

Программа учебной дисциплины
«Теория вероятностей и математическая статистика»
2019-2020 учебный год

Утверждена
Академическим советом ООП
Протокол 2019 г.

Автор	Пашкевич Анна Валерьевна, к. соц. н, доцент факультета социальных наук, apashkevich@hse.ru
Число кредитов	5
Контактная работа (час.)	94
Самостоятельная работа (час.)	96
Курс	2
Формат изучения дисциплины	Без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Теория вероятностей и математическая статистика» являются изучение и практическое освоение ключевых понятий и подходов теории вероятностей, математической статистики и базовых моделей статистического анализа данных в социальных науках, приобретение концептуального понимания специфики работы с количественными данными в социальных науках, понимания типов задач, которые могут быть решены с помощью математико-статистических методов.

В результате освоения дисциплины студент должен:

- **знать** основные понятия теории вероятностей, математической статистики, методы статистического анализа данных в пределах программы курса;
- **уметь** ставить и понимать социологические задачи, которые могут быть решены с помощью методов теории вероятностей и математической статистики; понимать специфику данных, используемых в математико-статистическом анализе.
- **владеть** навыками самостоятельного статистического анализа данных как на бумаге, так и на компьютере (в программе SPSS).

Изучение дисциплины «Теория вероятностей и математическая статистика» базируется на следующих дисциплинах:

- математика в объеме средней школы;
- линейная алгебра и анализ в объеме вуза;

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- знать стандартные методы решения математических (арифметических) задач;
- знать основные понятия и инструменты классической алгебры;
- обладать элементарными навыками в области решения комбинаторных задач.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- 1 Анализ социологических данных;
- 2 Методология и методы социологического исследования;
- 3 Анализ данных в социальных науках: методы факторизации и кластеризации.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Предмет теории вероятностей. Случайный эксперимент. Выбор из конечной совокупности. Пространство элементарных исходов (элементарных событий).

Классическое и статистическое определение понятия вероятности. Случайность как предмет изучения. Предмет теории вероятностей. Случайный эксперимент. Пространство элементарных исходов (событий). Различные подходы к изучению вероятности. Классическое и статистическое определение понятия вероятности. Относительная частота и вероятность.

Тема 2. События и операции над ними.

Правила сложения и умножения вероятностей. Условная вероятность. Зависимость/независимость событий. Совместимые и несовместимые события.

Тема 3. Формула полной вероятности и формула Байеса.

Полная группа событий. Формула полной вероятности. Априорные и апостериорные вероятности. Формула Байеса

Тема 4. Случайные величины: дискретные и непрерывные. Характеристики случайной величины: математическое ожидание и дисперсия.

Дискретные и непрерывные случайные величины. Основные социологические шкалы. Математические операции со случайными величинами. Характеристики случайной величины: математическое ожидание и дисперсия.

Тема 5. Испытания Бернулли. Биномиальное распределение.

Вероятность в дискретных пространствах. Последовательность испытаний Бернулли.

Тема 6. Функция и плотность распределения случайной величины.

Функция и плотность распределения. Квантили распределения. Равномерное распределение, его числовые характеристики и квантили.

Тема 7. Нормальное распределение: произвольное и стандартное.

Нормальное распределение: ситуации возникновения, функция и плотность распределения. Произвольное и стандартное нормальное распределение. Работа с таблицами нормального распределения. Квантили распределения.

Тема 8. Теорема Муавра – Лапласа. Центральные предельные теоремы.

Теорема Муавра – Лапласа. Центральные предельные теоремы.

Тема 9. Закон больших чисел.

Закон больших чисел.

Тема 10. Введение в статистический анализ данных. Выборка и генеральная совокупность. Вероятностный отбор. Простая случайная выборка.

Выборка и генеральная совокупность. Вероятностный отбор. Простая случайная выборка (SRS). Примеры исследований, построенных на статистическом анализе данных. Источники данных для статистического анализа: опрос, перепись, агрегированная статистика, большие данные. Основные методы анализа. Обзор содержания курса. Основы работы в SPSS. Архитектура пакета. Синтаксис SPSS.

