Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

сборник учебно-методических материалов

для направления подготовки 37.03.01 – Психология

2017 г.

Печатается по решению
редакционно-издательского совета
факультета математики и информатики
Амурского государственного
Университета

Составители: Двоерядкина Н.Н., Гришкина Т.Е.

Математическая статистика: сборник учебно-методических материалов для направления подготовки 37.03.01. – Благовещенск: Амурский гос. ун-т, 2017.

Рассмотрен на заседании кафедры общей математики и информатики 04.12.2017, протокол № 4.

ВВЕДЕНИЕ

Дисциплина «Математическая статистика» направлена на развитие у студентов умения использовать математические методы систематизации, интерпретации и обработки статистических данных для научных и практических выводов.

Задачи дисциплины:

- изучение основных разделов математической статистики, овладение понятиями, утверждениями, выводами данных разделов и методами математического исследования;
- овладение методами математического описания типовой математической модели процесса или явления, навыками разработки плана математической обработки экспериментальных данных; методами математической статистики; методами математической обработки экспериментальных данных, полученных в разных сериях экспериментов, методикой составления приближенной модели зависимости практических величин на основании имеющихся экспериментальных данных.

В результате освоения обучающийся должен демонстрировать следующие результаты образования:

1) знать: математические и статистические методы, для обработки данных, полученных при решении различных профессиональных задач;

2) уметь: получать, обрабатывать и интерпретировать данные исследований с помощью математико-статистического аппарата;

3) владеть: навыками применения стандартных статистических пакетов для обработки данных, полученных при решении различных профессиональных задач.

Сборник учебно-методических материалов в своей структуре состоит из краткого конспекта лекций, методических указаний для практических, лабораторных занятий и самостоятельной работы.

Лекции, практические и лабораторные занятия и самостоятельная работа содействуют развитию познавательной активности студентов, формированию необходимых компетенции. Представленные виды занятий предусматривают как продуктивную, так и репродуктивную деятельность студента. При применении активных методов обучения доминирующими видами деятельности являются частично-поисковые, творческие, исследовательские.

Система задач и упражнений на практических и лабораторных занятиях обеспечивает формирование у студентов системы математических знаний, умений, компетенции, познавательного интереса; проверку качества усвоения изучаемого материала.

Задания для самостоятельной работы включают в себя задачи и упражнения тренировочного типа (в форме домашних заданий к практическим занятиям; самостоятельная работа над книгой или конспектом лекции по отбору и систематизации учебного материала) и реконструктивно-вариативного типа (при выполнении этих заданий студенты применяют правила, теоремы в различных ситуациях; реконструируют известный учебный материал или способы решения задач с целью их приложения к решению заданной задачи с измененными условиями).

Лабораторные занятия по данной дисциплине способствуют развитию аналитических и вычислительных способностей и формированию навыков, связанных с анализом, представлением и обработкой реальных эмпирических статистических данных. Основным требованием преподавателя к студентам является обязательное присутствие студентов на всех практических и лабораторных занятиях, а также выполнение всех заданий преподавателя, как текущих, так и контрольных.

1. КРАТКОЕ ИЗЛОЖЕНИЕ ЛЕКЦИОННОГО МАТЕРИАЛА

Раздел 1. Количественные характеристики случайных событий

Ключевые вопросы

Понятие о событии. Случайные и неслучайные события. Частота, частость и вероятность. Статистическое определение вероятности. Совместное появление событий. Зависимость между событиями. Преобразования событий. Уровни количественного определения событий. Распределения вероятностей событий. Ранжирование событий в системе по вероятностям. Меры связи между классифицированными событиями. Последовательности событий. Ранжирование событий по величине. Распределение вероятностей ранжированной системы упорядоченных событий. Количественные характеристики распределения вероятностей системы упорядоченных событий. Меры корреляции рангов.

Основные определения, теоремы и методы решения задач

Математическая статистика – раздел математической науки, изучающий методы обработки большого количества экспериментальных данных.

Математическая статистика опирается на теорию вероятностей, позволяющую оценить надёжность и точность выводов, делаемых на основании ограниченного статистического материала.

Классической вероятностью P(A) события A называется отношение числа благоприятствующих этому событию исходов (т.е. таких, при которых событие A обязательно произойдет) к общему числу несовместных единственно возможных и равновозможных исходов.

Геометрической вероятностью события A называется отношение меры области, благоприятствующей появлению события A, к мере пространства элементарных исходов.

Статистической вероятностью события A называется число, около которого колеблются частости W(A) появления этого события во многих сериях выборочных испытаний больших объемов, проводимых в одинаковых условиях.

К основным типам комбинаторных задач относятся отыскание числа перестановок, размещений, сочетаний, перестановок с повторениями, размещений с повторениями, сочетаний с повторениями.

Объединением (или суммой) нескольких случайных событий называется событие, состоящее в осуществлении, по крайней мере, одного из данных событий.

Совмещением (или произведением) двух событий A и B называется событие, состоящее в одновременном осуществлении обеих событий.

Вероятность суммы несовместных событий равна сумме вероятностей этих событий:

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

Вероятность суммы совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:

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

Вероятность совместного появления нескольких событий, независимых в совокупности, равна произведению вероятностей этих событий:

$$P(A_1, A_2, \ldots, A_n) = P(A_1) \cdot P(A_2) \ldots P(A_n).$$

Вероятность совместного появления нескольких зависимых событий равна произведению вероятности одного из них на условные вероятности всех остальных, причем вероятности каждого последующего события вычисляются в предположении, что все предыдущие события произошли:

$$P(A_1, A_2, ..., A_n) = P(A_1) \cdot P_{A_1}(A_2) P_{A_1 A_2}(A_3) \dots P_{A_1 A_2 \dots A_{n-1}}(A_n).$$

Если в результате испытания может появиться n событий, независимых в совокупности, то вероятность появления хотя бы одного из них равна $P(A) = 1 - q_1 q_2 ... q_n$, здесь событие А обозначает наступление хотя бы одного из событий A_i, а q_i – вероятность противоположных событий $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$.

