

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 88.9
С60

Рецензенты:

– доктор психологических наук, профессор кафедры «Прикладная психология» ПГУПС
М. А. Бендюков

– кандидат педагогических наук, доцент кафедры психологии Санкт-Петербургской Академии постдипломного педагогического образования А. Г. Думчева

Соломин И. Л. Математические методы в психологии: учебное пособие / И. Л. Соломин. – СПб.: ФГБОУ ВО ПГУПС. 2021. – 50 с.

ISBN 978-5-7641-1681-5

В пособии описываются базовые математические методы, необходимые к освоению в процессе подготовки и для дальнейшей научной и практической профессиональной деятельности психологов. Представлены теоретические основы планирования эксперимента, измерения и статистической обработки данных. 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».

Оглавление

Введение.....	4
1. Теоретические основы.....	6
1.1. Основы планирования эксперимента.....	6
1.2. Основы измерения и шкалирования	8
1.3. Основы статистической обработки данных.....	11
1.4. Основы математического моделирования	23
2. Основы работы в программе PAST.....	25
2.1. Запуск, интерфейс программы и группы инструментов	25
2.2. Пункт меню File (Файл) и ввод данных	28
2.3. Пункт меню Edit (Редактировать).....	30
2.4. Пункт меню Plot (Графики).....	32
2.5. Пункт меню Univariate (Одномерные методы).....	38
2.6. Пункт меню Multivariate (Многомерные методы).....	48
2.7. Пункт меню Model (Модели)	59
2.8. Другие пункты меню	63
Литература	63

Введение

Развитие и современное состояние психологической науки и практики тесно связано с использованием математических методов. В 19 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 20 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 разработал теорию уровней измерения и шкалирования.

Во второй половине 20 века психологи стали 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 20 – начале 21 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 главным образом 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 свидетельствует о том, что использование математических методов для обработки данных вызывает у многих студентов 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 значительной степени затрудняется тем, что значительное число соответствующих учебников для психологов написаны профессиональными математиками и основаны на сложном аппарате теории вероятностей и математической статистики. Классические учебники [2, 4, 7], основанные на трудоемких вычислениях 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, 6], позволяют эффективно вводить, редактировать и обрабатывать данные, вычислять параметры распределений, считать t-критерии Стьюдента и коэффициенты линейной корреляции Пирсона. Однако использовать Excel для вычисления непараметрических критериев и ранговых корреляций может быть затруднительно, а проводить факторный, кластерный, регрессионный, дискриминантный и другие многомерные методы – невозможно. Наконец, ряд современных и, несомненно, качественных 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». Однако 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/resources/past/>

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

- некоммерческий характер и возможность бесплатного использования,
- маленький размер программы, не превышающий 30 Мб, возможность запуска без установки на компьютер с любого носителя, например, с USB-флэшки, в то числе 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, а также реализовать многообразные формы графического представления результатов.

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

Все это побудило меня 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 проведения других научных исследований.

В разделе «Теоретические основы» представлен материал, позволяющий получить представления о принципах планирования эксперимента, основах измерения и шкалиро-

вания, методах статистической обработки данных и некоторых средствах математического моделирования в психологии. В разделе «Основы работы в программе PAST» содержится описание программы, 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, основ 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. Теоретические основы

1.1. Основы планирования эксперимента

Эксперимент является наиболее сложным методом сбора данных в психологических исследованиях по сравнению с наблюдением, опросом и тестированием. *Эксперимент* – основной общенаучный эмпирический метод исследования, основанный на контролируемом изменении одного или нескольких условий при сохранении постоянными других условий и измерении показателей, характеризующих изучаемое явление. Только эксперимент позволяет с высокой степенью достоверности выявлять причинно-следственные связи, зависимости изучаемых явлений от тех или иных условий. В структуре любого эксперимента, включая психологический, выделяют независимые, зависимые и дополнительные переменные.

Независимые переменные – активно изменяемые экспериментатором условия. В психологии в качестве независимых переменных могут использоваться, например:

- качественные и количественные характеристики стимулов,
- окружающие физические условия,
- физиологические факторы: режим сна, применяемые лекарственные препараты и др.
- факторы деятельности: содержание и сложность задач, нагрузки, помехи, уровень сформированности навыков,
- факторы мотивации: предшествующие успехи или неудачи, награды и наказания, инструкции,

- социальные факторы: присутствие, действия и реакции других людей и т.д.

Зависимые переменные – регистрируемые и измеряемые экспериментатором показатели, которые изменяются под влиянием независимых переменных, например:

- параметры физиологических, двигательных и речевых реакций, сенсорно-перцептивных, мнемических и мыслительных процессов,
- скорость и точность деятельности, длительность, темп и частота различных действий,
- психические состояния,
- субъективные оценки,
- содержание представлений и отношений.

Дополнительные переменные – другие условия, которые также могут оказывать влияние на зависимые переменные, и которые необходимо учитывать, например:

- демографические характеристики испытуемых: пол, возраст, национальность, образование, профессия, семейное положение, род занятий и др.,
- индивидуальные особенности испытуемых: интеллект, черты личности,
- имеющийся опыт, знания и навыки,
- состояние здоровья,
- психическое состояние и установки испытуемых,
- содержание и условия деятельности, параметры заданий.

Планирование эксперимента – это определение схемы эксперимента, позволяющей получить достоверную информацию о зависимости изучаемых явлений от отдельных условий и их сочетаний. План эксперимента (Таблица 1) определяет:

- число независимых переменных (факторов),
- число значений каждой независимой переменной (уровней факторов),
- число повторений измерений зависимой переменной при различных сочетаниях значений независимых переменных.

Таблица 1. Пример двухфакторного плана эксперимента

Уровни независимой переменной А	Уровни независимой переменной В		
	В1	В2	В3
А1	N=5	N=5	N=5
А2	N=5	N=5	N=5

В результате эксперимента матрица сочетаний независимых переменных и их уровней заполняется соответствующим числом измеренных значений показателя зависимой переменной. Если значения зависимой переменной различаются при разных уровнях независимых переменных и их комбинаций, делается вывод о влиянии этих независимых переменных на зависимую.

Значения дополнительных переменных должны быть случайно распределены по различным сочетаниям уровней независимых переменных, иначе их необходимо рассматривать как самостоятельные независимые переменные. Если при различных сочетаниях уровней независимых переменных обнаруживаются различные значения дополнительных переменных, например, разные типы и категории испытуемых, то изменения показателей изучаемого явления вследствие различия в дополнительных переменных могут ошибочно относиться за счет различий в уровне независимых переменных.

Планы эксперимента делятся на простые и сложные, полные и дробные, равномерные и неравномерные. В простых планах значения дополнительных переменных распределяются по уровням независимых переменных в случайном порядке (рандомизированно). В сложных планах значения дополнительных переменных распределяются по уровням независимых переменных на основе предварительного деления на однородные группы по какому-либо признаку (стратифицированно). В полных планах зависимая переменная из-

меряется при всех сочетаниях уровней независимой переменной, в дробных планах зависимую переменную измеряют не при всех сочетаниях уровней. В равномерных планах число повторных измерений при различных сочетаниях уровней независимых переменных одинаково, в неравномерных – различно.

1.2. Основы измерения и шкалирования

В результате эксперимента, наблюдения, опроса, тестирования и других методов сбора данных мы получаем значения различных показателей. **Показатель** – наблюдаемый, измеряемый или вычисляемый признак, позволяющий выявить наличие или величину какого-либо явления. В том числе показатели могут характеризовать различные психические явления: состояния, представления, отношения, способности и т.п. В психологии могут использоваться различные по содержанию показатели:

- физические (время реакции, пороги чувствительности, сила или амплитуда движений и т.п.),
- демографические (пол, возраст, образование, род занятий, семейное положение, национальная принадлежность и т.п.),
- антропометрические (рост, вес, тип телосложения и т.п.),
- физиологические (частота сердечных сокращений, частота дыхания, артериальное давление, электрическое сопротивление кожи, выраженность альфа-ритма ЭЭГ и т.п.),
- биохимические (количество адреналина или сахара в крови или моче и т.п.),
- поведенческие (скорость и точность решения различных задач, частота определенных ответов на вопросы, частота тех или иных действий и т.п.).

Основным способом получения значений показателей является измерение. **Измерение** – способ определения значения показателя с помощью установления его отношения к некоторому образцу или эталону. Для измерения показателей используются шкалы. **Шкала** – система измерения, устанавливающая соответствие между значениями показателей и наличием или величиной признаков. Американский психолог С. Стивенс разработал теорию уровней измерения и шкалирования, описывающую шкалы наименований, порядка, равных интервалов и равных отношений.

1. Шкала наименований, или номинальная шкала:

- характеризует наличие или отсутствие признака,
- устанавливает, равно или не равно одно значение показателя другому значению, но не говорит, какое значение больше или меньше,
- допускает любые преобразования, сохраняющие различие между значениями, в том числе, перемену мест значений.

Примеры показателей в шкале наименований: пол, род занятий, название или тип профессии, название национальности, цвет кожи, тип телосложения, характер заболевания, тип темперамента, переживаемая эмоция, способ реагирования, стратегия поведения, правильность решения задачи, воспринимаемый цвет, воспринимаемая идентичность раздражителей.

2. Шкала порядка, или ранговая шкала:

- характеризует величину признака, но не имеет единицы измерения,
- устанавливает, больше или меньше одно значение показателя относительно другого значения, но не говорит, насколько,
- допускает любые монотонные преобразования, сохраняющие порядок значений, поэтому значения нельзя менять местами.

Примеры показателей в шкале порядка: уровень образования, уровень должности, балльная оценка академической успеваемости, величина привлекательности, сила эмоци-

онального переживания, воспринимаемая яркость или громкость раздражителя, воспринимаемое неравенство раздражителей.

3. Шкала равных интервалов, или интервальная шкала:

- характеризует величину признака, имеет единицу измерения, но не имеет точки отсчета (нулевой точки, характеризующей отсутствие признака),
- устанавливает, на сколько единиц больше или меньше одно значение показателя относительно другого значения, но не говорит, во сколько раз,
- определяет, равна или не равна разность между одной парой значений разности между другой парой,
- допускает любые линейные преобразования, сохраняющие разности между значениями, в том числе, добавление ко всем значениям постоянной величины.

Примеры показателей в шкале равных интервалов: уровень интеллекта в шкале IQ, степень выраженности черт личности в шкале Т-баллов или шкале стенов,

4. Шкала равных отношений, или абсолютная шкала:

- характеризует величину признака, имеет единицу измерения и точку отсчета (нулевую точку, характеризующую отсутствие признака),
- устанавливает, во сколько раз больше или меньше одно значение показателя относительно другого значения,
- определяет, равно или не равно отношение между одной парой значений отношению между другой парой.

Примеры показателей в шкале равных отношений: время реакции, порог чувствительности, сила и амплитуда движений, возраст, рост, вес, частота сердечных сокращений, артериальное давление, количество сахара в крови, скорость решения различных задач, количество ошибок, частота определенных ответов на вопросы, частота тех или иных действий.

Можно установить соответствие между различными уровнями шкалирования, качественными и количественными показателями, неметрическими и метрическими показателями (Таблица 2).

Таблица 2. Виды показателей по типу шкал

По наличию или величине признака	По наличию единицы измерения	Тип шкалы по С. Стивенсу
Качественные показатели (наличие признака)	Неметрические показатели (отсутствие единицы измерения)	Шкала наименований
Количественные показатели (величина признака)		Шкала порядка
	Метрические показатели (наличие единицы измерения)	Шкала равных интервалов
		Шкала равных отношений

Кроме того, показатели мы можем разделить на дискретные и континуальные, прямые и косвенные, объективные и субъективные. Количественные показатели делятся на дискретные и континуальные:

- дискретные (прерывистые) показатели – число явлений, обладающих определенным признаком, выраженное натуральным числом, например, число лиц определенного пола, число определенных ответов на вопрос, число ошибочных двигательных реакций и т.п.,
- континуальные (непрерывные) показатели – число условных единиц определенного признака, выраженное действительным числом, например, возраст, рост, время двигательной реакции и т.п.

По степени непосредственности показатели делятся на прямые и косвенные:

- прямые показатели – основаны на использовании в качестве эталона признака той же природы, например, измерение расстояния в километрах, измерение роста, веса, силы, времени реакции в соответствующих единицах измерения,
- косвенные показатели – основаны на использовании в качестве эталона признака другой природы, например, измерение расстояния в днях пути, измерение скорости мышления в количестве правильно решенных задач в единицу времени, измерение черты личности в количестве определенных ответов на вопросы и т.п.

Наконец, показатели делятся на объективные и субъективные:

- объективные показатели – основаны на использовании в качестве эталонов объектов или единиц измерения, существующих независимо от сознания, например, расстояние, длительность, сила, порог чувствительности, время реакции и т.п.,
- субъективные показатели – основаны на использовании в качестве эталонов представлений, существующих только в сознании, например, субъективные оценки расстояния, длительности, яркости, громкости, цвета, запаха, вкуса, привлекательности, интереса, важности, трудности, уверенности и т.п.

Особой задачей измерения в психологии является **субъективное многомерное шкалирование** – определение, на основе матрицы попарного сходства объектов, минимального числа субъективных признаков этих объектов и значений этих признаков для каждого объекта. Метод многомерного шкалирования основан на представлении субъективных признаков как координатных осей многомерного пространства и представлении объектов как точек в этом пространстве признаков. В таком случае расстояния между точками обратно пропорциональны степени сходства между объектами по множеству признаков. Человеку предъявляется множество объектов, например, понятий или изображений, и предлагается попарно оценивать степень их сходства или различия. На основе этого можно построить пространство минимальной размерности и определить в нем координаты точек так, чтобы матрица попарных расстояний между ними соответствовала матрице попарных различий между объектами.

В процессе любого измерения показателей неизбежны ошибки. **Ошибки измерения** – отклонение измеренных значений показателей от действительных. Можно выделить систематические и случайные ошибки измерения. **Систематические ошибки измерения:**

- не изменяются от измерения к измерению или изменяются предсказуемым образом,
- проявляются в регулярном смещении, сдвиге результатов в какую-то сторону на какую-то величину,
- обусловлены неслучайным изменением условий измерения: состава испытуемых, их состояния, содержания заданий и т.п.

Случайные ошибки измерения:

- изменяются от измерения к измерению непредсказуемым образом,
- проявляются в рассеивании, разбросе результатов в разные стороны на разную величину,
- обусловлены случайными изменениями различных условий измерения.

Основным требованием к измерению является максимально возможная точность, минимизация систематических и случайных ошибок. Точность измерения характеризуется валидностью и надежностью. **Валидность:**

- степень соответствия измеряемых показателей изучаемым психическим явлением, соответствия между тем, что должно измеряться, и тем, что измеряется в действительности,
- отсутствие ошибок, связанных с непригодностью средства измерения.

Выделяются различные виды валидности.

Содержательная валидность – соответствие содержания заданий представлению о реальной учебной или профессиональной деятельности, соответствие содержания во-

просов представлению об измеряемом свойстве. Например, корректурная проба позволяет измерять концентрацию и устойчивость внимания, а скорость ответов и количество ошибок в тесте сходных фигур Кагана являются валидными показателями импульсивности личности.

Конструктивная валидность – соответствие результатов измерения представлению об измеряемом свойстве, его структуре или динамике, соответствие результатов измерения гипотезам о наличии корреляций со сходными показателями и отсутствии корреляций с несходными показателями. Например, ответы на ряд вопросов личностного опросника Айзенка коррелируют между собой, образуя фактор экстраверсии, и не коррелируют с фактором нейротизма, а ответы на определенные вопросы личностного опросника Кеттелла коррелируют между собой и группируются в 16 независимых факторов, представляющих черты личности.

Критериальная валидность – соответствие результатов измерения показателя результатам измерения этого свойства с помощью других методов, которые считаются достоверными, таких как демографические или биографические данные, успешность учебной или трудовой деятельности, врачебный диагноз, экспертные оценки, данные другого теста, валидность которого установлена. Например, показатель интеллекта теста Амтхауэра коррелирует с академической успеваемостью, показатели клинических шкал личностного опросника ММРІ коррелируют с психиатрическим диагнозом.

Надежность:

- степень соответствия значений измеряемого показателя значениям того же показателя при других измерениях в одинаковых условиях,
- отсутствие ошибок измерения, возникающих вследствие различных случайных факторов (ситуации, состояния, заданий).

Выделяют различные виды надежности:

- устойчивость (ретестовая надежность) – корреляция между результатами повторного измерения тем же методом,
- однородность (внутренняя согласованность, надежность параллельных форм или частей) – корреляция между результатами измерения эквивалентными формами теста или расщепленными половинами теста, например, включающими четные и нечетные задания или вопросы.

В качестве наглядной метафоры отношения валидности и надежности измерения можно представить стрельбу по мишеням. Валидность означает, что стрелок попадает именно по той мишени, по которой нужно, а не по другим мишеням. В свою очередь, надежность означает, что при повторной стрельбе по какой-либо мишени разброс небольшой. Иногда с высокой степенью надежности может измеряться не тот показатель, который необходимо измерить.

1.3. Основы статистической обработки данных

Полученные данные или результаты измерения значений показателей нуждаются в статистической обработке. **Обработка данных** – представление значений показателей в форме таблиц, графиков и формул, характеризующих наличие и величину признаков изучаемых явлений в определенных условиях. Целями обработки данных являются:

- упорядочивание, упрощение и повышение наглядности полученных результатов,
- обнаружение новых фактов, выявление скрытых от непосредственного восприятия закономерностей,
- оценка достоверности полученных результатов, обнаружение и исправление ошибок.

