

**МИНОБРНАУКИ РОССИИ**



Федеральное государственное бюджетное образовательное учреждение  
высшего образования

«Российский государственный гуманитарный университет»  
(ФГБОУ ВО «РГГУ»)

ИСТОРИКО-АРХИВНЫЙ ИНСТИТУТ

ФАКУЛЬТЕТ АРХИВОВЕДЕНИЯ И ДОКУМЕНТОВЕДЕНИЯ

*Кафедра источниковедения*

**Рабочая программа дисциплины**

**МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ  
СО СТАТИСТИЧЕСКИМИ ИСТОЧНИКАМИ**

**Направление подготовки 46.03.02 – Документоведение и архивоведение**

**Направленность – Электронные архивы и документы**

**Уровень квалификации выпускника – бакалавр**

**Форма обучения – очно-заочная**

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

**Москва 2021**

**МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ СО  
СТАТИСТИЧЕСКИМИ ИСТОЧНИКАМИ**

Рабочая программа курса

Составитель:

д.и.н., доц. И.М. Гарскова

к.ф.-м.н., доц. С.В. Шпирко

УТВЕРЖДЕНО

Протокол заседания кафедры

№\_\_4\_\_ от\_\_17.05.2021\_\_

## ОГЛАВЛЕНИЕ

1. Пояснительная записка .....	4
1.1. Цель и задачи дисциплины .....	4
1.2. Формируемые компетенции, а также перечень планируемых результатов обучения .....	5
1.3. Место дисциплины в структуре основной образовательной программы .....	6
2. Структура дисциплины .....	7
3. Содержание дисциплины .....	8
4. Образовательные технологии .....	9
5. Оценка планируемых результатов обучения .....	12
5.1. Система оценивания .....	12
5.2. Критерии выставления оценок .....	12
5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине .....	13
6. Учебно-методическое и информационное обеспечение дисциплины .....	16
6.1. Список источников и литературы .....	13
6.2. Перечень ресурсов информационно-телекоммуникационной сети Интернет .....	21
7. Материально-техническое обеспечение дисциплины .....	22
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов .....	22
9. Методические материалы .....	23
9.1. Планы практических (семинарских, лабораторных) занятий .....	23
9.2. Методические рекомендации по подготовке письменных работ .....	28
Приложения	
Приложение 1. Аннотация дисциплины .....	30
Приложение 2. Лист изменений .....	33

## **1. Пояснительная записка**

### **1. Цель и задачи дисциплины**

Дисциплина «МЕТОДЫ И ТЕХНОЛОГИИ РАБОТЫ СО СТАТИСТИЧЕСКИМИ ИСТОЧНИКАМИ» входит в состав части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение.

**Цель дисциплины:** познакомить студентов с возможностями визуализации, обработки и анализа статистических источников с помощью различных методов и технологий; дать студентам навыки использования стандартных и специализированных компьютерных программ.

#### **Задачи дисциплины:**

- характеристика основных содержательных проблем в исторических исследованиях, для решения которых используются данные статистических источников;
- изучение основных методов визуализации, обработки и анализа данных статистических источников с помощью табличных процессоров;
- изучение основных методов математической статистики для работы со статистическими источниками;
- знакомство со специализированным программным обеспечением для анализа статистических источников.

### **1.2. Формируемые компетенции, а также перечень планируемых результатов обучения**

<b>Коды компетенции</b>	<b>Содержание компетенций</b>	<b>Перечень планируемых результатов обучения по дисциплине</b>
ПК-1 Способность применять научно-методические	ПК-1.3 Решает стандартные задачи профессиональной	<b>Знать:</b> направления и опыт применения методов и технологий работы со статистическими источниками в исторических исследованиях; возможности и специфику визуализации, обработки

основы документоведения и архивоведения в профессиональной сфере	деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом требований информационной безопасности	и анализа информации статистических источников с помощью табличных процессоров ; основные методы, включенные пакеты статистических программ; логику стандартных и специализированных математических методов анализа информации статистических источников Уметь: формализовать содержательную задачу, требующую применения математических методов анализа статистических источников; выбрать адекватные методы и технологии работы с источниками, соответствующие решаемым исследовательским задачам; готовить материалы источников для компьютерного анализа с помощью соответствующих программ ; <i>Владеть:</i> навыками работы со стандартным и специализированным программным обеспечением ; аппаратными и программными средствами перевода текстов в электронный формат; математическими методами и компьютерными технологиями визуализации, обработки и анализа информации статистических источников ;
--	--	--

### ***1.3. Место дисциплины в структуре ООП***

Дисциплина "Методы и технологии работы со статистическими источниками" относится к части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение. Дисциплина реализуется кафедрой источниковедения факультета архивоведения и документоведения Историко-архивного института РГГУ. и адресована студентам 4 курса. Дисциплина реализуется кафедрой источниковедения факультета архивоведения и документоведения историко-архивного института РГГУ.

