



НАРОДНАЯ УКРАИНСКАЯ АКАДЕМИЯ

МАТЕМАТИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ В СОЦИОЛОГИИ

Методические рекомендации для студентов,
обучающихся по специальности 054 – Социология

Издательство НУА

НАРОДНАЯ УКРАИНСКАЯ АКАДЕМИЯ

**МАТЕМАТИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ
В СОЦИОЛОГИИ**

Методические рекомендации для студентов,
обучающихся по специальности 054 – Социология

Харьков
Издательство НУА
2017

УДК 316:[51:371] (072+075.8)
ББК 60.5в631р30-2
М34

*Утверждено на заседании кафедры социологии
Народной украинской академии.
Протокол № 17 от 8.05.2017*

Автор-составитель канд. социол. наук, доц. *И. С. Нечитайло*
Рецензент канд. физ.-мат. наук, доц. *Е. В. Свищева*

Дане видання складено згідно з вимогами щодо організації навчального процесу в кредитно-модульній системі підготовки спеціалістів і включає матеріали, які допоможуть студентам закріпити теоретичні знання з математичних методів в соціології та розвинути навички їх використання на практиці в системі роботи зі спеціальності «Соціологія».

М34

Математико-статистический анализ в социологии : метод. рекомендации для студентов, обучающихся по специальности 054 – Социология / Нар. укр. акад., [каф. социологии ; авт.-сост. И. С. Нечитайло]. – Харьков : Изд-во НУА, 2017. – 67 с.

Настоящее издание составлено в соответствии с требованиями по организации учебного процесса в системе подготовки бакалавров и включает материалы, позволяющие студентам закрепить теоретические знания по применению математических методов в социологии и развить навыки их использования на практике в системе работы по специальности «Социология».

УДК 316:[51:371] (072+075.8)
ББК 60.5в631р30-2

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
Рейтинговая система по курсу. Распределение баллов по содержательным модулям курса.....	6
Контроль выполнения индивидуального учебного плана студента.....	8
ТЕМАТИЧЕСКИЙ ПЛАН КУРСА.....	9
УЧЕБНАЯ ПРОГРАММА КУРСА «Математико-статистический анализ в социологии».....	11
МОДУЛЬ 1. Применение процедур математической статистики в эмпирической социологии. Числовые характеристики выборки.....	18
Тема 1. Основные принципы измерения в социологии. Типы переменных.....	18
Тема 2. Перекрестная классификация. Графическое представление социологической информации.....	20
Тема 3. Характеристики положения (среднее арифметическое, мода, медиана и др.).....	23
Тема 4. Характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации).....	27
Тема 5. Проверка процедуры первичного измерения на надежность с использованием математических методов.....	30
МОДУЛЬ 2. Измерение связи между признаками с использованием математических методов.....	34
Тема 6. Нормальное распределение как модель вариации. Критерий линейной взаимосвязи.....	34
Тема 7. Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных.....	38
Тема 8. Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии.....	40
Тема 9. Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция.....	44
Тема 10. Корреляция рангов: коэффициент r_s	48
Тема 11. Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции.....	50
МОДУЛЬ 3. Математические процедуры в сетевом анализе и анализе социометрических данных.....	52
Тема 12. Элементы теории графов в сетевом анализе.....	52
Тема 13. Обработка данных в социометрическом исследовании: вычисление индексов, построение социограмм.....	55
ТИПОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ И ИНДИВИДУАЛЬНОЙ РАБОТЫ СТУДЕНТОВ.....	59
Перечень возможных тем эссе и сообщений по курсу.....	62
Перечень вопросов для подготовки к экзамену.....	63

ВВЕДЕНИЕ

Пожалуй, согласимся с высказыванием Г. Г. Татаровой¹ о том, что у студентов, специализирующихся в области социологии, наблюдается «синдром аллергии на формулы, графики и т. д.». Вряд ли найдется преподаватель, читающий математические курсы для «студентов-гуманитариев», который захочет оспаривать это высказывание. Вместе с тем очевиден тот факт, что именно математическая грамотность является неотъемлемой составной частью профессиональной подготовки каждого специалиста, имеющего дело с анализом массовых явлений, будь то социальные, экономические и др. Социологическое изучение любого сложного процесса или явления не может ограничиться чистым теоретизированием. В противном случае социология рискует слиться с той же социальной философией и другими «пограничными» науками. Проникновение математических методов в разнообразные сферы человеческой деятельности предоставляет социологии весьма эффективные средства для исследования сложных социальных объектов. Необходимо понимать, что подобные исследования требуют определенного математического формализма для того, чтобы их результаты могли считаться достоверными, а выводы обоснованными. Это особенно актуально на сегодняшний день, когда насущная необходимость получения адекватной и однозначной информации социологического характера систематически возрастает. В то же время все чаще приходится сталкиваться с сомнениями относительно результатов тех или иных социологических опросов. И такие сомнения не безосновательны. Проблема в том, что многие из подобных исследований носят далеко не научный характер и проводятся на низком профессиональном уровне. Для обеспечения высокого качества результатов исследовательской деятельности социологу необходимы математические знания.

Необходимость использования разнообразных математических методов в социологии диктуется нынешним этапом развития социальных процессов в нашей стране, динамикой социальных трансформаций. Современные задачи социального управления, прогнозирования и планирования невозможно решать без глубоких знаний в области сбора, обработки и анализа социологической информации. При этом необходимо использовать устойчивые математические методы, имеющие и без того довольно широкие границы применения.

Именно поэтому в настоящее время от работника, занятого в любой области, связанной с изучением массовых социальных явлений, в том числе и от социолога, требуется определенный уровень математической грамотности. Невозможно успешно специализироваться в области социологических дисциплин без усвоения математико-статистического курса. В связи с этим для студентов большое значение имеет ознакомление с общими категориями, принципами и методологией математического анализа. Для того чтобы студенты высших учебных заведений обладали необходимыми знаниями и умениями, в образова-

¹ Татарова Галина Галеевна – известный специалист в области применения математико-статистических методов в социологии, доктор социологических наук, профессор.

тельную программу введены специальные дисциплины, позволяющие их получить, одной из которых является «Математико-статистический анализ в социологии».

В целом данный курс направлен на развитие у студентов навыков применения математических знаний при работе с эмпирическим материалом. В своей структуре он имеет три модуля.

Первый модуль посвящен детальному рассмотрению методов математической статистики, применимых в эмпирической социологии. Особое внимание при этом уделяется измерительным процедурам и определению числовых характеристик выборочной совокупности.

Второй модуль посвящен изучению математических методов, позволяющих определить связь между признаками, ее тип, характер и т. п. Изучаются различные меры взаимосвязи для количественных и качественных переменных.

Третий модуль знакомит студентов с теми математическими процедурами, которые могут быть эффективно использованы в исследованиях малых социальных групп, а также в сетевом анализе.

В результате изучения курса студенты приобретают знания относительно теоретических законов распределения выборочной совокупности и навыки вычисления количественных характеристик социальных объектов, а также интерпретации полученных данных, определения взаимосвязи между различными признаками, характеризующими тот или иной процесс или явление, визуализации и сравнения особенностей протекания этих процессов и развития явлений между собой.

Все вычисления проводятся с использованием компьютерной техники, в том числе при помощи программы «EXCEL» – для оперативного овладения основными задачами курса и достижения поставленной цели.

ЦЕЛЬ И ЗАДАЧИ КУРСА

Цель данного курса – формирование целостного видения логики исследовательского процесса, связанного с анализом социологических данных, путем усвоения возможностей современных математических методов при их обработке и интерпретации. Достижение поставленной цели видится возможным путем решения следующих *задач*:

- усвоить ряд математических процедур, применение которых представляется целесообразным и функциональным в эмпирической социологии;
- научиться вычислять числовые характеристики выборки;
- научиться измерять связи между признаками, усвоить коэффициенты корреляции, вычисление которых наиболее эффективно при анализе данных социологического исследования, уметь интерпретировать их значения;
- определить возможности применения математических процедур при анализе данных, собранных посредством методов, не являющихся собственно социологическими, научиться вычислять и интерпретировать значимые для исследования социальных явлений и процессов коэффициенты и индексы.

Междисциплинарные связи

Курс «Математико-статистический анализ в социологии» основывается на предварительном освоении студентами таких учебных дисциплин, как «Введение в специальность», «Математика для социологов», «Методология и методы социологического исследования», «Методы компьютерной обработки социологических данных», «Социальная психология» и др. Изучение данного курса должно обеспечить последующее освоение более сложных процедур, осуществляемых с эмпирическими данными, например, в рамках такого курса, как «Социологический практикум».

РЕЙТИНГОВАЯ СИСТЕМА ПО КУРСУ «Математико-статистический анализ в социологии»

№ п/п	Виды работ	Кол-во баллов за вид работы	Кол-во вида работы за учебный курс	Максимальное кол-во баллов
<i>Основные виды работы, осуществляемой студентами по курсу</i>				
1.	Решение задач на практическом занятии, в т. ч. выполнение индивидуальных и самостоятельных заданий	3–5	21	105
2.	Дополнения в ходе лекционного занятия, ответы на вопросы по теме	2	21	42
3.	Выполнение модульных контрольных работ	7; 7; 9 (соответственно, по кажд. модулю)	3	23
4.	Экзамен	30	1	30
Всего				200
<i>Дополнительные виды работ, выполняемые в случае необходимости корректировки рейтинга</i>				
1.	Решение задач повышенной сложности (не более 8 по всему курсу)	3 балла за одну задачу		
2.	Составление библиографии по одному из модулей курса	5 баллов		
3.	Написание эссе и подготовка сообщений по тематике курса (см. стр. 62 данного методического пособия) с использованием рекомендованных информационных источников (не более одного эссе по каждому модулю)	2–5 баллов (в зависимости от качества работы)		
4.	Составление кроссвордов по тематике курса (не более десяти слов и не более одного кроссворда по каждому модулю)	0,5 балла за слово		
5.	Подготовка медиа-презентации по результатам социометрического исследования.	5 баллов		

ШКАЛА ОЦЕНИВАНИЯ

Критерии оценки

в системе ECTS	по национальной шкале	по 100-балльной системе	по рейтинговой системе курса
A	5 отлично	85–100	170–200
B	4 хорошо	75–84	151–169
C		65–74	132–150
D	3 удовлетворительно	57–64	117–131
E		50–56	104–116
FX	2 неудовлетворительно	25–49	55–103
F		0–24	0–54

КОНТРОЛЬ ВЫПОЛНЕНИЯ ИНДИВИДУАЛЬНОГО УЧЕБНОГО ПЛАНА СТУДЕНТА

Знания, полученные студентом по конкретной теме в рамках данного учебного курса, контролируются посредством:

а) работы на практических занятиях (решение задач, выступления с сообщениями, докладами по написанным эссе, презентация кроссвордов, результатов социометрии и др.);

б) ответов на дополнительные вопросы в ходе лекционных занятий.

Знания, полученные студентом по конкретному модулю в рамках курса, контролируются посредством написания контрольной работы на последнем практическом занятии модуля. Контрольная работа предполагает решение ряда задач и оценивается в зависимости от правильности решения и количества выполненных заданий. При этом студенческая аудитория делится на несколько групп для решения разных вариантов задач, аналогичных по содержанию и сложности.

Знания, полученные студентом по всему курсу «Математико-статистический анализ в социологии», контролируются посредством экзамена. Экзамен осуществляется в устной форме и предполагает ответы студентов на вопросы и представление результатов выполнения практического задания. Как правило, в билете содержится два теоретических вопроса и одно практическое задание.

Если студента не устраивает количество баллов, набранных посредством основных видов работ по конкретному модулю или всему курсу, он может повысить свой рейтинг путем выполнения заданий из перечня дополнительных видов работ (см. выше).

К экзамену допускаются студенты, набравшие по всем видам работ не менее 100 баллов рейтинговой системы курса, при условии отработки всех заданий по каждому модулю учебной дисциплины.

ТЕМАТИЧЕСКИЙ ПЛАН КУРСА

	ТЕМА	Количество часов				Форма контроля
		Всего	Лекции	Практические занятия	Самостоятельная и индивидуальная работа	
МОДУЛЬ 1. Применение процедур математической статистики в эмпирической социологии. Числовые характеристики выборки						
1.	Основные принципы измерения в социологии. Типы переменных	12	2	2	8	Решение задач, ответы на контрольные вопросы
2.	Перекрестная классификация. Графическое представление социологической информации	12	2	2	8	Решение задач, ответы на контрольные вопросы
3.	Характеристики положения (среднее арифметическое, мода, медиана и др.)	15	4	4	7	Решение задач, ответы на контрольные вопросы
4.	Характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации)	13	4	2	7	Решение задач, ответы на контрольные вопросы
5.	Проверка процедуры первичного измерения на надежность с использованием математических методов	12	2	2	8	Контрольная работа
МОДУЛЬ II. Измерение связи между признаками с использованием математических методов						
6.	Нормальное распределение как модель вариации. Критерии линейной взаимосвязи	15	4	4	7	Решение задач, ответы на контрольные вопросы

7.	Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных	15	4	4	7	Решение задач, ответы на контрольные вопросы
8.	Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии	15	4	4	7	Решение задач, ответы на контрольные вопросы
9.	Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция	15	4	4	7	Решение задач, ответы на контрольные вопросы
10.	Корреляция рангов: коэффициент r_s	12	2	2	8	Решение задач, ответы на контрольные вопросы
11.	Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции	15	4	4	7	Контрольная работа
МОДУЛЬ III. Математические процедуры в сетевом анализе и анализе социометрических данных						
12.	Элементы теории графов в сетевом анализе	13	2	4	7	Решение задач, ответы на контрольные вопросы
13.	Обработка данных в социометрическом исследовании: вычисление индексов, построение социограмм	16	4	4	8	Контрольная работа
		180	42	42	96	

УЧЕБНАЯ ПРОГРАММА КУРСА «Математико-статистический анализ в социологии»

Содержательный модуль 1

*«Применение процедур математической статистики
в эмпирической социологии. Числовые характеристики выборки»*

Тема 1. Основные принципы измерения в социологии. Типы переменных

Предмет, цель и задачи, структура курса. Понятие измерения в социологии. Методологические основы измерений в социологии. Процедуры измерения, применяемые в ходе социологических исследований, возможности математических методов в осуществлении этих процедур. Уровни и шкалы измерения: номинальная, порядковая, метрическая, интервальная и шкала отношений; их особенности. Возможности измерения для каждой из шкал. Виды ошибок измерения. Типы переменных: количественные, качественные, дискретные, непрерывные переменные. Меры измерения социологических данных: абсолютная, относительная и накопительная частоты. Вариационный ряд. Статистические таблицы. Возможности программы «EXCEL» для обработки статистических данных, особенности ее использования в рамках курса «Математические методы в социологии».

Тема 2. Перекрестная классификация. Графическое представление социологической информации

Типы перекрестной классификации социологической информации и их интерпретация. Особенности построения и анализ динамических рядов. Кумулятивные динамические ряды. Возможности математического анализа в изучении совпадающих тенденций развития социальных процессов. Назначение графика. Общие принципы построения графиков (гистограмма, полигон, кумулята). Диаграмма полос; круговая диаграмма (гарттовская диаграмма); статистическая карта; временные диаграммы; многозначный график.

Неравные интервалы группировки. Отображение двух, трех и более полигонов на одном графике. Возможности «EXCEL» при построении различных видов графиков.

Тема 3. Характеристики положения (среднее арифметическое, мода, медиана и др.)

Основные числовые характеристики анализа одномерного распределения: максимум; минимум; среднее. Среднее арифметическое: определение и правила вычисления. Вычисление среднего для сгруппированных и несгруппированных данных. Взвешивание среднего. Свойства невзвешенного среднего. Применение среднего.

Принцип порядкового расположения. Вычисление медианы для сгруппированных и несгруппированных данных. Медиана дискретных данных. Квантили и другие меры усреднения в качестве нормирующих критериев. Мода или вероятностное среднее, правила вычисления моды. Оценка моды и бимодальность. Значение моды и медианы в социологических исследованиях. Сопоставимость средних. Выбор среднего в зависимости от цели усреднения, вида распределения и технических соображений, особенностей вычисления того или иного среднего. Сравнительные характеристики средних.

