

**Санкт-Петербургское государственное бюджетное
профессиональное образовательное учреждение
«Училище олимпийского резерва № 1»**

**УТВЕРЖДАЮ
ДИРЕКТОР СПб ГБПОУ «УОР № 1»**

_____ **В.А. КУЗНЕЦОВ**

«31» августа 2022 г.

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ
ПРАКТИЧЕСКИХ ЗАДАНИЙ МДК. 03.01. ТЕОРЕТИЧЕСКИЕ
И ПРИКЛАДНЫЕ АСПЕКТЫ МЕТОДИЧЕСКОЙ РАБОТЫ
ПЕДАГОГА ПО ФИЗИЧЕСКОЙ КУЛЬТУРЕ И СПОРТУ
РАЗДЕЛ «КОМПЬЮТЕРНАЯ ОБРАБОТКА ДАННЫХ
ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ»**

программа подготовки специалистов среднего звена

49.02.01 Физическая культура

Санкт-Петербург

2023 год

Организация-разработчик: Санкт-Петербургское государственное бюджетное профессиональное образовательное учреждение Училище олимпийского резерва № 1.

Разработчик: Цивинская Е.В., преподаватель КОДЭИ.

Рассмотрено на заседании
предметно-цикловой комиссии дисциплин
профессионального цикла
СПб ГБПОУ «УОР № 1»

Протокол № 1 от 30 августа 2023 г.

Председатель ПЦК дисциплин профессионального
цикла– С.Н. Бекасова

Утверждено приказом СПб ГБПОУ «УОР № 1»
от 31.08.2023 № 287 «Об утверждении учебного
плана, графика учебного процесса, рабочих
программ учебных дисциплин (модулей) и
практик, фондов оценочных средств, учебно-
методических рекомендаций, рабочей программы
воспитания, календарного плана воспитательной
работы на 2023-2024 учебный год по
специальности 49.02.01 Физическая культура»

СОДЕРЖАНИЕ

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА	3
ПЕРЕЧЕНЬ ПРКТИЧЕСКИХ ЗАДАНИЙ	6
Практическое задание № 1 «Работа с программой статистической обработки данных STATGRAPHICS Plus 5.0».....	7
Практическое задание № 2 «Построение частотной таблицы».....	10
Практическое задание № 3 «Расчёт числовых характеристик выборки»	16
Практическое задание № 4 «Построение графиков частотной таблицы».....	21
Практическое задание № 5 «Расчёт точечных и интервальных оценок».....	26
Практическое задание № 6 «Проверка гипотезы о нормальном распределении»	28
Практическое задание № 7 «Сравнение независимых выборок»	32
Практическое задание № 8 «Сравнение связанных выборок»	45
Практическое задание № 9 «Решение регрессионных задач в спорте»	48
Практическое задание № 10 «Расчёт коэффициентов корреляции».....	56
ПРИЛОЖЕНИЯ	62
<i>Приложение 1 Критические значения F-критерия Фишера</i>	62
<i>Приложение 2 Критические значения t-критерия Стьюдента</i>	63
<i>Приложение 3 Критические значения W-критерия Вилкоксона для независимых выборок</i>	64
<i>Приложение 4 Критические значения U-критерия Манна-Уитни для независимых выборок</i>	65
<i>Приложение 5 Критические значения W-критерия Вилкоксона для сопряженных пар</i> ..	66
<i>Приложение 6 Критические значения коэффициентов корреляции r-Пирсона (r-Спирмена)</i>	67
ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ УЧЕБНЫХ ИЗДАНИЙ, ИНТЕРНЕТ-РЕСУРСОВ, ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ	68

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Методические рекомендации по выполнению практических заданий по разделу Компьютерная обработка данных экспериментальных исследований Междисциплинарного курса 03.01 Теоретические и прикладные аспекты методической работы педагога по физической культуре и спорту Профессионального модуля 03 Методическое обеспечение организации физкультурной и спортивной деятельности предназначены для студентов 2 курса на базе среднего (полного) общего образования и студентов 3 курса на базе основного общего образования, обучающихся по программе среднего профессионального образования по специальности 49.02.01 Физическая культура.

Целью методических рекомендаций является повышение эффективности учебного процесса, а также оказание помощи учащимся в выполнении практических заданий.

Выполнение практических заданий является неотъемлемым этапом изучения дисциплины. Практическое задание выполняется студентом самостоятельно или с помощью преподавателя во время учебного процесса согласно календарно-тематическому плану на основании нормативных документов, методических указаний, полученных теоретических знаний и умений.

В соответствии с учебным планом на практических занятиях отводится 38 часов.

Методические рекомендации составлены для формирования практического опыта, умений и навыков по следующим темам:

- Основные понятия математической статистики и теории вероятности;
- Работа с программой статистической обработки данных STATGRAPHICS Plus 5.0;
- Первичная обработка результатов эксперимента;
- Проверка правдоподобия статистических гипотез;
- Регрессионный и корреляционный анализ.

Основные этапы выполнения практических заданий:

- Изучение темы и цели практического занятия;
- Изучение или повторение теоретической части к практическому заданию;
- Выполнение расчетной части (если она предусмотрена);
- Оформление графиков, диаграмм (если они предусмотрены);
- Формулирование статистических выводов;
- Выполнение и оформление практического задания (в электронном виде, согласно заданию).

Контроль выполнения практических заданий осуществляется во время проведения аудиторных занятий, после ее предоставления в письменном и электронном видах, оформленных согласно заданию. Критериями оценивания результатов выполнения практического задания студентами являются:

- уровень освоения студентом учебного материала;
- умение студента использовать теоретические знания при выполнении практических заданий;
- обоснованность и четкость представления ответов и выводов;
- оформление материала в соответствии с требованиями.

Оценка "5" ставится в следующем случае:

- работа выполнена полностью;
- при решении задач сделан перевод единиц всех физических величин в "СИ", все необходимые данные занесены в условие, правильно выполнены чертежи, схемы, графики, рисунки, сопутствующие решению задач, сделана проверка по наименованиям, правильно записаны исходные формулы, записана формула для конечного расчета, проведены математические расчеты и дан полный ответ;

- на качественные и теоретические вопросы дан полный, исчерпывающий ответ литературным языком с соблюдением технической терминологии в определенной

логической последовательности, учащийся приводит новые примеры, устанавливает связь между изучаемым и ранее изученным материалом, а также с материалом, усвоенным при изучении других предметов, умеет применить знания в новой ситуации;

- учащийся обнаруживает верное понимание физической сущности рассматриваемых явлений и закономерностей, законов и теорий, дает точное определение и истолкование основных понятий, законов, теорий, а также правильное определение физических величин, их единиц и способов измерения.

Оценка "4" ставится в следующем случае:

- работа выполнена полностью или не менее чем на 80 % от объема задания, но в ней имеются недочеты и несущественные ошибки: правильно записаны исходные формулы, но не записана формула для конечного расчета; ответ приведен в других единицах измерения.

- ответ на качественные и теоретические вопросы удовлетворяет вышеперечисленным требованиям, но содержит неточности в изложении фактов, определений, понятий, объяснении взаимосвязей, выводах и решении задач;

- учащийся испытывает трудности в применении знаний в новой ситуации, не в достаточной мере использует связи с ранее изученным материалом и с материалом, усвоенным при изучении других предметов.

Оценка "3" ставится в следующем случае:

- работа выполнена в основном верно (объем выполненной части составляет не менее 2/3 от общего объема), но допущены существенные неточности; пропущены промежуточные расчеты.

- учащийся обнаруживает понимание учебного материала при недостаточной полноте усвоения понятий и закономерностей;

- умеет применять полученные знания при решении простых задач с использованием готовых формул, но затрудняется при решении качественных задач и сложных количественных задач, требующих преобразования формул.

Оценка "2" ставится в следующем случае:

- работа в основном не выполнена (объем выполненной части менее 2/3 от общего объема задания);

- учащийся показывает незнание основных понятий, непонимание изученных закономерностей и взаимосвязей, не умеет решать количественные и качественные задачи.

Оценка "1" ставится в следующем случае: работа полностью не выполнена.

Выполнение практических заданий помогут студентам сформировать следующие общие и профессиональные компетенции:

ОК 1. Понимать сущность и социальную значимость своей будущей профессии, проявлять к ней устойчивый интерес.

ОК 2. Организовывать собственную деятельность, определять методы решения профессиональных задач, оценивать их эффективность и качество.

ОК 4. Осуществлять поиск, анализ и оценку информации, необходимой для постановки и решения профессиональных задач, профессионального и личностного развития.

ОК 5. Использовать информационно-коммуникационные технологии для совершенствования профессиональной деятельности.

ПК 1.5. Анализировать учебно-тренировочные занятия, процесс и результаты руководства соревновательной деятельностью.

ПК 3.4. Оформлять методические разработки в виде отчетов, рефератов, выступлений.

ПК 3.5. Участвовать в исследовательской и проектной деятельности в области образования, физической культуры и спорта.

Выполнение практических заданий также помогут студентам получить следующий практический опыт и умения, а также закрепить теоретические знания:

Умения:

У.5. определять цели, задачи, планировать учебно-исследовательскую работу с помощью руководителя;

У.8. оформлять результаты исследовательской работы;

У.9. готовить и оформлять отчеты, рефераты, конспекты.

Знания:

З.4. основы организации опытно-экспериментальной работы в сфере физической культуры и спорта;

З.12. статистические методы обработки результатов исследований.

ПЕРЕЧЕНЬ ПРАКТИЧЕСКИХ ЗАДАНИЙ

№ п/п	Тема работы
1	Работа с программой статистической обработки данных STATGRAPHICS Plus 5.0
2	Построение частотной таблицы
3	Расчёт числовых характеристик выборки
4	Построение графиков частотной таблицы
5	Расчёт доверительного интервала
6	Проверка гипотезы о нормальном распределении
7	Сравнение независимых выборок
9	Сравнение связанных выборок
10	Решение регрессионных задач в спорте
11	Расчёт коэффициентов корреляции

А

Практическое задание № 1

«Работа с программой статистической обработки данных STATGRAPHICS Plus 5.0»
по теме «Работа с программой статистической обработки данных».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: выработать основные навыки работы с программы статистической обработки данных STATGRAPHICS Plus 5.0.

Теоретические сведения к практическому занятию:

STATGRAPHICS Plus – программа для решения задач статистического анализа.

Панели и инструменты пакета STATGRAPHICS Plus 5.0 (рис.1):

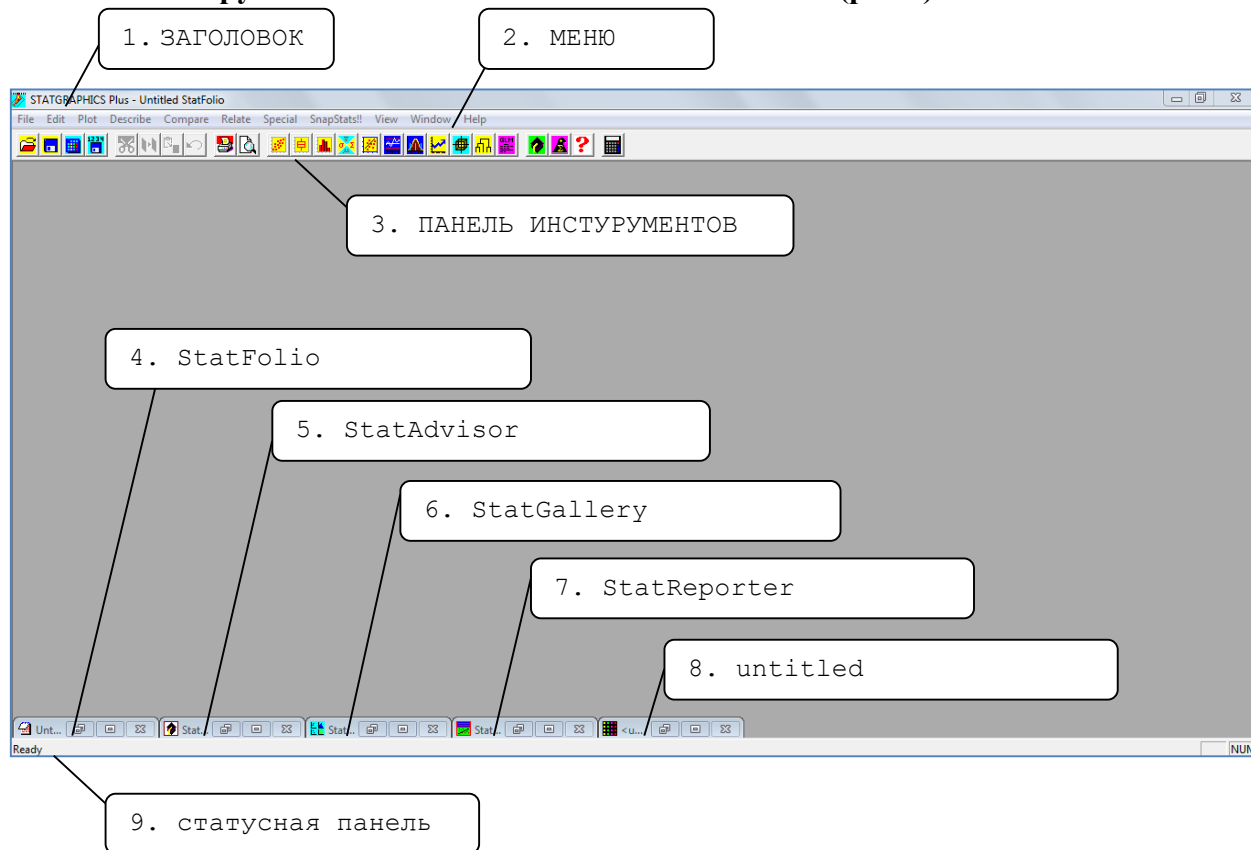


Рис.1. Интерфейс статистического пакета STATGRAPHICS Plus 5.0

Назначение инструментов пакета STATGRAPHICS Plus:

1. Заголовок – названия открытого приложения и документа.

