



Правительство Российской Федерации

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет прикладной политологии

**Программа дисциплины
"Многомерный статистический анализ в политологии"**

для направления 030200.62 «Политология» подготовки бакалавра

Автор программы:

Стукал Д.К., преподаватель, denis.stukal@gmail.com

Одобрена на заседании кафедры высшей математики «___»_____ 2011 г
Зав. кафедрой Макаров А.А.

Москва, 2011

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.



1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов направления подготовки 030200.62 «Политология» подготовки бакалавра, изучающих дисциплину «Многомерный статистический анализ в политологии».

Программа разработана в соответствии со здравым смыслом и образцом на сайте НИУ ВШЭ.

2 Цели освоения дисциплины

Цель освоения дисциплины «Многомерный статистический анализ в политологии» – выработать базовые компетенции по решению задач, связанных с анализом эмпирических данных с помощью методов многомерной статистики.

В соответствии с поставленной целью, курс решает следующие задачи:

1. формирование у студентов знания понятий и идей, лежащих в основе многомерной математической статистики;
2. освоение основных статистических моделей социально-экономических и политических процессов и явлений;
3. овладение основными методами многомерной математической статистики и эконометрики, позволяющими решать различные социально-экономические и политологические исследовательские задачи;
4. формирование у студентов понимания перспектив использования статистических методов анализа данных в прикладной политологии.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать методы многомерной статистике в объеме данной программы.
- Уметь строить диаграмму рассеяния, осуществлять кластеризацию наблюдений любой размерности с использованием изученных метрик и алгоритмов, получать главные компоненты по представленному массиву данных малой размерности и/или корреляционной/ковариационной матрице, строить уравнение парной регрессии с помощью МНК.
- Иметь навыки анализа эмпирических данных в пакете Stata с использованием изученных методов.

В результате освоения дисциплины студент осваивает следующие компетенции:

Компетенция	Код по ФГОС/ НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
Умение логически верно, аргументировано и ясно строить устную и письменную речь	ОНК-2	Демонстрирует способность написания профессиональных текстов в виде эссе/статьи	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов Эссе: самостоятельная подготовка эссе/статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов



Компетенция	Код по ФГОС/ НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
Использование основных положений и методов социальных, гуманитарных и экономических наук при решении профессиональных задач, способность анализировать социально-значимые проблемы и процессы	ОНК-4	Демонстрирует способность отбирать положения политической науки, а также смежных наук, позволяющие интерпретировать результаты анализа данных	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов Домашнее задание: самостоятельная подготовка эссе/статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов
Способность применять математические методы анализа к социальным, политическим и экономическим явлениям	ОНК-7	Оценивает и интерпретирует регрессионные модели, реализует и интерпретирует результаты кластерного и компонентного анализа	Семинары: реализация методов многомерной статистики «руками», а также на компьютерах; обсуждение статей ведущих зарубежных реферируемых журналов
Способность работать с информацией в глобальных компьютерных сетях	ИК-2	Демонстрирует способность к нахождению статистической информации в сети Интернет	Домашнее задание: самостоятельная подготовка эссе/статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов
Владение иностранным языком на уровне, достаточном для разговорного общения, а также для поиска и анализа иностранных источников информации	ИК-3	Демонстрирует способность к чтению профессиональной литературы на английском языке.	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов Домашнее задание: самостоятельная подготовка эссе/статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов
Способность использования основ защиты производственного персонала и населения от возможных последствий аварий, катастроф, стихийных бедствий и применения современных средств поражения, основных мер по ликвидации их последствий, способность к общей оценке условий безопасности жизнедеятельности	СЛК-8	Демонстрирует способность к безопасному включению и выключению компьютера	Семинары: реализация методов многомерной статистики в компьютерных классах

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к федеральному компоненту общих профессиональных дисциплин направления 030200.62 «Политология» подготовки бакалавра. Для указанного направления дисциплина является базовой.



Изучение данной дисциплины базируется на следующих дисциплинах:

- «Алгебра и анализ» (1 курс)
- «Теория вероятностей, начала статистики и обработки данных» (2 курс)

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

- Знать содержательный смысл следующих понятий: функция, производная, интеграл; матрица, линейное преобразование, определитель матрицы, обратная матрица, собственный вектор матрицы, собственное значение матрицы.
- Уметь дифференцировать и вычислять определенный интеграл элементарных функций; совершать матричные операции (сложение, вычитание, умножение матрицы на число, вектор и другую матрицу), находить обратную матрицу, определитель матриц 2x2, 3x3, собственный вектор матрицы.
- Владеть знаниями по теории вероятностей и математической статистике в объеме дисциплины «Теория вероятностей, начала статистики и обработки данных» 2010 – 2011 уч.года.

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

- «Сравнительная политология»
- «Анализ региональных и электоральных процессов»
- «Методы социально-экономического прогнозирования»
- «Многомерные типологии политических систем государств мира»

5 Тематический план учебной дисциплины

Наименование разделов	Аудиторные часы		Самостоятельная работа	Всего
	Лекции	Практические занятия		
Введение	2	2	14	8
Классификация: иерархический кластерный анализ	4	4	10	18
Построение индексов: метод главных компонент	4	6	18	28
Введение в измерение латентных признаков: обзор факторного анализа *	2	0	2	4
Введение в регрессионный анализ	8	8	24	40
ИТОГО	20	20	68	108

* Тема рассматривается, если изложение темы №3 укладывается в 4 лекционных академических часа.



