

Правительство Российской Федерации

**Государственное образовательное бюджетное учреждение высшего
профессионального образования**

**«Государственный университет -
Высшая школа экономики»**

Факультет социологии

Программа дисциплины

Теория вероятностей и математическая статистика

для направления/ специальности 040100.62 «Социология»

подготовки бакалавров

Авторы:

Толстова Ю.Н., Макаров А.А., Пашкевич А.В., Хавенсон Т.Е., Назаров Б.В.

Рекомендована секцией УМС

_____социология_____
Председатель
_____Ледяев В.Г._____
« _____ » _____ 20 г.

Одобрена на заседании кафедры

_____Высшей математики_____
Зав. кафедрой
_____Макаров А.А._____
« _____ » _____ 20 г.

Утверждена УС факультета

_____социологии_____

Ученый секретарь

_____Надеждина Е.В._____
« _____ » _____ 20 г.

Одобрена на заседании кафедры

_____методов сбора и анализа
социологической информации

Зав. кафедрой

_____Козина И.М._____
« _____ » _____ 20 г.

Москва, 2010 г.

Аннотация

Курс состоит из двух взаимосвязанных разделов: теории вероятностей и математической статистики.

Основная цель Раздела I познакомить студентов с основными понятиями классической теории вероятностей. Научить выявлять различные вероятностные понятия в исследовательской практике и применять их. Заложить основы для изучения курсов по анализу социологических данных.

После изучения курса студенты будут

- знакомы с основными подходами к определению понятия вероятности (классический, субъективный, байесовский)
- знать основные исторические предпосылки зарождения понятий ТВ.
- уметь использовать в своей работе основные понятия классической теории вероятностей
- иметь необходимые навыки и знания для дальнейшего обучения

Раздел II посвящен изложению основ математической статистики и призван предложить студентам *методологию* первичной обработки данных и разные *практические задачи и упражнения*, учитывающие тематику, релевантную специальности «Социология», для того чтобы учащиеся могли закрепить знания на примерах и исследованиях, находящихся в поле их будущей профессионально-деловой деятельности. Успешное овладение инструментарием математической статистики предопределяет в дальнейшем приобретение учащимися высокого уровня квалификации по большинству других смежных курсов, касающихся стратегии обработки и анализа информации.

Как известно, подобные курсы традиционно читаются студентам самых разных специальностей. Объясняется это тем, что изучение статистических закономерностей требуется практически в любой отрасли человеческого знания. Отечественная литература в соответствующем отношении очень богата, имеется множество учебников (в том числе переводных) и методических пособий самого разного плана: с разной широтой охвата проблематики, рассчитанных на читателей с различной подготовкой и т.д. Казалось бы, преподавание математической статистики для студентов - прикладников стало рутинным делом. Тем не менее, предлагаемая программа имеет ряд особенностей, позволяющих считать ее в некоторых отношениях оригинальной. Особенности эти вызваны желанием авторов сделать курс хорошо воспринимаемым именно социологами.

Особенность преподавания математической статистики социологам состоит в том, что нами уделяется довольно большое внимание проблеме измерения исходных данных. В предлагаемом курсе это находит отражение прежде всего в том, что, говоря, к примеру, о

параметрах распределений, мы соотносим их с типами шкал, используемых при получении исходных данных. Определенное внимание также уделяется описанию роли статистического подхода в социологии, анализу ситуаций, при которых его возможности ограничены: указывается, что социолог теряет, опираясь лишь на статистическую парадигму; обсуждается вопрос о возможности обеспечения того комплекса условий, реализация которого приводит к появлению интересующих социолога случайных событий; в частности, затрагивается проблема существования случайных величин. Рассматривается ряд часто встречающихся в социологии ситуаций, в которых могут не выполняться условия реализации известных математико-статистических подходов. Показывается, как может действовать социолог в таких случаях. В общем и целом в результате изучения основ математической статистики студенты будут знать:

- континуум основных понятий, относящихся к статистической парадигме, проблематике изучения социальных закономерностей
- правила и техники расчета мер средней тенденции, мер разброса, коэффициентов связи; соотнесение с типами социологических шкал, уровнями измерения
- правила и техники построения доверительных интервалов для параметров, оцененных по выборке; правила расчета объема простой случайной выборки
- принципы проверки статистических гипотез; используемые статистические критерии и таблицы
- правила и техники реализации однофакторного и двухфакторного дисперсионного анализа в социологии, основы изучения причинно-следственных отношений между переменными (признаками)
- основы метода максимального (наибольшего) правдоподобия, оценка параметров моделей

Специфика представления библиографии в программе

В отечественной литературе имеется очень много работ (в том числе переводных), прекрасно описывающих основные положения теории вероятностей и математической статистики. Список этих работ приведен в конце программы. В них можно найти материал почти по все темам. После некоторых разделов (тем) приведены списки книг, содержание которых более узко - касается только соответствующего раздела (темы). К обязательной литературе мы отнесли работы, либо выпущенные в последние несколько лет, либо ориентированные на социологов или по способу изложения, или по специфике рассматриваемых аспектов (к сожалению, эти работы в большинстве случаев опубликованы довольно давно).

Курсы-пререквизиты: дискретная математика, алгебра и анализ, методология и методы социологических исследований в объеме, преподаваемом на социологическом факультете.

Базовые учебники

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009
2. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007
3. *Айвазян С. А., Мхитарян В. С.* Теория вероятностей и прикладная статистика, ч.1. М.: Юнити, 2001
4. *Математическая статистика для социологов. Задачник.* М.: ИД ГУ-ВШЭ, 2010
5. *Bluman A.* Elementary statistics. McGraw-Hill, 2008
6. *Pawitan Y.* In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Oxford Science Publications, 2005

Тематический план учебной дисциплины

№	Название темы	Всего часов по дисциплине	Аудиторные часы		Самостоятельная работа
			Лекции	Сем. и практ. занятия	
1.	Предмет теории вероятностей. Случайный эксперимент. Выбор из конечной совокупности. Пространство элементарных исходов (эл.событий). Классическое и статистическое определение понятия вероятности.	6	2	1	3
2.	События и операции над ними. Правила сложения и умножения вероятностей. Условная вероятность. Зависимость/независимость событий	10	2	3	5
3.	Формула полной вероятности. Формула Байеса.	9	2	2	5
4.	Испытания Бернулли. Биномиальное распределение.	9	2	2	5

5.	Случайные величины - дискретные и непрерывные. Характеристики случайной величины – мат.ожидание, дисперсия, асимметрия.	11	2	2	7
6.	Функция и плотность распределения. Нормальное распределение. Примеры непрерывных распределений: равномерное, экспоненциальное, нормальное.	10	2	2	6
7.	Произвольное и стандартное нормальное распределение. Стандартизация. Работа с таблицами нормального распределения. Квантили распределения.	12	2	3	7
8.	Дискретные распределения. Биномиальное. Пуассона.	7	2	1	4
9.	Двумерные распределения. Ковариация и корреляция случайных величин.	5	1	1	3
10.	Центральные предельные теоремы. Закон больших чисел.	6	1	1	4
11.	Общее представление о математической статистике (объект, предмет, цели, задачи). История развития математической статистики как науки	6	2	0	4
12.	Частотное распределение. Разбиение на интервалы. Полигон, гистограмма, кумулята. Меры средней тенденции (мода, медиана, среднее арифметическое). Меры разброса (размах, дисперсия, коэффициент вариации). Соотнесение с типами социологических шкал	14	2	4	8
13.	Коэффициенты связи: коэффициенты корреляции Пирсона, Спирмена, Кендала	9	2	2	5
14.	Оценивание параметров: точечное и интервальное. Доверительные интервалы для среднего арифметического, для медианы, для доли. Средняя и предельная ошибки выборки	10	2	2	6
15.	Объем простой случайной выборки	8	1	2	5
16.	Виды распределений случайных величин: Хи-квадрат-распределение (Пирсон), t-распределение (Стьюдент), F-распределение (Фишер).	3	1	0	2

	Стандартные таблицы функций этих распределений				
17.	Понятие статистической гипотезы. Направленные и ненаправленные альтернативные гипотезы. Односторонние и двусторонние критерии. Ошибки 1-го, 2-го рода	5	1	1	3
18.	Проверка статистических гипотез об отсутствии связи, о значении мат. ожидания определенному числу, о равенстве матожиданий, о равенстве долей, о равенстве дисперсий, о равенстве нулю коэффициента корреляции	16	3	4	9
19.	Проверка статистических гипотез о виде распределения данных (критерий согласия Пирсона). Равномерное распределение. Нормальное распределение. Распределение Пуассона. Биномиальное распределение	16	3	4	9
20.	Изучение причинно-следственных отношений в социологии с помощью математических методов	6	2	0	4
21.	Корреляционное отношение. Однофакторный дисперсионный анализ	10	2	2	6
22.	Двухфакторный дисперсионный анализ	10	2	2	6
23.	Классическая и субъективистская вероятностные концепции. Парадигма Неймана-Пирсона, парадигма правдоподобия, байесовская парадигма. Место подтверждения в теориях.	9	2	1	6
24.	Определение правдоподобия. Объединение правдоподобий. Закон правдоподобия. Принцип правдоподобия. Кривизна функции правдоподобия. Свойства правдоподобия. Интерпретация правдоподобия.	16	3	4	9
25.	Отношение правдоподобий, критерий Вальда и множитель Лагранжа.	8	2	2	4
26.	Оценивание параметров моделей. Метод наибольшего правдоподобия и его свойства.	14	3	2	9
27.	Правдоподобие и байесовская теория. Простейшие байесовские модели.	8	1	2	5

Итого:	252	52	52	148
--------	-----	----	----	-----

Формы контроля знаний студентов:

Итоговая оценка по учебной дисциплине складывается из следующих элементов:

1. Оценка по материалу Раздела I (теория вероятностей), форма итогового контроля – зачет после 1 модуля, вес оценки – 30%
2. Оценка по материалу Раздела II (математическая статистика), форма итогового контроля - экзамен в конце курса, вес оценки – 70%

Таким образом, итоговая оценка за курс рассчитывается по формуле:

$$\text{Оценка}_{\text{итог}} = 0,3\text{оценка}_{\text{разделI}} + 0,7\text{оценка}_{\text{разделII}}$$

Получение неудовлетворительной оценки на зачете или экзамене блокирует получение положительной оценки за курс.

В свою очередь обе оценки учитывают следующие виды активности студентов:

Оценка по материалу Раздела I (теория вероятностей)

Текущий контроль осуществляется на семинарских занятиях, в первую очередь учитывается выполнение домашних и семинарских заданий. Также на семинарах студенты пишут небольшие самостоятельные работы. *Задачи для домашних работ* раздаются в электронном виде. Выполненные работы должны сдавать на следующем семинаре или присылаться по эл. почте до семинара. Работы сданные на неделю позже, оцениваются в 0,5 оценки, просроченные больше, чем на неделю – в 0 баллов.