2. Меню - элемент интерфейса пользователя, позволяющий выбрать одну из нескольких перечисленных опций программы.

3. Панель инструментов - элемент графического интерфейса пользователя, предназначенный для размещения на нём нескольких других элементов (наиболее часто используемых команд меню).

4. StatFolio – сохранение не только результатов статистического анализа, но и выбранных методов, параметров статистических процедур, видов графических отображений результатов анализа, табличных форм, комментариев.

5. StatAdvisor – встроенная экспертная система для интерпретации результатов, полученных в результате статистического анализа и формулирования статистических выводов.

6. StatGallery – инструмент комбинирования текста и графики

7. StatReporter – инструмент создания отчетов

8.untitled – окно электронной таблицы для создания нового файла исходных данных, а также модификации существующих.

Основные функции меню статистического пакета STATGRAPHICS Plus 5.0 (рис.2):

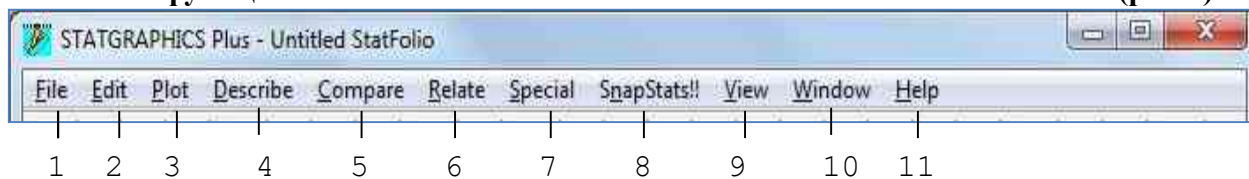


Рис.2. Меню статистического пакета STATGRAPHICS Plus 5.0

- 1.
2. File (Файл) – открытие, закрытие, создание файлов, предварительный просмотр и печать, настройка параметров страницы.
3. Edit (Правка) – редактирование содержимого файла.
4. Plot (График) – строить различные графики.
5. Describe (Описание данных) – статистические методы анализа по одной и множеству переменных, процедуры подбора распределений и кросстабуляции данных.
6. Compare (Сравнение данных) – методы сравнения двух и более выборок данных, процедуры одно- и многофакторного дисперсионного анализа.
7. Relate (Отношение данных) – выполнение простого (Simple Regression), полиномиального (Polynomial Regression) и множественного (Multiple Regression) регрессионные анализы.
8. Special (Специальный) – использование дополнительных модулей.
9. SnapStats!! (Горячие команды) – наиболее часто используемые команды.
10. View (Вид) – настройка главного окна программы.
11. Window (Окно) – расположение окон.
12. Help (Помощь) – вызов справочной системы.

Работа с электронной таблицей.

Окно электронной таблицы Untitled находится в свернутом положении на панели задач. В развернутом виде оно состоит из столбцов, имеющих названия Col_1, Col_2 и т.д. и строк, обозначенных порядковыми числительными: 1, 2, 3 и т.д. На пересечении столбца и строки находится ячейка.

Для ввода числа в ячейку его необходимо набрать на числовой клавиатуре и нажать клавишу Enter. Если число в виде десятичной дроби, то между целой и дробной частями ставится символ «,», в противном случае на экране будет выдано сообщение «Invalid entry in numeric cell (неправильный ввод в числовую ячейку)».

Для выделения ячейки необходимо установить в ней курсор.

Для выделения столбца необходимо щелкнуть левой кнопкой мыши на его заголовке (например, Col_1), столбец выделится черным цветом.

Для выделения строки необходимо щелкнуть на номере строки (например, 5), строка выделится черным цветом.

Для выделения блока ячеек необходимо установить курсор мыши в левый верхний угол выделяемого блока и, не отпуская кнопку мыши, перетащить указатель в правый нижний угол выделяемого блока. При этом блок выделится черным цветом.

Для копирования необходимого элемента таблицы (ячейки, строки, столбца) необходимо его выделить. Затем щелкнуть правой кнопкой мыши для вызова контекстного меню и выбрать команду Copy.

Для вставки скопированного элемента электронной таблицы необходимо установить курсор в ячейку, вызвать контекстное меню и выбрать команду Paste или Insert.

Для удаления выделенного элемента таблицы выбрать в контекстном меню команду Delete.

Для переименования столбца, написания комментариев, изменения его ширины или изменения типа вводимых данных необходимо выделить столбец, вызвать контекстное меню и выбрать команду Modify Column.

Для ввода формул необходимо выделить столбец, вызвать контекстное меню и выбрать команду Modify Column и щелкнуть на кнопке Formula (ввод формул) и затем на кнопку Define (определить). Появится диалоговое окно Generate Data (модификация данных). В поле Expression (выражение) диалогового окна вводится формула (рис.3):

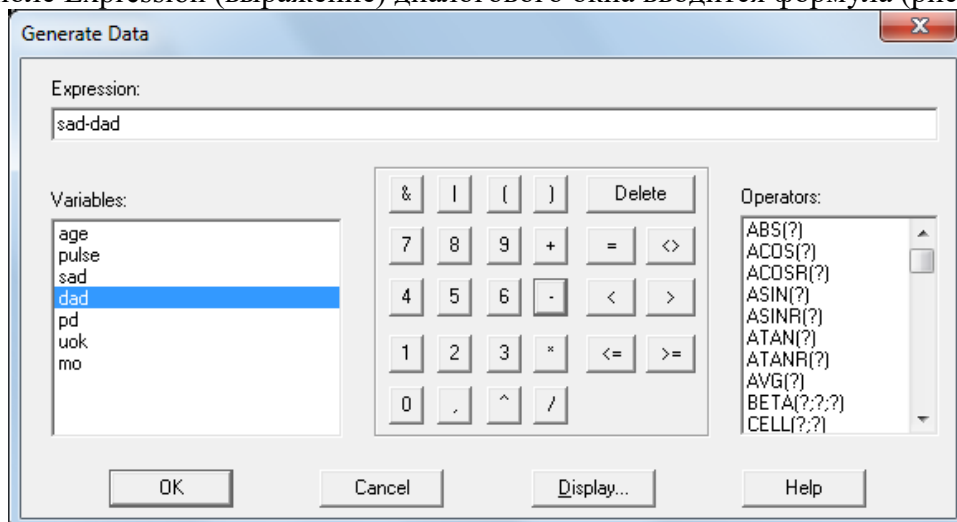


Рис.3. Диалоговое окно Generate Data

После окончания ввода формулы необходимо нажать кнопку ОК и произойдет автоматический расчёт данных и занесение их в предварительно выделенный столбец. Рассчитанные данные отображаются в электронной таблице серым цветом.

Сохранение файла данных.

Для сохранения файла или электронной таблицы в меню File выбирается команда Save As / Save Data File As. В появившемся диалоговом окне Save Data File As в поле Имя Файла внести название файла и нажать кнопку Сохранить.

Для сохранения статистического проекта (не только электронной таблицы, но и всех действий с исходными данными. Например, построение частотной таблицы, расчёт числовых характеристик, построение графиков) в меню File выбирается команда Save As / Save StatFolio As. В появившемся диалоговом окне Save StatFolio As в поле Имя Файла внести название файла и нажать кнопку Сохранить.

Форма выполнения практического занятия: решение профессиональной задачи на компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0

Задача: В исследованиях у испытуемых измерялись следующие показатели: возраст (age), ЧСС (pulse), систолическое артериальное давление (sad) и диастолическое артериальное давление (dad). Согласно представленным данным вычислить пульсовое давление (pd), ударный объем крови (uok) и минутный объем (mo).

Таблица 1: Данные измерений экспериментальных исследований.

age	pulse	sad	dad
14	115	170	70
18	111	190	70
15	94	150	70
19	115	130	60
18	115	150	50
18	115	140	70
17	100	180	65

age	pulse	sad	dad
18	100	140	70
22	136	160	70

Ход выполнения практического занятия:

Математическое решение задачи:

Формулы используемые в задаче:

$pd = sad - dad$ (пульсовое давление)

$uok = 100 + 0,5 \cdot pd - 0,6 \cdot dad - 0,6 \cdot age$ (ударный объем крови)

$mo = uok \cdot pulse$ (минутный объем)

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. Запустить программу STATGRAPHICS Plus.
2. Открыть окно электронной таблицы untitled.
3. Переименовать столбцы Col_1, Col_2, Col_3, Col_4, Col_5, Col_6, Col_7 в age, pulse, sad, dad, pd, uok, mo соответственно.
4. Если необходимо, ввести в поле Comment (комментарий) описания переменной (можно на русском языке), а в поле Width (ширина столбца) установить требуемую ширину столбца (в количестве символов).
5. В созданную таблицу ввести исходные данные (Таблицы 1).
6. В столбцы pd, uok и mo ввести формулы расчёта пульсового давления, ударного объема крови и минутного объема соответственно.
7. Добавьте в таблицу строку с вашими данными (ваш возраст, пульс и артериальное давление) и сделайте перерасчет результатов в таблице

Примечание: если изменить исходные данные, а также добавить новые, автоматического пересчета не происходит. Чтобы повторить пересчет нужно:

- Выделить столбец, например, pd
- Вызвать контекстное меню
- Выбрать команду Update Formulas/ произойдет пересчет всех столбцов: pd, uok и mo.

8. Сохранить файл с данными в папку со своей фамилией и номером группы на рабочем столе с именем «PR_1».

Практическое задание № 2

«Построение частотной таблицы»

по теме «Первичная обработка результатов эксперимента».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение проводить первичную обработку результатов исследования и предоставлять полученный материал в виде частотной таблицы.

Теоретические сведения к практическому занятию:

Формы представления статистического материала.

Полученные данные в ходе экспериментальной работы представлены в виде неупорядоченного набора чисел. Этот набор данных нуждается в обработке, которая всегда начинается с их группировки.

Группировка – процесс систематизации, или упорядочения, первичных данных с целью извлечения содержащейся в них информации. Группировка выполняется различными методами в зависимости от целей исследования, вида изучаемого признака и объема выборки, но наиболее часто группировка сводится к представлению данных в виде статистических таблиц.

Группировка производится для того, чтобы построить эмпирическое распределение и сформировать с его помощью предположения о форме распределения изучаемого признака в генеральной совокупности, из которого взята выборка.

Группировка заключается в построении вариационного ряда.

Вариационный ряд – двойной числовой ряд, показывающий, каким образом численные значения изучаемого признака связаны с их повторяемостью в выборке.

Вариационные ряды бывают интервальными и безынтервальными.

В интервальном вариационном ряду частоты (или частости), характеризующие повторяемость вариант в выборке, распределяются по интервалам группировки. Данный ряд строится, если изучаемый признак варьирует непрерывно, но используется и для дискретно варьирующихся признаков в тех случаях, когда признак варьирует в широких пределах.

В безынтервальном вариационном ряду частоты (или частости) распределяются непосредственно по значениям варьирующего признака. Данный ряд применяется в тех случаях, когда исследуемый признак варьирует дискретно и слабо.

Пример: Двенадцать стрелков, выполняя упражнение лёжа, из 10 выстрелов показали следующие результаты (очки): 94, 91, 96, 94, 94, 92, 91, 92, 91, 95, 94, 94. Произвести первичную обработку данных (построить частотную таблицу (безынтервальный вариационный ряд)).

Таблица 2. Частотная таблица

№ варианты	x_i	n_i	p_i^*

Решение: Переформируем ряд, данный в условии задачи, записав полученные значения x_i в порядке возрастания с абсолютными частотами n_i (числа, показывающие, сколько раз варианты, относящиеся к каждому интервалу группировки, встречаются в выборке), а потом с относительными частотами p_i^* (отношение абсолютной частоты n_i к общей серии испытаний n) (см. таблица 3).

