МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт сервиса, туризма и дизайна (филиал) СКФУ в г. Пятигорске

УТВЕРЖДАЮ
Зам. директора по учебной работе
ИСТиД (филиал) СКФУ в г. Пятигорске
___________________ М.В. Мартыненко
«____» ______________________ 2020 г.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
по выполнению практических работ
по дисциплине
ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
(Наименование дисциплины)

Направление подготовки	**09.03.02 Информационные системы и технологии**
Направленность (профиль)	Информационные системы и технологии
Квалификация выпускника	Бакалавр
Форма обучения	Очная
Год начала обучения	2020
Изучается в 4 семестре	

Пятигорск 2020 г.

Методические указания по выполнению практических работ по дисциплине «Теория вероятностей и математическая статистика» рассмотрены и утверждены на заседании кафедры физики, электротехники и электроэнергетики (протокол №___ от «____» ______ 2020 г.).

Зав. кафедрой физики, электротехники и электроэнергетики ___________ А.В.Пермяков

Содержание

1.ЦЕЛЬ И СОДЕРЖАНИЕ

Целью освоения дисциплины «Теория вероятностей и математическая статистика» является формирование набора компетенций бакалавра по направлению подготовки 09.03.02 «Информационные системы и технологии».

Задачи освоения дисциплины: формирование представлений о роли и месте теории вероятностей и математической статистики в современном мире, этапах развития, универсальности ее понятий и представлений; формирование умений конструирования и анализа статистических моделей объектов, систем и процессов при решении задач, связанных со сферой будущей профессиональной деятельности; овладение навыками точного и сжатого выражения математической мысли в устном и письменном изложении, с использованием соответствующей символики.

Целью проведения практических занятий по дисциплине является развитие логического и алгоритмического мышления, формирование знаний по основным разделам дисциплины, необходимым студентам как для освоения базовых и вариативных дисциплин, так и для дальнейшей самостоятельной работы.

В ходе практического занятия студент учится логично, ясно, четко, грамотным математическим языком излагать свои мысли, приводить доводы, формулировать аргументы в защиту своей позиции.

На занятии студент опирается на свои конспекты, сделанные на лекции, собственные выписки из учебников и словарно-справочную литературу.

2. НАИМЕНОВАНИЕ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

№ темы	Наименование тем практических занятий	Объем часов	Форма проведения
	3 семестр		
1	Решение комбинаторных задач: перестановки, размещения, сочетания.	1,5	Решение разноуровневых задач
2	Вероятности и случайные процессы. Решение задач на непосредственное вычисление вероятности события.	1,5	Решение разноуровневых задач
3	Основные теоремы теории вероятностей. Применение теорем сложения и умножения при решении задач.	1,5	Решение разноуровневых задач
3	Решение задач на применение формул полной вероятности, формулы Байеса.	1,5	
3	Схема испытаний Бернулли. Наивероятнейшее число появлений события в схеме повторных независимых испытаний.	1,5	Решение разноуровневых задач
3	Приближенные формулы в схеме Бернулли.	1,5	
4	Закон распределения дискретной случайной величины.	1,5	Решение разноуровневых задач

4	Вычисление числовых характеристик дискретной случайной величины.	1,5	Решение разноуровневых задач
5	Функции распределения и плотности распределения вероятностей непрерывных случайных величин, их свойства.	1,5	
5	Числовые характеристики непрерывных случайных величин.	1,5	Решение разноуровневых задач
7	Вариационные ряды и их графическое изображение.	1,5	Решение разноуровневых задач
8	Выборочный метод. Точечные оценки.	1,5	
9	Понятие интервального оценивания. Доверительная вероятность и предельная ошибка выборки. Оценка характеристик генеральной совокупности по малой выборке.	1,5	
10	Статистическая проверка гипотез.	1,5	
11	Корреляционная зависимость. Коэффициент корреляции.	1,5	Решение разноуровневых задач
12	Линейная регрессия. Коэффициенты регрессии.	1,5	
	Итого за 4 семестр	**24**	

3.ВОПРОСЫ И ЗАДАЧИ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие 1. Решение комбинаторных задач: перестановки, размещения, сочетания.

Цель: формирование основных понятий комбинаторики и методов решения комбинаторных задач в приложении к теории вероятностей.

Теоретическая часть:

Комбинаторика - ветвь математики, изучающая комбинации и перестановки предметов, - возникла в XVII в. Долгое время казалось, что комбинаторика лежит вне основного русла развития математики и ее приложений. Положение дел резко изменилось после появления быстродействующих вычислительных машин и связанного с этим расцвета конечной математики. Сейчас комбинаторные методы применяются в теории случайных процессов, статистике, вычислительной математике, планировании экспериментов и т. д.

Комбинаторикой (от латинского combinare – соединять, сочетать) называют раздел математики, в котором изучаются задачи следующего типа: сколько комбинаций, удовлетворяющих тем или иным условиям, можно составить из элементов данного множества.

Элементарными комбинаторными конфигурациями являются сочетания, размещения, перестановки. Для подсчёта числа этих конфигураций используются правила суммы и произведения.

Правило суммы:

Если элемент A можно выбрать m способами, а элемент B можно выбрать k способами, то выбор элемента A или B можно осуществить m + k способами.

Обобщением правила суммы является правило произведения.

Правило произведения:

Если элемент A можно выбрать m способами, а после каждого выбора элемента A элемент B можно выбрать k способами, тогда, упорядоченную пару элементов (A, B) можно выбрать m*k способами.

Пример 1:

Сколько трехзначных четных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, 6, если цифры могут повторяться?

Решение: n_1=6 (т.к. в качестве первой цифры можно взять любую цифру из 1, 2, 3, 4, 5, 6), n_2=7 (т.к. в качестве второй цифры можно взять любую цифру из 0, 1, 2, 3, 4, 5, 6), n_3=4 (т.к. в качестве третьей цифры можно взять любую цифру из 0, 2, 4, 6). Итак, $N=n_1*n_2*n_3=6*7*4=168$.

Пример 2:

Сколько всех четырехзначных чисел можно составить из цифр 1, 5, 6, 7, 8?

Решение: Для каждого разряда четырехзначного числа имеется пять возможностей, значит N=5*5*5*5=54=625.

Размещения

Назовём множество, содержащее n элементов, n-множеством.

Последовательность (x_1, x_2, …, x_k) длины k без повторяющихся элементов из элементов данного n-множества назовём k-размещением.

Обозначим символом A_n^k число размещений из n по k элементов (от фран. "arrangement" - размещение). Используя правило произведения, вычислим число A_n^k.

Пусть произвольное размещение длины k имеет вид:

(x_1, x_2, …, x_k).

Элемент x_1 можно выбрать n способами. После каждого выбора x_1 элемент x_2 можно выбрать (n-1) способами. После каждого выбора элементов x_1 и x_2 элемент x_3 можно выбрать (n-2) способами, и т.д. После каждого выбора элементов x_1 , x_2, …, x_{k-1} элемент x_k можно выбрать n-(k-1)) = (n-k+1) способами. Тогда, по правилу произведения, последовательность (x_1; x_2; , …, x_k) можно выбрать числом способов, равным $n(n-1)(n-2) \ldots (n-k + 1) = A_n^k$.

Произведение в левой части равенства умножим и разделим на (n - k)!, получим

$$A_n^k = \frac{n!}{(n-k)!}$$

Если в форуме k = n, то A_n^k есть число P_n перестановок из n элементов $P_n=n!$ (от "permutation"- перестановка).

Пример 3:

Сколько существует двузначных чисел, в которых цифра десятков и цифра единиц различные и нечетные?

Решение: т.к. нечетных цифр пять, а именно 1, 3, 5, 7, 9, то эта задача сводится к выбору и размещению на две разные позиции двух из пяти различных цифр, т.е. указанных чисел будет:

$$A_2^5 = 4 \cdot 5 = 20\,.$$

Сочетания

k-подмножество данного n-множества называется k-сочетанием.

Обозначим через C_n^k число k-сочетаний из данных n элементов. Формулу для числа C_n^k получим, рассуждая следующим образом. Если каждое сочетание упорядочить всеми возможными способами, то получим все k-последовательностей из n элементов, без повторений, то есть все k-размещения.

Иными словами,

$$C_n^k \bullet k! = A_n^k.$$

Откуда

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{(n-k)!\,k!},$$

$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}.$$

Пример 4: Сколькими способами читатель может выбрать две книжки из шести имеющихся?

Решение: Число способов равно числу сочетаний из шести книжек по две, т.е. равно:

$$C_6^2 = \frac{6!}{2!4!} = \frac{5 \cdot 6}{2!} = 15.$$

Предполагая, что n и k - целые положительные числа и 0!=1, сформулируем основные свойства сочетаний.

Основные свойства сочетаний

1. Условились, что $C_n^0 = 1$
2. $C_n^1 = n$
3. $C_n^k = C_n^{n-k}$
4. $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$
5. $\sum_{i=0}^{n} C_n^i = 2^n$

Задачи и упражнения:

1.Определить, какой объект комбинаторики применяется для решения задачи, обосновать выбор.

2.Решить задачу.

1. Сколько трехзначных четных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, 6, если цифры могут повторяться?

2. Сколько существует пятизначных чисел, которые одинаково читаются слева направо и справа налево?

3. В классе десять предметов и пять уроков в день. Сколькими способами можно составить расписание на один день?

4. Сколькими способами можно выбрать 4 делегата на конференцию, если в группе 20 человек?

5. Сколькими способами можно разложить восемь различных писем по восьми различным конвертам, если в каждый конверт кладется только одно письмо?

6. Из трех математиков и десяти экономистов надо составить комиссию, состоящую из двух математиков и шести экономистов. Сколькими способами это можно сделать?

7. В киоске продают 5 видов конвертов и 4 вида марок. Сколькими способами можно купить конверт и марку?

8. Сколькими способами можно выбрать гласную и согласную буквы из слова КОНВЕРТ?

9. Начальник транспортного цеха пригласил несколько человек на совещание. Каждый участник совещания, входя в кабинет, пожимал руки всем присутствующим. Сколько человек участвовали в совещании, если было всего 78 рукопожатий?

10. Крыса бежит по лабиринту, который устроен так, что сначала она должна выбрать одну из двух дверей, затем одну из трёх дверей, а за каждой из них её ожидают четыре двери. Пройдя дверь, крыса не может вернуться через неё обратно. Сколькими различными путями крыса может пройти лабиринт от начала до конца?

Практическое занятие 2. Вероятности и случайные процессы. Решение задач на непосредственное вычисление вероятности события.

Цель: формирование понятий классической, геометрической, статистической вероятности события и способах их вычисления.

Теоретическая часть:

Вероятность - одно из основных понятий теории вероятностей. Существует несколько определений этого понятия. Приведем определение, которое называют классическим. Далее укажем слабые стороны этого определения и приведем другие определения, позволяющие преодолеть недостатки классического определения.

Рассмотрим пример. Пусть в урне содержится 6 одинаковых, тщательно перемешанных шаров, причем 2 из них - красные, 3 - синие и 1 - белый. Очевидно, возможность вынуть наудачу из урны цветной (т. е. красный или синий) шар больше, чем возможность извлечь белый шар. Можно ли охарактеризовать эту возможность числом? Оказывается, можно. Это число и называют вероятностью события (появления цветного шара). Таким образом, вероятность есть число, характеризующее степень возможности появления события.

Поставим перед собой задачу дать количественную оценку возможности того, что взятый наудачу шар цветной. Появление цветного шара будем рассматривать в качестве события А. Каждый из возможных результатов испытания (испытание состоит в извлечении шара из урны) назовем*элементарным исходом (элементарным событием)*. Элементарные исходы обозначим через ω_1, ω_2, ω_3 и т.д. В нашем примере возможны следующие 6 элементарных исходов: ω_1 - появился белый шар; ω_2, ω_3 - появился красный шар; ω_4, ω_5, ω_6 - появился синий шар. Легко видеть, что эти исходы образуют полную группу попарно несовместных событий (обязательно появится только один шар) и они равновозможны (шар вынимают наудачу, шары одинаковы и тщательно перемешаны).