Вид деятельности	Баллы
Письменный зачет в конце 1 модуля	5
Домашние работы и самостоятельные работы на семинарах	4
Активность на семинарских занятиях (в том числе посещаемость)	1

$$\text{Оценка}_{\text{раздел I}} = 0,5*\text{зач.} + 0,4*\text{ДР} + 0,1*\text{АС}$$

Оценка по материалу Раздела II (математическая статистика)

Текущий (основной) контроль осуществляется на семинарских занятиях.

Кроме посещения лекций и работы на семинарских занятиях, студентами будут выполняться следующие работы:

1. написана контрольная работа (1 промежуточная работа).
2. выполнены домашние задания (одно из домашних заданий предполагает написание творческого эссе);

Список ориентировочных тем эссе и примерный перечень задач для каждой темы приведены в базовом учебнике и задачнике.

- текущий контроль – оценка посещаемости лекций и семинарских занятий, оценка работы слушателей на семинарских занятиях; оценка домашних заданий и контрольной работы.

Итоговый контроль – письменный экзамен.

Вид деятельности	Баллы
Письменный экзамен в конце 3-го модуля	5
Контрольная работа	2
Домашние работы (в т.ч. творческое эссе)	1,5
Активность на семинарских занятиях	1
Посещение лекций и семинарских занятий	0,5

Оценка достигнутых студентом результатов по математической статистике будет складываться следующим образом:

Оценка Π раздел $\Pi = 0,5*Э + 0,2*КР + 0,15*ДР + 0,1*АС + 0,05*П$, где

Э – письменный экзамен

КР – контрольная работа

ДР – домашние работы

АС – активность на семинарских занятиях

П – посещение лекций и семинарских занятий

Содержание программы

Раздел I «Теория вероятностей»

Тема 1. Предмет теории вероятностей. Случайный эксперимент. Выбор из конечной совокупности. Пространство элементарных исходов (эл.событий). Классическое и статистическое определение понятия вероятности.

Случайность как предмет изучения. Предмет теории вероятностей. Случайный эксперимент. Пространство элементарных исходов (эл.событий).

Различные подходы к изучению вероятности. Объективизм и субъективизм. Классическое и статистическое определение понятия вероятности. Относительная частота и вероятность.

Литература основная

1. *Мостеллер Ф., Рурке Р., Томас Дж.* Вероятность. М.: Мир, 1969. С. 15-17
2. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.29-52
3. *Bluman A.* Elementary statistics. McGraw-Hill, 2008. 2nd ed. Pp.153-163

Литература дополнительная

4. *Мизес Р.* Вероятность и статистика. М.: КомКнига, 2007
5. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.15

Тема 2. События и операции над ними. Правила сложения и умножения вероятностей. Условная вероятность. Зависимость/независимость событий. Совместимые и несовместимые события.

Литература основная

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.60-63, 68-77
2. *Bluman A.* Elementary statistics. McGraw-Hill, 2008. Pp.163-174

Тема 3. Формула Бейеса.

Полная группа событий. Формула полной вероятности. Априорные и апостериорные вероятности. Формула Бейеса.

Литература основная

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.77-79

2. *Bluman A. Elementary statistics. McGraw-Hill, 2008. Pp.174-182*

Тема 4. Испытания Бернулли. Биномиальное распределение.

Вероятность в дискретных пространствах. Схема испытаний Бернулли.

Литература основная

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.63-68*
2. *Bluman A. Elementary statistics. McGraw-Hill, 2008. Pp.207-214*

Литература дополнительная

3. *Сигел Э. Практическая бизнес-статистика. М.: Издательский дом "Вильямс", 2004*

Тема 5. Случайные величины - дискретные и непрерывные. Характеристики случайной величины – мат.ожидание, дисперсия, асимметрия.

Дискретные и непрерывные СВ. Основные социологические шкалы. Математические операции со случайными величинами. Характеристики случайной величины – мат.ожидание и дисперсия. Дисперсия для выборки. Корректировка на несмещенность оценки. Асимметрия.

Литература основная

1. *Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.20-22*
2. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.84-112*
3. *Bluman A. Elementary statistics. McGraw-Hill, 2008. Pp. 69-90, 199-206,*

Тема 6. Функция и плотность распределения случайной величины. Нормальное распределение. Примеры непрерывных распределений: равномерное, экспоненциальное, нормальное.

Основные свойства нормального распределения. Его роль при изучении социальных явлений.

Литература основная

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.185-195
2. *Bluman A.* Elementary statistics. McGraw-Hill, 2008. Pp.194-199, 227-242

Литература дополнительная

3. *Хили Дж.* Статистика. Социологические и маркетинговые исследования. СПб.: Питер, 2005

Тема 7. Произвольное и стандартное нормальное распределение. Стандартизация. Работа с таблицами нормального распределения. Квантили распределения.

Литература основная

1. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.51-53
2. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.173-185,
3. *Bluman A.* Elementary statistics. McGraw-Hill, 2008. Pp.242-251

Литература дополнительная

4. *Хили Дж.* Статистика. Социологические и маркетинговые исследования. СПб.: Питер, 2005

Тема 8. Дискретные распределения. Биномиальное. Пуассона.

Основные параметры биномиального распределения. Понятия математического ожидания и дисперсии. Аппроксимация биномиального распределения нормальным.

Основные параметры распределения Пуассона. Понятия математического ожидания и дисперсии.

Литература основная

1. *Тюрин Ю.Н., Макаров А.А., Симонова Г.И.* Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.126-136, 148-156
2. *Bluman A.* Elementary statistics. McGraw-Hill, 2008. Pp.207-214, 214-220

Литература дополнительная

3. Сигел Э. Практическая бизнес-статистика. М.: Издательский дом "Вильямс", 2004

Тема 9. Двумерные распределения. Ковариация и корреляция случайных величин.

Литература основная

1. Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.112-126
2. Bluman A. Elementary statistics. McGraw-Hill, 2008. Pp.383-393

Тема 10. Центральные предельные теоремы. Закон больших чисел. Неравенство Чебышева. Теорема Ляпунова.

Литература основная

1. Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.64-68
2. Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей: учебник для экономических и гуманитарных специальностей. М.: МЦНМО, 2009. С.206-221, 226-233
3. Bluman A. Elementary statistics. McGraw-Hill, 2008. Pp.251-263

Раздел II «Математическая статистика»

Тема 11. Общее представление о математической статистике (объект, предмет, цели, задачи). История развития математической статистики как науки

Понятия выборки, генеральной совокупности, случайной величины, в особенности понятие статистической закономерности, параметра распределения, выборочной статистики. Статистические закономерности как основные закономерности, выявляемые в процессе социологических исследований (исторический экскурс).

Случайная величина как основной объект изучения математической статистики. Параметры её распределения как предмет математической статистики. Перенос результатов с выборки на генеральную совокупность как основная задача (цель) математической статистики. Основные способы такого переноса как задачи математической статистики. Основные методологические принципы использования в социологии математических методов.

Литература основная

1. Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.19-36
2. Толстова Ю.Н. Анализ социологических данных: методология, дескриптивная статистика, изучение связей между номинальными признаками. М.: Научный мир, 2000

Литература дополнительная

(о методологических принципах использования математики в социологии)

3. Загоруйко Н.Г. Эмпирическое предсказание. Новосибирск: Наука, 1979. С.105-118
4. Рабочая книга социолога / отв. ред. Г.В.Осипов. М.: Наука, 1983
5. Толстова Ю.Н. Методология математического анализа данных // Толстова Ю.Н. Социология и математика. М.: Научный мир, 2003. С.80-94.
6. Толстова Ю.Н. Методология математического анализа данных // Социологические исследования, 1990. №6. С.77-87.

Тема 12. Частотное распределение. Разбиение на интервалы. Меры средней тенденции (мода, медиана, среднее арифметическое). Меры разброса (размах, дисперсия, коэффициент вариации). Соотнесение с типами социологических шкал

Выборочные представления генеральных распределений. Частотные таблицы. Принципы разбиения диапазона изменения признака на интервалы. Определение величин интервалов. Вариационный ряд. Графические методы изображения вариационных рядов: полигон, гистограмма, кумулята. Статистики, используемые для оценивания математического ожидания, моды, медианы, дисперсии, коэффициента вариации. Выполнение (невыполнение) для них определенных свойств.

Общие принципы понимания измерения в социологии. Определение понятия шкалы и её допустимых преобразований. Номинальная, порядковая, интервальная шкалы и соотнесение с допустимым применением мер средней тенденции и мер разброса. Проблема адекватности математического метода относительно типа используемых шкал.

Литература основная

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С.151-157
2. Девятко И.Ф. Методы социологического исследования. / И.Ф. Девятко. – 4-е изд. – М.: КДУ, 2006, С.166-198, 240-248

3. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.43-49
4. *Толстова Ю.Н.* Измерение в социологии: учеб. пособие. М.: КДУ, 2009. С. 9-37.
5. *Шмойлова Р.А., Минашкин Н.А., Садовникова Н.А.* Практикум по теории статистики: учеб. пособие / под ред. Шмойловой. – 3-е изд. – М.: Финансы и статистика, 2008. С. 124-139
6. *Bluman A.G.* Elementary statistics. Wm. C. Brown Publishers, 1995, pp.19-40, 69-109

Литература дополнительная

7. *Агабекян Р.Л., Кириченко М.М., Усатилов С.В.* Математические методы в социологии. Анализ данных и логика вывода в эмпирическом исследовании: учеб. пособие для вузов. Ростов н/Д: Феникс. 2005. С. 100-114.
8. *Клигер С.А., Косолапов М.С., Толстова Ю.Н.* Шкалирование при сборе и анализе социологической информации. М.: Наука, 1978 (глава.1)
9. *Орлов А.И.* Асимптотика квантований и выбор числа градаций в социологических анкетах // Математические методы и модели социологии. М.: ИСИ АН СССР, 1977. С.42-55
10. *Палий И.А.* Прикладная статистика: Учеб. пособие для вузов. / И.А. Палий. – М.: Высш. шк., 2004. С.16-20
11. *Пасхавер Б.* Проблема интервалов в группировках // Вестник статистики. 1972. №6
12. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.36-40
13. *Чупров А.А.* Об приемах группировки статистических наблюдений // Известия Санкт-Петербургского политехнического института. 1904. Т.1. Вып.1-2.
14. *Doane D.P.* Aesthetic Frequency Classification // American Statistician. 1976. Vol.30. P.181-183
15. *Freedman D., Diaconis P.* On this Histogram as a Density Estimator: L2 theory // Zeit. Wahr. Ver. Geb. 1981. Vol. 57. P.453-476.
16. *Scott D.W.* Multivariate Density Tstimation: Theory, Practice, and Visualization. N.Y.: John Wiley & Sons, 1992.
17. *Wand M.P.* Data-based Choice of Histogram Bin-width. Technical Report, Australian Graduate School of Management, University of NSW, 1995

Тема 13. Коэффициенты связи: коэффициенты корреляции Пирсона, Спирмена, Кендала

Коэффициенты корреляции: коэффициент линейной корреляции Пирсона (для переменных, измеренных на уровне шкал высокого порядка), коэффициенты ранговой корреляции Спирмена и Кендала (для переменных, измеренных на порядковом уровне). Соотнесение с ограничениями социологических шкал.

