

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Декан

Дата подписания: 30.01.2024 17:27:03

Уникальный программный ключ:

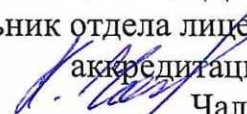
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Начальник отдела лицензирования и аккредитации



Чаленко К.Н.

« 30 » января 2021 г.

**Рабочая программа дисциплины  
Многомерные статистические методы**

по профессионально-образовательной программе направление 01.03.05 "Статистика"  
профиль 01.03.05.01 "Анализ больших данных"

Для набора 2021 года

Квалификация  
Бакалавр

КАФЕДРА

Статистики, эконометрики и оценки рисков

Семестр (<Курс>.<Семестр на курсе>)	6 (3.2)		Итого	
	Неделя			
Вид занятий	уп	рп	уп	рп
Лекции	32	32	32	32
Лабораторные	32	32	32	32
Практические	16	16	16	16
Итого ауд.	80	80	80	80
Контактная работа	80	80	80	80
Сам. работа	172	172	172	172
Часы на контроль	36	36	36	36
Итого	288	288	288	288

### ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 30.08.2021 протокол № 1.

Программу составил(и): д.э.н., профессор, Арженовский С.В.; к.э.н., доц., Трегубова А.А.

Зав. кафедрой: д.э.н., профессор Ниворожкина Л.И.

Методическим советом направления: к.э.н., доц., Кислая И.А.

**1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

1.1	Цель изучения дисциплины: овладение методологией многомерного статистического анализа, инструментальными средствами обработки данных, навыками использования современного программного обеспечения для построения многомерных моделей.
-----	--

**2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

ПК-6:Способен осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных
ОПК-3:Способен осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов

**В результате освоения дисциплины обучающийся должен:**

<b>Знать:</b>
методы математической и дескриптивной статистики для анализа данных; способы снижения размерности исследуемых многомерных признаков и отбора наиболее информативных показателей
<b>Уметь:</b>
содержательно интерпретировать результаты расчетов и обосновывать полученные при моделировании выводы, использовать современное программное обеспечение для построения многомерных моделей; критически оценивать полученные при моделировании результаты
<b>Владеть:</b>
методами многомерного анализа с использованием специализированных программных средств; инструментальными средствами обработки данных

**3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ**

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	<b>Раздел 1. Регрессионный анализ и классификация.</b>				
1.1	Тема "Первичная обработка данных". Многомерное признаковое пространство. Многомерное нормальное распределение. Методы шкалирования при обработке качественных признаков. Проблема размерности в многомерных методах исследования. Статистическое оценивание и сравнение многомерных, генеральных совокупностей. Распределение и характеристики многомерной совокупности. Многомерное нормальное распределение. Статистические оценки многомерной генеральной совокупности. /Лек/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.2	Тема "Первичная обработка данных". Многомерное признаковое пространство. Многомерное нормальное распределение. Методы шкалирования при обработке качественных признаков. Проблема размерности в многомерных методах исследования. Статистическое оценивание и сравнение многомерных, генеральных совокупностей. Распределение и характеристики многомерной совокупности. Многомерное нормальное распределение. Статистические оценки многомерной генеральной совокупности. Работа в MS Excel, EViews, Statistica. Знакомство с источниками данных: База данных Центрального банка РФ <a href="http://cbr.ru/hd_base/">http://cbr.ru/hd_base/</a> База статистических данных <a href="https://rosstat.gov.ru/databases">https://rosstat.gov.ru/databases</a> Единая межведомственная информационно-статистическая система <a href="https://www.fedstat.ru/">https://www.fedstat.ru/</a> /Лаб/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3



1.3	Тема "Первичная обработка данных". Многомерное признаковое пространство. Многомерное нормальное распределение. Методы шкалирования при обработке качественных признаков. Проблема размерности в многомерных методах исследования. Статистическое оценивание и сравнение многомерных, генеральных совокупностей. Распределение и характеристики многомерной совокупности. Многомерное нормальное распределение. Статистические оценки многомерной генеральной совокупности. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.4	Тема "Корреляционно-регрессионный анализ". Корреляционный анализ. Построение и интерпретация модели множественной линейной регрессии. Ранговая корреляция. Корреляция категоризованных переменных. Статистический анализ экспертных оценок. /Лек/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.5	Тема "Корреляционно-регрессионный анализ". Корреляционный анализ. Построение и интерпретация модели множественной линейной регрессии. Ранговая корреляция. Корреляция категоризованных переменных. Статистический анализ экспертных оценок. Работа в MS Excel, EViews, Statistica. Знакомство с источниками данных: База данных Центрального банка РФ <a href="http://cbr.ru/hd_base/">http://cbr.ru/hd_base/</a> База статистических данных <a href="https://rosstat.gov.ru/databases">https://rosstat.gov.ru/databases</a> Единая межведомственная информационно-статистическая система <a href="https://www.fedstat.ru/">https://www.fedstat.ru/</a> /Лаб/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.6	Тема "Корреляционно-регрессионный анализ". Корреляционный анализ. Построение и интерпретация модели множественной линейной регрессии. Ранговая корреляция. Корреляция категоризованных переменных. Статистический анализ экспертных оценок. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.7	Тема "Дискриминантный анализ". Построение и интерпретация модели линейного дискриминантного анализа. Пошаговый дискриминантный анализ. Оценка качества дискриминантной функции. /Лек/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.8	Тема "Дискриминантный анализ". Построение и интерпретация модели линейного дискриминантного анализа. Пошаговый дискриминантный анализ. Оценка качества дискриминантной функции. Работа в Statistica. /Лаб/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.9	Тема "Дискриминантный анализ". Построение и интерпретация модели линейного дискриминантного анализа. Пошаговый дискриминантный анализ. Оценка качества дискриминантной функции. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.10	Тема "Кластерный анализ". Непараметрический случай классификации без обучения: кластерный анализ. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Иерархические кластер-процедуры. Метод k-средних. Расщепление смесей вероятностных распределений. /Лек/	6	6	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.11	Тема "Кластерный анализ". Непараметрический случай классификации без обучения: кластерный анализ. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Иерархические кластер-процедуры. Метод k-средних. Расщепление смесей вероятностных распределений. Работа в Statistica. /Лаб/	6	6	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3

1.12	Тема "Кластерный анализ". Непараметрический случай классификации без обучения: кластерный анализ. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Иерархические кластер-процедуры. Метод k-средних. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.13	Тема "Первичная обработка данных". Многомерное признаковое пространство. Многомерное нормальное распределение. Методы шкалирования при обработке качественных признаков. Проблема размерности в многомерных методах исследования. Знакомство с источниками данных: База данных Центрального банка РФ <a href="http://cbr.ru/hd_base/">http://cbr.ru/hd_base/</a> , База статистических данных <a href="https://rosstat.gov.ru/databases">https://rosstat.gov.ru/databases</a> , Единая межведомственная информационно-статистическая система <a href="https://www.fedstat.ru/">https://www.fedstat.ru/</a> Работа с Консультант + /Ср/	6	14	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.14	Тема "Корреляционно-регрессионный анализ". Корреляционный анализ. Построение и интерпретация модели множественной линейной регрессии. Ранговая корреляция. Корреляция категоризованных переменных. Статистический анализ экспертных оценок. /Ср/	6	18	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.15	Тема "Дискриминантный анализ". Построение и интерпретация модели линейного дискриминантного анализа. Пошаговый дискриминантный анализ. Оценка качества дискриминантной функции. /Ср/	6	22	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
1.16	Тема "Кластерный анализ". Непараметрический случай классификации без обучения: кластерный анализ. Расстояние между объектами. Меры близости между объектами. Меры близости между кластерами. Иерархические кластер-процедуры. Метод k-средних. Расщепление смесей вероятностных распределений. /Ср/	6	20	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
	<b>Раздел 2. Снижение размерности. Комплексный многомерный анализ.</b>				
2.1	Тема "Снижение размерности исследуемых многомерных признаков". Метод главных компонент. Собственные векторы и собственные значения и их использование для получения матрицы весовых коэффициентов. Построение и интерпретация модели главных компонент. /Лек/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.2	Тема "Снижение размерности исследуемых многомерных признаков". Метод главных компонент. Собственные векторы и собственные значения и их использование для получения матрицы весовых коэффициентов. Построение и интерпретация модели главных компонент. Работа в Statistica. /Лаб/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.3	Тема "Снижение размерности исследуемых многомерных признаков". Метод главных компонент. Собственные векторы и собственные значения и их использование для получения матрицы весовых коэффициентов. Построение и интерпретация модели главных компонент. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.4	Тема "Факторный анализ". Модель ортогональных факторов. Определение факторных нагрузок методом главных факторов. Вращение пространства общих факторов. Статистическая оценка надежности решений методом факторного анализа. Построение сводного (интегрального) показателя качества сложной системы. /Лек/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3



2.5	Тема "Факторный анализ". Модель ортогональных факторов. Определение факторных нагрузок методом главных факторов. Вращение пространства общих факторов. Статистическая оценка надежности решений методом факторного анализа. Построение сводного (интегрального) показателя качества сложной системы. Работа в Statistica. /Лаб/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.6	Тема "Факторный анализ". Модель ортогональных факторов. Определение факторных нагрузок методом главных факторов. Вращение пространства общих факторов. Статистическая оценка надежности решений методом факторного анализа. Построение сводного (интегрального) показателя качества сложной системы. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.7	Тема "Многомерное шкалирование". Многомерное шкалирование: алгоритм и примеры. /Лек/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.8	Тема "Многомерное шкалирование". Многомерное шкалирование: алгоритм и примеры. /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.9	Тема "Комплексный многомерный анализ". Регрессия на главные компоненты/общие факторы. Кластерный анализ на главных компонентах/общих факторах. /Лек/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.10	Тема "Прикладной многомерный анализ". Решение практических задач с помощью инструментов многомерного статистического анализа (первичная обработка данных, корреляционно-регрессионный анализ, методы снижения размерности и классификации). Работа в EViews, Statistica. Знакомство с источниками данных: База данных Центрального банка РФ <a href="http://cbr.ru/hd_base/">http://cbr.ru/hd_base/</a> , База статистических данных <a href="https://rosstat.gov.ru/databases">https://rosstat.gov.ru/databases</a> , Единая межведомственная информационно-статистическая система <a href="https://www.fedstat.ru/">https://www.fedstat.ru/</a> Работа с Консультант + /Лаб/	6	6	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.11	Тема "Прикладной многомерный анализ". Решение практических задач с помощью инструментов многомерного статистического анализа (первичная обработка данных, корреляционно-регрессионный анализ, методы снижения размерности и классификации). /Пр/	6	2	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.12	Тема "Прикладной многомерный анализ". Пример решения практической задачи с помощью инструментов многомерного статистического анализа (первичная обработка данных, корреляционно-регрессионный анализ, методы снижения размерности и классификации). /Лек/	6	4	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.13	Тема "Снижение размерности исследуемых многомерных признаков". Метод главных компонент. Собственные векторы и собственные значения и их использование для получения матрицы весовых коэффициентов. Построение и интерпретация модели главных компонент. /Ср/	6	28	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.14	Тема "Факторный анализ". Модель ортогональных факторов. Определение факторных нагрузок методом главных факторов. Вращение пространства общих факторов. Статистическая оценка надежности решений методом факторного анализа. Построение сводного (интегрального) показателя качества сложной системы. /Ср/	6	24	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3

