

**Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Московский физико-технический институт  
(национальный исследовательский университет)»**

**УТВЕРЖДЕНО**  
**Проректор по учебной работе**

**А.А. Воронов**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Статистическое мышление
<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Технологическое предпринимательство (on-line-программа) кафедра технологического предпринимательства кафедра технологического предпринимательства
<b>курс:</b>	2
<b>квалификация:</b>	магистр

Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Количество контрольных работ, заданий: 3

Программу составил: С.В. Багузин, канд. физ.-мат. наук

Программа обсуждена на заседании кафедры технологического предпринимательства 04.06.2020

## Аннотация

Цель курса – сформировать у слушателей статистическую парадигму восприятия действительности (в качестве альтернативы причинной), а также навыки для успешной деятельности в современном мире, имеющем статистическую (вероятностную) природу. В рамках курса слушатели научатся строить контрольные карты Шухарта и применять их для совершенствования процессов. Использовать Excel для анализа и визуализации данных. Подтверждать или отвергать нулевую гипотезу. Количественно оценивать изменение априорной вероятности на основе новых данных. Понимать различие между явлениями, подчиняющимися нормальному распределению и степенному. Освоить ряд методов моделирования и прогнозирования.

### 1. Цели и задачи

#### Цель дисциплины

Цель курса – сформировать у слушателей статистическую парадигму восприятия действительности (в качестве альтернативы причинной), а также навыки для успешной деятельности в современном мире, имеющем статистическую (вероятностную) природу.

#### Задачи дисциплины

- сформировать у слушателей понимание базовых концепций статистики, теории вероятностей, теории информации, обработки больших данных, планирования и оценки результатов эксперимента;
- сформировать у слушателей базовые навыки обработки данных в Excel;
- сформировать у слушателей базовые навыки презентации итогов анализа средствами Excel;
- сформировать у слушателей навыки сбора данных, и управления качеством на основе контрольных карт Шухарта.

### 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной задачи	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
	УК-3.2 Учитывает в своей социальной и профессиональной деятельности интересы, особенности поведения и мнения (включая критические) людей, с которыми работает/взаимодействует, в том числе посредством корректировки своих действий

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- принципы построения контрольных карт, области их применения;
- основные направления разведочного анализа данных, виды выборок, шкалы данных, типы диаграмм, и уместность их использования, принципы визуализации и презентации итогов анализа;
- меры среднего и разброса, принципы их использования, отличия смещенной и несмещенной оценки, коэффициент корреляции Пирсона;
- основные типы распределений случайной величины: нормальное, Пуассона, биномиальное, центральную предельную теорему, закон больших чисел, эвристику доступности;
- основные понятия теории вероятностей, правила сложения и умножения вероятностей;
- формулу Байеса, и ее применения, взаимосвязь априорной вероятности, новой информации и апостериорной вероятности, что такое идеальный байесовский наблюдатель;
- суть проверки гипотез, t-распределение Стьюдента, проблему выбора критерия значимости, ошибки первого и второго типа, основы байесовского статистического вывода;
- степенные распределения, и их повсеместность в реальной жизни, закон Бенфорда или закон первой цифры, закон Ципфа, принцип Парето, понятие фракталов, ведение бизнеса в Интернете с использованием толстого хвоста;
- основы анализа больших данных: кластерный анализ, линейное программирование, регрессионный анализ, прогнозирование на основе экспоненциального сглаживания;
- формулу Шеннона, связь информации и энтропии, алгоритмическое определение информации, что не измеряется в битах.

уметь:

- использовать основные статистические функции в Excel;
- устанавливать и использовать Пакет анализа;
- формировать случайную выборку данных с использованием функций Excel;
- обрабатывать сырые данные из файлов и Интернета с использованием Power Query;
- настраивать наивный байесовский классификатор;
- обрабатывать ошибки измерений в Excel на основе байесовской вероятности;
- моделировать системную динамику в iThink;
- моделировать методом Монте-Карло в Excel;
- оценивать прогноз методом Брайера;
- калибровать экспертов.

