

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Владимирский государственный университет  
имени Александра Григорьевича и Николая Григорьевича Столетовых»  
(ВлГУ)

ИНСТИТУТ ТУРИЗМА И ПРЕДПРИНИМАТЕЛЬСТВА  
(Наименование института)



УТВЕРЖДАЮ:  
Директор института  
Яресь О.Б.  
2021 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

ИНСТРУМЕНТЫ И ПРОГРАММНЫЕ СРЕДСТВА АНАЛИЗА ДАННЫХ  
(наименование дисциплины)

**направление подготовки / специальность**

38.04.06 «ТОРГОВОЕ ДЕЛО»

(код и наименование направления подготовки (специальности))

**направленность (профиль) подготовки**

КОММЕРЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ НА РЫНКЕ ТОВАРОВ И УСЛУГ

(направленность (профиль) подготовки)

Владимир

2021

## 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

**Цель освоения дисциплины** – научиться решать практические задачи с использованием инструментария профессионального статистического анализа данных, моделирования и прогнозирования социально-экономических явлений и процессов, содержательной интерпретации результатов статистических расчетов.

### Задачи:

- изучение статистических методов и вычислительных алгоритмов, позволяющих извлекать знания из результатов наблюдений;
- формирование у студента основы аналитического мышления для последующего его применения в экономике, социальной сфере и бизнесе;
- формирование умения применять методы анализа данных для анализа реальных данных с использованием вычислений на современных вычислительных устройствах.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина «Инструменты и программные средства анализа данных» к обязательной части учебного плана программы магистратуры по направлению 38.04.06 «Торговое дело».

## 3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Планируемые результаты обучения по дисциплине «Инструменты и программные средства анализа данных», соотнесенные с планируемыми результатами освоения ОПОП (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	
ОПК-2. Способен применять инструментальные методы сбора, обработки и анализа данных, необходимые для стратегического планирования и координации деятельности торговых структур	ОПК-2.1. Применяет инструментальные методы сбора, обработки и анализа данных для стратегического планирования деятельности торговых структур	<b>Знает:</b> - основные понятия и положения, связанные со сбором, систематизацией, обработкой и анализом статистических данных; <b>Умеет:</b> - определять методы анализа, необходимые для оценки степени и вида зависимостей между переменными, снижения размерности признакового пространства и многомерной классификации данных; - использовать методы устойчивого, робастного оценивания параметров и непараметрического моделирования; <b>Владеет:</b> - навыками описательной статистики, табличного и графического представле-	Тестовые вопросы Практико-ориентированные задания

		<ul style="list-style-type: none"> <li>- ния данных, их содержательной интерпретации;</li> <li>- навыками применения многомерных статистических методов исследования зависимостей, снижения размерностей и классификации при анализе социально-экономических данных с использованием аналитического программного обеспечения;</li> </ul>	
	ОПК-2.2. Разрабатывает стратегические планы развития торговых структур на основе результатов анализа	<p>Знает:</p> <ul style="list-style-type: none"> <li>- основные подходы к анализу данных с использованием описательных и вероятностно-статистических методов;</li> </ul> <p>Умеет:</p> <ul style="list-style-type: none"> <li>- анализировать временные данные и прогнозировать;</li> </ul> <p>Владеет:</p> <ul style="list-style-type: none"> <li>- навыками анализа динамики временных данных и прогнозирования социально-экономических процессов с использованием статистических методов и аналитического программного обеспечения;</li> </ul>	Тестовые вопросы Практико-ориентированные задания
ОПК-5. Способен применять современные информационные технологии и программные средства, в том числе использовать интеллектуальные информационно-аналитические системы, при решении профессиональных задач	ОПК-5.1. Использует современные информационные технологии и программные средства при решении профессиональных задач	<p>Знает:</p> <ul style="list-style-type: none"> <li>- современные программные средства, применяемые при анализе данных;</li> </ul> <p>Умеет:</p> <ul style="list-style-type: none"> <li>- применять современные программные средства для анализа данных;</li> </ul> <p>Владеет:</p> <ul style="list-style-type: none"> <li>- навыками применения современных программных средств для анализа данных о деятельности торговых структур.</li> </ul>	Тестовые вопросы Практико-ориентированные задания
	ОПК-5.2. Использует электронные библиотечные системы для поиска необходимой научной литературы и социально-экономической статистики	<p>Знает:</p> <ul style="list-style-type: none"> <li>- электронные источники, онлайн базы данных для поиска научной информации и статистических данных для анализа деятельности торговых структур;</li> </ul> <p>Умеет:</p> <ul style="list-style-type: none"> <li>- использовать электронные библиотечные системы для поиска необходимой научной литературы и социально-экономической</li> </ul>	Тестовые вопросы Практико-ориентированные задания

		статистики для анализа деятельности торговых структур; Владеет: - навыками применения информационно-аналитических систем для анализа деятельности торговых структур.	
--	--	--	--

#### 4. ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 8 зачетных единиц, 288 часов

##### Тематический план форма обучения – заочная (2 г, 6 мес)

№ п/п	Наименование тем и/или разделов/тем дисциплины	Семестр	Неделя семестра	Контактная работа обучающихся с педагогическим работником				Самостоятельная работа	Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные работы	В форме практической подготовки		
1	Предварительный анализ данных. Описательная статистика	1		2				19	
2	Генеральная и выборочная совокупности	1						19	
3	Корреляционный анализ	1			2			19	Рейтинг-контроль №1
4	Регрессионный анализ. Двумерная линейная модель регрессии	1		2				19	
5	Множественная линейная модель регрессии	1			4			19	Рейтинг-контроль №2
6	Нелинейные модели регрессии и их линеаризация	1						19	
7	Регрессионные модели с фиктивными переменными	1						20	Рейтинг-контроль №3
	<b>Всего за 1 семестр:</b>			<b>4</b>	<b>6</b>			<b>134</b>	<b>Зачет</b>
8	Снижение размерности признакового пространства	2						15	
9	Классификация многомерных наблюдений	2						15	Рейтинг-контроль №1
10	Кластерный анализ, непараметрическая классификация без обучения	2						15	
11	Классификация с обучением, дискриминантный анализ	2						16	Рейтинг-контроль №2
12	Робастное оценивание параметров и непараметрические модели генеральной совокупности	2						16	
13	Анализ временных данных	2		2	4			16	
14	Использование моделей авторегрессии	2			2			16	Рейтинг-контроль №3
	<b>Всего за 2 семестр:</b>			<b>2</b>	<b>6</b>			<b>109</b>	<b>Экзамен (27)</b>
	<b>Наличие в дисциплине КР/КР</b>								-
	<b>Итого по дисциплине</b>			<b>6</b>	<b>12</b>			<b>243</b>	<b>Экзамен (27)</b>

