

**МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ)
МАИ**

Рассмотрено и одобрено
Редакционным Советом
Факультета № 8
« » _____ 2005г.
протокол №

Рассмотрено и одобрено
Редакционным Советом
Факультета № 10
« » _____ 2005г.
протокол №

**Методическое пособие
Практикум по SPSS**

Для студентов дневной и вечерней форм обучения
по специальности 061100 «Менеджмент организации» в курсе дисциплины
«Информационные технологии»
(Базовый уровень высшего профессионального образования)

Автор: Малинина Н.Л.

Москва, 2010 г.

Лабораторная работа № 1. Введение в SPSS. Частотный анализ данных

Во время проведения лабораторной работы все действия, особенно результаты, оформляются в виде отчета в текстовом процессоре *Word*. Для этого на сетевом диске STUD в директории SPSS каждая группа заводит папку группы, а каждая пара студентов – папку со своим личным именем. В этой папке будут храниться все результаты их работы. Таблицы и графики, полученные в результате работы программы *SPSS* вводятся в отчет через пункт меню: **Правка – Специальная вставка**.

1. Запуск пакета

Для запуска системы *SPSS* надо дважды щелкнуть мышью по фирменному значку *SPSS*, находящемуся в программной группе или непосредственно в Диспетчере программ. После инициализации в общем случае будут открыты:

1. Окно приложения *SPSS* с главным меню в верхней части и панелью инструментов под строкой главного меню.

2. Окно **Вывода**.

3. Окно **Редактора Данных** с пустой электронной таблицей, готовой для ввода данных.

2. Опросный лист:

Sex – пол респондента: 1 – мужской, 2 – женский. **Age** – возраст респондента. **Educat** – образование: полное число лет обучения. **Job** – занятость: 6 – полная занятость, 4 – частичная занятость, 2 – пенсионер, 3 – безработный, 1 – учащийся. **Pets** – наличие домашних животных: 1 – есть, 2 – нет. **Hobby** – хобби: 1 – коллекционер, 2 – меломан, 3 – автолюбитель, 4 – спортсмен, 5 – книголюб, 6 – юзер, 7 – экстрим, 8 – хобби нет. **Children** – количество детей. **Salary** – месячный доход в условных единицах. **Nation** – национальность: 1 – русский, 2 – татарин, 3 – украинец, 4 – белорус, 5 – болгарин, 6 – чех. **Car** – наличие машины: 1 – есть, 2 – нет.

3. Определение переменных:

В редакторе данных дважды щелкнуть мышью по ярлычку с надписью *Variable View* (определение переменных) в левом нижнем углу Редактора Данных. Вы перейдете в режим определения переменных.

В поле *Name* введите название переменной и подтвердите ввод. Имена переменных могут содержать только буквы латинского алфавита и цифры. Имя переменной начинается с буквы. Последний символ не может быть точкой или знаком подчеркивания. Длина переменной не может быть больше 8-ми символов.

Тип переменной задается при переходе на закладку *Type*. Переменные по умолчанию являются численными с максимальной длиной восемь знаков, из них два знака после запятой. Для задания типа переменной надо щелкнуть в поле *Type* на кнопке с тремя точками. Откроется окно *Define Variable Type*. Примите предлагаемую настройку *Numeric* и установите длину «8». Подтвердите настройку кнопкой *Continue*.

Всего в *SPSS* существует восемь типов переменных, однако мы пока будем использовать только числовые и строковые (*String*), которые представляют собой строку символов. К допустимым значениям относятся буквы, цифры и специальные символы. Со строковыми переменными нельзя проводить никаких вычислительных операций, но можно проводить подсчет повторяемости.

В поле *Columns* мы определяем ширину столбца. Подтвердите настройку по умолчанию «8» клавишей **OK**.

На закладке *Label* вводится метка переменной – название, позволяющее описать переменную более подробно. Метка переменной может содержать до 256 символов. Различаются

прописные и строчные буквы. Метки переменных желательно вводить на латинице, поскольку кириллица в полном объеме программой не воспринимается.

На закладке *Missing Values* по умолчанию не должно быть отсутствующих величин. Введите единичное отсутствующее значение «0».

Расположение переменной внутри колонки можно выровнять, пользуясь командой *Alignment*.

Поле *Decimals* предназначено для определения количества знаков после запятой. У вас это значение должно быть равно нулю.

В колонке *Measure* можно задать шкалу переменной. По умолчанию автоматически предлагается метрическая шкала. Подтвердите настройку по умолчанию *Nominal* или выберите соответствующую настройку.

Переменная	Тип	Ширина (Width)	Метка	Шкала
Sex	Numeric	8	Пол респондента	Номинальная
Age	Numeric	8	Возраст респондента	Отношений
Educat	Numeric	8	образование	Отношений
Job	Numeric	8	Занятость респондента	Порядковая
Salary	Numeric	8	Доход респондента	Отношений
Nation	Numeric	8	Национальность	Номинальная
Car	Numeric	8	Наличие машины	Номинальная
Hobby	Numeric	8	Хобби	Номинальная
Children	Numeric	8	Наличие детей	Отношений
Pets	Numeric	8	Наличие домашних животных	Номинальная

4. Ввод данных:

Приступим к вводу данных. Данные можно вводить отдельно по строкам или по столбцам. Действуйте следующим образом: Щелкните на ячейке. Вокруг ячейки появится рамка. Таким образом, ячейка обозначается как активная. Введите в ячейку значение. Это значение отобразится в редакторе ячеек в верхней части окна редактора данных. Щелкните мышью или нажмите клавишу <Таб>. Значение из редактора отобразится в ячейке.

После ввода данных сделайте визуальную проверку правильности введенных данных.

№	sex	age	educat	job	salary	nation	car	hobby	children	pets
1	1	34	15	6	500	1	1	3	0	2
2	1	18	11	1	14	1	2	5	0	2
3	1	28	15	6	680	1	1	2	1	1
4	2	46	16	6	320	1	1	6	0	1
5	1	22	14	4	200	1	2	2	0	2
6	2	31	16	6	170	1	2	4	1	2
7	1	26	15	6	500	1	1	8	0	2
8	1	39	17	3	12	1	2	4	0	2
9	1	28	9	6	300	1	2	2	1	1
10	2	18	11	1	10	1	2	6	0	1
11	1	37	17	6	210	1	1	5	1	2
12	2	36	15	3	0	1	2	4	0	2
13	1	23	15	1	20	1	2	4	0	2
14	2	43	10	6	260	1	1	2	2	1
15	2	68	9	6	230	1	1	8	3	2
16	1	29	16	6	145	1	2	2	1	1
17	1	40	19	6	1000	1	1	2	1	1
18	2	34	15	6	350	1	1	2	0	1
19	1	53	11	6	510	1	2	1	1	2
20	1	18	10	1	10	1	2	7	0	1

21	1	19	11	1	10	1	2	7	0	1
22	1	27	15	6	100	1	1	1	2	1
23	2	41	18	3	20	1	1	5	1	2
24	2	65	9	2	35	1	2	3	3	1
25	1	63	18	2	35	1	2	2	1	2
26	1	40	12	6	250	1	1	5	3	1
27	1	53	9	6	100	1	2	1	2	2
28	2	69	5	2	30	1	2	8	1	2
29	2	25	15	4	300	1	1	2	1	1
30	2	40	12	4	200	1	1	1	2	1
31	1	34	10	6	50	1	2	4	3	2
32	2	43	15	3	10	1	1	3	1	1
33	1	19	12	1	10	1	2	7	0	2
34	2	73	7	2	30	1	2	5	2	1
35	1	17	10	1	0	1	2	6	0	2
36	1	26	15	4	250	1	1	3	1	1
37	2	28	15	3	15	1	1	2	0	1
38	2	37	10	6	540	1	1	7	2	1
39	2	53	15	4	310	1	2	8	2	1
40	2	32	18	6	300	1	2	1	2	2
41	1	22	13	2	10	1	2	2	2	1
42	2	47	15	3	15	1	2	5	2	1
43	1	34	17	6	550	1	1	1	2	2
44	2	18	11	1	10	1	2	4	0	1
45	1	43	16	6	600	1	1	3	1	2
46	1	17	10	1	10	1	2	7	0	2
47	2	24	15	6	500	1	2	4	0	1
48	1	32	15	6	700	1	1	2	1	2
49	1	39	17	6	430	1	1	2	2	2
50	1	35	16	6	1000	1	1	3	2	1
51	2	28	15	6	760	1	2	8	2	1
52	2	30	16	6	450	1	1	8	2	1
53	1	32	18	3	20	1	2	5	2	2
54	2	45	14	6	790	1	1	1	1	2
55	1	56	18	2	765	1	1	1	1	1
56	1	45	17	6	120	1	2	4	1	1
57	1	57	9	2	50	2	1	1	3	1
58	2	36	15	3	20	2	1	5	1	1
59	2	26	14	6	500	2	1	8	0	2
60	1	33	15	6	450	2	2	6	1	2
61	1	26	15	6	100	2	1	4	1	2
62	2	20	10	3	20	2	2	2	0	2
63	1	36	15	4	650	2	1	5	1	1
64	1	34	12	3	15	2	1	3	2	1
65	2	23	15	4	100	2	1	3	1	1
66	1	23	15	6	450	2	1	4	1	2
67	2	19	10	1	20	2	2	7	0	1
68	1	28	15	6	650	2	1	6	1	1
69	2	21	14	1	10	2	2	4	0	2
70	2	65	15	2	30	2	2	5	3	2
71	1	63	9	2	50	2	2	5	0	2
72	1	40	14	6	400	2	2	7	1	1
73	2	53	17	6	300	2	1	1	2	2
74	2	69	16	6	700	2	1	3	1	1
75	1	25	15	4	300	3	1	6	1	2
76	1	40	10	4	65	3	2	3	2	2
77	2	34	11	4	45	3	2	8	1	1
78	1	43	10	4	50	3	2	8	0	2
79	1	19	12	1	10	3	2	3	0	1

80	2	73	15	6	70	3	1	2	2	2
81	2	17	10	4	35	3	2	3	0	1
82	1	26	15	6	50	3	2	5	2	2
83	1	28	15	3	20	3	2	8	0	2
84	2	37	15	3	15	3	2	5	2	1
85	1	53	11	1	10	3	2	6	0	1
86	2	32	16	4	90	3	1	2	1	2
87	2	22	15	4	65	3	1	8	2	1
88	1	47	15	6	50	3	1	4	1	2
89	2	34	15	6	200	3	1	4	1	2
90	1	18	10	1	10	3	1	1	1	1
91	2	43	16	6	75	3	1	3	1	2
92	1	17	10	6	50	3	1	8	1	2
93	2	67	9	2	30	4	1	7	4	1
94	2	45	15	6	70	4	1	4	2	1
95	1	61	10	2	80	4	2	1	1	1
96	11	28	15	4	65	4	2	6	2	1
97	2	31	13	3	20	4	1	8	1	2
98	1	19	12	1	10	4	2	3	0	2
99	1	33	15	4	70	4	1	1	3	2
100	2	70	8	2	30	4	2	7	2	2
101	2	34	12	6	80	4	1	8	1	2
102	1	54	8	4	60	4	2	3	3	1
103	1	21	14	1	10	4	2	4	0	2
104	2	17	10	1	0	4	2	8	0	1
105	1	28	15	6	90	4	2	8	1	1
106	2	39	15	4	50	5	2	5	2	1
107	2	23	15	4	120	5	1	8	0	1
108	2	54	12	2	50	5	2	1	5	1
109	1	32	10	6	100	5	2	3	3	1
110	1	45	8	4	60	5	2	8	3	1
111	1	26	15	4	200	5	2	8	1	2
112	2	19	12	1	13	6	2	2	0	1
113	1	19	12	1	10	6	1	7	0	2
114	2	42	10	6	100	6	2	5	3	2
115	2	19	11	1	10	6	1	6	0	2
116	1	46	15	6	1000	6	1	8	0	2
117	2	20	11	1	15	6	2	2	0	2
118	1	60	16	2	130	6	2	8	3	2
119	2	38	16	6	520	6	1	7	1	1
120	1	31	16	4	80	6	2	3	2	1

5. Сохранение файла данных:

Выберите в меню команды **File** (Файл) ⇒ **Save as** (Сохранить как...). Откроется диалоговое окно **Save Data as** (Сохранить данные как...). По умолчанию **SPSS** сохраняет файл данных в текущем каталоге с расширением **.sav**. Если Вы следовали указаниям и создали рабочий каталог **\SPSSBOOK**, он будет предлагаться по умолчанию. Задайте имя файла, соответствующее соглашению об именах в **DOS**. Расширение **.sav SPSS** присваивает файлам данных по умолчанию.

