

Программа учебной дисциплины «Статистический анализ данных (SPSS)»

Утверждена
Академическим руководителем ООП
А.А. Бляхман
«30» января 2018 г.

Автор	Д.ф.-м.н., профессор О.Н. Савина
Число кредитов	5
Контактная работа (час.)	24
Самостоятельная работа (час.)	166
Курс	4
Формат изучения дисциплины	Без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины "Статистический анализ данных (SPSS)" является овладение студентами с основными методами прикладной статистики и эконометрики, приобретение навыков работы с данными в пакета SPSS и умение применять их при анализе и интерпретации данных отечественной и зарубежной статистики о социально-экономических процессах и явлениях.

В результате освоения дисциплины студент должен:

знать:

- основные возможности и функции пакета SPSS, необходимые для проведения статистического анализа данных.
- основные понятия и теоретические положения, связанные с анализом статистических данных.
- основные методы и подходы статистического анализа данных для количественного статистического описания эксперимента

уметь:

- применять пакет SPSS при решении классических задач математической статистики
- применять изученные методы статистического анализа данных для количественного статистического описания эксперимента
- оценить границы применимости и достоверность выводов проведенного анализа
- анализировать степень зависимости между переменными, определять ее тип
- проводить классификацию многомерных наблюдений

владеть:

- навыками обработки данных и их интерпретации при проведении научных исследований (экспериментов, наблюдений и количественных измерений), связанных с объектами профессиональной деятельности.
- методами статистической оценки информации в социальных науках, бизнесе и управлении
- навыками использования среды SPSS для получения простейших прогнозов в экономических ситуациях при принятии решений в управлении экономикой

Изучение дисциплины «Статистический анализ данных (SPSS)» базируется на следующих дисциплинах:

- Математика (математический анализ и линейная алгебра)
- Теория вероятности и математическая статистика

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

- навыками работы с персональным компьютером на уровне пользователя;
- уметь применять знания по математическому анализу и линейной алгебре к решению конкретных задач;
- владеть навыками решения задач теории вероятности и математической статистики.

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

- Стратегический менеджмент
- Экономическая статистика
- Методы научных исследований в менеджменте

Основные положения дисциплины могут быть использованы в проектной и исследовательской работе студента, а также в будущей научно-исследовательской и педагогической деятельности выпускника.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Элементы описательной статистики (SPSS)

Общее представление о среде SPSS как универсальной системе статистического анализа данных. Основные правила выполнения операций в среде SPSS: выбор процедуры анализа; определение параметров в диалоговых окнах; исполнение, просмотр и редактирование данных, представление результатов. Ознакомление со справочной системой SPSS.

Порядковые, номинальные и интервальные данные. Вычисление статистических характеристик: выборочное среднее, мода, медиана и квартили. Показатели изменчивости данных: дисперсия, стандартное отклонение, вариация, межквартильный размах. Графическое отображение данных (гистограммы, графики, диаграммы). Графическое представление и вычисление статистических характеристик в среде SPSS.

Тема 2. Инструменты SPSS для решения задач теории вероятностей математической статистики

Дискретные и непрерывные случайные величины. Законы распределения дискретных и непрерывных случайных величин. Предельные распределения для биномиального распределения. Нормальное распределение. Стандартизация нормального распределения. Распределение χ^2 с n степенями свободы. Функции распределения многомерных случайных величин. Числовые характеристики случайных величин. Закон больших чисел. Распределения случайных величин в библиотеке SPSS. Решение задач теории вероятности в среде SPSS.

Числовые характеристики генеральной совокупности и выборки. Доверительные интервалы. Доверительные интервалы для математического ожидания и дисперсии. Простая случайная выборка. Таблица случайных чисел. Стратифицированная, систематическая и кластерная выборки. Точечные и интервальные оценки параметров распределений. Ошибка оценки. Стандартная ошибка среднего и доли признака.

Тема 3. Доверительные интервалы. Проверка статистических гипотез (SPSS)

Доверительные интервалы. Доверительные интервалы для математического ожидания и дисперсии. Доверительный интервал для среднего генеральной совокупности, имеющей нормальный закон распределения (случаи известной и неизвестной дисперсии).

