# 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 и студентов направления 040100.68 подготовки магистров по программе «Современный социальный анализ».

Программа разработана в соответствии с:

* Образовательным стандартом государственного образовательного бюджетного учреждения высшего профессионального образования ВЫСШЕЙ ШКОЛЫ ЭКОНОМИКИ, в отношении которого установлена категория «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ» <http://www.hse.ru/org/spb/orkko/structure%20standards-hse>
* Образовательной программой для направления 040100.68 «Социология»
* Рабочим учебным планом НИУ ВШЭ – Санкт -Петербург для направления 040100.68 «Социология».

# Цели освоения дисциплины

Целью дисциплины «Анализ социологических данных» является обучение формальному статистическому моделированию в социальных науках с использованием языка R.

Задачами освоения дисциплины «Анализ социологических данных» являются: 1) развитие навыка применения количесвтенного анализа данных для исследования социальной реальности; 2) приобретение опыта уверенного использования R и современных статистических методов; 3) развитие навыка интерпретации формальных моделей и получении на их основе содержательных выводов.

*Форма проведения занятий*: лекции и семинары.

# Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

***Знать:***

* Механизмы работы основных методов анализа данных;
* Различия между методами, сферы их применения;
* Достоинства и ограничения изучаемых статистических методов.

***Уметь:***

* выбрать подходящий метод для исследования;
* статистически и содержательно интерпретировать результаты анализа данных.

***Иметь навыки (приобрести опыт):***

* аналитической работы с разными данными;
* использования R и статистических пакетов для построения моделей.

В результате освоения дисциплины студент осваивает следующие компетенции:

* 1. **Системные компетенции**

|  |  |
| --- | --- |
| Код компетенции по ЕК | Формулировка компетенции |
| СК-М1 | Способен оценивать и перерабатывать освоенные научные методы и способы деятельности |
| СК-М2 | Способен изобретать и апробировать способы и инструменты профессиональной деятельности |
| СК-М3 | Способен к самостоятельному освоению новых методов исследования, изменению научного и научно-производственного профиля своей деятельности |
| СК-М8 | Способен вести профессиональную, в том числе научно- исследовательскую деятельность в международной среде |

* 1. **Профессиональные компетенции**

|  |  |
| --- | --- |
| Код компетенции по ЕК | Формулировка компетенции |
| ***Инструментальные компетенции*** | |
| ИК-М1.3.ЭД\_5.6 КД\_5.6 С | способен использовать методы и теории социальных и гуманитарных наук при осуществлении экспертной, консалтинговой и аналитической деятельности |
| ИК-М1.2.НИД\_ 7.1 С | способен самостоятельно ставить конкретные задачи научных исследований в фундаментальных и прикладных областях социологии и решать их с помощью современных исследовательских методов с использованием новейшего отечественного и зарубежного опыта и с применением современной аппаратуры, оборудования, информационных технологий |
| ИК - М1.2.ПД\_1.2\_1.3 С | способен составлять и представлять проекты научно-исследовательских и аналитических разработок в соответствии с нормативными документами |
| ИК – М7.1. С | Способен использовать социологические методы исследования для изучения актуальных социальных проблем, для идентификации потребностей и интересов социальных групп |
| ***Социально-личностные компетенции*** | |
| СЛК – М8 | способен порождать принципиально новые идеи и продукты, обладает креативностью, инициативностью |
| СЛК – М9 | Способен создавать, описывать и ответственно контролировать выполнение технологических требований и нормативов в профессиональной деятельности |

# Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к базовой части профессионального цикла по направлению 040100.68 «Социология».

Изучение данной дисциплины базируется на следующих дисциплинах:

* Современная прикладная статистика с использованием языка "R" (рекомендуется)
* Методология и методы исследований в социологии: обработка и анализ информации (обязательно)
* Английский язык, социологическая терминология и основы перевода и написания научных текстов (обязательно)

# 5. Тематический план изучения дисциплины

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Тема | Лекции | Семинары | Сам. раб | Всего |
|  | Linear regression. Interaction effects | 2 | 3 | 8 | 13 |
|  | Collinearity. Model selection using the step () function | 2 | 3 | 8 | 13 |
|  | Individual project proposals | 2 | 2 | 8 | 12 |
|  | Model diagnostics | 2 | 3 | 8 | 13 |
|  | Mid-term test | 2 | 2 | 8 | 12 |
|  | Binary logistic regression | 2 | 3 | 8 | 13 |
|  | Multinomial logistic regression | 2 | 3 | 8 | 13 |
|  | Multinomial ordered logistic regression | 2 | 3 | 8 | 13 |
|  | Poisson regression | 2 | 3 | 8 | 13 |
|  | Loglinear regression | 2 | 3 | 8 | 13 |
| **Итого** | | **20** | **28** | **80** | **128** |

# Формы контроля знаний студентов

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 4-й модуль | Параметры \*\* |
|  |  |
| Текущий  (неделя) | Контрольная работа | \* | Две контрольные работы по 90 минут |
|  |  |
| Эссе |  |  |
| Реферат |  |  |
| Коллоквиум |  |  |
| Домашнее задание |  |  |
| Итоговый | Зачет | \* | Текст итогового индивидуального проекта (10 страниц, двойной интервал, 12 pt), презентация, ответы на вопросы по проекту |

**6.1. Критерии оценки знаний, навыков**   
***Для итогового зачёта:***

Студент должен продемонстрировать:

* понимание изученных методов;
* способность применить и объяснить изученные методы.