Те элементарные исходы, в которых интересующее нас событие наступает, назовем *благоприятствующими* этому событию. В нашем примере благоприятствуют событию А (появлению цветного шара) следующие 5 исходов: ω_2, ω_3, ω_4, ω_5, ω_6.

Таким образом, событие А наблюдается, если в испытании наступает один, безразлично какой, из элементарных исходов, благоприятствующих А; в нашем примере А наблюдается, если наступит ω_2, или ω_3, или ω_4, или ω_5, или ω_6. В этом смысле событие А подразделяется на несколько элементарных событий (ω_2, ω_3, ω_4, ω_5, ω_6); элементарное же событие не подразделяется на другие события. В этом состоит различие между событием А и элементарным событием (элементарным исходом).

Отношение числа благоприятствующих событию А элементарных исходов к их общему числу называют вероятностью события А и обозначают через Р (А). В рассматриваемом примере всего элементарных исходов 6; из них 5 благоприятствуют событию А. Следовательно, вероятность того, что взятый шар окажется цветным, равна $P(A) = 5 / 6$. Это число и дает ту количественную оценку степени возможности появления цветного шара, которую мы хотели найти. Дадим теперь определение вероятности.

Вероятностью события А называют отношение числа благоприятствующих этому событию исходов к общему числу всех равновозможных несовместных элементарных исходов, образующих полную группу. Итак, вероятность события А определяется формулой

$$P(A) = m / n,$$

где m - число элементарных исходов, благоприятствующих А; n - число всех возможных элементарных исходов испытания.

Здесь предполагается, что элементарные исходы несовместны, равновозможны и образуют полную группу. Из определения вероятности вытекают следующие ее свойства:

С в о й с т в о 1. *Вероятность достоверного события равна единице.*

Действительно, если событие достоверно, то каждый элементарный исход испытания благоприятствует событию. В этом случае $m = n$, следовательно,

$$P(A) = m / n = n / n = 1.$$

С в о й с т в о 2. *Вероятность невозможного события равна нулю.*

Действительно, если событие невозможно, то ни один из элементарных исходов испытания не благоприятствует событию. В этом случае $m = 0$, следовательно,

$$P(A) = m / n = 0 / n = 0.$$

С в о й с т в о 3. *Вероятность случайного события есть положительное число, заключенное между нулем и единицей*.

Действительно, случайному событию благоприятствует лишь часть из общего числа элементарных исходов испытания. В этом случае $0 < m < n$, значит, $0 < m / n < 1$, следовательно,

$$0 < P(A) < 1.$$

Итак, вероятность любого события удовлетворяет двойному неравенству

$$0 <= P(A) < 1.$$

Задачи и упражнения:

1. В ящике имеется 50 одинаковых деталей, из них 5 окрашенных. Наудачу вынимают одну деталь. Найти вероятность того, что извлеченная деталь окажется окрашенной.

2. Брошена игральная кость. Найти вероятность того, что выпадет четное число очков.

3. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100. Найти вероятность того, что номер первого наудачу извлеченного жетона не содержит цифры 5.

4. В мешочке имеется 5 одинаковых кубиков. На всех гранях каждого кубика написана одна из следующих букв: о, п, р, с, т. Найти вероятность того, что

на вынутых по одному и расположенных "в одну линию" кубиков можно будет прочесть слово "спорт".

5. На каждой из шести одинаковых карточек напечатана одна из следующих букв: а, т, м, р, с, о. Карточки тщательно перемешаны. Найти вероятность того, что на четырех, вынутых по одной и расположенных "в одну линию" карточках можно будет прочесть слово "трос".

6. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера, которые затем тщательно перемешаны. Найти вероятность того, что наудачу извлеченный кубик будет иметь окрашенных граней: а) одну; б) две; в) три.

7. Из тщательно перемешанного полного набора 28 костей домино наудачу извлечена кость. Найти вероятность того, что вторую наудачу извлеченную кость можно приставить к первой, если первая кость: а) оказалась дублем; б) не есть дубль.

8. В замке на общей оси пять дисков. Каждый диск разделен на шесть секторов, на которых написаны различные буквы. Замок открывается только в том случае, если каждый диск занимает одно определенное положение относительно корпуса замка. Найти вероятность того, что при произвольной установке дисков замок можно будет открыть.

9. Восемь различных книг расставляются наудачу на одной полке. Найти вероятность того, что две определенные книги окажутся поставленными рядом.

10. Библиотечка состоит из десяти различных книг, причем пять книг стоят по 4 рубля каждая, три книги — по одному рублю и две книги — по 3 рубля. Найти вероятность того, что взятые наудачу две книги стоят 5 рублей.

11. В партии из 100 деталей отдел технического контроля обнаружил 5 нестандартных деталей. Чему равна относительная частота появления нестандартных деталей?

12. При стрельбе из винтовки относительная частота попадания в цель оказалась равной 0,85. Найти число попаданий, если всего было произведено 120 выстрелов.

Практическое занятие 3. Основные теоремы теории вероятностей. Применение теорем сложения и умножения при решении задач.

Цель: формирование представления об основных теоремах теории вероятностей и методах их применения для решения задач.

Теоретическая часть:

Суммой двух событий А и В называется событие С, состоящее в появлении хотя бы одного из событий А или В.

Теорема сложения вероятностей

Вероятность суммы двух несовместимых событий равна сумме вероятностей этих событий:

$P(A + B) = P(A) + P(B).$

В случае, когда события A и B совместны, вероятность их суммы выражается формулой

$P(A+B) = P(A) + P(B) - P(AB)$,

где AB – произведение событий A и B.

Два события называются зависимыми, если вероятность одного из них зависит от наступления или не наступления другого. В случае зависимых событий вводится понятие условной вероятности события.

Условной вероятностью $P(A/B)$ события A называется вероятность события A, вычисленная при условии, что событие B произошло. Аналогично через $P(B/A)$ обозначается условная вероятность события B при условии, что событие A наступило.

Произведением двух событий A и B называется событие C, состоящее в совместном появлении события A и события B.

Теорема умножения вероятностей

Вероятность произведения двух событий равна вероятности одного из них, умноженной на условную вероятность другого при наличии первого:

$P(AB) = P(A) \cdot P(B/A)$, или $P(AB) = P(B) \cdot P(A/B)$.

Следствие. Вероятность совместного наступления двух независимых событий A и B равна произведению вероятностей этих событий:

$P(AB) = P(A) \cdot P(B)$.

Следствие. При производимых n одинаковых независимых испытаниях, в каждом из которых событие A появляется с вероятностью p, вероятность появления события A хотя бы один раз равна 1 - (1 - p)n.

Задачи и упражнения:

1. Менеджер имеет 8 кандидатов для взятия в штат фирмы, при этом 6 женщин и 3 мужчин. Требуется взять 3 человека. Какова вероятность того, что при произвольном выборе 3 – х людей будет выбрана хотя бы одна женщина.
2. В ящике 12 шаров, из которых 3 белых, 4 красных и 5 голубых. Наудачу извлечены 3 шара. Найти вероятность, что все 3 шара разного цвета.
3. Из группы студентов, в которой 18 юношей и 12 девушек, в совет факультета избираются два человека. Какова вероятность того, что среди избранных окажется хотя бы один юноша.
4. В корзине 6 белых и 8 черных мячей. Наудачу извлекается 3 мяча. Рассчитать вероятности:

   A - все 3 мяча белые;

   B - один мяч белый а два черных;

   C - все три мяча черных
5. Какова вероятность того,что при подбрасывании двух монет выпадет хотя бы один герб? Выписать пространство элементарных исходов.

6. Среди 30 билетов 5 выигрышных. Наити вероятность того, что среди двух выбранных наугад билетов хотя бы один выигрышный

7. Имеется 5 белых и 8 черных шаров. Вынимают два. Найти вероятность того ,что они разного цвета.

8. Имеется 2 урны: в первой a белых и b черных шаров; во второй c белых и d черных. Из каждой урны вынимают по шару. Найти вероятность того, что оба шара будут белыми (событие A) и вероятность того что оба шара будут разного цвета (событие B).

9. Какова вероятность что при одновременном бросании двух правильных игральных кубиков выпавшее число очков составит величину равную 9.

10. В корзине 6 белых и 8 черных мячей Рассчитаь вероятность что наудачу взятые два мяча белые.

11. В корзине 6 белых и 8 черных одинаковых по размеру и весу шариков. Наугад выбираются 3 шарика. Рассчитать вероятности следующих событий:

- Все три шарика белые;
- Один шарик белый, а два других черные;
- Все три шарика черные

12. Правильный игральный кубик бросается 3 раза. Посчитать вероятности следующих событий:

- Все три раза выпадет шестерка;
- Все три раза выпадет одно и тоже число.

13. Три раза бросается монета достоинством 5 сантимов. Найти вероятности следующих случайных событий:

- A –“все три раза выпадет цифра”,
- B - “цифра выпадет 1 раз”,
- C - “цифра выпадет большое число раз, чем герб”,
- D - “цифра выпадет 2 раза”.

14. Из букв слова “полюс” наугад выбираются две буквы дна за другой. Выписать пространство элементарных исходов.

15. Контролер, проверяя качество 400 изделий, установил, что 20 из них относятся ко 2 сорту, а остальные к первому. Найти частоту изделий первого сорта, частоту изделий 2 сорта.

16. Студент знает 14 вопросов из 20. В билете содержится 3 вопроса. Найти вероятность того, что студент ответит хотя бы на один из них.

17. Из ящика, в котором 90 годных деталей и 10 негодных деталей наудачу выбраны 10 деталей. Рассчитать вероятность того, что все выбранные детали годные.

18. В ящике 20 деталей, 15 из которых окрашены. Из ящика выбираются произвольным образом 4 детали. Найти вероятность, что они все окрашены

19. В ящике 5 белых шаров, 2 черных и 3 красных. Какова вероятность, что два вынутых шара будут одного цвета.

20.Устройство состоит из пяти элементов, из которых два изношены. При включении устройства включаются случайным образом два элемента. Найти вероятность того, что включенными окажутся изношенные элементы.

21.Из n изделий, среди которых k бракованных, наудачу берется k изделий. Какова вероятность, что все они бракованные

22.В урне a белых и b черных шаров (a $\geq 2, b \geq 3$)Из урны вынимают сразу 5 шаров. Найти вероятность p того, что 2 из них будут белыми и 3 черными.

23.Игральная кость бросается один раз. Найти вероятности событий -

- а: на кости выпало четное число очков
- б:на кости выпало не менее 5 очков
- с: на кости выпало не более 5 очков

Практическое занятие 4. Решение задач на применение формул полной вероятности, формулы Байеса.

Цель: формирование представления об основных теоремах теории вероятностей и методах их применения для решения задач.

Теоретическая часть:

Вероятность события A, которое может наступить лишь при появлении одного из несовместных событий-гипотез B_1, B_2,…,B_n, образующих полную группу, равна сумме произведений вероятностей каждой из гипотез на соответствующую условную вероятность события A:

$P(A) = P(B_1)*P_{B1}(A) + P(B_2)*P_{B2}(A) +…+ P(Bn)*P_{Bn}(A)$,

где $P(B_1) + P(B_2) +…+ P(Bn) = 1$. Приведенную формулу называют *формулой полной вероятности.*

Если событие A уже произошло, то вероятности гипотез могут быть переоценены по формулам Байеса:

Пример.

$$P_A(B_i) = \frac{P(B_i) * P_{B_i}(A)}{P(A)} \quad (i = 1,2,...,n)$$

Пять курсантов одинаковое число раз стреляют по отдельным пяти мишеням. Результат стрельбы таков:

- первый и второй курсанты поражают три мишени и имеют два промаха (событие B_1);
- третий курсант поражает две мишени и имеет три промаха (событие B_2);
- четвертый и пятый курсанты поражают одну мишень и имеют четыре промаха (событие B_3).

Наугад берется мишень, она оказалась пораженной (событие A). Определить вероятность того, что эта мишень поражена первым курсантом.