Вероятность события А, которое может наступить при условии появления одного из несовместных событий, образующих полную группу событий, равна сумме произведений вероятности каждого из событий на соответствующую условную вероятность события А:

$$P(A) = P(H_1) P_{H_2}(A) + P(H_2) P_{H_2}(A) + \dots + P(H_n) P_{H_n}(A).$$

Условные вероятности гипотез определяются по формулам Байеса:

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

Вероятность появления события А m раз в n независимых испытания, в каждом из которых вероятность этого события постоянна и равна p, находится по формуле Бернулли:

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

Если число испытаний n велико, то применение формулы Бернулли приводит к очень громоздким вычислениям. В таких случаях пользуются приближенными формулами, основанными на теоремах Лапласа и Пуассона.

Локальная теорема Лапласа: Вероятность $P_n(m)$ того, что в n испытаниях, в каждом из которых вероятность появления события равна p, событие наступит ровно m раз приближённо равна

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}, \text{ где } x = \frac{m - np}{\sqrt{npq}}, \ \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Формула Пуассона: Если вероятность p наступления события постоянна и мала, а число испытаний n велико, то $P_n(m) \approx \frac{\lambda^m \cdot e^{-\lambda}}{m!}$, где $\lambda = np$.

Интегральная теорема Лапласа: Если вероятность p наступления некоторого события в каждом испытании постоянна (0 < p <1), а число испытаний n достаточно велико, то вероятность того, что это событие наступит не менее a раз и не более b приближенно равна

$$P(a \le m \le b) = \Phi(\beta) - \Phi(\alpha), \text{ где } \alpha = \frac{a - np}{\sqrt{npq}}, \quad \beta = \frac{b - np}{\sqrt{npq}},$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz \text{ – функция Лапласа.}$$

Раздел 2. Количественные характеристики случайной величины

Ключевые вопросы

Случайная величина. Распределение вероятностей значений случайной величины. Основные свойства распределений. Меры положения. Меры асимметрии и эксцесса. Исходные положения. Вычисление мер положения, рассеивания, асимметрии и эксцесса по несгруппированным данным. Группировка данных и получение эмпирических распределений. Вычисление мер положения, рассеивания, асимметрии и эксцесса по эмпирическому распределению. Общие положения. Нормальный закон. Нормализация распределений. Другие законы распределения, важные для психологии.

Основные определения, теоремы и методы решения задач

Случайная величина X – это некоторая функция элементарного события $\omega: X = \varphi(\omega)$, где $\omega \in U$. Значение этой функции зависит от того, какое элементарное событие ω появилось в результате опыта. Случайные величины часто обозначают большими буквами, а их значения малыми. Случайная величина называется дискретной, если ее возможные значения могут быть пронумерованы числами натурального ряда. Непрерывной называют случайную величину, которая может принимать все значения из некоторого конечного или бесконечного промежутка.

Законом распределения случайной величины называется любое правило, позволяющее находить вероятность возможных событий (значений), связанных со случайной величиной. Закон распределения дискретных случайных величин может быть задан в форме ряда, многоугольника и функции (интегрального закона) распределения.

Непрерывную случайную величину можно задать функцией распределения F(x) действительной переменной x, определяемой формулой

$$F(x) = P(X < x)$$

и функцией, которая называется плотностью распределения или дифференциальной функций и определяется формулой

$$f(x) = F'(x).$$

Основными числовыми характеристиками являются математическое ожидание, дисперсия и среднее квадратичное отклонение. Математическое ожидание определяется по формулам:

$$M(X) = \sum_{i=1}^{n} x_i \cdot p_i \quad \text{для дискретных случайных величин,}$$

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$$ - для непрерывных.

Дисперсия: $D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 \cdot p_i$ – для дискретных случайных величин,

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f(x)dx$$ – для непрерывных.

Дисперсию также можно определять по единой формуле:

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

Среднее квадратичное отклонение $\sigma_X = \sqrt{D(X)}$

Центральным моментом порядка *k* случайной величины Х называется математическое ожидание величины $(X - m_x)^k$. $\mu_k = M[(X - m_x)^k]$.

Для дискретной случайной величины: $\mu_k = \sum_{i=1}^{n} (x_i - m_x)^k p_i$, для непрерывной случайной величины: $\mu_k = \int_{-\infty}^{\infty} (x - m_x)^k f(x)dx$.

Центральный момент первого порядка всегда равен нулю, а центральный момент второго порядка равен дисперсии. Центральный момент третьего порядка характеризует асимметрию распределения.

Отношение центрального момента третьего порядка к среднему квадратическому отклонению в третьей степени называется коэффициентом асимметрии $a_x = \frac{\mu_3}{\sigma_x^3}$.

Для характеристики островершинности и плосковершинности распределения используется величина, называемая эксцессом $E_x = \frac{\mu_4}{\sigma_x^4} - 3$.

Дискретная случайная величина Х имеет биномиальный закон распределения, если она принимает значения 0,1,2,…,m,…,n с вероятностями $P_n(m) = C_n^m \cdot p^m \cdot q^{n-m}$, где 0 < p < 1, q=1-p, m=0,1,…,n.

