

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Волков В.В.
Должность: Ректор
Дата подписания: 11.09.2024 15:38:56
Уникальный программный ключ:
ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»

УТВЕРЖДАЮ

Ректор  В.В. Волков

« 28 » августа 2024 г.

Протокол Ученого Совета

№ 7 от 28 августа 2024 г.



Рабочая программа дисциплины
«Статистика»

дополнительная профессиональная программа
«Аналитика данных в социогуманитарных науках»

вид программы
программа повышения квалификации

язык обучения – русский
форма обучения – очная

Санкт-Петербург

Авторы:

Тушканова О.Н., к.т.н., доцент факультета социологии АНООВО «ЕУСПб»

Рабочая программа дисциплины «Статистика», входящая в состав дополнительной профессиональной программы повышения квалификации «Аналитика данных в социогуманитарных науках» утверждена на заседании Ученого совета.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ.....	4
2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ.....	4
3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ.....	5
4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ	8
5. ОЦЕНОЧНЫЕ СРЕДСТВА.....	9
6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	11
7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ.....	12

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ

Цель обучения: освоения дисциплины «Статистика» — изучение ключевых методологий в области теории вероятностей и математической статистики.

Задачи обучения:

- формирование навыков практического применения освоенных методов;
- изучение таких категорий как: переменная, распределение, статистическая значимость, p-value, гипотеза, регрессия;
- приобретение навыков формулирования научных задач и гипотез в количественной парадигме.

Изучение данной дисциплины способствует формированию профессиональных навыков по разработке теоретических моделей исследуемых процессов, явлений и объектов, относящихся к сфере профессиональной деятельности, их исследованию, оценке и интерпретации полученных результатов.

Отличительной особенностью реализуемого подхода к преподаванию дисциплины является разнообразных практических иллюстраций основных теоретических положений.

Общая трудоемкость дисциплины составляет 46 часов.

2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ

В результате освоения дисциплины слушатель должен приобрести следующие знания и умения, необходимые для качественного изменения профессиональных компетенций:

слушатель должен знать:

- особенности применения различных теоретико-методологических концепций с использованием технологий прикладного анализа данных;
- общие правила ведения научных дискуссий.

слушатель должен уметь:

- использовать основы статистики для реализации задач профессиональной деятельности;
- использовать полученные знания и умения в профессиональной деятельности;
- правильно оформлять и представлять результаты исследований.

слушатель должен владеть:

- практического применения освоенных методов;
- формулирования научных задач и гипотез в количественной парадигме;
- способностью к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности.

В результате изучения дисциплины «Статистика» слушатель приобретает следующие профессиональные компетенции (Таблица 1):

Планируемые результаты обучения по дисциплине

Таблица 1

Код и название компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапам формирования компетенций
ПК-3	способен применить анализ данных к научным и общественным задачам	Знать: З (ПК-3) – основы анализа данных
		Владеть: В (ПК-3) – навыками анализа данных с учетом специфики научных и общественных задач

3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) в соответствии с таблицей 2.

Таблица 2

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Коды ЗУВ (в соответствии с табл. 1)
1	Нормальное распределение	<ul style="list-style-type: none"> – статистический бум конца XIX века и полидетерминистическая интерпретация нормального распределения; – понятие семейства распределений и нормировка: формула приведения любой нормально распределенной случайной величины к стандартному нормальному распределению; – задачи на вероятности интервалов для нормально распределенных величин с использованием его симметричности; – типы выборок, концепция статистического вывода о генеральной совокупности по выборке; Центральная предельная теорема в формулировке распределения выборочного среднего для случайных выборок. 	ПК-3	З (ПК-3) В (ПК-3)
2	Тестирование гипотез	<ul style="list-style-type: none"> – логика статистического теста, формулировка нулевой и альтернативной гипотез; – уровень статистической значимости; – ошибки 1 и 2 рода; – мощность теста; – двусторонние и односторонние тесты для дискретной биномиально распределенной случайной величины; – двусторонние и односторонние тесты для непрерывной нормально распределенной случайной величины; – стандартная ошибка среднего и её зависимость от размера выборки; – распределение Стьюдента и учёт размера выборки при проведении t-теста; проверка гипотез для пропорций. 	ПК-3	З (ПК-3) В (ПК-3)
3	Доверительные интервалы	<ul style="list-style-type: none"> – построение доверительного интервала для среднего; – доверительная вероятность: почему степень уверенности нельзя интерпретировать как вероятность; расчёт размера выборки, необходимого для заданной точности оценки. 	ПК-3	З (ПК-3) В (ПК-3)
4	Зависимые случайные величины	<ul style="list-style-type: none"> – парные переменные и их отображение на диаграмме рассеяния; – понятие корреляции и её графическая репрезентация, зависимые и независимые переменные; – расчёт ковариации как меры линейной зависимости случайных величин; 	ПК-3	З (ПК-3) В (ПК-3)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Коды ЗУВ (в соответствии с табл. 1)
		<p>– нормированная ковариация — коэффициент корреляции Пирсона;</p> <p>– ограничения применимости коэффициента корреляции Пирсона (формальные ограничения метода, независимость от наклона прямой, нечувствительность к кластерам, спонтанные корреляции, квартет Энскомба)</p> <p>– линейная регрессия для двух переменных: расчёт уравнения, интерполяция и экстраполяция;</p> <p>– многомерная линейная регрессия; логистическая регрессия: сведение к линейной через логарифм отношения шансов, почему используется сигмоид для бинарной объясняемой переменной.</p>		

