МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МУРМАНСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра математики, информационных систем и программного обеспечения

Методические указания к выполнению работы по темам «Теория вероятностей. Математическая статистика» для обучающихся заочной формы обучения направлений подготовки естественно-технологического института

Мурманск

2020 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В настоящем пособии содержатся методические рекомендации к изучению теоретического материала и выполнению работы по темам «Теория вероятностей. Математическая статистика», варианты работ и список рекомендуемой литературы.

В результате изучения этой темы студенты должны:

- знать основные положения теории вероятностей; основные положения математической статистики; основы теории случайных процессов;
- уметь применять методы теории вероятностей, математической статистики и теории случайных процессов для решения практических задач.

Данные методические рекомендации включают краткий справочный материал для выполнения контрольной работы, который нужно рассматривать как дополнение к имеющимся учебникам по теории вероятностей и математической статистике. Так же в пособии содержится подробное решение примерного варианта контрольной работы со ссылками на используемый справочный материал.

ЗАДАНИЯ ДЛЯ ВЫПОЛНЕНИЯ КОНТРОЛЬНОЙ РАБОТЫ

Перед выполнением контрольной работы необходимо ознакомится сначала со справочным материалом, затем более глубоко изучить теоретический материал и закрепить его решением задач.

Задача 1.

Вариант 1. При увеличении напряжения может произойти разрыв электрической цепи из-за выхода из строя одного из трех элементов, Вероятности выхода из строя элементов 0,3, 0,4 и 0,5 соответственно. Какова вероятность того, что не будет разрыва сети?

Вариант 2. Радист 3 раза вызывает корреспондента. Вероятность того, что будет принят первый вызов, равна 0,2, второй – 0,3, третий – 0,4. События, состоящие в том, что данный вызов будет услышан, независимы. Найти вероятность того, что корреспондент услышит вызов радиста хотя бы один раз.

Вариант 3. Вероятности своевременного выполнения студентом контрольных работ по каждой из трех дисциплин равны соответственно 0,6, 0,5 и 0,8. Найти вероятность своевременного выполнения студентом контрольных работ по двум дисциплинам.

Вариант 4. Вероятности своевременного выполнения студентом контрольных работ по каждой из трех дисциплин равны соответственно 0,6, 0,5 и 0,8. Найти вероятность своевременного выполнения студентом контрольных работ хотя бы по двум дисциплинам.

Вариант 5. Экспедиция издательства отправила газеты в три почтовых отделения. Вероятность своевременной доставки газет в первое отделение равна 0,95, во второе отделение – 0,85 и в третье – 0,7. Найти вероятность, того, что хотя бы одно отделение получит газеты вовремя.

Вариант 6. Экспедиция издательства отправила газеты в три почтовых отделения. Вероятность своевременной доставки газет в первое отделение равна 0,95, во второе отделение – 0,9 и в третье – 0,8. Найти вероятность, того, что только два отделения получат газеты вовремя.

Вариант 7. Экспедиция издательства отправила газеты в три почтовых отделения. Вероятность своевременной доставки газет в первое отделение равна 0,9, во второе отделение – 0,95 и в третье – 0,85. Найти вероятность, того, что только одно отделение получит газеты вовремя.

Вариант 8. По цели стреляют из трех орудий. Вероятность попадания для первого орудия равна 0,9, для второго – 0,8, для третьего – 0,7. Найти вероятность того, что только два орудия попадут в цель.

Вариант 9. По цели стреляют из трех орудий. Вероятность попадания для первого орудия равна 0,9, для второго – 0,8, для третьего – 0,6. Найти вероятность того, что только одно орудие попадет в цель.

Вариант 10. По цели стреляют из трех орудий. Вероятность попадания для первого орудия равна 0,6, для второго – 0,8, для третьего – 0,7. Найти вероятность того, что хотя бы одно орудие попадет в цель.

Задача 2.

В каждом варианте для заданной случайной величины ξ составить закон распределения, построить многоугольник распределения вероятностей, вычислить математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины.

Вариант 1. Вероятность отказа каждого прибора при проведении испытания равна 0,4, для испытания было отобрано 4 прибора, случайная величина ξ – число приборов, отказавших при проведении испытаний.

Вариант 2. Вероятность совершить покупку для каждого покупателя магазина равна 0,3, в магазин пришли 4 покупателя, случайная величина ξ – число покупателей, совершивших покупку.

Вариант 3. По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,515, случайная величина ξ – число мальчиков в семье из 4 детей.

Вариант 4. Вероятность того, что корреспондент примет вызов радиста, равна 0,4, случайная величина ξ – число вызовов, принятых корреспондентом, если радистом было передано 4 вызова.

Вариант 5. В рекламных целях торговая фирма вкладывает в единицу товара денежный приз размером 1 тыс. руб., случайная величина ξ – размер выигрыша при четырех сделанных покупках, если вероятность выигрыша в каждой покупке равна 0,1.

Вариант 6. В контрольной работе 4 задачи, вероятность правильного решения учеником каждой задачи 0,7, случайная величина ξ – число правильно решенных задач.

Вариант 7. Торговый агент имеет четырех потенциальных покупателей, вероятность того, что потенциальный покупатель сделает заказ, равна 0,4, случайная величина ξ – число покупателей, сделавших заказ.

Вариант 8. Студент должен сдать в сессию 4 экзамена, вероятность успешной сдачи каждого экзамена 0,7, случайная величина ξ – число экзаменов, которые сдал студент в сессию.

Вариант 9. Телевизионный канал рекламирует новый вид детского питания. Вероятность того, что телезритель увидит эту рекламу, оценивается в 0,2. В случайном порядке выбраны четыре телезрителя, случайная величина ξ – число лиц, видевших рекламу.

Вариант 10. Вероятность того, что расход электроэнергии в продолжение одних суток не превысит установленной нормы, равна 0,75, контроль расхода электроэнергии производится в течение четырех суток, случайная величина ξ – число дней, в которые расход электроэнергии был выше установленной нормы.

Задача 3

Вариант 1. Значения теста IQ (коэффициента интеллекта) Стэнфорда – Бине распределены приблизительно по нормальному закону с математическим ожиданием $a = 100$ и средним квадратическим отклонением $\sigma = 16$. Найти вероятность того, что коэффициент интеллекта у случайно отобранного для тестирования человека окажется меньше 95.

Вариант 2. Заряд охотничьего пороха отвешивается на весах. Вес заряда - нормально распределенная случайная величина с параметрами a = 2,3 г и σ=150 мг. Найти вероятность повреждения ружья при выстреле, если максимально допустимый вес заряда пороха равен 2,5 г.

Вариант 3. Размер детали подчинен нормальному закону с параметрами a= 30 см и σ= 5 см. Детали считаются годными, если их размер находится в пределах от 20 до 40 см. Если размер детали больше 40 см, то она подлежит переделке. Найти вероятность того, что случайно отобранная деталь подлежит переделке.

Вариант 4. Значения теста IQ (коэффициента интеллекта) Стэнфорда – Бине распределены приблизительно по нормальному закону с математическим ожиданием $a = 100$ и средним квадратичным отклонением $\sigma = 16$. Найти вероятность того, что коэффициент интеллекта у случайно отобранного для тестирования человека окажется в пределах от 80 до 120.

Вариант 5. Средняя длина взрослой рыбы оценивается в 65 см. со стандартным отклонением в 5 см. Считая распределение длины рыбы нормальным, найдите вероятность того, что длина конкретной рыбы будет больше 70 см.

Вариант 6. Спортсмен бросает копье. Дальность полета копья – нормально распределенная случайная величина со средним значением 70 м и средним квадратическим отклонением σ=5 м. Найти вероятность того, что дальность полета копья будет от 65 до 72 м.

Вариант 7. На рынок поступила крупная партия говядины. Предполагается, что вес туш – случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием $a=950$ кг и средним квадратическим отклонением σ=150 кг. Определить вероятность того, что вес случайно отобранной туши будет между 800 и 1300 кг.

Вариант 8. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед. Найти вероятность того, что цена акции будет не выше 15,3 ден. ед.

Вариант 9. Текущая цена акции может быть смоделирована с помощью нормального закона распределения с математическим ожиданием 15 ден. ед. и средним квадратическим отклонением 0,2 ден. ед. Найти вероятность того, что цена акции будет в интервале от 14,9 до 15,3 ден. ед.

Вариант 10. На рынок поступила крупная партия говядины. Предполагается, что вес туш – случайная величина, подчиняющаяся нормальному закону распределения с математическим ожиданием $a=950$ кг и средним квадратическим отклонением σ=150 кг. Найти вероятность того, что вес случайно отобранной туши окажется больше 1250 кг.