Дискретная случайная величина Х имеет закон распределения Пуассона, если она принимает значения 0,1,2,…,m,…(бесконечное, но счетное множество значений) с вероятностями

$$P_n(m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!}, \quad \lambda = n \cdot p.$$

Дискретная случайная величина Х имеет геометрическое распределение, если она принимает значения 0,1,2, …,m,… (бесконечное, но счетное множество значений) с вероятностями

$$P_n(m) = p \cdot q^{m-1}, \text{ где } 0 < p < 1,\ q = 1 - p,\ m = 1,2,....$$

Распределение вероятностей называют нормальным, если оно описывается дифференциальной функцией $\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$.

Равномерным распределением называют распределение вероятностей непрерывной случайной величины X, дифференциальная функция f (x) которой сохраняет постоянное значение на сегменте [a; b] и равна нулю вне этого сегмента, т. е.

$$f(x)=\begin{cases}0 & \text{при} \quad x<a,\\ \frac{1}{b-a} & \text{при} \quad a\le x\le b,\\ 0 & \text{при} \quad x>b.\end{cases}$$

Непрерывная случайная величина x распределена по показательному закону, если ее плотность вероятности имеет вид:

$$f(x)=\begin{cases}0 & \text{при} \;\; x<0,\\ \lambda e^{-\lambda x} & \text{при} \;\; x\ge 0. \quad \text{где} \;\; \lambda>0.\end{cases}.$$

Раздел 3. Выборочный метод и статистическое оценивание

Ключевые вопросы

Выборочный метод. Количественные характеристики генеральной совокупности и выборки. Свойства точечных оценок. Методы вычисления. Ошибка выборки. Определение объема выборки. Интервальные оценки среднего, доли, дисперсии.

Основные определения, теоремы и методы решения задач

Выборкой или выборочной совокупностью называется совокупность случайно отобранных объектов. Генеральной совокупностью называют совокупность объектов, из которых производится выборка.

Вариационным рядом называется ряд значений исследуемого признака с указанием соответствующих весов (частот или относительных частот). Выделяют дискретные и вариационные ряды.

Гистограммой частот сгруппированной выборки называется кусочно-постоянная функция, постоянная на интервалах группировки и принимающая на каждом из них соответственно значения $\frac{n_i}{\Delta x_i}$, i = 1,2, …, k. Площадь ступенчатой фигуры под графиком гистограммы равна объему выборки n. Полигоном частот называется ломаная с вершинами в точках $(x_i,\frac{n_i}{\Delta x_i})$, i – 1,2 ,…, k, где k – число интервалов вариационного ряда.

Эмпирической функцией распределения случайной величины X называется функция $F_n(x)=\frac{n_x}{n}$, где n – объем выборки; n_x – число значений x_i из выборки, удовлетворяющих неравенству $x_i < x$.

Выборочная средняя вычисляется по формуле:

$$\overline{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{n}.$$

Выборочная дисперсия вычисляется по формуле:

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

Выборочные начальные моменты вычисляются по формулам:

$$\overline{m}_к(x) = \frac{\sum_{i=1}^{k} (x_i)^к n_i}{n}.$$

Выборочные центральные моменты вычисляются по формулам:

$$\mu_к(x) = \frac{\sum_{i=1}^{k} (x_i - \overline{x}_b)^к n_i}{n}.$$

Эксцесс: $E_x = \dfrac{\mu_4}{S^4} - 3$.

Асимметрия: $a_x = \dfrac{\mu_3}{S^3}$.

Точечной оценкой параметра называется оценка, определяемая одним числом.

Оценка $\tilde{\Theta} = \tilde{\Theta}$ (x1,x2,…,x ^{n})–должна приближаться к оцениваемому параметру Θ по мере увеличения объема выборки. Если оценка стремится по вероятности к оцениваемому параметру, то она называется состоятельной.

Оценка параметра $\tilde{\Theta}$ называется несмещенной, если она при любом объеме выборки имеет математическое ожидание, совпадающее с оцениваемым параметром, т.е. М ($\tilde{\Theta}$) = Θ .

Несмещенная оценка $\tilde{\Theta}$ параметра Θ называется эффективной, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра Θ, вычисленных по выборкам одного и того же объема n.

Доверительным интервалом для параметра θ называется интервал (θ_1, θ_2), содержащий (покрывающий) истинное значение θ с заданной вероятностью $\gamma = 1 - \alpha$. Число γ называется доверительной вероятностью, а значение γ − уровнем значимости, границы θ_1 и θ_2 являющимися случайными величинами − соответственно нижней и верхней границами доверительного интервала.

Величина $\alpha = 1 - \gamma$ указывает на то, что те значения параметра θ, для которых $|\theta - \tilde{\theta}| > \varepsilon$ нужно признать противоречащими опытным данным.

Доверительный интервал для оценки математического ожидания а при неизвестном σ :

$$\left(\overline{x}_в - \frac{t\sigma}{\sqrt{n}}; \ \overline{x}_в + \frac{t\sigma}{\sqrt{n}} \right)$$

с надежностью γ покрывает неизвестный параметр a; точность оценки $\delta = \dfrac{t\sigma}{\sqrt{n}}$.

Доверительный интервал для оценки математического ожидания а при неизвестном σ для заданной доверительной вероятности γ вычисляется по формуле:

$$\bar{x}_в - t_\gamma S/\sqrt{n} < a < \bar{x}_в + t_\gamma S/\sqrt{n}, \text{ где } \bar{x}_в = \frac{\sum_{i-1}^{n} x_i}{n},\ S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_в)^2 n_i}{n-1}.$$

Доверительный интервал для дисперсии записан в скобках:

$$P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2};n-1}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \right) = \gamma$$

Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Нулевой (основной) называют выдвинутую гипотезу H0. Конкурирующей (альтернативной) называют гипотезу H1, которая противоречит нулевой.