Решение: $P(B_1) = 2/5$, $P(B_2) = 1/5$, $P(B_3) = 2/5$, $P_{B1}(A) = 3/5$, $P_{B2}(A) = 2/5$, $P_{B3}(A) = 1/5$. По формуле Байеса определим искомую вероятность:

$$P_A(B_1) = \frac{P(B_1) * P_{B_1}(A)}{P(A)}$$

$$P_A(B_1)=\frac{\frac{2}{5}*\frac{3}{5}}{\frac{2}{5}*\frac{3}{5}+\frac{1}{5}*\frac{2}{5}+\frac{2}{5}*\frac{1}{5}}=0{,}6$$

Задачи и упражнения:

Вариант 1. При исследовании криминогенной обстановки в некотором городе все правонарушения, совершенные за последние 5 месяцев, были распределены по трем группам в соответствии с районами города. В первой группе оказалось 70 %, во второй – 23 %, в третьей – 7 % всех правонарушений. Вероятность того, что правонарушителем является человек младше 25 лет для каждой группы равна соответственно 0,6; 0,35 и 0,1. Определить вероятность того, что выбранному произвольно правонарушителю окажется меньше 25 лет.

Выбранному произвольно правонарушителю оказалось меньше 25 лет. Найти вероятность того, что этот человек из первой группы.

Вариант 2. Стрелковое отделение получило 10 винтовок, из которых 8 пристрелянных, две нет. Вероятность попадания в цель из пристрелянной винтовки равна 0,6, а из не пристрелянной 0,4. Какова вероятность, что стрелок из наудачу взятой винтовки попадет в цель при одном выстреле?

Стрелок поразил цель. Какова вероятность, что он стрелял из пристрелянной винтовки?

Вариант 3. В районе 24 человека обучаются на заочном факультете института, из них 6 – на физико-математическом факультете, 12 – на агрономическом факультете, 6 – на экономическом факультете. Вероятность успешно сдать все экзамены на предстоящей сессии для студентов физико-математического факультета равна 0,6, агрономического факультета – 0,76 и экономического факультета – 0,8. Найти вероятность того, что наудачу взятый студент, сдавший успешно все экзамены, окажется студентом экономического факультета.

Какова вероятность, что студент сдаст экзамен?

Вариант 4. В группе милицейского колледжа г. Брюково, состоящей из равного количества юношей и девушек, 5% всех юношей и 0,25% всех девушек страдают дальтонизмом. Какова вероятность того, что наугад выбранное лицо страдает дальтонизмом? Если наугад выбранное лицо страдает дальтонизмом, то какова вероятность того, что это юноша?

Вариант 5. На предприятии, изготавливающем дверные замки для автомобилей, первый цех производит 25%, второй – 35%, третий – 40% всех замков. Вероятность, что замок будет вскрыт угонщиками, по статистике составляет 5%, 4%, 2% соответственно. Найти вероятность того, что в случайно выбранном автомобиле с замками данного завода, будет вскрыта дверь. Если замок оказался взломан, какова вероятность того, что он был произведен в третьем цехе завода?

Вариант 6. На промышленной зоне ИТК 023/1 работают две бригады рабочих: первая производит в среднем ¾ продукции с процентом брака 4%, вторая – ¼ продукции с процентом брака 6%. Найти вероятность того, что взятое наугад изделие а) окажется бракованным; б) изготовлено второй бригадой при условии, что изделие окажется бракованным.

Вариант 7. В пирамиде 10 винтовок, из которых 4 снабжены оптическим прицелом. Вероятность того, что стрелок поразит мишень при выстреле из винтовки с оптическим прицелом, равна 0,95; для винтовки без оптического прицела – 0,8. Стрелок поразил мишень из наудачу взятой винтовки. Что вероятнее: стрелок стрелял из винтовки с оптическим прицелом или без него?

Вариант 8. В специализированную больницу поступают в среднем 50% с заболеванием Т, 30% - с заболеванием К, 20% - с заболеванием М. Вероятность полного излечения болезни Т равна 0,7; для болезней К и М эти вероятности соответственно равны 0,8 и 0,9. Какова вероятность того, что поступивший больной будет выписан здоровым? Если больной выписан здоровым, какова вероятность того, что этот больной страдал заболеванием М?

Вариант 9. В компьютерной базе данных содержатся данные об угонах автомашин по трем районам города. В среднем за месяц из первого района поступает 10 сообщений, из второго – 6, из третьего – 4. Вероятность того, что угнан легковой автомобиль для каждого из районов соответственно – 0,9; 0,8; 0,7. Какова вероятность того, что очередное сообщение – об угоне легкового автомобиля? Какова вероятность того, что угнанный легковой автомобиль из третьего района?

Вариант 10. Команда стрелков милицейского колледжа города Брюково состоит из 5 человек, трое из них попадают с вероятностью 0,8, а двое – с вероятностью 0,6. Наудачу из команды берется стрелок и производит выстрел. Какова вероятность того, что стрелок попадет? Если стрелок попал в цель, то какова вероятность, что это один из трех (один из двух)?

Практическое занятие 5. Схема испытаний Бернулли. Наивероятнейшее число появлений события в схеме повторных независимых испытаний.

Цель: формирование представления об основных теоремах теории вероятностей и методах их применения для решения задач.

Теоретическая часть:

Испытание, в результате которого ожидается наступление интересующего нас события, можно многократно повторять. Главный вопрос каждого повторения – произойдет или не произойдет это событие? А во всей серии повторений важно знать, сколько именно раз оно может произойти или не произойти. Например, какова вероятность, что при пяти бросаниях монеты «орел» выпадет ровно 4 раза? Швейцарский математик начала XVIII века Якоб Бернулли объединил подобные вопросы в единую вероятностную схему, которую принято называть схемой Бернулли.

Рассматривают n независимых повторений одного и того же испытания с двумя возможными исходами: «успехом» и «неудачей». Вероятность «успеха» равна p, а вероятность неудачи равна q. Известна формула p+q=1. Требуется найти вероятность $P_n(x)$ того, что в этих повторениях произойдет ровно x «успехов».

Формула Бернулли:

$$P_n(x) = C_n^x p^x q^{n-x}.$$

Пример. Из урны, содержащей 2 белых и 6 черных шаров, наудачу выбирается с возвращением 5 раз подряд один шар. Подсчитать вероятность того, что 4 раза появится белый шар.

Решение: В приведенных выше обозначениях n=8; p=1/4; q=3/4; x=5. Искомую вероятность вычисляем по формуле Бернулли:

$$P_8(5) = C_5^4 \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right) = \frac{15}{1024}$$

Формула Бернулли при заданных числах p и n позволяет рассчитывать вероятность любой частоты x ($0 \le x \le n$). Возникает естественный вопрос, какой частоте будет соответствовать наибольшая вероятность?

Предположим, что такая частота существует, и попытаемся ее определить из условия, что вероятность этой частоты не меньше вероятности "предыдущей" и "последующей" частот:

$$P_n(x) \ge P_n(x-1);\ P_n(x) \ge P_n(x+1) \qquad (*)$$

Первое неравенство (*) представляется в виде:

$$C_n^x p^x q^{n-x} \ge C_n^{x-1} p^{x-1} q^{n-x+1},$$

что эквивалентно $\frac{q}{n-x+1} \le \frac{p}{x}$ или $qx \le pn - px + p$. Отсюда следует:

$$(p+q)x \le np + p$$

$$x \le np + p$$

Решая второе неравенство (*), получим

$$x \ge np - q.$$

Таким образом, частота, имеющая наибольшую вероятность (наивероятнейшая частота), определяется двойным неравенством:

$$np - q \le x \le np + p.$$

Если np+p – целое число (тогда и np–q – целое число), то две частоты: x=np–q и x=np+p обладают наибольшей вероятностью.

Задачи и упражнения:

1. Каждый день акции корпорации АВС поднимаются в цене или падают в цене на один пункт с вероятностями соответственно 0,75 и 0,25. Найти вероятность того, что акции после шести дней вернутся к своей первоначальной цене. Принять условие, что изменения цены акции вверх и вниз – независимые события.

2. Моторы многомоторного самолёта выходят из строя во время полёта независимо один от другого с вероятностью p. Многомоторный самолёт продолжает лететь, если работает не менее половины его моторов. При каких значениях p двухмоторный самолёт надёжней четырёхмоторного самолёта?

3. Бригада из десяти человек идёт обедать. Имеются две одинаковые столовые, и каждый член бригады независимо один от другого идёт обедать в любую из этих столовых. Если в одну из столовых случайно придёт больше посетителей, чем в ней имеется мест, то возникает очередь. Какое наименьшее число мест должно быть в каждой из столовых, чтобы вероятность возникновения очереди была меньше 0,15?

4.Пусть всхожесть семян пшеницы составляет 90%. Чему равна вероятность того, что из 7 посеянных семян взойдут 5?

5. Если в семье четыре ребёнка, что вероятнее: это два мальчика и две девочки, или три ребёнка одного пола и один другого пола? Принять вероятность того, что данный ребёнок – мальчик, равной 0,5.

6. Капитан корабля перед высадкой десанта приказал выпустить по береговой полосе длиной 200 метров 20 реактивных снарядов, опасаясь замаскированных огневых точек. Вдоль берега в землю был врыт бункер длиной 20 метров.

a) Найти вероятность того, что 4 снаряда попали в бункер.

b) Найти наивероятнейшее число снарядов, попавших в бункер.

Практическое занятие 6. Приближенные формулы в схеме Бернулли.

Цель: формирование представления об основных теоремах теории вероятностей и методах их применения для решения задач.

Теоретическая часть:

Непосредственное применение формулы Бернулли при большом числе испытаний связано с громоздкими вычислениями. Поэтому при больших n вместо нее, как правило используют приближенные формулы Пуассона и Муавра-Лапласа.

Формула Пуассона

Если число испытаний n достаточно велико, а вероятность p достаточно мала, причем их произведение $a = n \cdot p$ не мало и не велико (обычно достаточно условий $p < 0{,}1;\ \ npq < 10$), то вероятность $P_n(m)$ можно приближенно найти по формуле Пуассона:

$$P_n(m) = \frac{a^m \cdot e^{-a}}{m!}$$

Локальная теорема Муавра-Лапласа

Если число испытаний n достаточно велико, а вероятности p и q не очень близки к нулю (обычно достаточно условий $n > 100$; $npq > 20$), то вероятность $P_n(m)$ можно приближенно найти по формуле Муавра-Лапласа:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x),$$

где $x = \frac{m - np}{\sqrt{npq}}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функция Гаусса.

Интегральная теорема Муавра-Лапласа

В условиях локальной формулы Муавра-Лапласа вероятность $P_n(m_1 \le m \le m_2)$ того, что число успехов m заключено между m_1 и m_2, можно приближенно найти по интегральной формуле Муавра-Лапласа

$$P_n(m_1 \le m \le m_2) = \Phi(x_2) - \Phi(x_1),$$

где $x_1 = \frac{m_1 - np}{\sqrt{npq}}, x_2 = \frac{m_2 - np}{\sqrt{npq}}$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа.

Значение функций Гаусса и Лапласа находятся по таблицам.

Пример. Вероятность «сбоя» в работе телефонной станции при каждом вызове равна $0,007$. Поступило 1000 вызовов. Определить вероятность девяти «сбоев».

Решение: По условию задачи $n = 1000,\ m = 9, p = 0,007$. Поскольку n – достаточно велико, p – достаточно мало $(npq < 7)$, то для вычисления $P_{1000}(9)$ можно использовать формулу Пуассона. Имеем $a = n \cdot p = 1000 \cdot 0,007 = 7$, откуда $P_{1000}(9) \approx \frac{7^9 \cdot e^{-7}}{9!} \approx 0,1014$.

Пример. Вероятность того, что при автоматической штамповке изделий отдельное изделие окажется бракованным, постоянна и равна $0,05$. Какова вероятность того, что в партии из 1000 изделий ровно 40 окажется бракованными?

