



**Правительство Российской Федерации**

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

Факультет социальных наук

**Программа дисциплины  
«Дополнительные главы теории вероятностей и математической  
статистики»**

для направления 41.03.04 «Политология» подготовки бакалавра

Авторы программы:

Макаров А.А., к.ф.-м.н., профессор, [amakarov@hse.ru](mailto:amakarov@hse.ru)

Сальникова Д.В., преподаватель [dsalnikova@hse.ru](mailto:dsalnikova@hse.ru)

Одобрена на заседании кафедры высшей математики

«\_\_» \_\_\_\_\_ 2017 г.

Зав. Кафедрой, к.ф.-м.н., проф. Макаров А.А. \_\_\_\_\_

Утверждена Академическим советом образовательной программы

«\_\_» \_\_\_\_\_ 201\_ г., № протокола \_\_\_\_\_

Академический руководитель образовательной программы

Локшин И.М. \_\_\_\_\_

Москва, 2017

*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 41.03.04 «Политология» подготовки бакалавра, изучающих дисциплину «Теория вероятностей и математическая статистика».

## 2 Цели освоения дисциплины

Цель освоения дисциплины «Дополнительные главы теории вероятностей и математической статистики» (2 курс бакалавриата) – овладеть знаниями в области математической статистики и многомерных методов анализа, необходимыми в социальных науках.

В соответствии с поставленной целью, курс решает следующие задачи:

1. формирование у студентов углубленных знаний о моделировании в социальных науках
2. освоение основных статистических моделей и методов анализа данных, отражающих различные социально-экономические и политические процессы и явления;
3. формирование у студентов понимания перспектив использования методов анализа данных в прикладной политологии.

## 3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать теорию вероятностей и математическую статистику в объеме программы курса «Теория вероятностей и математическая статистика» 2 курса бакалавриата, а также программы дисциплины «Математика и статистика» 1 курса бакалавриата.
- Уметь применять изученные в рамках дисциплины методы анализа данных к решению содержательных социально-экономических и политологических задач.

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 мини-эссе	Написание мини-эссе к семинарам
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 результатов применения изученных методов к решению содержательных задач
Способность применять математические методы анализа к социальным, политическим и	ОНК-7	Оценивает и интерпретирует результаты статистического анализа	Семинары: реализация методов статистики «руками»



Компетенция	Код по ФГОС/ НИУ	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции
экономическим явлениям			
Способность работать с информацией в глобальных компьютерных сетях	ИК-2	Демонстрирует способность к нахождению статистической информации в сети Интернет	Домашние задания: самостоятельный поиск данных в сети Интернет
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.	Домашние задания: изучения обязательной и дополнительной литературы на английском языке

#### 4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к математическому и естественнонаучному циклу дисциплин направления 41.03.04 «Политология» подготовки бакалавра и предназначена для студентов профиля «Политический анализ». Для указанного направления дисциплина является базовой.

Изучение данной дисциплины базируется на следующих дисциплинах:

- «Математика и статистика» (1 курс)
- «Теория вероятностей и математическая статистика» (2 курс)

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

- Знать содержательный смысл следующих понятий: функция, производная, интеграл; функция распределения и функция плотности вероятности, условная вероятность, математическое ожидание и дисперсия, квантиль, корреляция.
- Знать основы теории статистического оценивания и теории проверки статистических гипотез в объеме дисциплины «Математика и статистика» (1 курс).
- Уметь дифференцировать и вычислять определенный интеграл элементарных функций.

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

- «Многомерный статистический анализ в политологии» (3 курс)
- «Анализ данных с многоуровневой структурой» (3 курс)
- «Анализ категориальных данных в статистических пакетах» (3 курс)
- «Электоральный анализ» (4 курс)
- «Сравнительная политика» (2 курс)



## 5 Тематический план учебной дисциплины

№	Наименование разделов	Аудиторные часы		Самостоятельная работа	Всего
		Лекции	Семинары		
1	Модель множественной линейной регрессии. Эффекты взаимодействия переменных и фиктивные переменные. Регрессионный анализ в политологических исследованиях.	4	4	28	36
2	Методы классификации. Иерархический кластерный анализ. Виды расстояний и методы агломерации.	6	6	22	34
3	Факторный анализ. Разведывательный и конфирматорный факторный анализ. Введение в моделирование структурными уравнениями.	6	6	32	44
4	Задача снижения размерности. Метод главных компонент.	4	4	30	38
	<b>ИТОГО</b>	<b>20</b>	<b>20</b>	<b>112</b>	<b>152</b>

## 6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Параметры
		1	2	3	4	
Текущий	Контрольная работа				*	Письменная работа
Итоговый	Экзамен				*	Письменная работа

## 7 Содержание дисциплины

### Тема 1. Модель множественной линейной регрессии. Эффекты взаимодействия переменных и фиктивные переменные. Регрессионный анализ в политологических исследованиях.

