ЦЕНТРОСОЮЗ РОССИЙСКОЙ ФЕДЕРАЦИИ

СИБИРСКИЙ УНИВЕРСИТЕТ
ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ

МАТЕМАТИКА
ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания и задания контрольной работы для студентов заочной формы обучения экономических и коммерческих специальностей

Новосибирск 2004

Кафедра высшей математики

Методические указания и задания контрольной работы: Математика. Теория вероятностей и математическая статистика / Сост.: доценты Г. И. Шегурова, Т. П. Руднева. – Новосибирск: СибУПК, 2004. – 72 с.

Рецензент доцент Баланчук Т. Т.

Утверждено и рекомендовано к изданию кафедрой высшей математики, протокол № 2 от 10 октября 2003 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Предлагаемая методическая разработка содержит задания контрольной работы по одному из разделов математики: «Теория вероятностей и математическая статистика». Контрольная работа включает восемь типовых задач, из которых шесть – по теории вероятностей и две – по математической статистике. В работе кратко изложен теоретический материал по основным вопросам программы, необходимый для решения всех задач, а также приведены примеры, которые являются образцами решения этих задач.

Основные вопросы программы

Теория вероятностей

1. Основные понятия теории вероятностей: испытание, событие. Несовместимые, единственно возможные, равновозможные события.
2. Классическое определение вероятности.
3. Статистическое определение вероятности.
4. Численные значения вероятностей.
5. Теоремы сложения вероятностей для несовместимых и совместимых событий.
6. Теоремы умножения вероятностей для зависимых и независимых событий. Условная вероятность.
7. Полная группа событий, свойство их вероятностей. Противоположные события.
8. Вероятность появления только одного и хотя бы одного события из двух или трех зависимых или независимых событий.
9. Формула полной вероятности. Формула Байеса.
10. Повторение независимых испытаний. Формула Бернулли.
11. Локальная и интегральная формулы Лапласа.
12. Наивероятнейшее число наступлений события и соответствующая ему вероятность.
13. Дискретная случайная величина, закон ее распределения.

14. Числовые характеристики дискретной случайной величины: математическое ожидание, дисперсия, среднее квадратическое отклонение, их смысл и свойства.

15. Непрерывные случайные величины. Дифференциальная и интегральная функции распределения вероятностей непрерывной случайной величины, их вероятностный смысл, свойства и графическое изображение.

16. Математическое ожидание и дисперсия непрерывной случайной величины.

17. Нормальный закон распределения непрерывной случайной величины, его особенности.

18. Вероятность попадания нормально распределенной случайной величины в заданный промежуток.

19. Вероятность заданного отклонения нормально распределенной случайной величины.

20. Понятие о законе больших чисел.

Математическая статистика

1. Генеральная совокупность и выборка. Статистическое распределение выборки, его графическое представление.

2. Генеральная и выборочная средняя, генеральная и выборочная дисперсии.

3. Оценки характеристик генеральной совокупности по выборочным данным. Точечные и интервальные оценки. Требования к точечным оценкам: несмещенность, состоятельность и эффективность.

4. Доверительный интервал для оценки математического ожидания нормально распределенной случайной величины.

5. Статистическая проверка гипотезы о нормальном распределении случайной величины. Критерий согласия Пирсона.

6. Функциональная, статистическая и корреляционная зависимости.

7. Основные задачи корреляционного анализа.

8. Линейная корреляция. Нахождение параметров прямой регрессии по методу наименьших квадратов.

9. Оценка тесноты линейной связи по коэффициенту линейной корреляции.

10. Понятие о нелинейной корреляции. Корреляционное отношение.

ЛИТЕРАТУРА

1. Теория вероятностей и математическая статистика / Под ред. Н. Ш. Кремера. – М.: ЮНИТИ, 2001.
2. Гмурман В. Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1999.
3. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1997.
4. Колемаев В. А. Староверов О. В., Турундаевский Б. В. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1997.
5. Колемаев В. А. Калинина В. Н. Теория вероятностей и математическая статистика. – М.: ИНФРА-М., 1997.
6. Кудрявцев В. А. Демидович Б. П. Краткий курс высшей математики. – М.: Наука, 1986.
7. Карасев А. И., Аксютина З. М., Савельева Т. И. Курс высшей математики для экономических вузов. – М.: Высшая школа, 1986.
8. Венецкий И. Г., Кильдишев Г. С. Основы теории вероятностей и математической статистики. – М.: Статистика, 1988.
9. Теория вероятностей и математическая статистика: Справочный материал и методические указания для самостоятельной работы студентов. – Новосибирск: СибУПК, 1997.
10. Теория вероятностей: Задания для практических занятий. – Новосибирск: СибУПК, 2002.

ПРАВИЛА ОФОРМЛЕНИЯ КОНТРОЛЬНОЙ РАБОТЫ

При выполнении контрольной работы нужно придерживаться перечисленных ниже правил.

1. Каждую контрольную работу следует выполнять в отдельной тетради чернилами любого цвета, кроме красного, оставляя поля для замечаний рецензента. Распечатки на компьютере не принимаются.
2. На обложке тетради должны быть чётко написаны фамилия студента, его учебный шифр, название дисциплины и контрольной работы.
3. В работу должны быть включены все задачи, указанные в задании, строго по своему варианту. Контрольные работы, содержащие не все задачи или задачи не своего варианта, не рассматриваются.
4. Решения задач следует располагать в порядке, указанном в заданиях, сохраняя номера задач.
5. Перед решением каждой задачи надо выписать полностью ее условие. Если задача имеет общую формулировку для всех вариантов, то при переписывании её условия следует заменить общие данные конкретными данными из своего варианта.
6. Решения задач необходимо излагать подробно и аккуратно, объясняя и мотивируя все действия по ходу решения и делая необходимые чертежи.
7. В конце работы нужно указать использованную литературу, дату выполнения и расписаться.
8. После получения прорецензированной работы как допущенной, так и недопущенной к собеседованию, студент должен исправить все отмеченные рецензентом ошибки и недочеты. Для этого рекомендуется оставлять в конце тетради несколько чистых листов. Если работа отправлена на доработку, то эту доработку нужно выполнить в той же тетради в короткий срок и сдать работу на повторную проверку.

Студент, правильно выполнивший все задания контрольной работы, допускается к устному собеседованию, по результатам которого выставляется зачет по контрольной работе. Собеседование проводится во время плановых субботних консультаций в межсессионный период или во время сессии. Без зачтенной контрольной работы студент к экзамену не допускается.

Студенту, не выполнившему контрольную работу до начала экзаменационной сессии, может быть предложена аудиторная контрольная работа.

ПРАВИЛО ВЫБОРА ВАРИАНТА

Вариант контрольной работы определяется по таблице в зависимости от двух последних цифр шифра личного дела студента. В колонке таблицы по вертикали расположены цифры от 0 до 9, каждая из которых – предпоследняя цифра шифра. В верхней строке по горизонтали размещены цифры от 0 до 9, каждая из которых – последняя цифра шифра.

Пересечение вертикальной и горизонтальной линий определяет номера заданий контрольной работы. Например, по последним двум цифрам шифра «18» находим вариант контрольной работы на пересечении строк с цифрой 1 и столбца с цифрой 8. Это номера: 5, 19, 25, 39, 46, 57, 65, 72.

Будьте внимательны при выборе варианта. Работа, выполненная не по своему варианту, возвращается без проверки.

ТАБЛИЦА ВЫБОРА ВАРИАНТА КОНТРОЛЬНОЙ РАБОТЫ

Предпоследняя цифра		Последняя цифра шифра									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	0	2	1	3	6	4	5	10	8	7	9
		15	14	13	12	11	16	20	19	18	17
		25	27	28	30	29	21	22	23	24	26
		40	32	33	34	35	36	37	31	38	39
		48	47	50	49	44	45	46	41	43	42
		51	55	60	52	59	57	53	58	56	54
		67	69	61	63	66	62	64	70	68	65
		72	78	77	75	73	79	71	74	80	76
	1	4	3	2	1	8	9	10	6	5	7
		14	15	13	12	17	16	11	20	19	18
		26	28	29	30	21	22	23	24	25	27
		32	31	34	35	36	37	38	33	39	40
		47	48	49	50	42	43	44	45	46	41
		53	60	59	52	58	56	55	54	57	51
		69	61	63	66	63	64	70	68	65	67
		77	75	73	79	71	74	80	76	72	78

Предпоследняя цифра шифра		Последняя цифра шифра									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	2	1	2	3	10	4	5	6	7	8	9
		11	12	13	20	14	15	16	17	18	19
		30	21	28	29	27	26	24	25	23	22
		39	40	31	38	32	33	34	35	36	37
		44	41	46	45	49	47	50	48	43	42
		60	57	56	54	51	52	53	55	58	59
		62	63	65	69	68	66	67	61	64	70
		74	79	80	77	72	71	78	76	75	73
	3	10	9	8	7	6	5	4	3	2	1
		11	12	13	14	15	16	17	18	19	20
		21	22	23	24	25	26	27	28	29	30
		40	39	38	37	36	35	34	33	32	31
		43	44	45	46	47	48	49	50	41	42
		54	56	55	60	57	53	51	58	59	52
		65	69	68	66	67	61	64	70	62	63
		80	77	72	71	78	76	75	73	74	79
	4	2	1	6	7	3	8	4	5	10	9
		12	13	14	11	19	16	15	17	20	18
		28	30	22	23	24	25	26	27	29	21
		40	38	39	36	31	37	32	33	34	35
		45	46	42	41	43	44	50	47	48	49
		51	54	58	53	55	59	52	57	56	60
		68	66	67	64	70	62	63	65	69	61
		77	72	71	78	76	75	73	74	79	80
	5	9	10	7	6	5	4	3	1	8	2
		19	18	20	17	16	15	14	12	13	11
		30	29	28	27	26	25	24	23	21	22
		39	40	37	38	31	32	33	34	35	36
		42	41	43	44	45	46	47	48	49	50
		54	56	55	59	58	57	60	52	51	53
		66	67	64	70	62	63	65	69	61	68
		71	78	76	75	73	74	79	80	77	72

Предпоследняя цифра шифра		Последняя цифра шифра									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	6	3	2	1	4	5	7	9	10	6	8
		13	17	15	16	14	12	18	11	20	19
		27	29	30	21	22	24	25	26	28	23
		31	33	32	34	35	36	39	37	40	38
		46	45	47	48	50	49	41	42	43	44
		51	56	57	54	55	53	60	52	58	59
		65	64	68	62	63	70	69	67	66	61
		75	76	80	78	77	72	74	79	71	73
	7	8	7	10	5	4	6	2	1	9	3
		18	19	17	20	12	14	13	15	11	16
		22	23	24	25	21	27	28	29	30	26
		38	31	32	33	34	35	36	37	39	40
		42	46	45	49	48	47	50	41	43	44
		58	56	57	60	52	59	53	55	51	54
		68	62	63	70	69	67	66	61	65	64
		76	80	78	77	72	74	79	71	73	75
	8	7	6	9	10	3	2	1	8	5	4
		17	18	19	16	20	11	12	14	13	15
		23	24	25	26	27	28	29	30	22	21
		37	38	31	32	33	34	35	36	40	39
		49	50	48	47	41	42	43	44	45	46
		54	55	58	56	60	52	59	53	51	57
		63	70	69	67	66	61	65	64	68	62
		80	78	77	72	74	79	71	73	75	76
	9	6	10	8	9	7	1	2	3	4	5
		16	15	14	13	12	20	18	19	17	11
		24	26	27	28	29	30	21	22	23	25
		36	37	40	31	32	33	34	35	39	38
		49	50	41	42	43	44	45	46	47	48
		54	60	57	55	53	58	52	59	56	51
		69	67	66	61	65	64	68	62	63	70
		78	77	72	74	79	71	73	75	76	80

ЗАДАНИЯ КОНТРОЛЬНОЙ РАБОТЫ

Задачи 1–10

Для сигнализации на складе установлены три независимо работающих устройства. Вероятность того, что при необходимости первое устройство сработает, составляет p_1, для второго и третьего устройства эти вероятности равны соответственно p_2 и p_3. Найти вероятность того, что в случае необходимости сработают:

а) все устройства;

б) только одно устройство;

в) хотя бы одно устройство.

1. $p_1 = 70\%, \quad p_2 = 85\%, \quad p_3 = 90\%$;

2. $p_1 = 75\%, \quad p_2 = 90\% \quad p_3 = 80\%$;

3. $p_1 = 95\%, \quad p_2 = 90\%, \quad p_3 = 75\%$;

4. $p_1 = 98\%, \quad p_2 = 85\%, \quad p_3 = 80\%$;

5. $p_1 = 75\%, \quad p_2 = 80\%, \quad p_3 = 95\%$;

6. $p_1 = 85\%, \quad p_2 = 95\%, \quad p_3 = 80\%$;

7. $p_1 = 90\%, \quad p_2 = 85\%, \quad p_3 = 95\%$;

8. $p_1 = 95\%, \quad p_2 = 75\%, \quad p_3 = 70\%$;

9. $p_1 = 80\%, \quad p_2 = 85\%, \quad p_3 = 90\%$;

10. $p_1 = 70\%, \quad p_2 = 90\%, \quad p_3 = 98\%$.

Задачи 11–20

В партии, состоящей из $\boldsymbol{n}$ одинаково упакованных изделий, смешаны изделия двух сортов, причем $\boldsymbol{k}$ из этих изделий – первого сорта, а остальные изделия – второго сорта. Найти вероятность того, что взятые наугад два изделия окажутся:

а) одного сорта;

б) разных сортов.

11. $n = 20$, $k = 15$;

12. $n = 25$, $k = 10$;

13. $n = 30$, $k = 20$;

14. $n = 55$, $k = 23$;

15. $n = 40$, $k = 25$;

16. $n = 45$, $k = 15$;

17. $n = 50$, $k = 30$;

18. $n = 55$, $k = 35$;

19. $n = 60$, $k = 40$;

20. $n = 70$, $k = 45$.

Задачи 21–30

21. Вероятность того, что новый товар будет пользоваться спросом на рынке, если конкурент не выпустит в продажу аналогичный продукт, равна 0,75 , а при наличии конкурирующего товара она равна 0,35. Вероятность выпуска конкурентом товара равна 0,45.

а) Найти вероятность того, что товар будет пользоваться спросом.

б) Товар пользуется спросом на рынке. Какова вероятность, что это произошло в условиях конкуренции?

22. Курс доллара повышается в течение квартала с вероятностью 0,9 и понижается с вероятностью 0,1. При повышении курса доллара фирма рассчитывает получить прибыль с вероятностью 0,85; при понижении – с вероятностью 0,5.

а) Найти вероятность того, что фирма получит прибыль.

б) Фирма в течение квартала получила прибыль. Какова вероятность, что это произошло при повышении курса доллара?

23. На строительство объекта поступают железобетонные плиты от четырех цементных заводов в количестве 50, 10, 40 и 30 штук соответственно. Каждый из заводов допускает при изготовлении плит брак (несоответствие ГОСТу), составляющий соответственно 1%, 5%, 2% и 3%.

а) Какова вероятность того, что наугад взятая плита будет удовлетворять требованиям ГОСТа?

б) Наугад взятая плита удовлетворяет требованиям ГОСТа. Какова вероятность, что она произведена на четвертом заводе?

24. В цехе трудятся три мастера и шесть их учеников. Мастер допускает брак при изготовлении изделия с вероятностью 0,05; а ученик – с вероятностью 0,15.

