**КУРС**

**"Анализ данных в SPSS/Data analysis in SPSS"**

Автор

Кучмаева Оксана Викторовна, д.э.н., профессор, кафедра народонаселения

kuchmaeva@yandex.ru

**Общий объем курса - 24 часа.**

**Периодичность - 1 раз в неделю.**

**Формат проведения - практико-ориентированный.**

**Образовательные технологии: дискуссионные лекции с элементами мастер-классов, разбор творческих заданий, решение практических задач и пр.**

**Аудитория: любой курс обучения.**

**Тема 1. Сущность и основные направления обследований населения. Возможности использования специальных ППП для обработки данных.**

 Методы сбора количественной информации. Выборочные исследования. Понятие совокупности: изучаемая и выборочная. Вероятностные и невероятностные выборки. Выборочные оценки. Способы формирования выборки. Единицы наблюдения. Определение размера выборки, необходимого для достижения заданной точности оценок. Ошибка выборки. Особенности формирования инструментария исследования.

История проведения обследований населения. Выборочные исследования в России и за рубежом.

Выборочные социально-демографические обследования в России. Обследования бюджетов домашних хозяйств. Выборочные обследования рабочей силы. Российский мониторинг экономического положения и здоровья населения (РМЭЗ). Выборочные обследования Росстата.

Основные пакеты статистических прикладных программ для социальных исследований. Функции специальных ППП (Statistica, SPSS) в обработке данных выборочных исследований. Структура, модули SPSS. Подготовка данных. Ввод и сохранение данных. Шкалы измерений (количественная, порядковая, номинальная). Иерархическая упорядоченность шкал. Свойства шкал и их допустимые преобразования. Типы категоризации данных.

**Тема 2. Подготовка данных. Отбор и модификация данных.**

Выбор наблюдений. Редактор условий. Логические функции. Извлечение случайной выборки. Сортировка наблюдений. Разделение наблюдений на группы. Модификация данных. Вычисление новых переменных. Подсчет частоты появлений определенных значений. Перекодирование значений. Ручное перекодирование. Автоматическое перекодирование. Вычисление новых переменных в соответствии с определенными условиями. Формулировка условий. Создание индекса. Агрегирование данных. Ранговые преобразования. Типы рангов. Веса случаев. Коррекция при отсутствии репрезентативности. Причины и механизмы порождения пропусков данных. Возможности игнорирования пропусков. Методы заполнения пропущенных значений. Методы выявления аномальных значений. Применение устойчивых процедур оценивания. Анализ множественных ответов.

**Тема 3. Дескриптивная статистика. Таблицы сопряженности.**

Роль статистики в обработке результатов выборочных обследований. Микро- и метаданные. Области применения и границы применимости математико-статистических методов. Классификация статистических данных по числу переменных для каждого объекта, типу измерения этих переменных, наличию упорядочения результатов наблюдения во времени, по источнику данных. Сводка наблюдений. Вариационный ряд по исходным и сгруппированным данным. Табличное и графическое представление распределения данных. Описательная статистика. Показатели вариации. Дисперсия, вариационный размах, среднее абсолютное отклонение, квантильные размахи. Построение таблиц сопряженности. Графическое представление таблиц сопряженности.

**Тема 4. Параметрические и непараметрические критерии.**

Независимость переменных. Основные характеристики связи. Критерий независимости (критерий согласия χ2). Сравнение двух и нескольких выборок (зависимых и независимых). t-критерий. Статистические критерии для таблиц сопряженности. Коэффициенты корреляции (для номинальных и ранговых шкал). Меры тесноты связи между переменными. Требования, предъявляемые к мерам связи. Меры, основанные непосредственно на *χ*2: коэффициент Пирсона, коэффициент Чупрова, коэффициент Крамера. Свойства коэффициентов сопряженности. Максимальные достижимые значения коэффициентов сопряженности. Стандартные ошибки коэффициентов и их интервальные оценки. Недостатки коэффициентов сопряженности.

Простейшие меры тесноты связи (для дихотомических переменных). Отношение Связь между простейшими мерами связи и различия между ними. Стандартные ошибки простейших мер связи.

Коэффициенты с теоретико-вероятностной интерпретацией. Общий подход к построению коэффициентов. Интервальные оценки коэффициентов с теоретико-вероятностной интерпретацией. Меры связи для таблиц с порядковыми данными. *t*-меры Кендэла и их свойства. *d*-меры Сомерса. Мера Гудмена-Краскала и ее свойства.