Простой называют гипотезу, содержащую только одно предположение, сложной – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

В результате проверки правильности выдвинутой нулевой гипотезы (такая проверка называется статистической, так как производится с применением методов математической статистики) возможны ошибки двух видов: ошибка первого рода, состоящая в том, что будет отвергнута правильная нулевая гипотеза, и ошибка второго рода, заключающаяся в том, что будет принята неверная гипотеза.

Вероятность ошибки первого рода называется уровнем значимости α.

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Статистическим критерием называется случайная величина К с известным законом распределения, служащая для проверки нулевой гипотезы.

Критической областью называют область значений критерия, при которых нулевую гипотезу отвергают, областью принятия гипотезы – область значений критерия, при которых гипотезу принимают.

Критерий для проверки гипотезы о математическом ожидании.

Пусть генеральная совокупность X имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу a_0.

Пусть известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема n найдем выборочное среднее $\bar{x}_B$ и проверим нулевую гипотезу

H_0: M(X) = a_0.

Учитывая, что выборочное среднее $\bar{X}$ является несмещенной оценкой M(X), то есть M($\bar{X}$)=M(X), можно записать нулевую гипотезу так: M($\bar{X}$)=a_0. Для ее проверки выберем критерий

$$U = \frac{\bar{X} - a_0}{\sigma(\bar{X})} = \frac{(\bar{X} - a_0)\sqrt{n}}{\sigma}.$$

Это случайная величина, имеющая нормальное распределение, причем, если нулевая гипотеза справедлива, то M(U) = 0, σ(U) = 1.

Выберем критическую область в зависимости от вида конкурирующей гипотезы:

если H_1: $M(\overline{X}) \neq a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \dfrac{1-\alpha}{2}$, критическая область двусторонняя, $U_{набл} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$, и, если $|U_{набл}| < u_{кр}$, то нулевая гипотеза принимается; если $|U_{набл}| > u_{кр}$, то нулевая гипотеза отвергается.

если H_1: $M(\overline{X}) > a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \dfrac{1-2\alpha}{2}$, критическая область правосторонняя, и, если $U_{набл} < u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} > u_{кр}$, то нулевая гипотеза отвергается.

если H_1: $M(\overline{X}) < a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \dfrac{1-2\alpha}{2}$, критическая область левосторонняя, и, если $U_{набл} > -u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} < -u_{кр}$, то нулевая гипотеза отвергается.

Достоинством критерия Пирсона является его универсальность: с его помощью можно проверять гипотезы о различных законах распределения.

Проверка гипотезы о нормальном распределении.

Пусть получена выборка достаточно большого объема n с большим количеством различных значений вариант. Для удобства ее обработки разделим интервал от наименьшего до наибольшего из значений вариант на s равных частей и будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку:

варианты………..x_1 x_2 … x_s

частоты…………n_1 n_2 … n_s ,

где x_i – значения середин интервалов, а n_i – число вариант, попавших в i-й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_B$ и выборочное среднее квадратическое отклонение σ_B. Проверим предположение, что генеральная совокупность распределена по нормальному закону с параметрами $M(X) = \bar{x}_B$, $D(X) = \sigma_B^2$. Тогда можно найти количество чисел из выборки объема n, которое должно оказаться в каждом интервале при этом предположении (то есть теоретические частоты).

Для этого по таблице значений функции Лапласа найдем вероятность попадания в i-й интервал: $p_i = \Phi\left(\dfrac{b_i - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\dfrac{a_i - \bar{x}_B}{\sigma_B}\right)$, где a_i и b_i - границы i-го интервала. Умножив полученные вероятности на объем выборки n, найдем теоретические частоты: $n_i = n \cdot p_i$. Наша цель – сравнить эмпирические и теоретические частоты, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о нормальном распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе. Для этого используется критерий в виде случайной величины

$$\chi^2 = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}.$$

Для выбранного критерия строится правосторонняя критическая область, определяемая условием

$$p(\chi^2 > \chi^2_{kp}(\alpha, k)) = \alpha,$$

где α – уровень значимости. Следовательно, критическая область задается неравенством $\chi^2 > \chi^2_{kp}(\alpha, k),$ а область принятия гипотезы - $\chi^2 < \chi^2_{kp}(\alpha, k)$.

Итак, для проверки нулевой гипотезы H_0: генеральная совокупность распределена нормально – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{набл} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'},$$

а по таблице критических точек распределения χ^2 найти критическую точку $\chi^2_{кр}(\alpha, k)$, используя известные значения α и k = s – 3. Если $\chi^2_{набл} < \chi^2_{kp}$ - нулевую гипотезу принимают, при $\chi^2_{набл} > \chi^2_{kp}$ ее отвергают.

Пусть составляющие X и Y двумерной случайной величины (X, Y) зависимы. Будем считать, что одну из них можно приближенно представить как линейную функцию другой, например Y ≈ g(X) = α + βX, и определим параметры α и β с помощью метода наименьших квадратов.

Определение. Коэффициент $\beta = r\frac{\sigma_y}{\sigma_x}$ называется коэффициентом регрессии Y на X, а прямая $y - m_y = r\frac{\sigma_y}{\sigma_x}(x - m_x)$ - прямой среднеквадратической регрессии Y на X.

Аналогично можно получить прямую среднеквадратической регрессии X на Y: $x - m_x = r\frac{\sigma_x}{\sigma_y}(y - m_y)$.

Рассмотрим выборку объема n, извлеченную из нормально распределенной двумерной генеральной совокупности (X, Y). Вычислим выборочный коэффициент корреляции $r_В$. Пусть он оказался не равным нулю. Это еще не означает, что и коэффициент корреляции генеральной совокупности не равен нулю. Поэтому при заданном уровне значимости α возникает необходимость проверки нулевой гипотезы H_0: r = 0 о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе H_1: r ≠ 0. Таким образом, при принятии нулевой гипотезы X и Y некоррелированы, то есть не связаны линейной зависимостью, а при отклонении H_0 они коррелированы.