владеть:

- навыками поиска в Интернет;
- терминологией в областях статистики, управления качеством, больших данных, теории информации;
- навыками визуализации данных, сторителлинга;
- навыками письменной коммуникации в части презентации результатов анализа.

#### 4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

##### 4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Статистическое мышление против причинного	2	4		6
2	Разведочный анализ	4	4		6
3	Описательные статистики	4	4		6
4	Статистический вывод	4	4		6
5	Теория вероятностей	4	4		6
6	Распределения с толстыми хвостами	4	2		6
7	Прогнозирование	4	4		4

8	Теория информации	4	4		5
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

#### 4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 3 (Осенний)

##### 1. Статистическое мышление против причинного

Эволюция и причинное мышление. Системное и статистическое мышление. От данных к пониманию. Инструменты системного подхода. Системная динамика. Модель черного ящика. Эмерджентные (возникающие) свойства. Кейс: в чем причина роста просроченной дебиторской задолженности? Измерение эмерджентных свойств. История возникновения контрольных карт. Кейс: управление размещением товаров на складе с использованием контрольных карт. Контрольные границы. Практикум. Построение контрольной карты в Excel. Кейс: управление затратами с использованием контрольной карты. Два подхода к потерям.

##### 2. Разведочный анализ

Сбор данных. Генеральная совокупность и выборка. Методы выборки: детерминированная и вероятностная. Простая случайная выборка. Стратифицированная выборка. Опросы населения. Типы и шкалы данных: номинальная, порядковая, интервальная и шкала отношений. Упорядочение данных. Диаграмма «ствол и листья». Распределение частот, формула Стерджесса. Типы диаграмм: столбчатая, линейчатая, гистограмма, круговая, кольцевая, пузырьковая, лепестковая, диаграмма рассеивания, дерево, солнечные лучи, ящик с усами, каскадная. «Говорящие» заголовки. Принцип Эдварда Тафти минимизации количества элементов диаграммы. Фактор лжи Эдварда Тафти. Мусор на входе – мусор на выходе. Мухи отдельно, котлеты отдельно. Дашборды. Сторителлинг на основе данных. Обработка «сырых» данных с помощью Power Query в Excel. Эпидемии и Google Trends.

##### 3. Описательные статистики

Сводки данных. Что-то теряется. Описательные статистики. Меры центральной тенденции: медиана, среднее арифметическое, мода. 5-числовая сводка. Инклюзивная и эксклюзивная оценка. Меры разброса: размах, среднее квадратичное отклонение, дисперсия. Смещенная и несмещенная оценка. Основные типы распределений случайной величины: биномиальное, Пуассона, нормальное. Основные свойства нормального распределения. Качество шести сигм. Центральная предельная теорема. Закон больших чисел. Эвристика доступности или закон малых чисел.

##### 4. Статистический вывод

Двойной слепой рандомизированный метод. Определение объема выборки. Понятия статистики и параметра. Цикл: выборка, измерение, расчеты, оценка гипотезы. Вклад Рональда Фишера. t-статистика Стьюдента. Критерии значимости. Пресловутые 5%. Ошибки первого и второго типа. Корреляционный и регрессионный анализ. Статистический вывод в байесовском подходе.

##### 5. Теория вероятностей

Три типа вероятностей: априорная, эмпирическая и субъективная. Вероятность события. Выборочное пространство и события. Таблица сопряженности признаков. Безусловная вероятность. Вероятность совместных событий. Правило сложения и умножения вероятностей.

Этические проблемы и вероятность. Правила счета. Условная вероятность. Дерево решений. Исторический очерк: Байес и Лаплас. Формулу Байеса, и ее применения, взаимосвязь априорной вероятности, новой информации и апостериорной вероятности, что такое идеальный байесовский наблюдатель; фильтрация спама с помощью наивного байесовского классификатора в Excel.

