Санкт-Петербургский государственный университет

Н. М. Москалева

Практикум по математической статистике

часть 1

Учебно-методическое пособие

Санкт-Петербург

2018

ПРЕДИСЛОВИЕ

Практикум по математической статистике предназначен для студентов второго курса Института наук о Земле, обучающимся по направлениям «География» и «Экология и природопользование». Пособие является дополнением к электронной версии лекций по теории вероятностей и математической статистике. В пособие включены задачи, связанные с методами описательной статистики и оцениванием числовых характеристик исследуемой модели. Задачи подобраны и составлены с учетом области приложения теории. В предлагаемых задачах и примерах в основном использованы реальные статистические данные. Пособие предназначено для самостоятельной работы и практических занятий, проводимых в компьютерном классе. Практические задания, требующие трудоемких вычислений, выполняются с помощью статистического программного пакета Excel.

Тема 1: Построение гистограммы частот.

Пусть $x_1, x_2, \dots, x_n$ – наблюдаемые значения случайной величины X. Если n велико, то выборочные значения обычно подвергают группировке. Интервал, содержащий все n наблюдений, разбивают на k непересекающихся интервалов $\Delta_1, \Delta_2, \dots \Delta_{k.}$ Для каждого интервала Δ_j подсчитывают частоту ν_j наблюдений, попавших в этот интервал. Удобно, но не обязательно, брать интервалы одинаковой длины h.

Соответствие между интервалами Δ_j и частотами ν_j (относительными частотами $\gamma_j\ /\ n$) называют *группированным статистическим рядом.*

$\Delta_1 \sim (d_0, d_1]$	$\Delta_2 \sim (d_1, d_2]$	…	$\Delta_k \sim (d_{k-1}, d_k]$
ν_1	ν_2	…	ν_k
$\gamma_1\ /\ n$	$\gamma_2\ /\ n$	…	$\gamma_k\ /\ n$

(1)

Группированный статистический ряд можно графически изобразить в виде *гистограммы частот***,** т.е. графика кусочно-постоянной функции со значениями ν_j на интервале Δ_j.

Построение гистограммы частот по выборке $x_1, x_2, \dots, x_n$ с помощью инструмента АНАЛИЗ ДАННЫХ (пакет MS Excel).

1) Предварительно найти:
   - x_{min} и x_{max} – наименьший и наибольший из элементов выборки $x_1, x_2, \dots, x_n$;
   - размах выборки $R = x_{max} - x_{min}$ и при заданном количестве интервалов $\boldsymbol{k}$ длину интервала $h = \frac{R}{k}$;
   - границы интервалов: $d_0 = x_{min}$, $d_k = x_{max}$, $d_j = x_{min} + jh$, $j = 1, \dots, k$;

2) Для определения частоты ν_j и построения гистограммы частот обратиться к программе пакета MS Excel: ДАННЫЕ — АНАЛИЗ ДАННЫХ — ГИСТОГРАММА. В качестве входных данных ввести: в поле ВХОДНОЙ ИНТЕРВАЛ − адреса ячеек, содержащих выборку $x_1, x_2, \dots, x_n$, в поле ИНТЕРВАЛ КАРМАНОВ − адреса ячеек, содержащих границы интервалов $d_1, \dots, d_{k-1}$. Указать выходные параметры: необходимость ВЫВОДА ГРАФИКА и в поле ВЫХОДНОЙ ИНТЕРВАЛ − адреса ячеек для результатов, получаемых в виде группированного статистического ряда.

Замечание. Количество интервалов k выбирают в зависимости от объема выборки n. Рекомендуемое число интервалов может быть найдено по одной из формул: $k = [1 + 3{,}322 \cdot \lg n]$ (формула Старджесса) или $k = \left[1{,}72 n^{\frac{1}{3}}\right]$, где $[\alpha]$ − целая часть числа $\alpha > 0$.

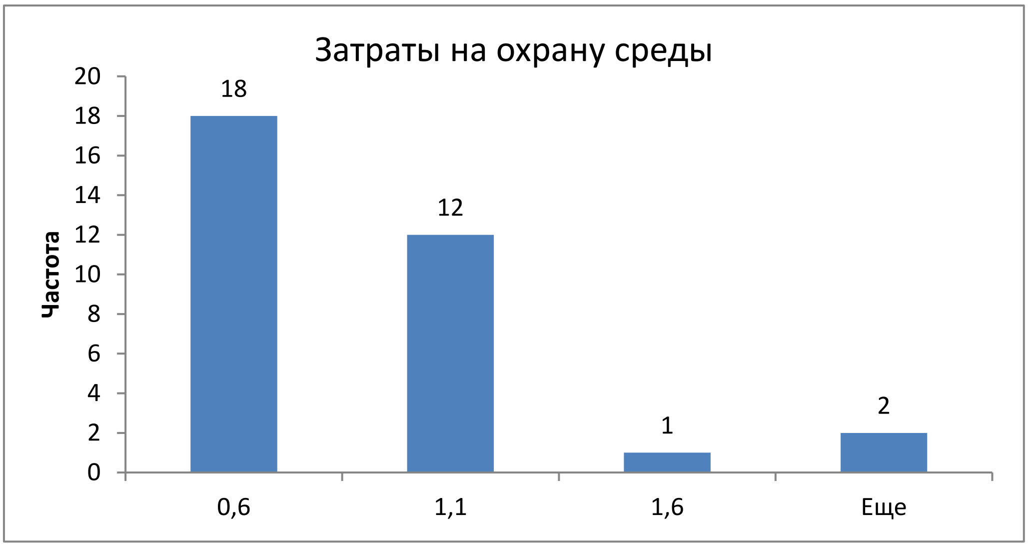
Пример. В таблице приводятся затраты стран мира на охрану окружающей среды.

Удельный вес расходов на охрану окружающей среды (в процентах к ВВП) - 2008

Россия	0.9	Испания	0.9	Франция	0.9
Австрия	0.4	Италия	0.8	Чехия	1.9
Беларусь	0.4	Люксембург	1.1	Швейцария	0.6
Бельгия	0.6	Молдова	0.1	Швеция	0.4
Болгария	2.1	Нидерланды	0.8	Израиль	0.7
Великобритания	1	Норвегия	0.6	Иран	0.5
Венгрия	0.7	Польша	0.6	Казахстан	0.1
Германия	0.5	Португалия	0.5	Япония	1.2
Греция	0.5	Румыния	0.4	Канада	0.7
Дания	0.5	Украина	0.2	Австралия	0.5
Ирландия	0.7	Финляндия	1.1	Новая Зеландия	0.5

По данной выборке построить гистограмму частот для затрат стран мира с заданным числом интервалов $k = 4$.

В результате обращения к программе ДАННЫЕ — АНАЛИЗ ДАННЫХ — ГИСТОГРАММА получаем гистограмму



и статистический ряд

Карман	*Частота*
0,6	18
1,1	12
1,6	1
Еще	2

Из таблицы следует, что в 18 странах затраты на охрану окружающей среды не превышают 0.6, затраты 12 стран находятся в промежутке (0.6; 1.1], затраты одной страны лежат в промежутке (1.1; 1.6] и затраты двух стран больше 1,6.