6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Параметры
		1	2	3	4	
Текущий	Контрольная работа			*		Письменная работа 160 минут
	Домашнее задание				*	См. ниже.
Итоговый	Зачет				*	

Параметры домашнего задания.

Домашнее задание предполагает написание мини-исследования по заданной теме по итогам самостоятельного поиска и изучения соответствующей литературы и анализа полученного массива социально-экономических и политологических данных методами многомерной статистики. Структура домашнего задания в обязательном порядке должна включать в себя следующие разделы:

1. Аннотация (от 500 до 800 знаков с пробелами)
2. Введение (постановка задачи, обзор литературы и краткая «приманка» для читателя в виде формулировки основных полученных в работе результатов)
3. Гипотезы (формулируются и обосновываются гипотезы работы)
4. Данные и используемые методы (описываются имеющиеся данные и используемые методы; выбор методов обосновывается)
5. Результаты (в табличной форме, соответствующей зарубежным публикационным стандартам¹, представляются полученные результаты, а также предлагается их интерпретация)
6. Обсуждение полученных результатов (описывается, в какой мере полученные результаты соответствуют существующей литературе; обсуждаются возможные причины расхождений; обсуждается устойчивость полученных результатов к изменению массива данных и изменению методов анализа)
7. Заключение (формулируются выводы работы, обсуждаются теоретические следствия из полученных выводов, указываются возможные направления дальнейшего уточнения результатов)
8. Список использованной литературы

Объем домашнего задания – 15 – 30 страниц. Требования к оформлению: шрифт Times New Roman, 12 кегль, полуторный интервал; нумерация страниц в правом нижнем углу с отсутствием номера на титульной странице. Текст отформатирован по ширине.

¹ Примеры публикационных стандартов оформления таблиц студенты найдут в зарубежных статьях, указанных в списке обязательной литературы по темам дисциплины.

7 Содержание дисциплины²

Тема 1. Введение

Стохастическая природа показателей социальной и политической статистики. Взгляд на объекты социальной реальности как случайные векторы. Понятие признаков пространства. Задачи статистического вывода (инференции) и нестатистические задачи анализа данных в социальных науках. Числовые характеристики случайных величин в одномерном и многомерном случаях (математическое ожидание, дисперсия; вектор математических ожиданий, ковариационная матрица, ее след и определитель). Выборочные оценки числовых характеристик. Ковариация и корреляция как меры линейной взаимосвязи, их выборочные оценки. Проверка значимости коэффициента корреляции К.Пирсона и логика проверки статистических гипотез.

Основная литература:

1. † King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press. Chapters 1 and 2.
2. † Мангейм Дж. Б., Рич Р.К. Политология: Методы исследования. М.: Издательство «Весь Мир», 1997. – С. 19 – 117.

Дополнительная литература:

1. Попова О.В. Политический анализ и прогнозирование: Учебник. М.: Аспект Пресс, 2011. С. 62 – 73, 230 – 264, 311 – 329.
2. Толстова Ю.Н. Математико-статистические модели в социологии (математическая статистика для социологов): учеб. пособие. М.: Издательский дом ГУ ВШЭ, 2008. – С. 19 – 36.
3. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С.38 – 81, 271 – 295, 253 – 270.
4. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. - М.: ЮНИТИ, 2001. – С. 26 – 50.

Тема 2. Классификация: иерархический кластерный анализ

Классификация как одна из общенаучных задач. Роль классификации в политической науке по Дж. Сартори. Типология задач классификации: с обучением и без обучения, параметрическая и непараметрическая постановка задачи. Постановка задачи иерархической кластеризации. Понятия расстояния и его свойства. Виды расстояний и проблема их выбора. Понятие типа (алгоритма) агломерации и его виды. Проблема выбора алгоритма агломерации.

Определение числа кластеров в задаче иерархического кластерного анализа (дендрограмма и spider-web). Проблема устойчивости результатов кластеризации к смене алгоритма агломерации и метрики. Применение кластерного анализа к классификации траекторий режимных трансформаций посткоммунистических государств в проектах ЦФИ НИУ ВШЭ ТЗ-9.3 «Анализ режимных изменений в современном мире: новые демократии и/или новые автократии?» и ТЗ-47.0 «Анализ факторов развития и упадка государственности в социалистических и постсоциалистических странах Европы и Азии второй половины XX – начала XXI вв. с помощью количественных и качественных методов (включая QCA, Qualitative Comparative Analysis)».

Основная литература:

1. † Cluster Analysis. In *Analysis of Multivariate Social Science Data* edited by David J. Batholomew, Fiona Steele, Irini Moustaki and Jane I. Galbraith. Boca Raton, London, New York: CRC Press, 2008. Pp. 17 – 53. (материал размещен на корпоративной

² Символом † в списке литературы отмечен материал, освоение которого проверяется на практических занятиях, в контрольной работе и на зачете.

странице преподавателя; доступен в центральной библиотеке ВШЭ: полочный индекс – 519 А53)

2. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. - М.: ЮНИТИ, 2001. – С. 488 – 518.