Для освоения дисциплины необходимы компетенции, сформированные в ходе изучения следующих дисциплин: "Математические методы в исторических исследованиях" и "Компьютерные технологии в исторических исследованиях". Данная дисциплина дополняет их навыками владения компьютерными методами визуализации, обработки и анализа информации статистических источников.

В результате освоения дисциплины "Методы и технологии работы со статистическими источниками" формируются компетенции, необходимые для изучения следующих дисциплин: "Источниковедение электронных документов" и "Электронные документы в составе историко-культурного наследия современной России".

## 2. Структура дисциплины

### Структура дисциплины для очно-заочной формы обучения

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 152 час., в том числе контактная работа обучающихся с преподавателем 48 час., самостоятельная работа обучающихся 104 час.

№ п/п	Раздел дисципли- ны/темы	Виды учебной работы (в часах)						Формы текуще- го контроля успеваемости, форма проме- жуточной атте- стации (по се- местрам)
		Контактная				Промежуточ- ная аттестация	Самостоятель- ная работа	
		Лекции	Семинар	Практические занятия	Лабораторные занятия			
1	Тема 1. Введение	2	4				12	доклад
2	Тема 2. Методология и методика работы с текстовыми докумен- тами	2	4				12	дискуссия
3	Тема 3. Концепции электронного текста	4	6				10	контрольная
4	Тема 4. Традиционные и формализованные методы анализа со- держания. Контент- анализ	4	6				30	дискуссия
5	Тема 5. Компьютеризован- ный анализ содержа-	4	4				20	Доклад

	ния текстов							
<b>6</b>	Тема 6. Опыт применения контент-анализа при изучении различных типов источников	<b>4</b>	<b>4</b>				<b>20</b>	дискуссия
	Зачет							защита проекта
	ИТОГО:	<b>20</b>	<b>28</b>				<b>104</b>	<b>152</b>

### **3. Содержание дисциплины**

#### ***Тема 1. Введение***

- Математизация научного знания: закономерности и этапы.
- Математические методы и анализ данных в исторических исследованиях: история становления и тенденции развития.
- Возможности и ограничения использования математических методов в гуманитарных науках.
- Проблема измерения исторических явлений; шкалы измерения.
- Основные понятия математической статистики и анализа данных.
- Понятие о теории статистического оценивания.
- Понятие о теории статистической проверки гипотез.

#### ***Тема 2. Возможности табличного процессора и СУБД для визуализации и анализа данных***

- Возможности табличного процессора MS Excel как средства обработки и анализа данных.
- Использование формул и функций.
- MS Excel как средство поиска ошибок в статистических источниках.
- Графические инструменты MS Excel.
- Вычисляемые поля и групповые операции в запросах MS Access.
- Сводные таблицы и сводные диаграммы.
- Построение таблиц сопряженности с помощью перекрестных запросов.

#### ***Тема 3. Статистическое описание (дескриптивная статистика)***

- Общая характеристика пакетов статистических программ.
- Классификация основных программных "модулей".
- Знакомство с пакетом Statistica: интерфейс и организация данных.
- Методы дескриптивной статистики и основные статистические характеристики количественных и качественных показателей.



- Визуализация данных.

#### ***Тема 4. Выборочный метод***

- Основные понятия выборочного метода.
- Генеральная совокупность и выборка.
- Случайные и систематические ошибки выборки.
- Точность и надежность выборочного метода; доверительный интервал.
- Понятие о доверительных интервалах. t-статистика.
- Проверка гипотез о равенстве выборочных средних и выборочных долей.

#### ***Тема 5. Корреляционный и регрессионный анализ***

- Статистический анализ связей – основные понятия.
- Различия в подходах и интерпретациях корреляционного и регрессионного анализа связей.
- Корреляционный анализ: линейная корреляция; коэффициент парной корреляции; коэффициент детерминации и его интерпретация.
- Понятие статистической значимости коэффициента корреляции.
- Линейная регрессия (парная и множественная): уравнение линейной регрессии; интерпретация коэффициента регрессии; его статистическая значимость.
- Коэффициент множественной корреляции.
- Уравнение множественной регрессии как статистическая объясняющая модель.

#### ***Тема 6. Анализ взаимосвязи качественных признаков (таблицы сопряженности)***

- Анализ взаимосвязей качественных данных. Типы качественных данных.

- Взаимосвязи ранговых признаков; коэффициенты ранговой корреляции, их статистическая значимость.
- Взаимосвязи номинальных признаков; таблицы сопряженности.
- Критерий Хи-квадрат и коэффициент Крамера.

#### ***Тема 7. Анализ динамических рядов***

- Компоненты динамического ряда.
- Методы построения и удаления тренда.
- Циклическая и сезонная составляющие ряда.
- Анализ случайной компоненты динамического ряда. Автокорреляция.
- Взаимосвязи динамических рядов.

#### ***Тема 8. Методы многомерного анализа данных***

- Специфика факторного и кластерного анализа.
- Методы многомерной классификации. Основные понятия.
- Кластерный анализ: иерархический метод, многомерная группировка.
- Многомерная классификация с использованием концепции нечетких множеств.
- Факторный анализ, цели его использования. Метод главных компонент.
- Понятие о факторных нагрузках и факторных весах, их смысл и пределы значений.
- Факторный анализ как способ классификации.

