

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА АЛЕКСАНДРА I»
(ФГБОУ ВО ПГУПС)

И. Л. Соломин

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ

Учебное пособие

САНКТ-ПЕТЕРБУРГ

2021

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА АЛЕКСАНДРА I»
(ФГБОУ ВО ПГУПС)

И. Л. Соломин

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ

Учебное пособие

САНКТ-ПЕТЕРБУРГ

2021

УДК 159.9:51
ББК 88.3
С60

Рецензенты:
доцент кафедры «Психология»,
кандидат педагогических наук СПбАППО
А. Г. Думчева;
профессор кафедры «Прикладная психология»
ФГБОУ ВО ПГУПС, доктор психологических наук
М. А. Бендюков

Соломин, И. Л.

С60 Математические методы в психологии : учебное пособие / И. Л. Соломин. – Санкт-Петербург : ФГБОУ ВО ПГУПС. 2021. – 50 с.

ISBN 978-5-7641-1681-5

В пособии описываются базовые математические методы, необходимые в процессе подготовки и дальнейшей научной и практической профессиональной деятельности психологов, для статистической обработки данных эмпирического психологического исследования. Пособие основано на использовании пакета компьютерных статистических программ PAST, описывает средства ввода и редактирования данных, графические средства, одномерные и многомерные математические методы, включая описательную статистику, параметрические и непараметрические критерии, корреляционный, регрессионный, дисперсионный, факторный, кластерный, дискриминантный и другие виды анализа.

Пособие предназначено для обучающихся по направлению «Психология».

УДК 159.9:51
ББК 88.3

© Соломин И. Л., 2021
© ФГБУ ВО ПГУПС, 2021

ISBN 978-5-7641-1681-5

Введение

Развитие и современное состояние психологической науки и практики тесно связано с использованием математических методов. В XIX веке голландский физиолог Франц Дондерс впервые решил задачу измерения времени простых и сложных сенсомоторных реакций. Немецкий психолог Густав Фехнер разработал методы измерения порогов чувствительности и открыл психофизический закон, описывающий логарифмическую зависимость силы ощущений от силы раздражителя. Герман Эббингауз в результате экспериментального изучения получил известную кривую забывания, а американский психолог Эдвард Торндайк – кривую научения. Английский антрополог и психолог Френсис Гальтон внес значительный вклад в развитие дифференциальной психологии, психометрии и статистики, а его ученик Карл Пирсон предложил способ вычисления коэффициента корреляции.

В XX веке английский психолог Чарльз Спирмен и американские психологи Луис Терстоун и Джой Гилфорд разработали методы факторного анализа, с помощью которых были созданы факторные теории и методики диагностики интеллекта. На основе факторного анализа Гансом Айзенком и Реймондом Кеттеллом были созданы известные личностные опросники, Джордж Келли разработал методику репертуарных решеток для анализа личностных конструктов, а Чарльз Осгуд – методику семантического дифференциала для вычисления расстояний между понятиями в многомерном пространстве субъективных признаков. Американский психолог и психофизик Стэнли Стивенс стал автором теории уровней измерения и шкалирования.

Во второй половине XX века психологи начали использовать в своих работах теорию информации, теорию обнаружения сигналов, теорию автоматического управления, кибернетику, теорию алгоритмов, теорию игр, разрабатывали математические модели когнитивных процессов, а в конце XX – начале XXI века – методы и технологии экспертных систем, распознавания образов, машинного обучения, нейронных сетей, искусственного интеллекта, «больших данных». В настоящее время математические методы используются психологами-исследователями и практиками для планирования эксперимента, измерения и шкалирования, моделирования психических явлений, но главным образом для статистической обработки эмпирических данных.

Опыт преподавания различных дисциплин для студентов-психологов, а также практика руководства курсовыми и дипломными работами свиде-

тельствуют о том, что использование математических методов для обработки данных вызывает у многих обучающихся значительные затруднения. И даже достаточно квалифицированные специалисты-психологи при подготовке к защите кандидатских или докторских диссертаций, в ходе проведения научно-исследовательских работ, в процессе обработки, описания и интерпретации эмпирических результатов иногда испытывают проблемы. Отчасти это связано с тем, что психология многими представляется гуманитарной наукой, не предполагающей точных измерений и количественных расчетов. Практическая психология в форме консультирования и тренингов требует коммуникативных способностей и навыков, вербального и эмоционального интеллекта, интуиции. Однако решение практических задач диагностики и прогнозирования, а также прикладные и фундаментальные исследования в области психологии, связанные с тестированием, опросами, оценками, проведением экспериментов невозможны без наличия у специалиста числовых и пространственных интеллектуальных способностей, знаний основ математики и навыков обработки количественных данных.

Освоение математических методов психологами в большой степени затрудняется тем, что значительное число учебников для психологов написаны профессиональными математиками и основаны на сложном аппарате теории вероятностей и математической статистики. Классические учебники [2, 4, 6], основанные на трудоемких вычислениях, сделанных вручную, позволяют освоить методы описательной статистики, параметрические и непараметрические критерии и коэффициенты корреляции, но на практике мало полезны, поскольку ручные расчеты требуют больших затрат времени и сил. Современные учебные пособия, активно использующие электронные таблицы MS Excel [3], позволяют эффективно вводить, редактировать и обрабатывать данные, вычислять параметры распределений, рассчитывать t -критерии Стьюдента и коэффициенты линейной корреляции Пирсона. Однако использовать Excel для вычисления непараметрических критериев и ранговых корреляций может быть затруднительно, а факторный, кластерный, регрессионный, дискриминантный и другие многомерные методы – невозможны. Наконец, ряд современных и несомненно высококачественных учебных пособий [1, 5] рассматривают использование специализированных коммерческих пакетов компьютерных программ для статистической обработки данных Statistica фирмы StatSoft или SPSS фирмы IBM, легальное приобретение которых специалистами и тем более студентами невозможно вследствие их гигантской стоимости, доходящей до нескольких сотен тысяч рублей в год.

Данное учебное пособие ориентировано на использование пакета компьютерных программ PAST. Пакет статистических программ PAST разработан под руководством профессора Оувинда Хаммера и других сотрудников Музея естественной истории Университета Осло в Норвегии. Изначально пакет был предназначен для анализа палеонтологических данных, отсюда

его название – PAleontological STatistics. Представленные в пакете методы статистического анализа универсальны и позволяют обрабатывать любые данные: геологические, биологические, психологические, социологические, экономические и т. д. Первая версия PAST была разработана в 1999 году. На начало 2021 года доступна версия 4.05. Официальная ссылка, с которой можно скачать последнюю версию пакета и руководство пользователя: <https://www.nhm.uio.no/english/research/infrastructure/past>.

Пакет программ может эффективно использоваться студентами, преподавателями и научными сотрудниками. Достоинствами данного пакета являются:

- некоммерческий характер и возможность бесплатного использования;
- небольшой размер программы, не превышающий 30 Мб, возможность запуска без установки на компьютер с любого носителя, например с USB-флешки, на компьютерах с минимальными техническими характеристиками;
- огромное количество алгоритмов статистической обработки данных, сопоставимое с такими коммерческими пакетами, как Statistica и SPSS, позволяющее проводить одномерный и многомерный анализ данных, включая вычисление параметров распределений, оценку достоверности различий с помощью параметрических и непараметрических критериев, линейный и ранговый корреляционный анализ, дисперсионный, факторный, кластерный, регрессионный и дискриминантный анализ, реализовать многообразные формы графического представления результатов.

Единственным ограничением PAST является то, что программа и документация написаны на английском языке – это может вызвать определенные затруднения у русскоязычных пользователей.

Все это побудило автора пособия написать методическое руководство для русскоязычных студентов и научных руководителей, позволяющее осуществлять полноценную статистическую обработку данных для курсовых и выпускных квалификационных работ. Впрочем, программа может быть полезна также и в процессе анализа данных для защиты кандидатских и докторских диссертаций и других научных исследований.

Для овладения пакетом программ PAST от пользователя требуется предварительное знакомство с теоретическими основами планирования эксперимента, измерения и шкалирования и статистической обработки данных.

Данное учебное пособие не заменяет, а дополняет имеющуюся учебную литературу [1–6]. В пособии содержится описание программы PAST, способов ввода и редактирования данных, получения графиков, использования методов одномерной и многомерной статистической обработки.

В результате освоения дисциплины «Математические методы в психологии» обучающиеся должны

знать:

- понятия экспериментальных планов, независимых, зависимых и дополнительных переменных;
- уровни измерения и шкалирования показателей, характеристики шкал наименований, порядка, равных интервалов и равных отношений, понятия валидности и надежности измерения;
- основные методы одномерной, двухмерной и многомерной обработки данных, описательной статистики и вычисления параметров распределений, оценки достоверности различий с помощью параметрических и непараметрических критериев, выявления взаимосвязей между показателями, вычисления корреляций, основы дисперсионного, регрессионного, факторного, дискриминантного и кластерного анализа;
- возможности стандартных пакетов компьютерных программ статистической обработки данных;

уметь:

- планировать исследование, выбирать методы статистической обработки в зависимости от решаемых задач, используемых шкал и полученных данных;
- вводить данные в компьютер, использовать стандартные программы статистической обработки данных;
- представлять результаты статистической обработки данных в форме таблиц и графиков, оценивать достоверность, описывать и интерпретировать полученные результаты.

1. Запуск программы PAST, интерфейс программы, группы инструментов и пункты меню

Для работы программы PAST требуется операционная система Windows 7-й, 8-й или 10-й версии. Единственный исполнимый exe-файл программы просто копируется в любую папку на диск компьютера или внешний накопитель и запускается. Последняя, 4-я версия, программы PAST (past4.exe) предназначена только для работы под 64-разрядной версией Windows, которая наиболее часто устанавливается на современных мощных компьютерах. Предыдущая, 3-я, версия программы (past3.exe) работает как под 64-разрядной, так и под 32-разрядной версией Windows, которая предназначена для работы на более старых и слабых компьютерах. Узнать тип операционной системы Windows (32- или 64-разрядная), установленной на компьютере, можно с помощью меню «Пуск» и далее «Параметры» → «Система» → «О системе» или с помощью программы «Панель управления» и далее «Система».

Интерфейс программы похож на стандартную электронную таблицу (рис. 1). Строки таблицы соответствуют объектам, людям, наблюдениям, столбцы – признакам, показателям, переменным. Данные в таблицу могут вводиться вручную или копироваться из другой электронной таблицы, например MS Excel, и вставляться. В качестве разделителей целых и дробных значений могут использоваться как десятичные запятые, так и точки.

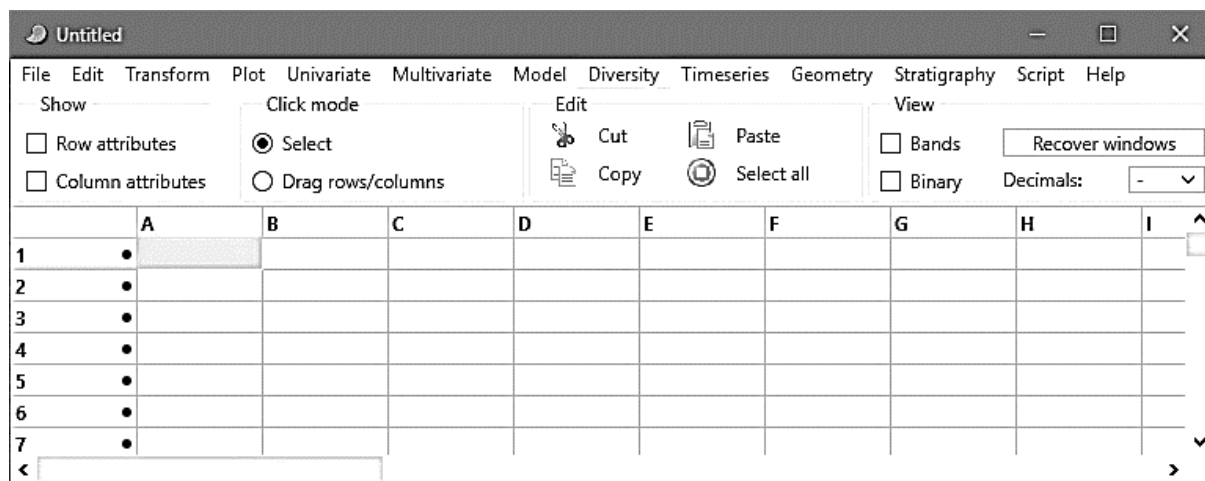


Рис. 1. Интерфейс программы PAST

В заголовке программы указывается имя открытого файла. Ниже представлена строка меню. Под строкой меню – лента с группами основных инструментов.

Группа Show (Показать)

Row attributes (Атрибуты строк) – для каждой строки в шапке открываются три дополнительных поля: *Color (Цвет)*, *Symbol (Символ)* и *Name (Имя)*, которые позволяют обозначить соответствующий объект, наблюдение или испытуемого выбранным цветом, символом (точкой, плюсиком, квадратиком и т. п.) и именем, которые будут отображаться в диаграммах рассеивания (рис. 2).

Column attributes (Атрибуты столбцов) – для каждого столбца в шапке открываются два дополнительных поля: *Type (Тип)* и *Name (Имя)*, которые позволяют для соответствующих признаков, показателей или переменных выбрать тип и ввести имя (рис. 3, 4). По умолчанию типом является любое числовое значение, обозначенное дефисом. Если значение переменной не является числом, следует указать другой тип: *Group (Группа)*, *Ordinal (Порядковое значение или ранг)*, *Nominal (Наименование)*, *Binary (Значение 0 или 1)* или *String (Строка символов)*. Наиболее полезным является тип *Group*, позволяющий указать, например, пол (мужской или женский),

направление подготовки (техническое или гуманитарное), категорию должности (руководитель или специалист), семейное положение, наличие детей, род занятий и др. Тип *Group* необходимо использовать для оценки достоверности различий между группами, для дисперсионного или дискриминантного анализа.

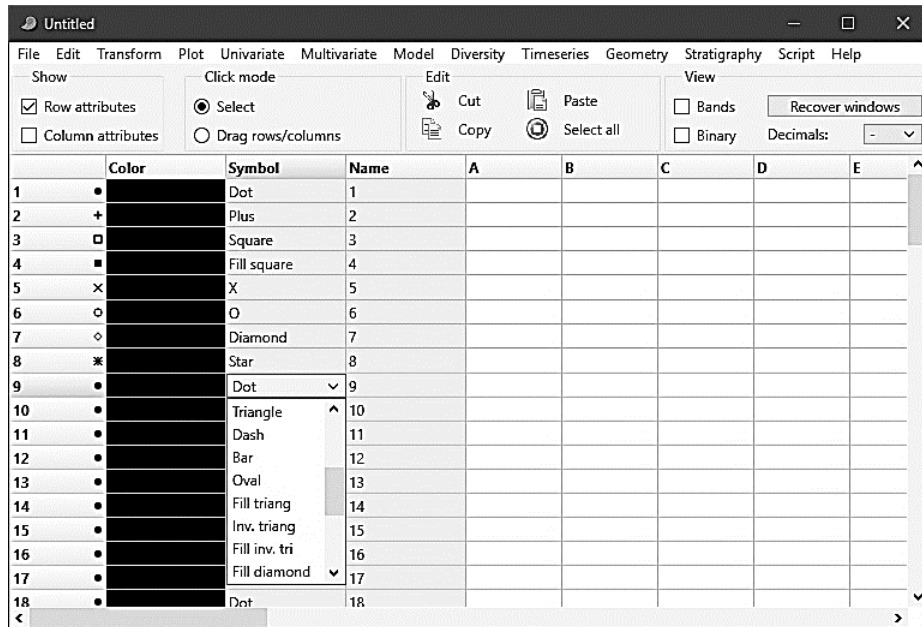


Рис. 2. Атрибуты строк

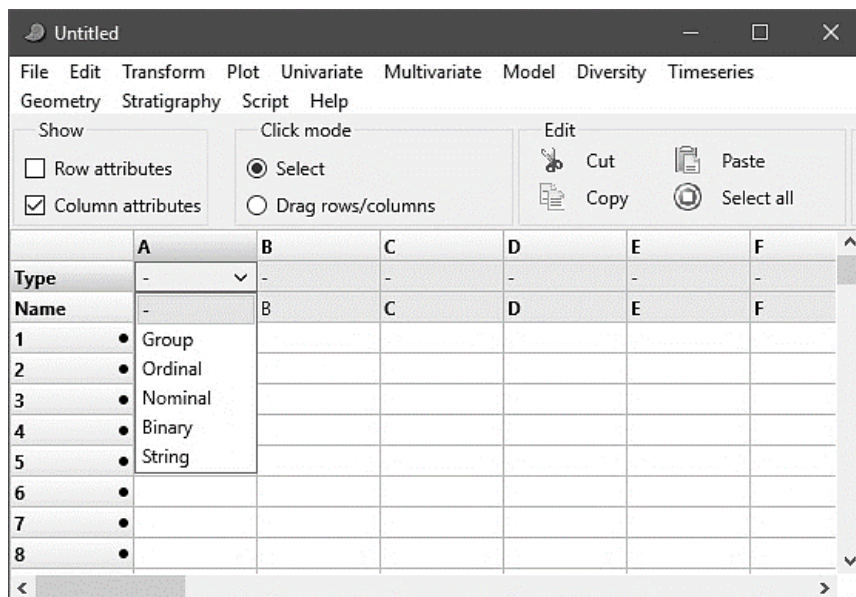


Рис. 3. Атрибуты столбцов. Выбор атрибута столбца

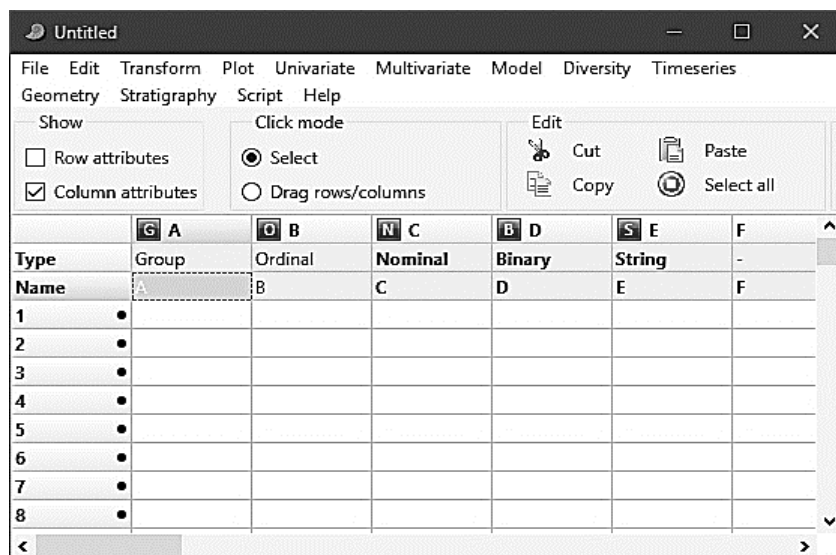


Рис. 4. Атрибуты столбцов. Выбранные атрибуты столбцов

Группа Click Mode (Режим работы мыши)

Select (Выбор) – выделение. Щелчок левой клавишей мыши по заголовку столбца или первому столбцу строки выделяет соответствующий столбец или строку. При нажатой клавише Ctrl, щелкая левой клавишей мыши, можно выделить любое количество столбцов или строк, даже если они не будут смежными. При нажатой клавише Shift, щелкая левой клавишей мыши по первому и последнему столбцу интервала или манипулируя стрелками на клавиатуре влево или вправо, можно выделить нужный интервал смежных столбцов.