Дополнительная литература:

1. Попова О.В. Политический анализ и прогнозирование: Учебник. М.: Аспект Пресс, 2011. – С. 380 – 389.
2. Ахременко А.С. Политический анализ и прогнозирование: учеб. пособие. М.: Гардарики, 2006. С. 146 – 162.
3. Gore Paul A., Jr. Cluster Analysis. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. Academic Press, 2000. Pp. 297 – 321. (материал размещен на корпоративной странице преподавателя)
4. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 2003. С. 241 – 254.
5. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях: Учебник для вузов. М.: ЮНИТИ-ДАНА, 2001. С. 172 – 205.
6. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; Под ред. И.С. Енюкова. М.: Финансы и статистика, 1989. С. 139 – 201.

Тема 3. Построение индексов: метод главных компонент

Индекс как результат снижения размерности многомерного признакового пространства. Объяснение потребности в индексах. Размерность каких признаков пространств может быть снижена? Геометрический смысл корреляции. Постановка задачи метода главных компонент (МГК). Ковариационная / корреляционная матрица как основной объект. Идея МГК. Алгоритм МГК. Понятия собственного вектора и собственного числа матрицы, характеристического уравнения, определителя матрицы.

Главная компонента vs. среднее. Свойства главных компонент (взаимная ортогональность; наименьшее искажение геометрической структуры данных, наименьшая ошибка автопрогноза). Оценка качества снижения размерности: доля объясненной вариации и понятие информативности главной компоненты, методы Г. Кайзера (1960 г.) и Р.Б. Кеттелла (1966 г.).

Основная литература:

1. † Principal Component Analysis. In *Analysis of Multivariate Social Science Data* edited by David J. Batholomew, Fiona Steele, Irini Moustaki and Jane I. Galbraith. Boca Raton, London, New York: CRC Press, 2008. Pp. 117 – 144. (материал размещен на корпоративной странице преподавателя; доступен в центральной библиотеке ВШЭ: полочный индекс – 519 А53)
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. М.: ЮНИТИ, 2001. С. 520 – 550.

Дополнительная литература:

1. 18.06 Linear Algebra MIT Video Course by Gilbert Strang. Lectures 18 – 21. URL: <http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/>
2. Simon Carl, and Lawrence Blume. (1994). *Mathematics for economists*. Cambridge: W.W.Norton & Company (книга доступна в библиотеке ВШЭ)
3. Jolliffe I.T. 2002. Principal Component Analysis. New York: Springer. Chapters 1 – 2, 4 – 6. (книга доступна в системе электронных ресурсов библиотеки ВШЭ, а также по адресу: [http://kolxo3.tiera.ru/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%2002\)\(518s\)_MVsa_.pdf](http://kolxo3.tiera.ru/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%2002)(518s)_MVsa_.pdf))

4. Попова О.В. Политический анализ и прогнозирование: Учебник. М.: Аспект Пресс, 2011. С. 265 – 293, 364 – 379.
5. Ахременко А.С. Политический анализ и прогнозирование: учеб. пособие. М.: Гардарики, 2006. С. 169 – 180.
6. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. М.: Изд. дом ГУ ВШЭ, 2007. С. 115 – 146.
7. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 2003. С. 134 – 177.
8. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях: Учебник для вузов. М.: ЮНИТИ-ДАНА, 2001. С. 140 – 171.
9. Иберла К. Факторный анализ / Пер. с нем. В.М. Ивановой; Предисл. А.М. Дуброва. М.: Статистика, 1980. С. 54 – 349.

Тема 4. Введение в измерение латентных признаков: обзор факторного анализа *

* *Тема рассматривается, если изложение темы №3 укладывается в 4 лекционных академических часа.*

Понятие латентного признака как формализации концепта. Подходы к измерению латентных признаков: прокси переменные vs. моделирование с учетом стохастической природы наблюдаемых индикаторов. Латентный признак как «причина» индикаторов. Почему латентный признак не следует отождествлять с индексом, полученным с помощью МГК? Проблема заменимости индикаторов. Сравнение моделей ФА и МГК. Разведывательный («эксплораторный») и подтверждающий («конфирматорный») ФА. Программное обеспечение для ФА.

Основная литература:

1. Factor Analysis. In *Analysis of Multivariate Social Science Data* edited by David J. Bartholomew, Fiona Steele, Irini Moustaki and Jane I. Galbraith. Boca Raton, London, New York: CRC Press, 2008. Pp. 175 – 207. (*материал размещен на корпоративной странице преподавателя; доступен в центральной библиотеке ВШЭ: полочный индекс – 519 А53*)
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. М.: ЮНИТИ, 2001. С. 551 – 569.

Дополнительная литература:

1. Ахременко А.С. Политический анализ и прогнозирование: учеб. пособие. М.: Гардарики, 2006. С. 169 – 180.
2. Попова О.В. Политический анализ и прогнозирование: Учебник. М.: Аспект Пресс, 2011. С. 265 – 293, 364 – 379.
3. Fabrigar Leandre R., Duane T. Wegener, Robert C. MacCallum, and Erin J. Strahan. (1999). “Evaluating the Use of Exploratory Factor Analysis in Psychological Research.” *Psychological Methods*, Vol.4. No. 3, pp. 272 – 299.
4. Widaman K.F. Common Factor Analysis versus Principal Component Analysis: Differential Bias in Representing Model Parameters? // *Multivariate Behavioral Research*. – 1993. – Vol.28, No.3. – Pp. 263 – 311.
5. Velicer W.F. and Jackson D.N. Component Analysis versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure // *Multivariate Behavioral Research*. – 1990. – Vol.25. – Pp. 1 – 28.
6. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; Под ред. И.С. Енюкова. М.: Финансы и статистика, 1989. С. 11 – 69.
7. Иберла К. Факторный анализ / Пер. с нем. В.М. Ивановой; Предисл. А.М. Дуброва. М.: Статистика, 1980. С. 54 – 349.