## Содержание лекционных занятий по дисциплине

### **Тема 1. Предварительный анализ данных. Описательная статистика**

Классификация статистических данных. Анализ одномерных категориальных данных. Анализ одномерных количественных данных. Предварительный анализ временных данных.

### **Тема 4. Регрессионный анализ. Двумерная линейная модель регрессии**

Основные понятия регрессионного анализа. Оценивание параметров регрессии. Определение интервальной оценки для  $b_0$ . Определение интервальной оценки и проверка значимости  $b_1$ . Определение интервальной оценки для условного математического ожидания. Модель регрессии в случае двумерной нормальной генеральной совокупности.

### **Тема 13. Анализ временных данных**

Введение в анализ временных данных. Методы сглаживания временных данных и моделирования тенденции развития. Статистический анализ и прогнозирование сезонных колебаний во временных данных. Применение адаптивных моделей, основанных на экспоненциальном сглаживании, для краткосрочного прогнозирования.

## Содержание практических/лабораторных занятий по дисциплине

### **Тема 3. Корреляционный анализ**

Основные понятия корреляционного анализа. Корреляционный анализ взаимосвязи количественных признаков. Корреляционный анализ взаимосвязи качественных признаков. Канонические корреляции и канонические величины генеральной совокупности. Оценка канонических корреляций и канонических величин - *практико-ориентированные задания.*

### **Тема 5. Множественная линейная модель регрессии**

Оценивание параметров линейной модели регрессии и анализ свойств оценок. Проверка значимости уравнения и коэффициентов регрессии. Доверительные интервалы для параметров регрессионной модели - *практико-ориентированные задания.*

### **Тема 13. Анализ временных данных**

Методы сглаживания временных данных и моделирования тенденции развития. Прогнозирование сезонных колебаний во временных данных. Краткосрочное прогнозирование - *практико-ориентированные задания.*

### **Тема 14. Использование моделей авторегрессии**

Модели стационарных временных рядов. Методология применения моделей ARIMA - *практико-ориентированные задания.*

## **5. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ**

### **5.1. Текущий контроль успеваемости.**

Текущий контроль успеваемости по дисциплине «Инструменты и программные средства анализа данных» проводится в форме рейтинг-контроля три раза в семестр. Типовые тестовые вопросы и задачи для текущего контроля приведены ниже.

### **1-ый семестр**

#### **Рейтинг-контроль №1**

1. Переменная «число дождливых дней в октябре» относится к типу:

- а) категориальная номинальная;
- б) категориальная порядковая;
- в) количественная дискретная;
- г) количественная непрерывная.

2. Переменная «число преступлений в течение года в Москве за 1990—2015 гг.» относится к типу:

- а) интервальный временной ряд;
- б) моментный временной ряд;
- в) пространственная переменная;
- г) пространственно-временная переменная.

3. Понятие «мода» применимо к данным:

- а) только количественным;
- б) только категориальным;
- в) количественным и категориальным;
- г) временным.

4. С помощью гистограммы можно изобразить графически данные:

- а) только количественные;
- б) только категориальные;
- в) количественные и категориальные;
- г) временные.

5. Значению признака, левее и правее которого находится половина всех значений, соответствует следующая числовая характеристика:

- а) мода;
- б) медиана;
- в) среднее;
- г) дисперсия.

6. Определить, является ли гистограмма распределения признака скошенной в какую-либо сторону, можно с помощью следующей числовой характеристики:

- а) дисперсия;
- б) коэффициент вариации;
- в) коэффициент асимметрии;
- г) коэффициент эксцесса.

7. Определить, является ли гистограмма распределения признака плоско- или островершинной по сравнению с кривой нормального распределения, можно с помощью следующей числовой характеристики:

- а) дисперсия;
- б) коэффициент вариации;
- в) коэффициент асимметрии;
- г) коэффициент эксцесса.

8. На главной диагонали ковариационной матрицы вектора  $X$  находятся следующие моменты элементов вектора:

- а) центральные моменты первого порядка;
- б) центральные моменты второго порядка;
- в) начальные моменты второго порядка;
- г) начальные моменты первого порядка.

9. Уровень значимости статистического критерия  $\alpha$  выбирают в интервале:

- а) от 0 до 0,5;
- б) от -1 до 0;
- в) от 0,5 до 1;
- г) от -1 до 1.

10. Зависимость, при которой каждому значению величины  $X$  соответствует единственное значение величины  $Y$  и наоборот, называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

11. Зависимость, при которой каждому фиксированному значению независимой переменной  $X$  соответствует не одно, а множество значений переменной  $Y$ , называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

12. Зависимость, при которой функциональной зависимостью связаны фактор  $X$  и среднее значение результирующего показателя  $K$ , называется:

- а) статистической;
- б) корреляционной;
- в) функциональной;
- г) математической.

13. Графическое изображение реальных статистических данных в виде точек в декартовой системе координат называется:

- а) диаграммой рассеивания;
- б) корреляционным полем;
- в) круговой диаграммой;
- г) верификацией модели.

14. Коэффициент детерминации между  $x$  и  $y$  характеризует:

- а) долю дисперсии  $y$ , обусловленную влиянием не входящих в модель факторов;
- б) долю дисперсии  $y$ , обусловленную влиянием  $x$ ;
- в) долю дисперсии  $x$ , обусловленную влиянием не входящих в модель факторов;
- г) направление зависимости между  $x$  и  $y$ .

