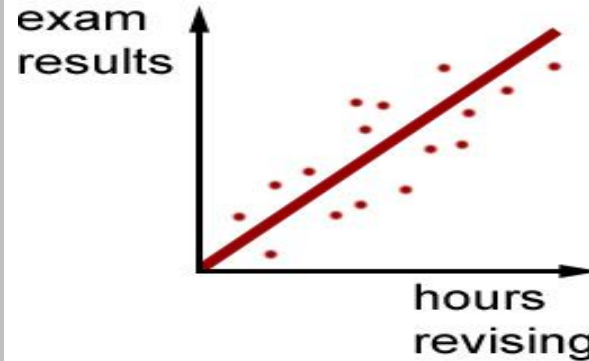
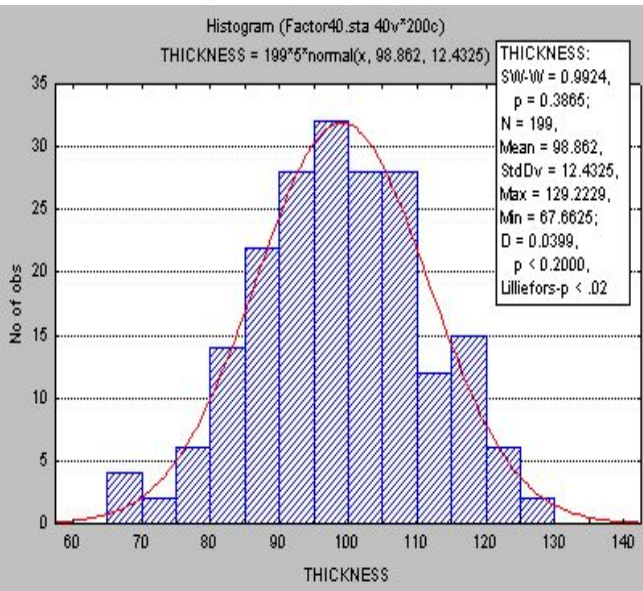


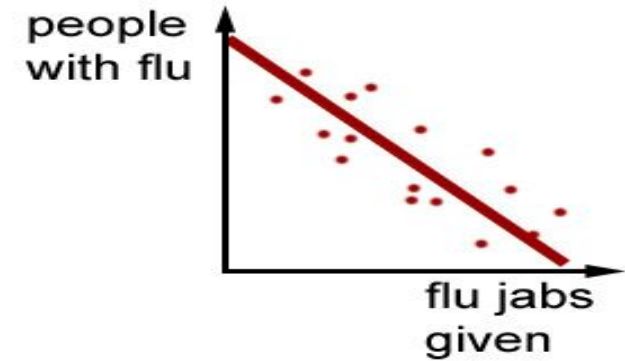
Многомерные статистические

методы



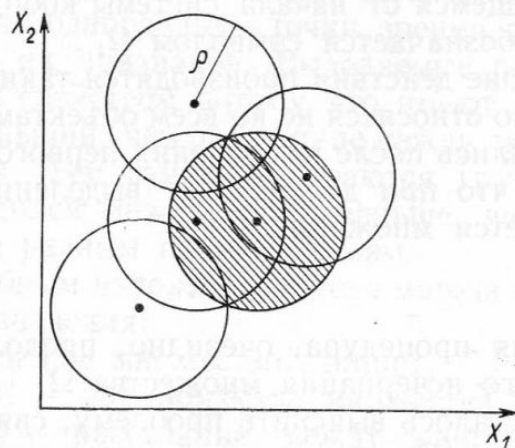
POSITIVE CORRELATION

- people who do more revision get higher exam results.
- revising increases success.



NEGATIVE CORRELATION

- when more jabs are given the number of people with flu falls.
- flu jabs prevent flu.



Архипова Марина Юрьевна
департамент статистики
и анализа данных
Шаболовка 31, каб. 714
marchipova@hse.ru

Форма контроля

Форма текущего контроля -

- Выступление на семинарах с докладом по выбранной тематике;
- Контрольная работа по итогам 1-го модуля;
- Активность на семинарах;
- Подготовка и презентация итоговой работы;
- Написание работы по выбранной тематике на основе реальных статистических данных (исследовательская работа).

Форма контроля

Форма итогового контроля – экзамен.

$$O_{\text{текущий}} = 0,3 \cdot O_{\text{к/р}} + 0,5 \cdot O_{\text{ир}} + 0,2 \cdot O_{\text{доклад.}};$$

$$O_{\text{накопленная}} = 0,8^* O_{\text{текущий}} + 0,2^* O_{\text{ауд}}$$

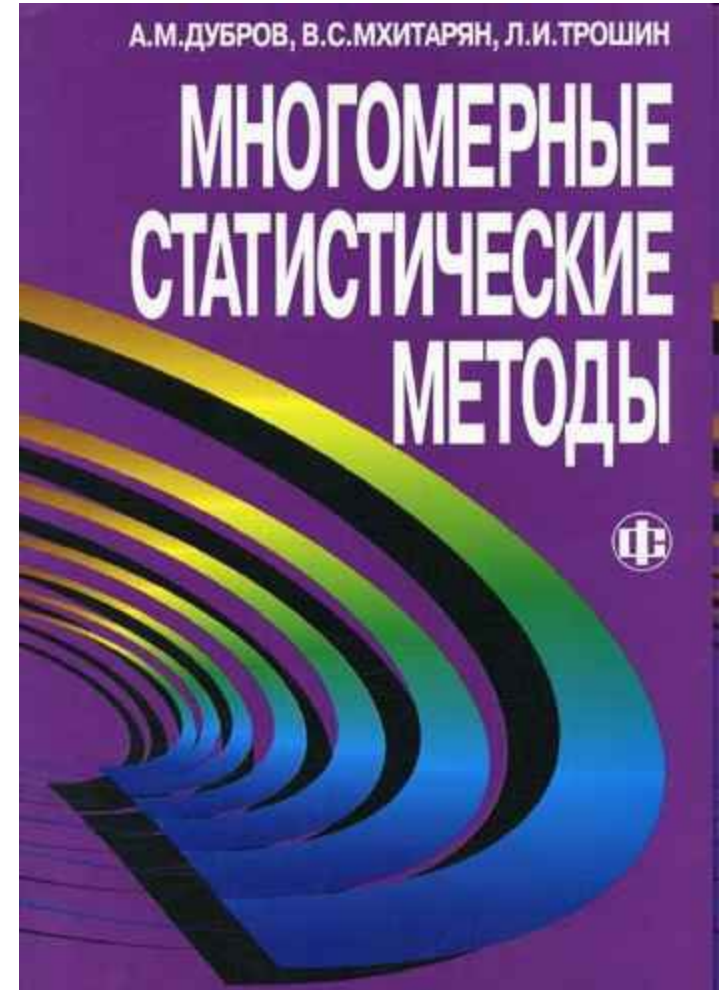
$$O_{\text{результ}} = 0,6^* O_{\text{накопл}} + 0,4^* \cdot O_{\text{экз}}$$

Предостережение от формального использования математических методов в статистике

Ш.В.Шарлье «математическая статистика не представляет собой какого-то автомата, в который достаточно заложить статистический материал, чтобы в результате нескольких манипуляций, как на счетной машине, получить готовый результат».