### **4. Образовательные технологии**

Образовательные технологии направлены на формирование общекультурных и профессиональных компетенций и ориентированы как на традиционные формы обучения, так и на интерактивный режим аудиторных и внеаудиторных занятий, коллективное и индивидуальное общение студентов с

преподавателем и консультации в режиме удаленного доступа (форумы, электронная почта).

Изложение материала является проблемно-ориентированным. На лекциях студенты знакомятся с основными направлениями в исторических исследованиях, где эффективно используются методы работы со статистическими источниками, с отечественным и зарубежным опытом исследований. Изучение отдельных методов и технологий визуализации, обработки и анализа статистических данных, их логики, возможностей и ограничений сочетает формы теоретического обучения и практических занятий, базируется на лекциях и компьютерных презентациях (может частично осуществляться в дистанционном режиме обучения).

Интерактивные формы обучения составляют 15–20% аудиторной нагрузки по данной дисциплине. Студенты вместе с преподавателем обсуждают наиболее сложные темы, получают коллективные и индивидуальные консультации и собеседования. Обсуждения и собеседования проводятся как в аудитории, так и в дистанционном режиме. Практическое освоение стандартного и специализированного программного обеспечения проводится в форме лабораторных работ на базе компьютерного класса, укомплектованного соответствующим программным обеспечением. Эта часть курса включает значительное число упражнений учебного характера, а также контрольных заданий, выполняемых самостоятельно. На семинарах происходит коллективное обсуждение результатов контрольных работ с анализом методических ошибок.

Самостоятельная работа студентов заключается в знакомстве с научной и методической литературой, в том числе с научно-образовательными интернет-ресурсами, и освоении программы Statistica для подготовки к лекциям, лабораторным и контрольным работам, а также к зачету по курсу.

Учебно-методические и справочные материалы по курсу, а также рабочие файлы для выполнения учебных и самостоятельных заданий размещены в Интернете на сайте исторического факультета МГУ.

№ п/п	Наименование раздела	Виды учебной работы	Формируемые компетенции	Информационные и образовательные технологии
1	2	3	4	5
1.	Введение	Лекция 1 Пр. занятие 1-2  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами
2.	Возможности табличного процессора и СУБД для визуализации и анализа данных	Лекция 2 Лаб. работа 1  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций
3.	Статистическое описание (дескриптивная статистика)	Лекция 3 Лаб. работа 1  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций
4.	Выборочный метод	Лекции 4-5 Пр. занятие 3 Лаб. работа 2  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций Подготовка к контрольной работе Контрольная работа 1
5.	Корреляционный и регрессионный анализ	Лекции 6-7 Пр. занятие 4 Лаб. работы 3-4  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций Подготовка к контрольной работе Контрольная работа 2
6.	Анализ взаимосвязи качественных признаков (таблицы сопряженности)	Лекции 8-9 Лаб. работа 5  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций
7	Анализ динамических рядов	Лекция 10 Пр. занятие 5 Лаб. работа 6  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций Подготовка к контрольной работе Контрольная работа 3
8	Методы многомерного анализа данных	Лекции 11-12 Пр. занятие 6 Лаб. работа 7  Самостоятельная работа	ПК-4  ПК-4	Компьютерная презентация Чтение литературы, знакомство с интернет-ресурсами Подготовка к лаб. работе с использованием электронного курса лекций Подготовка к контрольной работе Контрольная работа 4

## 5. Фонд оценочных средств для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

### 5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
– контрольная	10,25 баллов	50 баллов
– опрос	5 баллов	10 баллов
Промежуточная аттестация		40 баллов
Итого за семестр		100 баллов

### 5.2. Критерии выставления оценки по дисциплине

В качестве *текущего контроля* используются такие формы, как контрольные работы и итоговое тестирование. Формой *промежуточной аттестации* является зчет.

Для получения зачета обучающийся в сумме должен набрать не менее 50 баллов из 100 возможных (итоговое тестирование – 40 баллов, четыре контрольные работы – 60 баллов).

Совокупный результат (максимум 100 баллов) конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

### Текущий контроль.

При оценивании контрольной работы учитываются полнота выполнения заданий, логичность рассуждений и обоснованность выводов:

- если задание выполнено не полностью и (или) допущены серьезные ошибки – оценка до 5 баллов;
- если задание выполнено, но имеются ошибки в рассуждениях и интерпретации выводов – оценка 5–10 баллов;
- если задание выполнено полностью, в рассуждениях и интерпретации выводов нет принципиальных ошибок, при том, что возможны небольшие неточности – оценка 10–15 баллов.

### Промежуточная аттестация (зачет)

При проведении промежуточной аттестации студент должен ответить на два вопроса, первый – теоретического характера, второй – практического характера).