Таблица 3. Частотная таблица (безынтервальный вариационный ряд)

№ варианты	x_i	n_i	p_i^*
1	91	3	0,250
2	92	2	0,167
3	94	5	0,417
4	95	1	0,083
5	96	1	0,083
		$n = \sum_{i=1}^5 n_i = 12$	$\sum_{i=1}^5 p_i^* = 1$

Таким образом, нами получены безынтервальные (дискретные) вариационные ряды с абсолютными и относительными частотами. Эти дискретные вариационные ряды называют – эмпирическим распределением.

В случае непрерывных случайных величин и большого количества испытаний предпочтительным является интервальный (сгруппированный) вариационный ряд. Он представляется в виде таблицы, у которой в столбце вариант ставятся промежутки (диапазоны) значений измеряемого признака.

Алгоритм процесса группировки для построения интервального (сгруппированного) вариационного ряда:

1. Нахождение числа группировки l по формуле Стерджеса $l = 1 + 3,32 \cdot \lg n$, где n – объём выборки. Результат округляют до целого числа и всегда в большую сторону.

Также для вычисления числа интервалов используют таблицу 4:

Таблица 4. Нахождение количества интервалов для построения интервального вариационного ряда.

Объем выборки, n	Число интервалов, l
25-40	5-6
40-60	6-8
60-100	7-10
100-200	8-12
>200	10-15

2. Определение размаха варьирования (R): $R = x_{max} - x_{min}$

3. Нахождение ширины интервалов группировки (h): $h = \frac{R}{l}$

Для удобства расчётов данную величину округляют, но всегда в большую сторону.

4. Построить интервал ($H_{low}; H_{upp}$): $\Delta_1 + \Delta_2 = l \cdot h - R$, где Δ_1 и Δ_2 - положительные, равные или неравные друг другу.

$$H_{low} = x_{min} - \Delta_1$$

$$H_{upp} = x_{max} + \Delta_2$$

5. Частота интервалов (n_i) – числа, показывающие, сколько раз варианты, относящиеся к каждому интервалу группировки, встречаются в выборке. общая сумма всех частот всегда равна объему выборки n .

6. Накопленная частота интервала (n_{x_i}) – число, полученное последовательным суммированием частот в направлении от первого интервала к последнему, до того интервала включительно, для которого определяется накопленная частота.

7. Частость (относительная частота) (f_i): $f_i = \frac{n_i}{n}$

8. Накопленная частость (F_i): $F_i = \frac{n_{x_i}}{n}$, Сумма всех частостей всегда равна 1.

Форма выполнения практического занятия: решение профессиональной задачи:

1. Письменно в тетради рассчитать количество интервалов группировки l по формуле Стерджеса, размах варьирования R , ширину интервала h , интервал ($H_{low}; H_{upp}$);

2. Построить интервальный вариационный ряд (частотную таблицу) на компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0.

Задача: В таблице 5 представлены результаты бега на 100 м в секундах (юноши, 9 класс). Произвести первичную обработку данных (построить интервальный вариационный ряд (частотную таблицу)).

Таблица 5. Результаты бега на 100 м в секундах (юноши, 9 класс).

16,2	15,4	15,5	14,7	14,3	12,8	16,6	16,9	15,8	15,0
15,4	16,8	14,5	16,0	14,8	14,6	16,1	15,6	15,8	16,1
15,3	17,8	16,0	15,6	13,7	15,0	16,1	15,6	16,2	15,0
15,3	16,2	15,5	15,5	14,8	13,6	14,2	16,4	16,2	15,2
15,3	15,9	15,8	15,0	14,2	14,2	15,8	16,4	14,2	14,2

Ход выполнения практического занятия:

Математическое решение задачи (письменно в тетради):

1. Объем выборки $n=50$.

2. Количество интервалов группировки: $l = 1 + 3,32 \cdot \lg n = 1 + 3,32 \cdot \lg 50 = 6,64 \approx 7$

3. Размах варьирования: $R = x_{max} - x_{min} = 17,8 - 12,8 = 5,0$ секунд.

4. Ширина интервала группировки: $h = \frac{R}{l} = \frac{5,0}{7} = 0,714 \approx 0,8$ секунд.

5. Интервал вариационного ряда:

$$\Delta_1 + \Delta_2 = l \cdot h - R = 7 \cdot 0,8 - 5,0 = 0,6 \text{ секунд.}$$

Пусть $\Delta_1 = 0,3$ секунды и $\Delta_2 = 0,3$ секунды.

$$H_{low} = x_{min} - \Delta_1 = 12,8 - 0,3 = 12,5 \text{ секунд}$$

$$H_{upp} = x_{max} + \Delta_2 = 17,8 + 0,3 = 18,0 \text{ секунд}$$

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. Запустить программу STATGRAPHICS Plus и создать файл с исходными данными beg100. Название переменной-x1.

2. Выбирать в меню пункты: Describe/Numeric Data/One Variable Analysis. Появится диалоговое окно One-Variable Analysis (рис.4).

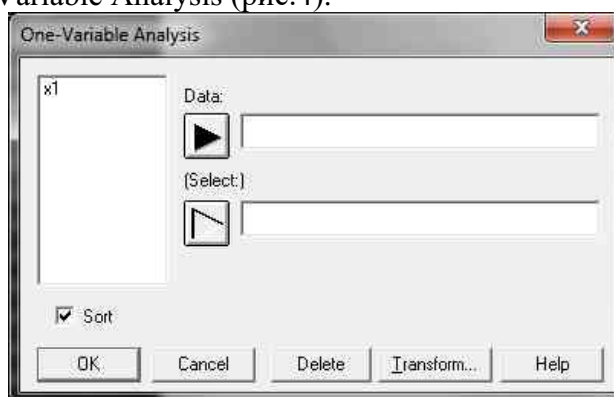


Рис. 4. Диалоговое окно One-Variable Analysis

Перенести в поле Data имя переменной (x1). Для этого щелкнуть на ее имени (переменная изменит цвет), затем нажать кнопку со стрелкой, при этом выделенная переменная перейдет в поле Data, нажать кнопку ОК. В окне анализа (One-Variable Analysis) появится текстовое окно общего анализа (Analysis Summary) (рис.5).

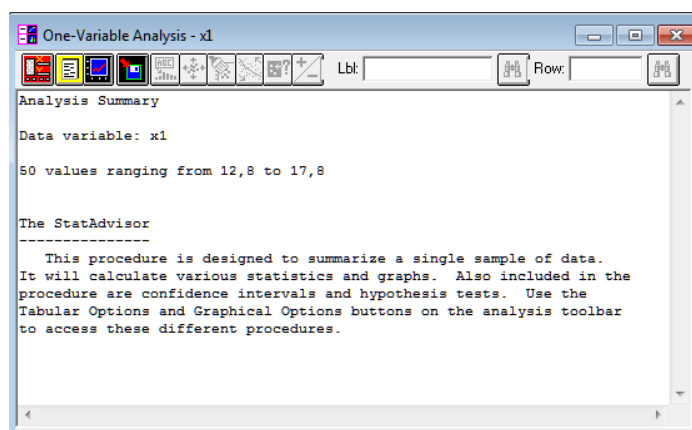


Рис.5. Окно Анализа

В этом окне содержится информация о переменной, об объеме выборки а также о минимальном и максимальном значениях выборки.

Например, 50 values ranging from 12,8 to 17,8, означает что, 50 вариант изменяются от 12,8 до 17,8; т.е. минимальное значение равно 12,8, а максимальное значение равно 17,8; объем выборки равен 50.

Затем нажать кнопку табличных опций Tabular options – появится одноименное диалоговое окно (рис.6), снять флажок с окна Analysis Summary и выбрать Frequency Tabulation (частотная таблица), после чего нажать кнопку ОК.

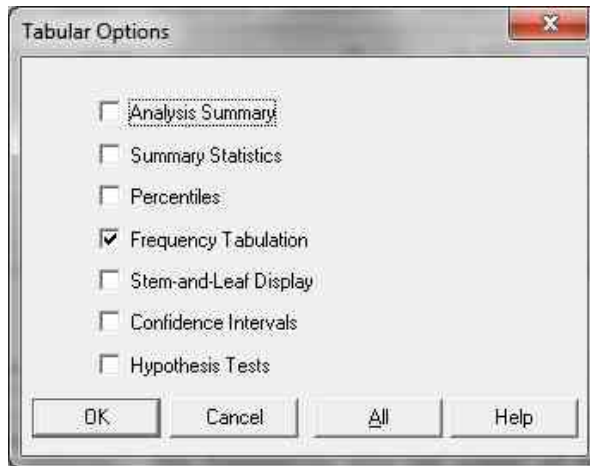


Рис. 6. Диалоговое окно Tabular Options

3. Получаем частотную таблицу (рис.7), которую STATGRAPHICS Plus 5.0 выдает по умолчанию:

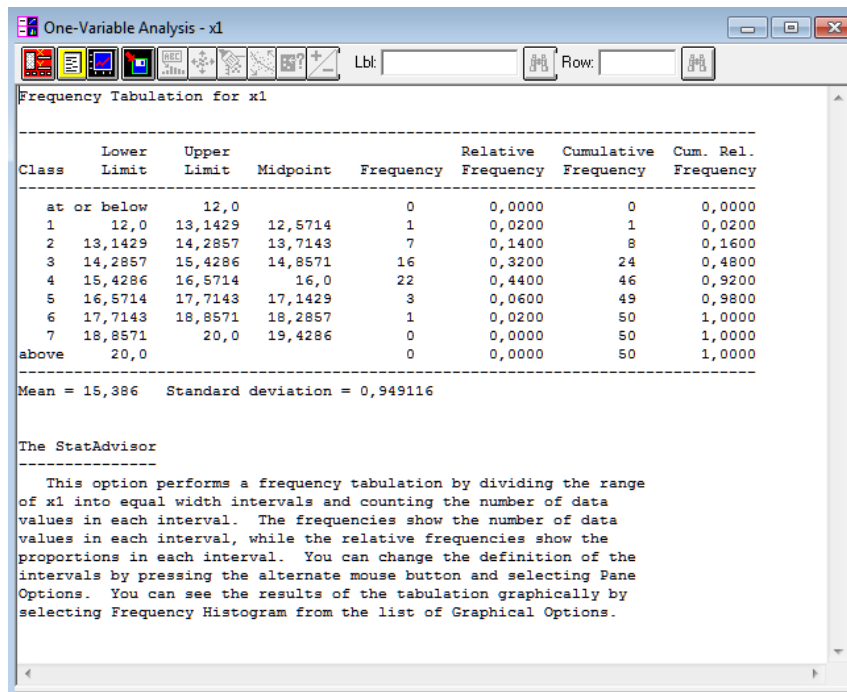


Рис. 7. Частотная таблица, полученная по умолчанию

В этой таблице указаны следующие параметры:

- Class* – номер интервала группировки
- Lower Limit* – нижняя граница интервала группировки
- Upper Limit* – верхняя граница интервала группировки
- Midpoint* – срединное значение интервала группировки
- Frequency* – частота
- Relative frequency* – относительная частота
- Cumulative frequency* – накопленная частота
- Cum. Rel. Frequency* – накопленная относительная частота

Под частотной таблицей указывается среднее арифметическое (*Mean*) и стандартное отклонение (*Standard deviation*).

В нашем примере среднее арифметическое равно 15,382 с, а стандартное отклонение – 0,946031 с.

4. Изменяем границы интервалов группировки в соответствии с результатами, полученными при расчете интервального (сгруппированного) вариационного ряда.

Группировка, выполняемая STATGRAPHICS Plus по умолчанию не очень рациональна, т.к. при ее выполнении в последний интервал не попадает ни одной варианты.

Для этого щелкнуть на частотной таблице, а затем выполнить щелчок правой кнопкой мыши. В появившемся контекстном меню выбрать Pane Option (опции окна). После этого появится диалоговое окно Frequency Tabulation Options (опции окна частотной таблицы) (рис.8).