Для обработки данных используются одномерные, двумерные и многомерные методы. Одномерные методы включают:

- описательную статистику – анализ частот значений показателя и параметров его распределения (среднее значение, дисперсия и др.) на одном множестве объектов,
- проверку статистических гипотез – оценку достоверности различий между значениями показателя для двух множеств объектов с помощью различных параметрических и непараметрических критериев,
- однофакторный дисперсионный анализ – оценку степени влияния какой-либо независимой переменной на зависимую переменную или изучаемый показатель.

Двумерные методы позволяют выявить статистическую взаимосвязь между двумя показателями с помощью:

- корреляционного анализа – вычисления коэффициентов корреляции, описывающих степень и знак взаимосвязи показателей,
- регрессионного анализа – получения параметров уравнения регрессионной функции, характеризующего форму и величину зависимости одного показателя от другого.

Наконец, многомерные методы дают возможность выявить взаимосвязь между несколькими показателями, структуру этих показателей, сгруппировать их. К ним относятся:

- множественный линейный регрессионный анализ – получение коэффициентов линейного уравнения зависимости одного показателя от нескольких других,
- многофакторный дисперсионный анализ – оценка степени влияния нескольких независимых переменных и их сочетаний на зависимую переменную или изучаемый показатель,
- факторный анализ – объединение большого количества показателей в меньшее количество групп сходных коррелирующих между собой показателей или факторов, позволяющее уменьшить количество переменных,
- дискриминантный анализ – получение коэффициентов линейного уравнения, максимально разделяющего два или несколько множеств объектов по нескольким показателям,
- кластерный анализ – объединение большого количества объектов в несколько сходных между собой по множеству показателей групп или кластеров, позволяющее выделить типы объектов,
- многомерное шкалирование – получение минимального количества и структуры признаков, образующих многомерное пространство, на основе матрицы попарного сходства каждого объекта с каждым.

Объектом любого исследования является определенная категория людей или животных. В исследовании выделяют генеральную совокупность и выборку. **Генеральная совокупность** – множество всех существующих объектов, на изучение которых направлено исследование. Например, это могут быть все люди определенного пола, возраста, национальности, гражданства, рода занятий, профессии, семейного положения, состояния здоровья. Или это могут быть все животные определенного вида, живущие на определенной территории. Однако мы редко имеем возможность исследовать всю генеральную совокупность. **Выборка** – подмножество генеральной совокупности, которое непосредственно исследуется в опыте. Выборка должна быть репрезентативной генеральной совокупности, то есть быть достаточной по объему и соответствовать генеральной совокупности по составу. Как правило, результаты измерения выборочных значений показателей могут в какой-то степени отличаться от значений генеральной совокупности, зависеть от большого количества условий, содержать ошибку, то есть могут рассматриваться как случайные значения. Для количественной оценки случайности – возможности того или иного значения показателя или события – используется мера вероятности. **Вероятность** – отношение числа случаев данного значения к общему числу случаев исключающих друг друга и одинаково возможных значений. Вероятность может принимать значения от 0 до 1, характеризуя значения от совершенно невозможных до полностью достоверных. О ве-

роятности значения можно судить по его частоте. Для анализа частот значений показателя осуществляется группировка данных:

- значения показателя делятся на категории или упорядочиваются по величине и делятся на интервалы,
- по каждой категории или интервалу подсчитывается количество значений, или их частота,
- сгруппированные данные представляются в виде таблицы частот или гистограммы (Рисунок 1).

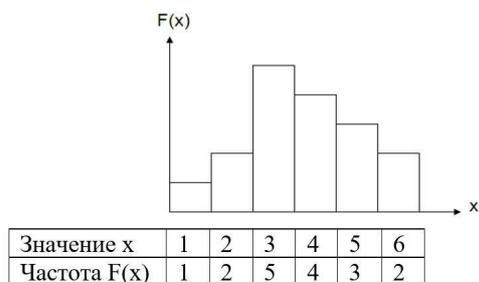


Рисунок 1. Пример таблицы и гистограммы частот

При анализе частот используются следующие термины. **Ранг** – это порядковый номер значения в упорядоченном по величине ряду значений. **Распределение значений** – математическая функция, характеризующая зависимость частоты встречаемости значений показателя от величины этого значения. **Нормальное распределение** – симметричное распределение значений показателя, при котором средние по величине значения встречаются часто, а крайние – редко (Рисунок 2). Нормальное распределение характеризует зависимость измеряемой величины от большого числа условий.

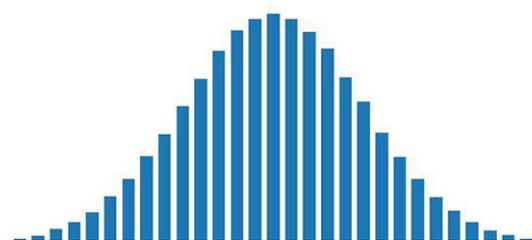


Рисунок 2. Пример нормального распределения.

Для описания распределения используются **параметры распределения** – количественные показатели различных признаков распределения: **положения** (среднее арифметическое, медиана, мода), **рассеивания** (среднеквадратичное отклонение, дисперсия), **асимметрии, эксцесса (выпуклости или остроты)**. Любые два распределения могут различаться (Рисунок 3) средними, дисперсиями, асимметриями, эксцессами.

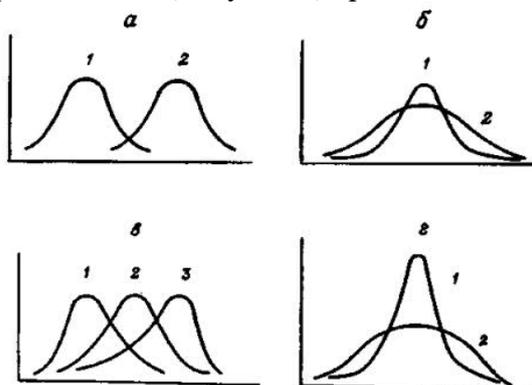


Рисунок 3. Пример различий параметров распределений (а – различие средних, б – различие дисперсий, в – различие асимметрий, г – различие эксцессов).

В зависимости от типа шкалы используются различные параметры распределений (Таблица 3).

Таблица 3. Параметры распределения для различных типов шкал

Шкала	Параметры распределения	
	Положения	Рассеивания
Метрическая (при нормальном распределении)	Среднее арифметическое (M)	Среднеквадратичное отклонение (σ)
Порядка	Медиана (Me)	Размах (D), квартильный размах
Наименований	Мода (Mo)	

В качестве параметров положения используются мода, медиана и среднее арифметическое. **Мода (Mo)** – наиболее часто встречающееся значение в распределении показателя. Может использоваться при любом уровне измерения и типе шкал и при любой форме распределения. **Медиана (Me)** – значение, меньше и больше которого встречается 50% всех значений в распределении показателя. Может использоваться только в случае использования количественных показателей и шкал и при любой форме распределения. **Среднее арифметическое (M)** – отношение суммы значений к количеству значений в распределении показателя:

$$M = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

где x_i – i -значение показателя, n – количество значений. Может использоваться только в случае использования метрических показателей и шкал и только при нормальной форме распределения.

В качестве параметров рассеивания используются размах, квартильный размах и среднеквадратичное отклонение. **Размах (D)** – разность между максимальным и минимальным значением распределения. **Квартильный размах** – разность между значениями первого и третьего квартиля, где первый квартиль – значение, меньше которого встречается четверть всех значений распределения, а третий квартиль – значение, меньше которого встречается три четверти всех значений. Размах и квартильный размах могут использоваться при любых количественных показателях и шкалах и любой форме распределения. **Среднеквадратичное (стандартное) отклонение (σ)** – корень квадратный из отношения суммы квадратов отклонений отдельных значений от среднего арифметического к количеству значений, средняя величина отклонения отдельных значений от среднего арифметического без учета знака:

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{n-1}}$$

Среднеквадратичное отклонение может использоваться только при метрических показателях и нормальном распределении. Квадрат стандартного отклонения σ^2 называется дисперсией.

Значения показателей можно представить в форме **нормированных значений** – отклонения значений от среднего арифметического в долях среднеквадратичного отклонения:

$$Zx_i = \frac{x_i - M_x}{\sigma_x}$$

При нормальном распределении в интервал от $M-\sigma$ до $M+\sigma$ попадает 68% всех значений, в интервал от $M-2\sigma$ до $M+2\sigma$ попадает 95% всех значений, в интервал от $M-3\sigma$ до $M+3\sigma$ попадает 99% всех значений (Рисунок 4).

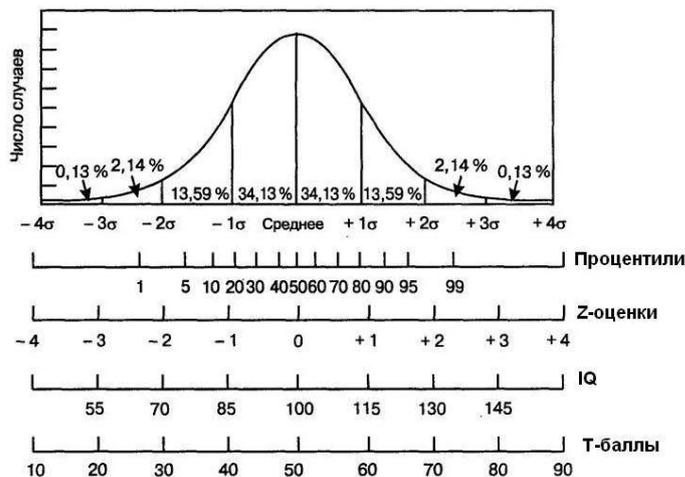


Рисунок 4. Интервалы нормального распределения.

Примечание: **процентили** – значения, меньше которых в распределении встречается соответствующее число процентов значений (например, 10 процентиль – значение, меньше которого имеется 10% всех значений), **z-оценки** – нормированные значения, **IQ** – единицы измерения ряда интеллектуальных тестов, **Т-баллы** – единицы измерения некоторых личностных опросников.

Поскольку значение выборочного среднего значения показателя может отличаться от истинного среднего генеральной совокупности, бывает полезно определить **теоретическую ошибку выборочного среднего (m)** – оценку величины возможного отклонения среднего арифметического значения показателя выборки от среднего арифметического значения показателя генеральной совокупности. Ошибка среднего повышается при увеличении рассеивания и уменьшается при увеличении числа измерений:

$$m = \frac{\sigma}{\sqrt{n}}$$

Доверительный интервал – область значений, в которой с вероятностью не менее заданной находится истинное значение среднего. Так, с вероятностью $p \geq 0.95$ истинное значение среднего лежит в интервале от $M-2m$ до $M+2m$, а с вероятностью $p \geq 0.99$ – от $M-3m$ до $M+3m$.

Иногда средние значения двух совокупностей отличаются друг от друга, но отдельные их значения пересекаются (Рисунок 5). В таком случае необходимо определить, насколько случайно различие между средними. Если доверительные интервалы средних пересекаются, следует признать, что различия между средними значениями случайно.

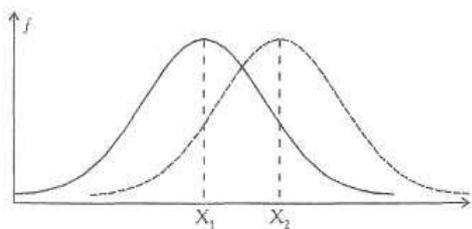


Рисунок 5. Пересечение значений двух распределений

Нулевая гипотеза – предположение о том, что различия между параметрами двух распределений значений являются случайными. **Уровень статистической значимости** – это вероятность сходства, вероятность того, что нулевая гипотеза верна, различия случайны. Считается, что, если уровень значимости $p > 0.05$, вероятность сходства слишком велика, поэтому различия считаются недостоверными. Если же уровень значимости $p < 0.05$ или даже $p < 0.01$, вероятность сходства настолько мала, что ею можно пренебречь, поэтому делается вывод о достоверности различий.

Критерий оценки достоверности различий – количественный показатель отвержения или не отвержения нулевой гипотезы при заданном уровне значимости. Если значение критерия больше некоторого критического значения, которое определяется по таблице для соответствующего уровня значимости с учетом числа степеней свободы или количества значений совокупности, то нулевая гипотеза отвергается. Для оценки достоверности различий между средними или дисперсиями нормально распределенных метрических показателей используются параметрические критерии Стьюдента или Фишера. Для оценки достоверности различий между медианами распределений количественных показателей, представленных в шкалах порядка, равных интервалов или отношений, независимо от формы распределения используются непараметрические критерии Манна-Уитни, Вилкоксона или другие. Для оценки достоверности различий между частотами значений показателей, представленных в любых шкалах независимо от формы распределения используется непараметрический критерий Хи-квадрат и др. (Таблица 4).

Таблица 4. Критерии достоверности различий между распределениями для различных типов шкал

Шкала	Критерии	
	Для оценки разницы между двумя группами	Для оценки сдвигов в одной группе
Метрическая (при нормальном распределении)	Параметрический t-критерий Стьюдента для независимых выборок	Параметрический t-критерий Стьюдента для зависимых выборок
Порядка	Непараметрические: U-критерий Манна-Уитни, критерий Уайта, Q-критерий Розенбаума, λ критерий Колмогорова-Смирнова и др.	Непараметрические: T-критерий Вилкоксона, критерий знаков и др.
Наименований	Непараметрический критерий Хи-квадрат (χ^2) Пирсона	

Параметрический t-критерий Стьюдента для независимых выборок используется для оценки достоверности различий между средними арифметическими значениями нормально распределенного метрического показателя двух различных выборок:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

Параметрический t-критерий Стьюдента для зависимых выборок используется для оценки достоверности различий между средними в одной группе при повторных измерениях нормально распределенного метрического показателя, то есть для оценки достоверности отличия среднего сдвига показателя от нуля, где M_d - среднее значение разности, m_d – теоретическая ошибка среднего значения разности:

$$t = \frac{M_d}{m_d}$$

Непараметрический **критерий Вилкоксона-Манна-Уитни** используется для оценки достоверности различий между медианами двух независимых распределений значений показателей, выраженных в шкалах порядка, равных интервалов или равных отношений при любой форме распределения. Непараметрический **критерий Вилкоксона** используется для оценки достоверности сдвигов медианы при повторном измерении показателей, выраженных в шкалах порядка, равных интервалов или равных отношений при любой форме распределения. Непараметрический **критерий Хи-квадрат** используется для оценки достоверности различий между частотами двух распределений значений показателей, выраженных в любой шкале, включая шкалу наименований, и при любой форме распределения.

Для оценки достоверности различий между дисперсиями нормально распределенных метрических показателей используется параметрический критерий Фишера:

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

На основе критерия Фишера может осуществляться **дисперсионный анализ** – метод определения влияния различных независимых переменных и их сочетаний на значение показателя зависимой переменной. Дисперсионный анализ основан на сравнении дисперсии значений показателя, варьирующих под влиянием определенной независимой переменной (межгрупповая или факторная дисперсия) и дисперсии значений показателя, варьирующих под влиянием остальных контролируемых и не контролируемых переменных (внутригрупповая или остаточная дисперсия). Сила влияния независимой переменной определяется отношением факторной дисперсии к общей дисперсии, которая для равномерных планов равна сумме факторной и остаточной дисперсии. Однофакторный и многофакторный дисперсионный анализ дает информацию о том, влияет ли отдельный фактор или сочетание факторов на изучаемое явление, но не говорит о форме этого влияния. Дисперсионный анализ используется при условии метрических шкал и нормальном распределении показателя зависимой переменной.

Часто необходимо определить, насколько согласовано изменяются значения двух показателей. Поскольку значения величин являются случайными, зависимость между ними является многозначной, каждому значению одной величины может соответствовать несколько значений другой величины. Поэтому графическим выражением статистической зависимости между двумя величинами является рассеянное множество точек, где каждая точка соответствует паре значений величин (Рисунок 6):

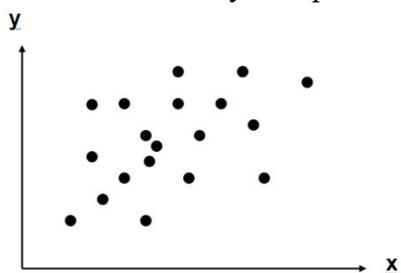


Рисунок 6. Пример графического выражения корреляционной связи между двумя показателями

Корреляция – это статистическая связь между двумя величинами, выраженная в долях единицы. Значение коэффициента корреляции изменяется от -1 до 1. Величина коэффициента корреляции выражает степень связи, а знак – направление связи (Рисунок 7). Положительный коэффициент корреляции указывает на то, что чем больше значение одного показателя, тем больше значение другого показателя. Напротив, при отрицательном коэффициенте корреляции, чем больше значение одного показателя, тем меньше значение другого. При коэффициенте корреляции, приближающемся к 1, мы можем говорить о связи, близкой к функциональной, а не статистической. Коэффициент корреляции, близкий к 0, говорит об отсутствии связи.