При оценивании ответа на теоретический вопрос применяются следующие критерии оценки:

- теория освоена фрагментарно, имеются грубые ошибки в ответе – оценка 5–10 баллов;
- теория освоена не полностью, допущен ряд ошибок – оценка 10–15 баллов);
- теория освоена в полном объеме, – оценка 15–20 баллов.

При оценивании ответа на практический вопрос применяются следующие критерии оценки:

- навыки практической работы освоены недостаточно, ответ содержит грубые ошибки – оценка 5–10 баллов;
- навыки практической работы имеются, но ответ неполон и (или) содержит ряд ошибок – оценка 10–15 баллов;

- ответ полный, показывает высокий уровень освоения навыков практической работы, хотя могут присутствовать небольшие неточности – оценка 15–20 баллов.

### ***5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине***

#### Вопросы к зачету:

- Опыт использования методов математической статистики для анализа информации исторических источников.
- Возможности табличных процессоров и СУБД для визуализации и анализа данных.
- Статистическая совокупность. Объекты и признаки. Количественные и качественные признаки.
- Дескриптивная статистика. Меры среднего уровня и меры разброса. Гистограммы.
- Выборочный метод. Ошибки выборки. Доверительный интервал и доверительная вероятность.
- Корреляционный и регрессионный анализ. Парная и множественная регрессия.
- Интерпретация коэффициентов корреляции и регрессии.
- Анализ взаимосвязи ранговых и номинальных признаков.
- Динамические ряды и их компоненты.
- Методы многомерного статистического анализа.

#### Примеры вариантов контрольных заданий:

- Контрольная работа по темам 3 и 4.
  1. Открыть файл Turpol.sta с показателями аграрного развития 50 губерний Европейской России на рубеже XIX – XX вв.

- ✓ Определить, по какому из двух показателей (файл) степень неоднородности совокупности губерний выше: var19 (осенние цены ржи в коп/пуд), var13 (доля безлошадных дворов). Дать ответ с указанием коэффициентов.
- 2. Сохранились фрагменты источника, в котором зарегистрированы данные о росте рекрутов в 1877 и 1914 гг. (ввести данные таблицы, создать новый файл в программе STATISTICA).
  - ✓ Можно ли по этим данным сделать вывод о том, что рост призывников действительно увеличился, то есть, является ли статистически значимой разница среднего роста рекрутов в указанные годы? Дать обоснование ответа для доверительной вероятности  $P=95\%$ .
  - ✓ Дать графическую интерпретацию результата, используя доверительные интервалы.
- Контрольная работа по теме 5.
  1. Открыть файл Turpol.sta с показателями аграрного развития 50 губерний Европейской России на рубеже XIX – XX вв.
    - ✓ Построить матрицу корреляции первых шести признаков.
    - ✓ Перечислить те коэффициенты корреляции, которые являются статистически значимыми. Объяснить смысл статистической значимости по отношению к самому высокому из положительных коэффициентов корреляции, используя понятие доверительного интервала.
    - ✓ Дать графическую интерпретацию доверительного интервала.
    - ✓ Дать содержательную интерпретацию самой сильной связи в этой матрице, используя названия двух соответствующих признаков.
  2. Открыть файл Turpol.sta с показателями аграрного развития 50 губерний Европейской России на рубеже XIX – XX вв.



- ✓ Построить регрессионное уравнение, оценивающее зависимость сбора хлебов в пудах на душу (признак var7) от обеспеченности лошадьми (признак var12, лошадей на душу) и доли наемных с/х рабочих (признак var1, в %).
  - ✓ Дать точную (количественную) интерпретацию коэффициентов регрессии  $b_1$  и  $b_{12}$ , то есть, объяснить, что означают конкретные числовые значения этих коэффициентов и их знаки. Указать, в каких единицах измеряются эти коэффициенты в данном случае.
  - ✓ Какие из этих двух коэффициентов являются статистически значимыми? На основании чего можно сделать такой вывод?
  - ✓ Каково значение коэффициента множественной корреляции в данной регрессионной модели? Какой процент вариаций зависимого признака (сбор хлебов на душу) объясняют два учтенных независимых признака? Достаточно ли этих двух признаков?
- Контрольная работа по темам 6 и 7.
    1. Открыть файл General.sta, содержащий просопографические данные о генералах – участниках Великой Отечественной войны.
      - ✓ Построить таблицу сопряженности двух признаков: социальное происхождение (строки таблицы сопряженности) и должность в Первой мировой войне (столбцы таблицы сопряженности).
      - ✓ Отобрать только 3 категории по социальному происхождению и 4 – по воинской должности (исходя из наиболее высоких частот встречаемости).
      - ✓ По какой клетке таблицы сопряженности реальная частота наиболее сильно отличается от ожидаемой частоты?