Литература основная

1. *Агабекян Р.Л., Кириченко М.М., Усатигов С.В.* Математические методы в социологии. Анализ данных и логика вывода в эмпирическом исследовании: учеб. пособие для вузов. Ростов н/Д: Феникс. 2005. С. 129-160.
2. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С.190-196, 201-206
3. *Девятко И.Ф.* Методы социологического исследования. / И.Ф. Девятко. – 4-е изд. – М.: КДУ, 2006, С. 253-257, 270-276.
4. *Шмойлова Р.А., Минашкин Н.А., Садовникова Н.А.* Практикум по теории статистики: учеб. пособие / под ред. Шмойловой. – 3-е изд. – М.: Финансы и статистика, 2008. С. 204-212, 217-222
5. *Bluman A.G.* Elementary statistics. Wm. C. Brown Publishers, 1995, pp.383-393

Литература дополнительная

6. *Палий И.А.* Прикладная статистика: Учеб. пособие для вузов. / И.А. Палий. – М.: Высш. шк., 2004. С.150-154
7. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.38-40

Тема 14. Оценивание параметров: точечное и интервальное. Доверительные интервалы для среднего арифметического, для медианы, для доли. Средняя и предельная ошибки выборки

Два вида оценивания: точечное и интервальное. Понятие доверительного интервала и принципы его построения (на примере математического ожидания, медианы, доли). Уровень значимости, его содержательный смысл. Разные виды доверительного интервала для математического ожидания признака в зависимости от того, известна или неизвестна исследователю заранее генеральная дисперсия этого признака.

Оценка доли как частный случай оценки математического ожидания для специальным образом организованных дихотомических переменных. Обобщение подхода на другие способы многомерного статистического анализа.

Требования, предъявляемые к статистикам, используемым для оценок: несмещенность, состоятельность, эффективность. Содержательный смысл этих требований. Модель, на которую опираются все способы оценивания. Понятия средней и предельной ошибки выборки для оцениваемого параметра.

Литература основная

1. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.71-95
2. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. С.145-152
3. *Bluman A.G.* Elementary statistics. Wm. C. Brown Publishers, 1995, pp. 273-294

Литература дополнительная

4. *Агабекян Р.Л., Кириченко М.М., Усатииков С.В.* Математические методы в социологии. Анализ данных и логика вывода в эмпирическом исследовании: учеб. пособие для вузов. Ростов н/Д: Феникс. 2005. С. 124-129.
5. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С. 157-163, 174-180
6. *Ниворожкина Л.И., Морозова З.А.* Теория вероятностей и математическая статистика / Л.И. Ниворожкина, З.А. Морозова. – М.: Эксмо, 2008. С.249-261
7. *Палий И.А.* Прикладная статистика: Учеб. пособие для вузов. / И.А. Палий. – М.: Высш. шк., 2004. С.102-116

Тема 15. Объем простой случайной выборки

Определение объема выборки с помощью доверительного интервала для математического ожидания признака. Основные способы получения предварительных сведений о том, чему равна дисперсия изучаемого признака в исследовании. Плюсы и минусы такого подхода.

Литература основная

1. *Девятко И.Ф.* Методы социологического исследования. / И.Ф. Девятко. – 4-е изд. – М.: КДУ, 2006, с. 199-209.
2. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. С.445-463

3. Чуриков А.В. Основы формирования выборки: лекции для студентов направления 040100.62 (Социология) – М.: ГУ-ВШЭ, 2005, с. 8-38.
4. *Bluman A.G. Elementary statistics.* Wm. C. Brown Publishers, 1995, pp.273-294

Литература дополнительная

5. *Давыдов А. А.* Репрезентативность выборки // Социологические исследования. – 1990, № 1, с. 115-121.
6. *Рабочая книга социолога* / отв. ред. Г.В.Осипов. М.: Наука, 1983, с.200-236

Тема 16. Виды распределений случайных величин: Хи-квадрат-распределение (Пирсон), t-распределение (Стьюдент), F-распределение (Фишер). Стандартные таблицы функций этих распределений

Общая характеристика нормального распределения (кратко повторение положений теории вероятностей). Роль нормального распределения для социолога.

Распределения, основанные на нормальном: распределения «Хи-квадрат», Стьюдента, Фишера. Для всех распределений - формула функции плотности распределения, характер графиков функции распределения и функции плотности. Изменение характера кривых в зависимости от числа степеней свободы (для последних трех распределений). Работа с таблицами распределений.

Литература основная

1. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.51-64, С. 222-243 (таблицы распределений)
2. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. С.67-85
3. *Bluman A.G. Elementary statistics.* Wm. C. Brown Publishers, 1995, pp.608-622 (tables)

Литература дополнительная

4. *Агабекян Р.Л., Кириченко М.М., Усатилов С.В.* Математические методы в социологии. Анализ данных и логика вывода в эмпирическом исследовании: учеб. пособие для вузов. Ростов н/Д: Феникс. 2005. С. 114-118.
5. *Ниворожкина Л.И., Морозова З.А.* Теория вероятностей и математическая статистика / Л.И. Ниворожкина, З.А. Морозова. – М.: Эксмо, 2008. С.241-249
6. *Оуэн Д.Б.* Сборник статистических таблиц. 2-е изд., испр. М.: ВЦ АН СССР, 1973.

7. *Хастингс Н., Пикок Дж.* Справочник по статистическим распределениям. М.: Статистика, 1980

Тема 17. Понятие статистической гипотезы. Направленные и ненаправленные альтернативные гипотезы. Односторонние и двусторонние критерии. Ошибки 1-го, 2-го рода.

Общее представление о статистической гипотезе, способах и целях ее проверки. Общее представление о направленных и ненаправленных гипотезах. Подходы к определению того, с какой гипотезой исследователь имеет дело. Связь с содержательным характером задачи. Определение односторонних и двусторонних (однохвостовых и двуххвостовых) критериев. Различие в их использовании при проверке статистических гипотез. Специфика учета двух «хвостов» при использовании распределения Фишера.

Ошибки первого и второго рода. Разъяснение сути подхода на примерах проверки рассмотренных статистических гипотез. Мощность критерия.

Литература основная

1. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.101-103, 120-123
2. *Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. С.93-107
3. *Bluman A.G.* Elementary statistics. Wm. C. Brown Publishers, 1995, pp.300-311

Литература дополнительная

4. *Ниворожкина Л.И., Морозова З.А.* Теория вероятностей и математическая статистика / Л.И. Ниворожкина, З.А. Морозова. – М.: Эксмо, 2008. С.284-289, 292-293, 295-297
5. *Палий И.А.* Прикладная статистика: Учеб. пособие для вузов. / И.А. Палий. – М.: Высш. шк., 2004. С.127-130

Тема 18. Проверка статистических гипотез об отсутствии связи, о значении мат. ожидания определенному числу, о равенстве математических ожиданий, о равенстве долей, о равенстве дисперсий, о равенстве нулю коэффициента корреляции.

Понятие о двумерной частотной таблице и способах отражения в ее виде независимости рассматриваемых признаков. Теоретическая частота и формула для ее вычисления (с доказательством). Принцип невозможности реализации маловероятных событий и его использование при проверке статистической гипотезы. Правило «трех сигм». Роль выбора уровня значимости.

Логика проверки гипотезы об отсутствии связи между двумя номинальными (или рассматриваемыми как номинальные) признаками на основе критерия «Хи-квадрат».

Зависимые и независимые выборки. Проверка гипотезы для независимых выборок. Учет того, известны или нет генеральные дисперсии. В случае, если они неизвестны – проверка гипотезы об их равенстве. Цели такой проверки. Учет объема выборки. В каждой рассматриваемой ситуации – вид критерия для проверки гипотезы, функция его распределения, в случае необходимости – определение числа степеней свободы.

Проверка гипотезы для зависимых выборок.

Формула для критерия во всех остальных рассматриваемых случаях. Способ определения числа степеней свободы (в случае необходимости). Логика проверки всех рассматриваемых гипотез. Примеры социологических задач, где такая проверка требуется.

Влияние содержательного характера задачи на выбор уровня значимости. Примеры. Различие между статистической и содержательной гипотезами. Примеры возможности проверки одной и той же содержательной гипотезы с помощью проверки разных статистических гипотез и наоборот.

Литература основная

1. Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.108-120, 125-144
2. Bluman A.G. Elementary statistics. Wm. C. Brown Publishers, 1995, pp.311-337, 344-372, 419-422

Литература дополнительная

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С.206-229, 239-244

Тема 19. Проверка статистических гипотез о виде распределения данных (критерий согласия Пирсона). Равномерное распределение. Нормальное распределение. Распределение Пуассона. Биномиальное распределение

Логика проверки рассматриваемых гипотез о том, с каким из распределений согласуются данные. Примеры социологических задач, где такая проверка требуется. Формула для критерия Пирсона. Способ определения числа степеней свободы.

Литература основная

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С.251-283
2. Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.123-125

Литература дополнительная

3. Палий И.А. Прикладная статистика: Учеб. пособие для вузов. / И.А. Палий. – М.: Высш. шк., 2004. С.84-96

Тема 20. Изучение причинно-следственных отношений в социологии с помощью математических методов.

Методологические аспекты изучения причинно-следственных отношений: исторический аспект. Коротко – о понимании каузальных отношений у Ньютона, Чупрова, Лазарсфельда. Возможность связать методологические принципы использования математического аппарата для изучения причин с некоторыми идеями Юма (индуктивное понимание причинности, переход к нестатистическим моделям) и Канта (априорная модель понимания причинности, заложенная в используемом методе). Невозможность формализовать понятия причины и следствия. Выделение двух основных направлений изучения каузальных отношений с помощью математических методов: построение систем структурных уравнений (SEM) и эксперимент.

Роль математической статистики в проведении эксперимента. Специфика эксперимента в социологии. Нестатистические методы проведения эксперимента. Основные идеи Милля и их развитие в настоящее время.

Литература основная

1. Толстова Ю.Н. Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.144-163

О понятии причины

1. Давыдов Ю.Н. Макс Вебер и современная теоретическая социология. М.: Мартис, 1998. С. 150-190.
2. Лаплас. Опыт философии теории вероятностей // Вероятность и математическая статистика. Энциклопедия. М.: Большая российская энциклопедия, 1999.