а) Какова вероятность, что взятое наугад изделие будет бракованным?

б) Поступившее из цеха изделие оказалось бракованным. Какова вероятность, что его изготовил мастер?

25. В данный район изделия поставляются двумя фирмами их объем находится в соотношении 5:8. Среди продукции первой фирмы стандартные изделия составляют 90%, у второй фирмы этот показатель 85%.

а) Какова вероятность, что взятое наугад изделие оказалось стандартным?

б) Взятое наугад изделие оказалось стандартным. Найти вероятность того, что оно изготовлено первой фирмой.

26. В магазине имеются телевизоры с импортными и отечественными трубками в соотношении 2:9. Вероятность выхода из строя в течение гарантийного срока телевизора с импортной трубкой равна 0,005; с отечественной трубкой она равна 0,01.

а) Найти вероятность того, что купленный в магазине телевизор выдержит гарантийный срок.

б) Купленный телевизор выдержал гарантийный срок. Какова вероятность, что он с отечественной трубкой?

27. По результатам проверки контрольных работ оказалось, что в первой группе получили положительную оценку 20 студентов из 30, а во второй группе 15 из 25.

а) Какова вероятность, того что взятая наугад работа из наугад выбранной группы оценена положительно?

б) Найти вероятность того, что наугад выбранная работа, имеющая положительную оценку, написана студентом первой группы.

28. Страховая компания разделяет застрахованных по классам риска: I класс – малый риск, II класс – средний риск, III класс – большой риск. Среди клиентов компании 50 % – клиенты первого класса риска, 30% – второго и 20% – третьего. Вероятность необходимости выплачивать страховое вознаграждение для первого класса риска равна 0,01, второго 0,03, третьего 0,08.

а) Какова вероятность того, что застрахованный получит денежное вознаграждение за период страхования?

б) Найти вероятность того, что получивший денежное вознаграждение застрахованный относится к группе малого риска.

29. Вся продукция цеха проверяется двумя контролерами, причем первый контролер проверяет 55% изделий, а второй – остальные. Вероятность того, что первый контролер пропустит нестандартное изделие, равна 0,01, а для второго контролера эта вероятность 0,02.

а) Какова вероятность, что взятое наугад изделие, маркированное как стандартное, оказалось нестандартным?

б) Взятое наугад изделие, маркированное как стандартное, оказалось нестандартным. Найти вероятность того, что это изделие проверялось вторым контролером.

30. Вероятность изготовления изделия с браком на данном предприятии равна 0,04. Перед выпуском изделие подвергается упрощенной проверке, которая в случае бездефектного изделия пропускает его с вероятностью 0,98, а в случае изделия с дефектом – с вероятностью 0,05.

а) Определить, какая часть изготовленных изделий выходит с предприятия, после упрощенной проверки.

б) Какова вероятность того, что изделие, выдержавшее упрощенную проверку, оказалось дефектным?

Задачи 31–40

Вероятность того, что в результате проверки изделию будет присвоен знак «изделие высшего качества» равна $\boldsymbol{p}$.

1) На контроль поступило $\boldsymbol{n}$ изделий. Какова вероятность того, что знак высшего качества будет присвоен:

а) ровно $\boldsymbol{m}$ изделиям;

б) более чем $\boldsymbol{k}$ изделиям;

в) хотя бы одному изделию;

г) указать наивероятнейшее количество изделий, получивших знак высшего качества, и найти соответствующую ему вероятность.

2). При тех же условиях найти вероятность того, что в партии из $\boldsymbol{N}$ изделий знак высшего качества получает:

а) ровно половина изделий;

б) не менее чем $\boldsymbol{k_1}$, но не более, чем $\boldsymbol{k_2}$ изделий.

31. n=8; p=0,4; m=5; k=6; N=20; k_1= 5; k_2=10.

32. n=7; p=0,3; m=4; k=5; N=24; k_1= 5; k_2=15.

33. n=6; p=0,2; m=3; k=4; N=28; k_1= 4; k_2=14.

34. n=5; p=0,3; m=2; k=3; N=30; k_1= 8; k_2=20.

35. n=4; p=0,6; m=1; k=2; N=32; k_1=10; k_2=25.

36. n=9; p=0,2; m=6; k=7; N=34; k_1= 5; k_2=20.

37. n=7; p=0,5; m=3; k=4; N=36; k_1=15; k_2=30.

38. n=6; p=0,4; m=1; k=3; N=38; k_1=12; k_2=30.

39. n=8; p=0,6; m=4; k=5; N=40; k_1=20; k_2=30.

40. n=5; p=0,5; m=3; k=2; N=26; k_1=10; k_2=20.

Задачи 41–50

В лотерее на каждые 100 билетов приходится $\boldsymbol{m_1}$ билетов с выигрышем $\boldsymbol{a_1}$ тыс. рублей, $\boldsymbol{m_2}$ билетов с выигрышем $\boldsymbol{a_2}$ тыс. рублей, $\boldsymbol{m_3}$ билетов с выигрышем $\boldsymbol{a_3}$ тыс. рублей и т.д. Остальные билеты из сотни не выигрывают.

Составить закон распределения величины выигрыша для владельца одного билета и найти его основные характеристики: математическое ожидание, дисперсию и среднее квадратическое отклонение. Пояснить смысл указанных характеристик.

41. a_1=20; a_2=10; a_3= 5; a_4= 3; a_5= 1;
m_1=1; m_2=2; m_3=8; m_4=10; m_5=15.

42. a_1=18; a_2=15; a_3=10; a_4=35;
m_1=2; m_2=3; m_3=5; m_4=20.

43. a_1=15; a_2=12; a_3= 8; a_4= 4;
m_1=3; m_2=10; m_3=15; m_4=20.

44. a_1=16; a_2=10; a_3= 6; a_4= 3; a_5= 2; a_6= 1;
m_1=2; m_2=5; m_3=8; m_4=10; m_5=15; m_6=20.

45. a_1=10; a_2= 8; a_3= 6; a_4= 4; a_5= 2; a_6= 1;
m_1=5; m_2=10; m_3=12; m_4=15; m_5=18; m_6=20.

46. a_1=6; a_2=5; a_3= 4; a_4= 3; a_5= 2; a_6= 1;
m_1=2; m_2=4; m_3=6; m_4=10; m_5=15; m_6=20.

47. a_1=14; a_2=12; a_3= 8; a_4= 5; a_5= 1;
m_1=2; m_2=8; m_3=15; m_4=20; m_5=30.

48. $a_1=12$; $a_2=10$; $a_3=6$; $a_4=3$; $a_5=1$;
$m_1=5$; $m_2=8$; $m_3=14$; $m_4=25$; $m_5=30$.

49. $a_1=8$; $a_2=5$; $a_3=4$; $a_4=2$;
$m_1=4$; $m_2=6$; $m_3=12$; $m_4=20$.

50. $a_1=5$; $a_2=4$; $a_3=3$; $a_4=2$;
$m_1=8$; $m_2=10$; $m_3=15$; $m_4=25$.

Задачи 51–60

Вес изготовленного серебряного изделия должен составлять a граммов.

При изготовлении возможны случайные погрешности, в результате которых вес изделия случаен, но подчинен нормальному закону распределения со средним квадратическим отклонением σ граммов.

Требуется найти вероятность того, что:

а) вес изделия составит от α до β граммов;

б) величина погрешности в весе не превзойдет δ граммов по абсолютной величине.

51. $a=50$; $\sigma=2$; $\alpha=30$; $\beta=55$; $\delta=5$;
52. $a=60$; $\sigma=2$; $\alpha=56$; $\beta=62$; $\delta=6$;
53. $a=70$; $\sigma=3$; $\alpha=64$; $\beta=80$; $\delta=7$;
54. $a=80$; $\sigma=3$; $\alpha=75$; $\beta=92$; $\delta=8$;
55. $a=90$; $\sigma=4$; $\alpha=78$; $\beta=95$; $\delta=9$;
56. $a=100$; $\sigma=4$; $\alpha=80$; $\beta=110$; $\delta=10$;
57. $a=120$; $\sigma=5$; $\alpha=100$; $\beta=150$; $\delta=10$;
58. $a=130$; $\sigma=5$; $\alpha=125$; $\beta=140$; $\delta=12$;
59. $a=140$; $\sigma=6$; $\alpha=130$; $\beta=155$; $\delta=14$;
60. $a=150$; $\sigma=6$; $\alpha=145$; $\beta=160$; $\delta=15$.

Задачи 61–70

По итогам выборочных обследований для некоторой категории сотрудников величина их дневного заработка $\boldsymbol{X}$ руб. и соответствующее количество сотрудников $\boldsymbol{n_i}$ представлены в виде интервального статистического распределения.

а) Построить гистограмму относительных частот распределения.

б) Найти основные характеристики распределения выборочных данных: среднее выборочное значение, выборочную дисперсию и выборочное среднее квадратическое отклонение.

в) Оценить генеральные характеристики по найденным выборочным характеристикам.

г) Считая, что значения признака X в генеральной совокупности подчинены нормальному закону распределения, найти доверительный интервал для оценки математического ожидания (генерального среднего значения) с надежностью γ, считая, что генеральная дисперсия равна исправленной выборочной дисперсии.

61.

X	50-60	60-70	70-80	80-90	90-100
n_i	5	10	20	15	10

$\gamma = 0{,}95$

62.

X	30-34	34-38	38-42	42-46	46-50
n_i	2	5	15	10	8

$\gamma = 0{,}90$

63.

X	40-46	46-52	52-58	58-64	64-70
n_i	5	10	20	15	10

$\gamma = 0{,}92$

64.

X	42-44	44-46	46-48	48-50	50-52
n_i	7	12	18	13	5

$\gamma = 0{,}94$

65.

X	60-64	64-68	68-72	72-76	76-80
n_i	5	12	20	15	8

$\gamma = 0{,}91$

66.

X	66-70	70-74	74-78	78-82	82-86	86-90
n_i	7	15	22	18	5	3

$\gamma = 0{,}93$

67.

X	36-42	42-48	48-54	54-60	60-66	66-72
n_i	8	13	15	15	7	2

$\gamma = 0{,}85$

68.

X	52-56	56-60	60-64	64-68	68-72	72-76	76-80
n_i	5	15	25	18	12	8	2

$\gamma = 0{,}99$

69.

X	42-46	46-50	50-54	54-58	58-62	62-66	66-70
n_i	8	15	19	22	12	5	1

$\gamma = 0{,}98$

70.

X	80-82	82-84	84-86	86-88	88-90
n_i	3	7	20	15	5

$\gamma = 0{,}88$

Задачи 71–80

С целью анализа взаимного влияния прибыли предприятия и его издержек выборочно были проведены наблюдения за этими показателями в течение ряда месяцев: ***X*** – величина месячной прибыли в тыс. руб., ***Y*** – месячные издержки в процентах к объему продаж.

Результаты выборки сгруппированы и представлены в виде корреляционной таблицы, где указаны значения признаков ***X*** и ***Y*** и количество месяцев, за которые наблюдались соответствующие пары значений названных признаков.

а) По данным корреляционной таблицы найти условные средние $\overline{y}_x$ и $\overline{x}_y$.

б) Оценить тесноту линейной связи между признаками ***X*** и ***Y***.

в) Составить уравнения линейной регрессии ***Y*** по ***X*** и ***X*** по ***Y*** .

г) Сделать чертеж, нанеся на него условные средние и найденные прямые регрессии.

д) Оценить силу связи между признаками с помощью корреляционного отношения.

71.

$Y \diagdown X$	20	30	40	50	60	n_y
5	3					
10	5	4				
15		4	2			
20			5	4	5	
25			3	1	6	
30					3	
n_x						

72.

$Y \backslash X$	15	25	35	45	55	n_y
10	1					
15	4	5				
20		3	2			
25			5	6	1	
30			2	1	2	
35					3	
n_x						

73.

$Y \backslash X$	30	40	50	60	70	n_y
5	1					
10	5	5				
15		3	2	4		
20			4	1	4	
25			2	7	6	
30					3	
n_x						

74.

$Y \backslash X$	25	35	45	55	65	n_y
15	4					
20	2	6				
25		4	6	2		
30			5	8	4	
35			2	6	7	
40					4	
n_x						

75.

Y \ X	25	35	45	55	65	n_y
12	2					
17	4	6				
22		3		2		
27			5	6	2	
32			4	1	4	
37					3	
n_x						

76.

Y \ X	110	120	130	140	150	n_y
2	2					
7	4	6				
12		2	3	1		
17			5	4	1	
22			2	3	4	
27					3	
n_x						

77.

Y \ X	20	30	40	50	60	n_y
5	2					
10	4	3				
15		7	5			
20			2	7	5	
25			3	2	6	
30					3	
n_x						

78.

$Y \diagdown X$	35	45	55	65	75	n_y
10	5					
15	1	6				
20		2	5	1		
25			8	10	1	
30			5	2	4	
35					8	
n_x						

79.

$Y \diagdown X$	35	45	55	65	75	n_y
5	4					
10	2	5				
15		3	3			
20			5	1	1	
25			2	4	2	
30					3	
n_x						

80.

$Y \diagdown X$	30	40	50	60	70	n_y
4	3					
9	3	2				
14		4	1			
19			2	4	4	
24			8	2	7	
29					3	
n_x						

ВОПРОСЫ ДЛЯ ЗАЩИТЫ КОНТРОЛЬНОЙ РАБОТЫ

Выполненную контрольную работу студент-заочник должен защитить. Ниже предлагаются вопросы, которые могут быть заданы студенту в процессе защиты. Ответы на эти вопросы можно найти в учебниках, при этом достаточно использовать только один из перечисленных в разделе «Литература» учебников. На большую часть вопросов студент сможет найти ответ, внимательно изучив методические указания, приведенные далее. Эти вопросы полезно использовать также при подготовке к экзаменам.

Вопросы к задачам 1–10, 11–20, 21–30

1. Поясните понятие «испытание».
2. Что называется событием? Как обозначаются события?
3. Дайте определения несовместимых, единственно возможных, равновозможных событий.
4. Какие события образуют систему элементарных исходов?
5. Дайте классическое определение вероятности.
6. Что называется частотой появления события? В чем заключается свойство устойчивости относительных частот?
7. Дайте статистическое определение вероятности.
8. Какие численные значения может принимать вероятность?
9. Дайте определение достоверного, невозможного, случайного события. Чему равны их вероятности?
10. Сформулируйте теоремы сложения вероятностей для несовместимых и совместимых событий.
11. Сформулируйте теоремы умножения вероятностей для зависимых и независимых событий. Что такое условная вероятность?
12. Дайте определение полной группы событий. Каким свойством обладают вероятности событий, образующих полную группу?
13. Какие события называются противоположными? Как найти вероятность противоположного события?
14. Как найти вероятность появления только одного из двух независимых или зависимых событий?
15. Как найти вероятность появления только одного из трех независимых или зависимых событий?
16. Как найти вероятность появления только двух событий из трех независимых или зависимых событий?

17. Как найти вероятность появления хотя бы одного из двух или трех независимых или зависимых событий?

18. Запишите формулу полной вероятности. Что вычисляют по этой формуле?

19. Для чего используется формула Байеса?