- ✓ Какая должность наиболее часто встречается в социальной группе выходцев из крестьян (указать %)?
  - ✓ К какой социальной группе чаще всего относятся рядовые (указать %)?
  - ✓ Получить значение коэффициента Крамера.
2. Открыть файл Index1.sta, содержащий значения индексов цен (индексов инфляции) в России в 1885–1914 гг.
    - ✓ Построить и интерпретировать уравнение линейного тренда для показателя Var4. Удалить тренд из исходного ряда.
    - ✓ Для этого же показателя выполнить сглаживание ряда с использованием скользящей средней по пяти точкам.
    - ✓ С помощью автокорреляции найти "память" процесса (при каком лаге коэффициент автокорреляции равен нулю?).
- Контрольная работа по теме 8.
    1. Открыть файл Family.sta, содержащий данные о среднем размере семьи по республикам СССР за 1959, 1970 и 1979 гг.
      - ✓ Провести кластер-анализ данных по методу Joining (иерархическому). Описать состав групп, когда их число равно двум.
      - ✓ Провести кластер-анализ данных по методу K-means, задав число групп, равное двум.
      - ✓ Найти для каждого класса самый типичный объект.
      - ✓ Сравнить средние значения классов для 1979 г.
      - ✓ Сравнить результаты обоих методов кластер-анализа. Сделать вывод.
    2. Открыть файл Turpol.sta с показателями аграрного развития 50 губерний Европейской России на рубеже XIX – XX вв.
      - ✓ Провести факторный анализ (использовать для анализа признаки переменные 1, 2, 3, 9, 13, 14, 15, 17.). Подобрать число факторов так, чтобы суммарный вклад факторов в объяснен-

ную дисперсию был не менее 50%. Указать вклад каждого фактора.

- ✓ Указать для каждого фактора по три признака, наиболее существенных для его интерпретации. Привести соответствующие значения факторных нагрузок.
- ✓ Как можно назвать эти факторы?
- ✓ Для первого фактора определить две губернии, которые являются наиболее развитыми по этому фактору, и две губернии, которые являются наименее развитыми.

## **6. Учебно-методическое и информационное обеспечение дисциплины:**

### **6.1. Список источников и литературы**

#### ***а) основная литература:***

- Боровиков В.П. Популярное введение в программу STATISTICA. М., 1998.
- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998.
- Бородкин Л.И. Многомерный статистический анализ в исторических исследованиях. М., 1986.
- Бородкин Л.И., Гарскова И.М. Программное обеспечение FuzzyClass в историко-типологическом исследовании // История и компьютер: новые информационные технологии в исторических исследованиях и образовании. Goettingen, 1993.
- Историческая информатика / Под ред. Л.И. Бородкина и И.М. Гарсковой, М., 1996.
- Количественные методы в исторических исследованиях. Учебное пособие / Отв. ред. И.Д. Ковальченко. М., Высшая школа, 1984.
- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999.

- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.Ж Бином-Пресс, 2008.

#### ***б) дополнительная литература***

- Историческая информатика / Под ред. Л.И. Бородкина и И.М. Гарсковой. М., 1996. 400 С.
- Информатика для гуманитариев / Под ред. Л.И. Бородкина и И.М. Гарсковой, М., 1997. 220 С.
- История. Статистика. Информатика. Барнаул, АГУ, 1995. 238 С.
- Источник. Метод. Компьютер. Барнаул, АГУ, 1996. 228 С.
- Компьютер и экономическая история // Отв. ред. Л.И. Бородкин и В.Н. Владимиров. Барнаул: АГУ, 1997. 228 С.
- Миронов Б.Н. История в цифрах. Математика в исторических исследованиях. Л., Наука, 1991. 168 С.
- Серия "Математические методы в исторических исследованиях". М., 1972, 1975, 1977, 1981, 1985, 1989, 1992;
- Серия "ЭВМ и математические методы". М. 1993, 1994.

#### ***в) программное обеспечение***

- MS Office
- StatSoft Statistica

### ***6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимый для освоения дисциплины***

#### ***г) ресурсы Интернета***

- <http://www.hist.msu.ru/Departments/Inf/Stud/index.html> — учебно-методические и справочные материалы по курсу, а также рабочие файлы для выполнения учебных и контрольных заданий на сайте исторического факультета МГУ.

## 7. Материально-техническое обеспечение дисциплины

- Компьютерный класс с установленным программным обеспечением MS Office, StatSoft Statistica
- Подключение компьютеров к Интернету.
- Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочные систем (ИСС) (2018 г.)

### • 1. Перечень БД и ИСС (к п. 6.2)

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru

### • Состав программного обеспечения (ПО) (к п. 7)

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
11	Microsoft Office 2013	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

## **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
  - письменные задания выполняются на компьютере в письменной форме;
  - экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
- письменные задания выполняются на компьютере со специализированным программным обеспечением;
- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата:

- в печатной форме;
- в форме электронного документа;
- в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - передвижными, регулируемые эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.

## **9. Методические материалы**

**9.1. Планы практических занятий** (с указанием тематических разделов и количества часов)

### Указания

Лабораторные работы выполняются с помощью стандартного (программа Statistica) и специализированного программного обеспечения (программа



FuzzyClass). Рабочие файлы и файлы контрольных заданий используют фрагменты реальных исторических источников. Коллекция файлов доступна в Интернете (<http://www.hist.msu.ru/Departments/Inf/Stud/index.html>).

