



НАРОДНАЯ УКРАИНСКАЯ АКАДЕМИЯ

АНАЛИЗ КАТЕГОРИАЛЬНЫХ ДАННЫХ

Методические рекомендации
для слушателей факультета последипломного
образования, обучающихся по специальности
7.03010101 – Социология

Издательство НУА

НАРОДНАЯ УКРАИНСКАЯ АКАДЕМИЯ

АНАЛИЗ КАТЕГОРИАЛЬНЫХ ДАННЫХ

Методические рекомендации
для слушателей факультета последипломного
образования, обучающихся по специальности
7.03010101 – Социология

Харьков
Издательство НУА
2015

УДК 316:51(072+075.8)
ББК 60.504в631р30-2
А64

*Утверждено на заседании кафедры социологии
Народной украинской академии.
Протокол № 13 от 06.04.2015*

Автор-составитель *И. С. Нечитайло*

Рецензенты: канд. социол. наук *Е. В. Бирченко*
канд. социол. наук *О. С. Овакимян*

Дане видання складено згідно з вимогами щодо організації навчального процесу з підготовки спеціалістів, які отримують післядипломну освіту, і включає матеріали, що дозволяють слухачам закріпити теоретичні знання з аналізу категоріальних даних, розвинути навички застосування цих знань на практиці та відповідні компетенції в системі роботи за спеціальністю «Соціологія».

А64 **Анализ** категориальных данных : метод. рекомендации для слушателей фак. последиплом. образования, обучающихся по специальности 7.03010101 – Социология / Нар. укр. акад., [каф. социологии ; авт.-сост. И. С. Нечитайло]. – Харьков : Изд-во НУА, 2015. – 52 с.

Настоящее издание составлено в соответствии с требованиями по организации учебного процесса подготовки специалистов, получающих последипломное образование, и включает материалы, позволяющие слушателям закрепить теоретические знания по анализу категориальных данных, развить навыки применения этих знаний на практике и соответствующие компетенции в системе работы по специальности «Социология».

УДК 316:51(072+075.8)
ББК 60.504в631р30-2

© Народная украинская академия, 2015

ВВЕДЕНИЕ

Интенсивное развитие отечественного социогуманитарного знания с неизбежностью ведет к повышению требований к качеству и содержанию программ и курсов подготовки специалистов высшей квалификации социально-гуманитарного профиля. Неотъемлемой составляющей разносторонней подготовки социологов выступает дисциплина «Анализ категориальных данных».

Учебная дисциплина «Анализ категориальных данных» читается для слушателей факультета последипломного образования, обучающихся по специальности 7.03010101 – Социология и входит в цикл дисциплин профессиональной и практической подготовки.

Учебный курс «Анализ категориальных данных» включает в себя лекционные и практические занятия и самостоятельную работу над материалом.

В ходе *лекций* раскрывается предметно-понятийное содержание дисциплины, рассматриваются наиболее значимые проблемы анализа категориальных данных; анализируются основные приемы и методы работы с категориальными переменными, модели связи между переменными, подходы к их анализу.

Практические занятия дают возможность получить навыки самостоятельной работы с компьютерным пакетом статистического анализа данных «ОСА» и «SPSS». Желательным является предварительное знакомство с курсом «Математические методы в социологии», «Информатика и АРМ социолога».

В ходе *самостоятельной подготовки индивидуальной работы* уточняются, развиваются и находят реализацию индивидуальные предпочтения и интересы слушателей, что является чрезвычайно важным с точки зрения эффективности образовательного процесса.

Цель курса – сформировать целостное видение логики исследовательского процесса, связанного с анализом социологических данных, обработкой категориальных переменных и интерпретации категориальных данных, а также развитие навыков использования этих знаний и умений в социологической исследовательской практике.

Задачи:

1. Научиться отличать категориальные от некатегориальных переменных.
2. Изучить ряд математических процедур, обращение к которым представляется целесообразным при анализе категориальных данных.
3. Научиться анализировать данные, полученные по одной категориальной переменной.
4. Научиться определять характер и силу связи между двумя категориальными переменными.
5. Научиться определять характер и силу связи между тремя и более категориальными переменными.

По окончании курса у слушателей должно сложиться целостное представление о комплексе существующих проблем, связанных с анализом категориальных данных, приемами и методами их обработки, определением характе-

ра взаимосвязей между переменными (двумя и более); об особенностях интерпретации коэффициентов связи.

В методических рекомендациях детально рассмотрены темы содержательных модулей: «Анализ одной категориальной переменной», «Модели связи между двумя категориальными переменными» и «основы многомерного анализа категориальных данных», которые направлены как на овладение теоретическим материалом, так и на формирование практических умений анализа категориальных данных.

Наряду с этим пособие дополняется словарем основных понятий курса «Анализ категориальных данных», что позволяет еще раз закрепить лекционный материал во время подготовки к семинарским занятиям.

В результате изучения учебной дисциплины слушатели должны *знать*:

- методологические основы применения аналитических процедур при обработке данных, полученных по категориальным переменным;
- отличие категориальных переменных от некатегориальных;
- особенности построения шкал, отражающих специфику категориальных переменных, а также обработки этих переменных;
- основные характеристики таблиц сопряженности двух категориальных переменных;
- коэффициенты связи между двумя категориальными переменными и алгоритмы расчета этих коэффициентов;
- основные характеристики многомерных таблиц сопряженности категориальных переменных;
- методологические и алгоритмические основы процедур измерения связи между тремя и большим количеством категориальных переменных (в частности, причинного, факторного и кластерного анализа)

уметь:

- организовывать процесс определения и кодирования переменных, строить категориальные переменные;
- приводить примеры одномерного распределения и признаков объекта исследования, измеряемых с помощью шкалирования категориальных переменных;
- осуществлять вычислительные процедуры, связанные с анализом одной, двух, трех и более категориальных переменных;
- владеть основными практическими приемами применения соответствующих алгоритмов вычисления важных (с точки зрения социологии) показателей связи между категориальными переменными.

В общем указанная цель дисциплины «Анализ категориальных данных» направлена на формирование системы компетенций, необходимых для полноценной социально-профессиональной деятельности социолога. В частности, речь идет о таких компетенции, как:

- социально-личностные (организация личной деятельности, отражается умением учитывать изменения, происходящие в обществе);
- общенаучные (предполагающие умение использовать в социальных и профессиональных практиках базовые знания в области информатики и современных информационных технологий, умение обрабатывать и обновлять социологические базы данных; навыки использования программных средств);
- инструментальные (общие навыки работы с компьютером);
- общепрофессиональные (качество социологической деятельности, отображаемое умением осуществлять экспертизу проведенных социологических исследований);
- специализированно-профессиональные (глубокое понимание сути и социальной значимости, основных задач социолога в условиях современного общества, обеспечение необходимого уровня индивидуальной подготовки в области профессиональных знаний, использование на практике собственных знаний и наблюдений, а также соответствующих процедур социологического анализа).

Межпредметные связи

Содержательно курс «Анализ категориальных данных» связан с рядом учебных дисциплин, которые читаются на факультете последипломного образования, для слушателей, обучающихся по специальности «Социология». А именно, речь идет о следующих: «Методология и методы социологических исследований», «Математические методы в социологии», «Информатика и АРМ социолога», «Социологический практикум», «Надежность социологической информации». Однако специфика дисциплины «Анализ категориальных данных» заключается в изучении более сложных механизмов анализа взаимосвязей между переменными (двумя и более) и выработке навыков интерпретации коэффициентов подобной связи.

Контроль полученных знаний

Предварительный – осуществляется в устной и письменной форме в начале изучения курса либо перед изучением новой темы с целью выяснения общего уровня знаний.

Текущий – проводится с целью диагностики уровня знаний, умений и навыков по темам практических занятий. Предполагает выполнение практических заданий. По содержанию включает два аспекта: качество усвоения материала и самостоятельность слушателей при выполнении заданий. Слушатель может получить максимально 5 баллов за устные ответы и выполнение задания на практическом занятии.

Итоговый – осуществляется в форме зачета на последнем занятии по курсу «Анализ категориальных данных» на основе специально подготовленных вопросов, либо в форме выполнения индивидуальных заданий.

Оценки «зачтено», «не зачтено» детерминируются показателями контроля усвоения материала лекций, практических занятий и всего курса в целом.

ТЕМАТИЧЕСКИЙ ПЛАН КУРСА

ТЕМА	Количество часов				Форма контроля
	Всего	Лекционные занятия	Практические занятия	Самостоятельная/индивидуальная работа	
<i>Модуль 1. Анализ одной категориальной переменной</i>					
1. Дискретные и непрерывные, категориальные и некатегориальные, зависимые и независимые переменные	9	-	1	4/4	опрос
2. Процедуры построения, кодирования и интерпретации категориальных переменных	7	1	-	2/4	опрос
3. Важные характеристики одной категориальной переменной	6	1	1	2/2	решение задач
<i>Модуль 2. Анализ двух категориальных переменных</i>					
4. Модели связи двух категориальных переменных	12	2	2	4/4	решение задач
5. Связь двух категориальных переменных (как отличие от статистической независимости)	12	2	2	4/4	решение задач
<i>Модуль 3. Основы многомерного анализа категориальных переменных</i>					
6. Варианты взаимодействия трех и большего количества категориальных переменных	8	2	2	2/2	опрос
7. Подходы к анализу связи трех и большего количества категориальных переменных	6	1	1	2/2	решение задач
ЗАЧЕТ	12	1	1	4/6	опрос, вып. индивид. заданий
Итого	72	10	10	24/28	

ПРОГРАММА КУРСА

Учебный курс «Анализ категориальных данных» включает в себя три модуля, содержание которых заключается в следующем.

Первый модуль предполагает изучение особенностей анализа одной категориальной переменной, а именно – дискретных и непрерывных, зависимых и независимых переменных; дихотомических переменных (их определение, кодирование, интерпретация фиктивных переменных). Кроме того, содержание данного модуля включает в себя повторение знаний об одномерных таблицах абсолютных и относительных частот и интерпретации представленных в них данных, а также значения мер вариации и центральной тенденции для номинальных признаков. Цель данного модуля – сформировать у слушателей четкое представление в отношении анализа одной переменной.

В соответствии с этим, слушатели должны уметь описать основные формы существования информации (динамический ряд, вариационный ряд, матрица данных вида «объект-признак»); охарактеризовать процесс определения и кодирования переменных; привести примеры одномерного распределения; произвести вычислительные процедуры по определению мер центральной тенденции; знать социологический смысл понятий «измерение», «одномерное распределение», «мера центральной тенденции»; характеристику дискретных и непрерывных, зависимых и независимых переменных; сущность интерпретации фиктивных переменных.

Второй модуль предполагает изучение моделей связи между двумя категориальными переменными, а именно – определение специфики отличия подобной связи от статистической независимости, а также ее производных. В результате освоения материала студенты должны знать основные характеристики таблиц сопряженности двух эмпирических индикаторов; определение функции «Хи-квадрат», сущность нумерации значений функции «Хи-квадрат»; теоретическое обоснование коэффициентов Чупрова и Крамера; уметь привести примеры таблиц сопряженности; обосновывать связи типа «признак-признак»; на основе функции «Хи-квадрат», с использованием реальных примеров, осуществлять проверку статистических гипотез; с использованием индивидуальных примеров вычислять коэффициенты Чупрова, Крамера и др., объяснять полученные результаты.

Третий модуль предполагает подробное рассмотрение разных подходов к анализу многомерных таблиц, в частности – различных вариантов взаимодействия переменных. Кроме того, содержание данного модуля предусматривает изучение таких процедур многомерного анализа как факторный и кластерный анализ. Слушатели должны знать сущность и особенности анализа многомерных таблиц, особенности построения тримерных таблиц, уметь проводить факторный анализ, интерпретировать факторы, факторные нагрузки, а также владеть процедурами кластерного анализа.

Содержательный модуль 1 *«Анализ одной категориальной переменной»*

Тема 1. Дискретные и непрерывные, категориальные и некатегориальные, зависимые и независимые переменные

Роль эмпирических данных в социологии. Эмпирические основы для изучения социальных явлений. Основные цели и задачи анализа данных. Характеристика дискретных и непрерывных переменных. Зависимые и независимые переменные. Матрица «объект-признак», ее роль.