## 6. Распределения с толстыми хвостами

Закон Бенфорда или закон первой цифры. Закон Ципфа. Принцип Парето. Историческая справка: Бенуа Мандельброт. Самоподобие. Фракталы. Фрактальная размерность. Фрактальная природа финансовых рынков. Основы R/S-анализа. Случайное блуждание и эффект долгосрочной памяти. Риск и неопределенность. Мудрость толпы. Поведенческая экономика. Рынки, как конкурс красоты. Роль психологии. Кейс: две трети от среднего. Два слова о теории игр. Толстый хвост открывает возможности для бизнеса в Интернете.

## 7. Прогнозирование

Моделирование методом Монте-Карло в Excel и с использованием надстройки Crystal Ball. Анализ инвестиционного проекта. Моделирование системной динамики в iThink. Мировая динамика Форрестера. Современное состояние. Прогнозирование на основе экспоненциального сглаживания. Кластерный анализ. Метод Энрике Ферми. Калибровка экспертов. Кейс: линейное программирование команды Fantasy Premier League.

## 8. Теория информации

История: от Чарлза Бэббиджа до Андрея Колмогорова. Количество информации: формула Шеннона. Зависимость количества информации от вероятности событий. Связь информации и энтропии. Алгоритмическое определение информации. Машина и тест Тьюринга. История: вскрытие кода Энигмы. Что не измеряется в битах. Кейс: победный алгоритм игры «Быки и коровы».

## 5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Необходимое оборудование для лекций и практических занятий: компьютер, микрофон, наушники.

Microsoft Excel 365 или более ранние версии (чем ниже версия Excel, тем меньше возможностей у слушателя).

## 6. Перечень рекомендуемой литературы

### Основная литература

Рекомендованная литература для самостоятельного изучения

Деминг Уильям Эдвардс. Выход из кризиса: Новая парадигма управления людьми, системами и процессами. Альпина Паблишер, 2019

Адлер Юрий, Шпер Владимир. Практическое руководство по статистическому управлению процессами. Альпина Паблишер, 2019

Бослаф Сара. Статистика для всех. М.: ДМК Пресс, 2017

### Дополнительная литература

Рекомендованная литература для самостоятельного изучения

Хафф Дарелл. Как лгать при помощи статистики. Альпина Паблишер, 2020

Хургин Яков Да, нет или может быть... Рассказы о статистической теории управления и эксперимента. М.: ЛЕНАНД

## **7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)**

ABC-анализ и принцип Парето для бизнеса: <http://baguzin.ru/wp/?p=310>

t-статистика Стьюдента в Excel: <http://baguzin.ru/wp/?p=19184>

## **8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)**

ABC-анализ и принцип Парето для бизнеса: <http://baguzin.ru/wp/?p=310>

t-статистика Стьюдента в Excel: <http://baguzin.ru/wp/?p=19184>

Введение в теорию статистического вывода: <http://baguzin.ru/wp/?p=16808>

Визуализация статистических данных с помощью диаграммы ящик с усами:  
<http://baguzin.ru/wp/?p=17422>

ГОСТ Р 50779.42-99. Статистические методы. Контрольные карты Шухарта:  
<http://docs.cntd.ru/document/gost-r-50779-42-99>

Закон Бенфорда или закон первой цифры: <http://baguzin.ru/wp/?p=4367>

Закон Ципфа и фрактальная природа социальных и экономических явлений:  
<http://baguzin.ru/wp/?p=1716>

Идеи Байеса для менеджеров: <http://baguzin.ru/wp/?p=6355>; в конце заметки представлена подборка литературы по Байесовскому подходу.

Как с помощью диаграммы приукрасить действительность? или о факторе лжи Эдварда Тафти:  
<http://baguzin.ru/wp/?p=2086>.