Рекомендуемая литература

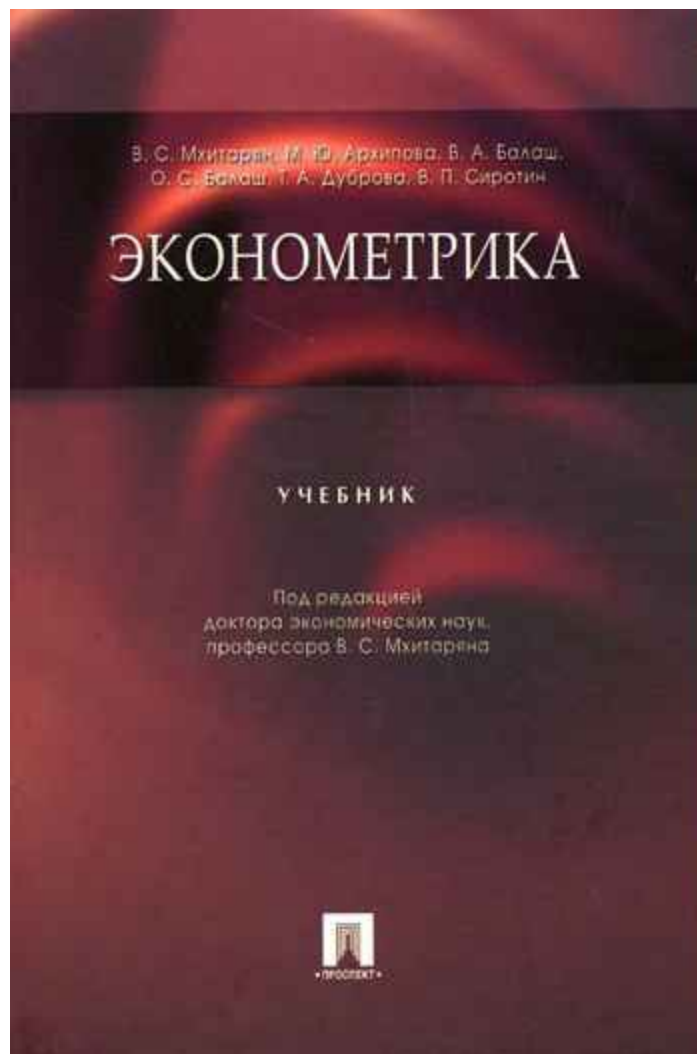
Дубров А.М., Мхитарян В.С.,
Трошин Л.И.
***Многомерные
статистические методы.***
– М.: Финансы и статистика,
2014.



Рекомендуемая литература

Мхитарян В.С., Архипова М.
Ю., Сиротин В.П. и др.

Эконометрика: учеб. / под
ред. В.С. Мхитаряна. – М.:
Проспект, 2014.



Рекомендуемая литература

Айвазян С.А., Мхитарян В.С. *Прикладная статистика. Основы эконометрики*

В 2 т. 2-е изд. – М :

ЮНИТИ-ДАНА, 2011.



Список литературы

1. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. Из-е 2-е. – М.: ЮНИТИ, 2009.
2. Ниворожкина Л.И., Арженовский С.В. **Многомерные статистические методы в экономике**. Учебник. «Дашков и К; Ростов н/Д: Наука-Спектр, 2007.
3. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. И.И. Елисеевой. — 6-е издание, переработанное и дополненное. — Москва: Финансы и Статистика, 2012.

Список литературы

1. *Айвазян С.А., Мхитарян В.С.* Теория вероятностей и прикладная статистика. Из-е 2-е. – М.: ЮНИТИ, 2009.
2. *Ниворожкина Л.И., Арженовский С.В.* **Многомерные статистические методы в экономике.** Учебник. «Дашков и К; Ростов н/Д: Наука-Спектр, 2007.
3. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики: Учебник / Под ред. И.И. Елисеевой. — 5-е издание, переработанное и дополненное. — Москва: Финансы и Статистика, 2006.
4. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: исследование зависимостей. – М.: Финансы и статистика, 1985.
5. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: основы моделирования и первичная обработка данных – М.: Финансы и статистика, 1983.
6. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: исследование зависимостей. – М.: Финансы и статистика, 1985.

Рекомендуемая литература



Журнал «*Вопросы статистики*» -
научно-информационное издание
(ФСГС)

<http://www.infostat.ru/ru/catalog.html?page=info&id=113>



Журнал «*Вопросы экономики*»
-теоретический и научно-практический
журнал

общеекономического содержания
Главный редактор: Л. И. АБАЛКИН

<http://www.vopreco.ru/>

Рекомендуемая литература:

- Журнал «Прикладная эконометрика»

Главный редактор АЙВАЗЯН Сергей Артемьевич

<http://www.marketds.ru/?sect=journal&id=econometrics&item=board>



Рекомендуемая литература:

Журнал «Форсайт»

«Форсайт» – научный журнал, выпускаемый Институтом статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики».

<http://ecsocman.hse.ru/mags/foresight/archive.html>

Основные темы:

- Результаты Форсайт-исследований, выполненных в России и за рубежом;
- Долгосрочные приоритеты социального, экономического и научно-технологического развития;
- Тенденции и индикаторы развития науки, технологий и инноваций;
- Научно-техническая и инновационная политика;
- Стратегические программы инновационного развития на национальном, региональном, отраслевом и корпоративном уровнях

.....



Рекомендуемая литература:

«Квантиль» - международный электронный научный
эконометрический журнал, распространяемый бесплатно в сети.

квантиль

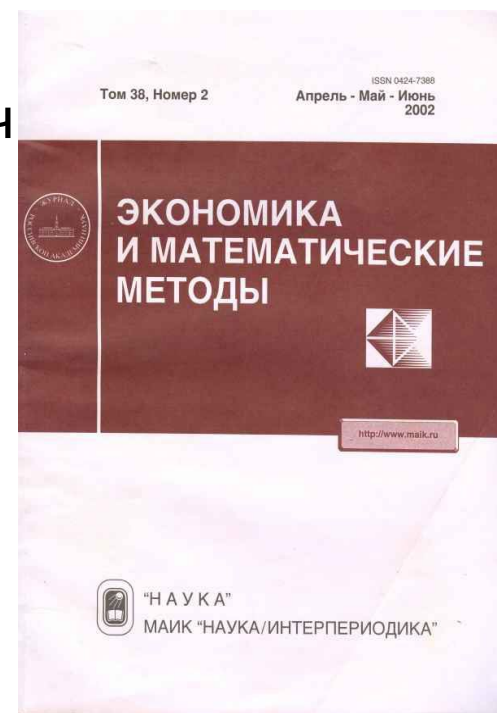
Гл. редактор: профессор РЭШ С.А. Анатольев.

<http://quantile.ru/>

- Журнал **«Экономика и математические методы»**

научный журнал ЦЭМИ РАН,
с 1965, гл. редактор Макаров Валерий Леонидович

<http://www.cemi.rssi.ru/emm/home.htm>



Рекомендуемая литература:



JSTOR полнотекстовая база данных
англоязычных научных журналов.