Тема 2. Процедуры построения, кодирования и интерпретации категориальных переменных

Процесс определения и кодирования переменных. Интерпретация фиктивных переменных. Дихотомические переменные: определение сущности. Метод множественной дихотомии. Особенности кодирования дихотомических переменных. Частотные таблицы для дихотомических и категориальных переменных. Сравнение дихотомического и категориального методов. Специфика анализа множественных ответов.

Тема 3. Важные характеристики одной категориальной переменной

Одномерные частотные распределения как метод анализа социологических данных (в частности, категориальных). Различные методы вычисления процентных распределений и анализа одной категориальной переменной: процент от числа опрошенных, процент от числа ответивших, кумулятивный процент. Статистические характеристики одномерных частотных распределений. Меры центральной тенденции для категориальных (качественных, номинальных) признаков. Вариация категориальных (качественных) переменных. Возможности «трансформации» категориальных переменных в некатегориальные для вычисления средних, модальных и медиальных показателей. Особенности использования графических форм представления одномерных частотных распределений по категориальным переменным.

Содержательный модуль 2 *«Модели связи между двумя категориальными переменными»*

Тема 4. Модели связи двух категориальных переменных

Методы оценки и измерения статистических связей между признаками. Таблица сопряженности как инструмент анализа взаимосвязи между категори-

альными переменными. Методы получения различных типов процентов, их познавательные возможности. Анализ вопросов множественного выбора. Проверка статистических гипотез о взаимозависимости категориальных переменных. Характеристика связей типа «признак-признак». Особенности построения и анализа двумерных таблиц для категориальных (номинальных, качественных), порядковых и метрических переменных, в том числе и средствами пакета «ОСА».

Тема 5. Связь двух категориальных переменных (как отличие от статистической независимости)

Процедура определения связи между двумя категориальными переменными: возможности пакета «ОСА». Помощь видов критериев для проверки гипотезы об отсутствии связи между двумя категориальными переменными. Понимание отсутствия связи между признаками как их статистической независимости; свойства частотных таблиц, отражающих такую независимость; использование критерия Хи-квадрат для проверки статистической гипотезы о независимости признаков. Коэффициенты связи, основанные на критерии Хи-квадрат. Коэффициенты корреляции для категориальных (номинальных, качественных) признаков, их специфика и познавательные возможности. Коэффициенты Пирсона, Чупрова и Крамера, лямбда-Гудмана и другие).

Содержательный модуль 3

«Основы многомерного анализа категориальных переменных»

Тема 6. Варианты взаимодействия тремя и большего количества категориальных переменных

Трехмерные и многомерные таблицы сопряженности (связности): особенности построения и анализа. Условная и многомерная энтропия. Понятие многомерной связи. Примеры необходимости поиска связи между несколькими (или многими) категориальными переменными.

Тема 7. Подходы к анализу связи трех и большего количества категориальных переменных

Методологические основы многомерного анализа переменных. Основные виды многомерного анализа, в том числе и категориальных, переменных: причинный, факторный, кластерный и др. Практики применения некоторых методов многомерного анализа категориальных данных: метод «кластеров на факторах». Графическое представление связи между несколькими (или многими) категориальными данными: практика построения графов.

СПИСОК РЕКОМЕНДУЕМЫХ ЛИТЕРАТУРНЫХ И ИНФОРМАЦИОННЫХ ИСТОЧНИКОВ

Основные источники

1. Айвазян С. А. Инструменты статистического анализа данных / С. А. Айвазян, В. С. Степанов. – М., 2000. – 320 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004. – 250 с.
3. Гайдышев И. Анализ и обработка данных : спец. справ. / И. Гайдышев. – СПб. : Питер, 2002. – 752 с.
4. Гаптон А. Анализ таблиц сопряженности [Электронный ресурс] / Гаптон А. // Электронная библиотека МГУ им. М. В. Ломоносова. – Режим доступа: <http://lib.socio.msu.ru/library>
5. Калинина В. Н. Введение в многомерный статистический анализ / В. Н. Калинина, В. И. Соловьев. – М. : Госуниверситет управления, 2003. – 289 с.
Косова Л. Б. Как оценить неравенство: аналитический обзор подходов к измерению социальной стратификации / Л. Б. Косова // Социология 4М. – 2013. – № 37. – С. 151–178.
6. Лук'янова В. В. Комп'ютерний аналіз даних : посіб. / В. В. Лук'янова. – К. : Академія, 2003. – 344 с.
7. Нечитайло И. С. Математические методы в социологии : учеб. для студентов вузов / И. С. Нечитайло, М. В. Бирюкова ; Нар. укр. акад. – Харьков : Изд-во НУА, 2014. – 320 с.
8. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
9. Програма обробки соціологічних анкет «ОСА» : метод. рек. для студ., які навчаються за спец. 6.030101 – Соціологія / Нар. укр. акад. [каф. соціології ; авт.-уклад. О. С. Овакімян]. – Х. : Вид-во НУА, 2007. – 44 с.
Руднев М. Г. Инвариантность измерения базовых ценностей по методике Шварца среди русскоязычного населения четырех стран / М. Г. Руднев // Социология 4М. – 2013. – № 37. – С. 7–38.
10. Татарова Г. Г. Анализ дисперсионный. Анализ типологический. Матрица данных. Мера близости. Меры рассеяния. Методы классификации. Методы социологического исследования / Г. Г. Татарова // Социологический словарь. – М. : НОРМА, 2008. – 535 с.
11. Татарова Г. Г. Методология анализа данных в социологии : учеб. пособие / Г. Г. Татарова. – М. : NOTA BENE, 1999. – 224 с.
12. Толстова Ю. Н. Анализ социологических данных (Методология, дескриптивная статистика, изучение связей между номинальными признаками) / Ю. Н. Толстова. – М. : Науч. мир, 2003. – 364 с.
13. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003. – 330 с.

14. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

Дополнительные источники

15. Анализ данных и логика вывода в эмпирическом исследовании : учеб. пособие для вузов. – Ростов н/Д : Феникс, 2005. – 192 с.

16. Анурин В. Ф. Эмпирическая социология : учеб. пособие для вузов / В. Ф. Анурин. – М. : Академ. проект, 2003. – 287 с.

17. Аргунова К. Д. Качественный регрессионный анализ в социологии / Аргунова К. Д. – М. : ИСАН СССР, 1999. – 180 с.

18. Батыгин Г. С. Ремесло Пауля Лазарсфельда (Введение в научную биографию) / Г. С. Батыгин // Вестник АН СССР. – 1990. – № 8. – С. 31–40.

19. Бююль А. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цефель. – М., СПб., К. : Торгово-издат. дом Dia Soft, 2002. – С. 82–104, 117–129, 143, 164–206, 220–255.

Гаврилов К. А. «Шкала мировоззрения»: возможности и ограничения / К. А. Гаврилов // Социология 4М. – 2014. – № 38. – С. 128–164.

20. Голузинський Г. Г. Сучасні технологічні засоби обробки інформації : навч. посіб. / Г. Г. Голузинський, І. В. Гордієнко. – К. : КНЕУ, 1998. – 224 с.

21. Горбачик А. П. Руководство пользователя системы ОСА / А. П. Горбачик. – К. : Бюро стат. анализа, 2004. – 52 с.

22. Девятко И. Ф. Модели объяснения и логика социологического исследования / И. Ф. Девятко. – М. : Ин-т социологии и образования, 1996. – 420 с.

23. Коченков А. И. Идеи Лазарсфельда в современной России / А. И. Коченков, Ю. Н. Толстова // Социология: 4М. – 2003. – № 16. – С. 70–176.

Кузьмина Ю. В. Метод разрывной регрессии и метод отбора подобного по вероятности для оценки эффекта одного года обучения: опыт применения на примере данных PISA 2009 / Ю. В. Кузьмина // Социология 4М. – 2014. – № 38. – С. 7–37.

24. Кулешков Ю. В. Статистичні методи обробки та аналізу економічних даних : навч. посіб. / Ю. В. Кулешков, В. В. Аулін, М. М. Петренко, Л. В. Саловська, А. А. Олійник. – Кіровоград : КДТУ, 2003. – 240 с.

25. Міщенко Г. Автоматизована система інформаційно-аналітичного забезпечення маркетингових досліджень / Г. Міщенко. – К., 2000. – 150 с.

26. Наследов А. Д. SPSS: Компьютерный анализ данных в психологии и социальных науках / А. Д. Наследов. – СПб. : Питер, 2005. – 280 с.

27. Плотинский Ю. М. Визуализация информации / Ю. М. Плотинский. – М. : изд-во МГУ, 1994. – 230 с.

28. Плотинский Ю. М. Теоретические и эмпирические модели социальных процессов : учеб. пособие / Ю. М. Плотинский. – М. : Логос, 1998. – 188 с.

29. Сажин Ю. В. Многомерные статистические методы : учеб. пособие / Ю. В. Сажин, В. А. Басова. – М. : Спутник, 2002. – 310 с.

30. Таганов Д. Н. SPSS: Статистический анализ в маркетинговых исследованиях / Д. Н. Таганов. – СПб. : Питер, 2005. – 192 с.

31. Татарова Г. Г. Основы типологического анализа в социологических исследованиях / Г. Г. Татарова. – М. : Изд. дом «Высш. образование и наука», 2007. – 210 с.

32. Татарова Г. Г. Типологический анализ как исследовательская стратегия / Г. Г. Татарова // Социологические методы в современной исследовательской практике: материалы Всерос. науч. конф. памяти А. О. Крыштановского. – М. : ГУ ВШЭ, 2007.

33. Тимченко А. А. Основи системного проектування та системного аналізу складних об'єктів : навч. посіб. для студ. вищ. навч. закл. напрям. «Комп'ютерні науки» / А. А. Тимченко. – К. : Либідь, 2004. – 280 с.

34. Толстова Ю. Н. Измерение в социологии / Ю. Н. Толстова. – М.: Инфра-М, 1998. – 280 с.

35. Толстова Ю. Н. Классификация методов анализа социологических данных / Ю. Н. Толстова // Глобализация и социальные изменения в современной России: тез. докл. и выступлений Всерос. социол. конгресса. – М. : Альфа-М, 2006. – Т. 5. – С. 78–81.

36. Толстова Ю. Н. Обобщенный подход к определению понятия социологического измерения / Ю. Н. Толстова // Методология и методы социологического исследования (итоги работы поисковых проектов 1992–1996). – М. : ИСРАН, 1996. – С. 66–95.

37. Тюрин Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. – М. : Инфра-М, 1998.

38. Шудло С. А. Соціологія: структурно-логічні схеми і таблиці / С. А. Шудло. – Дрогобич, 2002. – 220 с.

39. Ядов В. А. Рассуждения о теоретических предпочтениях / В. А. Ядов // Социологический журнал. – 1995. – № 2. – С. 70–72.

40. Ядов В. А. Стратегия социологического исследования: Описание, объяснение, понимание социальной реальности : учеб. пособие / Ядов В. А. – М. : Омега-Л, 2012. – 567 с.

41. Ярская-Смирнова Е. Социокультурный анализ нетипичности / Е. Ярская-Смирнова. – Саратов : Саратовский технический ун-т, 1997. – 130 с.

Информационные ресурсы

1. Центр Разумкова. Официальный сайт. – Режим доступа: http://www.ucerps.org/ukr/expert.php?news_id=3293.

2. Офіційний сайт Головного управління статистики в Харківській області. – Режим доступу: <http://kh.ukrstat.gov.ua/>.

ПЛАНЫ СЕМИНАРСКИХ ЗАНЯТИЙ И МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ

Содержательный модуль 1 «Анализ одной категориальной переменной»

Тема 1. Дискретные и непрерывные, категориальные и некатегориальные, зависимые и независимые переменные

План

1. Построение дискретных и непрерывных переменных и соответствующих шкал.
2. Определение особенностей категориальных и некатегориальных переменных.
3. Практики определения зависимых и независимых переменных.