Задачи для самостоятельного решения

Задача 1. 1. Рейтинг 100 стран мира по уровню безработицы составлен на основе подтвержденных данных Всемирного банка и Международной организации труда. Уровень безработицы выражается в процентах и рассчитывается как отношение числа безработных к общей численности экономически активного населения страны. (2017 г.)

№	Страна	% безработных	№	Страна	% безработных
1	Катар	0,2	51	Чили	7
2	Беларусь	0,5	52	Монголия	7
3	ОАЭ	1,7	53	Литва	7,1
4	Сингапур	2	54	Виргинские о-ва	7,1
5	Кувейт	2,1	55	Узбекистан	7,2
6	Вьетнам	2,1	56	Киргизия	7,3
7	Куба	2,6	57	Бельгия	7,4
8	Япония	2,8	58	Словакия	7,9
9	Исландия	2,9	59	Уругвай	8,1
10	Чехия	3,1	60	Венесуэла	8,1
11	Гонконг	3,2	61	Ангола	8,2
12	Малайзия	3,4	62	Ирак	8,2
13	Туркменистан	3,4	63	Аргентина	8,7
14	Индия	3,5	64	Финляндия	8,7
15	Мексика	3,5	65	Афганистан	8,8
16	Германия	3,7	66	Колумбия	9
17	Перу	3,7	67	Португалия	9
18	Южная Корея	3,8	68	Латвия	9,1

19	Пакистан	4	69	Марокко	9,3
20	Норвегия	4,2	70	Украина	9,5
21	Великобритания	4,3	71	Франция	9,7
22	Венгрия	4,3	72	Алжир	10
23	Индонезия	4,3	73	Таджикистан	10,3
24	Израиль	4,3	74	Кипр	10,8
25	США	4,4	75	Хорватия	10,8
26	Молдова	4,5	76	Италия	11,3
27	Китай	4,7	77	Турция	11,3
28	Швейцария	4,8	78	Грузия	11,5
29	Северная Корея	4,8	79	Египет	12,1
30	Казахстан	4,9	80	Ямайка	12,4
31	Нидерланды	4,9	81	Багамские о-ва	12,6
32	Новая Зеландия	4,9	82	Судан	12,8
33	Азербайджан	5	83	Бразилия	12,9
34	Мальдивы	5	84	Иран	13,1
35	Польша	5	85	Йемен	13,8
36	Эфиопия	5,2	86	Албания	13,9
37	Румыния	5,2	87	Сербия	14,1
38	Россия	5,2	88	Иордания	14,9
39	Австрия	5,5	89	Сирия	15,2
40	Панама	5,6	90	Тунис	15,2
41	Австралия	5,7	91	Черногория	16
42	Люксембург	5,7	92	Оман	16
43	Саудовская Аравия	5,7	93	Испания	17,4
44	Дания	5,8	94	Ливия	17,7
45	Болгария	6,3	95	Армения	18,2
46	Канада	6,4	96	Греция	21,4
47	Ирландия	6,4	97	Македония	22,9
48	Эстония	6,8	98	Босния и Герцеговина	25,6
49	Швеция	6,8	99	ЮАР	27,7
50	Словения	6,9	100	Палестина	27,9

По данным таблицы построить гистограмму частотного распределения уровня безработицы при числе интервалов $k = 8$.

Задача 1. 2.В таблице приведены данные о плотности населения по субъектам Российской Федерации. Три субъекта (Москва, Санкт- Петербург, Севастополь) в таблице отсутствуют.

№ п/ п	Субъект России	Плотность населения, чел/ км²	№ п/п	Субъект России	Плотность населения чел/ км²
1	Московская область	163,06	41	Свердловская обл.	22,27
2	Республика Ингушетия	127,73	42	Ленинградская обл.	21,14
3	Северная Осетия	88,3	43	Астраханская обл.	20,85
4	Чеченская Республика	87,57	44	Смоленская обл.	19,37
5	Республика Крым	72,59	45	Пермский край	16,46
6	Краснодарский край	72,25	46	Оренбургская обл.	16,18
7	Кабардино-Балкарская Республика	69,03	47	Тверская область	15,62
8	Чувашская Республика	67,5	48	Новосибирская обл.	15,45
9	Калининградская обл.	64,02	49	Алтайский край	14,2
10	Самарская область	59,96	50	Омская область	14,02
11	Республика Дагестан	59,48	51	Курганская обл.	12,17
12	Тульская область	58,93	52	Псковская обл.	11,75
13	Республика Адыгея	57,64	53	Приморский край	11,74
14	Белгородская область	57,04	54	Новгородская обл.	11,35
15	Республика Татарстан	56,82	55	Костромская обл.	10,87
16	Ивановская область	48,38	56	Кировская область	10,84
17	Владимирская область	48,33	57	Тюменская обл.	8,93
18	Липецкая область	48,17	58	Вологодская обл.	8,24
19	Воронежская область	44,65	59	Сахалинская обл.	5,61
20	Нижегородская область	42,68	60	Мурманская обл.	5,29
21	Ставропольский край	42,32	61	Еврейская авт. обл.	4,64
22	Ростовская область	42,02	62	Республика Калмыкия	3,75
23	Челябинская область	39,49	63	Республика Карелия	3,5
24	Курская область	37,25	64	Томская область	3,42
25	Удмуртская Республика	36,07	65	Иркутская область	3,12
26	Брянская область	35,37	66	Ханты-Мансийский а. о.	3,02
27	Ярославская область	35,16	67	Республика Бурятия	2,79
28	Ульяновская область	33,96	68	Республика Алтай	2,3
29	Калужская область	33,91	69	Амурская область	2,24
30	Карачаево-Черкесская Республика	32,87	70	Республика Коми	2,07
31	Пензенская область	31,28	71	Республика Тыва	1,86
32	Орловская область	31,08	72	Хабаровский край	1,7
33	Республика Мордовия	30,98	73	Красноярский край	1,21
34	Тамбовская область	30,83	74	Ямало-Ненецкий а. о.	0,7
35	Республика Марий Эл	29,42	75	Камчатский край	0,68
36	Рязанская область	28,68	76	Магаданская обл.	0,32
37	Республика Башкортостан	28,48	77	Республика Саха	0,31

38	Кемеровская область	28,47	78	Ненецкий а. о.	0,25
39	Саратовская область	24,63	79	Чукотский а. о.	0,07
40	Волгоградская область	22,66			

По данной выборке построить гистограмму частотного распределения плотности населения при числе интервалов $k = 7$.

Задача 1. 3. В таблице приведены данные о среднедушевых денежных доходах населения по итогам 2017 года. Доходы Чукотского автономного округа (139195 руб.) не включены в таблицу.