<http://www.jstor.org/>

Участие в студенческой конференции

VII Международная (Межвузовская) студенческая научно-практическая конференция **«Статистические методы анализа экономики и общества»**

Требования к тезисам докладов:

- Объем – 2 полностью заполненные страницы
- Формат статей: **параметры страницы:** А4,
- **отступы:** сверху: 6.1, снизу: 7.1 (Высота строго 16,5)
- слева: 4.8, справа: 4.7 (Ширина строго 11,5)
- **Шрифт** Times New Roman Cyr 10,
- межстрочное расстояние - одинарное.

Три базовые задачи прикладного МСА

1. Статистическое исследование структуры и характера взаимосвязей, существующих между анализируемыми признаками.
2. Разработка статистических методов классификации объектов и признаков.
3. Снижение размерности признакового пространства с целью лаконичного объяснения природы анализируемых многомерных данных.

Проблемы прикладной статистики и эконометрики:

1. Статистическое исследование структуры и характера взаимосвязей, существующих между анализируемыми количественными переменными.

«Существует ли зависимость между доходом семьи и ее расходами на питание?»

«Связан ли уровень безработицы в стране с ВВП?»

«Как влияет потребление алкоголя на заработную плату мужчины?»

Проблемы прикладной статистики и эконометрики:

2. Разработка статистических методов классификации объектов и признаков.



3. Снижение размерности исследуемого признакового пространства

Влияние удовлетворенности работой на заработную плату

Вопросы анкеты:

Насколько Вы удовлетворены или не удовлетворены Вашей работой в целом?

Насколько Вы удовлетворены или не удовлетворены условиями Вашего труда?

Насколько Вы удовлетворены или не удовлетворены оплатой Вашего труда?

Насколько Вы удовлетворены или не удовлетворены возможностями Вашего профессионального роста?

Ответы:

1 Полностью удовлетворены

2 Скорее удовлетворены

3 И да, и нет

4 Не очень удовлетворены

5 Совсем не удовлетворены

fac1_1 – интегральный показатель «удовлетворенность работы» (область



Переменные (суждения)		3 компоненты (фактора)			Извлеченные общности	
Накопленный % объясненной дисперсии		24%	36%	46%		
1.	Цена	0,5	0,06	0,23		0,31
2.	Пунктуальность	0,07	0,02	0,63		0,40
3.	Скорость	0,05	-0,11	0,51		0,27
4.	Удобство расписания	-0,01	0,07	0,68		0,46
5.	Ожидание в очереди	0,29	0,5	0,1		0,35
6.	Информированность	0,08	0,12	0,61		0,39
7.	Безопасность в электричке	0,71	0,08	0		0,51
8.	Состояние салона	0,62	0,32	-0,09		0,49
9.	Подозрительные лица	0,53	0,34	0,06		0,40
10.	Давка	-0,06	0,82	0,02		0,68
11.	Удобство посадки	0,13	0,81	0		0,67
12.	Общий эмоциональный фон	0,38	0,54	-0,01		0,44
13.	Безопасность на станции	0,73	-0,01	0,09		0,55
% дисперсии фактора, объясненной интерпретационными переменными		87%	94%	97%		

Три выделенных фактора объясняют в совокупности **половину информации**, что может считаться удовлетворительным результатом. Отдельные переменные (например, «скорость», «цена») характеризуются относительно небольшим процентом объясненной моделью дисперсии (столбец «Извлеченные общности»).

1

Переменные для интерпретации: 7, 8, 9 и 13. Этот фактор отражает восприятие личной безопасности пассажира



2

Переменные для интерпретации: 5,10,11,12. Определяют отношение пассажира к количеству пользователей транспорта



3

Переменные для интерпретации: 2,3,4,6. Отражают **совокупные временные затраты** на всех этапах поездки.



Компьютерные семинары

Все задания выполняются в SPSS, Statistica, Stata!

SPSS - **Statistical Package for the Social Science (1967)** – одна из самых распространённых программ для обработки статистической информации и выполнения всех этапов статистического анализа: от просмотра данных, создания таблиц, вычисления дескриптивных статистик до применения сложных статистических методов.



Многомерные статистические методы (МСМ) –

это дисциплина, объединяющая совокупность теоретических результатов, методов и приёмов, позволяющих на базе

- экономической теории,**
- общей теории статистики**
- ТВиМС**
- математико-статистического инструментария**
получать количественное выражение качественным закономерностям.

Экономическая теория используется на этапе содержательной постановки задачи и интерпретации полученных результатов.

Примеры

1. Падение кирпича на голову – случайное событие.

Однако движение человека в этом месте и температура воздуха, падение камня – законом. Совмещение этих событий – случайность.

2. Травмы на улицах Москвы. Каждый конкретный случай – это случайность. Но когда мы начинаем их исчислять, то при большом n можем увидеть в какие часы они происходят чаще, с кем чаще

Примеры

1. Падение кирпича на голову – случайное событие.

Однако движение человека в этом месте и температура воздуха, падение камня – законом. Совмещение этих событий – случайность.

2. Травмы на улицах Москвы. Каждый конкретный случай – это случайность. Но когда мы начинаем их исчислять, то при большом n можем увидеть в какие часы они происходят чаще, с кем чаще

3. Выброс мебели из окна дома - НЗР. Холмообразность связана с законом всемирного тяготения, с точкой сбрасывания, с твердостью предметов. Но не связана с индивидуальными особенностями предметов. Это статистическое свойство.

4. Роман «Евгений Онегин» – уникальное явление. А пушкинские рифмы можно исследовать, изучать отношение людей к ним (по возрасту, соц. статусу...)

Экономическая теория используется на этапе

- содержательной постановки задачи и
- интерпретации полученных результатов.

Статистические методы используются для разработки показателей, а также на этапе сбора информации.

ТВМС используются для проверки статистических гипотез о значимости построенных моделей и их коэффициентов, построения интервальных оценок

Методы математико-статистического инструментария используются для

- сжатия информации до размеров, допускающих принятие решения,
- для построения эконометрической модели, позволяющей получать различные сценарии развития экономических явлений.

Шкалы измерений

Каждое измерение над объектом производится в определенной шкале.

- социальная принадлежность семьи,
- пол и профессия главы семьи,
- качество жилищных условий,
- число членов семьи, количество детей,
- среднегодовой доход и т. п..

Шкалы измерений

Среди всевозможных шкал (число которых теоретически бесконечно) наиболее широко применяются шкалы, составляющие определенную иерархическую систему, в которой каждая последующая шкала включает в себя предыдущую как частный случай:

- номинальная,
- порядковая,
- количественная
 - интервальная,
 - относительная.

Первые две шкалы являются неколичественными.

Каждому типу шкал соответствуют специфические методы исследования.

Шкалы измерений

Номинальная шкала (шкала наименований) представляет собой простое перечисление различающихся между собой явлений или объектов.

Объекты могут быть пронумерованы, причем цифры, присваиваемые различным градациям, служат лишь для отличий их друг от друга.

Шкалы измерений

Номинальная шкала (шкала наименований) представляет собой простое перечисление различающихся между собой явлений или объектов.