В качестве критерия примем случайную величину

$$T = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B^2}},$$

которая при справедливости нулевой гипотезы имеет распределение Стьюдента с k=n–2 степенями свободы. Из вида конкурирующей гипотезы следует, что критическая область двусторонняя с границами ± $t_{кр}$, где значение $t_{кр}(\alpha, k)$ находится из таблиц для двусторонней критической области.

Вычислив наблюдаемое значение критерия

$$T_{набл} = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B^2}}$$

и сравнив его с $t_{кр}$, делаем вывод: если |Тнабл| < tкр – нулевая гипотеза принимается (корреляции нет); если |Тнабл| > tкр – нулевая гипотеза отвергается (корреляция есть).

Пусть изучается двумерная случайная величина (X, Y), и получена выборка из n пар чисел $(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$. Будем искать параметры прямой линии среднеквадратической регрессии Y на X вида $Y = \rho_{yx}x + b$,

Подбирая параметры ρ_{yx} и b так, чтобы точки на плоскости с координатами (x_1, y_1), $(x_2, y_2),\ldots, (x_n, y_n)$ лежали как можно ближе к прямой. Используем для этого метод наименьших квадратов и найдем минимум функции $F(\rho,b)=\sum_{i=1}^{n}(Y_i-y_i)^2=\sum_{i=1}^{n}(\rho x_i+b-y_i)^2$.

Приравняем нулю соответствующие частные производные:

$$\begin{gathered}\frac{\partial F}{\partial \rho}=2\sum_{i=1}^{n}(\rho x_i+b-y_i)x_i=0\\ \frac{\partial F}{\partial b}=2\sum_{i=1}^{n}(\rho x_i+b-y_i)=0\end{gathered}.$$

В результате получим систему двух линейных уравнений относительно ρ и b:

$$\begin{cases}\left(\sum x^2\right)\rho+\left(\sum x\right)b=\sum xy\\ \left(\sum x\right)\rho+nb=\sum y\end{cases}.$$

Ее решение позволяет найти искомые параметры в виде:

$$\rho_{xy}=\frac{n\sum xy-\sum x\cdot\sum y}{n\sum x^2-\left(\sum x\right)^2};\quad b=\frac{\sum x^2\cdot\sum y-\sum x\cdot\sum xy}{n\sum x^2-\left(\sum x\right)^2}.$$

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ

Практические занятия сопровождают лекционный курс дисциплины. Практические занятия должны проводиться в логичном единстве с теоретическим курсом, подкрепляя и уточняя понятийный аппарат.

Каждое практическое занятие начинается с теоретического опроса необходимого материала и проверки домашнего задания. Далее на конкретных примерах рассматриваются пути и способы применения тех математических методов, которые не требуют использования электронных вычислительных машин. При этом необходимо активизировать самостоятельную работу студентов. Задания и методические указания к ним выдаются студентам, каждый из которых выбирает оптимальный для себя темп работы. Преподавателю отводится роль консультанта и помощника. Задания, вызвавшие трудности у большинства студентов, разбираются на доске.

В конце занятия выдается домашнее задание, состоящее из теоретических вопросов, уяснение которых необходимо для следующего занятия и практических заданий по пройденному материалу.

При выполнении домашнего задания решать задачи удобнее поэтапно, в той последовательности, в какой эти задания сформулированы. В этом случае при возникновении трудностей будет легче обратиться к анализу тех тем, которые изложены в лекции и задач, разобранных на практическом занятии.

После выполнения практической части задания следует найти ответы на теоретические вопросы, заданные преподавателем и таким образом подготовится к осознанному восприятию следующего материала.

Активная, регулярная самостоятельная работа над домашним заданием – путь к успешному усвоению дисциплины.

Раздел 1. Количественные характеристики случайных событий

Основные вопросы

Понятие о событии. Случайные и неслучайные события. Частота, частость и вероятность. Статистическое определение вероятности. Распределения вероятностей событий. Ранжирование событий в системе по вероятностям. Меры связи между классифицированными событиями. Последовательности событий. Ранжирование событий по величине. Распределение вероятностей ранжированной системы упорядоченных событий.

Типовые задания

1. В цехе работают 6 мужчин и 4 женщины. По табельным номерам отобраны 7 человек. Найти вероятность того, что среди отобранных лиц окажутся три женщины.

2. Два студента условились встретиться в определенном месте между 8 и 9 часами. Пришедший первым ждет второго в течение 10 минут. Чему равна вероятность встречи студентов, если приход каждого из них в течение указанного часа может произойти наудачу и момент прихода независим.

3. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго 0,8. Найти вероятность того, что при одном залпе в мишень: попадет только один, попадут оба, ни один не попадет.

4. Одинаковые детали обрабатываются тремя рабочими на трёх станках. Вероятность брака у них равна 0,01; 0,02; 0,03. Обработанные детали откладываются в один ящик. Какова вероятность того, что наугад взятая деталь будет бракованной, если производительности станков относятся как 2:3:5. Иначе говоря, каков общий процент бракованных деталей производится тремя рабочими?

5. Завод выпускает за три декады месяца соответственно 20 %, 30 %, 50 % задания, причём вероятности брака соответственно составляют 0,01, 0,012, 0,015. Найти вероятность того, что изделие выпущено в первой декаде, если в нём обнаружен дефект.

6. Два равносильных соперника играют в шахматы. Что вероятнее выиграть:

а) одну партию из двух или две из четырёх?

б) не менее двух партий из четырёх или не менее трёх партий из пяти?