Данные следует сохранять в своей личной папке в на сетевом диске.

6. Выбор статистической процедуры:

Меню статистических процедур открывается по команде меню **Analyze** (Анализ), в котором содержится список статистических методов. Стрелки указывают на существование следующего подуровня меню. В качестве примера попробуем построить частотное распределение. Выберите в меню команды **Analyze** (Анализ) ⇒ **Descriptive statistics** (Описательная

статистика) ⇒ *Frequency...* (Частотное распределение...). Появится диалоговое окно *Frequency* (см. рис. 3).

Диалоговые окна всех статистических процедур содержат следующие компоненты:

Список исходных переменных – в данный момент в списке исходных переменных присутствуют следующие переменные: пол, возраст, занятость, образование, наличие детей и т.д. Перед именем каждой переменной стоит значок, который позволяет определить, является ли эта переменная численной или строковой.

Список выбранных переменных – должен содержать все переменные, отобранные для анализа. Его также называют целевым списком или списком тестируемых переменных. Он имеет заголовок *Variable(s)* (Переменная).

Командные кнопки – кнопки, при щелчке на которые выполняются определенные действия:

Кнопка **OK** – запускает соответствующую процедуру. Одновременно она закрывает диалоговое окно.

Paste (Вставить) – переносит выбранный в диалоговом окне синтаксис команды в редактор синтаксиса, где можно отредактировать синтаксис команды и дополнить его другими опциями, недоступными в данном диалоговом окне.

Reset (Сброс) – отменяет перенос выделенной переменной в целевой список переменных.

Cancel – отменяет все изменения и закрывает диалоговое окно.

Help – выводит контекстно-чувствительную справку.

Существуют вспомогательные диалоговые окна: **Statistics** (Статистика), **Charts** (Диаграммы и графики) и **Format** (Формат). Кнопки вспомогательных диалоговых окон отличаются троеточием после названия.

7. Частотное распределение:

Построим частотное распределение. Выполняем следующие действия. Выберите в меню команды **Analyze** (Анализ) ⇒ **Descriptive statistics** (Описательная статистика) ⇒ **Frequency...** (Частотное распределение...). Выделяем переменную «**educat**» в списке исходных переменных. Щелкнем на стрелке. Переменная будет перенесена из списка исходных переменных в список выбранных. Подтверждаем операцию щелчком на **OK**. Результаты отображаются в окне просмотра **Viewer**. Окно просмотра разделено на две части. В левой части отображается структура ввода, в правой части – собственно выводимые данные. В разделе вывода отображаются как таблицы, так и графики. Самостоятельно вернуться в редактор данных и выбрать для частотного распределения переменные **age, educat, salary, job, children, nation**, содержащиеся в файле данных.

В группе: **Статистика** просчитайте для всех переменных квартили, процентилю, точки раздела (подгруппа **Percentile values**), стандартное отклонение, дисперсию, размах, минимум, максимум, стандартную ошибку (подгруппа **Dispersion**), среднее значение (только для переменных **age, educat, salary**, почему?), медиану, моду (подгруппа **Central tendency**), коэффициент асимметрии и коэффициент вариации только для переменных **age, educat, salary** (подгруппа **Distribution**).

Проделайте частотный анализ всех переменных, чтобы проверить правильность ввода данных.

Вопрос: Что надо контролировать, чтобы сделать такую проверку?

8. Завершение сеанса работы:

Выберите в меню команды **File** (Файл) \Rightarrow **Exit** (Выход). Для каждого из открытых окон **SPSS** спрашивает, надо ли сохранять его содержимое. Если щелкнуть на кнопке **OK** (Да), или нажать **<Enter>**, **SPSS** открывает специальное диалоговое окно, в котором надо указать тип сохраняемого файла.

9. Вопросы для контроля:

Вопрос: входит ли номер респондента в число переменных?

Вопрос: как мы проверяем правильность ввода данных?

Вопрос: приведите свой пример данных номинальной шкалы

Вопрос: приведите свой пример данных порядковой шкалы

Вопрос: приведите свой пример данных интервальной шкалы

Вопрос: приведите свой пример данных шкалы отношений

Вопрос: почему мы считаем коэффициент асимметрии и коэффициент вариации только для переменных **age**, **educat**, **salary**?

Вопрос: для каких переменных считают моду, а для каких – среднее значение?

Продолжая в том же духе, займемся анализом данных из Всероссийского центра изучения общественного мнения. В Независимом институте социальной политики есть архив, который содержит около 500 исследований, проведенных ведущими социологическими агентствами России.

Архив предоставляет свои коллекции для научной работы некоммерческим организациям, творческим коллективам и отдельным исследователям на безвозмездной основе. Итак, мы начинаем наше изучение программы **SPSS**, пользуясь чужими опросами.

Из папки на сетевом диске надо скопировать файл данных **BL97_10.sav** и файл «Вопросы к полевому материалу **BL97_10.doc**» в свою папку под новыми именами, чтобы каждая группа могла работать с ними, не мешая окружающим.

Проанализируйте данные в файле «вопросы к полевому материалу **BL97_10**» и таблицу полевого материала (файл данных **BL97_10.sav**), и впишите имена своих переменных для отчета из файла **BL97_10.sav** по каждому вопросу в соответствующую таблицу (примеры таблиц следуют ниже), а также внесите в таблицы типы шкал для всех переменных. Например:

Вопрос 4:

Тип шкалы: ???	Имя:	Частота	%
1. только учиться			
2. только работать			
3. совмещать учебу и работу			
4. дожидаться призыва в армию			
5. пока ничего			
6. другое			
7. затрудняюсь ответить			

Таблицы можно создать в своем отчете и заполнить их. Недостающие таблицы построить самим.

1. Частотное (линейное) распределение:

Исследуйте выборку по каждому вопросу отдельно и внесите данные по результатам работы **SPSS** в соответствующие таблицы отчета. Сделайте дополнительные таблицы и сравните частотное распределение, разделив выборку по переменной «пол». Не забудьте проверить данные (какие?) на наличие нормального распределения. Внесите результаты исследования в отчет.

Частотное распределение каждая пара студентов строит для своих переменных:

Варианты для задания 1

Вариант	Переменные для исследования	Переменная для выборки или группировки	Вопрос из переменной для фильтра
1	q1	q16	3
2	q6	q14	1
3	q7	q14	2
4	q4	q14	3
5	q8	q14	4
6	q12	q15	1
7	q13	q15	2
8	q14	q15	3
9	q15	q19	1
10	q16	q19	2
11	q17	q19	3
12	q13	q19	4
13	q19	q15	4
14	q5	q16	1
15	q9	q16	2

1. Подумайте, какого типа переменные достались вам для анализа.
2. Подумайте, можно ли изменить тип этих переменных?

В зависимости от типа переменных сделайте оба варианта, и по возможности посчитайте максимально возможный пакет данных. В группе: *Статистика* можно просчитать для всех переменных квартили, процентиля, точки раздела (подгруппа *Percentile values*), стандартное отклонение, дисперсию, размах, минимум, максимум, стандартную ошибку (подгруппа *Dispersion*), среднее значение, медиану, моду (подгруппа *Central tendency*), а также коэффициент асимметрии и коэффициент вариации.

2. Проверка соответствия нормальному распределению:

Для проверки в окне *Frequencies: Statistics* следует щелкнуть на *Charts*, выбрать *Histograms*, установить флажок *With normal curve*, подтвердить кнопкой *Continue*, вернуться в диалоговое окно *Frequencies*, снять флажок *Display frequency tables*. Щелкните на *OK*. В окне просмотра увидим гистограмму с кривой нормального распределения.

Диаграммы и гистограммы строить для своих вариантов, однако для того, чтобы вы имели право делать проверку нормальному распределению, следует привести данные как минимум к порядковой шкале. Как это сделать рассказывается ниже. Графики перенести в отчет.

Вы сделаете проверку соответствия нормальному распределению для тех переменных, которые вам достались, однако желательно знать ответ на *Вопрос*: для переменных какой шкалы можно делать это исследование?

Найдите в файле данных те переменные, для которых можно сделать проверку наличию нормального распределения и проделайте эту проверку.

3. Выбор случаев:

Выберите в меню команду *Data, Select Cases*. В диалоговом окне *Select Cases* выберите пункт меню *If condition is satisfied* и щелкните на кнопке *If*. Отберите данные таким образом, чтобы построить частотные таблицы только для мужчин (или женщин), ответивших на вопрос № (см. таблицу вариантов) соответствующим образом (см. таблицу вариантов) с применением булевой алгебры. Переменная переносится в редактор условий. С клавиатуры копируется знак равенства или другой. Так создается условие для выборки. В результате для ка-

кой выборки вы получите частотные таблицы? Описать в отчете подробно. Постройте необходимые графики. Сохраните файл.

С помощью этого пункта меню можно также привести свою переменную к порядковой шкале.

4. Разделение случаев на группы:

Для разделения выборки на группы следует выбрать *Data, Split file*. В открывшемся диалоговом окне выберите *Organize output by groups* и перенесите переменную **pets** в поле *Groups based on*. Щелкните на кнопке **ОК**. Частотный анализ провести для тех же переменных, что и в предыдущем пункте. Результаты частотного анализа будут выведены в разных таблицах. Выберем пункт *Compare groups*. Снова выполним частотный анализ переменных. Получим другую результирующую таблицу. Скопируйте результаты в отчет и прокомментируйте результаты в отчете.

Чтобы вернуть данные в исходное состояние в *Data, Split file* выберите опцию *Analyze all cases, do not create groups*, снимите сортировку.

Проанализируйте полученные данные и ответьте на вопросы:

Вопрос: для каких переменных правомочно считать коэффициенты асимметрии и вариации и почему?

Вопрос: для каких переменных считают моду, для каких – медиану, а для каких – среднее значение?

Вопрос: какие переменные из обработанных вами соответствуют нормальному распределению, а какие – нет?