Тема 11. Данные и переменные. Описательная статистика: меры центральной тенденции и меры разброса для разных типов шкал, статистические графики.

Шкалы измерений (интервальная, порядковая, категориальная). Среднее, медиана, мода. Числовые характеристики распределений (квартили, децили). Дисперсия, стандартное отклонение, квартильный размах, размах вариации, коэффициент вариации, и другие меры оценки разреженности

данных. Распределения и их характеристики, типология распределений. Статистическая визуализация. График плотности распределения. Гистограммы. Коробчатые диаграммы. Диаграммы рассеяния. Столбиковые диаграммы. Линейные диаграммы. Круговые диаграммы.

Тема 12. Точечное и интервальное оценивание. Доверительные интервалы для среднего и процентной доли. Размер выборки.

Два вида оценивания: точечное и интервальное. Понятие стандартной ошибки. Доверительные интервалы для среднего и для процентной доли. Уровень доверительной вероятности, его содержательный смысл. Связь точности исследования и объема выборки исследования.

Тема 13. Статистические гипотезы и их проверка. Основные понятия и алгоритмы. Статистическая значимость.

Введение в проверку статистических гипотез. Базовые понятия: статистическая гипотеза, нулевая гипотеза, альтернативная гипотеза, направленные и ненаправленные альтернативные гипотезы, статистика критерия, фиксированный уровень значимости, минимальный уровень значимости (p -value), критическая область, ошибки 1 и 2 рода, статистический вывод и др. Примеры.

Тема 14. Совместное распределение переменных. Таблицы сопряженности. Критерий Хи-квадрат. Анализ стандартизированных остатков.

Понятие о двумерной частотной таблице и способах отражения в ее виде независимости рассматриваемых признаков. Ожидаемые частоты и наблюдаемые частоты. Логика проверки гипотезы об отсутствии связи между двумя номинальными (или рассматриваемыми как номинальные) признаками на основе критерия Хи-квадрат. Основная модель. Интерпретация остатков. Примеры.

Тема 15. Парные коэффициенты корреляции

Диаграмма рассеяния, визуальное «прочтение» характера связи между переменными. Коэффициента линейной корреляции Пирсона (для переменных, измеренных на уровне шкал высокого порядка), коэффициенты ранговой корреляции Спирмена и Кендалла (для переменных, измеренных на порядковом уровне). Проверка статистических гипотез о корреляционной независимости. Соотнесение с ограничениями социологических шкал.

Тема 16. Проверка гипотезы о равенстве математического ожидания (среднего) определенному числу и t-критерий Стьюдента (для одной выборки). Гипотеза о равенстве процентной доли определенному числу и z-критерий. Гипотеза об отсутствии среднего эффекта воздействия для случая парных (связных) выборок и t-критерий Стьюдента.

Алгоритм (правила) проверки гипотез. Статистический критерий проверки гипотез о равенстве среднего (математического ожидания) и процентной доли определенному числу (константе). Статистический критерий проверки гипотезы об отсутствии среднего эффекта воздействия для случая парных (связных) выборок. Содержательный характер исследовательских задач. Примеры.

Тема 17. Проверка гипотезы о равенстве дисперсий в двух группах и F-критерий Фишера. Гипотеза о равенстве средних в случае двух независимых выборок и t-критерий Стьюдента. Гипотеза о равенстве процентных долей в двух группах и z-критерий.

Алгоритм (правила) проверки гипотез. Статистический критерий проверки гипотезы о равенстве дисперсий, дисперсионное отношение. Статистический критерий проверки гипотезы о равенстве двух средних (математических ожиданий) для случая независимых выборок. Статистический критерий проверки гипотезы о равенстве процентных долей в двух группах. Содержательный характер исследовательских задач. Примеры.

Тема 18. Дисперсионный анализ (ANOVA).