Задача 4.

Из генеральной совокупности, распределенной по нормальному закону, сделана выборка. Найти: 1) числовые характеристики выборки – выборочную среднюю, выборочную дисперсия, выборочное среднее квадратическое отклонение; 2) несмещенные оценки для генеральной средней и генеральной дисперсии; 3) доверительный интервал для оценки генеральной средней с заданной надежностью γ.

Вариант 1.

x_i	54-58	58-62	62-66	66-70	70-74	74-78	78-82
n_i	12	16	22	24	12	10	4

$\gamma = 0{,}93$.

Вариант 2

x_i	154-158	158-162	162-166	166-170	170-174	174-178	178-182
n_i	10	14	26	28	12	8	2

$\gamma = 0{,}94$.

Вариант 3.

x_i	0-200	200-300	300-400	400-500	500-600	600-700
n_i	3	5	9	14	8	3

$\gamma = 0{,}97$.

Вариант 4.

x_i	33,2	38,2	43,2	48,2	53,2
n_i	1	2	18	3	1

$\gamma = 0{,}92$.

Вариант 5.

x_i	15,4	18,4	21,4	24,4	27,4
n_i	2	4	11	5	3

$\gamma = 0{,}96$.

Вариант 6.

x_i	15	20	16	17	19	18
n_i	3	9	2	7	6	8

$\gamma = 0{,}91$.

Вариант 7.

x_i	4 - 9	9 - 14	14 - 19	19 - 24	24 - 29
n_i	5	9	13	6	7

$\gamma = 0{,}98$.

Вариант 8.

x_i	12,4	16,4	20,4	24,4	28,4	32,4	36,4
n_i	5	15	40	25	8	4	3

$\gamma = 0{,}93$.

Вариант 9.

x_i	1,7– 2,8	2,8– 3,9	3,9-5,0	5,0-6,1	6,1-7,2	7,2-8,3
n_i	8	10	22	10	6	4

$\gamma = 0{,}97$.

Вариант 10.

x_i	3 - 7	7 - 11	11- 15	15 - 19	19 - 23
n_i	1	5	11	7	3

$\gamma = 0{,}94$

Задача 5.

Имеются две нормально распределенные генеральные совокупности X и Y, из которых были сделаны выборки. По полученным выборкам на уровне значимости α проверить гипотезу $H_0: \ \bar{x}_0 = \bar{y}_0$, считая дисперсии неизвестными, но равными.

Вариант 1.

x_i	51	44	47	24	43	34	60
y_i	40	38	37	52	42		

Уровень значимости $\alpha = 0{,}05$, альтернативная гипотеза $H_1: \bar{x}_0 \neq \bar{y}_0$

Вариант 2.

x_i	113	120	113	109	111	102	116

y_i	97	104	105	103	122	128	113

Уровень значимости $\alpha = 0{,}04$, альтернативная гипотеза $H_1 : \bar{x}_0 > \bar{y}_0$

<u>Вариант 3.</u>

x_i	49	51	46	49	56		
y_i	57	58	50	51	46	39	67

Уровень значимости $\alpha = 0{,}06$, альтернативная гипотеза $H_1 : \bar{x}_0 < \bar{y}_0$

<u>Вариант 4.</u>

x_i	35	65	50	46			
y_i	38	33	37	65	78	66	31

Уровень значимости $\alpha = 0{,}05$, альтернативная гипотеза $H_1 : \bar{x}_0 \neq \bar{y}_0$

<u>Вариант 5.</u>

x_i	58	56	53	53	56	52	52
y_i	50	54	51	56	53	70	68

Уровень значимости $\alpha = 0{,}1$, альтернативная гипотеза $H_1 : \bar{x}_0 < \bar{y}_0$

<u>Вариант 6.</u>

x_i	93	94	93	92	90	65	87
y_i	93	92	103	95			

Уровень значимости $\alpha = 0{,}01$, альтернативная гипотеза $H_1 : \bar{x}_0 \neq \bar{y}_0$

<u>Вариант 7.</u>

x_i	35	34	47	46			
y_i	65	45	34	53	68	70	70

Уровень значимости $\alpha = 0{,}015$, альтернативная гипотеза $H_1 : \bar{x}_0 < \bar{y}_0$

<u>Вариант 8.</u>

x_i	99	97	95	94	90	91	89
y_i	93	96	94	95			

Уровень значимости $\alpha = 0{,}02$, альтернативная гипотеза $H_1 : \bar{x}_0 \neq \bar{y}_0$

<u>Вариант 9.</u>

x_i	47	59	61	60			

y_i	55	55	61	62	67	75	64

Уровень значимости $\alpha = 0{,}01$, альтернативная гипотеза $H_1 : \bar{x}_0 < \bar{y}_0$

Вариант 10.

x_i	16	17	14	8	20		
y_i	27	24	18	15	5	7	30

Уровень значимости $\alpha = 0{,}025$, альтернативная гипотеза $H_1 : \bar{x}_0 < \bar{y}_0$

Задача 6.

Была исследована зависимость признака Y от признака X. В результате проведения 10 измерений была получена таблица.

Требуется: 1) оценить тесноту и направление связи между признаками с помощью коэффициента корреляции; 2) найти уравнение линейной регрессии Y на X; 3) в одной системе координат построить эмпирическую и теоретическую линии регрессии.

Вариант 1.

x_i	9	12	13	14	15	17	18	19	21	23
y_i	69	73	95	87	96	98	105	111	107	129

Вариант 2.

x_i	6,0	6,5	6,8	7,0	7,4	8,0	8,2	8,7	9,0	10,0
y_i	10	11	12	13	15	17	18	20	20	25

Вариант 3.

x_i	39,0	38,7	38,9	40,1	39,4	39,4	39,5	39,1	40,4	39,5
y_i	4	3	4	6	6	5	4	6	7	5

Вариант 4.

x_i	85	88	85	106	100	97	105	106	105	103
y_i	45	49	50	56	53	55	56	58	60	62

Вариант 5.

x_i	28	25	33	49	32	24	32	24	36	32
y_i	34	28	38	47	36	27	28	29	31	37

Вариант 6.

x_i	3,5	4,0	3,8	4,6	3,9	3,0	3,5	3,9	4,5	4,1
y_i	4,2	3,9	3,8	4,5	4,2	3,4	3,8	3,9	4,6	3,0

Вариант 7.

x_i	2	3	4	4	5	5	7	10	10	12
y_i	2	2	3	2,5	3,5	4	4	5	6	8

Вариант 8.

x_i	30	41	52	60	73	80	92	100	112	125
y_i	19	25	30	32	37	40	45	47	51	53

Вариант 9.

x_i	2,8	2,2	3,0	3,5	3,2	3,7	4,0	4,8	6,0	5,4
y_i	6,7	6,9	7,2	7,3	8,4	8,8	9,1	9,8	10,6	10,7

Вариант 10.

x_i	3,2	3,7	4,0	4,8	6,0	5,4	5,2	5,4	6,0	9,0
y_i	8,4	8,8	9,1	9,8	10,6	10,7	11,1	11,8	12,1	12,4

СПРАВОЧНЫЙ МАТЕРИАЛ К ВЫПОЛНЕНИЮ КОННТРОЛЬНОЙ РАБОТЫ

1. Теория вероятностей

1.1. Случайные события

В основе теории вероятностей лежит следующая модель: имеется комплекс условий, который можно воспроизводить, хотя бы принципиально, неограниченное число раз. Каждое его воспроизведение называется *опытом*, *испытанием* или *экспериментом*. Предполагается, что в каждом опыте обязательно происходит одно и только одно так называемое *элементарное событие* (*элементарный исход*) ω. Все множество элементарных событий, которые могут происходить в результате опыта, называется *пространством элементарных* *событий* (исходов) Ω[1].

Случайное событие – это некоторое множество, состоящее из элементарных исходов $A=\{\omega_1,\omega_2,...,\omega_k\}$. При этом исходы $\omega_1,\omega_2,...,\omega_k$ называются *благоприятствующими событию* A. Случайные события, так же как и множества, обычно обозначают заглавными буквами латинского алфавита с индексами или без: A, B, C_1, D_2 и т.д.

Говорят, что в результате эксперимента *произошло событие* $A \subseteq \Omega$, если в эксперименте произошел один из элементарных исходов, благоприятствующих событию A.

Достоверным называется событие, которое всегда происходит в результате рассматриваемого эксперимента. Следовательно, оно включает в себя <u>все</u> элементарные исходы, т.е. достоверным событием является пространство элементарных исходов Ω.

Событие называется *невозможным*, если оно заведомо не может произойти в результате рассматриваемого эксперимента.