Лабораторная работа № 2. Модификация данных

Для проведения анализа часто бывает необходимо выполнить преобразование данных. Новые переменные можно создавать на основе первоначально собранных данных, так же можно изменять кодирование переменных. Все это называется модификацией данных.

Существует много возможностей для модификации:

- Вычисление новых переменных путем использования различных арифметических выражений (математических формул)
- Подсчет частоты появления определенных значений
- Перекодирование значений
- Вычисление новых переменных при выполнении определенного условия
- Агрегирование данных
- Ранговые преобразования
- Вычисление суммы весов случаев

1. Подсчет частоты появлений определенных значений

В SPSS есть возможность подсчитать количество появлений одного и того же значения или значений для определенной переменной. Допустим, мы хотим подсчитать, сколько респондентов ответили в переменной qM на вопрос №X при условии, что по переменной qN они ответили по вопросу №Y.

Как это сделать? Выберите в меню команды *Transform ⇒ Count*. Откроется диалоговое окно *Count Occurrences of Values within Cases*. Оно разделено на части:

Target Variable - в этом поле указывается имя переменной, в которой будут содержаться подсчитанные значения.

Target Label - указывается метка выходной переменной.

Variables – список содержит переменные, выбранные из списка исходных данных. Список хранится в файле данных, для которых нужно подсчитать определенные значения, он не может одновременно содержать численные и строковые переменные.

Внеся в список переменные, присвойте выходной переменной имя и метку. От щелчка на кнопке *Define Values...* Откроется диалоговое окно *Count Values within Cases: Values to*

Count. Это диалоговое окно служит для определения подсчитываемых значений. Можно задать отдельное значение, диапазон или сочетание того или другого. В группе **Value** можно выбрать следующие варианты:

Value – вводится отдельное значение, частоту которого необходимо подсчитать.

System Missing (системное отсутствующее) – можно подсчитать количество появлений системного отсутствующего значения. Для строковых переменных этот вариант не применим.

System-or user- missing – подсчитывается количество появлений всех отсутствующих значений, как пользовательских, так и системных.

Range Through (диапазон) – подсчитывается количество появлений всех значений в определенном диапазоне. Также не применим для строковых переменных.

Range: Lowest Through - подсчитывается количество значений в диапазоне от наименьшего наблюдаемого до указанного (не применим к строковым переменным).

Range: Through Highest - подсчитывается количество значений в диапазоне от указанного до наибольшего наблюдаемого (не применим к строковым переменным).

Если потребуется подсчитать повторяемость нескольких значений, щелкните после выбора соответствующей опции на кнопке **Add**. В этом случае будет подсчитана частота повторений каждого значения, присутствующего в списке **Values to Count**. Ввод отдельного значения подтверждается кнопкой **Continue**, а затем – **OK**.

Варианты для задания 2

Вариант	Переменная (qM) и подсказка для подсчета (№X)	При условии, что переменная (qN) и подсказка (№Y)
1	q5, подсказка 6	q9, подсказка 1, q9, подсказка 2
2	q5, подсказка 4	q9, подсказка 1, q9, подсказка 2
3	q 5, подсказка 3	q9, подсказка 1, q9, подсказка 2
4	q 5, подсказка 2	q9, подсказка 1, q9, подсказка 2
5	q5, подсказка 5	q9, подсказка 1, q9, подсказка 2
6	q6, подсказка 1	q9, подсказка 1, q9, подсказка 2
7	q6, подсказка 2	q9, подсказка 1, q9, подсказка 2
8	q6, подсказка 3	q9, подсказка 1, q9, подсказка 2
9	q6, подсказка 4	q9, подсказка 1, q9, подсказка 2
10	q 6, подсказка 5	q9, подсказка 1, q9, подсказка 2
11	q5, подсказка 6	q9, подсказка 1, q9, подсказка 3
12	q5, подсказка 4	q9, подсказка 1, q9, подсказка 3
13	q5, подсказка 3	q9, подсказка 1, q9, подсказка 3
14	q5, подсказка 2	q9, подсказка 1, q9, подсказка 3
15	q5, подсказка 5	q9, подсказка 1, q9, подсказка 3

2. Перекодирование значений

Первоначально собранные данные можно перекодировать с помощью средств SPSS. Это бывает необходимо, когда первоначальное разнообразие исходных данных не нужно для последующего анализа. Перекодирование в таком случае означает уменьшение объема обрабатываемой информации. Перекодирование можно выполнить автоматически или вручную.

Ручное перекодирование

Выберите в меню команды **Transform** ⇒ **Recode...** Выберите в подменю пункт: **Into Different Variables**. Откроется диалоговое окно **Recode into Different Variables**. Перенесите вашу переменную в поле **Input Variable** ⇒ **Output Variable**. Выходной переменной надо задать имя. В поле **Name** надо ввести имя. Щелкните на кнопке **Change**. В поле **Label** вводится обозначение переменной (метка). Чтобы установить значения, которые следует перекодировать, следует щелкнуть на кнопке **Old and New Variables**. Откроется диалоговое окно **Recode into Different Variables: Old and New Variables**.

Для осуществления каждого перекодирования надо указать значение или диапазон входной переменной и соответствующее значение выходной переменной. Перекодирование завершается щелчком на кнопке *Add*.

Диалоговое окно разделено на части:

В группе *Old Value* можно выбрать один из следующих вариантов:

Value – вводится отдельное значение.

System missing – с помощью этой опции значение входной переменной обозначается, как системное отсутствующее. Этот вариант не применим для строковых переменных. В списке значений переменных они указываются как SYSMIS.

System-or user-missing – опция служит для обозначения всех пользовательских или системных отсутствующих значений. В списке значений переменных они указываются как MISSING.

Range through – здесь можно задать замкнутый интервал значений. Не применим для строковых переменных.

Range: Lowest Through - в этом случае будут перекодированы все значения в диапазоне от наименьшего наблюдаемого до указанного (не применим к строковым переменным).

Range: through highest – в этом случае будут перекодированы все значения от указанного до наибольшего наблюдаемого. Этот вариант не применим для строковых переменных.

All other values – эта опция касается всех еще не указанных значений. В списке значений переменных они указываются как ELSE.

В группе *New Value* можно выбрать один из следующих вариантов:

Value – вводится новое значение.

System missing – служит для обозначения значения выходной переменной как системного отсутствующего значения. Значение появляется в списке значений в виде SYSMIS. Этот вариант не применим для строковых переменных.

Copy old values – значения входной переменной сохранятся без изменений.

Если новые выходные переменные являются строковыми, следует установить флажок *Output variables are string*.

Старое значение переменной вводится в поле *Value* в группе *Old Value*, новое значение – в поле *Value* в группе *New Value*, не забывайте щелкать на кнопке *Add*.

Сделайте перекодирование переменной КУРЕНИЕ. Эта переменная очень характерна тем, что у нее небольшие значения переменной соответствуют большому объему курения. Запишите результаты в отчет. Сравните результаты гистограмм для старой переменной и новой переменной. Гистограммы должны быть зеркальными друг относительно друга.

Автоматическое перекодирование

Если категории не были закодированы непрерывно, начиная с 1, то это может привести к негативным последствиям при решении многих задач в SPSS. Для преобразования значений численных или строковых переменных в непрерывную последовательность целых чисел в SPSS реализована возможность автоматического перекодирования. Чаще всего это применяется в тех случаях, когда надо перекодировать строковые переменные в численные.

3. Вычисление новых переменных в соответствии с определенными условиями

Допустим, нам надо провести частотный анализ для некоторых переменных, но желательно делать это для разных возрастных групп. Образование новых переменных не затронет основной файл данных.

В файле ex_09.sav и подсказок к нему в файле «Вопросы и подсказки для отчета ВЦИОМ 2003-9.doc» содержится много данных, но нас будет волновать ответы на вопросы для разных возрастных групп.

Формулировка условий: IF («возраст» < 30) «старая переменная» = «новая переменная»

Выберите команды меню: *Transform* ⇒ *Compute...* В открывшемся диалоговом окне в поле выходной переменной задайте ее имя, а для численного выражения напишите формулу. Щелкните на кнопке *If...* Откроется диалоговое окно *Compute Variable: If Cases*. Измените

начальную настройку *Include All cases* на *Include If Case satisfies condition*. В поле под этой опцией введите условие. Закройте диалоговое окно. Теперь в файле появилась новая переменная, которая в случаях, когда заданное условие не выполняется, содержит системное отсутствующее значение. Предлагаются следующие варианты.

Варианты для задания 3

Вариант	Переменная и подсказка	Возраст – это условие для выбора
1	q1 и №1,2,3	До 30
2	q1 и №1,2,3	От 30 до 40
3	q1 и №1,2,3	От 40
4	q9a и № 1,2	До 30
5	q9a и № 1,2	От 30 до 40
6	q9a и № 1,2	От 40
7	q50a и №1,2,3,4	До 30
8	q50b и №1,2,3,4	От 30 до 40
9	q52 и №1,2,3,4	От 40
10	q54b и №1,2,3	До 30
11	q54c и №1,2,3	От 30 до 40
12	q65B и №1,2,3	От 40
13	q68 и № 1,2	До 30
14	q69 и № 1,2	От 30 до 40
15	qd5 и № 1,2,3,4	От 40

4. Вычисление новой переменной

Проанализируйте переменную q50b из полевого материала ex_09: Как бы вы оценили в целом отношения между Россией и США? У нее 7 градаций, включая и отказ от ответа. В некоторых случаях было бы хорошо иметь меньше градаций ответов. Допустим, надо сократить 6 ответов до трех. Это можно сделать с помощью процедуры *Recode*, а также с помощью процедуры *Compute*.

1. дружественные
2. хорошие, соседские
3. нормальные, спокойные
4. прохладные
5. напряженные
6. враждебные
7. затрудняюсь ответить

Объединение вопросов можно сделать через редактор условий в пункте меню процедуры *Compute*. Для этого следует выбрать *Transform* ⇒ *Compute*. Вы попадете в Редактор процедуры *Compute*. В окне *Target Variable* следует задать имя для новой переменной, например, это будет q1050b. В окне *Numeric Expression* следует задать первую кодировочную цифру, например это будет 1. Нажмите курсором кнопку *If*. Вы попадете в редактор условий. Следует написать условие. Как это сделать? Допустим, вы хотите объединить первый и второй ответы по переменной q50b. Для этого следует объединить множества респондентов, которые по этой переменной ответили либо по первому вопросу, либо по второму вопросу в одно множество респондентов. Вам следует в окне написать фразу: q50b=1 or q50b=2 и нажать на кнопки **Continue** и **OK**. После этого в файле данных в колонке с именем q1050b в тех строках, которые соответствуют единицам и двойкам переменной q50b, будут стоять единицы.

Доделать до конца самостоятельно. Получите новую переменную. У нее будут единицы в тех местах, где у старой переменной стояли единицы и двойки. У новой переменной будут стоять двойки в тех строках, где у старой переменной были тройки и четверки. И так далее. Отказ от ответа или оставить как есть, или вообще забыть про него. В результате должно получиться две гистограммы.

5. Вычисление новой переменной

В полевом материале ex_09 найти переменную ОБЩИЙ ДОХОД семьи и переменную КОЛИЧЕСТВО ЧЕЛОВЕК в семье. Задание: создать новую переменную СРЕДНИЙ ДОХОД НА ЧЕЛОВЕКА через процедуру *Compute*. В результате следует получить статистику.

