



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

Факультет социальных наук

**Программа дисциплины  
«Избранные главы анализа данных: причинность и пространственное  
моделирование»**

для направления 41.03.04 «Политология» подготовки бакалавра

Авторы программы:

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

Одобрена на заседании кафедры высшей математики «\_\_\_»\_\_\_\_\_ 2016 г  
Зав. кафедрой к.ф.-м.н., проф. Макаров А.А.

Москва, 2016

*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 «Избранные главы анализа данных: причинность и пространственное моделирование».

## 2 Цели освоения дисциплины

Цель освоения дисциплины «Избранные главы анализа данных: причинность и пространственное моделирование» (4 курс бакалавриата) – овладение методами анализа длительности ожидания до наступления события.

## 3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать содержательные задачи, при решении которых обоснованно применение моделей для анализа дожития
- Уметь применять изученные в рамках дисциплины методы анализа данных к решению содержательных социально-экономических и политологических задач
- Корректно интерпретировать результаты оценивания моделей для анализа времени дожития

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
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 экономическим явлениям	ОНК-7	Оценивает и интерпретирует результаты статистического анализа	Решение задач на семинарах, интерпретация результатов оценивания моделей для анализа дожития
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:

- «Алгебра и анализ» (1 курс)
- «Дополнительные главы алгебры и анализа» (1 курс)
- «Теория вероятностей и математическая статистика» (1 курс)
- «Вероятностно-статистические модели в политологии» (2 курс)
- «Анализа политологических данных в пакете Stata» (2 курс)
- «Многомерный статистический анализ» (3 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, условное математическое ожидание, понятие классической линейной парной и множественной регрессии модели регрессии
- Знать основы теории статистического оценивания и теории проверки статистических гипотез в объеме дисциплины «Теория вероятностей и математическая статистика» (1 курс), «Вероятностно-статистические модели в политологии» (2 курс).
- Знать основы многомерного статистического анализа в объеме дисциплины «Многомерный статистический анализ» (3 курс)
- Уметь дифференцировать и вычислять определенный интеграл элементарных функций



## 5 Тематический план учебной дисциплины

№	Тема	Лекции	Семинары	Самостоятельная работа	Итого часов
1	Анализ панельных данных: повторение	2	2	4	8
2	Введение в анализ дожития	4	6	12	22
3	Непараметрические методы оценивания	4	4	24	32
4	Полупараметрические методы оценивания	8	10	34	52
5	Параметрические методы оценивания	6	4	20	30
	<b>ИТОГ</b>	<b>24</b>	<b>26</b>	<b>94</b>	<b>144</b>

## 6 Формы контроля знаний студентов

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

## 7 Содержание дисциплины

### Тема 1. Анализ панельных данных: повторение

Модель с фиксированными эффектами. Модель со случайными эффектами.

Автокорреляция: диагностика, последствия. Подход к автокорреляции в рамках классической эконометрической школы (serial correlation as a nuisance). Процедура Кохрейна-Оркатта, процедура Прайс-Уинстена. Стандартные ошибки Ньюи-Уэста.

Динамические модели. Функция импульсного отклика. Модели с включением лагированных независимых переменных. Модели с включением лагированного отклика в качестве предиктора. Процедура Ареллано – Бонда. Обобщенный метод моментов.

Пространственная корреляция: суть, условия возникновения, последствия.

#### **Основная литература:**

Beck N., Katz J. N. (2011). Modeling Dynamics in Time-Series-Cross-Section Political Economy Data. Annual Review of Political Science, Vol. 14. Pp. 331 – 352.

Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003.

#### **Дополнительная литература:**

Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008.

Ратникова Т.А., Фурманов К.К. Анализ панельных данных и данных о длительности состояний. Уч. Пособие. – М.: изд. Дом Высшей школы экономики, 2014.



Boef de S., Keele L. (2008). Taking Time Seriously. American Journal of Political Science, Vol. 52, No. 1, pp. 184 – 200.

## **Тема 2. Введение в анализ дожития**

Постановка задачи. Содержательные задачи, решаемые посредством моделей анализа дожития: примеры применения в социальных науках.

Специфика данных. Цензурирование и усечение данных.

Функция дожития (функция надежности): аналитический вид, содержательная интерпретация, свойства. Функция риска: аналитический вид, содержательная интерпретация, свойства. Связь между функцией дожития и функцией риска.

### **Основная литература:**

Kleinbaum D.G., Klein M. 2005. Survival Analysis. A Self-Learning Text. NY: Springer. Ch. 1. Pp. 2 – 44.

Hosmer D.W., Lemeshow S., May S. Applied Survival Analysis. Regression Modeling of Time to Event Data. 2<sup>nd</sup> edition. NY: John Wiley & Sons, Inc., 2008. Ch. 1, 2. Pp. 1 – 66.

