

Курс «Статистика»

Автор: И.В. Щуров

Общая информация о дисциплине

Курс «Статистика» рассчитан на 1 модуль по 4 акад. часа в неделю и является обязательным для студентов 3-го курса экономического бакалавриата. Читается в третьем модуле. Все математические понятия и навыки, необходимые для этого курса, содержатся в курсах «Математический анализ», «Линейная алгебра», «Теория вероятностей», «Эконометрика».

Цели и задачи курса

Задача курса состоит в получении навыков обработки и статистического анализа данных с применением специализированных компьютерных инструментов. В качестве основных инструментов используются статистические библиотеки языка *Python* и специализированный пакет *R*. Использование языка *Python* обусловлено тем фактом, что существенной частью любой практической задачи по анализу данных является их извлечение и приведение к виду, пригодному для применения статистических методов. Эта часть обычно значительно упрощается при использовании языков программирования общего назначения. Языки *Python* и *R* являются стандартом де-факто в современной *Data Science*.

Структура и содержание дисциплины

1. Базовые понятия программирования. Переменные, операторы, управляющие конструкции, циклы.
2. Типы данных в *Python*. Функции и элементы функционального программирования.
3. Извлечение данных из различных источников (HTML, XML, CSV, JSON).
4. Дескриптивная статистика. Построение графиков и диаграмм в *Python* и *R*.
5. Математические инструменты *Python*. Библиотеки *numpy* и *Scipy*.
6. Статистические инструменты *Python*. Библиотеки *pandas*, *Statsmodels*, *scikit-learn*.
7. Проверка гипотез и построение статистических моделей в *R*.

Описание методологии

Занятия проходят в компьютерном классе. Примерно половину времени преподаватель рассказывает новый материал, одновременно демонстрируя его на экране, параллельно слушатели воспроизводят действия преподавателя на своих компьютерах. Оставшееся время

студенты самостоятельно решают задачи, имея возможность тут же получить консультацию преподавателя. Существенной частью образовательного процесса является самостоятельное решение домашних заданий.

Курс может осваиваться удалённо. Преподаватель публикует на странице курса необходимые для этого учебные материалы.

Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература (обязательная)

1. Официальная документация по языку *Python* и используемым библиотекам.
2. Официальная документация по пакету *R*.

Дополнительная литература (необязательная)

1. Ajith Kumar B.P. Python for Education. (<http://www.iuac.res.in/~elab/phoenix/docs/Maths/mapy.pdf>)

Программное обеспечение и Интернет-ресурсы

1. [Python](#).
2. [R](#).
3. Система [Jupyter](#).
4. Пакеты для [математических вычислений в Python](#).
5. Библиотека [pandas](#).
6. Библиотека [Statsmodels](#).
7. Библиотека [scikit-learn](#).

Формы и методы контроля знаний студентов

Предусмотрено две основные формы контроля знаний студентов: домашние задания и проектная работа.

В курсе запланировано проведение 6-7 обязательных домашних работ. Общий вес текущего контроля в итоговой оценке — 80%.

Также в ходе курса студенты готовят индивидуальный или групповой проект, посвященный решению какой-либо практически-значимой задачи с использованием знаний и навыков, полученных в ходе курса. Вес проекта в итоговой оценке составляет 20%.

В случае получения неудовлетворительной итоговой оценки, студент пересдаёт курс путём

написания контрольной работы. Контрольная работа на пересдачу пишется 120 минут в компьютерном классе. В ходе выполнения работы можно пользоваться любыми средствами, кроме помощи другого человека.