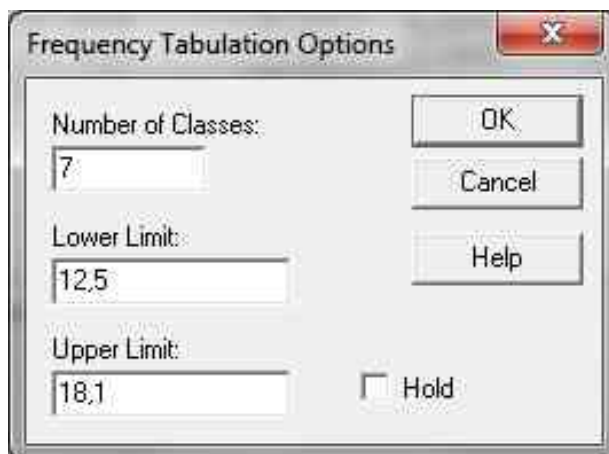


Рис. 8. Диалоговое окно Frequency Tabulation Options

В поле Number of Classes (число классов) нужно указываем число классов группировки. В нашем примере их $l=7$. В поле Lower Limit (нижняя граница первого интервала) необходимо указать $N_{low}=12,5$ с; в поле Upper Limit (верхняя граница последнего интервала), необходимо указать $N_{upp}=18,1$ с. Нажать кнопку ОК. После выполненных действий частотная таблица примет вид (рис.9).

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below	12,5			0	0,0000	0	0,0000
1	12,5	13,3	12,9	1	0,0200	1	0,0200
2	13,3	14,1	13,7	2	0,0400	3	0,0600
3	14,1	14,9	14,5	11	0,2200	14	0,2800
4	14,9	15,7	15,3	16	0,3200	30	0,6000
5	15,7	16,5	16,1	16	0,3200	46	0,9200
6	16,5	17,3	16,9	3	0,0600	49	0,9800
7	17,3	18,1	17,7	1	0,0200	50	1,0000
above	18,1			0	0,0000	50	1,0000

Mean = 15,386 Standard deviation = 0,949116

The StatAdvisor

 This option performs a frequency tabulation by dividing the range of x1 into equal width intervals and counting the number of data values in each interval. The frequencies show the number of data values in each interval, while the relative frequencies show the proportions in each interval. You can change the definition of the intervals by pressing the alternate mouse button and selecting Pane Options. You can see the results of the tabulation graphically by selecting Frequency Histogram from the list of Graphical Options.

Рис. 9. Частотная таблица

5. Сохранить файлы с данными и результатами статистической обработки в папку со своей фамилией и номером группы на рабочем столе с именем «Beg100»:

Практическое задание № 3

«Расчёт числовых характеристик выборки»

по теме «Первичная обработка результатов эксперимента».

Время, отведенное на выполнение практического занятия: 4 часа.

Цель: сформировать умение рассчитывать числовые характеристики выборки.

Теоретические сведения к практическому занятию:

Числовые характеристики выборки.

Числовые характеристики выборки дают количественное представление об эмпирических данных и позволяют сравнить их между собой. наибольшее значение имеют характеристики положения, рассеяния и формы.

Числовые характеристики положения.

Среднее арифметическое - сумма всех зафиксированных значений, деленная на их количество. Среднее принято обозначать той же буквой, что и варианты выборки, с той лишь разницей, что над буквой ставится символ усреднения – черта. Например, если обозначить исследуемый признак X , а его числовые значения – через x_i , то среднее арифметическое имеет обозначение \bar{x} .

Для несгруппированных данных среднее арифметическое определяется по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1)$$

где n – объем выборки, x_i – варианты выборки, i – индекс (порядковый номер).

Если данные сгруппированы, то:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i, \quad (2)$$

где n – объем выборки, k – число интервалов группировки, n_i – частоты интервалов, x_i – срединные значения интервалов, i – индекс (порядковый номер).

Мода – значение признака, встречающееся в выборке наиболее часто. Обозначается – Mo .

Медиана – полусумма средних значение ранжированного ряда. обозначение – Me .

Числовые характеристики рассеяния.

Ошибка среднего арифметического – отличие выборочного среднего от генерального среднего.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}, \quad (3)$$

Размах вариации – разность между максимальной и минимальной вариантами выборки.

$$R = x_{max} - x_{min}, \quad (4)$$

где x_{max} - максимальное значение выборки, x_{min} - минимальное значение выборки.

Дисперсия – средний квадрат отклонения значения признака от среднего арифметического (мера разброса случайной величины относительно среднего арифметического).

$$D^* = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (5)$$

Стандартное отклонение – рассеивание значений случайной величины относительно среднего арифметического.

$$\sigma^* = S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (6)$$

Размерность среднего квадратического или стандартного отклонения в отличие от размерности дисперсии совпадает с единицами измерения экспериментальных данных, поэтому на практике обычно используют S , а не S^2 .

Коэффициент вариации – сравнение наблюдаемой в выборке изменчивости (вариативности) значений признаков, измеренных в различных единицах. Величина безразмерная и измеряется в процентах.

$$V = \frac{S}{\bar{x}} \cdot 100\% \quad (7)$$

Коэффициент вариации используется для оценки однородности выборки. Если $V < 10\%$, то выборка считается однородной, т.е. полученной из одной генеральной совокупности.

Числовые характеристики формы.

Коэффициент асимметрии – безразмерная величина, мера отклонения распределения от симметрии (относительно среднего арифметического).

$$A_s = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{nS^3} \quad (8)$$

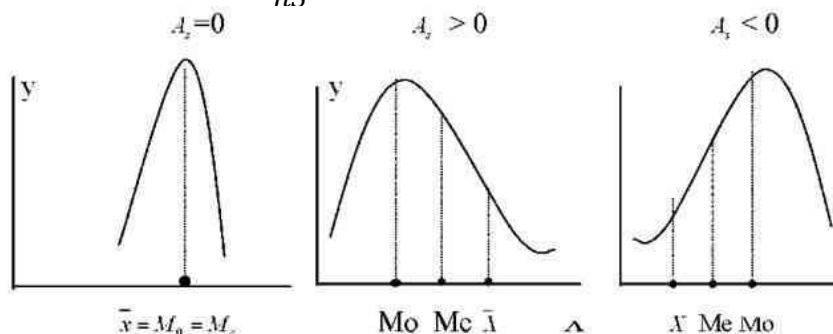


Рис.10. График асимметрии относительно среднего арифметического, моды и медианы.

Коэффициент эксцесса – мера остроты пика распределения случайной величины.

$$E_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{nS^4} - 3 \quad (9)$$



Рис.11. График эксцесса относительно среднего арифметического и моды.

Форма выполнения практического занятия: решение профессиональной задачи:

1. Письменно в тетради рассчитать следующие числовые характеристики: среднее арифметическое, дисперсию, среднеквадратичное отклонение, ошибку среднего арифметического, размах вариации, а также найти моду и медиану.

2. На компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 рассчитать все числовые характеристики выборки.

Задача: Рассчитать числовые характеристики, характеризующие техническую и функциональную подготовленность юношей-пловцов. Данные даны в таблице 6.

Таблица 6. Техническая и функциональная подготовленность юношей-пловцов ($n=13$).

№	Темп гребка/мин	Время прохождения дистанции кролем на груди			
		25 м, с	50 м, с	100 м, с	200 м, с
1	52	14,65	33,90	75,9	161,8
2	51	15,00	33,65	74,8	161,7
3	52	14,20	32,90	73,5	157,8
4	54	15,60	34,50	75,2	160,0
5	50	14,65	33,50	74,5	160,0
6	49,5	14,30	33,20	74,0	160,2
7	49	16,20	36,30	74,1	167,6
8	49,5	16,00	35,90	77,3	167,0
9	49	13,95	30,85	69,2	152,0
10	48,5	13,60	30,50	68,8	151,2
11	49,5	14,75	34,95	76,4	162,8
12	50,5	15,20	34,60	74,6	161,2
13	52,5	14,45	33,60	74,0	160,8

Ход выполнения практического занятия:

Математическое решение задачи (письменно в тетради):

Произвести расчёт числовых характеристик выборки по следующим формулам:

1. среднее арифметическое $M^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
2. медиана $P^*(Me^* \leq X) = P^*(Me^* > X)$
3. мода Mo^* – значение признака, встречающееся в выборке наиболее часто
4. дисперсия $D^* = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
5. среднеквадратическое отклонение $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
6. ошибку среднего арифметического $S_{\bar{x}} = \frac{S}{\sqrt{n}}$
7. коэффициент вариации $V = \frac{S}{\bar{x}} 100\%$

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. В пустой электронной таблице создать 5 столбцов, имеющих следующие имена:
temp – темп, гребки/мин;
kr25 – время прохождения дистанции 25 м кролем на груди, с;
kr50 – время прохождения дистанции 50 м кролем на груди, с;
kr100 – время прохождения дистанции 100 м кролем на груди, с;
kr200 – время прохождения дистанции 200 м кролем на груди, с;
2. В созданные столбцы занести данные таблицы 6 (из условия задачи).
3. Сохранить данный в файле «NumChars».
4. Для расчета числовых характеристик выборки, следует выбрать Describe/Numeric Data/Multiple-Variable Analysis. Появится диалоговое окно (рис.12).

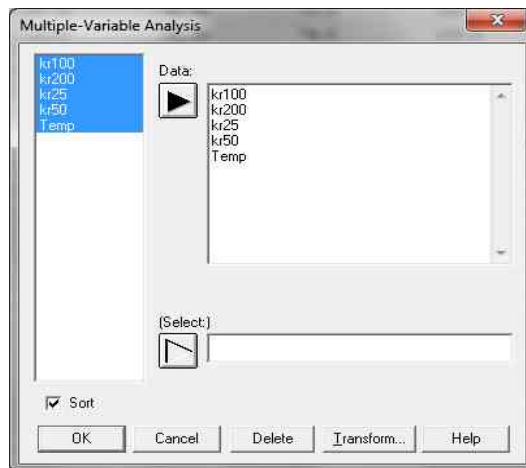


Рис. 12. Диалоговое окно Multiple-Variable Analysis

Чтобы заполнить диалоговое окно нужно выделить переменную левой кнопкой мыши и нажать на стрелку, после этого переменная появится в поле **Data**. Последовательно выполните эту процедуру со всеми переменными, которые вы хотите проанализировать и нажмите кнопку **OK**. В окне анализа (*Multiple-Variable Analysis*) появится текстовое окно общего анализа (**Analysis Summary**) (рис.13).

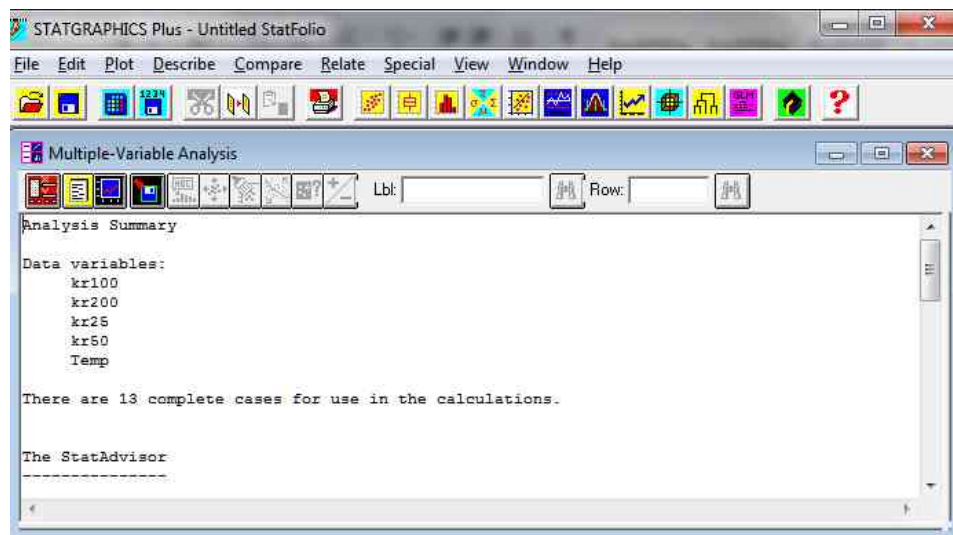


Рис. 13. Окно Анализа

В этом окне сообщаются названия переменных, а также число вариантов, относящихся к каждой переменной.

5. Для продолжения вычислений необходимо щелкнуть на кнопке табличных опций (**Tabular options**). После этого появится диалоговое окно табличных опций рис.14.

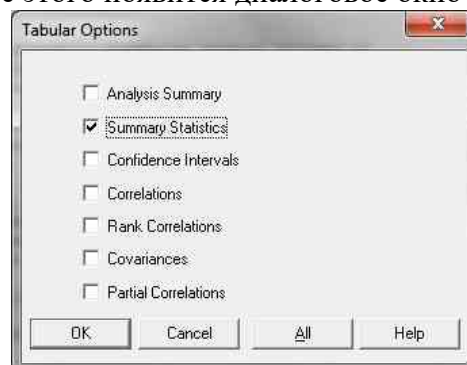


Рис.14. Диалоговое окно табличных опций (Tabular options)

6. В диалоговом окне табличных опций поставьте флажок напротив надписи **Summary Statistics** (описательная статистика), нажмите ОК. После появится окно **Summary Statistics**, в котором рассчитаны числовые характеристики выборки, заданные STATGRAPHICS Plus 5.0 по умолчанию (рис.15).