***Оценки за зачётную работу***выставляются по 10–балльной шкале по следующим критериям:

10 – 9 – отличная работа, выполнена самостоятельно, изложена логично, продемонстрировано уверенное владение материалом, работа грамотно оформлена и результаты анализа грамотно представлены.

8 – очень хорошая работа, продемонстрированы навыки самостоятельного анализа материала и аргументирования, но есть маленькие недочеты, не повлиявшие на правильность выводов

7 – 6 – хорошая работа, продемонстрированы навыки анализа материала и самостоятельного мышления, но есть существенные недочеты в постановке проблемы, графиках или результатах анализа.

5 – 4 – удовлетворительная работа, оформленная правильно, но нет достаточного понимания или интерпретации результатов. Возможны ошибки.

3 – 1 – неудовлетворительная работа, отсутствие понимания, допущены грубые ошибки (незачет).

0 – неудовлетворительная работа, отсутствие ответа или неоригинальный текст (незачет).

**6.2. Порядок формирования оценок по дисциплине**

## Преподаватель не оценивает работу студентов на семинарских и практических занятиях: активность студентов в дискуссиях, правильность решения задач на семинаре, знание материала.

Накопленная оценка за текущий контроль учитывает результаты студента по текущему контролю следующим образом:

О*накопленная*= 0,5\*Ок.р.1 + 0,5\* Ок.р.2

Результирующая оценка за дисциплину рассчитывается следующим образом:

Орезульт =0,5\* Онакопл + 0,5 \*·Озач

Способ округления накопленной оценки результирующей (итогового) контроля в форме зачета: арифметический.

Оценка за итоговый контроль (зачёт) **блокирующая,** при неудовлетворительной итоговой оценке результирующая оценка - **неудовлетворительная**.

На пересдаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль.

# 7. Содержание дисциплины

Основные темы курса и рекомендуемая литература

Тема 1: Linear regression. Interaction effects

Тема 2: Collinearity. Model selection using step () function

Тема 3: Individual project proposals

Тема 4: Model diagnostics

Тема 5: Mid-term test

Тема 6: Binary logistic regression

Тема 7: Multinomial logistic regression

Тема 8: Multinomial ordered logistic regression

Тема 9: Poisson regression

Тема 10: Log-linear regression

Тематика лекций, семинарских занятий и литература к ним

**Тема 1: Linear regression. Interaction effects**

Общий смысл регрессионного уравнения. Описанный процент дисперсии, R2 модели. Значимость коэффициентов. Интерпретация интерактивных эффектов в модели.

Литература

An Introduction to Statistical Learning, with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (Springer 2013)

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 2: Collinearity. Model selection using the step () function**

Проблемы, связанные с коллинеарностью, диагностика коллинеарности. Выбор лучшей модели.

Литература

An Introduction to Statistical Learning, with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (Springer 2013)

Fox J. An R and S-Plus companion to applied regression.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 3: Individual project proposals**

**Тема 4: Model diagnostics**

Литература

An Introduction to Statistical Learning, with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (Springer 2013)

Fox J. An R and S-Plus companion to applied regression.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 5: Mid-term test**

**Тема 6: Binary logistic regression**

Литература

An Introduction to Statistical Learning, with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (Springer 2013)

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 7: Multinomial logistic regression**

Литература

Fox J. An R and S-Plus companion to applied regression.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 8: Multinomial ordered logistic regression**

Литература

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 9: Poisson regression**

Литература

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**Тема 10: Loglinear regression**

Литература

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

# 8. Образовательные технологии

Занятия проводятся в форме лекций, включающих в себя демонстрацию R-скриптов и практических занятий в компьютерном классе.

# 9. Оценочные средства для текущего контроля и аттестации студента

**9.1. Тематика заданий текущего контроля**

**Контрольная работа**

* 1. На предоставленной базе данных, проведите отбор наилучшей модель при помощи функции step(), продиагностируйте модель.
  2. Для предложенной переменной, определите подходящий вид регрессии, постройте модель, проинтерпретируйте получившиеся коэффициенты статистически и содержательно.

**9.2. Вопросы для оценки качества освоения дисциплины**

1. Какие недостатки существуют у бинарной логистической регрессии?
2. Каким образом интерпретируются интерактивные эффекты?
3. Опишите способы отбора наилучшей регрессионной модели?
4. Какие критерии существуют для оценки качества модели?

# 10. Учебно-методическое и информационное обеспечение дисциплины

**10.1 Базовый учебник**

Базового учебника на английском языке, который удовлетворяет требованиям НИУ-ВШЭ, не существует. В основном материал курса состоит из материала книг, которые доступны через электронные ресурсы нашей библиотеки.

**10.2 Основная литература**

An Introduction to Statistical Learning, with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (Springer 2013)

Fox J. An R and S-Plus companion to applied regression.

Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical models, Cambridge University Press 2006.

Venables W.N., Ripley B.D. Modern Applied Statistics with S. (Springer 2002).

**10.5 Программные средства**

Студенты пользуются свободнораспространяемой программной средой R.

**10.6 Дистанционная поддержка дисциплины**

Доступ к материалам курса для студентов производится через LMS (Learning Management System)

# Материально-техническое обеспечение дисциплины

Практические занятия проводятся в компьютерном классе. Для лекций используется проектор, презентации, а также примеры скриптов, реализованные в R.