Основная литература

1. Лук'янова В. В. Компьютерний аналіз даних : посіб. / В. В. Лук'янова. – К. : Вид. центр «Академія», 2003. – 344 с.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Толстова Ю. Н. Анализ социологических данных (Методология, дескриптивная статистика, изучение связей между номинальными признаками) / Ю. Н. Толстова. – М. : Науч. мир, 2003. – 364 с.
4. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.

Дополнительная литература

1. Анализ данных и логика вывода в эмпирическом исследовании : учеб. пособие для вузов. – Ростов н/Д : Феникс, 2005. – 192 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004.
3. Гайдышев И. Анализ и обработка данных : спец. справ. / И. Гайдышев. – СПб. : Питер, 2002. – 752 с.
4. Толстова Ю. Н. Классификация методов анализа социологических данных / Ю. Н. Толстова // Глобализация и социальные изменения в современной России: тез. докл. и выступлений Всерос. социол. конгресса. – М. : Альфа-М, 2006. – Т. 5. – С. 78–81.
5. Ядов В. А. Стратегия социологического исследования: Описание, объяснение, понимание социальной реальности : учеб. пособие / Ядов В. А. – М. : Омега-Л, 2012. – 567 с.

Тема 3. Важные характеристики одной категориальной переменной

План

1. Построение категориальных и некатегориальных переменных и соответствующих шкал.
2. Построение и особенности анализа таблиц одномерного распределения.
3. Вычисление мер центральной тенденции и вариации для номинальных признаков.
4. Практики «трансформации» категориальных переменных в некатегориальные.

Основная литература

1. Лук'янова В. В. Комп'ютерний аналіз даних : посіб. / В. В. Лук'янова. – К. : Вид. центр «Академія», 2003. – 344 с.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Толстова Ю. Н. Анализ социологических данных (Методология, дескриптивная статистика, изучение связей между номинальными признаками) / Ю. Н. Толстова. – М. : Науч. мир, 2003. – 364 с.
4. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.

Дополнительная литература

1. Анализ данных и логика вывода в эмпирическом исследовании : учеб. пособие для вузов. – Ростов н/Д : Феникс, 2005. – 192 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004.
3. Гайдышев И. Анализ и обработка данных : спец. справ. / И. Гайдышев. – СПб. : Питер, 2002. – 752 с.
4. Толстова Ю. Н. Классификация методов анализа социологических данных / Ю. Н. Толстова // Глобализация и социальные изменения в современной России: тез. докл. и выступлений Всерос. социол. конгресса. – М. : Альфа-М, 2006. – Т. 5. – С. 78–81.
5. Ядов В. А. Стратегия социологического исследования: Описание, объяснение, понимание социальной реальности : учеб. пособие / Ядов В. А. – М. : Омега-Л, 2012. – 567 с.

Методические рекомендации по изучению тем модуля 1 «Анализ одной категориальной переменной»

Тема «Дискретные и непрерывные переменные. Зависимые и независимые переменные» представляет собой вводную в курсе «Анализ категориальных данных». Ее целью является ознакомление с основными задачами анализа данных, значимыми характеристиками дискретных и непрерывных, за-

зависимых и независимых переменных. Планируется также определение роли матрицы «объект-признак» при анализе переменных.

Безусловно, важно обратить внимание на то, что анализ данных – область социологии, математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных. В социологии анализ данных применяется в основном к информации, полученной в процессе проведения опросов. Основными формами представления информации по одной категориальной переменной выступают динамический ряд, вариационный ряд, матрица данных вида «объект-признак». При этом динамическим является ряд последовательных значений какого-либо статистического показателя, изменяющихся во времени. Динамический ряд широко используется при обработке материалов социологических исследований. Вариационный ряд представляет собой совокупность каких-либо величин, расположенных в порядке их возрастания; полностью определяется указанием различных значений, входящих в него величин и числа членов ряда. Матрица данных вида «объект-признак» выглядит в виде прямоугольной таблицы, которая используется в качестве стандартного способа представления данных в программных средствах, предназначенных для статистической обработки и анализа данных, в том числе в пакете «ОСА». Каждая строка такой матрицы отводится одному объекту, каждый столбец – одному признаку; на пересечении строки и столбца указывается значение данной переменной для соответствующего объекта.

При изучении первой темы модуля важно обратить внимание на то, что категориальные переменные различаются по типам. В зависимости от того, какие значения может потенциально принимать переменная, выделяют два типа количественных переменных: дискретные и непрерывные.

Дискретная – это такая переменная, которая может принимать значения только из некоторого списка определенных чисел. Примерами дискретной переменной являются число детей в семье; число вызовов «скорой помощи», поступающих в больницу; число клиентов, обратившихся в фирму за определенный промежуток времени, и т. д. Непрерывной будем считать любую переменную, не являющуюся дискретной. Такая переменная принимает значения из некоторого промежутка. Примерами непрерывной переменной является рост взрослого человека (например, от 140 до 230 см), фактическая масса буханки хлеба (например, от 750 до 830 грамм) и т. п.

Таким образом, дискретная переменная представляет собой числовую переменную, которая может принимать значения только в целых числах и иметь только два возможных значения. Непрерывная переменная в отличие от дискретной может иметь любое значение. Она является количественной переменной, которая постоянно возрастает в своем значении с непрерывным рядом (серией) градаций возможных величин.

Кроме того, есть данные, которые регистрируют определенное качество, которым обладает объект. Такие данные называют качественными. Даже если

значениям этого качества можно приписать числа (например, полу человека приписать соответственно числа 0 и 1), то обрабатывать эти числа как количественные данные нельзя. Примерами качественных данных являются тип школы, где обучается ребенок (лицей, гимназия, специализированная школа); должность, которую занимает сотрудник на предприятии; названия газет, которые читают в определенном городе, и т. п.

Качественные данные бывают двух типов: порядковые, для которых характерен имеющий содержательный смысл порядок, и номинальные, для которых не характерен содержательно интерпретируемый порядок.

Порядковые данные можно ранжировать и использовать это ранжирование при проведении статистического анализа. Примером порядковых данных являются ответы на вопросы анкеты, содержащей следующие варианты ответов: да; скорее да, чем нет; скорее нет, чем да; нет. Хотя и можно выразить эти ответы числами (например, 4, 3, 2, 1), но предложенная шкала оценок носит субъективный характер. Нельзя считать, что разница между ответами 4 и 3 такая же, как и между ответами 2 и 1. Также нельзя считать, что ответ 3 в три раза лучше ответа 1.

Для номинальных данных не характерны числовые значения и нет основы для ранжирования. Примерами номинальных данных являются регионы Украины, из которых приехали студенты ХГУ «НУА»; названия организаций, где работают выпускники; пол работников фирмы и т. д.

Зависимая переменная в моделях причинных связей измеряет следствие и применяется в дисперсионном регрессионном анализе. К зависимой переменной предъявляется требование надежности, то есть повторения результата при тех же условиях (возможно с теми же участниками). Если зависимая переменная не изменяется при изменении независимой переменной, то говорят о «нулевом результате». Таким образом, на зависимую переменную обязательно будет оказывать влияние другой фактор (независимая переменная). Независимая переменная (объясняющая) также называется предикторной переменной или ковариатой. В моделях причинных связей она измеряет причину. Если зависимая переменная обычно обозначается как Y , то независимая как X . Последняя применяется для прогноза зависимой переменной в регрессионном анализе.

Вторая тема модуля «Дихотомические переменные: определение, кодирование, интерпретация фиктивных переменных» дает возможность познакомиться с процессом определения и кодирования переменных. Еще одним важным направлением в изучении данной темы является вопрос об интерпретации фиктивных переменных. Так, фиктивная переменная используется для определения дихотомических данных, а также для определения характеристик, которые не могут быть оценены количественно. Такая переменная, кодируется значениями «0» или «1». При этом дихотомическая переменная (бинарная) представляет собой качественную переменную с двумя категориями.

При осуществлении анализа одной категориальной переменной следует помнить, что одномерные частотные распределения используются как метод анализа социологических и опросных данных. Частотное распределение пред-

ставляет собой распределение, которое показывает частоту появления каждого возможного наблюдения, класса наблюдений или категорий.

Одномерная статистика представляет собой статистику, описывающую одну переменную при помощи построения простейшего типа перекрестной таблицы. Одномерная таблица используется для анализа распределения данных одного типа по одному фактору. При этом используются различные методы вычисления процентных распределений – процент от числа опрошенных, процент от числа ответивших, кумулятивный процент.

Графическими формами представления одномерных частотных распределений являются: столбиковая диаграмма, точечный график, форест-график, круговой график, гистограмма.

Столбиковая диаграмма показывает распределение качественной или дискретной переменной горизонтально или вертикально для каждой категории, причем длина пропорциональна (относительной) частоте этой категории. Каждый столбец на графике соответствует одному значению признака, а его высота пропорциональна частоте встречаемости этого значения.

Точечный график представляет собой диаграмму, в которой каждое наблюдение за переменной представлено одной точкой на горизонтальной (или вертикальной) линии.

Форест-график выглядит как диаграмма, применяемая в мета-анализе и показывающая оцененный эффект в каждом исследовании и их средние (с доверительными интервалами).

Круговой график показывает частотное распределение количественной или дискретной переменной. Круг делится на секции по одной для каждой категории. При этом площадь каждой секции пропорциональна частоте в этой категории.

Гистограмма представляет собой диаграмму, которая показывает относительное распределение частоты непрерывной переменной путем присоединения столбчатых диаграмм. Область такой диаграммы пропорциональна (относительной) частоте в интервале, определенном границами этой диаграммы. При этом горизонтальная ось графика соответствует значениям переменной, а вертикальная – частотам.

Таким образом, одномерные данные представляют собой наборы данных, содержащих только один признак для каждого объекта. Эти данные позволяют определить типичное значение признака, насколько значения отличаются друг от друга, требуют ли отдельные данные особого внимания. Примером одномерных данных является информация о средней зарплате в регионе по отраслям. Она позволяет назвать отрасль с самым высоким уровнем зарплаты, понять, насколько отличаются уровни средних зарплат в различных отраслях друг от друга, обратить внимание на отрасли, где уровень зарплаты самый низкий.

Кодирование переменных – это процесс преобразования количественных и качественных данных в числовые или буквенные символы. Такой процесс предполагает выполнение операций отождествления символов или групп сим-

волов одного кода с символами или группами символов другого кода. Обратите внимание на то, что, используя процедуру кодирования переменных на этапе анализа одной категориальной переменной, необходимо осуществить подготовку информации к формализованной обработке.

Следует помнить, что анализ одной категориальной переменной будет не полным без обращения к изучению вопросов, освещаемых в теме 4 «Меры центральной тенденции и меры вариации для номинальных признаков». Мера центральной тенденции выступает как характеристика совокупности переменных (признаков), указывающих на наиболее типичный, репрезентативный для изучаемой выборки результат. В социологии мерами центральной тенденции принято считать моду, медиану и среднее.

Из курса «Математические методы в социологии» Вам уже известно, что мода выступает величиной отдельной переменной, которая наиболее часто появляется в группе данных и представляет собой наиболее часто повторяющееся значение распределения переменной. Медиана является значением переменной, которое способно разделить упорядоченное множество всех значений выборки ровно пополам: у половины объектов выборки значения переменной больше, а у другой половины меньше медианы.

Вариация же ориентирует нас на фиксацию разброса значений. За меру вариации может быть принята дисперсия: мера рассеяния, равная квадрату стандартного отклонения.

Обращаясь к определению понятия «энтропия» и объяснению вычисленного энтропийного коэффициента разброса, следует помнить, что оно представляет собой меру отклонения от эталонного (нормального, ожидаемого) результата.

Таким образом, в результате изучения тем первого модуля Вы познакомитесь с особенностями анализа одной категориальной переменной, а именно – дискретной и непрерывной, зависимой и независимой; дихотомической переменной (ее определением, кодированием, интерпретацией фиктивных переменных). Кроме того, Вы сможете продемонстрировать уровень остаточных знаний из курса «Информатика и АРМ социолога» при обсуждении вопросов об одномерных таблицах частот процентов и интерпретации представленных в них данных, а также из курса «Математические методы в социологии» при рассмотрении тем, связанных с определением значений мер вариации и центральной тенденции для номинальных признаков.