Суммой событий A и B называется событие $A+B$, состоящее в наступлении хотя бы одного из событий A или B, т.е. это событие состоит из элементарных исходов, которые принадлежат либо A, либо B, либо двум событиям одновременно.

Произведением событий A и B называется событие AB, состоящее в том, что оба события A и B произошли одновременно, т.е. это событие состоит из элементарных исходов, принадлежащих и A, и B.

Два события A и B называются *несовместными*, если A и B не могут

---

[1] Понятия “элементарное событие” и “происходит” являются первоначальными неопределяемыми понятиями, подобно геометрическим понятиям “точка” и “лежит”. При общих рассуждениях полезно иметь в виду какой-либо простой конкретный эксперимент типа общепонятного бросания монеты, игральной кости, извлечения карты из колоды и т.п.

произойти одновременно. Несовместные события не имеют ни одного общего благоприятствующего исхода, следовательно, $AB = \varnothing$.

Противоположным событию A называется событие $\overline{A}$, которое происходит тогда и только тогда, когда не происходит событие A. Для противоположных событий одновременно выполняются два условия: их сумма является достоверным событием, т.е. $A + \overline{A} = \Omega$, а произведение – невозможным событием, т.е. $A\overline{A} = \varnothing$.

1.2. Вероятность события

Для количественной оценки возможности появления случайного события A в рассматриваемом эксперименте вводится специальная числовая функция $P(A)$, называемая *вероятностью события* A, которая каждому событию ставит в соответствие число.

Например, вероятность события A можно найти, используя *классическое определение вероятности*: вероятность случайного события A равна отношению числа k элементарных равновозможных исходов, благоприятствующих событию, к числу n всех возможных элементарных исходов эксперимента, т.е.

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

Заметим, что вероятность достоверного события $P(\Omega) = 1$, вероятность невозможного события $P(\varnothing) = 0$. Кроме того, из определения вероятности следует, что для любого события A выполняется неравенство:

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

1.3. Вероятности сложных событий

Для вычисления вероятностей сложных событий используются следующие теоремы теории вероятностей.

<u>Теорема сложения вероятностей.</u> Если события A и B несовместны, то вероятность их суммы равна сумме их вероятностей: $P(A+B)=P(A)+P(B)$ (1)

Сформулированная теорема справедлива для любого конечного числа слагаемых.

Если же события A и B совместны, то $P(A+B)=P(A)+P(B)-P(AB)$ (2)

Из теоремы сложения вероятностей следует, что если A и $\overline{A}$ – противоположные события, то $P(A)+P(\overline{A})=1$ или $P(A)=1-P(\overline{A})$ (3)

<u>Теорема умножения вероятностей.</u> Если события A и B независимы (т.е. вероятность одного из событий не зависит от появления или непоявления другого), то вероятность произведения событий равна произведению их вероятностей: $P(AB)=P(A)\cdot P(B)$ (4)

Сформулированная теорема также справедлива для любого конечного числа сомножителей.

1.4. Формула Бернулли

Последовательность n независимых испытаний, в каждом из которых может произойти некоторое событие A (его называют *успехом*) с вероятностью $p=P(A)$, постоянной при любом испытании, или противоположное ему событие $\overline{A}$ (его называют *неудачей*) с вероятностью $q=P(\overline{A})=1-p$, называется *схемой Бернулли*.

Если производится n независимых испытаний, в каждом из которых вероятность появления события A постоянна и равна p, то вероятность того, что событие A произойдет ровно m раз определяется *формулой Бернулли*:

$$P_n(m)=C_n^m p^m q^{n-m}, \qquad (5)$$

где C_n^m – число сочетаний из n элементов по m элементам, которое вычисляется по формуле: $C_n^m = \dfrac{n!}{m!\cdot(n-m)!}$ (6)

1.5. Случайные величины

Случайной величиной ξ называется числовая функция, определенная на пространстве элементарных исходов Ω, которая каждому элементарному исходу ω ставит в соответствие число $\xi(\omega)$.

Условимся в дальнейшем, как правило, случайные величины обозначать греческими буквами: ξ, η, …, а принимаемые ими значения – строчными латинскими (с индексами или без): x, y_1 и т.д.

Случайные величины делятся на *дискретные* и *непрерывные*.

Случайная величина называется *дискретной*, если значения, которые может принимать данная случайная величина, образуют дискретное (конечное или бесконечное) множество чисел x_1, x_2, …, x_n, …

Под *непрерывной* случайной величиной будем понимать величину, значения которой, заполняют конечный или бесконечный промежуток числовой оси.

Например, число студентов на лекции – дискретная случайная величина, продолжительность лекции – непрерывная.

1.6. Дискретная случайная величина и ее числовые характеристики

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Закон распределения дискретной случайной величины обычно задается *рядом распределения*, который представляется в виде таблицы:

x_i	x_1	x_2	…	x_n

p_i	p_1	p_2	$\cdots$	p_n

где в первой строке перечислены все возможные значения случайной величины, а во второй – соответствующие им вероятности $p_i = P(\xi = x_i)$, удовлетворяющие соотношению $\sum_{i=1}^{n} p_i = 1$.

Закон распределения может быть задан графически в виде *многоугольника распределения вероятностей*, т.е. в виде ломаной, соединяющей точки с координатами (x_i, p_i) для $i = \overline{1, n}$.

Математическим ожиданием или *средним значением* дискретной случайной величины ξ называется сумма произведений всех ее возможных значений на соответствующие им вероятности: $M\xi = \sum_i x_i p_i$ (7)

Дисперсией случайной величины ξ называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания, т.е.

$$D\xi = M(\xi - M\xi)^2 = \sum_i (x_i - M\xi)^2 \cdot p_i . \tag{8}$$

Для вычисления дисперсии на практике бывает удобнее использовать другую формулу, которую можно получить из формулы (8) с помощью простых преобразований:

$$D\xi = M\xi^2 - (M\xi)^2 = \sum_i x_i^2 \cdot p_i - (M\xi)^2 . \tag{9}$$

Средним квадратическим отклонением или *стандартным отклонением* случайной величины называется корень квадратный из ее дисперсии:

$$\sigma_\xi = \sqrt{D\xi} . \tag{10}$$

1.7. Непрерывная случайная величина и ее числовые характеристики

Закон распределения непрерывной случайной величины ξ удобно задавать с помощью *функции плотности вероятности* $f(x)$. Вероятность $P(a<\xi<b)$ того, что значение, принятое случайной величиной ξ, попадет в интервал $(a;b)$, определяется равенством: $P(a<\xi<b)=\int\limits_a^b f(x)dx$.

Математическим ожиданием или *средним значением* непрерывной случайной величины ξ с плотностью распределения вероятности $f(x)$ называется число: $M\xi=\int\limits_{-\infty}^{\infty} x\cdot f(x)dx$, если этот интеграл сходится абсолютно. В противном случае математическое ожидание случайной величины ξ не существует.

Дисперсия непрерывной случайной величины определяется также, как и для дискретной. Для непосредственного вычисления дисперсии используют формулы:

$$D\xi=\int\limits_{-\infty}^{\infty}(x-M\xi)^2 f(x)dx \text{ или } D\xi=\int\limits_{-\infty}^{\infty}x^2 f(x)dx-(M\xi)^2.$$

1.8. Нормальное распределение

Непрерывная случайная величина имеет *нормальное* или *гауссовское распределение* с параметрами a и σ^2, где $-\infty<a<\infty$ и $\sigma>0$ (пишут $\xi\sim N(a,\sigma^2)$), если функция плотности случайной величины имеет вид:

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

для любого $x\in R$.

При $a=0$ и $\sigma^2=1$, т.е. $\xi\sim N(0,1)$, нормальный закон распределения называется *стандартным* или *нормированным*.

Если случайная величина ξ имеет нормальное распределение с параметрами a и σ^2, то

$$M\xi = a \text{ и } D\xi = \sigma^2.$$

Найти вероятность попадания случайной величины ξ, распределенной по нормальному закону с параметрами a и σ^2, в интервал (α, β) можно по формуле:

$$P(\alpha < \xi < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right), \qquad (11)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функции Лапласа, значения которой определяются по таблице. При использовании таблицы необходимо учитывать, что функция $\Phi(x)$ является нечетной и при $x > 4$ значения $\Phi(x)$ считаются равными 0,5.

2. Математическая статистика

2.1. Генеральная и выборочная совокупности

Генеральной совокупностью называется вся подлежащая изучению совокупность объектов (наблюдений). Один или несколько элементов, взятых из генеральной совокупности для получения информации о ней, называется *выборочной совокупностью* или *выборкой*. *Объемом совокупности* (выборочной или генеральной) называется число элементов этой совокупности.