Статистическая гипотеза. Этапы проверки статистических гипотез. Проверяемая и альтернативная гипотезы. Уровень значимости, критическая область. Проверка гипотезы о среднем значении математического ожидания и дисперсии нормально распределенной случайной величины. Проверка гипотез о функциях распределения (критерий Пирсона и критерий Колмогорова-Смирнова). Построение и анализ таблиц сопряженности признаков. Графические методы проверки гипотез о распределении с помощью Q-Q и P-P процедур. Формат отчета о проверке гипотез в SPSS.

Тема 4. Элементы корреляционного и регрессионного анализа (SPSS)

Понятие независимой и зависимой переменной. Виды зависимости. Диаграмма рассеяния. Коэффициент корреляции Пирсона. Ранговая корреляция (по Спирмену и Кендаллу). Критерий для проверки значимости выборочного коэффициента корреляции. Линейная регрессия, коэффициенты регрессии, их интерпретация. Коэффициент детерминации. Доверительный интервал коэффициентов регрессии. Статистические критерии двумерных случайных величин Гипотеза об отсутствии корреляционной зависимости. t-критерий, F-критерий, критерий Уилкоксона

Введение в дисперсионный анализ (ANOVA). Критерий Фишера. Однофакторный дисперсионный анализ. Проверка гипотезы о влиянии фактора на независимую переменную с использованием F-критерия. Интерпретация результатов однофакторного дисперсионного анализа. Элементы двухфакторного дисперсионного анализа.

Тема 5. Понятие факторного и кластерного анализа (SPSS)

Назначение факторного и кластерного анализа. Корреляционная матрица, Определение факторов. Кластеры, методы кластеризации, интерпретация и профилирование кластеров. Примеры проведения факторного и кластерного анализа данных в среде SPSS.

При реализации учебной работы предполагается разбор теоретических вопросов и практических задач в рамках теоретических и практических занятий.

III. ОЦЕНИВАНИЕ

Формы контроля знаний студентов

Тип контроля	Форма контроля	Параметры
Текущий контроль	Контроль активности студентов на занятиях	Устный опрос, письменный опрос 2-10 минут
Текущий контроль	Проверочная работа	Письменная работа 80 минут
		Самостоятельная работа в компьютерном классе 80 минут
Итоговый	Экзамен	Письменная работа 80 минут

По формам итогового контроля при выставлении оценок учитывается способность студента распознавать тип поставленной задачи, обосновывать применимость метода решения, применить необходимый метод, интерпретировать полученный результат, оценить влияние внешних воздействий на полученное решение поставленной задачи.

Оценки по всем формам текущего и итогового контроля выставляются по 10-ти балльной шкале:

высшая оценка в 9 баллов (10 баллов проставляется в исключительных случаях)

проставляются при отличном выполнении заданий: полных (с детальными или многочисленными примерами и возможными обобщениями) ответах на вопросы, правильном решении задачи и четком и исчерпывающем ее представлении,

почти отличная оценка в 8 баллов проставляется при полностью правильных ответах и решении задач, но при отсутствии какого-либо из выше перечисленных отличительных признаков, как, например: детальных примеров или обобщений, четкого и исчерпывающего представления решаемой задачи,

оценка в 7 баллов проставляется при правильных ответах на вопросы и правильном решении задачи, но при отсутствии пояснений, примеров, обобщений, без представления алгоритма или последовательности решения задач,

оценка в 6 баллов проставляется при наличии отдельных неточностей в ответах на вопросы (включая грамматические ошибки) или неточностях в решении задачи не принципиального характера (описки и случайные ошибки арифметического характера),

оценка в 5 баллов проставляется в случаях, когда в ответах и в решении задач имеются неточности и ошибки, свидетельствующие о недостаточном понимании вопросов и требующие дополнительного обращения к тематическим материалам,

оценка в 4 балла проставляется при наличии серьезных ошибок и пробелов в знании по контролируемой тематике,

оценка в 3 балла проставляется при наличии лишь отдельных положительных моментов в ответах на вопросы и в решении задач, говорящих о потенциальной возможности в последующем более успешно выполнить задания; оценка в 3 балла, как правило, ведет к повторному написанию ответов на вопросы или решению дополнительной задачи,

оценка в 2 балла проставляется при полном отсутствии положительных моментов в ответах на вопросы и решении задач и, как правило, ведет к повторному написанию контрольной работы в целом,

оценка в 1 балл проставляется, когда неправильные ответы и решения, кроме того, сопровождаются какими-либо демонстративными проявлениями безграмотности или неэтичного отношения к изучаемой теме.