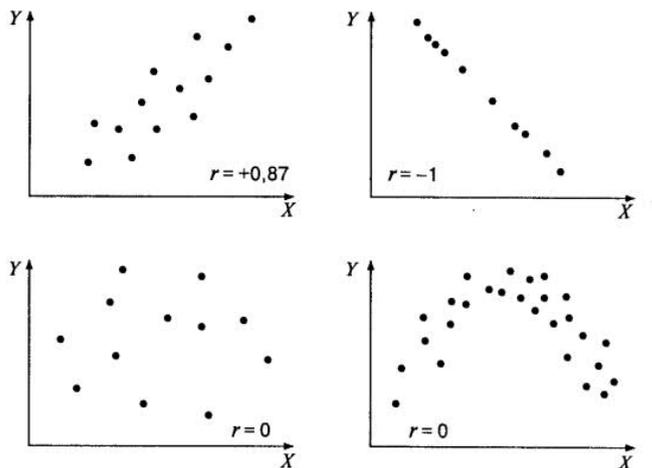


Рисунок 7. Пример различных линейных и нелинейных корреляций

В зависимости от уровня и типа шкал используются различные коэффициенты корреляции (Таблица 5).

Таблица 5. Коэффициенты корреляции для различных типов шкал

Шкала	Коэффициенты корреляции
Метрическая (при нормальном распределении)	Коэффициент линейной корреляции Пирсона r
Порядка	Коэффициент ранговой корреляции Спирмена ρ
Наименований	Коэффициент корреляции ϕ

Коэффициент линейной корреляции Пирсона выражает линейную связь между нормированными значениями двух метрических нормально распределенных показателей, где Zx_i и Zy_i - нормированные значения x_i и y_i :

$$r = \frac{Zx_1Zy_1 + Zx_2Zy_2 + \dots + Zx_nZy_n}{n} = \frac{\sum_{i=1}^n Zx_iZy_i}{n}$$

Коэффициент ранговой корреляции Спирмена выражает линейную и нелинейную связь между рангами двух количественных показателей независимо от формы распределения, где d – разности рангов, n – число пар значений:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)}$$

Коэффициент корреляции ϕ выражает связь между двумя показателями, измеренными с помощью 2-балльной шкалы, где a, b, c и d - частоты комбинаций дихотомических значений 1 и 2 признака в 4-клеточной матрице сочетаний значений признаков:

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Значения 1 показателя	Значения 2 показателя	
	0	1
0	a	b
1	c	d

Если коэффициенты корреляции указывают только на степень и знак связи, то регрессия сообщает также о форме связи. **Регрессия** – статистическая зависимость средних значений одного показателя от значений другого или нескольких других показателей. Выражается в табличной форме, графической форме или в форме алгебраического уравнения. Параметры регрессии – коэффициенты уравнения, характеризующего статистическую зависимость значений одного показателя от значений другого. В статистике разработаны методы аппроксимации регрессии. **Аппроксимация** – преобразование эмпирической регрессии, имеющей, как правило, неправильную форму, в теоретическую регрессию, характеризующуюся более простым математическим уравнением (Рисунки 8-10). Аппроксимация чаще всего проводится при помощи метода наименьших квадратов – подбора теоретического уравнения с минимальным квадратом разностей между соответствующими значениями теоретического уравнения и эмпирической зависимости. Регрессионный анализ используется при использовании метрических шкал и нормальном распределении показателей.

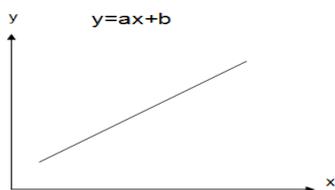


Рисунок 8. Пример линейной функции

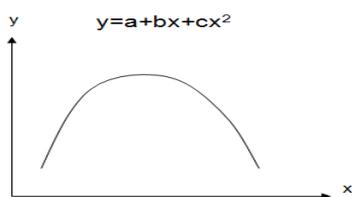


Рисунок 9. Пример полиномиальной функции 2 степени

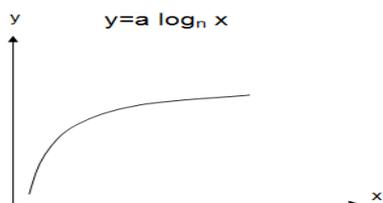


Рисунок 10. Пример логарифмической функции

Зависимость среднего значения одного показателя от значений множества других показателей характеризуется множественной регрессией. Наиболее простой и часто ис-

пользуемой формой множественной регрессии является *множественная линейная регрессия*: $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$

Факторный анализ предназначен для объединения сходных между собой показателей в факторы. **Фактор** – это группа показателей, коррелирующих между собой и не коррелирующих с другими показателями. Если представить показатель как вектор, задающий определенное измерение пространства, то коэффициент корреляции между признаками будет представлен как косинус угла между соответствующими векторами. Тогда коррелирующие показатели будут представлены косоугольными (облическими) векторами, а не коррелирующие – прямоугольными (ортогональными). Если представить фактор как ось пучка векторов, то проекция вектора показателя на ось фактора будет *нагрузкой* этого показателя на фактор, которая тем больше, чем сильнее коррелирует с фактором данный показатель. Если множество коррелирующих показателей разделяется на несколько подмножеств, образуя несколько пучков векторов, то их можно объединить в несколько соответствующих факторов, которые объясняют определенную долю дисперсии данных. В каждый фактор будут входить показатели, имеющие максимальную абсолютную нагрузку на данный фактор.

Графическая иллюстрация принципа факторного анализа представлена на рисунке 11. Рисунок наглядно демонстрирует корреляцию показателей V1, V2 и V3 между собой, что позволяет объединить их в фактор F1, а также слабую корреляцию показателей V1, V2 и V3 с показателями V4 и V5. При этом, например, поскольку показатель V3 в большей степени коррелирует с фактором F1, чем показатель V4, проекция вектора V3 на ось F1 больше, чем проекция вектора V4.

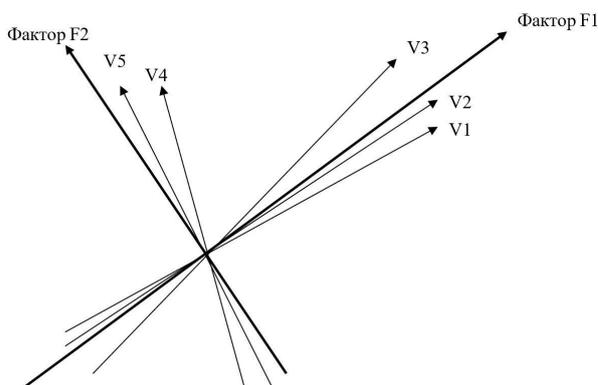


Рисунок 11. Пример графического выражения факторного анализа

Существует большое количество алгоритмов факторного анализа, наиболее распространенным из которых является метод главных компонент, основанный на вычислении собственных значений и собственных векторов матрицы корреляций между признаками. В результате факторного анализа матрица коэффициентов корреляции между показателями $M \times M$ преобразуется в матрицу факторных весов или нагрузок M признаков на L факторов, где $L \ll M$ (Рисунок 12). Факторы упорядочены по уменьшению процента вклада в суммарную дисперсию или количества признаков, входящих в них, то есть первый фактор является наиболее информативным и включает наибольшее число признаков, второй – меньшее число и т. д. В каждый фактор включается столько процентов показателей, сколько процентов дисперсии он описывает.

Матрица коэффициентов корреляции между признаками

Показатели	1	2	M
1					
2					
...					
...					
M					

Матрица факторных нагрузок показателей

Показатели	Факторы	
	F1	F2
1		
2		
...		
...		
M		
% вклада		



Рисунок 12. Табличное выражение факторного анализа

Фактор идентифицируется показателями, входящими в него с максимальными по абсолютной величине факторными весами. Показатели, входящие в один фактор, сильно коррелируют с другими показателями, входящими в этот фактор, и слабо коррелируют с показателями, включенными в другие факторы. Таким образом, факторы независимы между собой, то есть ортогональны, и образуют прямоугольные оси многомерного пространства. Поскольку, чем больше номер фактора, тем меньше показателей в него входит, количество факторов ограничивается номером фактора, в который еще входит более одного показателя.

Факторный анализ используется при использовании метрических шкал и нормальном распределении показателей.

Кластерный анализ предназначен для объединения объектов, сходных по множеству показателей, в группы, или так называемые **кластеры**, с помощью методов автоматической классификации, или «распознавания образов без учителя». Свое название кластерный анализ получил от английского слова «cluster», означающего «гроздь», «пучок», «ветвь», «скопление». Если объекты располагаются в пространстве показателей неоднородно, то их можно сгруппировать. Существует большое количество алгоритмов кластерного анализа, к сожалению, не дающих одинаковые результаты. Простая кластеризация позволяет получить множество непересекающихся классов, исключая друг друга, иерархическая кластеризация дает множество кластеров, включенных друг в друга в виде дерева, объединяя объекты в классы при различных уровнях сходства.

В качестве меры сходства объектов может использоваться расстояние между ними в многомерном пространстве признаков. В математике разработаны различные меры расстояний, наиболее простым из которых является евклидово расстояние. Показатели можно представить как координатные оси многомерного пространства, объекты как точки в этом пространстве. В этом случае

$$D_{ij} = \sqrt{\sum_{k=1}^r (X_{ik} - X_{jk})^2}$$

где D_{ij} – евклидово расстояние между точками i и j в r -мерном пространстве показателей, k – порядковый номер измерения или показателя, X_{ik} и X_{jk} – координаты точек i и j по k -измерению или значения k -показателя i и j объектов. Например (Рисунок 13), для двумерного пространства расстояние между точками 1 и 2 равно корню квадратному из суммы квадратов разностей по каждой из двух координат (поскольку квадрат гипотенузы равен сумме квадратов катетов). Чем больше сходство между объектами, тем меньше расстояние между соответствующими точками.

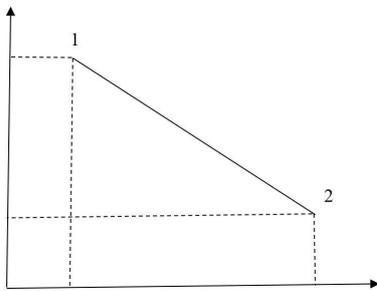


Рисунок 13. Графическое выражение расстояния между двумя объектами

Результаты кластерного анализа могут быть наглядно представлены в виде **дендрограммы**, или дерева (Рисунок 14), где различные ветви дерева соответствуют различным объектам и кластерам. Последовательное соединение ветвей выражает объединение соответствующих объектов или кластеров. Чем ближе к вершине соединяются между собой ветви, тем более сходны между собой соответствующие этим ветвям кластеры. При этом не исключено, что кластер может состоять из одного-единственного объекта, если он слишком отличается от всех остальных. Перемещаясь от основания дерева к вершине, можно последовательно выделять группы все более сходных объектов. При этом количество необходимых для интерпретации кластеров определяется исходя из некоторого порогового критерия.

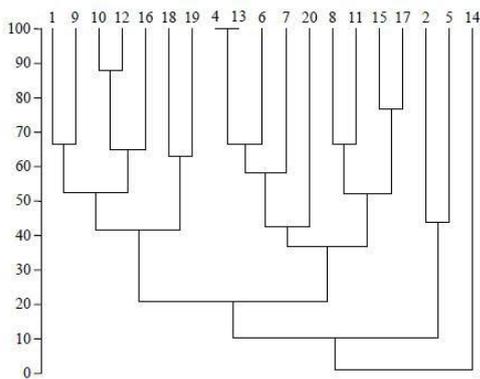
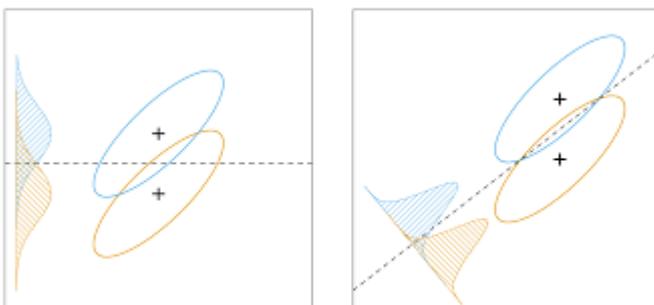


Рисунок 14. Пример результатов кластерного анализа в виде дендрограммы

Дискриминантный анализ представляет метод классификации или «распознавания образов с учителем» на основе получения параметров линейной функции, максимально дифференцирующей два или несколько множеств объектов в многомерном пространстве показателей. Если по каждому отдельно взятому показателю два множества объектов могут статистически достоверно не различаться, то по значению дискриминантной функции суммы показателей с определенными коэффициентами $D = a_1x_1 + b_2x_2 + \dots + b_nx_n$ различие может быть достоверным (Рисунок 15). Дискриминантный анализ используется при использовании метрических шкал, нормальном распределении показателей и их независимости друг от друга.



1.4. Основы математического моделирования

Математическое моделирование в психологии – это описание психических явлений с помощью математических средств. Для моделирования различных форм отношений между психическими явлениями используются различные математические средства (Таблица 6).

Таблица 6. Математические средства, используемые для моделирования психических явлений

Исследуемые явления	Математические средства
Состав и элементы психических явлений	Теория множеств
Структура психических явлений, связи между элементами	Теория матриц и графов, топология, векторное исчисление
Функции психических явлений, зависимость от физических и физиологических явлений	Алгебра и математический анализ
Рассуждения и умозаключения	Математическая логика
Случайные явления	Теория вероятностей
Процесс передачи информации	Теория информации
Процесс ответного реагирования	Теория автоматического управления
Последовательность действий	Теория алгоритмов

В качестве иллюстраций рассмотрим отдельные математические теории, созданные в 20 веке и используемые для моделирования некоторых психических явлений.

Теория информации – математическая теория, описывающая передачу информации по каналам связи. Под информацией имеется в виду сообщение об определенном событии, которым может быть любое значение признака какого-либо явления. Можно представить объект как источник информации, сигнал как носитель информации, субъект как приемник информации. Информация об объекте используется субъектом для управления объектом и регуляции своей деятельности. Психическая деятельность субъекта представляет собой прием, хранение и переработку информации. Следовательно, психические явления зависят от характеристик информации. Информация может характеризоваться содержанием, формой и количеством. Содержание информации определяется значением и значимостью признаков объекта, форма – способом кодирования сигналов, а количество – длиной алфавита сигналов. Статистическая теория информации абстрагируется от содержания и формы сигналов, а описывает лишь количественные характеристики информации. Источник информации может находиться в одном из нескольких состояний. Поскольку существует разнообразие возможных состояний, существует неопределенность возможных событий. Энтропия – это мера неопределенности: $H = -\log_2 N$, где N – число независимых и равновероятных событий. При получении сообщения об определенном событии неопределенность этого события уменьшается. Количество информации в сигнале равно разности между энтропией до и после сигнала: $I = H - H_0$. Для независимых и равновероятных сигналов при отсутствии шумов количество информации в сигнале $I = \log_2 N$, где N – число сигналов в алфавите. Единицей количества информации в сигнале является двоичная единица или бит. Один бит информации содержится в сообщении об одном событии из двух возможных. Например, в сообщении об одном событии из 4-х возможных содержится 2 бита информации, а из 8-ми возможных – 3 бита.

Для адекватного использования теории информации необходимо при равной вероятности сигналов определить длину их алфавита, при разной вероятности сигналов – распределение их абсолютных вероятностей, при зависимости одних сигналов от других – распределение их условных вероятностей. Однако это определение возможно лишь при

большом числе повторений сигналов. Другими условиями применимости теории информации являются следующие допущения:

- состояния источника дискретны,
- состояния источника изменяются только по одному признаку,
- смысл и ценность сигналов примерно одинаковы,
- субъективная длина алфавита и вероятность сигналов соответствуют объективным,
- отсутствует временная неопределенность сигналов, известно время их поступления,
- характеристики переработки информации неизменны во времени, отсутствует упрямление, утомление и т. п.

Поэтому использование теории информации ограничено описанием процесса передачи информации по человеческому каналу, который характеризуется объемом в битах и скоростью в битах в секунду. Максимальная скорость передачи информации по каналу связи называется пропускной способностью канала. Эмпирическими показателями пропускной способности человека могут быть:

- минимальное время сигнала, необходимое для его опознания,
- время двигательного или речевого ответа на сигнал,
- число сигналов, принимаемых в единицу времени.

Информационный поток должен соответствовать пропускной способности человека, иначе возникает избыток или недостаток информации. Однако информационный поток характеризуется не только информационными параметрами (длиной алфавита, абсолютными и условными вероятностями сигналов), но и энергетическими параметрами (силой сигнала и шума), и временными параметрами (длительностью сигналов и интервалов между сигналами), и пространственными параметрами (числом каналов)

Теория автоматического управления – это математическая теория, описывающая зависимость параметров выходного ответа системы управления от параметров входного сигнала, изменяющегося во времени (Рисунок 16).



Рисунок 16. Система управления

Для каждой системы управления существует единственная функция, которая называется передаточной и определяет зависимость y от x . Различные характеристики и коэффициенты уравнения передаточной функции определяют параметры переходного процесса – изменения выхода системы в ответ на ступенчатое изменение входного сигнала. В зависимости от характеристик передаточной функции мы можем определить задержку и усиление переходного процесса, характер этого процесса (не колебательный, колебательный затухающий или колебательный незатухающий), периодичность колебаний и скорость затухания. Линейные системы управления характеризуются постоянным периодом и скоростью затухания. Позиционные системы управления характеризуются ответом на непосредственное значение сигнала, в то время как дифференцирующие системы реагируют на изменение значений входного сигнала, а интегрирующие системы – на накопленное значение входного сигнала. Человек как звено системы управления также может характеризоваться передаточной функцией. При этом выход человека характеризуется нестабильностью задержки и усиления, фильтрацией высокочастотных составляющих сигнала, нелинейностью затухания переходного процесса, стохастичностью, дискретностью, экстраполяцией, адаптивностью и т. д.