Метод статистического исследования, состоящий в том, что на основе изучения выборки делается заключение обо всей генеральной совокупности, называется *выборочным*. Для того чтобы по отобранным значениям некоторого показателя можно было достаточно уверенно судить обо всей совокупности, полученная выборка должна быть *репрезентативной (представительной),* т.е. правильно отражать пропорции генеральной совокупности. Выборка будет представительной лишь тогда, когда все объекты генеральной совокупности имеют *одинаковую вероятность попасть в выборку.*

В дальнейшем под генеральной совокупностью будем подразумевать не само множество объектов, а множество значений случайной величины, принимающей числовое значение на каждом из объектов. В статистике обычно исследуемые случайные величины называют *признаками* и обозначают большими латинскими буквами X, Y и т.д.

Выборку будем рассматривать как совокупность независимых случайных величин $X_1, X_2, ..., X_n$, распределенных так же, как и случайная величина (признак) X, представляющая генеральную совокупность. Конкретные значения, которые приняли эти случайные величины в результате эксперимента, называют *реализацией выборки* или *значениями признака* и обозначают строчными буквами $x_1, x_2, \ldots, x_n$. Различные значения признака называют *вариантами*.

2.2. Вариационные ряды

Числа, показывающие, сколько раз встречаются варианты x_i в выборке, называются *частотами* (обозначаются n_i).

Вариационным (*статистическим*) *рядом* называется расположенный в порядке возрастания или убывания ряд вариант с соответствующими им частотами. Вариационный ряд часто называют *рядом распределения выборки*.

Вариация (изменение) количественных признаков может быть дискретной, например, академическая система успеваемости: 5 – отлично, 4 – хорошо и т.д., или непрерывной, например, возраст, рост или вес человека. В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды.

Вариационный ряд называется *дискретным*, если любые его варианты отличаются на постоянную величину. Он представляет собой таблицу, состоящую из двух строк: конкретных значений признака и их частот:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

где $\sum_{i=1}^{k} n_i = n$.

Вариационный ряд называется *интервальным*, если варианты могут отличаться один от другого на сколь угодно малую величину[2]. Он представляет собой таблицу, состоящую из двух строк – интервалов значений признака и числа значений выборки, попадающих в этот интервал:

Интервал, $x_i - x_{i+1}$	$x_1 - x_2$	$x_2 - x_3$	…	$x_k - x_{k+1}$
Частота, n_i	n_1	n_2	…	n_k

2.3. Числовые характеристики

Вариационный ряд содержит достаточно полную информацию об изменчивости признака. Однако обилие числовых данных, с помощью которых он задается, усложняет их использование. В то же время на практике часто оказывается достаточным знание лишь сводных числовых характеристик выборочной совокупности. Рассмотрим наиболее часто используемые числовые характеристики вариационных рядов: среднюю арифметическую, дисперсию и среднее квадратическое отклонение.

Средняя арифметическая

Средние величины характеризуют значения признака, вокруг которого концентрируются наблюдения. Наиболее распространенной из средних величин является *средняя арифметическая*. Для ее расчета используют формулу:

$$\bar{x} = \frac{1}{n}\sum_i x_i n_i , \tag{14}$$

[2] Построение интервальных вариационных рядов целесообразно не только при непрерывной вариации признака, но и если дискретная вариация проявляется в широких пределах, т.е. число вариантов дискретного признака достаточно велико.

где x_i – варианты, n_i – соответствующие им частоты, n – объем совокупности.

Если средняя арифметическая рассчитывается по всей генеральной совокупности в целом, то ее называют *генеральной средней*, а если по выборке – *выборочной средней*.

Если статистический материал представлен в виде интервального вариационного ряда, то при расчете выборочной средней сначала необходимо вычислить середины каждого интервала x_i^*, которые рассчитываются по формуле: $x_i^* = \frac{x_i + x_{i+1}}{2}$. Далее расчеты ведутся, как и для дискретного вариационного ряда, но в качестве вариантов используем x_i^*.

Дисперсия и среднее квадратическое отклонение

Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины и вычисляется по формуле:

$$\sigma_x^2 = \frac{\sum_i (x_i - \bar{x})^2 \cdot n_i}{n}. \tag{15}$$

Дисперсия, рассчитанная для генеральной совокупности, называется *генеральной дисперсией*, а для выборки – *выборочной дисперсией*.

При вычислении выборочной дисперсии для интервальных вариационных рядов в качестве x_i, как и при вычислении выборочной средней, используются середины соответствующих интервалов.

Иногда, особенно если дисперсию приходится рассчитывать «вручную», удобнее использовать другую формулу, которая легко получается из формулы (15) с помощью несложных математических преобразований:

$$\sigma_x^2 = \frac{\sum x_i^2 \cdot n_i}{n} - (\bar{x})^2. \tag{16}$$

Среднее квадратическое отклонение представляет собой квадратный корень из дисперсии:

$$\sigma_x = \sqrt{\sigma_x^2}\ .$$

Среднее квадратическое отклонение (СКО) выражается в тех же единицах измерения, что и признак.

2.4. Точечные оценки

Большинство случайных величин имеют распределения, зависящие от одного или нескольких параметров. Так, например, нормальное распределение зависит от параметров a и σ^2.

Выборочная характеристика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется *статистической точечной оценкой* этого параметра. Статистическая оценка неизвестных параметров теоретического распределения генеральной совокупности (или просто параметров генеральной совокупности) – одна из основных задач математической статистики.

Обозначим через θ некоторый неизвестный параметр генеральной совокупности, а через θ_n^* – точечную оценку этого параметра. Оценка θ_n^* есть функция $f(X_1, X_2, ..., X_n)$ от выборки объема n из независимых случайных величин $X_1, X_2, ..., X_n$, каждая из которых имеет тот же закон распределения, что и генеральная совокупность. Поэтому оценка θ_n^*, как функция случайных величин, также является случайной величиной, в отличие от оцениваемого параметра θ, который является величиной неслучайной, детерминированной.

Оценка θ_n^* для параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, т.е. $M\theta_n^* = \theta$. В противном случае оценка называется *смещенной*.

Несмещенность – свойство оценок при фиксированном n. Оно означает отсутствие ошибки "в среднем", т.е. при систематическом использовании данной оценки.

Рассмотрим некоторые наиболее часто встречающиеся точечные оценки параметров генеральной совокупности.

1. Выборочная средняя $\bar{x}$ есть несмещенная оценка для генеральной средней $\bar{x}_0$, причем $\sigma_{\bar{x}}^2 = \dfrac{\sigma_0^2}{n}$, где n – объем выборки, σ_0^2 – генеральная дисперсия признака.
2. Выборочная дисперсия σ_x^2 является смещенной оценкой генеральной дисперсии σ_0^2.
3. Исправленная дисперсия s^2, вычисляемая по формуле

$$s_x^2 = \frac{n}{n-1}\sigma^2 \qquad (17)$$

или

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2, \qquad (18)$$

является несмещенной оценкой для генеральной дисперсии σ_0^2.

Разница между σ_x^2 и s_x^2 заметна при небольшом числе наблюдений n. При $n > 40-50$ получим, что $\sigma_x^2 \approx s_x^2$, т.е. в качестве оценки σ_0^2 вполне можно использовать выборочную дисперсию σ_x^2.

2.5. Интервальные оценки

Точечные оценки параметров генеральной совокупности могут быть приняты в качестве ориентировочных, первоначальных результатов обработки выборочных данных. Их недостаток заключается в том, что неизвестно, с какой точностью оценивается параметр. Если для выборок большого объема точность обычно бывает достаточной, то для выборок небольшого объема вопрос точности оценок становится очень важным.

Чтобы получить представление о точности и надежности оценки θ_n^* параметра θ, используют интервальную оценку параметра.

Интервальной оценкой параметра θ называется числовой интервал (θ_1^*, θ_2^*), который с заданной вероятностью γ накрывает неизвестное значение параметра θ, т.е. $P(\theta \in (\theta_1^*, \theta_2^*)) = \gamma$. Такой интервал (θ_1^*, θ_2^*) называется *доверительным*, а вероятность γ – *доверительной вероятностью* или *надежностью оценки*.

Обычно надежность оценки γ задается заранее величиной, близкой к единице, например: 0,9, 0,95, 0,99 или 0,999.

Границы доверительного интервала и его длина находятся по выборочным данным и поэтому являются случайными величинами. Длина доверительного интервала уменьшается с ростом объема выборки n и увеличивается с ростом доверительной вероятности γ.