Объекты могут быть пронумерованы, причем цифры, присваиваемые различным градациям, служат лишь

Семейное положение:

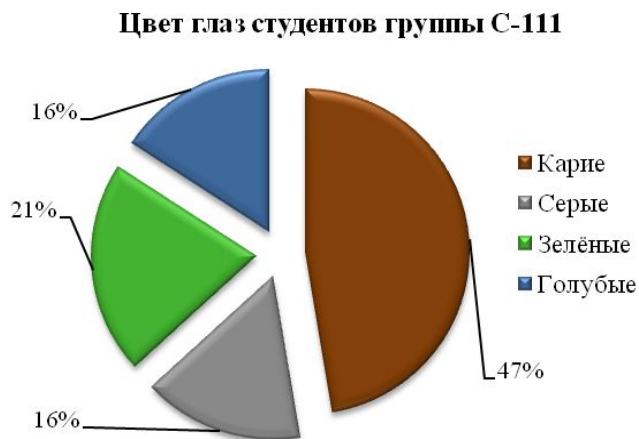
- состоит в зарегистрированном браке,
- состоит в гражданском браке,
- никогда не состоял в браке,
- разведен,
- вдовец и т.д.

Номинальные признаки:

раса, национальность, цвет глаз, волос, пол и т.д.

Примеры номинальных данных

- группа крови человека;
- банк России (Сбербанк, ВТБ, Газпромбанк и т.д.);
- Цвет глаз студентов некой группы



Базовой операцией с данными в номинальной шкале является установление тождества или различия между объектами в смысле наличия определенного свойства, отличающего их от остальных объектов.

Шкалы измерений

Если категории признака можно упорядочить друг относительно друга, то в этом случае используется **порядковая шкала**.

Шкалы измерений

Если категории признака можно упорядочить друг относительно друга, то в этом случае используется **порядковая шкала**.

Места, занимаемые величинами в шкале порядка, называются рангами, а сама шкала – ранговой (неметрической).

Основные типы исходных данных

Примеры:

- образование (начальное, среднее, ...);
- место на спортивных соревнованиях;
- возраст респондента:
 - ✓ «меньше 10 лет»,
 - ✓ «11– 20»,
 - ✓ «21–30»
 - ✓ «старше 30»,

Шкалы измерений

Шкалы количественных признаков.

Количественным является признак, значение которого имеют числовое выражение и отражают размеры, масштабы некоторого объекта или явления.

Примеры:

Шкалы измерений

Шкалы количественных признаков.

Количественным является признак, значение которого имеют числовое выражение и отражают размеры, масштабы некоторого объекта или явления.

Примеры:

- доход домохозяйства,
- площадь жилого помещения,
- цена товара,
- стаж работы.

Количественные признаки могут быть

- интервальными
- относительными.

Количественные признаки могут быть

- интервальными
- относительными.

Интервальные измеряют величину различия, т.е.
на сколько один объект превосходит другой объект.

Интервальная шкала может иметь произвольные
начало отсчета и масштаб.

Количественные признаки могут быть

- интервальными
- относительными.

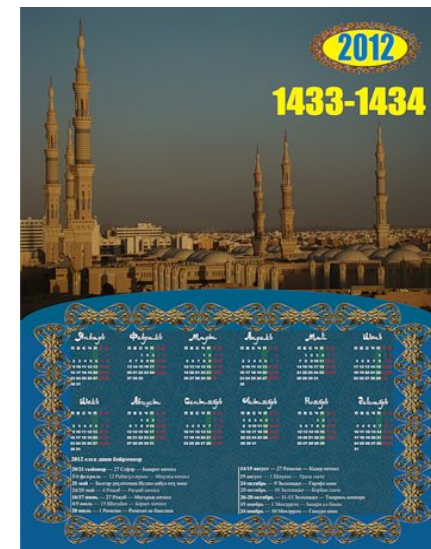
Интервальные измеряют величину различия, т.е. на сколько один объект превосходит другой объект.

Интервальная шкала может иметь произвольные начало отсчета и масштаб.

Примеры: **календари, шкалы температур.**

По григорианскому/юлианскому календарю, сегодня **2015** год, а вот что говорят о порядковом номере этого года другие календари:

- 7524 — по Византийскому (Православной церкви) календарю;
- 4647 — по китайскому календарю;
- 1732 — по эре Диоклетиана;
- 2329 — по эллинской эре Селевкидов;
- 1937 — по индийскому календарю Сака;
- 2675 — по японскому календарю;
- 2765 — по эре Набунасара;
- 2769 — от основания Рима;
- 1437-1438 — по исламскому.



Шкала Кельвина В термодинамике используется шкала Кельвина, в которой температура отсчитывается от абсолютного нуля (состояние, соответствующее минимальной теоретически возможной внутренней энергии тела), а один кельвин равен $1/273,16$ расстояния от абсолютного нуля до тройной точки воды (состояния, при котором лёд, вода и водяной пар находятся в равновесии).


Шкала Цельсия В быту используется шкала Цельсия, в которой за 0 принимают точку замерзания воды, а за 100° точку кипения воды при атмосферном давлении.

Шкала Фаренгейта В Англии и, в США используется шкала Фаренгейта. В этой шкале на 100 градусов разделён интервал от температуры самой холодной зимы в городе, где жил Фаренгейт, до температуры человеческого тела. Ноль градусов Цельсия - это 32 градуса Фаренгейта, а градус Фаренгейта равен $5/9$ градуса Цельсия.

Шкала Реомюра Предложена в 1730 году Р. А. Реомюром, который описал изобретённый им спиртовой термометр.

Количественные интервальные признаки

дают ответ на вопрос «**на** сколько больше»?

 если температура воздуха повысилась с 4 до 8 градусов по Цельсию, то нельзя сказать, что стало **в** два раза теплее.

Множество допустимых преобразований шкалы интервалов составляют ***все линейные преобразования.***

Основным свойством шкалы интервалов является ***сохранение отношения длин интервалов.***

Относительная шкала (Шкала отношений)

сохраняет отношения свойств объектов.

Показывает **во** сколько раз свойства одного объекта превосходит свойства другого объекта.

Всегда существует точка с отсутствием измеряемого свойства объекта (количество наличных денег, шкала измерения веса, длины и т.д.).

Шкала отношений не накладывает никаких ограничений на математический аппарат, используемый для обработки результатов измерений.

Шкалы количественных признаков

Чем выше тип шкалы, тем больше действий можно производить над переменными.