Реальные денежные доходы населения

№	Субъект РФ	доходы населения (в тысячах руб.)	№	Субъект РФ	доходы населения (в тыс. руб.)
1	Республика Крым	28,586	43	Башкортостан	37,36
2	Республика Алтай	27,329	44	Краснодарский край	45,499
3	Республика Адыгея	35,152	45	Забайкальский край	33,384
4	Кабардино-Балкарская Респуб.	30,367	46	Воронежская обл.	41,161
5	Амурская область	44,048	47	Липецкая область	33,903
6	Ленинградская обл.	36,304	48	Респ. Северная Осетия	33,576
7	Республика Дагестан	49,92	49	Кировская область	29,848
8	Оренбургская обл.	29,95	50	Тульская область	34,709
9	Курская область	38,184	51	Удмуртская Респуб.	30,897
10	Псковская область	32,307	52	Респ. Марий Эл	24,151
11	Респ. Ингушетия	20,03	53	Республика Саха	65,252
12	Ивановская область	33,584	54	Камчатский край	59,698
13	Брянская область	33,904	55	Хабаровский кр.	54,954
14	Владимирская обл.	32,701	56	Красноярский край	41,939
15	Калининградская обл.	39,253	57	Чувашская Респ.	24,147
16	Волгоградская обл.	30,733	58	Новгородская обл.	35,088
17	Костромская обл.	30,755	59	г. Севастополь	35,763
18	Белгородская обл.	42,307	60	Новосибирская обл.	35,889
19	Кемеровская обл.	28,573	61	г. Санкт-Петербург	66,151
20	Смоленская область	34,187	62	Иркутская область	31,682
21	Мурманская обл.	67,251	63	Омская область	32,725
22	Республика Карелия	37,968	64	Самарская область	36,925
23	г. Москва	96,43	65	Свердловская обл.	46,201
24	Ставропольский кр.	32,101	66	Рязанская область	36,177

25	Респуб. Мордовия	24,069	67	Сахалинская область	68,306
26	Приморский край	45,087	68	Респ. Татарстан	42,96
27	Орловская область	33,162	69	Ханты-Мансийский авт. окр.	60,866
28	Тверская область	35,683	70	Челябинская область	30,866
29	Ненецкий авт. округ	97,238	71	Чеченская Респуб.	31,233
30	Карачаево-Черкесская Респ.	26,386	72	Нижегородская обл.	40,857
31	Республика Калмыкия	17,382	73	Ямало-Ненецкий авт. окр.	99,163
32	Саратовская область	28,278	74	Калужская область	38,553
33	Алтайский край	35,238	75	Тамбовская область	37,112
34	Московская область	59,932	76	Астраханская обл.	32,064
35	Курганская область	28,086	77	Республика Тыва	22,511
36	Архангельская обл.	42,204	78	Республика Бурятия	33,813
37	Ростовская область	39,859	79	Ярославская область	34,511
38	Пензенская область	29,193	80	Магаданская область	67,194
39	Ульяновская область	31,346	81	Республика Коми	42,578
40	Тюменская область	42,159	82	Еврейская авт. обл.	32,899
41	Республика Хакасия	25,81	83	Томская область	32,996
42	Пермский край	40,503	84	Вологодская область	32,308

По данной выборке построить гистограмму частотного распределения денежных доходов населения.

Задача 1.4. Импорт высокотехнологичных товаров (% от импорта товаров) 2016 г.

Импорт информационных и коммуникационных технологий товаров включает импорт телекоммуникаций, аудио и видео, компьютеров и компьютерной техники; электронных компонентов; и другие информационно-коммуникационные товары. Программное обеспечение исключается.

№	Страна	% ВВП	№	Страна	% ВВП
1	Китай	23,8	22	Австрия	7,4
2	Япония	16,2	23	Финляндия	7,3
3	Южная Корея	15,7	24	Румыния	7,3
4	Словакия	14,4	25	Канада	7,1
5	Чехия	14,2	26	Норвегия	6,9
6	США	14,1	27	Франция	6,7
7	Голландия	13,7	28	Турция	6,7
8	Венгрия	12,4	29	Португалия	5,6

9	Эстония	11,5	30	Литва	5,5
10	Израиль	10,9	31	Болгария	5,4
11	Латвия	10,1	32	Казахстан	5,2
12	Австралия	10	33	Казахстан	5,2
13	Швеция	9,8	34	Италия	5,1
14	Индия	9,3	35	Испания	5,1
15	Польша	9,2	36	Греция	4,7
16	Россия	8,9	37	Грузия	4,7
17	Ирландия	8,6	38	Египет	4,1
18	Германия	8,4	39	Словения	4,1
19	Бразилия	8,4	40	Швейцария	3,7
20	Дания	7,7	41	Беларусь	3,3
21	Великобритания	7,6	42	Бельгия	3

По данной выборке построить гистограмму частотного распределения.

Задача 1.5. В таблице приводятся данные о вкладе туризма в экономику стран мира в 2017 г. (в % ВВП).

№	Страна	Общий вклад туризма % ВВП	№	Страна	Общий вклад туризма % ВВП
1	Грузия	31	24	Венгрия	8
2	Греция	19,7	25	Бразилия	7,9
3	Португалия	17,3	26	Чехия	7,8
4	Мексика	16	27	Дания	7,7
5	Испания	14,9	28	США	7,7
6	Австрия	14,8	29	Япония	6,8
7	Италия	13	30	Сербия	6,7
8	Словения	11,9	31	Канада	6,5
9	Турция	11,6	32	Словакия	6,3
10	Болгария	11,5	33	Беларусь	6,2
11	Китай	11	34	Израиль	6
12	Египет	11	35	Казахстан	6
13	Австралия	11	36	Румыния	5,3
14	Германия	10,7	37	Голландия	5,2
15	Великобритания	10,5	38	Россия	4,8
16	Аргентина	10,3	39	Литва	4,8
17	Швеция	9,5	40	Южная Ко-рея	4,7
18	Индия	9,4	41	Тайвань	4,3
19	Вьетнам	9,4	42	Польша	4,5
20	Швейцария	9,1	43	Люксембург	4,3

21	Норвегия	9	44	Киргизия	3,9
22	Франция	8,9	45	Молдавия	3,3
23	Финляндия	8,3	46	Узбекистан	2,8

По данной выборке построить гистограмму частотного распределения.

Задача 1.6. Затраты стран мира на фундаментальные и прикладные исследования и экспериментальные разработки (% ВВП) (2015 г.)

№	Страна	% ВВП	№	Страна	% ВВП
1	Израиль	4,3	21	Люксембург	1,3
2	Южная Корея	4,2	22	Португалия	1,3
3	Япония	3,3	23	Испания	1,2
4	Швеция	3,3	24	Словакия	1,2
5	Австрия Дания	3,1	25	Россия	1,1
6	Дания	3,0	26	Литва	1,0
7	Финляндия	2,9	27	Польша	1,0
8	Германия	2,9	28	Греция	1,0
9	США	2,8	29	Болгария	1,0
10	Бельгия	2,5	30	Сербия	0,9
11	Франция	2,2	31	Египет	0,7
12	Исландия	2,2	32	Индия	0,6
13	Китай	2,1	33	Латвия	0,6
14	Голландия	2,0	34	Украина	0,5
15	Чехия	1,9	35	Беларусь	0,5
16	Норвегия	1,9	36	Грузия	0,3
17	Великобритания	1,7	37	Казахстан	0,2
18	Эстония	1,5	38	Монголия	0,2
19	Венгрия	1,4	39	Киргизия	0,1
20	Италия	1,3			

По данной выборке построить гистограммы частотного распределения.