Вопросы к задачам 31–40

1. Запишите формулу Бернулли. Для чего она используется?
2. Для чего служит локальная формула Лапласа? Когда она применяется?
3. Для чего служит интегральная формула Лапласа? Когда ее применяют?
4. Дайте понятие наивероятнейшего числа появлений события в $\boldsymbol{n}$ независимых испытаниях. Как найти это число? Как найти соответствующую ему вероятность?
5. Функция $\varphi(x)$, ее свойства и графическое изображение. Как пользоваться таблицей для функции $\varphi(x)$?
6. Функция Лапласа $\Phi(x)$, ее свойства и график. Как пользоваться таблицей значений для функции $\Phi(x)$?

Вопросы к задачам 41–50

1. Сформулируйте понятие случайной величины. Когда случайная величина называется дискретной? Приведите примеры.
2. Как задать закон распределения дискретной случайной величины?
3. Назовите основные характеристики закона распределения случайной величины.
4. Как находится математическое ожидание дискретной случайной величины? В чем смысл этой характеристики?
5. Дайте определение дисперсии дискретной случайной величины. В чем смысл этой характеристики?
6. Как находится среднее квадратическое отклонение случайной величины? В чем его смысл?
7. Запишите две формулы для вычисления дисперсии дискретной случайной величины.

Вопросы к задачам 51–60

1. Сформулируйте понятие непрерывной случайной величины. Приведите примеры.
2. Как задать закон распределения непрерывной случайной величины?
3. В чем заключается вероятностный смысл интегральной функции распределения?
4. В чем заключается вероятностный смысл дифференциальной функции распределения?
5. Как связаны между собой дифференциальная и интегральная функции распределения?
6. Назовите свойства интегральной функции распределения. Каковы особенности ее графика?
7. Назовите свойства дифференциальной функции распределения. Каковы особенности ее графика?
8. Как найти вероятность попадания непрерывной случайной величины в заданный промежуток с помощью интегральной и с помощью дифференциальной функций распределения?
9. Назовите основные числовые характеристики непрерывной случайной величины. По каким формулам они находятся?
10. Как задается нормальный закон распределения? В чем смысл его параметров?
11. Каковы особенности нормального закона распределения? Поясните их на кривой Гаусса.
12. Как находится вероятность попадания нормально распределенной случайной величины в заданный промежуток?
13. Как находится вероятность заданного отклонения нормально распределенной случайной величины от ее математического ожидания?
14. Сформулируйте правило «трех сигм».

Вопросы к задачам 61–70

1. В чем состоит сущность выборочного метода?
2. Как записать результаты выборки?
3. Назовите основные характеристики выборки.

4. Как вычисляется выборочная средняя?

5. Выборочная дисперсия и среднее квадратическое отклонение, в чем заключается смысл этих характеристик?

6. Способы вычисления выборочной дисперсии.

7. Как оценить генеральную среднюю и генеральную дисперсию признака?

8. Для чего служит доверительный интервал?

9. Как вычислить длину доверительного полуинтервала, если значения признака X распределены по нормальному закону? От чего зависит эта длина?

10. Как влияет заданный уровень надежности на точность оценки?

Вопросы к задачам 71–80

1. Назовите виды зависимостей между признаками.

2. Объясните устройство корреляционной таблицы.

3. Поясните на корреляционной таблице, как задается статистическая зависимость между признаками.

4. В каком случае статистическая зависимость является корреляционной?

5. Условные средние, их вычисление.

6. Общие средние, их вычисление.

7. Среднее квадратическое отклонение, его смысл и вычисление.

8. Как оценить тесноту (силу) линейной связи?

9. Какие значения может принимать коэффициент линейной корреляции? Как по значению коэффициента корреляции делается вывод о направлении и тесноте линейной связи?

10. В чем заключается сущность метода наименьших квадратов? Для чего он применяется?

11. Запишите уравнения регрессий $\boldsymbol{Y}$ по $\boldsymbol{X}$ и $\boldsymbol{X}$ по $\boldsymbol{Y}$. Для чего они служат?

12. Каковы особенности расположения прямых регрессии на графике?

13. Что характеризуют внутригрупповая и межгрупповая дисперсии? Как они вычисляются?

14. Для чего служит корреляционное отношение? Как оценить силу связи по корреляционному отношению?

15. Назовите виды криволинейной регрессии.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К РЕШЕНИЮ ЗАДАЧ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Теория вероятностей – это раздел математики, где изучаются закономерности массовых случайных явлений.

Тема 1. Основные понятия и теоремы теории вероятностей

ИСПЫТАНИЕ – это изначальное понятие**,** разъясняется как наблюдение, явление, действие, опыт и прочее.

СОБЫТИЕ – это результат (исход) испытания.

Пусть в результате испытания могут появиться несколько событий (исходов), эти исходы называются:

– ***несовместимыми*** – если появление одного из них исключает появление других;

– ***единственно-возможными*** – если в результате испытания появится хотя бы одно из них (других нет);

– ***равновозможными*** – если нет преимуществ у каждого исхода перед другими при испытании.

Например: испытание – бросание монеты;

событие $Г$ – выпал герб;

событие $Ч$ – выпало число.

События $Г$ и $Ч$ несовместимые, единственно-возможные, равновозможные.

События, которые могут появиться в результате испытания, образуют ***полную систему элементарных*** исходов, если они:

1) несовместимы;
2) единственно-возможны;
3) по каждому из этих исходов можно судить о появлении или непоявлении любого события, которое может произойти в результате данного испытания.

Рассмотрим пример с игральной костью.

Игральная кость – это однородный кубик, на гранях которого изображено количество очков от 1 до 6, так как всего граней 6.

Испытание – бросание игральной кости.

Событие – выпадение определенного количества очков на верхней грани.

Полную систему элементарных исходов образуют шесть событий: E_1, E_2, E_3, E_4, E_5, E_6, которые заключаются в том, что количество выпавших очков составит соответственно 1, 2, 3, 4, 5 или 6 очков.

Действительно, эти шесть исходов удовлетворяют трем перечисленным выше условиям полной системы элементарных исходов.

Пусть событие *A* – «выпало четное число очков». Это означает, что появились события E_2 или E_4 или E_6. По этим трем элементарным исходам можно судить о появлении события *A*. Они называются благоприятствующими событию *A*.

Пусть событие *B* – «количество выпавших очков больше четырех». Это означает, что появились события E_5 или E_6, то есть эти два элементарных исхода благоприятствуют событию *B*.

Отметим также, что, так как кубик однородный, то все шесть элементарных исходов являются **равновозможными**.

Классическое определение вероятности

Вероятностью события *A* называется число *P(A)*, равное отношению количества благоприятствующих событию *A* элементарных исходов ***m*** к общему количеству элементарных **равновозможных** исходов ***n***:

$$\boxed{P(A) = \frac{m}{n}}$$

В примере с игральной костью всего элементарных равновозможных исходов $n = 6$, поэтому

$$P(A) = \frac{m_A}{n} = \frac{3}{6} = \frac{1}{2} = 0{,}5 \ ;$$

$$P(B) = \frac{m_B}{n} = \frac{2}{6} = \frac{1}{3} \ .$$

Заметим, что $P(E_1) = P(E_2) = \ldots = P(E_6) = 1/6$, так как любому элементарному событию благоприятствует один исход (одна грань из шести).

В примере с бросанием монеты $P(Г) = 1/2$ и $P(Ч) = 1/2$, так как каждому из событий «выпал герб» и «выпало число», благоприятствует один элементарный исход из двух равновозможных элементарных исходов *Г* и *Ч*.

Рассмотрим ещё пример с урновой схемой.

Урна – это ёмкость с шарами. Пусть всего в урне 20 одинаковых на ощупь шаров (по размеру, температуре, гладкости), которые отличаются только цветом, например, 12 из них красные, а остальные – белые.

Испытание – извлечение наугад одного шара.

Событие K – появление красного шара.

Очевидно, что всего элементарных исходов n = 20 (по количеству шаров), причём все эти исходы равновозможны. Событию K благоприятствуют 12 исходов (по количеству красных шаров), поэтому

$$P(K)=\frac{m_K}{n}=\frac{12}{20}=0{,}6.$$

Внимание! Вероятность любого события может принимать значения только от 0 до 1:

$$0\le P(A)\le 1.$$

Вероятность можно задать в процентах, например $P(A)$= 0,8 или $P(A)$= 0,8 · 100% = 80 %.

Достоверное событие – обязательно произойдет в результате испытания, так как все исходы благоприятные, то есть $m = n$:

$$P(A)=\frac{m}{n}=\frac{n}{n}=1.$$

Невозможное событие – не может произойти в результате испытания, так как благоприятных исходов нет, то есть m = 0:

$$P(A)=\frac{m}{n}=\frac{0}{n}=0.$$

Случайное событие – может произойти или не произойти в результате испытания: $0<P(A)<1$.

Вероятность является числовой мерой объективной возможности наступления события.

Статистическое понятие вероятности

На практике чаще пользуются статистическим понятием вероятности.

Вероятность события A – это число, около которого колеблются устойчивые значения относительных частот W при многократном повторении испытаний:

$$P(A) \approx W(A) = \frac{m}{n},$$

где n – количество испытаний,

m – количество появлений события A.

Устойчивость означает, что относительные частоты незначительно изменяются в различных сериях испытаний.

Например, если подбрасывать монету много раз, то, как показывают опыты, примерно в половине случаев выпадет герб, то есть $P(Г) = \frac{1}{2}$, что соответствует результату, полученному по классическому определению вероятности.

Аналогично можно подбрасывать много раз игральную кость, или много раз извлекать шар из урны, возвращая его каждый раз обратно, и по результатам этих опытов получить значения вероятностей для конкретных событий.

Статистическое понятие вероятности применимо лишь в тех случаях, когда испытания можно воспроизводить многократно.

Например, результаты статистических наблюдений за рождением детей показывают, что на каждую 1000 детей мальчиков приходится около 515. Поэтому принято считать, что вероятность рождения мальчика $P(М)$=0,515 (51, 5%), а для девочки $P(Д)$=0,485 (48,5%).

Аналогично по результатам статистических наблюдений можно установить: вероятность попадания в цель каждым стрелком; вероятность выпуска бракованных изделий каждым изготовителем; вероятность поступления заявки на товар определенного вида и т. п.

Основные теоремы теории вероятностей

Теорема сложения вероятностей **несовместимых** событий:

$$\boxed{P(\,A \text{ или } B) = P(A) + P(B)}$$

Теорема умножения вероятностей **независимых** событий:

$$\boxed{P(\,A \text{ и } B) = P(A) \cdot P(B)}$$

События называются **независимыми**, если вероятность одного из них не зависит от того, произошло или нет другое событие.

Теорема умножения вероятностей **зависимых** событий:

$$\boxed{P(\,A \text{ и } B) = P(A) \cdot P_A(B)}$$

где $P_A(B)$ – условная вероятность, то есть вероятность события $\boldsymbol{B}$ при условии, что событие $\boldsymbol{A}$ уже произошло.

Эти теоремы остаются справедливыми для трех и более событий.

Теорема сложения вероятностей **совместимых** событий:

$$\boxed{P(\,A \text{ или } B) = P(A) + P(B) - P(\,A \text{ и } B)}$$

Эта теорема позволяет находить вероятность появления хотя бы одного из двух событий A или B.

Ниже на примерах будет показано, как эти теоремы применяют для решения конкретных задач, в частности, для нахождения вероятности появления только одного события, хотя бы одного события и в других случаях.

Полная группа событий

События $H_1, H_2, \ldots, H_n$ образуют полную группу, если они:

а) несовместимы попарно;

б) единственно возможны.

Для таких событий справедливо равенство:

$$P(H_1) + P(H_2) + \ldots + P(H_n) = 1.$$

Если в полной группе лишь два события: $\boldsymbol{A}$ и $\overline{A}$ (не $\boldsymbol{A}$), то они называются противоположными, очевидно, при этом

$$P(A)+P(\overline{A})=1, \text{ то есть } \quad P(\overline{A})=1-P(A)\cdot$$

Формула полной вероятности

Пусть событие ***A*** наступает с одним из событий (гипотез) H_i , тогда полная вероятность события ***A*** находится по формуле

$$P(A) = P(H_1) \cdot P_{H_1}(A) + P(H_2) \cdot P_{H_2}(A) + \ldots + P(H_n) \cdot P_{H_n}(A),$$

где события $H_1, H_2, \ldots, H_n$ образуют полную группу.

Формула Байеса

Эта формула служит для переоценки вероятности гипотезы H_k при условии, что событие A произошло:

$$P_A(H_k) = \frac{P(H_k)\cdot P_{H_k}(A)}{\sum_{i=1}^{n} P(H_i)\cdot P_{H_i}(A)} .$$

Здесь в знаменателе стоит полная вероятность события A, а в числителе – одно из ее слагаемых.

Задачи 1–10

Пример 1. Три стрелка делают по одному выстрелу в цель. Вероятность попадания для первого стрелка 0,8, для второго – 0,7, для третьего – 0,6. Найти вероятность того, что:

а) все стрелки попадут в цель;

б) только один стрелок попадет в цель;

в) только два стрелка попадут в цель;

г) все стрелки промахнутся;

д) цель будет поражена.

Решение.

Обозначим: событие A – первый стрелок попадет в цель,

$P(A)=0{,}8$; $P(\overline{A})=1-0{,}8=0{,}2$;

событие B – второй стрелок попадет в цель,

$P(B)=0{,}7$; $P(\overline{B})=1-0{,}7=0{,}3$;

событие C – третий стрелок попадет в цель,

$P(C)=0{,}6$; $P(\overline{C})=1-0{,}6=0{,}4$.

События A, B, C – независимые.

а) P (все попадут) = $P(A$ и B и $C)$ = $P(A) \cdot P(B) \cdot P(C)$ =
= 0,8 · 0,7 · 0,6 = 0,336.

б) P (только 1 стрелок попадет) =
= $P(A$ и $\overline{B}$ и $\overline{C}$ или $\overline{A}$ и B и $\overline{C}$ или $\overline{A}$ и $\overline{B}$ и $C)$ =
= $P(A) \cdot P(\overline{B}) \cdot P(\overline{C}) + P(\overline{A}) \cdot P(B) \cdot P(\overline{C}) + P(\overline{A}) \cdot P(\overline{B}) \cdot P(C)$ =
= 0,8 · 0,3 · 0,4 + 0,2 · 0,7 · 0,4 + 0,2 · 0,3 · 0,6 =
= 0,096 + 0,056 + 0,036 = 0,188.

в) P (только 2 стрелка попадут) = $P(A$ и B и $\overline{C}$ или A и $\overline{B}$ и C или $\overline{A}$ и B и $C)$ =

$=P(A)\cdot P(B)\cdot P(\overline{C})+P(A)\cdot P(\overline{B})\cdot P(C)+P(\overline{A})\cdot P(B)\cdot P(C)=$
= 0,8 · 0,7 · 0,4 + 0,8 · 0,3 · 0,6 +0,2 · 0,7 · 0,6 =
= 0,224 +0,144 + 0,084 = 0,452.

г) P (все промахнутся) = $P(\overline{A}$ и $\overline{B}$ и $\overline{C})$=

$=P(\overline{A})\cdot P(\overline{B})\cdot P(\overline{C})=0{,}2\cdot 0{,}3\cdot 0{,}4=0{,}024.$

д) P (цель поражена) = P (попадет хотя бы один стрелок) =
= 1 – P (все стрелки промахнутся) = 1 – P ($\overline{A}$ и $\overline{B}$ и $\overline{C}$) =
$=1-P(\overline{A})\cdot P(\overline{B})\cdot P(\overline{C})$ = 1 – 0,2 · 0,3 · 0,4 = 1 – 0,024 = 0,976.