2.15	Тема "Многомерное шкалирование". Многомерное шкалирование: алгоритм /Ср/	6	26	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.16	Тема "Комплексный многомерный анализ". Регрессия на главные компоненты/общие факторы. Кластерный анализ на главных компонентах/общих факторах /Ср/	6	20	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3
2.17	/Экзамен/	6	36	ОПК-3 ПК-6	Л1.1 Л1.2 Л1.4 Л1.3Л2.4 Л2.1 Л2.7 Л2.6 Л2.2 Л2.5 Л2.3

#### 4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

#### 5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

##### 5.1. Основная литература

	Авторы,	Заглавие	Издательство, год	Колич-во
Л1.1	Ниворожкина Л. И.	Статистические методы анализа данных: учеб.	М.: РИОФ, 2016	105
Л1.2	Ниворожкина Л. И., Арженовский С. В.	Многомерные статистические методы в экономике: учеб. для вузов	М.: Дашков и К, 2008	196
Л1.3	Александровская, Ю. П.	Многомерный статистический анализ в экономике: учебное пособие	Казань: Казанский национальный исследовательский технологический университет, 2017	<a href="http://www.iprbookshop.ru/79330.html">http://www.iprbookshop.ru/79330.html</a> неограниченный доступ для зарегистрированных пользователей
Л1.4	Афанасьев В. Н., Леушина Т. В., Лебедева Т., Цыпин А. П., Афанасьев В. Н.	Эконометрика: учебник	Оренбург: Оренбургский государственный университет, 2012	<a href="https://biblioclub.ru/index.php?page=book&amp;id=260747">https://biblioclub.ru/index.php?page=book&amp;id=260747</a> неограниченный доступ для зарегистрированных пользователей

##### 5.2. Дополнительная литература

	Авторы,	Заглавие	Издательство, год	Колич-во
Л2.1	Елисеева И. И.	Эконометрика: учеб. для бакалавриата и магистратуры	М.: Юрайт, 2016	60
Л2.2		Журнал "Вопросы статистики"	,	1
Л2.3	Шорохова, И. С., Кисляк, И. В., Мариев, О. С.	Статистические методы анализа: учебное пособие	Екатеринбург: Уральский федеральный университет, ЭБС АСВ, 2015	<a href="http://www.iprbookshop.ru/65987.html">http://www.iprbookshop.ru/65987.html</a> неограниченный доступ для зарегистрированных пользователей
Л2.4	Арженовский С. В.	Применение многомерных методов анализа в оценке рисков с использованием ППП: метод. указания к лаборатор. занятиям	Ростов н/Д: Изд-во РГЭУ (РИНХ), 2015	95
Л2.5	Зехин В. А., Мхитарян В. С., Айвазян С. А.	Практикум по многомерным статистическим методам: учебное пособие	Москва: Московский государственный университет экономики, статистики и информатики, 2003	<a href="https://biblioclub.ru/index.php?page=book&amp;id=90409">https://biblioclub.ru/index.php?page=book&amp;id=90409</a> неограниченный доступ для зарегистрированных пользователей



Приложение 1

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
<b>ОПК-3: способность осознанно применять методы математической и дескриптивной статистики для анализа количественных данных, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ, содержательно интерпретировать полученные результаты, готовить статистические материалы для докладов, публикаций и других аналитических материалов</b>			
<i>Знания:</i> методов математической и дескриптивной статистики для анализа данных	Формулирует ответы на поставленные вопросы коллоквиума в части методов регрессионного анализа и классификации	Полнота и содержательность ответа; умение приводить примеры	Коллоквиум (вопросы 1-23) Экзаменационные билеты (1-10)
<i>Умения:</i> содержательно интерпретировать результаты расчетов и обосновывать полученные при моделировании выводы; использовать современное программное обеспечение для построения многомерных моделей	Решает разноуровневые задачи, анализирует и интерпретирует полученные результаты. Формирует отчет по заданию к лабораторной работе с использованием стандартных пакетов прикладных программ	Полнота и правильность решений; обоснованность обращения к базам данных; содержательность выводов и интерпретации полученных результатов. Правильность использования методов обработки данных, корреляционно-регрессионного анализа при выполнении задания к лабораторной работе, качество анализа и интерпретации полученных результатов, обоснованность выводов; качество оформления	Разноуровневые задачи (задачи 1-17) Лабораторное задание (задания 1-4) Экзаменационные билеты (1-10)
<i>Навыки:</i> владения методами многомерного анализа с использованием специализированных программных средств	Формирует отчет по заданию к лабораторной работе в части методов многомерного анализа с использованием стандартных пакетов прикладных программ	Правильность использования методов обработки данных, стандартных прикладных программ, методов многомерного анализа при выполнении задания к лабораторной работе, качество анализа и интерпретации полученных результатов, обоснованность выводов; качество оформления	Лабораторное задание (задания 3-7) Экзаменационные билеты (1-10)
<b>ПК-6: способность осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных</b>			
<i>Знания:</i> способов снижения размерности многомерных признаков и отбора наиболее информативных показателей	Формулирует ответы на поставленные вопросы коллоквиума в части методов снижения размерности многомерных признаков, комплексного многомерного анализа. Формирует отчет по заданию к лабораторной работе с использованием стандартных пакетов прикладных программ	Полнота и содержательность ответа; умение приводить примеры. Правильность использования методов обработки данных, снижения размерности многомерных признаков, комплексного многомерного анализа при выполнении задания к лабораторной работе, качество анализа и интерпретации полученных результатов, обоснованность выводов; качество оформления.	Коллоквиум (вопросы 24-38) Лабораторное задание (задания 5-7) Экзаменационные билеты (1-10)

<i>Умения:</i> критически оценивать полученные при моделировании результаты	Решает разноуровневые задачи, анализирует, оценивает и интерпретирует полученные результаты	Полнота и правильность решений; обоснованность обращения к базам данных; содержательность выводов и интерпретации полученных результатов.	Разноуровневые задачи (задачи 18-19) Экзаменационные билеты (1-10)
<i>Навыки:</i> владения инструментальными средствами обработки данных	Решает разноуровневые задачи, анализирует и интерпретирует полученные результаты в части обработки данных с помощью инструментальных средств. Формирует отчет по заданию к лабораторной работе с использованием стандартных пакетов прикладных программ	Полнота и правильность решений; обоснованность обращения к базам данных; содержательность выводов и интерпретации полученных результатов. Правильность использования методов обработки данных, корреляционно-регрессионного анализа, методов снижения размерности многомерных признаков, комплексного многомерного анализа при выполнении задания к лабораторной работе, качество анализа и интерпретации полученных результатов, обоснованность выводов; качество оформления.	Разноуровневые задачи (задачи 20-23) Лабораторное задание (задания 1-7) Экзаменационные билеты (1-10)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»)

67-83 баллов (оценка «хорошо»)

50-66 баллов (оценка «удовлетворительно»)

0-49 баллов (оценка «неудовлетворительно»)

2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Экзаменационные билеты

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Содержание и основные этапы многомерного статистического анализа.
2. Параметрический случай классификации без обучения: расщепление смесей вероятностных распределений.
3. Задача.  
Фирма изучает спрос на автомобили. Сформирована по опросам покупателей выборка по переменным: цена автомобиля, объем двигателя, расход бензина, безопасность, фирма производитель. Предложите многомерный статистический метод для сегментирования рынка автомобилей. Обоснуйте свое решение.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 2

1. Постановка задачи корреляционного анализа многомерной генеральной совокупности.



2. Непараметрический случай классификации без обучения: кластерный анализ.

Постановка задачи автоматической классификации.

3. Задача.

Получите дискриминантную функцию Фишера для следующих исходных данных:

$$X_1 = \begin{pmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{pmatrix}, X_2 = \begin{pmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{pmatrix}, \bar{x}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \bar{x}_2 = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$$

и общая ковариационная матрица

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 3

1. Корреляционный анализ количественных признаков: множественные и частные коэффициенты корреляции.

2. Кластерный анализ: расстояние между объектами и меры близости объектов друг к другу.

3. Задача.

Получите дискриминантную функцию Фишера для следующих исходных данных:

$$X_1 = \begin{pmatrix} 3 & 7 \\ 2 & 5 \\ 4 & 7 \end{pmatrix}, X_2 = \begin{pmatrix} 6 & 9 \\ 4 & 7 \\ 3 & 8 \end{pmatrix}, \bar{x}_1 = \begin{pmatrix} 2 \\ 6 \end{pmatrix}, \bar{x}_2 = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

и общая ковариационная матрица

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 4

1. Проверка статистических гипотез о параметрах многомерной нормально распределенной генеральной совокупности

2. Кластерный анализ: расстояние между классами объектов.

3. Задача.

Получены следующие результаты дискриминантного анализа

Discriminant Function Analysis Summary (Lab 3)

No. of vars in model: 3; Grouping: Var4 (2 grps)

Wilks' Lambda: .14454 approx. F(3,5)=9,8640 p= .0153

N=9	Wilks'	Partial	F-remove	p-level	Toler.	1-Toler.
Производительность труда	0.181529	0.796246	1.279471	0.309315	0.687015	0.312985
Удельный вес потерь от брака	0.153153	0.943771	0.297895	0.608664	0.877639	0.122361
Фондоотдача	0.163325	0.884994	0.649754	0.456810	0.627656	0.372345

Проинтерпретируйте результаты и сделайте соответствующие выводы.

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 5

1. Ранговая корреляция: по Спирмену, Кендаллу.

2. Кластерный анализ: оценка качества разбиения объектов на классы.

3. Задача. Получены следующие результаты дискриминантного анализа

Classification Functions: grouping: Var4 (Lab 3)

	G_1:1 P=0.44	G_2:2 P=0.56
Производительность труда	9.1345	6.4151
Удельный вес потерь от брака	6.3900	11.0694
Фондоотдача	10.5584	3.3542
Constant	-54.8703	-25.1824

Posterior Probabilities (Lab 3)

Incorrect classifications are marked with \*

Case	Observed	G_1:1	G_2:2	Case	Observed	G_1:1	G_2:2
1	G_1:1	0.999868	0.000132	7	G_2:2	0.009825	0.990175
2	G_1:1	0.999568	0.000432	8	G_2:2	0.010251	0.989749
3	G_1:1	0.999740	0.000260	9	G_2:2	0.000000	1.000000
4	G_1:1	0.999989	0.000011	10	---	0.000492	0.999508
5	G_2:2	0.000478	0.999522	11	---	0.000935	0.999065
6	G_2:2	0.000000	1.000000	12	---	1.000000	0.000000

Проинтерпретируйте результаты и сделайте соответствующие выводы.

