1601.61	18738.9
27	181	32761	12.7	12.7	161.29	2048.38	26014.5
28	182	33124	13.7	13.7	187.69	2571.35	35227.5
29	183	33489	14.7	14.7	216.09	3176.52	46694.9
30	188	35344	19.7	19.7	388.09	7645.37	150614
Σ	5049	852431	0	221.6	2684.3	-3859.68	692965

В результате получаем точечные оценки:

1. Меры центральной тенденции.

   (a) Выборочная средняя арифметическая:

$$\bar{X} = \frac{5049}{30} = 168.3.$$

(b) Выборочная медиана. Используем вариационный ряд, представленный в табл. 2. Так как n – четное, то

$$M_{\mathrm{e}} = \frac{1}{2}(x_{15} + x_{16}) = \frac{1}{2}(167 + 168) = 167.5.$$

(c) Выборочная мода:

$$M_{\mathrm{O}} = 168.3 + 3(167.5 - 168.3) = 165.9.$$

2. Абсолютные и средние показатели вариации данных.

(a) Размах вариации:

$$X_R = 188 - 146 = 42.$$

(b) Среднее линейное отклонение:

$$\bar{D} = \frac{221.6}{30} = 7.38667.$$

(c) Выборочная дисперсия и исправленная выборочная дисперсия:

$$S^2 = \frac{852431}{30} - 168.3^2 = 89.4767, \quad \bar{S}^2 = 92.5621.$$

(d) Соответствующие выборочные средние квадратичные отклонения:

$$S = \sqrt{89.4767} = 9.45921, \quad \bar{S} = \sqrt{92.5621} = 9.62092.$$

(e) Ошибка выборочного среднего

$$S_{\bar{X}} = \frac{9.62092}{\sqrt{30}} = 1.75653\ .$$

3. Показатели относительного рассеивания.

(a) Коэффициент вариации и исправленный коэффициент вариации:

$$V = \frac{9.45921}{168.3} \cdot 100 = 5.62\%, \quad \bar{V} = \frac{9.62092}{168.3} \cdot 100 = 5.72\%.$$

(b) Относительное линейное отклонение:

$$K_D = \frac{7.38667}{168.3} \cdot 100 = 4.39\%.$$

(c) Коэффициент осциляции

$$K_O = \frac{42}{168.3} \cdot 100 = 25\%.$$

4. Для вычисления моментов и показателей асимметрии и эксцесса используем табл. 3. В результате получим:

   (a) Точечные оценки асимметрии и ошибки оценок

   $$A = -0.152008, \quad S_A = 0.412417, \quad \bar{A} = -0.160128, \quad S_{\bar{A}} = 0.419468.$$

   (b) Точечные оценки эксцесса и ошибки оценок

   $$E = -0.114837, \quad S_E = 0.74858, \quad \bar{E} = 0.0936, \quad S_{\bar{E}} = 0.832746.$$

4.4. Вычисление интервальных оценок

При $p = 0.95$ и $\nu = n - 1 = 29$ находим

$$t((1+p)/2, n-1) = 2.04572, \quad S_{\bar{X}} t((1+p)/2, n-1) = 3.59337.$$

Отсюда доверительный интервал для математического ожидания

$$164.707 < M(X) < 171.893.$$

При $p = 0.95$ и $\nu = n - 1 = 29$ находим

$$\chi^2((1+p)/2; n-1) = 45.7295, \quad \chi^2((1-p)/2; n-1) = 16.8719.$$

Отсюда доверительный интервал для дисперсии

$$58.6995 < D(X) < 159.099.$$

5. Эмпирические характеристики распределения

5.1. Эмпирическая плотность распределения и гистограмма

Для построения эмпирических характеристик распределения превратим данный дискретный вариационный ряд в интервальный. Для этого разобьем весь диапазон наблюдаемых значений на k интервалов (классов). Число классов определим по правилу Штюргеса

$$k = [1 + 3.32 \lg n].$$

Число классов рекомендуется выбирать нечетным. Имеем

$$1 + 3.32 lg 30 = 5.90404,$$

отсюда $k = 5$. Длину частичных интервалов определим по формуле

$$h = \frac{x_{max} - x_{min}}{k} = 8.4.$$

Границы интервалов ξ_i, $i = 0, \ldots, k$ определим соотношением

$$\xi_i = x_{min} + ih, \quad \xi_0 = x_{min}, \quad \xi_k = x_{max}.$$

Подсчитаем частоты n_i попадания наблюдаемых значений случайной величины в частичные интервалы ($\sum n_i = n$). Все значения признака в пределах интервала приравняем его срединному значению, считая, что частоты n_i относятся к середине интервала

$$x_i^* = \frac{1}{2}(\xi_i + \xi_{i+1}).$$

Обозначим через $w_i = n_i/n$ относительные частоты, через w_i/h плотности относительных частот, через $c_i = \sum\limits_{j=1}^{i} w_j$ относительные накопленные частоты, через c_i/h плотности относительных накопленных частот, где h – величина шага (в нашем случае $h = 8.4$). Результаты представлены в табл. 4.