Критерии оценки знаний, навыков

Текущий контроль: Контроль активности студента на занятии, проверочная работа, самостоятельная работа

Под Контролем активности студента на занятии подразумевается а) регулярный, как правило - письменный, реже – устный опрос студентов по формулировкам основных понятий, теорем, методов, по решению типовых задач. Длительность опроса – 2 – 10 минут. б) Самостоятельное решение и объяснение задач текущей темы у доски студентом перед своей группой.

Под проверочной работой подразумевается письменное решение набора задач текущей темы. Возможно – с ответами на вопросы теории по текущей теме. Длительность контрольной работы не более 80 минут, Преподаватель может сократить длительность выполнения контрольной.

Под самостоятельной работой подразумевается работа на компьютере в компьютерном классе по решению набора задач текущей темы в программе SPSS. Длительность самостоятельной работы не более 80 минут, Преподаватель может сократить длительность выполнения работы.

При проведении проверочной/самостоятельной студенты могут по решению преподавателя использовать только учебные программы, справочники, таблицы и прочие источники информации, перечень которых установлен преподавателем. Использование

материалов, не предусмотренных этим перечнем, попытка общения с другими студентами или иными лицами, наличие электронных средств связи/гаджетов, несанкционированные перемещения студентов, и т.п. являются основанием для выставления соответствующему студенту оценки 0 баллов и удаления его из аудитории. Студент обязан по окончании времени выполнения работы сдать письменную работу преподавателю независимо от степени ее готовности. Если студент не сдал вовремя работу, ему за эту работу выставляется оценка 0 баллов. При обнаружении подложных работ, плагиата в работе, идентичных (списанных), и т.п. работ – всем связанным с этими работами студентам ставится оценка 0.

Преподаватель оценивает как выполнение студентом задания экзамена ($O_{\text{ЭКЗАМЕН}}$), так и активность студента на практических занятиях ($O_{\text{ауд}}$). Оценка $O_{\text{ауд}}$ учитывает факт выступления студента и качество выступления, результаты проверочных опросов. Накопленная оценка рассчитывается по формуле:

$$O_{\text{накопленная}} = 0,2 * O_{\text{п/р}} + 0,6 * O_{\text{с/р}} + 0,2 * O_{\text{ауд}}$$

Здесь $O_{\text{п/р}}$ - оценка проверочной работы, $O_{\text{с/р}}$ - средняя оценка самостоятельной работы. Результирующая оценка по дисциплине рассчитывается следующим образом:

$$O_{\text{результат}} = 0,3 * O_{\text{ЭКЗАМЕН}} + 0,7 * O_{\text{накопленная}}$$

Способ округления оценок – арифметический.

В зависимости от уровня подготовленности потока студентов коэффициенты в указанных формулах могут быть изменены.

Полученный после округления этой величины до целого значения результат и **выставляется** как **результирующая оценка по 10-балльной шкале** по учебной дисциплине «Статистический анализ данных. SPSS» в экзаменационную ведомость. В экзаменационную ведомость выставляется также и оценка по данной дисциплине **по 5-и балльной системе**, получаемая из оценки по десятибалльной шкале в соответствии со следующей таблицей соответствия

Таблица соответствия оценок по десятибалльной и пятибалльной системам.