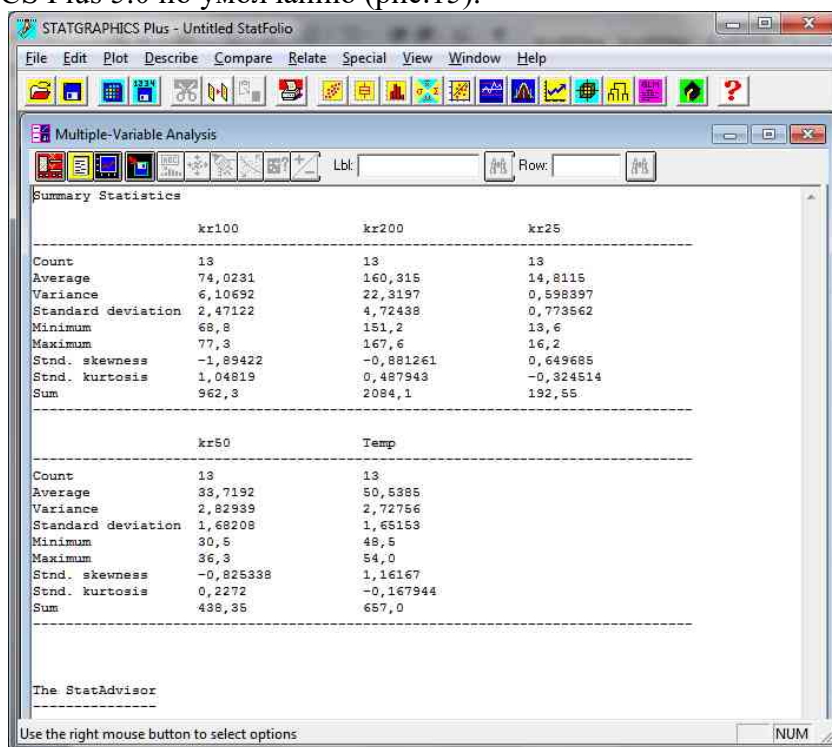


Рис.15. Окно анализа с результатами расчета числовых характеристик, заданных по умолчанию

7. Чтобы задать для анализа необходимые числовые характеристики, вызываем правым щелчком мыши контекстное меню и выбираем опцию **Pane Options**. После этого появляется диалоговое окно описательной статистики (рис.16).

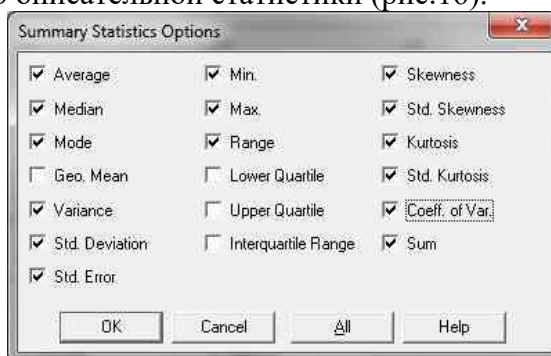


Рис.16. Опции диалогового окна описательной статистики

В данном окне перечислены числовые характеристики выборки:

Average – среднее арифметическое
Median – медиана
Mode – мода
Variance – дисперсия
Standard deviation – стандартное отклонение
Standard error – ошибка среднего арифметического
Minimum – минимальное значение признака
Maximum – максимальное значение признака
Range – размах варьирования
Skewness – коэффициент асимметрии
Stand.skewness – нормированный коэффициент асимметрии
Kurtosis – коэффициент эксцесса
Stand.kurtosis – нормированный коэффициент эксцесса
Coeff.of Variation – коэффициент вариации
Sum – сумма
Count – объем выборки

8. На рис.17 приведены результаты расчёта числовых характеристик выборки.

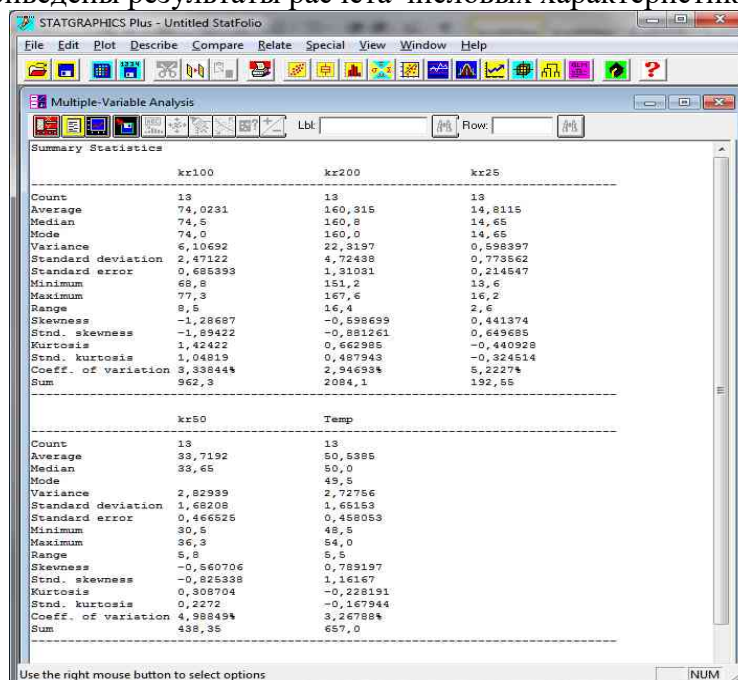


Рис.17. Результаты расчёта числовых характеристик выборки.

9. Сохранить файлы с данными и результатами статистической обработки в папку со своей фамилией и номером группы на рабочем столе с именем «NumChars».

Практическое задание № 4

«Построение графиков частотной таблицы»

по теме «Первичная обработка результатов эксперимента».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение предоставлять статистически обработанный материал в графическом виде.

Теоретические сведения к практическому занятию:

Для повышения наглядности эмпирических распределений используют их графическое представление. Наиболее распространенными способами графического

представления являются гистограмма, полигон частот и полигон накопленных частот (кумулята).

Гистограмма.

Гистограмма - способ представления статистических данных в виде столбчатой диаграммы (рис.18).

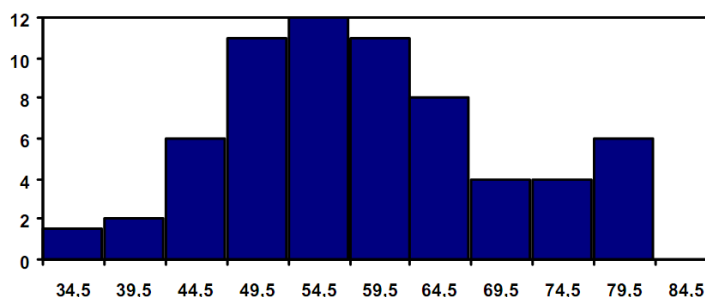


Рис.18. Гистограмма

Основание каждого прямоугольника равно ширине интервала группировки, а высота его такова, что площадь прямоугольника пропорциональна частоте (или частости) попадания в данный интервал. Таким образом, высоты прямоугольников должны быть пропорциональны величинам $p_i = \frac{n_i}{h}$, где n_i – частота i -того интервала группировки, h_i – ширина i -того интервала группировки.

Полигон частот.

Полигон частот – ломаная линия, соединяющая точки, соответствующие средним значениям интервалов группировки и частотам этих интервалов (рис.19). Средние значения откладываются по оси x , а частоты – по оси y .

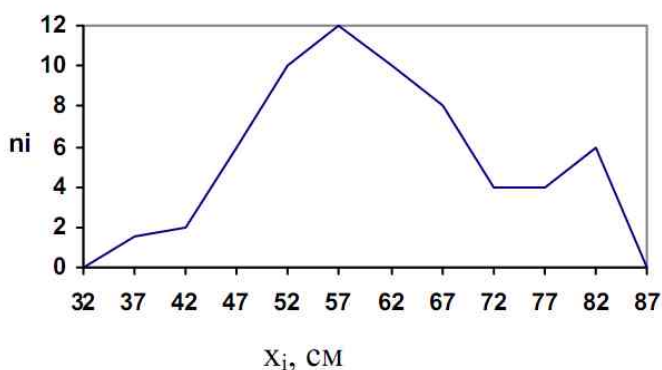


Рис.19. Полигон частот

Полигон частот используется для представления распределений как непрерывных, так и дискретных признаков. В случае непрерывного распределения полигон частот является более предпочтительным способом графического представления, чем гистограмма, если график эмпирического распределения описывается плавной зависимостью.

Полигон накопленных частот (кумулята).

Полигон накопленных частот (кумулята) получается при соединении отрезками прямых точек, координаты которых соответствуют верхним границам интервалов группировки и накопленным частотам (рис.20).

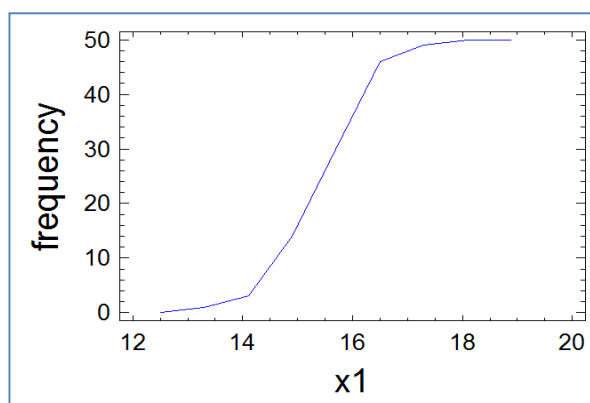


Рис.20. Полигон накопленных частот (кумулята).

Кумулята в основном используется для предоставления дискретных данных.

Форма выполнения практического занятия: решение профессиональной задачи:

1. В тетради построить следующие графики: гистограмма, полигон частот и полигон накопленных частот (кумулята);

2. На компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 построить следующие графики: гистограмма, полигон частот и полигон накопленных частот (кумулята).

Задача: Дана частотная таблица (по результатам решения задачи из практического занятия № 2).

Таблица 7. Частотная таблица.

Номер интервала группировки	Нижняя граница интервала группировки	Верхняя граница интервала группировки	Среднее значение интервала группировки	Частота	Относительная частота	Накопленная частота	Накопленная относительная частота
1	12,5	13,3	12,9	1	0,02	1	0,02
2	13,3	14,1	13,7	2	0,04	3	0,06
3	14,1	14,9	14,5	11	0,22	14	0,28
4	14,9	15,7	15,3	16	0,32	30	0,60
5	15,7	16,5	16,1	16	0,32	46	0,92
6	16,5	17,3	16,9	3	0,06	49	0,98
7	17,3	18,1	17,7	1	0,02	50	1

Представить данный обработанный статистический материал в графическом виде (построить графики гистограммы, полигона частот и кумуляты).

Ход выполнения практического занятия:

Математическое решение задачи (в тетради):

1. **Гистограмма:** в декартовой системе координат по оси x отложить значения ширины i -того интервала, а по оси y – значения частоты i -того интервала. Значения соединить в виде прямоугольных столбцов;

2. **Полигон накопленных частот:** в декартовой системе координат по оси x отложить средние значения интервала группировки, а по оси y – значения частоты i -того интервала. Значения соединить ломаной линией;

3. **Кумулята:** в декартовой системе координат по оси x отложить верхние значения интервала группировки, а по оси y – значения накопленной частоты i -того интервала. Значения соединить ломаной линией.

Решение задачи в пакете STATGRAPHICS Plus 5.0:

1. Запустить программу STATGRAPHICS Plus и открыть файл «Beg100».

2. Щелкнуть на кнопке графических опций (Graphical Options), появится соответствующее диалоговое окно (рис.21).

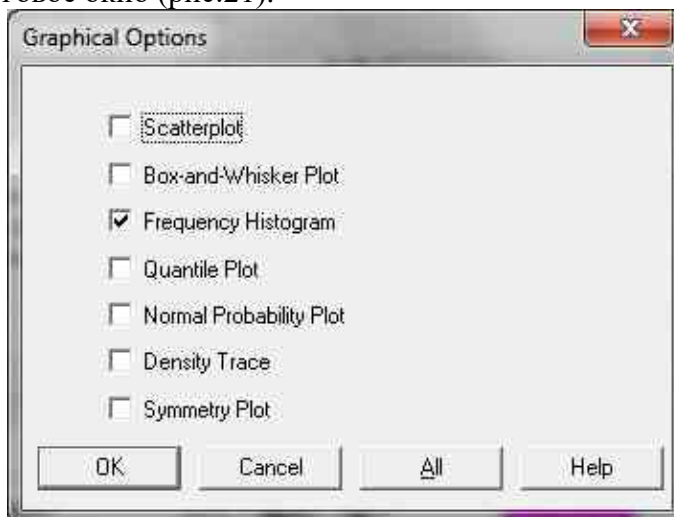


Рис. 21. Диалоговое окно Graphical Options

3. В этом окне установим флажок на поле Frequency Histogram. Получим гистограмму (рис.22). Скопируйте ее в буфер обмена.