Тема 2. Оценивание числовых характеристик случайной величины. Выборочные числовые характеристики

Пусть случайная величина X исследуется на основании случайной выборки X_1, $\mathrm{X}_2,\ldots,\mathrm{X}_\mathrm{n}$, реализация которой $x_1, x_2,\ldots, x_n$. Пусть θ – некоторый параметр распределения случайной величины X, значение которого неизвестно. Например, предполагается, что X имеет показательный закон распределения $f(x) = \lambda e^{-\lambda x}, x \geq 0$, но значение параметра $\theta = \lambda$ неизвестно. Задача оценивания параметра θ состоит в нахождении его приближенного значения по результатам наблюдений $\mathrm{X}_1, \mathrm{X}_2,\ldots, \mathrm{X}_\mathrm{n}$.

Функцию от случайной выборки $\theta_n^* = \theta_n^*(\mathrm{X}_1, \mathrm{X}_2,\ldots, \mathrm{X}_\mathrm{n})$ называют *статистикой*. Очевидно, что для оценивания параметра θ следует рассматривать только те статистики θ_n^*, которые в определенном смысле близки к истинному значению параметра θ. Статистику θ_n^*, принимаемую в качестве приближенного значения θ, называют *оценкой* параметра θ. Например, для оценивания неизвестного математического ожидания $\mathrm{E(X)} = m$ используется оценка $\overline{\mathrm{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i$, называемая *выборочным средним.* Оценкой для дисперсии $\mathrm{D(X)} = \sigma^2$ является *выборочная дисперсия* $\mathrm{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. На реализации x_1, x_2,…, x_n случайной выборки оценки $\overline{\mathrm{X}}$ и S^2 принимают числовые значения $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ и $\mathrm{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Согласно теории, когда n велико, с большой вероятностью $\overline{x}$ и s^2 будут близки к m и σ^2 соответственно.

Основные числовые характеристики генеральной совокупности и соответствующие оценки.

1. Характеристики положения

Числовая характеристика X	Оценка числовой характеристики
Математическое ожидание $E(X) = \sum_i x_i p_i$ – для дискретной случайной величины. $E(X) = \int xf(x)dx$ – для непрерывной случайной величины.	*Выборочное среднее* $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Медиана непрерывной случайной величины – это такое ее значение M_e, для которого $P\{X \le M_e\} = P\{X > M_e\} = \frac{1}{2}$.	*Выборочная медиана* определяется как такое значение M_e^*, меньше и больше которого оказывается одинаковое число наблюдений. $M_e^* = \begin{cases} \frac{1}{2}\left(X_{(n/2)} + X_{((n+2)/2)}\right), & \text{если } n - \text{четное} \\ X_{((n+1)/2)}, & \text{если } n - \text{нечетное.} \end{cases}$ $X_{(k)}$ – *k-й* элемент вариационного ряда.
Мода M_o – значение случайной величины X, соответствующее локальному максимуму плотности $f(x)$ для непрерывной случайной величины или максимуму вероятности для дискретной случайной величины.	*Выборочная мода* – это наиболее часто встречающееся значение в выборке. Если данные сгруппированы и построено распределение частот, модой является значение, имеющее наибольшую частоту.

2. Основные характеристики разброса

Дисперсия $D(X) = E(X - EX)^2$, среднее квадратическое отклонение $\sigma(X) = \sqrt{D(X)}$	*Выборочная дисперсия* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ S – выборочное стандартное отклонение (выборочное среднее квадратическое отклонение).
p- квантиль функции распределения $F(x)$ –это значение x_p случайной величины X, задаваемое уравнением $x_p = \inf\{x: F(x) > p\}$.	*Выборочная p- квантиль* $\hat{x}_p$ определяется равенством $\hat{x}_p = \begin{cases} \frac{1}{2}\left(X_{(np)} + X_{(np+1)}\right), & \text{если } np - \text{целое} \\ X_{([np]+1)}, & \text{если } np - \text{не целое число.} \end{cases}$

3. Характеристики формы распределения

Мерой отклонения от симметрии плотности $f(x)$ являются 3-й момент $\mu_3 = \int (x - EX)^3 f(x)dx$ и коэффициент асимметрии $A = \mu_3/\sigma^3$	*Выборочный коэффициент асимметрии* $a_s = \sum_{i=1}^{n}(X_i - \overline{X})^3/nS^3$
Мерой отклонения плотности $f(x)$ от нормального распределения является эксцесс $E = \mu_4/\sigma^4 - 3$;	*Выборочный эксцесс* $E^* = \sum_{i=1}^{n}(X_i - \overline{X})^4/nS^4 - 3$ – мера отклонения эмпирического распределения от нормального.

Для нахождения выборочных характеристик могут быть использованы встроенные функции категории «статистические» из электронных таблиц Excel. Средство Excel ОПИСАТЕЛЬНАЯ СТАТИСТИКА позволяет вычислить важнейшие числовые характеристики выборки и представить их в виде таблицы.

Пример 1 Случайная величина X, характеризующая уровень воды в реке по отношению к номиналу, измерялась в течение 36 весенних паводков. Результаты измерений приведены в таблице:

№ измерения	Уровень (в см)	№ измерения	Уровень (в см)	№ измерения	Уровень (в см)	№ измерения	Уровень (в см)
1	47	10	164	19	93	28	115
2	151	11	158	20	121	29	171
3	52	12	243	21	118	30	205

4	163	13	190	22	110	31	61
5	77	14	85	23	173	32	174
6	156	15	139	24	243	33	148
7	205	16	179	25	254	34	217
8	181	17	257	26	307	35	149
9	311	18	143	27	99	36	187

a. Найдем оценки среднего значения, дисперсии и среднего квадратического отклонения случайной величины X по формулам:
выборочное среднее $\overline{x} = \frac{1}{36}\sum_{i=1}^{36} x_i \approx 162{,}4$;
выборочная дисперсия $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2 = \frac{1}{35}\sum_{i=1}^{36} x_i^2 - \frac{36}{35}\overline{x}^2 \approx 4295{,}2$;
выборочное среднее квадратическое отклонение $s \approx 65{,}5$.
b. Оценим эти же характеристики с помощью функций пакета Excel. Предположим, что элементы выборки $x_1, x_2, \ldots, x_{36}$ находятся в ячейках с адресами A1:A36. В ячейки B1, B2, B3, предназначенные для результатов $\overline{x},\ s^2, s$, нужно ввести формулы: =СРЗНАЧ(A1:A36), =ДИСП.В(A1:A36), = СТАНДОТКЛОН.В(A1:A36) соответственно.
c. Определим основные числовые характеристики выборки с помощью процедуры ДАННЫЕ – АНАЛИЗ ДАННЫХ – ОПИСАТЕЛЬНАЯ СТАТИСТИКА. В качестве ВХОДНОГО ИНТЕРВАЛА укажем адреса ячеек, содержащих выборку $x_1, x_2, \ldots, x_{36}$. Результат получим в виде таблицы

Среднее	162,4
Стандартная ошибка	10,9
Медиана	160,5
Мода	205,0
Стандартное отклонение	65,5
Дисперсия выборки	4295,2
Эксцесс	0,02
Асимметричность	0,4
Интервал	264,0
Минимум	47,0
Максимум	311,0
Сумма	5846,0
Счет	36,0

Здесь *стандартная ошибка* − это оценка параметра $\sigma(\overline{X})$, равная $S/\sqrt{n}$.