Тема 5. Введение в регрессионный анализ

Каузальность и регрессия: почему нам нужна множественная регрессия? МНК как метод построения регрессии по генеральной совокупности. Повторение: выборочные данные и необходимость статистического вывода (инференции). Свойства точечных оценок.

Простейшая вероятностная модель, необходимая для инференции (классическая нормальная модель множественной регрессии – КНММР). Свойства МНК в КНММР: теорема Гаусса-Маркова. Проверка гипотез о модели: F-критерий, t-критерий. Меры качества модели: R^2 , скорректированный R^2 .

Типичные проблемы в регрессионном анализе: влиятельные наблюдения (методы диагностики), автокорреляция (суть, последствия, генезис, диагностика), гетероскедастичность (суть, последствия, генезис, диагностика).

Типичные проблемы в регрессионном анализе (*продолжение*): мультиколлинеарность (суть, генезис, последствия, как бороться), эндогенность (суть, последствия), ошибки измерения в регрессорах (суть, последствия, как бороться).

Основная литература:

1. † King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press. Chapters 3 – 6.
2. † Freedman D.A. (1991). “Statistical Models and Shoe Leather.” *Sociological Methodology*, Vol. 21, pp. 291 – 313.
3. † Blalock H.M., Jr. (1991). “Are There Really Any Constructive Alternatives to Causal Modeling.” *Sociological Methodology*, Vol. 21, pp. 325 – 335.
4. † Casper, Gretchen and Claudiu Tufis. (2003). “Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets.” *Political Analysis*, Vol. 11, pp. 196 – 203.

Дополнительная литература:

1. Доугерти К. Введение в эконометрику: Учебник. 2-е изд. М.: ИНФРА-М, 2007. С. 49 – 153.
2. Gujarati Damodar. Basic Econometrics. Mcgraw-Hill, 2004. Pp. 20 – 302. (*материал размещен на корпоративной странице преподавателя*)
3. Krueger J., Lewis-Beck M. (2008). “Is OLS Dead?” *The Political Methodologist*, Vol.15, No.2, pp. 2 – 4.
URL: http://polmeth.wustl.edu/methodologist/tpm_v15_n2.pdf
4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ / Пер. с англ. М.: Издательский дом «Вильямс», 2007. 912 с.
5. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. М.: Дело, 2007. 504 с.
6. Вербик М. Путеводитель по современной эконометрике / Пер. с англ. В.А. Банникова. Научн. ред. и предисл. С.А. Айвазяна. М.: Научная книга, 2008. 616 с.
7. Wooldridge J. Introductory Econometrics: A Modern Approach, 4th Edition. South-Western College Publishing, 2009. 865 p.
8. Greene W.H. Econometric Analysis, 7th Edition. Prentice Hall, 2011. 1232 p.
9. Hayashi F. Econometrics. Princeton: Princeton University Press, 2000. 690 p.

8 Образовательные технологии

Изучение дисциплины проводится в режиме лекций и практических занятий. Практические занятия делятся на 3 типа:

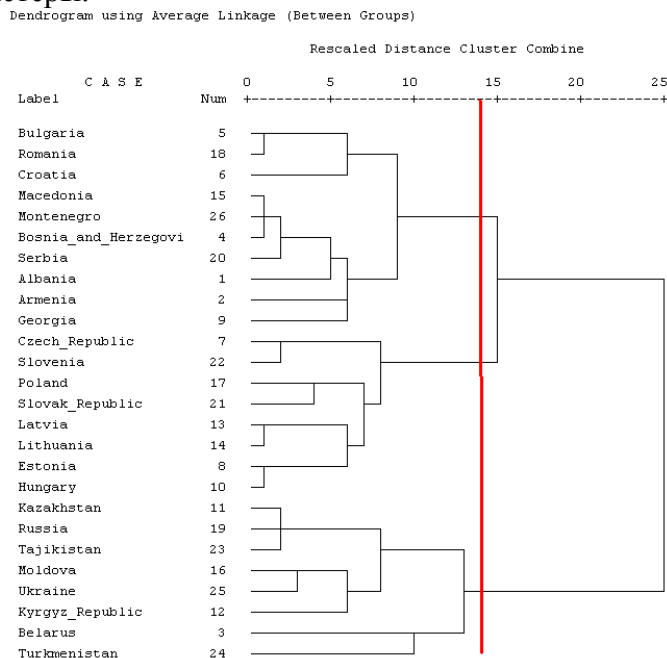
- обсуждение профессиональных текстов (статей и книг)
- решение задач «руками»
- анализ данных в пакете Stata