### **Дополнительная литература:**

Ратникова Т.А., Фурманов К.К. Анализ панельных данных и данных о длительности состояний. Уч. Пособие. – М.: изд. Дом Высшей школы экономики, 2014. С. 247 – 260; 282 – 287.

Klein J. P., Moeschberger M.L. Survival Analysis. Techniques for Censored and Truncated Data. 2<sup>nd</sup> edition. NY: Springer, 2003. Ch. 1 – 3. Pp. 1 – 90.

Springate D. Introduction to Survival Analysis in R. URL: <http://rpubs.com/daspringate/survival>

Statistical Computing Seminars. Survival analysis with Stata. Institute for Digital Research and Education. UCLA. URL: [http://www.ats.ucla.edu/stat/stata/seminars/stata\\_survival/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/default.htm)

## **Тема 3. Непараметрические методы оценивания**

Оценка Каплана – Майера. Оценка Нельсона – Аалена. Случай цензурированных данных справа. Построение доверительных интервалов для функций дожития и интерпретация. Описательные статистики. Визуализация результатов. Сравнение функций дожития для разных групп: логарифмически-ранговый тест, критерий Уилкоксона – Гехана, критерий Тарона – Вэра.

### **Основная литература:**

Klein J. P., Moeschberger M.L. Survival Analysis. Techniques for Censored and Truncated Data. 2<sup>nd</sup> edition. NY: Springer, 2003. Ch. 4. Pp. 104 – 138.

Kleinbaum D.G., Klein M. 2005. Survival Analysis. A Self-Learning Text. NY: Springer. Ch. 2. Pp. 43 – 82.

### **Дополнительная литература:**



Ратникова Т.А., Фурманов К.К. Анализ панельных данных и данных о длительности состояний. Уч. Пособие. – М.: изд. Дом Высшей школы экономики, 2014. С. 287 – 293; 298 – 312.

Hosmer D.W., Lemeshow S., May S. Applied Survival Analysis. Regression Modeling of Time to Event Data. 2<sup>nd</sup> edition. NY: John Wiley & Sons, Inc., 2008. Ch. 2.

Gross, S.T., Lai, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. Journal of the American Statistical Association. Vol. 91. Pp. 1166 – 1180.

Statistical Computing Seminars. Survival analysis with Stata. Institute for Digital Research and Education. UCLA. URL: [http://www.ats.ucla.edu/stat/stata/seminars/stata\\_survival/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/default.htm)

#### **Тема 4. Полупараметрические методы оценивания**

Модель пропорциональных рисков Кокса. Допущения модели. Спецификация модели. Оценивание модели: метод частичного правдоподобия. Интерпретация результатов оценивания. Оценки отношения рисков. Меры качества модели. Регрессионные остатки Кокса – Снелла. Выявление влиятельных наблюдений.

Расширенная модель Кокса: включение в модель изменяющихся во времени предикторов.

##### **Основная литература:**

Hosmer D.W., Lemeshow S., May S. Applied Survival Analysis. Regression Modeling of Time to Event Data. 2<sup>nd</sup> edition. NY: John Wiley & Sons, Inc., 2008. Ch. 3 – 6.

Kleinbaum D.G., Klein M. 2005. Survival Analysis. A Self-Learning Text. NY: Springer. Ch. 3 – 4. Pp. 83 – 170.

##### **Дополнительная литература:**

Родригес Г. Модели выживаемости // Квантиль. №5. 2008. Стр. 1 – 27.

Klein J. P., Moeschberger M.L. Survival Analysis. Techniques for Censored and Truncated Data. 2<sup>nd</sup> edition. NY: Springer, 2003. Ch. 8 – 9.

#### **Тема 5. Параметрические методы оценивания**

Введение в параметрические методы оценивания в контексте анализа дожития. Экспоненциальное распределение, распределение Вейбулла, логнормальное распределение. Интерпретация результатов. Сравнение с непараметрическими методами оценивания: в каких случаях параметрические методы оказываются более предпочтительными.

##### **Основная литература:**

Ратникова Т.А., Фурманов К.К. Анализ панельных данных и данных о длительности состояний. Уч. Пособие. – М.: изд. Дом Высшей школы экономики, 2014. С. 325 – 344.

Kleinbaum D.G., Klein M. 2005. Survival Analysis. A Self-Learning Text. NY: Springer. Ch. 3 – 4. Pp. 257 – 327.

##### **Дополнительная литература:**

Hosmer D.W., Lemeshow S., May S. Applied Survival Analysis. Regression Modeling of Time to Event Data. 2<sup>nd</sup> edition. NY: John Wiley & Sons, Inc., 2008. Ch. 9.