Контроль. События а), б), в) и г) несовместимы попарно и единственно возможны, следовательно, они образуют полную группу событий, действительно:

$$0{,}336+0{,}188+0{,}452+0{,}024=1.$$

Задачи 11–20

Пример 2. В партии из 100 одинаковых по внешнему виду изделий смешаны 40 шт. I сорта и 60 шт. II сорта. Найти вероятность того, что взятые два изделия окажутся:

а) одного сорта;

б) разных сортов.

Решение.

Обозначим: событие A_1 – первое взятое изделие I сорта,
событие A_2 – второе взятое изделие I сорта,
событие B_1 – первое взятое изделие II сорта,
событие B_2 – второе взятое изделие II сорта.

Эти события зависимые. Используем теорему сложения вероятностей для несовместимых событий и теорему умножения вероятностей для зависимых событий.

а) P (изделия одного сорта) = $P(A_1$ и A_2 или B_1 и $B_2) =$
$= P(A_1$ и $A_2) + P(B_1$ и $B_2) = P(A_1) \cdot P_{A1}(A_2) + P(B_1) \cdot P_{B1}(B_2) =$
$= (40/100) \cdot (39/99) + (60/100) \cdot (59/99) = 5100/9900 = 51/99$;

б) P (изделия разных сортов) = $P(A_1$ и B_2 или B_1 и $A_2) =$
$=P(A_1) \cdot P_{A1}(B_2) + P(B_1) \cdot P_{B1}(A_2) = (40/100) \cdot (60/99) +$
$+(60/100) \cdot (40/99) = 4800/9900 = 48/99$.

Контроль. События под буквами а) и б) противоположны, следовательно, сумма их вероятностей должна быть равна единице. Действительно: $51/99 + 48/99 = 99/99 = 1$.

Задачи 21–30

Пример 3. Два консервных завода поставляют в магазин мясные и овощные консервы, причем первый завод поставляет продукции в три раза больше второго. Доля овощных консервов в продукции первого завода составляет 60%, а второго 70%. Для контроля в магазине взято наугад одно изделие.

а) Какова вероятность того, что это окажутся мясные консервы?

б) Взятое изделие оказалось мясными консервами. Какова вероятность, что оно изготовлено вторым заводом?

Решение. Обозначим: событие A – взяты мясные консервы;
событие H_1 – изделие изготовлено I заводом;
событие H_2 – изделие изготовлено II заводом.

По условию задачи первый завод поставляет продукции в три раза больше, чем второй, то есть $P(H_1) > P(H_2)$ в три раза, или

$$P(H_1) = 3 \cdot P(H_2).$$

Отсюда имеем:

$$\begin{cases} P(H_1) + P(H_2) = 1 \\ P(H_1) = 3P(H_2) \end{cases} \Rightarrow \begin{cases} 3P(H_2) + P(H_2) = 1 \\ P(H_1) = 3P(H_2) \end{cases} \Rightarrow \begin{cases} P(H_2) = 1/4 \\ P(H_1) = 3/4. \end{cases}$$

Вероятность того, что консервы мясные, для первого завода составляет 40%, то есть $P_{H1}(A) = 0{,}4$, для второго завода 30%, то есть $P_{H2}(A) = 0{,}3$.

а) Учитывая, что событие A произойдет обязательно с одним из событий (гипотез) H_i, образующих полную группу, применим формулу полной вероятности:

$$P(A) = P(H_1) \cdot P_{H_1}(A) + P(H_2) \cdot P_{H_2}(A) = \frac{3}{4} \cdot 0{,}4 + \frac{1}{4} \cdot 0{,}3 = \frac{1{,}2 + 0{,}3}{4} = 0{,}375.$$

б) По условию событие A произошло, то есть взяты мясные консервы. Тогда вероятность гипотезы H_2 – консервы изготовлены вторым заводом – находим по формуле Байеса

$$P_A(H_2) = \frac{P(H_2) \cdot P_{H2}(A)}{P(H_1) \cdot P_{H1}(A) + P(H_2) \cdot P_{H2}(A)} = \frac{\frac{1}{4} \cdot 0{,}3}{\frac{3}{4} \cdot 0{,}4 + \frac{1}{4} \cdot 0{,}3} =$$

$$= \frac{0{,}075}{0{,}3 + 0{,}075} = \frac{0{,}075}{0{,}375} = 0{,}2 \quad (20\%).$$

Замечание. Задача решается аналогично, если количество заводов будет три или более. Соответственно увеличится число слагаемых в формуле полной вероятности.

Тема 2. Повторение независимых испытаний

Пусть известна вероятность появления события A в одном испытании: $P(A) = \boldsymbol{p}$, причем $\boldsymbol{p} \neq 0$ и $\boldsymbol{p} \neq 1$, тогда

$P(\overline{A}) = 1 - \boldsymbol{p} = \boldsymbol{q}$ – вероятность непоявления события A.

Испытание повторяется $\boldsymbol{n}$ раз. Требуется найти вероятность того, что событие A наступит при этом ровно $\boldsymbol{k}$ раз:

$P_n(k)$ – вероятность, что в $\boldsymbol{n}$ испытаниях событие наступит $\boldsymbol{k}$ раз.

Эта вероятность находится по формуле Бернулли

$$P_n(k) = \frac{n!}{k!(n-k)!} \cdot p^k \cdot q^{n-k},$$

! – знак факториала, математической операции такой, что

$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$, например,

$1! = 1$

$2!=1\cdot2=2$

Внимание: **0!=1**

$3!=1\cdot2\cdot3=6$

$4!=1\cdot2\cdot3\cdot4=24$

$5!=1\cdot2\cdot3\cdot4\cdot5=120$ и т. д.

Формулу Бернулли удобно применять, если число повторных испытаний невелико ($n \le 10$).

Формулы Лапласа

Если число испытаний велико ($n > 10$), то вместо формулы Бернулли используется так называемая локальная формула Лапласа, которая является приближенной:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x_0), \quad \text{где} \quad x_0 = \frac{k-np}{\sqrt{npq}}.$$

Значение функции $\varphi(x_0)$ берется по таблице (см. приложение 1).

Если число испытаний велико ($n > 10$) и нужно найти вероятность того, что при $\boldsymbol{n}$ испытаниях событие A наступит от $\boldsymbol{k_1}$ до $\boldsymbol{k_2}$ раза, то применяют интегральную формулу Лапласа, которая также является приближенной:

$$P_n(k_1;k_2) \approx \Phi(x_2) - \Phi(x_1), \quad \text{где} \quad x_1 = \frac{k_1-np}{\sqrt{npq}}; \quad x_2 = \frac{k_2-np}{\sqrt{npq}}.$$

Значения функции Лапласа $\Phi(x)$ берутся по таблице (приложение 2).

Приближенные формулы Лапласа тем точнее, чем больше количество повторных испытаний $\boldsymbol{n}$.

Наивероятнейшее число появлений события

Пусть в $\boldsymbol{n}$ повторных испытаниях событие A появляется $\boldsymbol{k}$ раз, где $\boldsymbol{k}$ может принимать значения: 0; 1; 2; 3; …; $\boldsymbol{n}$ (то есть $0 \le \boldsymbol{k} \le \boldsymbol{n}$). Для каждого из этих значений $\boldsymbol{k}$ можно найти соответствующую ему вероятность по формулам Бернулли или Лапласа.

То значение $\boldsymbol{k}$, которому соответствует самая большая вероятность, называется НАИВЕРОЯТНЕЙШИМ числом появления события A.

Наивероятнейшее число $\boldsymbol{k_0}$ находится как ***целое число*** из промежутка:

$$np - q \le k_0 \le np + p\ ,$$

$\boldsymbol{k_0}$ может принимать либо одно значение, либо два соседних целых значения, когда вероятности их одинаковы.

Вероятность $P_n(k_o)$, соответствующую значению $k = k_0$, находим либо по формуле Бернулли, при $n \le 10$, либо по локальной формуле Лапласа, при $n > 10$.

Появление события хотя бы один раз

Вероятность появления события «хотя бы один раз» в $\boldsymbol{n}$ испытаниях находится с помощью противоположного ему события «ни одного раза»:

P_n (событие наступит хотя бы один раз) = $1 - P_n$ (ни разу) =

$$= 1 - P_n(0) = 1 - \frac{n!}{0!\cdot n!}\cdot p^0 \cdot q^{n-0} = 1 - q^n\ ,$$

при этом учтено, что $0! = 1$ и $p^0 = 1$.

Событие наступит «хотя бы один раз» означает, что оно наступит один или более раз, поэтому можно записать

$$P_n(k \ge 1) = 1 - q^n.$$

Задачи 31–40

Пример 4. Стрелок поражает цель с вероятностью 0,7.

1) С какой вероятностью в серии из 5 выстрелов он поразит мишень:

а) ровно два раза;

б) хотя бы один раз;

в) не менее четырех раз.

г) каково наивероятнейшее число попаданий и соответствующая ему вероятность?

2) При тех же условиях стрелок выполняет серию из 20 выстрелов.

а) Чему равна вероятность того, что попаданий будет ровно половина?

б) Найти вероятность того, что число попаданий будет не менее 12 и не более 18.

Решение. По условию задачи: $\boldsymbol{p} = 0{,}7$; $\boldsymbol{n} = 5$; $\boldsymbol{k} = 2$; $\mathbf{m} = 4$;
$\boldsymbol{N} = 20$; $\boldsymbol{k_1} = 12$; $\boldsymbol{k_2} = 18$.

Вероятность промаха $\boldsymbol{q} = 1 - \boldsymbol{p} = 1 - 0{,}7 = 0{,}3$.

а) Вероятность попадания ровно два раза в серии из пяти выстрелов находим по формуле Бернулли, так как число испытаний $n = 5$ невелико ($n \le 10$):

$$\mathrm{P}_5(2) = \frac{5!}{2!(5-2)!} \cdot p^2 \cdot q^{5-2} = \frac{5!}{2!\cdot 3!} \cdot 0{,}7^2 \cdot 0{,}3^3 =$$

$$= \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 1\cdot 2\cdot 3} \cdot 0{,}49 \cdot 0{,}027 = 10 \cdot 0{,}01323 = 0{,}1323.$$

б) Событию D – «стрелок поразит мишень хотя бы 1 раз», – противоположно событие $\overline{D}$ – «не поразит ни разу», то есть стрелок промахнется все пять раз, следовательно, число попаданий $k = 0$:

$$P(D) = 1 - P(\bar{D}) = 1 - P_5(0) = 1 - \frac{5!}{0!\cdot 5!} \cdot p^0 \cdot q^5 = 1 - q^5 =$$

$$= 1 - 0{,}3^5 = 1 - 0{,}00243 = 0{,}99757 .$$

Здесь учтено, что $0! = 1$ и $p^0 = 1$.

в) Событие «стрелок поразит мишень не менее четырех раз» запишем в виде: $m \ge 4$, тогда

$$P_5(m \ge 4) = P_5(4 \text{ или } 5) = P_5(4) + P_5(5) .$$

Здесь применена теорема сложения вероятностей несовместимых событий. Используя формулу Бернулли, найдем:

$$P_5(4)=\frac{5!}{4!\cdot(5-4)!}\cdot p^4\cdot q^1=\frac{5!}{4!\cdot1!}\cdot0{,}7^4\cdot0{,}3=5\cdot0{,}2401\cdot0{,}3=0{,}36015\ ;$$

$$P_5(5)=\frac{5!}{5!\cdot0!}\cdot p^5\cdot q^0=1\cdot p^5\cdot1=0{,}7^5=0{,}16807\ ;$$

$$P_5(m\ge4)=0{,}36015+0{,}16807=0{,}52822\ .$$

г) Наивероятнейшее число попаданий $\boldsymbol{k_o}$ находим как ***целое*** число из промежутка:

$$np-q\le k_o\le np+p\ ;$$

$$5\cdot0{,}7-0{,}3\le k_o\le5\cdot0{,}7+0{,}7\ ;$$

$$3{,}2\le k_o\le4{,}2\ ;$$

$$k_o=4\ .$$

Соответствующую ему вероятность $P_5(4)$ вычислим по формуле Бернулли. В данной задаче она уже была найдена выше:

$$P_5(4)=\frac{5!}{4!\cdot1!}\cdot0{,}7^4\cdot0{,}3=0{,}36015\ .$$

2) В этой задаче число испытаний N = 20 достаточно велико ($N>10$), поэтому используем приближенные формулы Лапласа.

а) Число попаданий равно половине из 20, то есть $k=\dfrac{N}{2}=10.$ Соответствующую вероятность находим по локальной формуле Лапласа.

$$P_{20}(10)\approx\frac{1}{\sqrt{Npq}}\cdot\varphi(x_0),\qquad \text{где}$$

$$x_0=\frac{k-Np}{\sqrt{Npq}}=\frac{10-20\cdot0{,}7}{\sqrt{20\cdot0{,}7\cdot0{,}3}}=\frac{10-14}{\sqrt{4{,}2}}\approx-1{,}95\ .$$

Результат вычислений для $\boldsymbol{x_0}$ округляем с точностью до 0,01, так как значения функции $\varphi(x_0)$ табулируются в приложении 1 с такой точностью. По таблице приложения 1, учитывая четность функции $\varphi(x_0)$, находим:

$$\varphi(x_0)=\varphi(-1{,}95)=\varphi(1{,}95)=0{,}0596\ .$$

Тогда $P_{20}(10) \approx \dfrac{1}{\sqrt{20 \cdot 0{,}7 \cdot 0{,}3}} \cdot \varphi(1{,}95) = \dfrac{1}{\sqrt{4{,}2}} \cdot 0{,}0596 = 0{,}0291$.

б) Вероятность того, что число попаданий будет не менее 12 и не более 18, вычисляем по интегральной формуле Лапласа:

$$P_{20}(k_1;k_2) \approx \Phi(x_2) - \Phi(x_1)\text{, где}$$

$$x_1 = \frac{k_1 - Np}{\sqrt{Npq}} = \frac{12 - 20 \cdot 0{,}7}{\sqrt{20 \cdot 0{,}7 \cdot 0{,}3}} = \frac{-2}{\sqrt{4{,}2}} = -0{,}9759 \approx -0{,}98;$$

$$x_2 = \frac{k_2 - Np}{\sqrt{Npq}} = \frac{18 - 20 \cdot 0{,}7}{\sqrt{4{,}2}} = \frac{4}{2{,}04939} \approx 1{,}9518 \approx 1{,}95 .$$

Значения для $\boldsymbol{x_1}$ и $\boldsymbol{x_2}$ округляем до 0,01, так как таблица значений функции Лапласа $\Phi(x)$ предусматривает такую точность для $\boldsymbol{x}$. По таблице приложения 2, учитывая нечетность функции $\Phi(x)$, находим: $\Phi(-0{,}98) = -\Phi(0{,}98) = -0{,}3365$; $\Phi(1{,}95) = 0{,}4744$.

Тогда $P_{20}(12;18) = \Phi(1{,}95) - \Phi(-0{,}98) = 0{,}4744 + 0{,}3365 = 0{,}8109$.

Ответы: 1) а) 0,1323; б) 0,99757; в) 0,52822; г) $k_0 = 4$; 0,36015;
2) а) 0,0291; б) 0,8109.

Тема 3. Дискретная случайная величина

Случайной величиной называется переменная, принимающая свои возможные числовые значения с определенной вероятностью.

Например: X – балл, полученный на экзамене;
Y – число студентов, явившихся на лекцию;
Z – величина выигрыша в лотерее;
U – рост человека и т.п.