Решение: По условию задачи $n = 1000,\ m = 40,\ \ p = 0,05,\ q = 1 - p = 0,95$. Теоретически можно использовать формулу Бернулли. Однако полученное выражение будет слишком громоздко, поэтому удобней применить локальную формулу Муавра-Лапласа. $(npq = 1000 \cdot 0,05 \cdot 0,95 = 47,5 > 20)$. Так как $\sqrt{npq} = \sqrt{47,5} \approx 6,829$, то $x = \frac{40 - 1000 \cdot 0,05}{\sqrt{47,5}} = \frac{40 - 50}{6,892} \approx -1,45$. $\varphi(x) = \varphi(-1,45) = \varphi(1,45) \approx 0,1394$

(значение функции находим по таблице). Следовательно,

$$P_{1000}(40) \approx \frac{1}{\sqrt{47,5}} \cdot \varphi(-1,45) \approx \frac{0,1394}{6,892} = 0,02$$

Пример. Вероятность попадания в цель из скорострельного орудия при отдельном выстреле равна 0,75. найти вероятность того, что при 300 выстрелов число попаданий будет не менее 210, но не более 230?

Решение: По условию задачи $n = 300, p = 0,75, q = 1 - p = 0,25, \quad m_1 = 210, m_2 = 230$. Для нахождения вероятности $P_{300}(210 \le m \le 230)$ воспользуемся интегральной формулой Муавра-Лапласа.

$n \cdot p = 300 \cdot 0,75 = 225$, $\sqrt{npq} = \sqrt{300 \cdot 0,75 \cdot 0,25} = \sqrt{56,25} \approx 7,5$, тогда

$$x_1 = \frac{210 - 225}{\sqrt{56,25}} = \frac{-15}{7,5} \approx -2; \qquad x_2 = \frac{230 - 225}{\sqrt{56,25}} = \frac{5}{7,5} \approx 0,67$$

Следовательно,

$$P_{300}(210 \le m \le 230) = \Phi(0,67) - \Phi(-2) = \Phi(0,67) + \Phi(2) = 0,2486 +$$
$$+0,4772 = 0,7258$$

Задачи и упражнения:

Внимательно прочтите условия задачи и выберите нужную формулу (Бернулли, Пуассона или Муавра-Лапласа)

1. Завод изготовитель отправил на базу 12000 доброкачественных изделий. Число изделий поврежденных при транспортировке, составляет в среднем $0,05\%$. Найти вероятность того, что на базу поступит: а) не более трех поврежденных изделий; б) хотя бы две поврежденные.
2. Вероятность появления события А в каждом из 900 независимых испытаний равна $p = 0,8$. Найти вероятность того, что событие А произойдет: а) 750 раз; б) 710 раз.
3. Всхожесть семян данного растения равна 0,9. Найти вероятность того, что из 900 посаженных семян число проросших заключено между 790 и 830.
4. Мишень состоит из трех попарно непересекающихся зон. При одном выстреле по мишени вероятность попадания в первую зону для данного стрелка 0,5. Для второй и третей зон эта вероятность соответственно равны 0,3 и 0,2. Стрелок производит 6 выстрелов по мишени. Найти вероятность того, что при этом окажется 3 попадания в первую зону, 2 попадания во вторую и 1 попадания в третью зону.
5. Вероятность выхода из строя одного элемента устройства, в течении t часов работы, равна 0,002. Какова вероятность того, что за время t из 1500 независимо работающих элементов выйдет из строя: а) 4 элемента; б) не более двух элементов?
6. Сколько раз с вероятностью 0,0484 можно ожидать появления события А в 100 независимых испытаниях, если вероятность его появления в отдельном испытании равна 0,5.
7. Вероятность появления события А в каждом из 900 независимых испытаний равна $p = 0,8$. Найти вероятность того, что событие А произойдет не менее 710 раз и не более 740.

8. Два равносильных шахматиста играют в шахматы. Что вероятнее: а) выиграть одну партию из двух или две партии из четырех; б) выиграть не менее двух партий из четырех или не менее трех партий из пяти? Ничьи во внимание не принимаются.
9. Вероятность допустить ошибку при наборе текста, состоящего из 1200 знаков, составляет 0,005. Найти вероятность того, что при наборе будет допущено: а) 6 ошибок; б) хотя бы одна ошибка.
10. В городе N из каждых 100 семей 85 имеют цветные телевизоры. Какова вероятность того, что из 400 семей 340 имеют такие же телевизоры.
11. какова вероятность того, что из 2450 ламп, освещающих улицу, к концу года будут гореть от 1500 до 1600 ламп? Считать, что каждая лампа будет гореть в течении года с вероятностью 0,64.
12. В семье 6 детей. Найти вероятность, что в данной семье не менее двух мальчиков, но не более четырех. Считать вероятность рождения мальчика и девочки 0,5.

Практическое занятие 7. Закон распределения случайной величины.

Цель: формирование представления о способах построения закона распределения случайной величины.

Теоретическая часть:

Наряду со случайными событиями, одним из основных понятий теории вероятностей является понятие случайной величины – величины, численное значение которой принимает различные значения в зависимости от результата стохастического эксперимента или опыта. Примерами случайных величин могут являться: число дорожно-транспортных происшествий за сутки; количество студентов, опоздавших на занятия; число дактилокарт, поступивших в течение дня на экспертизу. Приведенные примеры характеризуют случайные величины, имеющие количественные параметры, т.е. изолированные значения с определенными вероятностями. Такие случайные величины называют дискретными.

Но если оценить количественно как дискретную величину число ДТП, то величина интервала времени между последовательными моментами наличия ДТП может принимать любое значение в некотором временном интервале – есть непрерывное множество возможных значений рассматриваемой величины. Такие величины называют непрерывными.

Законом распределения дискретной случайной величины называют число ее возможных значений и соответствующих им вероятностей. Для примера, если взять монету и в течение некоторого времени подбрасывать ее, например 20 раз, то в этом случае могут быть ситуации, что все 20 раз монета будет обращена к вам гербом, или всего 19, 18 и т. д. раз, вплоть до того, что может выпасть 20 раз и решка. Можно вычислить вероятность того, что случайная величина выпадения герба примет значения от нуля до двадцати. Но, как нам известно, вероятность появления герба или решки одна и та же и равна 1/2. Тогда вероятности появления герба б, 7, 8 раз и соответственно решек — 14, 13, 12 раз будут равны появлению герба — 14, 13, 12 раз и соответственно б, 7, 8 раз — решки. Таким образом, для дискретной случайной величины существует определенный закон распределения, который может быть задан графически, аналитически и таблично. В последнем случае это распределение задается таблично, где в одном столбце записаны все возможные значения случайной величины, а в другом - соответствующие им вероятности.

Число выпадений герба	Вероятность	Число выпадений решки
0	0,000	20
1	0,000	19
2	0,000	18
3	0,001	17
4	0,005	16
5	0,015	15
6	0,037	14
…	…	…

Задачи и упражнения:

Составить закон распределения вероятностей случайной величины X, если:

1) вероятность работы каждого из четырех автомобилей автопарка ОВД города Урюпина без поломок в течение определенного времени равна 0,9; случайная величина X – число автомобилей, работавших безотказно;

2) вероятность рождения мальчика равна 0,5; случайная величина X – число мальчиков в семьях, имеющих четырех детей;

3) вероятность того, что покупатель совершит покупку в магазине 0,4; случайная величина X – число покупателей, совершивших покупку, если магазин посетило 3 покупателя;

4) вероятность выпадения герба при бросании равна 0,5; случайная величина X – число выпадений герба, если монета брошена 5 раз;

5) в лотерее на каждые 100 билетов приходится 15 выигрышей (количество и размер выигрышей приведены в таблице); случайная величина X – размер выигрыша в лотерее, приходящегося на один билет;

Размер выигрыша, $	20	5	1
Количество выигрышей	1	4	10

6) среди 10 лотерейных билетов имеется 4 билет с выигрышем, наудачу покупают два билета; случайная величина X – число выигрышных билетов среди купленных;

7) баскетболист делает три штрафных броска, вероятность попадания мяча в корзину при каждом броске равна 0,7; случайная величина X – число попаданий мяча в корзину;
8) в партии из 25 кожаных курток 5 имеют скрытый дефект, покупают три куртки; случайная величина X – число дефектных курток среди купленных;
9) вероятность того, что при составлении бухгалтерского баланса допущена ошибка, равна 0,3; аудитору на заключение представлено три баланса предприятия; случайная величина X – число положительных заключений на проверяемые балансы;
10) из 20 заключенных, находящихся в следственном изоляторе, четверо – моложе 25 лет; произвольно выбирают 2 заключенных; случайная величина X – число заключенных моложе 25 лет среди выбранных.

Практическое занятие 8. Вычисление числовых характеристик дискретной случайной величины.

Цель: формирование представлений о значении и способах вычисления числовых характеристик дискретной случайной величины.

Теоретическая часть:

Характеристикой среднего значения случайной величины служит математическое ожидание. Математическим ожиданием дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности:

$$M(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n.$$

Дисперсией случайной величины X называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

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

Средним квадратическим отклонением случайной величины называют квадратный корень из дисперсии.

Задачи и упражнения: Для случайной величины из задач практического занятия 7 найти ее математическое ожидание, дисперсию и среднее квадратическое отклонение.

Практическое занятие 9. Функции распределения и плотности распределения вероятностей непрерывных случайных величин, их свойства.

Цель: формирование основных представлений о назначении и способах вычисления функций распределения и плотности непрерывной случайной величины.

Теоретическая часть:

Для непрерывной случайной величины теряет смысл понятие вероятности каждого конкретного значения, поскольку таких значений бесконечно много, и из условия, что сумма вероятностей всех значений равна 1, следует, что вероятность каждого фиксированного значения равна нулю. Поэтому основными характеристиками, описывающими поведение непрерывной случайной величины, являются *функция распределения* (интегральная функция распределения) и *плотность распределения* вероятностей (плотность вероятности, дифференциальная функция распределения).

Функция распределения $F(x)$ представляет собой вероятность того, что случайная величина примет значение, меньшее аргумента этой функции:

$$F(x) = P\{X < x\}.$$

Плотность вероятности (плотность распределения) $f(x)$ является первой производной от функции распределения: $f(x) = F'(x)$. График функции $f(x)$ называют *кривой распределения*.

Понятие вероятности сохраняется для непрерывной случайной величины как вероятность ее попадания в некоторый интервал, которую можно определить так:

$$P\{a < X < b\} = F(b) - F(a) \quad \text{или} \quad P\{a < X < b\} = \int_a^b f(x)dx.$$

Пример. Плотность вероятности случайной величины X имеет вид:

$$f(x) = \begin{cases} 0, & x \le 0, \\ C(2x - x^2), & 0 < x < 2, \\ 0, & x \ge 2. \end{cases}$$

Найти:

1) C; 2) $F(x)$; 3) $P\{0{,}5 < X < 1{,}5\}$.

Решение. 1) Из условия нормированности плотности вероятности следует, что $\int_{-\infty}^{+\infty} f(x)dx = 1.$ В нашем случае

$$\int_0^2 C(2x - x^2)dx = C(x^2 - x^3/3)\Big|_0^2 = C(4 - 8/3) = 4C/3 = 1,$$

откуда $C = 3/4$.

2) Связь между $F(x)$ и $f(x)$ задается формулой $F(x) = \int_{-\infty}^{x} f(t)dt.$

Поэтому при $x \le 0$ $$F(x) = \int_{-\infty}^{x} 0dx = 0,$$

при $0 < x < 2$ $$F(x) = \frac{3}{4}\int_0^x (2t - t^2)dt = \frac{3}{4}x^2 - \frac{1}{4}x^3,$$

а для $x \ge 2$ $$F(x) = \frac{3}{4}\int_0^2 (2t - t^2)dt + \int_2^x 0 \cdot dx = 1.$$

Следовательно, $$F(x) = \begin{cases} 0, & x \le 0, \\ \frac{3}{4}x^2 - \frac{1}{4}x^3, & 0 < x < 2, \\ 1, & x \ge 2. \end{cases}$$

3) $$P\{0{,}5 < X < 1{,}5\} = \frac{3}{4}\int_{0{,}5}^{1{,}5} (2x - x^2)dx = \frac{3}{4}\left(x^2 - x^3/3\right)\Big|_{0{,}5}^{1{,}5} = 0{,}6875.$$

Задачи и упражнения

1.Плотность вероятности непрерывной случайной величины X имеет вид:

$$f(x) = \begin{cases} ax^2, & -3 < x < 0, \\ 0, & \text{в остальных случаях.} \end{cases}$$

а) Найти значение параметра a. б) Построить график функции распределения $F(x)$. в) Найти вероятность того, что случайная величина X примет значения из интервала (–1; 1).