Структура дисциплины

Таблица 3.

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.				Форма текущего контроля успеваемости, промежуточной аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР ¹	
				Л	СЗ ²		
1.	Нормальное распределение	<p>– статистический бум конца XIX века и полидетерминистическая интерпретация нормального распределения;</p> <p>– понятие семейства распределений и нормировка: формула приведения любой нормально распределенной случайной величины к стандартному нормальному распределению;</p> <p>– задачи на вероятности интервалов для нормально распределенных величин с использованием его симметричности;</p> <p>– типы выборок, концепция статистического вывода о генеральной совокупности по выборке;</p> <p>Центральная предельная теорема в формулировке распределения выборочного среднего для случайных выборок.</p>	11	4	4	3	
2.	Тестирование гипотез	– логика статистического теста, формулировка нулевой и	12	4	4	4	

¹ Самостоятельная работа, включает в себя часы на промежуточный контроль

² Могут включать в себя: лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.				Форма текущего контроля успеваемости, промежуточной аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР ¹	
				Л	СЗ ²		
		альтернативной гипотез; – уровень статистической значимости; – ошибки 1 и 2 рода; – мощность теста; – двусторонние и односторонние тесты для дискретной биномиально распределенной случайной величины; – двусторонние и односторонние тесты для непрерывной нормально распределенной случайной величины; – стандартная ошибка среднего и её зависимость от размера выборки; – распределение Стьюдента и учёт размера выборки при проведении t-теста; проверка гипотез для пропорций.					
3.	Доверительные интервалы	– построение доверительного интервала для среднего; – доверительная вероятность: почему степень уверенности нельзя интерпретировать как вероятность; расчёт размера выборки, необходимого для заданной точности оценки.	10	4	4	2	
4.	Зависимые случайные величины	– парные переменные и их отображение на диаграмме рассеяния; – понятие корреляции и её графическая репрезентация, зависимые и независимые переменные; – расчёт ковариации как меры линейной зависимости случайных величин; – нормированная ковариация – коэффициент корреляции Пирсона; – ограничения применимости коэффициента корреляции Пирсона (формальные ограничения метода, независимость от наклона прямой, нечувствительность к кластерам, спонтанные корреляции, квартет Энскомба) – линейная регрессия для двух переменных: расчёт уравнения, интерполяция и экстраполяция; – многомерная линейная	11	4	4	3	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.			Форма текущего контроля успеваемости, промежуточной аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР ¹
				Л	СЗ ²		
		регрессия; логистическая регрессия: сведение к линейной через логарифм отношения шансов, почему используется сигмоид для бинарной объясняемой переменной.					
5.	Промежуточная аттестация	Опрос по вопросам (устно)	2	-	-	2	
Всего:			46	16	16	14	

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Общие положения.

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения исследовательской литературы (из списков основной, дополнительной), статей по проблематике занятия и их анализа.

Самостоятельная работа обучающегося представляет самостоятельное изучение дополнительных материалов, Интернет-ресурсов и пр. Подготовка к семинарским занятиям, создание докладов, проектов и презентаций также является важной формой работы обучающихся. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя. Вопросы и замечания, возникшие в ходе самостоятельного внеаудиторного чтения рекомендованной литературы, обсуждаются с преподавателем и другими обучающимися. Выносятся на обсуждение, как правило, актуальные проблемы и предлагается их рассмотреть с точки зрения того или иного теоретического подхода.

На занятиях материал излагается в проблемной форме. Основной упор в преподавании делается на изучение теоретических понятий и возможности их применения на конкретных примерах, в том числе в устных выступлениях обучающихся.

Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся.