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 6

1. Корреляция категоризованных переменных: таблицы сопряженности и меры степени тесноты статистической связи.

2. Кластерный анализ: принцип построения агломеративных иерархических процедур классификации.

3. Задача.

Получены следующие результаты факторного анализа

$X_6$  - удельный вес покупных изделий;  $X_{11}$  - среднегодовая численность ППП;  $X_{12}$  - среднегодовая стоимость ОПФ;  $X_{14}$  - фондовооруженность труда;  $X_{15}$  - оборачиваемость нормируемых оборотных средств;  $X_{17}$  - непроизводственные расходы.

Factor Loadings (Varimax normalized) (Lab 7)

Extraction: Principal components

(Marked loadings are > .700000)

Variables	Factor 1	Factor 2	Factor 3
X6	0.088887	0.002317	0.902880
X11	0.755472	-0.010037	0.395243
X12	0.957894	-0.086188	0.164689
X14	0.760477	-0.176215	-0.350881
X15	-0.219036	0.858033	-0.165881
X17	0.023075	0.888763	0.168818
Expl. Var	2.123035	1.564707	1.177664
Prp Totl	0.353839	0.260785	0.196277

Проинтерпретируйте результаты и сделайте соответствующие выводы.

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 7

$x_2$	4	-2	1	1
-------	---	----	---	---

Необходимо классифицировать объекты на две группы методом  $k$ -средних.

1. Многомерная классификация: постановка задачи, основные определения. Классификации с обучением и без обучения.
2. Кластерный анализ: последовательные кластер-процедуры, метод  $k$ -средних.
3. Задача.

По данным о 5 домохозяйствах провести компонентный анализ на основе показателей удельного веса доходов, не связанных с основной работой, в общей сумме доходов  $x_1$  и удельного веса расходов на питание  $x_2$ , если ковариационная

$$\text{матрица } \Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix}.$$

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 8

1. Многомерная классификация: оптимальная (байесовская) процедура классификации.
2. Снижение размерности многомерных признаков: метод главных компонент.
3. Задача.

Для данных по 138 индивидам выполните корреляционный анализ и сделайте выводы.

Доход	Удовлетворение работой	
	Полное неудовлетворение	Полное удовлетворение
< \$250	42	27
> \$500	7	62

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 9

1. Параметрический дискриминантный анализ в случае нормальных классов. Линейная дискриминантная функция Фишера.
2. Алгоритм вычисления главных компонент.
3. Задача.

Пусть  $X$  имеет многомерное нормальное распределение  $N_3(\mu, \Sigma)$ ,

$$\text{где } \mu^T = [1, -1, 2] \text{ и } \Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$

Какие из следующих случайных величин являются независимыми? Объясните.

- а).  $X_1$  и  $X_2$ .
- б).  $X_1$  и  $X_3$ .
- в).  $(X_1, X_3)$  и  $X_2$ .

### ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 10

1. Алгоритм дискриминантного анализа в случае двух нормальных классов.
2. Главные компоненты многомерной нормально распределенной совокупности. Главные компоненты стандартизованных переменных.
3. Задача.

Пусть переменные  $x_1$  и  $x_2$  измерены на четырех объектах А, В, С и D:

Объекты	А	В	С	Д
$x_1$	5	1	-1	3

### Критерии оценивания:

Экзаменационный билет оценивается максимально в 100 баллов:

- 84-100 баллов (оценка «отлично»)
- 67-83 баллов (оценка «хорошо»)
- 50-66 баллов (оценка «удовлетворительно»)
- 0-49 баллов (оценка «неудовлетворительно»)

Задача оценивается максимально в 50 баллов. Критерии оценивания задачи:

- Отлично – 42-50 баллов. Задача решена в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.
- Хорошо – 33,5-41 балла. Задача решена в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.
- Удовлетворительно – 25-33,4 балла. Задача решена частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты, сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.
- Неудовлетворительно – 0-24 балла. Задача не решена или решена частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

Каждый вопрос оценивается отдельно, максимально в 25 баллов. Максимальная общая оценка – 50 баллов. Критерии оценивания отдельного вопроса:

- Отлично – 21-25 баллов. Ответ на вопрос верный; продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- Хорошо 16,75-20 балла. Ответ на вопрос верный, но с отдельными погрешностями и ошибками, уверенно исправленными после дополнительных вопросов; продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- Удовлетворительно – 12,5-16,74 балла. Ответ на вопрос частично верен, продемонстрирована некоторая неточность ответов на дополнительные и наводящие вопросы.
- Неудовлетворительно – 0-12,4 балла. Ответ на вопрос не верен, продемонстрирована неуверенность и неточность ответов на дополнительные и наводящие вопросы.



## Вопросы для коллоквиума

### Раздел 1 «Регрессионный анализ и классификация»

1. В чем особенности МСА?
2. Основные этапы МСА.
3. ~~Формы представления данных, используемых в МСА.~~
4. Понятие признакового пространства. Приведите примеры.
5. Виды зависимостей исследуемых многомерными статистическими методами.
6. Кратко поясните логическую схему построения статистического критерия для проверки однородности нормальной выборочной совокупности.
7. Каковы основные характеристики многомерной случайной величины?
8. Кратко поясните особенности множественного коэффициента корреляции, частного коэффициента корреляции.
9. В чем особенности измерения степени тесноты статистической связи между категоризованными переменными?
10. В чем особенности дискриминантного анализа?
11. Как определяется качество дискриминантных функций?
12. В чем суть непараметрического дискриминантного анализа?
13. Приведите пример (графически), когда дискриминантная функция будет нелинейной.
14. Как определяется количество дискриминантных функций?
15. Суть оптимального байесовского правила классификации.
16. Какие задачи решаются с помощью кластерного анализа?
17. Какие меры сходства используются при проведении кластерного анализа?
18. Особенности параметрической классификации без обучения.
19. Какие меры расстояний между объектами используются в кластерном анализе?
20. Как оценивается качество полученного разбиения на классы?
21. Принцип "работы" иерархических процедур классификации.
22. Особенности метода Уорда.
23. Алгоритм метода  $k$ -средних.

### Раздел 2 «Снижение размерности. Комплексный многомерный анализ»

24. Какие задачи решаются с помощью компонентного анализа?
25. Как находятся главные компоненты?
26. Как интерпретируются результаты компонентного анализа?
27. В чем суть факторного анализа? Какие виды факторного анализа используются на практике?
28. Как определить достаточное число факторов для характеристики изучаемого явления или процесса?
29. Модель ортогональных факторов.
30. Метод главных факторов.
31. Вращение системы факторов.
32. В чем отличие факторного анализа от компонентного?
33. Как проверить надежность результатов факторного анализа?
34. В чем суть задачи многомерного шкалирования?
35. Как решается задача метрического шкалирования по Торгерсону?
36. В чем отличие метрического шкалирования от неметрического?

37. Как строится матрица различий объектов?
38. Каков алгоритм решения задачи неметрического шкалирования?

### Критерии оценивания:

Каждый вопрос оценивается отдельно, максимально в 1 балл. Максимальная общая оценка – 38 баллов. Критерии оценивания отдельного вопроса:

- 0,84-1 балл. Ответ на вопрос верный, продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- 0,67-0,83 балла. Ответ на вопрос верный, но с отдельными погрешностями и ошибками, уверенно исправленными после дополнительных вопросов; продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- 0,5-0,66 балла. Ответ на вопрос частично верен, продемонстрирована некоторая неточность ответов на дополнительные и наводящие вопросы.
- 0-0,49 балла. Ответ на вопрос не верен, продемонстрирована неуверенность и неточность ответов на дополнительные и наводящие вопросы.

## Комплект разноуровневых задач

### 1 Задачи репродуктивного уровня

1. Имеются данные о нескольких индивидах

№ п п	Число лет образования	Возраст	Логарифм доходов, руб./месяц
1	10	67	6,55
2	15	23	8,16
3	15	42	7,24
4	10	60	6,40
5	13	29	7,90
6	10	27	9,29
7	10	59	6,68
8	17	30	7,31
9	13	20	0,00
10	13	55	7,78
11	18	60	8,07

Вычислите вектор средних значений и ковариационную матрицу. Прокомментируйте результаты.

2. Для данных предыдущей задачи в проверьте гипотезу о равенстве вектора средних, вычисленного по первым 5 наблюдениям, вектору средних, вычисленного по последним 6 наблюдениям.
3. Совместная плотность двумерной случайной величины  $(X_1, X_2)$  задана формулой

$$f(X_1, X_2) = \frac{1}{1,6\pi} e^{-\frac{1}{128}[(X_1 - 2)^2 - 1,2(X_1 - 2)(X_2 + 3) + (X_2 + 3)^2]}. \text{ Найдите } \bar{X}_1, \bar{X}_2, \sigma_{11}, \sigma_{22} \text{ и } r_{12}.$$

4. Пусть  $\mathbf{X}$  имеет нормальное распределение  $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  с  $\boldsymbol{\mu} = [-3, 1, 4]$  и  $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ .

Какие из следующих случайных величин являются независимыми? Поясните.

- а)  $X_1$  и  $X_2$   
 б)  $X_2$  и  $X_3$   
 в)  $(X_1, X_2)$  и  $X_3$   
 г)  $\frac{X_1 + X_2}{2}$  и  $X_3$ .

5. Вычислите статистику Хоттелинга для проверки гипотезы  $H_0: \mu = [7, 11]$ , используя

данные  $X^T = \begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}$ . Проверьте гипотезу на 5% уровне значимости.

6. Постройте 90% доверительную область для  $\mu$  из предыдущей задачи.

7. Оцените значимость различий двух рынков сбыта бытовой техники. На первом рынке (число наблюдений – 5) средний уровень цены реализации составил 5 тыс. руб., а экспертная оценка качества обслуживания – 3,4 балла; на втором рынке (число наблюдений – 7) соответственно 7 тыс. руб. и 4,3 балла. Объединенная ковариационная матрица имеет вид:  $S = \begin{bmatrix} 9,3 & 0,26 \\ 0,26 & 2,0 \end{bmatrix}$ .

8. Имеется корреляционная матрица признаков  $X_1$  – доход за месяц,  $X_2$  – число лет образования,  $X_3$  – процент времени, который занимает основная работа  $R = \begin{bmatrix} 1 & 0,24 & 0,01 \\ 0,24 & 1 & 0,09 \\ 0,01 & 0,09 & 1 \end{bmatrix}$ , вычисленная по массиву данных msm.sta (число наблюдений

2874). Рассчитайте парные и множественные (в том числе частные) коэффициенты корреляции.

Принять уровень надежности – 95%.

9. Два эксперта проранжировали 10 предложенных им проектов по степени эффективности:  $X_1 = (1; 2; 4; 6; 7; 3; 5; 8; 9; 10)$ ,  $X_2 = (2; 3; 1; 4; 6; 5; 9; 7; 8; 10)$ . Оцените степень согласованности мнений экспертов, вычислив ранговые коэффициенты корреляции Спирмена и Кендалла.