2.Плотность вероятности непрерывной случайной величины X имеет вид:

$$f(x) = \begin{cases} 0, & x \le 2, \\ ax^{-9/2}, & x > 2. \end{cases}$$

а) Найти значение параметра a.

б) Построить график функции распределения.

в) Найти вероятность того, что случайная величина X примет значения из интервала (3; 4).

3. Независимые случайные величины $X_1, X_2, \ldots, X_8$ имеют нормальный закон распределения с параметрами $m = 1$, $\sigma = \sqrt{2}$. Рассматривается случайная величина $Y = X_1 + X_2 + \ldots + X_8$. С помощью неравенства Чебышева оценить вероятности $P\{3 < Y < 13\}$, $P\{|Y - 8| > 8\}$.

4. Плотность вероятности непрерывной случайной величины X имеет вид:

$$f(x)=\begin{cases} ax(4-x), \ 1<x<4, \\ 0, \ \text{в остальных случаях.} \end{cases}$$

а) Найти значение параметра a.

б) Построить график функции распределения.

в) Найти вероятность того, что случайная величина X примет значения из интервала (3; 5).

5.Случайная величина X имеет нормальный закон распределения. Известно, что $M(X)=-2$, $D(X)=1$. Найти: а) плотность вероятности случайной величины X и ее значения в точках $x=-1$, $x=0$, $x=2$; б) вероятности $P\{-2<X<0\}$, $P\{X>1\}$.

Практическое занятие 10. Числовые характеристики непрерывных случайных величин.

Цель: формирование основных представлений о назначении и способах вычисления числовых характеристик непрерывной случайной величины.

Теоретическая часть:

Для непрерывной случайной величины математическое ожидание, дисперсия и среднее квадратическое отклонение определяются следующим образом:

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx, \quad D(X)=\int_{-\infty}^{+\infty} x^2 f(x)dx-(M(X))^2, \quad \sigma(X)=\sqrt{D(X)}.$$

Пример. Плотность вероятности случайной величины X имеет вид:

$$f(x)=\begin{cases} 0, & x\le 0, \\ C(2x-x^2), & 0<x<2, \\ 0, & x\ge 2. \end{cases}$$

Вычислить: $M(X)$, $D(X)$, $\sigma(X)$.

Решение:

$$M(X)=\frac{3}{4}\int_0^2 x(2x-x^2)dx=\frac{3}{4}\left(\frac{2}{3}x^3-\frac{1}{4}x^4\right)\Bigg|_0^2=\frac{3}{4}(16/3-4)=1;$$

$$D(X)=\frac{3}{4}\int_0^2 x^2(2x-x^2)dx-1=\frac{3}{4}\left(x^4/2-x^5/5\right)\Bigg|_0^2-1=\frac{3}{4}(8-32/5)-1=0{,}2;$$

$$\sigma(X)=\sqrt{0{,}2}\approx 0{,}447.$$

Задачи и упражнения:

1.Плотность вероятности непрерывной случайной величины X имеет вид:

$$f(x)=\begin{cases} ax^2, \ -3<x<0, \\ 0, \ \text{в остальных случаях.} \end{cases}$$

Найти $M(X)$, $D(X)$ и $\sigma(X)$.

2. Найти математическое ожидание и дисперсию нормально распределенной случайной величины X, если известно, что

$P\{X<1\}=0{,}1 \qquad P\{X>5\}=0{,}2$.

Построить кривую распределения и найти ее максимум.

3.Плотность вероятности непрерывной случайной величины X имеет вид:

$$f(x)=\begin{cases} 0, \ x\le 2, \\ ax^{-9/2}, \ x>2. \end{cases}$$

Найти $M(X)$, $D(X)$ и $\sigma(X)$.

Практическое занятие 11. Вариационные ряды и их графическое изображение.

Цель: формирование основных понятий математической статистики.

Теоретическая часть:

Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений. Двумя основными задачами математической статистики являются:

- определение способов сбора и группировки этих статистических данных;
- разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости от других случайных величин и т.д.;

б) проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения.

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Определим основные понятия математической статистики.

Генеральная совокупность – все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности N и *объем выборки* n – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была *репрезентативной* (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Пусть интересующая нас случайная величина X принимает в выборке значение x_1 n_1 раз, x_2 – n_2 раз, …, $x_к$ – $n_к$ раз, причем $\sum_{i=1}^{k} n_k = n$, где n – объем выборки. Тогда наблюдаемые значения случайной величины $x_1, x_2, \ldots, x_к$ называют вариантами, а $n_1, n_2, \ldots, n_к$ – частотами. Если разделить каждую частоту на объем выборки, то получим *относительные частоты* $\omega_i = \frac{n_i}{n}$.

Последовательность вариант, записанных в порядке возрастания, называют вариационным рядом, а перечень вариант и соответствующих им частот или относительных частот – статистическим рядом.

Пример 1:

При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1, 1, 4, 0, 1, 2, 1, 2, 2, 0 , 5, 3, 3, 1, 0, 2, 2, 3, 4, 1.

Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид:

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
w_i	0,15	0,3	0,25	0,15	0,1	0,05

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать группированную выборку. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал. Составленная по этим результатам таблица называется *группированным статистическим рядом*.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – полигон частот: ломаная, отрезки которой соединяют точки с координатами (x_1, n_1), $(x_2,$

n_2),…, (x_k, n_k), где x_i откладываются на оси абсцисс, а n_i – на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные (w_i) частоты, то получим полигон относительных частот.

Этапы первичной обработки выборки:

1.Ранжирование опытных данных (расположение значений признака по убыванию или возрастанию).

2.Частотный анализ (построение статистического ряда, определение относительных частот).

3.Группировка (частотная табуляция выборки).

Задачи и упражнения:

1. Имеются следующие данные об успеваемости 20 студентов группы по статистике в летнюю сессию: 4, 4, 3, 5, 2, 2, 5, 3, 3, 3, 5, 3, 2, 4, 3, 2, 3, 5, 5, 4. Постройте: ряд распределения студентов по баллам оценок; ряд распределения студентов пол уровню успеваемости, выделив в нем две группы студентов: неуспевающие и успевающие. Изобразите каждый из рядов графически.

2.Построить кривую и гистограмму суммы налоговых неуплат, зафиксированных по условным регионам страны, данные по которым приведены в таблице (в млн. руб.):

21	43	72	84
22	54	75	32
26	49	77	45
27	53	78	65
28	54	81	12
32	58	83	34
34	61	84	54
37	65	84	34
39	68	88	41

3.Дана исходная выборка по росту и весу студентов группы:

Рост	160	168	175	169	170	169	162	166	163	160	158	173	162	173	156
Вес	48	58	69	64	69	70	51	60	67	54	48	58	44	50	56

Определите объем выборки. Составьте ранжированный и вариационный ряды по каждому из признаков.

4.Построить гистограмму нагрузки на одного следователя по расследованным уголовным делам по данным таблицы:

Годы	1995	1996	1997	1998	1999	2000	2001	2001	2003
По линии МВД	63,2	49,6	41	31,5	32,3	32,4	37,9	34,2	36,2
По линии прокуратуры	17,2	16,9	15,5	15,1	16	17,7	18,2	18,3	19,7

5.Три варианта исходных данных таблицы ниже – результаты телефонных переговоров в минутах сотрудников трех отдельных служб в течение рабочего дня. Требуется составить интервальный ряд распределения.

№ выборки	1	2	3	4	5	6	7	8	9
Служба 1	1,9	2,6	3,1	3,3	2,1	2,0	4,7	0,9	2,8
Служба 2	6,9	1,1	3,6	1,0	7,0	2,1	8,7	1,0	8,0
Служба 3	4,7	6,8	3,8	3,1	1,8	7,1	9,2	8,1	4,0

6. Дана выборка 7,3,3,6,4,3,5,1,2,1,3. Построить вариационный ряд. Определить размах выборки.

Практическое занятие 12. Выборочный метод. Точечные оценки.

Цель: формирование основных понятий математической статистики при обработке результатов эксперимента.

Теоретическая часть:

Числовые показатели, характеризующие генеральную совокупность, называют *параметрами,* а числовые показатели, характеризующие выборку,— *выборочными характеристиками* или *статистиками.* Выборочные характеристики являются приближенными оценками генеральных параметров. Это величины случайные, варьирующие вокруг своих параметров. Оценки генеральных параметров по выборочным характеристикам могут быть точечными и интервальными.

Генеральные характеристики, или параметры, принято обозначать буквами греческого алфавита, а выборочные характеристики - латинского. Выборочная средняя $\bar{x}$ является оценкой генеральной средней *a*, выборочная дисперсия s_x^2— оценкой генеральной дисперсии σ_x^2, а среднее квадратическое отклонение s_x — оценкой стандартного отклонения σ_x, характеризующего генеральную совокупность. Это *точечные оценки,* представляющие собой не интервалы, а числа («точки»), вычисляемые по случайной выборке.

Требования, предъявляемые к точечным оценкам.

Выборочные характеристики как величины случайные, варьирующие вокруг своих генеральных параметров, в основном не совпадают с ними по абсолютной величине. Для того чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям. Оценки должны удовлетворять по меньшей мере следующим требованиям: быть состоятельными, эффективными и несмещенными.

Точечная оценка статистики называется *состоятельной,* если при увеличении объема выборки она стремится к величине генерального параметра. Так, для генеральной средней состоятельной оценкой является выборочная средняя, для генеральной дисперсии состоятельной оценкой будет выборочная дисперсия.

Точечная оценка называется *эффективной,* если она имеет наименьшую дисперсию выборочного распределения по сравнению с другими аналогичными оценками, т. е. обнаруживает наименьшую случайную вариацию.

Оценка называется *несмещенной,* если математическое ожидание ее выборочного распределения совпадает со значением генерального параметра.

Выборочная средняя

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объёма n.

Выборочной средней $\bar{x}_в$ называют среднее арифметическое значение признака выборочной совокупности.

Если все значения $x_1, x_2, ..., x_n$ признака выборки объёма n различны, то $\bar{x}_в = \dfrac{x_1 + x_2 + ... + x_n}{n}$.

Если же значения признака $x_1, x_2, ..., x_k$ имеют соответственно частоты $f_1, f_2, ... f_k$, то $\bar{x}_в = \dfrac{\sum_{i=1}^{k} x_i f_i}{n}$.

Замечание. Выборочная средняя, найденная по данным одной выборки, есть, очевидно, определённое число. Если же извлекать другие выборки того же объёма из той же генеральной совокупности, то выборочная средняя будет изменяться от выборки к выборке. Таким образом, выборочную среднюю можно рассматривать как случайную величину, а, следовательно, можно говорить о распределениях выборочной средней и о числовых характеристиках этого распределения (его называют выборочным), в частности, о математическом ожидании и дисперсии выборочного распределения.

Выборочная дисперсия

Для того, чтобы охарактеризовать рассеяние наблюдаемых значений количественного признака выборки вокруг своего среднего значения $\bar{x}_в$, вводят сводную характеристику – выборочную дисперсию.

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{x}_в$.

Если все значения $x_1, x_2, ..., x_n$ признака выборки объёма n различны, то

$$D_в = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_в)^2}{n}.$$

Если значения признака $x_1, x_2, ..., x_k$ имеют соответственно частоты $f_1, f_2, ..., f_k$, причём $f_1 + f_2 + ... + f_k = n$, то

$$D_в = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n}.$$

Легко «исправить» выборочную дисперсию так, чтобы её математическое ожидание было равно генеральной дисперсии. Достаточно для этого умножить $D_в$ на дробь $\frac{n}{n-1}$. Сделав это, получим исправленную дисперсию, которую обычно обозначают через s^2:

$$s^2 = \frac{n}{n-1} D_в = \frac{n}{n-1} \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n} = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n-1}.$$

Исправленная дисперсия является, несмещённой оценкой генеральной дисперсии.

Итак, в качестве оценки генеральной дисперсии принимают исправленную дисперсию $s^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_k)^2 f_i}{n-1}$.