Очень часто (но не всегда) доверительный интервал выбирается симметричным относительно несмещенной точечной оценки θ_n^*, т.е. выбирается интервал вида $(\theta_n^* - \Delta; \theta_n^* + \Delta)$. Число Δ при этом называется *точностью оценки*.

Так, например, интервальная оценка (доверительный интервал) для генеральной средней $\bar{x}_0$ исследуемого признака X, имеющего нормальное распределение, может быть найдена по формуле:

$$\bar{x} - \Delta < \bar{x}_0 < \bar{x} + \Delta. \tag{19}$$

В случае, когда генеральная дисперсия σ_0^2 известна (например, это заранее заданная ошибка измерительного прибора), то точность оценки Δ находится по формуле:

$$\Delta = \frac{t_\gamma \cdot \sigma_0}{\sqrt{n}}, \qquad (20)$$

где n – объем выборки, а число t_γ определяется из равенства $2\Phi(t_\gamma) = \gamma$, т.е. по таблице значений функции Лапласа $\Phi(x)$ находится значение аргумента t_γ, которому соответствует значение функции $\Phi(x)$, равное $\gamma/2$.

В случае, когда генеральная дисперсия σ_0^2 неизвестна, а известна лишь ее оценка – исправленная выборочная дисперсия s^2, то точность оценки Δ находится по формуле:

$$\Delta = \frac{t_{\gamma,k} \cdot s}{\sqrt{n}}, \qquad (21)$$

где значение числа $t_{\gamma,k}$ определяется по таблице критических точек распределения Стьюдента при доверительной вероятности γ и числе степеней свободы $k = n - 1$.

<u>Замечание.</u> Если выборка $X_1, X_2, \ldots, X_n$ объема n представляет собой набор независимых одинаково распределенных случайных величин, то, согласно центральной предельной теореме, распределение $\dfrac{\bar{x} - \bar{x}_0}{\sigma_0 / \sqrt{n}}$ при больших n близко к стандартному нормальному. Это позволяет строить доверительный интервал для генеральной средней $\bar{x}_0$ по формулам (19) и (20) при любом распределении признака, если объем выборки является достаточно большим ($n > 100$), при этом в качестве σ_0^2 используется ее любая оценка.

2.6. Проверка статистических гипотез

Статистической гипотезой называется любое предположение о виде или о параметрах неизвестного распределения генеральной совокупности.

Не располагая сведениями обо всей генеральной совокупности, высказанную гипотезу сопоставляют по определенным правилам с выборочными данными и делают вывод о том, можно принять гипотезу или нет. Эта процедура сопоставления называется *проверкой гипотезы.*

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

1. Располагая выборочными данными и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0, которую называют *основной* или *нулевой,* и гипотезу H_1, *конкурирующую* с гипотезой H_0. Гипотезу H_1 называют также *альтернативной*, она является логическим отрицанием гипотезы H_0. Выбор тех или иных нулевых или альтернативных гипотез определяется решаемыми исследователем прикладными задачами.

2. Задается вероятность α, которую называют *уровнем значимости.*

Уровень значимости α определяет вероятность так называемой ошибки первого рода, которая совершается при отвержении гипотезы H_0, т.е. принимается конкурирующая гипотеза H_1, тогда как на самом деле гипотеза H_0 верна. Вероятность α задается заранее малым числом: 0,1; 0,05, 0,001 и т.д.

3. Выбирается статистический критерий проверки гипотезы – K. *Статический критерий* – это случайная величина, закон распределения которой при условии справедливости проверяемой гипотезы H_0 известен.

После выбора критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при котором нулевая гипотеза отвергается – *критическая область* Q, а другое содержит те значения критерия, при которых гипотеза принимается – *область принятия гипотезы*. *Критическими точками* $K_{кр}$ называются точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области. *Правосторонней* (*левосторонней*) называют критическую область, определяемую неравенством $K > K_{кр}$ ($K < K_{кр}$). *Двусторонней* называют критическую область, определяемую неравенствами $K < K_{кр1} \cup K > K_{кр2}$.

4. По результатам эксперимента находят эмпирическое (наблюдаемое) значение статистического критерия K. Если наблюдаемое значение критерия принадлежит критической области, то нулевую гипотезу отвергают в пользу конкурирующей гипотезы; если наблюдаемое значение критерия принадлежит области принятия гипотезы, то нулевую гипотезу принимают.

5. Результат проверки гипотезы формулируется следующим образом: гипотеза H_0 проверена по критерию K на уровне значимости α и принята (не противоречит имеющимся экспериментальным данным) или отвергнута.

Проверка гипотезы о равенстве средних двух нормально распределенных совокупностей с неизвестными, но равными дисперсиями по малым выборкам ($n \le 25$)

Пусть имеются две нормально распределенные генеральные совокупности X и Y, характеризуемые генеральными средними $\bar{x}_0$ и $\bar{y}_0$. Для проверки гипотезы из этих совокупностей берутся две независимые выборки объемов n_1 и n_2, по которым находят выборочные средние $\bar{x}$, $\bar{y}$ и исправленные выборочные дисперсии s_x^2, s_y^2.

1. Нулевая гипотеза H_0: $\bar{x}_0 = \bar{y}_0$.

Альтернативная гипотеза H_1: а) $\bar{x}_0 > \bar{y}_0$ ($\bar{x}_0 < \bar{y}_0$);

б) $\bar{x}_0 \neq \bar{y}_0$.

2. Уровень значимости α.

3. Статистический критерий: $$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{n_1 s_x^2 + n_2 s_y^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (22)$$

Критерий t имеет распределение Стьюдента с $k = n_1 + n_2 - 2$ степенями свободы.

а) При альтернативной гипотезе $\bar{x}_0 > \bar{y}_0$ ($\bar{x}_0 < \bar{y}_0$) критическая область является односторонней и определяется неравенством $|t| > t_{1-2\alpha,k}$. Критическая точка определяется по таблице значений $t_{\gamma,k}$ распределения Стьюдента, где $k = n_1 + n_2 - 2$, $\gamma = 1 - 2\alpha$.

б) При альтернативной гипотезе $\bar{x}_0 \neq \bar{y}_0$ критическая область является двусторонней и определяется неравенством $|t| > t_{1-\alpha,k}$. Критическая точка определяется по таблице значений $t_{\gamma,k}$ распределения Стьюдента, где $k = n_1 + n_2 - 2$, $\gamma = 1 - \alpha$.

4. По формуле (22) определяем эмпирическое значение t-критерия.

Гипотеза H_0 принимается, если: а) $|t| \leq t_{1-2\alpha,k}$;

б) $|t| \leq t_{1-\alpha,k}$.

5. Делается вывод о результатах проверки гипотезы H_0.

2.7. Основные понятия корреляционного и регрессионного анализов

При одновременном изучении нескольких признаков какого-либо объекта или учете нескольких показателей в эксперименте возникает вопрос о взаимосвязях между исследуемыми величинами. Наиболее разработанными в математической статистике методами анализа взаимосвязей являются *корреляционный* и *регрессионный анализы*.

При изучении взаимосвязи признаки делятся на два класса:

- признаки, обуславливающие изменения других признаков, называются *факторными*, или *факторам*;
- признаки, изменяющиеся под действием факторных признаков, называются *результативными*.

Связь называется *статистической,* если каждому значению факторного признака соответствует определенное (условное) распределение результативного признака. *Корреляционной связью* называется частный случай статистической связи, состоящий в том, что разным значениям факторного признака соответствуют *различные средние значения* результативного.

Корреляционная зависимость между признаками X и Y может быть представлена в виде уравнения:

$$M_x(Y) = \varphi(x),$$

где $M_x(Y)$ – условное математическое ожидание признака Y при заданном $X = x$. Это уравнение называется *теоретическим уравнением регрессии* (или *функцией регрессии*) Y на X, а его график – *теоретической линией регрессии*.

2.8. Парная регрессия

В зависимости от вида функции $\varphi(x)$ различают *линейную* и *нелинейную* регрессию.

Для отыскания теоретического уравнения регрессии необходимо знать закон распределения двумерной случайной величины (X,Y). Но на практике исследователь располагает выборкой пар значений (x_i, y_i) ограниченного объема n. В этом случае можно построить лишь наилучшую оценку для функции регрессии, которой является *выборочное уравнение регрессии* $y_x = \varphi(x, a_0, a_1, \ldots, a_p)$ Y *на* X (или просто *уравнение регрессии*), где y_x – условная средняя признака Y при фиксированном значении признака $X = x$, $a_0, a_1, \ldots, a_p$ – параметры уравнения регрессии.

Так, например, оценкой линейного уравнения регрессии Y на X является выборочное уравнение регрессии $y_x = a_0 + a_1 x$.