Дисперсионный анализ как проверка статистической гипотезы о равенстве средних в k группах (формулировка гипотезы, используемый критерий, логика проверки). Однофакторный

дисперсионный анализ. Основная модель. Класс решаемых задач. Понятия «межгрупповая дисперсия» и «внутригрупповая дисперсия». Понимание причинно-следственной связи при использовании однофакторного дисперсионного анализа (априорная модель такой связи). Примеры.

III. ОЦЕНИВАНИЕ

Аттестация знаний осуществляется в 10-балльной шкале.

Курс состоит из двух, тесно связанных между собой, частей – теория вероятностей (первый модуль) и математическая статистика (второй модуль). Каждый модуль завершается экзаменом.

Формами итогового контроля выступают: письменный экзамен (проводится в *сессию 1 модуля*) по материалам теории вероятностей и письменный экзамен (проводится в *сессию 2 модуля*) по материалам всего курса целиком, в который включаются задания по теории вероятностей и математической статистике. Оценка за первый модуль участвует в расчете итоговой оценки во втором модуле, с соответствующим весовым коэффициентом.

Оценка выставляется в ведомость в 1-м модуле и 2-м модуле (итоговая).

Оценка в первом (1) модуле курса определяется следующим образом:

$$O_{(1)} = 0,6 * O_{\text{ЭКЗАМЕН (1)}} + 0,3 * O_{\text{КР (1)}} + 0,1 * O_{\text{АКТИВНОСТЬ (1)}}$$

Оценка во втором (2) модуле курса (является итоговой оценкой за весь курс) определяется следующим образом:

$$O_{\text{ИТОГОВАЯ (2)}} = 0,4 * O_{\text{ЭКЗАМЕН (2)}} + 0,3 * O_{\text{КР (2)}} + 0,1 * O_{\text{АКТИВНОСТЬ (2)}} + 0,2 * O_{(1)}$$

Пояснения к элементам контроля знаний:

- ✓ Основным элементом контроля является письменный экзамен.
- ✓ Экзамен не является блокирующим и подлежит передаче, в случае если результирующая оценка ниже удовлетворительного уровня. Первая передача проводится по материалам экзамена как основного элемента контроля в период передач. Вторая передача (комиссия) проводится по материалам экзамена с соблюдением аналогичных правил, тоже письменно.
- ✓ В случае неудовлетворительной результирующей оценки за первый модуль курса предусмотрена передача в установленный период передач. Передача по материалам экзамена за первый модуль предшествует передаче по материалам экзамена за второй модуль, согласно хронологическому порядку.
- ✓ Контрольная работа (КР №1) в первом модуле состоит из двух частей, распределенных во времени, – **контрольная А** и **контрольная В**. Контрольная работа (КР №2) во втором модуле состоит из двух частей, распределенных во времени, – **контрольная С** и **контрольная D**. Академический обобщенный результат за контрольные работы (КР №1 и КР №2) оценивается в 10-балльной целочисленной шкале (с грейдами: 10, 9, 8, 7, ..., 0 баллов) в каждом модуле.
- ✓ Контрольные работы (как элементы текущего контроля) не являются блокирующими и не подлежат передаче. Выполняются в ходе семинарского занятия в заранее установленные даты.
- ✓ Даты написания контрольных работ назначаются преподавателем. Для студентов, пропустивших контрольную работу по причине болезни или другой уважительной причине (подтвержденной соответствующими документами), предусмотрена возможность повторной сдачи. В случае

пропуска контрольной работы по неуважительной причине или отсутствия соответствующих документов повторное проведение контрольной работы не производится.