Тема 4. Характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации)

Дисперсия и среднее квадратическое отклонение. Меры протяженности, промежуточные диапазоны и измерение размаха вариации. Отклонение от среднего как мера вариации. Выбор нормы и построение меры вариации. Среднее линейное отклонение (d_{cp}). Коэффициент вариации. Квадратичные отклонения как мера вариации. Вычисление среднего квадратического отклонения (σ). Характеристики среднего квадратического отклонения. Вариация качественных переменных. Коэффициент качественной вариации. Элементарное нормирование. Другие коэффициенты качественной вариации: процентные отношения, пропорции, степени (коэффициенты), индекс. Нормировка посредством подклассификации.

Тема 5. Проверка процедуры первичного измерения на надежность с использованием математических методов

Основные показатели надежности измерения в социологической практике. Основные математические процедуры проверки шкал на надежность. Правильность измерения: выявление систематических ошибок. Устойчивость измерения: показатель абсолютной устойчивости W , средняя квадратическая ошибка, относительные показатели ошибок. Обоснованность измерения как завершающий этап подтверждения надежности измерения.

Содержательный модуль 2

*«Измерение связи между признаками с использованием
математических методов»*

Тема 6. Нормальное распределение как модель вариации. Критерии линейной взаимосвязи

Понятие нормального частотного распределения. Закон распределения. Примеры социальных явлений, описываемых нормальным, биномиальным и пуассоновым распределением. Характеристики нормальной кривой. Стандартное

отклонение. Вычисление нормированного отклонения. Особенности эмпирических распределений (асимметрия) и сопоставимость средних величин. Сравнение эмпирических и теоретических распределений. Необходимость и возможности такого сравнения. Критерий линейной взаимосвязи Хи-квадрат (χ^2).

Тема 7. Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных

Выборочное распределение и ошибки выборки. Построение доверительного интервала (большая выборка). Некоторые актуальные проблемы выборки. Нестабильность генеральной совокупности. Гетерогенность социального универсума. Проверка статистических гипотез и сравнимые оценки. Принцип проверки нуль-гипотезы. Оценка персонального риска. Другие применения нуль-гипотезы. Сравнение двух процентных отношений. Сравнение трех или более процентных отношений. Хи-квадрат (χ^2) как тест значимости.

Тема 8. Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии

Понятие статистической связи. Особенности восприятия принципа взаимной сопряженности. Виды функциональной зависимости социальных признаков. Корреляционное поле как форма графического представления корреляционной зависимости. Виды корреляционного рассеивания. Скедастичность (вариабельность). Корреляционная таблица как форма представления сгруппированных данных.

Необходимость общей меры корреляции. Отклонения от среднего арифметического: объясняемые и необъясняемые остаточные отклонения. Измерение линейной корреляции. Вычисление наклона линии регрессии. Коэффициент детерминации (r^2). Сравнение r^2 и r . Расчетные формулы: несгруппированные и сгруппированные данные. Назначение корреляционной таблицы. Коэффициент корреляции как мера тесноты, типа и направления связи между двумя признаками. Уравнение регрессии.

Тема 9. Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция

Особенности нелинейной регрессии. Вычисление корреляционного отношения. Сравнение статистических показателей r^2 и η^2 . Условия применимости критерия корреляционного отношения. Принципы интерпретации корреляционного отношения. Предосторожности в применении критерия η^2 . Виды нелинейной формы связи. Корреляция между двумя и более величинами. Частная и множественная регрессии. Множественная корреляция.

Тема 10. Корреляция рангов: коэффициент r_s

Измерение взаимосвязи признаков с помощью рангов. Метод корреляции рангов (r_s) – коэффициент Спирмена. Техника вычисления коэффициента Спирмена в случае объединенных рангов. Корреляция между упорядоченными переменными. Анализ и полезность r_s – интерпретация коэффициентов ранговой корреляции. Мера соответствия для трех и более ранговых рядов – коэффициент множественной корреляции признаков для порядкового уровня измерения.

Тема 11. Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции

Особенности вычисления коэффициента множественной корреляции для номинального уровня измерения. Коэффициент взаимной сопряженности Пирсона (C): основные характеристики и принципы вычисления.

Социологический и статистический смысл корреляции: случайная связь; бессмысленная корреляция; корреляция как доказательство наличия причины и следствия; корреляция как мера общих факторов; ложная корреляция. Причины использования большого числа критериев корреляции: вид зависимости; форма представления данных; требуемая степень точности; обратимость индексов. «Ловушки» при интерпретации корреляции: смещение во времени; меры корреляции, применяемые к неоднородным данным; определенность индекса; корреляция между синоптическими мерами. Общие социологические принципы интерпретации коэффициентов корреляции.

Содержательный модуль 3

*«Математические процедуры в сетевом анализе
и анализе социометрических данных»*

Тема 12. Элементы теории графов в сетевом анализе

Сетевой анализ как метод планирования работ проектного характера. Сетевой анализ в управленческой деятельности: определение вероятной продолжительности выполнения работ, их стоимости, возможных размеров экономии времени; выявление того, выполнение каких операций нельзя отсрочить, не задержав при этом срок выполнения проекта в целом и т. п. Составление списка операций и необходимость иллюстрации их логической последовательности. Некоторые понятия теории графов: вершина, дуга, граф, ориентированный граф, нагруженный граф. Стрелочные и вершинные графы. Анализ критического пути с применением графов. Метод прогнозного графа и его применение в социологических исследованиях и управленческой практике.

Тема 13. Обработка данных в социометрическом исследовании: вычисление индексов, построение социограмм

Социометрия как метод изучения отношений в малых группах. Социологическая специфика данных, полученных в ходе социометрического исследования. Матрица данных вида «объект-объект». Бинарные отношения. Визуализация социометрических данных. Социограммы: групповые и индивидуальные; социограммы-мишени и конвенциональные социограммы. Карта-монограмма – изображение отношений каждого члена группы с остальными ее участниками. Представление данных социометрии в виде индексов. Персональные индексы: индекс социометрического статуса члена группы, его эмоциональной экспансивности, объема интенсивности, а также концентрации взаимодействия с другими членами и др. Групповые индексы: индекс уровня благополучия взаимоотношений в группе, индекс групповой сплоченности и др.

СПИСОК ИНФОРМАЦИОННЫХ ИСТОЧНИКОВ ПО КУРСУ

Основные

1. Нечитайло І. С. Математичні методи в соціології : підручник для студентів вищ. навч. закл. [Електронне видання] / І. С. Нечитайло, М. В. Бірюкова. – Харків, 2013. – 248 с.
2. Нечитайло И. С. Математические методы в социологии : учебник для студентов вузов / И. С. Нечитайло, М. В. Бирюкова. – Харьков : Изд-во НУА, 2014. – 319 с.
3. Нечитайло И. С. Математические методы в социологии: метод. рекомендации для студентов, обучающихся по направлению подготовки 6.030101 – Социология (кредитно-модульная система) / И. С. Нечитайло. – Харьков : Изд-во НУА, 2011. – 63 с.
4. Сечко В. В. Математические методы обработки психологических данных / В.В. Сечко. – Минск, 2002. – 80 с.
5. Социометрия: исследование межличностных отношений в группе [Электронный ресурс]. – Режим доступа: <http://psyfactor.org/moreno.htm>
6. Телейко А. Б. Математико-статистичні методи в соціології та психології / А. Б. Телейко, Р. К. Чорней. – Київ : МАУП, 2007. – 424 с.

Дополнительные

7. Амберкромби Н. Социологический словарь: [Пер. с англ.] / Н. Аберкромби, С. Хилл, Б. С. Тернер. – М. : ЗАО «Издательство «Экономика»», 2004. – 509 с.
8. Барвенков С. А. Лекции по высшей математике [Электронный ресурс] / С. А. Барвенков. – Режим доступа: <http://www.kursach.com/urista.htm>.

9. Борисова Е. В. Формирование и математическая обработка данных в социологии: Учебное пособие ; [1-е изд.] / Е. В. Борисова. – Тверь : ТГТУ, 2006. – 120 с.
10. Гаптон А. Анализ таблиц сопряженности [Электронный ресурс] / Гаптон А. – Режим доступа: <http://lib.socio.msu.ru/l/library>.
11. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высшая школа, 2003. – 480 с.
12. Золотовицкий Р. Социометрия: мера общения [Электронный ресурс] / Р. Золотовицкий // Флогистон. – 2002, 20 апреля. – Режим доступа: http://flogiston.ru/articles/social/moreno_zolotovitsky.
13. Ибрагимов Н. М. Регрессионный анализ / Н. М. Ибрагимов и др. – TEMPUS, Экономический факультет НГУ, 1997. – 120 с.
14. Клюенко Э. Политическое участие: теория, методология и измерение с применением метода шкалограммирования по Гуттману / Э. Клюенко // Социология: теория, методы, маркетинг. – 2005. – №4. – С. 46-72.
15. Колесов В. М. Коэффициенты связи для совокупности номинальных признаков / В. М. Колесов // Социология: 4М. – 2012. – Том 1. – №1. – С. 62-74.
16. Лемешко Б. Ю. Сравнительный анализ критериев проверки отклонения распределения от нормального закона / Б. Ю. Лемешко, С. Б. Лемешко // Метрология. – 2005.– № 2. – С.3–23.
17. Метод социометрических измерений [Электронный ресурс] // Энциклопедия психодиагностики. – Режим доступа: <http://psylab.info>.
18. Морено Я. Социометрия [Электронный ресурс] / Я. Морено. – Режим доступа: http://polbu.ru/moreno_sociometry.
19. Наследов А. Д. Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие; 4-е изд. / А. Д. Наследов. – СПб. : Речь, 2012.–392 с.
20. Осипов Г. В. Методы измерения в социологии / Г. В. Осипов. – М. : Наука, 2003. – 124 с.
21. Пак Т. В. Эконометрика. Учебное пособие / Т. В. Пак, Я. И. Еремеева. – Владивосток : Изд-во Дальневосточного ун-та, 2009. – 70 с.
22. Письменный Д. Т. Конспект лекций по теории вероятностей / Д. Т. Письменный. – М. : Айрис-пресс, 2008. – 256 с.
23. Татарова Г. Г. Методология анализа данных в социологии : Учебник для вузов [2-е издание] / Г. Г. Татарова. – М. : Note Vene, 1999. – 180 с.
24. Толстова Ю. Н. Анализ социологических данных. Методология, дескриптивная статистика, изучение связей между номинальными признаками / Ю. Н. Толстова. – М. : Научный мир, 2000. – 352 с.
25. Толстова Ю. Н. Измерение в социологии: курс лекций / Ю. Н. Толстова. – М., 1998. – 224 с.
26. Толстова Ю. Н. Кризис социологического измерения в начале нашего века и пути выхода из него / Ю. Н. Толстова // Социология.– 1995.– № 5/6.–

С. 103–117.

27. Шеломовский В. В. Математическая статистика [Электронный ресурс] / В. В. Шеломовский. – МФГПУ. – Режим доступа: <http://www.exponenta.ru/educat/systemat/shelomovsky>
28. Ядов В. А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В. А. Ядов. – М., 2007. – 596 с.
29. Ястребов А. В. Междисциплинарный подход в преподавании математики / А. В. Ястребов // Ярославский педагогический вестник. – 2004. – № 3 (40). – С. 5-15.

Информационные ресурсы

1. Центр Разумкова. Офіційний сайт. – Режим доступу: http://www.uceps.org/ukr/expert.php?news_id=3293.
2. Офіційний сайт Головного управління статистики в Харківській області. – Режим доступу: <http://kh.ukrstat.gov.ua> .

**ПЛАНЫ ПРАКТИЧЕСКИХ ЗАНЯТИЙ И МЕТОДИЧЕСКИЕ УКАЗАНИЯ
ПО ИЗУЧЕНИЮ КУРСА
«Математико-статистический анализ в социологии»**

**МОДУЛЬ I. Применение процедур математической статистики
в эмпирической социологии. Числовые характеристики выборки**

Тема 1. Основные принципы измерения в социологии. Типы переменных

План

1. Предмет, задачи и структура курса.
2. Понятие измерения в социологии: методологические основы, методы и уровни измерения.
3. Шкала – эталон измерения. Виды шкал в социологическом исследовании, их особенности, математические процедуры для каждой из них.
4. Виды ошибок измерения.
5. Типы переменных.
6. Особенности статистических рядов в ходе измерения социологической информации.

*Методические указания по изучению темы
«Основные принципы измерения в социологии. Типы переменных»*

Тема «Основные принципы измерения в социологии. Типы переменных» является вводной темой курса «Математические методы в социологии». Учитывая тот факт, что для «студентов-социологов» не является очевидной целесообразность изучения математики, в рамках данной темы следует обратить особое внимание на цель и задачи курса, подчеркнуть необходимость его изучения для будущей профессиональной деятельности.

Конечно же, студентам известно о том, что предметом социологии является изучение общества или его отдельных фрагментов сквозь призму их социальной организации, социальных связей и отношений. Необходимо вспомнить о том, что в организационной структуре социологических исследований существует три независимых уровня: уровень фундаментальных исследований (приращение научного знания путем построения теорий); уровень прикладных исследований (изучение актуальных проблем на основе фундаментальных знаний); социальная инженерия – уровень практического внедрения научных знаний. В этом контексте особую актуальность приобретают проблемы измерения социальных процессов и явлений. Несмотря на то что само понятие «измерение» уходит корнями в точные (математические) науки и имеет дело с конкретными числовыми значениями, основой прикладной социологии являются именно измерительные операции. Студентам необходимо запомнить, что для социоло-

лога измерение – это процедура, с помощью которой объекты социологического исследования отображаются в определенной числовой и/или графической системе, что осуществляется посредством математических методов. Такими системами являются шкалы. Представление социологической информации в виде шкал открывает возможности осуществления ряда математических операций, что в итоге позволяет исследователю сделать конкретные выводы в отношении предмета и объекта изучения. Студентам должно быть известно из содержания других курсов о том, что в социологии, как правило, используются следующие шкалы: номинальная; порядковая; метрическая. Однако, следует знать, что существуют и другие варианты измерения при помощи иных шкал.

Важно помнить, что в любой науке, и тем более гуманитарной, все исследовательские процессы требуют человеческой деятельности. В связи с этим каждое изменение подвержено ошибкам, свойственным человеку, а следовательно, каждое наблюдение выражается в численном виде с большей или меньшей ошибкой. Однако подобные ошибки не следует путать с погрешностями. Основное различие между ошибкой и погрешностью заключается в том, что для оценки и исключения погрешностей существует статистическая техника, в то время как ошибки исправить невозможно.

Кроме вопросов о видах шкал, ошибках и погрешностях, в рамках изучения данной темы студенты должны усвоить материал, связанный с рассмотрением различных типов переменных. В этой связи следует запомнить, что любой объект, который в ходе последовательных наблюдений меняется качественно или количественно, может быть назван переменной. Соответственно, переменные классифицируются как качественные или количественные. Помимо этого, в математической статистике различаются также дискретные и непрерывные переменные. *Дискретные* представляют события, которые считаются «неделимыми», которые просто существуют или отсутствуют (все качественные переменные по своей природе дискретны). Количественные же могут быть как дискретными (прерывными), так и непрерывными.

Студентам следует также усвоить, что в соответствии с приведенной выше классификацией существует два основных типа статистической группировки: количественная (величины объединяются в упорядоченные интервалы – классы) и качественная (признаки дифференцируются по категориям разного порядка). При этом первым шагом к упорядочению результатов измерения является расположение их в зависимости от величины – в порядке возрастания или убывания. Для дальнейшего упрощения восприятия данных и их анализа необходима группировка близких величин в интервалы (классы родственных объектов).

Нужно помнить, что социологические данные, как правило, представляются в виде таблиц частот. Для того чтобы сделать материал, содержащийся в подобных таблицах, максимально доходчивым, следует придерживаться определенных правил, касающихся как содержания, так и общего вида таблиц. Например, заголовков таблицы должен в краткой и лаконичной форме доносить ее содержание (то есть отвечать по порядку на вопросы: «Что?», «Как?», «Где?» и «Когда?»).