10. Имеются результаты опроса группы из 10 экспертов по трем вопросам социально-экономической политики:

Значение экспертной оценки по вопросам	Порядковый номер эксперта									
	1	2	3	4	5	6	7	8	9	10
1	0,35	0,52	0,48	0,64	0,69	0,2	0,36	0,65	0,78	0,54
2	30	35	38	40	45	31	33	46	42	37
3	2,5	3,1	3	3,6	4,5	2,8	3,7	4,2	4,7	2,9

Требуется оценить согласованность мнений экспертов с помощью коэффициента конкордации и проверить его значимость на уровне 5%. Сделать вывод.

11. Проанализируйте связь между полом работника и характером труда в сезонных отраслях:

Пол	Численность занятых в отраслях		
	Сезонных	Не сезонных	Всего
Мужчины	187	265	452
Женщины	307	272	579
Всего	494	537	1031

12. Оценка студентами профессиональных качеств преподавателей представлена в таблице ниже. Рассчитайте коэффициент взаимной сопряженности Пирсона, Крамера.

Критерии оценки качества преподавателей	Оценка				
	Высокая	Средняя	Низкая	Затрудняюсь ответить	Итого
Знание предмета	62	26	1	11	100
Умение обучать	21	61	8	10	100
Восприимчивость к новому	20	51	10	19	100
Способность к саморазвитию	25	51	10	14	100
Итого	128	189	29	54	400

13. Деятельность предприятий региона характеризуется четырьмя показателями. При проведении компонентного анализа получены собственные числа: 1,2; 0,8; 0,4 и одно из них оказалось пропущенным. Чему оно равно? Чему равен относительный вклад двух первых компонент?

14. Для ковариационной матрицы трех переменных постройте факторную модель с одним фактором: найдите нагрузки фактора, значения общностей и специфических дисперсий:

$$\Sigma = \begin{pmatrix} 1 & 0,4 & 0,9 \\ 0,4 & 1 & 0,7 \\ 0,9 & 0,7 & 1 \end{pmatrix}$$

15. По известной матрице факторных нагрузок  $\begin{pmatrix} 0,76 & 0,42 \\ 0,45 & 0,21 \\ 0,65 & 0,37 \\ 0,38 & 0,19 \end{pmatrix}$  воспроизведите матрицу парных корреляций.

16. Имеется матрица факторных нагрузок  $L = \begin{pmatrix} 0,86 & -0,25 \\ 0,61 & -0,31 \\ 0,44 & 0,51 \\ -0,47 & -0,28 \\ 0,38 & 0,08 \end{pmatrix}$  и матрица вращения

$$T = \begin{pmatrix} 0,574 & 0,819 \\ -0,819 & 0,574 \end{pmatrix}$$

Матрица  $T$  – ортогонального или косоугольного вращения?

Вращение осуществляется по часовой или против часовой стрелке? Какой принят угол вращения?

17. Дана матрица различий по восьми странам. Вычислить матрицу скалярных произведений.

№ п п	Ангола	Аргентина	Австралия	Китай	Куба	Япония	США	Зимбабве
Ангола	0,00	1,41	1,00	1,00	1,41	1,41	1,73	0,71
Аргентина	1,41	0,00	1,00	1,73	1,41	1,41	1,00	1,41
Австралия	1,00	1,00	0,00	1,41	1,73	1,00	1,41	1,00
Китай	1,00	1,73	1,41	0,00	1,00	1,00	1,41	1,00
Куба	1,41	1,41	1,73	1,00	0,00	1,41	1,00	1,41
Япония	1,41	1,41	1,00	1,00	1,41	0,00	1,00	1,41
США	1,73	1,00	1,41	1,41	1,00	1,00	0,00	1,73
Зимбабве	0,71	1,41	1,00	1,00	1,41	1,41	1,73	0,00



## 2 Задачи реконструктивного уровня

18. Имеется два набора данных  $X_1 = \begin{pmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{pmatrix}$  и  $X_2 = \begin{pmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{pmatrix}$ , для которых  $\bar{x}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$ ,

$\bar{x}_2 = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$  и  $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ . Вычислите линейную дискриминантную функцию.

Классифицируйте наблюдение  $x_0 = (2 \ 7)$ .

19. По данным о 5 домохозяйствах провести компонентный анализ на основе показателей удельного веса доходов, не связанных с основной работой, в общей сумме доходов  $x_1$  и удельного веса расходов на питание  $x_2$

№ п.п.	$x_1$	$x_2$
1	0,23	0,40
2	0,24	0,26
3	0,19	0,40
4	0,17	0,50
5	0,23	0,40

## 3 Задачи творческого уровня

20. Деятельность двенадцати машиностроительных предприятий характеризуется показателями рентабельности ( $x_1$ , %) и производительности труда ( $x_2$ , тыс. руб./чел.). Первые 4 предприятия имеют высокий уровень организации управления, а следующие 5 предприятий – низкий. Требуется обобщить и реализовать подходящий метод классификации последних трех предприятий.

№ п.п.	Группы предприятий	$x_1$	$x_2$
1	Высокий	23,4	9,1
2		19,1	6,6
3		17,5	5,3
4		17,2	10,0
5	Низкий	5,4	4,3
6		6,6	5,5
7		8,0	5,7
8		9,7	5,5
9		9,1	6,6
10	Подлежат классификации	9,9	7,4
11		14,2	9,4
12		12,9	6,7

21. По 15 фирмам имеются следующие данные

№ п.п.	Фондовооруженность труда млн. руб./чел.	Фондоотдача основных фондов, руб./руб.	Удельный вес рабочих в составе персонала
1	4,82	1,67	0,46
2	3,85	1,78	0,72
3	4,75	1,23	0,67
4	5,35	1,32	0,71
5	8,7	0,75	0,66
6	7,3	1,15	0,72
7	6,4	1,26	0,7
8	5,9	1,43	0,75
9	6	1,28	0,63
10	8,95	0,95	0,76
11	7,41	1,18	0,69

№ п.п.	Фондовооруженность труда млн. руб./чел.	Фондоотдача основных фондов, руб./руб.	Удельный вес рабочих в составе персонала
12	4,71	1,9	0,71
13	5,03	1,81	0,72
14	6,94	1,29	0,73
15	7,95	0,98	0,7

Используя алгоритм кластерного анализа, сформируйте из первых десяти наблюдений две обучающие выборки. На основании полученных выборок проведите классификацию пяти оставшихся фирм. Дайте экономическую интерпретацию результатов дискриминантного анализа.

22. Шесть домохозяйств характеризуются показателями  $x_1$  – потребление фруктов (кг/месяц) и  $x_2$  – потребление молока (л/месяц):

№ п.п.	1	2	3	4	5	6
$x_1$	21,4	16,5	9,7	18,2	6,6	8,0
$x_2$	8,1	4,2	5,5	9,4	7,5	5,7

Требуется: с помощью иерархического агломеративного алгоритма провести классификацию этих предприятий и построить дендрограмму:

1) при использовании обычной евклидовой метрики – методом а) ближайшего соседа, б) дальнего соседа, в) центра тяжести, г) средней связи;

2) при использовании взвешенной евклидовой метрики (с весами 0,2 и 0,8) методом ближайшего соседа.

23. Имеются данные по пяти социально-экономическим параметрам 12 населенных пунктов. Проведите компонентный анализ по исходным данным, дайте интерпретацию полученным результатам.

№ п.п.	Численность населения	Образование (число лет обучения)	Общее число занятых	Число занятых в сфере услуг	Средняя стоимость жилья (долларов)
1	5700	12,8	2500	270	25000
2	1000	10,9	600	10	10000
3	3400	8,8	1000	10	9000
4	3800	13,6	1700	140	25000
5	4000	12,8	1600	140	25000
6	8200	8,3	2600	60	12000
7	1200	11,4	400	10	16000
8	9100	11,5	3300	60	14000
9	9900	12,5	3400	180	18000
10	9600	13,7	3600	390	25000
11	9600	9,6	3300	80	12000
12	9400	11,4	4000	100	13000

### Критерии оценивания:

Каждая задача **репродуктивного уровня** оценивается отдельно, максимально в 1 балл. Максимальная общая оценка – 17 баллов. Критерии оценивания задачи:

- 0,84-1 баллов. Задача решена в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.
- 0,67-0,83 балла. Задача решена в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.

## Лабораторные задания

- 0,5-0,66 балла. Задача решена частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты, сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.
- 0-0,49 балла. Задача не решена или решена частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

Каждая задача **реконструктивного уровня** оценивается отдельно, максимально в 2 балла. Максимальная общая оценка – 4 балла. Критерии оценивания задачи:

- 1,68-2 баллов. Задача решена в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.
- 1,34-1,67 балла. Задача решена в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.
- 1-1,33 балла. Задача решена частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты, сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.
- 0-0,9 балла. Задача не решена или решена частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

Каждая задача **творческого уровня** оценивается отдельно, максимально в 5 баллов. Максимальная общая оценка – 20 баллов. Критерии оценивания задачи:

- 4,2-5 баллов. Задача решена в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.
- 3,35-4,1 балла. Задача решена в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.
- 2,5-3,34 балла. Задача решена частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты, сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.
- 0-2,4 балла. Задача не решена или решена частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

## Раздел 1 «Регрессионный анализ и классификация»

### Лабораторное задание 1. «Первичная обработка данных»

Данные об издержках на транспортировку продуктов питания 10-ти фирм, занимающихся снабжением, представлены в таблице. Необходимо построить 95% доверительные интервалы для средних значений трех имеющихся признаков в предположении, что они имеют нормальное распределение, а также доверительную область для первых двух признаков.

№ п.п.	Затраты топлива, л, $X_1$	Затраты на ремонт, у.е., $X_2$	Капитал фирмы, тыс. у.е., $X_3$
1	16,44	12,43	11,23
2	7,19	2,70	3,92
3	9,92	1,35	9,75
4	4,24	5,78	7,78
5	11,20	5,05	10,67
6	14,25	5,78	9,88
7	13,50	10,98	10,60
8	13,32	14,27	9,45
9	29,11	15,09	3,28
10	12,68	7,61	10,23

### Лабораторное задание 2. «Корреляционно-регрессионный анализ»

Запустите приложение *Statistica*. Откройте файл *Lab1.sta* воспользовавшись меню *File\Open*. В окне переменных вы увидите четыре переменные. Файл содержит подвыборку 4794 наблюдений из массива *msm.sta* по индивидам. Описание переменных: *ln* – логарифм заработной платы, *edu* – число лет образования, *expr* – опыт работы, *expr2* – квадрат переменной *expr*. Просмотрите дескриптивные статистики переменных в выборке. Сделайте выводы об эмпирическом распределении каждой переменной. Постройте корреляционную матрицу переменных. Постройте уравнение линейной множественной регрессии переменной заработной платы *ln* от переменных *edu*, *expr* и *expr2*. Сделайте выводы по результатам всех расчетов.