Основные виды случайных величин: 1) непрерывная;
2) дискретная.

Непрерывная случайная величина может принимать все значения из некоторого промежутка.

Дискретная случайная величина X принимает отдельные числовые значения. Закон распределения дискретной случайной величины записывается в виде таблицы, где перечислены все значения случайной величины X и соответствующие им вероятности:

X	x_1	x_2	x_3	…	x_n
P(X)	p_1	p_2	p_3	…	p_n

Следует иметь в виду, что всегда $\sum_{i=1}^{n} p_i = p_1 + p_2 + ... + p_n = 1.$

Основные числовые характеристики закона распределения дискретной случайной величины:

1) **Математическое ожидание** (ожидаемое среднее значение случайной величины):

$$M(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + ... + x_n \ p_n = a.$$

2) **Дисперсия** (мера рассеяния значений случайной величины X от среднего значения a):

$$D(X) = \sum_{i=1}^{n} (x_i - a)^2 p_i = (x_1 - a)^2 p_1 + (x_2 - a)^2 p_2 + ... + (x_n - a)^2 p_n .$$

Второй способ вычисления дисперсии:

$D(X) = M(X^2) - M^2(X),$ где

$$M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum x_i p_i ,$$

$$M(X^2) = x_1^2 p_1 + x_2^2 p_2 + \ldots + x_n{}^2 p_n = \sum x_i^2 p_i .$$

3) **Среднее квадратичное отклонение** (характеристика рассеяния в единицах признака X):

$$\sigma(X) = \sqrt{D(X)}.$$

Задачи 41–50

Пример 5. В лотерее на каждые 100 билетов приходится 2 билета с выигрышем по 50 тыс. рублей, 5 билетов по 20 тыс. рублей, 10 билетов по 10 тыс. рублей, 20 билетов по 5 тыс. рублей и 25 билетов по 3 тыс. рублей. Остальные билеты не выигрывают.

Составить закон распределения величины выигрыша для владельца одного билета и найти его основные характеристики.

Решение. Обозначим X тыс. рублей – величина выигрыша на один билет.

Очевидно, что X – случайная дискретная величина. Составим закон распределения этой случайной величины, перечислив все ее возможные значения и найдя соответствующие им вероятности. Число выигрышных билетов из 100 составляет: 2 + 5 + 10 + 20 + 25 = 62, значит, число невыигрышных билетов: 100 – 62 = 38.

Располагая величины возможного выигрыша $\boldsymbol{x_i}$ в порядке возрастания, получим следующую таблицу:

x_i	0	3	5	10	20	50
p_i	0,38	0,25	0,20	0,10	0,05	0,02

где $p_1 = P(X=0) = \frac{38}{100} = 0{,}38$; $p_2 = P(X=3) = \frac{25}{100} = 0{,}25$ и т. д.

Отметим, что $\sum p_i = 0{,}38 + 0{,}25 + 0{,}20 + 0{,}10 + 0{,}05 + 0{,}02 = 1.$

а) Математическое ожидание случайной величины X:

$$M(X) = \sum x_i p_i = 0\cdot 0{,}38 + 3\cdot 0{,}25 + 5\cdot 0{,}2 + 10\cdot 0{,}1 + 20\cdot 0{,}05 + 50\cdot 0{,}02 = 4{,}75.$$

Таким образом, ожидаемый средний выигрыш на 1 билет составляет 4,75 тыс. рублей.

б) Дисперсию случайной величины найдем двумя способами:

1) $$D(X) = \sum_{i=1}^{6} \left[x_i - M(X)\right]^2 \cdot p_i =$$

$$= (0-4{,}75)^2 \cdot 0{,}38 + (3-4{,}75)^2 \cdot 0{,}25 + (5-4{,}75)^2 \cdot 0{,}2 +$$

$$+ (10-4{,}75)^2 \cdot 0{,}1 + (20-4{,}75)^2 \cdot 0{,}05 + (50-4{,}75)^2 \cdot 0{,}02 =$$

$$= 8{,}57375 + 0{,}765625 + 0{,}0125 + 2{,}75625 + 11{,}628125 + 40{,}95125 =$$
$$= 64{,}6875.$$

2) $$D(X) = M(X^2) - M^2(X).$$

$$M(X^2) = \sum x_i^2 \cdot p_i = 0^2 \cdot 0{,}38 + 3^2 \cdot 0{,}25 + 5^2 \cdot 0{,}20 + 10^2 \cdot 0{,}1 + 20^2 \cdot 0{,}05 + 50^2 \cdot 0{,}02 =$$
$$= 0 + 2{,}25 + 5 + 10 + 20 + 50 = 87{,}25.$$

Тогда:

$$D(X) = 87{,}25 - (4{,}75)^2 = 87{,}25 - 22{,}5625 = 64{,}6875 .$$

Результаты вычислений по обоим способам совпадают.

в) Среднее квадратичное отклонение:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{64{,}6875} \approx 8{,}04285.$$

Таким образом, $\boldsymbol{\sigma}$ = 8,04285 тыс. рублей – характеристика разброса фактических значений выигрыша от найденного среднего значения a = 4,75 тыс. рублей, то есть основные значения случайной величины выигрыша находятся в диапазоне (4,75 ± 8,04285) тыс. руб., что соответствует таблице данных.

Тема 4. Нормальный закон распределения непрерывной случайной величины

Непрерывная случайная величина X может принимать любые значения из некоторого промежутка. Распределение вероятностей ее значений на этом промежутке задается дифференциальной функцией распределения $f(x)$. Эта функция называется также функцией плотности распределения вероятностей.

Исследования показали, что в большом числе встречающихся на практике случаев с достаточным основанием можно считать, что случайные величины подчиняются нормальному закону распределения. Дифференциальная функция нормального распределения имеет вид

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\cdot\sigma^2}} ,$$

где $\boldsymbol{a}$ – математическое ожидание; $\boldsymbol{\sigma}$ – среднее квадратическое отклонение случайной величины X.

Графиком этой функции является кривая Гаусса, которая наглядно показывает, как распределена вероятность между возможными значениями x.

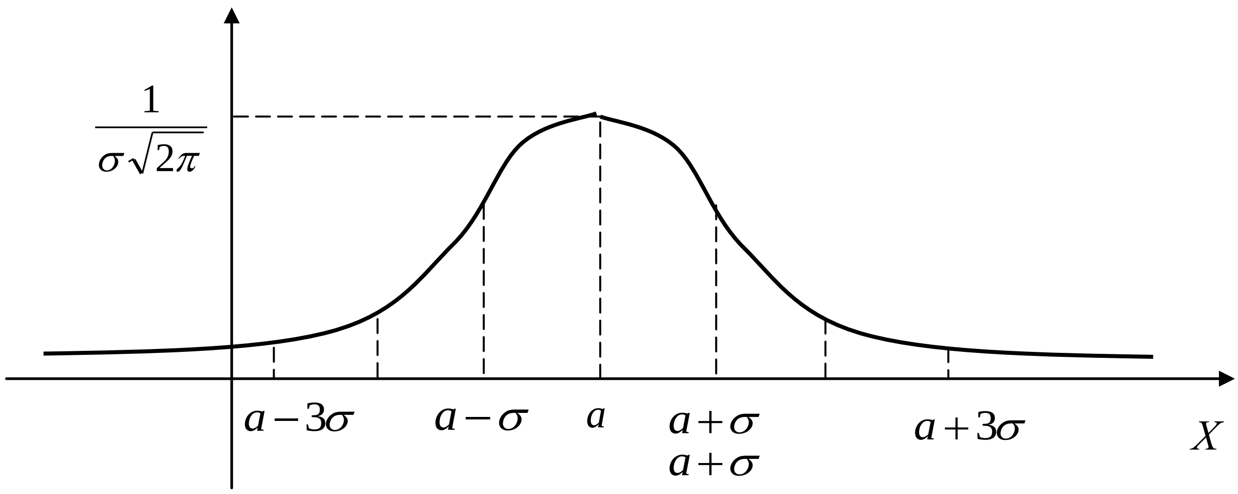

Рис.1

Особенности нормального распределения

а) Наиболее вероятны значения ***x***, близкие к ожидаемому среднему значению ***a.***

б) Отклонения от среднего значения ***a*** в обе стороны равновероятны.

в) Большие отклонения ***x*** от среднего значения ***a*** маловероятны.

Площадь под кривой Гаусса всегда равна единице, что соответствует полной вероятности. Поэтому при уменьшении $\boldsymbol{\sigma}$ увеличивается вероятность значений, близких к ***a***, рассеяние уменьшается, кривая Гаусса сжимается. При увеличении $\boldsymbol{\sigma}$ график кривой Гаусса становится более расплывчатым, что говорит об увеличении рассеяния.

Вероятность попадания нормально распределенной случайной величины X в интервал $(\alpha\ ;\ \beta)$ находят по формуле

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Вероятность того, что абсолютная величина отклонения «*X–a*» меньше δ, составляет:

$$P(|X - a| < \delta) = 2 \cdot \Phi(\delta/\sigma).$$

В частности, если δ=3σ, получим:

$$P(\left| X - a \right| < 3\sigma) = 2 \cdot \Phi(3\sigma/\sigma) = 2\Phi(3) = 2 \cdot 0{,}49865 \approx 0{,}9973\ (99{,}7\,\%).$$

Этот результат называют правилом «трех сигм»: почти достоверно (на 99,7%), что значения нормально распределенной случайной величины отличаются от своего среднего значения ***a*** менее, чем на 3*σ*, то есть практически все значения нормально распределенной случайной величины *X* попадают в интервал: (*a* - 3σ; *a*+3σ), см. рис. 1.

Задачи 51–60

Пример 6. Заданы математическое ожидание ***a*** = 2 и среднее квадратическое отклонение ***σ*** =5 нормально распределенной случайной величины *X*. Требуется найти:

а) вероятность того, что *X* примет значение, принадлежащее интервалу (1; 4);

б) вероятность того, что абсолютная величина отклонения «*X–a*» окажется меньше ***δ*** = 3.

Решение.

а) Воспользуемся формулой

$$P\left(\alpha < X < \beta\right) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

По условию задачи ***a***=2, ***α*** = 1, ***β*** = 4, ***σ*** = 5. Следовательно,

$$P\left(1 < X < 4\right) = \Phi\left(\frac{4-2}{5}\right) - \Phi\left(\frac{1-2}{5}\right) = \Phi(0{,}4) - \Phi(-0{,}2) = \Phi(0{,}4) + \Phi(0{,}2).$$

По таблице приложения 2: Ф(0,4) = 0,1554; Ф(0,2) = 0,0793, поэтому искомая вероятность попадания нормально распределенной случайной величины в интервал (1; 4) составит:

$$P(1 < X < 4) = 0{,}1554 + 0{,}0793 = 0{,}2347.$$

б) Вероятность того, что абсолютная величина отклонения «*X–a*» меньше δ = 3, равна: $P\left(\left|X - a\right| < \delta\right) = 2 \cdot \Phi\left(\delta / \sigma\right)$,

$$P\left(\left|X - 2\right| < 3\right) = 2 \cdot \Phi\left(3/5\right) = 2 \cdot \Phi(0{,}6) = 2 \cdot 0{,}2257 = 0{,}4514.$$

Это вероятность попадания случайной величины случайной величины *X* в промежуток (2 ± 3), то есть от –1 до 5.

Полученные результаты поясним графически с помощью кривой Гаусса для данного распределения:

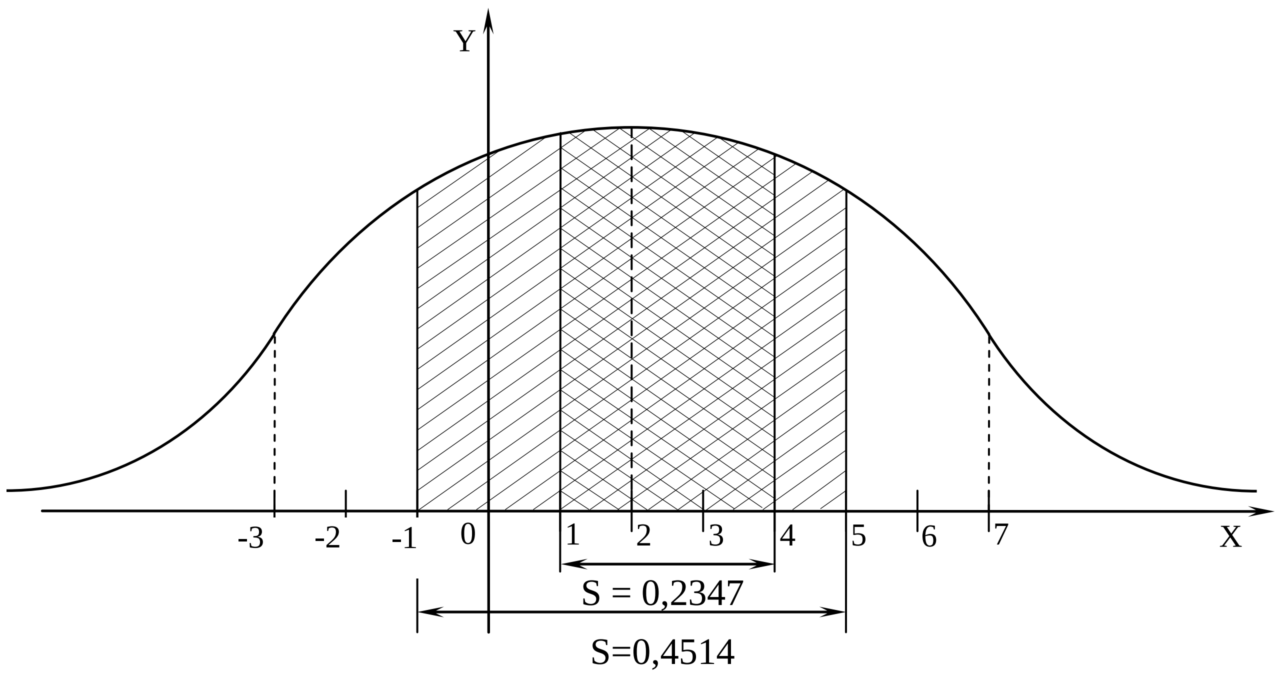

Рис. 2

Вероятности попадания в промежутки равны соответствующим площадям под кривой Гаусса. Напомним, что полная площадь под кривой равна 1.

Ответы: а) 0,2347; б) 0,4514.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К РЕШЕНИЮ ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Математическая статистика базируется на теории вероятностей и является теоретической основой всей статистики. Ее задачей является создание способов сбора и методов обработки статистической информации.

Тема 5. Статистическое распределение выборки и его основные числовые характеристики

Выборочный метод – один из основных методов математической статистики. Его сущность заключается в том, что изучение большой совокупности объектов относительно некоторого количественного признака X производится по сравнительно небольшому числу случайно отобранных объектов.

Генеральной совокупностью называется множество всех изучаемых объектов, из которых производится выборка.

Выборочной совокупностью (выборкой) называется множество объектов, отобранных для изучения из генеральной совокупности.

Выборка должна быть организована случайным образом, чтобы правильно представлять генеральную совокупность.

Объемом совокупности называется количество объектов в совокупности. Объем выборки $\boldsymbol{n}$, как правило, значительно меньше объема $\boldsymbol{N}$ генеральной совокупности: $\boldsymbol{n} \ll \boldsymbol{N}$.