Студенты должны всегда иметь в виду, что любое частотное распределение приобретает смысл только в сопоставлении с другими распределениями. Однако при сравнении двух и более распределений часто возникают трудности, связанные с существующими различиями в природе соответствующих величин. Для того чтобы подобное сравнение стало возможным, опять же, необходимо обратиться к математике и перевести абсолютные величины в относительные (в доли или проценты). Эта процедура представляет собой деление частоты в каждой клетке таблицы на сумму всех величин (для получения долей) и умножение этого результата на 100 (для получения процентов).

Вопросы и задачи для самоконтроля

1. Определите процесс измерения с различных позиций.
2. Дайте определение шкалам, их типам. Проанализируйте возможности и различия шкал для измерения социологических процессов и явлений.
3. Индивид А одобрил 50 из 100 признаков, подтверждающих наличие расовых предрассудков, в то время как индивид В одобрил только 25. Можно ли утверждать, что расовых предрассудков у А в два раза больше, чем у В? Приведите все доводы «за» и «против».
4. Каковы возможные ошибки, погрешности и их источники в следующих гипотетических измерениях: а) число разводов увеличилось на 20% за последние 50 лет; б) 20% рабочей силы являются безработными.
5. Ошибки и погрешности являются отклонениями от «истинного» значения. В каких случаях это истинное значение наиболее очевидно?
6. Размер семьи является дискретной переменной, однако средний размер семьи в Украине составляет 3,4 человек, а в США – 7,8 человек. В чем заключается противоречие и каковы возможности его устранения?
7. Семейный доход измеряется в дискретных гривнях и копейках, тем не менее доход определяется как непрерывная переменная. Объясните кажущееся несоответствие этих утверждений.

Тема 2. Перекрестная классификация. Графическое представление социологической информации

План

1. Типы перекрестной классификации социологической информации и их интерпретация. Особенности построения и анализа динамических рядов.
2. Кумулятивные динамические ряды. Возможности математического анализа в изучении совпадающих тенденций развития социальных процессов.
3. Назначение графика. Общие принципы и частные примеры построения графиков при визуализации социологической информации.
4. Отображение двух, трех и более полигонов на одном графике. Возможности «EXCEL» при построении различных видов графиков.

5. Графики качественных данных: диаграмма полос; круговая диаграмма – гарттовская диаграмма; статистическая карта; временные диаграммы; многозначный график.

*Методические указания по изучению темы
«Перекрестная классификация. Графическое представление
социологической информации»*

В ходе изучения данной темы необходимо усвоить ее основные понятия, а именно: гистограмма, полигон распределения, правило нулевого начала, разрыв шкалы, многозначный график, арифметическая шкала, арифметическая временная диаграмма, график отношений, кумулята, диаграмма полос; круговая диаграмма.

Важным является усвоение основных форм и методов представления данных, собранных в ходе социологического исследования. Большие массивы данных, с которыми обычно работает социолог, довольно сложны для целостного восприятия человеческим сознанием. Однако именно целостность необходима для полноценного анализа этих данных. В связи с этим социология заимствует у математической статистики ряд методов группировки и представления информации. Изначально данные группируются в таблицы. Необходимо обратить внимание на тот факт, что любой табличный материал можно представить в графической форме, которая (обычно) более наглядно, чем таблица, выражает картину общего распределения. Однако следует учитывать, что не всегда подобная визуализация действительно необходима. Главное назначение графика – дать наиболее точное представление о форме частотного распределения – представление, понятное каждому.

Студентам, наверняка, уже известно, что существует множество видов графического представления. И все же, нужно понять, что каждое из них полезно в своем конкретном приложении, и научиться определять, какой вид или тип графика будет более уместен в конкретном случае. Некоторые из них весьма сложны, но для студентов, обучающихся социологическим специальностям, достаточно усвоить основные и простейшие: 1) гистограмма, 2) кумулята, 3) полигон распределения, 4) диаграмма полос, 5) статистическая карта, 6) временная диаграмма.

Первые три из указанных шести типов применимы только к количественным данным. Для визуализации качественных данных можно использовать диаграмму полос. Статистическая карта представляет распределение событий по географической площади, а временная диаграмма является графическим вариантом динамического ряда.

Следует всегда иметь в виду, что графики эквивалентны таблицам, а следовательно, они должны иметь аналогичные названия и обозначения и подчиняться критериям доступности, простоты и ясности. График нельзя построить до тех пор, пока не будет приготовлена соответствующая таблица.

Необходимо помнить, что гистограмма состоит из ряда соприкасающихся столбцов, высота которых пропорциональна частоте соответствующего класса событий, а ширина пропорциональна величине интервала группировки переменной. Гистограмма является не только графической записью абсолютных частот группировок, но и наглядным изображением значения каждой частоты относительно всех других.

Полигон распределения предназначен для изображения динамики изменения данных в виде сглаженной кривой, которая, возможно, возникла бы, если бы размеры интервалов стремились к нулю, а число наблюдений неограниченно возрастало. Полигон распределения можно получить из гистограммы, проводя прямые линии через средние точки верхних частей смежных столбцов.

Построение кумуляты несколько сложнее, однако она весьма полезна, так как позволяет расчленять частотное распределение в любых точках в зависимости от необходимости.

Студентам следует знать, что графическое изображение качественных данных отличается от графиков количественных данных. Для графического изображения качественных данных используется длина отрезка, площадь фигуры или интенсивность оттенка цвета. Здесь представляется только три простейших обычно встречаемых типа: 1) диаграмма полос; 2) гартовская диаграмма; 3) статистическая карта.

Важно также помнить, что при некоторых обстоятельствах возможно начертить на одной и той же арифметической шкале два или больше временных ряда (чтобы более ясно выявить соотношения между ними). Такой график легко понятен, потому что две переменные имеют приблизительно одну и ту же область и общее расположение на шкале. Однако студенты должны иметь в виду: эта процедура должна производиться с некоторой предосторожностью, так как может ввести в заблуждение в тех случаях, когда переменные располагаются неидентично. Ложное впечатление возникает в результате того, что два ряда данных расположены на неравных расстояниях от начала отсчета.

В целом, для выработки практических навыков в рамках данной темы, предлагается выполнить задание, заключающееся в следующем: каждый студент получает опросный лист, в котором уже зафиксированы результаты одномерного частотного распределения по каждому признаку; необходимо внимательно изучить этот опросный лист и постараться максимально полно визуализировать содержащиеся в нем данные (в табличной и/или графической форме – на усмотрение студента).

Вопросы и задания для самоконтроля

1. Определите следующие понятия: гистограмма, полигон распределения, правило нулевого начала, разрыв шкалы, многозначный график, арифметическая шкала, арифметическая временная диаграмма, график отношений, кумулята, диаграмма полос; круговая диаграмма.

2. Объясните, что более целесообразно: вычерчивать относительные или абсолютные частоты? Изменяется ли внешний вид гистограммы, когда на график вместо абсолютных частот наносятся относительные частоты?

3. В каких случаях полигон распределения наилучшим образом представляет дискретные данные? Рассмотрите его соответствие для наименьших возможных интервалов и для интервалов увеличенной ширины.

4. Если попытаться сравнить два распределения путем наложения графиков один на другой, какой тип представления данных лучше выбрать: гистограмму или полигон распределения? Объясните почему?

5. Шкала частот обычно начинается с нуля. Является ли это существенным при построении полигона распределения и гистограммы? Объясните ответ.

6. Можно ли полигон распределения превратить в гистограмму и наоборот?

7. Опишите отличительные особенности графиков качественных данных.

Тема 3. Характеристики положения (среднее арифметическое, мода, медиана и др.)

План

1. Основные числовые характеристики анализа одномерного распределения: максимум; минимум; средние величины.

2. Среднее арифметическое: определение, правила вычисления и применения.

3. Мода, медиана и другие меры усреднения в качестве нормирующих критериев: определения, правила вычисления и применения.

4. Мода или вероятностное среднее, правила вычисления моды.

5. Значение среднего арифметического, моды и медианы в социологических исследованиях.

6. Сопоставимость средних величин.

Методические указания по изучению темы «Характеристики положения (среднее арифметическое, мода, медиана и др.)»

В ходе изучения данной темы необходимо научиться легко владеть ее основными понятиями: среднее арифметическое, медиана, мода, срединное положение, ранг, медианный и модальный ранг, медианная и модальная точка, медианный и модальный интервал группировки, модальная частота, дециль, метод разностей, бимодальность.

В рамках данной темы рассматриваются характеристики, определяющие положение центра эмпирического распределения. В описательной статистике, чаще всего, употребляются такие характеристики положения, как среднее арифметическое, медиана и мода. Определение этих характеристик может быть весьма результативным и в эмпирической социологии. Из предыдущих тем

студенты должны были усвоить, что распределение частот количественных данных представляет собой результат большого числа наблюдений. По существу это ряд величин, расположенных на непрерывной шкале. Любая величина в распределении может описывать всю совокупность, если известно ее относительное положение в распределении. Наиболее удобными для вычисления и полезными для исследователя являются такие величины, как: 1) максимум, 2) минимум и 3) центральные или типичные величины, известные как средние.

Нужно понимать, что в некоторых случаях максимумы и минимумы не дают необходимой информации. Тогда и возникает необходимость обращения к более общим мерам положения, которыми являются средние. Явная тенденция многих статистических совокупностей концентрироваться вокруг центра часто называется «центральной тенденцией», а значение величины в этом центре – «мера центральной тенденции» – обычно называется средней.

Одной из основных характеристик выборки является **среднее арифметическое**. Это такое значение признака, сумма отклонений от которого выборочных значений признака равна нулю (с учетом знака отклонения). Студентам следует обратить внимание, что вычисление среднего необходимо для осуществления более сложных математических операций, связанных с анализом социологических данных, в чем можно будет убедиться на последующих занятиях.

Очень важно запомнить, что процедуры вычисления среднего для негруппированных и сгруппированных данных несколько отличаются.

Для негруппированных данных среднее вычисляется путем суммирования отдельных величин и последующего деления на N . Формула вычисления имеет следующий вид: $\bar{X} = \frac{\sum x_i}{N}$, где \bar{X} – среднее арифметическое, $\sum x_i$ – сумма переменных, x_i – переменная или величина, N – сумма частот, или число событий, объем распределения.

Для сгруппированных данных переменные x_i умножаются на свои частоты (n_i), суммируются, а затем делятся на N , что имеет вид следующей формулы: $\bar{X} = \frac{\sum x_i \cdot n_i}{N}$. В этой связи важно понимать, что *если применить две разные формулы к одному и тому же распределению, значения среднего в обоих случаях отличаться не будут*.

Студенты также должны усвоить процедуру вычисления среднего арифметического *для интервальных рядов данных*, которая несколько отличается от процедур, описанных выше. При этом необходимо помнить, что для интервальных непрерывных рядов каждая из рассматриваемых переменных находится в средней точке соответствующего интервала. Когда величина интервала больше единицы, сначала находятся границы интервалов и их средние точки по формуле: $x_{ci} = \frac{(x_i + x_{i+1})}{2}$, а затем каждая средняя точка взвешивается (то есть

перемножается на частоту интервала) и вычисляется среднее по формуле:

$$\bar{X} = \frac{\sum x_{ci} \cdot n_i}{N}$$

Необходимо запомнить основное свойство среднего арифметического, которое заключается в том, что оно представляет величины каждого события в распределении. В связи с этим оно подвержено влиянию как очень больших, так и крайне малых величин, что особенно заметно в несимметричных распределениях. Необходимо всегда иметь в виду, что для таких распределений более информативными могут быть иные меры усреднения, такие как медиана и мода.

Медианой называется точка, которая рассекает таблицу на две равные части так, что одна половина событий точно находится ниже, а другая – выше этой точки. Опять-таки, важно помнить, что, по аналогии со средним арифметическим, существуют разные формулы вычисления медианы для несгруппированных и сгруппированных данных. *Для несгруппированных данных* придерживаются следующей формулы: – $\frac{N}{2}$. *Для сгруппированных* – вычисления вы-

глядят так: $Me = x_o + \delta \frac{1/2 N - n_H}{n_{me}}$, где x_o – нижняя граница медианного интер-

вала, δ – ширина медианного интервала, n_{me} – частота медианного интервала, N – объем выборки, n_H – частота, накопленная до медианного интервала.

Следует знать, что для большей точности можно разделить совокупность не на две равные части (как в случае с медианой), а на меньшие доли, то есть локализовать данные в меньшем интервале – верхней четверти, десятой или даже сотой. Для этого существуют другие меры. Речь идет о вычислении **квартилей, децилей и центилей**. Таким образом, *первый квартиль* в распределении

будет равен: $Q_1 = x_o + \delta \frac{N/4 - n_{H1}}{n_{Q1}}$. Если потребуется выделить точку, ниже ко-

торой находятся 75% событий (или Q_3), то производят следующую подстановку

в формулу: $Q_3 = x_o + \delta \frac{3N/4 - n_{H3}}{n_{Q3}}$. *90-й центиль* (или C_{90}) может быть найден

по формуле: $C_{90} = x_o + \delta \frac{90N/100 - n_{H90}}{n_{C90}}$.

Медиана, квартили, квантили, децили и центили, которые, согласно своему определению, указывают на долю событий, расположенных ниже или выше данной величины, носят обобщенное название **квантили**.

Мода представляет собой наиболее часто повторяющуюся величину в упорядоченном распределении; она характеризует то место распределения, где концентрация событий максимальна. Студентам необходимо усвоить **фор-**

мулу вычисления моды: $M_o = x_o + \delta \frac{n^{mo} - n^-}{2n_{mo} - n^+ - n^-}$, где x_o – нижняя граница модального интервала, δ – ширина модального интервала, n_{mo} – частота модального интервала, n^- – частота интервала, предшествующего модальному интервалу, n^+ – частота последующего интервала.

Следует знать, что некоторые распределения обнаруживают два максимума и поэтому называются **бимодальными**, в отличие от **унимодальных** распределений. Не менее важным является понимание того, что значения мер положения (\bar{X} , Me и M_o) существенно отличаются в случае асимметричного распределения. Чем больше между ними разница – тем сильнее асимметрия.

Необходимо помнить о существовании трех основных критериев, помогающих решить вопрос о применимости того или иного типа среднего: 1) цель усреднения, 2) вид распределения данных, 3) различные технические соображения, главным образом, арифметического характера, которые ограничивают выбор типа усреднения.

В целом, для выработки практических навыков по данной теме, усвоения формул и доведения до автоматизма вычислительных процедур студентам предлагается целый ряд простейших таблиц (не менее 10), для которых они должны определить характеристики положения и сделать соответствующие выводы. Все вычислительные процедуры осуществляются в «EXCEL», что существенно экономит время, позволяя сконцентрироваться не на процессе вычисления, а на его алгоритме, что, несомненно, является более важным для освоения материала.

Вопросы и задачи для самоконтроля

1. Определите следующие понятия: характеристики положения, среднее арифметическое, взвешивание.
2. Приведите несколько примеров, в которых уместно было бы применять арифметическое среднее, даже если разброс событий сильно асимметричен.
3. Покажите, что крайние величины в асимметричном распределении дают непропорциональный вклад в арифметическое среднее.
4. Опишите схематично процедуру нахождения среднего двух и более групп, когда даны только сумма частот и арифметическое среднее каждой группы. Выразите эту процедуру в символах.
5. Каково будет комбинированное среднее двух групп, если среднее из 100 событий первой будет равно 10, а среднее из 50 событий второй будет равно 15? Каково было бы комбинированное среднее, если бы каждая группа состояла из 50 событий? Из 100 событий?

6. Определите следующие понятия: медиана, срединное положение, ранг, медианный ранг, медианная точка, медианный интервал группировки, дециль, модальная частота, метод разностей, бимодальность.

7. Какое делается предположение относительно распределения событий в пределах медианного интервала, когда рассчитывается медиана сгруппированных данных?

8. Каково соотношение между модой, медианой и средним в распределении?

9. Изложите в общих чертах, как влияет правая асимметрия на три основных типа средних; то же самое – при левой асимметрии.

10. В каких случаях максимум наилучшим образом представляет распределение частот, а в каких – минимум?

11. Как видоизменились бы медианная и средняя оценки, если бы:

а) наихудшие студенты были удалены из класса?