Выбор данных для анализа

Статистические издания

Библиотека официальных публикаций **Федеральной службы государственной статистики (ФСГС)**:

<http://www.gks.ru/wps/wcm/connect/rosstat/rosstatsite/main/publishing/catalog/statisticCollections>

- Россия в цифрах
- Российский статистический ежегодник
- Промышленность России
- Регионы России. Социально-экономические показатели
- Россия и страны мира
- Россия и страны – члены Европейского союза

Статистические издания

- Статистические данные **Организации экономического сотрудничества и развития** (сокр. ОЭСР, Organization for Economic Co-operation and Development, OECD):
<http://www.oecd.org/>
- ESS - Европейское Социальное Исследование:
<http://www.ess-ru.ru/>

Статистические издания

Статистические сборники ГУ-ВШЭ (совместно с Росстатом)

<http://www.stat.hse.ru/>

- Индикаторы инновационной деятельности
- Наука. Инновации. Информационное общество
- Наука в Российской Федерации
- Индикаторы науки
- Наука Москвы

Источники данных

Единый архив экономических и социологических данных
Высшей школы экономики

<http://econhist.ru/>

ЕДИНЫЙ АРХИВ ЭКОНОМИЧЕСКИХ
И СОЦИОЛОГИЧЕСКИХ ДАННЫХ

ОБ АРХИВЕ БЮЛЛЕТЕНЬ «СОФИСТ» НАШИ ДЕПОЗИТОРЫ НАШИ ПОЛЬЗОВАТЕЛИ ПО ДАННЫМ ОПРОСОВ... ПОЛЕЗНЫЕ ССЫЛКИ ЗАЯВКИ КОНТАКТЫ

ДОСТУП К ДАННЫМ

РЕЗУЛЬТАТЫ ОПРОСОВ

СТАТИСТИЧЕСКИЕ РЯДЫ

ЭЛЕКТРОННЫЕ ТАБЛИЦЫ

ДОСТУП К ВНЕШНИМ РЕСУРСАМ

Доступ к данным

Единый архив предоставляет доступ к своим коллекциям на безвозмездной основе. Коллекции Е.

Результаты социологических исследований	Более 950 опросов, проведенных ведущими социологическими центрами по выборкам.
Статистические ряды	Более 1000 основных статистических показателей экономики России по мере публикации новых значений каждого показателя, включая значения, производимых источником данных.

Источники данных

Рейтинг инновационного развития субъектов РФ

<http://www.hse.ru/primarydata/>

▶ СТАТИСТИЧЕСКИЕ СБОРНИКИ

- ▼ Индикаторы науки
 - ▼ Индикаторы образования
 - ▼ Индикаторы инновационной деятельности
 - ▼ Индикаторы информационного общества
 - ▼ Информационное общество: тенденции развития
 - ▼ Информационное общество: тенденции развития в субъектах РФ
 - ▼ Наука. Инновации. Информационное общество
 - ▼ Образование в цифрах
 - ▼ Образование в Российской Федерации
- ### ▶ АНАЛИТИЧЕСКИЕ ДОКЛАДЫ
- ▼ Рейтинг инновационного развития субъектов РФ
 - ▼ Информационное общество: востребованность информационно-коммуникационных технологий



Источники данных

<http://www.globalinnovationindex.org/content.aspx?page=data-analysis> – Глобальный инновационный индекс (бизнес,

С The Global Innovation Index



QUICK LINKS

- ▶ HOME
- ▶ **GII 2015 REPORT**
- ▶ GII ADVISORY BOARD
- ▶ GII FRAMEWORK
- ▶ DATA ANALYSIS
- ▶ MEDIA LIBRARY
- ▶ PRESS RELEASES
- ▶ GII BLOG
- ▶ PAST EVENTS
- ▶ PAST REPORTS

55

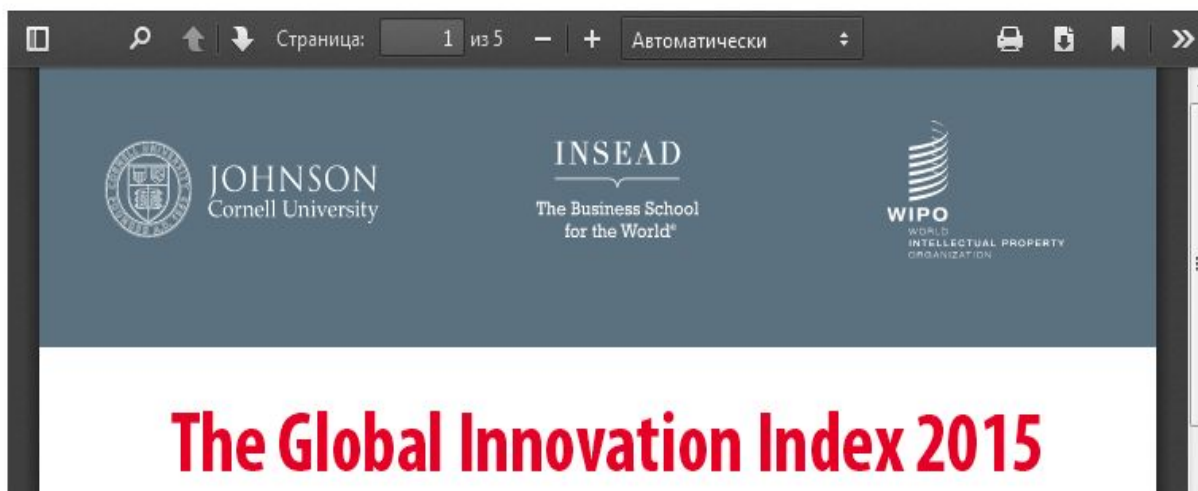
151

Tweet

Share

2

G+ Поделиться



Статистические сборники

- Выборочное обследование бюджетов домашних хозяйств (ФСГС): <http://www.micro-data.ru>
- **РМЭЗ** - Российский мониторинг экономического положения и здоровья населения (Russian Longitudinal Monitoring Survey)
<http://hse.ru/science/rlms> (описание)
- **ESS** - Европейское Социальное Исследование - исследование изменения установок, взглядов, ценностей и поведения населения Европы
<http://www.ess-ru.ru/>
- ...

Источники данных

- Официальный сайт Банка России [www.cbr.ru];
- Росбизнесконсалтинг [www.rbc.ru];
- Рейтинговое агентство ЭкспертРа [www.raexpert.ru];
- Портал статистических данных Статистика.RU [www.stat.gov.ru];

Источники данных

Science Direct	Web of Knowledge/Web of Science
EBSCO	Scopus
ProQuest	Oxford Journals Online
EastView	Cambridge Journals Online
Project Muse	DigiZeitschriften
Emerald Management Xtra	APA PsycARTICLES
Springer Link	Статистические ресурсы МВФ
InfoTrac	БД Factiva
Taylor & Francis	Ресурсы OECD
SAGE Journals Online	
Elibrary.ru	

<http://www.sciencedirect.com/>

<http://www.ebscohost.com/public/>

<http://www.proquest.co.uk/en-UK/>

www.eastview.com/

<http://www.emeraldinsight.com/>

Формы представления исходных статистических данных

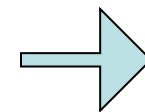
Матрица наблюдений «объект – свойство»

а) Пространственная выборка (статистическая таблица)

(cross-section data) - является наиболее общим типом представления исходных данных. Соответствует ситуации, когда исходные данные регистрируются *только* «в пространстве», но не во времени (**n-число объектов** ($i=1,2,\dots,n$), **к - число переменных** ($j=1,2,\dots,k$)).

Допустимо предположение о взаимной независимости наблюдений.