Теория алгоритмов – это математическая теория, описывающая последовательность выполнения операций в зависимости от определенных условий. Алгоритм – это последовательность операторов в определенных логических условиях, операторы – это действия, логические условия – это критерии их выполнения. Операторы могут графически выражаться прямоугольниками, логические условия – ромбами, а алгоритм – стрелками между соответствующими фигурами. Например, после выполнения оператора 1 при выполнении условия выполняется оператор 2, а при невыполнении условия вновь выполняется оператор 1 (Рисунок 17).

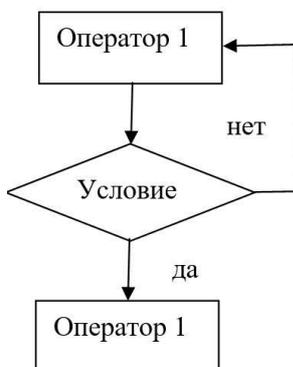


Рисунок 17. Пример простейшего алгоритма

Теория алгоритмов позволяет описать деятельность человека. По количеству операторов и логических условий определяется сложность, стереотипность и другие характеристики деятельности. Однако алгоритмическое описание деятельности основывается лишь на использовании уже известных способов, в то время как эвристическое описание включает поиск еще не известных способов. Другими условиями применимости теории алгоритмов являются:

- последовательность действий, одновременность выполнения операторов,
- детерминированность действий, однозначная зависимость операторов от логических условий.

2. Основы работы в программе PAST

2.1. Запуск, интерфейс программы и группы инструментов

Для работы программы PAST требуется операционная система Windows 7, 8 или 10 версии. Единственный исполнимый exe-файл программы просто копируется в любую папку на диск компьютера или внешний накопитель и запускается. Имеются версии программы PAST для 64-битных и 32-битных версий Windows. Первая предназначена только для работы под 64-разрядной версией Windows, которая наиболее часто устанавливается на современных мощных компьютерах. Вторая 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 (Рисунок 18). 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 разделителя

телей целых и дробных значений могут использоваться как десятичные запятые, так и точки.

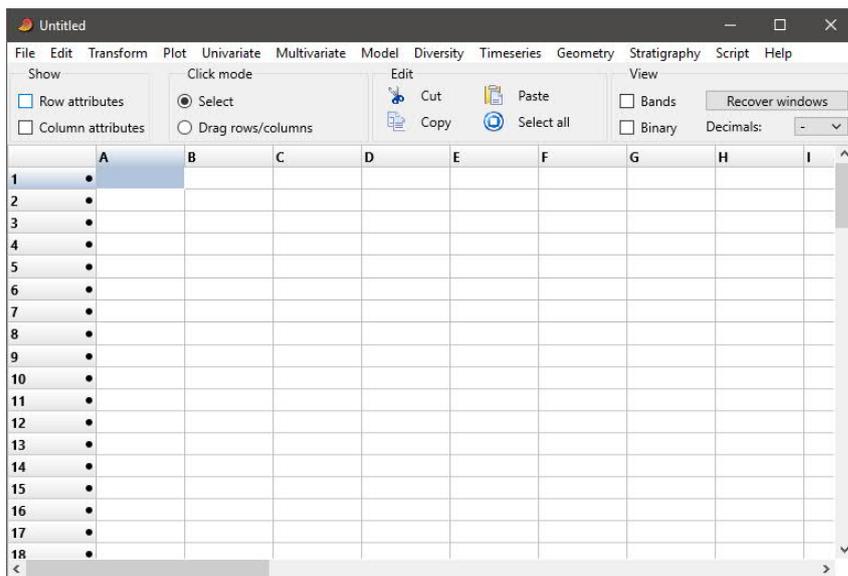


Рисунок 18. Интерфейс программы PAST

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 3 дополнительных поля: *Color (Цвет)*, *Symbol (Символ)* и *Name (Имя)*, которые позволяют соответствующий объект, наблюдение или испытуемого обозначить выбранным цветом, символом (точкой, плюсиком, квадратиком и т.п.) и именем, которые будут отображаться в диаграммах рассеивания (Рисунок 19).

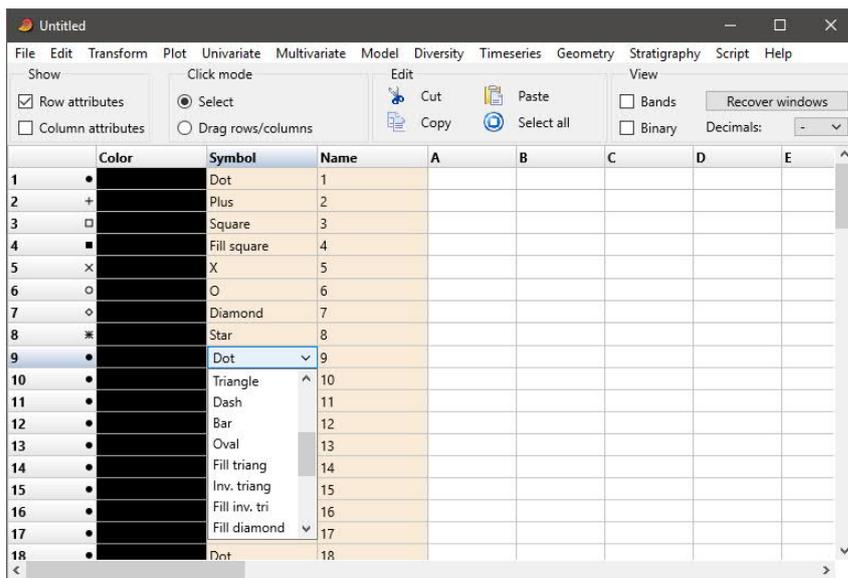


Рисунок 19. Атрибуты строк

Column attributes (Атрибуты столбцов) – для каждого столбца в шапке открываются 2 дополнительных поля: *Type (Тип)* и *Name (Имя)*, которые позволяют для соответ-

ствующих признаков, показателей или переменных выбрать тип и ввести имя (Рисунок 20, 21). 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 т.п. Тип *Group* необходимо использовать для оценки достоверности различий между группами, для дисперсионного или дискриминантного анализа.

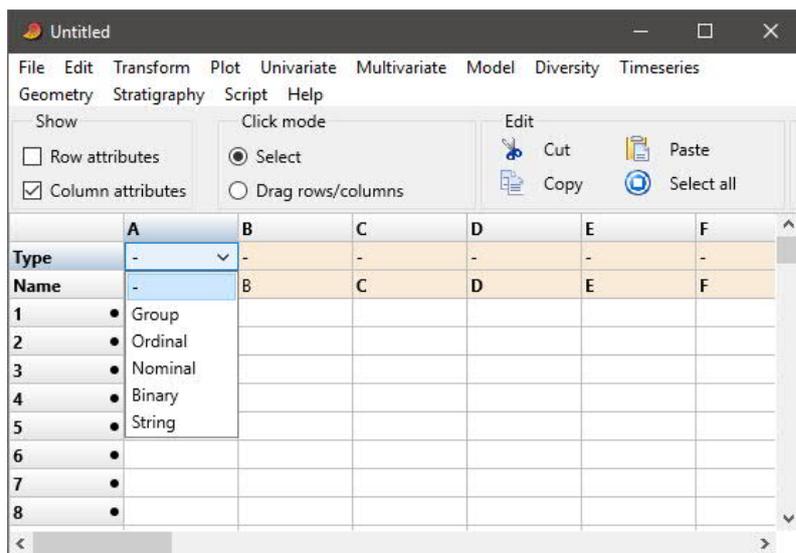


Рисунок 20. Атрибуты столбцов. Выбор атрибута столбца

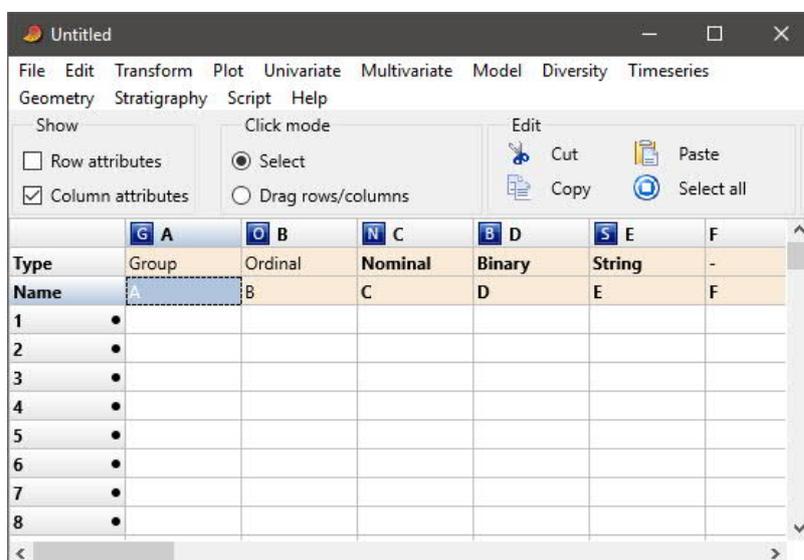


Рисунок 21. Атрибуты столбцов. Выбранные атрибуты столбцов

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, щел-

кая левой клавишей мыши по первому и последнему столбцу интервала или 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 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 различения.

Binary – выделяет черной заливкой ненулевые значения.

Decimals – определяет количество выводимых знаков после десятичной точки или запятой.

2.2. 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 текущий открытый файл с тем же именем.

Save as... (Сохранить как...) – сохраняет данные в файл 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 (Рисунок 22). В данном примере представлены значения показателей 2, 4, 7 и 10 субтестов универсального интеллектуального теста (УИТ), 2 показателей теста сходных рисунков Кагана и 11 показателей модифицированного патохарактерологического диагностического опросника (МПДО) школьников 9 и 11 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	T	U
1		Класс																			
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		
11	11 м		5	5	2	9	58	0	9	5	8	1	5	2	4	6	7	3	3		
12	11 м		7	3	6	7	89	1	12	8	5	4	6	4	6	6	10	6	6		
13	9 ж		8	6	8	9	47	1	5	5	9	5	6	4	2	3	3	5	2		
14	9 ж		7	5	6	9	84	2	7	5	6	3	4	0	6	5	5	3	3		
15	9 ж		6	5	7	5	88	2	6	6	8	5	8	3	4	6	6	6	2		
16	11 ж		6	4	5	8	56	1	2	3	4	2	3	2	3	1	4	2	1		
17	11 ж		4	2	6	7	136	4	2	1	3	0	0	3	3	1	3	5	2		
18	9 ж		3	5	7	8	58	1	9	9	7	10	8	9	10	7	7	6	3		
19	11 м		8	6	9	5	94	4	5	0	4	0	1	4	2	5	4	2	1		
20	9 ж		8	2	8	8	49	0	12	8	9	7	9	9	9	9	10	10	1		
21	9 ж		6	8	4	6	58	2	2	5	6	8	7	6	6	5	8	4	8		
22	11 ж		5	11	10	7	83	1	11	1	5	1	4	5	0	4	6	4	7		
23	11 ж		7	6	11	8	112	1	2	4	7	2	2	8	4	8	2	2	3		
24	11 м		6	9	9	8	91	0	9	4	4	3	2	4	1	4	4	5	5		

Рисунок 22. 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 (Рисунок 23) указать, что *Row contains – Only data cells* (Строки содержат только ячейки данных) и *Columns contain – Names, data* (Столбцы содержат имена переменных и данные).

Рисунок 23. 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 (Рисунок 24). После импорта файла данных из Excel в PAST следует в нужных столбцах указать тип переменных *Group*, а затем сохранить полученный файл PAST в выбранную папку и дать этому файлу имя.

	Класс	Пол	Зрительное	Числовой ин	Вербальный	Пространств	Время перво	Количество	Гипертимный
Типе	Group	Group	-	-	-	-	-	-	-
Name	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
9	9	ж	6	10	4	8	36	0	5
10	11	м	5	5	2	9	58	0	9
11	11	м	7	3	6	7	89	1	12
12	9	ж	8	6	8	9	47	1	5
13	9	ж	7	5	6	9	84	2	7
14	9	ж	6	5	7	5	88	2	6
15	11	ж	6	4	5	8	56	1	2
16	11	ж	4	2	6	7	136	4	2
17	9	ж	3	5	7	8	58	1	9
18	11	м	8	6	9	5	94	4	5
19	9	ж	8	2	8	8	49	0	12

Рисунок 24. Пример данных в PAST

2.3. 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 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) (Рисунок 25).

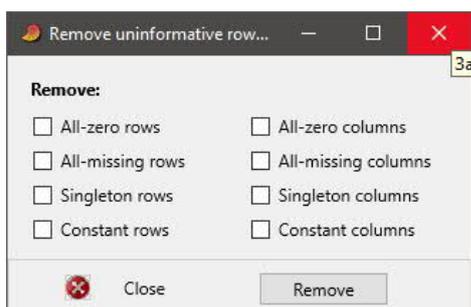


Рисунок 25. Удаление неинформативных строк или столбцов

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 0 и 1, другой 1, 2 и 3 в таблицу частот сочетаний 2x3), 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» (Рисунок 27), выбрать пункт меню «Value pairs to matrix» и получить таблицу частот сочетаний (Рисунок 28).

	Класс	Пол	Зрительное в	Числовой ин	Вербальный	Пространств	Время перво	Количество	Гипе
Типе	Group	Group	-	-	-	-	-	-	-
Name	Класс	Пол	Зрительное в	Числовой инт	Вербальный в	Пространств	Время перво	Количество о	Гипе
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
10	11	м	5	5	2	9	58	0	9
11	11	м	7	3	6	7	89	1	12

Рисунок 26. Выделение столбцов для таблицы частот сочетаний

	9	11
ж	19	13
м	5	11

Рисунок 27. Таблица частот сочетаний

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.

2.4. Пункт меню Plot (Графики)

XY graph (*Двухкоординатный график*) – строит двумерную диаграмму рассеивания (скатерграмму) по двум выделенным столбцам. Например, выделяем столбцы «Числовой интеллект» и «Пространственный интеллект» (Рисунок 28)

Type	Класс	Пол	Зрительное в	Числовой инт	Вербальный и	Пространств	Время перво	Количество с	Гипе
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
10	11	м	5	5	2	9	58	0	9
11	11	м	7	3	6	7	89	1	12

Рисунок 28. Выделение двух столбцов для получения двумерной диаграммы рассеивания

Результаты выводятся в окно «XY graph» (Рисунок 29).

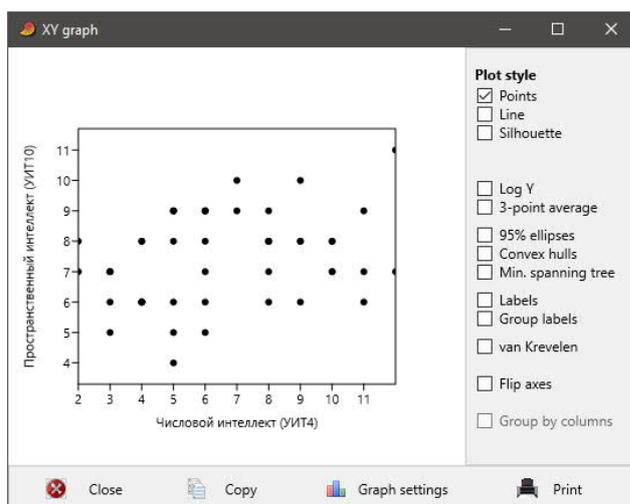


Рисунок 29. Окно XY graph

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, то есть поворачивает график на 90 градусов (Рисунок 30).

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

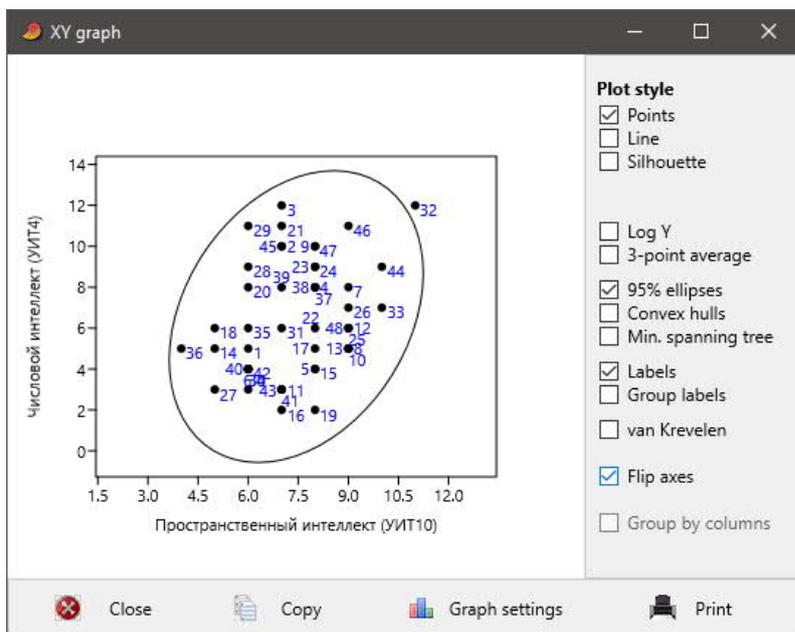


Рисунок 30. Модификация XY graph

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) (Рисунок 31).