Оценивание генеральных числовых характеристик по выборке, представленной в виде группированного статистического ряда

Если в группированной выборке (1) все выборочные элементы из интервала $(d_{i-1}, d_i]$ положить равными величине $x_i^* = \frac{d_{i-1}+d_i}{2}$, то распределение выборки принимает вид:

x_1^*	x_2^*	. . .	x_k^*
ν_1/n	ν_2/n	. . .	ν_k/n

. (2)

С помощью группированного статистического ряда (2) приближенное вычисление выборочных моментов порядка l выполняется по формуле $\overline{a_l} = \frac{1}{n}\sum_{i=1}^{k} \nu_i \, (x_i^*)^l$. Формулы, определяющие выборочное среднее и выборочную дисперсию, принимают вид:
$\overline{x} = \overline{a}_1 \approx \frac{1}{n}\sum_{i=1}^{k} \nu_i \, x_i^*,\;\; s^2 = \frac{n}{n-1}\left(\overline{a}_2 - \overline{a}_1^{\,2}\right) = \frac{1}{n-1}\sum_{i=1}^{k} \;(x_i^*)^2\nu_i - \frac{n}{n-1}\overline{(x)}^2.$ (2. 1)
Усреднение по интервалам вносит ошибку, особенно заметную при малом числе интервалов. Для уменьшения ошибок, вносимых подобной группировкой, применяют «поправки Шеппарда».

Пример 2. Распределение средних температур июня в Стокгольме в течение 100 лет дано в виде интервального статистического ряда в табл. 5 (столбцы 2-4). Найти оценки математического ожидания и дисперсии случайной величины X, характеризующей среднюю июньскую температуру. (Данные взяты из книги «Математические методы статистики», Г. Крамер.)

В примере по условию дано: объем выборки $n = 100$, наименьший и наибольший из элементов выборки равны: $x_{min} = 12{,}07$ и $x_{max} = 16{,}94$, $k = 10$. Отрезок [12, 17], содержащий все выборочные значения, разбит на 10 интервалов $\Delta_i = (d_{i-1}, d_i]$ длины $h = 0.5$ с границами: $d_0 = 12,\; d_i = d_0 + ih,\; i = 1,2, \ldots 10$.

Вычислим значения выборочного среднего и выборочной дисперсии по формулам (2.1).

№ интервала Δ_i	d_{i-1}	d_i	Частота ν_i	$x_i^* = \dfrac{d_{i-1}+d_i}{2}$	$x_i^*\nu_i$	$(x_i^*)^2\nu_i$
1	12	12,5	10	12,25	122,5	1500,625
2	12,5	13	12	12,75	153	1950,75
3	13	13,5	9	13,25	119,25	1580,063
4	13,5	14	10	13,75	137,5	1890,625
5	14	14,5	19	14,25	270,75	3858,188
6	14,5	15	10	14,75	147,5	2175,625
7	15	15,5	9	15,25	137,25	2093,063
8	15,5	16	6	15,75	94,5	1488,375
9	16	16,5	7	16,25	113,75	1848,438

10	16,5	17	8	16,75	134	2244,500
			∑=100		∑= 1430	∑=20630,252

Выборочное среднее $\overline{x} = \frac{1}{100}\sum_{i=1}^{10} x_i^* \, \nu_i$ =14,30;

выборочная дисперсия $s^2 = \frac{1}{99}\sum_{i=1}^{10} \ (x_i^*)^2\nu_i - \frac{100}{99}\overline{(x)}^2 \approx 1{,}83;\ s \approx$1,35.

Задачи для самостоятельного решения

Задача 2.1. В таблице приводятся данные об ожидаемой продолжительности жизни (2017 г.)

№	Субъект РФ	Ожидаемая продолжительность жизни при рождении, лет	№	Субъект РФ	Ожидаемая продолжительность жизни при рождении, лет
1	Алтайский край	71	43	Пермский край	71
2	Амурская область	69	44	Приморский край	70
3	Архангельская обл.	72	45	Псковская область	70
4	Астраханская обл.	73	46	Республика Адыгея	73
5	Белгородская обл.	74	47	Республика Алтай	71
6	Брянская обл.	71	48	Башкортостан	72
7	Владимирская обл.	71	49	Республика Бурятия	71
9	Волгоградская обл.	74	50	Республика Дагестан	78
9	Вологодская обл.	71	51	Республика Ингушетия	82
10	Воронежская обл.	73	52	Республика Калмыкия	74
11	г. Москва	78	53	Республика Карелия	71
12	Санкт-Петербург	75	54	Республика Коми	71
13	г. Севастополь	73	55	Республика Крым	72
14	Еврейская авт. обл.	69	56	Республика Марий Эл	72
15	Забайкальский край	70	57	Республика Мордовия	73
16	Ивановская обл.	71	58	Республика Саха	72
17	Иркутская обл.	69	59	Республика Северная Осетия	76
18	Кабардино-Балкарская Респ.	76	60	Республика Татарстан	74

19	Калининградская обл.	73	61	Республика Тыва	66
20	Калужская обл.	72	62	Республика Хакасия	70
21	Камчатский край	70	63	Ростовская обл.	73
22	Карачаево-Черкесская Респ.	76	64	Рязанская область	73
23	Кемеровская обл.	69	65	Самарская обл.	72
24	Кировская обл.	73	66	Саратовская обл.	73
25	Костромская обл.	72	67	Сахалинская обл.	70
26	Краснодарский край	73	68	Свердловская область	71
27	Красноярский край	71	69	Смоленская область	71
28	Курганская обл.	71	70	Ставропольский край	74
29	Курская область	72	71	Тамбовская обл.	73
30	Ленинградская обл.	73	72	Тверская область	70
31	Липецкая обл.	72	73	Томская область	72
32	Магаданская область	69	74	Тульская область	71
33	Московская обл.	73	75	Тюменская обл.	72
34	Мурманская обл.	72	76	Удмуртская Республика	72
35	Ненецкий авт. окр.	72	77	Ульяновская область	72
36	Нижегородская обл.	72	78	Хабаровский край	70
37	Новгородская область	70	79	Ханты-Мансийский ао	74
38	Новосибирская обл.	72	80	Челябинская область	72
39	Омская область	71	81	Чеченская Республика	75
40	Оренбургская область	71	82	Чувашская Республика	73
41	Орловская обл.	72	83	Чукотский авт. округ	66
42	Пензенская обл.	73	84	Ямало-Ненецкий	74
			85	Ярославская обл.	72

Вычислить важнейшие числовые характеристики выборки с помощью процедуры ДАННЫЕ – АНАЛИЗ ДАННЫХ – ОПИСАТЕЛЬНАЯ СТАТИСТИКА.

Задача 2.2. По данным, приведенным в задаче 1.3, оценить важнейшие числовые характеристики случайной величины X, характеризующей денежные доходы населения.

Задача 2.3. Результаты исследования 70 коммерческих фирм по затратам на рекламу представлены в таблице:

Расходы на рекламу, млн. руб.	0−20	20−40	40−60	60−80	80-100	120 и более
Количество фирм	2	8	12	27	16	5

Оценить среднее и дисперсию случайной величины X, выражающей затраты фирм на рекламу в год.