Данные выборки записываются в виде таблицы, называемой статистическим распределением выборки:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

В первой строке перечислены все наблюдаемые значения признака X в порядке их возрастания (или убывания). Они называются вариантами x_i ($i = 1, 2, 3, \ldots, k$). Во второй строке указаны частоты n_i соответствующих вариант x_i. Они показывают, сколько раз наблюдалось каждое значение признака X.

Очевидно, что сумма всех частот n_i равна объему выборки n:

$$n_1 + n_2 + \ldots + n_k = \sum_{i=1}^{k} n_i = n.$$

Основные числовые характеристики выборки

1. **Средняя выборочная** (среднее взвешенное значение признака в выборке):

$$\bar{x}_в = \frac{\sum_{i=1}^{к} x_i n_i}{n} = \frac{1}{n}\cdot(x_1 \cdot n_1 + x_2 n_2 + \ldots + x_k n_k).$$

2. **Дисперсия выборочная**. Характеризует разброс (рассеяние) значений вариант $\boldsymbol{x_i}$ от выборочного среднего значения $\overline{x_в}$ и измеряется в квадратных единицах признака X:

$$D_в = \frac{1}{n}\cdot\sum_{i=1}^{k}(x_i - \bar{x}_в)^2 \cdot n_i = \frac{1}{n}\left[(x_1 - \bar{x}_в)^2 n_1 + (x_2 - \bar{x}_в)^2 \cdot n_2 + \ldots + (x_к - \bar{x}_в)^2 n_k\right]$$

Для вычисления дисперсии используется также другая, часто более удобная формула

$$D_в = \overline{x_в^2} - (\bar{x}_в)^2,$$

где

$$\overline{x_в} = \frac{1}{n}\cdot\sum_{i=1}^{k} x_i n_i; \qquad \overline{x_в^2} = \frac{1}{n}\cdot\sum_{i=1}^{k} x_i^2 n_i.$$

3. **Среднее квадратическое отклонение выборки** – характеристика рассеяния значений признака в выборке от среднего выборочного в единицах признака *X:*

$$\sigma_в = \sqrt{D_в}.$$

С помощью найденных выборочных характеристик $x_в, D_в, \sigma_в$ оцениваются соответствующие генеральные характеристики

$\bar{x}$ – генеральная средняя;

D – генеральная дисперсия;

σ – генеральное среднее квадратическое отклонение.

Оценки имеют следующий вид:

$$\bar{x} \approx \overline{x_в}; \qquad D \approx \frac{n}{n-1}\cdot D_в = S_в^2; \qquad \sigma \approx S_в = \sqrt{\frac{n}{n-1}D_в},$$

где $S_в^2$ - так называемая исправленная выборочная дисперсия.

Приведенные оценки носят случайный характер, так как зависят от выборки. Они называются точечными и удовлетворяют следующим требованиям:

- ***несмещенность*** (отсутствие систематических ошибок);
- ***состоятельность*** (увеличение объема выборки повышает вероятность правильности оценки);
- ***эффективность*** (имеют самый незначительный разброс по сравнению с другими возможными оценками).

Основные характеристики выборки $\overline{x}_в, D_в, \sigma_в$ лишь приближенно характеризуют генеральную совокупность и могут оказаться далекими от соответствующих характеристик генеральной совокупности: $\overline{x} = a$, D, σ. Поэтому для последних используют интервальные оценки, когда неизвестная характеристика заключена в некотором интервале с заданной надежностью (вероятностью) **γ.** Такой интервал называется доверительным. Значения надежности берутся, как правило, высокими: 0,9; 0,95; 0,99 или 0,999, что соответствует 90; 95; 99 или 99,9%.

Если количественный признак X в генеральной совокупности распределен по нормальному закону, причем среднее квадратическое отклонение $\boldsymbol{\sigma}$ этого распределения известно, то с вероятностью $\boldsymbol{\gamma}$ доверительный интервал, заданный выражением:

$$(\overline{x_в} - t\cdot\sigma/\sqrt{n};\quad \overline{x_в} + t\cdot\sigma/\sqrt{n}),$$

покрывает неизвестное математическое ожидание $\boldsymbol{a}$. Здесь параметр t находится из соотношения $2\cdot\Phi(t) = \gamma$ с помощью таблицы значений для интегральной функции Лапласа (приложение 2).

Часто статистическое распределение выборки носит интервальный характер. В этом случае указывают числовые частичные интервалы, куда попадают значения признака X, и $\boldsymbol{n_i}$ – количество значений, попавших в интервал с номером i. В качестве значений $\boldsymbol{x_i}$ выбирают середины частичных интервалов.

Значения $\boldsymbol{n_i}$ называются абсолютными частотами, их сумма равна объему выборки: $\sum n_i = n$. Относительные частоты $w_i = \dfrac{n_i}{n}$

показывают долю значений x_i в общем объеме выборки. Очевидно, что сумма всех относительных частот (долей) равна 1: $\sum w_i = 1$.

Графически дискретное статистическое распределение изображается в виде **полигона** частот, обычно относительных. **Полигон** представляет собой ломаную линию, соединяющую соседние точки с координатами (x_i ; w_i).

Интервальное статистическое распределение изображается на графике в виде **гистограммы** относительных частот. **Гистограмма** – это ступенчатая фигура, состоящая из прямоугольников. В основании каждого прямоугольника лежит частичный интервал, а высота прямоугольника определяется относительной частотой w_i , а чаще величиной $\dfrac{w_i}{h_i}$, где h_i – длина частичного интервала. При таком построении площадь каждого частичного прямоугольника равна относительной частоте w_i , а сумма всех площадей, то есть площадь ступенчатой фигуры, равна единице: $\sum w_i = 1$.

Задачи 61–70

Пример 7. В результате выборочного наблюдения за вкладами клиентов банка получено следующее распределение клиентов по величине вклада X в тыс. руб.:

X	До 100	100-200	200-300	300-400	400-500
n_i	10	18	20	32	28

где n_i – количество клиентов с величиной вклада в заданном интервале.

а) Изобразить данное распределение графически, построив гистограмму относительных частот.

б) Найти основные характеристики выборки: среднее значение, дисперсию и среднее квадратическое отклонение.

в) Оценить генеральные характеристики по найденным выборочным характеристикам.

г) С надежностью 95% указать доверительный интервал для генеральной средней, приняв гипотезу о нормальном распределении признака *X*.

Решение. Найдем объем выборки ***n***:

$$n = \sum n_i = 10 + 18 + 20 + 32 + 28 = 108,$$

то есть для обследования выбрано 108 клиентов.

а) Вычислим относительные частоты для каждого частичного интервала:

$$w_1 = \frac{n_1}{n} = \frac{10}{108} = 0{,}093; \qquad w_2 = \frac{n_2}{n} = \frac{18}{108} = 0{,}167;$$

$$w_3 = \frac{n_3}{n} = \frac{20}{108} = 0{,}185; \qquad w_4 = \frac{n_4}{n} = \frac{32}{108} = 0{,}296;$$

$$w_5 = \frac{n_5}{n} = \frac{28}{108} = 0{,}259\,.$$

Рекомендуем все вычисления вести с точностью до 0,001.

Контроль: $\sum w_i = 0{,}093 + 0{,}167 + 0{,}185 + 0{,}296 + 0{,}259 = 1$.

В итоге получено следующее интервальное распределение относительных частот признака *X:*

X	0-100	100-200	200-300	300-400	400-500
w_i	0,093	0,167	0,185	0,296	0,259

Шаг разбиения, то есть длина каждого частичного интервала $h = 100$.

Построим гистограмму относительных частот (рис. 3), откладывая по оси *OX* значения признака *X*, а по вертикальной оси значения $\frac{w_i}{h} = \frac{w_i}{100}$.

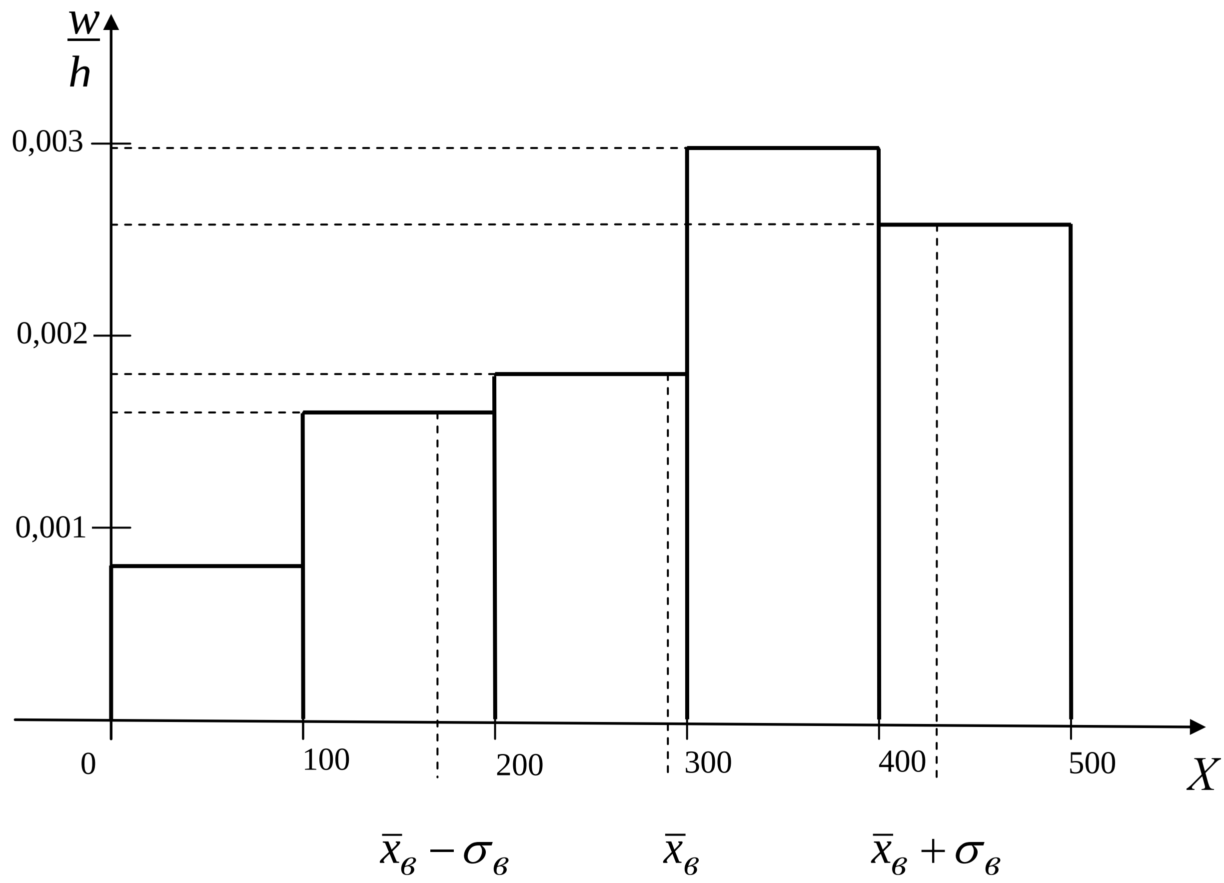

Рис. 3

б) Для нахождения характеристик выборки от интервального распределения признака X перейдем к дискретному, выбирая в качестве значений признака $\boldsymbol{x_i}$ середины частичных интервалов:

$\boldsymbol{x_i}$	50	150	250	350	450
$\boldsymbol{n_i}$	10	18	20	32	28

Найдем основные характеристики этого распределения.

Средняя выборочная (средняя величина вклада в тыс. рублей):

$$\overline{x_в} = \frac{1}{n} \cdot \sum x_i n_i = \frac{1}{108}(50 \cdot 10 + 150 \cdot 18 + 250 \cdot 20 + 350 \cdot 32 + 450 \cdot 28) =$$

$$= \frac{1}{108}(500 + 2700 + 5000 + 11200 + 12600) = \frac{1}{108} \cdot 32000 \approx 296{,}296.$$

Выборочная дисперсия:

$$D_в = \frac{1}{n}\sum (x_i - \overline{x_в})^2 \cdot n_i = \frac{1}{108}\cdot\big[(50-296{,}296)^2\cdot 10 + (150-296{,}296)^2\cdot 18 +$$

$$+(250-296{,}296)^2\cdot 20 + (350-296{,}296)^2\cdot 32 + (450-296{,}296)^2\cdot 28\big] =$$

$$= \frac{1}{108}(606617{,}196 + 385245{,}353 + 42866{,}392 + 92291{,}828 +$$

$$+ 661497{,}749) = \frac{1}{108}\cdot 1788518{,}518 = 16560{,}357.$$

Второй способ вычисления дисперсии.

Найдем среднюю квадратов значений признака:

$$\overline{x_в^2} = \frac{1}{n}\cdot\sum x_i^2 n_i = \frac{1}{108}(50^2\cdot 10 + 150^2\cdot 18 + 250^2\cdot 20 + 350^2\cdot 32 + 450^2\cdot 28) =$$

$$= \frac{1}{108}(25000 + 405000 + 1250000 + 3920000 + 5670000) =$$

$$= \frac{1}{108}\cdot 11270000 \approx 104351{,}852.$$

$$D_в = \overline{x_в^2} - (\overline{x_в})^2 = 104351{,}852 - (296{,}296)^2 =$$

$$= 104351{,}852 - 87791{,}495 = 16560{,}357.$$

Этот результат совпадает с результатом первого способа (иногда приближенно из-за округлений).

Среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в} = \sqrt{16560{,}357} \approx 128{,}687,$$

то есть, в среднем разброс вкладов составляет $\pm$ 128,687 тыс. рублей от среднего значения 296,296 тыс. рублей (см. рис. 3, пунктирные вертикальные линии).

в) Оценим неизвестные генеральные характеристики:

генеральная средняя: $\overline{x} \approx \overline{x_в} = 296{,}296$ тыс. рублей;

генеральная дисперсия:

$$D \approx \frac{n}{n-1} D_в = \frac{108}{108-1}\cdot 16560{,}357 \approx 16715{,}127;$$

генеральное среднее квадратическое отклонение:

$$\sigma = \sqrt{D} = \sqrt{16715{,}127} \approx 129{,}287 \text{ (тыс. рублей).}$$

г) Доверительный интервал для оценки генеральной средней $\boldsymbol{a}$ (среднего вклада) с надежностью $\boldsymbol{\gamma}$ находим по формуле

$$a \in (\overline{x_в} - t \cdot \sigma / \sqrt{n}; \quad \overline{x_в} + t \cdot \sigma / \sqrt{n}).$$

По условию задачи $n = 108$, $\overline{x_в} = 296{,}296$; $\sigma = 129{,}287$, $\gamma = 0{,}95$. Неизвестный параметр t находим из условия: $2\Phi(t) = \gamma$. Поскольку в данной задаче $\gamma = 0{,}95$, то есть $2\Phi(t) = 0{,}95$, то $\Phi(t) = 0{,}475$. Из таблицы (приложение 2) берем соответствующее значение t: $t = 1{,}96$.

Вычислим по этим данным доверительный интервал:

$$(296{,}296 - 1{,}96 \cdot \frac{129{,}287}{\sqrt{108}}; \quad 296{,}296 + 1{,}96 \cdot \frac{129{,}287}{\sqrt{108}}),$$

$$(296{,}296 - 24{,}384; \quad 296{,}296 + 24{,}384),$$

$$(271{,}912; \; 320{,}680).$$

Таким образом, с вероятностью 95% неизвестная генеральная средняя (математическое ожидание) находится в этом интервале:

$$\overline{x} = a \in (271{,}912; \quad 320{,}680).$$

Замечание. Доверительный интервал можно изобразить графически.