Drag row/columns (Перетаскивание строк/столбцов) – перемещение любого столбца или строки мышью за заголовок. Особенно полезно, если необходимо изменить последовательность столбцов.

Группа Edit (Редактирование)

Cut (Вырезать) вырезает выделенные элементы (ячейки, строки или столбцы) и помещает в буфер обмена.

Copy (Копировать) копирует выделенные элементы в буфер обмена.

Paste (Вставить) вставляет вырезанные или скопированные элементы из буфера обмена.

Select all (Выделить все) выделяет все строки и столбцы.

Группа View (Вид)

Band (Полоса) выделяет желтой заливкой все нечетные строки для лучшего визуального различия.

Binary выделяет черной заливкой ненулевые значения.

Decimals определяет количество выводимых знаков после десятичной точки или запятой.

Пункты меню

Для статистического анализа данных психологического исследования могут использоваться пункты меню *File (Файл)*, *Edit (Редактирование)*, *Plot (Графики)*, *Univariate (Одномерные методы)*, *Multivariate (Многомерные методы)*, *Model (Модели)*.

Пункт меню *Transform (Преобразовать)* позволяет осуществлять математические преобразования данных, такие как получение логарифмов, вычитание средних, получение рангов значений, получение процента от суммарного значения, вычисление выражений по введенным формулам и т. д. Возможности данного пункта меню PAST для обработки данных психологического исследования либо неприменимы, либо используются редко, либо более удобно реализованы в MS Excel.

Пункт меню *Script (Язык сценариев)* используется для написания собственных программных модулей и требует высокого уровня компетентности в области программирования.

Другие пункты меню PAST, используемые в геологии, биологии и прочих естественных науках для обработки психологических данных не применяются:

- *Diversity (Анализ разнообразия)*;
- *Timeseries (Анализ временных рядов)*;
- *Geometry (Пространственный анализ)*;
- *Stratigraphy (Анализ геологических пластов)*.

2. Пункт меню File (Файл) и ввод данных

New (Новый) создает новую пустую таблицу для ввода данных.

Open... (Открыть...) открывает имеющийся файл данных. Можно открывать как файлы данных в формате PAST с расширением «.dat», так и файлы данных в формате MS Excel 97–2003 с расширением «.xls», но не в формате MS Excel 2007–2019 с расширением «.xlsx».

Save (Сохранить) сохраняет данные в текущем открытом файле с тем же именем.

Save as... (Сохранить как...) сохраняет данные в файле под другим именем. Можно сохранять файлы данных в формате PAST с расширением «.dat» или в формате MS Excel с расширением «.xls».

Exit (Выход) закрывает программу.

Для ввода и редактирования данных удобнее всего использовать программу MS Excel. Лист с таблицей данных в файле Excel должен быть единственным или самым первым. В первой строке таблицы вводятся названия показателей, в следующих строках – значения этих показателей по каждому испытуемому (рис. 5). В данном примере представлены значения показателей 2, 4, 7 и 10-го субтестов универсального интеллектуального теста (УИТ),

2-го показателя теста сходных рисунков Кагана и 11 показателей модифицированного патохарактерологического диагностического опросника (МПДО) школьников 9-го и 11-го классов. Шапка таблицы с названиями показателей должна состоять только из одной строки. Недопустимо использовать две и более строки в шапке, например, первую строку с названиями методик, вторую строку – с названиями показателей.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Класс	Пол	Зрительное восприятие (ЗИТ2)	Числовой интеллект (ЧИТ4)	Вербальный интеллект (ВИТ7)	Пространственный интеллект (ПИТ10)	Время первого ответа (Тест Кагана)	Количество ошибок (Тест Кагана)	Гипертимный (МПДО1)	Циклоидный (МПДО2)	Лабильный (МПДО3)	Астено-невротический (МПДО4)	Сензитивный (МПДО5)	Психастенический (МПДО6)	Шизоидный (МПДО7)	Эпилептоидный (МПДО8)	Истероидный (МПДО9)	Неустойчивый (МПДО10)	Неискренность (МПДО11)
2	9	Ж	10	5	7	6	27	4	9	2	6	3	2	3	3	4	9	4	4
3	9	Ж	6	10	8	7	57	0	10	5	6	3	4	0	3	5	8	4	1
4	11	М	8	12	5	7	80	5	12	3	5	1	4	2	7	5	7	3	5
5	9	Ж	8	8	5	8	117	3	8	9	12	4	9	3	8	8	4	5	3
6	11	Ж	8	4	4	8	46	4	11	9	9	8	5	5	7	8	10	4	0
7	11	М	6	3	5	6	157	6	11	3	3	4	3	1	3	5	8	2	1
8	9	М	10	8	7	9	50	5	5	6	7	2	1	2	7	1	6	4	0
9	11	Ж	4	5	7	9	131	0	9	6	5	1	2	3	1	1	5	6	4
10	9	Ж	6	10	4	8	36	0	5	0	5	1	4	7	2	2	1	1	3

Рис. 5. Пример данных в MS Excel

Полученный файл следует сохранить в формате MS Excel 97–2003 «.xls» (не «.xlsx»). После этого данный файл можно открыть с помощью пункта меню «File/Open...» программы PAST, при этом в появившемся окне (рис. 6) указать, что *Row contains – Only data cells* (Строки содержат только ячейки данных) и *Columns contain – Names, data* (Столбцы содержат имена переменных и данные).

Рис. 6. Выбор содержания строк и столбцов при открытии файла MS Excel

Другим способом переноса введенных в Excel данных в программу PAST является использование буфера обмена. Для этого следует выделить все необходимые в Excel строки, включая первую с названиями показателей, и столбцы, скопировать их в буфер обмена, открыть программу PAST, поставить флажок в поле *Column attributes*, поставить курсор в ячейку *Name A* и вставить скопированную в Excel таблицу (пример – на рис. 7).

После импорта файла данных из Excel в PAST следует в нужных столбцах указать тип переменных *Group*, а затем сохранить полученный файл PAST в выбранной папке и дать этому файлу имя.

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество с	Гипертимны
1	9	ж	10	5	7	6	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9
9	9	ж	6	10	4	8	36	0	5

Рис. 7. Данные в PAST

3. Пункт меню Edit (Редактировать)

Undo (Вернуть назад) отменяет последнюю операцию.

Redo (Повторить) возвращает и повторяет отмененную операцию.

Cut (Вырезать) вырезает выделенные элементы (ячейки, строки или столбцы) и помещает в буфер обмена.

Copy (Копировать) копирует выделенные элементы в буфер обмена.

Paste (Вставить) вставляет вырезанные или скопированные элементы из буфера обмена.

Select all (Выделить все) выделяет все строки и столбцы.

Insert more rows... (Вставить больше строк...) вставляет указанное количество строк выше выделенной.

Insert more columns... (Вставить больше столбцов...) вставляет указанное количество столбцов левее выделенного.

Remove (Удалить) удаляет выделенные строки или столбцы.

Remove uninformative row/columns... (Удалить неинформативные строки/столбцы...) удаляет строки или столбцы *All-zero* (Только с нулевыми значениями), *All-missing* (Только с отсутствующими значениями), *Singleton* (Только с одним ненулевым значением), *Constant* (Только с постоянным одинаковым значением) (рис. 8).

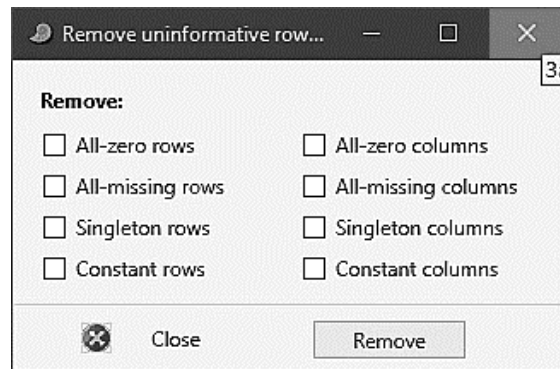


Рис. 8. Удаление неинформативных строк или столбцов

Find... (Найти...) ищет заданные значения.

Replace... (Найти и заменить...) ищет и меняет заданные значения.

Fill... (Заполнить...) заполняет выделенные ячейки заданным значением.

Rearrange → (Изменить порядок).

Rearrange → ***Transpose*** (Транспонировать) поворачивает таблицу на 90 градусов, меняет местами строки и столбцы, когда строки становятся столбцами, а столбцы – строками.

Rearrange → ***Observations to contingency table*** (Наблюдения к таблице сопряженности) преобразует данные двух столбцов, имеющих целые значения, в таблицу частот сочетаний (например, один столбец со значениями 0 и 1, другой – 1, 2 и 3 в таблицу частот сочетаний 2×3), которую можно обработать с помощью *Contingency table* (Таблица сопряженности) в пункте меню *Univariate*.

Rearrange → ***Value pairs to matrix*** (Пары значений в матрицу) преобразует данные двух столбцов, имеющих значения типа *Group* или *Nominal*, в таблицу частот сочетаний, которую можно обработать с помощью *Contingency table* (Таблица сопряженности) в пункте меню *Univariate*.

Например, в предыдущем файле «Школьники.dat» можно выделить первые два столбца – «Класс» и «Пол» (рис. 9), выбрать пункт меню *Value pairs to matrix* и получить таблицу частот сочетаний (рис. 10).

Rearrange → ***Sort ascending*** (Сортировать в возрастающем порядке) сортирует всю таблицу в возрастающем порядке по выделенному столбцу.

Rearrange → **Sort descending** (*Сортировать в убывающем порядке*) сортирует всю таблицу в убывающем порядке по выделенному столбцу.

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество	Гипертимный
1	9	ж	10	5	7	6	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9

Рис. 9. Выделение столбцов для таблицы частот сочетаний

	9	11
ж	19	13
м	5	11

Рис. 10. Таблица частот сочетаний

4. Пункт меню Plot (Графики)

X_Y graph (*Двухкоординатный график*) строит двумерную диаграмму рассеивания (скатерграмму) по двум выделенным столбцам. Например, выделяем столбцы «Числовой интеллект» и «Пространственный интеллект» (рис. 11).

Результаты выводятся в окно *X_Y graph* (рис. 12).

Отметка *Labels* (*Метки*) позволяет вывести имена или номера точек, *95% ellipses* (*95 % эллипс*) описывает множество точек эллипсом с 95 % доверительным интервалом, *Flip axes* (*перевернуть оси*) меняет местами оси, т. е. поворачивает график на 90 градусов (рис. 13).

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество	Гипертимный
1	9	ж	10	5	7	6	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9

Рис. 11. Выделение двух столбцов для получения двумерной диаграммы рассеивания

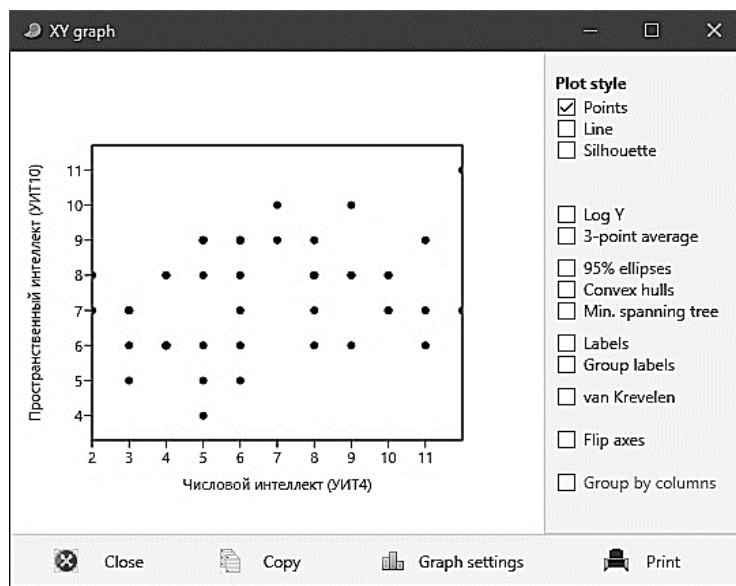


Рис. 12. Окно XY graph

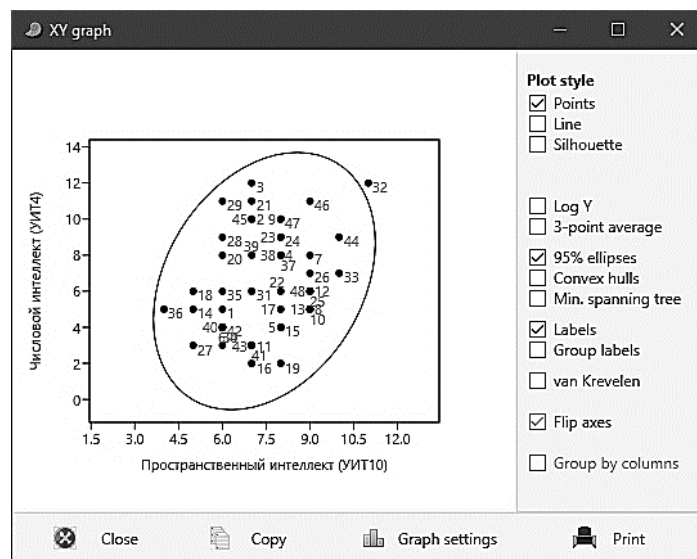


Рис. 13. Модификация XY graph

Здесь и далее во всех окнах графиков кнопка *Copy* копирует полученный график в буфер обмена, после чего график можно вставить из буфера в любой документ, в том числе MS Excel или MS Word.

Кнопка *Graph settings* (*Параметры графика*) открывает окно *Graph preferences* (*Предпочтения графика*), которое позволяет изменить *Font* (*Тип и размер шрифта*), *Symbol size* (*Размер точек*), *X label* и *Y label* (*Подписи осей X и Y*), *Minimal tick marks x* и *Minimal tick marks y* (*Шаг делений осей X и Y*) (рис. 14).