Контрольные карты Шухарта. Правила определения отсутствия управляемости.  
<http://baguzin.ru/wp/?p=2101>

Лучшая команда Fantasy Premier League сезона 2017/18: <http://baguzin.ru/wp/?p=18673>

Малые выборки в конкурентной разведке: <http://baguzin.ru/wp/?p=2880>

Нормальное распределение. Построение графика в Excel. Концепция шести сигм:  
<http://baguzin.ru/wp/?p=1170>

Оценка прогноза. Количественное измерение неопределенности: <http://baguzin.ru/wp/?p=21941>

Пример построения контрольной карты Шухарта в Excel. <http://baguzin.ru/wp/?p=1479>

Разработка оптимальной стратегии игры «Быки и коровы» на основе теории информации:  
<http://baguzin.ru/wp/?p=2574>

СТАНДОТКЛОН.В и СТАНДОТКЛОН.Г: в чем различие? <http://baguzin.ru/wp/?p=15634>

Эффект Даннинга-Крюгера и регрессия к среднему:

<https://medium.com/@flavorchemist/dunning-kruger-regression-7d747eecb806>

## **9. Методические указания для обучающихся по освоению дисциплины (модуля)**

### **Основная литература**

Левин, Дэвид М., Стефан, Дэвид, Кребель, Тимоти С., Беренсон, Марк Л. Статистика для менеджеров с использованием Microsoft Excel, 4-е изд. — М.: Издательский дом «Вильямс», 2004. — 1312 с. Конспект: <http://baguzin.ru/wp/?p=5285>

### **Дополнительная литература**

Альфред Реньи. Дневник. Записки студента по теории информации // Альфред Реньи. Трилогия о математике. — М.: Мир, 1980. — С. 199–284. Конспект: <http://baguzin.ru/wp/?p=2562>

Андрей Колмогоров. Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965, т. 1, вып. 1, стр. 3–11

Бенуа Мандельброт. (Не)послушные рынки: фрактальная революция в финансах. — М.: Издательский дом Вильямс, 2006. — 400 с. Конспект: <http://baguzin.ru/wp/?p=1604>

Гектор Макдональд. Правда. Как политики, корпорации и медиа формируют нашу реальность, выставляя факты в выгодном свете. — М.: Альбина Паблишер, 2019. — 368 с. Конспект: <http://baguzin.ru/wp/?p=20135>

Генри Нив. Организация как система: Принципы построения устойчивого бизнеса Эдвардса Деминга. — М.: Альпина Паблишер, 2017. — 368 с. Конспект: <http://baguzin.ru/wp/?p=253>

Дарелл Хафф. Как лгать при помощи статистики — М.: Альпина Паблишер, 2015. — 168 с. Конспект: <http://baguzin.ru/wp/?p=12682>

Дэниэль Канеман. Думай медленно... решай быстро. — М.: АСТ, 2013. — 656 с. Конспект: <http://baguzin.ru/wp/?p=7840>

Джеймс Глик. Информация. История. Теория. Поток. — М.: АСТ, Corpus, 2013. — 576 с. Конспект: <http://baguzin.ru/wp/?p=13062>

Джон Тьюки. Анализ результатов наблюдений. Разведочный анализ. — М.: Мир, 1981. — 696 с. Конспект: <http://baguzin.ru/wp/?p=15897>

Джон Форман. Много цифр: Анализ больших данных при помощи Excel. — М.: Альпина Паблишер, 2016. — 464 с. Конспект: <http://baguzin.ru/wp/?p=13802>

Джордан Элленберг. Как не ошибаться. Сила математического мышления. — М.: Манн, Иванов и Фербер, 2017. — 576 с. Конспект: <http://baguzin.ru/wp/?p=18371>

Дональд Уилер, Дэвид Чамберс. Статистическое управление процессами. Оптимизация бизнеса с использованием контрольных карт Шухарта. М.: Альпина Паблишер, 2016. — 410 с. Конспект: <http://baguzin.ru/wp/?p=15577>

Дональд Уилер, Дэвид Чамберс. Рациональная группировка данных при построении контрольной карты Шухарта (кейс из книги Статистическое управление процессами). <http://baguzin.ru/wp/?p=15617>

Дуглас Хаббард. Как измерить всё, что угодно. Оценка стоимости нематериального в бизнесе. — М.: Олимп-Бизнес, 2009. — 320 с. Конспект: <http://baguzin.ru/wp/?p=2511>

Исаак Яглом, Акива Яглом. Вероятность и информация. М.: Наука, 1973. — 512 с. Конспект: <http://baguzin.ru/wp/?p=18059>

