**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования**

**Национальный исследовательский университет
"Высшая школа экономики"**

**Факультет Бизнес-информатики**

**Программа дисциплины**

**Анализ данных**

**для направления 38.04.05 «Бизнес- информатика»
подготовки магистра**

для магистерской программы «Электронный бизнес»

Автор программы: **к. ф.-м. н., доцент Горяинова Елена Рудольфовна**

Одобрена на заседании кафедры высшей математики на факультете экономики 28.08.2014 г.

 Зав. кафедрой Алескеров Ф.Т.

Рекомендована секцией УМС «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_ 20 г

 Председатель

Утверждена Ученым Советом факультета экономики «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_20 г.

 Ученый секретарь

**Москва, 2014**

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

# 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 **38.04.05 «Бизнес- информатика»**, обучающихся по магистерской программе **«Электронный бизнес»** и изучающих дисциплину «Анализ данных».

Программа разработана в соответствии с:

* Образовательным стандартом государственного образовательного бюджетного учреждения высшего профессионального образования «Государственный университет – Высшая школа экономики», в отношении которого установлена категория «Национальный исследовательский университет»;
* Рабочим учебным планом университета подготовки магистра по направлению **38.04.05 «Бизнес- информатика»,** магистерская программа **«Электронный бизнес»**, утвержденным в 2014 г.

# Цели освоения дисциплины

Целью освоения дисциплины «Анализ данных» является обучить студентов применять основные модели и методы математической статистики для обработки реальных социально-экономических данных.

# Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения учебной дисциплины «Анализ данных» обучающийся должен

**знать:**

**-**алгоритм проверки статистических гипотез;

- основные методы проверки однородности экспериментальных данных;

- методы построения доверительных интервалов параметров случайных величин;

- методы проверки независимости признаков, измеренных в различных шкалах;

- коэффициенты, измеряющие силу связи показателей;

- принципы сравнения статистических критериев;

**уметь:**

– строить математические модели, адекватно описывающие социально-экономические явления;

- использовать статистические критерии для проверки гипотез относительно наблюдаемых случайных данных;

- оценивать неизвестные параметры статистической модели;

- готовить научно-технические отчеты и научные публикации по результатам выполненных исследований;

- обрабатывать статистическую информацию в условиях априорной стохастической неопределённости;

- проводить сравнительный анализ различных статистических методов;

**владеть:**

- основными определениями, методами и алгоритмами анализа данных, содержащих случайную составляющую;

- стандартными инструментариями обработки статистической информации.

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** |
| --- | --- | --- | --- |
| Системная | СК-2 | Способен создавать новые теории, изобретать новые способы и инструменты профессиональной деятельности | Стандартные (лекционно-семинарские). Самостоятельные внеаудиторные занятия |
| Системная | СК-3 | Способен к освоению новых методов исследований, изменению научного и производственного профиля своей деятельности | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Системная | СК-6 | Способен анализировать, верифицировать, оценивать полноту информации в ходе профессиональной деятельности, при необходимости восполнять и синтезировать недостающую информацию | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-9 | Способен организовать научно-исследовательскую деятельность | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-10 | Способен анализировать и воспроизводить смысл междисциплинарных текстов с использованием языка и аппарата прикладной математики | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-11 | Способен создавать междисциплинарные тексты с использованием языка и аппарата прикладной математики | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-12 | Способен публично представлять результаты профессиональной деятельности | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-14 | Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат прикладной математики при решении междисциплинарных проблем | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-16 | Способен использовать в профессиональной деятельности знание естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-17 | Способность строить и решать математические модели в соответствии с направлением подготовки и специализацией  | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-18 | Способен понимать и применять в исследовательской и прикладной деятельности современный математический аппарат  | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-19 | Способен в составе научно-исследовательского и производственного коллектива решать задачи профессиональной деятельности в соответствии с профилем подготовки, общаться с экспертами в других предметных  | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |
| Инструментальные | ПК-20 | Способен применять в исследовательской и прикладной деятельности современные языки программирования и манипулирования данными, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии и т.п. | Стандартные (лекционно-семинарские).Самостоятельные внеаудиторные занятия |

# Место дисциплины в структуре образовательной программы

Для специализации «Электронный бизнес» настоящая дисциплина является адаптационной дисциплиной, которая согласно пункту 5.5 «Регламента планирования и организации дисциплин по выбору и факультативов», утвержденным ученым советом НИУ ВШЭ 24 июня 2011 года ([http://www.hse.ru/docs/33592234.html)](http://www.hse.ru/docs/33592234.html%29), является дисциплиной по выбору для выпускников НИУ ВШЭ по данному направлению обучения и обязательной дисциплиной для прочих студентов.

Изучение данной дисциплины базируется на следующих дисциплинах:

* Математический анализ;
* Геометрия и алгебра;
* Теория вероятностей и математическая статистика.

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

* Знаниями основных определений и теорем перечисленных выше дисциплин;
* Навыками решения типовых задач этих дисциплин.

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

* Эконометрика;
* Компьютерное моделирование;
* При подготовке магистерской работы.