б) наихудшие студенты стали лучше заниматься?

в) средние студенты стали лучше заниматься?

г) были облегчены экзамены?

12. В населенном пункте «А» модальная продолжительность жизни равна 55 годам, медианная равна 60 годам и средняя – 60. В населенном пункте «В» модальная продолжительность жизни составляет 70 лет, медианная – 65 лет, а средняя – 60. Воспроизведите кривые частот по этой информации. В каком населенном пункте более здоровое население?

Тема 4. Характеристики рассеяния

(дисперсия, среднее квадратическое отклонение, коэффициент вариации)

План

1. Дисперсия и среднее квадратическое отклонение.

2. Отклонение от среднего как мера вариации. Коэффициент вариации.

3. Вычисление дисперсии и среднего квадратического отклонения (σ).

Характеристики среднего квадратического отклонения.

4. Вариация качественных переменных. Коэффициент качественной вариации.

5. Элементарное нормирование. Другие коэффициенты качественной вариации: процентные отношения, пропорции, степени, индексы.

Методические указания по изучению темы «Характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации)»

В рамках данной темы необходимо усвоить ее основные понятия, а именно: варьированность, вариация, полный диапазон (размах вариации), промежуточный диапазон, интерквартильный диапазон, 10–90% диапазон, дисперсия, рассеяние, среднее квадратическое отклонение, коэффициент вариации.

На примере предыдущей темы студенты должны были убедиться в том, что характеристики положения, хоть и являются чрезвычайно важными при изучении варьирующего признака, все же не дают полной информации о нем. Нетрудно представить себе два эмпирических распределения, у которых средние одинаковы, но при этом у одного из них значения признака рассеяны в узком диапазоне вокруг среднего, а у другого – в широком. Поэтому наряду со средними значениями вычисляют и характеристики рассеяния выборки. В рамках данной темы рассматриваются наиболее употребляемые из них: размах вариации, дисперсия, среднее линейное и квадратическое отклонение, коэффициент вариации. Студентам необходимо запомнить определения этих характеристик и усвоить процедуры их вычисления.

Размах вариации – это интервал, заключающий в себе все значения. Можно вычислить разность между истинными крайними значениями множества переменных, которые в этом случае устанавливают границы размаха вариации. Иногда эта величина обозначается как «включающий» диапазон и определяется: $(H - L + 1)$; например $(12 - 2) + 1$.

Еще более ограниченным является промежуточный диапазон между первым и третьим квартилями (интервал, который содержит средние 50% случаев). Обычно он называется **интерквартильным диапазоном**. Подобная процедура обнаруживает степень группировки случаев, относящихся к среднему интервалу, вокруг медианы.

Дисперсией называется средний квадрат отклонения значений признака от среднего арифметического. Дисперсия, вычисляемая по выборочным данным, называется выборочной дисперсией и обозначается σ^2 .

Выборочную дисперсию вычисляют по приведенным ниже формулам:

Для *несгруппированных данных*
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$
. В этой формуле

$\sum_{i=1}^n (x_i - \bar{x})^2$ – сумма квадратов отклонений значений признака x_i от среднего арифметического \bar{x} . Для получения среднего квадрата отклонений эта сумма поделена на объем выборки n .

Для *сгруппированных данных*:

$$\sigma^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x})^2}{n}$$

где: x_i – срединные значения интервалов группировки;

$\sum_{i=1}^k n_i (x_i - \bar{x})^2$ – взвешенная сумма квадратов отклонений.

Стандартное или **среднее квадратическое отклонение** (сигма) обозначается с помощью греческой буквы σ и вычисляется путем нахождения корня квадратного из дисперсии по формуле: $\sigma = \sqrt{\sigma^2}$.

Для получения σ сгруппированных данных применяется формула:
$$\sigma = \sqrt{\frac{\sum (x_{ci} - \bar{x})^2 \cdot n_i}{N}}$$
, где n_i – частота, $(x_{ci} - \bar{x})^2$ – отклонение средней точки интервала от среднего арифметического.

Стандартное отклонение может быть превращено в меру относительной вариации посредством нормирования его по отношению к собственному началу отсчета (то есть среднему арифметическому), которая называется **коэффициентом вариации** и вычисляется по формуле $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%$.

Студентам следует запомнить, что *малое значение σ при большом среднем указывает на большую однородность данных и в силу этого на типичность среднего, что в некоторых условиях крайне существенно.*

Вариацию качественных переменных нельзя измерять тем же способом, который был разработан для количественных. Вместо вычисления величин подсчитываются различия в качествах.

Любое событие исследователь рассматривает не изолированно, а в сравнении с конкретной нормой, вытекающей из социальной основы данного события. Совокупность процедур можно назвать операциями нормировки, поскольку они устанавливают определенные стандарты наблюдаемых величин. Можно осуществлять нормировку приблизительно в следующем порядке сложности: 1) процентные отношения; 2) пропорции; 3) степени; 4) индексы; 5) подклассификация; 6) стандартизация.

Определение первых четырех терминов не должно вызывать сложности у студентов. Что же касается пятого и шестого, то следует запомнить, что **подклассификация** подразумевает разделение факторов на «внешние» и «внутренние» (причем внешние факторы не должны изменяться в ходе исследования). **Стандартизация** же кажется полностью формализованной процедурой. Любой вид арифметического среднего может быть стандартизован при наличии необходимых данных.

В целом, для выработки практических навыков по данной теме, усвоения формул и доведения до автоматизма вычислительных процедур студентам предлагается ряд таблиц, с которыми осуществлялась работа на предыдущем практическом занятии (по нахождению характеристик положения). Необходимо для этих же таблиц определить характеристики рассеяния. На основе совокупных результатов сделать соответствующие выводы. Так же как и на предыдущем занятии, все вычислительные процедуры осуществляются в «EXCEL», с учетом всех преимуществ работы с данной программой.

Вопросы и задачи для самоконтроля

1. Определите следующие понятия: варьированность, вариация, полный диапазон (размах вариации), промежуточный диапазон, интерквартильный диапазон, 10–90% диапазон, дисперсия, рассеяние, среднее квадратическое отклонение, коэффициент вариации.
2. Объясните, почему среднее квадратическое отклонение, а не вариация, обычно используется как мера дисперсии.
3. Можно ли вычислить σ качественных данных?
4. Если средний возраст студентов вуза равен 20 годам, а среднее квадратическое отклонение равно 2, то каким будет среднее и сигма (σ) этой группы двадцать лет спустя? Чему будет равен V_σ ?
5. Население конкретного города состоит из 50% мужчин и 50% женщин; 70% – украинцев – 30% россиян. Можно ли представить переменные одним V_σ ? Поясните.
6. Население конкретного города распределено по четырем этническим группам следующим образом: украинцы – 60%, россияне – 20%, татары – 15%, румыны – 5%. Итого: 100%. Вычислите V_σ .

Тема 5. Проверка процедуры первичного измерения на надежность с использованием математических методов

План

1. Основные показатели надежности измерения в социологической практике. Математические процедуры проверки шкал на надежность.
2. Правильность измерения, выявление систематических ошибок.
3. Устойчивость измерения: показатель абсолютной устойчивости W , средняя квадратическая ошибка, относительные показатели ошибок.
4. Обоснованность измерения как завершающий этап подтверждения надежности измерения.

Методические указания по изучению темы «Проверка процедуры первичного измерения на надежность с использованием математических методов»

Центральным понятием данной темы является понятие «надежность». Следует обратить внимание на факт отсутствия единообразия в толковании данного термина применительно к социологической информации. Одни авторы трактуют надежность слишком широко, подразумевая качество всего исследования, его итогов. Другие, наоборот, отождествляют надежность лишь с тем или иным ее проявлением (либо с устойчивостью данных, либо с их адекватностью целям исследования и т. п.). К моменту изучения данной темы, основыва-

ясь на базисе накопленных социологических знаний (как из предыдущих тем курса «Математические методы в социологии», так и из других дисциплин, ему предшествующих), каждый студент может иметь свою точку зрения относительно изложенного выше. Было бы интересно подискутировать на эту тему в рамках практического занятия. И все же, необходимо знать, что в строгом смысле слова **понятие надежности измерения**, как правило, относится именно к инструменту, с помощью которого производится измерение, но не к самим данным, подлежащим измерению. В отношении данных и заключительных выводов исследования используется термин «достоверность».

Из предыдущих курсов студентам уже должно быть известно о том, что в целом достоверность результатов исследования зависит от многих составляющих, начиная с того, насколько обоснована его общая концепция и все компоненты теоретико-методологического раздела программы, от качества исходных данных, системы их сбора (соответствия типа выборки и ее организации целям исследования), качества анализа данных, глубины интерпретации полученных зависимостей и связей. Однако следует запомнить, что важной предпосылкой получения достоверных данных, которые бы максимально точно описывали или объясняли существующие социальные реалии, является надежность шкал. В рамках курса «Математические методы в социологии» рассматриваются, соответственно, те операции повышения надежности первичного измерения, которые осуществляются посредством математических вычислений. Следует обратить внимание на то, что такие операции используются лишь на стадии отработки инструмента измерения в процессе пилотажа. Итак, в рамках данного курса рассматриваются обобщающие понятия надежности инструмента измерения (и, соответственно, надежности данных, фиксируемых этим инструментом), что подразумевает под собой три составляющие: обоснованность, устойчивость и правильность измерения.

Правильность устанавливает общую приемлемость данного метода (способа) измерения и проявляется в выявлении систематических ошибок. Систематические ошибки – это ошибки, которые проявляются постоянно или в соответствии с определенным законом. К систематическим ошибкам относятся: отсутствие разброса ответов по значениям шкалы; использование части шкалы; неравномерное использование отдельных пунктов шкалы; определение грубых ошибок.

В процессе измерения иногда возникают грубые ошибки, причиной которых могут быть неправильные записи исходных данных, плохие расчеты, неквалифицированное использование измерительных средств и т. д., в связи с чем в рядах измерений могут попадаться данные, резко отличающиеся от совокупности всех остальных значений. Чтобы выяснить, являются ли подобные расхождения результатом грубых ошибок, устанавливается критическая граница таким образом, чтобы вероятность превышения ее крайними значениями была достаточно малой и соответствовала некоторому уровню значимости α . Это правило основано на том, что появление в выборке чрезмерно больших значений хотя и возможно как следствие естественной вариабельности значе-

ний, но маловероятно.

Например, в случае неравномерного использования отдельных пунктов шкалы для выявления аномалий равномерного распределения по шкале можно предложить следующее правило: для достаточно большой доверительной вероятности ($1-\alpha \geq 0,99$) и, следовательно, в достаточно широких границах наполнение каждого значения не должно существенно отличаться от среднего из соседних наполнений. Соответствующий статистический критерий таков:

$$\chi^2 = \frac{(n_i - \tilde{n}_i)^2 (2n-1)}{(n_i + \tilde{n}_i)(2n - n_i - \tilde{n}_i)}. \text{ Эта величина имеет Хи-квадрат (далее по тексту - } \chi^2 \text{)}$$

распределение с одной степенью свободы ($df = 1$). Здесь i – номер значения признака, который подвергается анализу; n_i – наблюдаемая частота для этого

значения; $\tilde{n}_i = \frac{n_{i-1} + n_{i+1}}{2}$ – ожидаемая частота, как средняя из двух соседних.

Устойчивость характеризует совпадение результатов измерения при повторных применениях измерительной процедуры и описывается случайными ошибками. Необходимо знать *основное правило проведения повторного исследования*: оно проводится на одной и той же выборке таким образом, чтобы временной промежуток между исследованиями не был ни слишком коротким (чтобы респонденты не смогли воспроизвести ответы по памяти), ни слишком длинным (чтобы мнение респондентов в отношении предмета исследования не изменилось под влиянием объективных обстоятельств). Как правило, в зависимости от объекта исследования, этот промежуток составляет 2–3 недели.

Результаты обоих исследований заносятся в специальную таблицу. В качестве меры устойчивости шкалы определяют несколько коэффициентов, одним из которых является *показатель абсолютной устойчивости* шкалы (W).

Он рассчитывается по следующей формуле: $W = \frac{\sum_{i=j=1}^k n_{ij}}{n} = \frac{n_{11} + n_{22} + \dots + n_{kk}}{n}$, где n_{ij} –

количество совпадающих и в первом и во втором опросе, n – количество опрошенных. $W_{\max} = 1$ – в случае, когда различий между первым и вторым опросом нет, то есть все ответы совпадают. Студентам следует запомнить, что данный коэффициент применяется в основном для качественных признаков номинальной шкалы. Для всех остальных необходимо считать коэффициенты несовпадающих ответов. В этом случае вычисляются *показатели неустойчивости*, то есть величины ошибки, учитывающие не только факт несовпадения ответов, но и степень этого несовпадения. Линейной мерой несовпадения оценок является средняя арифметическая ошибка, показывающая средний сдвиг в ответах в расчете на одну пару последовательных наблюдений:

$|\Delta| = \frac{1}{n} \sum_{i=1}^n |x_i^{II} - x_i^I|$, где x_i^I, x_i^{II} – ответы по анализируемому вопросу i -порядка в I и II пробах соответственно. В качестве **показателя для нормирования абсолютной ошибки** можно использовать максимально возможную ошибку в рассматриваемой шкале Δ_{\max} . Если число делений шкалы k , тогда Δ_{\max} равна разнице между крайними значениями шкалы ($x_{\max} - x_{\min}$), то есть $k-1$, и относительная ошибка имеет вид: $\Delta_{\text{отн}} = \frac{|\Delta|}{\Delta_{\max}}$, где $|\Delta|$ – средняя арифметическая ошибка измерения. Важно понять, что для повышения устойчивости шкалы необходимо выяснить различительные возможности ее пунктов. Высокой различимости соответствует малая ошибка.

Обоснованность связана с доказательством того, что был измерен именно тот объект и то свойство, которые требовались, и является самой тяжелой процедурой. Правильность и устойчивость определяются с помощью математических формул, а обоснованность доказывается логическим путем. При доказательстве обоснованности необходимо точно определить предмет и объект, методы и средства измерения.

Проверку надежности необходимо начать с проверки правильности и устойчивости, если они подтверждаются, то, как правило, подтверждается и обоснованность.

В целом, для закрепления полученных знаний по данной теме и выработки практических навыков на практическом занятии студентам предлагается решить ряд задач, направленных на: 1) нахождение статистических критериев правильности; 2) определение показателей устойчивости.

Вопросы и задачи для самоконтроля

1. Дайте определение надежности измерения. Каковы основные ее составляющие?
2. Что подразумевается под правильностью измерения?
3. Какие типы ошибок могут свидетельствовать о неправильности измерения. Приведите примеры.
4. Назовите основное отличие систематических ошибок от грубых.
5. Дайте определение устойчивости измерения.
6. С разрешением каких ошибок связан вопрос об устойчивости измерения?
7. Измерьте устойчивость проведенного измерения по результатам таблицы, приведенной ниже.

Проба I	Проба II					Сумма
	1	2	3	4	5	
1	3	5	1			9
2		3	1	1		5
3		7	6	2	2	17
4	1	3	4	6	1	15
5		1		1	2	4
Σ	4	19	12	10	5	50

8. В чем заключается суть обоснованности измерения и как эта процедура осуществляется?

МОДУЛЬ II. Измерение связи между признаками с использованием математических методов

Тема 6. Нормальное распределение как модель вариации. Критерий линейной взаимосвязи

План

1. Понятие нормального частотного распределения. Характеристики нормальной кривой.
2. Особенности эмпирических распределений.
3. Сравнение эмпирических и теоретических распределений, необходимость и возможности такого сравнения.
4. Критерий линейной взаимосвязи.

Методические указания по изучению темы «Нормальное распределение как модель вариации. Критерий линейной взаимосвязи»

В рамках данной темы следует обратить внимание на необходимость сравнения эмпирических распределений с неким эталоном, называемым идеальным распределением. Студенты должны убедиться в том, что такое сравнение, в исследовательских целях, может быть просто необходимо для: возможности спрогнозировать дальнейшее поведение и развитие того или иного феномена (в случае если различия между эмпирическими и теоретическими распределениями невелики); выявления причин, влияющих на проявление отличий между теоретическими и эмпирическими распределениями (если таковые наблюдаются).