Кроме дисперсии для характеристики рассеяния значений признака выборочной совокупности вокруг своего среднего значения пользуются сводной характеристикой – средним квадратическим отклонением.

Выборочным средним квадратическим отклонением (стандартным) называют квадратный корень из выборочной дисперсии: $\sigma_в = \sqrt{D_в}$.

Для оценки же среднего квадратического отклонения генеральной совокупности используют «исправленное» среднее квадратическое отклонение, которое равно квадратному корню из исправленной дисперсии: $s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n-1}}$.

Замечание. Сравнивая формулы $D_в = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n}$ и $s^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_k)^2 f_i}{n-1}$, видим, что они отличаются лишь знаменателями. Очевидно, при достаточно больших значениях n объёма выборки выборочная и исправленная дисперсии различаются мало. На практике пользуются исправленной дисперсией, если примерно n<30.

Пример. Среди свиноматок хозяйства было выбрано 64 свиноматки, данные по опоросам среди которых следующие:

8; 10; 6; 10; 8; 5; 11; 7; 10; 6; 9; 7; 8; 7; 9;11; 8; 9; 10; 8; 7; 8; 11; 8; 7; 10; 8; 8; 5; 11; 8; 10; 12; 7; 5; 7; 9; 7; 10; 5; 8; 9; 7; 12; 8; 9; 6; 7; 8; 7; 11; 8; 6; 7; 9; 10.

Определим среднее выборочное число поросят в пометах 64 свиноматок и вычислим показатели вариации для этой выборки.

Решение:

Выборочная средняя числа поросят в опоросах:

$$\overline{x_в} = \frac{\sum x_i f_i}{\sum f_i} = 528/64 = 8{,}25$$
;

Выборочная дисперсия числа поросят:

$$D_в = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 f_i}{n} = 218/64 = 3{,}41.$$

Выборочное среднее квадратическое отклонение числа поросят в опоросах:

$\sigma_в = \sqrt{D_в} = \sqrt{3{,}41} = 1{,}85$ поросят.

Исправленная выборочная дисперсия числа поросят:

$$s^2 = \frac{n}{n-1} D_в = 64/63 * 218 = 3{,}46.$$

Исправленное среднее квадратическое отклонение числа поросят в опоросах:

$s = \sqrt{s^2} = \sqrt{3{,}46} = 1{,}86$ поросят.

Итак, нами найдены следующие точечные оценки по данной выборке:

точечная оценка среднего числа поросят в опоросах для генеральной совокупности (то есть по всему хозяйству): 8,25 поросят;

точечная оценка среднего квадратического отклонения числа поросят в генеральной совокупности: 1,86 поросят.

Задачи и упражнения:

По данным выборки, удовлетворяющей нормальному закону распределения, вычислить:

1) выборочное среднее;
2) исправленное выборочное среднее квадратическое отклонение.

А) 8.0 -1.1 13.5 10.0 2.4 4.1 20.0 12.4 13.4 4.8 7.8 0.0 10.9 13.7 6.6

Б) 31.6 34.9 46.9 42.8 36.0 26.2 28.6 48.5 27.7 45.8 32.0 41.2 39.8 33.1 36.3 53.5 43.9 35.8 32.9 34.4

В) 25.4 31.1 13.2 23.0 19.1 26.5 23.2 29.2 24.8 26.6 29.3 21.4 28.2 38.2 19.9 30.6 24.5 23.2

Г) 10.5 5.5 12.6 27.0 25.0 31.2 15.9 15.3 17.4 32.8 30.3 9.5 17.7 16.4 15.9.

Практическое занятие 13. Понятие интервального оценивания. Доверительная вероятность и предельная ошибка выборки. Оценка характеристик генеральной совокупности по малой выборке.

Цель: формирование основных понятий интервального оценивания и методов оценки характеристик генеральной совокупности по малой выборке.

Теоретическая часть:

При выборке малого объёма точечная оценка может значительно отличаться от оцениваемого параметра, т. е. приводить к грубым ошибкам. По этой причине при небольшом объёме выборки следует пользоваться интервальными оценками. Интервальной называют оценку, которая определяется двумя числами – концами интервала.

Пусть найденная по данным выборки статистическая характеристика Θ^* служит оценкой неизвестного параметра Θ. Будем считать Θ постоянным числом (Θ может быть и случайной величиной). Ясно, что Θ^* тем точнее определяет параметр Θ, чем меньше абсолютная величина разности $\left|\Theta - \Theta^*\right|$. Другими словами, если $\delta > 0$ и $\left|\Theta - \Theta^*\right| < \delta$, то чем меньше δ, тем оценка точнее. Таким образом, положительное число δ характеризует точность оценки.

Однако статистические методы не позволяют категорически утверждать, что оценка Θ^* удовлетворяет неравенству $\left|\Theta - \Theta^*\right| < \delta$; можно лишь говорить о вероятности γ, с которой это неравенство осуществляется.

Надёжностью (доверительной вероятностью) оценки Θ по Θ^* называют вероятность γ, с которой осуществляется неравенство $\left|\Theta - \Theta^*\right| < \delta$. Обычно надёжность оценки задаётся наперёд, причём в качестве γ берут число, близкое к единице. Наиболее часто задают надёжность, равную 0,95; 0,99 и 0,999.

Пусть вероятность того, что $\left|\Theta - \Theta^*\right| < \delta$, равна γ: $P\left[\left|\Theta - \Theta^*\right| < \delta\right] = \gamma$.

Заменив неравенство $\left|\Theta - \Theta^*\right| < \delta$ равносильным ему двойным неравенством $-\delta < \Theta - \Theta^* < \delta$, или $\Theta^* - \delta < \Theta < \Theta^* + \delta$, имеем $P\left[\Theta^* - \delta < \Theta < \Theta^* + \delta\right] = \gamma$.

Это соотношение следует понимать так: вероятность того, что интервал $(\Theta^* - \delta, \Theta^* + \delta)$ заключает в себе (покрывает) неизвестный параметр Θ, равна γ. Доверительным называют интервал $(\Theta^* - \delta, \Theta^* + \delta)$, который покрывает неизвестный параметр с заданной надёжностью γ.

Доверительные интервалы для оценки математического ожидания нормального распределения при известном σ.

Пусть количественный признак X генеральной совокупности распределён нормально, причём среднее квадратическое отклонение σ этого распределения известно. Требуется оценить неизвестное математическое ожидание a по выборочной средней $\bar{x}$. Поставим своей задачей найти доверительные интервалы, покрывающие параметр a с надёжностью γ.

Если случайная величина X распределена нормально, то выборочная средняя $\bar{X}$, найденная по независимым наблюдениям, также распределена нормально. Параметры распределения $\bar{X}$ таковы:

$$M(\bar{X}) = a, \;\; \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Приняв во внимание, что по условию нам задана вероятность γ, получаем следующую формулу (чтобы получить рабочую формулу, выборочную среднюю вновь обозначим через $\bar{x}$)

$$P\left(\bar{x} - \frac{t\sigma}{\sqrt{n}} < a < \bar{x} + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi_0(t) = \gamma.$$

Смысл полученного соотношения таков: с надёжностью γ можно утверждать, что доверительный интервал $\left(\bar{x} - \frac{t\sigma}{\sqrt{n}}, \bar{x} + \frac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр a; точность оценки $\delta = \frac{t\sigma}{\sqrt{n}}$.

Укажем ещё, что число t определяется из равенства $2\Phi_0(t) = \gamma$, или $\Phi_0(t) = \frac{\gamma}{2}$; по таблице функции Лапласа находят аргумент t, которому соответствует значение функции Лапласа, равное $\frac{\gamma}{2}$.

Надёжность γ=0,95 указывает, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключён; лишь в 5 % случаев он может выйти за границы доверительного интервала.

Доверительные интервалы для оценки математического ожидания нормального распределения при неизвестном σ

Пусть количественный признак X генеральной совокупности распределён нормально, причём среднее квадратическое отклонение σ неизвестно. Требуется оценить неизвестное математическое ожидание a с помощью доверительных интервалов. Разумеется, невозможно воспользоваться результатами предыдущего параграфа, в котором σ предполагалось известным.

Оказывается, что по данным выборки можно построить случайную величину $T = \frac{\bar{X} - a}{S/\sqrt{n}}$, которая имеет распределение Стьюдента с k = n-1 степенями свободы;

здесь $\overline{X}$ - выборочная средняя, S – «исправленное» среднее квадратическое отклонение, n – объём выборки.

Пользуясь распределением Стьюдента, находим:

$$P\left(\overline{X}-\frac{t_\gamma S}{\sqrt{n}}<a<\overline{X}+\frac{t_\gamma S}{\sqrt{n}}\right)=\gamma .$$

Значит, доверительный интервал $\left(\overline{x}-\frac{t_\gamma s}{\sqrt{n}},\overline{x}+\frac{t_\gamma s}{\sqrt{n}}\right)$, покрывает неизвестный параметр a с надёжностью γ.

Пример. Случайная величина X – вес полугодовалого поросенка в хозяйстве (то есть в генеральной совокупности) - распределена нормально. По выборке объёма n = 16 найдены выборочная средняя $\overline{x}$=20,2 кг и «исправленное» среднее квадратическое отклонение s=0,8 кг. Оценить неизвестное математическое ожидание при помощи доверительного интервала с надёжностью 0,95.

Решение: Найдём t_γ. Пользуясь таблицей, по γ=0,95 и n = 16 находим t_γ =2,13.

Найдём доверительные границы:

$$\overline{x}-\frac{t_\gamma s}{\sqrt{n}}=20{,}2-\frac{2{,}13\cdot 0{,}8}{\sqrt{16}}=19{,}774\text{кг},$$

$$\overline{x}+\frac{t_\gamma s}{\sqrt{n}}=20{,}2+\frac{2{,}13\cdot 0{,}8}{\sqrt{16}}=20.626\text{кг}.$$

Итак, с надёжностью 0,95 неизвестный параметр a заключён в доверительном интервале 19,774< a <20,626 (кг).

Доверительные интервалы для оценки среднего квадратического отклонения σ нормального распределения

Пусть количественный признак X генеральной совокупности распределён нормально. Требуется оценить неизвестное генеральное среднее квадратическое отклонение σ по «исправленному» выборочному среднему квадратическому отклонению s.

Доверительный интервал, покрывающий параметр σ с заданной надёжностью γ находят по следующей формуле:

$$s(1-q)<\sigma<s(1+q).$$

Здесь параметр q определяют, пользуются таблицей приложения 2, а s находят по выборке.

Пример. Случайная величина X – вес полугодовалого поросенка в хозяйстве – (то есть в генеральной совокупности) распределён нормально. По выборке объёма n=25 найдено «исправленное» среднее квадратическое отклонение s=0,8 кг. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надёжностью 0,95.

Решение: По таблице по данным γ=0,95 и n=25 найдём q=0,32.

Искомый доверительный интервал таков:

0,8 (1 - 0,32)< σ<0,8 1(1+0,32), или

0,544<σ<1,056 (кг).

Замечание. Если q>1, то неравенство примет вид

0<σ<s(1+q).

Задачи и упражнения:

Задача 1. По данным выборки, удовлетворяющей нормальному закону распределения, вычислить:

1) доверительный интервал для математического ожидания при доверительной вероятности γ;

2) доверительный интервал для среднего квадратического отклонения для того же значения γ.

А) γ = 0.999

8.0 -1.1 13.5 10.0 2.4 4.1 20.0 12.4 13.4 4.8 7.8 0.0 10.9 13.7 6.6

Б) γ = 0.95

31.6 34.9 46.9 42.8 36.0 26.2 28.6 48.5 27.7 45.8 32.0 41.2 39.8 33.1 36.3 53.5 43.9 35.8 32.9 34.4

В) γ = 0.999

25.4 31.1 13.2 23.0 19.1 26.5 23.2 29.2 24.8 26.6 29.3 21.4 28.2 38.2 19.9 30.6 24.5 23.2.

Задача 2. По данным выборки, удовлетворяющей нормальному закону распределения со средним квадратическим отклонением s , вычислить доверительный интервал для математического ожидания при доверительной вероятности γ.