матрица типа «объект – свойство»



$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(k)} \\ \dots & \dots & \dots & \dots \\ x_i^{(1)} & x_i^{(2)} & \dots & x_i^{(k)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(k)} \end{pmatrix}$$

Пример 1 (трехмерные данные)

Характеристика стран мира на 1 июля 2009 г. по показателям:

- *площадь территории,*
- *численность населения*
- *плотность населения на 1 км²*

Страна	Территория, км ²	Население, тыс. человек	Человек на 1 км ²
Монако	2	33	16350
Сингапур	705	4615	6545
Бермудские острова	54	65	1190
Сан-Марино	61	31	512
Китай	9 596 961	1337411	139
США	9 629 091	311666	32
Россия	17 098 240	141394	8
Канада	9 984 670	33259	3
Монголия	1564100	2641	2
Гренландия	2166086	57	0,03

Пример 2

В работе [Айвазян С.А.] приведены значения восьми показателей для 266 крупных американских компаний ($n = 266$).

Матрица X , представляющая эти данные, имеет размерность 266×8 .

В качестве анализируемых показателей рассмотрены:

- $x(1)$ – капитальные вложения за год (млн. долл.);
- $x(2)$ – годовой фонд оплаты труда (млн. долл.);
- $x(3)$ – расходы на нематериальные активы за год (млн. долл.);
- $x(4)$ – годовые расходы на рекламу и маркетинг (млн. долл.);
- $x(5)$ – годовые расходы на исследования и развитие (млн. долл.);
- $x(6)$ – численность работников (тыс. чел.);
- $y(1)$ – объем продаж за год (млн. долл.);
- $y(2)$ – себестоимость проданного товара (млн. долл.).

б) временная выборка

Если зафиксировать

- номер переменной j и
- номер статистически обследуемого объекта i ,
то расположенную в хронологическом порядке
последовательность значений называют *одномерным
временным рядом*.

В этом случае $x_t^{(j)}$ - значение
 j -го показателя в
 t - момент времени.

Пример Индексы цен на
первичном рынке жилья
(на конец года, в процентах
к концу предыдущего года)



**Основные этапы
прикладного
статистического анализа**

Компьютерная работа

Часть 1

**Корреляционно-
регрессионный анализ**

Основные этапы прикладного статистического анализа

1. Постановка задачи исследования

Определение цели, объекта и предмета исследования.

Главным предметом изучения в любом исследовании является та или иная совокупность объектов.

(Руководители малых предприятий, избиратели Москвы, субъекты РФ, студенты старших курсов и др.)

Основные этапы прикладного статистического анализа

1. Постановка задачи исследования

Пример 1 **Статистическое исследование и моделирование инновационного поведения предприятия**

Основные этапы прикладного статистического анализа

1. Постановка задачи исследования

Пример 1 Статистическое исследование и моделирование инновационного поведения предприятия

Объектом исследования являются промышленные предприятия.

Предметом исследования служат формы и результаты инновационной деятельности промышленных предприятий.

Задачи исследования.

.....

Основные этапы прикладного статистического анализа

1. Постановка задачи исследования

Пример 2

**Статистическое исследование влияния потребления
алкогольных напитков на занятость и заработную плату в
России**

Основные этапы прикладного статистического анализа

1. Постановка задачи исследования

Пример 2

Статистическое исследование влияния потребления алкогольных напитков на занятость и заработную плату в России

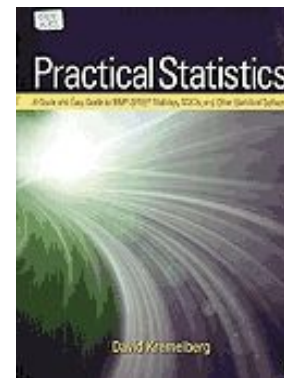
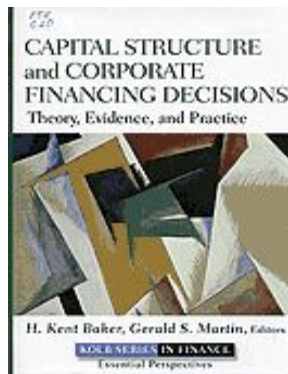
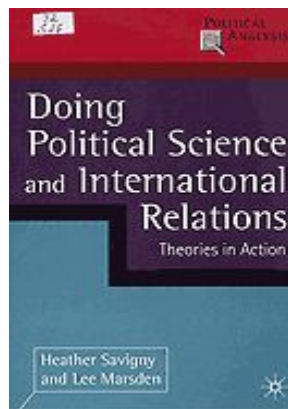
- **Объектом исследования** - экономически активное население России.
- **Предметом исследования** – совокупность показателей, характеризующих занятость и заработную плату в зависимости от социальной патологии населения.

Структура компьютерной работы

2. Обзор литературы по тематике исследования

(рассмотреть 2-3 источника). Что сделано по данной теме до Вас, насколько далеко продвинулась теория в России (странах мира), какие методы используются, система показателей и т.д.

Англоязычные статьи можно брать, например, <http://library.hse.ru/info/JSTORinfo.htm>



Основные этапы прикладного статистического анализа

3. Выбор признаков (показателей) для проведения исследования.

В зависимости от цели исследования выбирают **K-показателей**.

Например, возраст, затраты на проведение ИиР, число предприятий и т.д.

N - число объектов в генеральной совокупности;

n - число объектов в выборке.

Измерение и описание показателей. Каждое измерение производится с определенной точностью (например, возраст фиксируется с точностью до одного года ...).

Основные этапы прикладного статистического анализа

Формулировка рабочих гипотез исследования.

- можно придумать самим,
- можно найти в литературе.

Проверка статистических гипотез

- «Чем больше времени студент тратит на подготовку к экзамену, тем выше финальная оценка».
- «Доллар - это товар, поэтому должен действовать закон спроса. Чем выше цена, тем меньше спрос».
- «Объём покупок в долларах убывает при увеличении курса продажи» (в общем и среднем при прочих равных условиях, *ceteris paribus* - Кейнс).
-

Основные этапы прикладного статистического анализа

4. Первичная статистическая обработка данных.

- Отображение признаков в номинальной, порядковой или количественной шкале;
- Восстановление пропущенных наблюдений;
- Проверка однородности выборки;

Основные этапы прикладного статистического анализа

4. Первичная статистическая обработка данных.

- Отображение признаков в номинальной, порядковой или количественной шкале;
- Восстановление пропущенных наблюдений;
- Проверка однородности выборки;
- Анализ выбросов (ранжировки, использование статистических критериев, напр., критерия Титъена-Мура),
- Экспериментальный анализ закона распределения исследуемой ГС (вычисление основных числовых характеристик, численный и графический анализ одномерных ЗР) и параметризация сведений о природе изучаемых распределений (процесс сводки и группировки)..

Основные этапы многомерного статистического анализа

- 5 . Определение формы модели, уточнение методов анализа.**
- 6. Оценка параметров модели.**

Основные этапы многомерного статистического анализа

7. Диагностика модели

Полученные количественные оценки - удовлетворяют ли они разумным требованиям?

- Если модель проходит диагностику

→ модель адекватна, нормально передаёт реальность, не противоречит имеющимся данным.

Это значит, что с помощью модели получили DGP (Data Generated Process).

Основные этапы многомерного статистического анализа

Если модель не проходит диагностику

(например, в регрессии $\beta < 0$, то показатель впервые в истории станет товаром Гиффена)? Нет? Значит, что-то недоучли.