Необходимо знать, что к идеальным распределениям можно отнести **закон нормального распределения или гауссовский закон распределения**. Функ-

ция плотности гауссовского распределения имеет вид $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$,

где σ – дисперсия случайной величины, μ или (\bar{x}) – среднее значение (математическое ожидание).

По определению, **нормальная кривая** состоит из бесконечного числа точек, унимодальна, симметрична и неограничена в обоих направлениях.

Следует запомнить, что различают несколько вариантов эмпирических распределений при сравнении их с теоретическими, а именно: *симметрические* и *скошенные*. Студенты должны научиться определять величину скошенности

или асимметрии. Для этого вычисляется коэффициент $As = \frac{\bar{x} - Mo}{\sigma}$,

где $-3 < As < 3$. Необходимо запомнить *основные свойства асимметрии*:

1) коэффициент As изменяется от -3 до $+3$. Чем ближе As к граничным значениям (-3 , $+3$), тем больше скошенность; 2) если значение As положительно, то говорят, что распределение вправо скошено, если отрицательно, то – влево скошено, если $As=0$ – то асимметрии нет, то есть распределение симметрично и $\bar{x} = Mo = Me$.

Нужно знать, что для оценки различий между теоретическим и эмпирическим распределениями существует **критерий** Хи-квадрат (далее по тексту – χ^2). Если закон распределения признака неизвестен, но есть основание предположить, что он имеет определенный вид A , то χ^2 позволяет проверить гипотезу: исследуемая совокупность распределена по закону A (это – Нуль-гипотеза (H_0), подробно рассматриваемая в рамках следующей темы). Критерий χ^2 отвечает на вопрос, случайно или нет такое расхождение частот. Важно помнить, что как любой критерий, χ^2 не доказывает справедливость гипотезы, а лишь с определенной вероятностью альфа (α) устанавливает ее согласие или несогласие с данными наблюдениями. Если $\chi^2_{\text{набл}} < \chi^2_{\text{табл.}}$, то говорят, что распределение приближено к теоретическому с определенной вероятностью альфа (α). Другими словами, α – вероятность ошибки утверждения. Чем она меньше – тем лучше, то есть вероятность правильности выбора типа распределения больше, а различия между наблюдаемыми и теоретическими частотами меньше.

Критерий χ^2 имеет вид: $\chi^2 = \sum_{i=1}^k \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$. Критическая точка распределения

χ^2 находится по заданному уровню значимости α и числу степеней сво-

боды df . Число степеней свободы находится по формуле: $df=(k-1)*(l-1)$, где k – число строк матрицы, а l – число столбцов матрицы.

Показатель χ^2 имеет довольно сложное математическое обоснование, однако это очень распространенная величина, полезная в тех статистических ситуациях, когда необходимо измерить расхождение между наблюдаемыми и ожидаемыми частотами. Последовательность операций в вычислении такова:

1. Вычислить ожидаемые частоты (E):

$$E_{ij} = \frac{r_i c_j}{N},$$

где: r_i – маргинал i -й строки, c_j – маргинал j -го столбца, а N – объем выборочной совокупности.

2. Вычесть ожидаемые частоты из фактически наблюдаемых частот: $(O-E)$.

3. Каждую полученную разность возвести в квадрат $(O-E)^2$.

4. Разделить каждое отклонение, возведенное в квадрат, на соответствующую ожидаемую частоту E (тем самым нормируя каждое абсолютное расхождение на его собственную базу).

$$\frac{(O-E)^2}{E}$$

вующую ожидаемую частоту E (тем самым нормируя каждое абсолютное расхождение на его собственную базу).

5. Сложив все нормированные отклонения, получим χ^2 . Таким образом, формула будет иметь следующий вид: $\chi^2 = \sum \frac{(O-E)^2}{E}$.

Ожидаемая частота в любой клетке вычисляется обычно путем перемножения соответствующих маргиналов и последующего деления их произведения на N . В символах это выглядит так:

$$E_{ij} = \frac{c_i \cdot m_j}{N}, \text{ где } c_i \text{ – маргинал } i\text{-й строки, } m_j \text{ – маргинал } j\text{-го столбца.}$$

Важно знать, что чисто случайное распределение свидетельствует об отсутствии связи. А вообще, чем больше наблюдаемое распределение приближается к случайному распределению – тем слабее связь между признаками. Подобно этому, чем больше расхождение между наблюдаемым и случайным распределениями, тем сильнее связь или зависимость между переменными. Так как χ^2 выбран в качестве меры этого расхождения, то чем больше значение χ^2 , тем теснее взаимосвязь. Таким образом, χ^2 можно было бы принять как приближенную меру корреляции. Однако значения χ^2 ненормированы; эти значения не изменяются в пределах от нуля до единицы и, таким образом, непригодны для измерения корреляции в общепринятом смысле слова.

Следует запомнить, что для качественных признаков считается **индекс**

качественной вариации:
$$I = \frac{\sum n_i n_j}{(k_i k_j) m} 100\% ,$$

Где: n_i, n_j – частоты качественного признака; k_i, k_j – теоретические частоты; m – количество градаций.

Этот индекс указывает степень неоднородности полученных ответов, так как для качественных признаков теоретическим является равномерное распределение.

Для альтернативных качественных признаков вариация обычно рассчитывается по следующей формуле:
$$S^2 = \frac{n_1 n_2}{(n_1 + n_2)^2} .$$

Не менее важным является знание другой меры вариации признака, которая носит название «**энтропия**», представляет собой **меру неопределенности** и вычисляется по формуле:
$$H = - \sum_{i=1}^n m_i \log m_i .$$

Знак минус в этой формуле отражает тот факт, что вероятности всегда меньше единицы, а логарифм может быть взят по любому основанию.

Из предыдущих тем данного курса студенты должны были усвоить, что среднее квадратическое отклонение показывает, насколько в среднем отличаются все варианты выборки от среднего арифметического. В этой связи необходимо уточнить, что в случае нормального распределения ответы респондентов попадают в отрезок $(\bar{x} - 3\sigma \dots \bar{x} + 3\sigma)$. Для качественных признаков, если σ близка к нулю, то рассеивания нет. Для центральной вариации справедливо следующее неравенство: $X_{min} < \bar{X} , Mo, Me < \sigma < X_{max}$.

Для количественных признаков чаще всего информационно значимой является \bar{X} (среднее арифметическое), но близка к нему и медиана, которая является вспомогательным коэффициентом и вычисляется в случае, если наблюдается большая скошенность ряда.

В целом, для закрепления знаний по данной теме и выработки практических навыков обработки данных, полученных в ходе социологического исследования, студентам предлагается решить ряд задач, направленных на вычисление χ^2 и интерпретацию полученных значений. Содержание задач имеет социологический характер, а результаты их разрешения могут рассматриваться в качестве примера подтверждения первичных и/или постановки вторичных гипотез социологического исследования. Вычислительные процедуры осуществляются в «EXCEL», с учетом всех преимуществ работы в данной программе.

Вопросы и задания для самоконтроля

1. Определите следующие понятия: нормальное распределение, нормальная площадь, нормальная ордината, нормальное отклонение, колоколообразное распределение.
2. Объясните, в каком смысле нормальная кривая является «нормальной».
3. Объясните сущность энтропии.
4. Каким образом распределения делятся по их симметрии? О чем свидетельствуют отклонения от идеальных распределений?
5. Между какими двумя сигма-точками на основной линии нормальной кривой лежат средние 50% случаев?
6. Объясните, почему доля случаев между 0 и 1,0 сигма не равна доле между 1,0 и 2,0?
7. Опишите правила вычисления Хи-квадрат (χ^2) и его сущность.

Тема 7. Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных

План

1. Понятие статистической гипотезы. Проверка статистических гипотез и сравнимые оценки.
2. Принцип проверки Нуль-гипотезы (H_0). Хи-квадрат (χ^2) как тест значимости.
3. Нуль-гипотеза (H_0): некоторые современные проблемы, связанные с формулировкой, доказательством/опровержением.

Методические указания по изучению темы

«Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных»

В рамках предыдущей темы студенты должны были усвоить, что критерий Хи-квадрат (далее по тексту – χ^2) позволяет сделать вывод относительно закона распределения, которому подчиняется наблюдаемая случайная величина. Однако важно знать, что данный вывод основывается на проверке Нуль-гипотезы (далее по тексту – гипотеза H_0).

Следует запомнить, что в наиболее распространенном варианте гипотеза H_0 утверждает, что данные выборки получены из статистически идентичных совокупностей, а, следовательно, любое различие между выборками является случайной вариацией. Необходимо усвоить, что гипотеза H_0 выдвигается для того, чтобы потом, как правило, аннулировать. По своей природе, она тесно связана с более конструктивными статистическими гипотезами, называемыми иногда альтернативными гипотезами. Студенты должны знать, что гипотеза

H_0 идентифицируется с двумя типами исследовательских процедур: 1) сравнением двух или более универсумов в отношении заданного свойства, 2) установлением корреляции между двумя или более свойствами данного универсума. В первом случае она отрицает различие между параметрами совокупностей; во втором – предполагает случайное соотношение или нулевую корреляцию между исследуемыми переменными.

При испытании гипотезы H_0 наблюдаемое различие рассматривается как отдельное значение в нормальном выборочном распределении, среднее которого равно нулю (что указывает на тождество средних значений генеральных совокупностей), а его квадратическая ошибка при этом оценивается по формуле:

$$S_D = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}.$$

В процессе принятия статистического решения необходимо различать два связанных этапа: 1) оценка полученной вероятности; 2) оценка последствий правильного решения.

Отвергая гипотезу H_0 , мысленно принимаем некоторую неконкретизированную альтернативную гипотезу, которая имеет свое собственное выборочное распределение. Таким образом, можно сформулировать дилемму: когда риск отклонения правильной гипотезы уменьшается, риск принятия ошибочной гипотезы, соответственно, увеличивается. Какой же вид риска следует предпочесть? Студенты должны понять, что математическая статистика может измерять риск неправильного решения, но она не может посоветовать заинтересованному лицу, принимать или нет этот риск. Принятие измеренного риска будет зависеть от соображений, имеющих субъективный, этический, экономический и т. п. характер. Статистический вывод, тем самым, основывается на весьма нестатистических соображениях.

Следует остановить внимание на основных принципах проверки гипотезы H_0 : 1) формулируется гипотеза о том, что генеральные совокупности являются однородными; 2) вычисляется комплексный показатель χ^2 величины наблюдаемых различий, который 3) позволяет определить вероятность получения заданных разностей процентных отношений в предположении гипотезы H_0 .

Принятие или отклонение χ^2 зависит от величины этой вероятности – малая вероятность указывает на ее отклонение, а большая – на принятие.

Студентам важно понять и запомнить, что значение χ^2 не является мерой степени связи; его уровень значимости указывает лишь на вероятность существования этой связи. Для измерения степени корреляции необходимо обратиться к другим методам (о которых речь пойдет в последующих темах данного курса). Причина этого состоит в том, что, как уже говорилось, χ^2 выражен в абсолютных отклонениях и является переменной величиной, а не нормированным показателем, изменяющимся в стандартных пределах от 0 до 1.

Подобно многим другим математическим методам, используемым при статистическом испытании гипотез, методы с использованием χ^2 применимы лишь тогда, когда выборки сделаны случайно и независимо.

В целом, для закрепления знаний и выработки практических навыков на практическом занятии по теме «Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных» студентам предлагается ряд задач по вычислению χ^2 , аналогичных тем, с которыми велась работа на предыдущем практическом занятии. Однако, помимо осуществления вычислительных процедур, студенты должны самостоятельно сформулировать гипотезу для каждого конкретного случая, подтвердить либо опровергнуть ее и сделать соответствующие выводы относительно возможной пользы проделанных операций для изучения того либо иного социального процесса или явления (исходя из содержания конкретной задачи).

Вопросы и задачи для самоконтроля

1. Определите следующие понятия: гипотеза, статистическая гипотеза, нуль-гипотеза, принятие решения, отношение значимости (критическое), выборочное распределение разности, Хи-квадрат (χ^2), проверка независимости, проверка однородности.

2. Кандидат заявил, что 60% избирателей должны голосовать за него. В выборке из 1000 зарегистрированных бюллетеней оказалось 55% голосов, отданных за него. С помощью метода χ^2 определите правдоподобность заявления кандидата.

3. Заданная выборка имеет следующие характеристики: $\bar{X} = 11$, $S = 3$, $N = 100$. Если известно, что истинное среднее значение равно 12, какова ошибка выборки для наблюдаемого среднего? Какова оцениваемая квадратическая ошибка среднего?

4. Может ли случайная выборка, правильно рассчитанная технически, оказаться нерепрезентативной?

Тема 8. Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии

План

1. Корреляционное поле как форма графического представления корреляционной зависимости. Виды корреляционного рассеивания.
2. Скедастичность (вариабельность).
3. Корреляционная таблица: техника группирования и основные функции.
4. Общая мера корреляции и ее необходимость.
5. Процедуры измерения линейной корреляции и вычисление наклона

линии регрессии.

6. Коэффициент корреляции как мера тесноты, типа и направления связи между двумя признаками.

7. Построение и вычисление уравнения регрессии.

Методические указания по изучению темы «Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии»

В рамках обозначенной темы следует обратить внимание на необходимость и информативность изучения взаимосвязи между признаками в социологическом исследовании. Мерой такой взаимосвязи является коэффициент корреляции. Студенты должны понять смысл термина «correlation» (корреляция). Данный термин состоит из приставки «со-», обозначающей совместность происходящего (по аналогии с «координация») и корня «relation», переводимого как «отношение» или «связь». Дословно correlation – взаимосвязь.

В эмпирической социологии измерение взаимосвязей осуществляется с использованием математических методов. Однако важно знать, что способы измерения связи могут различаться в зависимости от того, будут данные представлены в форме качественных признаков, которые просто перечислены, или в виде количественных измерений, и будет ли вид зависимости между переменными простым или сложным.

Студентам необходимо иметь в виду, что формулы измерения связи могут быть удобно сгруппированы на основе двух принципов связи: совместного появления и ковариации. Измерение ковариации (влияния) можно произвести предварительно с помощью корреляционного поля, которое относится к двумерной совокупности данных точно так же, как гистограмма к одномерной совокупности. Любая линия концентрации данных называется линией регрессии.

Следует помнить, что вычисление коэффициентов корреляции возможно с использованием стандартных программ. Итогом данной процедуры будет число, показывающее меру взаимообусловленности в распределении частот появления соответствующих признаков. Студентам необходимо знать, что **анализ коэффициентов связи позволяет:** 1) выделить факторы, уровень влияния которых слишком низок, что дает возможность исключить их из дальнейшего анализа (гипотеза о наличии связи отрицается); 2) проранжировать оставшиеся связи по уровню взаимной сопряженности (при этом следует иметь в виду, что уровень взаимной сопряженности может определяться как влиянием данного фактора на процесс, так и взаимным изменением данного фактора и процесса под влиянием третьего фактора).

Важно помнить, что тенденция рассеяния не всегда бывает линейной, чаще всего она бывает криволинейной и принимает форму одного из многочисленных видов кривых. Рассеяние значений Y , соответствующих данному значению X , называется **скедастичностью**. Если степень вариации значений (ши-

рина зоны рассеяния) одинакова для всех значений X , то можно говорить о том, что переменная Y *гомоскедастична* по отношению к X . *Гетероскедастичность* означает, что степень корреляции неодинакова для всей совокупности.

Связь между признаками изучает корреляционный анализ, а закономерность изменения – регрессионный анализ. **Корреляционная зависимость** – взаимосвязь между признаками, состоящая в том, что с изменением величины одного признака меняется величина другого. Как правило, при изучении взаимозависимости двух признаков различают: независимые признаки (факторные), которые, чаще всего, обозначаются – X и зависимые признаки (результатирующие) – Y . В ходе корреляционного анализа необходимо узнать, как под влиянием факторных признаков изменяется результирующий (если он изменяется вообще) и по какому закону. Корреляционный анализ фиксирует форму, направление и тесноту связи. По типу корреляционная связь может быть прямой или обратной, по форме – прямолинейной или криволинейной, по тесноте – тесной или слабой. Корреляция может быть также парной или множественной. Парная связь устанавливается между двумя признаками (факторным и результирующим). Множественная связь – между большим количеством факторных признаков и результирующим. Корреляционный анализ применяется для объектов, измеренных по интервальной или порядковой шкале, для количественных признаков. Студентам важно знать, что **все характеристики корреляционного анализа определяются тремя коэффициентами: коэффициент корреляции – r ; коэффициент регрессии $R_{y/x}$; корреляционное отношение η^2** (данный коэффициент будет подробно рассматриваться в рамках следующей темы).