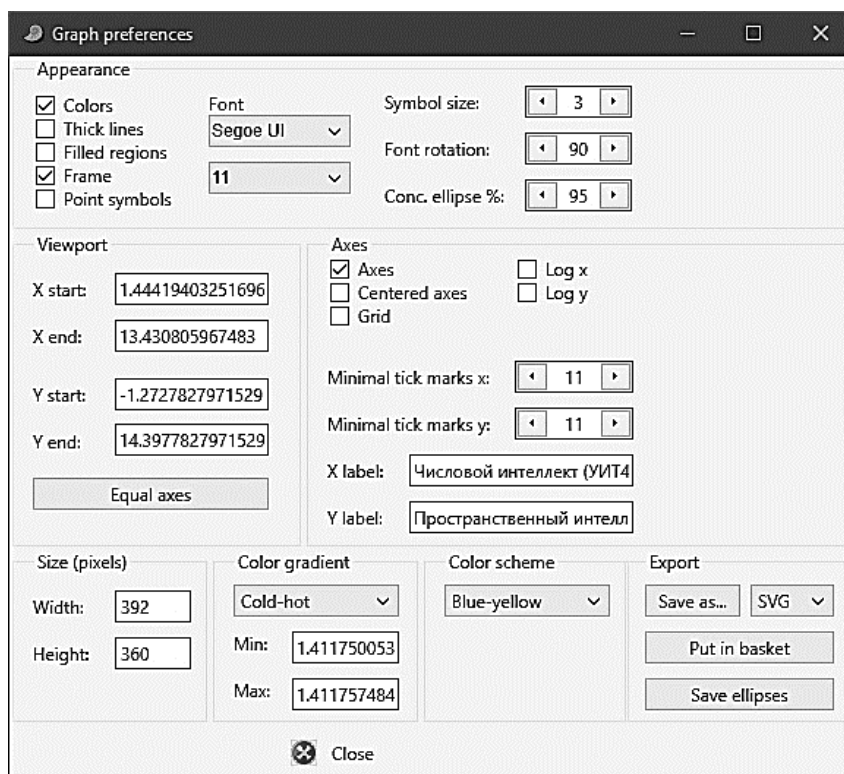


Рис. 14. Параметры графика

Histogram (**Гистограмма**) строит гистограмму распределения частот значений по выделенному столбцу. Например, выделяем столбец «*Пространственный интеллект*» (рис. 15).

Результаты выводятся в окно *Histogram* (рис. 16).

Параметр *Bins* позволяет изменить число интервалов группировки значений, *Fit normal* – наложить кривую нормального распределения, *Flip axes* – повернуть график на 90 градусов (рис. 17). Кнопка *Graph settings* (*Параметры графика*) дает возможность *Font* (*Тип и размер шрифта*), *X label* и *Y label* (*Подписи осей X и Y*), *Minimal tick marks x* и *Minimal tick marks y* (*Шаг делений осей X и Y*), *Color scheme* (*Цветовая схема*). В данном графике полезно в параметрах графика вручную указать подпись оси *X* с названием показателя.

Школьники.dat

File Edit Transform Plot Univariate Multivariate Model Diversity Timeseries Geometry Stratigraphy Script Help

Show

Click mode

Row attributes Column attributes

Select Drag rows/columns

Edit

Cut Copy Paste Select all

View

Bands Binary Decimals: -

Recover windows

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество	Гипертимны
1	9	ж	10	5	7	5	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9

Рис. 15. Выделение столбца для получения гистограммы

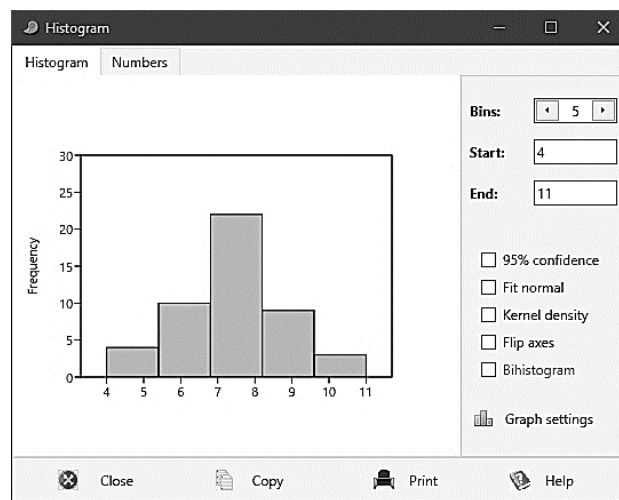


Рис. 16. Окно Histogram

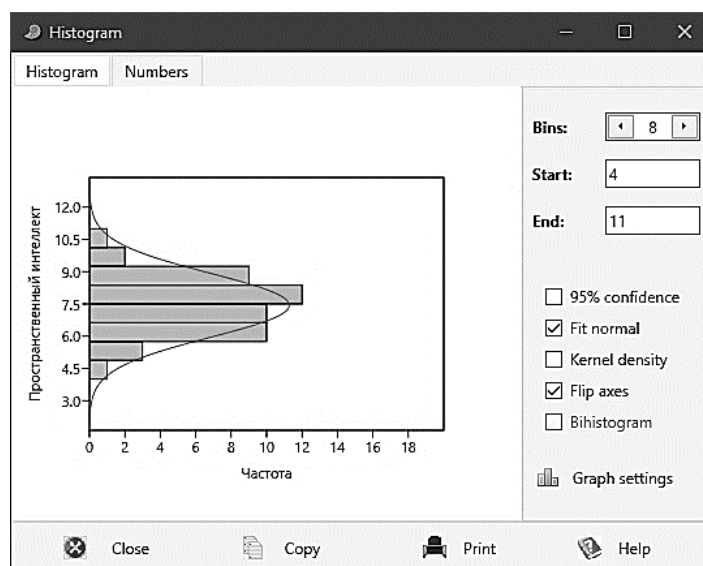


Рис. 17. Модификация Histogram

Для двух выделенных столбцов, один из которых соответствует типу *Group* (рис. 18), можно получить *Bihistogram* (Двойную гистограмму) (рис. 19).

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество с	Гипертимны
1	9	ж	10	5	7	5	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9

Рис. 18. Выделение столбцов для двойной гистограммы, так называемой диаграммы средних с усиками, двухвыборочного теста или однофакторного дисперсионного анализа

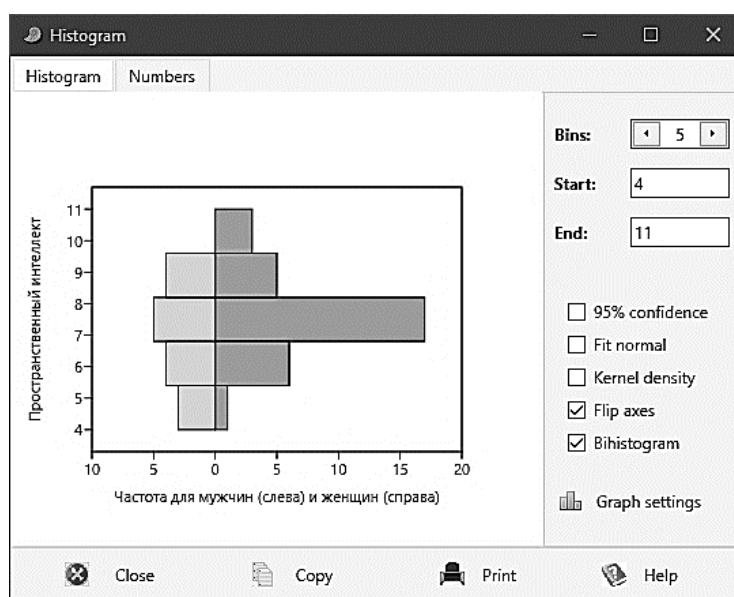


Рис. 19. Двойная гистограмма

***Barchart/boxplot* (Столбиковая/прямоугольная диаграмма)** строит график параметров распределений для одного или нескольких выделенных однотипных показателей (рис. 20).

Результаты выводятся в окно *Box plot* (рис. 21–23). Можно выбрать *Plot type* (Тип графика) – *Bar chart* (Столбиковая диаграмма), *Whisker type* (Тип усов) – *Standard error* (Стандартная ошибка) или *Standard deviation* (Стандартное отклонение), *Whisker length* (Длина усов) – *One sigma* (Одна

сигма) или 95% interval (95% доверительный интервал). Для того чтобы названия показателей помещались в данные графики, они должны быть достаточно короткими в таблице данных.

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество с	Гипертимный
1	9	ж	10	5	7	6	27	4	9
2	9	ж	6	10	8	7	57	0	10
3	11	м	8	12	5	7	80	5	12
4	9	ж	8	8	5	8	117	3	8
5	11	ж	8	4	4	8	46	4	11
6	11	м	6	3	5	6	157	6	11
7	9	м	10	8	7	9	50	5	5
8	11	ж	4	5	7	9	131	0	9

Рис. 20. Выделение нескольких столбцов для столбиковой диаграммы, вычисления параметров распределений или корреляций

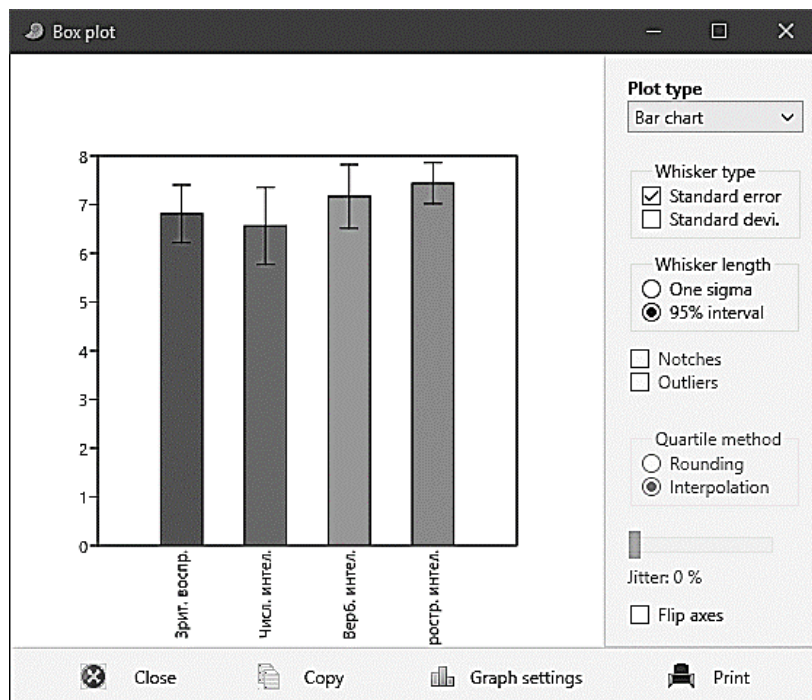


Рис. 21. Окно Box Plot, тип графика Bar chart (Столбиковая диаграмма)

Можно построить график для двух показателей, один из которых имеет тип *Group*, например зависимость пространственного интеллекта от пола (см. рис. 18, 24).

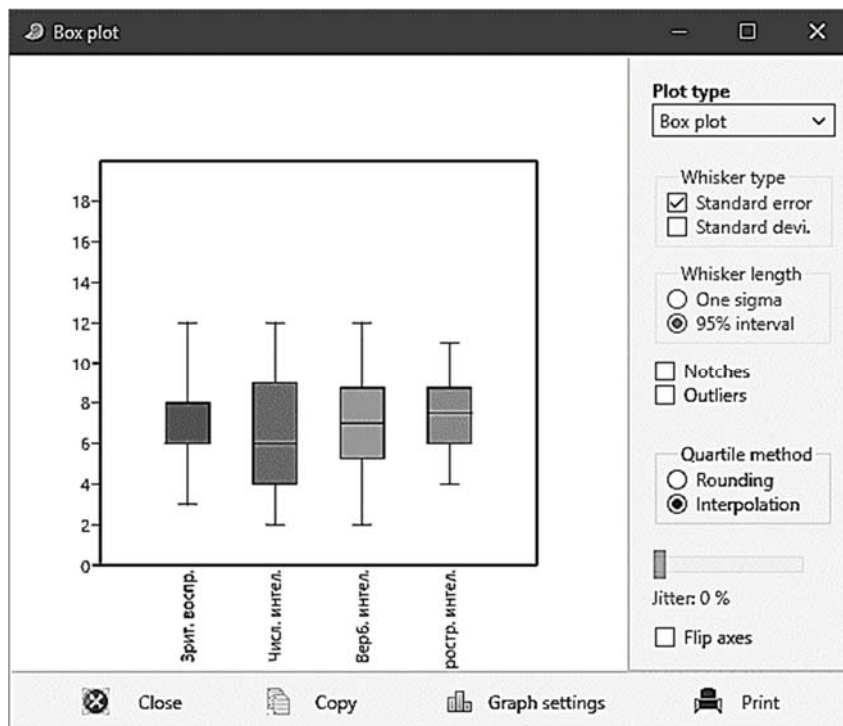


Рис. 22. Тип графика Box plot
(Прямоугольная диаграмма)

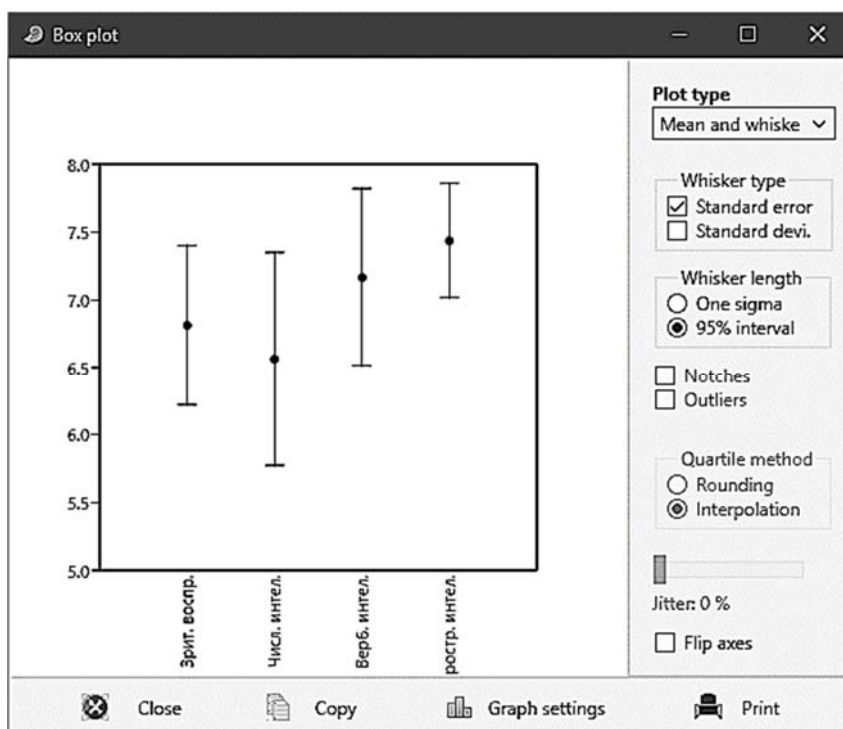


Рис. 23. Тип графика Mean and whiskers
(Среднее с усиками)

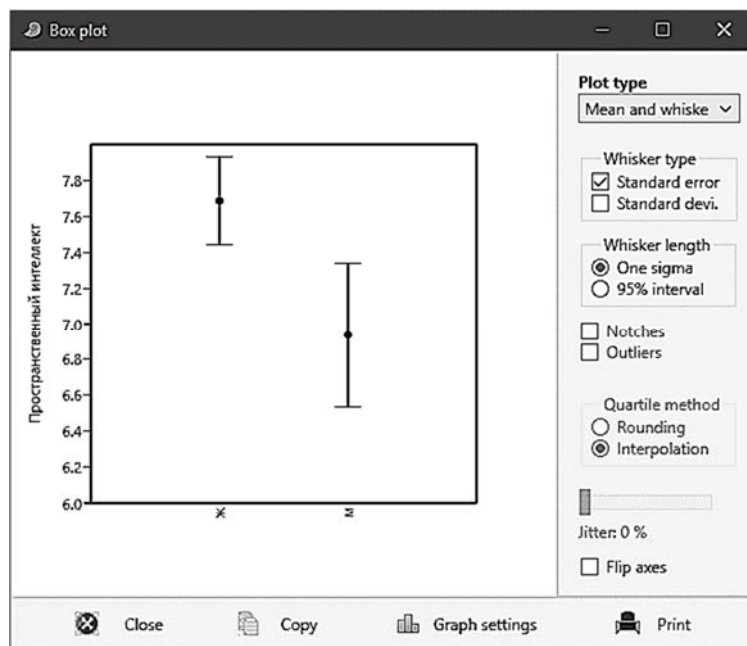


Рис. 24. Тип графика Mean and whiskers (Среднее с усиками)

5. Пункт меню Univariate (Одномерные методы)

Summary statistics (Описательная статистика) дает возможность вычислить параметры распределений показателей по выделенным столбцам (см. рис. 20, 25):

- N – количество значений в выборке;
- Min – минимальное значение;
- Max – максимальное значение;
- Sum – суммарное значение;
- $Mean$ – среднее арифметическое, отношение суммы всех значений к количеству значений;
- $Std. error$ – стандартная ошибка среднего, отношение стандартного отклонения к корню квадратному из количества значений;
- $Variance$ – дисперсия, квадрат стандартного отклонения;
- $Stand. dev.$ – *standard deviation*, стандартное (среднеквадратичное) отклонение, корень квадратный из отношения суммы квадратов отклонений значений от среднего к сумме значений;
- $Median$ – медиана, 50 перцентиль, значение, меньше которого встречается половина всех значений;
- $25\ prcntil$ – 25 перцентиль, значение, меньше которого встречается четверть всех значений;
- $75\ prcntil$ – 75 перцентиль, значение, меньше которого встречаются три четверти всех значений;

- *Skewness* – асимметрия, нулевая для нормального распределения, положительная при смещении моды влево и пологой части справа, отрицательная при смещении моды вправо и пологой части слева;
- *Kurtosis* – эксцесс, нулевой при нормальном распределении, положительный при более остром распределении, отрицательный при более плоском распределении;
- *Geom. mean* – среднее геометрическое, корень степени n из произведения всех n чисел;
- *Coeff. var* – коэффициент вариации, отношение стандартного отклонения к среднему арифметическому в процентах.