Непараметрические критерии. Сравнение двух независимых выборок. Сравнение двух зависимых выборок. Сравнение более чем двух независимых выборок. Сравнение более чем двух зависимых выборок. Биномиальный тест. Анализ последовательностей.

**Тема 5. Корреляционно-регрессионный анализ.**

Сущность и задачи корреляционного анализа. Диаграммы рассеяния. Парные коэффициенты корреляции. Измерение степени тесноты статистической связи, «очищенной» от влияния посторонних признаков с помощью частных коэффициентов корреляции. Проверка существенности связи признаков. Доверительные интервалы для коэффициентов корреляции. Множественный коэффициент корреляции. Коэффициент детерминации. Проверка существенности статистической связи с совокупностью признаков.

Двухмерная модель регрессионного анализа: линейная и нелинейная модели регрессии. Методы линеаризации. Расчёт уравнения регрессии. Сохранение новых переменных. Построение регрессионной прямой.

Кривые роста в задачах прогнозирования, «фиктивные» переменные и их применение. Множественная линейная модель регрессии. Нелинейная регрессия (бинарная логистическая регрессия, мультиномиальная логистическая регрессия, порядковая регрессия, пробит-анализ, приближение с помощью кривых).

**Тема 6. Методы снижения размерности.**

Статистический подход в методе главных компонентов. Линейная модель метода главных компонентов. Вычисление главных компонентов и их графическая интерпретация. Информативность редуцированного признакового пространства. Регрессия на главные компоненты. Процедуры факторного анализа на основе метода главных компонентов в пакетах прикладных программ.

Интегральные показатели и их использование. Построение интегральных показателей на основе количественной информации. Экспертно-статистический метод построения интегральных показателей.

Роль и место непараметрических методов в структурном моделировании. Двухэтапный кластерный анализ. Классификация без обучения. Иерархический кластерный анализ. Метрики признакового пространства. Принципы измерения расстояния между группами объектов. Алгоритмы быстрого кластерного анализа, метод k-средних.

**Методическое обеспечение:**

**Список литературы**

**Основная литература**

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики: учебник для вузов: т.1, т.2. – М.: ЮНИТИ-ДАНА, 2001.
2. Бююль А., Цёфель П. SPSS: искусство обработки информации. СПб.: ДиаСофтЮП, 2005.
3. Крыштановский А.О. Анализ социологических данных. М.:ВШЭ, 2006
4. Наследов А. SPSS 19: профессиональный статистический анализ данных. – Спб.: Питер, 2011.
5. Статистика: учебник./ Под редакцией Мхитаряна В.С.- М.: Издательство Юрайт, 2016
6. Плис А.И., Сливина Н.А. Практикум по прикладной статистике в среде SPSS. – М.: Финансы и статистика, 2004.