Ничьи во внимание не принимаются.

7. Вероятность попадания в цель при одном выстреле равна 0,4. Найти вероятность 100 попаданий из 320 выстрелов.

8. Торговая база получила 10 000 электрических лампочек. Вероятность повреждения электрической лампочки в пути равна 0,0001. Определить вероятность того, что в пути повреждено 4 электрических лампочки.

9. Вероятность наступления события в каждом из одинаковых и независимых испытаний равна 0,8. Найти вероятность того, что в 120 испытаниях событие наступит не менее 70 раз и не более 90 раз.

10. Два стрелка сделали по два выстрела по мишени. Вероятность попадания в мишень для первого стрелка равна 0,6, для второго - 0,7. Необходимо: а) составить закон распределения общего числа попаданий; б) найти математическое ожидание и дисперсию этой случайной величины.

Раздел 2. Количественные характеристики случайной величины

Основные вопросы

Случайная величина. Распределение вероятностей значений случайной величины. Основные свойства распределений. Меры положения. Меры асимметрии и эксцесса. Исходные положения. Вычисление мер положения, рассеивания, асимметрии и эксцесса по несгруппированным данным. Группировка данных и получение эмпирических распределений. Вычисление мер положения, рассеивания, асимметрии и эксцесса по эмпирическому распределению. Нормальный закон. Нормализация распределений. Другие законы распределения, важные для психологии.

Типовые задания

1. В лотерее разыгрывается: автомобиль стоимостью 5000 долл., 4 телевизора стоимостью 250 долл. каждый, 5 видеомагнитофонов стоимостью 200 долл. каждый. Всего продается 1000 билетов по 7 долл. Составить закон распределения чистого выигрыша, полученного участником лотереи, купившим один билет.

2. Монету бросают 6 раз. Случайная величина X – число выпадений герба. Составить ее закон распределения.

3. В партии из семи деталей имеются 3 стандартных. Наудачу отобраны четыре детали. Составьте закон распределения случайной величины X – числа стандартных деталей среди отобранных.

4. Случайная величина X в интервале (3, 5) задана дифференциальной функцией $f(x)=\frac{3}{4}x^2+6x-\frac{45}{4}$; вне этого интервала $f(x)=0$. Найти математическое ожидание X.

5. Дана функция распределения случайной величины X:

$$F(x)=\begin{cases}0 & при\ x\le 0,\\ x^2/4 & при\ 0<x\le 2,\\ 1 & при\ x>2.\end{cases}$$

Найти плотность вероятности *f(x)*; построить графики *f (x)* и *F(x)*; вычислить математическое ожидание M(X), дисперсию D(X), эксцесс, асимметрию.

6. Средний рост взрослых женщин некоторого национального округа равен 167,3 см; σ = 5,8 см. Каков общий процент женщин ростом: не превышающим 170 см, не меньшим, чем 165 см.

7. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее 3 мин.

8. Количество электроэнергии, потребляемой поселком в течение суток, является случайной величиной, математическое ожидание которой равно 4 тыс. кВт-ч. Оценить вероятность того, что в ближайшие сутки потребление энергии: а) превысит 8 тыс. кВт- ч.; б) не превысит 6 тыс. кВт- ч.

9. Пользуясь неравенством Чебышева, оценить вероятность того, что из посеянных 5000 семян число взошедших окажется от 3750 до 4250, если известно, что M(X)=4000. Определить вероятность попадания случайной величины в заданный интервал.

10. Для определения средней урожайности на площади 100 000 га взято на выборку по одному гектару от каждого участка размером 100 га. Определите вероятность того, что средняя выборочная вероятность будет отличаться от действительной средней по всей площади не более чем на 0,5 ц, если дисперсия урожайности на отдельных участках (по 100 га) не превышает 2 ц.

Раздел 3. Выборочный метод и статистическое оценивание

Основные вопросы

Свойства точечных оценок. Методы вычисления. Ошибка выборки. Определение объема выборки. Интервальные оценки среднего, доли, дисперсии.

Типовые задания

1. Предприятие получает сырье, характеризующееся процентным содержанием в нем полезного вещества А, а также стабильностью этого показателя. Для поступившей партии сырья были произведены анализы концентрации вещества А по 40 опытам: 2,17; 2,27; 2,60; 2,76; 2,81; 3,00; 3,05; 3,06; 3,22; 3,28; 3,37; 3,60; 3,63; 3,68; 3,69; 3,90; 4,27; 4,39; 4,67; 4,69; 4,84; 5,02; 5,02; 5,02; 5,17; 5,29; 5,35; 5,61; 5,62; 5,62; 6,01; 6,64; 6,68; 7,07; 7,12; 7,35; 7,36; 7,40; 7,99; 8,19.

Пусть концентрация вещества А может быть рассмотрена как случайная величина X, распределенная по нормальному закону.

Найти выборочные точечные характеристики: выборочную среднюю, выборочную дисперсию, эксцесс, асимметрию, моду, коэффициент вариации; выдвинуть гипотезу относительно близости распределения к нормальному; записать плотность вероятности и функцию распределения; найти доверительный интервал, покрывающий математическое ожидание с заданной доверительной вероятностью $\gamma = 1 - \alpha = 0{,}95$, считая σ неизвестным; найти доверительный интервал, покрывающий среднеквадратическое отклонение σ с заданной доверительной вероятностью $\gamma = 1 - \alpha = 0{,}95$.

2. Студенты группы из 30 человек написали контрольную работу. Каждый студент получил определённое количество баллов:75, 145, 150, 180, 125, 150, 150, 165, 95, 135, 130, 70, 130, 105, 135, 135, 100, 160, 60, 85, 120, 60, 145, 150, 135, 132, 140, 65, 170,155.