Модель множественной линейной регрессии: повторение. Эффекты взаимодействия переменных и фиктивные переменные. Регрессионный анализ в политологических исследованиях: дизайн исследования, подходы к анализу данных, интерпретация и репликация результатов.

#### Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 9. С. 262–285.
2. Brambor Th., Clark W., Golder M. (2006). “Understanding Interaction Models: Improving Empirical Analyses.” *Political Analysis*, Vol. 14, pp. 63–82.
3. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 158-160, 277-290.

#### Дополнительная литература

4. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 166-182.
5. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
6. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 9.
7. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 347-378.

## **Тема 2. Методы классификации. Иерархический кластерный анализ. Виды расстояний и методы агломерации.**

Классификация как одна из общенаучных задач. 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 сосульчатая диаграмма (icicle plot)). Проблема устойчивости результатов. Применение кластерного анализа к классификации поведения объектов в динамике. Методы валидации результатов кластеризации: дисперсионный анализ, лепестковые диаграммы.

### **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.
2. Gore Paul A., Jr. Cluster Analysis. In Handbook of Applied Multivariate Statistics and Mathematical Modeling. Academic Press, 2000. Pp. 297 – 321.
3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. – М.: ЮНИТИ, 2001 – С. 471 – 478, 479 – 488, 488 – 518.
4. Centellas, Miguel and Mihaiela Ristei Gugiu. (2013). The Democracy Cluster Classification Index. Political Analysis, 21: 334–349.

### **Дополнительная литература**

1. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 2003. – С. 241 – 254.
2. Ким О. Дж., и др. Факторный, дискриминантный и кластерный анализ. — М.: Финансы и статистика, 1989. — с. 78-138, 139-209.
3. Rui Xu, Don Wunsch. Clustering. Wiley-IEEE Press, 2009. Pp. 1-110.

## **Тема 3. Факторный анализ. Разведывательный и конфирматорный факторный анализ. Введение в моделирование структурными уравнениями.**

Факторный анализ. Разведывательный и конфирматорный факторный анализ. Введение в моделирование структурными уравнениями.

### **Основная литература**

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**

Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 2003. – 352 с.

#### **Тема 4. Задача снижения размерности. Метод главных компонент.**

Индекс как результат снижения размерности многомерного признакового пространства. Объяснение потребности в индексах. Размерность каких признаковых пространств может быть снижена? Постановка задачи метода главных компонент (МГК). Ковариационная / корреляционная матрица как основной объект. Алгоритм МГК. 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 информативность главной компоненты, методы Г. Кайзера и Р.Б. Кеттелла.

#### **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.
2. Jolliffe I.T. 2002. Principal Component Analysis. New York: Springer. Chapters 1, 6.

### **8 Образовательные технологии**

Изучение дисциплины проводится в режиме лекций и семинаров. На семинарах регулярно проводятся текущие проверочные работы, нацеленные на выявление глубины освоения как текущего материала, так и изученного ранее. Семинары проводятся в форме лекций и практических занятий в компьютерном классе.

### **9 Оценочные средства для текущего контроля и аттестации студента**

#### **9.1 Вопросы для оценки качества освоения дисциплины**

1. Множественная регрессия: проверка гипотез про коэффициенты
2. Интерпретация коэффициентов регрессии
3. Эффекты взаимодействия переменных и фиктивные переменные. Взаимодействия категориальных переменных. Взаимодействия непрерывных переменных.
4. Для решения какой задачи применяется кластерный анализ?
5. Сформулируйте свойства, которым должно удовлетворять любое расстояние. Какое из этих свойств выполняется не всегда (например, в психологических исследованиях)?
6. Какие виды метрики (расстояний) Вам известны?
7. Какую метрику следует использовать при кластеризации количественных признаков методом Варда (Ward)?
8. Какую метрику Вы бы предложили использовать для кластеризации выборки женщин по следующим переменным: курение табака (да, нет), семейный статус (не замужем, замужем/проживаем совместно, разведена, вдова), количество детей, должность (руководящая, не руководящая)? Объясните почему. Напишите формулу для расчета.
9. Почему не очень осмысленно применять алгоритмы кластерного анализа для классификации объектов в одномерном или двумерном признаковом пространстве?
10. Назовите задачи, которые решает метод главных компонент (МГК).
11. Дайте определение главной компоненте.
12. Чему равен коэффициент корреляции между главными компонентами?