Необходимо запомнить, что **коэффициент корреляции** – это мера тесноты и направления между двумя признаками при линейной связи, обладающая определенными *свойствами*:

1) $r_{y/x} = r_{x/y}$;

2) чем ближе $r_{y/x}$ к 1 или -1 , тем теснее связь. Чем ближе $r_{y/x}$ к 0, тем связь слабее. Если $r_{y/x} = 0$, то говорят, что связи нет. Для социальных процессов $r_{y/x}$ редко превышает $|0,75|$;

3) если $r_{y/x} > 0$, то связь прямая. Если $r_{y/x} < 0$, то связь обратная.

Формула вычисления r может быть представлена в следующем виде:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \cdot \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}}$$

Следует иметь в виду, что для сгруппированных данных формула выглядит несколько иначе, а именно:

$$r = \frac{\frac{1}{n} \sum (x_i - \bar{x}) \cdot (y_i - \bar{y}) \cdot n_{ixy}}{\sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot n_{ix}}{n}} \cdot \sqrt{\frac{\sum (y_i - \bar{y})^2 \cdot n_{iy}}{n}}}$$

Статистическая зависимость одного или большего числа признаков от остальных выражается с помощью уравнений регрессии. **Уравнение регрессии** описывает числовое соотношение между величинами, выраженное в виде тенденции к возрастанию (или убыванию) одной переменной величины при возрастании (убывании) другой. Для линейной регрессии уравнение имеет вид: $\hat{y}_x = R_{y/x} \cdot (x - \bar{x}) + \bar{y}$ или $\hat{x}_y = R_{x/y} \cdot (y - \bar{y}) + \bar{x}$, где $R_{y/x}$, $R_{x/y}$ – коэффициенты регрессии, которые показывают, насколько в среднем изменится результирующий признак, если факторный изменить на единицу.

Коэффициенты $R_{y/x}$ и $R_{x/y}$ вычисляются по следующим формулам:

$$R_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x}, \quad R_{x/y} = r \cdot \frac{\sigma_x}{\sigma_{yx}}$$

Следует запомнить основные *свойства коэффициента регрессии*:

- 1) коэффициент регрессии принимает любые значения;
- 2) коэффициент регрессии не симметричен, т.е. изменяется, если X и Y поменять местами;
- 3) единицей измерения коэффициента регрессии является отношение единицы измерения Y к единице измерения X ($[Y]/[X]$);
- 4) коэффициент регрессии изменяется при изменении единиц измерения X и Y .

В целом, для закрепления знаний, полученных в рамках изучения данной темы и выработки практических навыков поиска и определения взаимосвязей между признаками, на практическом занятии студентам предлагается решить ряд задач, направленных на вычисление коэффициента корреляции (r) и коэффициента регрессии $R_{y/x}$. Важно, чтобы студенты не только усвоили соответствующие математические процедуры вычисления, но и умели интерпретиро-

вать полученные числовые значения, показывающие факт наличия/отсутствия связи и описывающие ее характер (в случае наличия).

Вопросы и задания для самоконтроля

1. Определите следующие понятия: линия регрессии; линия наименьших квадратов; коэффициент корреляции, коэффициент регрессии.
2. Предположим, что $r = 0,3$ для связи между школьными оценками и часами подготовки. Проанализируйте эту «низкую корреляцию».
3. Для использования r требуется, чтобы рассеяние наблюдений относительно линии регрессии было гомоскедастичным. Аргументируйте это утверждение.
4. Возможно ли идеальное прямолинейное рассеяние, если маргинальные распределения являются несхожими? Проиллюстрируйте свой ответ с помощью таблицы или рисунка (графика).
5. Гарантирует ли наличие гомоскедастичности в рассеянии точек линейный вид зависимости? Коротко обоснуйте.

Тема 9. Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция

План

1. Особенности нелинейной регрессии.
2. Вычисление корреляционного отношения. Сравнение статистических показателей r^2 и η^2 .
3. Корреляция между двумя и более величинами. Частная и множественная регрессии.
4. Множественная корреляция.

Методические указания по изучению темы «Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция»

Из предыдущих занятий студенты должны были усвоить, что применимость r базируется в основном на следующих двух условиях: 1) отношение между переменными линейно; 2) двумерное распределение гомоскедастично.

Однако важно знать, что очень часто исходные данные приводят к нелинейным моделям. В рамках данной темы как раз и рассматриваются математические методы, с помощью которых можно измерить криволинейные зависимости, изучаемые социологами. Один из них – это метод корреляционного отношения, который часто обозначается через η (греческая строчная буква «эта»), но лучше выражается посредством η^2 .

$$\eta_{yx}^2 = \frac{\text{объясняемая_вариация}}{\text{полная_вариация}} = \frac{\sum (y_i - \mathcal{E})^2}{\sum (y_i - \bar{y})^2}, \text{ где } \mathcal{E} - \text{внутригрупповое среднее значение признака } y;$$

\bar{y} – общее среднее арифметическое признака y . Данная формула представляет собой вычисление η^2 для негруппированных данных.

В случае, если данные сгруппированы, эта формула выглядит следующим образом:

$$\eta_{yx}^2 = \frac{\text{объясняемая_вариация}}{\text{полная_вариация}} = \frac{\sum (y_i - \mathcal{E})^2 \cdot n_y}{\sum (y_i - \bar{y})^2 \cdot n_y}.$$

Студенты должны запомнить, что для вычисления η^2 необходимо найти среднее значение в каждом столбце (строке) корреляционной таблицы, общее среднее этих значений и отклонения средних значений для соответствующих колонок от общего среднего.

Для более эффективного усвоения материала данной темы следует провести **сравнение статистических показателей r^2 и η^2** :

$r^2 = 0$, если X и Y независимы.

$r^2 = \eta_{x/y}^2 = 1$, тогда и только тогда, когда имеется строгая линейная связь между X и Y .

$r^2 \leq \eta_{x/y}^2$ – когда имеется строгая нелинейная функциональная зависимость X и Y .

$r^2 = \eta_{x/y}^2 < 1$ – когда регрессия X и Y строго линейная, но нет функциональной зависимости.

$r^2 < \eta_{x/y}^2 < 1$ означает, что нет функциональной зависимости и существует нелинейная кривая регрессия.

Важно также знать, что в случае, когда $r^2 \leq \eta_{x/y}^2$ – между признаками существует строгая нелинейная связь. Задача социолога отыскать уравнение функциональной зависимости, определяющее эту связь, то есть уравнение регрессии. В рамках курса «Математические методы в социологии» рассматриваются только два типа криволинейной зависимости: параболический и гиперболический. Соответствующие уравнения имеют следующий вид:

$$\bar{y}_x = a + b \cdot x + c \cdot x^2 - \text{уравнение, определяющее } \textit{параболу}.$$

Для нахождения параметров a, b, c необходимо решить систему уравнений:

$$\begin{cases} \sum y = na + b \sum x + c \sum x^2 \\ \sum yx = a \sum x + b \sum x^2 + c \sum x^3 \\ \sum yx^2 = a \sum x^2 + b \sum x^3 + c \sum x^4. \end{cases}$$

В случае если корреляционное поле представляет собой кривую, которая может быть описана *гиперболой*, поиск ее уравнения ведется по формуле: $y_x = a + \frac{b}{x}$. Для нахождения параметров a, b необходимо решить систему уравнений:

$$\begin{cases} \sum \frac{y}{x} = a \sum \frac{1}{x} + b \sum \frac{1}{x^2} \\ \sum y = an + b \sum \frac{1}{x} \end{cases} \quad \text{— формулы для несгруппированных данных.}$$

Студенты должны знать, что в социологической практике часто возникает необходимость в изучении влияния нескольких переменных x_1, x_2, \dots, x_k на результирующий признак y . В этой связи нужно уметь строить регрессионное уравнение более общего вида, то есть $y = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$, где a, b_1, b_2, \dots, b_k — постоянные коэффициенты, называемые частными коэффициентами регрессии.

Следует рассмотреть *пример построения двухфакторного регрессионного уравнения*. Предположим, что изучается зависимость признака (y) от признаков (x_1) и (x_2). Нужно найти эту зависимость в виде линейного уравнения следующего вида: $y = a + b_1x_1 + b_2x_2$.

При расчете коэффициентов уравнения множественной регрессии полезно преобразовать исходные эмпирические данные следующим образом:

$z_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i}$, при этом уравнение множественной регрессии примет вид:

$y = c_1z_1 + c_2z_2$, где c_1 и c_2 находятся из системы уравнений

$$\begin{cases} c_1 + r_{12}c_2 = r_{1y} \\ c_1r_{12} + c_2 = r_{2y} \end{cases}, \text{ решая которую, получаем, что } c_1 = \frac{r_{1y} - r_{12}r_{2y}}{1 - r_{12}^2}, c_2 = \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2},$$

где r — коэффициент парной корреляции между признаками.

c_1 и c_2 называются стандартизированными коэффициентами регрессии. Следовательно, зная коэффициенты корреляции между изучаемыми признаками, можно подсчитать коэффициенты регрессии.

Коэффициенты a, b_1, b_2, \dots, b_k исходного регрессионного уравнения находятся по формулам:

$$b_1 = c_1 \frac{\sigma_y}{\sigma_1}; \quad b_2 = c_2 \frac{\sigma_y}{\sigma_2}; \quad a = \bar{y} - b_1 x_1 - b_2 x_2.$$

Коэффициенты b_1, b_2 можно рассматривать как показатели тесноты связи между переменными Y и, например, x_1 , при постоянстве x_2 .

Частный коэффициент корреляции – показатель, который характеризует тесноту и направление связи между результирующим признаком (Y) и факторным признаком (x_i) при элиминировании остальных признаков.

Частный коэффициент корреляции записывается $r_{y1.2}$ и вычисляется по формуле:

$$r_{y1.2} = \frac{r_{y1} - r_{y2} r_{12}}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}}.$$

Необходимо запомнить, что для определения степени связи результирующего признака Y с совокупностью независимых переменных вычисляется множественный коэффициент корреляции $R_{y(1...k)}^2$, который иногда выражается в процентах и вычисляется по формуле:

$$1 - R_{y(1...k)}^2 = (1 - r_{y1}^2)(1 - r_{y2.1}^2) \dots (1 - r_{yn.23...(k-1)}^2).$$

Важно также запомнить: множественный коэффициент корреляции показывает, что включение признаков x_1 и x_2 в уравнение на определенное количество процентов объясняет изменчивость результирующего фактора. Чем больше $R_{y(1...k)}^2$, тем полнее независимые переменные x_1, x_2, \dots, x_k описывают признак Y . Обычно $R_{y(1...k)}^2$ служит критерием включения или исключения новой переменной в регрессионное уравнение. Если $R_{y(1...k)}^2$ мало изменяется при включении новой переменной в уравнение, то такая переменная отбрасывается.

В целом, для закрепления знаний, полученных в рамках изучения данной темы и выработки соответствующих практических навыков, на практическом занятии студентам предлагается решить ряд задач, условие которых предполагает вычисление корреляционного отношения, а также нахождение нелинейной регрессии и определение множественной и частной корреляции (по аналогии с приведенным выше примером).

Вопросы и задания для самоконтроля

1. Допустимо ли вычисление η^2 , если линия регрессии плавная?
2. Можно ли вычислить η^2 для двух качественных переменных? Почему?

3. При каких табличных условиях числовые значения r^2 и η^2 совпадают?
4. Изобразите графически условия, при которых значение r^2 приблизительно равнялось бы нулю, а $\eta^2 = 1,00$.
5. Поясните, почему гомоскедастичность не может быть одинаковой в обоих направлениях, если регрессия криволинейна? Изобразите графически.
6. Определите, что фиксирует частный коэффициент корреляции и каковы его свойства?
7. Опишите, для каких целей применяется множественный коэффициент корреляции?

Тема 10. Корреляция рангов: коэффициент r_s

План

1. Измерение взаимосвязи признаков с помощью рангов.
2. Метод корреляции рангов (r_s) – коэффициент Спирмена.
3. Вычисление коэффициента Спирмена в случае объединенных рангов.
4. Коэффициент множественной корреляции качественных признаков.
5. Интерпретация коэффициентов ранговой корреляции.

Методические указания по изучению темы

«Корреляция рангов: коэффициент r_s »

Следует уточнить, что на предыдущих занятиях использование корреляционных методов ограничивалось измерением связи между количественными переменными на основе совместимых частот. Однако студентам важно знать, что качественные переменные также могут быть связаны (тогда говорят, что они ковариантны). Из всего множества известных корреляционных индексов в рамках данной темы рассматриваются только два – коэффициент парной ранговой корреляции (r_s) и (W) – коэффициент конкордации (определяющий степень множественной связи между качественными признаками). **Коэффициент ранговой корреляции** используется для измерения взаимозависимостей между качественными признаками, значения которых могут быть упорядочены или проранжированы по степени убывания (или возрастания) данного качества у исследуемых социальных объектов.

Следует знать, что формула для измерения степени корреляции между двумя рядами рангов была выведена Спирменом в 1904 году. Студентам нужно

запомнить эту формулу:
$$r_s = 1 - \frac{6 \cdot \sum d^2}{N \cdot (N^2 - 1)},$$

где: d – разность между парными рангами;

N – число ранжированных объектов.

Величина r_s может изменяться в пределах от +1 до -1, когда два ряда проранжированы в одном порядке. При полном взаимном беспорядочном расположении рангов r_s равен нулю. Значимость коэффициента корреляции Спирмена для $N \leq 100$ можно определить по таблице критических величин r_s . Если $N > 100$, то критические значения также находятся по таблице. При

этом наблюдаемые значения критерия вычисляются по формуле: $z = \frac{r_s}{\sqrt{N-1}}$.

Важно помнить, что ранги иногда могут объединяться. В таких случаях обоим наблюдениям приписывается значение среднего арифметического из двух объединенных рангов. Также необходимо понять, что r_s измеряет корреляцию между порядковыми рангами, а не между ранжируемыми величинами. В связи с этим нужно иметь в виду, что r_s преувеличивает степень связи между изучаемыми переменными.

Следует подчеркнуть тот факт, что формула для вычисления r_s пригодна только для двух ранговых рядов. Однако данные могут состоять из трех и более ранговых рядов. В такой ситуации одним из наиболее простых методов определения общей степени соответствия между тремя и более упорядоченными рядами данных является вычисление среднего арифметического из всех возможных значений r_s . Результат подобной операции носит название «**коэффициент конкордации**» (соответствия).

Студенты должны иметь в виду, что когда усредняется очень много значений r_s , вычисления становятся более трудоемкими. Для таких случаев разработан очень простой метод усреднения значений r_s , что имеет вид следующей формулы:

$W = \frac{12S}{k^2 \cdot N \cdot (N^2 - 1)}$, где k – число переменных; N – число

индивидов или категорий, которые ранжируются; $S = \sum (a - \bar{a})^2$ (сумма рангов по строке – a), \bar{a} – среднее из суммы рангов. Важно помнить, что значимость полученной величины W для $N > 7$ проверяется по критерию χ^2 : $\chi^2 = \frac{12S}{kN(N+1)}$ со степенью свободы $N - 1$.

Так как r_s применяется к порядковым данным, не имеющим определенных единиц измерения, этот показатель весьма полезен для социолога, которому часто приходится иметь дело с данными, носящими субъективный характер. Ранжирование предпочтений, эстетических суждений и других аналогичных явлений – все это можно коррелировать для того, чтобы определить степень со-

гласия в предпочтениях респондентов и оценках экспертов.