- ✓ *ОАКТИВНОСТЬ* – это оценка, подразумевающая вовлечённую академическую работу на протяжении всего курса. Фиксируется и включает в себя активную работу на семинарских занятиях (решение задач у доски), комплексное выполнение домашних заданий для самоподготовки и закрепления изученного материала к каждому семинару. Оценивается в 10-балльной целочисленной шкале (с грейдами: 10, 9, 8, 7, ..., 0 баллов) отдельно в первом модуле, и отдельно во втором модуле.
- ✓ Округление применяется к результирующей оценке, рассчитываемой с помощью формул (см. выше). Округление производится в сторону ближайшего целого (арифметический способ). Так, например, оценка 3,50 балла округляется до 4.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

- Оценочные средства для промежуточной аттестации:

Итоговый контроль представляет собой два письменных экзамена – в конце 1 модуля, в конце 2 модуля.

- Оценочные средства для текущего контроля студента:

В 1 и 2 модулях студентам предлагаются контрольные работы по задачам теории вероятностей и математической статистики, которые выполняются письменно или на компьютере в учебной аудитории в фиксированное время. Регламент проведения контрольных работ заранее сообщается студентам.

На протяжении учебного процесса студентам предлагаются дополнительные задания для самоподготовки, выполняемые в виде «*Тренировочных задач*». Они позволяют закрепить материал к каждому изученному разделу программы курса. Эти задания требуют математического решения (расчетов) «вручную» или на компьютере. Тренировочные задачи требуют самостоятельного решения и подразумевают развернутую содержательную интерпретацию результатов анализа, выводов. Выполненные домашние работы регулярно засчитываются преподавателем как компонента по линии «Академическая активность» студента и включаются в оценку.

Оценки за письменные работы выставляются, исходя из следующих критериев:

1. правильность решения задач,
2. полнота решения задач, аргументированность,
3. корректность оформления вывода, содержательная интерпретация к полученным расчетам.

V. РЕСУРСЫ

5.1 Основная литература (Библиотечный фонд НИУ «Высшая школа экономики»)

- *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей: учеб. Пособие для вузов. – М.: МЦНМО, 2009. – 256 с. - ISBN 978-5-940575-40-5. (420 экземпляров).
- *Пашкевич А.В.* Теория вероятностей и математическая статистика для социологов и менеджеров: учебник для вузов / Под ред. А.А. Макарова. – М.: Академия, 2014. – 332 с. – (Сер. «Высшее образование»). (Сер. «Бакалавриат»). - ISBN 978-5-446-80561-7. (59 экземпляров).

- *Макаров А.А., Пашкевич А.В.* Задачник по теории вероятностей для студентов социально-гуманитарных специальностей. М.: МЦНМО, 2015. – 158 с. - ISBN 978-5-443-90668-3. (199 экземпляров).
- *Макаров А.А., Пашкевич А.В., Тамбовцева А.А.* Задачник по математической статистике: для студентов социально-гуманитарных и управленческих специальностей М.: МЦНМО, 2018. – 253 с. - ISBN 978-5-4439-1369-8. (311 экземпляров).

5.2 Дополнительная литература (Библиотечный фонд НИУ «Высшая школа экономики»)