### Лабораторное задание 3. «Дискриминантный анализ»

В файле *firm.sta* имеются данные по 12 предприятиям, характеризующимся тремя экономическими показателями: *labor* – производительность труда, *defect* – удельный вес потерь от брака (%) и *fund* – фондоотдача активной части основных производственных фондов. Из этих предприятий выделены две обучающие выборки (переменная *firm*), первая из которых включает 4 предприятия группы А, а вторая 5 предприятий группы В. Требуется классифицировать в одну из групп А или В оставшиеся три предприятия.

### Лабораторное задание 4. «Кластерный анализ»

Запустите приложение *Statistica*. Откройте файл *Lab3.sta* воспользовавшись меню *File\Open*. Файл содержит подвыборку из массива *msm.sta* по индивидам. Описание переменных: *ln* – логарифм заработной платы, *edu* – число лет образования, *nhh* – число членов домохозяйства, *dd* – доля доходов главы домохозяйства в семейном бюджете, *age* – возраст, *pm* – процент заработков, который дает основная работа. Необходимо классифицировать наблюдения.



## Раздел 2 «Снижение размерности. Комплексный многомерный анализ»

### Лабораторное задание 5. «Снижение размерности исследуемых многомерных признаков»

Запустите приложение Statistica. Откройте файл Lab3.sta воспользовавшись меню File\Open. Файл содержит подвыборку из массива msm.sta по индивидам. Описание переменных: lw – логарифм заработной платы, edu – число лет образования, nhh – число членов домохозяйства, dd – доля доходов в семейном бюджете, age – возраст, pm – процент заработков, который дает основная работа. Необходимо провести компонентный анализ данных, а затем классифицировать наблюдения.

### Лабораторное задание 6. «Факторный анализ»

Запустите приложение Statistica. Рабочий файл данных тот же, что и в предыдущей лабораторной работе: Lab3.sta. Выполните факторный анализ по имеющимся выборочным данным.

### Лабораторное задание 7. «Прикладной многомерный анализ»

Деятельность предприятий характеризуется следующими показателями

№ п.п.	Трудоемкость единицы продукции, $x_1$	Удельный вес покупных изделий, $x_2$	Коэффициент сменности оборудования, $x_3$	Индекс снижения себестоимости продукции, $q$
1	0,51	0,20	1,47	21,9
2	0,36	0,64	1,27	48,4
3	0,23	0,42	1,51	173,5
4	0,26	0,27	1,46	74,1
5	0,27	0,37	1,27	68,6
6	0,29	0,38	1,43	60,8
7	0,01	0,35	1,50	355,6
8	0,02	0,42	1,35	264,8
9	0,18	0,32	1,41	526,6
10	0,25	0,33	1,47	118,6

Приняв за результативный признак  $q$ , построить уравнение регрессии на главные компоненты, наиболее тесно связанные с  $q$ . Дать экономическую интерпретацию результатов.

### Критерии оценки:

Каждое лабораторное задание оценивается отдельно, максимально в 3 балла.

Максимальная общая оценка – 21 балл. Критерии оценивания:

- 2,52-3 баллов. Задание решено в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.
- 2,01-2,51 балла. Задание решено в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.
- 1,5-2 балла. Задание решено частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты,

сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.

- 0-1,4 балла. Задание не решено или решено частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

## 3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

**Текущий контроль** успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

**Промежуточная аттестация** проводится в форме экзамена. Экзамен проводится по расписанию промежуточной аттестации в письменном виде. Количество теоретических вопросов в экзаменационном задании – 2, количество задач – 1. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в экзаменационную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Учебным планом предусмотрены следующие виды занятий:

- лекции,
- практические занятия;
- лабораторные занятия.

В ходе лекционных занятий рассматриваются основные теоретические положения и понятия, методы многомерного анализа данных, даются рекомендации для самостоятельной работы и подготовке к практическим и лабораторным занятиям.

В ходе практических и лабораторных занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки многомерного статистического анализа больших массивов данных в пакетах прикладных программ, а также самостоятельной работы и работы в коллективе.

При подготовке к практическим и лабораторным занятиям каждый студент должен:

- 1) изучить рекомендованную учебную литературу;
- 2) изучить конспекты лекций;
- 3) подготовить ответы на все вопросы по изучаемой теме.

В процессе подготовки к практическим и лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях, практических и лабораторных занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом устного опроса. В ходе самостоятельной работы каждый студент обязан прочитать основную и, по возможности, дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников. Выделить непонятные термины, найти их значение в энциклопедических словарях.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.

**Методические рекомендации по выполнению лабораторных заданий**

*Лабораторное задание 1. «Первичная обработка данных»*

Данные об издержках на транспортировку продуктов питания 10-ти фирм, занимающихся снабжением, представлены в таблице. Необходимо построить 95% доверительные интервалы для средних значений трех имеющихся признаков в предположении, что они имеют нормальное распределение, а также доверительную область для первых двух признаков.

Вычислим вектор средних значений и ковариационную матрицу для трехмерного нормального распределения.

№ п.п.	Затраты топлива, л, $X_1$	Затраты на ремонт, у.е., $X_2$	Капитал фирмы, тыс. у.е., $X_3$
1	16,44	12,43	11,23
2	7,19	2,70	3,92
3	9,92	1,35	9,75
4	4,24	5,78	7,78
5	11,20	5,05	10,67
6	14,25	5,78	9,88
7	13,50	10,98	10,60
8	13,32	14,27	9,45
9	29,11	15,09	3,28
10	12,68	7,61	10,23

Применяя формулу  $\bar{X}_l = \frac{1}{n} \sum_{i=1}^n X_{il}, l = 1, 2, 3$ , получим для, например,  $X_1$ :  $\bar{X}_1 = (16,44+7,19+9,92+4,24+...+12,68)/10=13,185$ . Аналогично получаем  $\bar{X}_2=8,104$ ,  $\bar{X}_3=8,679$  и вектор средних имеет вид:  $\bar{X} = [13,185, 8,104, 8,679]$ .

Элементы ковариационной матрицы вычисляем по формуле  $s_{ij} = \frac{1}{10} \sum_{l=1}^{10} (X_{il} - \bar{X}_l)(X_{jl} - \bar{X}_j), l, j = 1, 2, 3$ . Например, для  $i=j=1$ :  $s_{11} = [(16,44-13,185)^2 + (7,19-13,185)^2 + ... + (12,68-13,185)^2]/10 = 44,029$  и далее, для  $i=1, j=2$ :  $s_{12} = [(16,44-13,185)(12,43-8,104) + (7,19-13,185)(2,7-8,104) + ... + (12,68-13,185)(7,61-8,104)] / 10 = 20,615$  и т.д. Ковариационная матрица имеет вид:  $S = \begin{bmatrix} 44,029 & 20,615 & -4,735 \\ 20,615 & 23,222 & -0,547 \\ -4,735 & -0,547 & 8,039 \end{bmatrix}$  или несмещенная оценка ковариационной матрицы:

$$\hat{S} = \frac{10}{10-1} \begin{bmatrix} 44,029 & 20,615 & -4,735 \\ 20,615 & 23,222 & -0,547 \\ -4,735 & -0,547 & 8,039 \end{bmatrix} = \begin{bmatrix} 48,921 & 22,905 & -5,261 \\ 22,905 & 25,803 & -0,608 \\ -5,261 & -0,608 & 8,933 \end{bmatrix}$$

Вычислим  $T_{\alpha;(p,n-p)}^2 = \frac{p(n-1)}{n-p} F_{\alpha;(p,n-p)} = \frac{3(10-1)}{10-3} F_{0,05;(3,7)} = 3,857 \cdot 4,347 = 16,767$ .

Тогда получим доверительные интервалы для каждой из средних:

$$13,185 - \sqrt{16,767} \sqrt{\frac{48,921}{10}} \leq \mu_1 \leq 13,185 + \sqrt{16,767} \sqrt{\frac{48,921}{10}} \text{ и } 4,128 \leq \mu_1 \leq 22,242,$$

$$8,104 - \sqrt{16,767} \sqrt{\frac{25,803}{10}} \leq \mu_2 \leq 8,104 + \sqrt{16,767} \sqrt{\frac{25,803}{10}} \text{ и } 1,527 \leq \mu_2 \leq 14,681,$$

$$8,679 - \sqrt{16,767} \sqrt{\frac{8,933}{10}} \leq \mu_3 \leq 8,679 + \sqrt{16,767} \sqrt{\frac{8,933}{10}} \text{ и } 4,809 \leq \mu_3 \leq 12,549.$$



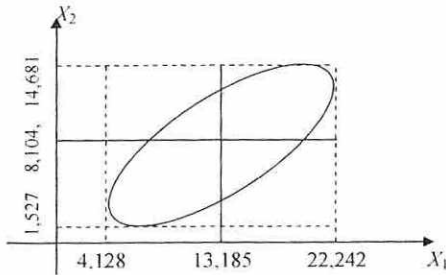
Для первых двух признаков доверительная область строится с учетом формулы

$$n(\bar{X} - \mu_0)^T S^{-1} (\bar{X} - \mu_0) = \frac{p(n-1)}{n-p} F_{\alpha, (p, n-p)}. \text{ Имеем:}$$

$$10 \begin{bmatrix} 13,185 - \mu_1, 8,104 - \mu_2 \end{bmatrix} \begin{bmatrix} 48,921 & 22,905 \\ 22,905 & 25,803 \end{bmatrix}^{-1} \begin{bmatrix} 13,185 - \mu_1 \\ 8,104 - \mu_2 \end{bmatrix} =$$

$$= 0,35(13,185 - \mu_1)^2 + 0,66(8,104 - \mu_2)^2 - 2 \cdot 0,31(13,185 - \mu_1)(8,104 - \mu_2) \leq 16,767.$$

Получившаяся доверительная область показана на рис. в виде эллипса.



С целью оценки воздействия состояния окружающей среды на здоровье населения собраны данные\* по двум Федеральным округам. Необходимо проверить при  $\alpha = 0,05$  существенность различий двух округов по выбранным двум показателям.