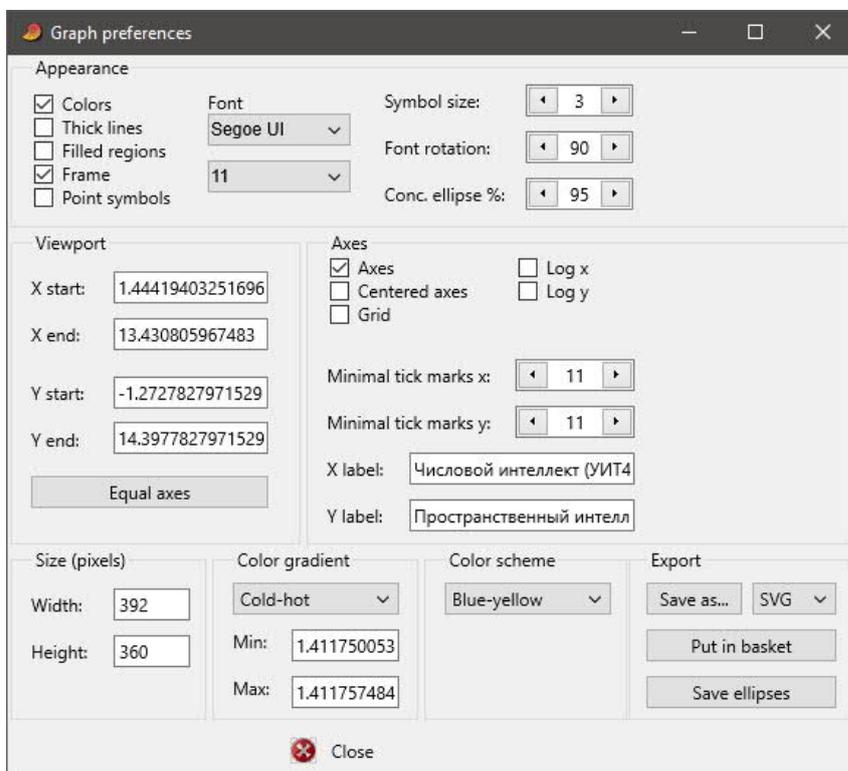


Рисунок 31. Параметры графика

Histogram (Гистограмма) – строит гистограмму распределения частот значений по выделенному столбцу. Например, выделяем столбец «Пространственный интеллект» (Рисунок 32).

	Класс	Пол	Зрительное в	Числовой инт	Вербальный	Пространств	Время перво	Количество о	Гипе
Тип	Group	Group	-	-	-	-	-	-	-
Имя	Класс	Пол	Зрительное в	Числовой инт	Вербальный	Пространств	Время перво	Количество о	Гипе
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
10	11	м	5	5	2	9	58	0	9
11	11	м	7	3	6	7	89	1	12

Рисунок 32. Выделение столбца для получения гистограммы

Результаты выводятся в окно «Histogram» (Рисунок 33).

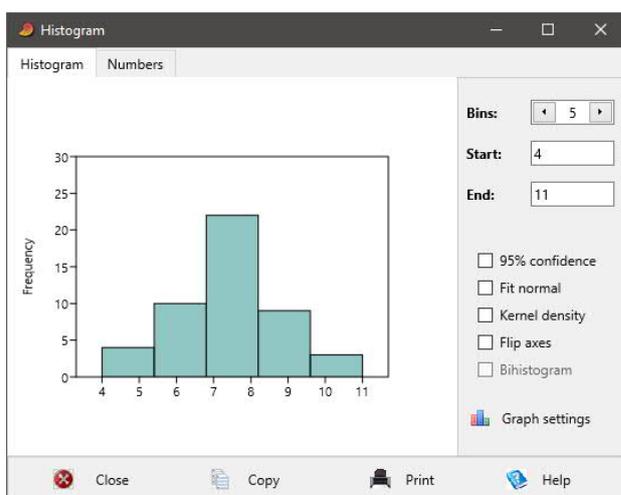


Рисунок 33. Окно Histogram

Параметр «Bins» позволяет изменить число интервалов группировки значений, «Fit normal» – наложить кривую нормального распределения, «Flip axes» – повернуть график на 90 градусов (Рисунок 34). Кнопка «Graph settings» (Параметры графика) позволяет изменить «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 (Цветовую схему). В данном графике полезно в параметрах графика вручную указать подпись оси X с названием показателя.

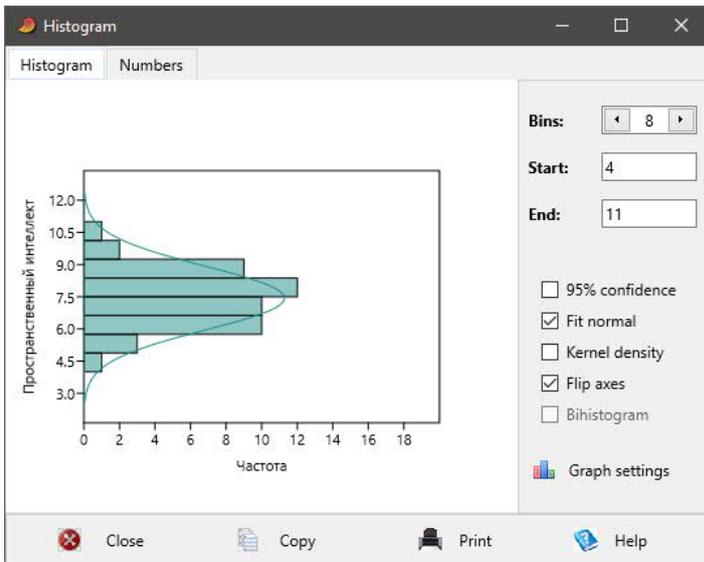


Рисунок 34. Модификация Histogram

Для двух выделенных столбцов, один из которых соответствует типу «Group» (Рисунок 35), можно получить «Bihistogram» (Двойную гистограмму) (Рисунок 36).

Type	Name	Класс	Пол	Зрительное в	Числовой инт	Вербальный	Пространств	Время перво	Количество о	Типе
1	9	9	ж	10	5	7	6	27	4	9
2	9	9	ж	6	10	8	7	57	0	10
3	11	11	м	8	12	5	7	80	5	12
4	9	9	ж	8	8	5	8	117	3	8
5	11	11	ж	8	4	4	8	46	4	11
6	11	11	м	6	3	5	6	157	6	11
7	9	9	м	10	8	7	9	50	5	5
8	11	11	ж	4	5	7	9	131	0	9
9	9	9	ж	6	10	4	8	36	0	5
10	11	11	м	5	5	2	9	58	0	9
11	11	11	м	7	3	6	7	89	1	12

Рисунок 35. Выделение столбцов для двойной гистограммы, диаграммы средних с усами, двухвыборочного теста или однофакторного дисперсионного анализа

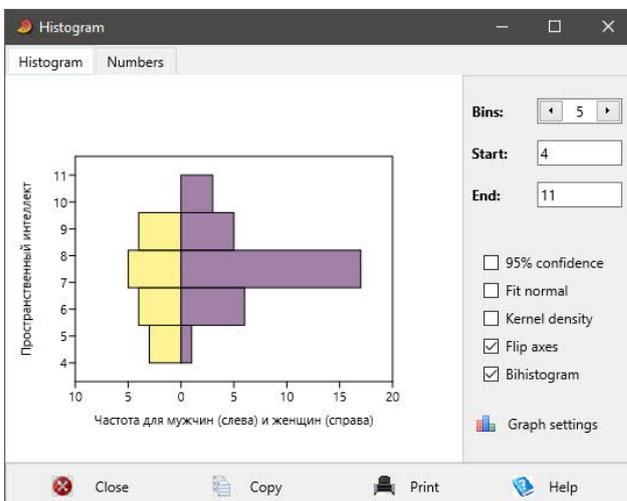


Рисунок 36. Двойная гистограмма

Barchart/boxplot (Столбиковая/прямоугольная диаграмма) – строит график параметров распределений для одного или нескольких выделенных однотипных показателей (Рисунок 37).

Type	Класс	Пол	Зрительное в	Числовой инт	Вербальный и	Пространств	Время перво	Количество о	Гипе
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
10	11	м	5	5	2	9	58	0	9
11	11	м	7	3	6	7	89	1	12

Рисунок 37. Выделение нескольких столбцов для столбиковой диаграммы, вычисления параметров распределений или корреляций

Результаты выводятся в окно «Box plot» (Рисунок 38, 39, 40). 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.

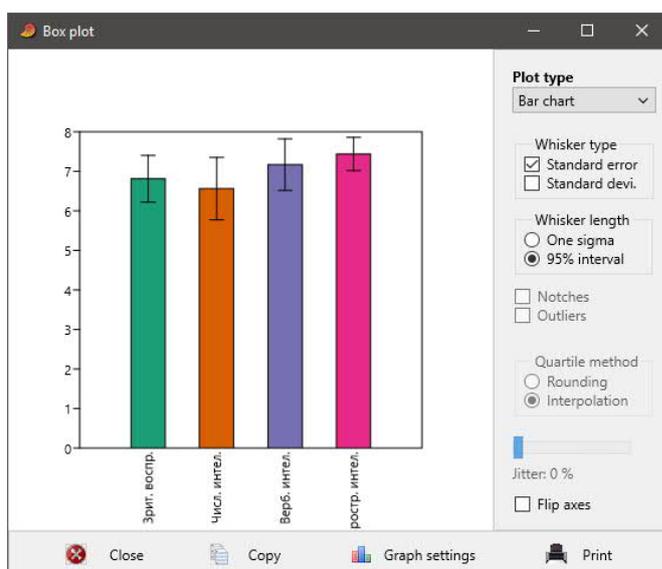


Рисунок 38. Окно Box Plot, тип графика Bar chart (Столбиковая диаграмма)

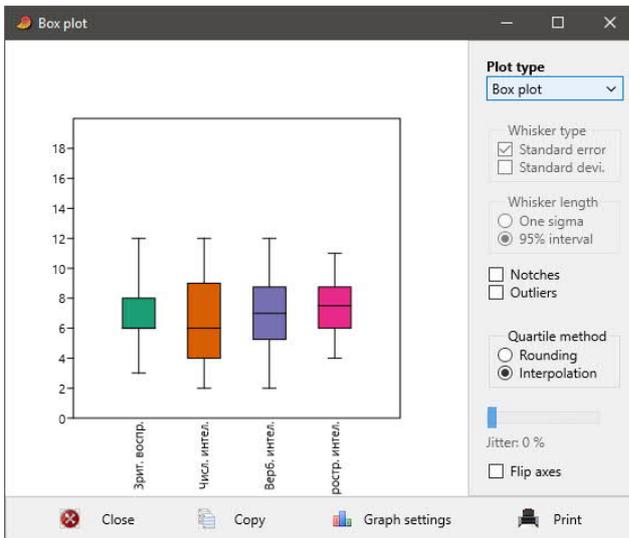


Рисунок 39. Тип графика Box plot (Прямоугольная диаграмма).

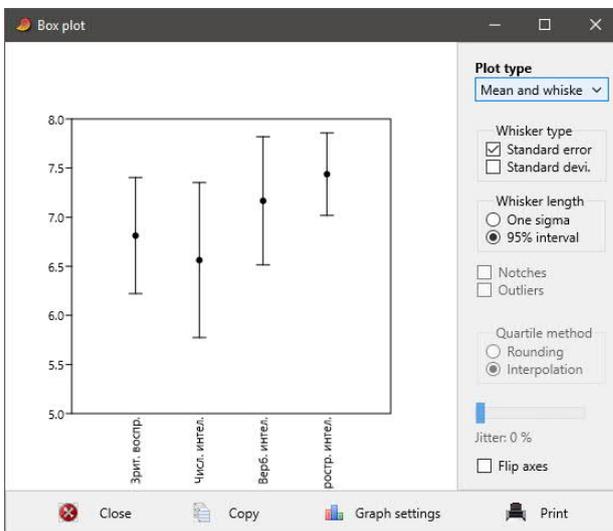


Рисунок 40. Тип графика Mean and whiskers (Среднее с усами)

Можно построить график для двух показателей, один из которых имеет тип Group, например, зависимость пространственного интеллекта от пола (Рисунок 35, 41)

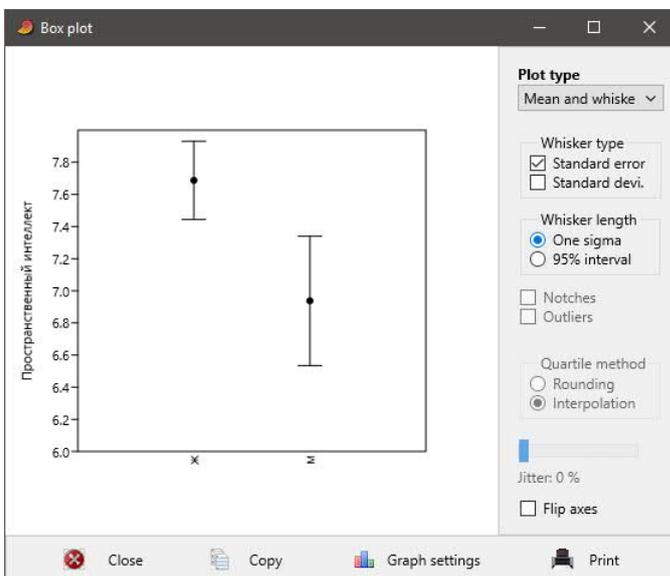


Рисунок 41. Тип графика Mean and whiskers (Среднее с усами)

2.5. Пункт меню Univariate (Одномерные методы)

Summary statistics (Описательная статистика) – дает возможность вычислить параметры распределений показателей по выделенным столбцам (Рисунок 37, 42).

- *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 встречается 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

Bootstrap
 Bootstrap type: Simple
 Bootstrap N: 9999
 Recompute

Рисунок 42. Окно параметров распределений показателей

Здесь и далее во всех окнах таблиц кнопка «Сору» 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 2 столбца, соответствующие двум показателям, один из которых может иметь тип Group, например, Пол и Пространственный интеллект (Рисунок 35). После выбора соответствующего пункта меню выводится окно с результатами.

Вкладка «t test» (Рисунок 43) 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.

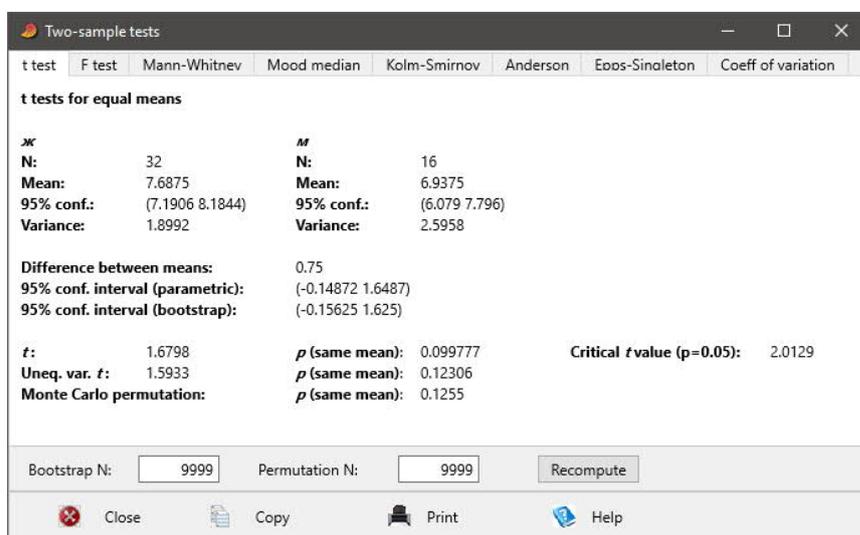


Рисунок 43. Окно Two-sample tests / t-test

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

Вкладка Mann-Whitney (Рисунок 44) 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.

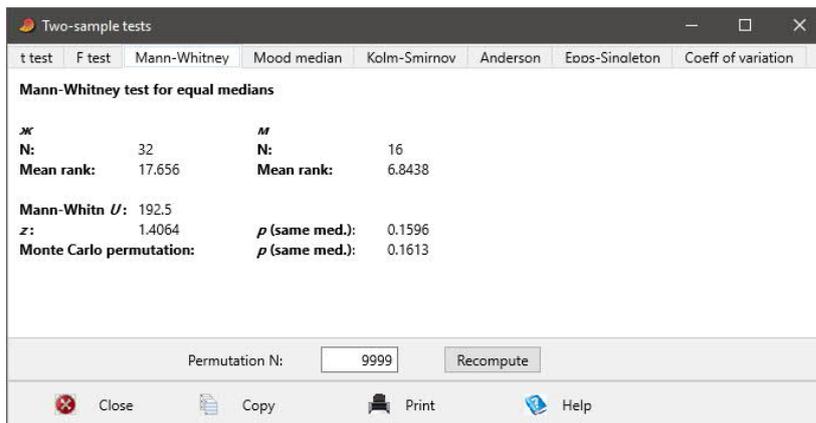


Рисунок 44. Окно Two-sample tests / Mann-Whitney

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

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 2 столбца, соответствующие выбранному показателю до и после какого-либо события, например, усталость до и после тренинга (Рисунок 45).