В соответствии с этим необходимо:

- уметь описать основные формы существования информации (динамический ряд, вариационный ряд, матрица данных вида «объект-признак»);
- четко знать социологический смысл понятия «измерение», определение «одномерное распределение»;
- уметь привести примеры одномерного распределения;
- знать характеристику дискретных и непрерывных, зависимых и независимых переменных;
- уметь охарактеризовать процесс определения и кодирования переменных;

- знать сущность понятия «фиктивная переменная»;
- знать определение меры центральной тенденции;
- уметь произвести вычислительные процедуры по определению мер центральной тенденции.

При изучении тем первого модуля «Анализ одной категориальной переменной» необходимо овладеть следующими основными понятиями: *измерение, динамический ряд, вариационный ряд, матрица данных вида «объект-признак, кодирование переменных, одномерное распределение, частотное распределение, дихотомическая переменная, фиктивная переменная, зависимая переменная, независимая переменная, мера центральной тенденции, мода, медиана, среднее значение.*

Содержательный модуль 2 «Модели связи между двумя категориальными переменными»

Тема 4. Модели связи двух категориальных переменных

План

1. Понятие связи между признаками (корреляция и сопряженность).
2. Случаи необходимости поиска связей между признаками.

Основная литература

1. Айвазян С. А. Инструменты статистического анализа данных / С. А. Айвазян, В. С. Степанов. – М., 2000. – 320 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004.
3. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Марков ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.
4. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

Дополнительная литература

1. Кулешков Ю. В. Статистичні методи обробки та аналізу економічних даних : навч. посіб. / Ю. В. Кулешков, В. В. Аулін, М. М. Петренко, Л. В. Саловська, А. А. Олійник. – Кіровоград : КДТУ, 2003.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Татарова Г. Г. Основы типологического анализа в социологических исследованиях / Г. Г. Татарова. – М. : Изд. дом «Высш. образование и наука», 2007.

Тема 5. Связь двух категориальных переменных (как отличие от статистической независимости)

План

1. Процедура определения связи между двумя категориальными переменными средствами пакета «ОСА».
2. Хи-квадрат в качестве критерия проверки статистической гипотезы о независимости категориальных переменных.
3. Вычисление и интерпретация коэффициентов корреляции, основанных на Хи-квадрат для категориальных (номинальных, качественных) признаков (Пирсона, Чупрова и Крамера, лямбда-Гуттмана и другие).

Основная литература

1. Айвазян С. А. Инструменты статистического анализа данных / С. А. Айвазян, В. С. Степанов. – М., 2000. – 320 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004.
3. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Маркаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.
4. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

Дополнительная литература

1. Кулешков Ю. В. Статистичні методи обробки та аналізу економічних даних : навч. посіб. / Ю. В. Кулешков, В. В. Аулін, М. М. Петренко, Л. В. Саловська, А. А. Олійник. – Кіровоград : КДТУ, 2003.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Татарова Г. Г. Основы типологического анализа в социологических исследованиях / Г. Г. Татарова. – М. : Изд. дом «Высш. образование и наука», 2007.
4. Толстова Ю. Н. Анализ социологических данных (Методология, дескриптивная статистика, изучение связей между номинальными признаками) / Ю. Н. Толстова. – М. : Науч. мир, 2003. – 364 с.

Методические рекомендации по изучению тем модуля 2

«Модели связи между двумя категориальными переменными»

В процессе изучения тем модуля «Модели связи между двумя категориальными переменными» особое внимание следует уделить следующим моментам.

Наборы двумерных данных содержат информацию о двух признаках для каждого из объектов. Кроме того, что они дают возможность получить два набора одномерных данных, двумерные данные позволяют установить, существу-

ет ли связь между двумя переменными, насколько сильно связаны переменные, можно ли предсказать значение одной переменной по значению другой и если да, то с какой надежностью. Например, данные опроса студентов о том, удовлетворены ли они уровнем теоретической и практической подготовки, получаемой в вузе (значения обеих переменных записываются в виде да/нет, или 1/0), позволяют установить, есть ли связь между уровнями теоретической и практической подготовки в вузе.

Представляется естественным использовать для оценки между двумя категориальными переменными так называемые частотные таблицы, или таблицы сопряженности. Такие таблицы несколько отличаются от обычной таблицы распределения (частот), представляющей собой таблицу, устанавливающую соотношение между категориями (значениями) признака и частотами их встречаемости. Таблица сопряженности выглядит как таблица совместного распределения частот для двух категориальных или дискретных переменных и дает выборочные оценки вероятностных распределений многомерных случайных величин. Таблица сопряженности является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения. Она выступает формой представления данных об объектах социологического исследования на основе группировки двух или более признаков по принципу их сочетаемости. Таблицы сопряженности применяются для определения процентного распределения градаций переменной в выборке.

Следует знать, что термин «таблица сопряженности» обязан своим происхождением именно тому обстоятельству, что на основе анализа подобных таблиц можно судить о сопряженности (совместной встречаемости) значений одних категориальных переменных с некоторыми значениями других переменных. Как мы увидим на практических занятиях, связь между номинальными переменными, собственно говоря, и выражается в виде подобных сопряженностей.

Предположим, что мы имеем две переменные X и Y , первая из которых принимает «г» значение $1, 2, \dots, r$, а вторая «с» значение $1, 2, \dots, c$. Назовем двумерной таблицей сопряженности (двумерной частотной таблицей) некоторую матрицу, на пересечении i -й строки и j -го столбца которой стоит число n_{ij} , означающее количество объектов, обладающих i -м значением первой переменной и j -м значением второй ($i = 1, \dots, r; j = 1, \dots, c$) (использование латинских букв r и c в указанном смысле принято в литературе; эти буквы сопрягаются с английскими словами *row* и *column*, означающими «строка» и «столбец» соответственно. При этом строки таблицы сопряженности соответствуют категориям (значениям) одной, а столбцы – другой переменной. Другими словами, таблица сопряженности выглядит так:

Общий вид таблицы сопряженности

X	Y						Маргиналы по строкам
	1	2	:	j	:	c	
1	n_{11}	n_{12}	:	n_{1j}	:	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	:	n_{2j}	:	n_{2c}	$n_{2.}$
:	:	:	:	:	:	:	:
i	n_{i1}	n_{i2}	:	n_{ij}	:	n_{ic}	$n_{i.}$
:	:	:	:	:	:	:	:
r	n_{r1}	n_{r2}	:	n_{rj}	:	n_{rc}	$n_{r.}$
Маргиналы по столбцам	$n_{.1}$	$n_{.2}$:	$n_{.j}$:	$n_{.c}$	n

В последние годы в литературе все больше используется расширительное понимание таблицы сопряженности. Предполагается, что в качестве ее элементов могут фигурировать не только частоты, но и многие другие числа: скажем, в клетках половозрастной таблицы могут стоять средние значения зарплаты тех людей, которые характеризуются отвечающим клетке значениям пола и возраста. Таким же образом в клетки таблицы могут быть помещены средние другого рода (мода, медиана), дисперсии, величины отклонений от средних по строке (столбцу), разница между эмпирической и теоретической частотой и т. д.

Изучая материал модуля, следует помнить, что практические занятия по курсу «Анализ категориальных данных» и выполнение индивидуальных заданий ориентировано на отработку навыков расчета и анализа связи для двух переменных. В том числе планируется расчет коэффициентов связи для номинальных переменных: коэффициента «Хи-квадрат», коэффициентов связи, основанных на «Хи-квадрате», коэффициентов связи, основанных на прогнозе; коэффициентов связи для порядковых и интервальных переменных; коэффициента корреляции Пирсона, рангового коэффициента корреляции Спирмена как способа измерения статистических взаимосвязей и проверки гипотез.

При этом Вам необходимо вспомнить информацию из курса «Математические методы в социологии» в отношении интерпретации подобного рода коэффициентов. Так, например, критерий «Хи-квадрат» позволяет определить статистическую значимость связи между двумя номинальными переменными. Он проверяет нулевую гипотезу о том, что нет связи между факторами, которые определяют таблицу сопряженности. Применяется также для тестирования разницы в долях (пропорциях).

Коэффициент корреляции Пирсона отражает количественную меру связи двух переменных. Обозначается он символом r и принимает значение между -1 и $+1$. Такая мера корреляции идеально подходит для двух непрерывных (метрических) переменных. Ранговый коэффициент корреляции Спирмена, напротив, является непараметрической альтернативой коэффициенту корреляции Пирсона и обеспечивает измерение связи между двумя переменными.

При социологическом исследовании обычно одновременно регистрируется множество разнообразных данных. Если данные об одной переменной представляют определенную информацию о другой, то эти переменные оказываются связанными друг с другом, или коррелируют. Корреляция является положительной, когда значения двух переменных имеют тенденцию возрастать или уменьшаться вместе. Отрицательная корреляция выявляется тогда, когда значение одной переменной возрастает, а значение другой – уменьшается. Существуют различные типы коэффициентов корреляции, используемые в отношении различных данных, причем все они направлены на измерение степени соответствия между двумя и более переменными. Сам факт, что две переменные коррелируют друг с другом, не говорит о том, какая из них предшествует или является причиной другой, или о том, что переменные вообще причинно связаны, поскольку обе они могут быть обусловлены каким-либо третьим фактором.

Следует также вспомнить, что корреляция – это мера степени и направления связи между значениями двух переменных. Она выступает показателем наличия линейной зависимости двух переменных. Коэффициент корреляции может быть использован только для интервальных переменных и не может быть интерпретирован как «степень взаимосвязи» переменных. Как правило, термин «корреляция» применяют к количественным, порядковым и дихотомическим переменным. Для количественных и порядковых переменных прямая (положительная) корреляция означает одновременное увеличение или уменьшение их значений; обратная (отрицательная) корреляция – возрастание значений одной переменной при уменьшении значений другой. Для дихотомических признаков прямая корреляция означает, что они чаще появляются или не появляются вместе, чем врозь; обратная – что признаки чаще появляются порознь, чем совместно. Для измерения корреляционной связи используются специальные меры связи.

Из курса «Информатика и АРМ социолога» Вам известно, что пакет «ОСА» обеспечивает возможность построения и анализа двумерных таблиц для дихотомических, номинальных, порядковых, метрических признаков и номинальных переменных с совместимыми альтернативами.

В рамках изучения теоретического материала нам предстоит определить-ся с сущностью понятий «статистическая зависимость» и «статистическая независимость». Итак, под статистической зависимостью мы будем понимать взаимосвязь двух признаков или величин. Статистической зависимостью будет такая связь двух случайных величин, при которой распределение вероятностей одной из них зависит от того, какие возможные значения приняла другая величина. И напротив, статистическая независимость представляет собой соотношение между двумя переменными, характеризующееся отсутствием статистической связи.

Коэффициенты Чупрова и Крамера являются производными от статистической независимости. Обращая внимание на характеристики и социологическое значение данных коэффициентов, следует помнить, что коэффициент Крамера представляет собой меру ассоциации между значениями двух катего-

риальных переменных. Значение коэффициента всегда варьируется от 0 до 1 и интерпретируется по аналогии с коэффициентом корреляции (исключая отсутствие отрицательных значений). При отсутствии статистической связи между переменными значение коэффициента равно 0; при полной связи (когда значение одной переменной полностью определяется значением второй переменной) коэффициент достигает 1. Таким образом, коэффициент Крамера является мерой связи двух номинальных переменных на основе критерия «Хи-квадрат». Применяется к таблицам сопряженности произвольной размерности.

Коэффициент Чупрова – тетракорический коэффициент корреляции для дихотомических переменных, основанных на нормальных распределениях. Изменяется в диапазоне от 0 до 1. Особенностью показателей, рассчитанных на основе коэффициента Чупрова, является то, что они строже оценивают тесноту связи.

Таким образом, в результате освоения материала необходимо:

- знать основные характеристики таблиц сопряженности двух эмпирических индикаторов,
- уметь привести примеры таблиц сопряженности;
- знать определение функции «Хи-квадрат»;
- уметь на основе функции «Хи-квадрат» с использованием реальных примеров осуществить проверку гипотезы об отсутствии связи, объяснить полученные результаты;
- знать сущность нумерации значений функции «Хи-квадрат»;
- уметь теоретически и практически обосновать связь типа «признак-признак»;
- знать теоретическое обоснование коэффициентов Чупрова и Крамера;
- уметь с использованием индивидуальных примеров вычислить коэффициенты Чупрова и Крамера, объяснить полученные результаты.

