Саратовский Государственный Технический Университет

имени Гагарина Ю. А.

Институт Прикладных Информационых технологий и комуникаций

Кафедра Прикладных Информационых Технологий

**ПРАКТИЧЕСКАЯ РАБОТА № 6**

«Анализ зависимости между переменными, измеренными в номинальных и ординальных данных

Таблицы сопряжённости»

Студент группа ИФСТ-21

Яновский Евгений Валерьевич

Преподаватель Печенкин Виталий Владимирович

**Задание для практической работы (6)**

**Основы интеллектуального анализа данных**

**бИФСТИнПИТ**

**ТЕМА: Анализ зависимости между переменными, измеренными в номинальных и ординальных данных**

**Таблицы сопряжённости**

**Начало работы**

**ВНИМАТЕЛЬНО прочитайте информацию и сформируйте файл отчета по этому занятию. Структура и содержание отчёта будут описана далее. Если в процессе выполнения задания возникают затруднения, задавайте вопросы преподавателю**

На этом занятии мы проанализируем реальные данные одного из исследований, которое проводилось в Российской Федерации. Набор данных содержит результаты опроса граждан РФ по одному из острых проблемных вопросов. В ходе занятия нужно будет установить наличие или отсутствие связи между несколькими переменными. Эти переменные измерены в номинальных или ординальных шкалах. На лекции мы обсуждали постановку этой задачи и критерий наличия связи между переменными (Хи квадрат Пирсона). Нашей целью является поиск зависимостей между переменными и обоснование их значимости.

**Выполните все представленные ниже этапы и подготовьте финальный отчёт по занятию.**

**Совет 1.** При работе с PSPP нужно периодически сохранять текущее состояние работы в файл. По этой причине заранее выберите папку, в которую будете сохранять файл по мере выполнения задания. Путь к этому файлу должен состоять только из **латинских букв** и **не содержать пробелы**. Одним из вариантов может быть сохранение работы на личный флеш-накопитель с последующим копированием в вашу индивидуальную сетевую папку (имя флеш-накопителя тоже не должно содержать русских букв и пробелов).

**Совет 2.** Просмотрите состав файла отчёта, описанный в конце этого задания, создайте заранее файл в программе MSWord и копируйтескриншоты всех этапов вашей работы.

Для выполнения работы воспользуемся установленной на компьютере программой PSPP.

**I. ПЕРВЫЙ ЭТАП.**

**Открытие набора данных в программе PSPP. Подготовка к анализу**

Загрузите на свой компьютер файл **Bureaucracy\_Crosstabs.sav**. Этот файл находится в ИОС, в папке, в которой вы берёте файлы заданий. Файл нужно разместить в какой-либо папке или в корне диска так, чтобы выполнялись условия, изложенные в Совете 1.

Откройте скачанный файл программой PSPP.

Для того, чтобы в закладке с данными были видны текстовые значения, нужно отметить элемент меню View→ValueLabel.

**Введите в значения переменной ID** (первыйстолбецназакладкеDataView) свой индивидуальный номер в журнале так, чтобы он был виден в начальном видимом отрезке набора данных – примерно 20 строк. **Сделайте скриншот** этой закладки и поместите его в отчёт с именем Report\_6\_FIO с соответствующим расширением. При написании фамилии используйте транслитерацию латинскими буквами. Например, если ваша фамилия «Иванов», имя файла будет Report\_6 \_Ivanov.

**ВАЖНО.**Сделайте скриншот результатов работы и поместите его в файл отчёта. Обратите внимание на то, что на скриншоте **должны быть видны значения вашего индивидуального номера в первом столбце**.

**Покажите сделанные изменения в наборе данных преподавателю.**

**II. ВТОРОЙ ЭТАП. Построение таблиц сопряженности в программе PSPP**

Для того, чтобы проверить наличие связи между переменными, мы должны выбрать те из них, для которых она может существовать, исходя из гипотезы исследования.