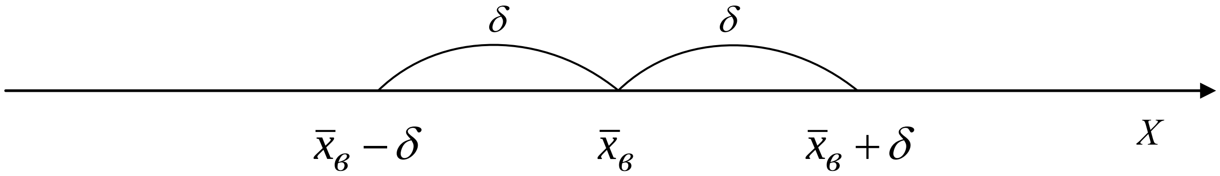

Рис. 4

Длина полуинтервала $\delta = t \cdot \sigma / \sqrt{n} = 24{,}384$ характеризует точность оценки. Величина $\boldsymbol{\delta}$ называется предельной ошибкой оценки. Оценка тем точнее, чем меньше $\boldsymbol{\delta}$ и, следовательно, доверительный интервал становится более узким. Величина $\boldsymbol{\delta}$ зависит от $\boldsymbol{n}$, $\boldsymbol{\sigma}$ и $\boldsymbol{t}$. Очевидно, с увеличением объема выборки $\boldsymbol{n}$ уменьшается $\boldsymbol{\delta}$ и повышается точность оценки. При увеличении рассеяния $\boldsymbol{\sigma}$ длина интервала $\boldsymbol{\delta}$ увеличивается, то есть оценка делается менее точной.

Увеличение надежности $\boldsymbol{\gamma}$ ведет к росту вспомогательного параметра $\boldsymbol{t}$ и расширению доверительного интервала (надежнее попасть в большой интервал). Это делает оценку менее точной. Таким образом, при повышении надежности оценки ухудшается ее точность.

Тема 6. Элементы теории корреляции

Пусть каждый из выбранных объектов характеризуется двумя количественными признаками X и Y. Между значениями этих признаков может существовать некоторая зависимость.

Функциональная зависимость – это такая зависимость, когда каждому значению x признака X соответствует единственное значение y признака Y. Эта зависимость является вполне определенной, однозначной и называется строгой (детерминированной). Она задается в виде функции $\boldsymbol{y = f(x)}$.

Статистическая зависимость – это такая зависимость, когда каждому значению x признака X соответствует статистическое распределение значений признака Y. Эта зависимость не является строгой и носит вероятностный (стохастический) характер, поскольку на величину признака $\boldsymbol{Y}$ влияют не только значения признака $\boldsymbol{X}$, но и другие случайные факторы.

Если случайные величины $\boldsymbol{X}$ и $\boldsymbol{Y}$ не являются взаимно независимыми, то в той или иной степени им свойственна стохастическая зависимость.

Корреляционная зависимость – это статистическая зависимость, обладающая тем свойством, что изменение значений x признака X приводит к изменению среднего значения признака Y, обозначаемого $\overline{y}_x$. Связь между x и условной средней $\overline{y}_x$ задаётся с помощью функции $f(x)$ и записывается в виде уравнения $\overline{y}_x = f(x)$, которое называется уравнением регрессии Y по X.

Аналогично, связь между значениями y признака Y и соответствующими условными средними значениями $\overline{x}_y$ записывается в виде уравнения $\overline{x}_y = \varphi(y)$, которое называется уравнением регрессии X по Y.

Практически наличие корреляционной связи между признаками $\boldsymbol{X}$ и $\boldsymbol{Y}$ прослеживается как изменение средних значений одного признака при изменении значений другого, причем эта связь может проявляться с различной степенью силы. Например, имеется корреляционная зависимость между ростом людей ***X*** и их весом ***Y***; между количеством внесенных удобрений ***X*** и урожайностью ***Y***; между успеваемостью студентов по математике в школе и в вузе и т. п.

Основные задачи теории корреляции состоят в том, чтобы по данным выборки:

1) оценить силу (тесноту) связи между признаками ***X*** и ***Y***;
2) найти вид (форму) этой связи в виде уравнения регрессии.

Уравнение регрессии выбирают по возможности простым, и оно, как правило, лишь приближенно описывает зависимость между значениями *x* одного признака и соответствующими средними значениями другого признака $\overline{y}_x$.

Наиболее простой и употребляемый вид зависимости – линейная зависимость. Она определяется уравнением линейной регрессии $\overline{y}_x = ax + b$ и изображается на графике в виде прямой регрессии. Уравнение регрессии называется выборочным, поскольку его параметры *a* и *b* находятся по результатам выборки (x_i, y_i), $i=1,2,\ldots n$, причем наилучшим образом в смысле метода наименьших квадратов. Сущность метода заключается в том, чтобы была наименьшей сумма квадратов отклонений наблюдаемых значений y_i от соответствующих значений $\overline{y_{x_i}}$, вычисленных по уравнению регрессии $\overline{y_{x_i}} = ax_i + b$, то есть $\sum_{i=1}^{n}(ax_i + b - y_i)^2 \to min.$

Проблема статистического исследования зависимостей является главной в решении многих типовых задач практики, таких как планирование, прогнозирование, нормирование, оценка эффективности функционирования или качества объекта, анализ систем и прочее.

Задачи 71–80

Пример 8. По данным корреляционной таблицы 1 найти условные средние $\overline{y}_x$ и $\overline{x}_y$. Оценить тесноту линейной связи между признаками *X* и *Y* и составить уравнения линейной регрессии *Y* по *X* и *X* по *Y*. Сделать чертеж, нанеся на него условные средние и найденные прямые регрессии. Оценить тесноту связи между признаками с помощью корреляционного отношения.

Таблица 1

$Y \diagdown X$	15	20	25	30	35	40	n_y
49				14	14	9	37
59			7	13	6		26
69			10	40	2		52
79	8	4	3				15
n_x	8	4	20	67	22	9	130

Решение. В таблице 1 приведены данные выборочных наблюдений за 130 объектами, обладающими признаками ***X*** и ***Y***. Каждому объекту соответствует пара значений (x, y), а частота n_{xy} показывает количество объектов с такой парой значений признаков.

Все возможные значения признака ***X*** перечислены в верхней горизонтальной строке таблицы 1, а для признака ***Y*** – в первом вертикальном столбце. В клетках на пересечении каждой строки и каждого столбца проставлена частота n_{xy}, с которой наблюдается каждая пара значений.

Например: пара значений (15;79) наблюдалась 8 раз, пара значений (30;69) наблюдалась 40 раз и т.д.

Пустые клетки означают, что соответствующие им пары значений не наблюдались.

В нижней итоговой строке данной таблицы напротив каждого значения признака X проставляется соответствующая ему частота n_x , равная сумме всех частот столбца и указывающая, сколько раз всего наблюдалось данное значение ***x***. Аналогично, в последнем итоговом столбце напротив каждого значения ***y*** записывают соответствующую ему частоту n_y, равную сумме частот по строке и указывающую, сколько раз всего наблюдалось данное значение ***y***. Очевидно, что суммы всех частот для n_x и для n_y должны быть равны между собой и равны объему выборки (количеству наблюдаемых пар):

$$\sum_i n_x = \sum_j n_y = \sum_i \sum_j n_{xy} = n .$$

Объем выборки **n** проставляется в последней клетке таблицы. В данной задаче **n** = 130. Решая аналогичную задачу из контрольной работы, необходимо самостоятельно заполнить последнюю строку и последний столбец.

В таблице 1 каждому значению *X* соответствует статистическое распределение признака *Y*.

Например, для $x = 30$:

Y	49	59	69	79
n_{xy}	14	13	40	-

Отсюда находим среднее значение y при условии, что $x = 30$, или условную среднюю:

$$\bar{y}_{x=30} = \frac{49 \cdot 14 + 59 \cdot 13 + 69 \cdot 40}{14 + 13 + 40} = 62{,}86.$$

Аналогично, каждому значению ***y*** соответствует статистическое распределение *X*. Например, для ***y*** = 49:

x	15	20	25	30	35	40
n_{xy}	-	-	-	14	14	9

Отсюда находим условную среднюю:

$$\bar{x}_{y=49} = \frac{30 \cdot 14 + 35 \cdot 14 + 40 \cdot 9}{14 + 14 + 9} = 34{,}32.$$

Не выписывая далее статистических распределений, а беря их непосредственно из данной корреляционной таблицы 1, найдем все условные средние по формулам

$$\boxed{\bar{y}_x = \frac{\sum y_j n_{xy_j}}{n_x}} \qquad \boxed{\bar{x}_y = \frac{\sum x_i n_{x_i y}}{n_y}}$$

$$\bar{y}_{x=15} = \frac{79 \cdot 8}{8} = 79; \qquad \bar{x}_{y=49} = \frac{30 \cdot 14 + 35 \cdot 14 + 40 \cdot 9}{37} = 34{,}32;$$

$$\bar{y}_{x=20} = \frac{79 \cdot 4}{4} = 79; \qquad \bar{x}_{y=59} = \frac{25 \cdot 7 + 30 \cdot 13 + 35 \cdot 6}{26} = 29{,}81;$$

$$\bar{y}_{x=25} = \frac{59 \cdot 7 + 69 \cdot 10 + 79 \cdot 3}{20} = 67; \qquad \bar{x}_{y=69} = \frac{25 \cdot 10 + 30 \cdot 40 + 35 \cdot 2}{52} = 29{,}23;$$

$$\bar{y}_{x=30} = \frac{49 \cdot 14 + 59 \cdot 13 + 69 \cdot 40}{67} = 62{,}88; \qquad \bar{x}_{y=79} = \frac{15 \cdot 8 + 20 \cdot 4 + 25 \cdot 3}{15} = 18{,}33;$$

$$\bar{y}_{x=35} = \frac{49 \cdot 14 + 59 \cdot 6 + 69 \cdot 2}{22} = 53{,}55;$$

$$\bar{y}_{x=40} = \frac{49 \cdot 9}{9} = 49.$$

Оценка тесноты линейной связи между признаками X и Y производится с помощью коэффициента линейной корреляции ***r***:

$$\boxed{r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}}$$

Коэффициент ***r*** может принимать значения от -1 до $+1$, то есть

$$\boxed{-1 \le r \le 1} \quad \text{или} \quad \boxed{|\,r\,| \le 1}$$

Знак ***r*** указывает на направление связи: прямая или обратная. Абсолютная величина $|\,r|$ указывает на силу (тесноту) связи.

Оценка тесноты линейной связи (шкала Чаддока)

Значение $\lvert r \rvert$	0-0,1	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99	1
Теснота линейной связи	Нет связи	Слабая	Умеренная	Заметная	Высокая	Очень высокая	Функциональная

При $r > 0$ связь **прямая**, то есть с ростом ***x*** растет ***y***.

При $r < 0$ связь **обратная**, то есть с ростом ***x*** убывает ***y***.

Для нахождения ***r*** вычислим указанные общие средние: $\bar{x}$, $\bar{y}$, $\overline{xy}$, а также средние квадратические отклонения σ_x и σ_y. Вычисления удобно поместить в таблицы 2 и 3, куда вписываем также найденные ранее условные средние.

Таблица 2

x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$	$\bar{y}_x$	$x \cdot n_x \cdot \bar{y}_x$
15	8	12	1800	79	9480
20	4	80	1600	79	6320
25	20	500	12500	67	33500
30	67	2010	60300	62,88	126390
35	22	770	26950	53,55	41230
40	9	360	14400	49	17640
$\sum$	130	3840	117550	-	234560

Таблица 3

y	n_y	$y \cdot n_y$	$y^2 \cdot n_y$	$\bar{x}_y$	$y \cdot n_y \cdot \bar{x}_y$
49	37	1813	88837	34,32	62230
59	26	1534	90506	29,81	45725
69	52	3588	247572	29,23	104880
79	15	1185	93615	18,33	21725
$\sum$	130	8120	520530	-	234560

Контроль: $\sum x \cdot n_x \cdot \bar{y}_x = \sum y \cdot n_y \cdot \bar{x}_y$.

В рассматриваемой задаче эта сумма в обеих таблицах равна 234560. Равенство может оказаться приближенным, что связано с приближенными вычислениями условных средних $\bar{y}_x$, $\bar{x}_y$.

С помощью таблиц 2 и 3 находим общие средние, средние квадратов, среднюю произведения и средние квадратические отклонения:

$$\bar{x} = \frac{\sum x \cdot n_x}{n} = \frac{3840}{130} = 29{,}54;$$

$$\overline{x^2} = \frac{\sum x^2 \cdot n_x}{n} = \frac{117550}{130} = 904{,}23;$$

$$\overline{xy} = \frac{\sum x \cdot n_x \cdot \overline{y_x}}{n} = \frac{234560}{130} = 1804{,}31;$$

$$\bar{y} = \frac{\sum y \cdot n_y}{n} = \frac{8120}{130} = 62{,}46;$$

$$\overline{y^2} = \frac{\sum y^2 \cdot n_y}{n} = \frac{520530}{130} = 4004{,}08;$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{904{,}23 - 29{,}54^2} = \sqrt{31{,}62} \approx 5{,}62;$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{4004{,}08 - 62{,}46^2} = \sqrt{102{,}83} \approx 10{,}14\ .$$

Отсюда коэффициент корреляции равен:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{1804{,}31 - 29{,}54 \cdot 62{,}46}{5{,}62 \cdot 10{,}14} = -\frac{40{,}76}{56{,}99} = -0{,}72\,.$$

Так как $\boldsymbol{r} < 0$, то связь обратная, то есть с ростом x убывает y.

Так как $|r| = 0{,}72$, то по шкале Чаддока определяем, что сила линейной связи высокая.

Находим линейное уравнение регрессии Y по X:

$$\boxed{\bar{y}_x - \bar{y} = r \cdot \frac{\sigma_x}{\sigma_y} \cdot (x - \bar{x})}$$

$$\bar{y}_x - 62{,}46 = -0{,}72 \cdot \frac{10{,}14}{5{,}62} \cdot (x - 29{,}54);$$

$$\bar{y}_x - 62{,}46 = -1{,}30(x - 29{,}54);$$

$$\bar{y}_x = -1{,}30 \cdot x + 100{,}83 \cdot$$

Аналогично находим линейное уравнение регрессии X по Y:

$$\boxed{\overline{x}_y - \overline{x} = r \cdot \frac{\sigma_x}{\sigma_y} \cdot (y - \overline{y})}$$

$$\overline{x}_y - 29{,}54 = -0{,}72 \cdot \frac{5{,}62}{10{,}14} \cdot (y - 62{,}46),$$

$$\overline{x}_y - 29{,}54 = -0{,}40 \cdot (y - 62{,}46),$$

$$\overline{x}_y = -0{,}40 \cdot y + 54{,}46.$$

Данные уравнения устанавливают связь между признаками X и Y и позволяют найти среднее значение признака $\overline{y}_x$ для каждого значения x и аналогично, среднее значение признака $\overline{x}_y$ для каждого значения y.