Конрад Карлберг. Регрессионный анализ в Microsoft Excel. — М.: Вильямс, 2017. — 400 с. Конспект: <http://baguzin.ru/wp/?p=18972>

Крис Фрит. Мозг и душа: Как нервная деятельность формирует наш внутренний мир. — М.: Астрель: CORPUS, 2010. — 336 с. Конспект: <http://baguzin.ru/wp/?p=11859>

Леонард Млодинов. (Не)совершенная случайность. Как случай управляет нашей жизнью. М.: Livebook/Гаятри, 2010. — 352 с. Конспект: <http://baguzin.ru/wp/?p=4219>

Нассим Николас Талеб. Антихрупкость. Как извлечь выгоду из хаоса. — М.: КоЛибри, 2014. — 768 с. Конспект: <http://baguzin.ru/wp/?p=7903>

Питер Панде, Ларри Холп. Что такое «шесть сигм»? Революционный метод управления качеством. — М.: Альпина Бизнес Букс, 2005. — 160 с. Конспект: <http://baguzin.ru/wp/?p=2405>

Питер Панде, Роберт П. Ньюмен, Роланд Р. Кэвенег. Курс на Шесть Сигм. Как General Electric, Motorola и другие ведущие компании мира совершенствуют свое мастерство. — М.: Издательство ЛОРИ, 2002. — 400 с. Конспект: <http://baguzin.ru/wp/?p=16912>

Пьер Симон Лаплас. Опыт философии теории вероятностей. — М.: Книжный дом «ЛИБРОКОМ», 2011. — 208 с. (книга впервые опубликована в 1814 г.; на старорусский язык книга переведена в 1908 г.; настоящее издание является репринтным). Конспект: <http://baguzin.ru/wp/?p=16017>

Сара Бослаф. Статистика для всех. — М.: ДМК Пресс, 2017. — 586. Конспект: <http://baguzin.ru/wp/?p=19047>

Сергей Багузин. Управление размещением товаров на складе с использованием контрольных карт Шухарта // Журнал «Менеджмент качества», 2010, №4, с. 300–310. См. <http://baguzin.ru/wp/?p=610>

Сергей Багузин. Управление логистическими издержками с использованием контрольных карт Шухарта // Системный анализ в экономике — 2016: сборник трудов IV Международной научно-практической конференции — биеннале (9–11 ноября 2016)/ под ред. Г.Б. Клейнера, С.Е. Щепетовой. Т.1. — М.: Финансовый университет, 2016. — С. 73–76. <http://baguzin.ru/wp/?p=16128>

Тим Филлипс. Управление на основе данных. Как интерпретировать цифры и принимать качественные решения в бизнесе. — М.: Манн, Иванов и Фербер, 2017. — 192 с. Конспект: <http://baguzin.ru/wp/?p=17283>

Уильям Эдвардс Деминг. Выход из кризиса: Новая парадигма управления людьми, системами и процессами. — М.: Альпина Паблишер, 2011. — 424 с. Конспект: <http://baguzin.ru/wp/?p=2138>

Уильям Эдвардс Деминг. Новая экономика. — М.: Эксмо, 2006. — 208 с. Конспект: <http://baguzin.ru/wp/?p=7791>

Филип Тетлок, Дэн Гарднер. Думай медленно - предсказывай точно. Искусство и наука предвидеть опасность. — М.: АСТ, 2018. — 384 с. Конспект: <http://baguzin.ru/wp/?p=21912>

Хитосе Кумэ. Статистические методы повышения качества. – М.: Финансы и статистика, 1990. – 304 с. Конспект: <http://baguzin.ru/wp/?p=15984>

Чарльз Уилан. Голая статистика. Самая интересная книга о самой скучной науке. – М.: Манн, Иванов и Фербер, 2016. — 352 с. Конспект: <http://baguzin.ru/wp/?p=14151>

Юрий Адлер, Владимир Шпер. Практическое руководство по статистическому управлению процессами – М.: Альпина Паблишер, 2019. – 234 с. Конспект: <http://baguzin.ru/wp/?p=20111>

Юрий Адлер и др. Контрольные карты Шухарта в России и за рубежом: краткий обзор современного состояния (статистические аспекты). – Журнал «Стандарты и качество», июль–август, 2011. Конспект: <http://baguzin.ru/wp/?p=15318>

Яков Хургин. Да, нет или может быть... Рассказы о статистической теории управления и эксперимента. М.: ЛЕНАНД, 2021. – 208 с.