- Объясните связь между распределением вероятностей случайной величины и кумулятивной функцией вероятностей для этой случайной величины.
- При каких условиях применимо моделирование результатов серии экспериментов биномиально распределенной случайной величиной?
- Каким образом формула числа сочетаний применяется при возведении суммы двух величин в целую положительную степень?
- Какими параметрами задаётся нормально распределённая случайная величина?
- Как выбранный уровень статистической значимости при проведении статистического теста влияет на вероятность совершения ошибки первого рода?
- Как выбор одностороннего или двустороннего теста влияет на удалённость границы критического интервала от предполагаемого в нулевой гипотезе значения?
- Почему при построении доверительного интервала для среднего используется доверительная вероятность, а не вероятность попадания величины в интервал?
- Каковы условия корректного применения коэффициента корреляции Пирсона для получения адекватной интерпретации связи между двумя переменными?

Источники для самостоятельной подготовки:

- Введение в статистическое обучение с примерами на языке R / Е. Джеймс, Д. Уитгон, Т. Хасты, Р.В. Тибширани ; пер. с англ. С.Э. Мاستицкого. - 2-е изд., испр. - Москва : ДМК Пресс, 2017. - 456 с. - ISBN 978-5-97060-495-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1027867>
- Социально-экономическая статистика: Учебное пособие [Электронный ресурс] / Я.С. Мелкумов. - М.: НИЦ Инфра-М, 2013. - 236 с. - URL: <http://znanium.com/bookread2.php?book=366756>
- Общая теория статистики: Учебное пособие / С.Н. Лысенко, И.А. Дмитриева. - Изд., испр. и доп. - М.: Вузовский учебник: НИЦ ИНФРА-М, 2014. - 219 с. - [Электронный ресурс]. URL: <http://znanium.com/bookread2.php?book=397795>
- Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python : [16+] / В.М. Волкова, М.А. Семенова, Е.С. Четвертакова, С.С. Вожов ; Новосибирский государственный технический университет. – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. : ил., табл. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=576496>

5. ОЦЕНОЧНЫЕ СРЕДСТВА

Проведение текущего контроля в рамках реализации данной дисциплины проходит в соответствии с Таблицей 3 данной рабочей программы дисциплины по основным понятиям (категориям) и проблемам, рассматриваемым в предложенных темах. Фиксация результатов текущего контроля в рамках реализации данной дисциплины не предусмотрена.

Типовые задания к текущей аттестации.

Тема 1. Нормальное распределение

Домашнее задание 1.

- рассчитайте вероятность того, что случайная величина со стандартным нормальным распределением примет значение а) меньше заданного, б) больше заданного, в) в заданном интервале;
- рассчитайте вероятность того, что случайная величина с нормальным распределением и заданными средним и дисперсией примет значение а) меньше заданного, б) больше заданного, в) в заданном интервале;
- по заданной вероятности попадания нормально распределённой случайной величины в заданный интервал и известному среднему найдите дисперсию этой величины.

Тема 2. Тестирование гипотез

Домашнее задание 2.

- рассчитайте стандартную ошибку оценки истинного среднего значения генеральной совокупности по заданной выборке;
- проверьте гипотезу о равенстве среднего значения заданному числу по заданной выборке и уровню значимости;
- при проведении статистического теста на равенство среднего значения заданному числу по заданной выборке получите и проинтерпретируйте p-value.

Тема 3. Доверительные интервалы:

Домашнее задание 3.

- постройте доверительный интервал для среднего по заданной выборке и доверительной вероятности;
- для заданной ширины доверительного интервала и доверительной вероятности рассчитайте необходимый размер выборки для его построения.

Тема 4. Зависимые случайные величины:

Домашнее задание 4.

- рассчитайте и проинтерпретируйте коэффициент корреляции Пирсона для заданного парного датасета;
- постройте уравнение регрессии по заданному парному датасету и рассчитайте оценённое значение зависимой переменной по заданному значению независимой.

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Домашнее задание	слушатель выполняет задание частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение задания в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

Форма промежуточной аттестации – зачет, выставляемый на основе устного опроса.

При аттестации используются система «зачтено» и «не зачтено» в соответствии с критериями оценивания.

В результате промежуточного контроля знаний обучающиеся получают аттестацию по дисциплине.

Показатели, критерии и оценивание компетенций по уровням их формирования в процессе промежуточной аттестации

Таблица 4

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Коды ЗУВ (в соответствии с Таблицей 1)	Критерии оценивания	Оценка
зачет / устный опрос	ПК-3	З (ПК-3) В (ПК-3)	– слушатель в ответах демонстрирует знание всех теоретических положений, (развернуто) отвечает на все поставленные вопросы, предлагает обоснования при ответе на все или большинство поставленных вопросов; несущественные ошибки не снижают качество ответа	зачтено
			– ответ отсутствует или является односложным, или содержит существенные ошибки	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций.

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине.

Типовые задания к промежуточной аттестации.