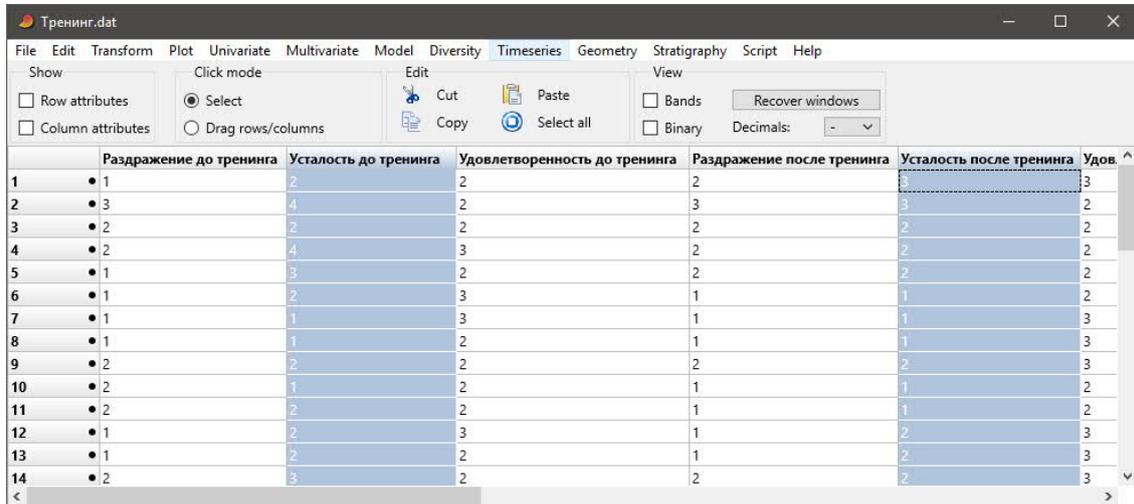


Рисунок 45. Выделение столбцов для попарно сопряженных критериев

После выбора соответствующего пункта меню выводится окно с результатами (Рисунок 46):

- 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.

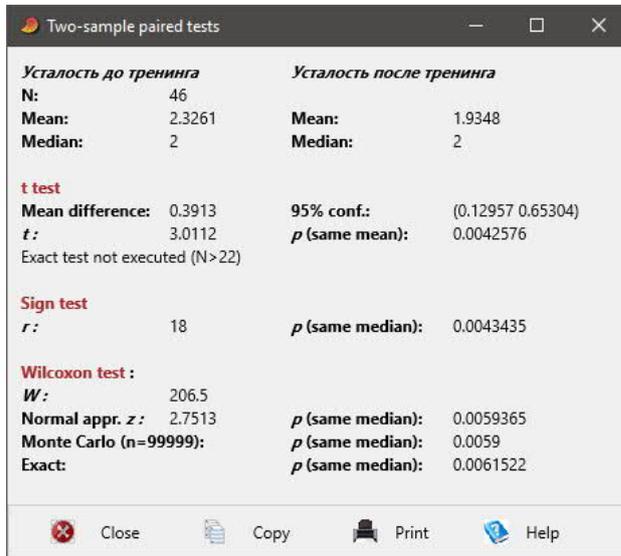


Рисунок 46. Окно Two-sample paired tests

В указанном примере сдвиг среднего значения и медианы показателя усталости до и после тренинга достоверно отличается от 0 ($p=0.004$ и $p=0.006$).

Two-sample tests >F and t tests with parameters – оценка достоверности различий между параметрами двух независимых выборок. Представляет простой калькулятор (Рисунок 47), в который вводятся значения средних значений (*Mean*), дисперсий (*Variance*) и объемов (N) 1 и 2 выборки (*Sample 1* и *Sample 2*), и при щелчке на «*Compute*» (*вычислить*) выводятся значения F-критерия Фишера и t-критерия Стьюдента с соответствующими уровнями значимости p .

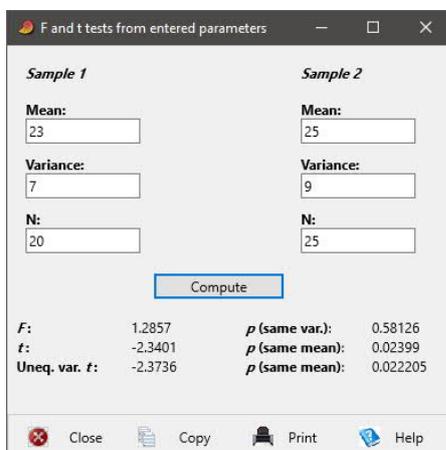


Рисунок 47. Критерии Фишера и Стьюдента

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

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 (Рисунок 35). После выбора соответствующего пункта меню выводится окно с результатами (Рисунок 48). Если показатель характеризуется нормальным распределением, а выборки равным объемом, используется F-критерий Фишера.

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	6	1	2,822	2,822	0,09978
Within groups:	97,8125	46	2,12636		Permutation p (n=99999)
Total:	103,813	47			0,121

Components of variance (only for random effects):			
Var(group):	0,181577	Var(error):	2,12636
		ICC:	0,078675
omega ² :	0,03656		

Levene's test for homogeneity of variance, from means	p (same):	0,3894
Levene's test, from medians	p (same):	0,3694

Welch F test in the case of unequal variances: F=2,538, df=26,28, p=0,1231

Рисунок 48. Окно Several sample test / One-way ANOVA

В данной форме:

- *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 (Рисунок 49).

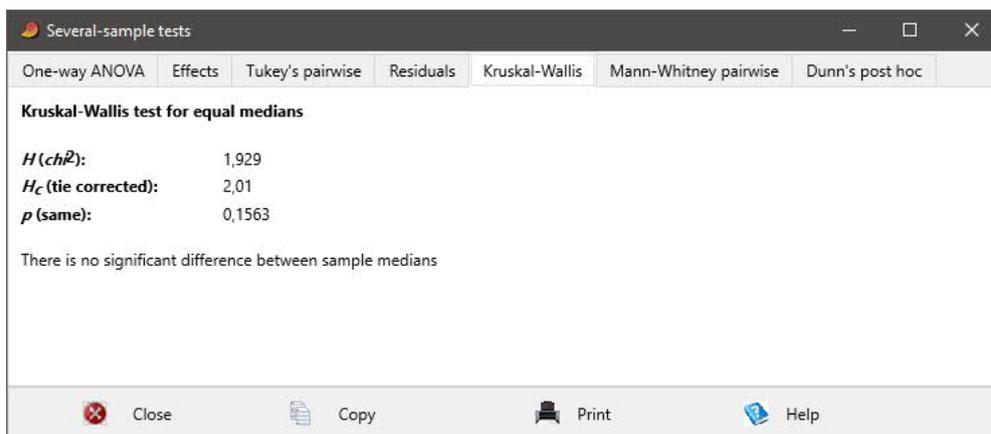


Рисунок 49. Окно Several sample test / Kruskal-Wallis

ANOVA etc. (several samples) > Two-way ANOVA – Двухфакторный дисперсионный анализ). Выделяются 3 столбца, 2 из которых имеют тип Group (Рисунок 50).

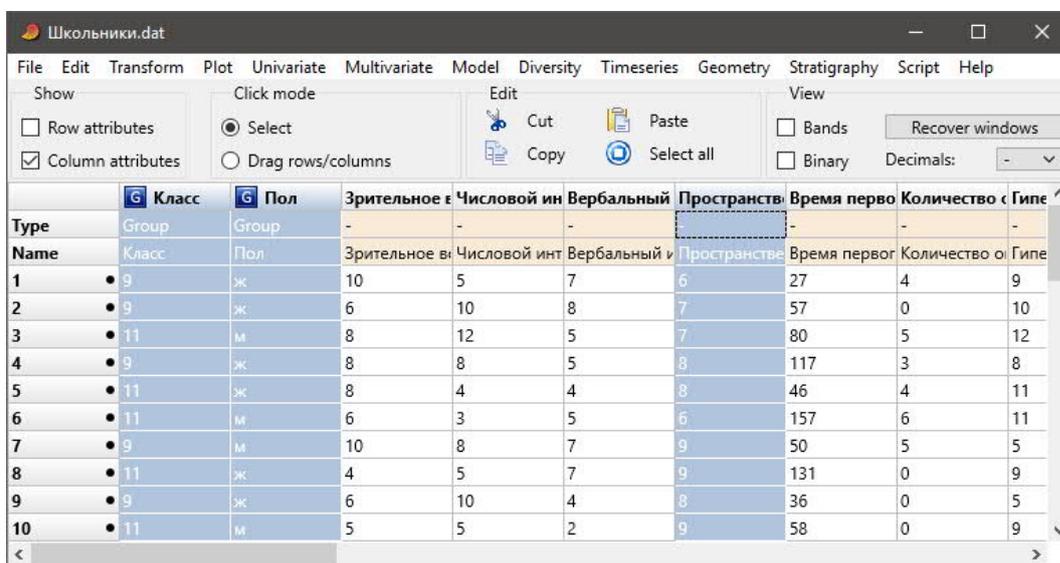


Рисунок 50. Выделение столбцов для двухфакторного дисперсионного анализа

В результате выводится следующее окно (Рисунок 51). 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.

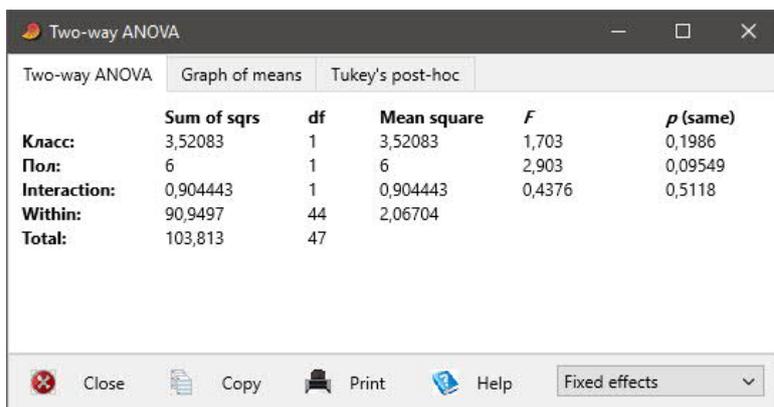


Рисунок 51. Окно Two-way ANOVA

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 (Рисунок 52, 53).

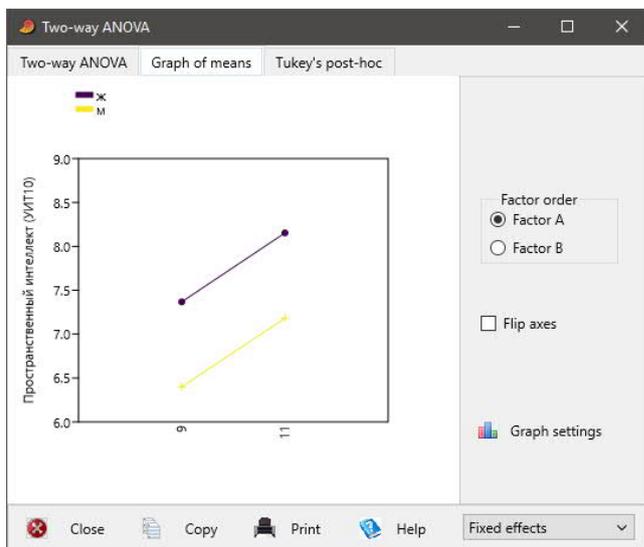


Рисунок 52. Окно Two-way ANOVA / Graph of means



Рисунок 53. Окно Two-way ANOVA / Graph of means

Correlation (Корреляция) – позволяет вычислять различные коэффициенты корреляции между выбранными показателями. Выделяются столбцы, соответствующие выбранным показателям (Рисунок 37). Выводится окно с таблицей корреляций показателей (Рисунок 54). 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(uncorr)* нижняя треугольная часть таблицы отображает значения коэффициентов корреляции, а верхняя треугольная – уровни значимости. При этом каждому значению коэффициента корреляции соответствует симметричное относительно главной диагонали значение уровня значимости. Если уровень значимости $p < 0,05$, то коэффициент корреляции достоверно отличается от 0 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(uncorr)* обе части таблицы отображают показатели уровней значимости.

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 0,000602. Иными словами, число после «E-» указывает, на сколько знаков влево нужно сместить десятичную запятую, или сколько нулей поставить перед имеющимся числом.

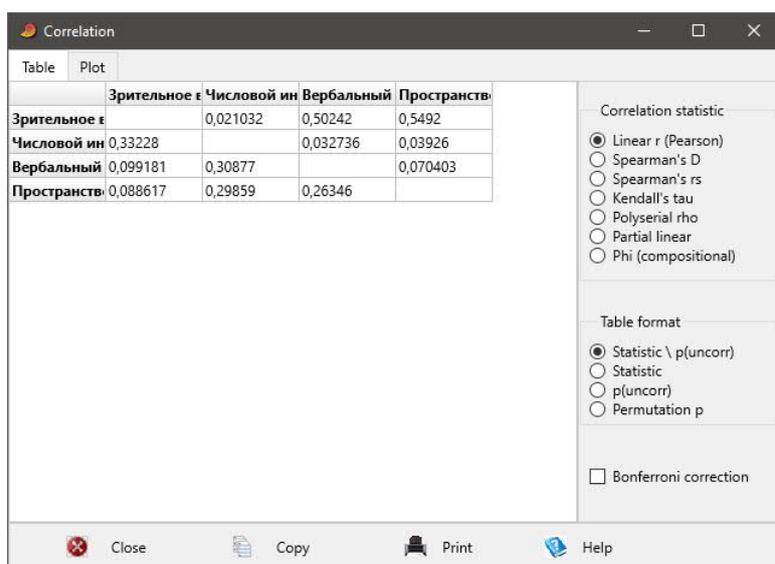


Рисунок 54. Окно Correlation

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 (Рисунок 55) выводятся в окно «Test for normal distribution» (Оценка нормальности распределения)

	Зрительное воспри:	Числовой интеллект	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

Рисунок 55. Окно Tests for normal distribution

- *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 в выбранном показателе. Результаты выводятся в окне «Outlier tests» (Рисунок 56). 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.

Single outlier	
Grubbs test	
G	2,397
p(no outlier)	0,6603
No significant outliers	
Dixon Q test	
Sample size too large for Dixon test (n>30)	

Рисунок 56. Окно Outlier tests / Single outlier

В указанном примере выскакивающие значения не выявлены. При наличии выброса сообщается, какое значение (*value*) в какой строке (*row*) является выбросом (Рисунок 57).

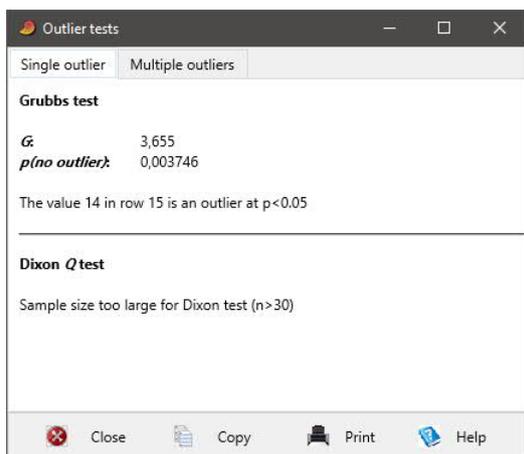


Рисунок 57. Окно Outlier tests / Single outlier

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 (Рисунок 58).

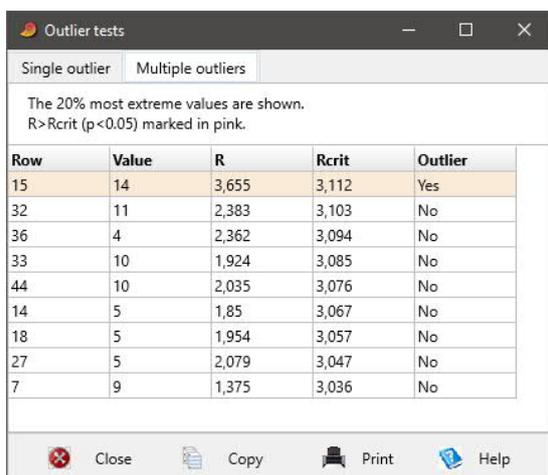


Рисунок 58. Окно Outlier tests / Multiple outlier

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 (Рисунок 27, 28) 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 (Рисунок 59) указывается:

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

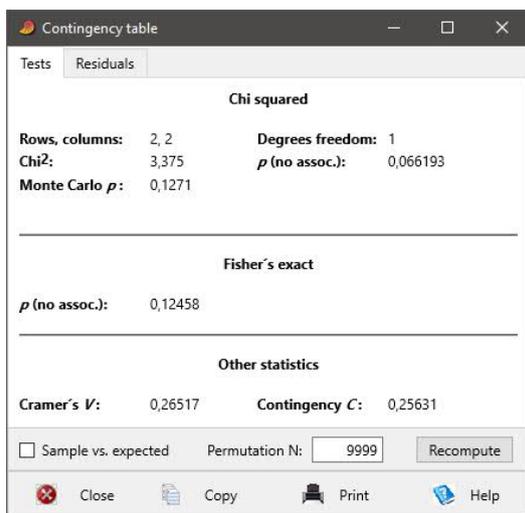


Рисунок 59. Окно таблицы сопряженности

В указанном 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.

2.6. 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 (Рисунок 60).