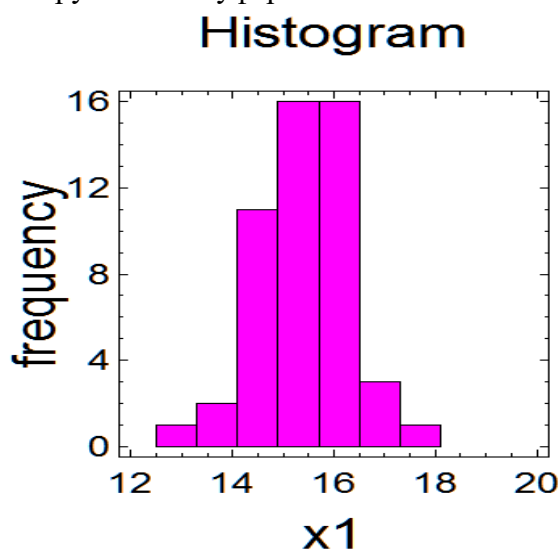


Рис. 22. Гистограмма результатов бега на 100 м

4. Открыть окно StatGallery. Вставить скопированную гистограмму в первый квадрат.

5. Чтобы построить график другого типа, необходимо вернуться в окно One Variable Analysis – x1 и щелкнуть правой клавишей мыши, в появившемся контекстном меню выбрать команду Pane Option (опции окна), появится диалоговое окно Frequency Plot Option (опции диалогового окна частотных графиков)(рис.23).

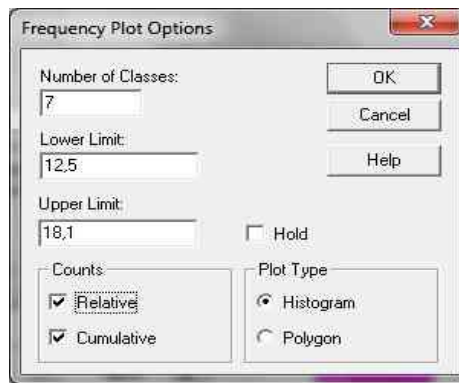


Рис.23. Диалоговое окно *Frequency Plot Options*

6. В поле Counts (частоты) отсутствие флажка у слова Relative (относительные частоты) позволит строить гистограммы, у которых по оси абсцисс будут отложены абсолютные частоты. Если флажок у слова Relative установлен, по оси абсцисс будут отложены относительные частоты в %. Наличие флажка у слова Cumulative свидетельствует о том, что вы работаете с накопленными частотами.

7. В поле Plot Type (тип графика) можно указать, какой график вы хотите построить: гистограмму (Histogram) или полигон (Polygon). Для построения полигона накопленных частот (кумуляты) необходимо поставить галочку Cumulative.

8. Щелкнуть на графике правой кнопкой мыши и в контекстном меню выбрать команду Graphics Options. В появившемся диалоговом окне установить параметры форматирования графиков по своему усмотрению.

9. Самостоятельно построить графики накопленных частот и кумуляты и скопировать в окно StatGallery.

10. Результат выполнения практического занятия должен выглядеть в окне StatGallery следующим образом (рис.24):

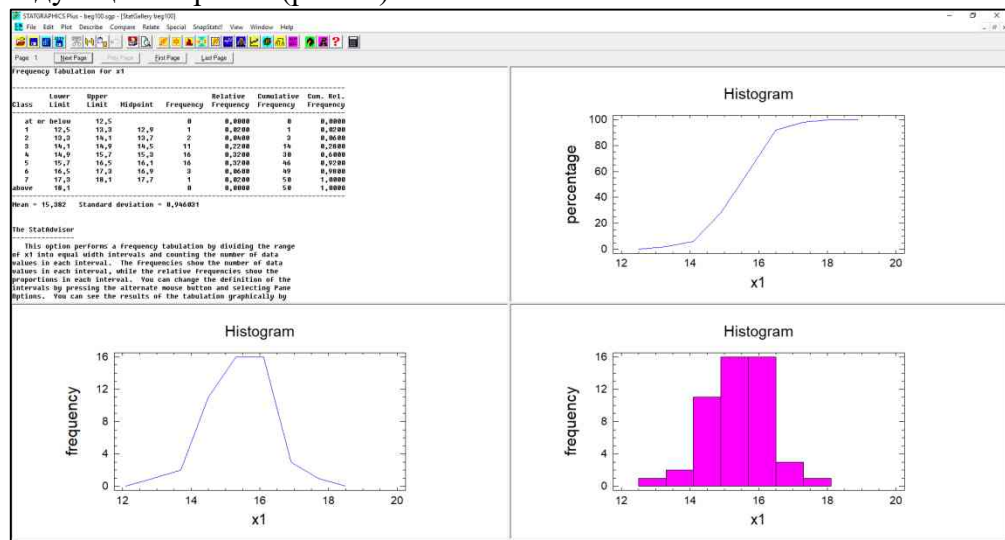


Рис. 24. Окно StatGallery. Графики: гистограмма, полигон частот и кумулята.

11. Сохранить файлы с данными, результатами статистической обработки и графического представления данных в папку со своей фамилией и номером группы на рабочем столе с именем «Beg100».

Практическое задание № 5

«Расчёт точечных и интервальных оценок»

по теме «Первичная обработка результатов эксперимента».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение рассчитывать доверительный интервал и формулировать выводы по данным расчетам.

Теоретические сведения к практическому занятию:

Под термином оценка в теории оценок принимают как сами значения параметров генеральной совокупности, полученные по выборке, так и процесс получения этих значений, т.е. правило, по которому они получены.

Оценки подразделяются на два класса: точечные и интервальные.

Точечные оценки представляют собой определенные значения параметров генеральной совокупности, полученные по выборочным данным. Эти значения должны быть максимально близки к значениям соответствующих параметров генеральной совокупности, которые являются истинными значениями оцениваемых параметров.

При формировании интервальных оценок определяют границы интервалов, между которыми с большой вероятностью находятся истинные значения параметров.

Точечная оценка.

Точечной оценкой числовой характеристики называют оценку, которая определяется одним числом, например, среднее арифметическое или стандартное отклонение.

Если выборка небольшого объема, то точечная оценка может значительно отличаться от оцениваемого параметра и её использование может привести к грубым ошибкам. Поэтому при небольшом объеме выборки используют другого типа оценки характеристик генеральной совокупности – *интервальные*.

Интервальная оценка.

Интервальной оценкой числовой характеристики называется интервал, который с *доверительной вероятностью* γ (задаваемой заранее) накрывает истинное значение числовой характеристики генеральной совокупности.

Доверительная вероятность - это уровень гарантии суждения о значениях генеральной характеристики на основании выборочных данных.

Как правило, в научных исследованиях в области спорта считается достаточной доверительная вероятность 0,95. В случаях, связанных с большей ответственностью при принятии решений, принимают $\gamma=0,99$ или $\gamma=0,999$.

Вероятность $\alpha=1-\gamma$ того, что построенный доверительный интервал не накроет значение генеральной характеристики, называется уровнем значимости, т.е. α – вероятность ошибки.

Доверительный интервал для математического ожидания μ генеральной совокупности в случае её нормального распределения по нормальному закону и неизвестным среднеквадратическим отклонением записывается в виде:

$$I_{\mu} = \left\langle \bar{x} - t_{\gamma} \frac{S}{\sqrt{n}}; \bar{x} + t_{\gamma} \frac{S}{\sqrt{n}} \right\rangle \quad , \quad (10)$$

то есть

$$P \left(\mu \in \left\langle \bar{x} - t_{\gamma} \frac{S}{\sqrt{n}}; \bar{x} + t_{\gamma} \frac{S}{\sqrt{n}} \right\rangle \right) = \gamma \quad , \quad (11)$$

где - *квантиль* распределения, который зависит объёма выборки n и доверительной вероятности γ .

Доверительный интервал для математического ожидания μ генеральной совокупности в случае её нормального распределения по нормальному закону с известным среднеквадратическим отклонением σ записывается в виде:

$$I_{\mu} = \left\langle \bar{x} - t_{\gamma} \frac{\sigma}{\sqrt{n}}; \bar{x} + t_{\gamma} \frac{\sigma}{\sqrt{n}} \right\rangle \quad , \quad (12)$$

$$P\left(\mu \in \left\langle \bar{x} - t_{\gamma} \frac{\sigma}{\sqrt{n}}; \bar{x} + t_{\gamma} \frac{\sigma}{\sqrt{n}} \right\rangle\right) = \gamma \quad , \quad (13)$$

то есть

где t_{γ} - квантиль распределения, который в этом случае зависит только от доверительной вероятности γ (находится из таблицы 8).

Таблица 8. Функция $\Phi_0(x)$ (значение t_{γ} в зависимости от объема выборки и доверительной вероятности).

n	γ		
	0,95	0,99	0,999
10	2,265	3,250	4,781
15	2,145	2,977	4,140
20	2,093	2,861	3,883
30	2,042	2,750	3,646
40	2,021	2,704	3,551
50	2,009	2,678	3,505
60	2,000	2,660	3,505
80	1,990	2,639	3,416

Пример: определить границы 95% доверительного интервала для среднего результата в беге на 100 м ($n=50$): $\bar{x} = 15,38$ с, $S_{\bar{x}} = 0,13$ с.

Решение: Из таблицы 8 для $n=50$ и $\gamma=0,95$ находим значение $t_{\gamma} = 2,009$. Следовательно, доверительный интервал будет следующим:

$$15,28 - 2,009 \cdot 0,13 \leq \mu \leq 15,38 + 2,009 \cdot 0,13 \text{ или } 15,12 \leq \mu \leq 15,64$$

После округления получим итоговый результат:

$$15,1 \leq \mu \leq 15,6 \text{ с.}$$

Вывод: с доверительной вероятностью $\gamma = 0,95$ можно утверждать, что генеральное среднее μ заключено в границах от 15,1 до 15,6 с.

Форма выполнения практического занятия: решение профессиональной задачи:

1. письменно в тетради рассчитать доверительный интервал и на основании вычислений сформулировать вывод.

2. на компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 рассчитать доверительные интервал.

Задача: Определить границы 99,9% ($\gamma=0,999$) доверительного интервала для среднего результата в беге на 100 м ($n=50$). Данные для расчета (среднее арифметическое и стандартное отклонение) взять из задачи, выполненной в практическом занятию № 3 (файл «Beg100»).

Ход выполнения практического занятия:

Математическое решение задачи (письменно в тетради):

1. Используя формулу $I_{\mu} = \langle \bar{x} - t_{\gamma} \frac{s}{\sqrt{n}}; \bar{x} + t_{\gamma} \frac{s}{\sqrt{n}} \rangle$ и данные таблицы 8 вычислить доверительный интервал;
2. На основании расчетов сделать вывод.

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. Запустить программу STATGRAPHICS Plus и открыть файл с исходными данными «Beg100».
2. Выбирать в меню пункты: **Describe/Numeric Data/One Variable Analysis**. Появится диалоговое окно **One-Variable Analysis**.
3. Перенести переменную x1 в поле **Data**, щелкнуть на кнопке ОК, появится окно общего анализа (**Analysis Summary**), в котором сообщается информация о переменной.
4. Щелкнуть на кнопке табличных опций (**Tabular option**), появится диалоговое окно табличных опций. В этом окне маркируем поле **Confidence Intervals** (доверительные интервалы). Щелкнуть на кнопке ОК. Появится диалоговое окно, в котором приводится информация о 95% доверительных интервалах для среднего арифметического (**mean**) и стандартного отклонения (**standard deviation**)

```
Confidence Intervals for x1
-----
95,0% confidence interval for mean: 15,382 +/- 0,26886 [15,1131;15,6509]
95,0% confidence interval for standard deviation: [0,790252;1,17888]
```

Рис.25. Значения доверительного интервала для переменной x1 с доверительной вероятностью $\gamma=0,95$.