В целом, для закрепления знаний, полученных в рамках изучения темы «Корреляция рангов: коэффициент r_s », и выработки навыков определения корреляции качественных переменных, на практическом занятии студентам предлагается ряд задач социологического содержания, направленных на вычисление коэффициентов парной ранговой корреляции (r_s) и коэффициента конкордации (W). Необходимо не только найти числовые значения этих коэффициентов, но и осуществить их социологическую интерпретацию в соответствии с условием задачи.

Вопросы и задания для самоконтроля

1. Определите следующие понятия: ранг, ранговый порядок, равные интервалы, порядковые числа, корреляция ранговых последовательностей.
2. Укажите два серьезных затруднения в интерпретации рангов в качестве аналогов социальных явлений, которые они призваны отобразить.
3. При каких обстоятельствах могут быть ранжированы качественные данные?
4. Являются ли различия между соседними рангами непременно одинаковыми? Приведите пример.
5. Определите различие между соответствием и корреляцией.
6. Проверьте на примере тот факт, что объединенные ранги уменьшают корреляцию.
7. Опишите алгоритм вычисления коэффициента Спирмена.
8. Объясните принцип вычисления множественного коэффициента корреляции для ранговых рядов.

Тема 11. Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции

План

1. Коэффициенты взаимной сопряженности для измерения связи между качественными признаками.
2. Вычисление коэффициента сопряженности К. Пирсона (С).
3. Общие принципы интерпретации коэффициентов корреляции.

Методические указания по изучению темы

«Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции»

К моменту изучения данной темы студенты должны были убедиться в том, что существует множество математических методов выявления корреляционных связей при анализе социологических данных. Причем одни пригодны для измерения взаимосвязей исключительно количественных признаков, другие –

для количественных и качественных, третьи – только для качественных. Данная тема посвящена последним в этом перечне, в частности, вычислению коэффициента взаимной сопряженности Пирсона, обычно обозначаемого как « C ».

Он основывается на отклонении наблюдаемых частот в клетках таблицы от ожидаемых частот и предполагает, что распределение носит случайный характер. Важно знать, что эти отклонения измеряются показателем Хи-квадрат (χ^2). Значение данного показателя и алгоритм его вычисления должны быть хорошо известны студентам, так как рассматривались в рамках одной из первых тем данного модуля. Студентам необходимо запомнить формулу, которая дает величину коэффициента сопряженности Пирсона (далее по тексту – C). Формула, в которой учтено изменяющееся число наблюдений N , дает нормализованный показатель связи:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Важно обратить внимание на тот факт, что если χ^2 велико по сравнению с N , то C будет стремиться к единице, так как числитель и знаменатель фактически будут равны; однако, если χ^2 мало по сравнению с N , то коэффициент C будет также мал и в пределе будет стремиться к нулю. Если $\chi^2 = 0$ (то есть если нет никакого расхождения между полученными данными и чисто случайным распределением), коэффициент C также будет равен нулю, потому что числитель равен нулю.

Следует запомнить формулу вычисления теоретического значения C_{\max} на основе информации только о числе строк (или столбцов) при условии, что таблица является квадратной, а распределения маргиналов идентичны. Для

квадратных таблиц уместно применить следующую формулу: $C_{\max} = \sqrt{\frac{t-1}{t}}$, где

t – число строк. C_{\max} является действительным эквивалентом полной корреляции; следовательно, этот показатель может быть использован в качестве стандарта для любой связи, которая меньше, чем полная при условии, что распределения маргиналов являются идентичными. Отношение данного значения к максимальному значению C является, таким образом, приблизительно эквивалентным условной мере связи, принимая значения от «0» до «1». Эта корректировка становится все меньше и меньше по мере того, как увеличиваются размеры таблицы.

В рамках заключительной темы модуля следует подвести итог, сделав вывод о научном и практическом значении корреляционного анализа в социологии. При этом студентам необходимо самостоятельно вывести некоторые общие принципы интерпретации всех изученных коэффициентов корреляции.

В целом, для закрепления знаний, полученных в рамках изучения данной темы и выработки соответствующих практических навыков, на практическом

занятия студентам предлагается ряд задач социологического содержания, направленных на выявление взаимосвязей между признаками путем вычисления коэффициента сопряженности Пирсона (C). Помимо нахождения числового значения этого коэффициента, необходимо осуществить его интерпретацию и сделать выводы согласно условию конкретной задачи.

Вопросы и задания для самоконтроля

1. Определите следующие понятия: коэффициент взаимной сопряженности, таблица взаимной сопряженности, многоуровневая группировка, C_{\max} .
2. Повторите алгоритм вычисления Хи-квадрат (χ^2). Как данный критерий связан с коэффициентом сопряженности Пирсона (C)?
3. Опишите формулу вычисления коэффициента сопряженности Пирсона (C).
4. Опишите формулу вычисления теоретического значения C_{\max} , при условии, что таблица является квадратной.
5. Каково научное и практическое значение коэффициентов корреляции в социологии?
6. Каковы общие принципы интерпретации коэффициентов корреляции?

МОДУЛЬ III. Математические процедуры в сетевом анализе и анализе социометрических данных

Тема 12. Элементы теории графов в сетевом анализе

План

1. Сетевой анализ как метод планирования работ проектного характера, его значение в управленческой деятельности.
2. Задачи, разрешаемые при помощи теории графов.
3. Основные понятия теории графов.
4. Проблема максимальных путей и контуров графов.
5. Свойства псевдопотенциальных графов.

Методические указания по изучению темы «Элементы теории графов в сетевом анализе»

Следует знать, что теория графов в качестве научной дисциплины может рассматриваться как раздел дискретной математики (точнее – теории множеств), исследующий свойства конечных множеств с заданными отношениями между их элементами. Начало теории графов датируют 1736 годом, и связывают с именем Л. Эйлера, разрешившего популярную на то время «задачу о ке-

нигсбергских мостах». Однако следует отметить, что термин «граф» впервые был введен лишь спустя 200 лет (в 1936 г.) Д. Кенигом.

Важно понять и запомнить, что язык графов оказывается удобным для описания многих физических, технических, экономических, биологических, социальных и других систем. К основным типам задач, решаемых при помощи теории графов, относятся следующие:

- транспортные задачи, в которых вершинами графа являются пункты, а ребрами – дороги (автомобильные, железные и др.) и/или другие транспортные (например, авиационные) маршруты;

- сети снабжения (энергоснабжения, газоснабжения, снабжения товарами и т. д.), в которых вершинами являются пункты производства и потребления, а ребрами – возможные маршруты перемещения (линии электропередач, газопроводы, дороги и т. д.);

- технологические задачи, в которых вершины отражают производственные элементы (заводы, цеха, станки и т. д.), а дуги – потоки сырья, материалов и продукции между ними, заключаются в определении оптимальной загрузки производственных элементов и обеспечивающих эту загрузку потоков;

- обменные схемы, являющиеся моделями таких явлений, как бартер, взаимозачеты и т. д. Вершины графа при этом описывают участников обменной схемы (цепочки), а дуги – потоки материальных и финансовых ресурсов между ними. Задача заключается в определении цепочки обменов, оптимальной с точки зрения, например, организатора обмена и согласованной с интересами участников цепочки и существующими ограничениями;

- управление проектами – раздел теории управления, изучающий методы и механизмы управления изменениями. **Проектом** называется целенаправленное изменение некоторой системы, осуществляемое в рамках ограничений на время и используемые ресурсы; характерной чертой любого проекта является его уникальность, то есть нерегулярность соответствующих изменений. С точки зрения теории графов проект – совокупность операций и зависимостей между ними. Примером является проект строительства некоторого объекта. Совокупность моделей и методов, использующих язык и результаты теории графов и ориентированных на решение задач управления проектами, получила название *календарно-сетевое планирование и управления* (КСПУ). В рамках КСПУ решаются задачи определения последовательности выполнения операций и распределения ресурсов между ними, оптимальных с точки зрения тех или иных критериев (времени выполнения проекта, затрат, риска и др.);

- модели коллективов и групп, используемые в социологии, основываются на представлении людей или их групп в виде вершин. Отношения между ними (например, отношения знакомств, доверия, симпатии и т. д.) представляются в виде ребер или дуг. В рамках подобного описания решаются задачи исследования структуры социальных групп, их сравнения, определения агрегированных показателей, отражающих степень напряженности, согласованности взаимодействия и др.;

• модели организационных структур, в которых вершинами являются элементы организационной системы, а ребрами или дугами – связи (информационные, управляющие, технологические и др.) между ними.

Необходимо запомнить определение термина «**граф**», описывающее его как систему, которая интуитивно может быть рассмотрена в виде множества кружков и соединяющих их линий (см. Рис. 1).

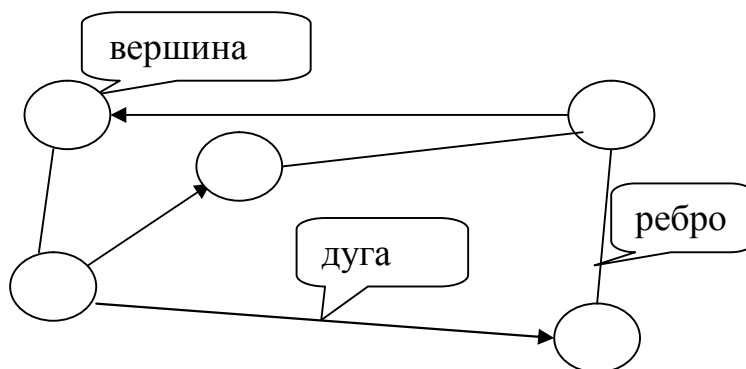


Рис.1. Пример графа

Кружки называются **вершинами графа**, линии со стрелками – **дугами**, без стрелок – **ребрами**.

Граф, в котором направление линий не выделяется (все линии являются ребрами), называется **неориентированным**; граф, в котором направление линий принципиально (линии являются дугами) называется **ориентированным**.

Подграфом называется часть графа, образованная подмножеством вершин вместе со всеми ребрами (дугами), соединяющими вершины из этого множества. Если из графа удалить часть ребер (дуг), то получим **частичный граф**.

Две вершины называются **смежными**, если они соединены ребром (дугой). Смежные вершины называются граничными вершинами соответствующего ребра (дуги), а это ребро (дуга) – инцидентным соответствующим вершинам.

Путем называется такая последовательность дуг (в ориентированном графе), что конец одной дуги является началом другой дуги. **Простой путь** – путь, в котором ни одна дуга не встречается дважды. **Элементарный путь** – путь, в котором ни одна вершина не встречается дважды. **Контур** – путь, у которого конечная вершина совпадает с начальной вершиной.

Длиной пути (контура) называется число дуг пути (или сумма длин его дуг, если последние заданы).

В целом, для закрепления лекционного материала по теме «Элементы теории графов в сетевом анализе» на практическом занятии студентам предлагается подготовить ряд сообщений, содержательно соответствующих основным вопросам, раскрываемым в рамках данной темы (см. выше). При этом студенты могут использовать литературные источники и Интернет-ресурсы, перечень которых дается на стр. 17–19 данного методического пособия. Для выработки практических навыков предлагается выполнить задание, целью которого является разрешение конкретной

социологической и/или управленческой проблемы с привлечением теории графов. Студенты должны самостоятельно определить соответствующую проблемную ситуацию из тех, которые являются актуальными на сегодняшний день (разрешение которой, возможно, представляет личный интерес). В качестве метода ее разрешения необходимо построить граф (один или несколько) и на этой основе сделать выводы.

Вопросы и задания для самоконтроля

1. Подумайте и опишите случаи использования теории графов для изучения конкретных социальных процессов или явлений, разрешения тех или иных социологических проблем.
2. Определите основные понятия теории графов.
3. Какова история развития теории графов?
4. В чем различия между стрелочными и вершинными графами?
5. В чем заключается метод прогнозного графа и как он может быть применим в эмпирической социологии?
6. Приведите примеры сетевого планирования с задействованием элементов теории графов.

Тема 13. Обработка данных в социометрическом исследовании: вычисление индексов, построение социограмм

План

1. Социометрия как метод изучения отношений в малых группах. Социологическая сущность данных, полученных в ходе социометрического исследования.
2. Новые методики социометрического сбора данных.
3. Структура данных, собранных методом социометрического опроса.
4. Бинарные отношения.
5. Социограмма и социометрические индексы.

Методические указания по изучению темы

«Обработка данных в социометрическом исследовании: вычисление индексов построение социограмм»

К моменту изучения данной темы студенты уже должны быть ознакомлены с методом социометрии, рассматриваемым также и другими дисциплинами. Однако следует подчеркнуть социологическую значимость этого метода: социометрия позволяет изучать отношения в малых группах, которые являются неотъемлемыми элементами общества. Непосредственное воздействие социальных факторов микросреды малых групп имеет большое значение, поскольку оно опосредует действие факторов макроуровня, преломляя их через свою призму. Следует помнить, что термин «социометрия» был предложен Дж. Морено в 30-е годы нашего столетия. Важно знать, что

Морено также разработал особую социопсихологическую теорию, согласно которой изменение психологических отношений в малой группе является якобы главным условием изменений во всей социальной системе.

Необходимо иметь в виду, что в целом социометрические методы направлены на исследование структуры межличностных отношений в малой социальной группе путем изучения выборов, сделанных членами группы по тем или иным социометрическим критериям. Социометрическая техника применяется для диагностики межличностных и межгрупповых отношений в целях их изменения, улучшения и совершенствования. С помощью социометрии можно изучать типологию социального поведения людей в условиях групповой деятельности, судить о социально-психологической совместимости членов конкретных групп, что особенно важно в социальном менеджменте. Важно знать, что, наряду с классическими, существуют новые методики проведения социометрии. Например, «референтометрия» (позволяющая выделять референтные группы, то есть определять для каждого респондента тех членов группы, мнение которых его в наибольшей степени интересует) или методика «выбора в действии» (исследователь получает информацию

о выборах членов группы путем наблюдения в реальной или специально сконструированной им экспериментальной ситуации).

Следует помнить, что социометрическая процедура может проводиться в двух формах: параметрической (когда число выборов ограничивается) и непараметрической (когда число выборов не ограничено). Также необходимо знать, что простейшими способами количественной обработки являются: *табличный* (социоматрица), *графический* (социограмма) и *индексологический* (персональные и групповые индексы).

Основное достоинство **социоматрицы** – возможность представить выборы в числовом виде, что в свою очередь позволяет проранжировать членов группы по числу полученных и отданных выборов, установить порядок влияний в группе. На основе социоматрицы строится **социограмма** – карта социометрических выборов – графическое изображение реакции испытуемых друг на друга при ответах на социометрический критерий. Социограмма позволяет произвести сравнительный анализ структуры взаимоотношений в группе в пространстве на некоторой плоскости («щите») с помощью специальных знаков. Анализ социограммы заключается в определении центральных, наиболее влиятельных членов, затем взаимных пар и группировок.

Студенты должны знать персональные (П.С.И.) и групповые (Г.С.И.) **социометрические индексы**. Первые характеризуют индивидуальные социально-психологические свойства личности в роли члена группы. Вторые дают числовые характеристики целостной социометрической конфигурации выборов в группе. Основными П.С.И. являются: индекс социометрического статуса i -члена; эмоциональной экспансивности j -члена, объема, интенсивности и концентрации взаимодействия ij -члена. Символы i и j обозначают одно и то же ли-

цо, но в разных ролях; i – выбираемый, j – он же выбирающий, ij – совмещение ролей.

Индекс социометрического статуса i -члена группы определяется по

формуле:
$$C_i = \frac{\sum_{i=1}^N (R_i^+ + R_i^-)}{N-1},$$

где: C_i – социометрический статус i -члена;

R^+ и R^- – полученные i -членом выборы;

N – число членов группы.

Возможен также расчет **C -положительного и C -отрицательного статуса** в группах малой численности (N).

Индекс эмоциональной экспансивности j -члена группы высчитывается

по формуле
$$E_j = \frac{\sum_{j=1}^N (R_{ji}^+ + R_j^-)}{N-1},$$

где: E_j – эмоциональная экспансивность j -члена;

R_j – сделанные /членом выборы (+, -).

С психологической точки зрения показатель экспансивности характеризует потребность личности в общении.