Параметры a_0 и a_1 выборочного уравнения регрессии находятся следующим образом:

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}; \qquad (23)$$

$$a_0 = \bar{y} - a_1 \bar{x}, \qquad (24)$$

где $\bar{x}$ – выборочная средняя факторного признака X, $\bar{y}$ – выборочная средняя результативного признака Y, $\overline{xy} = \frac{1}{n}\sum x_i y_i$ – средняя из произведений соответствующих значений факторного и результативного признаков, σ_x^2 – выборочная дисперсия факторного признака X.

Коэффициент a_1 в уравнении регрессии называется *коэффициентом регрессии* (выборочным). Он показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

2.9. Выборочный коэффициент корреляции

Измерение тесноты и направления связи является важной задачей изучения и количественного измерения взаимосвязи явлений.

Если известно (или предполагается), что между результативным и факторным признаками существует линейная связь, то для оценки ее тесноты используется *выборочный коэффициент корреляции* r (или просто *коэффициент корреляции*). Он чаще всего рассчитывается по формуле:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \sigma_y}. \tag{25}$$

Коэффициент корреляции изменяется в пределах от –1 до +1. Равенство коэффициента нулю свидетельствует об отсутствии линейной связи**.** Равенство коэффициента ± 1 показывает наличие функциональной связи. Знак «+» указывает на прямую связь (увеличение или уменьшение одного признака сопровождается аналогичным изменением другого признака), знак «–» – на обратную связь (увеличение или уменьшение одного признака сопровождается противоположным по направлению изменением другого признака).

В зависимости от того, насколько $|r|$ приближается к 1, различают линейную связь слабую – $0{,}1 < |r| < 0{,}3$, умеренную – $0{,}3 < |r| < 0{,}5$, заметную – $0{,}5 < |r| < 0{,}7$, достаточно тесную – $0{,}7 < |r| < 0{,}9$ и весьма тесную – $0{,}9 < |r| < 0{,}99$.

В отличие от коэффициента регрессии a_1 коэффициент корреляции r не зависит от принятых единиц измерения признаков, а, следовательно, он сравним для любых признаков.

Как любая статистическая величина, коэффициент корреляции подвержен случайным колебаниям в результате выборочности исследования.

Для оценки *значимости коэффициента корреляции* применяется t-критерий Стьюдента. При этом определяется эмпирическое значение критерия t_r:

$$t_r = |r| \sqrt{\frac{n-2}{1-r^2}}. \tag{26}$$

Вычисленное по формуле (27) значение t_r сравнивается с критическим, которое берется из таблицы значений $t_{\gamma,k}$ распределения Стьюдента с учетом заданного уровня значимости α $(\gamma = 1-\alpha)$ и числа степеней свободы $k = n-2$.

Если $t_r > t_{1-\alpha,k}$, то величина коэффициента корреляции признается значимой.

РЕШЕНИЕ ПРИМЕРНОГО ВАРИАНТА КОНТРОЛЬНОЙ РАБОТЫ

Задача 1. По каналу связи передаются три сообщения. Вероятность того, что первое сообщение будет искажено равна 0,1, второе – 0,2, третье – 0,3. Найти вероятности следующих событий: A – все три сообщения переданы без искажения; B – ровно одно сообщение передано без искажения; C – хотя бы одно сообщение искажено.

Решение.

Введем в рассмотрение вспомогательные события A_k – k-ое сообщение передано без искажений, $\overline{A}_k$ – k-ое сообщение искажено, $k = 1,2,3$. Согласно условию $P(\overline{A}_1) = 0{,}1$, тогда $P(A_1) = 1 - P(\overline{A}_1) = 1 - 0{,}1 = 0{,}9$. Аналогично, $P(\overline{A}_2) = 0{,}2$ и $P(A_2) = 0{,}8$, $P(\overline{A}_3) = 0{,}3$ и $P(A_3) = 0{,}7$.

Так как событие A можно представить в виде $A = A_1A_2A_3$ и события A_1, A_2, A_3 независимы, то вероятность события A можно найти по теореме умножения вероятностей для независимых событий:

$$P(A) = P(A_1A_2A_3) = P(A_1)\cdot P(A_2)\cdot P(A_3)$$

$$P(A) = 0{,}7\cdot 0{,}8\cdot 0{,}9 = 0{,}504.$$

Событие B можно представить следующим образом:

$$B = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3,$$

причем слагаемые $A_1\overline{A}_2\overline{A}_3$, $\overline{A}_1A_2\overline{A}_3$ и $\overline{A}_1\overline{A}_2A_3$ являются попарно несовместными событиями. Поэтому на основании теоремы сложения вероятностей (1) получаем:

$$P(B) = P(A_1\overline{A}_2\overline{A}_3) + P(\overline{A}_1 A_2\overline{A}_3) + P(\overline{A}_1\overline{A}_2 A_3).$$

Для вычисления вероятностей событий $A_1\overline{A}_2\overline{A}_3$, $\overline{A}_1 A_2\overline{A}_3$ и $\overline{A}_1\overline{A}_2 A_3$ используем теорему умножения вероятностей:

$$P(A_1\overline{A}_2\overline{A}_3) = P(A_1)\cdot P(\overline{A}_2)\cdot P(\overline{A}_3) = 0{,}9\cdot 0{,}2\cdot 0{,}3 = 0{,}054;$$

$$P(\overline{A}_1 A_2\overline{A}_3) = P(\overline{A}_1)\cdot P(A_2)\cdot P(\overline{A}_3) = 0{,}1\cdot 0{,}8\cdot 0{,}3 = 0{,}024;$$

$$P(\overline{A}_1\overline{A}_2 A_3) = P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot P(A_3) = 0{,}1\cdot 0{,}2\cdot 0{,}7 = 0{,}014.$$

Таким образом, окончательно получаем:

$$P(B) = 0{,}054 + 0{,}024 + 0{,}014 = 0{,}092.$$

События A и C являются противоположными, следовательно,

$$P(C) = 1 - P(A) = 1 - 0{,}504 = 0{,}496.$$

Ответы: $P(A) = 0{,}504$, $P(B) = 0{,}092$, $P(C) = 0{,}496$.

Задача 2. Стрелок ведет стрельбу по цели с вероятностью попадания при каждом выстреле 0,4. За каждое попадание он получает 5 очков, а в случае промаха очков ему не начисляют. Составить закон распределения случайной величины ξ – числа очков, полученных стрелком за 3 выстрела, построить многоугольник распределения, вычислить математическое ожидание, дисперсию и среднее квадратическое отклонение этой случайной величины.

Решение.

Случайная величина ξ может принимать 4 значения:

0 – если стрелок промахнулся 3 раза;

5 – если стрелок попал 1 раз при трех выстрелах;

10 – если стрелок попал 2 раза при трех выстрелах;

15 – если стрелок попал 3 раза.

Так как каждый выстрел можно рассматривать, как независимое испытание, в результате которого возможны только два исхода: попадание («успех») или промах («неудача»), то вероятности, соответствующие каждому значению случайной величины, можно найти по формуле Бернулли (5):

$$P(\xi=m)=P_n(m)=C_n^m p^m q^{n-m}.$$

По условию задачи имеем: число испытаний $n=3$, вероятность успеха $p=0,4$, $q=1-p=1-0,4=0,6$, значения m будут изменяться от 0 до 3. Т.о. имеем:

$$p_0=p(\xi=0)=C_3^0\,0,4^0\cdot 0,6^3=0,216,$$

$$p_1=p(\xi=5)=C_3^1\,0,4^1\cdot 0,6^2=0,432,$$

$$p_2=p(\xi=10)=C_3^2\,0,4^2\cdot 0,6^1=0,288,$$

$$p_3=p(\xi=15)=C_3^3\,0,4^3\cdot 0,6^0=0,064.\text{о}$$

Следовательно, окончательно закон распределения случайной величины ξ будет иметь вид:

x_i	0	5	10	15
p_i	0,216	0,432	0,288	0,064

Построим многоугольник распределения. Для этого по оси абсцисс отложим возможные значения случайной величины, а по оси ординат – соответствующие им вероятности и соединяем точки (x_i, p_i) отрезками прямых. Полученная при этом ломаная линия и есть многоугольник распределения вероятностей случайной величины ξ.