5. Если нужно изменить значение доверительной вероятности, нужно щелкнуть правой кнопкой мыши и в появившемся контекстном меню выбрать **Pane Options**, появится диалоговое окно **Confidence Intervals**, в котором необходимо установить требуемый уровень доверительной вероятности – 99,9% (рис.26)

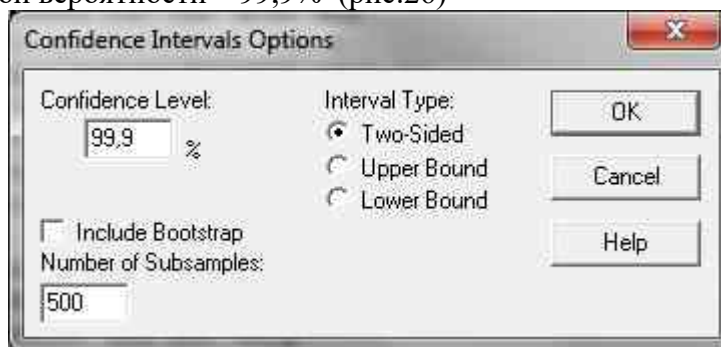


Рис.26. Диалоговое окно *Confidence Intervals*

6. После этого щелкнуть на кнопке ОК, в результате получаем новые значения доверительных интервалов

```
Confidence Intervals for x1
-----
99,9% confidence interval for mean: 15,382 +/- 0,468322 [14,9137;15,8503]
99,9% confidence interval for standard deviation: [0,705000;1,3872]
```

Рис.27. Значения доверительного интервала для переменной x1 с доверительной вероятностью $\gamma=0,999$.

7. Сохранить файлы с данными и результатами статистической обработки в папку со своей фамилией и номером группы на рабочем столе с именем «Beg100».

Практическое задание № 6

«Проверка гипотезы о нормальном распределении»

по теме «Проверка правдоподобия статистических гипотез».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение строить графики Normal Probability Plot и Symmetry Plot и по ним формулировать выводы о нормальном распределении экспериментальных данных.

Теоретические сведения к практическому занятию:

Полученные практические данные непрерывных величин в ходе различных экспериментов могут принимать различные значения от $-\infty$ до $+\infty$. Но любой из исследуемых признаков в виде отдельно взятых наблюдений подчиняется определенным закономерностям различных видов распределений. Наиболее применяемыми в практической статистике в области физической культуры и спорта являются: нормальное распределение, χ^2 – распределение, t – распределение Стьюдента, F- распределение. Основополагающая роль в математической статистике отводится *нормальному распределению*.

Нормальное распределение - это распределение вероятностей непрерывной случайной величины, которое описывается плотностью вероятности.

Плотность вероятностей нормально распределенной случайной величины записывается в виде математического выражения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (14)$$

где константа $\pi = 3,14\dots$, $e = 2,718\dots$ (основание натуральных логарифмов), x - переменная, показывающая значение признака (случайной величины), μ - математическое ожидание, σ - стандартное отклонение.

График плотности (нормальная кривая) показан на рис. 28.

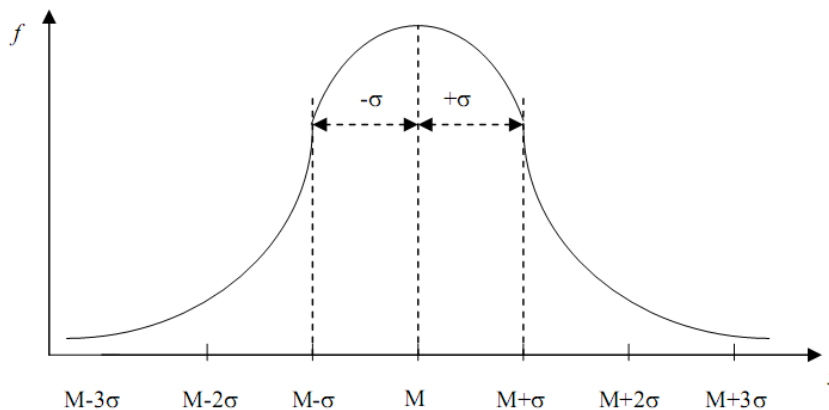


Рис.28. График плотности (нормальная кривая).

Теоретическая кривая нормального распределения имеет следующие свойства:

1. Нормальная кривая имеет колонообразную форму, симметричную относительно точки $x=\mu$, с точками перегиба, абсциссами которых являются значения: $x=\mu\pm\sigma$;
2. При стремлении x к $\pm\infty$ асимптотически приближается к оси абсцисс;
3. Максимальное значение функции, которую она отражает, достигается при $x = \mu$;
4. Коэффициенты асимметрии (As) и эксцесса (Ex) нормального распределения равны нулю. Это свойство используется для проверки предположения о нормальности распределения генеральной совокупности.

В пакете STATGRAPHICS Plus 5.0 проверка правдоподобия гипотезы о нормальном распределении состоит в графическом представлении данных (с помощью графиков Normal Probability Plot и Symmetry Plot).

При построении графика Normal Probability Plot идея заключается в том, чтобы вводом специальной системы координат на плоскости (т.е. преобразованием переменных), представить график гипотетической функции нормального распределения в виде прямой и набор экспериментальных точек, соответствующих выборочным значениям. *Чем ближе отдельно стоящие точки к прямой линии, тем правдоподобнее гипотеза о нормальном распределении экспериментальных данных (рис.29).*

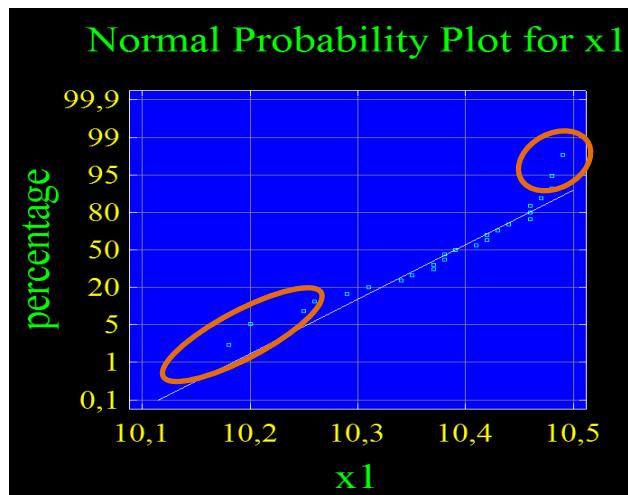


Рис.29. График Normal Probability Plot.

Вывод: Так как расположение экспериментальных точек (отмечены на графике овалами) сильно отличаются от прямой, то это позволяет отвергнуть гипотезу о нормальном распределении.

При построении графика Symmetry Plot используют свойство симметрии графика плотности нормального распределения.

Пример: Проверить гипотезу о нормальном распределении с помощью построения графика Symmetry Plot для экспериментальных данных, полученных в (таблица 9).

Таблица 9. Результат бега в секундах на 100 м (мужчины, $n=25$).

10,18	10,20	10,25	10,26	10,29
10,31	10,34	10,35	10,37	10,37
10,38	10,38	10,39	10,41	10,42
10,42	10,43	10,44	10,46	10,46
10,46	10,47	10,48	10,48	10,49

Решение: Алгоритм построения графика симметрии.

1. Перегруппировать экспериментальные данные по возрастанию или убыванию (ранжировать) и найти медиану (таблица 10, столбец 2);

2. Из каждого ранжированного ряда вычесть медиану (таблица 10, столбец 3);

3. Сформировать координаты точек и записать в столбца 5 и 6 соответственно: по x – значения столбца 3 по модулю с первого значения до медианы, по y – значения столбца 3 в обратном порядке (с последнего значения) до медианы;

Таблица 10. Нахождение координат точек для построения графика Symmetry Plot.

№ п/п	Ранжированный ряд x_i	Разность между значением ранжированного ряда и медианой	Координата по оси x	Координата по оси y
1	10,18	-0,21	0,21	0,10
2	10,20	-0,19	0,19	0,09
3	10,25	-0,14	0,14	0,09
4	10,26	-0,13	0,13	0,08
5	10,29	-0,10	0,10	0,07
6	10,31	-0,08	0,08	0,07
7	10,34	-0,05	0,05	0,07
8	10,35	-0,04	0,04	0,05
9	10,37	-0,02	0,02	0,04

10	10,37	-0,02	0,02	0,03
11	10,38	-0,01	0,01	0,03
12	10,38	-0,01	0,01	0,02
13	10,39- медиана	0		
14	10,41	0,02		
15	10,42	0,03		
16	10,42	0,03		
17	10,43	0,04		
18	10,44	0,05		
19	10,46	0,07		
20	10,46	0,07		
21	10,46	0,07		
22	10,47	0,08		
23	10,48	0,09		
24	10,48	0,09		
25	10,49	0,10		

4. В декартовой системе координат провести биссектрису первого координатного угла и нанести точки (рис.30);

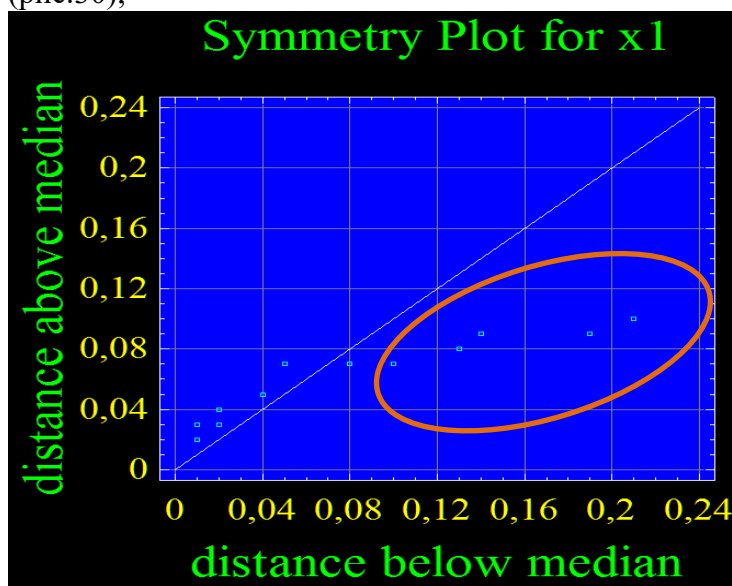


Рис.30. График асимметрии (Symmetry Plot).

5. Сделать вывод.

Вывод: Так как расположение экспериментальных точек (отмечены на графике овалом) сильно отличаются от прямой, то это позволяет отвергнуть гипотезу о нормальном распределении.

Форма выполнения практического занятия: решение профессиональной задачи:

1. В тетради построить график Symmetry Plot.

2. На компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 построить графики Normal Probability Plot и Symmetry Plot.

Задача: Проверить гипотезу о нормальном распределении по результатам бега на 100 м (юноши, $n=50$) с помощью графиков Normal Probability Plot и Symmetry Plot. Данные взять из задачи, выполненной в практическом занятию № 3 (файл «Beg100»).

Ход выполнения практического занятия:

Математическое решение задачи (в тетради):

1. Заполнить таблицу 11;

Таблица 11. Нахождение координат точек для построения графика Symmetry Plot.

№ п/п	Ранжированный ряд x_i	Разность между значением ранжированного ряда и медианой	Координата по оси x	Координата по оси y
1				
2				
...				
50				

2. Построить график в декартовой системе координат;
3. Сделать вывод.

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. Запустить программу STATGRAPHICS Plus и открыть файл с исходными данными «Beg100».
2. Выбирать в меню пункты: Describe/Numeric Data/One Variable Analysis. Занести название переменной в поле Data, щелкнуть на кнопке ОК. Появится окно анализа (One-Variable Analysis).
3. В окне анализа щелкнуть на кнопке табличных опций (Tabular option), в появившемся диалоговом окне выбрать Normal Probability Plot и Symmetry Plot.
4. В результате будут построены графики Normal Probability Plot и Symmetry Plot.
5. В окне StatReporter сформулировать выводы о нормальном распределении экспериментальных данных по построенным графикам в программе STATGRAPHICS Plus 5.0.
6. Сохранить файлы с данными, результатами статистической обработки и графического представления данных в папку со своей фамилией и номером группы на рабочем столе с именем «Beg100».

Практическое задание № 7

«Сравнение независимых выборок»

по теме «Проверка правдоподобия статистических гипотез».

Время, отведенное на выполнение практического занятия: 4 часа.

Цель: сформировать умение проводить сравнение независимых выборок и представлять статистически обработанный материал в графическом виде и формулировать выводы формулировать выводы на основе полученных результатов анализа и графикам.

Теоретические сведения к практическому занятию:

Данная тема рассматривает методы, применяемые тогда, когда предстоит проверить какие-то теоретические предположения, связанные с эффективностью педагогического процесса (усвоение учебного материала, результат тренировочного процесса и т.д.).

Статистическая гипотеза - утверждение о распределении генеральной совокупности, соответствующее некоторым представлениям об изучаемом явлении. В частности, это может быть утверждение о значениях параметров μ и σ нормально распределенной генеральной совокупности.

Нулевая гипотеза (H0) - гипотеза, основанная на утверждении, что между двумя генеральными совокупностями нет ожидаемого различия:

Например, $\mu_1 = \mu_2$, где μ_1 – взяты результаты прыжков в длину юных легкоатлетов, тренирующихся по традиционной методике; μ_2 – результаты прыжков в длину другой группы юных легкоатлетов, использующих новый комплекс специальных упражнений.