Литература дополнительная

О понятии причины

3. *Аристотель*. Метафизика // Аристотель. Сочинения. Т.1. М.: Мысль, 1976. С.63-367.
4. *И.Кант*. Теория познания // Антология мировой философии Т.3. М.: Мысль, 1971. С. 100-155.
5. *Милль Д.С.* Система логики символической и индуктивной, изд.2. М., 1914 // Антология мировой философии. Т.3. М.: Мысль, 1971. С. 594-605.
6. *Финн В.К.* Интеллектуальные системы и общество. М.: КомКнига, 2006
7. *Чупров А. А.* Нравственная статистика // Брокгауз Ф.А. (Лейпциг), Ефрон И.А. (СПб.). Энциклопедический словарь. Т. XXI. С.-Петербург: Типолитография И.А.Ефрона, 1897. С. 403–408.
8. *Чупров А.А.* Вопросы статистики. М.: Госстатиздат ЦСУ СССР, 1960.
9. *Юм Д.* Сокращенное изложение «Трактата о человеческой природе» // Антология мировой философии. Т.2. М.: Мысль, 1970. С. 575-593.

Об эксперименте в социологии

10. *Адлер Ю.П., Ковалёв А.Н.* Математическая статистика и планирование эксперимента в науке о человеке // Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. С. 477-490
11. *Андреева Г.М.* Социальная психология. М.: Наука, 1994. С.174-175 (Хоторнский эксперимент)
12. *Батыгин Г.С.* Лекции по методологии социологических исследований. М.: Аспект Пресс, 1995. С. 190-227
13. *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М.: Прогресс, 1976, с. 405-436 однофакторный и многофакторный дисперсионный анализ: случайные, смешанные и постоянные эффекты), с. 437-458 (Основы планирования эксперимента).
14. *Джонсон Н., Лион Ф.* Статистика и планирование эксперимента в технике и науке. М.: Мир, т.1 - 1980, т.2 - 1981
15. *Дэниел К.* Применение статистики в промышленном эксперименте. М.: Мир, 1979
16. *Кравченко А.И.* Хоторнский эксперимент //Справочное пособие по истории немарксистской западной социологии. М.: Наука, 1986. С.452-453
17. *Куприян А.П.* Методологические проблемы социального эксперимента. М.: Наука, 1971
18. *Куприян А.П.* Проблема эксперимента в системе общественной практики. М.:Наука, 1981

19. *Кемпбелл Д.* Модели экспериментов в социальной психологии и прикладных исследованиях. М.: Прогресс, 1980. Переиздано в СПб: Изд-во «Социально-психологический центр», 1996
20. *Киш Л.* Представительность, рандомизация и контроль // Математика в социологии. М.: Мир, 1977, с. 201-223.
21. *Методы сбора информации в социологических исследованиях.* М.: Наука, 1990.Т.2. С. 190-214
22. *Методы сбора данных: анализ документов, наблюдение, эксперимент.* М.: ИСИ АН СССР, 1985
23. *Методы социальной психологии.* Л.: изд-во ЛГУ, 1977. С.132-150
24. *Милграм С.* Эксперимент в социальной психологии. СПб: Питер, 2000
25. *Монсон П.* Современная западная социология. Теории, традиции, перспективы. С.Пб.: Нотабене, 1992. с.160-162 (эксперимент Цимбардо)
26. *Налимов В.В.* Теория эксперимента. М.: Наука, 1971
27. *Основы прикладной социологии.* М.: Интерпракс, 1996. С.68-72
28. *Рабочая книга социолога.* М.: Наука, 1983. С. 411-432
29. *Руткевич М.Н.* Макросоциология. Методологический очерк. М.: РАН, 1995. С.16-19
30. *Статистические методы анализа информации в социологических исследований.* М.: Наука, 1979. С.178-194
31. *Хагуров А.А.* Социальный эксперимент: логико-методологические и социальные проблемы. Ростов-на-Дону, 1989
32. *Хикс Ч.* Основные принципы планирования эксперимента. М.: Мир, 1967
33. *Ядов В.А.* Социологическое исследование: методология, программа, методы. Самара: «Самарский ун-т», 1995. С.220-231
34. *Blalock H.M.* Causal inferences in nonexperimental research. Chapel Hill: University of North Carolina Press, 1964

О причинном анализе

35. *Бестужев-Лада И. В. , Варыгин В. Н. , Малахов В. А.* Моделирование в социальных исследованиях. М. : Наука, 1978.
36. *Богомолова Е.А., Наумова Н.Ф.* Структурные модели как инструмент обобщения и интерпретации социальной информации на выходе системы моделирования // Неформализованные элементы системы моделирования. М.: ВНИИСИ, 1980

37. *Бородкин Ф.М.* Об одной схеме причинного анализа // Математика и социология. Новосибирск: ИЭиОПП СО АН СССР, 1970
38. *Бородкин Ф.М.* Научный эксперимент в социально-экономических исследованиях. Дисс. На соискание ученой степени доктора экономических наук. Новосибирск, 1975
39. *Бунге М.* Причинность. М.: Изд. Иностр. лит., 1962
40. *Волд Г.* Путевые модели с латентными переменными: подход NIPALS // Математика в социологии: моделирование и обработка информации. М.: Мир, 1977. С. 241-281.
41. *Гаврилец Ю. Н.* Структура связей и причинные зависимости между переменными // Математика в социологии: моделирование и обработка информации. М. : Мир, 1977. С. 135-169.
42. *Елисеева И. И.* Статистические методы измерения связей. Л. : Изд-во ЛГУ, 1982. С. 97-108. (Структурные модели. Путевой анализ).
43. *Елисеева И. И. , Рукавишников В. О.* Логика прикладного статистического анализа. М. : Финансы и статистика, 1982. С. 72-149 (Структурные и причинные модели).
44. *Левин К.* Теория поля в социальных науках. СПб: Сенсор, 2000, с. 7-14, 178-192, 213-220
45. *Математические методы анализа и интерпретация социологических данных.* М. : Наука, 1989. С.61-94 (Выбор стратегии анализа взаимосвязи признаков)
46. *Новак С.* Причинные интерпретации статистических связей в социальном исследовании // Математика в социологии: моделирование и обработка информации. М. : Мир, 1977. С. 76-123.
47. *Осипов Г. В. , Андреев Э. П.* Методы измерения в социологии. М. : Наука, 1977.
48. *Осипов Г.В., Андреев Э.П.* Эмпирическое обоснование гипотез в социологических исследованиях // Социс, 1974, № 1. С. 160-173.
49. *Рукавишников В.О.* Информационный подход к причинному анализу // Модели социально-экономических процессов и социальное планирование. М., 1979.
50. *Социальные исследования: построение и сравнение показателей.* М.: Наука, 1978. С.104-111 (Построение показателей в процессе применения метода причинных моделей).
51. *Статистические методы анализа информации в социологических исследованиях.* М. : Наука, 1979. С. 267-282 (Модели для анализа структуры причинных связей).
52. *Сунтес П.* Вероятностный анализ причинности // Математика в социологии: моделирование и обработка информации. М. : Мир, 1977. С. 50-75.
53. *Таганов И.Н.* Информационные меры причинного влияния // Математика в социологии: моделирование и обработка информации. М. : Мир, 1977. С. 124-134.

54. *Татарова Г. Г.* Структура многомерной случайной величины и проблема взаимосвязи признаков // Социологические исследования, 1986. №3. С. 142-148.
55. *Трофимов В.П.* Измерение взаимосвязей социально-экономических явлений. М.: Статистика. 1975. С. 15-29 (Соотношение причинной и корреляционной связи).
56. *Хейс Д.* Причинный анализ в статистических исследованиях. М. : Финансы и статистика, 1981.
57. *Blalock H. M.* Causal Inferences in Nonexperimental Research, Chapel Hill: university of North Carolina Press, 1964, с. 3-60, 95-96, 172-188
58. *Blalock H. M.* Causal models in the Social Sciences, 1970
59. *Blalock H. M.* Theory construction. From verbal to mathematical formulation. Prentice hall, New Jersey, 1969, p. 1-30
60. *Bollen, K. A.* Structural Equations with Latent Variables. NY: John Wiley & Sons, 1989
61. *Knoke D.* A causal model for the political party preferences of american men // Amer. Soc. Review, 1972. P. 679-689
62. *Spilerman S.* Forecasting social events // Social indicator model, S.N.Y. 1975
63. *Suppes P.* Probabylistic Theory of Causality. Amsterdam: North-Holl P/ Co.,1970

Sage University Paper series on Quantitative Applications in the Social Sciences:

64. n. 07-003. *Asher H.* Causal modeling, 1976,1980
65. n. 07-034. *Long.* Covariance Structure Models, 1983
66. n. 07-037. *Berry W.D.* Nonrecursive Causal Models, 1984
67. n. 07-055. *Davis.* The Logic of causal Order, 1985
68. n. 07-074. *Brown, Melamed.* Experimental design and analysis, 1990
69. n. 07-105. *Finkel S.E* Causal analysis of panel data, 1995
70. n. 07-114. *Jaccard J., Wan C.K.* LISREL Approaches to Interaction Effects in Multiple regression
71. n. 07-135. *Jaccard J.* Interaction effects in logistic regression, 2001

О SEM (структурные уравнения)

72. *Назаров Б.В., Мальцев В.Е.* Структурные ковариационные модели в социологии. М.: ИД ГУ-ВШЭ, 2008
73. *Golob Thomas F.* 2003. Structural Equation Modeling for Travel Behavior Research. Center for Activity Systems Analysis. Published in: Transportation Research, Vol. 37B, 2003, pp. 1-15.

74. *Hox J.J. Bechger T.M.* 1998. An Introduction to Structural Equation Modeling. *Family Science Review*, 11,354-373.
75. *McArdle John J. , Johnson Ronald C.* 2001. Structural Equation Modeling of Group Differences in CES-D Ratings of Native Hawaiian and Non-Hawaiian High School Students *Journal of Adolescent Research*, Vol. 16 No. 2, March 2001 108-149 Sage Publications, Inc.
76. *Mueller R.O.* 1996. Basic Principles of Structural Equation Modeling. Springer Verlag, January.

Тема 21. Корреляционное отношение. Однофакторный дисперсионный анализ

Общее представление о линейной и нелинейной связи между переменными. Границы применимости коэффициента корреляции как показателя линейной связи. Зависимая и независимая переменные. Внутригрупповой, межгрупповой, общий разбросы значений зависимой переменной. Связь между величинами этих разбросов. Корреляционное отношение: формула, содержательный смысл, значение для социолога. Проблемы, не решаемые с помощью корреляционного отношения

Однофакторный дисперсионный анализ. Основная модель. Класс решаемых задач. Представление данных. Дисперсионный анализ как проверка статистической гипотезы о равенстве средних (формулировка гипотезы, используемый критерий, логика проверки). Понимание причинно-следственной связи при использовании однофакторного дисперсионного анализа (априорная модель такой связи).