№ региона	Северо-Западный федеральн. округ		№ региона	Центральный федеральн. округ	
	Число умерших на 1000 чел. населения	Заболееваемость на 1000 чел. населения новорожденными		Число умерших на 1000 чел. населения	Заболееваемость на 1000 чел. населения новорожденными
1	16,6	9,6	1	16,1	12,3
2	12,5	7,6	2	17,6	8,5
3	15,3	8,1	3	19,2	10
4	17,1	7,6	4	18,2	8
5	16,3	7,3	5	20,2	10,9
6	20,1	6,4	6	18,1	7,5
7	11,6	10,3	7	19,2	8,2
8	20,9	8,7	8	18,2	6,8
9	22,5	7,4	9	17	10
10	16,4	8,6	10	18,1	9,3
11			11	17,7	10,9
12			12	19,7	9,8
			13	19,9	6,7
			14	18,2	8,9
			15	21,9	8
			16	21,5	9,2
			17	19,5	11,3
			18	15,6	10,3
$\bar{X}$	16,93	8,16	$\bar{X}$	18,66	9,26

1. Определим векторы средних и ковариационные матрицы.

$$\bar{X}_1 = [16,93, 8,16]; \bar{X}_2 = [18,66, 9,26];$$

$$S_1 = \begin{bmatrix} 13,38 & -1,86 \\ -1,86 & 1,50 \end{bmatrix}, S_2 = \begin{bmatrix} 2,93 & -0,76 \\ -0,76 & 2,62 \end{bmatrix}.$$

Объединенная ковариационная матрица рассчитывается так:

$$S_{12} = \frac{1}{10+18-2} ((10-1)S_1 + (18-1)S_2) = \begin{bmatrix} 6,54 & -1,14 \\ -1,14 & 2,24 \end{bmatrix}.$$

$$\text{Найдем обратную матрицу } S_{12}^{-1} = \begin{bmatrix} 0,17 & 0,09 \\ 0,09 & 0,49 \end{bmatrix}.$$

2. Рассчитаем фактическое значение критерия Хотеллинга.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)^T S_{12}^{-1} (\bar{X}_1 - \bar{X}_2) =$$

$$= \frac{10 \cdot 18}{10+18} [16,93 - 18,66, 8,16 - 9,26]^T \begin{bmatrix} 0,17 & 0,09 \\ 0,09 & 0,49 \end{bmatrix} [16,93 - 18,66, 8,16 - 9,26] = 1,42.$$

3. Найдем критическое значение критерия Хотеллинга и сравним с фактическим

$$T_{кр}^2 = \frac{p(n_X + n_Y - 2)}{n_X + n_Y - p - 1} F_{\alpha, (p, n_X + n_Y - p - 1)} = \frac{2(10+18-2)}{10+18-2-1} F_{0,05, (2, 25)} = 7,04.$$

Поскольку критическое значение больше фактического, то гипотеза о равенстве векторов средних значений признаков для двух округов не может быть отвергнута.

4. Проверим с помощью критерия Бартлетта равенство ковариационных матриц двух выборок

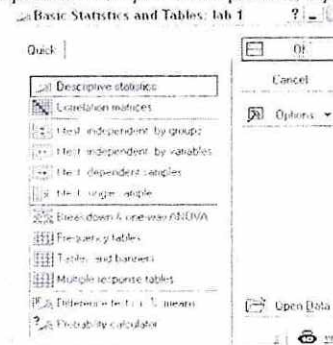
$$\text{Рассчитаем } b = 1 - \left( \frac{1}{n_X - 1} + \frac{1}{n_Y - 1} - \frac{1}{n_X + n_Y - 2} \right) \frac{2p^2 + 3p - 1}{6(p+1)} = 0,905.$$

$$a = (n_X + n_Y - 2) \ln \det S_{XY} - [(n_X - 1) \ln \det S_X + (n_Y - 1) \ln \det S_Y] = 8,835.$$

$W = ab = 7,99$ . По таблице находим  $\chi_{0,05,3}^2 = 7,81$ . Таким образом, фактическое значение критерия Бартлетта больше табличного и гипотеза о равенстве ковариационных матриц и, следовательно, однородности двух федеральных округов по выбранным двум признакам, отвергается с надежностью 95%.

### Лабораторное задание 2. «Корреляционно-регрессионный анализ»

1. Запустите приложение *Statistica*. Откройте файл *Lab1.sta* воспользовавшись меню File\Open. В окне переменных вы увидите четыре переменные. Файл содержит подвыборку 4794 наблюдений из массива *tsm.sta* по индивидам. Описание переменных: *lw* – логарифм заработной платы, *edu* – число лет образования, *expr* – опыт работы, *expr2* – квадрат переменной *expr*.

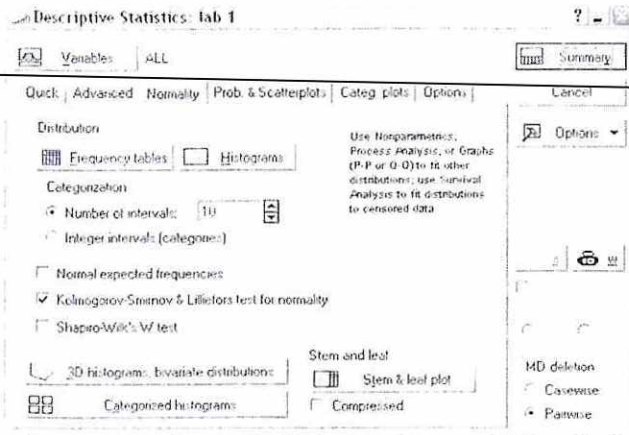


2. Просмотрите дескриптивные статистики переменных в выборке: Statistics\Basic Statistics and Tables\Descriptive statistics. Нажав вкладку Variables задайте переменные выбрав все переменные. Воспользовавшись вкладкой Advanced поставьте флажки для вывода медианы (Median), моды (Mode), дисперсии (Variance), эксцесса (Kurtosis) и асимметрии (Skewness). Нажмите кнопку Summary и получите таблицу результатов. Сделайте выводы.

Вернитесь к диалоговому окну Descriptive statistics и, выбрав вкладку Normality, нажмите кнопку Histograms. Просмотрите полученные диаграммы распределения для каждой из переменных, а также сравните их с кривой нормального распределения. Вверху каждой гистограммы приведено значение статистики Колмогорова-Смирнова для проверки соответствия эмпирического распределения нормальному закону распределения. Сделайте выводы об эмпирическом распределении каждой переменной.

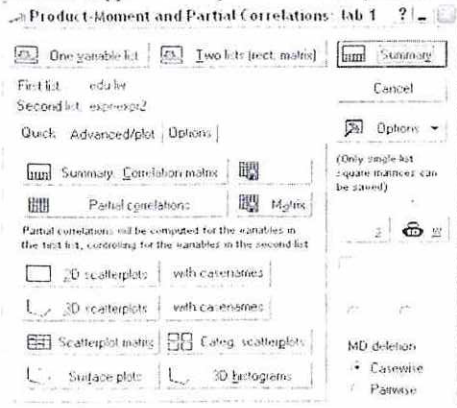
\* Данные за 2001 г. по: Регионы России. М.: Госкомстат, 2002.

3. Постройте корреляционную матрицу переменных. Для этого, находясь в меню окна **Basic Statistics and Tables**, выберите опцию **Correlation matrices**.



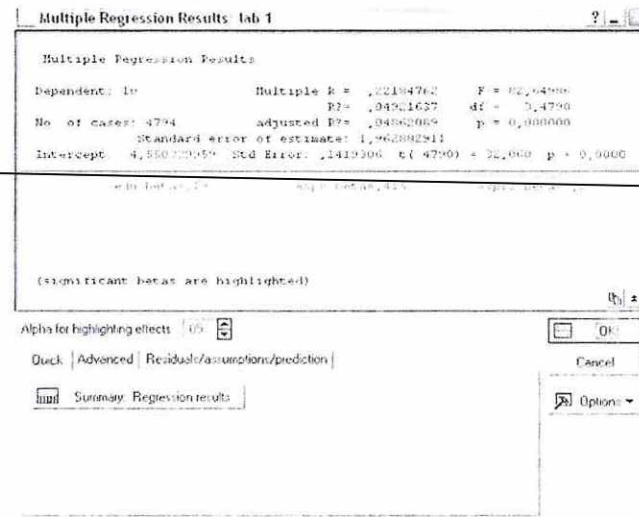
Задайте, нажав кнопку **One variable list** список переменных корреляционной матрицы и кликните **Summary**. В корреляционной матрице значимые (по умолчанию на 5% уровне) коэффициенты будут выделены красным цветом. Прокомментируйте результаты.

Также можно рассчитать частные коэффициенты корреляции. Например, для расчета частного коэффициента корреляции между переменными *edu* и *lw* при фиксированном значении переменных *expr* и *expr2* необходимо с помощью кнопки **Two lists** задать первый лист (слева) из переменных *edu* и *lw* и затем второй лист (справа) из переменных *expr* и *expr2*. Затем в окне **Product-Moment and Partial Correlation** необходимо выбрать вкладку **Advanced/plot** и в появившемся окне – опцию **Partial correlations**. Получим частный коэффициент корреляции. Аналогично получите остальные частные коэффициенты корреляции. Сравните их с парными коэффициентами корреляции и сделайте выводы.



4. Постройте уравнение линейной множественной регрессии переменной заработной платы *lw* от переменных *edu*, *expr* и *expr2*.

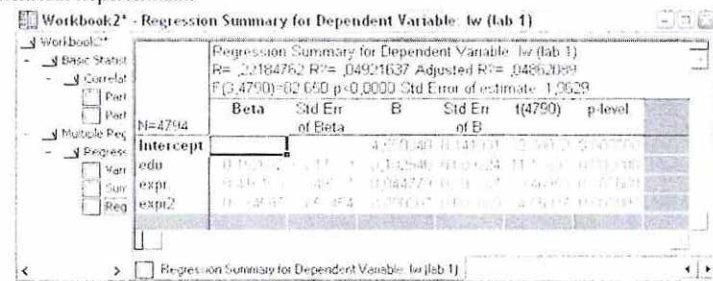
Для этого откройте меню **Statistics\Multiple Regression** и, выбрав опцию **Variables**, укажите в появившемся окне слева зависимую переменную (*lw*), в правом окне независимые переменные (*edu*, *expr* и *expr2*). Нажмите **OK**.



Появится окно результатов регрессии **Multiple Regression Results**. В окне результатов имеется значение множественного коэффициента корреляции зависимой переменной с независимыми, значение коэффициента детерминации, значения других статистик, относящихся к уравнению регрессии.

Для просмотра коэффициентов регрессии и относящихся к ним статистик нажмите в окне результатов на кнопку **Summary: Regression results**. Коэффициенты регрессии находятся в столбце с именем B, затем следует столбец стандартных ошибок коэффициентов и значения соответствующих t-статистик. Значимые коэффициенты регрессии выделены красным цветом. Вернитесь в окно **Multiple Regression Results** и выберите вкладку **Advanced** и затем опцию **ANOVA**. В появившемся окне результатов представлена таблица дисперсионного анализа для уравнения регрессии. В случае значимости уравнения в целом значения F-критерия будут выделены красным цветом.

Далее в окне **Multiple Regression Results** на вкладке **ANOVA** выберите опцию **Partial correlations**. В появившемся окне представлены частные коэффициенты корреляции между зависимой и каждой из независимых переменных.



Самостоятельно познакомьтесь с возможностями анализа остатков регрессии (окно **Multiple Regression Results**, вкладка **Residuals**, опция **Perform residual analysis**).

Сделайте выводы по результатам всех расчетов, выполненных в этом пункте задания.

5. Вы можете сохранить все результаты, запомненные программой в окне рабочей книги (Workbook), в файл в своем рабочем каталоге.