15. Парный коэффициент корреляции между факторами равен единице. Это означает:

- а) наличие нелинейной функциональной связи;
- б) отсутствие связи;
- в) наличие функциональной связи;
- г) отрицательную линейную связь.

### Рейтинг-контроль №2

1. Сравнить силу влияния факторов на исследуемый показатель можно, используя:

- а) коэффициенты регрессии в натуральном масштабе из многофакторной модели регрессии;
- б) частные коэффициенты эластичности;
- в) стандартизованные коэффициенты регрессии из многофакторного анализа.

2. Два фактора должны быть признаны коллинеарными, если

- а) величина коэффициента корреляции между ними больше 0,7;
- б) величина коэффициента корреляции между ними меньше 0,7;
- в) коэффициент корреляции между ними по модулю больше 0,7.

3. Совокупность наблюдений можно считать однородной по показателю  $x_j$ , если коэффициент вариации (в %)

- а) равен 33;
- б) меньше 33;
- в) больше 33;

4. Задачами корреляционно-регрессионного анализа являются:

- а) изучить оценки неизвестных параметров уравнения;
- б) оценить адекватность модели;
- в) качественно предсказать результат в будущем;
- г) определить тесноту связи между показателями.



5. Множественный коэффициент корреляции характеризует тесноту связи между

- а) несколькими зависимыми переменными и одной независимой переменной;
- б) несколькими зависимыми и несколькими независимыми переменными;
- в) одной зависимой переменной и несколькими независимыми переменными;
- г) одной зависимой и одной независимой переменными.

6. Укажите последовательность этапов проведения корреляционно-регрессионного анализа.

- а) проверка адекватности построенной модели;
- б) интерпретация полученных результатов;
- в) отбор факторов для построения модели;
- г) выбор вида регрессионной модели и оценка ее параметров;
- д) постановка задачи;
- е) сбор исходных данных их анализ.

7. Многофакторная регрессионная модель считается статистически значимой, если

- а)  $F_{табл} > F_{расч}$ ;
- б)  $F_{табл} = F_{расч}$ ;
- в)  $F_{табл} < F_{расч}$ .

8. К универсальным показателям тесноты связи относят:

- а) парный коэффициент корреляции;
- б) корреляционное отношение;
- в) множественный коэффициент детерминации;
- г) индекс корреляции.

9. Надежность модели определяется по:

- а) T-статистике Стьюдента;
- б) F-критерию Фишера;
- в) коэффициенту детерминации;
- г) средней относительной ошибке аппроксимации.

10. Число наблюдений считается достаточным, когда:

- а)  $n - p - 1 < 30$ ;
- б)  $n - p + 1 > 30$ ;
- в)  $n - p - 1 > 30$ ;
- г)  $n - p + 1 < 30$ .

11. Регрессионная связь – это

- а) связь между одной зависимой и несколькими другими, называемыми независимыми переменными, выраженная с помощью математической модели;
- б) связь, при которой на величину исследуемого показателя оказывают множество факторов, действующих в различных направлениях одновременно или последовательно;
- в) связь между несколькими зависимыми и независимыми показателями, действующих в различных направлениях одновременно и последовательно.

12. Коэффициент множественной детерминации показывает, какую часть вариации исследуемого показателя объясняют факторы

- а) неучтенные в многофакторной регрессионной модели;
- б) присутствующие в многофакторной регрессионной модели;
- в) как присутствующие, так и неучтенные в многофакторной регрессионной модели.

13. Исследование остатков предполагает проверку наличия следующих предпосылок метода наименьших квадратов:

- а) отсутствие автокорреляции остатков;
- б) остатки не случайны;
- в) гетероскедастичность;
- г) гомоскедастичность;
- д) случайный характер остатков;
- е) наличие зависимости остатков от изменения факторов;
- ж) нулевая средняя величина остатков;

з) остатки подчиняются нормальному закону распределения.

14. Значения частного коэффициента корреляции могут находиться в интервале:

- а) [-1; 0];
- б) [0, 4];
- в) [-1; 1];
- г) [0; 1].

15. Верные или ложные следующие утверждения:

- а) автокорреляция характерна в основном для временных рядов;
- б) статистика Дарбина-Уотсона лежит в пределах от 0 до 4;
- в) автокорреляция всегда является следствием неправильной спецификации модели;
- г) при наличии автокорреляции значение коэффициента детерминации  $R^2$  будет всегда существенно ниже единицы.

### Рейтинг-контроль №3

1. Фиктивные переменные – это

- а) независимые переменные;
- б) количественные переменные;
- в) качественные переменные;
- г) качественные переменные, преобразованные в количественные.

2. Мультиколлинеарность – это связь между:

- а) признаками;
- б) уровнями;
- в) явлениями.

3. Аналитическое выражение связи определяется с помощью метода анализа:

- а) корреляционного;
- б) регрессионного;
- в) группировок

4. Параметр  $a_1$  ( $a_1 = -1,04$ ) линейного уравнения регрессии  $y = 34,7 - 1,04x$  показывает, что:

- а) с увеличением признака  $X$  на 1 признак  $Y$  уменьшается на 1,04;
- б) с увеличением признака  $X$  на 1 признак  $Y$  уменьшается на 36,5;
- в) связь между признаками  $X$  и  $Y$  прямая;
- г) связь между признаками  $X$  и  $Y$  обратная.

5. В результате проведения регрессионного анализа получают информацию, описывающую:

- а) взаимосвязь показателей;
- б) соотношение показателей;
- в) структуру показателей;
- г) темпы роста показателей;
- д) темпы прироста показателей.

6. Если уравнение регрессии между себестоимостью единицы продукции ( $Y$ ) и производительностью труда одного работника ( $X$ ) выглядит следующим образом:  $Y = 320 - 0,2X$ , то при увеличении факторного признака результативный:

- а) не изменяется;
- б) увеличивается;
- в) изменяется произвольно;
- г) уменьшается.