При изучении тем второго модуля «Модели связи между двумя категориальными переменными» необходимо овладеть следующими основными понятиями: *таблица сопряженности, коэффициент корреляции Пирсона, ранговый коэффициент корреляции Спирмена, коэффициент Чупрова, коэффициент Крамера, матрица данных вида «признак-признак», функция «Хи-квадрат», статистическая зависимость, статистическая независимость.*

Содержательный модуль 3
«Основы многомерного анализа категориальных переменных»

Тема 6. Варианты взаимодействия трех и большего количества категориальных переменных

План

1. Процедура построения и особенности анализа трехмерных и многомерных таблиц сопряженности (связности).
2. Примеры необходимости поиска связи между несколькими (или многими) категориальными переменными при исследовании тех или иных социальных явлений и процессов.

Основная литература

1. Айвазян С. А. Инструменты статистического анализа данных / С. А. Айвазян, В. С. Степанов. – М., 2000. – 320 с.
2. Басегян А. А. Методы и модели анализа данных: OLAP и Data Mining / А. А. Басегян и др. – СПб. : БХВ – Петербург, 2004.
3. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.
4. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

Дополнительная литература

1. Кулешков Ю. В. Статистичні методи обробки та аналізу економічних даних : навч. посіб. / Ю. В. Кулешков, В. В. Аулін, М. М. Петренко, Л. В. Саловська, А. А. Олійник. – Кіровоград : КДТУ, 2003.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Татарова Г. Г. Основы типологического анализа в социологических исследованиях / Г. Г. Татарова. – М. : Изд. дом «Выш. образование и наука», 2007.
4. Толстова Ю. Н. Анализ социологических данных (Методология, дескриптивная статистика, изучение связей между номинальными признаками) / Ю. Н. Толстова. – М. : Науч. мир, 2003. – 364 с.
5. Толстова Ю. Н. Классификация методов анализа социологических данных / Ю. Н. Толстова // Глобализация и социальные изменения в современной России: тез. докл. и выступлений Всерос. социол. конгресса. – М. : Альфа-М, 2006. – Т. 5. – С. 78–81.

Тема 7. Подходы к анализу связи трех и большего количества категориальных переменных

План

1. Процедурные особенности причинной, факторного и кластерного анализа.
2. Практики применения некоторых методов многомерного анализа категориальных данных: метод «кластеров на факторах».
3. Практика построения графов.

Основная литература

1. Калинина В. Н. Введение в многомерный статистический анализ / В. Н. Калинина, В. И. Соловьев. – М. : Госуниверситет управления, 2003.
2. Коченков А. И. Идеи Лазарсфельда в современной России / А. И. Коченков, Ю. Н. Толстова // 4М. – 2003. – № 16. – С. 70–176.
3. Сажин Ю. В. Многомерные статистические методы : учеб. пособие / Ю. В. Сажин, В. А. Басова. – М. : Спутник, 2002.
4. Татарова Г. Г. Основы типологического анализа в социологических исследованиях / Г. Г. Татарова. – М. : Изд. дом «Высш. образование и наука», 2007.
5. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд. перераб. и доп. – М. : ИНФРА-М, 2003.
6. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

Дополнительная литература

1. Девятко И. Ф. Модели объяснения и логика социологического исследования / И. Ф. Девятко. – М. : Ин-т социологии и образования, 1996.
2. Паніотто В. І. Статистичний аналіз соціологічних даних / В. І. Паніотто, В. С. Максименко, Н. М. Марченко. – К. : Вид. дім «КМ Академія», 2004. – 270 с.
3. Толстова Ю. Н. Классификация методов анализа социологических данных / Ю. Н. Толстова // Глобализация и социальные изменения в современной России: тез. докл. и выступлений Всерос. социол. конгресса. – М. : Альфа-М, 2006. – Т. 5. – С. 78–81.

Методические рекомендации по изучению тем модуля 3

«Основы многомерного анализа категориальных переменных»

Третий модуль предполагает подробное рассмотрение различных вариантов взаимодействия переменных.

Приступая к изучению основных вопросов модуля следует обратить внимание на понятие многомерного шкалирования. Многомерное шкалирование относится к группе моделей, представляющих данные в виде совокупности точек в пространстве таким образом, что отношения в этой совокупности отражаются пространственно в геометрии модели. Данные наблюдений, имеющие различную степень связи в этой совокупности, изображаются в виде точек, раз-

деленных расстояниями, соответствующих коэффициентам связи. Многомерное шкалирование представляет сложные данные в визуальной форме, что облегчает их восприятие и интерпретацию по сравнению с данными, представленными в табличной форме. Изначально предложенное метрическое многомерное шкалирование в социологии использовалось ограниченно, поскольку оно предполагает интервальные или пропорциональные измерения. Неметрическое многомерное шкалирование представляет данные, которые просто определенным образом упорядочены с целью нанесения координат каждой точки. При таком шкалировании могут использоваться любые совокупности данных, содержащие меры сходства или различия. Этот метод все еще не применяется в социологии широко, хотя к его возможностям проявляется значительный интерес. Он оказался полезным в ходе различных исследований установок при изучении субъективных восприятий сходства и различия.

Следует запомнить, что многомерные данные содержат информацию о трех или более признаках для каждого объекта. В дополнение к той информации, которую можно извлечь из одномерных и двумерных наборов, многомерные данные можно использовать для получения информации о том, существует ли простая зависимость между этими признаками, насколько они взаимосвязаны (речь идет не только о попарной взаимосвязи признаков, но и о зависимости в совокупности), можно ли предсказать значение одной переменной на основании значений остальных.

При этом многомерное шкалирование представляет собой метод, позволяющий на основе матрицы различий между объектами построить одно-, двух- или трехмерное изображение, иллюстрирующее удаленность этих объектов друг от друга.

Таким образом, многомерный анализ – это анализ зависимостей и взаимозависимостей между несколькими переменными. Многопеременные методы используются при статистическом анализе данных, представляющих несколько переменных. Этот термин означает анализ отношений ряда независимых и зависимых переменных, в рамках которого используются методы, о которых речь пойдет ниже.

Итак, основными из методов анализа многомерных таблиц являются факторный метод, метод кластеризации и дискриминантный анализ.

Факторный анализ представляет собой метод многомерного статистического анализа, который позволяет установить скрытые закономерности в ансамбле переменных. Во введении в факторный анализ обратите внимание на следующие принципиальные моменты:

- опосредованность и относительность исходных переменных, не отражающих истинные действующие факторы;
- эти скрытые факторы могут проявляться в нескольких переменных (коррелированность);
- сила влияния фактора проявляется в диапазоне изменения переменных (дисперсия);

- основная задача анализа – найти главные действующие факторы.

Первой практически значимой задачей факторного анализа является нахождение системы существенных или действенных факторов в пространстве регистрируемых переменных. В процессе решения и развития этой задачи могут присутствовать несколько последовательных этапов:

- вычисление главных факторных компонент как новой координатной системы, расположенной по осям рассеяния анализируемых данных;
- волевой (содержательный) выбор в качестве факторов тех из главных компонент, которые отвечают за большую часть дисперсии (рассеяния) анализируемых данных;
- коррекция выделенных факторов специальными методами факторного анализа с целью достижения большей адекватности факторной модели исходным данным по ряду критериев;
- вращение выделенных факторов с целью обеспечения их лучшей проецируемости на исходные переменные для облегчения последующей предметной интерпретации;
- содержательная интерпретация факторов в предметных (физических) терминах, что является творческой задачей, выходящей за рамки формального метода, однако она может принести много полезного для дальнейшего понимания объекта исследования;
- исследование распределений объектов в их проекциях на плоскости главных факторов (в новых факторных координатах объектов).

При изучении данного аспекта темы следует знать о двух целевых направлениях факторного анализа. Первое из них – эксплораторный тип анализа (разведочный – следует посмотреть, может быть что-нибудь обнаружится интересное) и конфирматорный (подтверждающий гипотезу). Второе направление анализа часто используется на завершающей стадии практической работы и предполагает особые требования к планированию исследования по сбору первичных данных.

В первую очередь исследователь должен выработать гипотезу относительно того, какие факторы могли бы описывать предметную область. Статистически очень важно, чтобы экспериментальное исследование было достаточно широким и можно было бы выделить не менее пяти-шести гипотетических факторов, иначе будет трудно говорить об устойчивости решения. С другой стороны, потеря какого-то фактора, действительно участвующего (лежащего в основе, детерминирующего, связанного) в изучаемом процессе, может стать серьезным препятствием с точки зрения успешности исследования. Ошибки при измерении каких-либо важных факторов могут привести к неправильной трактовке очевидных взаимосвязей между измеряемыми факторами. Включение всех факторов, имеющих отношение к исследуемой области, это, в первую очередь, содержательная, а не статистическая проблема.

Далее исследователь должен выбрать переменные для наблюдения. Для каждого гипотетического фактора следует предусмотреть пять или шесть первичных наблюдаемых переменных, которые позволят измерить фактор в «чистом» виде

(исходя из предположения, что эти переменные детерминированы только этим фактором). Их называют маркерными переменными. Другими словами, маркерные переменные должны быть в высокой степени взаимосвязаны с одним и только одним фактором и иметь по нему высокие нагрузки вне зависимости от того, с помощью какого алгоритма выделялись и вращались факторы. Маркерные переменные четко определяют природу фактора. Добавление переменных, которые потенциально могут быть объяснены каким-то фактором, для того, чтобы точнее выявить его семантику, процесс гораздо более эффективный, если этот фактор с самого начала был однозначно определен маркерными переменными.

Следует учитывать также сложность переменной, характеризующую количеством факторов, с которыми эта переменная коррелирует (взаимосвязана). Переменные разной сложности, включенные в исследование, могут «связываться» в факторы, имеющие очень мало общего с исследуемым процессом. Переменные одинаковой сложности могут «связываться» между собой (иметь высокий показатель взаимосвязи) именно по этой причине, а не потому, что они относятся к одному и тому же семантическому фактору. Оценка сложности переменных – это неотъемлемая часть процесса формирования гипотез о факторах и выбора переменных для измерения факторов.

Очень важно, чтобы показатели объектов, составляющих экспериментальную выборку, покрывали возможно более широкий спектр значений как по наблюдаемым переменным, так и по латентным факторам. Действительно, если все исследуемые переменные имеют практически одинаковые значения по какому-то фактору, то взаимосвязь между наблюдаемыми переменными низка и фактор может не выделиться. Выбор объектов с ожидаемо высоким разбросом показателей по наблюдаемым переменным и латентным факторам – важная составная часть планируемого исследования.

Следует также внимательно относиться к объединению переменных для факторной обработки данных, полученных на нескольких выборках или на одной и той же, но в течение длительного временного интервала. Во-первых, объекты, различающиеся по какому-либо критерию (например, по социально-экономическому статусу), могут также характеризоваться различными факторами. Изучение групповых различий в большинстве случаев очень полезно для прояснения гипотезы. Во-вторых, структура латентного фактора может изменяться во времени для одних и тех же объектов (например, для респондентов по мере получения ими полной информации и опыта участия в исследовании). Анализ и исследование подобных различий может быть очень полезным, а объединение результатов разных групп в одну матрицу может скрыть (а не выявить) различия. Если же на различных выборках действительно получаются аналогичные факторы, то их объединение даже желательно, ибо это увеличивает статистическую устойчивость результатов.

Таким образом, факторный анализ – это метод многомерного анализа, позволяющий свести большое количество исходных переменных к значительно

меньшему числу факторов, каждый из которых объединяет исходные переменные, имеющие сходный смысл.

Кластерный анализ – группа статистических методов, используемых в рамках анализа данных с множественными переменными с целью определения их внутренней структуры. Кластерный анализ может применяться при выявлении некоторых образцов поведения на основании данных об определенном виде деятельности. Кластерный анализ используется для идентификации групп индивидов или переменных.