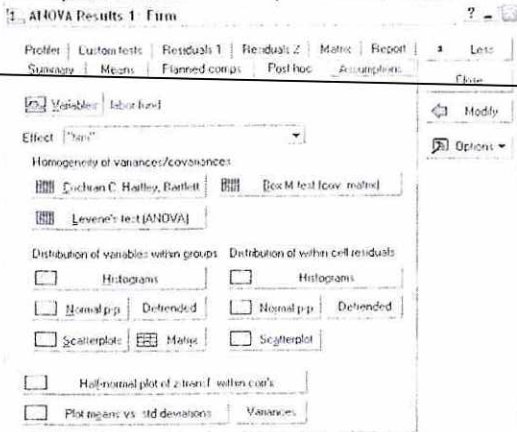
**Лабораторное задание 3. «Дискриминантный анализ»**

1. В файле *firm.sta* имеются данные по 12 предприятиям, характеризуемых тремя экономическими показателями: *labor* – производительность труда, *defect* – удельный вес потерь от брака (%) и *fund* – фондоотдача активной части основных производственных фондов. Из этих предприятий выделены две обучающие выборки (переменная *firm*), первая из которых включает 4 предприятия



группы А, а вторая 5 предприятий группы В. Требуется классифицировать в одну из групп А или В оставшиеся три предприятия.

2. Перед выполнением дискриминантного анализа необходимо убедиться



в том, что переменные характеризующие предприятия являются нормально распределенными и дисперсии и ковариации этих переменных внутри групп однородны. Для этого используется дисперсионный анализ. Необходимые опции реализованы в модуле ANOVA.

Выполнив опцию Statistics\ANOVA и выбрав One-way ANOVA, задайте лист зависимых (dependent) переменных labor, defect, fund и независимую переменную (factor) firm. Нажав ОК, выберите внизу появившегося окна опцию More results и затем вкладку Assumptions.

Затем, выбрав один из тестов в группе Homogeneity of variances/covariances (например, М-тест Бокса) путем нажатия соответствующей кнопки, получим результаты, которые убеждают нас в однородности дисперсий и ковариаций внутри двух групп.

Для проверки на нормальность распределения воспользуйтесь группой кнопок Distribution of variables within groups, например, графиками поля рассеяния: Scatterplots.

3. Выполните дискриминантный анализ имеющихся 9 предприятий, воспользовавшись меню Statistics\Multivariate Exploratory Techniques\Discriminant Analysis и указав в появившемся окне в качестве группировочной переменной (Grouping variable) firm, а в качестве независимых (Independent variable list) остальные labor, defect и fund. В появившемся окне нажмите кнопку Summary. Получим результаты дискриминантного анализа по каждой переменной, в частности, лямбды Уилкса как для всей дискриминации, так и отдельно для каждой переменной и значимость переменных для классификации.

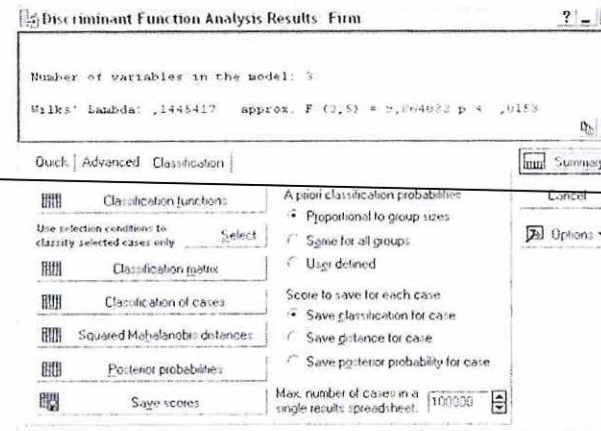
Discriminant Function Analysis Summary (Firm)						
No. of vars in model 3, Grouping firm (2 grps)						
Wilks' Lambda .14454 approx. F (3,5)=9.3640 p=.0153						
	Wilks' Lambda	Partial Lambda	F-remove (1,5)	p-level	Toler.	1-Toler. (R-Sqr)
labor	0,1815270	0,796246	1,279471	0,309315	0,687015	0,312985
defect	0,153153	0,943771	0,297895	0,608664	0,877639	0,122361
fund	0,163325	0,884994	0,649754	0,456810	0,627656	0,372345

Вернувшись в окно Discriminant Function Analysis Results выберите вкладку Classification и затем Classification functions. Получим значения коэффициентов дискриминантных функций с после опытными вероятностями попадания предприятия в одну из групп. Должны получиться следующие дискриминантные функции:

$$f_A = -54,87 + 9,13labor + 6,39defect + 10,56fund$$

$$f_B = -25,18 + 6,42labor + 11,07defect + 3,35fund$$

Если в этом же окне выбрать опцию Classification matrix, получим матрицу по строкам которой фактическая классификация, а по столбцам – полученная по модели. В идеальном случае они должны совпадать и матрица должна иметь диагональный вид при проценте корректных наблюдений 100%.



Далее, используя опции Classification of cases или Posterior probabilities получим соответственно классификацию по наблюдениям и вероятности отнесения каждого наблюдения к каждой из двух групп (А или В).

Posterior Probabilities (Firm)			
Incorrect classifications are marked with *			
Case	Observed	A	B
1	A	0,999868	0,000132
2	A	0,999568	0,000432
3	A	0,999740	0,000260
4	A	0,999899	0,000101
5	B	0,000476	0,999524
6	B	0,000000	1,000000
7	B	0,009825	0,990175
8	B	0,010251	0,989749
9	B	0,000000	1,000000
10	...	0,000492	0,999508
11	...	0,000935	0,999065
12	...	1,000000	0,000000

Причем классифицированы будут и последние 3 наблюдения, для которых мы не имели первоначально информации о том, к какой из групп они относятся (наблюдения 10 и 11 – к группе В, а 12 – к группе А). Также можно получить квадрат расстояния Махаланобиса от центра каждой из групп в помощью опции Squared Mahalanobis distances.

4. Выполните пошаговый дискриминантный анализ. Вернитесь к первоначальному окну дискриминантного анализа (Statistics\Multivariate Exploratory Techniques\Discriminant Analysis) и поставьте галочку напротив опции Advanced options. Нажмите ОК. Выберите вкладку Advanced (обратите внимание на возможность изменения значений F критерия для включения/исключения переменной и вида отображения – конечного результата или результатов по шагам) и метод (Method) пошагового анализа: Forward (включения) или Backward (исключения). При этом опция Standard относится к стандартному алгоритму анализа, выполненному нами в предыдущем пункте. Нажмите ОК и получите окно результатов, имеющее такой же как и в п. 3 вид.

Выполните пошаговый анализ методом последовательного включения переменных и методом исключения.

5. Получите результаты в п. 2–4. Сделайте содержательные выводы по результатам всех выполненных расчетов.

#### Лабораторное задание 4. «Кластерный анализ»

1. Запустите приложение Statistica. Откройте файл Lab3.sta воспользовавшись меню File\Open. Файл содержит подвыборку из массива msm.sta по индивидам. Описание переменных: lw – логарифм заработной платы, edu – число лет образования, nhh – число членов домохозяйства, dd – доля доходов главы домохозяйства в семейном бюджете, age – возраст, rpm – процент заработков, который дает основная работа. Необходимо классифицировать наблюдения.

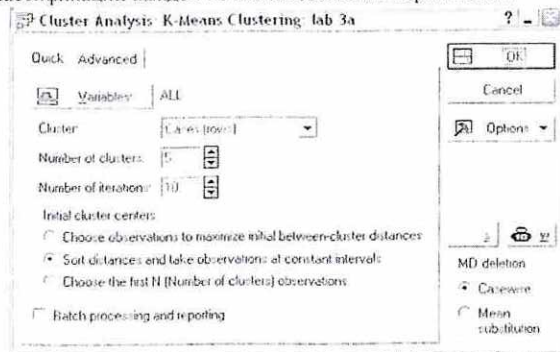
2. Выполните расчет описательных статистик по переменным выборки. Сделайте выводы. Почему нельзя использовать данные в натуральном виде? В файле Lab3a.sta содержатся стандартизованные переменные.



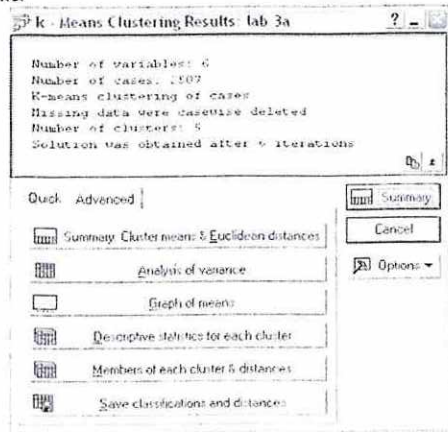
3. Выполните кластерный анализ имеющихся индивидов, воспользовавшись меню Statistics\Multivariate Exploratory Techniques\Cluster Analysis и методом автоматической классификации k-средних, указав в появившемся окне K-means clustering и нажав ОК. В окне кластерного анализа методом k-средних необходимо указать переменные классификации (укажите все имеющиеся переменные). Далее во вкладке Advanced выберите объекты классификации – в меню Cluster укажите Cases: задайте число кластеров – например, три; обратите внимание на возможность выбора начальных центров кластеров (Initial cluster centers).

После запуска вычислительной процедуры появится окно результатов, в верхней части которого содержится общая информация о классификации и количество итераций, после которых найдено решение. Выберите вкладку Advanced. Опция Cluster Means&Euclidean Distances позволяет получить таблицы с средними по переменным для каждого кластера и расстояниями между кластерами.

Опция Analysis of variance позволяет выполнить дисперсионный анализ для проверки значимости для классификации каждой из использованных переменных.



Кнопка Graph of means позволяет просмотреть средние значения для каждого кластера на графике.



Кнопка Descriptive statistics for each cluster позволяет получить описательные статистики (математическое ожидание, стандартное отклонение и дисперсия) для каждого кластера. Наконец опция Members of each cluster&distances дает объекты (наблюдения) каждого класса и расстояние от объектов до центра кластера, которому принадлежит этот объект. Кнопка Save позволяет сохранить результаты классификации.

Попробуйте изменить количество кластеров и состав переменных, по которым строится классификация. Как это влияет на результаты разбиения?

4. Организуйте случайную подвыборку наблюдений Data\Subset и далее, выбирая все переменные, укажите в Simple Random Sampling 2% percent of cases.

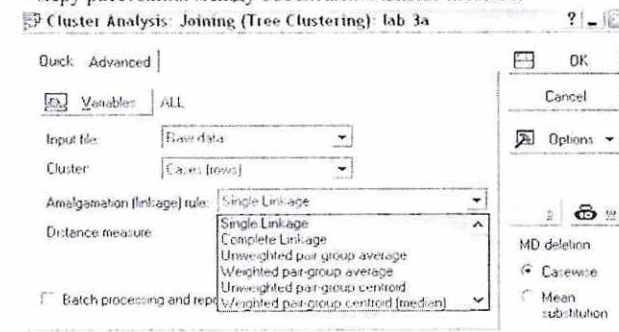
Выполните кластерный анализ имеющихся данных, воспользовавшись меню

Statistics\Multivariate Exploratory Techniques\Cluster Analysis несколькими методами агломеративной классификации, указав в появившемся окне Joining (tree clustering) и нажав ОК.