	Количество с	Гипертимный	Циклоидный	Лабильный (Астено-нерв)	Сензитивный	Психастенич	Шизоидный	Эпилептоид	Истероиды	Неустойчив	Неискренно Т
1	• 4	9	2	6	3	2	3	3	4	9	4
2	• 0	10	5	6	3	4	0	3	5	8	4
3	• 5	12	3	5	1	4	2	7	5	7	3
4	• 3	8	9	12	4	9	3	8	8	4	5
5	• 4	11	9	9	8	5	5	7	8	10	4
6	• 6	11	3	3	4	3	1	3	5	8	2
7	• 5	5	6	7	2	1	2	7	1	6	4
8	• 0	9	6	5	1	2	3	1	1	5	6
9	• 0	5	0	5	1	4	7	2	2	1	3
10	• 0	9	5	8	1	5	2	4	6	7	3
11	• 1	12	8	5	4	6	4	6	6	10	6
12	• 1	5	5	9	5	6	4	2	3	3	5
13	• 2	7	5	6	3	4	0	6	5	5	3
14	• 2	6	6	8	5	8	3	4	6	6	2
15	• 1	2	3	4	2	3	2	3	1	4	2
16	• 4	2	1	3	0	0	3	3	1	3	5
17	• 1	9	9	7	10	8	9	10	7	7	6
18	• 4	5	0	4	0	1	4	2	5	4	2
19	• 0	12	8	9	7	9	9	9	9	10	1
20	• 2	2	5	6	8	7	6	6	3	8	4
21	• 1	11	1	5	1	4	5	0	4	6	4
22	• 1	2	4	7	2	2	8	4	8	2	3

Рисунок 60. Выделение столбцов показателей для факторного анализа.

Результаты (Рисунок 61) выводятся в окно «*Principal component analysis*».

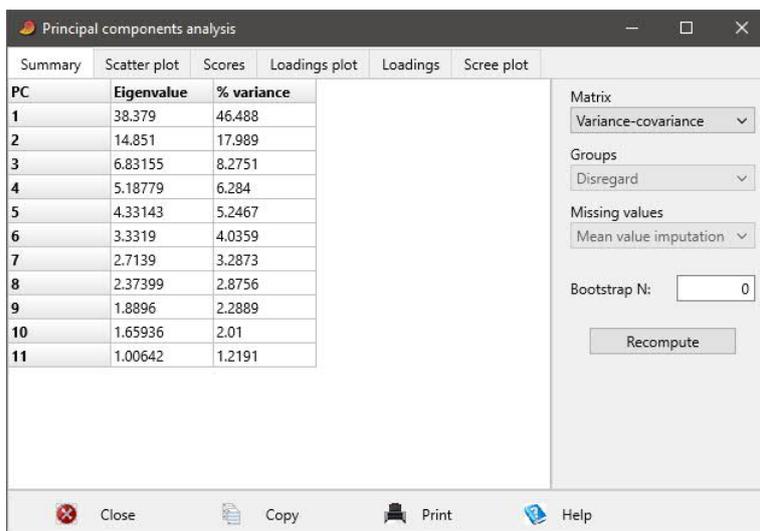


Рисунок 61. Окно Principal components analysis / Summary

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 (Рисунок 62).

- PC – номер фактора или главной компоненты.
- Eigenvalue – собственное значение, соответствующее количеству коррелирующих показателей, вошедших в данный фактор. Факторы, собственное значение которых менее 1, можно считать не информативными.
- % variance – процент дисперсии, доля показателей, вошедших в данный фактор.

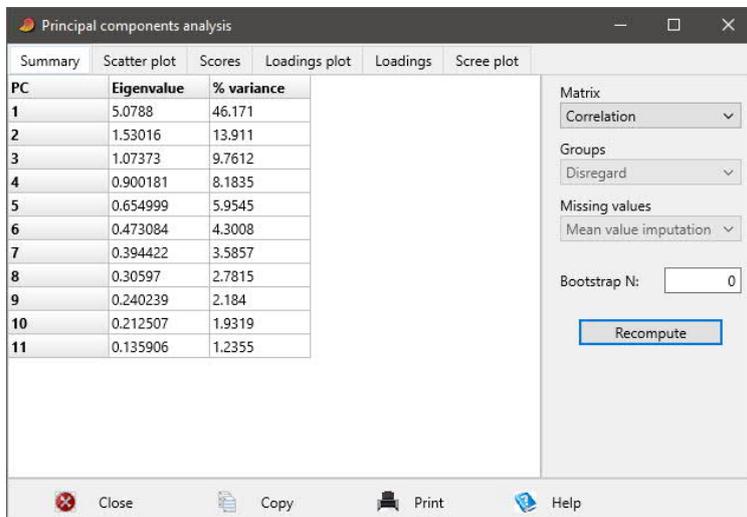


Рисунок 62. Окно Principal components analysis / Summary

В указанном примере в первый фактор можно включить 5 показателей, во второй фактор – 1 или 2 показателя, в 3 фактор – 1 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 (Рисунок 63). После того, как кривая перестает существенно падать, а также при значениях компоненты ниже кривой *Broken Stick* (*Сло-*

манная палка – кривая ожидаемых собственных значений при случайной модели), соответствующие факторы можно считать малоинформативными.

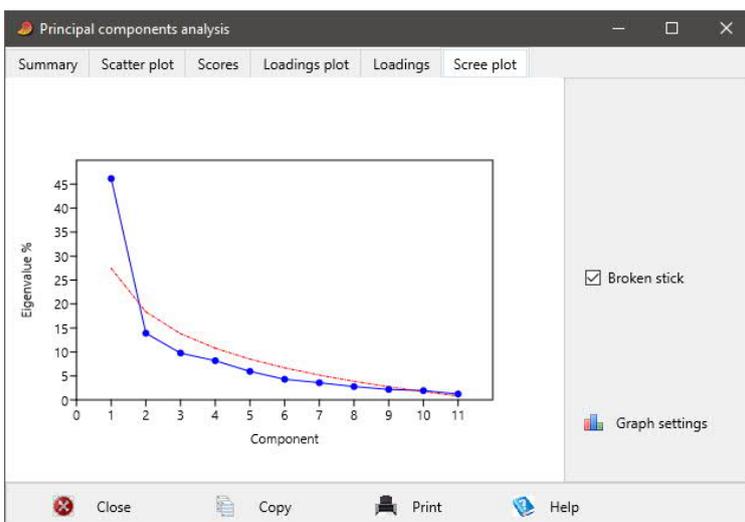


Рисунок 63. Окно Principal components analysis / Scree plot

Во вкладке «Loadings» (*Нагрузки*) представлены нагрузки или веса, с которыми соответствующие показатели вошли в каждый фактор (Рисунок 64).

	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
Неустойчивь	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

Рисунок 64. Окно Principal components analysis / Loadings

Полученную таблицу целесообразно скопировать и вставить в MS Excel (Рисунок 65). 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 столько показателей, каково собственное значение данного фактора.

	A	B	C	D	E	F	G	H	I	J	K	L
1		PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11
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	-0.0034905	-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

Рисунок 65. Таблица факторных нагрузок показателей на факторы в формате MS Excel

В данном примере в 1 фактор с наибольшими нагрузками вошли показатели циклоидного, астено-невротического, шизоидного, эпилептоидного и сензитивного типов акцентуаций личности, характеризующие эмоциональную нестабильность. 2 фактор включил показатели гипертимного и истероидного типов акцентуаций, свидетельствующие об активности и экстравертированности. Наконец, 3 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 (Рисунок 66, 67). 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.



Рисунок 66. Окно Principal components analysis / Loadings plot

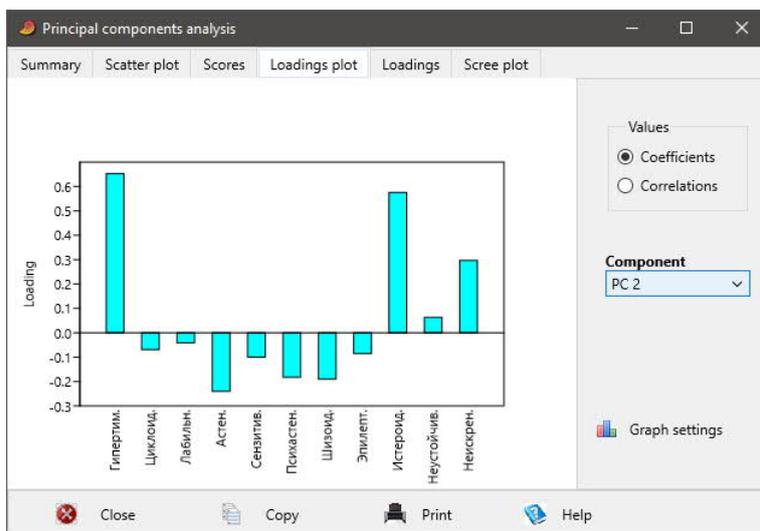


Рисунок 67. Окно Principal components analysis / Loadings plot

Во вкладке «Scores» (Оценки) представлены оценки каждого испытуемого по каждому фактору, соответствующие координатам этого испытуемого в пространстве факторов (Рисунок 68).

	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

Рисунок 68. Окно Principal components analysis / Scores

Во вкладке «Scatter plot» (двумерная диаграмма рассеивания, или скатерграмма) в виде точек представлены координаты каждого объекта в пространстве двух выбранных факторов (Рисунок 69). По умолчанию, если мы не выберем иначе, оси X (X axis) соответствует ось 1 фактора (Component 1), оси Y (Y axis) соответствует ось 2 фактора (Component 2). При желании названия осей (X label и Y label) можно вписать в параметрах графика (Graph settings).

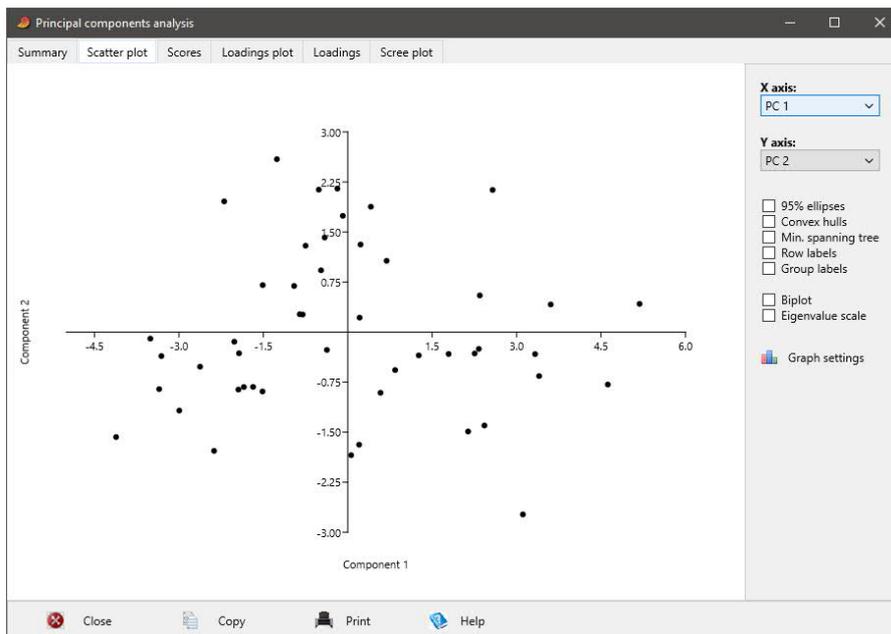


Рисунок 69. Окно Principal components analysis / Scatter plot

Можно на график вывести имена или номера объектов (Рисунок 70), отметив поле «Row labels» (*Метки строк*).

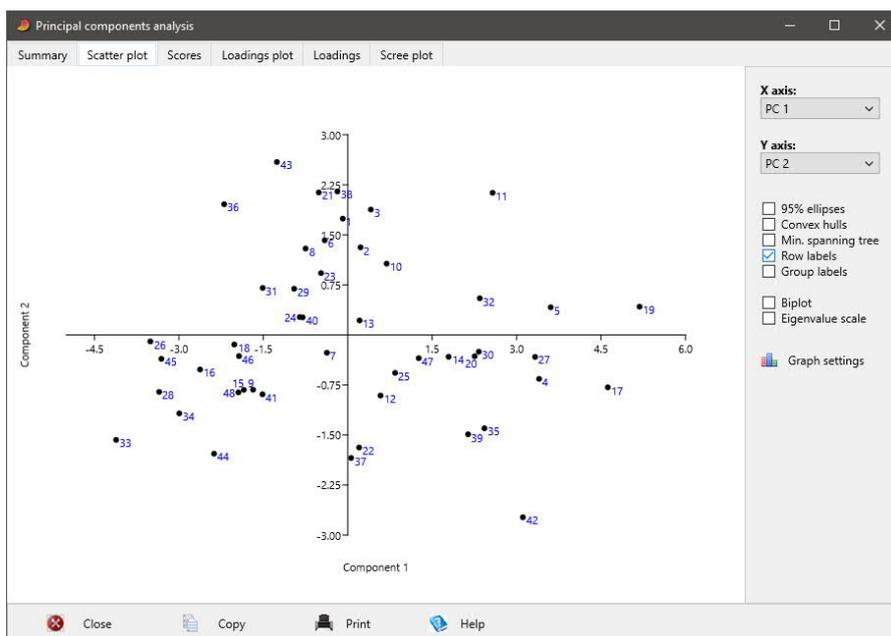


Рисунок 70. Окно Principal components analysis / Scatter plot

Наконец, на график можно вывести векторы показателей (Рисунок 71), отметив поле «Biplot» (*Двойной график*).

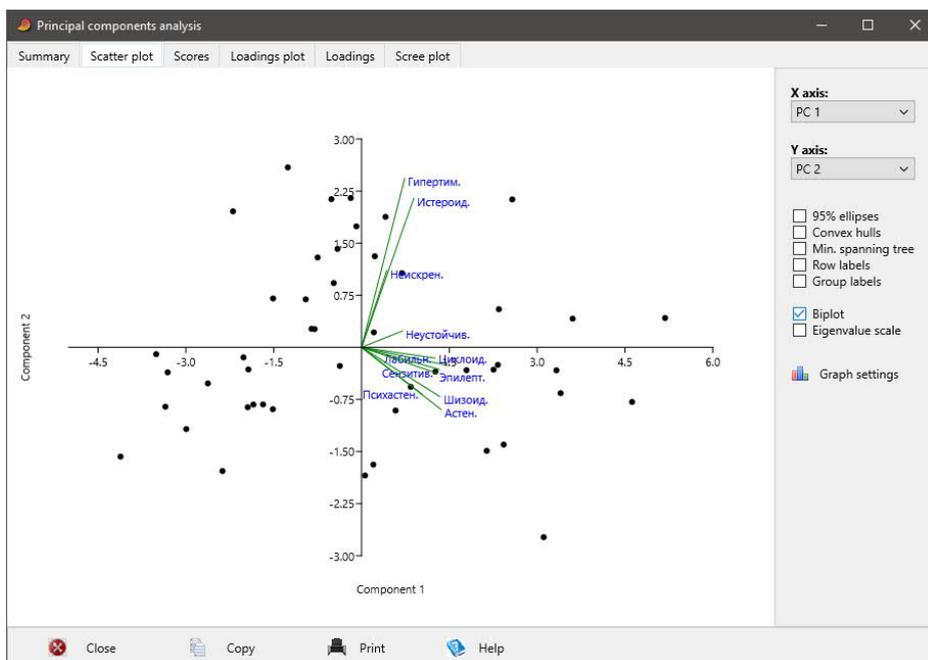


Рисунок 71. Окно Principal components analysis / Scatter plot

Discriminant analysis LDA (Дискриминантный анализ) – позволяет получить линейное уравнение, максимально дифференцирующее 2 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. То есть 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 – среднего балла 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 2 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, 8 показателей личностного типологического опросника (чувствительный, дистимический, демонстративный, возбудимый, застревающий, педантичный, замкнутый и гипертимный) и 4 – универсального интеллектуального теста (зрительное восприятие, числовой, вербальный и пространственный интеллект) в качестве независимых переменных (Рисунок 72).

	Направление	Пол	Успеваемость	Гр. успева: 1	Чувствител 2	Дистимиче 3	Демонстра 4	Возбудимь 5	Застревак ^
19	Техническое	м	1,50	1	1	4	4	5	4
20	Техническое	м	3,17	1	2	0	4	2	1
21	Техническое	м	3,50	1	2	3	1	2	2
22	Техническое	ж	4,00	2	3	4	2	1	2
23	Экономическ	м	4,17	2	5	2	3	4	1
24	Экономическ	м	4,50	2	2	1	2	8	4
25	Экономическ	м	5,00	2	3	1	1	3	4
26	Экономическ	м	3,83	2	2	3	0	2	3
27	Экономическ	м	3,50	1	2	1	1	5	0
28	Экономическ	ж	5,00	2	1	1	1	1	2
29	Экономическ	ж	4,33	2	4	3	3	4	4
30	Экономическ	ж	4,50	2	3	3	1	2	2
31	Экономическ	ж	5,00	2	1	1	2	5	2
32	Экономическ	ж	4,00	2	2	2	1	4	1
33	Экономическ	ж	3,67	1	2	1	1	5	4

Рисунок 72. 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 (Рисунок 73).

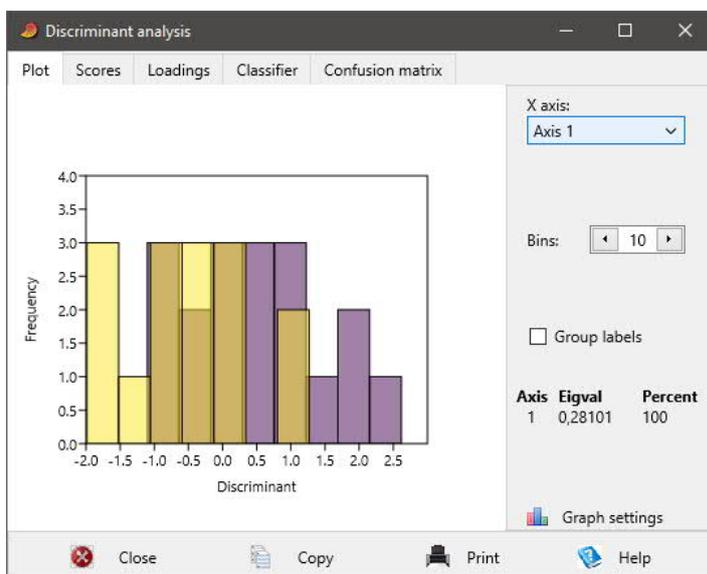


Рисунок 73. Окно Discriminant analysis / Plot

Во вкладке «Loadings» (Нагрузки) получаем весовые коэффициенты каждого показателя, характеризующие степень его влияния на группу успеваемости (Рисунок 74).