9 Оценочные средства для текущего контроля и аттестации студента

9.1 Вопросы для оценки качества освоения дисциплины

1. Рассмотрите любые 3 показателя социальной, политической или экономической статистики и покажите, что каждый из них следует рассматривать как случайную величину. Приведите примеры не менее 3 случайных факторов, влияющих на каждый из рассматриваемых Вами показателей.
2. Дайте определение теоретическому коэффициенту корреляции. Что он показывает? В каком диапазоне меняется?
3. Почему на практике исследователь не может вычислить сам теоретический коэффициент корреляции?
4. К. Пирсон предложил метод для приблизительного вычисления теоретического коэффициента корреляции на основе имеющихся выборочных эмпирических данных. Этот метод получил название «коэффициент корреляции К.Пирсона». Приведите его формулу. Примените ее для расчета коэффициента корреляции К.Пирсона по представленным Вам данным.
5. Что показывает коэффициент корреляции К.Пирсона?
6. Пусть коэффициент корреляции К.Пирсона равен 1 – что из этого следует? А если он равен -1?
7. Пусть коэффициент корреляции К.Пирсона равен 0.5 – что это значит?
8. Назовите не менее 2 недостатков коэффициента корреляции К.Пирсона.
9. Ч.Э. Спирмен предложил метод для приблизительного вычисления теоретического коэффициента корреляции на основе имеющихся выборочных эмпирических данных. Этот метод получил название «коэффициент ранговой корреляции Ч.Э. Спирмена». Приведите его формулу. Примените ее для расчета коэффициента корреляции Ч.Э. Спирмена по представленным Вам данным.
10. Что показывает коэффициент корреляции Ч.Э. Спирмена?
11. Пусть коэффициент корреляции Ч.Э. Спирмена равен 1 – что из этого следует? А если он равен -1?
12. Пусть коэффициент корреляции Ч.Э. Спирмена равен 0.5 – что это значит?
13. Назовите преимущества и недостатки коэффициента корреляции Ч.Э. Спирмена по сравнению с коэффициентом корреляции К. Пирсона.
14. Пусть на основе эмпирических данных Вы получили, что $R = 0.15$ (где R - коэффициент корреляции К. Пирсона). Требуется понять, можно ли на основании этого результата утверждать, что на самом деле есть корреляция между x и y . Как это сделать? Дайте ответ на поставленный вопрос, если число наблюдений $n = 25$.
15. Пусть на основе эмпирических данных Вы получили, что $\rho = 0.15$ (где ρ - коэффициент корреляции Ч.Э. Спирмена). Требуется понять, можно ли на основании этого результата утверждать, что на самом деле есть корреляция между x и y . Как это сделать? Дайте ответ на поставленный вопрос, если число наблюдений $n = 25$.
16. Для решения какой задачи применяется кластерный анализ?
17. Укажите информацию, требующуюся исследователю «на входе» для решения задачи кластеризации
18. Укажите, что является результатом кластеризации (что получается «на выходе»)
19. Какие виды кластерного анализа Вам известны
20. Как называется графическое отражения алгоритма иерархической кластеризации?
21. Сформулируйте свойства, которым должно удовлетворять любое расстояние. Какое из этих свойств выполняется не всегда (например, в психологических исследованиях)?
22. Какие виды метрики (расстояний) Вам известны?
23. Даны 2 четырехмерных наблюдения (2 точки в четырехмерном пространстве). Вычислите между ними расстояния: Евклида, Манхеттен, Чебышёва.
24. В каком случае применение евклидова расстояния не имеет под собой теоретического основания (хотя может давать разумные результаты)?

25. Назовите известные Вам алгоритмы агломерации.
26. Даны 4 трехмерных наблюдения. Реализуйте их кластеризацию на основе метода ближнего соседа (дальнего соседа, средней связи) и расстояния Евклида (Манхеттен, Чебышёва). Постройте дендрограмму.
27. Дана дендрограмма, на которой прочерчена вертикальная линия. Укажите получившиеся кластеры.



28. Какие методы приведения переменных к единой шкале Вы знаете? Даны 5 наблюдений. Приведите их к единой шкале измерения, используя линейное преобразование и z-шкалу.
29. Объясните, почему при кластеризации необходимо приводить все показатели к единой шкале измерения?
30. Укажите недостатки линейного преобразования как процедуры приведения показателей к единой шкале. Нарисуйте картинку, иллюстрирующую ситуацию, когда применение линейного преобразования окажется неудачным выбором.
31. Почему не очень осмысленно применять алгоритмы кластерного анализа для классификации объектов в одномерном или двумерном признаковом пространстве?
32. Как связаны понятия «регрессия игрека на икс» и «условное математическое ожидание игрека на икс»?
33. Какие соображения оправдывают использование линейных регрессионных моделей в качестве первого приближения заранее неизвестных нам функций условного математического ожидания игрека на икс?
34. Кто и почему впервые назвал регрессию регрессией?
35. В чем суть метода наименьших квадратов?
36. Формализуйте идею метода наименьших квадратов в ситуации парной линейной регрессии игрека на икс (в случае центрированных иксов)
37. Выведите МНК-оценки коэффициентов парной линейной регрессии (в случае центрированных иксов).
38. Как получить МНК-оценки коэффициентов парной линейной регрессии в обычном случае, зная МНК-оценки коэффициентов парной линейной регрессии в случае центрированных иксов?
39. Покажите связь между коэффициентом корреляции К.Пирсона и коэффициентом β_1 парной линейной регрессии.
40. Сформулируйте 4 условия Гаусса – Маркова.
41. Сформулируйте теорему Гаусса – Маркова