По десятибалльной шкале	По пятибалльной шкале
1 – неудовлетворительно 2 – очень плохо 3 – плохо	Неудовлетворительно - 2
4 – удовлетворительно 5 – весьма удовлетворительно	удовлетворительно -3
6 – хорошо 7– очень хорошо	хорошо - 4
8 – почти отлично 9 – отлично 10 – блестяще	отлично - 5

При результирующей оценке менее 4-х баллов (по 10 – ти балльной шкале) студент имеет право на одну пересдачу и на одну пересдачу с комиссией. При ранее полученной результирующей оценке 4 и более баллов пересдачи не допускаются.

Результирующая оценка по учебной дисциплине ставится в диплом

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Перечень вопросов к экзамену по всему курсу полностью определяется приведенным выше содержанием курса.

Оценочные средства для текущего контроля студента

Тематика заданий текущего контроля

Примерные задания для проверочных работ:

1. Использование правил выполнения операций в среде SPSS.
2. Нахождение числовых характеристик случайных величин.
3. Анализ выборки. Построение гистограммы, получение точечных и интервальных оценок.
4. Построение и интерпретация доверительного интервала.
5. Вычисление и интерпретация коэффициентов связи для номинальных данных.
6. Решение задачи на построение и анализ двумерной дискретной случайной величины с учетом зависимости и коррелированности одномерных компонент.
7. Предсказание значения изучаемой переменной методом регрессии.
8. Получение и интерпретация результатов дисперсионного анализа.
9. Проверка статистических гипотез.
10. Получение и интерпретация результатов факторного анализа.
11. Получение и интерпретация результатов кластерного анализа.
12. Составление отчета по проведенному анализу данных.

4.2. Вопросы для оценки качества освоения дисциплины

1. Создание и редактирование файлов данных.
2. Управление данными.
3. Редактирование графиков и диаграмм.
4. Описательные статистики.
5. Вычисление и интерпретация коэффициентов связи для номинальных данных.
6. Нормальное распределение случайной величины.
7. Регрессия. Линейная регрессия, Среднеквадратичная регрессия.
8. Закон больших чисел.
9. Центральная предельная теорема.
10. Выборочная и генеральная совокупности.
11. Эмпирическая функция распределения.
12. Выборочная средняя и выборочная дисперсия.
13. Точечная и интервальная оценки.
14. Доверительный интервал.
15. Понятие статистической гипотезы. Нулевая и альтернативная, простая и сложная гипотезы.
16. Ошибки первого и второго рода.
17. Проверка гипотезы о равенстве вероятностей.
18. Критерий согласия Хи-квадрат и таблицы сопряженности.
19. Ковариация и коэффициент корреляции двух случайных величин.
20. Доверительный интервал для коэффициента корреляции.
21. Статистические критерии двумерных случайных величин. Гипотеза об отсутствии корреляционной зависимости. t-критерий, F-критерий, критерий Уилкоксона.
22. Оценка регрессионных характеристик. Метод наименьших квадратов.
23. Однофакторный дисперсионный анализ.
24. Дисперсионный анализ с двумя факторами.

25. Понятие факторного и кластерного анализа.

Примеры заданий итогового контроля

В экзаменационный билет включены: теория и задачи. Студент письменно излагает ответ на вопрос теории, алгоритм решения задачи, интерпретирует полученные результаты и оценивает их статистическую значимость.

Примеры задач.

1. Студенту выдается таблица с данными (Например, в задаче о служащих некоторого офиса: заработная плата за год, пол, возраст и уровень подготовки).

- Охарактеризуйте переменные в этой задаче.
- Составьте таблицу сопряженности, используя процентные доли категорий по отношению общего количества сотрудников.
- Для переменной возраст постройте столбиковую диаграмму, секторную диаграмму и диаграмму Паретто. Какие еще виды диаграмм Вы знаете.
- Какая доля вариации заработной платы объясняется тем, что одни служащие имеют больший стаж работы, чем другие. Какую заработную плату можно ожидать для служащего со стажем работы 3 года?
- Постройте диаграмму рассеяния и опишите взаимосвязь переменных стаж работы и возраст. Найдите коэффициент корреляции.
- Найдите методом наименьших квадратов линию регрессии для прогнозирования связи стажа работы и возраста.
- Найдите 95% интервал для всех служащих, возраст которых составляет 50 лет.
- Наличие какого стажа работы можно ожидать у служащего в возрасте 40 лет?
- Проверьте по критерию согласия хи-квадрат соответствие экспериментального и нормального распределений для переменной возраст сотрудников. Интерпретировать результат.