	Зрительное восг	Числовой интел	Вербальный инт	Пространственн
N	48	48	48	48
Min	3	2	2	4
Max	12	12	12	11
Sum	327	315	344	357
Mean	6.8125	6.5625	7.166667	7.4375
Std. error	0.3016742	0.4026797	0.3333333	0.2145138
Variance	4.368351	7.783245	5.333333	2.208777
Stand. dev	2.09006	2.789847	2.309401	1.486195
Median	6	6	7	7.5
25 prcntil	6	4	5.25	6
75 prcntil	8	9	8.75	8.75
Skewness	0.9308646	0.283372	0.1882664	0.004913951
Kurtosis	0.8965597	-0.9626182	-0.2568975	-0.2741465
Geom. mean	6.519672	5.941816	6.765467	7.285524
Coeff. var	30.67978	42.51195	32.2242	19.98246

Рис. 25. Окно параметров распределений показателей

Здесь и далее во всех окнах таблиц кнопка *Copy* копирует полученную таблицу в буфер обмена, после чего таблицу можно вставить из буфера в любой документ MS Excel, позволяющий удобно форматировать результаты, добавлять комментарии, выделять и копировать необходимые элементы.

Two-sample tests → ***Two-sample tests (F, t, Mann-Wh, Colm-Sm etc.)*** (***Двухвыборочные тесты***) позволяют получить параметрические и непараметрические критерии оценки достоверности различий между двумя независимыми выборками.

Выделяются два столбца, соответствующие двум показателям, один из которых может иметь тип *Group*, например, *Пол* и *Пространственный интеллект* (см. рис. 18). После выбора соответствующего пункта меню выводится окно с результатами.

Вкладка *t test* (рис. 26) содержит оценку достоверности различий между средними с помощью параметрического *t*-критерия Стьюдента для двух независимых выборок:

- *t test for equal means* – *t*-критерий для оценки равенства средних;
- *N* – объемы выборок;
- *Mean* – средние арифметические значения;
- *95% conf.* – доверительные интервалы;
- *Variance* – дисперсии;
- *Difference between means* – разность между средними;
- *t* – значение *t*-критерия;
- *p (same mean)* – уровень значимости, вероятность равенства средних, подтверждения нулевой гипотезы;
- *Critical t value (p=0.05)* – критическое значение *t*-критерия при уровне значимости 0.05.

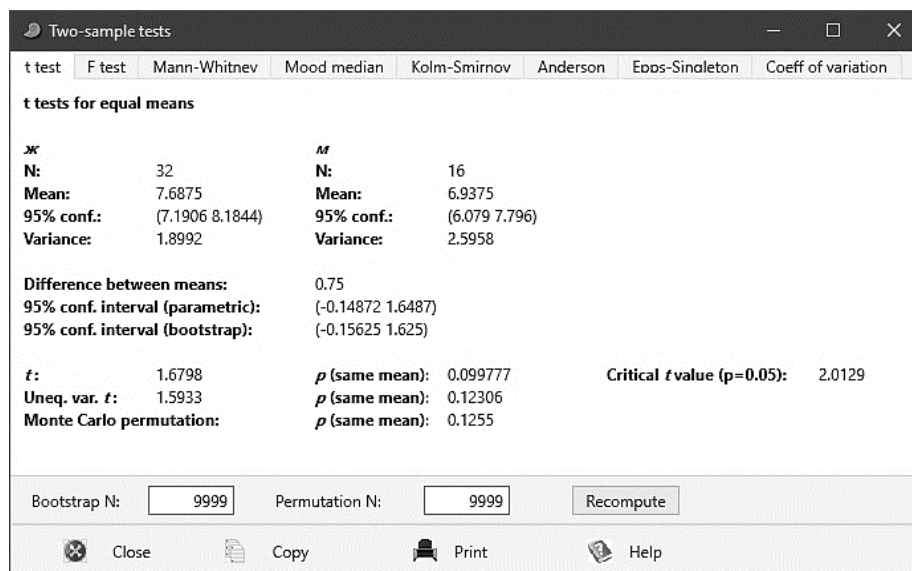


Рис. 26. Окно Two-sample tests/t-test

В указанном примере различия между средними значениями пространственного интеллекта для групп мужчин и женщин недостоверны ($p = 0.099$).

Вкладка *Mann – Whitney* (рис. 27) отображает оценку достоверности различий между медианами двух независимых выборок с помощью непараметрического U-критерия Манна – Уитни;

- *Mean rank* – средние ранги двух выборок;
- *Mann-Whitn U* – значение U-критерия Манна – Уитни;
- *p (same med.)* – уровень значимости, вероятность равенства медиан, подтверждения нулевой гипотезы.

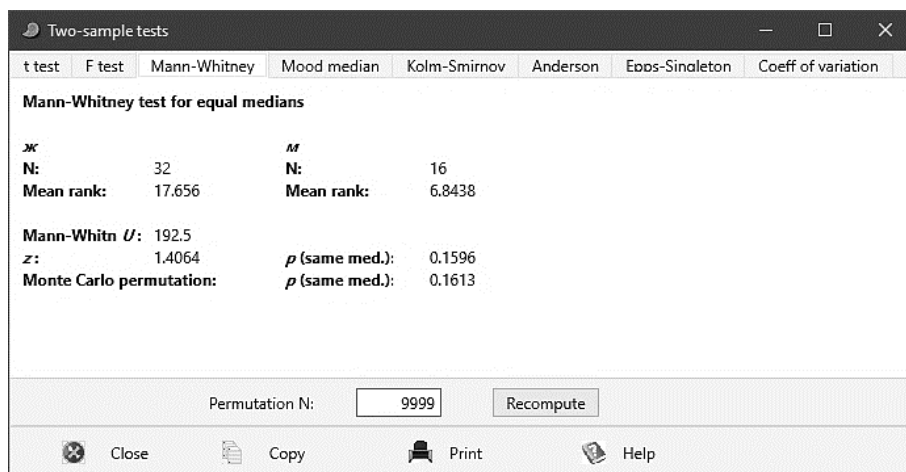


Рис. 27. Окно Two-sample tests/Mann – Whitney

В указанном примере различия между медианами пространственного интеллекта для групп мужчин и женщин недостоверны ($p = 0.159$).

Two-sample tests → **Two-sample paired tests (t, sign, Wilcoxon) (Двухвыборочные парные тесты)** позволяют получить параметрические и непараметрические критерии оценки достоверности различий между двумя зависимыми попарно сопряженными выборками, достоверности сдвигов. Выделяются два столбца, соответствующие выбранному показателю до и после какого-либо события, например усталость до и после тренинга (рис. 28).

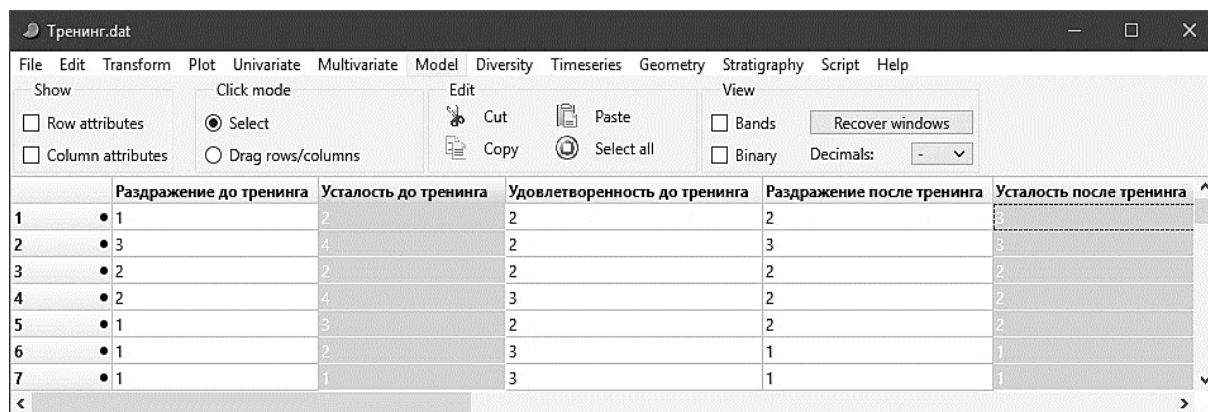


Рис. 28. Выделение столбцов для попарно сопряженных критериев

После выбора соответствующего пункта меню выводится окно с результатами (рис. 29):

- N – количество значений, объем выборки;
- $Mean$ – средние значения;
- $Median$ – медианы;
- t test – t -критерий Стьюдента для зависимых выборок.
- $Mean$ difference – разность средних;

- *95% conf.* – доверительные интервалы;
- *t* – значение *t*-критерия;
- *p (same mean)* – уровень значимости, вероятность равенства средних, подтверждения нулевой гипотезы;
- *Sign test* – критерий знаков;
- *r* – значение критерия знаков;
- *p (same median)* – уровень значимости, вероятность равенства медиан, подтверждения нулевой гипотезы;
- *Wilcoxon test* – критерий Вилкоксона;
- *W* – значение критерия Вилкоксона.

Усталость до тренинга		Усталость после тренинга	
N:	46	Mean:	1.9348
Mean:	2.3261	Median:	2
Median:	2		
t test			
Mean difference:	0.3913	95% conf.:	(0.12957 0.65304)
t:	3.0112	p (same mean):	0.0042576
Exact test not executed (N>22)			
Sign test			
r:	18	p (same median):	0.0043435
Wilcoxon test :			
W:	206.5	p (same median):	0.0059365
Normal appr. z:	2.7513	p (same median):	0.0059
Monte Carlo (n=99999):		p (same median):	0.0061522
Exact:			

Рис. 29. Окно Two-sample paired tests

В указанном примере сдвиг среднего значения и медианы показателя усталости до и после тренинга достоверно отличается от нуля ($p = 0.004$ и $p = 0.006$).

Two-sample tests → **F and t tests with parameters** – оценка достоверности различий между параметрами двух независимых выборок. Представляет собой простой калькулятор (рис. 30), в который вводятся средние значения (*Mean*), дисперсии (*Variance*) и объемы (*N*) 1-й и 2-й выборок (*Sample 1 u Sample 2*), и при щелчке на *Compute (Вычислить)* выводятся значения *F*-критерия Фишера и *t*-критерия Стьюдента с соответствующими уровнями значимости *p*.

ANOVA etc. (several samples) – дисперсионный анализ, позволяет оценить равенство средних значений или медиан одного показателя между несколькими выборками.

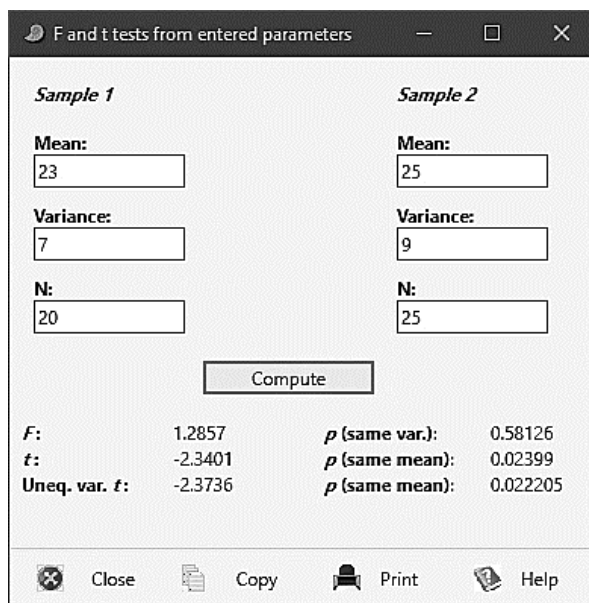


Рис. 30. Критерии Фишера и Стьюдента

ANOVA etc. (several samples) → Several sample test (ANOVA, Kruskal-Wallis) – однофакторный дисперсионный анализ. Выделяются два столбца, один из которых, характеризующий независимую переменную, может быть типа *Group* (см. рис. 18). После выбора соответствующего пункта меню выводится окно с результатами (рис. 31). Если показатель характеризуется нормальным распределением, а выборки – равным объемом, используется *F*-критерий Фишера.

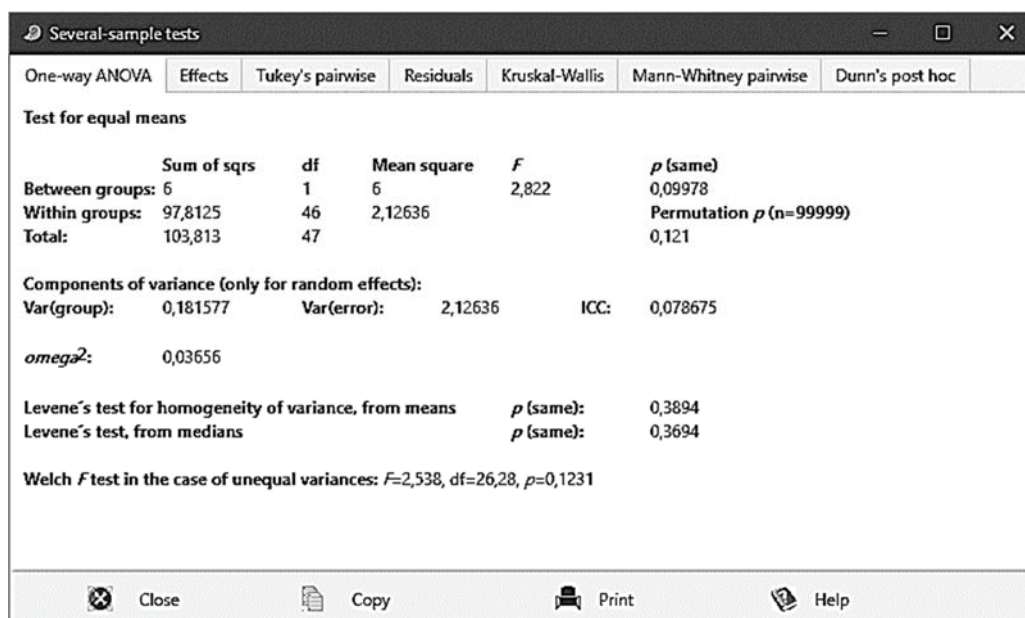


Рис. 31. Окно Several sample test/One-way ANOVA

В данной форме (см. рис. 31):

- *Test for equal means* – тест на равенство средних;
- *Between groups* – между группами;
- *Within group* – внутри групп;
- *Total* – общая;
- *Sum of sqrs* – сумма квадратов;
- *df* – число степеней свободы;
- *Mean square* – средний квадрат;
- *F* – значение *F*-критерия Фишера;
- *p (same)* – уровень значимости, вероятность сходства;
- *omega²* – величина эффекта от 0 до 1;
- *Levene's tests* – критерии однородности дисперсий для нормальных и любых распределений для двух выборок;
- *Welsh F test* – критерий однородности дисперсий для выборок с различными дисперсиями.

В указанном примере пол как независимая переменная не оказывает существенного влияния на пространственный интеллект как зависимую переменную ($p = 0.998$).

Если распределения показателя выборок отличаются от нормального, используется непараметрический критерий Краскала – Уоллиса для оценки сходства медиан нескольких выборок одного показателя (рис. 32).

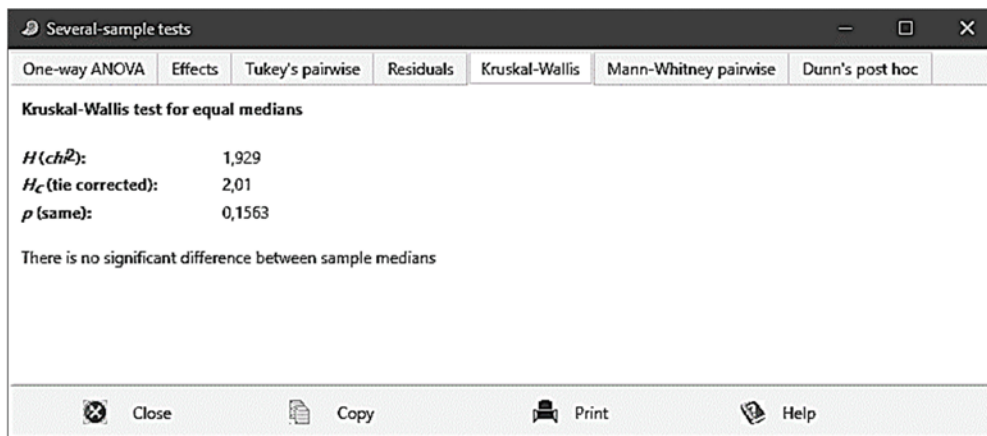


Рис. 32. Окно Several sample test/Kruskal-Wallis

ANOVA etc. (several samples) → *Two-way ANOVA* (Двухфакторный дисперсионный анализ). Выделяются три столбца, два из которых имеют тип *Group* (рис. 33). В результате выводится окно, показанное на рис. 34. Для выбранных групповых переменных и их взаимодействия (*Interactions*) получаем:

- *Sum of sqrs* – сумма квадратов;
- *df* – число степеней свободы;

- *Mean square* – средний квадрат;
- *F* – значение *F*-критерия Фишера;
- *p (same)* – уровень значимости.