Требуется:

1. Построить сгруппированный вариационный ряд
2. Построить сгруппированный статистический ряд
3. Построить эмпирическую функцию распределения
4. Построить гистограмму и полигон относительных частот
5. Найти выборочные точечные характеристики: выборочную среднюю, выборочную дисперсию, эксцесс, асимметрию, моду.
6. Выдвинуть гипотезу относительно близости распределения к нормальному.

3. Как известно, почерк человека, в том числе наклон букв, тесно связан с его характером. Низкий наклон ($30\text{-}40^0$) свидетельствует о вспыльчивости и возбудимости человека, излишней прямоте и торопливости в поступках; наклон в $40\text{-}50^0$ характеризует гармоничное развитие натуры; наклон $50\text{-}90^0$ свидетельствует о самообладании, узком диапазоне увлечений. Среди студентов

АмГУ выборочно был исследован почерк 50 человек. Оказалось, что у 30% присутствующих низкий наклон, у 50%-наклон 40-50^0 и у 20% наклон 50-90^0. Найти распределение частот. Относительных частот, построить полигон и гистограмму. Вычислить числовые характеристики сгруппированного ряда.

4. Курс «Социальная психология» прослушало 50 человек. Полученные студентами на экзамене оценки представляют собой следующий набор цифр: 3,4,5,4,3,3,5,…(составить самостоятельно). Построить вариационный статистический ряд. Вычислить моду, медиану, выборочную среднюю.

5. Результаты выборочных наблюдений над непрерывной случайной величиной X приведены ниже в виде интервалов одинаковой длины и соответствующих им частот:

$x_i - x_{i+1}$	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	12	18	45	15	10

Построить: а) гистограмму частот и б) гистограмму относительных частот случайной величины X.

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ ЗАНЯТИЯМ

Лабораторные работы предназначены для получения практических навыков студентами при изучении дисциплины. Предлагаемые задания охватывают основные вопросы математической статистики, рассматриваемые в рамках данного курса.

К выполнению лабораторной работы следует приступать после ознакомления с теоретической частью соответствующего раздела. Результаты всех лабораторных работ необходимо представить в письменном виде.

Типовые задания

Задание 1. Даны наблюдавшиеся значения некоторой случайной величины:

34	25	29	34	12	28	13	28	28	17
23	31	32	23	16	22	34	22	25	28
24	24	25	28	26	19	29	21	30	18
30	30	21	30	19	20	30	34	20	36
28	36	27	17	27	26	26	19	29	24
37	28	31	25	23	33	35	31	22	30
25	26	22							

Требуется:

1) построить сгруппированный статистический ряд;

2) построить гистограмму и полигон относительных частот;

3) найти выборочные точечные характеристики: среднюю, дисперсию, среднее квадратическое отклонение.

Задание 2. Даны наблюдаемые значения некоторой случайной величины.

34	25	29	34	12	28	13	28	28	17
23	31	32	23	16	22	34	22	25	28
24	24	25	28	26	19	29	21	30	18
30	30	21	30	19	20	30	34	20	36
28	36	27	17	27	26	26	19	29	24
37	28	31	25	23	33	35	31	22	30
25	26	22							

Требуется:

1) найти выборочные точечные характеристики: асимметрию, эксцесс, моду, коэффициент вариации;

2) проверить гипотезу относительно близости распределения к нормальному.

Задание 3. Приводятся результаты измерения некоторой величины, которую будем рассматривать как n реализаций случайной величины X.

20,58	22,80	25,4	26,08	23,25	21,42
23,1	24,09	24,02	26,12	22,84	25,10

В предположении, что X имеет нормальное распределение, требуется:

1) найти точечные несмещенные оценки математического ожидания a и среднего квадратического отклонения σ;

2) найти доверительный интервал, покрывающий математическое ожидание с заданной доверительной вероятностью γ=0,95;

3) найти погрешность, с которой среднее арифметическое оценивает математическое ожидание a случайной величины X, если доверительная вероятность γ=0,99; γ=0,999;

4) найти минимальный объем выборки, чтобы с доверительной вероятностью γ=0,95 можно было утверждать, что, принимая среднее арифметическое за математическое ожидание случайной величины X, допускаем погрешность $\varepsilon = \frac{1}{3}\sigma$.

Задание 4. Признак X распределен в генеральной совокупности нормально с известным средним квадратическим отклонением σ. Данные выборочного наблюдения представлены вариационным рядом.

Левая граница:	14,5	19,5	24,5	29,5	34,5	39,5	44,5	49,5	54,5
Правая граница:	19,5	24,5	29,5	34,5	39,5	44,5	49,5	54,5	59,5
Частота:	2	4	11	12	18	11	10	8	4

σ=10, α=29; β=47

По данным выборки требуется:

1) найти доверительный интервал для математического ожидания генеральной совокупности с надежностью 0,99;

2) составить функцию плотности распределения вероятности для предполагаемого нормального закона распределения;

3) определить надежность интервальной оценки (α; β) математического ожидания;

4) вычислите точечную оценку среднего квадратического отклонения двумя различными способами;

5) считая среднее квадратическое отклонение генеральной совокупности неизвестным, найти для него доверительный интервал с надежностью 0,95;

6) определить каким был объем выборки, если доверительный интервал (0, 2,33S) для среднего квадратического отклонения генеральной совокупности ожидается с надежностью 0,999.

Задание 5. Даны результаты эксперимента

X	7,5	7	8,3	8,4	6,9	7,7	8,1	7,6	7,9	8,2
Y	26	27	22	21	27	25	21	21	23	22

Требуется:

1) в предположении, что между x и y существует линейная зависимость, определить ее эмпирическое уравнение;

2) в предположении, что между x и y существует квадратичная зависимость, определить ее эмпирическое уравнение;

3) найти сумму квадратов отклонений для найденных зависимостей сравнить качество приближений.

