

## КОМПЬЮТЕРНАЯ ОБРАБОТКА АНТРОПОЛОГИЧЕСКИХ ДАННЫХ

### **1. Место дисциплины в структуре ОПОП**

Дисциплина «Компьютерная обработка антропологических данных» является дисциплиной профиля по выбору студента вариативной части для ОПОП «Общая биология и экология» мс\_антропология. Изучается в 7 семестре студентами кафедры антропологии отделения «Общая биология и экология».

Дисциплина «Компьютерная обработка антропологических данных» предназначена для подготовки специалистов по фундаментальной и прикладной биологии по профилю антропология. Курс позволяет обучающимся освоить основные методы обработки баз данных, прививает способность критически относится к выбору способов извлечения информации, формирует навык работы с основными прикладными программами для статистической обработки данных.

Освоение данной дисциплины необходимо как предшествующее для курса «Прикладная антропология», а также для выполнения выпускной квалификационной работы.

#### ***Цели освоения дисциплины***

Ознакомление обучающихся с наиболее распространенными пакетами прикладных статистических программ для ПК.

#### ***Задачи курса:***

- формирование устойчивого навыка работы с прикладными пакетами программ статистической обработки данных;
- формирование умения выбирать адекватные методы для извлечения информации из массивов данных;
- формирование навыка интерпретации результатов статистического анализа;
- формирование умения использовать средства визуализации цифровой информации.

### **2. Входные требования**

Перед началом освоения дисциплины «Компьютерная обработка антропологических данных» студент должен изучить курс «Статистический анализ антропологических данных».

### **3. Планируемые результаты изучения дисциплины, соотнесенные с требуемыми компетенциями выпускников**

- ***Компетенции выпускников (коды):***

**СПК-6.** Способность использовать современные компьютерные средства и специализированное программное обеспечение для статистической обработки антропометрических и других экспериментальных и обсервационных данных; владение навыками глубокого анализа и биологической интерпретации результатов статистической обработки антропологических данных; способность компьютерного моделирования антропологических и геногеографических интерполяционных карт.

— *Планируемые результаты обучения по модулю, сопряженные с компетенциями:*

Освоение базовых навыков работы со стандартными пакетами прикладных статистических программ для ПК, формирование устойчивого навыка работы с прикладными пакетами программ статистической обработки данных, умения выбирать адекватные методы для извлечения информации из массивов данных, формирование навыка интерпретации результатов статистического анализа и умения использовать средства визуализации цифровой информации.

— *Индикаторы (показатели) достижения компетенций:*

**Знает:**

- основные возможности прикладных программ пакетов (ППП) «Statistica», «PAST», «SPSS»;
- возможности использования ПО «MultiCan».

**Умеет:**

- применять адекватные методы получения информации с помощью пакетов статистических программ;
- визуализировать полученные данные с помощью графических возможностей ППС.

**Владеет навыками:**

- работы с прикладными статистическими программами;
- работы с различными графическими редакторами.

**Демонстрирует готовность:**

- применять полученные навыки работы в стандартных пакетах статистических программ в научной и профессиональной деятельности антрополога.

#### **4. Объем дисциплины «Компьютерная обработка антропологических данных»**

у обучающихся на ОПОП «Общая биология и экология» по подплану мс\_антропология:

- Общая трудоемкость дисциплины – 2 з.е. (72 ч).

- Аудиторная нагрузка – 56 ч. (4 ч. в неделю), из них семинары – 56 ч.
- Самостоятельная работа – 16 ч.
- Форма промежуточной аттестации – экзамен (7 семестр).

## 5. Форма обучения – очная

## 6. Содержание и структура дисциплины

№ п/п	Раздел дисциплины	Семинары (ч)	Самостоятельная работа (часы)
1	<u>Тема 1.</u> Первичная обработка массива данных в ППС	6	2
2	<u>Тема 2.</u> Проверка статистических гипотез для двух выборок	10	2
3	<u>Тема 3.</u> Методы классической биометрии для сравнения нескольких выборок	10	3
4	<u>Тема 4.</u> Расчет корреляций признаков	10	3
5	<u>Тема 5.</u> Методы многомерной биометрии для изучения внутригрупповой изменчивости	10	3
6	<u>Тема 6.</u> Методы многомерной биометрии для изучения межгрупповой изменчивости	10	3
	<b>Итого</b>	56	16

### 6.1. Программа дисциплины «Компьютерная обработка антропологических данных»

#### Тема 1. Первичная обработка массива данных в ППС

Создание файла данных в ППС (пакет программ «Статистика») и работа с ним. Форматы ячеек, кодировка пропущенных значений. Возможности пунктов меню «Data» и «Tools».