13. Для какого из наборов данных процедура МГК, основанная на ковариационной матрице исходных признаков, не может быть применена по техническим причинам?
- возраст; размер заработной платы; среднее количество часов, проверенное в интернете в день
  - размер заработной платы; общее количество лет обучения, включая школу; количество рабочих часов в неделю
  - размер заработной платы; общее количество лет обучения, включая школу; количество рабочих часов в неделю; характер занятости (самозанятость, работа по найму, госслужба)
  - доля затрат на образование; размер ВВП/чел.; уровень младенческой смертности
  - уровень безработицы; размер ВВП/чел.; размер дефицита гос. бюджета.
14. 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.
15. Чем отличаются конфирматорный и разведывательный факторный анализ?
16. Чем отличается факторный анализ от метода главных компонент?

## 10 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на семинарах, 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

Преподаватель оценивает работу студентов на семинарских занятиях: активность в дискуссиях, правильность решения задач на семинаре, правильность и своевременность решения задач в текущих домашних заданиях и прочих заданиях, которые выдаются на семинарских занятиях. Оценки за работу на семинарских занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских занятиях определяется перед итоговым контролем.

Накопленная оценка учитывает результаты студента по текущему контролю следующим образом:

$$\text{Онакопл.} = 0.6 \cdot \text{Ок/р} + 0.4 \cdot \text{Осм.р.}, \text{ где}$$

Ок/р – оценка за контрольную работу (в 10-балльной шкале)

Осм.р – оценка за работу на семинарских занятиях (в 10-балльной шкале)

Округление каждого компонента накопленной оценки производится в соответствии с правилами математики и происходит до расчета накопленной оценки. Округление накопленной оценки также производится в соответствии с правилами математики.

Результирующая оценка за дисциплину рассчитывается следующим образом:

$$\text{Результ} = 0.5 \cdot \text{Онакопл.} + 0.5 \cdot \text{Оэкзамен.}$$

Округление результирующей оценки также производится в соответствии с правилами математики. В диплом выставляется результирующая оценка по учебной дисциплине.



Экзамен проводится в письменной форме, предполагающей подготовку студентом письменного ответа на задания экзаменационной работы.

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

## 11 Учебно-методическое и информационное обеспечение дисциплины

### 11.1 Базовые учебники

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики, М.: ЮНИТИ, 1998 – 1006 с.
2. Доугерти К. Введение в эконометрику: Учебник. 2-е изд. – М.: ИНФРА-М, 1999. – 402 с.
3. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере/ Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА – М, 2003 – 544 с.
4. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Ященко И.В. Теория вероятностей и статистика. – М.:МЦНМО, 2014. – 248 с.
5. Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей: учебник для экономических и гуманитарных специальностей. – М.: МЦНМО, 2009 – 256 с.
6. Newbold P., Carlson W., Thorne B. Statistics for Business and Economics, 6th Edition. Prentice Hall. 1024 pp.
7. Stock J, Watson M. Introduction to Econometrics (2rd edition). Addison Wesley Longman; 2008.
8. Wackerly D., Mendenhall W., Scheaffer R., “Mathematical Statistics with Applications 7th Edition”, Duxbury, 2008 – 880 pp.

### 11.2 Дополнительная литература

1. Айвазян С.А. Методы эконометрики. Учебник. — М.: Инфра-М, 2010. — 512 с.
2. Вакуленко Е.С., Камалова Р.У., Кисельгоф С.Г., Смирнова Ж.И., Стукал Д.К., Хавенсон Т.Е. Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач. // Экономическая социология (электронный журнал), 2010. Т. 11. № 4. С. 190—194. ([http://ecsoc.hse.ru/data/2010/09/29/1234590525/ecsoc\\_t11\\_n4.pdf#page=190](http://ecsoc.hse.ru/data/2010/09/29/1234590525/ecsoc_t11_n4.pdf#page=190))
3. Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. — М.: Научная книга, 2008. — 616 с.
4. Единый архив социологических данных (рук. — Л. Б. Косова) // Экономическая социология (электронный журнал), 2010, т. 11, № 1, с. 133–140. ([http://www.hse.ru/data/2011/12/08/1208204985/ecsoc\\_t11\\_n1.pdf#page=133](http://www.hse.ru/data/2011/12/08/1208204985/ecsoc_t11_n1.pdf#page=133))
5. Коченков А. И., Толстова Ю.Н. Идеи латентно-структурного анализа Лазарсфельда в современной социологии// Социология, М., 2003. N.16 - С. 127 – 149
6. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. — М.: ГУ–ВШЭ, 2006. – 281 с.
7. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. 6-е изд., перераб. и доп. — М.: Дело, 2004. — 576 с.