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество	Типе
1	9	X	10	5	7	5	27	4	9
2	9	X	6	10	8	7	57	0	10
3	11	M	8	12	5	7	80	5	12
4	9	X	8	8	5	8	117	3	8
5	11	X	8	4	4	8	46	4	11
6	11	M	6	3	5	6	157	6	11
7	9	M	10	8	7	9	50	5	5

Рис. 33. Выделение столбцов для двухфакторного дисперсионного анализа

	Sum of sqrs	df	Mean square	F	p (same)
Класс:	3,52083	1	3,52083	1,703	0,1986
Пол:	6	1	6	2,903	0,09549
Interaction:	0,904443	1	0,904443	0,4376	0,5118
Within:	90,9497	44	2,06704		
Total:	103,813	47			

Рис. 34. Окно Two-way ANOVA

В указанном примере класс и пол как независимые переменные не оказывают существенного влияния на пространственный интеллект как зависимую переменную ни по отдельности ($p = 0.199$ и $p = 0.095$), ни вместе ($p = 0.512$).

Вкладка *Graph of means* показывает зависимость среднего значения выбранного показателя от первого фактора при разных значениях второго и зависимость от второго фактора при разных значениях первого (рис. 35, 36).

Correlation (Корреляция) позволяет вычислять различные коэффициенты корреляции между выбранными показателями. Выделяются столбцы, соответствующие выбранным показателям (см. рис. 20). Выводится окно с таблицей корреляций показателей (рис. 37). Наибольший интерес могут

представлять коэффициенты линейной корреляции Пирсона (*Linear r (Pearson)*) и коэффициент ранговой корреляции Спирмена (*Spearman's rs*).

В случае выбора формата таблицы (*Table format*) *Statistic \ p(uncor)* нижняя треугольная часть таблицы отображает значения коэффициентов корреляции, а верхняя треугольная – уровни значимости. При этом каждому значению коэффициента корреляции соответствует симметрично главной диагонали значение уровня значимости. Если уровень значимости $p < 0,05$, то коэффициент корреляции достоверно отличается от нуля с вероятностью ошибки менее 5 %. В случае выбора формата таблицы (*Table format*) *Statistic* обе части таблицы отображают значения коэффициентов корреляции/ В случае выбора формата таблицы (*Table format*) *p(uncor)* обе части таблицы отображают показатели уровней значимости.

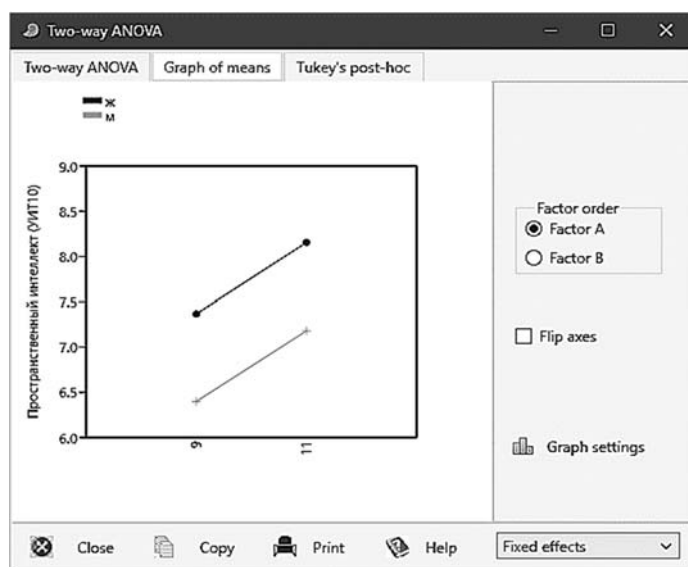


Рис. 35. Окно Two-way ANOVA/Graph of means

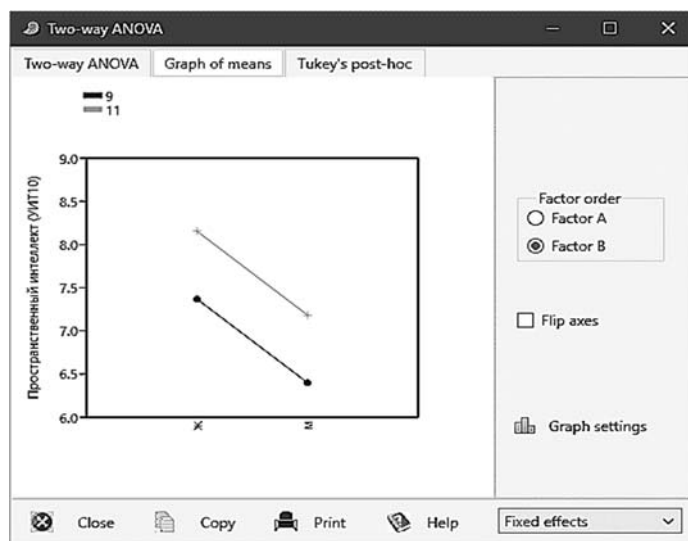


Рис. 36. Окно Two-way ANOVA/Graph of means

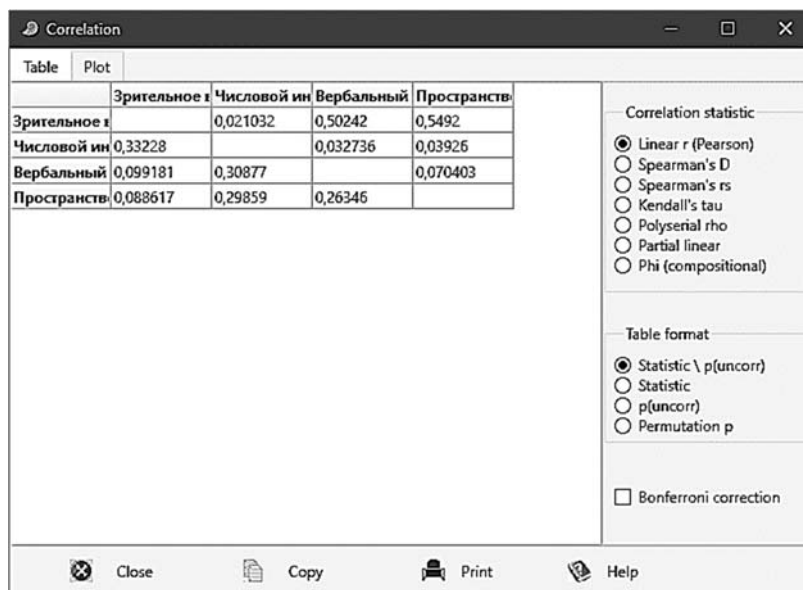


Рис. 37. Окно Correlation

Следует сказать, что при очень малых значениях уровней значимости эти числа могут быть представлены в экспоненциальной форме. Например, число $1,602E-4$ означает $1,602 \times 10^{-4}$ или $0,0001602$. Иными словами, число после «E-» указывает, на сколько знаков влево нужно сместить десятичную запятую или сколько нулей поставить перед имеющимся числом.

В указанном примере коррелируют между собой при уровне значимости $p < 0.05$ показатели зрительного восприятия и числового интеллекта, вербального и числового интеллекта, а также числового и пространственного интеллекта.

Normality tests (Критерии нормальности распределения) позволяют оценить соответствие распределения нормальному. Результаты (рис. 38) выводятся в окно *Test for normal distribution (Оценка нормальности распределения)*:

- *Shapiro-Wilk W* (критерий Шапиро – Уилка) и *Anderson-Darling A* (критерий Андерсона – Дарлинга) – наиболее точные критерии для оценки нормальности распределения;
- p (*normal*) – вероятность нормального распределения.

Принято считать, что если $p < 0.05$, распределение статистически достоверно отличается от нормального. Цветной заливкой выделяются показатели, распределение которых достоверно отличается от нормального.

В указанном примере показатель вербального интеллекта характеризуется нормальным распределением как по критерию Шапиро – Уилка ($p = 0.368$), так и по критерию Андерсона – Дарлинга ($p = 0.186$).

Outlier tests (Критерии выбросов или выскакивающих значений) позволяют выявить значения, существенно выходящие за пределы диапазона нормальных у выбранного показателя. Результаты выводятся в окне

Outlier tests (рис. 39). Во вкладке *Single outlier* (*Одиночный выброс*) выводятся значения *Grubbs test* (*критерия Граббса или Пирсона – Хартли*) и *p(no outlier)*, указывающие на вероятность отсутствия одного выброса. Сообщение *No significant outliers* указывает на отсутствие значимых выбросов.

	Зрительное воспри	Числовой интеллек	Вербальный интелл	Пространственный
N	48	48	48	48
Shapiro-Wilk W	0,9051	0,9511	0,9743	0,9595
p(normal)	0,0009196	0,04417	0,3678	0,0963
Anderson-Darling A	1,669	0,7811	0,5121	0,9775
p(normal)	0,0002389	0,03963	0,1857	0,01277
p(Monte Carlo)	0,0002	0,0398	0,1833	0,0131
Lilliefors L	0,1726	0,1424	0,1121	0,1475
p(normal)	0,0001	0,01595	0,1314	0,0106
p(Monte Carlo)	0,0006	0,0154	0,1383	0,0089
Jarque-Bera JB	7,436	2,554	0,5157	0,2723
p(normal)	0,02428	0,2789	0,7727	0,8727
p(Monte Carlo)	0,0245	0,1423	0,741	0,8603

Рис. 38. Окно Tests for normal distribution

Grubbs test	
G^*	2,397
$p(\text{no outlier})$	0,6603
No significant outliers	
Dixon Q test	
Sample size too large for Dixon test (n>30)	

Рис. 39. Окно Outlier tests/Single outlier

В данном примере выскакивающие значения не выявлены. При наличии выброса сообщается, какое значение (*value*) в какой строке (*row*) является выбросом (рис. 40).

В указанном примере значение 14 в строке 15 является выскакивающим.

Во вкладке *Multiple outliers* (*Множественные выбросы*) показывается, какие 20 % значений в каких строках являются наиболее крайними, и цветной заливкой выделяются те из них, которые являются выбросами (рис. 41).

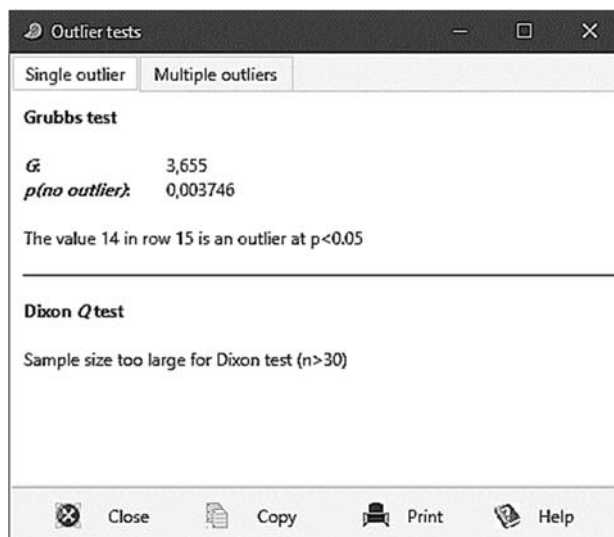


Рис. 40. Окно Outlier tests/Single outlier

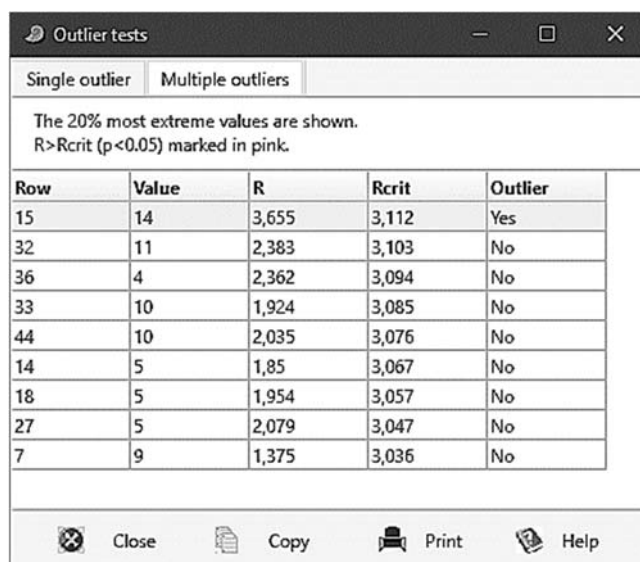


Рис. 41. Окно Outlier tests/Multiple outlier

Contingency table (χ^2 etc.) (Таблица сопряженности (Хи-квадрат и др.)) позволяет вычислить связь между двумя показателями, каждый из которых имеет тип *Group* или *Nominal*. Например, таблица, характеризующая связь между полом (м и ж) и классом (9 и 11), получается (см. рис. 9, 10) с помощью пункта меню *Edit/Rearrange* → *Value pairs to matrix*. Далее полученная таблица выделяется и выбирается пункт меню *Contingency table (χ^2 etc.)*. В полученной таблице (рис. 42) указываются:

- *Row, columns* – количество строк и столбцов;
- *Degrees freedom* – число степеней свободы;
- χ^2 – значение Хи-квадрат;
- *p (no assoc.)* – уровень значимости.

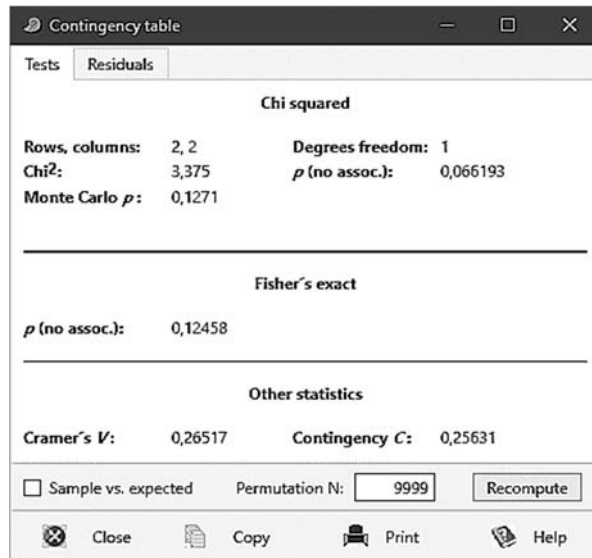


Рис. 42. Окно таблицы сопряженности

В данном примере связь между полом и классом не выявлена, так как $p = 0.066$.

Поле *Sample vs. expected* (*Выборочное значение против ожидаемого*) отмечается, если в первом столбце имеются частоты выборочных значений показателя, а во втором – частоты теоретического распределения или ожидаемые значения.

6. Пункт меню Multivariate (Многомерные методы)

Ordination → *Principal components (PCA)* (*Факторный анализ методом главных компонент*) позволяет сократить количество показателей, сгруппировав коррелирующие показатели. Фактор или главная компонента – это группа коррелирующих показателей. Выделяются столбцы, соответствующие выбранным для анализа показателям. Например, выделяются 11 столбцов показателей модифицированного патохарактерологического опросника МПДО (рис. 43).

	Количество	Гипертимны	Циклоидны	Лабильный	Астено-нерв	Сензитивны	Психастенич	Шизоидный	Эпилептоид	Истероидны	Неустойчив	Неискреннос
1	4	9	2	6	3	2	3	3	4	9	4	2
2	0	10	5	6	3	4	0	3	5	8	4	1
3	5	12	3	5	1	4	2	7	5	7	3	5
4	3	8	9	12	4	9	3	8	8	4	5	3
5	4	11	9	9	8	5	5	7	8	10	4	0
6	6	11	3	3	4	3	1	3	5	8	2	1
7	5	5	6	7	2	1	2	7	1	6	4	0

Рис. 43. Выделение столбцов показателей для факторного анализа

Результаты (рис. 44) выводятся в окно *Principal component analysis*.

PC	Eigenvalue	% variance
1	38.379	46.488
2	14.851	17.989
3	6.83155	8.2751
4	5.18779	6.284
5	4.33143	5.2467
6	3.3319	4.0359
7	2.7139	3.2873
8	2.37399	2.8756
9	1.8896	2.2889
10	1.65936	2.01
11	1.00642	1.2191

Рис. 44. Окно Principal components analysis/Summary

В поле *Matrix* (*Матрица*) следует заменить значение *Variance-Covariance* (*Дисперсия-ковариация*) на *Correlation* (*Корреляция*) и нажать кнопку *Recompute* (*Пересчитать*). Во вкладке *Summary* (*Общие результаты*) представлена информативность полученных главных компонент (рис. 45):

- *PC* – номер фактора или главной компоненты;
- *Eigenvalue* – собственное значение, соответствующее количеству коррелирующих показателей, вошедших в данный фактор (факторы, собственное значение которых менее единицы, можно считать неинформативными);
- *% variance* – процент дисперсии, доля показателей, вошедших в данный фактор.

