



Национальный исследовательский университет «Высшая школа экономики»  
Программа дисциплины «Прикладная статистика»  
для образовательной программы «Анализ данных в биологии и медицине» направления подго-  
товки магистра 01.04.02 Прикладная математика и информатика

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

Факультет компьютерных наук

**Рабочая программа дисциплины «Прикладная статистика»**

для образовательной программы «Анализ данных в биологии и медицине» направле-  
ния подготовки 01.04.02 Прикладная математика и информатика  
уровень магистр

Разработчик программы

С.А.Спирин, доцент кафедры технологий моделирования сложных систем, [sspirin@hse.ru](mailto:sspirin@hse.ru)

Одобрена на заседании кафедры технологий моделирования сложных систем

Зав. Кафедрой А.Н.Соболевский

Утверждена Академическим руководителем программы

Академический руководитель образовательной программы

М.С.Гельфанд

Москва, 2018

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



## 1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих курс «Прикладная статистика», учебных ассистентов и студентов направления подготовки 01.04.02 Прикладная математика и информатика, обучающихся по образовательной программе «Анализ данных в биологии и медицине».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ 01.04.02 Прикладная математика и информатика;
- Образовательной программой «Анализ данных в биологии и медицине».
- Объединенным учебным планом университета по образовательной программе «Анализ данных в биологии и медицине», утвержденным в 2018г.

## 2 Цели освоения дисциплины

Целями освоения дисциплины «Прикладная статистика» является знакомство с основными видами статистики;

## 3 Компетенции обучающегося, формируемые в результате освоения дисциплины

1. Понимание русской и английской терминологии, относящейся к тематике курса, способность читать и понимать литературу по теме;

## 4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к блоку обязательных дисциплин

Изучение данной дисциплины базируется на следующих дисциплинах:

- Математика
- Комбинаторика и теория вероятностей

## 5 Тематический план учебной дисциплины

| №  | Название раздела   | Общих часов | Аудиторные часы |          | Самостоятельная работа |
|----|--|-------------|-----------------|----------|------------------------|
|    |  |             | Лекция          | Семинары |                        |
| 1. | Планирование статистического эксперимента                                    | 13          | 1               | 4        | 8                      |
| 2. | Описательные статистики  | 13          | 1               | 4        | 8                      |
| 3. | Классические дискретные и непрерывные распределения в статистических задачах | 13          | 1               | 4        | 8                      |
| 4. | Проверка гипотез   | 13          | 1               | 4        | 8                      |
| 5. | Непараметрические тесты  | 13          | 1               | 4        | 8                      |
| 6. | Распределение хи-квадрат   | 11          | 1               | 2        | 8                      |
| 7. | Теория точечного оценивания  | 11          | 1               | 2        | 8                      |
| 8. | Принцип максимального правдоподобия  | 11          | 1               | 2        | 8                      |



|     | при получении оценок параметров  |     |    |    |     |
|-----|--|-----|----|----|-----|
| 9.  | Интервальное оценивание  | 11  | 1  | 2  | 8   |
| 10. | Распределение Стьюдента и условия его применимости                               | 11  | 1  | 2  | 8   |
| 11. | Свойства выборочного стандартного отклонения и интервальное оценивание дисперсии | 11  | 1  | 2  | 8   |
| 12. | Модель парной регрессии  | 11  | 1  | 2  | 8   |
| 13. | Стандартная ошибка коэффициента наклона в парной регрессии                       | 9   | 1  | 2  | 6   |
| 14. | Коэффициент корреляции и коэффициент детерминации, их интерпретация              | 9   | 1  | 2  | 6   |
| 15. | Дисперсионный анализ (ANOVA)   | 10  | 2  | 2  | 6   |
| 16. | Множественная линейная регрессия   | 10  | 2  | 2  | 6   |
| 17. | Поправки на множественное тестирование при проверке гипотез                      | 10  | 2  | 2  | 6   |
|     | ВСЕГО  | 190 | 20 | 44 | 126 |

## 6 Формы контроля знаний студентов

| Тип контроля | Форма контроля     | 2 год |   |    |   | Параметры                     |
|--------------|--------------------|-------|---|----|---|-------------------------------|
|              |                    | 1     | 2 | 3  | 4 |                               |
| Текущий      | Контрольная работа |       | * | ** |   | Письменная контрольная работа |
| Итоговый     | Экзамен            |       |   | *  |   | Устный экзамен                |

## 7 Критерии оценки знаний, навыков

В рамках курса студенты должны выполнить три контрольные работы и сдать экзамен. Каждая контрольная и экзамен оцениваются по 10-балльной шкале. Экзамен содержит набор теоретических вопросов и задач по различным темам курса. Суммарная оценка получается усреднением средней оценки за три контрольные и оценки экзамена, с округлением полученного дробного числа вверх.