8. Толстова Ю.Н. Математико-статистические модели в социологии (математическая статистика для социологов) : учеб. пособие. – М.: Изд. дом ГУ ВШЭ, 2008 – 244 с.
9. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – 368 с.
10. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Яценко И.В. Теория вероятностей и статистика. — М.: МЦНМО, 2004.
11. Учебно-методические материалы для преподавания курсов по применению количественного инструментария к решению социально-экономических задач (<http://www.hse.ru/jesda/mathbase/>).
12. A brief description of Stata. ([http://www.stata-press.com/manuals/stata10/guide\\_brief.pdf](http://www.stata-press.com/manuals/stata10/guide_brief.pdf)).
13. Bartels L.M. (1990). “Five Approaches to Model Specification.” *The Political Methodologist*, Vol. 3, No. 2, pp. 2–6.
14. Basics of Stata. (2008). (<http://www.ics.uci.edu/~dvd/Courses/Stata/stata.pdf>).
15. Baum C.F. *An Introduction to Modern Econometrics Using Stata*, 2006, Stata Press.
16. Bertsekas D. and Tsitsiklis J. “Introduction to Probability”, Athena Scientific, 2002 – 430 pp.
17. Brambor, T., W.R. Clark, M. Golder, 2006. Understanding Interaction Models: Improving Empirical Analysis, *Political Analysis* 14: 63-82.
18. Braumoeller B. (2006). “Explaining Variance.” *Political Analysis*, Vol.14, No.3, pp. 268–290.
19. Cameron, A. Colin and Pravin K. Trivedi. *Microeconometrics using Stata*, Revised Edition, 2010, Stata Press, 706 p.
20. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.
21. Fox, J., 2008. *Applied Regression Analysis and Generalized Linear Models*, London: Sage.
22. Fox, John. *Applied Regression Analysis and Generalized Linear Models*, 2nd Edition. Sage, 2008.
23. Fox, John. *Regression Diagnostics: An Introduction*. Sage, 1991.
24. Goertzel, T., 2002. “Myths of Murder and Multiple Regression”, *The Sceptical Inquirer* 26, 1: 19-23 (<http://crab.rutgers.edu/%7Egoertzel/mythsofmurder.htm>).
25. Graphics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/graphics.htm>).
26. Gujarati, D.N. “Basic econometrics”. New York McGraw-Hill, 2003. - 1002 с.
27. King G. (1986). “How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science.” *American Journal of Political Science*, Vol. 30, pp. 666–687.
28. King G. (1990). “When Not to Use R-Squared.” *The Political Methodologist*, Vol. 3, No. 2, pp. 9–11.
29. Kritzer, H.M., 1996. The Data Puzzle: The Nature of Interpretation in Quantitative Research, *American Journal of Political Science*, 40(1): 1-32.
30. Larocca, Roger. 2005. Reconciling Conflicting Gauss-Markov Conditions, *Political Analysis* 13:188–207.
31. Lewis-Beck, Michael S. and Andrew Skalaban. 1990. When to Use R-Squared. *The Political Methodologist* 3 (2): 11–12.
32. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press.
33. Long, J. Scott. And Jeremy Freese. *Regression Models for Categorical Dependent Variables Using Stata*, 2001, Stata Press.
34. Luskin R. (1991). “R-Squared Encore.” *The Political Methodologist*, Vol. 4, No. 1, pp. 21–23.
35. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/regression.htm>).



36. Resources to help you learn and use Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/>).

### **11.3 Программные средства**

Для изучения дисциплины студенты используют пакет для анализа данных R (RStudio).

### **11.4 Дистанционная поддержка дисциплины**

При выполнении домашних работ студентам рекомендуется пользоваться материалами

- Единого архива экономических и социологических данных НИУ ВШЭ (<http://sophist.hse.ru/>),
- Проекта «Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач» (<http://www.hse.ru/jesda/mathbase/>).

## **12 Материально-техническое обеспечение дисциплины**

Лекции и семинарские занятия проходят в компьютерном классе. На занятиях используется пакет для анализа данных R (RStudio).