Лабораторная работа № 3. Таблицы сопряженности

Таблицы сопряженности применяются, когда нас интересует двумерный анализ, а также когда необходимо выяснить, существует ли взаимосвязь между двумя переменными. Таблицы сопряженности будем строить, используя полевой материал из файла ex_09.sav и подсказки к нему в файле «Вопросы и подсказки для отчета ВЦИОМ 2003-9.doc». Эти файлы следует скопировать в свою папку с изменением имени. Не забудьте изменить тип переменных.

Вопрос: что надо сделать, чтобы изменить тип шкалы у переменной?

Варианты для задания 4

№ варианта	Номера вопросов	Условие выборки
1	q5a и q6	qd2 (№ 1, 2, 3)
2	q7 и qd7	qd2 (№ 4, 5, 6)
3	q9a и q9c	qd2 (№ 8, 9, 10)
4	q9b и q9c	qd2 (№ 1, 2, 3)
5	q9d и q 9c	qd2 (№ 1, 2, 3)
6	q9e и q 9c	qd2 (№ 4, 5, 6)
7	q29a и q9e	qd2 (№ 8, 9, 10)
8	q41a и q41b	qd2 (№ 1, 2, 3)
9	q 50a и q 52	qd2 (№ 8, 9, 10)
10	65b и 65c	qd2 (№ 4, 5, 6)
11	q13a и q14	qd2 (№ 8, 9, 10)
12	q7 и qd5	qd2 (№ 1, 2, 3)
13	q5b и q5c	qd2 (№ 8, 9, 10)
14	q5b и qd7	qd2 (№ 1, 2, 3)
15	q5b и qd5	qd2 (№ 4, 5, 6)

1. Создание таблиц сопряженности

Через команды меню *Analyze* ⇒ *Descriptive Statistics* ⇒ *Crosstabs* откроем диалоговое окно *Crosstabs*. Создайте таблицы сопряженности для переменных пар (смотри свой вариант) для всей выборки, аналогично для частичных выборок по полу.

Кроме разделения выборки по полу, следует еще посмотреть разделение выборки по переменным, указанным для каждого варианта в отдельности.

В итоге у вас получится несколько таблиц сопряженности. Внимательно рассмотрите их и скопируйте в отчет. Не забудьте отредактировать их в редакторе *Word*. Первая таблица содержит информацию о числе самих случаев. Вторая таблица – это собственно таблица сопряженности. Сопрягаемые переменные могут быть как столбцовыми, так и строковыми. Значение в каждой строке таблицы сопряженности – частота случаев. Если для таблицы сопряженности приняты параметры по умолчанию, в каждой ячейке отображается только абсолютная частота. Числа в последней строке и в последнем столбце показывают суммы значений соответственно по строкам и столбцам.

Для того чтобы провести исследование более детально, нам потребуется ответить на следующие вопросы:

- Существует ли зависимость между переменными вообще?
- Что можно сказать об интенсивности этой зависимости?
- Что можно сказать о направлении и характере этой зависимости?

Сделать такой анализ нам позволит вычисление значений ожидаемых частот.

Для этого выберите команды меню: *Analyze* ⇒ *Descriptive Statistics* ⇒ *Crosstabs...* В списке строк и в списке столбцов должны стоять переменные из вашего варианта. Не забыть

перекодировать их. Далее щелкните на кнопке **Cells...** Откроется диалоговое окно: **Crosstabs: Cell Display**. По умолчанию в ячейках таблицы сопряженности отображаются только наблюдаемые значения частот. В группе **Counts** можно выбрать один или более следующих вариантов отображения:

Observed – настройка по умолчанию, отображаются только наблюдаемые значения частот.

Expected – если установить этот флажок, то будут отображаться ожидаемые частоты. Они вычисляются как произведение сумм соответствующей строки и столбца, деленное на общую сумму частот.

Установив флажок **Expected** мы получим новую таблицу сопряженности. Теперь под наблюдаемыми частотами появились ожидаемые значения. Как можно интерпретировать результаты? Подумайте и напишите об этом в отчете.

Еще одну возможность существования зависимости между переменными дает вычисление остатков. Остатки являются показателем того, насколько сильно наблюдаемые и ожидаемые частоты отклоняются друг от друга. Чтобы получить остатки частот, выполните следующие действия.

Через команды меню **Analyze ⇒ Descriptive Statistics ⇒ Crosstabs ⇒ Cells**, флажки **Observed** и **Expected** оставьте помеченными, в группе **Residuals** выберите **Unstandardized**, щелкните на кнопке **Continue**, затем **OK**. Каждый ненормированный остаток равен разности наблюдаемой и теоретически ожидаемой частот в данной ячейке. Остатки делают еще более заметной наблюдаемые тенденции взаимосвязи между переменными.

Нормированные остатки полезны при проведении анализа тестов по критерию χ^2 . Полученные данные позволят скорректировать результаты анализа.

Недостаток таблиц сопряженности в том виде, в котором мы их рассматривали – это абсолютные значения. Чтобы узнать, насколько эти значения важны по отношению к общему количеству, надо определить их процентную долю. Для определения процентных долей следует через команды меню **Analyze ⇒ Descriptive Statistics ⇒ Crosstabs ⇒ Cells** в группе **Percentages** выбрать варианты **Row, Column, Total**. Выберите все. Щелкните на кнопке **Continue**, затем **OK**. В полученной таблице сопряженности дополнительно отобразятся процентные отношения частот по отношению к суммам строк, столбцов и общей сумме. Снова проанализируйте результаты расчетов и внесите в отчет.

Проанализируйте созданные таблицы сопряженности и ответьте на вопросы:

1. Существует ли зависимость между переменными?
2. Что можно сказать об интенсивности зависимости?
3. Что можно сказать о направлении и характере этой зависимости?

2. Форматы таблиц сопряженности

Сортировка. Можно изменить порядок сортировки строк в таблице сопряженности. Для этого следует открыть диалоговое окно **Crosstabs: Table Format** и в группе **Row (Column) Order** выбрать один из двух вариантов сортировки значений: **Ascending** или **Descending**.

Применение переменных групп и слоев. Допустим, вы хотите разделить созданные таблицы сопряженности в зависимости от пола респондента. В таком случае переменная ПОЛ будет играть роль переменной слоев. Анализ производится по группам – в нашем случае для каждого пола – составляется отдельная таблица сопряженности.

Чтобы задать переменную слоев, следует выбрать команды меню: **Analyze ⇒ Descriptive Statistics ⇒ Crosstabs**. В списках строк и столбцов оставьте ваши переменные, а переменную ПОЛ перенесите в список переменных слоев. Чтобы добавить новый слой, следует щелкнуть на кнопке **Next**. Каждый последующий уровень делит таблицу сопряженности на меньшие подгруппы. Переходить от слоя к слою можно при помощи кнопок **Next** и **Previous**. Самостоятельно проанализируйте результаты работы и сохраните в отчете.

3. Графическое представление таблиц сопряженности.

Данные, содержащиеся в таблицах сопряженности можно сделать более визуальными.

Для этого выберите команды меню: **Graphs** ⇒ **Bar**. Откроется диалоговое окно **Bar Charts**. Выберите пункт **Clustered**, оставьте предлагаемую по умолчанию опцию **Summaries for Groups of cases** и щелкните на кнопке **Define**. Откроется диалоговое окно **Define Clustered Bar, Summaries for Groups of cases**. Выберите пункт **% of cases**. Перенесите выбранную переменную в поле **Category Axis**, а вторую переменную – в поле **Define Clusters by**. Щелкните на кнопке **Titles**. Откроется диалоговое окно **Titles**. В нем можно ввести заголовок диаграммы, в поле **Subtitle** – подзаголовок. Кроме того, можно сделать нижний колонтитул. Кнопка **Options** откроет диалоговое окно **Options**. Снимите в нем флажок **Display groups defined by missing values**, щелкните на кнопке **Continue**, а затем на **OK**. В окне просмотра появится график.

Дважды щелкните на этом графике – откроется редактор диаграмм, в котором его можно править. Выберите в меню команды: **Format** ⇒ **Bar Label Style...** Откроется диалоговое окно **Bar Label Style**. Выберите пункт **Framed** и щелкните на кнопке **Apply all**, а затем на кнопке **Close**. Щелкните на одном из столбцов, он окажется выделен. Тогда выберите в меню команды: **Format** ⇒ **Color**. Откроется диалоговое окно **Colors**. Здесь можно изменить стандартный цвет столбцов, а также цвет их контура. Щелкните на сером поле, а затем на кнопках **Apply** и **Close**. В заключение вызовите команды меню: **Chart** ⇒ **Outer Frame**. Это даст возможность сделать красивое обрамление таблицы сопряженности. Сохраните результаты в отчете.

4. Статистические критерии для таблиц сопряженности

Продолжаем работать с уже полученными данными из варианта заданий №4.

Чтобы получить статистические критерии для таблиц сопряженности, щелкните на кнопке **Statistics** в диалоговом окне **Crosstabs**. Откроется диалоговое окно **Crosstabs: Statistics**. Флажки позволяют выбрать один или несколько критериев:

- Тест хи-квадрат (χ^2)
- Корреляции
- Меры связанности для переменных, относящихся к номинальной шкале
- Меры связанности для переменных, относящихся к порядковой шкале
- Меры связанности для переменных, относящихся к интервальной шкале
- Коэффициент каппа (κ)

Проведение теста χ^2 для проверки взаимной независимости

При проведении теста хи-квадрат проверяется взаимная независимость двух переменных таблицы сопряженности и благодаря этому косвенно выясняется зависимость обеих переменных. Две переменные считаются взаимно независимыми. Если наблюдаемые частоты в ячейках таблицы сопряженности совпадают с ожидаемыми частотами.

Для этого выберите в меню команды: **Analyze** ⇒ **Descriptive Statistics** ⇒ **Crosstabs**. Кнопкой **Reset** удалите возможные настройки. Перенесите переменные в списки строк и столбцов. Щелкните на **Cells**. Установите кроме предложенного по умолчанию флажка **Observed**, флажки **Expected** и **Standardized**, щелкните на **Continue**. Откроется диалоговое окно **Statistics**. Установите флажок **Chi-square**. Щелкните на кнопке **Continue**, затем **OK**. Вы получите таблицу сопряженности, кроме того, в окне просмотра вы найдете таблицу с результатами теста хи-квадрат, примерно такую:

	Значение		Асимптотическая значимость (двусторонняя)
Хи-квадрат по Пирсону			
Отношение правдоподобия			
Зависимость линейный-линейный			
Количество допустимых случаев			

Результаты интерпретируйте с учетом следующих таблиц:

Значение коэффициента корреляции r	Интерпретация
0 < r <= 0,2	Очень слабая корреляция
0,2 < r <= 0,5	Слабая корреляция
0,5 < r <= 0,7	Средняя корреляция
0,7 < r <= 0,9	Сильная корреляция
0,9 < r <= 1	Очень сильная корреляция

Вероятность ошибки	Значимость
p>0,05	Не значимая
P<=0,05	Значимая
P<=0,01	Очень значимая
P<=0,001	Максимально значимая

Для вычисления критерия Хи-квадрат по Пирсону применяется формула:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Корректно прокомментируйте целесообразность и правомерность применения теста Хи-квадрат для вашего варианта данных. Опишите полученные результаты в отчете. Не забывайте, какие данные вы обрабатываете.