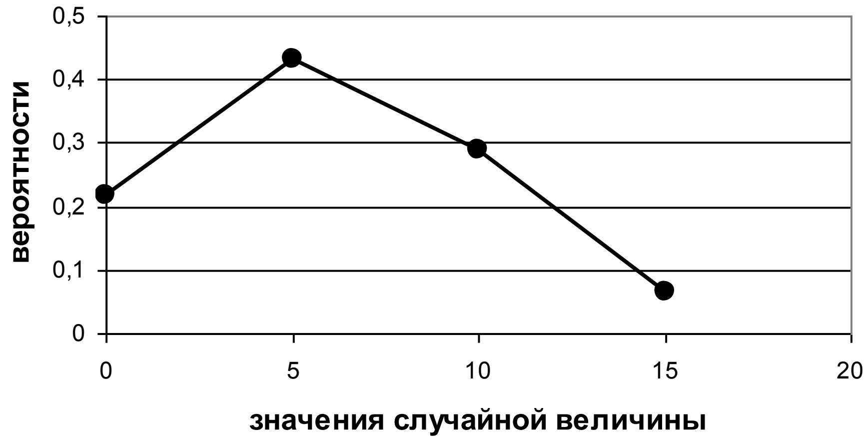

Рис. 1. Многоугольник распределения вероятностей

Рассчитаем числовые характеристики случайной величины ξ.

1. Математическое ожидание вычисляем по формуле (7)

$$M\xi = \sum_i x_i p_i = 0 \cdot 0{,}216 + 5 \cdot 0{,}432 + 10 \cdot 0{,}288 + 15 \cdot 0{,}064 = 6.$$

2. Дисперсия вычисляется по формуле (9):

$$D\xi = \sum_i x_i^2 \cdot p_i - (M\xi)^2 = 0^2 \cdot 0{,}216 + 5^2 \cdot 0{,}432 + 10^2 \cdot 0{,}288 + 15^2 \cdot 0{,}064 - 6^2 = 18.$$

3. Среднее квадратическое отклонение

$$\sigma_\xi = \sqrt{D\xi} = \sqrt{18} \approx 4{,}2.$$

Ответ. Закон распределения случайной величины ξ:

x_i	0	5	10	15
p_i	0,216	0,432	0,288	0,064

многоугольник распределения – на рисунке 1, $M\xi = 6$, $D\xi = 18$, $\sigma_\xi = \sqrt{18} \approx 4{,}2$.

Задача 3. Случайная величина ξ распределена по нормальному закону с математическим ожиданием $a = 10$ и дисперсией $\sigma^2 = 4$. Найти вероятность того, что в результате испытания ξ примет значение, заключенное в интервале $(12;14)$.

Решение.

Так как случайная величина ξ имеет нормальное распределение, то вероятность ее попадания в интервал можно найти по формуле (11). Учитывая, что по условию имеем: $\alpha = 12$, $\beta = 14$, $a = 10$, $\sigma^2 = 4$, то получим:

$$P(12 < \xi < 14) = \Phi\left(\frac{14-10}{2}\right) - \Phi\left(\frac{12-10}{2}\right) = \Phi(2) - \Phi(1).$$

По таблице значений функции Лапласа $\Phi(x)$ находим: Φ(2)=0,4772, Φ(1)=0,3413. Значит, получаем: $P(12 < \xi < 14) = 0{,}1359$.

Ответ: $P(12 < \xi < 14) = 0{,}1359$

Задача 4. По выборке из генеральной совокупности нормально распределенного количественного признака X найти: 1) числовые характеристики выборки – выборочную среднюю, выборочную дисперсию, среднее квадратическое отклонение; 2) несмещенные оценки для генеральной средней и генеральной дисперсии; 3) доверительный интервал для оценки генеральной средней с надежностью $\gamma = 0{,}99$.

x_i	33,2	38,2	43,2	48,2	53,2
n_i	2	4	10	3	1

Решение.

1. Сначала вычислим числовые характеристики выборки.

Выборочную среднюю найдем по формуле (14).

Учитывая, что объем выборки $n = 2 + 4 + 10 + 3 + 1 = 20$, получаем:

$$\bar{x} = \frac{1}{20}(33{,}2 \cdot 2 + 38{,}2 \cdot 4 + 43{,}2 \cdot 3 + 48{,}2 \cdot 3 + 53{,}2 \cdot 1) = 42{,}45.$$

Выборочную дисперсию удобнее вычислять по формуле (16):

$$\sigma_x^2 = \frac{1}{20}(33{,}2^2 \cdot 2 + 38{,}2^2 \cdot 4 + 43{,}2^2 \cdot 3 + 53{,}2^2 \cdot 1) - 42{,}45^2 = 23{,}1875.$$

Выборочное СКО:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{23{,}1875} \approx 4{,}82.$$

2. Несмещенной оценкой для генеральной средней $\bar{x}_0$ является выборочная средняя $\bar{x} = 42{,}45$.

Несмещенной оценкой дисперсии σ_0^2 генеральной совокупности является исправленная выборочная дисперсия s_x^2, которая вычисляется по формуле (17):

$$s_x^2 = \frac{20}{19} \cdot 23{,}1875 \approx 24{,}41.$$

3. Так как генеральная дисперсия σ_0^2 неизвестна, а известна лишь ее оценка – исправленная выборочная дисперсия s_x^2 и данная выборка имеет небольшой объем ($n < 30$), то доверительный интервал для генеральной средней можно найти, используя формулы (19) и (21).

Значение $t_{\gamma,n-1}$ находим по таблице распределения Стьюдента, где $\gamma = 0{,}99$ – доверительная вероятность, $n = 20$ – объем выборки, $n-1=19$ - число степеней свободы.

Учитывая, что $\bar{x} = 42{,}45$, $s_x = \sqrt{24{,}41} \approx 4{,}94$, $t_{0{,}99;19} = 2{,}86$, находим сначала точность оценки по формуле (21):

$$\Delta = 2{,}86 \cdot \frac{4{,}94}{\sqrt{20}} \approx 3{,}16.$$

Теперь искомый доверительный интервал определяем по формуле (19):

$$42{,}45 - 3{,}16 < \bar{x}_0 < 42{,}45 + 3{,}16$$

$$\text{или } 39{,}39 < \bar{x}_0 < 45{,}61.$$

Ответы: 1. $\bar{x} = 42{,}45$, $\sigma_x^2 = 23{,}1875$, $\sigma_x \approx 4{,}82$; 2. $\bar{x}_0 \approx \bar{x} = 42{,}45$, $\sigma_0^2 \approx s_x^2 \approx 24{,}41$; 3. $39{,}21 < \bar{x}_0 < 45{,}69$.

Задача 5. Массовую долю (%) оксида меди в минерале определили методом иодометрии и методом комплексометрии. По первому методу получили результаты: 38,20; 38,00; 37,66, а по второму: 37,70; 37,65; 37,55. Проверить, различаются ли средние результаты данных методов на уровне значимости $\alpha = 0{,}05$, если известно, что результаты измерений имеют нормальный закон распределения с неизвестными, но равными дисперсиями.

Решение.

Вычисляем для каждого метода числовые характеристики, учитывая, что объем каждой выборки равен $n_x = n_y = 3$:

- выборочные средние значения по формуле (14):

$$\bar{y} = \frac{1}{n_y}\sum y_i = \frac{1}{3}(37{,}70 + 37{,}65 + 37{,}55) = 37{,}63;$$

- исправленные выборочные дисперсии по формуле (18):

$$s_x^2 = \frac{1}{n_x - 1}\sum (x_i - \bar{x})^2,\ s_y^2 = \frac{1}{n_y - 1}\sum (y_i - \bar{y})^2$$

$$s_x^2 = \frac{1}{2}((38{,}20 - 37{,}95)^2 + (38{,}00 - 37{,}95)^2 + (37{,}66 - 37{,}95)^2) = 0{,}07453;$$

$$s_y^2 = \frac{1}{2}((37{,}70 - 37{,}63)^2 + (37{,}65 - 37{,}63)^2 + (37{,}55 - 37{,}63)^2) = 0{,}00583.$$

Теперь проверим гипотезу о равенстве средних двух совокупностей.

1. Нулевая гипотеза: H_0: $\bar{x}_0 = \bar{y}_0$.

Альтернативная гипотеза: H_1: $\bar{x}_0 \neq \bar{y}_0$

2. Уровень значимости $\alpha = 0{,}05$.

3. Проверку гипотезы будем проводить с помощью t-критерия, так как выборки маленькие и по условию дисперсии генеральных совокупностей неизвестны, но равны. По таблице значений $t_{\gamma,k}$ распределения Стьюдента при $\gamma = 1 - \alpha = 1 - 0{,}05 = 0{,}95$ и числе степеней свободы $k = 3 + 3 - 2 = 4$ находим критическое значение: $t_{0,95;4} = 2{,}78$.