Множественные сравнения. Суть метода. Цель и принципы его использования.

Литература основная

1. *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М.: Прогресс, 1976. С.138-141
2. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие. – 11-е изд., перераб. – М.: Высшее образование, 2006. С. 283-294 (Однофакторный дисперсионный анализ)
3. *Калинина В.Н., Панкин В.Ф.* Математическая статистика. М.: Высшая школа, 1998. С.244-266 (Однофакторный и двухфакторный дисперсионный анализ)
4. *Колемаева В.А., Калинина В.Н.* Теория вероятностей и математическая статистика. М.: Инфра-М., 1997. С.184-191; ЮНИТИ, 2003 (Однофакторный и двухфакторный дисперсионный анализ)
5. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. М.: Юнита-Дана, 2001. С.375-391 (Однофакторный и двухфакторный дисперсионный анализ)

6. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.168-190.

Литература дополнительная

7. *Haggard E.A.* Intra-class Correlation and the Analysis of Variance. N.Y.: Dryden Press, 1958
8. *Rozemboom W.W.* Foundations of the Theory of Prediction. Homewood, Illinois, Dorsey Press, 1966.

Тема 22. Двухфакторный дисперсионный анализ

Основная модель. Класс решаемых задач. Представление данных. Понятие взаимодействия. Его роль для социолога. Двухфакторный дисперсионный анализ как проверка статистических гипотез (формулировка гипотез, используемые критерии, логика проверки). Понимание причинно-следственной связи при использовании двухфакторного дисперсионного анализа (априорная модель такой связи).

Литература основная

1. *Горелова Г.В., Кацко И.А.* Теория вероятностей и математическая статистика в примерах и задачах с применением Excel. Ростов н/Д: Феникс, 2005. С.207-239 (Однофакторный, двухфакторный, трехфакторный дисперсионный анализ)
2. *Калинина В.Н., Панкин В.Ф.* Математическая статистика. М.: Высшая школа, 1998. С.244-266 (Однофакторный и двухфакторный дисперсионный анализ)
3. *Колемаева В.А., Калинина В.Н.* Теория вероятностей и математическая статистика. М.: Инфра-М., 1997. С.184-191; ЮНИТИ, 2003 (Однофакторный и двухфакторный дисперсионный анализ)
4. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. М.: Юнита-Дана, 2001. С.375-391 (Однофакторный и двухфакторный дисперсионный анализ)
5. *Толстова Ю.Н.* Математико-статистические модели в социологии. М.: ИД ГУ-ВШЭ, 2007. С.191-196

Тема 23. Статистическая неопределённость. Классическая и субъективистская вероятностные концепции. Парадигма Неймана-Пирсона, парадигма правдоподобия, байесовская парадигма. Место подтверждения в теориях.

Различия во взглядах на то, как должна быть определена вероятность, существуют десятки лет. Цель темы – показать различия теорий и решений простейших прикладных задач.

Последовательно сравниваются классическая (частотная), и субъективная (байесовская) теории. Рассматривается место правдоподобия: как третья парадигма (аргументация Фишера), как мощный инструмент классической теории и как необходимая составляющая субъективной теории.

Литература основная

1. Pawitan Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Oxford Science Publications, 2005, pp.1-19

Литература дополнительная

2. Royall R. Statistical Evidence: A Likelihood Paradigm. Chapman & Hall/CRC, 2000, pp.35-60

Тема 24. Определение правдоподобия. Объединение правдоподобий. Закон правдоподобия. Принцип правдоподобия. Кривизна функции правдоподобия. Свойства правдоподобия. Интерпретация правдоподобия.

Определение правдоподобия, функции правдоподобия. Элементы теории работы с правдоподобием: объединения правдоподобий, закон правдоподобия, принцип правдоподобия. Особое внимание уделяется форме функции правдоподобия, её изучению и интерпретации.

Литература основная

1. Pawitan Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Oxford Science Publications, 2005, pp.21-29

Литература дополнительная

2. Gill J. Bayesian Methods: A Social and Behavioral Science Approach. Chapman & Hall/CRC, 2002, pp. 31-48
3. Royall R. Statistical Evidence: A Likelihood Paradigm. Chapman & Hall/CRC, 2000, pp.1-34

Тема 25. Отношение правдоподобий, критерий Вальда и множитель Лагранжа.

Показывается взаимосвязь функции правдоподобия с критериями, позволяющими тестировать сложные гипотезы.

Рассматривается статистическая мощность и прочие свойства каждого из критериев. Особое внимание уделяется сложным гипотезам и применению критериев в линейных моделях.

Литература основная

1. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика, ч.1. М.: Юнити, 2001, с.290-300
2. Pawitan Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford: Oxford Science Publications, 2005, pp.29-43

Литература дополнительная

3. Green W. H. Econometric Analysis (5th edition). Pearson Education, Inc., 2003, pp.482-491

Тема 26. Оценивание параметров моделей. Метод наибольшего правдоподобия и его свойства.

Формально излагается метод наибольшего правдоподобия, его свойства. МНП сравнивается с поведением метода наименьших квадратов в линейных моделях и при нарушениях предположений применения методов.

Отдельно указываются и иллюстрируются место, преимущества и недостатки МНП перед альтернативами.

Литература основная

1. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика, ч.1. М.: Юнити, 2001, с.249-257

Литература дополнительная

2. Green W. H. Econometric Analysis (5th edition). Pearson Education, Inc., 2003, pp.468-524

Тема 27. Правдоподобие и байесовская теория. Простейшие байесовские модели.

Логика и простейшие модели байесовской теории. Показывается центральное место правдоподобия в байесовских моделях.

Отдельно рассматриваются простейшие задачи, демонстрирующие сложности, недостатки и преимущества байесовской теории перед альтернативами.

Литература основная

1. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика, ч.1. М.: Юнити, 2001, с.269-282

Литература дополнительная

2. Burnham K. P., Anderson D. R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd edition). Springer, 2002, pp.49-84, 267-304

3. *Gill J. Bayesian Methods: A Social and Behavioral Science Approach. Chapman & Hall/CRC, 2002, pp. 1-20, 65-83*
4. *Royall R. Statistical Evidence: A Likelihood Paradigm. Chapman & Hall/CRC, 2000, pp.167-176*

Литература ко всем темам по математической статистике

Литература основная

1. *Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика, ч.1. М.: Юнити, 2001*
2. *Андропов А.М., Копытов Е.А., Гринглаз Л.Я. Теория вероятностей и математическая статистика. Учебник для вузов. С.-Пб: Питер, 2004*
3. *Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. С.-Пб: Лань, 2004*
4. *Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1998*
5. *Гмурман В.Е. Руководство к решению задач по теории вероятностей математической статистике. М.: Высшая школа, 1998*
6. *Калинина В.Н., Панкин В.Ф. Математическая статистика. М.: Высшая школа, 1998*
7. *Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика, М.: Инфра-М,1997; Юнити, 2003*
8. *Кочетков Е.С., Смерчинская С.О., Соколов В.В. Теория вероятностей и математическая статистика. М.: Изд. дом «Форум», 2003*
9. *Кремер Н.Ш. Теория вероятностей и математическая статистика. М.: ЮНИТИ-ДАНА, 2001*
10. *Теория статистики с основами теории вероятностей / Под ред. И.И.Елисеевой. М.: Юнити, 2001*
11. *Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. М.: ИНФРА-М, 2003*
12. *Шведов А.С. Теория вероятностей и математическая статистика. М.: Изд. дом ГУ-ВШЭ, 2005*
13. *Эддоус М., Стэнсфилд Р. Методы принятия решения. М.: Финансы и статистика, 1997*
14. *Bluman A.G. Elementary statistics. A step by step. McGraw-Hill Companies. 1992, 1995, 1998, 2001*

Литература дополнительная

1. *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М.: Прогресс, 1976
2. *Горелова Г.В., Кацко И.А.* Теория вероятностей и математическая статистика в примерах и задачах с применением Excel. Ростов-на-Дону: Феникс, 2005
3. *Интерпретация и анализ данных в социологических исследованиях.* М.: Наука, 1987
4. *Паниотто В.И.* Количественные методы в социологических исследованиях. Киев: Наукова думка, 1982
5. *Рабочая книга социолога.* М.: Наука, 1983
6. *Статистические методы анализа социологической информации.* М.: Наука, 1989
7. *Татарова Г.Г.* Методология анализа данных в социологии. М., 1998
8. *Толстова Ю.Н.* Анализ социологических данных: методология, дескриптивная статистика, анализ связей номинальных признаков. М.: Научный мир, 2000
9. *Шеффе Г.* Дисперсионный анализ. М.: ГИФМЛ, 1963
10. *Hinton P.R.* Statistics Explained. A Guide for social Science Students. - N.-J.,L.: Routledge, 1995
11. *Kachigan S.K.* Statistical analysis. An interdisciplinary introduction to univariate and multivariate methods. - N.Y.: Radius Press, 1986
12. *Sirkin R.M.* Statistical for the social sciences. L., New Delhi: Sage publ., 1995
13. *Walsh A.* Statistical for the social sciences: with computer - based applications. - N.Y.: Harper & Row, Publishers, 1990.

Справочники, энциклопедии

1. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М., 1983
2. *Вероятность и математическая статистика.* Энциклопедия. М.: БРЭ, 1999
3. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. М., 1978
4. *Математическая энциклопедия,* в 5-ти томах. М.: БСЭ, 1977-1985
5. *Поллард Дж.* Справочник по вычислительным методам статистики. М.: Финансы и статистика, 1982
6. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. Основные понятия. Предельные теоремы. Случайные процессы (справочник). М., 1967
7. *Рунион Р.* Справочник по непараметрической статистике. Современный подход. М.: Финансы и статистика, 1982
8. *Социологическая энциклопедия,* в 2-х томах. М.: Мысль, 2003

9. *Справочник по прикладной статистике* (под ред. Э.Ллойда, У.Ледермана). В 2-х томах. М.: Финансы и статистика, 1989, 1990
10. *Справочник по теории вероятностей и математической статистике*. Киев, 1978
11. *Хастингс Н., Пикок Дж.* Справочник по статистическим распределениям. М.: Статистика, 1980
12. *Энциклопедический социологический словарь*. М., 1996
13. *Handbook of survey research* (Quantitative studies in social relations) (ed.by P.H.Rossi, J.D.Wright, A.V.Anderson). Academic Press, inc. LTD, 1983

Вопросы для оценки качества освоения дисциплины

Раздел I Теория вероятностей:

Примеры задач

1. Преподаватель готовит билеты к экзамену. К экзамену по предмету А преподаватель заготовил 8 теоретических вопросов и 4 задачи. К экзамену по предмету Б – 7 теоретических вопросов и 5 задач. К какому экзамену можно составить большее количество билетов, если в билет включать 1 теоретический вопрос и 1 задачу?