Задача 2.4. В воде мелководного озера в течение года были измерены концентрации общего фосфора (в мкг/л):

46	41	153	98	140	95	208	88	65	108
60	42	179	320	176	118	191	108	62	91
90	66	189	274	170	95	62	108	45	58
90	83	202	134	166	82	117	62	91	37
80	45	111	83	120	108	91	241	90	66
163	110	117	91	180	104	91	134	92	83

По полученной выборке оценить важнейшие числовые характеристики случайной величины X, выражающей концентрацию общего фосфора в озере.

Тема 3. Первоначальная проверка выборки на соответствие нормальному закону.

При проведении статистического анализа важно знать, насколько близок закон распределения выборки к нормальному закону. Выборочные асимметрия и эксцесс характеризуют степень отличия эмпирического распределения от нормального. Коэффициент асимметрии и эксцесс нормального распределения равны нулю. Поэтому, достаточно малые значения соответствующих выборочных величин дают основание предполагать, что генеральная совокупность распределена по нормальному закону. Для первоначальной проверки выборки на соответствие нормальному закону можно применить *экспресс-метод* [6]: предположение о близости выборочного распределения к нормальному не отвергается при условии $a_s < 4s/\bar{x}$. Здесь a_s − *выборочный коэффициент асимметрии,* s − выборочное стандартное отклонение, $\bar{x}$ − выборочное среднее. Величина $s/\bar{x}$ называется *коэффициентом вариации*.

Контрольные задания

По каждой выборке, приведенной ниже в задачах 3.1-3.4, требуется:

a) построить гистограмму частот;

b) найти числовые характеристики выборки, используя средство Excel ОПИСАТЕЛЬНАЯ СТАТИСТИКА;

c) проверить выборку на соответствие нормальному закону с помощью *экспресс-метода*. Сделать заключение.

Задача 3.1. Для обследования качества воды хозяйственно-бытового назначения из источника взяты 40 проб. Анализ проб показал наличие свинца в воде. Значения концентрации свинца в пробах приведены в таблице.

0,05	0,04	0,10	0,07	0	0,04	0,02	0,11
0,03	0,01	0,08	0,06	0,02	0,03	0	0,09
0	0,06	0,05	0,12	0,03	0	0,09	0,06
0,07	0,09	0,1	0,05	0,05	0,06	0,7	0,12
0,05	0,07	0,11	0,03	0,05	0,07	0,08	0,05

Задача 3.2 Дана выборка высот 40 деревьев из чистого соснового древостоя:

20,5	23,5	20,6	19,5	22,5	22,5	18,5	24,5
23,0	21,0	22,0	19,8	19,5	21,8	19,5	20,5
23,5	21,0	21,5	21,5	22,7	21,6	25,5	17,4
25,5	21,0	21,9	23,5	18,5	24,5	21,4	23,0
21,5	22,0	21,5	22,5	21,4	22,1	18,7	20,9

Задача 3.3. Получены следующие данные о весе 64 взрослых землероек:

9,2	11,6	8,1	9,1	10,1	9,6	9,3	9,7
7,6	10,0	9,7	8,4	8,6	9,0	8,8	8,6
9,2	10,2	11,2	8,1	10,3	9,2	9,8	9,9
9,6	7,3	8,3	8,8	9,2	8,0	8,6	8,8
8,5	8,8	9,7	11,5	10,5	9,8	10,0	9,4
8,6	8,7	9,1	8,2	9,2	9,4	8,8	9,8
9,9	9,3	9,3	9,0	8,7	9,6	9,9	9,2
9,7	9,5	10,0	9,1	11,8.	9,9	9,6	8,9

Задача 3.4. В задача 2.1. приводятся данные об ожидаемой продолжительности жизни (2017 г.)

Задача 3.5. Средняя температура июня в Ярославле измерялась в течение 60 лет. Данные измерений приведены в таблице:

12,4	14,1	15,2	13,8	17,2	14,9
11,3	13,8	16	15	18	16

12	13,7	16,2	16	17,8	17,8
13,1	15	17,2	13,6	19,2	13,8
14,9	15,7	16,8	14,8	19,3	15,1
14	14,9	16,6	14,9	20,1	16
13,5	15,9	17	16	20	15,7
13	16	16,7	17	14	14,5
14,1	15,5	17,3	16,1	13,6	14,8
15	16	17	16	14,1	15,2

Задача 3.6. Проведено испытание нового сорта зерновой культуры на 56 участках одинаковой площади и получены следующие значения урожайности (ц/г):

46	46	44	55	39	58	49	51
48	47	46	49	40	50	44	47
46	46	43	42	47	48	42	43
49	49	47	43	50	49	48	49
47	47	46	48	52	44	49	46

Тема 4 . Интервальное оценивание параметров.

Пусть θ – неизвестный параметр распределения случайной величины X, θ_n^*- оценка этого параметра. Для обоснованного использования этой оценки на практике необходимо выяснить, насколько θ_n^* может отличаться от истинного значения параметра θ, т.е. насколько точна оценка. Представление о точности оценивания можно получить, указав границы интервала, в котором с заданной степенью достоверности лежит неизвестный параметр. Границы интервала определяются по выборке: $\underline{\theta} = \underline{\theta}(X_1, X_2, ..., X_n)$ и $\overline{\theta} = \overline{\theta}(X_1, X_2, ..., X_n)$, $\underline{\theta} < \overline{\theta}$ для всех $X_1, X_2, ..., X_n$.

Доверительным интервалом для оценки параметра θ называется такой интервал $(\underline{\theta}, \overline{\theta})$, который с заданной вероятностью γ содержит (накрывает) оцениваемый параметр θ, т.е. $P\{\underline{\theta} < \theta < \overline{\theta}\} = \gamma$. Число γ называется *доверительной вероятностью*. Значения γ определяются конкретными практическими задачами. Обычно берутся значения γ: 0.9, 0.95, 0.99 и т.п.

Доверительный интервал для неизвестного математического ожидания m нормального распределения.

Пусть $X \in N(m, \sigma)$ − нормально распределенная генеральная совокупность и $X_1, X_2, ..., X_n$ − случайная выборка из нее. В качестве оценки неизвестного математического ожидания m возьмем выборочное среднее $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Для построения доверительного интервала естественно выбирать наиболее вероятные значения отклонения $\overline{X}-m$. Поскольку плотность распределения случайной величины $\overline{X}-m$ симметрична относительно нуля, то это будут значения из симметричного интервала $-\varepsilon<\overline{X}-m<\varepsilon,\ \varepsilon>0$. Таким образом, задача построения доверительного интервала для m сводится к нахождению ε такого, что $P\{|\overline{X}-m|<\varepsilon\}=\gamma$. Число $\varepsilon>0$ характеризует точность оценивания параметра. Чем меньше ε, тем точнее оценка $\overline{\mathrm{X}}$. Определение ε возможно, если известен закон распределения оценки $\overline{\mathrm{X}}$.

Построим доверительный интервал с заданной вероятностью γ при известной и неизвестной дисперсии σ^2.

Случай 1: дисперсия $DX=\sigma^2$ известна.