Задание 6. Дана выборка

X	51	50	33	40	42	51	52	51	55	36	53
У	70	56	31	48	73	72	40	66	76	34	63

По заданной выборке:
1) найти уравнение прямой линии регрессии У на Х;
2) оценить тесноту линейной связи, вычислив выборочный коэффициент корреляции;
3) проверить гипотезу о значимости коэффициента корреляции при уровне значимости 0,1.

Задание 7. По корреляционной таблице:

У \ *X*	8	9	10	11	12	n_y
10	2	3	5			10
11	2	6	20	7		35
12	1	3	10	9	5	28
13	1	2	5	4	7	19
14			2	3	3	8
n_x	6	14	42	23	15	100

По данным выборки требуется
1) в прямоугольной системе координат построить эмпирическую линию регрессии *У* на *X*;
2) оценить тесноту линейной связи;
3) провести гипотезу о значимости коэффициента корреляции при уровне значимости 0,05;
4) составить линейное уравнение регрессии *У* на *X* и построить линию регрессии в той же системе координат, где построена эмпирическая линия регрессии;
5) используя полученное уравнение регрессии, оценить ожидаемое среднее значение признака *У* при $x = x_0$.

4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Самостоятельная работа студентов предназначена для углубления сформированных знаний, умений, навыков. Самостоятельная работа развивает мышление, позволяет выявить причинно-следственные связи в изученном материале, решить теоретические и практические задачи. Самостоятельная работа студентов проводится с целью: систематизации и закрепления полученных теоретических знаний и практических умений студентов; углубления и расширения теоретических знаний; формирования умений использовать справочную документацию и специальную литературу; развития познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;- развития исследовательских умений. Роль самостоятельной работы возрастает, т.к. перед учебным заведением стоит задача в т. ч. и по формированию у студента потребности к самообразованию и самостоятельной познавательной деятельности

Студентами практикуется два вида самостоятельной работы: аудиторная; внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию. В этом случае студенты обеспечиваются преподавателем необходимой учебной литературой, дидактическим материалом, в т. ч. методическими пособиями и методическими разработками.

Внеаудиторная самостоятельная работа выполняется студентом по заданию преподавателя, но без его непосредственного участия. Видами заданий для внеаудиторной самостоятельной работы могут быть:

- для овладения знаниями: чтение текста (учебника, методической литературы); составления плана текста; графическое изображение структуры текста, выполнение лабораторных работ; конспектирование текста; выписки из текста; работа со словарями и справочниками; учебно-исследовательская работа; использование компьютерной техники, интернета и др.;
- для закрепления систематизации знаний: работа с конспектом лекции (обработки текста); повторная работа над учебным материалом (учебника, первоисточника, дополнительной литературы); составление плана выполнения работы в соответствие с планом, предложенным преподавателем; ответы на контрольные вопросы; тестирование, выполнение упражнений и лабораторных работ;
- для формирования умений: решение задач и упражнений по образцу; решение задач и упражнений; выполнение чертежей, схем.

Основное содержание самостоятельной работы составляет выполнение домашних заданий, подготовку к контрольным работам и зачету.

Прежде чем приступать к выполнению лабораторной работы, необходимо ознакомиться с содержанием теоретических вопросов по представленному списку литературы и по лекциям.

Работа пишется на стандартных листах писчей бумаги. Все листы заполняются только с одной стороны. Оформление лабораторной работы осуществляется в соответствии со стандартом. Каждая работа начинается с титульного листа, который служит обложкой работы. Сверху на нем указывается принадлежность студента к учебному заведению, факультету, специализации или кафедре. В середине листа указывается название изучаемой темы или раздела и название учебного задания, номер варианта. Ниже и справа указывается фамилия и инициалы студента, номер академической группы, фамилия и инициалы преподавателя. Внизу титульного листа отмечают год выполнения работы.

Эта страница служит также для отметок преподавателя о выполнении учебного задания и замечаний по поводу подготовленного студентом отчета.

При оформлении работы необходимо соблюдать нумерацию заданий. Задание переписывается полностью и ниже оформляется решение. Работа должна быть сдана на кафедру к назначенному преподавателем сроку.

Каждый учебный семестр заканчивается аттестационными испытаниями: зачетно - экзаменационной сессией.

Подготовка к экзаменационной сессии и сдача зачетов и экзаменов является ответственейшим периодом в работе студента. Серьезно подготовиться к сессии и успешно сдать все экзамены - долг каждого студента. Рекомендуется так организовать свою учебу, чтобы перед первым днем начала сессии были сданы и защищены все лабораторные работы, сданы все зачеты, выполнены другие работы, предусмотренные графиком учебного процесса.

Основное в подготовке к сессии - это повторение всего материала, курса или предмета, по которому необходимо сдавать зачет или экзамен. Только тот успевает, кто хорошо усвоил учебный материал.

Если студент плохо работал в семестре, пропускал лекции, слушал их невнимательно, не конспектировал, не изучал рекомендованную литературу, то в процессе подготовки к сессии ему придется не повторять уже знакомое, а заново в короткий срок изучать весь материал. А это зачастую, оказывается, невозможно сделать из-за нехватки времени. Для такого студента подготовка к экзаменам будет трудным, а иногда и непосильным делом, а финиш - отчисление из учебного заведения.

СОДЕРЖАНИЕ

Наталья Николаевна Двоерядкина,
доцент кафедры общей математики и информатики АмГУ, кандидат педагогических наук;
Татьяна Евгеньевна Гришкина,
старший преподаватель кафедры общей математики и информатики АмГУ