42. Назовите известные Вам свойства точечных оценок. Дайте определение каждому из этих свойств.
43. Приведите пример (например, в форме картинки), иллюстрирующий утверждение о том, что иногда несмещенные оценки хуже смещенных.
44. Почему требование состоятельности оценки очень важно и разумно?
45. Что такое гомоскедастичности остатков? Дайте формальное определение понятию. Нарисуйте картинку, отражающую ситуацию гомоскедастичности и гетероскедастичности.
46. О чем может свидетельствовать гетероскедастичность регрессионных остатков?
47. Что такое автокорреляция остатков? Когда она может возникать в регрессионной модели?
48. К каким последствиям для свойств МНК-оценок ведет нарушение каждого из условий Гаусса – Маркова?
49. Зачем в совокупность условий вероятностной модели регрессии включается условие нормальности?
50. Каково теоретическое обоснование требования нормальности распределения ошибок регрессии?
51. Укажите распределение (и его параметры) МНК-оценок коэффициентов регрессии при выполнении условий Гаусса – Маркова и условия нормальности.
52. Дано значение МНК-оценок коэффициентов парной регрессии. Известно число наблюдений. Проверьте, значимы ли коэффициенты регрессии.
53. Дана таблица результатов регрессионного анализа из некоторой статьи англоязычного политологического журнала. Укажите статистически значимые коэффициенты регрессии.
54. Даны несколько пар наблюдений над случайными величинами X и Y . Постройте уравнение регрессии Y на X и проверьте значимость коэффициентов.
55. Что такое коэффициент детерминации? Что он показывает?
56. Даны значения RSS и ESS . Найти R^2 .
57. Дано значение коэффициента корреляции между X и Y . Найти R^2 .
58. Дано значение коэффициента корреляции между \hat{Y} и Y . Найти R^2 .
59. Дано уравнение парной линейной регрессии. Даны выборочные оценки дисперсии x и y и R^2 . Найти R^2 .
60. Даны значения RSS и ESS , число наблюдений и число предикторов в регрессии. Проверить гипотезу о качестве модели на основе статистики Фишера.
61. Дано значение R^2 , число наблюдений и число предикторов в регрессии. Проверить гипотезу о качестве модели на основе статистики Фишера.
62. Дано значение t -статистики для коэффициента при предикторе в парной линейной регрессии. Проверьте гипотезу о качестве модели на основе статистики Фишера.
63. Дана таблица дисперсионного анализа. Заполнить пропуски в таблице.
64. Дана таблица выдачи регрессионного анализа. Запишите полученное уравнение регрессии.
65. Дана таблица результатов регрессионного анализа из некоторой статьи англоязычного политологического журнала. Запишите полученное уравнение регрессии.
66. Дана таблица результатов регрессионного анализа из некоторой статьи англоязычного политологического журнала. Дайте интерпретацию полученным результатам (что означают и о чем говорят полученные значения коэффициентов регрессии?).
67. Какова природа проблемы мультиколлинеарности в задачах множественной регрессии?
68. К чему ведет мультиколлинеарность?
69. Назовите хотя бы 1 способ выявления мультиколлинеарности.
70. Запишите формулу для вектора МНК-оценок коэффициентов множественной регрессии в векторно-матричной форме.
71. Опишите методы выявления влиятельных наблюдений в задаче множественной регрессии.

72. Почему возникает автокорреляция и чем она плоха в случае использования МНК-регрессии?
73. Каковы методы выявления автокорреляции в задаче регрессионного анализа?
74. В чем состоит проблема эндогенности в задаче регрессии? Каковы последствия эндогенности для МНК-регрессии?
75. В чем состоит проблема гетероскедастичности? Из-за чего она возникает и каковы ее последствия для МНК-оценок коэффициентов регрессии?
76. Опишите не менее 2 методов выявления гетероскедастичности остатков в задаче множественной регрессии.
77. Опишите известные Вам методы борьбы с гетероскедастичностью остатков в задаче множественной регрессии.
78. Назовите задачи, которые решает метод главных компонент (МГК).
79. Приведите примеры применения МГК в политологических задачах.
80. Размерность каких признаков пространств может быть снижена?
81. Когда целесообразно применять метод главных компонент? Сделайте «экспертный» вывод на основании представленной корреляционной матрицы иксов.
82. Какие критерии пригодности данных для процедур снижения размерности Вам известны? Сделайте вывод об обоснованности использования метода главных компонент на основании статистики критерия Кайзера–Мейера–Олкина.
83. Дайте определение первой главной компоненте.
84. Дайте определение главной компоненте.
85. Запишите в общем виде уравнение первой главной компоненты.
86. Дана таблица выдачи для МГК из пакета Stata. Какую долю дисперсии объясняет первая главная компонента? Какую долю дисперсии объясняют все главные компоненты?
87. Дана таблица выдачи для МГК из пакета Stata. Запишите выражение для определения меры информативности двух главных компонент.
88. Дана таблица выдачи для МГК из пакета Stata. Запишите полученное уравнение для первой главной компоненты.
89. Дана таблица выдачи для МГК из пакета Stata. Запишите полученное уравнение для второй главной компоненты.
90. Назовите оптимальные свойства главных компонент.
91. Объясните смысл характеристического уравнения для ковариационной матрицы:

$$|\Sigma - \lambda I| = 0$$
 или, в иной записи,
$$\det(\Sigma - \lambda I) = 0$$
92. Найдите компоненты собственного вектора ковариационной матрицы (I_1, \dots, I_p) , соответствующего наибольшему собственному числу λ_1 , располагая ковариационной матрицей центрированных иксов Σ .
93. Какое из p собственных значений ковариационной матрицы центрированных иксов нужно выбрать, чтобы обеспечить максимальную дисперсию первой главной компоненты?
94. Чему равен коэффициент корреляции между главными компонентами?
95. Объясните, как МГК может решить проблему мультиколлинеарности в задаче регрессионного анализа. Каковы ограничения в применении МГК в задаче регрессионного анализа?
96. Для какого из наборов данных процедура МГК, основанная на ковариационной матрице исходных признаков, не может быть применена по техническим причинам?
 - а) {рост, вес, объем талии, пол},
 - б) {рост, вес, ширина плеч},
 - в) {размер заработной платы, общее количество лет обучения, включая школу, количество рабочих часов в неделю},
 - г) {размер заработной платы, общее количество лет обучения, включая школу, количество рабочих часов в неделю, пол},