Значение ε находим по формуле

$$\varepsilon=\frac{z^*\sigma}{\sqrt{n}},\text{ где} \tag{4.1}$$

величина z^* − это квантиль порядка $(1+\gamma)/2$ стандартного нормального распределения, т.е. $z^*=z_{(1+\gamma)/2}$.

Доверительный интервал для математического ожидания *m,* соответствующий заданной доверительной вероятности γ, имеет вид

$$(\overline{X}-\frac{z^*\sigma}{\sqrt{n}},\overline{X}+\frac{z^*\sigma}{\sqrt{n}}).$$

Квантиль $z_{(1+\gamma)/2}$ может быть найдена из таблицы значений функция стандартного нормального распределения.

В Excel квантиль $\mathrm{z_p}$ порядка p и значение ε можно получить с помощью следующих статистических функций:

$$z_p=\text{НОРМ.СТ.ОБР}(p),\ \ \varepsilon=\text{ДОВЕРИТ.НОРМ}\ (1-\gamma,\sigma,n).$$

Замечание. Исходя из соотношения (4.1) можно оценить минимальный объем выборки n, обеспечивающий заданную точность ε оценки с вероятностью γ:

$$n\geq\left(\frac{z^*\sigma}{\varepsilon}\right)^2.$$

Случай 2: дисперсия $DX=\sigma^2$ неизвестна.

Значение ε находим по формуле

$$\varepsilon=\frac{t^*s}{\sqrt{n}},\text{ где} \tag{4.2}$$

Здесь величина t^* является квантилью порядка $(1+\gamma)/2$ распределения Стьюдента с $(n-1)$ степенями свободы: $t^*=t_{(1+\gamma)/2,\ n-1}$.

Доверительный интервал принимает вид

$$\left(\overline{X}-\frac{t^*s}{\sqrt{n}},\overline{X}+\frac{t^*s}{\sqrt{n}}\right).$$

Квантиль $t_{(1+\gamma)/2,n-1}$ может быть получена из таблиц квантилей распределения Стьюдента.

В Excel: значения $t_{(1+\gamma)/2,\ n-1}=\text{СТЬЮДЕНТ.ОБР.2Х}(1-\gamma,n-1)$,

$\varepsilon=$ ДОВЕРИТ.СТЬЮДЕНТ $(1-\gamma,s,n)$.

В рассмотренных случаях число наблюдений может быть небольшим, поскольку при построении интервалов не используются приближенные формулы.

Замечание 1. При большом числе наблюдений ($n > 30$) распределение Стьюдента практически не отличается от стандартного нормального распределения. В этом случае в результате замены в формуле (4.1) неизвестного параметра σ его оценкой S получаем

$$\varepsilon = \frac{z^*S}{\sqrt{n}}. \quad (4.3)$$

Тогда для доверительного интервала будет выполняться приближенное равенство

$P\{\overline{X} - \frac{z^*S}{\sqrt{n}} < m < \overline{X} + \frac{z^*S}{\sqrt{n}}\} \approx \gamma$.

Замечание 2. Пусть распределение выборки $X_1, X_2, ..., X_n$ отлично от нормального закона. Тогда, в силу асимптотической нормальности выборочного среднего $\overline{X}$, можно построить доверительный интервал для m с доверительной вероятностью, приблизительно равной γ. В этом случае требуются выборки больших объемов.

Пример. Для обследования качества воды хозяйственно-бытового назначения из источника взяты 40 проб. Анализ проб показал наличие свинца в воде. Значения концентрации свинца в пробах приведены в таблице

a) Построить доверительный интервал для математического ожидания концентрации свинца в источнике с доверительной вероятностью 0,95.

b) Определить, сколько проб надо взять для анализа, чтобы доверительный интервал обеспечивал точность оценки $\varepsilon = 0{,}025$ при доверительной вероятности 0,95.

0,05	0,04	0,10	0,07	0	0,04	0,02	0,11
0,03	0,01	0,08	0,06	0,02	0,03	0	0,09
0	0,06	0,05	0,12	0,03	0	0,09	0,06
0,07	0,09	0,1	0,05	0,05	0,06	0,7	0,12
0,05	0,07	0,11	0,03	0,05	0,07	0,08	0,05

Решение. Случайная величина X − концентрации свинца в источнике.

$E(X) = m$ и $D(X) = \sigma^2$ – неизвестны.

a. По имеющимся выборочным значениям x_i вычислим выборочное среднее и выборочную дисперсию:

$$\overline{x} = \frac{1}{40}\sum_{i=1}^{40} x_i \approx 0{,}073; \quad s^2 = \frac{1}{39}\sum_{i=1}^{40} x_i^2 - \frac{40}{39}\overline{x}^2 \approx 0{,}011;$$

При данном объеме выборки ($n = 40$) для определения ε можно воспользоваться формулой $\varepsilon = \frac{z^*s}{\sqrt{n}}$. Из таблицы значений функции стандартного нормального распределения находим $z^* = z_{(1+\gamma)/2} = z_{0,975} \approx 1{,}96$. Тогда $\varepsilon = \frac{1{,}96 * 0{,}105}{\sqrt{40}} \approx 0{,}033$ и границы искомого интервала:

$$\overline{x} - \varepsilon \approx 0{,}073 - 0{,}033 = 0{,}040; \quad \overline{x} + \varepsilon \approx 0{,}073 + 0{,}033 = 0{,}106.$$

Таким образом, для данных 40 проб интервал (0.040, 0.106) содержит математическое ожидание m с вероятностью, приблизительно равной 0,95.

В Excel: $z_{0,975}$ = НОРМ.СТ.ОБР(0,975) $\approx 1{,}96$,

ε = ДОВЕРИТ.НОРМ (0,05; 0,107; 40) $\approx 0{,}033$.

b. Дано $\varepsilon = 0{,}025$, по формуле $n \geq \left(\frac{z^*s}{\varepsilon}\right)^2 = \frac{1{,}96^2 \cdot 0{,}011}{0{,}025^2} \approx 67{,}6$. Следовательно, число проб, обеспечивающих заданную точность $\varepsilon = 0{,}025$ при доверительной вероятности 0,95 должно быть не меньше 68.

Задачи для самостоятельного решения

Задача 4.1. Уровень воды в реке по отношению к номиналу измерялся в течение 36 весенних паводков. По результатам измерений вычислены выборочное среднее $\overline{x} = 162{,}4$ и выборочная дисперсия $s^2 = 4295{,}2$. При доверительной вероятности 0,9 построить доверительный интервал для среднего уровня воды в реке, используя для определения ε формулы (4.2) и (4.3). Сравнить результаты.

Задача 4.2. По данным о продолжительности жизни в России (задача 2.1.) требуется оценить среднюю продолжительность жизни и построить для нее доверительный интервал с доверительной вероятностью 0,95.

Задача 4.3. В крупном фермерском хозяйстве проведено выборочное обследование урожайности пшеницы на 1000 участках. Все участки имеют одинаковую площадь, равную одному гектару. Данные об урожайности приведены в таблице 9.