# Тематический план учебной дисциплины

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***№*** | ***Название раздела*** | ***Всего часов***  | ***Аудиторные часы*** | ***Самостоятельная работа*** |
| ***Лекции*** | ***Семинары*** | ***Практические занятия*** |
|  | Проверка статистических гипотез. Репрезентативная выборка  | 18 | 4 | 4 |  | 10 |
|  | Исследование однородности двух выборок.  | 34 | 4 | 4 |  | 26 |
|  | Анализ статистической взаимосвязи случайных величин  |  46 | 6 | 6 |  | 34 |
|  | Регрессионный анализ |  10 | 2 | 2 |  | 6 |
|  | Итого | 108 | 16 | 16 |  | 76 |

# Формы контроля знаний студентов

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 1 год | Параметры  |
| 1 | 2 | 3 | 4 |
| Текущий(неделя) | Контрольная работа | 4 |  |  |  | Письменная работа на 30 минут |
| Итоговый | Экзамен | э |  |  |  |  письменная работа 80 минут |

## 6.1 Критерии оценки знаний, навыков

Для прохождения контроля студент должен, как минимум, продемонстрировать знания основных определений; умение решать типовые задачи, разобранные на семинарских занятиях.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

**6.2 Порядок формирования оценок по дисциплине**

Результирующая оценка за дисциплину рассчитывается следующим образом:

*Орезульт = 0.2\* Онакопл + 0.8 \*·Оэкз*

где О*накопленная*=  *Отекущий* , а *Отекущий* = *Ок/р*

Способ округления результирующей оценки – арифметический.

# Содержание дисциплины

***Раздел I.***

**Репрезентативная выборка (Л.-4ч., С.-4ч., СРС-10ч.: проработка лекционного материала -6ч., подготовка к семинарским занятиям -4ч.)**

Основные понятия математической статистики (выборка, оценка, свойства статистических оценок, квантиль, простые и сложные статистические гипотезы, статистический критерий, доверительная и критические области, уровень значимости, функция мощности). Биномиальный критерий. Понятие репрезентативной выборки. Выбор с возвращением, выбор без возвращения. Определение объёма репрезентативной выборки для однородной и стратифицированной генеральной совокупности.

***Литература:*** *основная*: [1] глава 1, дополнительная: [13] глава 2.

***Раздел 2.***

 **Исследование однородности двух выборок. (Л.-4ч., С.-4ч., СРС – 26ч. : проработка лекционного материала -12ч., подготовка к семинарским занятиям -14ч.)**

Понятие об однородности выборок.

Выявление неоднородности, связанной со сдвигом (классический критерий Стьюдента, критерий Вилкоксона) или масштабом (классический F- критерий, критерий Ансари-Брэдли). Проверка однородности против альтернатив общего вида (критерий Колмогорова – Смирнова).

Понятие робастности (устойчивости) статистической оценки.

***Литература:*** *основная*: [1], с. 59-89; дополнительная: [2],[4],[8],[15]

 ***Раздел 3.***

**Анализ статистической взаимосвязи случайных величин (Л.-6ч., С.-6 ч., СРС -34ч. .: проработка лекционного материала -18ч., подготовка к семинарским занятиям -16ч.)**

Шкалы измерений (количественная, порядковая, номинальная).

Исследование связи между номинальными переменными (таблица сопряженности признаков, критерий хи-квадрат, меры связи признаков: среднеквадратической сопряженности, Пирсона, Крамера). Меры прогноза Гутмана.

Исследование связи между порядковыми переменными (ранговый коэффициент корреляции Спирмена, коэффициент согласованности Кендалла, коэффициент конкордации).

Случайные векторы. Ковариационная матрица. Независимость и некоррелированность компонент случайного вектора. Выборочный коэффициент корреляции.

Анализ структуры и тесноты связи между количественными переменными. Критерий хи-квадрат. Критерий, основанный на выборочном коэффициенте корреляции.

***Литература:*** *основная*: [1] с. 113-152; дополнительная: [2], [3],[5],[7],[15].

***Раздел 4.***

 **Регрессионный анализ (Л.-2ч., С.-2 ч., СРС -6ч. .: проработка лекционного материала -4ч., подготовка к семинарским занятиям -2ч.)**

Задача линейной регрессии. Проблема выбора вида функции регрессии.

Методы оценивания параметров.

Свойства МНК оценок.

Некоторые критерии проверки адекватности в нормальной регрессионной модели.

***Литература:*** *основная*: [1], с. 152-169; дополнительная: [2],[6],[14]

# Оценочные средства для текущего контроля и аттестации студента

## Тематика заданий текущего контроля

## 9.2 Вопросы для оценки качества освоения дисциплины

**Тема 1.**

1. Что такое статистическая гипотеза?

2. В чем состоят ошибки I и II рода?

3. Дайте определение квантили. Чему равна 0,05-квантиль стандартного гауссовского распределения, если 0,95-квантиль этого распределения равна 1,65?

4. Каков порядок проверки параметрических статистических гипотез?

5. Какую выборку называют репрезентативной?