Проведение теста χ^2 с поправкой на правдоподобие.

Альтернативой формуле Пирсона для вычисления критерия Хи-квадрат является поправка на правдоподобие: $\chi^2 = -2 * \sum f_o * l_n * \frac{f_e}{f_o}$

При большом объеме выборки формула Пирсона и подправленная формула дают очень близкие результаты.

Тест Мантеля-Хэнзеля (*Linear-by-linear*)

Дополнительно выводится значение теста Мантеля-Хэнзеля. Эта форма критерия Хи-квадрат с поправкой Мантеля-Хэнзеля – еще одна мера линейной зависимости между строками и столбцами таблицы сопряженности. Она определяется как произведение коэффициента корреляции Пирсона на количество случаев, уменьшенное на единицу. Метод Мантеля-Хэнзеля используется всегда. Если установлен флажок *Chi-square*.

Свое мнение о зависимости или независимости переменных отразите в отчете.

ПРЕДУПРЕЖДЕНИЕ: у нас нет данных, принадлежащих к интервальной шкале или шкале отношений, поэтому следует запомнить, что вычисление коэффициента корреляции Пирсона – не правомерно, однако кнопки следует изучить. Хотя, если вы можете доказать, что переменная принадлежит к номинальной шкале и имеет две категории, то применение коэффициента Пирсона – правомерно.

Итак, если переменные принадлежат к порядковой шкале, то правильнее применять коэффициент Спирмена. Для этого: *Analyze* ⇒ *Descriptive Statistics* ⇒ *Crosstabs*. Кнопкой *Reset* удалите возможные настройки. Перенесите переменные в списки строк и столбцов. Щелкните на *Statistics*. В диалоге *Crosstabs: Statistics* установите флажок *Correlations*. Подтвердите выбор кнопкой *Continue*. Откажитесь от вывода таблиц, установив флажок *Suppress tables*. Щелкните на *OK*.

Будут вычислены коэффициенты Спирмена и Пирсона, а также проведена проверка их значимости. Скопируйте полученную таблицу в отчет и прокомментируйте полученные результаты. Опишите взаимосвязь между переменными, если она есть.

		Значение	Асимптотическая стандартная ошибка	Приближительное T	Приближительная значимость
Интервальный-интервальный	Пирсон				
Порядковый-порядковый	Корреляция по Спирмену				
Количество допустимых случаев					

Меры связанности для переменных, относящихся к номинальной шкале

Если переменные принадлежат к номинальной шкале и имеют больше двух категорий, то нельзя применять коэффициент корреляции в качестве зависимости между переменными. Почему?? А потому, что между кодировками невозможно установить порядкового отношения, и переменные (их кодировки) не могут быть расположены в определенном, рационально объяснимом порядке.

Самое лучшее для этого – провести анализ наблюдаемых и ожидаемых частот, а также нормированных остатков. Все это мы уже делали, но это не дает нам количественной оценки связанности двух переменных, поставленных во взаимное соответствие. На самом деле такие меры найдены. И, если у вас есть две переменные номинальной шкалы, то можно попробовать. Полная независимость переменных = 0, максимальная зависимость = 1.

Для анализа возьмем данные из файла ex_09.sav и подсказки к нему в файле «Вопросы и подсказки для отчета ВЦИОМ 2003-9.doc».

Загрузите файл данных, создайте таблицу сопряженности, задайте вывод ожидаемых частот, стандартизированных остатков, процентов по столбцам и критерия Хи-квадрат. Задайте (кнопкой **Statistics**) вывод всех мер связанности для переменных, принадлежащих к номинальной шкале (флажки в группе **Nominal**).

Варианты для задания 5

№ варианта	Номера вопросов
1	q1 и qs1
2	q1 и q9a
3	q1 и q9b
4	q1 и q9c
5	q1 и 8a
6	q1 и q9d
7	q1 и q9e
8	q1 и q9g
9	q1 и q9h
10	q9a и q9b
11	q9a и q9c
12	q9a и q9e
13	q9b и q9c
14	q9c и q9e
15	q9a и q9g

В окне просмотра получите таблицу Симметричных мер:

		Значение	Приблизительное значение
Номинальный - номинальный	Критерий Фишера		
	Критерий Крамера		
	Коэффициент сопряженности признаков		
Число случаев			

а) Нулевая гипотеза не принимается или принимается

b) используется асимптотическая стандартная ошибка с принятием нулевой гипотезы

И таблицу Направленных мер:

			Значение	Асимптотич. стандарт ошибка	Прибл. Т	Прибл. Знач.
Номинальный- номинальный	Лямбда	Симметрическая				
		В зависимости от столб- цовой величины				
		В зависимости от строко- вой величины				
	Тау Гудмена- Крускала	В зависимости от столб- цовой величины				
		В зависимости от строко- вой величины				
	Коэффиц. Неопре- де- ленности	Симметрическая				
		В зависимости от столб- цовой величины				
		В зависимости от строко- вой величины				

a) Нулевая гипотеза не принимается или принимается

b) используется асимптотическая стандартная ошибка с принятием нулевой гипотезы

c) на основе аппроксимации по распределению Хи-квадрат

d) Степень правдоподобия по закону Хи-квадрат

Коэффициент сопряженности признаков (Пирсона)

Его величина находится в пределах от 0 до 1 (аналогично для критериев Фишера и

Крамера) и вычисляется с использованием значения критерия Хи-квадрат: $c = \sqrt{\frac{\chi^2}{\chi^2 + N}}$

N – общая сумма частот в таблице сопряженности.

Коэффициент сопряженности признаков для двух таблиц с разным количеством полей сопоставим.

Критерий Фишера

Он используется только для таблиц 2*2, поскольку в других случаях его значение может превысить единицу: $\varphi = \sqrt{\frac{\chi^2}{N}}$

жет превысить единицу: $\varphi = \sqrt{\frac{\chi^2}{N}}$

Критерий Крамера

Он представляет собой модификацию критерия Фишера и для любых таблиц сопря-

женности дает значение в пределах от 0 до 1: $\varphi = \sqrt{\frac{\chi^2}{N(k-1)}}$

k – наименьшее из количеств строк и столбцов.

Лямбда

Допустим, в исследовании вы приняли переменную X зависимой от переменной Y. Теперь, вам надо сделать предположение для отдельного респондента о том, что для него истинно X. Это справедливо, если в частотном анализе для X получено значение >50%, а если нет? Тогда наше предположение неверно. Но вероятность предположения можно повысить, если учитывать другую переменную – Y. При разбивке выборки по переменной Y, можно будет для отдельных категорий по Y достаточно достоверно говорить об истинности для них X. Сокращается первоначальная вероятность ошибки. Относительное сокращение ошибки называется «лямбда». При этом, если ошибка при втором прогнозе сокращается до 0, то «лямбда» = 1. Если ошибка при первом и при втором прогнозе одинакова, то «лямбда» = 0.

Вторая переменная не помогает в уточнении предсказания первой (зависимой) переменной, то есть выбранные переменные не зависят друг от друга. SPSS вычисляет оба значения, по очереди рассматривая обе переменные, как зависимые между собой переменные. Выводится среднее из двух значений.

Задание на дом: проработать этот вопрос более подробно.

Тау (τ) Гудмена-Крускала

Это вариант меры связанности «лямбда», который вычисляется вместе с ней. При определении этой меры связанности наблюдаемые частоты взвешиваются с учетом своих процентов и складываются. Выводится относительное сокращение ошибки. Задание на дом: проработать этот вопрос более подробно.

Коэффициент неопределенности

Представляет собой еще один вариант критерия «лямбда», при определении которого имеется в виду не ошибочное предсказание, а «неопределенность», то есть степень неточности предсказания. Коэффициент неопределенности принимает значение в диапазоне от 0 до 1. Значение 1 говорит о том, что одну переменную можно точно предсказать по значениям другой, и они взаимосвязаны.

Меры связанности для переменных, относящихся к порядковой шкале

Все эти меры основаны на количестве нарушений порядка (так называемых инверсий - I). Количество инверсий можно определить, если расположить в порядке возрастания значения одной из двух переменных, между которыми необходимо установить степень взаимосвязи, а рядом с ними записать соответствующие значения другой переменной. Число нарушений порядка расположения второй переменной и есть количество инверсий. Это количество вместе с количеством соблюдения порядка (проверсий, обозначаемых через P) используется в различных формулах для определения меры связанности, которые дают значения этого параметра в диапазоне от -1 до +1.

Гамма (γ)

Гамма вычисляется по формуле:
$$\gamma = \frac{P - I}{P + I}$$

Если инверсий не наблюдается ($I=0$), то мы имеем $\gamma=1$ (полную зависимость). Если же не встречается проверсий, а только инверсии ($P=0$), то говорят о максимально разнонаправленной зависимости ($\gamma=-1$). Если $P=I$, то зависимости вообще не существует ($\gamma=0$).

d Сомера

Существуют две асимметричных и симметричная меры связанности d Сомера. Для их вычисления используется формула для γ с корректирующим членом T_γ , который учитывает количество связей зависимых переменных (одинаковых значений, встречающихся в измерениях):

$$d = \frac{P - I}{P + I + T_\gamma}$$

Тау-б (τ_b) Кендалла

Этот коэффициент одновременно учитывает связи как зависимых, так и независимых переменных, и может приобретать значения -1 и +1 только для квадратных таблиц сопряженности.

Тау-б (τ_b) Кендалла

Этот критерий может достигать значений -1 и +1 для любых таблиц.

Варианты для анализа вышперечисленных мер связанности.

Сделайте анализ полученных результатов и занесите в отчет.

Ответьте на вопросы:

Являются ли переменные взаимно независимыми?

Как можно прокомментировать результаты анализа с точки зрения других предметов вашего курса?

Варианты для задания 6

№ варианта	Номера вопросов
1	q15 и №16а
2	q23а и №23b
3	q47а и №47b
4	q47а и №47с
5	q47b и №47с
6	q50а и №52
7	q54а и №54b
8	q54b и №54с
9	qд5 и №д7
10	q50а и №50b
11	q41d и №41e
12	q40b и №41а
13	q41а и №40с
14	q40а и № 41с
15	q41а и №41с

Лабораторная работа № 4. Анализ множественных ответов

Во многих анкетах имеются вопросы, на которые можно дать одновременно несколько ответов. Для кодировки и анализа таких множественных ответов SPSS представляет два различных метода.

1. Дихотомный метод

В методе множественной дихотомии для каждой из возможностей ответа определяется отдельная переменная. Мы имеем столько переменных, сколько есть подсказок на вопрос с кодовыми значениями 0 (если ответ – «нет») и 1 (если ответ – «да»). Сами кодовые значения выбираются произвольно, однако для всех ответов они должны быть одинаковыми, и введены в компьютер на правильном месте.

Ответы мы возьмем в файле ex_09.sav, подсказки к нему в файле «Вопросы и подсказки для отчета ВЦИОМ 2003-9.doc». Это вопрос № 18: Назовите, пожалуйста, 5-6 политиков, которым вы более всего доверяете?

Сначала мы построим частотную таблицу ответов на этот вопрос, а затем перекрестную таблицу этого вопроса и пола.