Урожайность пшеницы (ц/га)	23 - 25	25 - 27	27 - 29	29 – 31	31 - 33	33 - 35
Количество участков	40	100	200	400	200	60

a) Оценить среднюю урожайность m на всей посевной площади хозяйства и построить для m доверительный интервал при доверительной вероятности 0,95.

b) Найти вероятность того, что средняя урожайность пшеницы в хозяйстве отличается от выборочной средней урожайности не более чем на 10 кг.

Задача 4.4. Проделано 16 измерений одним прибором некоторой физической величины. По результатам измерений определены выборочное среднее $\overline{x} = -0{,}45$ и выборочная дисперсия $s^2 = 0{,}09$. Предполагая, что результаты измерений распределены по нормальному закону, оценить истинное значение измеряемой величины с доверительной вероятностью 0,99. Сколько надо сделать независимых измерений, чтобы построить доверительный интервал, обеспечивающий точность оценки $\varepsilon = 0{,}10$?

Задача 4.5. В задаче 1.3 приводятся данные о душевых денежных доходах населения (в тыс. руб.). При объеме выборки $n = 84$ вычислены выборочное среднее $\overline{x} = 39{,}54$ и выборочная дисперсия s^2 =244,39. Найти доверительный интервал для неизвестного среднего душевого дохода с доверительной вероятностью 0,95. Найти вероятность, с которой средний денежный доход будет отличаться от выборочной средней не более чем на 5 тыс. руб.

Задача 4.6. По измерениям средней июньской температуры в Стокгольме в течение 100 лет получены выборочное среднее $\overline{x} = 14{,}33$ и выборочная дисперсия s^2 =1,64. Предполагая, что средняя июньская температура имеет нормальное распределение, построить доверительный интервал для математического ожидания средней июньской температуры при доверительной вероятности 0,95.

Задача 4.7. Для определения среднего процента сырого белка m в зернах пшеницы было отобрано 625 зерен, обследование которых показало, что выборочное среднее $\overline{x} = 16{,}8$ и выборочная дисперсия $s^2 = 4$. Построить доверительный интервал для среднего процента сырого белка при доверительной вероятности 0,975. Чему равна вероятность того, что m отличается от $16{,}8$ по абсолютной величине меньше чем на 0,1?

Задача 4.8. Результаты замеров рейтинга определенной политической партии в 6 республиках и 8 областях Приволжского федерального округа и 17 областях Центрального округа приведены в таблице 8.

Центральный федеральный округ	Приволжский федеральный округ
15,1	15,6
18,9	13,3
20,7	15
21,1	17,6
20,2	16,3
17,6	14,9
21,5	16,8
19,7	12,8
16,9	15,4
17,6	19,7
21,3	20,9
20,4	17,6
18,2	14,3
19,5	14,8
17,3	
21,2	
18,9	

Оценить средние значения рейтингов в каждом из округов и построить для них доверительные интервалы при доверительной вероятности $\gamma = 0{,}95$.

Задача 4.9. Предполагается, что количество осадков, выпадающих в течение года в данной местности, распределено по нормальному закону. По результатам наблюдений (в см) в течение 12 лет найдены выборочное среднее $\overline{x} = 55$ и выборочное среднеквадратичное отклонение s=13. Построить доверительный интервал для среднего годового количества осадков m с доверительной вероятностью γ =0,95.

Задача 4.10. Составлен Рейтинг «100 лучших городов России» по версии Kommersant (2013 г.) Данные о покупательской способности жителей выбранных 30 городов приводятся в табл.

№	Город	Покупательская способность	№	Город	Покупательская способность
1	Калининград	0,7	16	Красноярск	0,4
2	Екатеринбург	0,8	17	Саратов	0,6

3	Белгород	0,6	18	Череповец	0,2
4	Краснодар	1	19	Калуга	0,7
5	Казань	0,9	20	Курск	0,6
6	Ярославль	0,5	21	Нижний Новгород	0,7
7	Химки	2,2	22	Ростов-на-Дону	0,6
8	Новосибирск	0,7	23	Петрозаводск	0,5
9	Домодедово	1,4	24	Липецк	0,7
10	Тюмень	1,7	25	Старый Оскол	0,5
11	Обнинск	0,6	26	Владимир	0,8
12	Пермь	0,5	27	Тула	0,6
13	Челябинск	0,5	28	Иркутск	0,5
14	Самара	0,8	29	Тверь	0,7
15	Альметьевск	0,5	30	Подольск	0,3

По приведенным данным оценить среднее значение покупательской способности m и построить для m доверительный интервал при доверительной вероятности 0,95.

Задача 4.11. Глубина моря в данном месте измеряется прибором, систематическая ошибка которого равна нулю, а случайные ошибки измерения распределены нормально с $\sigma = 24$ м. По 36 измерениям вычислено выборочное среднее значение глубины моря $\overline{x}$=2250 *м*.

a. Построить доверительный интервал для математического ожидания m измеряемой величины с доверительной вероятностью γ =0,95.

b. Сколько надо сделать независимых измерений, чтобы определить глубину моря с ошибкой не более 5 м при $\gamma = 0{,}95$?

Задача 4.12. С целью изучения рентабельности производства продукции растениеводства было отобрано случайным образом 70 предприятий. Результаты представлены в таблице:

Рентабельность (в %)	0 - 10	10 -20	20- 30	30 -40	40 -50	50 -60	60 -70	70 -80	80 -90
Число предприятий	5	8	12	13	10	9	6	4	3

a. Найти границы, в которых с вероятностью 0,9 заключена средняя рентабельность производства.
b. Найти вероятность, с которой средняя рентабельность будет отличаться от выборочной средней не более чем на 5 %.

Ответы к задачам

4.1. (144.43; 180.37), (143,94; 180,85).
4.3. а) $\overline{x} = 29{,}58$; b) $\mathrm{P}\{|m - \overline{x}| < 0{,}1\} \approx 0{,}83$.
4.4. $-0{,}68 < m < -0{,}22$; $n \geq 79$.
4.6. (13,99; 14,62).
4.7. а) $16{,}59 < m < 17{,}01$; b)$P\{|16{,}8 - m|<0{,}1\} \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{s}\right) \approx 0{,}789$.
4.9. $46{,}74 < m < 63{,}26$.
4.10. $0{,}58 < m < 0{,}88$.
4.11. а) $(2242{,}16;\ 2257{,}84)$; $b)$ $n \geq 89$.
4.12. а) (35.68, 44.03); b) $\gamma \approx 0{,}951$.

Литература

1. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. – СПб.: Изд-во «Лань», 2011.
2. Кремер Н.Ш. Теория вероятностей и математическая статистика. Изд. ЮНИТИ-ДАНА, 2010.
3. Ермаков М.С., Сизова А.Ф., Товстик Т.М. Элементы математической статистики. – СПб.: Изд-во С.-Петерб. ун-та, 2001.
4. Алексеева Н.П., Товстик Т.М. Практикум по математической статистике.– СПб.: Изд-во С.-Петерб. ун-та, 2002.

Дополнительная литература

5. Самыловский А.И. Математические модели и методы для социологов. Книга 2. Математическая статистика. – М.: Изд-во «КДУ», 2009.
6. Третьяков В.Ю., Кулеш В.П. Автоматическая обработка экологической информации. – СПб.: Изд-во С.-Петерб. ун-та, 2005.

Содержание