Что делать дальше?

Мы можем

- учесть новые переменные, получить дополнительные данные;
- сменить МНК на другой метод оценки параметров;
- изменить спецификацию модели;
- проверить правильность выдвинутой гипотезы.

□ **Товар Гиффена** — товар, потребление которого увеличивается при повышении цены. Обычно это связано с тем, что эффект замещения от изменения цены перевешивается действием эффекта дохода.

Основные этапы многомерного статистического анализа

8. Интерпретация результатов.

Прогноз по модели.

Ошибки в статистических данных

1. Ошибки при сборе статистического материала возникают:

- из-за нечеткости формулировки задач;
- неточности в измерениях и классификациях;
- ошибочного выбора объектов;
- неточности в определениях;
- сознательного искажения данных при опросах;
- пропусков в данных, дублирования информации и др.

2. Ошибки при обработке и представлении статистического материала

- Нечеткая группировка данных;
- Неправильное толкование корреляции;
- Игнорирование рассеяния;
- Неточное графическое представление (вводящее в заблуждение) и др.

Результирующая переменная	Объясняющие переменные	Раздел многомерного статистического анализа
Количественная	Количественные	Регрессионный анализ и СОУ
Количественная	Единственная количественная (t)	Анализ временных рядов
Количественная	Ординарные или номинальные	Дисперсионный анализ
Количественная	Смешанные (количественные и неколичественные)	Ковариационный анализ, модели типологической регрессии
Ординарные	Ординарные и номинальные	Анализ ранговых корреляций и таблиц сопряженности
Номинальные	Количественные	Дискримин. анализ, логит- и пробит-модели, кластер-анализ, расщепление смесей
Смешанные	Смешанные	Аппарат логических решающих функций, Data Mining

Спасибо за внимание!

Основные формы представления исходных данных

Частоты и таблицы сопряженности

(для качественных и категоризованных переменных)

Ранжировки (рейтинги, упорядочения) – используются в случае экспертных оценок

Форма ответа эксперта:

упорядочение оцениваемых объектов в порядке убывания их качества по анализируемому свойству.

Таким образом, если i -й объект поставлен в этом ряду на 1-е место, это означает, что он признан экспертом лучшим по анализируемому свойству в ряду из n - оцениваемых объектов

и ему приписывается 1-е место (или ранг R_i , равный единице).

Например

- уровень квалификации сотрудников,
- моральный климат в фирме
- репутация фирмы.

Эксперно-бальные оценки выходного качества

(наиболее информативный способ)

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix}$$

где y_{ij} оценка выходного качества объекта, полученная от j -го эксперта.

n – число оцениваемых объектов

m – число экспертов

Экспертные упорядочения

обследованных объектов по степени проявления в них анализируемого свойства, то есть ранжировки вида

$$R = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{pmatrix}$$

где R_{ij} - ранг (место), присвоенный объекту O_i j -м экспертом в ряду из n исследуемых объектов, упорядоченном этим экспертом по степени проявления анализируемого свойства.

(средний по информативности способ)


Построение булевой матрицы парных сравнений

(наименее информативности способ)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{im} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

где a_{ij} равно **1**, если i -й объект «не хуже» j -го,
 a_{ij} равно **0** в противном случае.

В аналитической работе при обращении к многомерному пространству признаков (объектов) принимают во внимание следующие особенности:

- Использование большого числа признаков, разнородных по своей природе  необходимость приведения данных к сопоставимому виду (например, нормирование данных)
- При работе с *к-мерными* данными совместно используются математические, абстрактные методы и методы экономико-статистического анализа, ориентированные на конкретные сферы приложения. Необходимо обращать внимание на *непротиворечивость результатов, полученных различными методами.*

Возникающие противоречия указывают на нарушение логики решения экономической задачи и часто приводят к ошибочным выводам!

Результирующая переменная	Объясняющие переменные	Раздел многомерного статистического анализа
Количественная	Количественные	Регрессионный анализ и СОУ
Количественная	Единственная количественная (t)	Анализ временных рядов
Количественная	Ординарные или номинальные	Дисперсионный анализ
Количественная	Смешанные (количественные и неколичественные)	Ковариационный анализ, модели типологической регрессии
Ординарные	Ординарные и номинальные	Анализ ранговых корреляций и таблиц сопряженности
Номинальные	Количественные	Дискримин. анализ, логит- и пробит-модели, кластер-анализ, расщепление смесей
Смешанные	Смешанные	Аппарат логических решающих функций, Data Mining

Ковариация

Важной характеристикой совместного распределения двух случайных величин является **ковариация** (или **корреляционный момент**). Ковариация определяется как математическое ожидание произведения отклонений случайных величин:

$$\text{cov}_{XY} = M[(X - M(X))(Y - M(Y))] = M(XY) - M(X)M(Y)$$

где M - математическое ожидание.

$$\text{cov} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) \quad - \text{выборочный к-т ковариации}$$

Свойства ковариации:

Ковариация двух независимых случайных величин равна нулю.

Основные числовые характеристики многомерной случайной величины

□ **вектор средних значений**

$$a = (a_1, a_2, \dots, a_k)^T, \quad \text{где} \quad a_j = M\xi^{(j)}$$

$$\overline{x^{(i)}} = (\overline{x^{(1)}}, \overline{x^{(2)}}, \dots, \overline{x^{(k)}})^T,$$

Основные числовые характеристики многомерной случайной величины

Ковариационная матрица

$$\Sigma = \{\sigma_{jl}\}, \text{ где } j, l = 1, 2, \dots, k; \sigma_{jl} = M((\xi^{(j)} - a_j)(\xi^{(l)} - a_l))$$

$$\text{cov}(x^{(i)}, x^{(j)}) = S_{ij} = \frac{1}{n} \sum_{l=1}^n (x_l^{(i)} - \overline{x^{(i)}}) \cdot (x_l^{(j)} - \overline{x^{(j)}})$$

$$S_{ii} = S_i^2 = \frac{1}{n} \sum_{l=1}^n (x_l^{(i)} - \overline{x^{(i)}})^2 = \text{Var}(x^{(i)}) - \text{дисперсия}$$

$$\sqrt{S_{ii}} = \sqrt{S_i^2} = S_i$$

Основные числовые характеристики многомерной случайной величины

Ковариационная матрица

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{l1} & S_{l2} & \cdots & S_{lk} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{pmatrix}$$

The element S_{11} is circled in red, and a red arrow points from the label S_1^2 above it to the circled element.