**Дополнительная литература**

1. Айвазян С.А., Бежаева Э.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Методы исследования зависимостей. – М.: Финансы и статистика, 1983.
3. Айвазян С. А., Мхитарян В.С. Теория вероятностей и прикладная статистика.- М.: ЮНИТИ- ДАНА, 2001- 256с.
4. Аптон Г. Анализ таблиц сопряженности./Пер. с англ. – М.: Финансы и статистика, 1982
5. Бахметова Г.Ш. Сбор и обработка данных о населении. М., 2000.
6. Борисов В.А. Демография. М.: NOTA BENE, 2003.
7. Денисенко М.Б., Калмыкова Н.М. Демография. Учебное пособие. М., ИНФРА-М,2007.
8. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 2003.
9. Дуброва Т.А. Прогнозирование социально-экономических процессов. Статистические методы и модели. – М.: Маркет ДС, 2007.
10. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистические функции MS Excel в экономико-статистических расчетах.- М.: ЮНИТИ-ДАНА, 2003, с. 156
11. Кокрен У. Методы выборочного исследования. – М.: «Статистика», 1976.
12. Медков В.М. Демография. Учебное пособие. М., 2014.
13. Микроэкономическая статистика. Уч./под ред. С.Д.Ильенковой. – М.: Финансы и статистика, 2004 – 544 с.
14. Моосмюллер Г., Ребик Н.Н. Маркетинговые исследования с SPSS: Учебное пособие. – М.: ИНФРА-М, 2007. – с.118-127.
15. Мхитарян В.С., Архипова М.Ю., Дуброва Т.А., Сиротин В.П. и др. Эконометрика: учеб. / под ред. В.С. Мхитаряна. – М.: Проспект, 2008.
16. Ноэль Э. Массовые опросы. Введение в методику демоскопии. М.: Ава-Эстра, 1993.
17. Прикладная социология: методология и методы: Интерактивное учебное пособие / М.К. Горшков, Ф.Э. Шереги. М.: Институт социологии РАН, 2011.
18. Пузанова Ж.В., Троцук И.В., Витковская М.И. Практикум по курсу «Методология и методика социологических исследований». Учебное пособие. М.: Высшее образование и наука, 2009.
19. Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике. М.: ЮНИТИ, 1999.
20. Статистические методы анализа информации в социологических исследованиях. М. : Наука, 1979.
21. Сваффорд М.С., Косолапов М.С., Козырева П.М. Российский мониторинг экономического положения и здоровья населения (РМЭЗ): измерение благосостояния россиян в 90-е годы. // Мир России. Социология. Этнология. 1999, № 3, с.153-172.
22. Статистическое моделирование и прогнозирование. Учебное пособие (Под ред. А.Г. Гранберга). М.:Финансы и статистика, 1990.
23. Толстова Ю.Н. Анализ социологических данных. М.: Научный мир, 2000
24. Хили Дж. Статистика. Социологические и маркетинговые исследования. СПб.: Питер, 2005
25. Чуриков А.В. Основы формирования выборки: лекции для студентов направления 040100.62 (Социология) – М.: ГУ-ВШЭ, 2005
26. Холлендер М., Вулф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983
27. Agresti, A. and Finlay, B. Statistical Methods for the Social Sciences. 4th ed. Pearson Prentice Hall, 2009. Pp.301-320.
28. Babbie E. The practice of social research. Cengage Learning. 2015. PP. 88-155.
29. Bhattacherjee A. Research Design / Social Science Research: Principles, Methods, and Practices // Textbooks Collection. Book 3. 2012. PP. 1-65.
30. Creswell J.W. Research design: Qualitative, Quantitative, and Mixed Methods Approaches. Los Angeles [etc.] SAGE Publications, 2014.
31. Field, A. Discovering Statistics Using SPSS. 3th Ed. Sage, 2009.
32. Kline P. An easy guide to factor analysis. //Routledge, London & N.Y., 1994
33. Meulman, J.J., A.J. Van der Kooij, and W.J. Heiser. 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In: Handbook of Quantitative Methodology for the Social Sciences, D. Kaplan, eds. Thousand Oaks, Calif.: Sage Publications, Inc.
34. Theory and Methods in Social Research / Ed. by B. Somekh, C. Lewin. 2th Edition. SAGE Publications, Inc., 2011.

***Интернет-ресурсы:***

## Процедуры и методы анализа данных <http://rotmistrov.com/methods>

## Иллюстрированный самоучитель по SPSS <http://computers.plib.ru/math/SPSS/>

* + Информационно-библиографическая система ООН <http://unbisnet.un.org>
	+ Международная организация труда <http://www.ilo.org>
	+ Институт сравнительных социальных исследований: <http://cessi.ru/>
	+ Архив данных по социологии www.icpsr.umich.edu
	+ Система организации фактографической информации по социологической тематике http://sofist.socpol.ru
	+ Бюро Цензов США, включая International Data Base: <http://www.census.gov>
	+ Всемирная организация здравоохранения (World Health Organization): <http://www.who.int/whosis>
	+ Статистическая служба Европейского союза (Eurostat) <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home>
	+ Мировой Банк (World Bank): <http://www.worldbank.org/data>
	+ Отдел народонаселения ООН (United Nations Population Division), данные и публикации, включая World Population Prospects (2004 Revision): <http://www.unpopulation.org>
	+ Фонд по народонаселению ООН (United Nations Fund for Population Activities): <http://www.unfpa.org/index.htm>
	+ Официальный сайт Росстата: <http://www.gks.ru>
	+ Информационно-библиографическая система ООН <http://unbisnet.un.org>
	+ Программа развития ООН: <http://www.undp.ru>
	+ Институт сравнительных социальных исследований: <http://cessi.ru/>
	+ Российский демографический интернет-еженедельник «Демоскоп»**:** <http://demoscope.ru>
	+ Фонд «Общественное мнение» <http://www.fom.ru>
	+ ВЦИОМ <http://www.wciom.ru/>
	+ Левада-центр <http://www.levada.ru/>
	+ Единый архив экономических и социологических данных <http://sophist.hse.ru/>
	+ Выборочные обследования Росстата <http://www.gks.ru/free_doc/new_site/inspection/itog_inspect1.htm>
	+ Данные РМЭЗ <https://www.hse.ru/rlms/spss>
	+ Обследования бюджетов домашних хозяйств <http://obdx.gks.ru/>
	+ Европейское социальное исследование [http://www.ess-ru.ru](http://www.ess-ru.ru/)