Для анализа предлагается использовать следующие переменные

Для столбцов (Columns)выберем переменные

ПОЛ (V58)

ВОЗРАСТНОЕ РАСПРЕДЕЛЕНИЕ (V59k)

ТИП ПОСЕЛЕНИЯ (V65)

В строках (Rows) будем размещать следующие переменные

B02. КАК ВЫ СЧИТАЕТЕ, В ЧЬИХ РУКАХ НАХОДИТСЯ РЕАЛЬНАЯ ВЛАСТЬ В СТРАНЕ? (V02)

B05. НАСКОЛЬКО ВЫ УДОВЛЕТВОРЕНЫ ТЕМ, КАК В РОССИИ РАБОТАЕТ ДЕМОКРАТИЯ: (V05)

B07. КАКИЕ ЭПОХИ ИСТОРИИ РОССИИ ХАРАКТЕРИЗУЮТСЯ НАИБОЛЬШИМ ЗАСИЛЬЕМ БЮРОКРАТИИ? (V07)

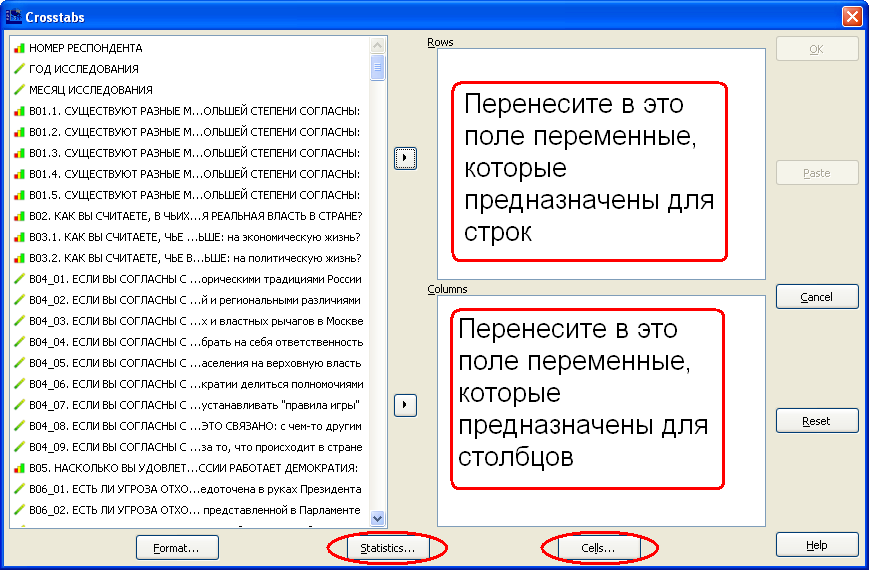
B13. КАК, НА ВАШ ВЗГЛЯД, НУЖНО ОТНОСИТЬСЯ К ЗАКОНУ? (V13)

B16. КАК В ЦЕЛОМ ВЫ ОТНОСИТЕСЬ К ЛЮДЯМ, РАБОТАЮЩИМ В ГОСАППАРАТЕ? (V16)

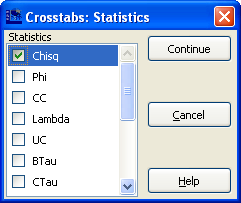
Выполните команду меню

Analyze→DescriptiveStatistics→Crosstabs

Перенесите переменные в соответствующие поля с помощью соответствующих кнопок так как это показано на рисунке

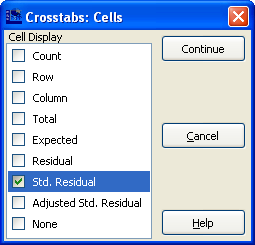


Щелкните по кнопке **Statistics…**, отметьте только чекбокс, соответствующий критерию Хи-квадрат Пирсона



На данном этапе анализа мы проанализируем стандартизированные остатки в таблицах сопряжённости и значимость критерия Хи-квадрат Пирсона, чтобы найти переменные, которые демонстрируют зависимость. Нажмите **Continue**для принятия сделанных изменений.