Табл. 4. Интервальный ряд распределения относительных частот

$[\xi_i, \xi_{i+1}]$	[146;154.4)	[154.4;162.8)	[162.8;171.2)	[171.2;179.6)	[179.6;188]
x_i^*	150.2	158.6	167	175.4	183.8
n_i	3	5	11	6	5
w_i	0.1	0.166667	0.366667	0.2	0.1667
w_i/h	0.0119048	0.0198413	0.0436508	0.0238095	0.01984
c_i	0.1	0.266667	0.633333	0.833333	1
c_i/h	0.0119048	0.031746	0.0753968	0.0992063	0.119

График эмпирической плотности распределения представляется гистограммой. Для построения гистограммы на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольник, высота которого равна w_i/h. Гистограмма изображена на рисунке (см. стр. 18).

5.2. Эмпирическая функция распределения и кумулята

Аналитическая формула для эмпирической функции распределения определяется множеством относительных накопленных частот и имеет вид

$$F(x) = \begin{cases} 0.1, & \textit{если} \quad 146 < x \leq 154.4 \\ 0.266667, & \textit{если} \quad 154.4 < x \leq 162.8 \\ 0.633333, & \textit{если} \quad 162.8 < x \leq 171.2 \\ 0.833333, & \textit{если} \quad 171.2 < x \leq 179.6 \\ 1, & \textit{если} \quad 179.6 < x \leq 188 \end{cases}$$

График эмпирической функции распределения представляется кумулятой. Для построения кумуляты на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольник, высота которого равна c_i/h. Кумулята изображена на рисунке (см. стр. 18).

6. Проверка гипотезы нормального распределения по точечным оценкам

Проверка соответствия результатов измерения закону нормального распределения является важным моментом предварительной обработки данных. Можно предположить, что значение случайной величины изменяется под влиянием большого числа факторов, примерно равнозначных по силе, поэтому распределение случайной величины является нормальным. Если эта гипотеза неприемлема, то следует определить, какому закону распределения подчиняются опытные данные и, если это возможно, преобразовать данное распределение к нормальному. Только после выполнения перечисленных выше операций можно перейти к построению эмпирических формул, применяя, например, метод наименьших квадратов.

6.1. Проверка гипотезы нормального распределения по оценке среднего линейного отклонения

Для не очень больших выборок ($n < 120$) проверку гипотезы нормальности распределения можно провести по среднему линейному отклонению. Для выборки, имеющей приближенно нормальный закон распределения, должно быть справедливо выражение

$$\left| \frac{\bar{D}}{\bar{S}} - 0.7979 \right| < \frac{0.4}{\sqrt{n}}.$$

Для исходных данных проверяем неравенство

$$\left|\frac{7.38667}{9.62092} - 0.7979\right| = 0.0301285 < \frac{0.4}{\sqrt{30}} = 0.0730297.$$

Так как неравенство выполняется, гипотеза нормальности может быть принята.

6.2. Проверка гипотезы нормального распределения по оценкам асимметрии и эксцесса

Некоторое представление о близости эмпирического распределения к нормальному может дать анализ показателей асимметрии и эксцесса.

Критерием нормальности распределения случайной величины является равенство нулю асимметрии и эксцесса. Если выполняются условия

$$|\bar{A}| \leq 3S_{\bar{A}}, \quad |\bar{E}| \leq 5S_{\bar{E}},$$

то гипотеза нормальности исследуемого распределения может быть принята. Имеем

$$\bar{A} = -0.160128, \quad 3S_{\bar{A}} = 1.2584,$$
$$\bar{E} = 0.0936, \quad 5S_{\bar{E}} = 4.16373.$$

В данном случае гипотеза нормальности исследуемого распределения может быгь принята.

6.3. Проверка гипотезы нормального распределения по правилу трех сигм

Проверку гипотезы нормальности распределения можно провести по правилу трех сигм, согласно которому при нормальном распределении признака $x_{min} > \bar{X} - 3\bar{S}$, $x_{max} < \bar{X} + 3\bar{S}$. Проверим это:

$$\bar{X} - 3\bar{S} = 139.437 < 146,$$
$$\bar{X} + 3\bar{S} = 197.163 > 188.$$

В данном случае гипотеза нормальности исследуемого распределения может быть принята.

7. Проверка гипотезы нормального распределения по критерию согласия Пирсона (критерий χ^2)

Рассмотрим критерий χ^2.

Можно предположить, что значение случайной величины изменяется под влиянием большого числа факторов, примерно равнозначных по силе, поэтому распределение случайной величины является нормальным.

Нормализуем случайную величину X, т.е. перейдем к случайной величине $Z = (X - \bar{X})/S$. Вычислим концы интервалов

$$z_i = \frac{\xi_i - \bar{X}}{S}, \quad z_{i+1} = \frac{\xi_{i+1} - \bar{X}}{S},$$

причем значение z_1 полагаем равным $-\infty$, а наибольшее значение z_{k+1} полагаем равным $+\infty$. Для этого составим расчетную табл. 5.