## ***Тема 2. Возможности табличного процессора и СУБД для визуализации и анализа (2 час.)***

1. Возможности табличного процессора MS Excel:
  - формулы и функции,
  - сортировка и фильтры,
  - автосуммирование,
  - математические, статистические и логические функции,
  - графики.
2. Возможности СУБД MS Access:
  - вычисляемые поля в запросах,
  - групповые операции в запросах,
  - сводные таблицы и сводные диаграммы,
  - перекрестные запросы.

### **Контрольные вопросы:**

1. Примеры статистических и логических функций в MS Excel.
2. Что такое фильтры в MS Excel?
3. Как использовать функцию автосуммирования для поиска ошибок в данных?
4. Основные типов графиков в MS Excel.
5. Какие графики нельзя использовать для визуализации динамических рядов?
6. Что такое вычисляемое поле?
7. Основные групповые операции в запросах.
8. Можно ли построить гистограмму средствами MS Access?
9. Как построить таблицу сопряженности в MS Access?

## Литература

- Информационные технологии для историков. Учебное пособие. М., МГУ, 2006. С. 70–93, 127–153.

### ***Тема 3. Статистическое описание (дескриптивная статистика) (2 час.).***

1. Знакомство со структурой и интерфейсом программы Statistica:
  - классификация основных программных "модулей",
  - переключение между модулями,
  - организация данных: таблицы и их элементы.
2. Организация данных:
  - команды для работы со строками и столбцами,
  - создание новых признаков,
  - формулы,
  - автоматическая кодировка качественных признаков.
3. Методы дескриптивной статистики и основные статистические характеристики (работа с файлом duma.sta):
  - меры среднего уровня и разброса,
  - вариационные ряды и частоты,
  - сравнение вариационных рядов.
4. Визуализация данных:
  - виды графического представления (гистограммы, круговые диаграммы,
  - категоризованные распределения.

### Контрольные вопросы:

1. Типы признаков.
2. Графическая интерпретация вариационного ряда.
3. Что такое гистограмма?
4. Меры среднего уровня.

5. Можно ли найти среднее арифметическое значение (медиану) для качественного признака?
6. Меры разброса.
7. В чем сходство и различие между  $\sigma$  и  $V$ ?
8. В каких единицах измеряется коэффициент вариации?

#### Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 6–31.
- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998. С. 64–75, 89–99, 105–125, 149–155, 351–359, 368–373.
- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.Ж Бином-Пресс, 2008. С. 20–32, 59–61, 81–84.

#### ***Тема 4. Выборочный метод (4 час.) – работа с файлом industry.sta.***

1. Методы построения выборки:
  - построение механической выборки.
2. Статистическое оценивание среднего значения количественного признака:
  - включение доверительного интервала и стандартной ошибки среднего в набор результатов дескриптивной статистики,
  - выбор доверительной вероятности,
  - построение доверительного интервала с заданной вероятностью.
3. Проверка гипотезы о статистической значимости различия двух выборочных средних.
4. Статистическое оценивание доли качественного признака:
  - вычисление стандартной ошибки доли качественного признака,
  - выбор доверительной вероятности,

- построение доверительного интервала с заданной вероятностью.
5. Проверка гипотезы о статистической значимости различия двух выборочных долей.

Контрольные вопросы:

1. Когда в историческом исследовании возникает проблема выборки?
2. Что такое репрезентативность?
3. "Естественная" выборка.
4. Случайные и систематические ошибки.
5. Может ли быть абсолютно точным результат выборочного исследования?
6. Верно ли, что выборка дает тем лучший результат, чем больше ее объем?
7. Что такое доверительный интервал?
8. Верно ли, что увеличение точности результата выборочного исследования связано с уменьшением надежности?
9. При увеличении объема выборки доверительный интервал:
  - увеличивается;
  - уменьшается;
  - не изменяется.

Укажите верный ответ.

Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 33–64.

***Тема 5. Корреляционный и регрессионный анализ (8 час.) – работа с файлами `industry.sta` и `duma.sta`***

1. Анализ парных взаимосвязей:
  - построение диаграмм рассеяния,

- построение уравнения линейной регрессии,
  - интерпретация коэффициента регрессии,
  - интерпретация коэффициентов парной корреляции и детерминации,
  - определение статистической значимости коэффициента корреляции.
2. Множественная корреляция и регрессия:
- построение уравнения множественной регрессии,
  - интерпретация коэффициентов регрессии,
  - определение единиц измерения коэффициентов регрессии,
  - интерпретация коэффициентов множественной корреляции и детерминации,
  - определение статистической значимости коэффициентов регрессии.

#### Контрольные вопросы:

1. Почему на диаграмме рассеяния точки отклоняются от линии регрессии?
2. Смысл коэффициента корреляции.
3. В каких границах заключен коэффициент корреляции?
4. Может ли значение  $r = 0$  говорить об отсутствии связи?
5. Как проверить значимость коэффициентов корреляции?
6. В выборке получен коэффициент корреляции, равный 0,68. Может ли это значение быть статистически незначимым?
7. Смысл коэффициента регрессии.
8. Смысл коэффициента множественной корреляции.
9. Как проверить значимость коэффициентов регрессии?

#### Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 65–86.
- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998. С. 30–53, 386–390.

- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.Ж Бином-Пресс, 2008. С. 85–91, 153–163.

***Тема 6. Анализ взаимосвязи качественных признаков (4 час.) – работа с файлами industry.sta и duma.sta***

1. Анализ взаимосвязей качественных данных для ранговых признаков: подсчет коэффициентов ранговой корреляции.
2. Анализ таблиц сопряженности для номинальных признаков:
  - построение таблицы сопряженности по всем категориям двух признаков,
  - выявление наиболее важных категорий каждого признака,
  - изменение размеров таблицы сопряженности,
  - анализ относительных частот в таблице сопряженности,
  - вычисление критерий Хи-квадрат и определение его статистической значимости.
3. Работа с бинарными признаками:
  - построение четырехклеточных таблиц,
  - подсчет коэффициента взаимосвязи.

**Контрольные вопросы:**

1. Типы качественных признаков.
2. Чем качественные признаки отличаются от количественных?
3. Привести примеры ранговых признаков из исторических источников.
4. Свойства коэффициентов ранговой корреляции.
5. Что такое номинальные признаки? Примеры.
6. Что такое таблица сопряженности.
7. В чем смысл критерия Хи-квадрат?
8. Свойства коэффициента Крамера V.
9. Какие меры связи можно использовать при изучении признаков разной природы?

## Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 87–102.
- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998. С. 380–383.
- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.Ж Бином-Пресс, 2008. С. 100–106.

## ***Тема 7. Анализ динамических рядов (4 час.) – работа с файлом cotton.sta***

1. Начало работы:
  - выбор переменных,
  - визуализация ряда.
2. Декомпозиция ряда:
  - удаление линейного тренда,
  - сглаживание ряда методом скользящей средней,
  - удаление сезонной "волны".
3. Работа со случайной компонентой динамического ряда:
  - вычисление автокорреляции остатков,
  - корреляция динамических рядов.

## Контрольные вопросы:

1. Компоненты динамического ряда.
2. Этапы анализа динамического ряда.
3. Что такое тренд?
4. Что такое сезонная компонента динамического ряда?
5. Метод скользящих средних. Что такое центрирование?
6. Сколько точек исходного ряда вы потеряете при использовании 7-членной скользящей средней?
7. Что такое случайная компонента ряда?

## 8. Для чего нужна автокорреляция?

### Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 133–152.
- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998. С. 380–383.
- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.Ж Бином-Пресс, 2008. С. 431–486.

### ***Тема 8. Многомерный статистический анализ (4 час.) – работа с файлами harvest.sta и tab1987.sta.***

1. Кластерный анализ: применение агломеративно-иерархического метода:
  - выбор меры расстояния между объектами,
  - выбор правила объединения объектов,
  - графическое представление результатов объединения,
  - интерпретация построенной классификации.
2. Кластерный анализ: применение метода k-средних:
  - задание числа классов,
  - выбор меры расстояния между объектами,
  - выбор правила объединения объектов,
  - определение состава классов,
  - определение центров классов,
  - графическое представление и интерпретация результатов.
3. Гибкая классификация: использование концепции нечетких множеств (работа с программой FuzzyClass):
  - выбор числа классов,
  - определение меры принадлежности объектов классам,
  - интерпретация ядра и периферии класса, переходных объектов.



#### 4. Факторный анализ: метод главных компонент:

- выбор способа представления данных,
- выбор метода,
- выбор числа факторов,
- определение вкладов факторов,
- интерпретация факторов с помощью матрицы факторных нагрузок,
- выбор объектов с наибольшими и наименьшими значениями факторных весов,
- кластерный анализ объектов с использованием значений их факторных весов.

#### Контрольные вопросы:

1. Какова содержательная гипотеза, лежащая в основе кластерного анализа?
2. Рассмотрим два основных метода кластерного анализа, используемых в статистических пакетах – иерархический и метод К-средних. В чем их главное различие?
3. Что такое матрица расстояний?
4. Изменяются ли результаты кластерного анализа, если удалить часть признаков из имеющегося набора?
5. Назовите основные меры расстояния между объектами (или кластерами) при использовании иерархического метода кластерного анализа.
6. Дать интерпретацию понятия "фактор".
7. Что такое факторные нагрузки? Существуют ли пределы их значений?
8. Что такое факторные веса? Существуют ли пределы их значений?
9. Что является показателем качества построенной факторной модели?

#### Литература

- Компьютеризованный статистический анализ для историков. Учебное пособие / Под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1999. С. 87–132.

- Боровиков В.П., Боровиков И.П. Statistica. Статистический анализ и обработка данных в среде Windows. М.: "Филинь", 1998. С. 491–507.
- Халафян А.А. STATISTICA 6. Статистический анализ данных. Учебник. 3-е изд. М.: Бином-Пресс, 2008. С. 289–305.