Варианты для задания 7

№ варианта	Номера вопросов	Вторая переменная для создания таблицы сопряженности
1	q27	q18
2	q11а	q11b
3	q11b	q11а
4	q34b	q35b
5	q18	q19
6	q19	q18
7	q27	q28
8	q28	q27
9	q32b	q33b
10	q33а	q32а
11	q34а	q35а
12	q35b	q34b
13	q35а	q34а
14	q18	q27
15	q11b	q11а

1. Определение набора

Мы должны сообщить компьютеру, что эти переменные принадлежат одному набору. Выберите в меню команды: *Analyze* ⇒ *Multiple Response* ⇒ *Define Sets...* Откроется диалоговое окно: *Define Multiple Response Sets*.

Выделите в списке исходных переменных анализируемые переменные и перенесите их в список *Variables in Set*.

Задайте дихотомную кодировку переменных (опция *Dichotomies* в группе *Variables Are Coded As*). Эта настройка по умолчанию. В поле *Counted Value* введите «1».

Присвойте набору имя и метку.

Щелкните на *Add*, и созданный набор будет внесен в список наборов множественных ответов: *Mult Response Sets*.

Щелкните на *Close*.

2. Частотные таблицы для дихотомных наборов

Выберите в меню команды: *Analyze* ⇒ *Multiple Response* ⇒ *Frequencies*. Откроется диалоговое окно *Multiple Response Frequencies*.

В списке *Mult Response Sets* этого диалога отображаются уже определенные наборы переменных.

Перенесите свой набор в список *Tables for* . Щелкните на кнопке *OK*.

В окне просмотра появятся результаты:

Первый столбец – приводятся метки переменных, принадлежащих к набору. Случай считается отсутствующим, если ни одна из переменных набора не имеет учитываемого значения (1).

Второй столбец – имена переменных.

Третий столбец – количество учитываемых случаев.

Четвертый столбец – процентное значение относительно общего числа «да» или «1».

Пятый столбец - процентное значение относительно общего числа допустимых случаев (количество респондентов вычесть количество случаев без ответов).

3. Таблицы сопряженности с дихотомными наборами

Таблицы сопряженности можно создавать между двумя наборами, а также между набором и «обычной» переменной. Попробуем сделать это для нашей переменной и переменной «пол».

Выберите команды меню: *Analyze* ⇒ *Multiple Response* ⇒ *Crosstabs...* Откроется диалоговое окно *Multiple Response Crosstabs*. В списке исходных переменных показаны переменные вашего файла данных. В списке наборов множественных ответов показан ранее определенный набор

Перенесите в список переменных строк ваш набор, а в список переменных столбцов – переменную «пол». Если таблица сопряженности строится между элементарными переменными (не являющимися наборами) и наборами, то для первых следует задать диапазон значений.

Щелкните на кнопке *Define Range...* Откроется диалоговое окно *Multiple Response Crosstabs: Define Variable Range*. Задайте минимальное значение: *Minimum* = «1», а максимальное *Maximum* = «2». Подтвердите выбор кнопкой *Continue*. Теперь вопросительные знаки заменены значениями «1» и «2». Щелкните на кнопке *Options*. Откроется диалоговое окно *Multiple Response Crosstabs: Options*. Абсолютные частоты в ячейках выводятся всегда. Дополнительно в группе *Cell Percentage* можно выбрать одну или несколько характеристик:

Row – отображаются проценты для строки.

Column – отображаются проценты для столбца.

Total - отображаются общие проценты для таблицы.

В группе *Percentage Based On* можно выбрать одну из следующих опций:

Cases – настройка по умолчанию.

Responses – основой расчета процентного отношения в ячейке является количество ответов. Для наборов множественных дихотомий количество ответов равно частоте учитываемого значения во всех случаях.

Флажок **Match variables across response sets** имеет смысл, только если первая таблица сопряженности строится на основе двух наборов переменных. В этом случае первая переменная из первого набора сочетается с первой переменной из второго набора и т.д.

В группе **Percentage based on** сохраните настройку по умолчанию **Cases**.

В группе **Cell Percentages** установите флажок **Column**.

Подтвердите ввод кнопкой **Continue**, а затем – **OK**.

В окне просмотра появится таблица.

Полученные проценты соответствуют отношению частот к числу допустимых случаев, длины меток ограничены 20 символами.

К сожалению, для множественных ответов не проводится проверка значимости по критерию Хи-квадрат. Если такая проверка необходима, следует сделать анализ концентрированных данных.

Сравнение средних

Сравнение средних – наиболее часто применяемый метод статистического анализа. Главный вопрос – можно ли объяснить имеющиеся различия статистическими колебаниями или нет?

При сравнении средних всегда предполагается, что обе выборки подчиняются нормальному распределению. Если это не так, то вычисляются медианы и применяется непараметрический тест. Существуют четыре различные тестовые ситуации:

Ситуация	Тест	
Две независимые выборки	t-тест для независимых выборок (тест Стьюдента)	Independent-Samples T Test
Две зависимые выборки	t-тест для зависимых выборок	Paired-Samples T Test
> двух независимых выборок	Однофакторный дисперсионный анализ	One-way ANOVA
> двух зависимых выборок	Однофакторный дисперсионный анализ с повторными измерениями	

Первые три из этих тестов вызываются с помощью меню: **Analyze** ⇒ **Compare** ⇒ **Means Repeated Measures**.

Сравнение двух независимых выборок

Выберите в подменю команду: **Independent-Samples T Test**. В открывшемся диалоговом окне перенесите переменную из вашего варианта в список тестируемых переменных, а переменную ПОЛ в поле группирующих переменных.

Варианты для задания 8

№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12	№13	№14	№15
q3	q5a	q5b	q5c	q6	q7	q8a	q8b	q16a	q16c	q15	q23a	q23b	q29a	q29b

Для работы загрузите файл данных BL97_10.sav. Не забудьте привести переменные хотя бы к порядковой шкале. Щелчком на кнопке **Define Groups** открывается диалоговое окно, в котором можно ввести значения двух категорий для группирующей переменной. Поскольку мы сравниваем респондентов по полу, то в поле **Group1** введите ПОЛ=1, а в поле **Group 2** введите ПОЛ=2. щелчком на кнопке **Continue** вернитесь в основное диалоговое окно, щелкните на кнопке **Options**, посмотрите настройки и, не меняя их, вернитесь через **Continue** в основное окно. Запустите t-тест, щелкнув на **OK**.

Выведенные результаты содержат: количества случаев, средние значения, стандартные отклонения и стандартные ошибки средних в обеих группах. Кроме того, есть результаты

теста Левеля на равенство дисперсий. Если в тесте Левеля $p < 0,05$, то не принимается гипотеза о равенстве или гомогенности дисперсий. Далее мы получаем результаты t -теста: значение распределения t , количество степеней свободы df , вероятность ошибки p , а также разницу средних значений, ее стандартную ошибку и доверительный интервал. Все результаты поместить в отчет с комментариями.

Вопрос: существует ли значимое различие между группами?

Сравнение двух зависимых выборок можно найти, если выбрать в меню команду: **Paired-Samples T Test**. Задание на дом: на основе данных в файлах *Malinina_2.rar* или *BL97_10.sav* придумать свой вариант, реализовать расчеты и прокомментировать их.

Сравнение более двух независимых выборок. Попробуем выяснить существует ли значимое различие в конкретных доходах между четырьмя разными группами респондентов по принадлежности к определенному классу по уровню дохода. Выберите в подменю команду: **One-way ANOVA...** Перенесите переменную *qd6* из файла данных *Malinina_2.rar* в список зависимых переменных: **Dependent List**, а переменную *qd5* в поле **Factor**. Посмотрите, какие параметры можно задать для этого теста (кнопка **Options**). Задайте вывод описательной статистики (флажок **Descriptive**) и проверку на гомогенность дисперсий (**Homogeneity-of-variance**). Чтобы выполнить апостериорный тест, вернувшись в основное диалоговое окно, щелкните на кнопке **Post Hoc...** Откроется диалоговое окно **One-way ANOVA: Post Hoc Multiple Comparisons** (Однофакторный дисперсионный анализ). Выберите тест Дункана (флажок **Duncan**) При значимом результате дисперсионного анализа этот тест показывает, какие именно группы по отношению к своим доходам значимо отличаются друг от друга. По умолчанию установлен уровень значимости 0,05; можно выбрать другое значение. Запустите тест, щелкнув на **OK**. В окне просмотра появятся следующие результаты:

Описательная статистика

	N	Среднее	Ст. Откл.	Стд. Ош.	95% доверит. интервал		Минимум	Максимум
					Нижняя гр.	Верхняя гр.		
D5-1								
D5-2								
D5-3								
D5-4								
D5-5								

Выведенные результаты содержат: количество случаев, средние значения, стандартные отклонения и стандартные ошибки средних, 95% доверительные интервалы, минимумы и максимумы для всех слоев фактора.

Дисперсионный анализ

	Сумма квадратов	Количество степеней свободы	Средний квадрат		Значимость
Между группами					
В группах					
Total					

Выведенные результаты содержат: типовую схему дисперсионного анализа, включая вероятность ошибки (значимость) для оценки общей значимости.

Апостериорные тесты

Класс дохода	Количество	Подгруппа для альфа =	
D5-1		1	2
D5-2			
D5-3			
D5-4			
D5-5			

Выведенные результаты содержат: результаты многогранного теста Дункана.

Статистика Левена	Количество степеней свободы 1	Количество степеней свободы 2	Значимость

Выведенные результаты содержат: результаты теста Левена на гомогенность дисперсий
Результаты прокомментировать в отчете по лабораторным работам.

Разложение на составляющие тренда. Сумму квадратов между группами можно разложить на линейные или полиномиальные составляющие тренда. В диалоговом окне *ANOVA* щелкните на кнопке **Contrasts**. Появится диалоговое окно **One-way ANOVA: Contrasts**. Установите флажок **Polinomial**, после этого в списке **Degree** (степень) можно будет выбрать порядок многочлена. Прделайте это самостоятельно и результаты прокомментируйте в отчете.

Лабораторная работа № 5. ИССЛЕДОВАНИЕ ДАННЫХ:

Корреляции

Расчеты подобных двухмерных критериев взаимосвязи основываются на формировании парных значений, которые образуются из рассматриваемых зависимых выборок. Для графического представления подобной связи используется диаграмма рассеяния. Статистик говорит о корреляции и указывает силу связи при помощи коэффициента корреляции. Он обычно принимает значение от -1 до +1, причем близость к 0 говорит о слабой связи. Методы:

- Переменные с интервальной или номинальной шкалой – коэффициент корреляции Пирсона.
- По меньшей мере, одна из двух переменных имеет порядковую шкалу либо не является нормально распределенной - ранговая корреляция по Спирману или Тау-гроссовая Кендалла.
- Одна из двух переменных является дихотомической – точечная двухрядная корреляция (в SPSS отсутствует). Может быть применен расчет ранговой корреляции.
- Обе переменные являются дихотомическими – четырехполевая корреляция. Этот вид корреляции рассчитывается на основании определения дистанционных мер и мер подобия.

Коэффициент корреляции Пирсона можно найти, если выбрать команды меню: **Analyze** ⇒ **Correlate** ⇒ **Bivariate**. Появится диалоговое окно **Bivariate Correlations**. Переменные из вашего варианта перенесите по очереди в поле тестируемых переменных. Начните расчет, нажатием кнопки **OK**.