- *Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р. Яценко И.В.* Теория вероятностей и статистика: учеб. пособие для 10 и 11 кл. общеобразоват. учреждений. – М.: МЦНМО, 2014. – 243 с. - ISBN 978-5-443-90724-6. (114 экземпляров).
- *Толстова Ю.Н.* Математическая статистика для социологов: учебник и практикум для академического бакалавриата. – 2-е изд. испр. и доп. – М.: Юрайт, 2017. – 258 с. – (Сер. «Бакалавр». Академический курс). - ISBN 9785534032444. (6 экземпляров).
- *Толстова Ю.Н.* Математико-статистические модели в социологии: математическая статистика для социологов: учеб. пособие. – М.: ГУ-ВШЭ, 2007. – 243 с. – (Сер. «Учебники Высшей школы экономики») (Приоритетные национальные проекты «Образование»). - ISBN 5-7598-0513-7. (304 экземпляра)
- *Толстова Ю.Н., Куликова А.А., Рыжова А.В., Юдин Г.Б.* Математическая статистика для социологов: задачник / Отв. ред. Ю.Н. Толстова. – М.: ГУ-ВШЭ, 2010. – 184 с. - ISBN 978-5-7598-0626-4. (122 экземпляра).
- *Толстова Ю.Н.* Анализ социологических данных: Методология, дескриптивная статистика, изучение связей между номинальными признаками: Учеб. пособие для вузов / М.: Научный мир, 2000. – 352 с. - ISBN 5-89176-086-X. (99 экземпляров).
- *Крыштановский А.О.* Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. М.: ГУ-ВШЭ, 2006. – 281 с. – (Сер. «Учебники Высшей школы экономики»). - ISBN 5-7598-0373-5. (312 экземпляров).
- *Бююль А., Цефель П.* SPSS: искусство обработки информации: анализ стат. данных и восстановление скрытых закономерностей: пер. с нем. М.; СПб.; Киев: Торг.-изд. дом «DiaSoft», 2002. – 601 с. - НП. - ISBN 5-937720-14-8. (116 экземпляров).
- *Наследов А.* SPSS 19: профессиональный статистический анализ данных / СПб.: Питер, 2011. – 399 с. - ISBN 978-5-459-00344-4. (60 экземпляров).
- *Наследов А.Д.* Математические методы психологического исследования: анализ и интерпретация данных: учеб. пособие для вузов / 3-е изд., стер. – СПб.: Речь, 2007. – 390 с. - ISBN 5-926802-75-7. (74 экземпляра)
- *Миркин Б.Г.* Введение в анализ данных: учебник и практикум для вузов / М.: Юрайт, 2015. – 174 с. – (Сер. «Авторский учебник»). - ISBN 978-5-9916500-9-0. (30 экземпляров).
- *Моосмюллер Г., Ребик Н.Н.* Маркетинговые исследования с SPSS: учеб. пособие для вузов / 2-е изд. – М.: ИНФРА-М, 2011. – 199 с. – (Сер. «Высшее образование»). - ISBN 978-5-16-004240-4. (7 и 3 экземпляров).
- *Мостеллер Ф., Рурке Р., Томас Дж.* Вероятность. / Пер. с англ. В.В. Фирсова; Под ред. И.М. Яглома. – М.: Мир, 1969. – 431 с. – (Попул. сер. «Современная математика»). (2 экземпляра).
- *Agresti A., Finlay B.* Statistical methods for the social sciences / 4th ed. – Upper Saddle River: Pearson Education International, 2009. – 609 с. – На англ. яз. - ISBN 978-0-13-713150-1. (10 экземпляров).
- *Bluman A.G.* Elementary statistics: a step by step approach / 2nd ed. – Dubuque: Wm. C. Brown Publishers, 1995. – 684 с. – На англ. яз. - ISBN 0-697-31478-2. (136 экземпляров).

5.3 Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 7 Professional RUS Microsoft Windows 10 Microsoft Windows 8.1 Professional RUS	<i>Из внутренней сети университета (договор)</i>
2.	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>
3	IBM SPSS	<i>Из внутренней сети университета (договор, лицензионное соглашение)</i>

Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа
	<i>Профессиональные базы данных, информационно-справочные системы</i>	
1.	Информационный портал Федеральной службы государственной статистики	URL: https://gks.ru/
2.	Единый архив экономических и социологических данных (ЕАЭСД)	URL: https://sophist.hse.ru/
3.	Информационный портал Фонда «Общественное мнение»	URL: https://fom.ru/
4.	Информационный портал Всероссийского центра изучения общественного мнения	URL: https://wciom.ru/
	<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>	
5.	Открытое образование	URL: https://openedu.ru/

5.4 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и презентацию тематических иллюстраций, соответствующих программе дисциплины в составе:

– ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы, статистический пакет для обработки и анализа социологической информации IBM SPSS);

– мультимедийный проектор с дистанционным управлением.

Учебные аудитории для практических занятий по дисциплине оснащены компьютерами с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ, с установленным программным обеспечением.