Изобразим полученные результаты графически (рис. 5). Нанесем на график точки $(x; \overline{y}_x)$, (их удобно взять из таблицы 2), отметив звездочками (*). Нанесем на график точки $(\overline{x}_y; y)$ (их удобно взять из таблицы 3), отметив кружочками (∘). Построим каждое из найденных уравнений регрессии по двум точкам:

$y = -1{,}30x + 100{,}83$

x	15	40
y	81,33	48,83

$x = -0{,}40y + 54{,}46$

x	34,86	22,86
y	49	79

Значения $x = 15$ и $x = 40$ для первого уравнения берем из таблицы 2, как крайние значения для признака X.

Аналогично $y = 49$ и $y = 79$ для второго уравнения берем из таблицы 3.

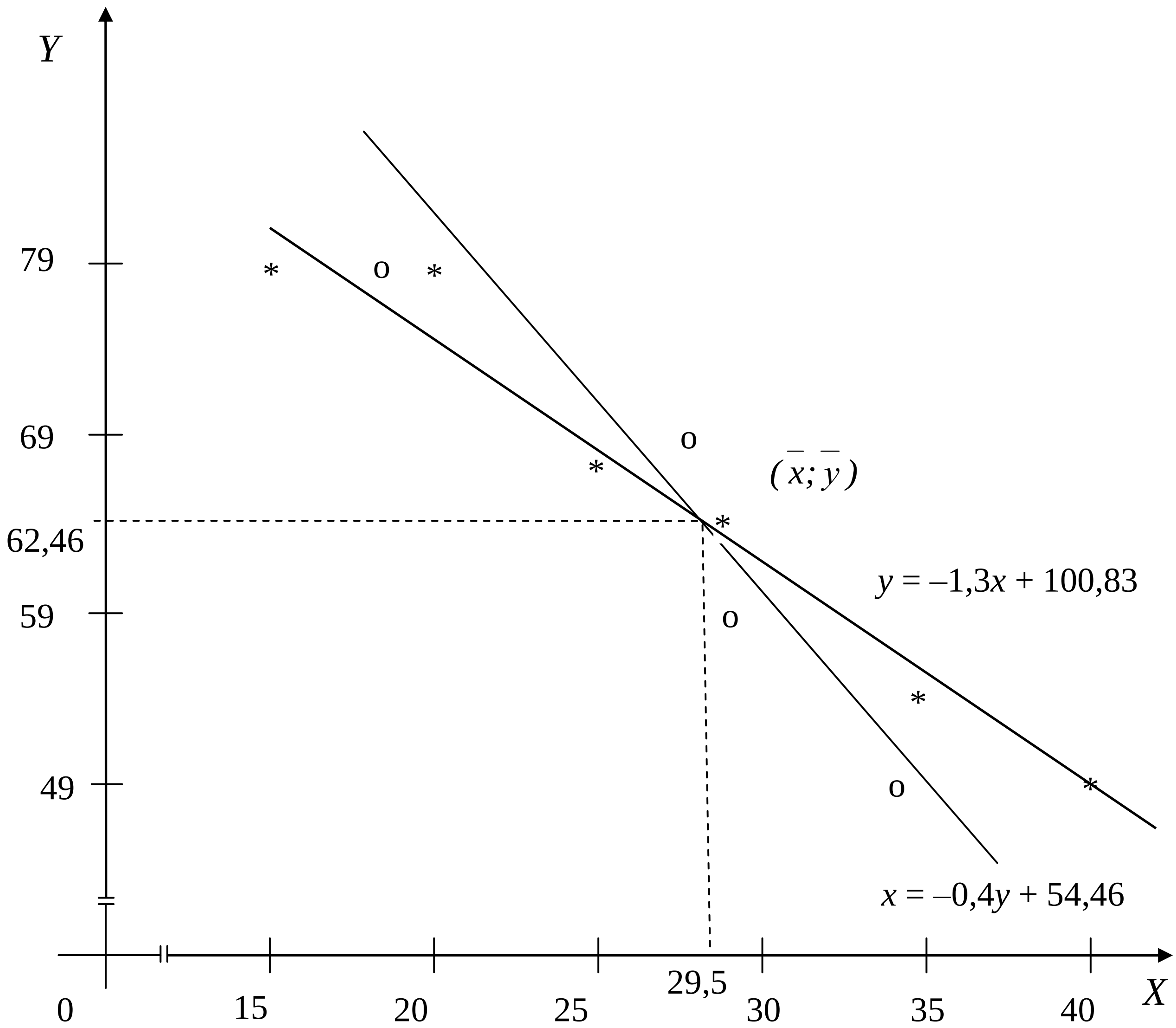

Рис. 5

Замечание. Если расчеты и построения выполнены верно, то прямая регрессии Y по X должна пройти вблизи всех точек ($x; \bar{y}_x$), отмеченных звездочками (*), так что эти точки расположены по обе стороны от прямой. Аналогично прямая регрессии X по Y наилучшим образом приближена к точкам ($\bar{x}_y; y$), отмеченным кружочками ($\circ$).

Обе прямые регрессии пересекаются в точке ($\bar{x}; \bar{y}$). В нашей задаче это точка (29,54; 62,46). Чем выше сила связи между признаками, тем ближе расположены обе прямые друг к другу (угол между ними мал).

Если линейная связь слабая, то это не исключает наличия между признаками *X* и *Y* нелинейной (криволинейной) связи. Оценка тесноты любой связи между признаками (линейной и нелинейной) производится с помощью корреляционных отношений *Y* по *X* и *X* по *Y*:

$$\eta_{yx} = \frac{\delta_y}{\sigma_y} \quad ; \qquad \eta_{xy} = \frac{\delta_x}{\sigma_x} \quad .$$

Дисперсии σ_x^2, σ_y^2, называемые внутригрупповыми, определены ранее. Их можно было также посчитать по формулам:

$$\sigma_x^2 = \frac{1}{n} \cdot \sum (x - \bar{x})^2 \cdot n_x \ ,$$

$$\sigma_y^2 = \frac{1}{n} \cdot \sum (y - \bar{y})^2 \cdot n_y \ .$$

Они характеризуют разброс фактических значений ***x*** и ***y*** от общих средних $\bar{x}$ и $\bar{y}$.

Величины δ_x^2, δ_y^2 называются межгрупповыми дисперсиями и вычисляются по формулам:

$$\delta_x{}^2 = \frac{1}{n} \cdot \sum (\bar{x}_y - \bar{x})^2 \cdot n_y \ ;$$

$$\delta_y{}^2 = \frac{1}{n} \cdot \sum (\bar{y}_x - \bar{y})^2 \cdot n_x \ .$$

Они характеризуют разброс условных средних $\bar{x}_y$ и $\bar{y}_x$ от общих средних $\bar{x}$ и $\bar{y}$.

В рассматриваемой задаче, используя данные из расчётных таблиц 3 и 2, находим:

$$\delta_x{}^2 = \frac{1}{130} \cdot \left[(34{,}32 - 29{,}54)^2 \cdot 37 + (29{,}81 - 29{,}54)^2 \cdot 26 + \right.$$
$$\left. + (29{,}23 - 29{,}54)^2 \cdot 52 + (18{,}33 - 29{,}54)^2 \cdot 15 \right] \approx 21{,}06 \ ;$$

$$\delta_y{}^2 = \frac{1}{130} \cdot \left[(79 - 62{,}46)^2 \cdot 8 + (79 - 62{,}46)^2 \cdot 4 + (67 - 62{,}46)^2 \cdot 20 + \right.$$
$$\left. + (62{,}88 - 62{,}46)^2 \cdot 67 + (53{,}55 - 62{,}46)^2 \cdot 22 + (49 - 62{,}46)^2 \cdot 9 \right] \approx 54{,}49 \, .$$

Вычисляем:

$$\delta_x = \sqrt{21{,}06} \approx 4{,}59;$$

$$\delta_y = \sqrt{54{,}49} \approx 7{,}38.$$

Тогда корреляционные отношения равны:

$$\eta_{yx} = \frac{\delta_y}{\sigma_y} = \frac{7{,}38}{10{,}14} \approx 0{,}73;$$

$$\eta_{xy} = \frac{\delta_x}{\sigma_x} = \frac{4{,}59}{5{,}62} \approx 0{,}82.$$

Замечание. Следует отметить, что корреляционное отношение всегда принимает значение от 0 до 1, причем оно не меньше, чем коэффициент корреляции, взятый по модулю, то есть,

$$0 \le |r| \le \eta \le 1.$$

В нашем примере:

$$0 < 0{,}72 < 0{,}73 < 1;$$

$$0 < 0{,}72 < 0{,}82 < 1.$$

Ответ: Корреляционная зависимость между признаками высокая, ее можно описать линейными уравнениями:

$$\bar{y}_x = -1{,}3 \cdot x + 100{,}83; \qquad \bar{x}_y = -0{,}4 \cdot y + 54{,}46.$$

Приложение 1

Таблица значений локальной функции Лапласа $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0.0	0.3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0.4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0.2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0631	0669
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2.0	0.0540	0529	2519	0508	0498	0488	0478	0468	0459	0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0.0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3.5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3.6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3.8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3.9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

$$\varphi(x \geq 4) = 0; \quad \varphi(-x) = \varphi(x).$$

Приложение 2

Таблица значений интегральной функции Лапласа

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{t^2}{2}}\,dt$$

x	**Ф(*x*)**	***x***	**Ф(*x*)**	***x***	**Ф(*x*)**	***x***	**Ф(*x*)**
0.00	0.0000	0.21	0.0832	0.42	0.1628	0.63	0.2357
0.01	0.0040	0.22	0.0871	0.43	0.1664	0.64	0.2389
0.02	0.0080	0.23	0.0910	0.44	0.1700	0.65	0.2422
0.03	0.0120	0.24	0.0948	0.45	0.1736	0.66	0.2454
0.04	0.0160	0.25	0.0987	0.46	0.1772	0.67	0.2486
0.05	0.0199	0.26	0.1026	0.47	0.1808	0.68	0.2517
0.06	0.0239	0.27	0.1064	0.48	0.1844	0.69	0.2549
0.07	0.0279	0.28	0.1103	0.49	0.1879	0.70	0.2580
0.08	0.0319	0.29	0.1141	0.50	0.1915	0.71	0.2611
0.09	0.0359	0.30	0.1179	0.51	0.1950	0.72	0.2642
0.10	0.0398	0.31	0.1217	0.52	0.1985	0.73	0.2673
0.11	0.0439	0.32	0.1255	0.53	0.2019	0.74	0.2703
0.12	0.0478	0.33	0.1293	0.54	0.2054	0.75	0.2734
0.13	0.0517	0.34	0.1331	0.55	0.2088	0.76	0.2764
0.14	0.0557	0.35	0.1368	0.56	0.2123	0.77	0.2794
0.15	0.0596	0.36	0.1406	0.57	0.2157	0.78	0.2823
0.16	0.0638	0.37	0.1443	0.58	0.2190	0.79	0.2852
0.17	0.0675	0.38	0.1480	0.59	0.2224	0.80	0.2881
0.18	0.0714	0.39	0.1517	0.60	0.2257	0.81	0.2910
0.19	0.0753	0.40	0.1554	0.61	0.2291	0.82	0.2939
0.20	0.0793	0.41	0.1591	0.62	0.2324	0.83	0.2967
0.84	0.2995	1.10	0.3643	1.36	0.4131	1.62	0.4474
0.85	0.3023	1.11	0.3665	1.37	0.4147	1.63	0.4484
0.86	0.3051	1.12	0.3686	1.38	0.4162	1.64	0.4495
0.87	0.3078	1.13	0.3708	1.39	0.4177	1.65	0.4505
0.88	0.3106	1.14	0.3729	1.40	0.4192	1.66	0.4515
0.89	0.3133	1.15	0.3749	1.41	0.4207	1.67	0.4525
0.90	0.3159	1.16	0.3770	1.42	0.4222	1.68	0.4535
0.91	0.3186	1.17	0.3790	1.43	0.4236	1.69	0.4545
0.92	0.3212	1.18	0.3810	1.44	0.4251	1.70	0.4554
0.93	0.3238	1.19	0.3830	1.45	0.4265	1.71	0.4564
0.94	0.3264	1.20	0.3849	1.46	0.4279	1.72	0.4573
0.95	0.3289	1.21	0.3869	1.47	0.4292	1.73	0.4582
0.96	0.3315	1.22	0.3883	1.48	0.4306	1.74	0.4591

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.97	0.3340	1.23	0.3907	1.49	0.4319	1.75	0.4599
0.98	0.3365	1.24	0.3925	1.50	0.4332	1.76	0.4608
0.99	0.3389	1.25	0.3944	1.51	0.4345	1.77	0.4616
1.00	0.3413	1.26	0.3962	1.52	0.4357	1.78	0.4625
1.01	0.3438	1.27	0.3980	1.53	0.4370	1.79	0.4633
1.02	0.3461	1.28	0.3997	1.54	0.4382	1.80	0.4641
1.03	0.3485	1.29	0.4015	1.55	0.4394	1.81	0.4649
1.04	0.3508	1.30	0.4032	1.56	0.4406	1.82	0.4656
1.05	0.3581	1.31	0.4049	1.57	0.4418	1.83	0.4664
1.06	0.3554	1.32	0.4066	1.58	0.4429	1.84	0.4671
1.07	0.3577	1.33	0.4082	1.59	0.4441	1.85	0.4678
1.08	0.3599	1.34	0.4099	1.60	0.4452	1.86	0.4686
1.09	0.3621	1.35	0.4115	1.61	0.4463	1.87	0.4693
1.88	0.4699	2.12	0.4830	2.48	0.4934	2.84	0.4977
1.89	0.4706	2.14	0.4838	2.50	0.4938	2.86	0.4979
1.90	0.4713	2.16	0.4846	2.52	0.4941	2.88	0.4980
1.91	0.4719	2.18	0.4854	2.54	0.4945	2.90	0.4981
1.92	0.4726	2.20	0.4861	2.56	0.4948	2.92	0.4982
1.93	0.4732	2.22	0.4868	2.58	0.4951	2.94	0.4984
1.94	0.4338	2.24	0.4875	2.60	0.4953	2.96	0.4985
1.95	0.4744	2.26	0.4881	2.62	0.4956	2.98	0.4986
1.96	0.4750	2.28	0.4887	2.64	0.4959	3.00	0.49865
1.97	0.4756	2.30	0.4893	2.66	0.4961	3.20	0.49931
1.98	0.4761	2.32	0.4898	2.68	0.4963	3.40	0.49966
1.99	0.4767	2.34	0.4904	2.70	0.4965	3.60	0.49984
2.00	0.4772	2.36	0.4909	2.72	0.4967	3.80	0.49993
2.02	0.4783	2.38	0.4913	2.74	0.4969	4.00	0.49997
2.04	0.4793	2.40	0.4918	2.76	0.4971	4.50	0.49999
2.06	0.0.480	2.42	0.4922	2.78	0.4973	5.00	0.50000
2.08	0.4812	2.44	0.4927	2.80	0.4974		
2.10	0.4821	2.46	0.4931	2.82	0.4976		

$$\Phi(x > 5) = 0{,}5 \;;$$

$$\Phi(-x) = -\Phi(x).$$

МАТЕМАТИКА
ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания и задания контрольной работы

Редактор О.В. Елистратова

Компьютерная верстка Т.М. Постниковой

Лицензия ИД № 01102 от 01.03. 2000

Подписано в печать 18.11.2004. Формат 60x84/16. Бумага офсетная. Тираж 1200 экз.
Печ.л. 4,5. Уч.-изд.л. 4,18. Изд. № 114. Заказ № 698.

Типография Сибирского университета потребительской кооперации.
630087, Новосибирск, пр.К.Маркса,26.