Табл. 5. Таблица для расчета нормализованных границ интервалов

i	ξ_i	ξ_{i+1}	$\xi_i - \bar{X}$	$\xi_{i+1} - \bar{X}$	z_i	z_{i+1}
1	146	154.4	-22.3	-13.9	$-\infty$	-1.46947
2	154.4	162.8	-13.9	-5.5	-1.46947	-0.581444
3	162.8	171.2	-5.5	2.9	-0.581444	0.306579
4	171.2	179.6	2.9	11.3	0.306579	1.1946
5	179.6	188	11.3	19.7	1.1946	$+\infty$

Вычислим теоретические частоты

$$n'_i = nP_i,$$

где n – объем выборки,

$$P_i = \Phi(z_{i+1}) - \Phi(z_i),$$

$\Phi(z)$ – функция Лапласа, определяемая соотношением

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z \exp\left(-\frac{x^2}{2}\right) dx.$$

Для этого составим расчетную табл. 6.

Табл. 6. Таблица для расчета теоретических частот

i	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = nP_i$
1	-0.5	-0.429147	0.0708526	2.12558
2	-0.429147	-0.219529	0.209618	6.28854
3	-0.219529	0.120418	0.339948	10.1984
4	0.120418	0.383879	0.263461	7.90383
5	0.383879	0.5	0.116121	3.48362

Вычислим наблюдаемое значение критерия Пирсона

$$\chi^2_{\text{набл}} = \sum \frac{(n_i - n'_i)^2}{n'_i}.$$

Для этого составим расчетную табл. 7.

Табл. 7. Таблица для расчета выборочного значения критерия Пирсона

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2/n'_i$
1	3	2.12558	0.874421	0.764613	0.35972
2	5	6.28854	-1.28854	1.66033	0.264025
3	11	10.1984	0.801566	0.642508	0.0630006
4	6	7.90383	-1.90383	3.62455	0.458582
5	5	3.48362	1.51638	2.2994	0.66006
Σ	30	30			$\chi^2_{\text{набл}}$=1.80539

Наблюдаемое значение критерия Пирсона равно 1.80539. По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0.05$ и числу степеней свободы $k - 3 = 2$ находим критическую точку правосторонней критической области $\chi^2_{\text{кр}}(0.05; 2) = 5.94$. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, принимаем гипотезу о нормальном распределении генеральной совокупности X.

Плотность вероятности нормального распределения имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\alpha)^2}{2\sigma^2}\right).$$

Точечные оценки параметров α и σ были найдены:

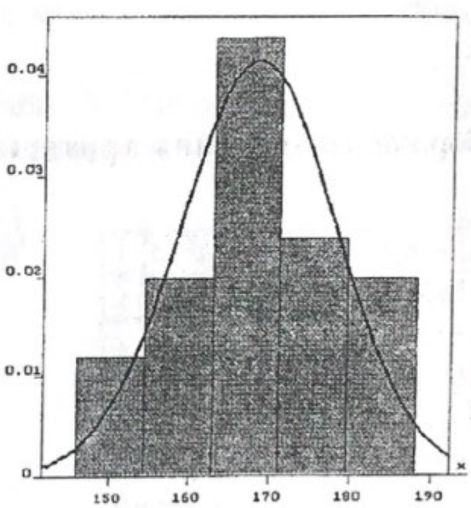
$$\alpha \approx \bar{X} = 168.3, \quad \sigma \approx \bar{S} = 9.62092.$$

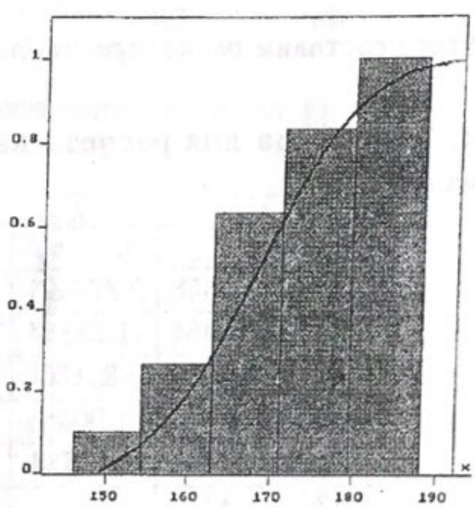
Гистограмма и нормальная кривая $f(x)$ с параметрами $\bar{X}$ и $\bar{S}$ изображены на рисунке. Максимальное значение кривой Гаусса равно 0.0414661.

Функция распределения предполагаемого нормального распределения имеет вид

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \frac{1}{2} + \Phi\left(\frac{x-\alpha}{\sigma}\right).$$

Кумулята и теоретическая кривая $F(x)$ с параметрами $\bar{X}$ и $\bar{S}$ изображены на рисунке.





Гистограмма эмпирического распределения и соответствующая нормальная кривая (a). Кумулята эмпирического распределения и соответствующая теоретическая кривая (b).