Получение гистограммы, вариационного ряда количественного признака. Выбор параметров гистограммы. Проверка распределения признака на нормальность. Критерии Колмогорова-Смирнова и Лилиефорса, их реализация в ППС.

Расчет базовых статистик, получение доверительного интервала для средней арифметической, показателей асимметрии и эксцесса, расчет перцентилей, получение сглаживающих кривых.

Тема 2. Проверка статистических гипотез для двух выборок

Применение t-критерия Стьюдента для независимых/зависимых выборок (возможности, варианты реализации, трактовка табличных и графических результатов). Т<sup>2</sup>-статистика Хотеллинга.

Непараметрический критерий Манна-Уитни. Реализация в ППС, графическое представление результатов.

Сравнение частот балловых (дискретных) признаков. Сравнение вариационных рядов в двух выборках. Возможности ППС для решения этих задач.

Тема 3. Методы классической биометрии для сравнения нескольких выборок.

Сравнение нескольких выборок с помощью дисперсионного анализа. Интерпретация табличных данных. Достоверность отличий, множественное сравнение по Шеффе. Варианты графического оформления результатов.

Непараметрический аналог дисперсионного анализа – метод Краскелла - Уоллиса. Алгоритм действий в ППС. Возможности модуля непараметрической статистики.

Тема 4. Расчет корреляций признаков

Получение матрицы пирсоновских корреляций. Ранговые корреляции Спирмена, показатели связи  $\gamma$  и  $\tau$ , расчет коэффициентов сопряженности С и V. Расчет показателей связи для смешанного набора признаков. Вычисление индекса корреляции с помощью модуля дисперсионного анализа. Формат .smx и возможности использования матрицы корреляций для проведения многомерных анализов.

Тема 5. Методы многомерной биометрии для изучения внутригрупповой изменчивости

Модуль регрессионного анализа, пошаговая регрессия, выбор критериев включения/исключения признаков. Канонические корреляции: возможности метода. Анализ главных компонент: основные принципы, результаты и интерпретация табличных данных. Возможности ППС по графическому представлению результатов анализа. Факторный анализ, обзор методов факторизации и представление результатов. Ортогональная ротация результатов факторного анализа. Реализация Р,Q,T-техник в модуле факторного анализа ППС.

Тема 6. Методы многомерной биометрии для изучения межгрупповой изменчивости

Дискриминантный анализ в ППС, возможности модуля дискриминантного анализа для решения задач классификации групп наблюдений. ПО «MultiCan», его возможности и ограничения применения. Кластерный анализ: иерархические и параллельные процедуры кластеризации в ППС. Получение матрицы расстояний в ППС, работа модуля многомерного шкалирования.

## **7. Фонд оценочных средств для оценивания результатов обучения по дисциплине:**

### **7.1. Типовые задания и иные материалы, необходимые для оценки результатов обучения**

*Примерный список заданий для проведения текущей аттестации (для подготовки к коллоквиумам, контрольным, опросам)*

По данной дисциплине в качестве текущего контроля успеваемости предусмотрено выполнение практических заданий по изученной теме и их обсуждение в рамках семинарского занятия.

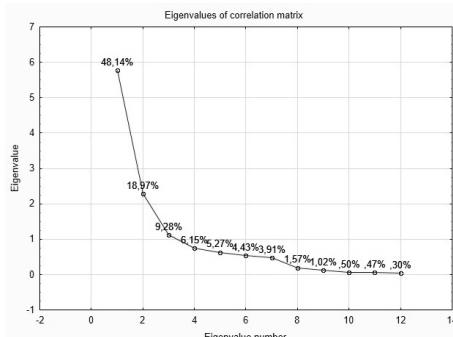
*Примерный список вопросов для промежуточной аттестации (экзамен)*

1. Каковы принципы создания файла данных в пакете программ «Статистика»? Как можно облегчить процесс внесения массивов данных?
2. Какие типы форматов файлов используются в пакетах статистических программ? Что такое формат матрицы? Какие типы матриц создаются в пакете программ «Статистика»?
3. В каких модулях пакета программ «Статистика» используются корреляционные матрицы как первичные данные для анализа?
4. Перечислите методы, необходимые для сравнения двух выборок между собой. Какие модули пакета программ «Статистика» необходимо использовать?
5. Какие методы, реализованные в пакете программ «Статистика» могут быть использованы для сравнения трех и более выборок между собой?
6. В каких модулях пакета программ «Статистика» можно реализовать сравнение частот описательных признаков по трем и более выборкам?
7. Какие типы показателей связи могут быть рассчитаны с помощью пакета программ «Статистика»? В каких модулях пакета программ «Статистика» эти показатели рассчитываются?