По проведенному анализу в среде SPSS составьте отчет и проверьте результаты с помощью аналитических выкладок.

2. Задано среднее значение случайной величины и ее дисперсия. Найти размер выборки, позволяющей построить критерий проверки гипотезы с уровнем значимости 0,01 и 5 %-й ошибкой второго рода для выявления 10 %-х отклонений от гипотетического значения. Построить область принятия гипотезы для данного критерия.

Задачу решите в среде SPSS и проверьте результат с помощью аналитических выкладок.

V. РЕСУРСЫ

5.1 Основная литература

1. Кремер, Н.Ш. Теория вероятностей и математическая статистика [Электронный ресурс]: учебник и практикум для бакалавриата и специалитета / Н.Ш.Кремер; ЭБС Юрайт. – 5-е изд., перераб. и доп. – М.: Юрайт, 2019. – 538 с. – (Серия: Бакалавр. Академический курс). – ISBN 978-5-534-10004-4. – Режим доступа: <https://www.biblio-online.ru/viewer/teoriya-veroyatnostey-i-matematicheskaya-statistika-431167#page/1>. Загл. с экрана.

2. Анализ данных [Электронный ресурс]: учебник для академического бакалавриата / В.С.Мхитарян [и др.]; ЭБС Юрайт. – М.: Юрайт, 2019. – 490 с. – (Серия: Бакалавр. Академический курс). – ISBN 978-5-534-00616-2. – Режим доступа: <https://www.biblio-online.ru/viewer/analiz-dannyh-432178#page/1>. Загл. с экрана.

5.2 Дополнительная литература

1. Гмурман, В.Е. Теория вероятностей и математическая статистика [Электронный ресурс]: учебник для прикладного бакалавриата / В.Е.Гмурман; ЭБС Юрайт. – 12-е изд. – М.: Юрайт, 2019. – 479 с. – (Серия: Бакалавр. Прикладной курс). – ISBN 978-5-534-

00211-9. – Режим доступа: <https://www.biblio-online.ru/viewer/teoriya-veroyatnostey-i-matematicheskaya-statistika-431095#page/1>. – Загл. с экрана

2. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров / В.Е.Гмурман. – 12-е изд. – М.: Юрайт, 2014. – 478 с. – (Бакалавр. Базовый курс).

3. Моосмюллер, Г. Маркетинговые исследования с SPSS [Электронный ресурс]: учебное пособие / Г.Моосмюллер, Н.Н.Ребик; ЭБС Znanium. – 2-е изд. – М.: ИНФРА-М, 2011. – 200 с. – (Высшее образование). – Режим доступа: <http://znanium.com/bookread2.php?book=257371>. Загл. с экрана.

5.3 Дополнительная литература для самостоятельного изучения дисциплины

1. Плис А.И., Сливина Н.А. Практикум по прикладной статистике в среде SPSS: Учеб. пособие / А.И. Плис, Н.А. Сливина. В 2-х ч. Ч. 1. Классические процедуры статистики – М.: Финансы и статистика, 2004. - 288 с.

5.4 Программное обеспечение

№ п/п	Наименование	Условия доступа
1	MS Office 2007 Prof +	<i>Из внутренней сети университета (договор)</i>
2	Windows Professional 8.1 Russian	<i>Из внутренней сети университета (договор)</i>
3	IBM SPSS Statistics Premium Campus Edition	<i>Из внутренней сети университета (договор)</i>

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

№ п/п	Наименование	Условия доступа/скачивания
1	Электронные образовательные ресурсы	<i>Договор на использование электронных баз данных/по подключению и обеспечению доступа к базам данных</i>

5.6 Материально-техническое обеспечение дисциплины

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

– ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);

– мультимедийный проектор с дистанционным управлением.

Учебные аудитории для самостоятельных занятий по дисциплине имеют возможность подключения к сети Интернет и доступ к электронной информационно-образовательной среде НИУ ВШЭ.