Щелкните по кнопке **Cells…**и оставьте активированным только вывод стандартизированныхостатков, как это показано на рисунке ниже. Нажмите **Continue**для принятия сделанных изменений.



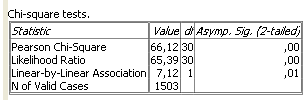
Щёлкните на кнопке OK в диалоге определения переменных для таблиц сопряжённости.

Подготовка к анализу закончена, перейдём непосредственно к анализу наличия зависимости.

**III. ТРЕТИЙ ЭТАП.**

**Анализ наличия зависимости.**

Просмотрите результаты вычислений, которые программа PSPPпоместила в окне Output-PSPP. Найдите пары переменных, для которых значение критерия Хи-квадрат Пирсона имеет высокую значимость (меньше 0,05). Обсудите с преподавателем правило выбора таких пар и необходимый уровень значимости соответствующего критерия. Таблица со значением критерия будет выглядеть так, как показано на рисунке



Просмотрите все таблицы для каждой пары переменных.

Перемещаем в отчёт скриншотами (при необходимости обрезайте их в программе Paint или непосредственно в программе MSWord)все таблицы сопряженности с выведеными стандартизированными остатками и соответствующие им таблицы со значением Хи-квадрат.

**Покажите текущее состояние своего отчёта преподавателю, объясните по каким причинам вы выбрали именно эти таблицы, а не другие.**

**IV. ЧЕТВЁРТЫЙ ЭТАП.**

**Подготовка данных для окончательного отчёта**

Повторите процедуру построения кросстаблиц, но на этот раз активируйте вывод в ячейки стандартизированных остатков и процентов по столбцу. Переместите все таблицы сопряженности с процентами в отчёт так, чтобы для каждой пары переменных, **между которыми была выявлена зависимость**, мы имели кросстаблицу с процентами по столбцу, стандартизированнымиостатками и значение Хи-квадрат Пирсона.

**ФИНАЛЬНЫЙ ЭТАП. Формирование окончательного отчёта.**

Сформируйте, используя MSWord, и покажите преподавателю Ваш отчёт по сегодняшнему занятию.

Имя файла отчёта Report\_6\_ФамилияСтудента. Сохраните отчёт в свою сетевую папку (рекомендуется папка МОМД на вашем сетевом диске)

**ВАЖНО.** Сохраняйте сформированные файлы данных и файл окончательного отчета. В конце семестра их нужно будет представить в семестровом отчёте по курсу для получения зачёта.

**СОСТАВ ОТЧЕТА**

1. Стандартный титульный лист практической работы с фамилией студента, преподавателя и подзаголовком «Подготовка набора данных и анализ корреляционной зависимости переменных»

2. Содержание файла задания

3. Скриншоты: набора данных в PSPP (скриншот закладки DataViewcвидимым вашим индивидуальным номером), для каждой пары зависимых переменных скриншоты кросстаблиц и значения Хи-квадрат, как это описано в четвёртом этапе задания.

Обратите внимание на то, чтобы разрешение и масштаб изображения позволили увидеть ваш индивидуальный номер на скриншоте, на котором он должен быть представлен.

**Покажите отчет преподавателю**

**ЗАМЕЧАНИЕ.** Преподаватель фиксирует в электронной таблице посещаемость занятий и выполнение задания на каждом занятии – предоставление отчёта студентом. Во избежание недоразумений с фиксацией отчётности проверяйте актуальность Ваших текущих оценок.

Просто подойдите к преподавателю и попросите показать отметки о выполнении вами заданий.

**ОТВЕТЬТЕ НА ВОПРОСЫ**

1. Как вычисляется значение Хи-квадрат?
2. Каким должно быть значение Хи-квадрат, чтобы говорить о наличии зависимости переменных?
3. Для чего нам нужно знать значимость критерия Хи-квадрат
4. Что является индикатором сильного смещения наблюдаемой частоты по отношению к ожидаемой в ячейках таблицы?