$s = 7$, γ = 0.99

13.4 8.6 22.1 2.3 14.6 13.0 11.1 29.4 23.3 1.7 13.6 2.1 21.6 6.1 8.6 6.6
16.0 11.6 16.6 1.6 15.8 18.9 10.6 11.9 0.1 10.7 3.8 -3.6 15.4 7.9
4.5 17.7 10.8 19.6 18.5 15.5 9.3 21.7 6.6 10.5 10.4 8.2 16.0 22.6 20.5 11.6
23.2 23.0 9.5 11.3 14.9 19.9 13.4 13.9 19.5 19.8 21.0 3.2 14.0 19.1
17.9 8.6 11.2 16.2 13.9 16.2 17.1 7.7 12.5 2.7 16.5 20.2 15.5 14.5 5.6 16.5
12.3 9.9 11.9 17.6 6.6 20.3 9.7 13.2 17.4 5.1 13.0 23.3 6.8 9.8
15.5 16.2 18.4 9.2 5.7 10.9 8.8 7.4 16.2 9.9

Практическое занятие 14. Статистическая проверка гипотез.

Цель: сформировать представление о методах и принципах проверки статистических гипотез.

Теоретическая часть:

Статистической называют гипотезу о виде неизвестного распределения, или о параметрах известных распределений. Наряду с выдвинутой гипотезой рассматривают и противоречащую ей гипотезу. Если выдвинутая гипотеза будет отвергнута, то имеет место противоречащая гипотеза. По этой причине эти гипотезы целесообразно различать.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки. Поскольку проверку производят статистическими методами, её называют статистической. В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, т. е. могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза. Ошибка второго рода состоит в том, что будет принята неправильная гипотеза.

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближённое распределение которой известно. Обозначим эту величину в целях общности через K.

Статистическим критерием (или просто критерием) называют случайную величину K, которая служит для проверки нулевой гипотезы.

После выбора определённого критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая – при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез можно сформулировать так: если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

В области статистики и биометрии в частности применяют два вида статистических критериев: *параметрические,* построенные на основании параметров данной совокупности (например, $\bar{x}$ и s^2_x) и представляющие функции этих параметров, и *непараметрические,* представляющие собой функции, зависящие непосредственно от вариант данной совокупности с их частотами. Первые служат для проверки гипотез о параметрах совокупностей, распределяемых по нормальному закону, вторые — для проверки рабочих гипотез независимо от формы распределения совокупностей, из которых взяты сравниваемые выборки. Применение параметрических критериев связано с необходимостью вычисления выборочных характеристик — средней величины и показателей вариации, тогда как при использовании непараметрических критериев такая необходимость отпадает.

t-критерий Стьюдента (t-распределение). Английский математик В. Госсет (печатавшийся под псевдонимом Стьюдент), в 1908 г. нашел закон распределения величины $t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$, в которой генеральный параметр σ заменен на его выборочную характеристику s_x, т. е. нашел *закон распределения значений*

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}.$$

Открытый Стьюдентом и теоретически обоснованный Р. Фишером *закон t-распределения* служит основой так называемой теории малой выборки, которая характеризует распределение выборочных средних в нормально распределяющейся совокупности в зависимости от объема выборки. t-распределение зависит только от числа степеней свободы $k = n—1$, причем с увеличением объема выборки n t-распределение быстро приближается к нормальному с параметрами $\mu = 0$ и $\sigma = 1$ и уже при $n>30$ не отличается от него. Это видно из таблицы ниже, в которой приведены табулированные значения t-распределения и нормального распределения для разных значений t.

Распределение	**Нормированное отклонение *t***						
	0,5	**1.0**	**1.5**	**2,0**	**2,5**	**3,0**	**3,5**
Нормальное	0,383	0,683	0,866	0,955	0,988	0,997	0,9995
Стьюдента при n = 3	0,333	0,577	0,728	0,816	0,870	0,905	0,927
n = 20	0,377	0,670	0,850	0,940	0,978	0,993	0,998
n = 30	0,383	0,683	0,866	0,955	0,988	0,997	0,9995

Оценка разности средних. Сравнивая друг с другом две независимые выборки, взятые из нормально распределяющихся совокупностей с параметрами μ_1 и μ_2. Разность μ_1- μ_2 этих параметров обозначим через D, то есть μ_1- μ_2=D, а дисперсию этой разности σ^2_D. Значения генеральных параметров неизвестны, однако по выборкам мы можем найти величины выборочных средних и разность между ними $\overline{x_1} - \overline{x_2}$, которую обозначим d, то есть $\overline{x_1} - \overline{x_2} = d$.

Нулевая гипотеза сводится к предположению, что $\mu_1=\mu_2$, то есть $D=0$. Критерием для проверки H_0-гипотезы служит отношение

$$t = \frac{d - (\mu_1 - \mu_2)}{s_d}$$

где t — переменная величина, следующая t-распределению Стьюдента с числом степеней свободы $k = n_1 + n_2 - 2$, а s_D - ошибка указанной разности, а n_1 и n_2 - объемы первой и второй выборок соответственно.

Так как, согласно H_0-гипотезе, $\mu_1=\mu_2$, то t-критерий выражается в виде отношения разности выборочных средних к своей ошибке, т. е.

$$t = \frac{d}{s_d}.$$

H_0-гипотезу отвергают, если фактически установленная величина t-критерия (обозначаемая $t_ф$) превзойдет или окажется равной критическому значению t_{kp} этой величины для принятого уровня значимости α и числа степеней свободы $k = n_1 + n_2 - 2$, т. е. при условии $t_ф \geq t_{kp}$.

Ошибку разности средних s_D определяют по формуле:

$$s_D = \sqrt{\frac{\sum(x_i - \overline{x_1})^2 + \sum(x_j - \overline{x_2})^2}{n_1 + n_2 - 2} \cdot \left(\frac{n_1 + n_2}{n_1 n_2}\right)}$$

Пример. Изучали влияние кобальта на массу тела кроликов. Опыт проводили на двух группах животных: опытной и контрольной. Были исследованы кролики в возрасте от полутора до двух месяцев, массой тела 500—600 г. Опыт продолжался полтора месяца. Животных обеих групп содержали на одном и том же кормовом рационе. Однако опытные кролики в отличие от контрольных ежедневно получали добавку к рациону в виде водного раствора по 0,06 г хлористого кобальта на 1 кг живой массы тела. За время опыта животные дали следующие прибавки живой массы тела:

Привесы, г		Отклонения от средней		Квадраты отклонений	
опыт	контроль	опыт $(x_i - \overline{x_1})$	контроль $(x_j - \overline{x_2})$	опыт $(x_i - \overline{x_1})^2$	контроль $(x_j - \overline{x_2})^2$
580	504	58	22	3364	484
692	560	54	34	2916	1 156
700	420	62	106	3844	11236
621	600	17	74	289	5 476
640	580	2	54	4	2916
561	530	77	4	5929	16
680	490	42	36	1764	1 296

630	580 470	8	54 56	64 ч	2916 3136
$\Sigma = 5104$	$\Sigma = 4734$	—	—	$\Sigma = 18\ 174$	$\Sigma = 28\ 632$
$\overline{x_1} = 638$	$\overline{x_2} = 526$	—	—	$\Sigma = 46\ 806$	

Средние арифметические привесов:

в опыте $\overline{x_1}$ = 5104/8= 638 г,

в контроле $\overline{x_2}$=4734/9 = 526 г. Разница $\overline{x_1} - \overline{x_2} = d$ = 112 г. Чтобы установить, достоверна или случайна эта разница, нужно определить ошибку разности средних:

$$s_D = \sqrt{\frac{46806}{8+9-2} \cdot \left(\frac{8+9}{8 \cdot 9}\right)} = \sqrt{736{,}8} = 27{,}14.$$

Отсюда $t_ф$ = 112/27,14= 4,1.

По таблице для уровня значимости α = 0,01 и числа степеней свободы k = 9+8—2= 15 находим t_{kp} = 2,95. Так как $t_ф > t_{kp}$, нулевая гипотеза опровергается на высоком уровне значимости (P<0,01). Разница между средними величинами опыта и контроля оказалась в высшей степени достоверной.

Правильное применение параметрических критериев для проверки статистических гипотез основано на предположении о нормальном распределении совокупностей, из которых взяты сравниваемые выборки. Однако это не всегда имеет место, так как не все биологические признаки распределяются нормально. Немаловажным является и то обстоятельство, что исследователю приходится иметь дело не только с количественными, но и с качественными признаками, многие из которых выражаются порядковыми номерами, индексами и другими условными знаками. В таких случаях необходимо использовать *непараметрические критерии.*

Известен целый ряд непараметрических критериев, среди которых видное место занимают так называемые *ранговые критерии,* применение которых основано на ранжировании членов сравниваемых групп. При этом сравниваются не сами по себе члены ранжированных рядов, а их порядковые номера, или ранги. Далее мы рассмотрим некоторые непараметрические критерии, применяемые для проверки нулевой гипотезы при сравнении как независимых, так и зависимых выборочных групп.

U **- критерий Уилкоксона (Манна—Уитни).** Гипотезу о принадлежности сравниваемых независимых выборок к одной и той же генеральной совокупности или к совокупностям с одинаковыми параметрами, т. е. H_0-гипотезу, можно проверить с помощью *рангового критерия Уилкоксона (Манна—Уитни).*

Для расчета U -критерия необходимо:

1. Расположить числовые значения сравниваемых выборок в возрастающем порядке в один общий ряд и пронумеровать члены общего ряда от 1 до $N = n_1 + n_2$. (Эти номера и будут «рангами» членов ряда.)

2. Отдельно для каждой выборки найти суммы рангов и определить величины,

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

и

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

которые отображают связь между суммами рангов первой и второй выборки.

3. В качестве U -критерия использовать меньшую величину $U_ф$, которую сравнить с табличным значением U_{kp} Условием для сохранения принятой H_0-гипотезы служит неравенство $U_ф > U_{kp}$. Критические точки U -критерия U_{kp} для n_1, n_2 и принимаемого уровня значимости α содержатся в таблице 3 Приложения.

Пример. На двух группах лабораторных мышей – опытной (n_1 = 9) и контрольной (n_2 = 11) – изучали воздействие на организм нового препарата. Испытание продолжалось один месяц. После этого масса тела животных, выраженная в граммах, варьировала следующим образом:

В опытной группе *80, 76, 75, 64, 70, 68, 72, 79, 83.*

В контрольной группе 70,78, 60, 80, 62, 68, 73, 60, 71, 66, 69.

Вычислим по выборкам: $\overline{x_1}$ =74,1 и $\overline{x_2}$ =68,8.

Проверим с помощью U -критерия, является ли разность в массе тела между опытной и контрольной группами мышей статистически достоверной.

Суммируя ранги отдельно для каждой группы, находим:

R_1= 4+6+9+12+14+15+17+19+20=112;

R_2= 1+2+3+5+7+8+10+11+13+16+18=94.

Подставляем эти данные в формулы:

$U_1 = 112 - \frac{9 \cdot 10}{2} = 67$; $U_2 = 94 - \frac{11 \cdot 12}{2} = 22$

Меньшую величину U_2 = 22 сравниваем с табличным U_{kp} значением для n_1=9, n_2=11 и уровня значимости α=0,01, которое равно U_{kp} =19.

Поскольку $U_{ф} > U_{kp}$, отвергнуть проверяемую H_0-гипотезу нельзя. Следовательно, различия, наблюдаемые между этими выборками, статистически недостоверны. Выборки не имеют значимых отличий.

Критерий знаков *z*. В тех случаях, когда результаты наблюдений выражаются не числами, а качественными признаками, принимающими два различных значения (помечаем их знаками плюс (+) и минус (—)), различия между попарно связанными членами сравниваемых выборок оценивают с помощью *критерия знаков z*. Конструкция этого критерия базируется на весьма простых соображениях: если попарно сравниваемые значения двух зависимых выборок существенно не отличаются друг от друга, то число плюсовых и минусовых разностей окажется совершенно одинаковым; если же заметно преобладают плюсы или минусы, это будет указывать на положительное или отрицательное действие изучаемого фактора на результативный признак. Большее число однозначных разностей служит в качестве фактически найденной величины *z*-критерия знаков. При этом нулевые разности, т. е. случаи, не давшие ни положительного, ни отрицательного результата, обозначаемые цифрой 0, в расчет не принимают и число парных наблюдений соответственно уменьшается.