Кластеризация является методом совместного хранения родственных данных (таблиц). Методы кластеризации являются способом хранения в одной области памяти таблиц, связанных внешними ключами (одна основная таблица, одна или несколько подчиненных таблиц).

Дискриминантный анализ представляет собой метод, который можно применять для идентификации факторов, существенно связанных с бинарным ответом (группирующим признаком). При этом он позволяет осуществить процедуру создания формулы регрессии, на основе которой производится разбиение объектов на группы, соответствующие категориям зависимой переменной.

Таким образом, факторный анализ является наиболее значимым и часто используемым из многомерных методов, а методы кластеризации и дискриминантного анализа дают более наглядные результаты, если они применяются к данным, уже прошедшим процедуру факторизации.

Следует обратить внимание на то, что построение тримерных таблиц, как результата проведения многомерного анализа, возможно при использовании знаний по работе с пакетом «ОСА». Анализ связи между тремя переменными будет более качественным, если Вы примените свои знания о методе «уточнения и объяснения» П. Лазарсфельда. Именно он интерпретируется как метод анализа взаимосвязи с введением дополнительных, контрольных переменных. Используется для анализа элементарных таблиц сопряженности и взаимосвязей номинальных признаков.

Кроме того, анализируя многомерные таблицы, следует помнить о так называемой фальшивой корреляции. Выполняя практические задания по курсу «Анализ категориальных данных», следует иметь в виду, что фальшивая корреляция – это двусмысленная связь между переменными, не предполагающая причинной зависимости.

Таким образом, содержание данного модуля предусматривает рассмотрение вариантов взаимодействия переменных, изучение особенностей построения и анализа тримерных таблиц, использования метода «уточнения и объяснения» (П. Лазарсфельда), определения и обоснования фальшивой корреляции.

В результате изучения материала третьего модуля необходимо:

- знать сущность и особенности анализа многомерных таблиц;
- уметь привести примеры случаев категориальных измерений;
- знать особенности построения тримерных таблиц;
- уметь привести собственные примеры тримерных таблиц и осуществить их анализ;

- уметь определить характер взаимосвязи между тремя переменными;
- с использованием собственных примеров уметь осуществить анализ связи между двумя переменными;
- знать определение понятия «фальшивая корреляция»;
- посредством метода «уточнения и объяснения» П. Лазарсфельда объяснить сущность фальшивой корреляции.

При изучении тем третьего модуля «Подходы к анализу многомерных таблиц» необходимо овладеть следующими основными понятиями: многомерное шкалирование, факторный анализ, метод кластеризации, дискриминантный анализ, метод «уточнения и объяснения» П. Лазарсфельда, фальшивая корреляция.

САМОСТОЯТЕЛЬНАЯ РАБОТА

№ п/п	Название темы	Кол-во часов
1	<p>Тема 1. Дискретные и непрерывные, категориальные и категориальные, зависимые и независимые переменные</p> <p><i>Подготовить сообщения по следующим темам:</i></p> <ol style="list-style-type: none"> 1) Основные формы существования данных (ряды данных, матрица данных вида «объект-признак»); 2) Понятие измерения и шкалы в социологии; 3) Кодирование как процесс измерения в социологии; 4) Эмпирические основы изучения социальных явлений и процессов; 5) Характеристики дискретных и непрерывных, зависимых и независимых переменных (привести собственные примеры); 6) Характеристика категориальных и некатегориальных переменных (привести собственные примеры). 	4
2	<p>Тема 2. Процедуры построения, кодирования и интерпретации категориальных переменных</p> <p><i>Подготовить сообщения по следующим темам:</i></p> <ol style="list-style-type: none"> 1) Дихотомические переменные: определение сущности. 2) Метод множественной дихотомии и категориальный метод. 3) Особенности кодирования дихотомических переменных. 4) Привести примеры шкал с помощью которых могут быть измерены категориальные переменные; 5) Определить суть кодирования категориальных переменных, описать процедуру кодирования (на примерах); 6) Сравнительный анализ дихотомического и категориального методов. 	2
3	<p>Тема 3. Важные характеристики одной категориальной переменной</p> <p><i>Выполнить следующие задания:</i></p> <ol style="list-style-type: none"> 1) Для конкретных распределений, полученных по категориальным переменным, построить таблицу одномерного распределения и проанализировать данные, которые она содержит; 2) Для конкретных распределений, полученных по категориальным переменным, вычислить меры центральной тенденции; 3) Для конкретных распределений, полученных по категориальным переменным, вычислить степени вариации; 4) Описать процедуру «трансформации» категориальных переменных в неговорящие, обосновать случаи необходимости проведения этой процедуры, привести примеры. 	2

4	<p>Тема 4. Модели связи двух категориальных переменных <i>Подготовить сообщения по следующим темам:</i></p> <ol style="list-style-type: none"> 1) Методы оценки и измерения статистических взаимосвязей; 2) Методы получения различных типов процентов, их познавательные возможности; 3) Понятие множественного выбора, анализ вопросов множественного выбора; 4) Характеристика связей типа «признак-признак»; 5) Таблица сопряженности (связанности) как инструмент анализа взаимосвязи между категориальными переменными; 6) Особенности формулировки и общая характеристика. <p><i>Выполнить следующее задание:</i> С помощью пакета «ОСА» выполнить процедуру двухмерного распределения для конкретных категориальных признаков, построить двухмерную таблицу, проанализировать данные, которые она содержит.</p>	4
5	<p>Тема 5. Связь двух категориальных переменных (как разницы от статистической независимости) <i>Подготовить сообщения по следующим темам:</i></p> <ol style="list-style-type: none"> 1) Понятие статистической гипотезы, применение статистических гипотез в эмпирической социологии, критерии проверки статистических гипотез; 2) Хи-квадрат в качестве критерия проверки статистической гипотезы о независимости категориальных переменных; 3) Коэффициенты связи между двумя категориальными переменными: общая характеристика процедуры вычисления и интерпретации. 4) Процедура определения связи между двумя категориальными. <p><i>Выполнить следующие задания:</i></p> <ol style="list-style-type: none"> 1) С помощью критерия Хи-квадрат, определить соответствие заданного эмпирического распределения, полученного по категориальным переменным, теоретическому. Сделать соответствующие выводы о вероятности взаимосвязи между этими переменными; 2) С помощью пакета «ОСА», выполнить процедуру двухмерного распределения для конкретных категориальных признаков, вычислить соответствующие коэффициенты корреляции (Пирсона, Чупрова и Крамера или лямбда-Гуттмана) и предоставить их интерпретацию. 	4
6	<p>Тема 6. Варианты взаимодействия трех и большего количества категориальных переменных <i>Подготовить сообщения по следующим темам:</i></p> <ol style="list-style-type: none"> 1) Описание процедуры построения и особенности анализа 	2

	<p>трехмерных и многомерных таблиц сопряженности (связности);</p> <p>2) Примеры необходимости поиска связи между несколькими (или многими) категориальными переменными при исследовании тех или иных социальных явлений и процессов;</p> <p>3) Условная и многомерная энтропия: определение и примеры.</p>	
7	<p>Тема 7. Подходы к анализу связи двух и большего количества категориальных переменных</p> <p><i>Подготовить сообщения по следующим темам:</i></p> <p>1) Процедурные особенности причинного анализа в эмпирической социологии;</p> <p>2) Факторный анализ как метод множественной корреляции;</p> <p>3) Кластерный анализ в социологии: сущность и описание процедуры;</p> <p>4) Суть, процедурная специфика и примеры необходимости применения метода «кластеров на факторах».</p> <p><i>Выполнить следующие задания:</i></p> <p>1) С помощью пакеты ОСА выполнить процедуру факторного анализа совокупности признаков (можно, чтобы к этой совокупности входили и не категориальные признаки), проанализировать и проинтерпретировать полученные результаты, сделать социологические выводы;</p> <p>2) Представить графически результаты факторного анализа;</p> <p>3) ЗАЧЕТНОЕ индивидуальное задание, ответы на вопросы (из списка вопросов к зачету, который будет дан нже)</p>	8
Итого		72

ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ

Индивидуальное научно-исследовательское задание (далее ИНДЗ) необходимо для систематизации, закрепления и расширения теоретических и практических знаний по дисциплине «Анализ категориальных данных».

Целью выполнения ИНДЗ является закрепление и углубление теоретического материала и развитие практических навыков по вопросам анализа категориальных данных, полученных в результате социологического исследования; усвоение правил применения соответствующих методов анализа в конкретных случаях работы с данными; изучение зарубежного и отечественного опыта анализа категориальных данных в социологии.

При выполнении ИНДЗ слушатель должен продемонстрировать умение в сфере научно-исследовательской деятельности, умение использовать соответствующие (учитывая поставленные социологические задачи) методы анализа данных, умение творческого и оригинального решения актуальных исследовательских проблем.

ИНДЗ слушатели выполняют самостоятельно, с проведением консультаций с преподавателем дисциплины, согласно графику учебного процесса.

ИНДЗ предполагает содержание следующих элементов научного исследования:

- практическая значимость;
- комплексный и системный подходы к решению поставленной задачи;
- использование современной исследовательской методологии по теме индивидуального задания;
- применение творческого подхода и отражение собственного видения решения поставленной задачи.

Требования к оформлению ИНДЗ

1. Объем работы – до 8 печатных страниц;
2. Шрифт – Times New Roman, 14,
3. Полуторный интервал;
4. Поля: 20 мм со всех сторон.

Темы для написания ИНДЗ

Тему ИНДЗ слушатель выбирает самостоятельно, ориентируясь на варианты, предложенные преподавателем. Темы индивидуальных работ не должны повторяться. Ниже приведен перечень вариантов тем ИНДЗ. Предложенные формулировки тем не являются неизменными. Слушатель может перефразировать формулировку заинтересовавшей его темы, тем самым сузив или расширив ее. Кроме того, слушатель может предложить свой собственный вариант темы ИНДЗ, если предложенные преподавателем варианты не отражают его индивидуальные научные интересы.

ВАРИАНТЫ ТЕМ ДЛЯ НАПИСАНИЯ ИНДЗ (КАК ЗАЧЕТНОГО ЗАДАНИЯ)

1. Эмпирические основы изучения социальных явлений.
2. Основные цели и задачи анализа данных (в частности, категориальных, на примерах).
3. Развернутая сравнительная характеристика дискретных и непрерывных, категориальных и некатегориальных, зависимых и независимых переменных (на примерах).
4. Характеристика процедур классификации и кластеризации, сходство и различие задач.
5. Типы шкал, используемых для анализа категориальных данных.
6. Определение и кодирование категориальных переменных: общие характеристики процедуры и реальные примеры.
7. Определение фиктивных переменных и их интерпретация.
8. Виды распределений, используемые для построения статистических критериев.
9. Представление одномерной случайной величины в выборочном социологическом исследовании, принципы и особенности анализа одномерного распределения (на примерах).
10. Одномерные частотные распределения как метод анализа категориальных переменных.
11. Статистические характеристики одномерных частотных распределений.
12. Способы установления взаимосвязи между признаками в эмпирической социологии.
13. Методы оценки корреляции между категориальными переменными.
14. Смысл таблиц сопряженности при установлении взаимосвязи категориальных переменных.
15. Параметрический и непараметрический критерии оценки связи между переменными: основные отличия.
16. Модели множественной корреляции.
17. Меры центральной тенденции и вариации в случае категориальных переменных: особенности определения и интерпретации (на примерах).
18. Определение энтропии, ее социологический смысл.
19. Графическое представление связи между несколькими (или многими) категориальными переменными.
20. Особенности построения и примеры анализа таблиц сопряженности (двухмерного распределения) между двумя категориальными переменными.
21. Особенности построения и примеры анализа многомерных таблиц сопряженности.
22. Понятие статистической гипотезы, особенности формулировки и отличия от обычной «исследовательской» гипотезы (на примерах).