7. Коэффициент регрессии в уравнении  $y = 6 + 0,4x$ , характеризующем связь между размером жилой площади квартиры (кв. м) и ценой квартиры (тыс. долл.), означает, что при увеличении жилой площади на 1 кв. м цена квартиры увеличится в среднем на:

- а) 0,4%;
- б) 6 тыс. долл.;
- в) 0,4 тыс. долл.

8. Какие формулы используются для аналитического выражения нелинейной связи между переменными:

а)  $y^c = a + bx$ ;

б)  $y = a + \frac{b}{x}$ ;

в)  $y^c = a + bx + cx^2$ ;

г)  $y^c = ax^b$ .

9. Связь между двумя признаками характеризуется линейным уравнением регрессии:

$$y = 0,68 + 0,02x.$$

Коэффициент регрессии показывает, что:

а) связь между признаками прямая;

б) связь между признаками обратная;

в) с увеличением признака  $x$  на единицу признак  $y$  в среднем уменьшается на 0,02;

г) с увеличением признака  $x$  на единицу признак  $y$  в среднем увеличивается на 0,02.

10. В модели  $ln Y = b_0 + bX + e$  коэффициент  $b$  имеет смысл:

а) абсолютного прироста;

б) темпа роста;

в) темпа прироста.

11. При анализе эластичности спроса по цене целесообразно использовать следующую модель:

а) линейную;

б) полиномиальную;

в) логарифмическую;

г) степенную;

д) экспоненциальную.

12. Академиком А.Н.Колмогоровым было предложено:

а) “обобщенное расстояние” между классами;

б) расстояние, измеряемое по принципу “ближайшего соседа”;

в) расстояние, измеряемое по принципу “дальнего соседа”;

г) расстояние, измеряемое по “центрам тяжести” групп;

д) расстояние, измеряемое по принципу “средней связи”.

13. Оценки неизвестных параметров  $A$ ,  $a$  и  $b$  в производственной функции Кобба - Дугласа можно найти с помощью:

а) метода наименьших квадратов;

б) принципа “ближнего соседа”;

в) дисконтированием множителей.

14. Если вектор ошибок имеет постоянную дисперсию, то это явление называется:

а) гомоскедастичностью;

б) гетероскедастичностью;

в) ситуация не определена.

15. С увеличением объема выборки:

а) увеличивается точность оценок;

б) уменьшается ошибка регрессии;

в) расширяются интервальные оценки;

г) уменьшается коэффициент детерминации.

## 2-ой семестр

### Рейтинг-контроль №1

1. Нормирование признаков производят с целью:

а) устранения влияния различных единиц измерения;

б) уменьшения размерности признакового пространства;

в) упрощения расчетов.

2. При исследовании взаимосвязи двух показателей  $x_1$  и  $x_2$  получен коэффициент корреляции  $r = 0,9$ . Тогда максимальное собственное число, соответствующее первой главной компоненте, равно:
- а) 0,1;
  - б) 1,8;
  - в) 0,2;
  - г) -0,9.
3. В двумерной совокупности коэффициент корреляции между величинами  $x$  и  $y$  равен 0,95. Тогда собственные значения  $\lambda_x$  и  $\lambda_y$  равны:
- а) 1,05 и 0,95;
  - б) 1,40 и 0,22;
  - в) 1,95 и 0,05;
  - г) 1,45 и 0,55.
4. Финансовая устойчивость предприятия характеризуется восемью показателями. В результате расчетов получены собственные значения трех первых главных компонент:  $\lambda_1 = 4,0$ ;  $\lambda_2 = 1,6$  и  $\lambda_3 = 0,8$ . Тогда относительный вклад двух первых главных компонент (в %) равен:
- а) 30;
  - б) 70;
  - в) 60;
  - г) 80.
5. При исследовании взаимосвязи двух показателей  $x^{(1)}$  и  $x^{(2)}$  получен коэффициент корреляции  $r_{12} = 0,9$ . Тогда собственное значение соответствующее первой главной компоненте, равно:
- а) 0,1;
  - б) 1,9;
  - в) 1,8;
  - г) 0,2.
6. Деятельность  $n$  предприятий региона характеризуется четырьмя показателями. При проведении компонентного анализа по матрице  $R$  получены собственные значения, одно из которых оказалось пропущенным: 1,2; 1,4 и 0,6. Собственное значение  $\lambda_3$ , соответствующее третьей главной компоненте, равно:
- а) 2,5;
  - б) 1,2;
  - в) 0,6;
  - г) 0,8.
7. В кластер  $S_1$  входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2,5,6,7. Расстояние от пятого объекта до кластера  $S_1$ , если исходить из метода «ближнего соседа», равно:
- а) 2;
  - б) 5
  - в) 6
  - г) 7.
8. В кластер  $S_1$  входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2, 5, 6, 7. Расстояние от пятого объекта до кластера  $S_1$ , если исходить из принципа «дальнего соседа», равно:
- а) 2;
  - б) 5;
  - в) 6;
  - г) 7.
9. В кластер  $S_1$  входят четыре объекта, расстояние от которых до пятого объекта составляет соответственно 2, 5, 6, 7. Расстояние от пятого объекта до кластера  $S_1$  если исходить из принципа «средней связи», равно:

- а) 2;
- б) 5
- в) 6
- г) 7

10. Процедура, при которой начальным является разбиение, состоящее из  $n$  одноэлементных классов, называется:

- а) дивизимной;
- б) корреляционной;
- в) агломеративной;
- г) линейной.

11. Кластерный анализ позволяет проводить:

- а) группировку объектов;
- б) группировку объектов и группировку признаков;
- в) группировку коэффициентов корреляции;
- г) группировку дисперсий;
- д) группировку признаков.

12. Нормирование признаков производят с целью:

- а) устранить влияние различных единиц измерения;
- б) уменьшить признаковое пространство;
- в) упростить расчеты;
- г) выделить латентные факторы.