- д) {доля затрат на образование³, размер ВВП/чел., уровень младенческой смертности},
 е) {уровень безработицы, размер ВВП/чел., присутствие дефицита государственного бюджета}.
97. Опишите методы Г. Кайзера (1974 г.) и Р.Б. Кеттелла (1966 г.) определения числа главных компонент извлекаемых из набора признаков, подлежащих анализу методом главных компонент.
 98. Сформулируйте основные идеи, высказанные в первой главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 99. Сформулируйте основные идеи, высказанные во второй главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 100. Сформулируйте основные идеи, высказанные в третьей главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 101. Сформулируйте основные идеи, высказанные в четвертой главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 102. Сформулируйте основные идеи, высказанные в пятой главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 103. Сформулируйте основные идеи, высказанные в шестой главе книги: King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
 104. Сформулируйте основные идеи, высказанные в главе «Процесс исследования» следующей книги: Мангейм Дж. Б., Рич Р.К. *Политология: Методы исследования*. М.: Издательство «Весь Мир», 1997.
 105. Сформулируйте основные идеи, высказанные в главе «Создание теории: понятия и гипотезы в политологии» следующей книги: Мангейм Дж. Б., Рич Р.К. *Политология: Методы исследования*. М.: Издательство «Весь Мир», 1997.
 106. Сформулируйте основные идеи, высказанные в главе «От абстрактного к конкретному: операционализация и измерение» следующей книги: Мангейм Дж. Б., Рич Р.К. *Политология: Методы исследования*. М.: Издательство «Весь Мир», 1997.
 107. Сформулируйте основные идеи, высказанные в статье Freedman D.A. (1991). "Statistical Models and Shoe Leather." *Sociological Methodology*, Vol. 21, pp. 291 – 313.
 108. Сформулируйте основные идеи, высказанные в статье Blalock H.M., Jr. (1991). "Are There Really Any Constructive Alternatives to Causal Modeling." *Sociological Methodology*, Vol. 21, pp. 325 – 335.
 109. Опишите основные положения дискуссии между Д. Фридманом и Г. Блэлоком (на материале работ: Freedman D.A. (1991). "Statistical Models and Shoe Leather." *Sociological Methodology*, Vol. 21, pp. 291 – 313; Blalock H.M., Jr. (1991). "Are There Really Any Constructive Alternatives to Causal Modeling." *Sociological Methodology*, Vol. 21, pp. 325 – 335)
 110. Сформулируйте основные идеи, высказанные в статье Casper, Gretchen and Claudiu Tufis. (2003). "Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets." *Political Analysis*, Vol. 11, pp.196 – 203.

³ При допущении, что большее значение свидетельствует о лучшем качестве показателя.

10 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на лекциях и практических занятиях, а также самостоятельную работу студентов по следующим параметрам:

- активность студентов
- правильность ответов на вопросы преподавателя
- квалифицированность ответов, требующих привлечения профессиональных знаний
- полнота, верность и своевременность выполнения текущих домашних заданий

Оценки за работу на лекциях и практических занятиях, а также за самостоятельную работу преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на лекциях, практических занятиях и за самостоятельную работу ($O_{\text{практич.}}$) определяется перед промежуточным или итоговым контролем.

Итоговая накопленная оценка, участвующая в расчете результирующей оценки, выставляемой в диплом, рассчитывается как среднее взвешенное трех оценок: накопленной оценки за работу на лекциях, практических занятиях и за самостоятельную работу ($O_{\text{практич.}}$), оценки за контрольную работу ($O_{\text{кр}}$) и оценки за домашнее задание ($O_{\text{дз}}$).

Итоговая накопленная оценка учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{накоп.}} = (W_{\text{практич.}} \times O_{\text{практич.}}) + (W_{\text{кр}} \times O_{\text{кр}}) + (W_{\text{дз}} \times O_{\text{дз}}), \text{ где}$$

$$W_{\text{практич.}} = 0.3$$

$$W_{\text{кр}} = 0.3$$

$$W_{\text{дз.}} = 0.4$$

Способ округления накопленной оценки текущего контроля: арифметический.

В диплом выставляет результирующая оценка по учебной дисциплине, которая формируется по следующей формуле:

$$O_{\text{результ.}} = (W_{\text{накоп.}} \times O_{\text{накоп.}}) + (W_{\text{зачет}} \times O_{\text{зачет}}), \text{ где}$$

$$W_{\text{накоп.}} = 0.6$$

$$W_{\text{зачет}} = 0.4$$

Способ округления результирующей оценки по учебной дисциплине: арифметический.

Студент не может пересдать низкие результаты за текущий контроль. На пересдаче (в том числе, на комиссии) студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль.

На пересдаче и комиссии результирующая оценка выставляется по той же самой формуле, что и до пересдач.