4. Рассчитаем эмпирическое значение t-критерия, используя формулу (22):

$$t = \frac{37{,}95 - 37{,}63}{\sqrt{\frac{3 \cdot 0{,}07453 + 3 \cdot 0{,}00583}{3 + 3 - 2}\left(\frac{1}{3} + \frac{1}{3}\right)}} = 1{,}96.$$

Сравним полученное значение t с табличным значением $t_{0,95;4}$. Так как $|t| < t_{0,95;4}$, то гипотеза H_0 принимается.

5. Гипотеза о равенстве средних значений двух методов проверена на уровне значимости $\alpha = 0{,}05$ с помощью t-критерия и принята. Следовательно, результаты обоих методов отражают истинное содержание CuO в минерале.

Ответ: гипотеза H_0 о равенстве средних проверена на уровне значимости $\alpha = 0{,}05$ с помощью t-критерия и принята.

Задача 6. Имеются следующие данные об уровне механизации работ X (%) и производительности труда Y (т/чел.) для 14 однотипных предприятий:

№ п/п	1	2	3	4	5	6	7
x_i	30	32	36	40	41	47	54
y_i	24	20	28	30	31	33	37

№ п/п	8	9	10	11	12	13	14
x_i	55	56	60	61	67	69	76
y_i	40	34	38	41	43	45	48

Требуется: 1) оценить тесноту и направление связи между признаками с помощью коэффициента корреляции; 2) найти уравнение линейной регрессии Y на X; 3) в одной системе координат построить эмпирическую и теоретическую линии регрессии.

Решение.

1. Для удобства проведем все необходимые предварительные расчеты в таблице.

Таблица 1

Расчетная таблица

№ п/п	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	30	24	900	576	720
2	32	20	1024	400	640
3	36	28	1296	784	1008
4	40	30	1600	900	1200
5	41	31	1681	961	1271
6	47	33	2209	1089	1551
7	54	37	2916	1369	1998
8	55	40	3025	1600	2200
9	56	34	3136	1156	1904
10	60	38	3600	1444	2280
11	61	41	3721	1681	2501
12	67	43	4489	1849	2881
13	69	45	4761	2025	3105
14	76	48	5776	2304	3648
Всего	**724**	**492**	**40134**	**18138**	**26907**

Рассчитаем числовые характеристики выборки, используя итоговую строку расчетной таблицы и учитывая, что объем выборки $n=14$:

- выборочные средние:

$$\bar{x}=\frac{1}{n}\sum x_i=\frac{1}{14}\cdot 724=51,71;$$

$$\bar{y}=\frac{1}{n}\sum y_i=\frac{1}{14}\cdot 492=35,14;$$

- средние по квадратам:

$$\overline{x^2}=\frac{1}{n}\sum x_i^2=\frac{1}{14}\cdot 40134=2866,71;$$

$$\overline{y^2}=\frac{1}{n}\sum y_i^2=\frac{1}{14}\cdot 18138=1295,57;$$

- средняя по произведениям:

$$\overline{xy}=\frac{1}{n}\sum x_i y_i=\frac{1}{14}\cdot 26907=1921,93;$$

- выборочные средние квадратические отклонения:

$$\sigma_x^2=\overline{x^2}-\bar{x}^2=2866,71-51,71^2=192,79;\ \sigma_x=\sqrt{192,79}=13,88;$$

$$\sigma_y^2=\overline{y^2}-\bar{y}^2=1295,57-35,14^2=60,75;\ \sigma_y=\sqrt{60,75}=7,79.$$

Вычислим выборочный коэффициент корреляции по формуле (26):

$$r=\frac{1921,93-51,71\cdot 35,14}{13,88\cdot 7,79}=0,970.$$

Т.к. $r>0$ и $r\in(0,9;0,99)$, то, следовательно, линейная связь между изучаемыми признаками является прямой и весьма тесной.

2. Найдем уравнение линейной регрессии Y на X: $y_x=a_0+a_1x$, вычислив параметры уравнения регрессии по формулам (23) и (24):

$$a_1=\frac{1921,93-51,71\cdot 35,14}{192,79}=0,54;$$

$$a_0 = 35{,}14 - 0{,}54 \cdot 51{,}71 = 7{,}22.$$

Следовательно, уравнение прямой регрессии имеет вид:

$$y_x = 0{,}54x + 7{,}22.$$

3) Построим в одной системе координат эмпирическую и теоретическую линии регрессии. Эмпирическая линия – это ломаная, соединяющая точки с координатами (x_i, y_i), а теоретическая – это график прямой регрессии, уравнение которой было получено в п. 2. Теоретическую линию регрессии можно построить по двум точкам, абсциссы которых выбираются произвольно, а ординаты находятся по построенному уравнению регрессии. Найдем координаты точек для построения теоретической линии регрессии: $x_1 = 30$, тогда $y_1 = 0{,}54 \cdot 30 + 7{,}22 = 23{,}42$; $x_2 = 76$, $y_2 = 48{,}26$. Значит, теоретическую линию регрессии будем строить по двум точкам с координатами $(30;48{,}26)$ и $(76;48{,}26)$.

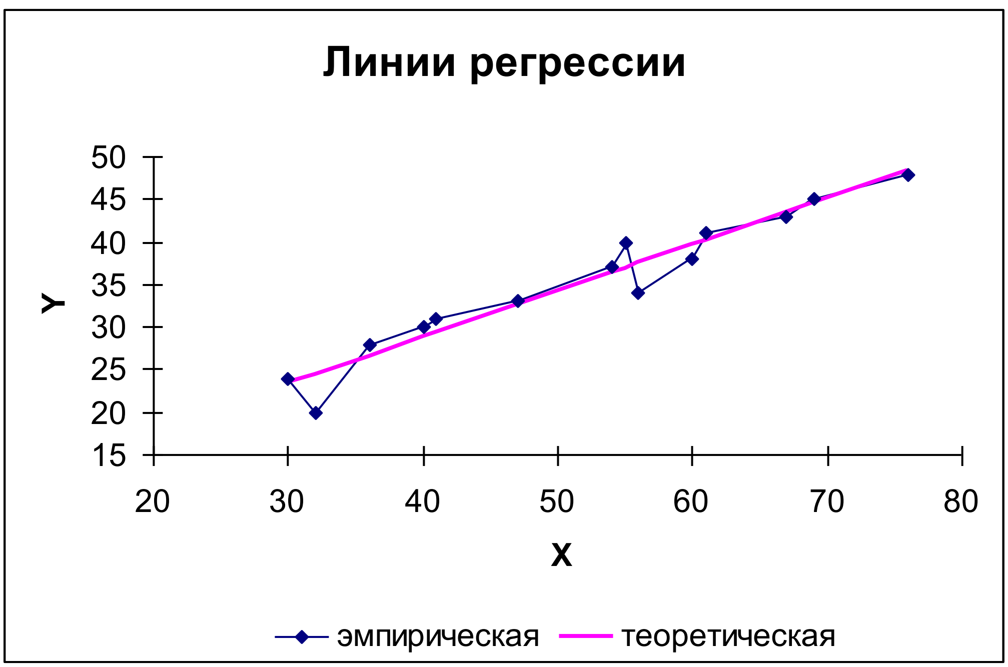

Рис. 2. Эмпирическая и теоретическая линии регрессии

Ответ: 1) $r = 0{,}970$, линейная связь прямая, весьма тесная, коэффициент корреляции значим на уровне значимости $\alpha = 0{,}05$; 2) выборочное уравнение прямой регрессии $y_x = 0{,}54x + 7{,}22$; 3) линии регрессии представлены на рис. 2.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная литература

1. Письменный, Д. Т. Конспект лекций по высшей математике : [полный курс] / Д. Т. Письменный. - 10-е изд., испр.- Москва : Айрис-пресс, 2011. - 602, [1] с. : ил. Количество экземпляров в библиотеке: абонемент – 212.
2. Сборник задач по курсу математического анализа : учеб. пособие / Г. Н. Берман. - [22-е изд., перераб.]. - Санкт-Петербург : Профессия, 2005, 2004, 2002, 2003, 2001. - 432 с. : ил. Количество экземпляров в библиотеке: абонемент – 781.

Дополнительная литература

1. Клетеник, Д.В. Сборник задач по аналитической геометрии : учеб. пособие для вузов / Д. В. Клетеник; под ред. Н. В. Ефимова. - 17-е изд., стер. - Санкт-Петербург : Профессия, 2007, 2003 ; Москва. - 200 с. : ил. Количество экземпляров в библиотеке: абонемент – 378.
2. Данко П. Е. , Попов А. Г., Кожевникова Т. Я., Данко С. П. Высшая математика в упражнениях и задачах: учеб. пособие / П. Е. Данко [и др.]. - 7-е изд., испр. - Москва: Оникс: Мир и Образование, 2008. - 815 с.: ил. Количество экземпляров в библиотеке: абонемент – 30.