**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Прикладные математика и физика
<b>профиль подготовки:</b>	Технологическое предпринимательство (on-line-программа) кафедра технологического предпринимательства кафедра технологического предпринимательства
<b>курс:</b>	2
<b>квалификация:</b>	магистр
Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен	
<b>Разработчик:</b>	С.В. Багузин, канд. физ.-мат. наук

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной задачи	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
	УК-3.2 Учитывает в своей социальной и профессиональной деятельности интересы, особенности поведения и мнения (включая критические) людей, с которыми работает/взаимодействует, в том числе посредством корректировки своих действий

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Статистическое мышление» обучающийся должен:

### знать:

- принципы построения контрольных карт, области их применения;
- основные направления разведочного анализа данных, виды выборок, шкалы данных, типы диаграмм, и уместность их использования, принципы визуализации и презентации итогов анализа;
- меры среднего и разброса, принципы их использования, отличия смещенной и несмещенной оценки, коэффициент корреляции Пирсона;
- основные типы распределений случайной величины: нормальное, Пуассона, биномиальное, центральную предельную теорему, закон больших чисел, эвристику доступности;
- основные понятия теории вероятностей, правила сложения и умножения вероятностей;
- формулу Байеса, и ее применения, взаимосвязь априорной вероятности, новой информации и апостериорной вероятности, что такое идеальный байесовский наблюдатель;
- суть проверки гипотез, t-распределение Стьюдента, проблему выбора критерия значимости, ошибки первого и второго типа, основы байесовского статистического вывода;
- степенные распределения, и их повсеместность в реальной жизни, закон Бенфорда или закон первой цифры, закон Ципфа, принцип Парето, понятие фракталов, ведение бизнеса в Интернете с использованием толстого хвоста;
- основы анализа больших данных: кластерный анализ, линейное программирование, регрессионный анализ, прогнозирование на основе экспоненциального сглаживания;
- формулу Шеннона, связь информации и энтропии, алгоритмическое определение информации, что не измеряется в битах.

### уметь:

- использовать основные статистические функции в Excel;
- устанавливать и использовать Пакет анализа;
- формировать случайную выборку данных с использованием функций Excel;
- обрабатывать сырые данные из файлов и Интернета с использованием Power Query;
- настраивать наивный байесовский классификатор;
- обрабатывать ошибки измерений в Excel на основе байесовской вероятности;
- моделировать системную динамику в iThink;
- моделировать методом Монте-Карло в Excel;
- оценивать прогноз методом Брайера;
- калибровать экспертов.

### владеть:

- навыками поиска в Интернет;
- терминологией в областях статистики, управления качеством, больших данных, теории информации;
- навыками визуализации данных, сторителлинга;
- навыками письменной коммуникации в части презентации результатов анализа.

### 3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Домашние задания к каждой лекции

#### 1) Лекция 1. Статистическое мышление против причинного

Задание 1. Прочитайте заметку Построение контрольной карты Шухарта в Excel. Разберите примеры в файле Excel. Изучите формулы, использованные для построения контрольных границ.

Откройте файл Excel «Тема 1. ККШ. Практикум». Выполните задания 2–4.

Задание 2. Воспользовавшись представленными данными о затратах логистики:

- Отрадите данные на диаграмме хода процесса (тип диаграммы в Excel – график с маркерами)
- Вынесите суждение, и разбейте все данные на области, соответствующие различным уровням затрат
- Для каждой области рассчитайте среднее, верхнюю и нижнюю контрольные границы, поместите эти значения в отдельные столбцы
- Оформите диаграмму по образцу, или предложите свой вариант наглядного представления данных.