В данном примере в первый фактор можно включить пять показателей, во второй фактор – один или два показателя, в третий фактор – один показатель.

К сожалению, если полученную таблицу скопировать и вставить в таблицу MS Excel, то значения *% variance* некоторых факторов ошибочно преобразуются MS Excel в формат календарных дат. Если вставить эту таблицу в MS Word или любой другой текстовый редактор, полученные результаты отображаются корректно.

Во вкладке *Scree plot* (*График осей*) представлена кривая процентов дисперсий собственных значений каждого фактора (рис. 46). После того, как кривая перестает существенно падать, а также при значениях компоненты ниже кривой *Broken Stick* (*Сломанная палка – кривая ожидаемых собственных значений при случайной модели*), соответствующие факторы можно считать малоинформативными.

Во вкладке *Loadings* (*Нагрузки*) представлены нагрузки или веса, с которыми соответствующие показатели вошли в каждый фактор (рис. 47).

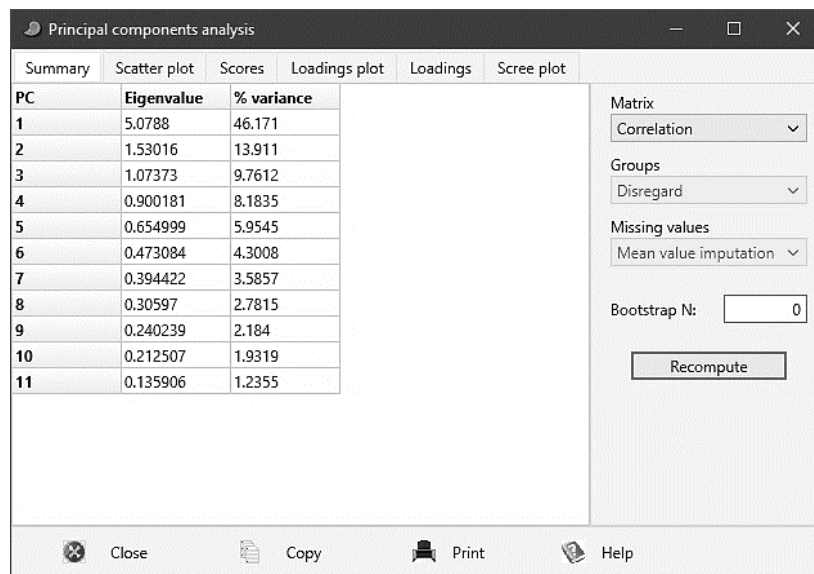


Рис. 45. Окно Principal components analysis/Summary

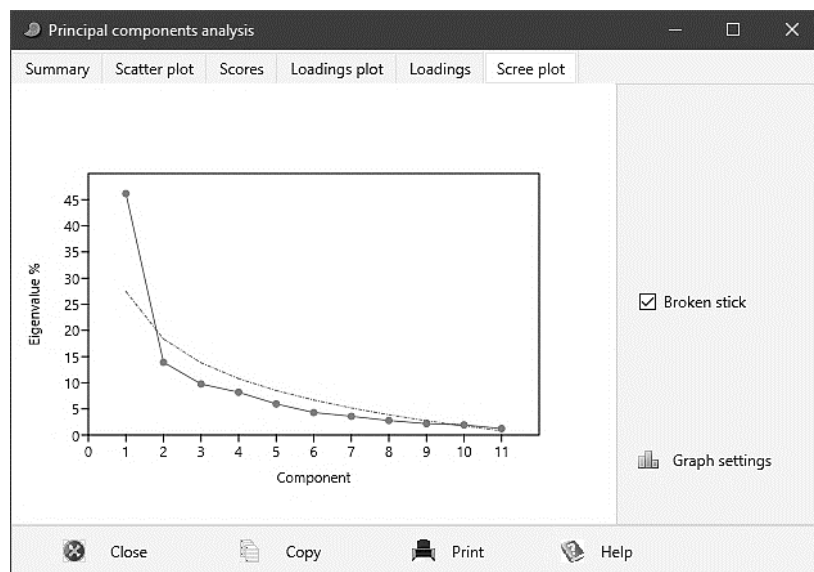


Рис. 46. Окно Principal components analysis/Scree plot

Полученную таблицу целесообразно скопировать и вставить в MS Excel (рис. 48). В формате электронной таблицы Excel для каждого фактора можно вручную выделить те показатели, которые вошли в данный фактор с максимальными по абсолютному значению (без учета знака) факторными нагрузками, что позволяет идентифицировать смысл каждого фактора. В каждый фактор следует включить такое количество показателей, какое содержит собственное значение данного фактора.

Principal components analysis							
Summary	Scatter plot	Scores	Loadings plot	Loadings	Scree plot		
	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
Гипертимный	0.19757	0.65242	-0.056663	-0.12338	0.21454	0.069391	0.067539
Циклоидный	0.39537	-0.069187	-0.046198	-0.037322	-0.015865	-0.27605	0.082026
Лабильный	0.33613	-0.041476	-0.16375	-0.22083	-0.48002	0.43415	0.37026
Астено-невр.	0.36468	-0.2401	0.075141	0.084015	0.089513	-0.25151	-0.55439
Сензитивный	0.345	-0.099573	-0.025224	0.019594	-0.57685	0.046507	-0.2954
Психастенич	0.27916	-0.18254	0.47005	0.0015052	0.38248	0.63393	-0.17261
Шизоидный	0.35875	-0.19014	-0.19459	0.0927	0.20259	-0.35297	0.24425
Эпилептоид	0.35697	-0.085031	0.16585	-0.2525	0.28454	-0.13533	0.42985
Истероидный	0.23916	0.57569	-0.049134	-0.24387	-0.0034909	-0.068261	-0.35481
Неустойчив	0.18731	0.06357	-0.58442	0.62879	0.21933	0.30429	-0.0089007
Неискреннос	0.11538	0.29684	0.57557	0.63243	-0.25713	-0.15185	0.23836

Рис. 47. Окно Principal components analysis/Loadings

	A	B	C	D	E	F	G	H	I	J	K	L
		PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11
1												
2	Гипертимный (МПДО1)	0.19757	0.65242	-0.056663	-0.12338	0.21454	0.069391	0.067539	0.31256	-0.12983	-0.48444	0.33504
3	Циклоидный (МПДО2)	0.39537	-0.069187	-0.046198	-0.037322	-0.015865	-0.27605	0.082026	-0.34844	-0.44168	-0.4141	-0.51402
4	Лабильный (МПДО3)	0.33613	-0.041476	-0.16375	-0.22083	-0.48002	0.43415	0.37026	-0.35668	-0.034418	0.076037	0.34573
5	Астено-невротический (МПДО4)	0.36468	-0.2401	0.075141	0.084015	0.089513	-0.25151	-0.55439	-0.11903	-0.26671	0.10805	0.56662
6	Сензитивный (МПДО5)	0.345	-0.099573	-0.025224	0.019594	-0.57685	0.046507	-0.2954	0.60566	0.13902	-0.091615	-0.23091
7	Психастенический (МПДО6)	0.27916	-0.18254	0.47005	0.0015052	0.38248	0.63393	-0.17261	-0.066902	0.18129	-0.15786	-0.16611
8	Шизоидный (МПДО7)	0.35875	-0.19014	-0.19459	0.0927	0.20259	-0.35297	0.24425	-0.05784	0.71539	-0.18928	0.11156
9	Эпилептоидный (МПДО8)	0.35697	-0.085031	0.16585	-0.2525	0.28454	-0.13533	0.42985	0.39838	-0.23829	0.52023	-0.062467
10	Истероидный (МПДО9)	0.23916	0.57569	-0.049134	-0.24387	-0.0034909	-0.068261	-0.35481	-0.30411	0.26983	0.42951	-0.26299
11	Неустойчивый (МПДО10)	0.18731	0.06357	-0.58442	0.62879	0.21933	0.30429	-0.0089007	0.062495	-0.14782	0.21062	-0.11425
12	Неискренность (МПДО11)	0.11538	0.29684	0.57557	0.63243	-0.25713	-0.15185	0.23836	-0.098803	0.018305	0.088126	0.058911

Рис. 48. Таблица факторных нагрузок показателей на факторы в формате MS Excel

В данном примере в первый фактор с наибольшими нагрузками вошли показатели циклоидного, астеноневротического, шизоидного, эпилептоидного и сензитивного типов акцентуаций личности, характеризующие эмоциональную нестабильность. Второй фактор включил показатели гипертимного и истероидного типов акцентуаций, свидетельствующие об активности и экстравертированности. Наконец, третий фактор оказался связан только с показателем психастенического типа акцентуации, который проявляется в сдержанности, самоконтроле и торможении импульсов.

Во вкладке *Loading plot* (*График нагрузок*) по каждой компоненте представлены факторные нагрузки каждого показателя (рис. 49, 50). Данные графики также позволяют идентифицировать смысл факторов по тому, какие показатели входят в них с максимальными по абсолютному значению факторными нагрузками. Для того чтобы названия показателей помещались в данные графики, они должны быть достаточно короткими в таблице данных.

Во вкладке *Scores* (*Оценки*) представлены оценки каждого испытуемого по каждому фактору, соответствующие координатам этого испытуемого в пространстве факторов (рис. 51).

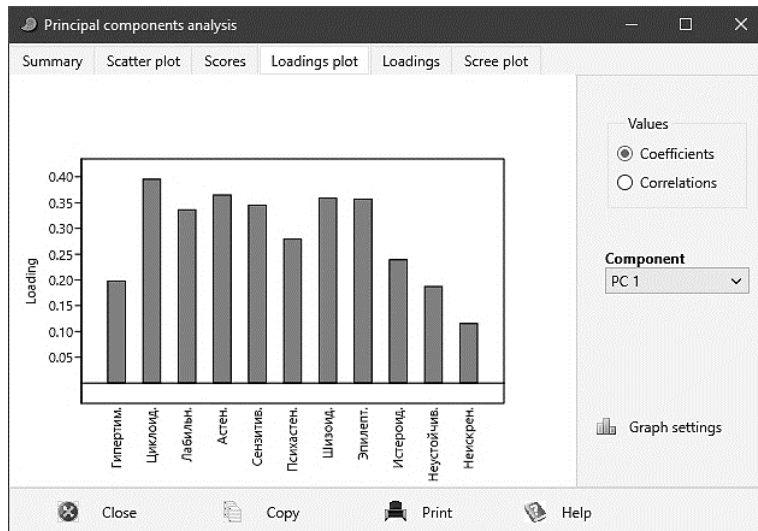


Рис. 49. Окно Principal components analysis/Loadings plot

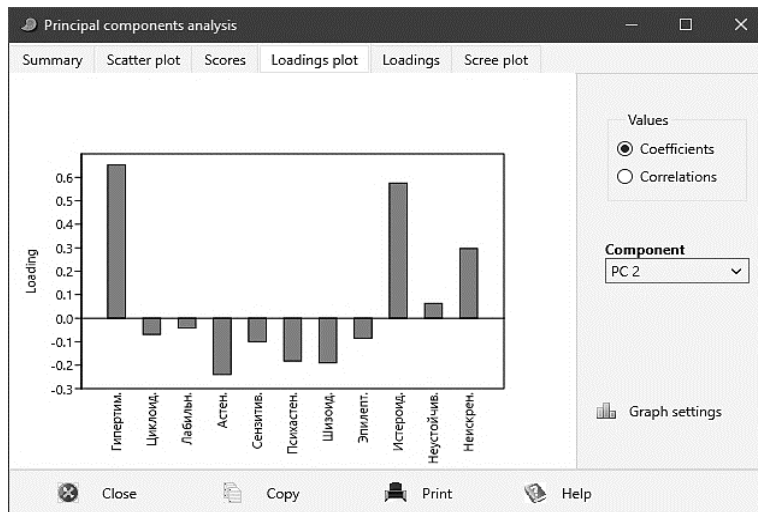


Рис. 50. Окно Principal components analysis/Loadings plot

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
1	-0.08836	1.7424	0.20888	-0.21094	0.11384	0.22985	-0.14626
2	0.22617	1.3121	-1.173	-1.2088	-0.21672	-0.55229	-0.11342
3	0.40962	1.879	0.29251	-0.033557	0.0042841	-0.66664	0.94745
4	3.4	-0.65884	-0.79223	-0.53203	-1.492	-0.024299	1.4791
5	3.6054	0.41399	-0.80151	-2.0168	0.7037	-0.33663	-0.60076
6	-0.41053	1.4182	-0.29886	-1.4416	0.60192	-0.91619	-0.72417
7	-0.3717	-0.26703	-1.5794	-0.80402	0.15997	-0.21898	0.17512
8	-0.75006	1.2953	-0.30408	0.81858	-0.089865	0.58472	0.10101
9	-1.6826	-0.82114	1.493	-0.20383	-0.26052	1.521	0.017661
10	0.69022	1.0682	-0.18155	-1.0007	-0.77576	0.021465	0.93245
11	2.574	2.1308	0.30016	0.66919	0.11369	-0.67161	-0.07461
12	0.58192	-0.90931	-0.44277	-0.046397	-1.1399	1.024	-0.37138
13	0.21019	0.21665	-0.48145	-0.38015	-0.52325	-1.1667	0.69446
14	1.7923	-0.32768	-0.83732	-0.24042	-0.93073	0.22546	-0.33101
15	-1.8466	-0.82323	-0.27834	-0.53997	-0.47049	-0.19548	-0.55478

Рис. 51. Окно Principal components analysis/Scores

Во вкладке *Scatter plot* (двумерная диаграмма рассеивания, или *скатер-грамма*) в виде точек представлены координаты каждого объекта в пространстве двух выбранных факторов (рис. 52). По умолчанию, если мы не выберем иначе, оси *X* (*X axis*) соответствует ось первого фактора (*Component 1*), оси *Y* (*Y axis*) соответствует ось второго фактора (*Component 2*). При желании названия осей (*X label u Y label*) можно вписать в параметрах графика (*Graph settings*). Можно вывести на график имена или номера объектов (рис. 53), отметив поле *Row labels* (*Метки строк*). Наконец, на график можно вывести векторы показателей (рис. 54), отметив поле *Biplot* (*Двойной график*).