8. Каким критерием следует воспользоваться для проверки статистической гипотезы о равенстве векторов средних в двух группах наблюдений?
9. Каким критерием следует воспользоваться для проверки статистической гипотезы о равенстве векторов средних в нескольких группах наблюдений?
10. Данна задача изучения коррелированности одного признака с набором других признаков. Какой метод анализа следует использовать для ее решения?
11. Данна задача изучения коррелированности двух наборов признаков. Какой метод анализа следует использовать для ее решения?
12. Данна задача изучения коррелированности единого набора признаков. Какой метод анализа следует использовать для ее решения?
13. Какой показатель, получаемый в ходе множественного регрессионного анализа, можно использовать для непосредственного суждения о доле вариации зависимой переменной, описываемой множественной связью с ней набора независимых признаков?
14. Для каких факторов в факторном анализе характерны нескоррелированность и совпадение матриц факторного отображения и факторной структуры?
15. Для какого способа факторизации характерно нахождение нагрузок как собственных векторов редуцированной корреляционной матрицы?
16. Для какого способа факторизации характерно нахождение нагрузок как наилучших статистических оценок?
17. В какой технике факторного анализа рассматривается один признак, измеренный в несколько моментов времени у многих индивидов?

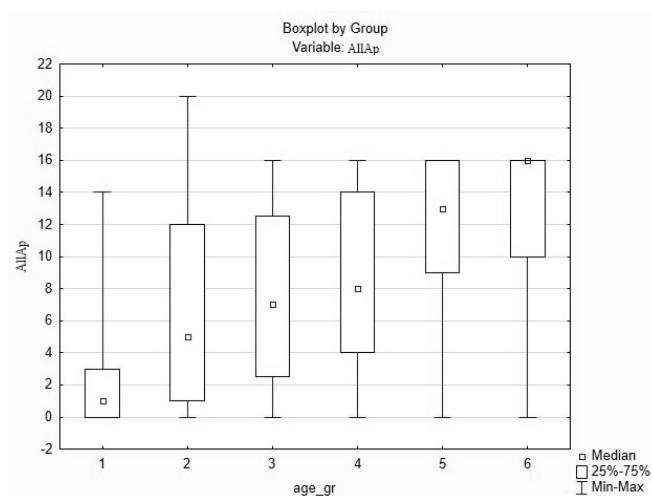
### *Примеры ситуационных кейс-заданий*

1. На рисунке показан так называемый график «каменистой осьпи», получаемый при проведении факторного анализа.



Какое количество новых переменных отражает реально существующие закономерности изменчивости?

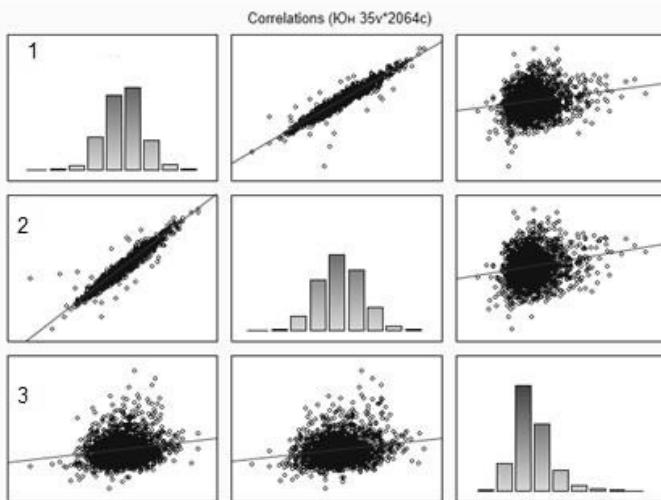
1. первые два
  2. первые три
  3. первые четыре
  4. все факторы
  5. нет верного ответа
2. На рисунке показаны проекции распределения признака в разных возрастных группах.