2. В анкете респондент должен расставить по привлекательности три профессии (А В С). Сколько возможных вариантов можно получить?

3. После ответов на вопросы скрининговой анкеты осталось 10 кандидатов, но все они мужчины. А для опроса нужно только 5 мужчин. Сколькими способами можно выбрать участников опроса?

4. Вопрос в опросе на портале <http://www.ecsocman.edu.ru> звучит следующим образом: Если вам необходима справочно-энциклопедическая информация, то к каким источникам вы предпочитаете обращаться? Выберите только один вариант ответа.

	Всего
К любым "бумажным" энциклопедиям и справочникам	206
К Wikipedia	214
К Большой Российской Энциклопедии	150
Всего опрошено:	470

Найти вероятность того, что респонденты предпочитают обращаться:

- А) к "бумажным" изданиям
- Б) к "бумажным" изданиям и Wikipedia
- В) к Wikipedia и БРЭ

Г) к БРЭ

5. Вопрос в опросе на портале <http://www.ecsocman.edu.ru> звучит следующим образом:

Если вам необходима справочно-энциклопедическая информация, то к каким источникам вы предпочитаете обращаться? Выберите все подходящие Вам варианты ответов.

	Всего
К любым "бумажным" энциклопедиям и справочникам	234
К Wikipedia	354
К Большой Российской Энциклопедии	250
Всего опрошено:	470

Найти вероятность того, что респонденты предпочитают обращаться:

А) к "бумажным" изданиям и Wikipedia

Б) к "бумажным" изданиям

В) к Wikipedia и БРЭ

6. Перед Вами результаты опроса на портале <http://www.ecsocman.edu.ru>. Вопросы и ответы представлены в таблицах:

1. Как вы оцениваете принятое решение Правительством РФ о введении обязательного полного среднего образования (11 лет обучения):

	%
положительно	47,0
скорее положительно	14,0
нейтрально	7,0
скорее отрицательно	14,0
отрицательно	18,0

2. Вы... –

	%
Школьник	9,0
Студент 1-3 курсов	21,0
Студент 4-5 курсов / бакалавриата	15,0
Студент магистратуры / аспирант	12,0
Преподаватель/исследователь	17,0
Родитель	13,0

Другое

13,0

Признаки независимы.

Найти вероятность того, что случайно выбранный респондент будет:

1. студентом 4-5 курса, положительно оценивающим решение.
2. не аспирантом, скорее отрицательно оценивающим решение.
3. студентом 1-3 курсов или студентом 4-5 курса отрицательно оценивающим решение.

7. В выборке 700 человек, 200 из которых имеют высшее образование. Вероятность того, что респондент с высшим образованием является противником отмены литературы как обязательного предмета в школе, равна 0,95. Для респондентов без высшего образования эта вероятность равна 0,75. А) Найти вероятность того, что произвольно взятый респондент будет противником указанного действия. Б) Оказалось, что респондент против отмены литературы, какова вероятность, что у него нет высшего образования.

8. В опросе на портале <http://www.ecsocman.edu.ru> задавалось два вопроса: 1-й: "Если вам необходима справочно-энциклопедическая информация, то к каким источникам вы предпочитаете обращаться?" Выберите только один вариант ответа. И 2-й: "Вы учитесь / работаете в области?" (Варианты ответа экономики, социологии, менеджмента, другое).

Было получено следующее распределение ответов:

	экономики	социологии	менеджмента	другое	Всего
К любым "бумажным" энциклопедиям и справочникам	40	10	35	22	107
К Wikipedia	50	23	18	44	135
К Большой Российской Энциклопедии	57	16	30	71	174
Всего:	147	49	83	137	416

А) Найти вероятность того, что респондент предпочитает обращаться к Wikipedia, учась/работая в сфере менеджмента.

Б) Если принять первый вопрос за событие А, а второй – за событие В, то проверьте равенство $P(A \text{ и } B) = P(A)P(B) = P(A)P(B|A) = P(B)P(A|B)$

9. Вероятность того, что студент сдаст экзамен на положительную оценку, равна 0,8, какова вероятность, что в группе из 18 человек экзамен сдадут:

- а) 14 человек
- б) не менее 16.

10. По результатам исследования оказалось, что 82% жителей района против строительства торгово-развлекательного центра в непосредственной близости от их дома. Найти вероятность того, что среди 6 случайно выбранных респондентов 3 будут против.

11. В лотерее участвуют всего 100 билетов. В 1 выигрыш 500р., в 10 – 100р, 89 пустых. Найти средний размер выигрыша.

12. В опросе на портале <http://www.ecsocman.edu.ru> задавалось два вопроса: 1-й: "Надо ли рассказывать детям правду о Дедушке Морозе?" Выберите только один вариант ответа. И 2-й: "Вы учитесь / работаете в области?" (Варианты ответа экономики, социологии, менеджмента, другое).

Было получено следующее распределение ответов:

	экономики	социологии	Итого
нет, не надо рассказывать, пускай верят, что он существует	78%	91%	
да, лучше рассказать, а то в будущем их ждет тяжелое разочарование	22%	9%	
Итого	72%	28%	100%

А) Найти безусловную вероятность того, что человек считает, что не надо говорить детям правду а Деде Морозе.

Б) Оказалось, что случайно выбранный респондент считает, что не надо говорить детям правду о Деде Морозе, какова вероятность, что он учится/работает в сфере экономики?

13. Дискретная случайная величина X принимает следующие значения:

X	3	4	7	10
p	0,2	0,1	0,4	0,3

Найти функцию распределения $F(x)$ и построить ее график.

14. По статистике 12% пользователей некоторого Интернет-магазина делают в нем «крупные» покупки. Если известно, что в течение суток в магазин обратились 500 человек, то какова вероятность того, что 55 из них сделали «крупные» покупки?

15. Потенциальный кандидат рассчитывает, что 30% голосующих членов его партии проголосуют за предложенный им проект закона о налогах. Если в голосовании принимает участие 400 человек, определите вероятность того, что как минимум 100 голосующих будут голосовать за его проект закона о налогах.

16. Значения в тесте были нормально распределены и имели среднее, равное 10, а стандартное отклонение — равное 3. Для каждого приведенного ниже значения определите Z-значение и процент области перед и после этого значения.

X	Z-значение	% области после	% области перед
3			
16			

17. Шкала, измеряющая предубеждение респондентов к какому-то явлению была использована для большой выборки. Распределение баллов по этой шкале приблизительно нормальное, среднее равно 31, а стандартное отклонение — 5. Какой процент выборки получили:

меньше 20 баллов;

18. Дано нормальное распределение, у которого $\mu=62$. Найти параметр σ , если 78,52% площади лежит справа от 55.

Определить диапазон значений для $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$.

Какой процент наблюдений попадет в эти интервалы?

Придумайте переменную, которая могла бы иметь нормальное распределение с данными параметрами.

19. Вероятность наступления события A в каждом из 1000 независимых испытаний равна 0,8. Оцените вероятность того, что событие A отклонится от математического ожидания случайной величины X – числа наступления события A – меньше, чем на 50.

20. Было проведено исследование с целью установить связь между возрастом водителя и количеством аварий за год. Данные приведены ниже. Если существует значительная связь, предскажите количество аварий для 28-летнего водителя.

Возраст водителя, x	16	24	18	17	23	27	32
Кол-во аварий, y	3	2	5	2	0	1	1

21. На основе анкетного опроса была получена следующая частотная таблица:

Возраст, (лет)	Зарплата, (у.е.)							Нет ответа
	300-400	400-500	500-600	600-700	700-800	800-900	900-1000	
15-20	20	10	0	5	15	10	5	0
20-25	50	100	100	20	0	0	0	10
25-30	0	50	50	20	5	0	5	0
30-35	5	10	20	30	0	2	10	5
35-40	2	5	10	40	50	30	20	0

Вычислить коэффициент корреляции между возрастом и зарплатой респондентов.

Если существует значительная связь между возрастом и зарплатой, то предскажите уровень заработной платы для 40 летнего человека.

22. Переменные X и Y независимые величины, распределенные нормально.

$X \sim N(10;9)$, $Y \sim N(8;16)$

$$V = 2X + 3Y + 1$$

Найти: EV , DV , $P(50 < V < 60)$

23. Переменные Z_1 и Z_2 независимые величины, распределенные по стандартному нормальному закону распределения

$$X = 2Z_1 + Z_2$$

$$Y = Z_1 - Z_2$$

Найти $\text{cov}(X,Y)$, $\text{cor}(X,Y)$

Раздел II Математическая статистика:

примерные экзаменационные вопросы

1. Основные задачи математической статистики
2. Виды случайных событий, изучаемых социологом. Примеры случайной величины в социологии
3. Функция распределения и функция плотности распределения. Их связь друг с другом.
4. Понятия выборки и генеральной совокупности. Способы переноса результатов с первой на вторую
5. Представление о параметрах распределениях и статистиках. Примеры
6. Основные параметры одномерных распределений – математическое ожидание, квантили, мода, медиана, среднее квадратическое отклонение, дисперсия
7. Основной параметр двумерного распределения – коэффициент корреляции. Его свойства, вид отражаемой им связи, недостатки.
8. Нормальное, равномерное распределение, распределение Пуассона, Бернулли. Основные параметры. Примеры важных для социолога случайных величин, имеющих названные распределения.
9. Стандартизированные случайные величины. Использование вероятностных таблиц нормального распределения.
10. Распределения, основанные на нормальном: хи-квадрат, Стьюдента, Фишера – аналитический вид случайной величины, графическое представление, расчет числа степеней свободы, математического ожидания и дисперсии этих величин. Использование вероятностных таблиц для этих распределений.
11. Определение номинальной, порядковой, интервальной шкалы. Понятие допустимого преобразования шкалы. Общее представление об адекватности метод типу шкал.
12. Обоснование адекватности (неадекватности) среднего арифметического для номинальной, порядковой, интервальной шкалы.
13. Для какого типа шкал коэффициент корреляции является адекватным и почему?
14. Для какого типа медиана является адекватной мерой средней тенденции и почему?
15. Среднее арифметическое и дисперсия для дихотомической шкалы
16. Дискретные и непрерывные признаки. Выборочные представления генеральных распределений. Частотные таблицы, полигоны, гистограмма, кумулята. Проблемы, возникающие при их построении. Гистограммы с неравными интервалами