13. Процедура, при которой начальным является разбиение, состоящее из одного класса, а конечное — из  $n$  одноэлементных классов, называется:

- а) дивизимной;
- б) корреляционной;
- в) агломеративной;
- г) линейной.

14. Большинство программ, реализующих алгоритм иерархической классификации, предусматривает графическое представление результатов классификации в виде:

- а) дендрограммы;
- б) блок-схемы;
- в) дерева решений.

15. В задачах многомерной классификации объектов при  $a = (3 = -8 = 1/2$  и  $y = 0$  расстояние между классами определяется по методу:

- а) «дальнего соседа»;
- б) «средней связи»;
- в) «ближнего соседа».

### Рейтинг-контроль №2

1. При оценке качества разбиения на классы принимается, что дисперсия внутри класса должна быть

- а) минимальной
- б) максимальной
- в) равной для всех классов
- г) не учитывается

2. Спорный объект кластеризации — это объект, который по мере сходства ...

- а) может быть отнесен к нескольким кластерам
- б) не может быть отнесен ни к одному кластеру
- в) может быть отнесен более чем к двум кластерам

3. Работа кластерного анализа опирается на предположения:

- а) рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры

- б) правильность выбора масштаба или единиц измерения признаков
  - в) отнесение всех объектов к одному из predetermined классов
4. Процедура, которая приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некоей величине, отражающей определенные свойства, это – ...
- а) стандартизация
  - б) нормирование
  - в) оба ответа верны
5. Работа кластерного анализа опирается на следующие предположения (выберите неверный ответ):
- а) рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры
  - б) правильность выбора масштаба или единиц измерения признаков
  - в) отнесение всех объектов к одному из predetermined признаков
6. Иерархические агломеративные методы характеризуются ...
- а) последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
  - б) делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
  - в) сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
7. Объект относится к кластеру, если ...
- а) расстояние от объекта до центра кластера меньше радиуса кластера
  - б) расстояние от объекта до центра кластера меньше диаметра кластера
  - в) расстояние от объекта до центра кластера больше радиуса кластера
8. Работа кластерного анализа опирается на предположения:
- а) рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры
  - б) отнесение всех объектов к одному из predetermined признаков
  - в) о сравнимости шкал
9. Иерархические дивизимные методы характеризуются ...
- а) последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
  - б) делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
  - в) сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
10. Назовите характеристики кластерного анализа:
- а) не требует априорных предположений о наборе данных
  - б) требует априорных предположений о наборе данных
  - в) не накладывает ограничения на представление исследуемых объектов
  - г) накладывает ограничения на представление исследуемых объектов
  - д) позволяет анализировать показатели различных типов данных.
11. При применении кластерного анализа переменные ...
- а) должны измеряться в сравнимых шкалах
  - б) могут измеряться в каких угодно шкалах
  - в) должны быть только числовыми
12. Характеристикой каких групп методов является последовательное объединение исходных элементов и соответствующее уменьшение числа кластеров?
- а) иерархические агломеративные методы
  - б) иерархические дивизимные (делимые) методы
  - в) и тех, и других

13. Назовите характеристики кластерного анализа:

- а) позволяет сокращать размерность данных
- б) позволяет делать данные более наглядными
- в) имеет в своем арсенале около десяти алгоритмов
- г) имеет в своем арсенале около сотни алгоритмов

14. Кластерный анализ ...

- а) может применяться к совокупностям временных рядов
- б) не может применяться к совокупностям временных рядов
- в) может определять группы временных рядов со схожей динамикой

15. Деление одного кластера на меньшие кластеры, в результате чего образуется последовательность расщепляющих групп. Характеристика каких групп методов описана выше?

- а) иерархические агломеративные методы
- б) иерархические дивизимные (делимые) методы
- в) и тех, и других

### Рейтинг-контроль №3

1. Выберите истинные утверждения

- а) В случае, когда анализируемая по панельным данным совокупность охватывает страны и регионы, рекомендуется использовать модель со случайными эффектами.
- б) Тест Хаусмана является асимптотическим, что обуславливает эффективность его применения при малых выборках.
- в) Within-оценки получают на основе уравнения, рассчитанного по групповым средним.
- г) При выборе между фиксированными и случайными эффектами в моделях панельных данных априорно тяжело доказать обоснованность применения модели со случайными эффектами.
- д) При альтернативной гипотезе в тесте Хаусмана состоятельны лишь оценки модели с фиксированными эффектами.

2. Какой тест, один или несколько, не используется для диагностики стационарности временного ряда?

- а) Грейнджера
- б) Чоу
- в) Хадри
- г) Левин-Ли-Чу
- д) Перрона

3. Какой тест, один или несколько, не используется для диагностики остатков регрессионной модели?

- а) Тест Бреуша-Пагана
- б) Тест Хаусмана
- в) Тест Бреуша-Годфрея
- г) Тест Вальда
- д) Тест Спирмена

4. Выберите истинные утверждения

- а) В тесте Квятковского-Филлипса-Шмидта-Шина проверяются три спецификации - с наличием тренда и константы, с наличием константы и так называемая спецификация None.
- б) В случае когда по тестам ADF и KPSS получают разные выводы относительно стационарности временного ряда, это объясняется либо тем, что временной ряд может содержать в себе нелинейный тренд, либо маломощностью обоих критериев (тестов).
- в) Расширенный тест Дикки-Фуллера используется для проверки нестационарности временного ряда в том случае, если для ряда характерны структурные сдвиги.
- г) Критические точки для теста "единичного корня" ADF содержатся в таблицах МакКиннона.

д) Для эффективности оценок модели с фиксированными эффектами необходимо, чтобы в тесте Хаусмана принималась нулевая гипотеза

5. Проинтерпретируйте результаты, полученные при построении модели на основании квартальных данных для показателей экспорта, ВВП, обменного курса (соответственно EXPORT, GDP, KURS). Выберите правильные с вашей точки зрения варианты ответов.