Klein J. P., Moeschberger M.L. Survival Analysis. Techniques for Censored and Truncated Data. 2<sup>nd</sup> edition. NY: Springer, 2003. Ch. 12.

Rodriguez G. Parametric survival models.

URL: <http://data.princeton.edu/pop509/ParametricSurvival.pdf>

## 8 Образовательные технологии

Изучение дисциплины проводится в режиме лекций и семинаров.

## 9 Оценочные средства для текущего контроля и аттестации студента

### 9.1 Вопросы для оценки качества освоения дисциплины

1. Специфика данных для анализа дожития
2. Цензурированные данные: цензурирование справа, слева, интервальное
3. Усеченные данные (truncated data): усечение справа, слева, интервальное
4. Функция дожития: аналитическая запись, интерпретация
5. Функция риска: аналитическая запись, интерпретация
6. Свойства функции риска и функции дожития
7. Как взаимосвязаны между собой функция риска и функция дожития?
8. Построение доверительных интервалов для функций дожития
9. Описательные статистики функции дожития, функции риска
10. Сравнение функций дожития для разных групп: логарифмически-ранговый тест, критерий Уилкоксона – Гехана, критерий Тарона – Вэра
11. Непараметрические методы оценивания: оценка Каплана – Майера
12. Непараметрические методы оценивания: оценка Нельсона – Аалена
13. Модель пропорциональных рисков Кокса: допущения
14. Меры качества модели пропорциональных рисков Кокса
15. Оценивание модели пропорциональных рисков Кокса
16. Сравнение моделей для анализа дожития
17. Регрессионные остатки Кокса – Снелла
18. Отношение рисков: интерпретация
19. Распределение Вейбулла
20. Логнормальное распределение и экспоненциальное распределение в контексте анализа дожития
21. Выбор распределения для модели дожития

## 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

Оценки за работу на семинарах, а также за самостоятельную работу преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-тибалльной шкале за работу на семинарах и за самостоятельную работу ( $O_{\text{накопл.}}$ ) определяется перед итоговым контролем.

Итоговая накопленная оценка, участвующая в расчете результирующей оценки, рассчитывается как среднее взвешенное трех оценок: накопленной оценки за работу на семинарах и за



самостоятельную работу ( $O_{\text{семин.}}$ ), оценки контрольную работу ( $O_{\text{кр}}$ ) и оценки за домашнее задание ( $O_{\text{дз}}$ ).

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

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

$$W_{\text{семин.}} = 0.4$$

$$W_{\text{дз}} = 0.3$$

$$W_{\text{кр.}} = 0.3$$

Способ округления накопленной оценки текущего контроля: арифметический.

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

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

$$W_{\text{накоп.}} = 0.6$$

$$W_{\text{экзамен}} = 0.4$$

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 Основная литература**

Beck N., Katz J. N. (2011). Modeling Dynamics in Time-Series-Cross-Section Political Economy Data. Annual Review of Political Science, Vol. 14. Pp. 331 – 352.

Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003.

Hosmer D.W., Lemeshow S., May S. Applied Survival Analysis. Regression Modeling of Time to Event Data. 2<sup>nd</sup> edition. NY: John Wiley & Sons, Inc., 2008.

Klein J. P., Moeschberger M.L. Survival Analysis. Techniques for Censored and Truncated Data. 2<sup>nd</sup> edition. NY: Springer, 2003.

Kleinbaum D.G., Klein M. 2005. Survival Analysis. A Self-Learning Text. NY: Springer.

### **11.2 Дополнительная литература:**

Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008.

Ратникова Т.А., Фурманов К.К. Анализ панельных данных и данных о длительности состояний. Уч. Пособие. – М.: изд. Дом Высшей школы экономики, 2014.

Boef de S., Keele L. (2008). Taking Time Seriously. American Journal of Political Science, Vol. 52, No. 1, pp. 184 – 200.

Gross, S.T., Lai, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. Journal of the American Statistical Association. Vol. 91. Pp. 1166 – 1180.



Rodriguez G. Parametric survival models. URL: <http://data.princeton.edu/pop509/ParametricSurvival.pdf>

Springate D. Introduction to Survival Analysis in R. URL: <http://rpubs.com/daspringate/survival>

Statistical Computing Seminars. Survival analysis with Stata. Institute for Digital Research and Education. UCLA. URL: [http://www.ats.ucla.edu/stat/stata/seminars/stata\\_survival/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/default.htm)

## **11.2 Программные средства**

Для изучения дисциплины студенты должны обладать базовыми навыками работы в программном пакете Stata или R.

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

Пакет для анализа данных Stata. На занятиях используется проектор.