## **9.2. Методические рекомендации по подготовке письменных работ**

**Самостоятельная работа** нацелена на расширение теоретических знаний и практических навыков на основе чтения рекомендуемой литературы и знакомства с научно-образовательными тематическими интернет-ресурсами.

Для подготовки к лекциям, практическим занятиям, лабораторным и контрольным работам, а также к зачету студенты используют список рекомендованной литературы, материалы лекций и презентаций по данным темам, список вопросов к зачету, планы лабораторных работ и списки контрольных вопросов к ним, примеры контрольных заданий.

Для подготовки к лабораторным и контрольным работам и выполнения практических заданий на компьютере используется программное обеспечение StatSoft Statistica.

Самостоятельная работа может выполняться студентом в библиотеке, в компьютерном классе или дома.

Вид работы	Содержание	Трудоемкость	Рекомендации
Подготовка к пр. занятию по теме 1	<b>Тема 1</b> (см. тематический план)	4 час.	См. список литературы и интернет-ресурсы
Подготовка к лекции и лабораторной работе по теме 2	<b>Тема 2</b> (см. тематический план)	4 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к лекции и лабораторной работе по теме 3	<b>Тема 3</b> (см. тематический план)	6 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к лекции, практ. занятию и лабораторной работе по теме 4	<b>Тема 4</b> (см. тематический план)	4 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к контрольной работе по темам 3–4	<b>Темы 3–4</b> (см. тематический план)	6 час.	См. планы лабораторных работ, примеры контрольных заданий и материалы лекций по темам
Подготовка к лекции, практ. занятию	<b>Тема 5</b> (см. тематический план)	7 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список кон-

и лабораторной работе по теме 4			трольных вопросов и материалы лекций по теме
Подготовка к контрольной работе по теме 5	<b>Тема 5</b> (см. тематический план)	8 час.	См. планы лабораторных работ, примеры контрольных заданий и материалы лекций по теме
Подготовка к лекциям и лабораторной работе по теме 6	<b>Тема 6</b> (см. тематический план)	6 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к лекции, практ. занятию и лабораторной работе по теме 7	<b>Тема7</b> (см. тематический план)	7 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к контрольной работе по темам 6–7	<b>Темы 6–7</b> (см. тематический план)	6 час.	См. планы лабораторных работ, примеры контрольных заданий и материалы лекций по темам
Подготовка к лекциям, практ. занятию и лабораторной работе по теме 8	<b>Тема 8</b> (см. тематический план)	4 час.	См. список литературы, интернет-ресурсы, планы лабораторных работ, список контрольных вопросов и материалы лекций по теме
Подготовка к контрольной работе по теме 8	<b>Тема 8</b> (см. тематический план)	6 час.	См. планы лабораторных работ, примеры контрольных заданий и материалы лекций по теме
Подготовка к промежуточной аттестации	Письменный зачет	18 час.	См. список литературы, список вопросов к зачету и материалы лекций
<b>Итого по курсу</b>		<b>87 час.</b>	

### Аннотация

Дисциплина "**Методы и технологии работы со статистическими источниками**" входит в состав части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение и адресована студентам 4 курса. Дисциплина реализуется кафедрой источниковедения факультета архивного дела историко-архивного института РГГУ.

Цель дисциплины: познакомить студентов с возможностями визуализации, обработки и анализа статистических источников с помощью различных методов и технологий; дать студентам навыки использования стандартных и специализированных компьютерных программ.

Задачи дисциплины: характеристика основных содержательных проблем в исторических исследованиях, для решения которых используются данные статистических источников; изучение основных методов визуализации, обработки и анализа данных статистических источников с помощью табличных процессоров; изучение основных методов математической статистики для работы со статистическими источниками; знакомство со специализированным программным обеспечением для анализа статистических источников.

Дисциплина направлена на формирование следующих компетенций выпускника: способность самостоятельно работать с различными источниками информации (ПК-1.3).

В результате освоения дисциплины обучающийся должен:

Знать направления и опыт применения методов и технологий работы со статистическими источниками в исторических исследованиях; возможности и специфику визуализации, обработки и анализа информации статистических источников с помощью табличных процессоров; основные методы, включенные пакеты статистических программ; логику стандартных и специализированных математических методов анализа информации статистических источ-

ников. Уметь формализовать содержательную задачу, требующую применения математических методов анализа статистических источников; выбрать адекватные методы и технологии работы с источниками, соответствующие решаемым исследовательским задачам; готовить материалы источников для компьютерного анализа с помощью соответствующих программ; интерпретировать результаты анализа и формулировать содержательные выводы. Владеть математическими методами и компьютерными технологиями визуализации, обработки и анализа информации статистических источников; навыками работы со стандартным и специализированным программным обеспечением; основами интерпретации результатов анализа.

Рабочей программой предусмотрены следующие виды контроля: текущий контроль успеваемости в форме контрольных работ; промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы (152 часа).