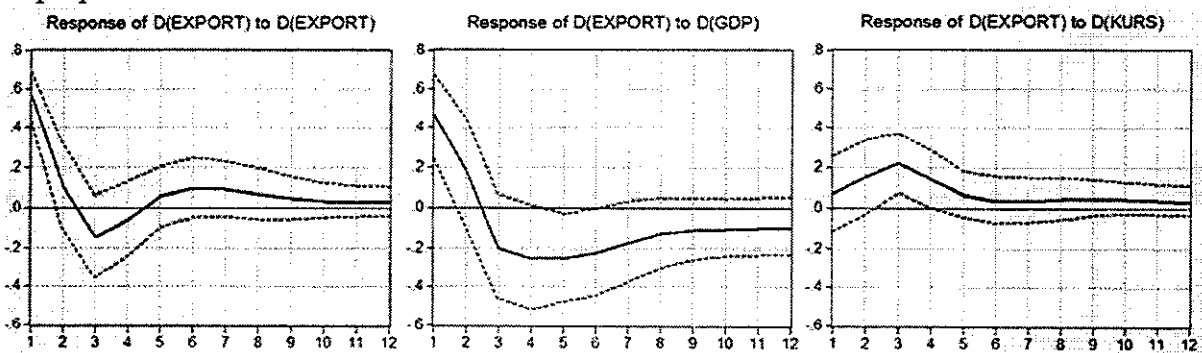
а) Инновация обменного курса приводит к максимальному отклику экспорта через три квартала.

б) Положительный отклик экспорта на шок ВВП наблюдается в течение полугодия и является статистически значимым.

в) Для показателя ВВП характерна автокорреляционная составляющая.

г) Инновация обменного курса приводит к максимальному отклику экспорта через два квартала.

д) Используемая модель является моделью Тейлора, основанной на системе векторных авторегрессий.



6. При моделировании зависимости индекса потребительских цен  $Y$  от экзогенных объемных макроэкономических показателей  $X_1$  и  $X_2$ , в ходе проверки временных рядов на нестационарность, было установлено, что  $Y$  представлен нестационарным временным рядом, интегрированным второго порядка, а показатели  $X_1$  и  $X_2$  представлены нестационарными временными рядами, интегрированными первого порядка. Выберите модели регрессионной зависимости для выбранных показателей из предложенных.

а)  $d(d(Y))=b_0+b_1*d(X_1)+b_2*d(X_2)+e$

б)  $Y=b_0+b_1*X_1+b_2*X_2+e$

в)  $d(Y)=b_0+b_1*d(X_1)+b_2*d(X_2)+e$

г)  $d(Y)=b_0+b_1*X_1+b_2*X_2+e$

$Y=b_0+b_1*d(X_1)+b_2*d(X_2)+e$

7. Проинтерпретируйте результаты тестирования эконометрической модели для переменных  $Y_1, Y_2, Y_3$ . Выберите правильные с вашей точки зрения варианты ответов.

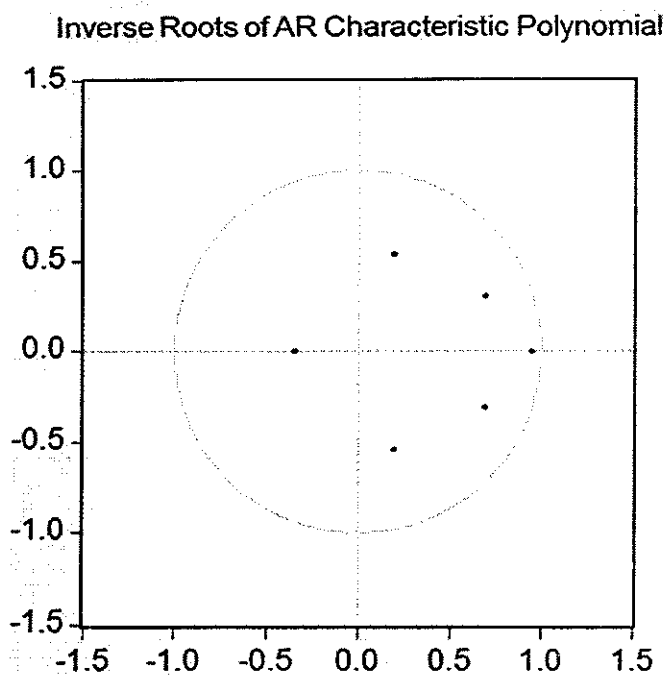
а) Представлено графическое отображение коэффициентов AR-схемы 6го порядка.

б) Представлены корни характеристического уравнения VAR-модели порядка  $k=6$ . Представлены корни характеристического уравнения VAR-модели порядка  $k=2$ .

в) Для модели выполняется условие обратимости.

г) Для модели не выполняется условие обратимости.





8. Проинтерпретируйте результаты тестирования эконометрической модели. Выберите правильные с вашей точки зрения варианты ответов.

- а) Построена модель ARMA(2;0).
- б) Построена модель ARIMA(2;1;0).
- в) Для модели выполняется условие обратимости.
- г) Для модели не выполняется условие обратимости.

Dependent Variable: LM1

Sample (adjusted): 2005M03 2012M04

Included observations: 86 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.528709	0.999510	0.528969	0.5982
AR(1)	1.757934	0.081701	21.51671	0.0000
AR(2)	-0.749470	0.087157	-8.599105	0.0000
R-squared	0.997139	Mean dependent var	1.622655	
Adjusted R-squared	0.997070	S.D. dependent var	0.729532	
S.E. of regression	0.039488	Akaike info criterion	-3.591356	
Sum squared resid	0.129425	Schwarz criterion	-3.505739	
Log likelihood	157.4283	F-statistic	14464.13	
Durbin-Watson stat	1.919953	Prob(F-statistic)	0.000000	
Inverted AR Roots	1.03	.73		

9. Для тестирования остатков авторегрессионной модели на наличие автокорреляции (серийной корреляции) исследователю необходимо:

- а) Проанализировать значения и доверительные вероятности статистики Жака-Бера.
- б) Проанализировать результаты теста Бреуша-Годфрея для разного количества лагов.
- в) Проанализировать значение статистики Дарбина-Уотсона DW.
- г) Проанализировать результаты ARCH теста для разного количества лагов.
- д) Проанализировать коррелограмму остатков авторегрессионной модели.