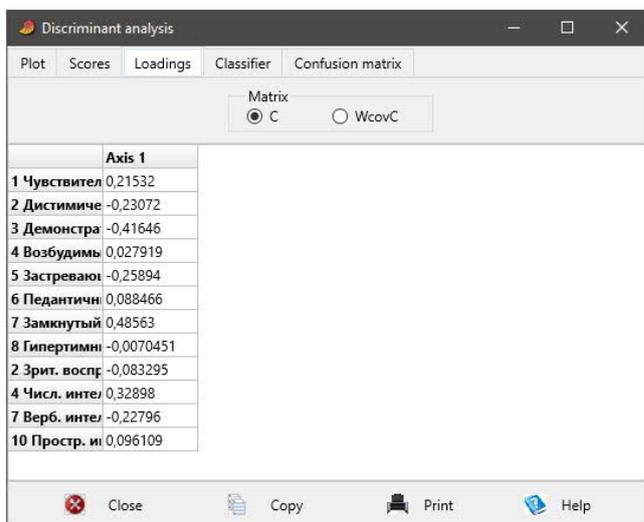


Рисунок 74. Окно Discriminant analysis / Loadings

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

Если значение дискриминантной функции менее 0, 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 (Рисунок 75). Чем выше значение дискриминантной функции, тем выше вероятность принадлежности испытуемого к группе 2.

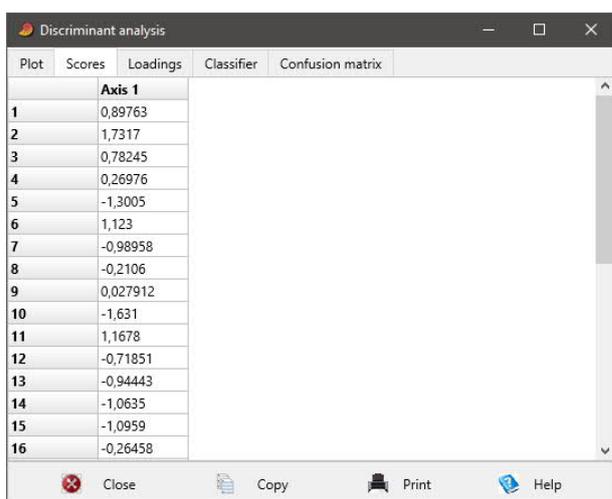


Рисунок 75. Окно Discriminant analysis / Scores

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) (Рисунок 76).

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

Рисунок 76. Окно Discriminant analysis / Classifier

Наконец, во вкладке «*Confusion matrix*» (*Матрица смешивания*) мы можем узнать (Рисунок 77), сколько испытуемых из 1 группы по результатам дискриминантного анализа отнесены ко 2 группе (ошибочно) и к 1 группе (правильно), а также, сколько испытуемых из 2 группы отнесены ко 2 группе (правильно) и к 1 группе (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

Рисунок 77. Окно Discriminant analysis / Confusion matrix

Clustering > Classical (*Кластеризация, иерархический кластерный анализ*) – позволяет сгруппировать 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*) (Рисунок 78).

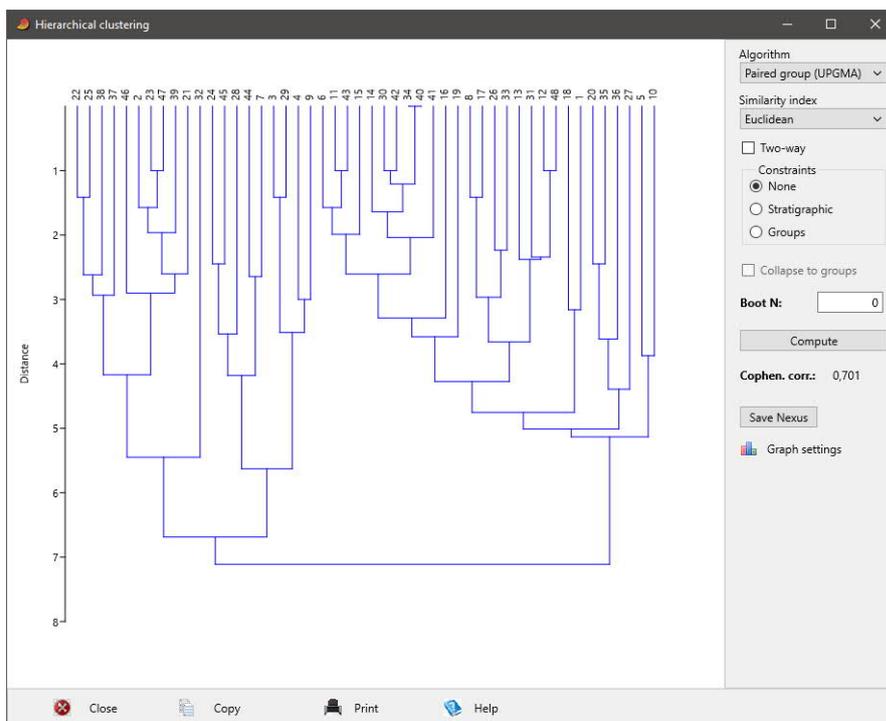


Рисунок 78. Окно Hierarchical clustering / Paired group

Мы можем выбрать один из 3-х 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-ти 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 (Рисунок 79).

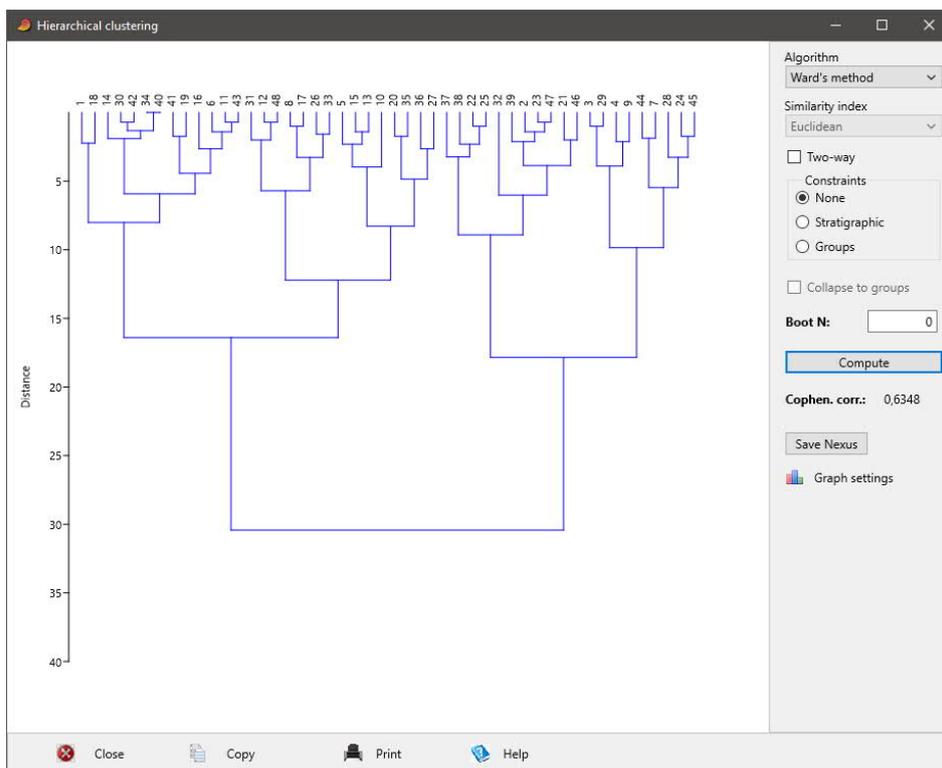


Рисунок 79. Окно Hierarchical clustering / Ward's method

2.7. 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*).

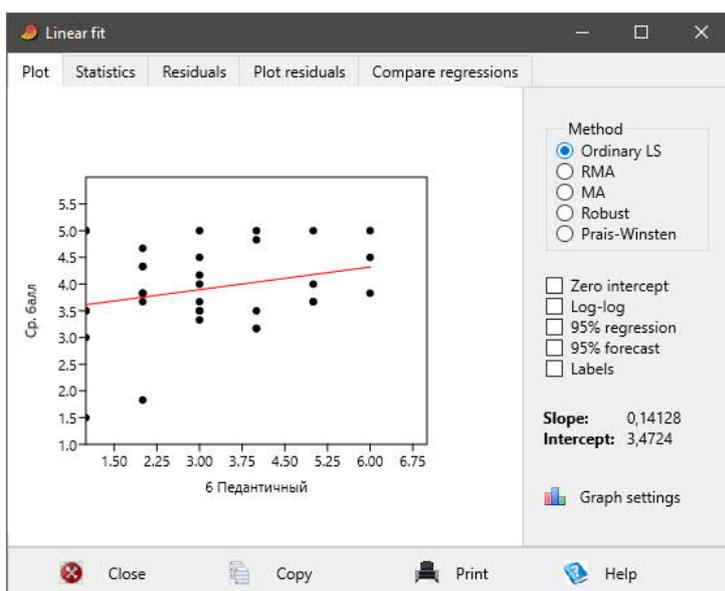


Рисунок 80. Окно Linear fit / Plot

Во вкладке «*Plot*» представлен график зависимости переменной Y от переменной X (Рисунок 80). В качестве метода регрессионного анализа (*Method*) можно выбрать «*Ordinary LS*».

inary LS» (*Ordinary Least Squares*) – простой метод наименьших квадратов), минимизирующий квадратичные ошибки Y .

Во вкладке «Statistics» выводятся параметры уравнения регрессии $Y=aX+b$, где

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

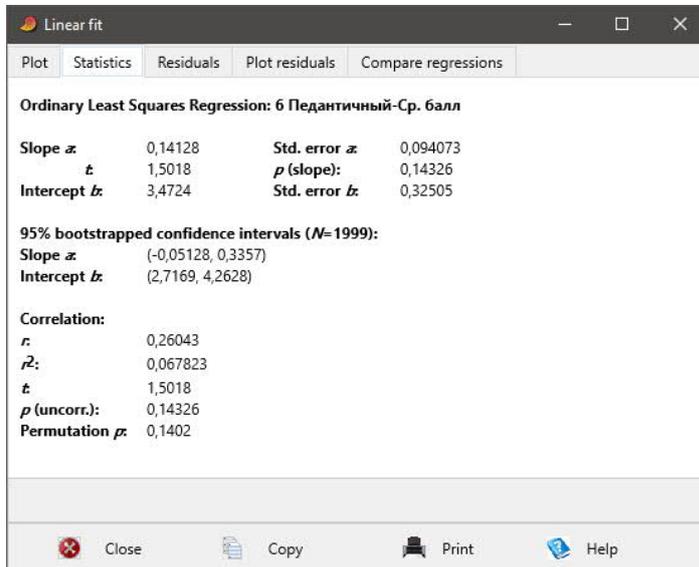


Рисунок 81. Окно Linear fit / Statistics

Во вкладке «Residuals» (*остатки*) указывается расстояние от каждой точки до линии регрессии (Рисунок 82). *Std. error of estimate* – стандартная ошибка оценки.

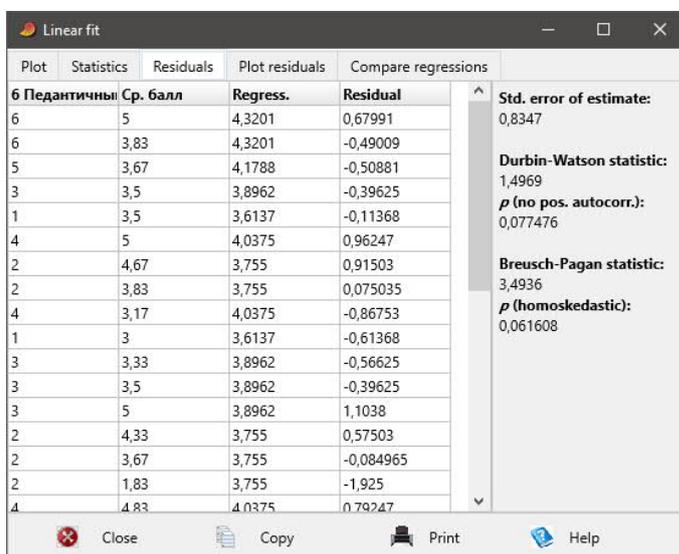


Рисунок 82. Окно Linear fit / Residuals

В данном примере формула регрессионной функции, позволяющая прогнозировать учебную успеваемость, будет: «Средний балл = 3.472 + 0.141 × Педагогический тип».

Multiple (1 dependent n independent) – множественная регрессия (1 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 (Рисунок 83).

	Ср. балл	Зрит. воспр.	Числ. интел.	Вербал. инте	Простр. инте
1	5,00	14	11	12	12
2	3,83	5	7	5	10
3	3,67	16	10	11	9
4	3,50	6	5	8	9
5	3,50	7	7	12	6
6	5,00	8	10	10	12
7	4,67	11	7	11	12
8	3,83	10	12	9	14
9	3,17	6	8	9	10
10	3,00	5	6	10	8
11	3,33	10	13	13	14
12	3,50	5	7	8	9
13	5,00	9	12	12	9
14	4,33	8	7	10	7
15	3,67	5	6	5	9
16	1,83	7	7	7	5

Рисунок 83. Выделение столбцов для множественного линейного регрессионного анализа

Результаты выводятся в окно «Multiple linear regression» (1 dependent n independent). Во вкладке «Statistics» (Рисунок 84) указываются:

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

Statistics	
Dependent variable:	Ср. балл
<i>N</i> :	33
Multiple <i>R</i> :	0,40977
Multiple <i>R</i> ² :	0,16791
Multiple <i>R</i> ² adj.:	0,049039
ANOVA	
<i>F</i> :	1,4125
<i>df1, df2</i> :	4, 28
<i>p</i> :	0,25558

Рисунок 84. Окно Multiple linear regression / Statistics

Во вкладке «Numbers» для константы «Constant» и каждого показателя представлены параметры уравнения регрессии:

- *Coeff.* – коэффициенты,
- *Std.err.* – стандартные ошибки,
- *t* – критерии Стьюдента,
- *p* – уровни значимости,

- R^2 – 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	
Зрит. 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

Рисунок 85. Окно Multiple linear regression / Numbers

В данном примере формула регрессионной функции будет: «Средний балл = 2.863 + 0.028 × Зрительное восприятие + 0.019 × Числовой интеллект – 0.028 × Вербальный интеллект + 0.105 × 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 дистримического типа личности) и столбец переменной Y (например, средний балл учебной успеваемости). Результат (Рисунок 86) выводится в окно «Polynomial fit» (Полиномиальная подгонка).

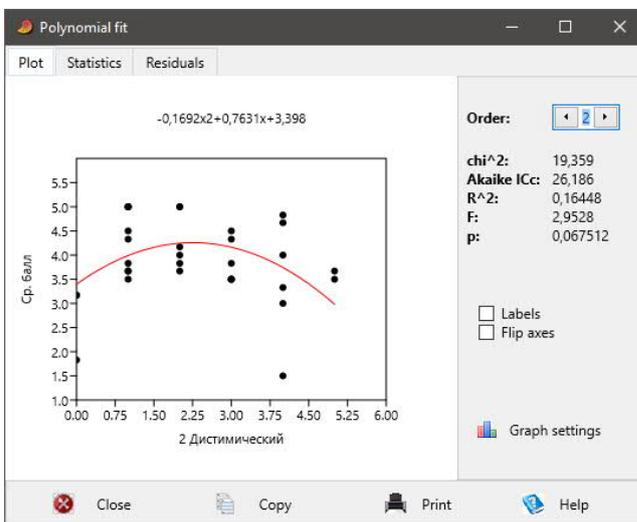


Рисунок 86. Окно Polynomial fit

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:

- chi^2 (Chi 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: «Средний балл = $0.169 \times \text{Дистимический тип}^2 + 0.763 \times \text{Дистимический тип} + 3.398$ ».

2.8. Другие пункты меню

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 не применимы, 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 других естественных науках:

Diversity (Анализ разнообразия),

Timeseries (Анализ временных рядов),

Geometry (Пространственный анализ),

Stratigraphy (Анализ геологических пластов)

– для обработки психологических данных не применяются.

Литература

1. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. Издание 2. – СПб.: Питер, 2003. – 688 с.
2. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 496 с.
3. Кедич С.И. Практикум по статистическим методам в психологии. Учебное пособие. – СПб.: ФГБОУ ВО ПГУПС, 2016. – 54 с.
4. Лакин Г.Ф. Биометрия. Издание 4. – М.: Высшая школа, 1990. – 350 с.
5. Наследов А. SPSS: Компьютерный анализ данных в психологии и социальных науках. – СПб.: Питер, 2005. – 416 с.
6. Некрасов С. Д. Математические методы в психологии (MS Excel): учебное пособие. – Краснодар: Кубанский государственный университет, 2014. – 147 с.
7. Сидоренко Е.В. Методы математической обработки в психологии. – СПб., Речь, 2003. – 350 с.