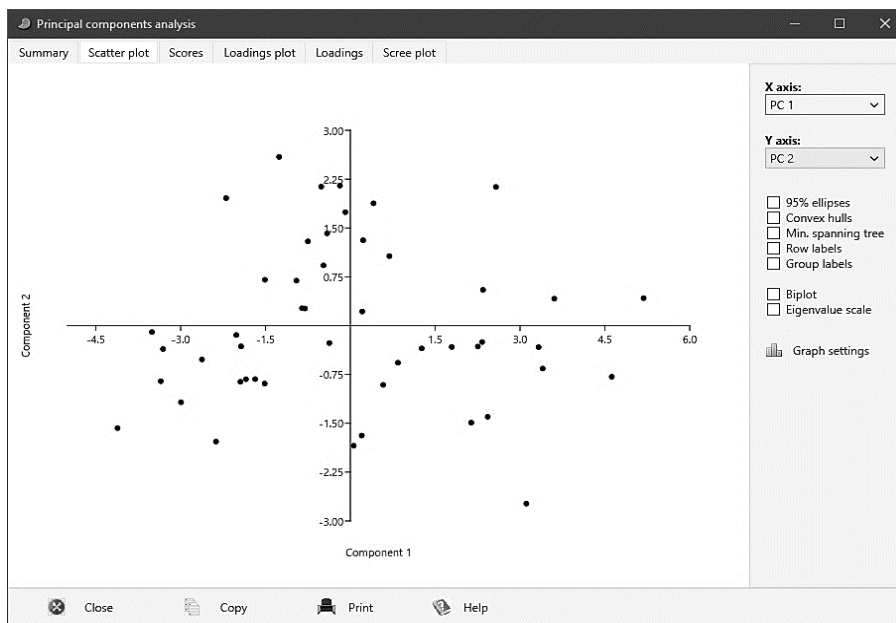


Рис. 52. Окно Principal components analysis/Scatter plot

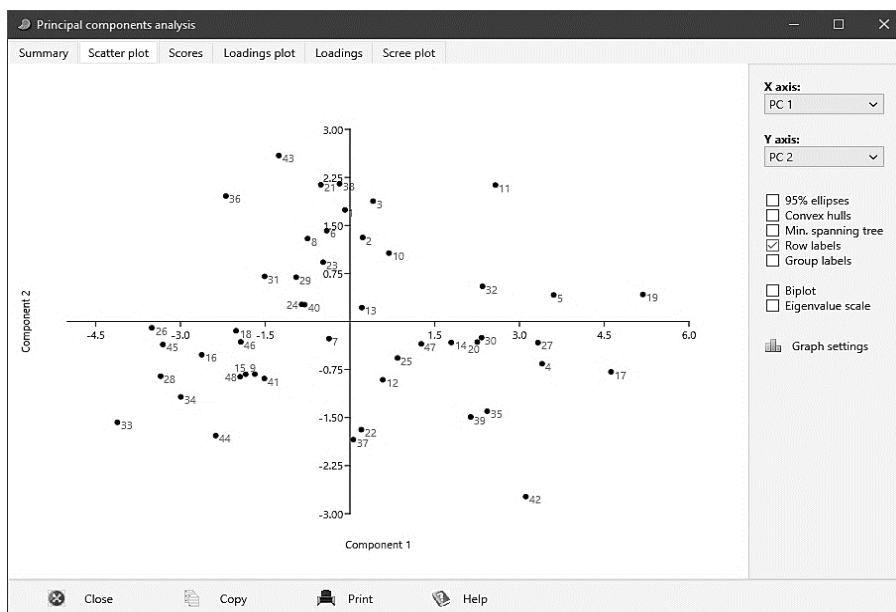


Рис. 53. Окно Principal components analysis/Scatter plot

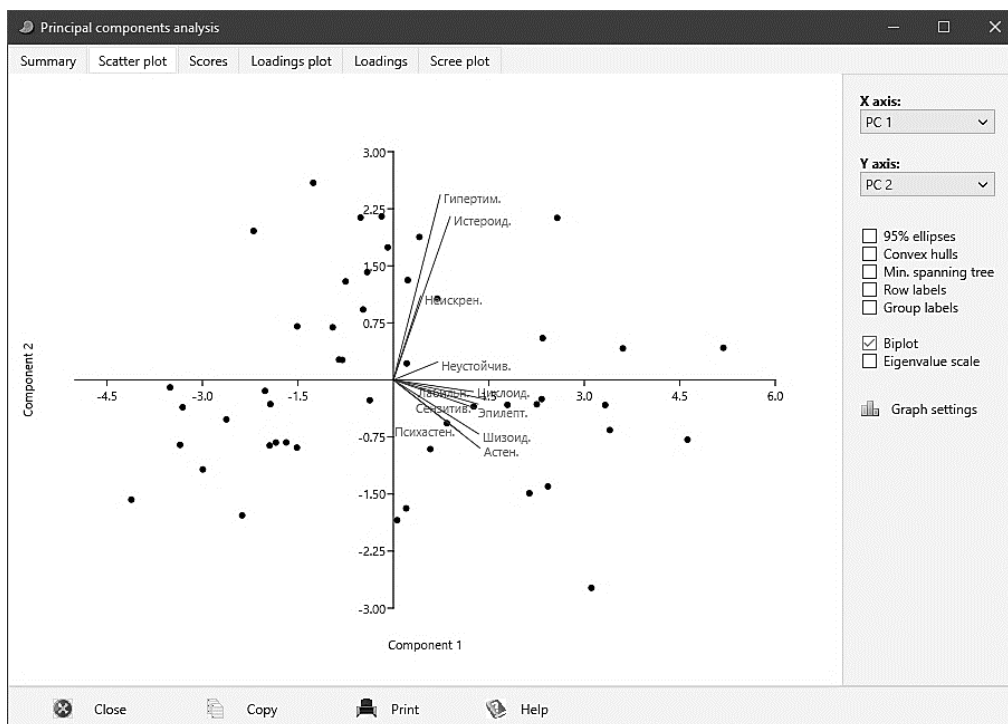


Рис. 54. Окно Principal components analysis/Scatter plot

Discriminant analysis LDA (Дискриминантный анализ) позволяет получить линейное уравнение, максимально дифференцирующее два множества объектов по совокупности показателей $D = A_1X_1 + A_2X_2 + \dots + A_nX_n$. Коэффициенты этого уравнения A_i представляют собой веса показателей X_i , указывающие на степень влияния показателей, т. е. мы можем выявить те показатели, которые вносят максимальный вклад в различие групп. В том случае, если группы статистически достоверно не различаются по каждому показателю отдельно, в некоторых случаях они могут более существенно различаться по совокупности показателей с учетом их весов.

Например, по каждому студенту в группе получен показатель успеваемости – средний балл за сессию по всем экзаменационным оценкам. Если успеваемость студента ниже средней по группе, то по показателю группы успеваемости он получает «1», если успеваемость выше средней по группе, получает «2». Иными словами, студенты разделены на две группы по успеваемости. Дискриминантный анализ позволит ответить на вопрос: в какой мере определяет принадлежность студента к группе с низкой или высокой успеваемостью ряд личностных и интеллектуальных показателей. Выделяем группу успеваемости в качестве зависимой переменной, восемь показателей личностного типологического опросника (чувствительный, дистимический, демонстративный, возбудимый, застревающий, педантичный, замкнутый и гипертимный) и четыре – универсального интеллектуального теста (зрительное восприятие, числовой, вербальный и пространственный интеллект) в качестве независимых переменных (рис. 55).

	Направление	Пол	Гр. усп.	Ср. балл	1 Чувствительный	2 Дистимический	3 Демонстративный	4 Возбудимый	5 Застревающий	6 Педантичный	7 Замкнутый	8 Гипертимный
1	Техническое	м	2	5,00	1	1	0	5	3	6	1	
2	Техническое	м	2	3,83	1	1	1	1	1	6	1	2
3	Техническое	м	1	3,67	2	5	1	2	2	5	4	0
4	Техническое	м	1	3,50	2	3	3	0	3	3	5	1
5	Техническое	м	1	3,50	5	5	2	2	1	1	1	1
6	Техническое	м	2	5,00	1	2	1	3	3	4	2	4
7	Техническое	ж	2	4,67	2	4	2	2	4	2	3	5
8	Техническое	м	2	3,83	2	2	4	1	3	2	0	5
9	Техническое	м	1	3,17	1	0	2	4	2	4	0	4

Рис. 55. Выбор столбцов для дискриминантного анализа

Выбираем соответствующий пункт меню. Во вкладке *Plot (График)* получаем гистограммы частот значений дискриминантной функции для двух групп успеваемости – 1 и 2, которые в определенной мере пересекаются (рис. 56).

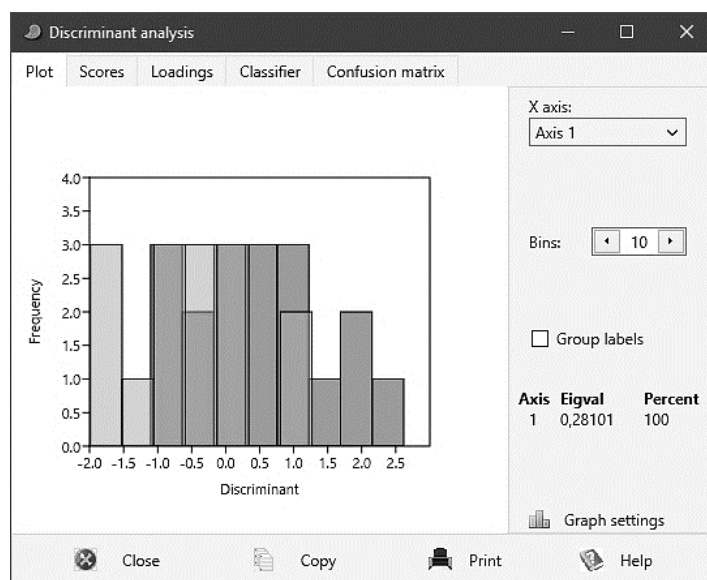


Рис. 56. Окно Discriminant analysis/Plot

Во вкладке *Loadings (Нагрузки)* получаем весовые коэффициенты каждого показателя, характеризующие степень его влияния на группу успеваемости (рис. 57).

В данном примере *Дискриминантная функция* = $0.215 \times \text{Чувствительный} - 0.231 \times \text{Дистимический} - 0.416 \times \text{Демонстративный} + 0.028 \times \text{Возбудимый} - 0.259 \times \text{Застревающий} + 0.088 \times \text{Педантичный} + 0.486 \times \text{Замкнутый} - 0.007 \times \text{Гипертимный} - 0.083 \times \text{Зрительное восприятие} + 0.329 \times \text{Числовой интеллект} - 0.228 \times \text{Вербальный интеллект} + 0.096 \times \text{Пространственный интеллект}$.

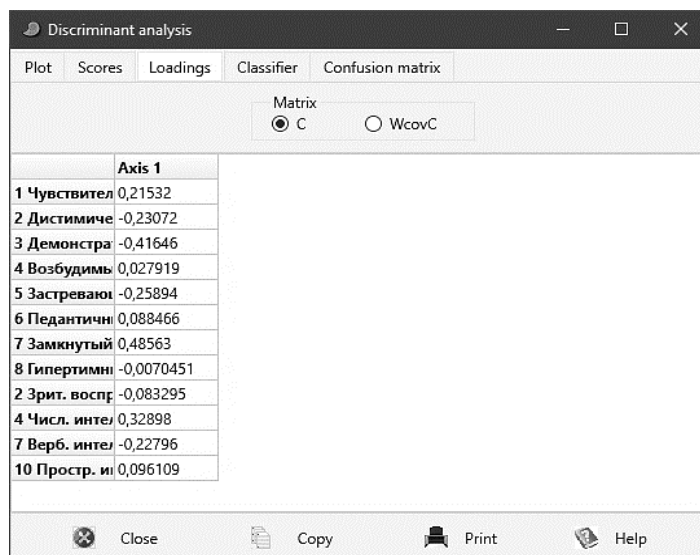


Рис. 57. Окно Discriminant analysis/Loadings

Если значение дискриминантной функции менее нуля, испытуемый с наибольшей вероятностью принадлежит к группе успеваемости 1, в другом случае – к группе успеваемости 2. Таким образом, с наибольшими весами определяют вероятность принадлежности к группе высокой успеваемости (группе 2) показатели высокой выраженности замкнутого типа акцентуации, низкой выраженности демонстративного типа акцентуации и высокой степени выраженности числового интеллекта.

Во вкладке *Scores (Оценки)* получаем значения дискриминантной функции для каждого испытуемого (рис. 58). Чем выше значение дискриминантной функции, тем выше вероятность принадлежности испытуемого к группе 2.

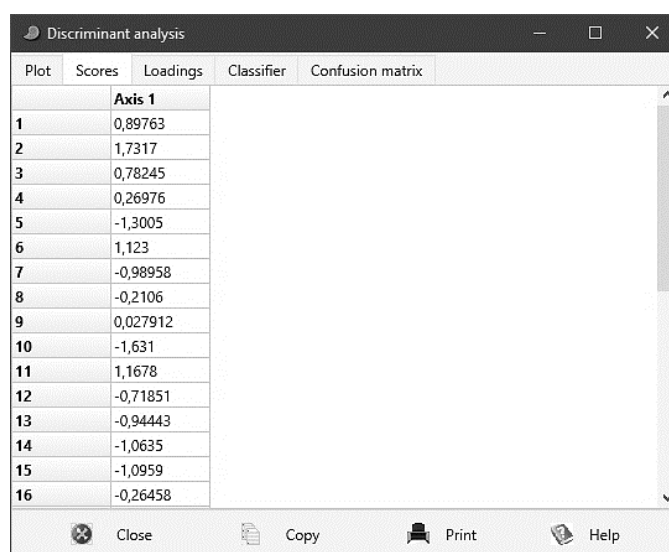


Рис. 58. Окно Discriminant analysis/Scores

Во вкладке *Classifier* (Классификация) для каждого испытуемого получаем информацию, к какой группе он относился (*Given group*) и к какой группе он отнесен по результатам дискриминантного анализа (*Classification*) (рис. 59).

Point	Given group	Classification	Jackknifed
1	2	2	2
2	2	2	2
3	1	2	2
4	1	2	2
5	1	1	2
6	2	2	2
7	2	1	1
8	2	1	1
9	1	2	2
10	1	1	1
11	1	2	2
12	1	1	1
13	2	1	1
14	2	1	1
15	1	1	1
16	1	1	2

Рис. 59. Окно Discriminant analysis/Classifier

Наконец, во вкладке *Confusion matrix* (Матрица смешивания) мы можем узнать (рис. 60), сколько испытуемых из группы 1 по результатам дискриминантного анализа отнесены к группе 2 (ошибочно) и к группе 1 (правильно), а также, сколько испытуемых из группы 2 отнесены к группе 2 (правильно) и к группе 1 (ошибочно). В главной диагонали представлено количество правильно определенных значений. В строках указано, из какой группы взяты испытуемые, в столбцах – к какой группе они отнесены. Показатель *% correctly classified* (Процент правильно классифицированных) указывает процент правильно предсказанных значений.

	2	1	Total
2	12	6	18
1	5	10	15
Total	17	16	33

Rows: Given groups
Columns: Predicted grps

Jackknifed

% correctly classified:
66,67

Рис. 60. Окно Discriminant analysis/Confusion matrix

Clustering → **Classical** (**Кластеризация, иерархический кластерный анализ**) позволяет объединять объекты в группы сходных объектов по множеству признаков. Результаты иерархического кластерного анализа могут быть представлены в виде дендрограммы или дерева. Для выделенных столбцов мы получаем в результате окно *Hierarchical clustering* (*Иерархическая кластеризация*) (рис. 61).

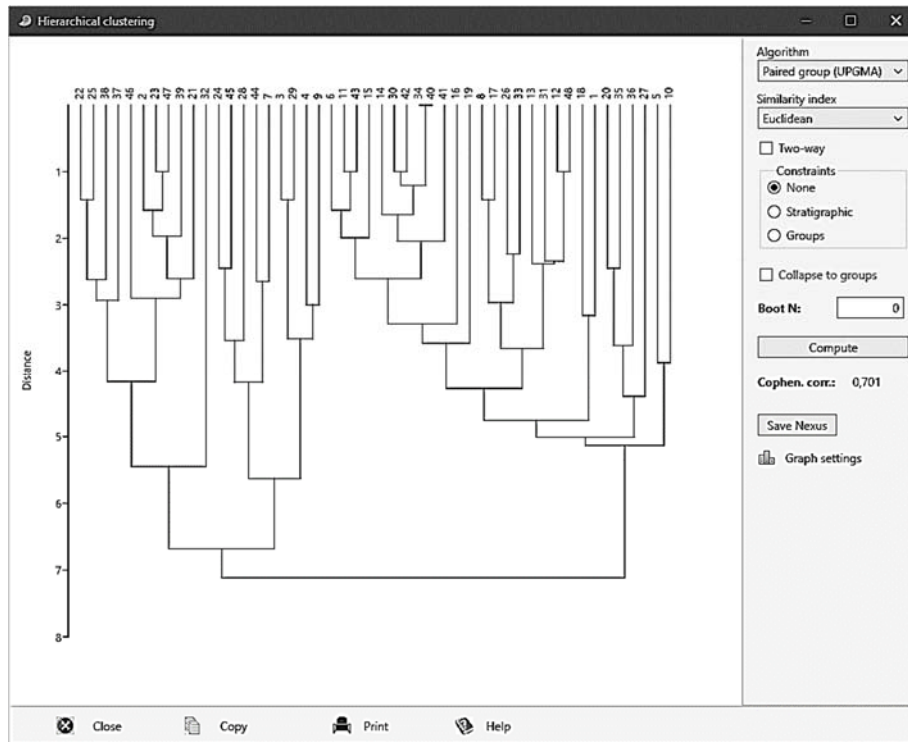


Рис. 61. Окно Hierarchical clustering/ Paired group

Мы можем выбрать один из трех алгоритмов кластеризации (*Algorithm*):

- *Paired group (UPGMA) Unweighted pair-group average (Невзвешенное парно-групповое среднее)* – кластеры объединяются на основе среднего расстояния между всеми членами двух групп;
- *Single linkage (Одиночная связь)* – кластеры объединяются по наименьшему расстоянию между двумя группами;
- *Ward's method (Метод Варда)* – объединение кластеров по критерию минимизация дисперсии внутри группы.

Мы можем выбрать одну из 25-и мер сходства или близости (*Similarity index*). *Euclidean (Эвклидово расстояние)* – расстояние между точками в многомерном эвклидовом пространстве, один из наиболее распространенных и понятных показателей. Другие показатели сходства, такие как манхеттенское расстояние, расстояние Махаланобиса, корреляции и т. д., используются в зависимости от теоретических представлений о метрике пространства и содержании признаков. Различные алгоритмы кластерного анализа

и меры сходства приводят к неодинаковым результатам. После выбора алгоритма и меры сходства следует щелкнуть по кнопке *Compute* для того, чтобы пересчитать результаты (рис. 62).