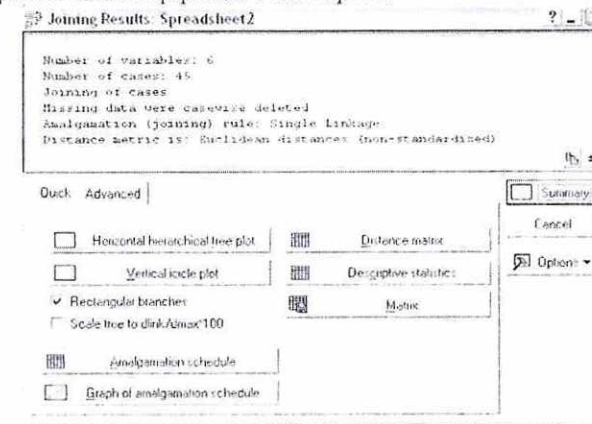
В появившемся окне кластерного анализа необходимо указать переменные классификации.

Далее во вкладке Advanced выберите:

- объекты классификации – в меню Cluster укажите Cases;
- задайте Amalgamation (linkage) rule, т.е. метод иерархического объединения кластеров: Single linkage – одиночной связи (ближайшего соседа), Complete linkage – полной связи (дальнего соседа), Unweighted pair-group average – невзвешенный метод средней связи, Weighted pair-group average – взвешенный метод средней связи, Unweighted pair-group centroid – невзвешенный центроидный метод, Weighted pair-group centroid – взвешенный центроидный метод (медианной связи), Ward's method – метод Уорда;
- меру расстояния между объектами Distance measure.



После запуска вычислительной процедуры появится окно результатов, в верхней части которого содержится общая информация о классификации.



Нажав на кнопку Vertical hierarchical tree plot, получим иерархическое дерево результатов агломеративной классификации.

Рекомендуется выбирать различные правила объединения кластеров и меры расстояний между объектами, сравнить результаты полученных классификаций, выбрать оптимальный по вашему мнению вариант и сделать содержательные экономические выводы по его результатам.

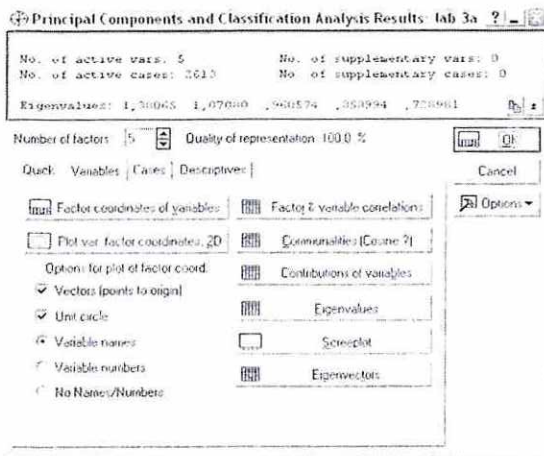
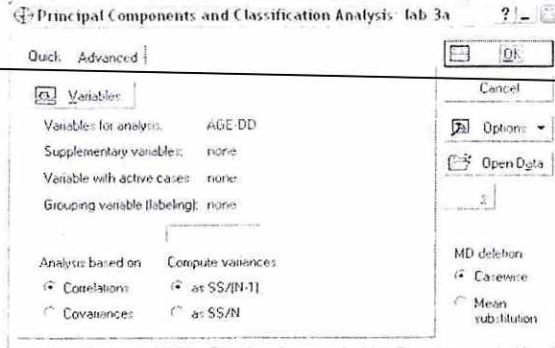
Также в пакете Statistica реализован метод классификации Two-way joining, в котором классифицируются и объекты и переменные одновременно.

Лабораторное задание 5. «Снижение размерности исследуемых многомерных признаков»

1. Запустите приложение Statistica. Откройте файл Lab3.sta воспользовавшись меню File\Open. Файл содержит подвыборку из массива msm.sta по индивидам. Описание переменных: lw – логарифм



заработной платы, edu – число лет образования, nhh – число членов домохозяйства, dd – доля доходов в семейном бюджете, age – возраст, pm – процент заработков, который дает основная работа. Необходимо провести компонентный анализ данных, а затем классифицировать наблюдения.



2. Выполните расчет описательных статистик по переменным выборки. Сделайте выводы. Рассчитайте корреляционную матрицу переменных. Необходимо ли стандартизировать переменные?

3. Анализ главных компонент выполняется в программе Statistica с помощью модуля Statistics\Multivariate Exploratory Techniques\Principal Components & Classification Analysis. В появившемся окне Principal Components and Classification Analysis укажите переменные классификации и нажмите ОК.

В верхней части появившегося окна результатов компонентного анализа дана общая информация и рассчитанные собственные числа корреляционной матрицы переменных.

Выберите число факторов (Number of factors) и получите соответствующий процент объясненной этими факторами вариации (Quality of representation). Просмотрите коэффициенты факторных нагрузок Factor&variable correlations. Далее можно получить график Screeplot собственных чисел и сами собственные числа Eigenvalues, соответствующие им собственные векторы Eigenvectors. Дайте интерпретацию факторам по значениям полученных собственных векторов.

Рекомендуется далее самостоятельно ознакомиться с другими опциями в представлении результатов компонентного анализа.

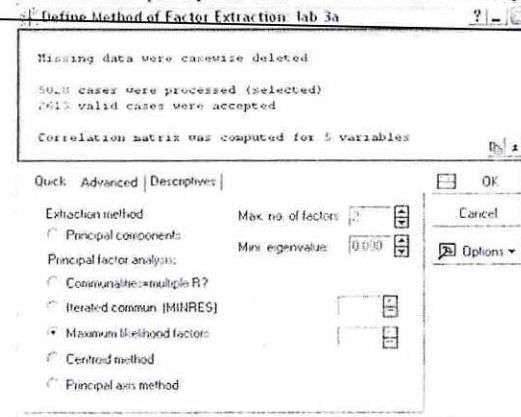
4. Для имеющихся данных определите и обоснуйте оптимальное число факторов и дайте их интерпретацию. Какой процент вариации они объясняют?

5. Выполните кластерный анализ по значениям выбранных в п. 4 главных компонент для выборки индивидов. Сделайте выводы.

6. Постройте уравнение множественной регрессии lw на выделенные в п. 4 главные компоненты. Дайте интерпретацию полученного результата.

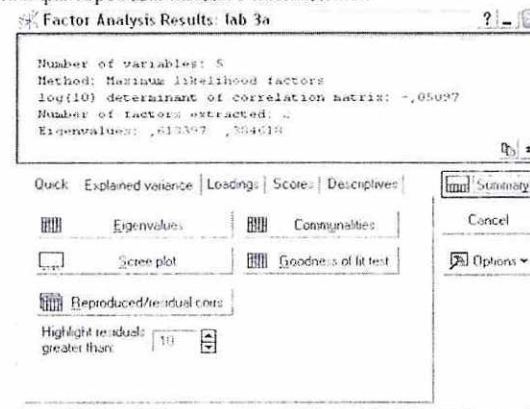
### Лабораторное задание 6. «Факторный анализ»

1. Рабочий файл данных тот же, что и в предыдущей лабораторной работе: Lab3.sta.
2. Выполним факторный анализ по имеющимся выборочным данным.



В программе Statistica запустите модуль Statistics\Multivariate Exploratory Techniques\Factor Analysis и в появившемся окне укажите переменные для анализа (укажите все имеющиеся переменные, кроме lw) и нажмите ОК. Далее во вкладке Advanced укажите метод поиска латентных факторов, например, Principal factor analysis\Maximum likelihood factors метод главных факторов (рекомендуется сравнить результаты, полученные разными методами оценивания), а также максимальное количество факторов, например 2, и минимальное собственное число фактора для включения его в анализ – 0. Нажмите ОК.

В появившейся таблице результатов после нажатия кнопки Summary на вкладке Quick получим факторные нагрузки, причем красным цветом выделены коэффициенты, большие по модулю 0,7. Также можно получить (опция Eigenvalues) значения собственных чисел и долю объясненной факторами вариации. Воспользовавшись вкладкой Explained variance и опцией Communalities получите значения общностей для каждого из фактора. С помощью вкладки Scores и кнопки Factor scores получим значения факторов для каждого наблюдения.



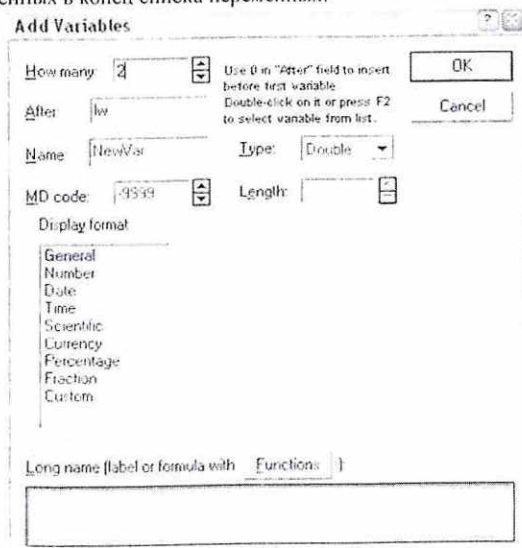
3. Осуществим вращение факторов. Для этого выберите в меню Factor rotation один из методов вращения: варимакс, биквартимакс, квартимакс, эквимакс, например, Varimax normalized. Посмотрите как изменились факторные нагрузки и другие результаты факторного анализа. Дайте интерпретацию

факторов.

Поэкспериментируйте, выбирая различные значения числа факторов на начальной стадии анализа и различные методы вращения факторов. Выберите оптимальный с вашей точки зрения результат.

4. Выполните кластерный анализ наблюдений (по объектам) по факторам, полученным после варимакс вращения в предыдущем пункте. Для этого выполните следующие действия.

В окне исходных данных, щелкнув правой кнопкой мыши по заголовку любой переменной, выберите Add variables и добавьте необходимое количество (по числу факторов в п. 3) новых переменных в конец списка переменных:



Вернувшись в окно факторного анализа с полученными как в пункте 3 факторами с помощью вкладки Scores и кнопки Factor scores получите таблицу значений факторов для каждого наблюдения. Скопируйте два столбца значений таблицы в окно исходных данных на место добавленных новых переменных (с именами по умолчанию NewVar – можно переименовать переменную дважды кликнув по заголовку). Далее воспользуйтесь меню Statistics\Multivariate Exploratory Techniques\Cluster Analysis и методом автоматической классификации k-средних (см. лабораторную работу по кластерному анализу) для вновь полученных переменных.

5. Постройте уравнение множественной регрессии lw на выделенные факторы.

6. Постройте необходимые графики и рассчитайте соответствующие статистики. Сделайте содержательные экономические выводы по результатам статистического анализа.