17. Способы нахождения моды и медианы для выборки и выборочного частотного распределения. Модели, используемые при расчете медианы
18. Формула расчета коэффициента корреляции для выборки и выборочного частотного распределения
19. Двумерные частотные распределения. Маргинальные частоты. Условные и безусловные распределения
20. Теоретические частоты и формулы для них. Вид частотной таблицы для независимых случайных величин
21. Свойства точечных оценок параметров (эффективность, несмещенность, состоятельность). Их определение и содержательный смысл
22. Доверительный интервал для математического ожидания. Принципы его построения. Связь с центральной предельной теоремой
23. Средняя и предельная ошибки выборки. Определение объема выборки на основе её средней ошибки. Плюсы и минусы такого подхода
24. Логика проверки статистической гипотезы. Уровень значимости. Принципы его определения
25. Нулевая и альтернативная (конкурирующая) гипотезы. Примеры
26. Направленные и ненаправленные альтернативные гипотезы. Односторонний и двусторонний критерий проверки нулевой гипотезы. Примеры.
27. Ошибки первого и второго рода. Мощность критерия.
28. Проверка гипотезы об отсутствии связи между двумя номинальными переменными.
29. Общая логика проверки гипотез о равенстве двух средних
30. Проверка гипотезы о равенстве математического ожидания определенному числу
31. Проверка гипотезы о равенстве двух средних для зависимых выборок
32. Проверка гипотезы о равенстве двух средних для независимых выборок
33. Проверка гипотезы о равенстве двух дисперсий для независимых выборок
34. Проверка гипотезы о равенстве двух долей для независимых выборок
35. Проверка гипотезы об отличии от нуля коэффициента корреляции
36. Проверка гипотезы о равномерном распределении
37. Проверка гипотезы о нормальном распределении
38. Проверка гипотезы о биномиальном распределении
39. Проверка гипотезы о распределении Пуассона
40. Корреляционное отношение. Его смысл, вид отражаемой им связи

41. Межгрупповая, внутригрупповая, общая дисперсии при расчете корреляционного отношения. Соотношение между ними (с доказательством)
42. Общее представление об эксперименте в социологии. Понятие причины
43. Методический эксперимент в социологии. Роль математической статистики в его проведении
44. Логические схемы экспериментальной проверки гипотез по Миллю
45. Модель, заложенная в однофакторном дисперсионном анализе. Выборочные оценки её параметров. Смысл решаемых с помощью однофакторного дисперсионного анализа задач
46. Проверка гипотез в однофакторном дисперсионном анализе
47. Понятие взаимодействия и роль его изучения в социологии.
48. Модель, заложенная в двухфакторном дисперсионном анализе.
49. Основные вероятностные концепции (парадигмы) – классическая и субъективистская. Парадигма правдоподобия, байесовская парадигма.
50. Понятие правдоподобия. Закон правдоподобия. Принцип правдоподобия
51. Функция правдоподобия. Кривизна этой функции. Свойства правдоподобия.
52. Какую интерпретацию имеют значения функции правдоподобия? Напишите общую формулу получения правдоподобия для зависимых наблюдений. Дайте интерпретацию всех составляющих формулы.
53. Отношение правдоподобий. Используемые критерии
54. Логика оценивания параметров моделей. Метод наибольшего правдоподобия: идеи и свойства
55. Правдоподобие и байесовская теория. Простейшие байесовские модели

Раздел II Математическая статистика:

Примеры задач

1. (*Меры средней тенденции и разброса*) Король некой страны издал указ о том, что разброс зарплат министров его правительства (состоящего из 10 человек) в среднем не должен превышать 67,1 у.е. Через некоторое время он решил проверить выполнение указа. Его помощники выяснили, что члены его правительства зарабатывают, соответственно, 240, 256, 334, 176, 254, 219, 277, 414, 215, 366 у.е. Докажите, что указ короля не соблюдается. Исправится ли положение, если король выгонит самого высокооплачиваемого министра?

2. (Меры средней тенденции и разброса) В произведении А.П.Чехова «Остров Сахалин (из путевых записок)» приводятся следующие данные о каторжных¹:

Срок осуждения	До 8 лет	От 8 до 12 лет	От 12 до 15 лет	От 15 до 20 лет	От 20 до 50 лет (рецидивисты)	Более 50 лет (бессрочные)
Число арестантов	2124	1567	747	731	175	386

Оцените, пожалуйста, среднее количество лет ссылки для арестантов. В качестве оценок используйте моду, медиану, среднее арифметическое. Рассчитайте характеристику разброса значений (для среднего арифметического).

3. (Построение доверительного интервала) 12 января 2010 г. Всероссийский центр изучения общественного мнения (ВЦИОМ) представил данные о том, представителям каких профессий россияне доверяют больше всего. Для опроса были выбраны четыре профессии (отражающие отношение людей к разным социальным институтам): журналист, учитель, священнослужитель и милиционер. Ниже приводится только фрагмент этих данных. Социологи задавали вопрос: «Насколько Вы доверяете представителям следующих профессий?» (закрытый вопрос, один ответ по каждой позиции). Всего опрошено 1600 человек. Оцените долю граждан, доверяющих представителям МВД (т.е. ответы «1» и «2»). Постройте, пожалуйста, 90%-ый доверительный интервал для генеральной доли.

Профессия	1 – «полностью доверяю»	2 – «скорее доверяю»	3 – «скорее не доверяю»	4 – «совершенно не доверяю»	9 – «затрудняюсь ответить»	Всего
Милиционер	7%	16%	20%	28%	29%	100%

4. (Проверка статистических гипотез) Социологическое выборочное исследование, проведенное менеджером университетского кафе, показало, что из 40 студентов 17 возмущает большое количество майонеза в блюдах. Выборка же 20 преподавателей показала, что только 7 человек крайне не удовлетворены этим фактом. Используя 96%-ую доверительную вероятность, определите, существует ли разница во мнениях студентов и преподавателей или они единодушны в этом вопросе?

5. (Проверка статистических гипотез) 10 ребятам детского сада были даны задания, направленные на изучение 1) наглядно-образного мышления и 2) вербального мышления. Измерялось среднее время решения заданий в секундах. Исследователя интересует

¹ Чехов А.П. Остров Сахалин. – Новосибирск: Сиб. Унив. Изд-во, 2009, с.146

вопрос: существует ли взаимосвязь между временем решения этих двух типов задач, можно ли считать эту связь доказанной? Переменная X – обозначает среднее время решения наглядно-образных, а переменная Y – среднее время решения вербальных заданий:

№	1	2	3	4	5	6	7	8	9	10
X	19	32	33	44	28	35	39	39	44	44
Y	17	7	17	28	27	31	20	17	35	43

6. (*Построение доверительного интервала*) В результате опроса учащихся выпускных классов школ города N . было выяснено, сколько денег ребята тратят ежемесячно на репетиторские занятия/ курсы довузовской подготовки и пр. Результаты такие (см. ниже). Для ответа на вопрос, сколько денег, в среднем, затрачивает выпускник 11-го класса этого города, постройте 95%-ый доверительный интервал.

Расходы, руб.	До 3000	3000-6000	6000-12000	12000-18000
Количество респондентов	18	36	7	4

7. (*Определение объема выборки*) На некотором предприятии работает 2100 человек. В целях командообразования и укрепления здоровья своих сотрудников, дирекция думает над тем, следует ли организовать корпоративное посещение фитнес-клуба. Принятие такого решения сопряжено не только с экономическими издержками, но и с эмоциональной готовностью самих сотрудников принять участие в таком корпоративном проекте. Для принятия решения руководство попросило своих социологов (отдел маркетинга) провести внутренний экспресс-опрос среди коллег – согласятся ли сотрудники ходить в фитнес-клуб регулярно. Выборку какого объема нужно взять, чтобы уровень доверия результатам составлял 85%, а доверительные интервал имел ширину не более $+5\%$. Из пилотажа известно, что примерно каждый третий соглашается (из 10 человек утвердительно ответили только 3).
8. (*Проверка статистических гипотез*) Менеджер цветочного магазина 14 февраля в День Святого Валентина решил выяснить, существует ли разница, сколько, в среднем, на букет тратят молодые мужчины (до 35 лет) и мужчины более старшего поколения (> 35 лет). Случайно организовав выборочное наблюдение за 10 покупателями в возрасте «до 35», он получил, что средняя стоимость букета составила 670 рублей, а выборочная дисперсия оказалась равной 14400 руб.². Выборка же 12 клиентов - старше 35 лет - выявила, что, в среднем, такой мужчина тратит 850 рублей, а выборочная дисперсия 22500 руб.². На уровне доверительной вероятности 95% определите, существует ли разница в «щедрости» молодых мужчин и мужчин более зрелого возраста.

9. (Определение объема выборки) Торговая компания, реализующая развивающие товары для детей – игрушки, книги, товары для детского творчества и пр., решила открыть магазин в новом развивающемся микрорайоне города. Это преимущественно новостройки, где большая часть жителей – это семейные пары, получившие жилье по кредитной программе. Корпорация хотела бы оценить планируемые ежемесячные расходы родителей на детские развивающие товары для своих чад. Для этого в местном парке социологи намереваются провести полноценный опрос семейных пар с маленькими детьми. Выборка какого объема необходима исследователям, если предъявляются следующие требования к точности: доверительная вероятность 0,95; а предельная ошибка выборки – не более +500 рублей? Пробный мини-опрос 7-ми пар (пилотаж) показал следующие ожидаемые суммы затрат (ожидаемая сумма затрат мерялась набором специальных вопросов):

№ пары	1	2	3	4	5	6	7
Затраты, руб.	6000	2500	3000	3000	4000	1000	6500

10. (Проверка статистических гипотез) Для сравнения показателей уровня агрессивности до и после психотренинга некоторой группы «трудных» подростков с помощью специальной методики эти показатели были измерены (в баллах) дважды у 9-ти случайно выбранных подростков. Результаты представлены ниже. Проверьте, имеется ли положительный эффект от тренинга или он оказался безрезультатным.

«Трудный» подросток	1	2	3	4	5	6	7	8	9
Агрессия «до»	31	35	38	34	39	40	27	27	32
Агрессия «после»	33	35	35	30	36	36	38	27	25

11. (Проверка статистических гипотез) Отношение зрителей к включению научно-популярной телепередачи X в вечернюю программу телеканала N выразилось следующими данными. Можно ли считать, что отношение к данной передаче не зависит от пола зрителей? Принять доверительную вероятность на уровне 90%

Отношение к передаче	Положительное	Безразличное	Отрицательное
Мужчины	14	24	2
Женщины	29	36	15

12. (Проверка статистических гипотез) Контрольную работу по курсу «История России» по индивидуальным вариантам выполняли студенты двух факультетов первого курса – социологии и философии. В первой группе (социологам) было предложено 105 заданий, в целом из которых верно решено 60. Во второй группе (философам) из 140 предложенных

вопросов верно решено 69. На уровне доверительной вероятности 98% проверить гипотезу об отсутствии существенных различий в усвоении учебного материала студентами обоих факультетов.

13. (*Проверка статистических гипотез*) Руководитель отдела аналитики проводит аттестацию молодых практикантов, подводя итоги их работы в течение 3-х месяцев в Компании. В качестве критериев он выбрал два: 1) тест на выявление теоретической проф. подготовки молодого специалиста, 2) практические результаты, достигнутые за время производственной практики (оценка получена по отзывам коллег о качестве исполненных обязанностей). Результаты следующие (см. ниже). Определите коэффициент линейной корреляции. Можно ли сказать, что теснота связи статистически доказана?