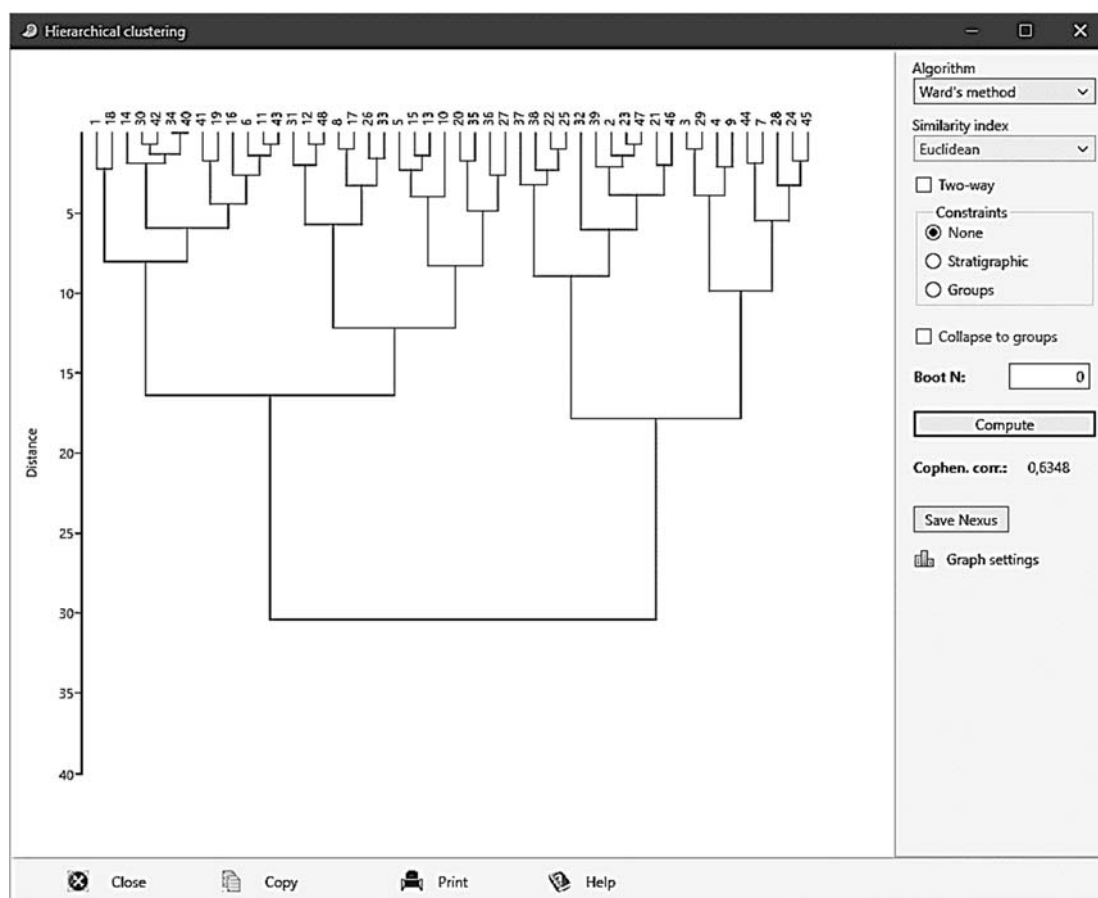


Рис. 62. Окно Hierarchical clustering / Ward's method

7. Пункт меню Model (Модели)

Linear (Линейные) → Bivariate (Двухмерные) позволяет получать линейные уравнения зависимости средних значений одной переменной от значений другой переменной. Выделяется столбец переменной X (например, степень выраженности черт педантичного типа личности) и столбец переменной Y (например, средний балл учебной успеваемости). Результат выводится в окно *Linear fit (Линейная подгонка)*.

Во вкладке *Plot* представлен график зависимости переменной Y от переменной X (рис. 63). В качестве метода регрессионного анализа (*Method*) можно выбрать *Ordinary LS (Ordinary Least Squares)* – *простой метод наименьших квадратов*, минимизирующий квадратичные ошибки Y .

Во вкладке *Statistics* (рис. 64) выводятся параметры уравнения регрессии $Y = aX + b$, где

- *Slope a* – наклон;
- *Intercept b* – пересечение;
- *Std.error* – стандартные ошибки.

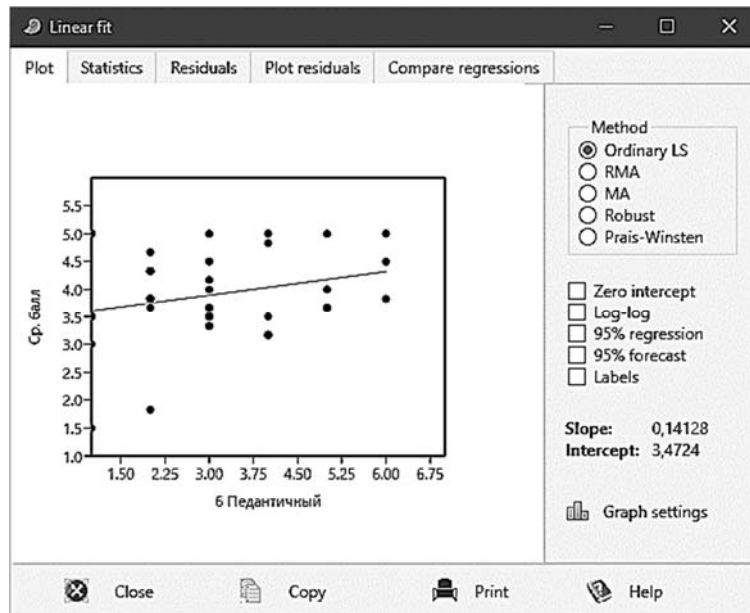


Рис. 63. Окно Linear fit/Plot

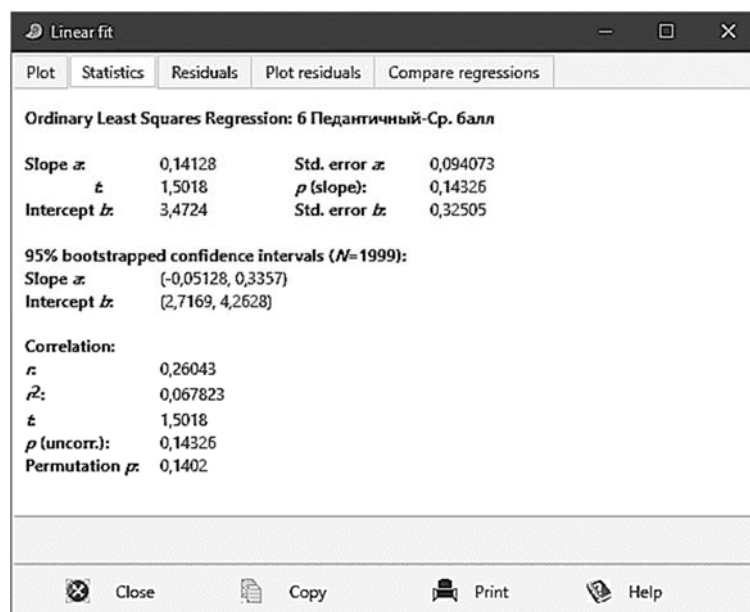


Рис. 64. Окно Linear fit/Statistics

Во вкладке *Residuals* (*Остатки*) указывается расстояние от каждой точки до линии регрессии (рис. 65). *Std. error of estimate* – стандартная ошибка оценки.

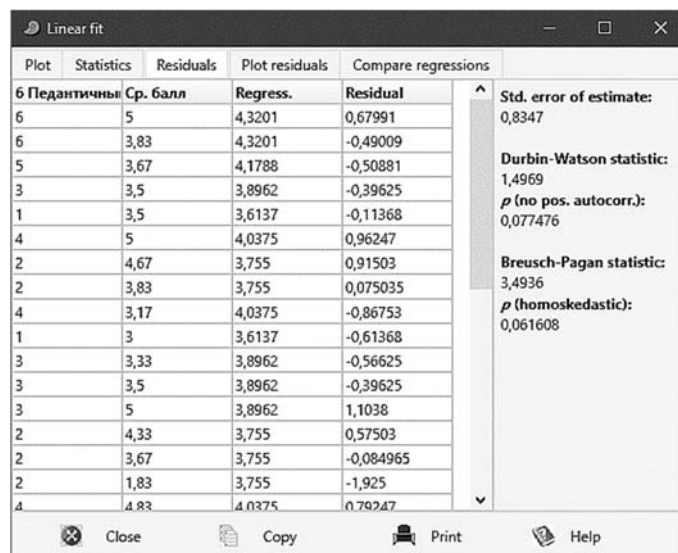


Рис. 65. Окно Linear fit/Residuals

В данном примере формула регрессионной функции, позволяющая прогнозировать учебную успеваемость, будет: $\text{Средний балл} = 3.472 + 0.141 \times \text{Педантичный тип}$.

Multiple (1 dependent n independent) – множественная регрессия (одна зависимая переменная, n независимых переменных) позволяет получить линейное уравнение регрессии, характеризующей зависимость одной зависимой переменной от нескольких независимых: $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$.

Выделяются столбец, соответствующий зависимой переменной, например средний балл учебной успеваемости, и правее него столбцы, соответствующие независимым переменным, например, показатели зрительного восприятия, числового, вербального и пространственного интеллекта (рис. 66).

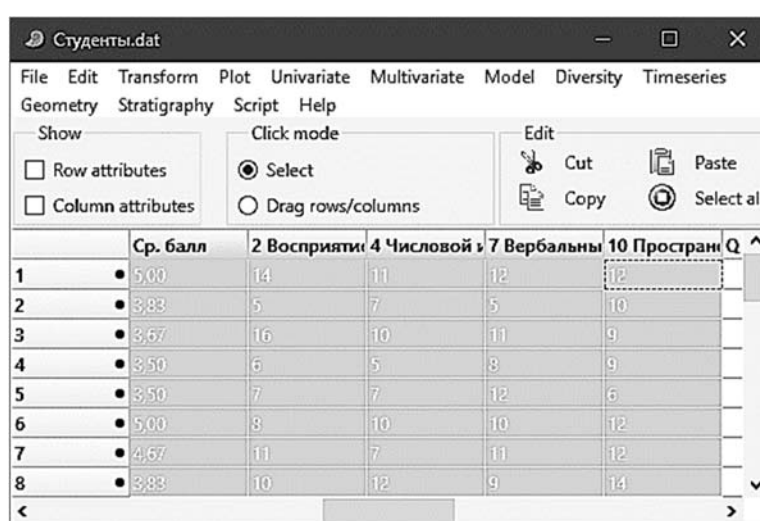


Рис. 66. Выделение столбцов для множественного линейного регрессионного анализа

Результаты выводятся в окно *Multiple linear regression (1 dependent n independent)*.

Во вкладке *Statistics* (рис. 67) указываются:

- *Dependent variable* – зависимая переменная;
- *N* – объем выборки;
- *Multiple R* – множественный коэффициент корреляции;
- *F* – критерий Фишера;
- *df1, df2* – число степеней свободы;
- *p* – уровень значимости.

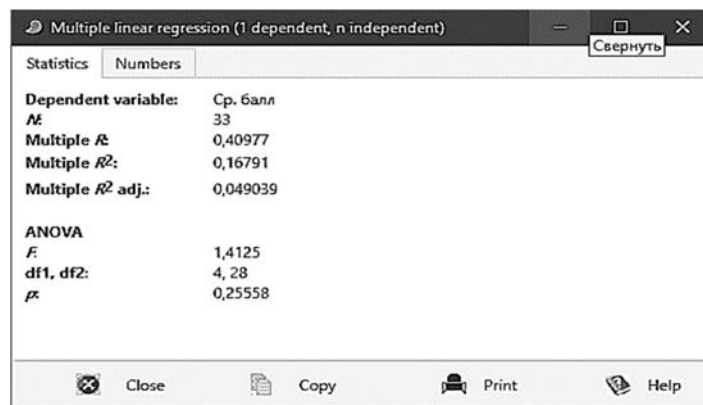


Рис. 67. Окно Multiple linear regression/Statistics

Во вкладке *Numbers* (рис. 68) для константы *Constant* и каждого показателя представлены параметры уравнения регрессии:

- *Coeff.* – коэффициенты;
- *Std.err.* – стандартные ошибки;
- *t* – критерии Стьюдента;
- *p* – уровни значимости;
- *R²* – квадрат коэффициента корреляции, коэффициент детерминации, доля дисперсии, объясняемой моделью.

	Coeff.	Std.err.	t	p	R ²
Constant	2,8629	0,60253	4,7514	5,4716E-05	
Эриг. воспр.	0,028532	0,056402	0,50586	0,61691	0,053783
Числ. интел.	0,019085	0,068208	0,2798	0,78169	0,053251
Вербал. интел.	-0,027546	0,060999	-0,45157	0,65505	0,0069461
Простр. интел.	0,10513	0,064951	1,6186	0,11673	0,15685

Рис. 68. Окно Multiple linear regression/Numbers

В данном примере формула регрессионной функции будет: «Средний балл = $2.863 + 0.028 \times \text{Зрительное восприятие} + 0.019 \times \text{Числовой интеллект} - 0.028 \times \text{Вербальный интеллект} + 0.105 \times \text{Пространственный интеллект}$ ».

Polynomial (Полиномиальная регрессия) позволяет получать уравнения полиномиальной зависимости n степени средних значений одной переменной от значений другой переменной $Y = a_n X^n + \dots + a_2 X^2 + a_1 X + a_0$.

Выделяется столбец переменной X (например, степень выраженности черт дистимического типа личности) и столбец переменной Y (например, средний балл учебной успеваемости). Результат (рис. 69) выводится в окно *Polynomial fit* (Полиномиальная подгонка).

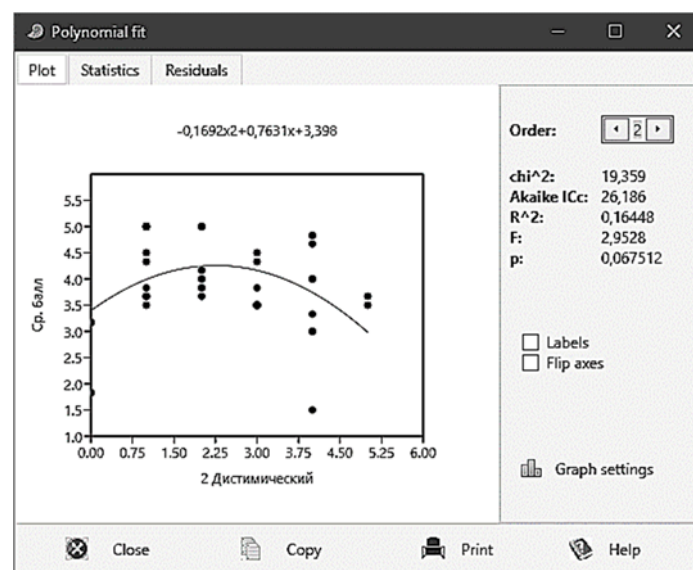


Рис. 69. Окно Polynomial fit

Во вкладке *Plot* выводится уравнение и график полиномиальной зависимости переменной Y от X . В поле *Order* выбирается степень полинома от 1 до 5. В данной форме (см. рис. 69):

- χ^2 (Хи-квадрат) – мера ошибки аппроксимации;
- R^2 – квадрат коэффициента корреляции, коэффициент детерминации, доля дисперсии, объясняемой моделью;
- F – критерий Фишера;
- p – уровень значимости.

В данном примере формула регрессионной функции будет: $\text{Средний балл} = -0.169 \times \text{Дистимический тип}^2 + 0.763 \times \text{Дистимический тип} + 3.398$.

Библиографический список

1. Боровиков, В. Statistica. Искусство анализа данных на компьютере: для профессионалов / В. Боровиков. – Изд. 2-е. – Санкт-Петербург : Питер, 2003. – 688 с. – Текст : непосредственный.
2. Гласс, Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стэнли. – Москва : Прогресс, 1976. – 496 с. – Текст : непосредственный.
3. Кедич, С. И. Практикум по статистическим методам в психологии : учебное пособие / С. И. Кедич. – Санкт-Петербург : ФГБОУ ВО ПГУПС, 2016. – 54 с. – Текст : непосредственный.
4. Лакин, Г. Ф. Биометрия / Г. Ф. Лакин. – Изд. 4-е. – Москва : Высшая школа, 1990. – 350 с. – Текст : непосредственный.
5. Наследов, А. SPSS: Компьютерный анализ данных в психологии и социальных науках. – Санкт-Петербург : Питер, 2005. – 416 с. – Текст : непосредственный.
6. Сидоренко, Е. В. Методы математической обработки в психологии. – ФГБОУ ВО ПГУПС : Речь Санкт-Петербург, 2003. – 350 с. – Текст : непосредственный.

Оглавление

Введение.....	3
1. Запуск программы PAST, интерфейс программы, группы инструментов и пункты меню.....	6
2. Пункт меню File (Файл) и ввод данных.....	10
3. Пункт меню Edit (Редактировать).....	12
4. Пункт меню Plot (Графики).....	14
5. Пункт меню Univariate (Одномерные методы).....	21
6. Пункт меню Multivariate (Многомерные методы).....	33
7. Пункт меню Model (Модели).....	44
Библиографический список.....	49

Учебное издание

СОЛОМИН Игорь Леонидович

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ

Учебное пособие

Редактор и корректор *И. А. Шабранская*

Компьютерная верстка *М. С. Савастеевой*

План 2021 г., № 109

Подписано в печать с оригинал-макета 19.11.2021.
Формат 60×84^{1/16}. Бумага для множ. апп. Печать ризография.
Усл. печ. л. 3,0. Тираж 75 экз.

Заказ 506.

ФГБОУ ВО ПГУПС. 190031, СПб., Московский пр., 9.
Типография ФГБОУ ВО ПГУПС. 190031, СПб., Московский пр., 9.