6. В каких случаях следует делать поправку на выбор без возвращения?

7. Что такое стратифицированная выборка?

**Тема 2.**

1. Какие выборки называют однородными?

2. Назовите основные типы неоднородности выборок.

3. Опишите условия применимости классических и ранговых критериев для проверки гипотезы об однородности.

4. Какие преимущества и какие недостатки имеют ранговые критерии по сравнению с классическими?

5. Какие критерии применяют для проверки гипотезы об однородности двух выборок?

**Тема 3.**

1. Опишите основные типы шкал измерений и допустимые преобразования в этих шкалах.

2. Что такое таблица сопряженности признаков?

3. Дайте определение независимости признаков, измеряемых в номинальной шкале.

4. Как проверить гипотезу о независимости признаков в номинальной шкале?

5. Что такое коэффициент корреляции? Каковы его основные свойства?

6. Как проверить гипотезу о некоррелированности признаков?

7. В каком случае проверка некоррелированности наблюдений эквивалентна проверке независимости?

**Тема 4.**

1. В чем состоит задача линейной регрессии?

2. В чем состоит идея метода наименьших квадратов (МНК)?

3. Какие методы оценивания параметров регрессии вам известны?

4. Какими свойствами обладает МНК-оценка параметров регрессии?

## 9.3 Примеры заданий итогового контроля

***Вариант письменной экзаменационной работы.***

1. В городскую администрацию поступила просьба о переносе автобусной остановки. Сколько жителей микрорайона надо опросить, чтобы с точностью до 0,03 оценить долю людей, считающих перенос остановки целесообразным? Результат должен быть гарантирован на уровне доверия 0,95. Известно, что в данном микрорайоне проживает 4000 человек.

2. Уровень гистамина в мокроте у 7 курильщиков, склонных к аллергии, составил (в микрограммах): 102,4; 100,0; 67,6; 65,9; 64,7; 39,6; 31,2, а у курильщиков, несклонных к аллергии: 48,1; 45,5; 41,7; 35,4; 29,1; 18,9; 58,3; 66,8; 71,3; 94,3. Верно ли предположение о том, что уровень гистамина у курильщиков, подверженных аллергии, выше, чем у неаллергенов? Принять уровень значимости равным 0,05.

3. В результате проведенного исследования было установлено, что у 309 светлоглазых мужчин жены также имеют светлые глаза, а у 214 светлоглазых мужчин жены темноглазые. У 119 темноглазых мужчин жены также темноглазые, а у 132 темноглазых мужчин жены светлоглазые. Имеется ли зависимость между цветом глаз мужей и их жен?

4. В таблице представлены данные за 1995 год показателей X (ВВП в паритетах покупательной способности) и Y (коэффициент детской смертности в %) для следующих стран: Бурунди, Чад, Индия, Египет, Мексика, Бразилия, Республика Корея, Канада, США, Швейцария.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 2.3 | 2.6 | 5.2 | 12.2 | 23.7 | 20 | 42.4 | 78.3 | 100 | 95.9 |
| Y | 98 | 117 | 68 | 16 | 33 | 44 | 10 | 6 | 8 | 6 |

Считая , что наблюдения имеют гауссовское распределение, выясните являются ли признаки X и Y зависимыми.

1. Используя данные предыдущей задачи, оцените по методу наименьших квадратов коэффициенты  линейной регрессии вида $Y=a\_{0}+a\_{1}X+ε, где Eε=0$.

# Учебно-методическое и информационное обеспечение дисциплины

## Базовый учебник

1.Горяинова Е.Р., Панков А.Р., Платонов Е.Н. Прикладные методы анализа статистических данных. М.: Изд. Дом ВШЭ,2012, 312 с.

.

## 10.2 Основная литература

2. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: Инфра. – М, 2003.

## 10.3 Дополнительная литература

3. Интерпретация и анализ данных в социологических исследованиях (под редакцией В.Г. Андреенкова и Ю.Н.Толстовой). – М.: Наука, 1987.

4. Холлендер М., Вулф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983.

5. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей. Справочное издание под ред. Айвазяна С.А. – М.: Финансы и статистика, 1985.

6. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высшая школа, 1992.

7. Кендел М. Ранговые корреляции. – М.: Статистика, 1975.

8. Леман Э. Проверка статистических гипотез. – М.: Наука, 1964.

9. Хьюбер П. Робастность в статистике. – М.: Мир, 1984.

10 Хампель Ф., Рончетти Э., Рауссей П., Штаэль В. Робастность в статистике. Подход на основе функции влияния. – М.: Мир, 1989.

11. Дэниел К. Применение статистики в промышленном эксперименте. – М.: Мир, 1979.

13. Кокрен У. Методы выборочного исследования. – М.: Статистика, 1976

14. Хеттманспергер Т. Статистические выводы, основанные на рангах. – М.: Финансы и статистика,1987.

15. Крыштановский А.О. Анализ социологических данных. - М.: ИД ГУ ВШЭ, 2007.

# Материально-техническое обеспечение дисциплины

Не требуется