## 8 Содержание дисциплины

1. Планирование статистического эксперимента. Эксперимент и наблюдательное исследование. Простая случайная выборка. Отклонения в выборках и их классификация.

2. Описательные статистики. Медиана, квартили, выборочное среднее, дисперсия. Способы графического изображения выборок: гистограмма, «ящик с усами» и другие. Способы представления двумерных данных.

3. Классические дискретные и непрерывные распределения в статистических задачах (биномиальное, пуассоновское, нормальное, экспоненциальное, равномерное и другие распределения). Приближение биномиального распределения пуассоновским. Аппроксимация биномиального распределения нормальным и поправка на непрерывность. Применимость аппроксимаций.

4. Проверка гипотез. Ошибки первого и второго рода. Уровень значимости и мощность критерия. Р-значение и его интерпретация. Проверка гипотез с использованием тестовых статистик. Критические значения. Биологические задачи.



5. Непараметрические тесты. Критерий знаков. Критерий Уилкоксона. Критерий Манна – Уитни.
6. Распределение хи-квадрат. Условия применимости в задачах. Критерий согласия Пирсона. Таблицы сопряженности. Выборки без возвращения и точный тест Фишера. Биологические задачи.
7. Теория точечного оценивания. Состоятельность, несмещенность и эффективность оценок. Среднеквадратичное отклонение. Теорема Штейнера. Математическое ожидание и дисперсия выборочного среднего. Примеры несмещенных и эффективных оценок.
8. Принцип максимального правдоподобия при получении оценок параметров. Элементы байесовой статистики.
9. Интервальное оценивание. Доверительные интервалы и их интерпретация. Уровень доверия. Согласованность гипотез и доверительных интервалов. Стандартная ошибка. Поправка на конечный размер популяции. Интервальное оценивание пропорций и разности пропорций. Оценки для среднего и разности средних в случае известных и неизвестных стандартных отклонений.
10. Распределение Стьюдента и условия его применимости. Одновыборочный критерий Стьюдента (t-тест). Двухвыборочный критерий Стьюдента в случае равных дисперсий.
11. Свойства выборочного стандартного отклонения и интервальное оценивание дисперсии. Распределение Фишера. Свойства квантилей. Проверка гипотез с использованием распределения Фишера.
12. Модель парной регрессии. Метод наименьших квадратов. Интерпретация коэффициента наклона. Остатки и их свойства. Выбросы и влиятельные значения.
13. Стандартная ошибка коэффициента наклона в парной регрессии. Доверительные интервалы и тестирование гипотез в модели регрессии. Доверительные и прогнозные интервалы.
14. Коэффициент корреляции и коэффициент детерминации, их интерпретация. Разложение суммы квадратов вычетов. Статистические свойства оценок коэффициента корреляции. Применение к анализу дисперсии зависимых выборок.
15. Дисперсионный анализ (ANOVA).
16. Множественная линейная регрессия. Многомерное нормальное распределение и предположения о нормальности. Остаточная дисперсия. Коэффициент детерминации и его интерпретация. Мультиколлинеарность.
17. Поправки на множественное тестирование при проверке гипотез. Контроль числа ошибочных отклонений нулевой гипотезы (FDR).

## 9 Порядок формирования оценок по дисциплине

| Тип контроля | Форма контроля     | Параметры                           |
|--------------|--------------------|-------------------------------------|
| Текущий      | Контрольная работа | Три контрольные, каждая с весом 1/6 |
| Итоговый     | Экзамен            | Устный экзамен, с весом 1/2         |

## 9 Рекомендуемая литература

### Литература:

1. Гнеденко Б.В. Курс теории вероятностей. – М: Наука, 1988. – 446 с.
2. Венцель Е.С. Теория вероятностей. – М.: Высшая школа, 1999. – 576
3. Diez, Barr, Cetinkaya-Rundel, Dorazio. Advanced High School Statistics (2015)
4. Phillip I. Good, James W. Hardin. Common Errors in Statistics (and How to Avoid Them) (2003)
5. М.Б. Лагутин. Наглядная математическая статистика (2007)