Задание 3. Исходные данные представляют собой отклонение от норматива веса 100 произведенных резиновых деталей. Все детали были разбиты на 20 групп по пять в каждой.

- Рассчитайте среднее в каждой группе.
- Рассчитайте размах по каждой группе.
- Найдите среднее средних и средний размах по 20 группам.
- Постройте контрольную карту средних и размахов
- Запишите формулы расчета контрольных границ карты средних и карты размахов

Задание 4. Постоянство толщины детали очень важно для выполнения последующих операций. Для измерения толщины отбирают 10 деталей из одной партии. Толщина каждой детали измеряется в восьми точках, как показано на схеме. Измеряемые детали были помещены в печь таким образом, что сторона с точками 1, 2 и 3 была у края печи, а сторона с точками 6, 7 и 8 – в глубине печи.

- Используя приведенные данные, постройте контрольную карту, которая отвечала бы на вопрос: «Есть ли разница между средней толщиной на концах деталей? То есть, между точками 1, 2, 3 и 6, 7, 8?»
- Постройте контрольную карту, отвечающую на вопрос: «Есть ли разница между средней толщиной десяти деталей?»

Задание 5. Сделайте обзор вашего проекта. Собирали ли вы ранее какие-либо данные? Если нет, можете ли вы в короткий срок собрать такие данные? Если нет идей, то найдите подходящие данные в Интернете. Желательно, чтобы это были данные какого-то процесса. Например, количество ДТП, или разница между временем прилета по расписанию и по факту. Подумайте, как сгруппировать данные. Выберите вид контрольной карты, и постройте ее.

#### 2) Лекция 2. Разведочный анализ.

Задание 1. Здесь представлен массив из 40 случайных чисел в интервале от 1 до 800. Вернитесь к рис. 2.4 лекции. Проанализируйте, как работает функция СЛУЧМЕЖДУ(). Что может пойти не так? Как вы исправите ситуацию?

Задание 2. Данные о среднемесячных температурах июля в Москве за период наблюдений с 1879 г. представьте в виде распределения частот:

- а) с помощью сводной таблицы и сводной диаграммы,
- б) с помощью таблицы, построенной на основе функции массива ЧАСТОТА(),
- в) с помощью статистической диаграммы, (доступна, начиная с Excel 2016, подробнее см. Новые диаграммы в Excel 2016).

#### 4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Итоговая работа выполняется в контексте учебного проекта студента. Итоговая работа должна включать минимум два аспекта из 8 тем дисциплины. Перечень типовых (примерных) аспектов итоговых работ:

1. Управление одним из процессов вашего проекта на основе контрольной карты
2. Сегментирование клиентской базы в вашем проекте методом k-средних
3. Презентация вашего проекта потенциальному инвестору с использованием сторителлинга
4. Разработка дашборда для руководителя вашего проекта
5. Сбор и первичная обработка Интернет-трафика вашего проекта (или иных структурированных данных) с использованием Excel Power Query
6. Анализ окружающей среды вашего проекта на основе Google Trends
7. Расчет пропускной способности обслуживания клиентов в вашем проекте на основе распределения Пуассона
8. Оценка числа ошибок в каком-либо процессе вашего проекта, и их сравнение с качеством шести сигм
9. Фильтрация спама с помощью наивного байесовского классификатора
10. Использование идей статьи Biases in Estimating Probabilities в разработке вашего проекта (предполагает полный перевод статьи).
11. Построение модели байесовской вероятности в Excel для проверки гипотез в вашем проекте (по типу модели в ранее рассматриваемой домашней задаче «Сколько граней в кубике?»).
12. Обработка ошибок измерений с использованием Excel и байесовской вероятности.

Билет №1 Построение модели финансовых результатов вашего проекта в iThink.

Билет №2 Разработка сценария рабочей сессии калибровки экспертов для составления ключевого прогноза в вашем проекте.

#### Критерии оценивания

- оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;

- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины.

## **5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности**

Во время проведения экзамена обучающиеся могут пользоваться программой дисциплины.

При проведении экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету не должен превышать одного астрономического часа.