23. Характеристика критериев проверки статистических гипотез, используемых в практической деятельности социолога.
24. Критерий Хи-квадрат: общая характеристика и примеры проверки с его помощью гипотез о зависимости/независимости категориальных переменных.
25. Коэффициенты связи, основанные на критерии Хи-квадрат: характеристика, алгоритмы вычисления и общие принципы интерпретации (на примерах).
26. Коэффициент Чупрова: характеристика, социологическое значение, примеры вычисления.
27. Коэффициент Крамера: характеристика, социологическое значение, примеры вычисления.
28. Коэффициент Пирсона: характеристика, социологическое значение, примеры вычисления.
29. Коэффициент лямбда-Гудмана: характеристика, социологическое значение, примеры вычисления.
30. Многомерные таблицы распределений: особенности анализа и конкретные примеры (по отношению к категориальным переменным).
31. Определение и характеристики условной и многомерной энтропии.
32. Суть, процедура и примеры факторного анализа в эмпирической социологии.
33. Интерпретация коэффициентов матрицы факторных нагрузок при факторном анализе категориальных данных в социологии (на примерах).
34. Кластерный анализ: постановка задачи, описание процедуры, особенности интерпретации данных.
35. Принципы кластеризации (с приведением примеров).
36. Характер взаимосвязи между тремя переменными.
37. Метод «уточнения и пояснения» П. Лазарсфельда.
38. Понятие фальшивой корреляции и ее суть.

ВОПРОСЫ К ЗАЧЕТУ ПО КУРСУ

1. Эмпирические основы для изучения социальных явлений.
2. Основные цели и задачи анализа данных.
3. Различие подходов при использовании универсальных и специализированных пакетов анализа данных.
4. Характеристика дискретных и непрерывных, зависимых и независимых переменных.
5. Проблемы при выполнении регрессионного анализа.
6. Сходство и различие задач классификации и кластеризации.
7. Методы анализа данных.
8. Типы шкал, используемых для анализа данных.
9. Процесс определения и кодирования переменных.
10. Интерпретация фиктивных переменных.
11. Виды распределений, используемые для построения статистических критериев.
12. Представление одномерной случайной величины в выборочном социологическом исследовании.
13. Кумулята и проблема пропущенных значений.
14. Одномерные частотные распределения как метод анализа социологических и опросных данных.
15. Статистические характеристики одномерных частотных распределений.
16. Способы установления взаимосвязи данных.
17. Методы оценки корреляций.
18. Смысл таблиц сопряженности при установлении взаимосвязи признаков.
19. Параметрический и непараметрический критерий: основные отличия.
20. Сущность метода дисперсионного анализа.
21. Основные непараметрические критерии и особенности их использования при проверке статистических гипотез.
22. Модель множественной линейной регрессии.
23. Коэффициенты регрессионного уравнения.
24. Преимущества стандартизованной модели регрессии.
25. Меры центральной тенденции и соответствующие им модели.
26. Меры разброса и соответствующие им модели.
27. Определение энтропии, ее социологический смысл.
28. Сущность и значение коэффициента детерминации.
29. Регрессионная модель для рядов динамики.
30. Особенности графической интерпретации дискриминантного анализа для случая двух переменных.
31. Факторные и результативные переменные дискриминантного анализа.
32. Критерии отбора дискриминирующих переменных.
33. Понятие таблицы сопряженности.
34. Характеристика связи типа «признак-признак».

35. Коэффициенты связи, основанные на критерии «Хи-квадрат».
36. Понимание отсутствия связи между признаками как статистическая независимость.
37. Коэффициент Чупрова, его характеристика и социологическое значение.
38. Коэффициент Крамера, его характеристика и социологическое значение.
39. Сущность анализа многомерных таблиц.
40. Условная и многомерная энтропия.
41. Случай категориальных измерений.
42. Сущность факторного анализа.
43. Сущность коэффициентов матрицы факторных нагрузок.
44. Сформулируйте постановку задачи кластерного анализа.
45. Принципы кластеризации в методе иерархического кластерного анализа.
46. Образование кластеров в методе К-средних.
47. Особенности построения тримерных таблиц.
48. Характер взаимосвязи между тремя переменными.
49. Метод «уточнения и объяснения» П. Лазарсфельда.
50. Понятие фальшивой корреляции и ее сущность.

СЛОВАРЬ ТЕРМИНОВ ПО КУРСУ «АНАЛИЗ КАТЕГОРИАЛЬНЫХ ДАННЫХ»

АНАЛИЗ ДАННЫХ	– область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных. В социологии анализ данных применяется в основном к анализу данных опросов.
БАЗА ДАННЫХ	– это информационная модель предметной области, совокупность взаимосвязанных, хранящихся вместе данных при наличии такой минимальной избыточности, которая допускает их использование оптимальным образом для одного или нескольких приложений. Данные (файлы) хранятся во внешней памяти и используются в качестве входной информации для решения задач.
БИНОМИАЛЬНЫЙ КРИТЕРИЙ	– непараметрический критерий, определяющий степень близости эмпирического распределения бинарной переменной к биномиальному распределению, для которого частоты двух категорий равны.
ВАРИАЦИОННЫЙ РЯД	– последовательность значений наблюдаемой величины, расположенных в порядке возрастания; – совокупность каких-либо величин, расположенных в порядке их возрастания; определяется указанием различных значений, входящих в него величин и числом членов ряда.
ВАРИАЦИЯ	– разброс значений; за меру вариации может быть принята дисперсия.
ГИСТОГРАММА	– диаграмма, которая показывает относительное распределение частоты непрерывной переменной путем присоединения столбчатых диаграмм. Область столбчатой диаграммы пропорциональна относительной частоте в интервале, определенном границами этой диаграммы; – столбиковая диаграмма для отображения распределения частот по категориям (диапазонам значений) переменной. Горизонтальная ось графика соответствует значениям переменной, а вертикальная – частотам.

**ДИАГРАММА
РАСSEИВАНИЯ**

– график для анализа связи между двумя переменными, на котором каждый объект представляет собой точку. Положение точки задано парой значений двух переменных для данного объекта.

**ДИАГРАММА
РЕГРЕССИИ**

– диаграмма разброса, включающая сдвиги точек от линии регрессии по вертикальной оси.

ДИНАМИЧЕСКИЙ РЯД

– ряд последовательных значений какого-либо статистического показателя, изменяющихся во времени; широко используется при обработке материалов социологических исследований.

**ДИСКРЕТНАЯ
ПЕРЕМЕННАЯ**

– числовая переменная, которая может принимать значения только в целых числах;
– переменная, имеющая только два возможных значения.

**ДИСКРИМИНАНТНЫЙ
АНАЛИЗ**

– метод, который можно применять для идентификации факторов, существенно связанных с бинарным ответом (группирующим признаком);
– процедура создания формулы регрессии, на основе которой производится разбиение объектов на группы, соответствующие категориям зависимой переменной.

ДИСПЕРСИЯ

– мера рассеяния, равная квадрату стандартного отклонения.

**ДИХОТОМИЧЕСКАЯ
ПЕРЕМЕННАЯ (бинарная)**

– качественная переменная с двумя категориями.

**ЗАВИСИМАЯ
ПЕРЕМЕННАЯ**

– переменная, на которую оказывает влияние другой фактор (независимая переменная);
– переменная, которая в моделях причинных связей измеряет следствие. Применяется в дисперсионном регрессионном анализе. К зависимой переменной предъявляется требование надежности, т. е. повторения результата при тех же условиях (возможно с теми же участниками). Если зависимая переменная не изменяется при изменении независимой переменной, то говорят о «нулевом результате»;
– та, которая была вызвана, зависела или выступала в качестве функций других явлений.

ИЗМЕРЕНИЕ

- определение соотношения какой-либо величины с однородной ей величиной, принимаемой за единицу меры;
- процедура, при помощи которой объекты исследования, рассматриваемые как носители определенных отношений между ними и как таковые составляющие эмпирическую систему, отображаются в некоторую математическую систему с соответствующими отношениями между ее элементами. Немалую роль при осуществлении измерения играет обеспечение его надежности, обоснованности, точности.

КАТЕГОРИАЛЬНАЯ ПЕРЕМЕННАЯ (номинативная)

- переменная, каждое значение которой указывает на принадлежность объекта к определенной группе (категории). Категориальная переменная не является количественной; она разделяет все объекты на непересекающиеся группы по определенному признаку (пол, хобби, класс и пр.), но не позволяет сравнивать объекты по уровню выраженности этого признака;
- переменная, чьи категории (уровни, градации) не имеют естественного упорядочения.

КЛАССИФИКАЦИЯ

- построение системы элементарных понятий и определяющих связей между ними.

КЛАСТЕРИЗАЦИЯ

- метод совместного хранения родственных данных (таблиц);
- способ хранения в одной области памяти таблиц, связанных внешними ключами (одна родительская таблица, одна или несколько подчиненных таблиц).

КЛАСТЕРНЫЙ АНАЛИЗ

- группа статистических методов, используемых в рамках анализа данных с множественными переменными с целью определения их внутренней структуры. Кластерный анализ может применяться при выявлении некоторых образцов поведения на основании данных об определенном виде деятельности. Кластерный анализ используется для идентификации групп индивидов или переменных.

КОВАРИАЦИОННЫЙ АНАЛИЗ

– специальная форма анализа дисперсии, которая сравнивает значения зависимой переменной между группами респондентов после корректировки эффекта одной или двух независимых переменных.

КОДИРОВАНИЕ ПЕРЕМЕННЫХ

– это процесс преобразования количественных и качественных данных в числовые или буквенные символы;

– операция отождествления символов или групп символов одного кода с символами или группами символов другого кода;

– процесс перевода необработанных данных исследования в ту форму, которая может использоваться при подсчетах (обычно производимых с помощью компьютера). Осуществляется посредством разбивки данных на категории и приписывания каждой категории числового значения;

– подготовка информации к формализованной обработке.

КОЛИЧЕСТВЕННАЯ ПЕРЕМЕННАЯ (числовая)

– переменная, которая принимает дискретные или непрерывные величины. Значения количественной переменной (в отличие от категориальной) отражают уровень выраженности у объектов соответствующего признака, измеренного в метрической или порядковой шкале.

КОРРЕЛЯЦИЯ

– понятие математической статистики, характеризующее вероятностную или статистическую зависимость;

– мера степени и направления связи между значениями двух переменных;

– показатель наличия линейной зависимости двух переменных. Может быть использован только для интервальных переменных и не может быть интерпретирован как «степень взаимосвязи» переменных;

– взаимная связь, взаимозависимость эмпирических признаков, свойств, черт, показателей;

– взаимосвязь между переменными, не предполагающая причинной зависимости. Как правило, термин применяют к количественным, порядко-

вым и дихотомическим переменным. Для количественных и порядковых переменных прямая (положительная) корреляция означает одновременное увеличение или уменьшение их значений; обратная (отрицательная) корреляция.

– возрастание значений одной переменной при уменьшении значений другой. Для дихотомических признаков прямая корреляция означает, что они чаще появляются или не появляются вместе, чем врозь; обратная – что признаки чаще появляются порознь, чем совместно. Для измерения корреляционной связи используются специальные меры связи.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ПИРСОНА

– количественная мера связи двух переменных, обозначаемая символом r и принимающая значения между -1 и $+1$; отражает степень, до которой точки в двумерном графике согласуются с прямой линией;

– мера корреляции, идеально подходящая для двух непрерывных (метрических) переменных.

КОЭФФИЦИЕНТ КРАМЕРА

– мера ассоциации между значениями двух категориальных переменных. Значение коэффициента всегда варьируется от 0 до 1 и интерпретируется по аналогии с коэффициентом корреляции (исключая отсутствие отрицательных значений). При отсутствии статистической связи между переменными значение коэффициента равно 0; при полной связи (когда значение одной переменной полностью определяется значением второй переменной) значение коэффициента достигает 1;

– мера связи двух номинальных переменных на основе критерия Хи-квадрат. Применяется к таблицам сопряженности произвольной размерности.

КОЭФФИЦИЕНТ ЧУПРОВА

– тетракорический коэффициент корреляции для дихотомических переменных, основанных на нормальных распределениях. Изменяется в диапазоне от 0 до 1. Строже оценивает тесноту связи.

КРУГОВОЙ ГРАФИК

– диаграмма, показывающая частотное распределение количественной или дискретной переменной. Круг делится на секции по одной для каждой категории; площадь каждой секции пропорциональна частоте в этой категории.