Из Г.С.И. наиболее важными являются: индекс эмоциональной экспансивности группы и индекс психологической взаимности. Индекс **эмоциональ-**

ной экспансивности группы вычисляется по формуле:
$$Ag = \frac{\sum_{i=1}^N (\sum_{j=1}^N R_{ij}^i)}{N-1},$$

где: Ag – экспансивность группы;

N – число членов группы;

R_j (+,-) – сделанные j -членом выборы.

Индекс показывает среднюю активность группы при решении задачи социометрического теста (в расчете на каждого члена группы).

Индекс психологической взаимности («сплоченности группы») в группе

высчитывается по формуле:
$$G_g = \frac{\sum_{ij=1}^N (\sum_{jj=1}^N A_{ij}^+)}{\frac{1}{2} \cdot N(N-1)},$$

где: G_g – взаимность в группе по результатам положительных выборов;

A_{ij}^+ – число положительных взаимных связей в группе;

N – число членов группы.

В целом, для закрепления знаний, полученных в рамках данной темы и выработки практических навыков, на практическом занятии студентам предлагается ряд заданий, выполнение которых предполагает: 1) построение социоматриц; 2) построение различных видов социограмм; 3) вычисление социометрических индексов. Помимо осуществления перечисленных выше операций, необходимо обосновать их необходимость и целесообразность в каждом кон-

кретном случае. Кроме того, возможно привлечение студентов к осуществлению самостоятельного исследования с использованием социометрических методик, рассмотренных в рамках лекционного занятия. Необходимо организовать и провести подобное исследование, а также осуществить глубокий анализ данных и сделать основные выводы с выходом на рекомендации.

Вопросы и задания для самоконтроля

1. Определите следующие понятия: социометрия, социоматрица, социограмма, социометрический индекс.
2. Объясните необходимость применения социометрического метода для изучения социальных процессов и явлений.
3. Покажите динамику изменения внутригрупповых предпочтений, если в группу попадает новый лидер, аутсайдер.
4. Обоснуйте специфику взаимоотношений объект-объект в коэффициентах социометрического анализа.
5. Перечислите основные социометрические индексы и объясните их значения.
6. Определите основные характеристики современных методов социометрии, проанализируйте их преимущества и недостатки.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ И ИНДИВИДУАЛЬНОЙ РАБОТЫ СТУДЕНТОВ

<i>Тема</i>	<i>Описание задания</i>
<i>МОДУЛЬ I. Применение процедур математической статистики в эмпирической социологии. Числовые характеристики выборки</i>	
1. Основные принципы измерения в социологии. Типы переменных	1. Подготовить эссе на тему: «Проблемы измерения в социологии» 2. Подготовить сообщение на тему: «Типы переменных в социологическом исследовании».
2. Перекрестная классификация. Графическое представление социологической информации	1. Найти дискретный ряд, который описывает объект исследования (далее объект). Построить для него гистограмму и кумуляту. 2. Найти непрерывный ряд, который описывает объект. Построить для него гистограмму, полигон и кумуляту. 3. Осуществить визуализацию данных конкретного социологического исследования с помощью таблиц и графиков.
3. Характеристики положения (среднее арифметическое, мода, медиана и др.)	Для конкретно заданного дискретного и непрерывного ряда вычислить: – среднее арифметическое (\bar{X}); – моду (Mo); – медиану (Me); – дисперсию (σ^2); – среднее квадратичное отклонение (σ); – коэффициент вариации (V); – асимметрию (As). Проанализировать полученные результаты.
4. Характеристики рассеяния (дисперсия, среднее линейное и квадратическое отклонение, коэффициент вариации)	
5. Проверка процедуры первичного измерения на надежность с использованием математических методов	Проверить надежность данных, полученных в ходе социологического исследования: 1) путем выявления систематических ошибок; 2) путем проверки шкал на устойчивость. Сделать соответствующие выводы, в том числе и относительно обоснованности процедуры первичного измерения.
<i>МОДУЛЬ II. Измерение связи между признаками с использованием математических методов</i>	
6. Нормальное распределение как модель вариации. Критерии линейной взаимосвязи	1. С помощью критерия Хи-квадрат (χ^2) определить соответствие заданного эмпирического распределения теоретическому. Сделать соответствующие выводы о вероятности взаимосвязи между признаками.

	2. Вычислить Хи-квадрат (χ^2) для квадратной матрицы. Проанализировать результаты.
7. Статистическая гипотеза. Проверка статистических гипотез при анализе социологических данных	1. Построить корреляционное поле. Сформулировать гипотезу относительно связи между признаками. 2. Сформулировать гипотезу для конкретно заданного случая, осуществить ее проверку с помощью критерия Хи-квадрат (χ^2).
8. Меры взаимосвязи для интервального уровня измерения. Корреляционный анализ. Уравнение регрессии	Найти коэффициент корреляции (r) между признаками X и Y, составить уравнение регрессии ($R_{y/x}$). Проанализировать полученные результаты и сделать выводы относительно силы и характера связи.
9. Корреляционное отношение. Нелинейная регрессия. Множественная и частная корреляция	1. Определить наличие и силу связи между признаками x и y путем вычисления корреляционного отношения (η^2); 2. Для конкретно заданного примера вычислить стандартизированные коэффициенты регрессии (C_1 и C_2); 3. Для конкретно заданного примера определить частный коэффициент корреляции ($r_{y1.2}$). По каждому из заданий проанализировать полученные результаты в контексте социологической науки, сделать соответствующие выводы.
10. Корреляция рангов: коэффициент r_s	Определить уровень согласованности между двумя событиями, вычислив коэффициент ранговой корреляции Спирмена (r_s). Сделать советующие выводы.
11. Коэффициент взаимной сопряженности. Некоторые общие принципы интерпретации коэффициентов корреляции	1. Найти трехмерное распределение, которое описывает объект исследования и имеет качественное измерение. 2. Найти коэффициент Пирсона (C) между значениями первого и второго признака заданного примера. Проанализировать полученные результаты.
<i>МОДУЛЬ III. Математические процедуры в сетевом анализе и анализе социометрических данных</i>	
12. Элементы теории графов в сетевом анализе	1. Составить граф, описывающий объект исследования, в котором не менее семи вершин. Проанализировать интенсивность и направленность информационных потоков внутри графа, оценить геометрию графа в целом.

	<p>2. Выбрать актуальную проблему, связанную с сетевым планированием, построить граф (один или несколько), на этой основе описать путь разрешения данной проблемы.</p>
<p>13. Обработка данных в социометрическом исследовании: вычисление индексов, построение социограмм</p>	<p>1. Для заданных примеров просчитать персональные и групповые социометрические индексы. Сделать соответствующие выводы.</p> <p>2. Провести социометрию в конкретной малой социальной группе: построить социоматрицу, социограмму, вычислить необходимые социометрические индексы. Сделать соответствующие выводы с выходом на рекомендации.</p>

ПЕРЕЧЕНЬ ВОЗМОЖНЫХ ТЕМ ЭССЕ И СООБЩЕНИЙ ПО КУРСУ

1. Математика в профессиональной деятельности социолога.
2. Возможности применения математических методов в практике управления персоналом.
3. Математические методы и социальное управление.
4. Основные виды шкал и типы переменных в социологии.
5. Правила вычисления моды и медианы: разные точки зрения.
6. Случаи необходимости визуализации социологической информации.
7. Правила графической и табличной визуализации результатов социологического исследования.
8. Необходимость определения характеристик положения в выборочном социологическом исследовании.
9. Необходимость определения характеристик рассеяния в выборочном социологическом исследовании.
10. От чего «отклоняется» среднее квадратическое отклонение?
11. Различные интерпретации понятия «надежность».
12. Роль математических методов при проверке процедуры первичного измерения на надежность.
13. Основные характеристики нормального распределения.
14. Может ли распределение быть «не нормальным»?
15. Понятие корреляции в социологии.
16. Зачем искать связь между признаками?
17. Статистические и нестатистические гипотезы в социологическом исследовании.
18. Линейная и нелинейная регрессия.
19. Суть вычислительных процедур измерения корреляции при линейной и нелинейной регрессии.
20. Карл Пирсон и его вклад в эмпирическую социологию.
21. «Регрессия», «корреляция», «детерминация», «сопряженность»: определение сущности понятий.
22. Коэффициенты корреляции, используемые при изучении социальных явлений и процессов: история применения.
23. Спирмен Ч. Э. и его вклад в изучение взаимосвязи между признаками.
24. Понятие микросоциологии в работах Дж. Морено.
25. Социометрия как один из способов получения информации о взаимоотношениях в малых группах.
26. «Социономия» и «социометрия» в понимании Дж. Морено.
27. Математические методы изучения отношений в малых группах.
28. Применение теории графов в социологии.
29. Особенности построения и типы графов.
30. Назначение и сущность сетевого анализа.
31. Применение теории графов в сфере социального управления.

ПЕРЕЧЕНЬ ВОПРОСОВ ДЛЯ ПОДГОТОВКИ К ЭКЗАМЕНУ

1. Охарактеризуйте процесс измерения с различных методологических позиций.
2. Дайте определения шкалам, их типам. Раскройте потенциал разных шкал в измерении тех или иных характеристик объекта исследования.
3. Назовите виды переменных и рядов распределения.
4. Обозначьте отличия в понятиях «ошибка» и «погрешность».
5. Определите сущность и виды статистической группировки.
6. Определите специфику и различия интервальных и дискретных рядов распределения.
7. Дайте определения частоты и частости, приведите примеры.
8. Определите понятие измерения в социологии.
9. Опишите в общих чертах методы измерения, которые используются в социологических исследованиях.
10. Охарактеризуйте типы перекрестной классификации, приведите примеры.
11. Обозначьте особенности построения и анализа динамических рядов.
12. Определите специфику кумулятивных динамических рядов.
13. Дайте перечень общих принципов построения графиков (гистограмма, полигон, кумулята).
14. Приведите примеры необходимости отображения двух, трех и больше полигонов на одном графике.
15. Обоснуйте предназначение графиков качественных данных. Изобразите в общих чертах, как выглядят: диаграмма полос, круговая диаграмма, временные диаграммы.
16. Назовите основные числовые характеристики исследуемой совокупности. Обоснуйте необходимость выбора среднего (в частности, среднего арифметического или моды, или медианы).
17. Дайте определение среднего арифметического, опишите алгоритм его вычисления.
18. Обоснуйте, в чем заключается отличие вычисления средних величин для сгруппированных и несгруппированных данных.
19. Опишите процедуру взвешивания среднего.
20. Опишите алгоритм вычисления медианы для сгруппированных и несгруппированных данных.
21. Опишите процедуру вычисления медианы дискретных данных.
22. Дайте определение моды (как вероятностного среднего); назовите основные правила вычисления моды.
23. Охарактеризуйте бимодальное распределение. Приведите пример и обоснуйте, почему в некоторых распределениях проявляется бимодальность.
24. Приведите пример нахождения модального значения по гистограмме.
25. Обоснуйте значение моды и медианы в социологических исследованиях.

26. Обоснуйте выбор среднего в зависимости от цели усреднения и вида распределения.
27. Дайте определение понятию дисперсии и приведите примеры ее вычисления.
28. Дайте определение понятию среднего квадратичного отклонения. Приведите формулу вычисления.
29. Дайте определение понятию коэффициента вариации и обоснуйте необходимость его вычисления.
30. Назовите характеристики среднего квадратичного отклонения.
31. Приведите примеры вариации качественных переменных.
32. Определите понятие нормального частотного распределения.
33. Определите, в чем заключается общий смысл всех законов распределения.
34. Дайте характеристику такой особенности эмпирического распределения как асимметрия.
35. Определите, в чем заключается сущность энтропии и коэффициентов вариации.
36. Опишите алгоритм вычисления χ^2 .
37. Определите критерий линейной взаимосвязи.
38. Дайте общее описание способам измерения связи между признаками.
39. Дайте развернутую характеристику корреляционного поля как формы графического представления корреляционной зависимости.
40. Дайте характеристику скедастичности как вариабельности (в том числе, гомоскедастичности и гетероскедастичности).
41. Приведите пример корреляционной таблицы как формы представления сгруппированных данных, назовите основную функцию корреляционной таблицы.
42. Опишите алгоритм измерения линейной корреляции (вычисление наклона линии регрессии, коэффициента детерминации).
43. Определите, в чем заключается суть коэффициента корреляции. Назовите коэффициенты корреляции, которые показывают тип и направление связи между двумя признаками.
44. Отметьте особенности нелинейной регрессии.
45. Назовите алгоритм вычисления корреляционного отношения.
46. Дайте сравнительную характеристику статистических показателей r и r^2 .
47. Дайте сравнительную характеристику статистических показателей r^2 и η^2 .
48. Опишите условия применимости критерия корреляционного отношения.
49. Определите основные принципы интерпретации корреляционного отношения.
50. Приведите примеры видов нелинейной формы связи.
51. Приведите примеры корреляции между двумя и более величинами.
52. В чем заключается суть частной и множественной регрессии?
53. Определите суть множественной корреляции.

54. Опишите процедуру измерения взаимосвязи признаков с помощью рангов.
55. Приведите пример вычисления коэффициента Спирмена (корреляции рангов r_s).
56. Приведите примеры необходимости вычисления коэффициентов ранговой корреляции.
57. Дайте характеристику коэффициенту множественной корреляции признаков для порядкового уровня измерения как меры соответствия для трех и более ранговых рядов.
58. Определите особенности вычисления коэффициента множественной корреляции для номинального уровня измерения.
59. Назовите основные характеристики и принципы вычисления коэффициента взаимной сопряженности (C).
60. Уточните социологическое и статистическое содержание корреляции (случайная связь; бессмысленная корреляция; корреляция как доказательство наличия причины и следствия; корреляция как мера общих факторов; ложная корреляция).
61. Назовите основные показатели надежности измерения.
62. Определите компоненты надежного измерения.
63. Определите, в чем заключается суть правильности измерения (выявление систематических ошибок).
64. Определите, в чем заключается суть устойчивости измерения (показатель абсолютной устойчивости W , средняя квадратичная ошибка, относительные показатели ошибок).
65. Дайте характеристику обоснованности измерения как завершающего этапа подтверждения надежности измерения.
66. Сравните понятие статистической гипотезы с понятием научной гипотезы.
67. Сформулируйте понятие статистической гипотезы с привязкой к процедуре определения неизвестного параметра эмпирического распределения.
68. Опишите общую процедуру проверки статистических гипотез.
69. Назовите основной принцип проверки нулевой гипотезы.
70. Дайте общую характеристику ошибок, которые возникают в ходе проверки статистической гипотезы.
71. Назовите основные показатели надежности измерения.
72. Обоснуйте необходимость сравнения трех или более процентных отношений в ходе проверки гипотез.
73. Определите суть критерия Хи-квадрат (χ^2), его роль в проверке статистических гипотез.
74. Назовите основные проблемы интерпретации, которые возникают в ходе проверки статистических гипотез.
75. Приведите пример процедуры измерения взаимосвязи ранжированных признаков.

76. Обоснуйте необходимость проведения социометрических исследований в социологии.
77. Опишите возможности математических методов в осуществлении сетевого анализа.
78. Опишите общую процедуру обработки данных в социометрическом исследовании.
79. Определите специфику структуры данных, собранных методом социометрического опроса.
80. Приведите пример матрицы данных вида «объект-объект».
81. Определите некоторые понятия теории графов: вершина, дуга, граф, ориентированный граф, нагруженный граф.
82. Определите роль социограммы в социометрическом исследовании .
83. Приведите примеры нулевой гипотезы.
84. Определите преимущества и недостатки включения статистических методов в сетевой анализ.
85. Обоснуйте необходимость графического и табличного представления социологических данных.
86. Дайте общую характеристику теории графов как раздела дискретной математики.

Навчальне видання

МАТЕМАТИКО-СТАТИСТИЧНИЙ АНАЛІЗ В СОЦІОЛОГІЇ

Методичні рекомендації для студентів,
які навчаються за спеціальністю
054 – Соціологія

(російською мовою)

Автор-упорядник НЕЧИТАЙЛО Ірина Сергіївна

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

Підписано до друку 03.06.2017. Формат 60×84/16.
Папір офсетний. Гарнітура «Таймс».
Ум. друк. арк. 3,95. Обл.-вид. арк. 4,12.
Наклад 5 пр. Зам. №

План 2016/17 навч. р., поз. № 8 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