Варианты для задания 9

№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12	№13	№14	№15
q9a	q9a	q9a	q9a	q9b	q9b	q9b	q9b	q9b	q9b	q9c	q9c	q9b	q9b	q9c
q9b	q9b	q9b	q9b	q9c	q9c	q9c	q9c	q9c	q9c	q9d	q9d	q9c	q9c	q9d
q9c	q9c	q9c	q9c	q9d	q9d	q9d	q9h	q9h	q9h	q9e	q9e	q9h	q9h	q9g
q9d	q9e	q9g	q9h	q9e	q9g	q9h	q9d	q9e	q9e	q9h	q9g	q9g	q9d	q9g

Полученные результаты содержат: корреляционный коэффициент Пирсона r , количество использованных пар значений переменных и вероятность ошибки p , соответствующая предположению о ненулевой корреляции. При помощи щелчка на кнопке **Options** можно организовать расчет среднего значения и стандартного отклонения для двух переменных. Можно также делать расчеты только для двух переменных.

Перенесите результаты в отчет и напишите комментарии.

Ранговые корреляции по Спирману и Кендалу

Для переменных, которые принадлежат к порядковой шкале или для переменных, не подчиняющихся нормальному распределению, а также для переменных, принадлежащих к интервальной шкале, вместо коэффициента Пирсона рассчитывается ранговая корреляция по

Спирману. Для этого отдельным значениям переменных присваиваются ранговые места, которые впоследствии обрабатываются с помощью соответствующих формул. Чтобы выявить ранговую корреляцию, уберите в диалоговом окне *Bivariate Correlations* метку для расчета корреляции по Пирсону и активируйте расчет корреляции по Спирману. Прodelайте это для следующих вариантов:

Варианты для задания 10

№1	№2	№3	№4	№5	№6	№7	№8
qd5	q65b	q54a	q54b	q54a	q50a	q47a	q47a
qd7	q65c	q54b	q54c	q54c	q52	q47b	q47c

№9	№10	№11	№12	№13	№14	№15
q47c	q41e	q41a	q41a	q41c	q40a	q40c
q47b	q41d	q41b	q41c	q41b	q40b	q40b

Не забудьте отфильтровать переменную q54a. Результаты внесите в отчет.

Аналогично для этих же переменных прodelайте ранговую корреляцию по Кендалу. Применение коэффициента Кендала является предпочтительным, если в исходных данных встречаются выбросы.

Дисперсионный анализ

С помощью Д.А. исследуют влияние одной или нескольких независимых переменных на одну зависимую переменную (одномерный анализ) или на несколько зависимых переменных. Дисперсионный анализ будем проводить в поле данных файла BL97_10.sav.

Дисперсионный анализ может быть вызван с помощью команд меню: *Analyze* ⇒ *General Linear Model* ⇒ *Univariate*. Откроется диалоговое окно *Univariate* (одномерная). Перенесите переменные ПОЛ и ВОЗРАСТ в поле фиксированных факторов. В поле зависимых переменных перенесите свой вариант:

Варианты для задания 11

№1	№2	№3	№4	№5	№6	№7	№8
q14	q15	q16	q17	q4	q5	q6	q7

№9	№10	№11	№12	№13	№14	№15
q12	q13	q18	q8	q19	q14	q15

Не забудьте поставить фильтры там, где это необходимо.

Пояснение: Фиксированным фактором или факторам с фиксированными эффектами называются такие факторы, которые охватывает все возможные классификационные слои одной независимой переменной, например: ПОЛ или ОБРАЗОВАНИЕ. Однако, если слои (подпопуляции) фактора выбираются случайным образом из бесконечного множества возможных подпопуляций, то говорят о факторах со случайными эффектами. В этом случае уместен расчет так называемых компонентов дисперсии.

Итак, щелкните по кнопке *Model*. Откроется диалоговое окно *Univariate: Model*. Модель дисперсионного анализа – это математическое соотношение, в котором каждая переменная представлена в виде суммы среднего значения и ошибки. В диалоговом окне оставьте все установки по умолчанию и покиньте его нажатием кнопки *Continue*. Щелкните на выключателе *Options...* Откроется диалоговое окно *Univariate: Options*. Перенесите **OVERALL** и обе переменные ПОЛ И ВОЗРАСТ в поле *Display means for*, в этом случае в качестве результатов будут выведены средние значения и стандартная ошибка для совокупной выборки **OVERALL** и для всех слоев по обоим факторам. Активируйте *Descriptive Statistics*, а затем опцию *Homogeneity Tests*. Таким образом проверяется однородность дисперсии. Покиньте диалоговое окно нажатием на *Continue*. При помощи выключателя *Plots...* откройте диалого-

вое окно **Univariate: Profile Plots**. В случае профильных диаграмм речь идет о графическом представлении средних значений слоев выбранных факторов в виде линейчатых диаграмм. Поместите переменную ВОЗРАСТ в поле **Horizontal Axis**, а переменную ПОЛ в поле **Separate Lines**. Можно указывать дополнительную переменную в поле **Separate Plots**. Тогда вы получите отдельные диаграммы. Щелкните на выключателе **Add** и уйдите через **Continue**. В заключение щелкните на выключателе **Post Hos...** Вы попадете в диалоговое окно **Univariate:Post Hos. Multiply Comparisons for Observed Means**. Поместите переменную ВОЗРАСТ в поле **Post Hos Tests for**. Активируйте тест Шефе и покиньте диалоговое окно через **Continue**. Начните расчет нажатие на **OK**. В окне вывода получите результаты:

Тест Левене на равенство дисперсии ошибок

			Значимость

Межсубъектные факторы

		Метка значений	N
ПОЛ	1	мужской	
	2	женский	
ВОЗРАСТ	14		
	15		
	16		
	17		
	18		

Дескриптивные статистики

ПОЛ	ВОЗРАСТ	Среднее значение	Стандартное отклонение	Количество
1	14			
	15			
	16			
	17			
	18			
2	14			
	15			
	16			
	17			
	18			
Сумма	14			

В графе **ЗНАЧИМОСТЬ** показано значение вероятности ошибки. Мы можем оценить однородность дисперсии между группами. Комментарий.

Тест межсубъектных объектов

Источник	Сумма квадратов III типа		Среднее значение квадрата		Значимость
Подправленная модель					
Отрезок					
ПОЛ					
ВОЗРАСТ					
ПОЛ*ВОЗРАСТ					
Ошибка					
Сумма					

Здесь можно оценить влияние ВОЗРАСТА и ПОЛА.

Далее следует вывод дескриптивных статистик для совокупности выборки и для отдельных слоев. Вывод результатов завершает профильная диаграмма.

Результаты проанализируйте дома, и объясните, отличаются ли возрастные группы и почему.

Факторный анализ

Главными целями факторного анализа являются: (1) *сокращение* числа переменных (редукция данных) и (2) *определение структуры* взаимосвязей между переменными, т.е. *классификация переменных*.

Варианты для задания 12

№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12	№13	№14	№15
q9a	q9a	q9a	q9a	q9b	q9b	q9b	q9b	q9b	q9b	q9c	q9c	q9b	q9b	q9c
q9b	q9b	q9b	q9b	q9c	q9c	q9c	q9c	q9c	q9c	q9d	q9d	q9c	q9c	q9d
q9c	q9c	q9c	q9c	q9d	q9d	q9d	q9h	q9h	q9h	q9e	q9e	q9h	q9h	q9g
q9d	q9e	q9g	q9h	q9e	q9g	q9h	q9d	q9e	q9e	q9h	q9g	q9g	q9d	q9g

Метод факторного анализа попробуем проиллюстрировать на примере исследования из файла *ex_09.sav* (подсказки к нему в файле «Вопросы и подсказки для отчета ВЦИОМ 2003-9.doc»). Опрашиваемые высказывали свое отношение в группе вопросов №9. Чтобы вы не запутались эти вопросы сгруппированы по четыре для каждого варианта.

Итак, откройте файл данных. Выберите в меню: *Analyze ⇒ Data Reduction ⇒ Factor...* Откроется диалоговое окно *Factor Analysis*. Переменные из вашего варианта поместите в поле тестируемых переменных и ознакомьтесь с возможностями, предлагаемыми различными кнопками меню. После щелчка на *Descriptive Statistics* оставьте вывод первичных результатов, которые включают в себя первичные относительные дисперсии простых факторов, собственные значения и процентные доли объединенной дисперсии. Довольно часто бывает необходим вывод одномерных статистик и корреляционных коэффициентов. С помощью кнопки *Extraction* вы можете выбрать метод отбора. Оставьте анализ главных компонент. Количество отображенных в этом случае факторов приравнивается к числу собственных значений, превосходящих единицу. Выключатель *Rotation* позволяет выбрать метод вращения. Активизируйте метод варимакса и оставьте активированным вывод повернутой матрицы факторов. Можно организовать вывод факторных нагрузок в графическом виде. Если вы хотите найти значения факторов и сохранить их в виде дополнительных переменных, то задействуйте выключатель *Scores* и отметьте *Save as variables*. По умолчанию установлен регрессионный метод. Выключатель *Options* предназначен для обработки отсутствующих значений. Для проведения расчетов щелкните на *OK*. В окне обзора появятся результаты в виде таблицы «Объясненная суммарная дисперсия»:

Компоненты	Первичные собственные значения			Повернутые суммы квадратов нагрузок		
	Сумма	% дисперсии	Совокупный %	Сумма	% дисперсии	Совокупный %
1						
2						

И так далее. По таблице можно увидеть, что некоторые факторы имеют значения превосходящие единицу. Для анализа отбираются только эти факторы.

Далее приводится повернутая матрица компонент:

	Компонент		
Переменная1	1	2	3
Переменная2			

Здесь начинается самое интересное: вы должны попытаться объяснить отобранные факторы. Для этого в каждой строке повернутой матрицы карандашом следует отметить ту факторную нагрузку, которая имеет наибольшее абсолютное значение. Факторные нагрузки следует понимать как корреляционные коэффициенты между переменными и факторами. В

большинстве случаев включение отдельной переменной в один фактор, осуществляемое на основании коэффициентов корреляции, является однозначным. Далее следует отнести переменные к своим факторам.

Последний и решающий шаг: необходимо обнаружить и описать смысловую связь факторов. Это вы сделаете дома.

Лабораторная работа № 6. Построение и редактирование графиков

1. Построение столбчатой диаграммы

Столбчатые диаграммы применяются:

- Отображение частот переменных, относящихся к номинальной или порядковой шкале.
- Отображение средних значений, сумм или других показателей последовательных переменных (шкала отношений или интервальная шкала)

Представим в графическом виде значения переменных вашего варианта. Выберите в меню: **Analyze** (Анализировать) ⇒ **Descriptive Statistics** (Описательные статистики) ⇒ **Frequencies** (Частоты). При помощи кнопки **Reset** (Сброс) удалите все предыдущие установки. Щелкните дважды на выбранной переменной, чтобы поместить ее в список отобранных переменных. Щелкните на выключателе **Charts** (Диаграммы). Откроется диалоговое окно **Frequencies: Charts** (Частоты: диаграммы). Щелкните на опции **Bar Charts** (Столбчатые диаграммы), в области **Chart Values** (Значения диаграммы) щелкните на опции **Percentages** (Проценты) и затем на **Continue** (Далее). В главном диалоговом окне деактивируйте опцию **Display frequency tables** (Показать частотные таблицы). Щелкните на **OK**. В окне просмотра появится столбчатая диаграмма. Этот график можно отредактировать.