Практикант	Дмитрий	Иван	Ирина	Павел	Алла	Ксения	Сергей	Михаил
Тест	5	8	7	10	4	7	6	6
Отзывы	7	9	6	8	6	7	7	8

14. (*Проверка статистических гипотез*) Фильм «2012», который вышел на экраны кинотеатров в конце 2009 г., стал одним из самых популярных в прокате. Оживленные дискуссии на тему апокалипсиса подхватил Фонд «Общественное мнение», адресовав жителям РФ вопрос о том, верят ли они в приближение конца света. По данным опроса Фонда составлена следующая таблица сопряженности. Вопрос: «Вы верите или не верите в то, что может наступить конец света?». Можно ли приведенные два признака считать статистически связанными? Категорию ответов «Затрудняюсь ответить» необходимо исключить из анализа

Мнение	Возрастные группы		
	18-35 лет	36-54 лет	55 лет и старше
Верю	24	21	21
Не верю	69	69	54
Затрудняюсь ответить	7	10	15

15. (*Однофакторный дисперсионный анализ*) Социологами проводилось исследование отношения населения трех крупных городов к приезжим трудовым мигрантам. Терпимость к мигрантам измерялась в баллах от 1 до 15 на основе специальной анкеты. Фрагмент результатов представлен в однофакторной таблице. Сделайте заключение о том, имеется ли статистически значимое различие в социальной толерантности населения в трех крупных городах страны.

Город А	Город Б	Город В
4, 5, 12, 10, 1	8, 9, 12, 11, 2, 9	6, 7, 13, 14, 3

16. (*Двухфакторный дисперсионный анализ*) Исследователей интересует, имеется ли различие в степени уверенности в себе детей обоего пола, выросших в благополучных семьях и в семьях, где родители в разводе. Результаты сведены в двухфакторной таблице. В каждой группе обследовано по 4 ребенка. Что можно сказать о влиянии вышеназванных факторов на самооценку детей. Проверьте только одну гипотезу – об отсутствии эффекта взаимодействия двух этих факторов.

Пол детей	Растут в семье, где родители вместе	Растут в семье, где родители в разводе
Мальчики	23, 35, 40, 43	25, 28, 38, 30
Девочки	25, 28, 43, 35	22, 27, 28, 21

17. (*Проверка статистических гипотез*) Приводится динамика численности незанятого населения по месяцам определенной профессии в городе N. Можно ли считать, что безработица одинаково наблюдается в течение всего последнего года, и нет никаких явных улучшений/ухудшений ситуации?

Месяц	Янв.	Фев.	Март	Апр.	Май	Июнь	Июль	Авг.	Сен.	Окт.	Ноя.	Дек.
незанятых	990	1040	1080	1080	1070	1030	990	900	910	1020	1070	1100

18. (*Меры средней тенденции*) В период проведения предвыборной кампании команда кандидата N задала россиянам вопрос о том, как они относятся к этому кандидату. Предлагался набор возможных ответов: 1 – очень хорошо, 2 – хорошо, 3 – средне (плюс/минус), 4 – плохо, 5 – очень плохо. Определите, пожалуйста, чему равна Медиана, характеризующая отношение населения к этому кандидату, в среднем.

Варианты ответов	1	2	3	4	5
Количество давших ответов	40	34	46	20	14

19. (*Проверка статистических гипотез*) Представлены данные опроса выпускников ГУ-ВШЭ (в 2008 г.). Можно ли считать на уровне доверительной вероятности 95%, что доля ребят, планирующих продолжать обучение в магистратуре ВШЭ и работающих в российских компаниях, равна доле тех, кто также планирует пойти учиться в магистратуру, но работает в иностранных корпорациях. В ячейках количество респондентов.

Ответы «Планирую продолжать образование в магистратуре ВШЭ»	Тип организации	
	Российская компания	Иностранная компания
«Нет»	39	11
«Да»	129	41

20. (Проверка статистических гипотез) Средний доход выпускника (4 курса), желающего продолжать обучение в магистратуре ВШЭ, равен 22860 рублей, в то время как выпускник, не желающий продолжать обучение в ГУ-ВШЭ, зарабатывает, в среднем, 28220 рублей. Выборочные дисперсии (на основе данных опроса) соответственно равны 19100^2 руб.² и 26400^2 руб.². Из опрошенных собираются пойти в магистратуру 165 человек, и не собираются 24 человека. Можно ли считать *средние показатели дохода по двум выборкам равными*? Уровень доверительной вероятности принять 1) 95%, 2) 90%.

21. (Проверка статистических гипотез) Было проведено социально-демографическое исследование репродуктивных планов у молодых женщин в возрасте 25-29 лет, занимающих высокие управленческие позиции в российских корпорациях. В опросе принимали участие только женщины, не имеющие детей, но имеющие позитивные установки на рождение детей (т.е. в их планах нет стратегии отказа от детей, а желаемое число детей >0). Задавался вопрос: «О каком количестве детей Вы мечтаете (хотели бы иметь)?» Проверьте гипотезу о том, согласуются ли данные о желаемом числе детей с распределением Пуассона или нет.

Желаемое число детей	1	2	3	4	Более 4
Число респондентов	210	140	110	30	10

22. (Проверка статистических гипотез) Измерен уровень обеспокоенности жителей определенными видами преступности. В каждом вопросе использовалась шкала от -3 до +3. Всего респондентам было предложено 10 утверждений, касающихся их обеспокоенности этими видами преступности в городе, где они живут. В результате агрегированные данные сведены в таблицу:

Уровень обеспокоенности	-30;-20	-20;-10	-10;0	0;10	10-20	20-30
Число респондентов	4	10	26	28	21	12

На уровне значимости 0,05 с помощью критерия Пирсона проверьте гипотезу о том, что уровень обеспокоенности населения, представленный в виде вышеуказанных сгруппированных значений, описывается нормальным законом распределения.

23. (Проверка статистических гипотез) В результате замеров верхнего давления респондентов, посетивших рок фестиваль, были получены следующие данные. Можно ли считать, что прослушивание концерта подобного рода, в среднем, повышает у людей верхнее давление?

№ респондента	Верхнее давление в спокойном состоянии	Верхнее давление при прослушивании концерта тяжелого рока
---------------	--	---

1	120	110
2	110	130
3	100	120
4	130	130
5	110	130

24. (*Построение доверительного интервала*) О работе общественного транспорта (московских электричек) социологи попросили высказать свое мнение 200 человек. 25 человек оказались крайне недовольны работой транспортного узла. Найдите доверительный интервал для оценки в генеральной совокупности доли жителей, возмущенных работой общественного транспорта. Уровень доверительной вероятности принять равным 99%.
25. (*Метод максимального правдоподобия*) Четверо историков гадают о танковых потерях немецкой армии в Курской битве. Результаты гадания: 1000-1200, 1400-1600, 1100-1300, 1200-1400. Как, на ваш взгляд, разумнее всего моделировать эти данные? Какие предположения вы делаете? Отвечайте развернуто, с аргументацией. Какой оценке потерь вы можете доверять?
26. (*Метод максимального правдоподобия*) Аналитический вид функции правдоподобия: $L(x) = 0.9^x \cdot 0.1^{(5-x)^2}$. Посчитайте отношение правдоподобий для оценки МНП (x_{ML}) и $x = 3$. Что означает полученное число? Проверьте гипотезу $H_0 : x = x_{ML}$.
27. (*Метод максимального правдоподобия*) Алла списывает на контрольных у Клары. За последние три контрольных производительность Аллы в этом вопросе была такая: на первой контрольной успевала списать 4 задачи в час, на второй контрольной – 2 задачи в час, и на третьей – 3 задачи в час. Как разумно (каким законом) моделировать оценку производительности Аллы в этом важном деле? Что бы Вы сказали, если бы Алла заявила, что может списывать и 7 задач в час? Оцените вероятность (правдоподобие) такого заявления, опираясь на выбранную Вами модель. Какую интерпретацию имеют значения функции правдоподобия?
28. (*Метод максимального правдоподобия*) Допустим, что после расчета функции правдоподобия мы обнаружили, что она монотонно возрастает. Объясните, что это означает. Нарисуйте пример такой функции. Что, на Ваш взгляд, мы должны сделать, если столкнулись с такой проблемой?
29. (*Метод максимального правдоподобия*) Девушка Катя, прогуливая семинар, накануне написала профессору И. искреннее письмо о том, что она заболела и не сможет прийти в университет, однако она сегодня дважды столкнулась с ним у лифта. При первом столкновении у неё из сумочки вылетело 20 мелких предметов. Во второй раз – 7 вещей.

Зная, что испуганная Катя (опасаясь быть укоренной во лжи), не подбирала вылетевших предметов, быстро убегала, сделайте разумные предположения (сформулируйте модель), оценив число предметов, которые изначально были в сумочке Кати – до того, как она первый раз столкнулась с профессором И.

30. (*Метод максимального правдоподобия*) После расчета функции правдоподобия мы обнаружили, что она бимодальна. Объясните, что это означает? Нарисуйте пример такой функции. Что на ваш взгляд мы должны сделать, если столкнулись с такой проблемой?
31. (*Метод максимального правдоподобия*) Пять маркетологов (специалистов в сфере российского рынка шампуней) спорят о доле рынка компании, выпускающей шампунь «Колокольчик». Споры такие: первый утверждает, что доля рынка лежит в диапазоне 0,07-0,13, второй говорит – 0,05-0,08, третий – 0,07-0,10, четвертый – 0,05-0,10, а пятый считает – 0,06-0,13. Как, на Ваш взгляд, разумнее всего моделировать эти данные? Какие предположения вы делаете? Отвечайте развернуто, с аргументацией. Какой оценке доли рынка можно доверять? Нарисуйте функцию правдоподобия этой оценки.
32. (*Метод максимального правдоподобия*) Гости на дне рождения у Миши играют в фанты. Миша, готовясь к этой вечеринке друзей, составил фанты двух типов – 1) рассказать анекдот, 2) станцевать. Всего в коробке 8 фант, но каково соотношение между ними мы не знаем. Первым пяти гостям досталось три фанты «расскажите анекдот» и две фанты «станцуйте». Согласно методу ММП, что будет наилучшей оценкой общего числа фант про анекдоты (изначально в коробке)? Ответ аргументируйте расчетами вероятности для каждой из потенциально возможных моделей.

Автор программы: _____	/Голстова Ю.Н./
Автор программы: _____	/Макаров А.А./
Автор программы: _____	/Пашкевич А.В./
Автор программы: _____	/Хавенсон Т.Е./
Автор программы: _____	/Назаров Б.В./