МАТРИЦА ДАННЫХ

– прямоугольная таблица, предназначенная для представления и хранения данных социологического исследования. В качестве признаков выступают сжатые факторы (главные компоненты).

МАТРИЦА ДАННЫХ ВИДА «ОБЪЕКТ- ПРИЗНАК»

– прямоугольная таблица, используемая в качестве стандартного способа представления данных в программных средствах, предназначенных для статистической обработки и анализа данных. Каждая строка такой матрицы отводится одному объекту, каждый столбец – одному признаку; на пересечении строки и столбца указывается значение данной переменной для соответствующего объекта.

МЕДИАНА

– мера положения, которая является срединным (расположенным посередине) значением упорядоченных наблюдений;
– значение переменной, делящее упорядоченное множество всех значений выборки ровно пополам: у половины объектов выборки значения переменной больше, а у другой половины – меньше медианы.

МЕРА ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

– характеристика совокупности переменных (признаков), указывающих на наиболее типичный, репрезентативный для изучаемой выборки результат. В социологии мерами центральной тенденции принято считать моду, медиану и среднее.

МЕТОД «УТОЧНЕНИЯ И ОБЪЯСНЕНИЯ» (П. ЛАЗАРСФЕЛЬД)

– метод анализа взаимосвязи с введением дополнительных, контрольных переменных. Применяется для анализа элементарных таблиц сопряженности и взаимосвязей номинальных признаков.

<p>МЕТРИЧЕСКАЯ ПЕРЕМЕННАЯ</p>	<p>– количественная переменная, соответствующая измерению признака в шкале интервалов или отношений. В отличие от ранговой (порядковой) переменной, при сравнении объектов позволяет судить не только о том, больше или меньше выражен признак, но и о том, насколько больше (меньше) он выражен.</p>
<p>МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ</p>	<p>– метод, позволяющий на основе матрицы различий между объектами построить одно-, двух- или трехмерное изображение, иллюстрирующее удаленность этих объектов друг от друга.</p>
<p>МНОГОМЕРНЫЙ АНАЛИЗ ДАННЫХ</p>	<p>– анализ зависимостей и взаимозависимостей между несколькими переменными.</p>
<p>МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ</p>	<p>– модель линейной регрессии, в которой одна зависимая переменная и две независимые переменные или более.</p>
<p>МОДА</p>	<p>– величина отдельной переменной, которая наиболее часто появляется в группе данных; – наиболее часто повторяющееся значение распределения переменной.</p>
<p>НАКОПЛЕННАЯ ЧАСТОТА</p>	<p>– суммарное число объектов, имеющих значение переменной, не большее, чем указано.</p>
<p>НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ (объясняющая)</p>	<p>– переменная, которая в моделях причинных связей измеряет причину; – переменная (обычно обозначаемая как X), которая применяется для прогноза зависимой переменной в регрессионном анализе. Также называется предикторной переменной или ковариатой; – наблюдаемые явления, параметры и характеристики, которые способны воздействовать на возникновение других явлений.</p>
<p>НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ</p>	<p>– серия критериев, каждый из которых применяется без предварительных допущений относительно нормальности распределения. Непараметрические критерии основаны на ранжировании, попарных сравнениях и других средствах, не требующих нормальности распределения переменных.</p>

НЕПРЕРЫВНАЯ ПЕРЕМЕННАЯ	<ul style="list-style-type: none"> – переменная, у которой может быть любое численное значение; – количественная переменная, которая постоянно возрастает в своем значении с непрерывным рядом (серией) градаций возможных величин.
НЕСТРУКТУРИРОВАННОЕ КОДИРОВАНИЕ	– операция преобразования количественных и качественных данных в числовые или буквенные символы, используемая для данных, полученных исследователем из вторичных источников. Примером применения является контент-анализ.
ОБРАТНАЯ СВЯЗЬ	– взаимосвязь, в которой значения переменных возрастают по мере уменьшения значений другой переменной.
ОДНОМЕРНАЯ СТАТИСТИКА	– статистика, описывающая одну переменную.
ОДНОМЕРНАЯ ТАБЛИЦА	– простейший тип перекрестной таблицы; используется для анализа распределения данных одного типа по одному фактору.
ОРДИНАЛЬНАЯ ПЕРЕМЕННАЯ (ранговая)	– переменная, чьи категории расположены упорядоченным образом.
ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ	– критерии, применяемые в предположении о нормальном распределении переменных в генеральной совокупности.
ПЕРЕМЕННАЯ	<ul style="list-style-type: none"> – любая величина, которая изменяется; – признак, который регистрируется для каждого из объектов.
ПЕРЕМЕННАЯ КАЧЕСТВЕННАЯ	– переменная, состоящая из различных категорий, а не из числовых единиц.
ПЕРЕМЕННАЯ КОЛИЧЕСТВЕННАЯ	– переменная, состоящая из ряда числовых единиц.
ПРЯМАЯ СВЯЗЬ	– взаимосвязь, в которой значение одной переменной возрастает по мере увеличения другой переменной.

<p>РАНГОВЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ СПИРМЕНА</p>	<p>– непараметрическая альтернатива коэффициенту корреляции Пирсона; он обеспечивает измерение связи между двумя переменными.</p>
<p>СЛУЧАЙНАЯ ПЕРЕМЕННАЯ</p>	<p>– величина, которая может принимать любое из группы взаимоисключающих значений с данной вероятностью.</p>
<p>СОЦИАЛЬНАЯ ЭНТРОПИЯ</p>	<p>– мера отклонения социальной системы либо ее подсистемы от эталонного (нормального, ожидаемого) состояния, когда отклонение проявляется в снижении уровня организации, эффективности функционирования, темпов развития системы.</p>
<p>СТАТИСТИЧЕСКАЯ ЗАВИСИМОСТЬ</p>	<p>– взаимосвязь двух признаков или величин; – такая связь двух случайных величин, при которой распределение вероятностей одной из них зависит от того, какие возможные значения приняла другая величина.</p>
<p>СТАТИСТИЧЕСКАЯ НЕЗАВИСИМОСТЬ</p>	<p>– соотношение между двумя переменными, характеризующееся отсутствием статистической связи.</p>
<p>СТОЛБИКОВАЯ ДИАГРАММА</p>	<p>– график распределения частот по категориям (значениям) переменной. Каждый столбец на графике соответствует одному значению признака, а его высота пропорциональна частоте встречаемости этого значения; – диаграмма, которая показывает распределение качественной или дискретной переменной горизонтально или вертикально для каждой категории, причем длина пропорциональна относительной частоте этой категории.</p>
<p>СТРУКТУРИРОВАННОЕ КОДИРОВАНИЕ</p>	<p>– операция преобразования количественных и качественных данных в числовые или буквенные символы, используемая для первичных данных. Область применения – анализ методического инструментария исследования.</p>

ТАБЛИЦА	– свод, перечень сведений, числовых данных о каких-либо объектах, расположенных в определенном порядке по графам, приведенным в определенную систему.
ТАБЛИЦА РАСПРЕДЕЛЕНИЯ (таблица частот)	– таблица, устанавливающая соотношение между категориями (значениями) признака и частотами их встречаемости.
ТАБЛИЦА СОПРЯЖЕННОСТИ (контингенции, кросстабу- ляции)	– таблица совместного распределения частот для двух категориальных или дискретных переменных; строки соответствуют категориям (значениям) одной, а столбцы – другой переменной; – средство представления совместного распределения двух переменных, предназначенное для исследования связи между ними. Таблица сопряженности является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения; – двухвходовая таблица, в которую вводят частоты; – форма представления данных об объекте социологического исследования на основе группировки двух или более признаков по принципу их сочетаемости. Применяются для определения процентного распределения градаций признака в выборке.
ТОЧЕЧНЫЙ ГРАФИК	– диаграмма, в которой каждое наблюдение за переменной представлено одной точкой на горизонтальной (или вертикальной) линии.
ФАКТОР	– в факторном анализе объединение нескольких переменных, чья взаимная корреляция исчерпывает определенную долю общей дисперсии. После процедуры вращения каждый фактор интерпретируется как некоторая общая причина взаимосвязи группы переменных.

ФАКТОРНЫЙ АНАЛИЗ	<ul style="list-style-type: none"> – метод, позволяющий свести большое количество исходных переменных к значительно меньшему числу факторов, каждый из которых объединяет исходные переменные, имеющие сходный смысл; – группа методов многомерного статистического анализа, которые позволяют установить скрытые закономерности в ансамбле переменных.
ФАЛЬШИВАЯ КОРРЕЛЯЦИЯ	– двусмысленная связь между переменными, не предполагающая причинной зависимости.
ФИКТИВНАЯ ПЕРЕМЕННАЯ	– переменная, кодируемая значениями «0» или «1». Используется для определения дихотомических данных, а также для определения характеристик, которые не могут быть оценены количественно.
ФОРЕСТ-ГРАФИК	– диаграмма, применяемая в мета-анализе и показывающая оцененный эффект в каждом исследовании и их средние (с доверительными интервалами).
ФУНКЦИЯ «ХИ-КВАДРАТ»	<ul style="list-style-type: none"> – критерий статистической значимости связи между двумя номинальными переменными; – распределение Пирсона; вытянутое вправо непрерывное распределение, характеризующееся степенями его свободы; полезно для анализа качественных данных. Критерий «Хи-квадрат» проверяет нулевую гипотезу о том, что нет связи между факторами, которые определяют таблицу сопряженности. Применяется также для тестирования разницы в долях (пропорциях).
ЧАСТОТА	– число тех или иных событий.
ЧАСТОТНОЕ РАСПРЕДЕЛЕНИЕ	– распределение, которое показывает частоту появления каждого возможного наблюдения, класса наблюдений или категории.
ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ	– наблюдаемое (фактическое) распределение переменной.
ЭНТРОПИЯ	– мера отклонения от эталонного (нормального, ожидаемого) результата.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
КОНТРОЛЬ ПОЛУЧЕННЫХ ЗНАНИЙ.....	5
ТЕМАТИЧЕСКИЙ ПЛАН КУРСА.....	6
ПРОГРАММА КУРСА	7
СПИСОК РЕКОМЕНДУЕМЫХ ЛИТЕРАТУРНЫХ И ИНФОРМАЦИОННЫХ ИСТОЧНИКОВ.....	10
ПЛАНЫ СЕМИНАРСКИХ ЗАНЯТИЙ И МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ.....	13
Методические рекомендации по изучению тем модуля 1 <i>«Анализ одной категориальной переменной»</i>	14
Методические рекомендации по изучению тем модуля 2 <i>«Модели связи между двумя категориальными переменными»</i>	20
Методические рекомендации по изучению тем модуля 3 <i>«Основы многомерного анализа категориальных переменных»</i>	26
САМОСТОЯТЕЛЬНАЯ РАБОТА.....	32
ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ.....	34
ВАРИАНТЫ ТЕМ ДЛЯ НАПИСАНИЯ ИНДЗ (КАК ЗАЧЕТНОГО ЗАДАНИЯ)	36
ВОПРОСЫ К ЗАЧЕТУ ПО КУРСУ	38
СЛОВАРЬ ТЕРМИНОВ ПО КУРСУ «АНАЛИЗ КАТЕГОРИАЛЬНЫХ ДАННЫХ»	40

Навчальне видання

АНАЛІЗ КАТЕГОРІАЛЬНИХ ДАНИХ

Методичні рекомендації
для слухачів факультету післядипломної освіти,
які навчаються за спеціальністю 7.03010101 – Соціологія

(російською мовою)

Автор-упорядник НЕЧИТАЙЛО Ірина Сергіївна

В авторській редакції
Комп'ютерний набір *І. С. Нечитайло*

Підписано до друку 26.05.2015. Формат 60×84/16.
Папір офсетний. Гарнітура «Таймс».
Ум. друк. арк. 3,02. Обл.-вид. арк. 3,21.
Тираж 60 пр. Зам. №

План 2014/15 навч. р., поз. № 16 в переліку робіт кафедри

Видавництво
Народної української академії
Свідоцтво № 1153 від 16.12.2002.

Надруковано у видавництві
Народної української академії

Україна, 61000, Харків, МСП, вул. Лермонтовська, 27.