Щелкните дважды на какой-нибудь точке в пределах графика. После этого он будет помещен в редактор диаграмм. Панель меню изменилась. Присутствуют только опции, предназначенные для обработки графиков. Претерпели изменения и панели инструментов. В меню редактора диаграмм выберите **Format** (Формат) ⇒ **Bar Style** (Вид столбца). Откроется диалоговое окно **Bar Style** (Виды столбцов). Щелкните на области **3-D effect** (3-Д эффект). В поле **Depth** (Глубина) введите число «40». Щелкните на **Apply All** (Применить для всех) и затем на выключателе **Close** (Заккрыть). В окне вывода появится столбчатая диаграмма.

Дадим графику название. Выберите в меню **Chart** (Диаграммы) ⇒ **Title** (Заголовок). Откроется диалоговое окно **Titles** (Заголовки). В поле **Title 1** (Заголовок 1) введите текст: «Job», а в поле **Title 2** «Data 2002 year». Заголовки обязательно вводите на латинице. Выберите для заголовка и подзаголовка центральное выравнивание - **Center** (Центр). Подтвердите нажатием **OK**. Выделим график при помощи рамки. Выберите в меню **Chart** (Диаграмма) ⇒ **Outer Frame** (Рамка снаружи). Пометим столбцы точными процентными показателями. Выберите в меню **Format** (Формат) ⇒ **Bar Label Styles** (Метки столбцов). Откроется диалоговое окно **Bar Label Styles** (Метки столбцов). Щелкните на области **Framed** (В рамке), затем на **Apply All** (Применить для всех) и в заключении на **Close** (Заккрыть). Диаграмма отредактирована.

Для сохранения полученного графика следует поступить: При помощи щелчка на значке **]** закройте редактор диаграмм. Отредактированный график останется в окне просмотра. Выберите в меню **File** (Файл) ⇒ **Save as** (Сохранить). Откроется диалоговое окно **Save As**. Согласно предварительным установкам, SPSS обозначает файлы, которые имеют формат средства просмотра, присваивая им расширение **.spo**. Задайте подходящее имя файла и щелкните на **OK**.

Аналогично сделайте для переменных из лабораторной работы №2, когда вы рассматривали выборку при соблюдении некоторых условий. Только в меню выберите круговую диаграмму.

На принтере можно распечатать график, если выбрать в меню **File** ⇒ **Print**. Подтвердите установки диалогового окна **Print** (Печать) при помощи кнопки **OK**.

2. Линейчатые диаграммы

Линейчатую диаграмму вместо столбчатой строят тогда, когда необходимо отобразить большое количество столбцов, или когда столбцы располагаются в определенной последовательности.

Выберите команды меню: **Graphs** ⇒ **Line...** Откроется диалоговое окно **Line Charts**.

Простую линейчатую диаграмму можно построить, если выбрать в области **Simple** опцию **Summaries for groups of cases**. После щелчка по выключателю **Define** откроется соответствующее диалоговое окно. В поле **Category Axis** введите переменную q1 из файла данных BL97_10.sav. В группе **Line Represent** активируйте **Other summary function** и в появившееся поле введите переменную q7. Вместо установленной по умолчанию функции **Mean of value**, пройдя выключатель **Change summary**, отметьте функцию суммы значений (**Sum of values**). С помощью выключателя **Titles** введите подходящий заголовок. Щелчком по **OK** начните построение диаграммы. Сделайте тоже самое при установленной по умолчанию функции **Mean of value**. Прокомментируйте в отчете разницу между графиками.

Сложную линейчатую диаграмму можно построить, если в диалоговом окне **Line Charts** щелкнуть на области **Multiple** и активировать опцию **Summaries of separate variables**. После щелчка по выключателю **Define** откроется соответствующее диалоговое окно. В поле **Category Axis** введите переменную q1 из файла данных BL97_10.sav. В группе **Line Represents** по очереди введите переменные q12 и q15. Не забудьте временно изменить метки этих переменных на латиницу. Прокомментируйте результат в отчете.

Диаграммы с областями являются разновидностью линейчатой диаграммы, в которой области, находящиеся под линиями, закрашиваются. Выберите в меню: **Graphs** ⇒ **Area**. Откроется диалоговое окно **Area Charts**. Вы можете построить простую или штабельную диаграмму с областями. Данные, отображаемые в диаграммах, могут быть заданы как категории одной переменной, как разные переменные или как значения отдельных случаев. В диалоговом окне **Area Charts** щелкните на области **Simple** и оставьте опцию **Summaries for groups of cases**, устанавливаемую по умолчанию. После щелчка по выключателю **Define** откроется соответствующее диалоговое окно. В поле **Category Axis** введите переменную из qs3 и в группе **Area Represents** установите маркер возле **Other summary function**. В появившееся поле введите переменную для вашего варианта:

Варианты для задания 13

№1	№2	№ 3	№ 4	№5	№ 6	№ 7	№8	№ 9	№10	№11	№12	№13	№14	№15
qd5	qd7	q65c	q65b	q52	q50a	q54a	q55	q65a	q50b	q41d	q39	q41a	q41b	q36b

Оставьте функцию **Mean of value**, устанавливаемую по умолчанию. С помощью выключателя **Titles** введите подходящий заголовок. Щелчком по **OK** начните построение диаграммы. Не забудьте исправить метки переменных на латиницу. Когда будете сохранять файл данных, можете исправить то, что вы изменили.

Попробуйте сделать штабельную диаграмму с областями. Следует не забывать, что этот вид диаграмм применяется только тогда, когда штабелируемые области дают не лишенный смысла эффект суммирования. Для этого снова обратимся к файлу данных BL97_10.sav. В диалоговом окне **Area Charts** щелкните на области **Stacked** и отметьте опцию **Summaries of separate variables**. После щелчка по выключателю **Define** откроется соответствующее диалоговое окно. В поле **Category Axis** введите переменную qs3 и в поле **Area Represents** введите переменные из вашего варианта:

Варианты для задания 14

№1	№ 2	№ 3	№ 4	№ 5	№ 6	№ 7	№ 8
q65b и q65c	q54b и q54c	q47a и q47c	q47c и q47b	q47a и q47b	q40a и q40c	q41a и q41b	q41b и q41c

№ 9	№ 10	№ 11	№ 12	№ 13	№ 14	№ 15
q41a и q41c	q40a и q40b	q23a и q23b	q16a и q16b	q5b и q5c	q2 и q3	q11a и q11b

И оставьте функцию **Sum of values**, устанавливаемую по умолчанию. Введите заголовок и постройте диаграмму. Внесите ее в отчет и напишите комментарий.

Диаграмма рассеяния отображает отношения между двумя переменными, которые как минимум относятся к интервальной шкале. Выберите в меню: **Graphs** ⇒ **Scatter**. Откроется диалоговое окно **Scatterplot**. Щелкните на области **Simple**, щелчком по выключателю **Define**, откройте следующее диалоговое окно. Поскольку в социологических исследованиях очень мало данных интервальной шкалы или шкалы отношений, то все делают один вариант. Переменную qd6 из списка исходных переменных перенесите в поле оси Y, а переменную qs2 – в поле оси X. Если теперь поместить какую-нибудь другую переменную в поле **Set Markers by**, то согласно принадлежности к этой переменной отдельные точки на диаграмме будут представлены окрашенными в другой цвет (можно пометить их при помощи маркировочного символа). Попробуйте сделать это, поместив в поле переменную qs3 или qd2a. Посмотрите, что получится. Скопируйте отредактированные графики в отчет и прокомментируйте их.

Круговые диаграммы

Представление данных в виде круговых диаграмм стоит выбирать тогда, когда частоты или значения переменных можно, не нарушая здравого смысла, сложить вместе и эта сумма будет соответствовать 100%.

Отобразите при помощи круговой диаграммы частоты категорий переменных из вашего варианта:

Варианты для задания 15

№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12	№13	№14	№15
qd5	qd7	q65c	q65b	q52	q50a	q54a	q55	q65a	q50b	q41d	q39	q41a	q41b	q36b

Выберите в меню: **Graphs** ⇒ **Pie**. Откроется диалоговое окно **Pie Charts**. Оставьте опцию **Summaries for groups of cases**, установленную по умолчанию и щелчком по кнопке **Define**, откройте следующее диалоговое окно. В поле **Define slices by** введите вашу переменную. Щелкните на выключателе **Options** и уберите маркер с опции **Display groups by missing values**. С помощью выключателя **Titles** введите заголовок и начните построение диаграммы щелчком на **OK**. Попробуйте самостоятельно придать этой диаграмме более интересный вид.

Построение интерактивных графиков

Диаграммы, которые можно построить с помощью интерактивного графического интерфейса, включают в себя следующие виды:

- Столбчатые диаграммы
- Линейные диаграммы
- Круговые диаграммы
- Коробчатые диаграммы
- Диаграммы величины ошибки
- Гистограммы
- Диаграммы рассеяния

Выберите в меню: **Graphs** (Графики) ⇒ **Interactive** ⇒ **Bar...** Откроется диалоговое окно, которое имеет пять регистрационных карт, первая из которых **Assign Variables** (Присвоить переменные) открывается по умолчанию. В ней находится список переменных и пять полей для ввода переменных, две кнопки с символами, соответствующими двум возможностям построения диаграммы и три выключателя. В окне можно выбрать тип сортировки переменных. Переменные, находящиеся в списке переменных, можно разделить на два типа: категориальные (номинальная или порядковая шкала) и метрические. Эти два типа идентифициру-

ются при помощи двух разных символов, устанавливаемых в начале имени. Тип переменной может быть изменен после ее активации и щелчка правой кнопки мыши. Разберитесь с диалоговым окном самостоятельно и постройте разнотипные графики для ваших переменных, учтите все возможные варианты. Работу над графиками и сами графики опишите в отчете.

Попробуйте активировать опции **Count**, **Value**, **Custom** и посмотрите, что получится. Запишите в отчет.

На регистрационной карте **Error Bars** можно организовать указание доверительных интервалов, стандартных отклонений и стандартных ошибок.

В регистрационной карте **Titles** можно указать заголовок. Подзаголовок и комментарии.

В зависимости от пола можно построить графики частотного распределения, используя введение легенды, отражающей принадлежность столбцов при помощи цвета и узора. Для этого следует активировать двумерное представление диаграммы и перенести переменную «пол» в поле **Color** группы **Legend Variable**. Обратите внимание, чтобы переключатель был установлен в положение **Cluster**. В поле **Panel Variable** можно поместить переменную в качестве разделителя полей, чтобы получить еще большую наглядность.

Вместо группированной столбчатой диаграммы можно построить штабельную столбчатую диаграмму, если переключатель установить в положение **Stack**.

Отображение информации в линейчатом виде целесообразно, если вы хотите показать изменение показателей с течением времени. Поскольку таких данных не имеется, то можно попробовать построить линейчатый график для переменных qd6 и qs2. Не перепутайте оси. Для этого выберите в меню: **Graphs** (Графики) \Rightarrow **Interactive** \Rightarrow **Line** (линия). Вы попадете в диалоговое окно построения интерактивных линейных графиков. В левой части окна находится перечень задействованных переменных, в правой – оси координат. Мышью перетаскиваем переменные на их места на осях координат.

Все построенные графики перенести в отчет (через специальную вставку) и прокомментировать полученные результаты.