**ВОПРОСЫ К ЗАЧЁТУ ПО МЕЖФАКУЛЬТЕТСКОМУ КУРСУ**

**"Анализ данных в SPSS"**

1. Основные направления и задачи проведения выборочных обследований населения.
2. Методы проведения выборочных обследований населения.
3. Исследование связи между переменными в ходе выборочных обследований.
4. Определение объема выборки при проведении выборочных исследований.
5. Распределение полученных в ходе выборочного обследования результатов на генеральную совокупность.
6. Многомерные группировки по результатам выборочных обследований. Методы кластерного анализа.
7. Какие шкалы измерений Вы знаете? Приведите примеры переменных, измеренных в различных шкалах.
8. Построение таблиц сопряженности.
9. Сформулируйте критерий согласия χ2 проверки гипотезы независимости признаков.
10. Показатели взаимосвязи.
11. Каково основное правило построения иерархических моделей кластерного анализа? Приведите примеры иерархических моделей.
12. В чем суть непараметрических критериев? В каких случаях они применяются.
13. Сущность и задачи корреляционного анализа. Специфика построения и интерпретация моделей по данным выборочных исследований.
14. В каких случаях производится дополнительный отбор и модификация данных? С помощью каких инструментов?
15. Что дает применение методов дескриптивной статистики для анализа данных выборочных исследований?
16. Какие показатели и для данных в каких шкалах можно рассчитать при анализе таблиц сопряженности?
17. Какие шкалы измерения вы знаете и в чем их отличие?
18. Что такое медиана и мода и как их определить для набора данных?
19. Сохранение и импорт данных.
20. Что такое среднеквадратическое отношение и коэффициент вариации? Для чего их используют в анализе данных?
21. Какими методами можно выявить факторы, влияющие на исследуемое явление?
22. Что понимается под генеральной совокупностью и выборкой, а также под репрезентативностью выборки?
23. Как определяется объем выборки? Какие факторы следует учитывать
24. Табличное и графическое представление распределения данных.
25. Дисперсионный анализ.
26. Методы регрессионного анализа в SPSS.
27. Формирование массива данных .
28. Методы снижения размерностей.
29. Построение дерева целей.
30. Одномерные распределения.

**Inter-Faculty Course « Data analysis in SPSS »**

**Summary**

In professional activities and everyday life, we are faced with a large flow of diverse data. In the digital economy, knowledge of the basics of analyzing data sets using the capabilities of special programs allows you to choose the right solution, justify it and evaluate its effectiveness.

The course "data Analysis in SPSS" is a presentation of methods and practices of statistical methods of collection, processing and analysis of quantitative data in the course of sample studies (population surveys). The aim of the course is to explore the specificity of the application of the methods of data collection and analysis in the process of conducting sample surveys of the population with the use of special SPSS program that includes a broad range of processing methods of quantitative and qualitative indicators.

The course will explore the main directions, methods and features of sample surveys; approaches to the formation of data sets and their processing using methods of information visualization, descriptive statistics and multivariate analysis. Students will learn how to apply this knowledge to the input and processing of research data, analysis of various problems of the population and socio-demographic groups, factors and determinants of social and economic policy.

The study of methods of data search and analysis will be based on a large number of practical examples from various fields of activity, the use of universal information technologies and the possibility of analysis tools in SPSS.

Students will be able to implement the author's research project in the chosen field, learn how to form arrays of data and carry out their analysis and processing in the SPSS program, interpret the results.

The course is intended for students who are interested in the possibilities of using data sets, various indicators obtained on the basis of their processing with the help of SPSS program to identify existing problems, solve problems and evaluate performance.