Какие выводы из предложенных ниже верны:

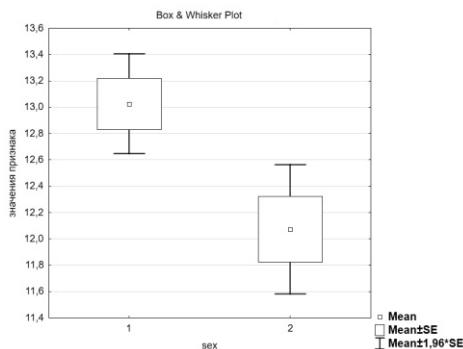
1. изучено шесть возрастных групп
2. изучены шесть проявлений признака в группе
3. асимметричность распределения отмечена только в первой и второй группе
4. асимметричность распределения отмечена в четырех группах
5. достоверные различия отмечаются между всеми возрастными группами
6. симметричность распределения есть в третьей и четвертой возрастной группе

3. На рисунке представлен отчет по результатам корреляционного анализа трех признаков



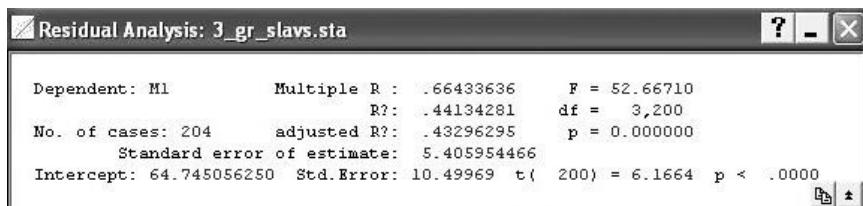
Какие выводы верны:

1. первый и второй признаки тесно связаны
  2. все признаки имеют нормальное распределение
  3. третий признак слабо связан с первым и вторым
  4. второй признак распределен асимметрично
  5. третий признак имеет правостороннюю асимметрию
  6. первый признак имеет слабую левостороннюю асимметрию
4. На рисунке представлен график, полученный по результатам применения t-критерия Стьюдента.



Какие из выводов, сделанных на основании анализа графика, верны:

1. группы 1 и 2 относятся к одной генеральной совокупности
  2. доверительные интервалы для средних по группам 1 и 2 практически не перекрываются
  3. группы 1 и 2 относятся к разным генеральным совокупностям
  4. по рисунку невозможно определить границы доверительных интервалов для групп 1 и 2
  5. квадратическая ошибка средней величины признака по группе 1 составляет не более 0.5 единицы измерения признака
5. По результатам применения регрессионного анализа получено диалоговое окно (рис)



Какие выводы из текстовой информации верны:

1. коэффициент множественной корреляции равен 0,66433636.
2. коэффициент множественной корреляции равен 0,44134281.
3. число индивидов в анализе не превышало 200 наблюдений.
4. полученный коэффициент множественной корреляции статистически недостоверен.
5. ошибка расчета признака по уравнению регрессии составляет 5,406 единиц.
6. ошибка расчета признака по уравнению регрессии составляет 10,500 единиц.

## 7.2. Описание критериев и шкал оценивания

### *Описание критериев оценивания выполнения задания*

Показатель	Баллы
Студент выполняет менее 50% задания	0-20
Задание студент выполняет все или большей частью, есть отдельные неточности, способен при направляющих вопросах исправить допущенные неточности	21-32
Задание выполнено студентом правильно, самостоятельно в полном объеме	33-40

### **Шкала оценивания сформированности компетенций**

Уровень сформированности компетенции	Баллы	Оценка в 5-балльной шкале
Недостаточный	Менее 20	неудовлетворительно
Базовый	20-26	удовлетворительно
Высокий (повышенный)	27-32	хорошо
Продвинутый (повышенный)	33-40	отлично

## **8. Ресурсное обеспечение:**

### **8.1. Перечень основной и дополнительной учебной литературы**

#### Основная литература

1. Боровиков В. STATISTICA Искусство анализа данных на компьютере. Для профессионалов. СПб, 2001, 650 с.
2. Боровиков В.П., Боровиков И.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. М. 1998.
3. Бюль А., Цефель П. SPSS. Искусство обработки информации. М. - Киев.: DiaSoft, 2005. 602 с.
4. Наследов А. SPSS: компьютерный анализ данных в психологии и социальных науках. СПб.: Питер, 2005. 416 с.
5. Поллард Дж. Справочник по вычислительным методам статистики. М.: Финансы и статистика, 1982, с.344.

### **8.2. Перечень лицензионного программного обеспечения**

1. Пакет программ «Statistica 6.0» и выше, пакет программ SPSS
2. Free-soft программное обеспечение PAST для обработки палеоантропологической информации
3. Яндекс Браузер

### **8.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

Электронная библиотека механико-математического факультета МГУ

<http://lib.mexmat.ru/books>

<http://pca.narod.ru/>

<http://statsoft.ru/home/textbook/default.htm>

**9. Язык преподавания**

Русский

**10. Преподаватель**

Гончарова Наталия Николаевна – кандидат биологических наук, доцент кафедры антропологии биологического факультета МГУ

**11. Автор программы**

Гончарова Наталия Николаевна – кандидат биологических наук, доцент кафедры антропологии биологического факультета МГУ