11 Учебно-методическое и информационное обеспечение дисциплины

11.1 Базовые учебники

- *Analysis of Multivariate Social Science Data*, edited by David J. Batholomew, Fiona Steele, Irini Moustaki and Jane I. Galbraith. Boca Raton, London, New York: CRC Press, 2008. (книга доступна в библиотеке ВШЭ; электронная версия нужных глав размещена на корпоративной странице преподавателя)
- King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press. (книга доступна в библиотеке ВШЭ; электронная версия книги размещена на корпоративной странице преподавателя)
- Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Издание 2-е. Том 1: Теория вероятностей и прикладная статистика. – М.: Юнити, 2001. – 656 с. (учебник доступен в библиотеке ВШЭ; электронная версия доступна в Интернете)

11.2 Основная литература

- Blalock H.M., Jr. (1991). “Are There Really Any Constructive Alternatives to Causal Modeling.” *Sociological Methodology*, Vol. 21, pp. 325 – 335.
- Casper, Gretchen and Claudiu Tufis. (2003). “Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets.” *Political Analysis*, Vol. 11, pp. 196 – 203.
- Freedman D.A. (1991). “Statistical Models and Shoe Leather.” *Sociological Methodology*, Vol. 21, pp. 291 – 313.
- King Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press. Chapters 1 – 6.
- Мангейм Дж. Б., Рич Р.К. Политология: Методы исследования. М.: Издательство “Весь Мир”, 1997. – С. 19 – 117.

11.3 Дополнительная литература

- Ахременко А.С. Политический анализ и прогнозирование: учеб. пособие. М.: Гардарики, 2006. 333 с.
- Вербик М. Путеводитель по современной эконометрике / Пер. с англ. В.А. Банникова. Научн. ред. и предисл. С.А. Айвазяна. М.: Научная книга, 2008. 616 с.
- Доугерти К. Введение в эконометрику: Учебник. 2-е изд. М.: ИНФРА-М, 2007. 432 с.
- Дрейпер Н., Смит Г. Прикладной регрессионный анализ / Пер. с англ. М.: Издательский дом «Вильямс», 2007. 912 с.
- Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 2003. 352 с.
- Иберла К. Факторный анализ / Пер. с нем. В.М. Ивановой; Предисл. А.М. Дуброва. М.: Статистика, 1980. 398 с.
- Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. М.: Изд. дом ГУ ВШЭ, 2007. 281 с.
- Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. М.: Дело, 2007. 504 с.
- Попова О.В. Политический анализ и прогнозирование: Учебник. М.: Аспект Пресс, 2011. 464 с.
- Толстова Ю.Н. Математико-статистические модели в социологии (математическая статистика для социологов): учеб. пособие. М.: Издательский дом ГУ ВШЭ, 2008. 243 с.

- Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. М.: ИД «ФОРУМ», 2008. 368 с.
- Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; Под ред. И.С. Енюкова. М.: Финансы и статистика, 1989. 215 с.
- 18.06 Linear Algebra MIT Video Course by Gilbert Strang. Lectures 18 – 21. URL: <http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/>
- Fabrigar Leandre R., Duane T. Wegener, Robert C. MacCallum, and Erin J. Strahan. (1999). “Evaluating the Use of Exploratory Factor Analysis in Psychological Research.” *Psychological Methods*, Vol.4. No. 3, pp. 272 – 299.
- Gore Paul A., Jr. Cluster Analysis. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. Academic Press, 2000. Pp. 297 – 321. (материал размещен на корпоративной странице преподавателя)
- Greene W.H. Econometric Analysis, 7th Edition. Prentice Hall, 2011. 1232 p.
- Gujarati Damodar. Basic Econometrics. Mcgraw-Hill, 2004. Pp. 20 – 302. (материал размещен на корпоративной странице преподавателя)
- Hayashi F. Econometrics. Princeton: Princeton University Press, 2000. 690 p.
- Jolliffe I.T. 2002. Principal Component Analysis. New York: Springer. Chapters 1 – 2, 4 – 6. (книга доступна в системе электронных ресурсов библиотеки ВШЭ, а также по адресу: [http://kolxo3.tiera.ru/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20application/s/Jolliffe%20I.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)_MVsa_.pdf](http://kolxo3.tiera.ru/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20application/s/Jolliffe%20I.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf))
- Krueger J., Lewis-Beck M. (2008). “Is OLS Dead?” *The Political Methodologist*, Vol.15, No.2, pp. 2 – 4. URL: http://polmeth.wustl.edu/methodologist/tpm_v15_n2.pdf
- Simon Carl, and Lawrence Blume. (1994). *Mathematics for economists*. Cambridge: W.W.Norton & Company (книга доступна в библиотеке ВШЭ)
- Velicer W.F. and Jackson D.N. Component Analysis versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure // *Multivariate Behavioral Research*. – 1990. – Vol.25. – Pp. 1 - 28.
- Widaman K.F. Common Factor Analysis versus Principal Component Analysis: Differential Bias in Representing Model Parameters? // *Multivariate Behavioral Research*. – 1993. – Vol.28, No.3. – Pp. 263 – 311.
- Wooldridge J. Introductory Econometrics: A Modern Approach, 4th Edition. South-Western College Publishing, 2009. 865 p.

11.4 Программные средства

В ходе изучения дисциплины студенты должны научиться работе в пакете Stata.

12 Материально-техническое обеспечение дисциплины

Для изучения дисциплины студентам необходим макет Stata. Возможно также самостоятельное использование других статистических пакетов (SPSS, Statistica, R).