Основные числовые характеристики многомерной случайной величины

□ вектор средних значений

$$a = (a_1, a_2, \dots, a_k)^T, \quad \text{где} \quad a_j = M\xi^{(j)}$$

□ Ковариационная матрица

$$\Sigma = \{\sigma_{jl}\}, \quad \text{где} \quad j, l = 1, 2, \dots, p; \quad \sigma_{jl} = M((\xi^{(j)} - a_j)(\xi^{(l)} - a_l))$$

- коэффициент ковариации (при $j=l \longrightarrow \sigma_{jj} = \sigma_j^2$)

это дисперсия величины

Таким образом, ковариационная матрица Σ есть неотрицательно определенная матрица, по главной диагонали которой находятся дисперсии элементов вектора

ξ_j

ξ

Основные числовые характеристики многомерной случайной величины

□ вектор средних значений

$$a = (a_1, a_2, \dots, a_k)^T, \text{ где } a_j = M\xi^{(j)}$$

□ Ковариационная матрица

$$\Sigma = \{\sigma_{jl}\}, \text{ где } j, l = 1, 2, \dots, p; \sigma_{jl} = M((\xi^{(j)} - a_j)(\xi^{(l)} - a_l))$$

- коэффициент ковариации (при $j=l \longrightarrow \sigma_{jj} = \sigma_j^2$)

это дисперсия величины

Таким образом, ковариационная матрица Σ есть неотрицательно определенная матрица, по главной диагонали которой находятся дисперсии элементов вектора

ξ_j

ξ

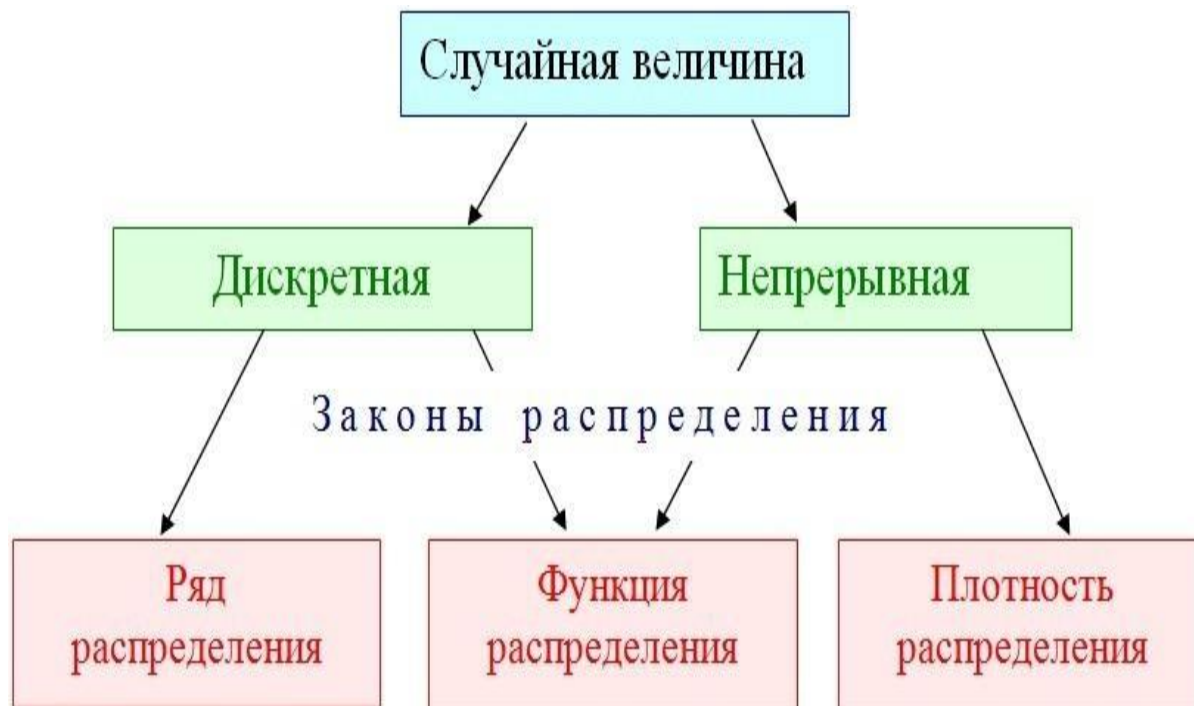
Способы описания поведения многомерного признака

1) Многомерный (совместный) закон распределения вероятностей

Задать случайную величину – это значит задать закон распределения случайной величины

Способы задания закона распределения вероятностей:

Способы описания поведения многомерного признака



Способы описания поведения многомерного признака

1) Многомерный (совместный) закон распределения вероятностей

Задать случайную величину – это значит задать закон распределения случайной величины

Способы задания закона распределения вероятностей:

- Аналитический
- Графический
- Табличный
- Алгоритмический

Примеры

Нормальный закон распределения

$$f(x^{(1)}, \dots, x^{(k)} \mid a; \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \cdot e^{-1/2(X-a)^T \Sigma^{-1}(X-a)}$$

где $a = (a_1, \dots, a_k)^T$ – вектор средних значений

$\Sigma = \{\sigma_{jl}\}$ $j, l = 1, 2, \dots, k$ – ковариационная матрица

анализируемой случайной величины

Основные числовые характеристики многомерной случайной величины

□ вектор средних значений

$$a = (a_1, a_2, \dots, a_k)^T, \quad \text{где} \quad a_j = M\xi^{(j)}$$

□ Ковариационная матрица

$$\Sigma = \{\sigma_{jl}\}, \quad \text{где} \quad j, l = 1, 2, \dots, p; \quad \sigma_{jl} = M((\xi^{(j)} - a_j)(\xi^{(l)} - a_l))$$

- коэффициент ковариации (при $j=l \longrightarrow \sigma_{jj} = \sigma_j^2$)

это дисперсия величины $\xi^{(j)}$

Таким образом, ковариационная матрица Σ есть неотрицательно определенная матрица, по главной диагонали которой находятся дисперсии элементов ξ_j вектора ξ

Три базовые задачи прикладного МСА

1. Статистическое исследование зависимостей

Множественный корреляционный и регрессионный анализ, каноническая корреляция

2. Многомерная классификация наблюдений и признаков

кластерный и дискриминантный анализ, многомерное шкалирование, расщепление смесей

3. Снижение размерности признакового пространства

Компонентный и факторный анализ, экспертно – статистические методы снижения размерности

Полезные материалы о Stata 7

Stata 7 - универсальный статистический пакет со специализацией в областях эконометрики и биометрики (нач. 80-х гг.)

Производитель - компания Stata Corporation (College Station), Техас, США.

<http://www.exponenta.ru/soft/Others/stata/stata.asp#1>

Полезные материалы о Stata

1. Официальный сайт Stata: <http://www.stata.com/>
2. Описание Stata на русском языке:
<http://www.exponenta.ru/soft/Others/stata/stata.asp>
3. Коленников С.О. Прикладной эконометрический анализ в статистическом пакете Stata.- М.: Российская экономическая школа, 2001.
4. Christopher F. Baum. An Introduction to Modern Econometrics Using Stata. Stata Press. Кристофер Баум. Введение в современную эконометрику с применением эконометрического пакета Stata.

Рекомендуемая литература

Берндт Э. Практика эконометрики
Классика и современность. М.:
Юнити, 2005.



Вербик Марно.
Путеводитель по
современной эконометрике.
М., «Научная книга», 2008.



Статистические методы в этом случае используются для решения следующих задач:

- определение основных характеристик по каждому одномерному признаку;
- выявление аномальных наблюдений;
- анализ наличия и степени зависимости между признаками;
- исследование вида зависимости одной переменной (результативной) от остальных (факторных);
- классификация наблюдений с целью получения однородных групп (кластеров);
- построение обобщающих, интегральных показателей с целью снижения размерности исходного признакового пространства.