10. По результатам проведения теста KPSS для временного ряда переводных депозитов населения РБ проверьте гипотезу о нестационарности временного ряда LDP, определив также спецификацию теста.

- а) Временной ряд LDP нестационарен на 10% уровне, спецификация ряда - тренд и константа.
- б) Временной ряд LDP нестационарен для 1% уровня, спецификация - константа и тренд.
- в) Временной ряд LDP стационарен для 1% уровня, спецификация - спецификация - константа и тренд.

- г) Временной ряд LDP нестационарен для 10% уровня, спецификация ряда - константа.  
 д) Временной ряд LDP стационарен на 5% уровне, спецификация ряда - тренд и константа.  
 Null Hypothesis: LDP is stationary

		LM-Stat.		
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.195507		
Asymptotic critical values*:	1% level	0.216000		
	5% level	0.146000		
	10% level	0.119000		
KPSS Test Equation				
Dependent Variable: LDP				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67.09002	0.838909	79.97290	0.0000
@TREND()	0.123976	0.036101	3.434150	0.0014
R-squared	0.232183	Mean dependent var	69.56954	
Adjusted R-squared	0.212496	S.D. dependent var	3.082107	
S.E. of regression	2.735106	Akaike info criterion	4.897768	
Sum squared resid	291.7514	Schwarz criterion	4.981357	
Log likelihood	-98.40424	F-statistic	11.79339	
Durbin-Watson stat	0.136448	Prob(F-statistic)	0.001423	

11. Выберите истинные утверждения

- а) Остатки модели коинтеграции должны иметь Гауссовское распределение, что проверяется с помощью ARCH теста.  
 б) Критические точки для теста "единичного корня" ADF имеют распределение Стьюдента.  
 в) Нестационарные временные ряды называются интегрированными первого порядка.  
 г) При построении коинтеграционного соотношения по Энглу-Грейнджеру для временных рядов строится модель векторной регрессии.  
 д) При исследовании вопроса нестационарности временных рядов необходимо проанализировать как коррелограмму самого временного ряда, так и временных рядов его разностей первого и второго порядка.

12. Выберите истинные утверждения

- а) Наличие ARCH эффектов в модели тестируется на основании коррелограммы остатков модели (функций ACF и PACF).  
 б) При тестировании модели Бокса-Дженкинса ARMA на адекватность требуется, чтобы в модели отсутствовала мультиколлинеарность.  
 в) Стационарные временные ряды называются интегрированными нулевого порядка.  
 г) При построении коинтеграционного соотношения по Энглу-Грейнджеру для временных рядов строится модель векторной регрессии.  
 д) VAR-модели представляют собой систему авторегрессионных уравнений, которая является системой независимых уравнений.

13. Выберите истинные утверждения:

- а) Для временных рядов некорректно применение традиционных методов построения регрессионных моделей.  
 б) Нестационарные временные ряды называются интегрированными.  
 в) Чем короче нестационарный временной ряд, тем выше, как правило, его порядок интегрированности.

г) Тесты "единичного" корня возможно использовать для установления факта коинтегрированности временных рядов.

д) Модели Бокса-Дженкинса ARMA возможно построить только для стационарных временных рядов.

14. Какие два теста "единичного корня" рекомендуется использовать совместно?

а) Тест Дикки-Фуллера (DF).

б) Расширенный тест Дикки-Фуллера (ADF).

в) Тест Нг-Перрона (Ng-P).

г) Тест Квятковского-Филлипса-Шмидта-Шина (KPSS).

д) Тест Бреуша-Пагана (BP).

15. Выберите утверждения, которые вы могли бы принять, основываясь на анализе представленной коррелограммы временного ряда:

а) Временной ряд имеет нормальное распределение.

б) Временной ряд содержит сезонные колебания.

в) Для временного ряда характерна условная гетероскедастичность.

г) Временной ряд содержит структурные сдвиги.

д) Временной ряд является стационарным.

Autocorrelation	Partial Correlation	AC	PAC	
█	█	1	-0.338	-0.338
█	█	2	-0.246	-0.406
█	█	3	-0.195	-0.607
█	█	4	0.725	0.445
█	█	5	-0.277	0.146
█	█	6	-0.239	-0.009
█	█	7	-0.142	-0.166
█	█	8	0.499	-0.263
█	█	9	-0.195	-0.183
█	█	10	-0.166	-0.002
█	█	11	-0.130	-0.050
█	█	12	0.390	0.022

## 5.2. Промежуточная аттестация по итогам освоения дисциплины.

Промежуточная аттестация по итогам освоения дисциплины «Инструменты и программные средства анализа данных» проводится в форме зачета в 1-ом семестре и в форме экзамена во 2-ом.

### Вопросы к зачету

1. Классификация статистических данных.
2. Анализ одномерных категориальных данных.
3. Анализ одномерных количественных данных.
4. Предварительный анализ временных данных.
5. Распределение генеральной совокупности
6. Характеристики генеральной совокупности
7. Выборка из генеральной совокупности
8. Статистическое оценивание параметров генеральных совокупностей
9. Статистическая проверка гипотез о параметрах генеральной совокупности
10. Корреляционный анализ взаимосвязи количественных признаков.
11. Корреляционный анализ взаимосвязи качественных признаков.
12. Канонические корреляции и канонические величины генеральной совокупности.
13. Оценка канонических корреляций и канонических величин.
14. Оценивание параметров регрессии.
15. Определение интервальной оценки для  $b_0$ .
16. Определение интервальной оценки и проверка значимости  $b_1$ .
17. Определение интервальной оценки для условного математического ожидания.

18. Модель регрессии в случае двумерной нормальной генеральной совокупности.
19. Оценивание параметров линейной модели регрессии и анализ свойств оценок.
20. Проверка значимости уравнения и коэффициентов регрессии.
21. Доверительные интервалы для параметров регрессионной модели.
22. Нелинейные модели регрессии и их линеаризация.
23. Регрессионные модели с фиктивными переменными.