Как и всякий другой выборочный показатель, *z*-критерий знаков является величиной случайной; он служит для проверки H_0-гипотезы, т. е. предположения о том, что совокупности, из которых взяты сравниваемые выборки, имеют одинаковые функции распределения. H_0-гипотеза отвергается, если $z_{ф} \geq z_{kp}$ для принятого уровня значимости α и числа парных наблюдений *n,* взятых без нулевых разностей. Критические точки z_{kp} для двух уровней значимость и числа парных наблюдений содержатся в таблице.

Задачи и упражнения:

1.Для изучения эффективности нового препарата железа были выбраны две группы пациентов с анемией. В первой группе пациенты в течение двух недель получали новый препарат, а во второй группе - получали плацебо. После этого было проведено измерение уровня гемоглобина в периферической крови. В первой группе средний уровень гемоглобина составил 115,4±1,2 г/л, а во второй - 103,7±2,3 г/л (данные представлены в формате $M \pm m$), сравниваемые совокупности имеют нормальное распределение. При этом численность первой группы составила 34, а второй - 40 пациентов. Необходимо сделать вывод о статистической значимости полученных различий и эффективности нового препарата железа.

2. После проведения вакцинации от гриппа среди студентов медицинского университета были подведены результаты: из 500 вакцинированных в период эпидемии заболели гриппом 20 человек, из 1600 отказавшихся от вакцинации гриппом заболели 200 человек. Оцените эффективность вакцинации от гриппа.

3. Учебной частью одной из кафедр медицинского университета было проведено исследование успеваемости студентов в зависимости от посещаемости лекций. Для студентов, посетивших менее половины лекционного курса (n=36), средняя оценка на экзамене составила 3,2, σ1=0,2. Для студентов, посетивших более 90% лекций по предмету (n=150), средняя оценка на экзамене составила 4,5, σ^2=0,5.Сделайте вывод о достоверности различий успеваемости студентов в зависимости от посещаемости лекций по предмету.

4. Результаты тестирования по 30-бальной шкале для группы X и группы Y представлены в таблице. Сравнить эффективность двух методов обучения студентов в двух группах для уровня статистической значимости β = 5%.

X	18	10	7	15	14	11	13			
Y	15	20	10	8	16	10	19	7	15	14	29

Практическое занятие 15. Корреляционная зависимость. Коэффициент корреляции.

Цель: формирование представления об элементах корреляционного анализа.

Теоретическая часть:

Зависимость между значениями одной случайной величины и условным математическим ожиданием другой случайной величины носит название статистической.

Чтобы изучить статистическую зависимость, нужно знать условное математическое ожидание случайной величины. Для его оценки необходимо знать аналитический вид двумерного распределения (X,Y). Однако, суждение об аналитическом виде двумерного распределения, сделанного по отдельной ограниченной по объёму выборке, может привести к серьёзным ошибкам. Поэтому идут на упрощение и переходят от условного математического ожидания случайной величины к условному среднему значению, т.е. принимают, что $M(Y)_x = \overline{Y}_x$

Зависимость между значениями одной случайной величины и условным средним значением другой случайной величины носит название **корреляционной** (от англ. correlation - согласование, связь, взаимосвязь, соотношение, взаимозависимость); термин впервые введен Гальтоном в 1888г.

Парный коэффициент корреляции Пирсона (1896 г.) изменяется в пределах от -1 до +1. Значение 0,00 интерпретируется как отсутствие корреляции. Корреляция определяет степень, с которой значения двух переменных пропорциональны друг другу.

Определяют две черты зависимости между переменными: величину зависимости и надежность зависимости.

Надежность зависимости – менее наглядные понятия, чем величина зависимости, однако чрезвычайно важна. Оно непосредственно связано с репрезентативностью той определенной выборки, на основе которой строятся выводы. Другими словами надежность говорит, насколько вероятно, что зависимость подобная найденной, будет вновь обнаружена (подтвердится) на данных другой выборки, извлеченной из той же самой популяции. Если исследование удовлетворяет некоторым специальным критериям, то надежность найденных зависимостей между переменными выборки можно количественно оценить и представить с помощью стандартной статистической меры (называемой р-уровень, или статистический уровень значимости).

Статистическая значимость результата представляет собой оцененную меру уверенности в его правильности. Уровень значимости или р-уровень, - это показатель, находящийся в убывающей зависимости от надежности результата. Более высокий р-уровень соответствует более низкому уровню доверия к найденной в выборке зависимости между переменными. Именно р-уровень представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю популяцию. Чем слабее зависимость между переменными, тем большего объема требуется выборка, чтобы значимо ее обнаружить. Другими словами, если зависимость между переменными почти отсутствует, объем выборки, необходимый для ее значимого обнаружения, почти равен объему всей популяции, которой предполагается бесконечным.

Рабочие формулы коэффициентов корреляции применяют с учетом того, с какой выборкой (большей или малой) мы имеем дело.

Например, <u>для малых выборок</u> удобнее всего пользоваться следующей формулой:

$$r=\frac{\sum xy-\frac{\sum x\sum y}{n}}{\sqrt{q_x\cdot q_y}} \quad (1)$$, где

$q_x=\sum x^2-\frac{\left(\sum x\right)^2}{n}$ и $q_y=\sum y^2-\frac{\left(\sum y\right)^2}{n}$, где x-варианты первого признака; y-варианты второго признака; n-число наблюдений в выборке.

Задачи и упражнения:

Определить зависимость содержания витамина С в продукте в зависимости от длительности его обработки при определенной температуре по результатам n=6 измерений (экспериментов). Какое количество витамина С (в долях к первоначальному) будет содержаться в продукте при обработке его в течение t^*_1 и t^*_2 часов. Построить график полученной зависимости и изобразить на координатной плоскости результаты измерений.

1.

x_i	0,11	0,21	0,31	0,41	0,51	0,61	t^*_1	t^*_2

y_i	0,97	0,94	0,9	0,86	0,83	0,78	0,25	0,74

2.

x_i	0,21	0,22	0,23	0,24	0,25	0,26	**t^*_1**	**t^*_2**
y_i	0,97	0,94	0,9	0,86	0,83	0,78	0,25	0,74

3.

x_i	0,14	0,24	0,34	0,44	0,54	0,64	**t^*_1**	**t^*_2**
y_i	0,97	0,94	0,9	0,86	0,83	0,78	0,25	0,74

Практическое занятие 16. Линейная регрессия. Коэффициенты регрессии.

Цель: формирование представления об элементах и принципах регрессионного анализа.

Теоретическая часть:

Для изучения корреляционных связей большое значение имеет коэффициент регрессии ρ, который показывает, насколько в среднем изменяется признак (X), если коррелирующий с ним признак (У) изменяется на определенную величину.

Коэффициент регрессии в конкретной выборке имеет два значения, а именно: ρ_{xy} и ρ_{yx}, т.е. прямое и обратное влияние признаков друг на друга. Формула для расчета коэффициента имеет вид:

$$\rho_{xy} = r\frac{S_x}{S_y};\ \rho_{yx} = r\frac{S_y}{S_x}$$

Статистическую зависимость Y от X описывают с помощью уравнения вида

$$M(Y)_x = f(x)$$

где $M(Y)_x$ - условное математическое ожидание величины Y, соответствующее данному значению x; x – отдельные значения величины X; $f(x)$ - некоторая функция. Это уравнение называется уравнением регрессии Y на X.

Обратную статистическую зависимость можно описать уравнением регрессии X на Y:

$$M(X)_y = \varphi(y)$$

где $M(X)_y$ - условное математическое ожидание величины X, соответствующее данному значению y случайной величины Y; $\varphi(y)$ - некоторая функция.

Функции $f(x)$ и $\varphi(y)$ называют соответственно регрессиями Y на X и X на Y, а их графики – линиями регрессии Y на X и X на Y. Уравнения регрессии выражают математическое ожидание случайной величины Y (или X) для случая, когда другая переменная принимает определенное число.

В зависимости от вида уравнений регрессии и формы соответствующих линий регрессии говорят о различной форме статистической зависимости между изучаемыми величинами – линейной, квадратичной, показательной и т.д.

Если функции $f(x), \varphi(y)$ линейные, т.е. уравнения регрессии можно представить в виде:

$$M(Y)_x = Ax + B$$

$$M(X)_y = Cy + D$$

где A,B,C,D – некоторые параметры, то описываемые этими уравнениями зависимости Y от X и X от Y называются линейными; линии регрессии при этом – прямые. Если линия регрессии не является прямой, то такую зависимость называют нелинейной.

Для характеристики формы связи между двумя случайными величинами, полученными в результате выборочных наблюдений, используют корреляционную зависимость $\overline{Y_x} = f(x)$ (или $\overline{X_y} = \varphi(y)$). Уравнения, описываемые подобной зависимостью, называют выборочными уравнениями регрессии.

Если функции $f(x)$, $\varphi(y)$ линейные, то выборочные уравнения линейной регрессии Y на Xи X на Y можно представить в виде:

$$\overline{Y_x} = \rho_{yx} x + b$$

$$\overline{X_y} = \rho_{xy} y + d$$

где Y_x и $\overline{X_y}$ - условные средние значения величин Y и X, параметры b и d - оценки B и D, ρ_{yx} и ρ_{xy} - выборочные оценки коэффициентов A и C.

Угловые коэффициенты ρ_{yx} и ρ_{xy} линий регрессии носят названия выборочных коэффициентов регрессии Y на X и X на Y соответственно. Они определяются как:

$$\rho_{yx} = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{S_x^2}; \quad \rho_{xy} = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{S_y^2},$$

где $S_x^2 = \overline{X^2} - (\overline{X})^2$, $S_y^2 = \overline{Y^2} - (\overline{Y})^2$.

Задачи и упражнения:

Найти выборочное уравнение прямой $\overline{y}_x - \overline{y} = r_B \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right)$ регрессии У на Х по данной корреляционной таблице.

1.

Y	X						n_y
	10	15	20	25	30	35	
6	4	2	–	–	–	–	6
12	–	6	2	–	–	–	8
18	–	–	5	40	5	–	50
24	–	–	2	8	7	–	17
30	–	–	–	4	7	8	19
n_x	4	8	9	52	19	8	n=100

2.

Y	X						n_y
	3	13	23	33	43	53	
8	4	2	–	–	–	–	6
10	–	6	2	–	–	–	8
12	–	–	5	40	5	–	50
14	–	–	2	8	7	–	17
16	–	–	–	4	7	8	19
n_x	4	8	9	52	19	8	n=100

3.

Y	X						n_y
	11	15	19	23	27	31	
10	4	2	–	–	–	–	6
20	–	6	2	–	–	–	8
30	–	–	5	40	5	–	50
40	–	–	2	8	7	–	17
50	–	–	–	4	7	8	19
n_x	4	8	9	52	19	8	n=100

4.

Y	X						n_y
	20	23	26	29	32	35	
16	4	2	–	–	–	–	6
19	–	6	2	–	–	–	8
22	–	–	5	40	5	–	50
25	–	–	2	8	7	–	17
28	–	–	–	4	7	8	19
n_x	4	8	9	52	19	8	n=100

5.

Y	X						n_y
	14	18	22	26	30	34	
5	4	2	–	–	–	–	6
10	–	6	2	–	–	–	8
15	–	–	5	40	5	–	50
20	–	–	2	8	7	–	17
25	–	–	–	4	7	8	19
n_x	4	8	9	52	19	8	n=100

4.СПИСОК ЛИТЕРАТУРЫ

Основная литература:

Гусак, А. А. Теория вероятностей. Примеры и задачи [Электронный ресурс]: учебное пособие / А. А. Гусак, Е. А. Бричикова. — Электрон. текстовые данные. — Минск: ТетраСистемс, 2013. — 287 с. — Режим доступа: http://www.iprbookshop.ru/28244.html

Дополнительная литература:

Климов, Г. П. Теория вероятностей и математическая статистика: учебник / Г. П. Климов. — Москва: Московский государственный университет имени М.В. Ломоносова, 2011. — 368 с. — Режим доступа: http://www.iprbookshop.ru/13115.html