- биномиальное распределение: расчет коэффициентов, интерпретация отдельных вероятностей и интервалов;
- некоторые часто употребляемые распределения с примерами;
- понятие семейства распределений и нормировка: формула приведения любой нормально распределенной случайной величины к стандартному нормальному

- распределению;
- задачи на вероятности интервалов для нормально распределенных величин с использованием его симметричности;
- типы выборок, концепция статистического вывода о генеральной совокупности по выборке;
- Центральная предельная теорема в формулировке распределения выборочного среднего для случайных выборок;
- логика статистического теста, формулировка нулевой и альтернативной гипотез;
- уровень статистической значимости;
- ошибки 1 и 2 рода;
- мощность теста;
- двусторонние и односторонние тесты для дискретной биномиально распределенной случайной величины;
- двусторонние и односторонние тесты для непрерывной нормально распределенной случайной величины;
- стандартная ошибка среднего и её зависимость от размера выборки;
- распределение Стьюдента и учёт размера выборки при проведении t-теста;
- проверка гипотез для пропорций;
- построение доверительного интервала для среднего;
- доверительная вероятность: почему степень уверенности нельзя интерпретировать как вероятность;
- расчёт размера выборки, необходимого для заданной точности оценки;
- парные переменные и их отображение на диаграмме рассеяния;
- понятие корреляции и её графическая репрезентация, зависимые и независимые переменные;
- расчёт ковариации как меры линейной зависимости случайных величин;
- нормированная ковариация — коэффициент корреляции Пирсона;
- ограничения применимости коэффициента корреляции Пирсона (формальные ограничения метода, независимость от наклона прямой, нечувствительность к кластерам, спонтанные корреляции, квартет Энскомба);
- линейная регрессия для двух переменных: расчёт уравнения, интерполяция и экстраполяция;
- многомерная линейная регрессия.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

- Введение в статистическое обучение с примерами на языке R: Практическое пособие / Д.Гарет и др.-М.:ДМК Пресс,2017.-456 с.(П)
- Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход/ЛемешкоБ.Ю., ЛемешкоС.Б., ПостоваловС.Н. и др. - Новосибирск : НГТУ, 2011. - 888 с.: ISBN 978-5-7782-1590-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/548140> (дата обращения: 06.07.2021). – Режим доступа: по подписке.

Дополнительная литература:

- Келлехер, Д. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни ; пер. с англ.. - Москва : Альпина Паблишер, 2020. - 222 с. - ISBN 978-5-9614-3170-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1221800> (дата обращения: 02.07.2021). – Режим доступа: по подписке.
- Статистический анализ данных, моделирование и исследование вероятностных

закономерностей. Компьютерный подход/ЛемешкоБ.Ю., ЛемешкоС.Б., ПостоваловС.Н. и др. - Новосибирск : НГТУ, 2011. - 888 с.: ISBN 978-5-7782-1590-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/548140> (дата обращения: 06.07.2021). – Режим доступа: по подписке.

Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения программы:

Информационно-справочные системы:

- Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
- Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
- Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
- Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
- Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
- Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
- Российское образование. Федеральный портал: <http://www.edu.ru>

Тематические системы:

- Google. Книги: <https://books.google.com>
- Internet Archive: <https://archive.org>
- Koob.ru. Электронная библиотека «Куб»: <http://www.koob.ru/philosophy/>
- Библиотека Ихтика [ihtik.lib.ru]: <http://ihtik.lib.ru/>
- Докусфера — Российская национальная библиотека: <http://leb.nlr.ru>
- ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>
- Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>
- Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
- Неприкосновенный запас: <http://magazines.russ.ru/nz/>
- Президентская библиотека: <http://www.prlib.ru>
- Российская государственная библиотека: <http://www.rsl.ru/>
- Российская национальная библиотека: <http://www.nlr.ru/poisk/>

7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ

В ходе реализации образовательного процесса используются многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа и семинарского типа обеспечивается демонстрационным оборудованием.

Для обучающихся с ограниченными возможностями здоровья и инвалидов (в случае необходимости) могут быть созданы специальные условия для получения образования.

Программное обеспечение

При осуществлении образовательного процесса в рамках Университета слушателям рекомендовано использовать следующее лицензионное программное обеспечение:

- Adobe Acrobat Professional 11.0 MLP AOO License RU
- Adobe CS5.5 Design Standart Win IE EDU CLP

- ABBYY FineReader 11 Corporate Edition
- ABBYY Lingvo x5
- Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
- Adobe Acrobat Reader DC /Pro – бесплатно
- Google Chrome – бесплатно
- Opera – бесплатно
- Mozilla – бесплатно
- VLC – бесплатно
- R – бесплатно
- Яндекс.Браузер – бесплатно