#### **Вопросы к экзамену**

1. Снижение размерности признакового пространства.
2. Компонентный анализ
3. Факторный анализ
4. Эвристические методы снижения размерности
5. Многомерное шкалирование
6. Классификация многомерных наблюдений.
7. Кластерный анализ.
8. Непараметрическая классификация без обучения.
9. Классификация с обучением, дискриминантный анализ.
10. Робастное оценивание параметров.
11. Аномальные значения. Методы обнаружения засорения выборки
12. Устойчивые параметрические методы оценивания
13. Оценки на основе порядковых статистик
14. Непараметрические модели распределений
15. Оценки методами бутстреп-анализа
16. Непараметрические модели генеральной совокупности.
17. Методы сглаживания временных данных и моделирования тенденции развития.
18. Статистический анализ и прогнозирование сезонных колебаний во временных данных.
19. Применение адаптивных моделей, основанных на экспоненциальном сглаживании, для краткосрочного прогнозирования.
20. Модели стационарных временных рядов.
21. Методология применения моделей ARIMA.

#### **5.3. Самостоятельная работа обучающегося предполагает:**

- 1) выполнение типового расчета в соответствии с методическими указаниями, доступными по ссылке [2021 \(vlsu.ru\)](http://2021.vlsu.ru).

Фонд оценочных материалов (ФОМ) для проведения аттестации уровня сформированности компетенций обучающихся по дисциплине оформляется отдельным документом.

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 6.1. Книгообеспеченность

Наименование литературы: автор, название, вид издания, издательство	Год издания	КНИГООБЕСПЕЧЕННОСТЬ
		Наличие в электронном каталоге ЭБС
Основная литература		
1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2021. — 490 с.	2021	<a href="https://urait.ru/bcode/469022">https://urait.ru/bcode/469022</a>
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с.	2021	<a href="https://urait.ru/bcode/450262">https://urait.ru/bcode/450262</a>
3. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel : учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2021. — 353 с.	2021	<a href="https://urait.ru/bcode/471895">https://urait.ru/bcode/471895</a>
Дополнительная литература		
1. Демидова, О. А. Эконометрика : учебник и практикум для вузов / О. А. Демидова, Д. И. Малахов. — Москва : Издательство Юрайт, 2021. — 334 с.	2021	<a href="https://urait.ru/bcode/469219">https://urait.ru/bcode/469219</a>
2. Вакуленко, Е. С. Эконометрика (продвинутый курс). Применение пакета Stata : учебное пособие для вузов / Е. С. Вакуленко, Т. А. Ратникова, К. К. Фурманов. — Москва : Издательство Юрайт, 2021. — 246 с.	2021	<a href="https://urait.ru/bcode/476410">https://urait.ru/bcode/476410</a>
3. Горленко, О. А. Дисперсионный анализ экспериментальных данных : учебное пособие для вузов / О. А. Горленко, Н. М. Борбаць, Т. П. Можяева. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2021. — 132 с.	2021	<a href="https://urait.ru/bcode/479041">https://urait.ru/bcode/479041</a>

### 6.2. Периодические издания

1. Бизнес-информатика\*

2. Вопросы экономики\*

\* – литература из фонда библиотеки ВлГУ

### 6.3. Интернет ресурсы

1. [www.consultant.ru](http://www.consultant.ru) – Справочно-правовая система «Консультант Плюс»
2. <http://library.vlsu.ru/> – сайт научной библиотеки ВлГУ
3. <https://office-guru.ru/excel/excel-dlya-analitika-4-tekhniki-analiza-dannyh-v-excel.html> – Excel для аналитика. 4 техники анализа данных в Excel
4. <https://histories.media/workshops/2020/04/30/analiz-dannikh-v-excel-chast/> - Анализ данных в Excel
5. <https://ranalytics.github.io/data-mining/012-R-Intro.html> – Статистическая среда R и ее использование в Data Mining
6. <https://www.retail.ru/theme/e-commerce-online-shopping/> - E-commerce, интернет-магазины
7. <http://www.scopus.com/> – библиографическая и реферативная база данных науч-

ных публикаций Scopus

8. <http://webofscience.com> – база данных международных индексов научного цитирования Web of Science

9. <http://journals.aps.org> - электронные научные информационные ресурсы American Physical Society;

10. <http://link.springer.com> - электронные научные информационные ресурсы издательства Springer

## **7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Для реализации данной дисциплины имеются специальные помещения для проведения занятий лекционного типа, занятий практического/лабораторного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы: аудитории, оснащенные мульти-медиа оборудованием, компьютерные классы с доступом в интернет, аудитории без специального оборудования.

Перечень используемого лицензионного программного обеспечения:

пакет MS-Office

Microsoft Windows

7-Zip

AcrobatReader

СПС «Консультант Плюс» (инсталлированный ресурс ВлГУ)

### Примечание.

В соответствии с нормативно-правовыми актами для инвалидов и лиц с ограниченными возможностями здоровья при необходимости тестирование может быть проведено только в письменной или устной форме, а также могут быть использованы другие материалы контроля качества знаний, предусмотренные рабочей программой дисциплины.

Рабочую программу составил Маркандеев М.М. доцент  
(ФИО, должность, подпись)

Рецензент  
(представитель работодателя) Руководитель отдела "Бизнес-процессы" ООО "Интерисофт" - Черкав  
(место работы, должность, ФИО, подпись)



Программа рассмотрена и одобрена на заседании кафедры «Коммерция и гостеприимство»  
Протокол № 1 от 25.08.21 года  
Заведующий кафедрой [подпись] О.Б.Яресь  
(ФИО, подпись)

Рабочая программа рассмотрена и одобрена  
на заседании учебно-методической комиссии направления 38.04.06. «Торговое дело»  
Протокол № 1 от 30.08.21 года  
Председатель комиссии [подпись]  
(ФИО, должность, подпись)