Таким образом, нулевая гипотеза делает предположение, что генеральные средние арифметические (результаты всех юных прыгунов данного класса, которые могли бы тренироваться по традиционной и новой программе) не отличаются $\mu_1 = \mu_2$.

Альтернативная гипотеза (H_1) – гипотеза с утверждением, обратным нулевой гипотезе, т.е. утверждение о том, что в действительности между генеральными совокупностями есть различие, например: $\mu_1 \neq \mu_2$ (результаты прыжков в длину с разбега юных легкоатлетов, занимающихся по традиционной методике и новой программе не равны).

Статистические гипотезы, в частности, нулевая и альтернативная, проверяются с помощью какого-то метода – критерия.

Существуют критерии, основанные на нормальном распределении данных (параметрические), к ним относятся: F - критерий Фишера; t-критерий Стьюдента; u-критерий.

Существуют критерии, которые сравнивают средние значения генеральных совокупностей μ_1 и μ_2 , распределение которых отклонилось от нормального или параметры тех совокупностей, которые измеряются в шкалах порядка данных или наименований (например, произвольная нумерация игроков футбольной команды или места, занятые спортсменами на соревнованиях и т.д.). К ним относятся: критерий Вилкоксона или Уайта.

Помимо этих критериев, существуют критерии, с помощью которых проверяется предположение о нормальном распределении генеральной совокупности. Они называются критериями согласия. К ним относятся: Асимметрия (As); Эксцесс (Ex); критерий χ^2 (хи – квадрат); критерий Шапиро-Уилки.

При проверке гипотез также применяются уровни значимости.

Уровни значимости

В ходе исследовательской работы очень важным моментом бывает установление наличия или отсутствия различий в полученных числовых характеристиках при изучении каких-то результатов, показанных спортсменами (испытуемыми) контрольной и экспериментальной групп.

Например, перед исследователем ставится задача – разработать экспериментальную методику обучения прыжкам в длину с разбега для учащихся общеобразовательной школы. После того как новая методика обучения разработана и применена в экспериментальной группе школьников, их средний результат вырос на 10 см ($-x = 10$ см), а в контрольной группе этот показатель увеличился всего на 4 см ($-y = 4$ см). Перед исследователем встает вопрос: можно ли утверждать, что нововведения эффективнее повлияли на процесс формирования изучаемого двигательного действия по сравнению с традиционной методикой или это случайность?

Отвечая на этот вопрос, исследователь перед проведением эксперимента формулирует гипотезы:

а) *Нулевая гипотеза (H_0)* – предполагается, что новый комплекс упражнений (методика обучения) недостаточно хорошо разработан и незначительно повлияет на результат прыжков в длину с разбега, а различия в средних значениях контрольной и экспериментальной групп (если они выявятся), будут обусловлены только действием случайностей.

б) *Альтернативная гипотеза (H_1)* – нововведения успешно решат задачу обучения в экспериментальной группе, а полученные данные будут превосходить результаты контрольной группы. Далее нужно доказать действительно ли, статистически достоверно, или, наоборот, недостоверно различие найденных средних приростов результатов прыжков в длину с разбега ($-x -y = 6$ см) контрольной и экспериментальной групп.

Для этого вычисляют значение некоторой величины, называемой критерием, которая чаще всего имеет стандартное распределение (u - распределение, t - распределение и т.п.). Найденная величина сравнивается с критическим (граничным) значением критерия,

взятым из соответствующих таблиц, и по результатам сравнения определяется статистическая достоверность наличия или отсутствия различий между двумя сравниваемыми параметрами. В области ФК и спорта достаточен уровень значимости $\alpha = 0,05$, более серьезные выводы рекомендуется давать, используя уровень значимости $\alpha = 0,01$ или $\alpha = 0,001$.

Чтобы избежать однозначных, убедительных ответов на поставленные серьезные вопросы, поступают следующим образом: уровень значимости до эксперимента не устанавливается точно, а по экспериментальным данным вычисляется вероятность p того, что критерий выйдет за пределы значения, рассчитанного в выборке. p – экспериментальный уровень значимости. Точное значение p также не указывают, обычно это делается следующим образом:

а) если вычисленное значение критерия (например, t - критерия Стьюдента) не превосходит критического значения (табличное, t - критерий Стьюдента) на уровне значимости $\alpha = 0,05$, то различия считаются статистически недостоверными, записывается - ($p > 0,05$)

б) если вычисленное значение критерия, превышает критические значения при $\alpha = 0,05$; $\alpha = 0,01$ или $\alpha = 0,001$, то записывается - ($p < 0,05$), ($p < 0,01$), ($p < 0,001$). Это означает, что наблюдаемые различия статистически достоверны на уровнях значимости - 0,05; 0,01 или 0,001.

Независимые выборки - две выборки, экспериментально не связанные между собой. Измерения, проведенные в одной выборке, не оказывают влияния на значения переменных в другой.

Рассмотрим 3 основных критерия для сравнения независимых выборок:

1. F-критерий Фишера;
2. T-критерий Стьюдента;
3. U-критерий Манна-Уитни и Вилкоксона.

F-критерий Фишера – проверка гипотезы о равенстве дисперсий генеральных совокупностей.

Условия применения: выборки независимы и получены из нормально распределенных генеральных совокупностей с параметрами $\mu_x, \sigma_x, \mu_y, \sigma_y$.

Гипотеза H_0 : $\sigma_x^2 = \sigma_y^2$

Альтернативная гипотеза: $\sigma_x^2 \neq \sigma_y^2$

Рассчитываем F-критерий Фишера по формуле:

$$F_{\text{набл}} = \frac{S_1^2}{S_2^2}, \quad (15)$$

и затем сравниваем $F_{\text{набл}}$ и $F_{\text{кр}}$ (см. Приложение 1). Если $F_{\text{набл}} \geq F_{\text{кр}}$, то гипотеза о равенстве дисперсий генеральных совокупностей отвергается на уровне значимости α .

t-критерий Стьюдента – проверка гипотезы о равенстве средних арифметических генеральных совокупностей.

Условия применения: выборки независимы и получены из нормально распределенных генеральных совокупностей с параметрами $\mu_x, \sigma_x, \mu_y, \sigma_y$.

Гипотеза H_0 : $\mu_x = \mu_y$

Альтернативная гипотеза: $\mu_x \neq \mu_y$

Рассчитываем t-критерий Стьюдента по формуле:

$$t_{\text{набл}} = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}, \quad (16)$$

и сравниваем $t_{\text{набл}}$ и $t_{\text{кр}}$ (см. Приложение 2). Если $t_{\text{набл}} \geq t_{\text{кр}}$, то гипотеза о равенстве генеральных средних отвергается на уровне значимости α .

Критерий Манна-Уитни и Вилкоксона – проверка гипотезы о равенстве медиан генеральных совокупностей.

Условия применения: выборки получены из однотипных непрерывных распределений. При этом закон распределения генеральных совокупностей может быть любым.

Гипотеза H_0 : $Me_x = Me_y$

Альтернативная гипотеза: $Me_x \neq Me_y$

Пример: Даны результаты бега на 100 м (в секундах) в контрольной (X) и экспериментальной (Y) группах юношей, тренирующимся по различным методикам. Проверить гипотезу о равенстве генеральных совокупностей и сделать вывод об эффективности данных методик.

Группы	Результаты бега на 100 м, с									
Контр.(X)	12,6	12,3	11,8	12,1	12,8	13,2	13,8	12,0	12,6	13,0
Эксп.(Y)	11,3	12,8	12,2	11,7	12,4	13,3	11,4	12,0	11,8	12,5

Решение:

1. Объединяем 2 выборки X и Y в одну и ранжируем их по возрастанию, при это выделяем данные, относящиеся к выборке Y (Таблица 12, столбец 3);

2. Присваиваем ранги объединенной выборки (Таблица 12, столбец 4);

3. После этого суммируем по отдельности ранги, относящиеся к каждой из выборок. Меньшая из сумм принимается в качестве критерия $W_{набл.}=83,5$. Его значение сравнивается с критическим $W_{кр.}=78$ (см. Приложение 3).

4. Формулируем вывод: если $W_{набл.} \leq W_{кр.}$, то нулевая гипотеза отвергается – различие считается статистически значимым на уровне значимости $\alpha=0,05$. В нашем случае $W_{набл.} > W_{кр.}$ ($83,5 > 78$), т.е. нулевая гипотеза принимается – различие считается статистически незначимым на уровне значимости $\alpha=0,05$.

Таблица 12. Расчёт критериев Манна-Уитни и Вилкоксона.

X	Y	Объединенная выборка	Ранги
12,6	11,3	11,3	1
12,3	12,8	11,4	2
11,8	12,2	11,7	3
12,1	11,7	11,8	4,5
12,8	12,4	11,8	4,5
13,2	13,3	12,0	6,5
13,8	11,4	12,0	6,5
12,0	12,0	12,1	8
12,6	11,8	12,2	9
13,0	12,5	12,3	10
		12,4	11
		12,5	12
		12,6	13,5
		12,6	13,5
		12,8	15,5
		12,8	15,5
		13,0	17
		13,2	18
		13,3	19
		13,8	20
		Сумма рангов	83,5
		Сумма рангов	126,5

5. После этого критерий Манна-Уитни U рассчитывается по формуле:

$$U = W - \frac{1}{2}m(m + 1) \quad , \quad (17)$$

где W – значение критерия Вилкоксона, m – объем выборки в группе с большей ранговой суммой.

$$U = W - \frac{1}{2}m(m + 1) = 83,5 - \frac{1}{2} \cdot 10 \cdot (10 + 1) = 28,5$$

6. Полученное значение $U_{\text{набл.}}=28,5$ сравнивается с $U_{\text{кр.}}=23$ (см. Приложение 4). Если $U_{\text{набл.}} \leq U_{\text{кр.}}$ – нулевая гипотеза отвергается на уровне значимости $\alpha=0,05$, различия результатов считается статистически значимым. В нашем случае $U_{\text{набл.}} > U_{\text{кр.}}$, т.е. нулевая гипотеза принимается на уровне значимости $\alpha=0,05$, различия результатов считается статистически незначимым.

Форма выполнения практического занятия: решение профессиональной задачи:

1. Письменно в тетради сравнить независимые выборки с помощью критерия Манна-Уитни и Вилкоксона и сформулировать статистический вывод.

2. На компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 сравнить независимые выборки с помощью следующих критериев: F-критерия Фишера, t-критерия Стьюдента и критерия Манна-Уитни и Вилкоксона; сформулировать и записать статистические выводы.

3. Построить все виды графиков (Frequency Histogram, Density Trace, Box-and-Whisker Plot, Quantile Plot, Quantile-Quantile Plot) и по ним сделать статистические выводы.

Задача: Проверить гипотезу о том, что вариативность роста мальчиков 15 лет такая же, как и девочек этого же возраста. Проверить гипотезу о том, что рост 15-летних девочек меньше роста 15-летних мальчиков.

Примечание: распределение роста девочек и мальчиков 15 лет считать соответствующим нормальному закону.

Таблица 13. Рост девочек и мальчиков Санкт-Петербурга 10, 12 и 15 лет ($n=50$)

№	dev10	mal10	dev12	mal12	dev15	mal15
1	146	142	153	142	179	167
2	138	130	153	160	171	174
3	144	143	149	155	175	163
4	142	142	155	153	163	175
5	141	136	164	132	153	176
6	144	144	151	160	163	160
7	145	145	155	154	160	158
8	142	142	153	154	167	159
9	135	147	159	158	167	160
10	136	140	156	147	170	164
11	142	148	158	145	168	161
12	143	137	158	144	159	175
13	138	136	160	147	157	164
14	143	151	156	158	165	168

<i>№</i>	<i>dev10</i>	<i>mal10</i>	<i>dev12</i>	<i>mal12</i>	<i>dev15</i>	<i>mal15</i>
15	146	129	164	163	162	160
16	150	142	155	153	170	165
17	139	140	159	144	157	166
18	146	132	153	147	163	169
19	140	135	161	155	166	168
20	149	149	160	160	163	171
21	146	150	155	145	167	169
22	146	141	155	156	164	162
23	137	146	149	137	170	167
24	146	144	157	143	147	166
25	143	129	162	150	168	169
26	144	142	157	144	165	170
27	144	138	139	145	159	170
28	141	137	139	141	166	175
29	132	144	144	141	173	165
30	142	146	142	151	168	175
31	143	133	157	146	155	173
32	149	144	153	130	162	169
33	131	137	156	145	169	172
34	136	138	150	147	162	177
35	136	134	160	142	163	175
36	145	140	160	143	168	176
37	140	139	151	139	164	182
38	138	132	153	150	162	174
39	144	139	149	146	163	160
40	145	142	148	149	164	180
41	138	139	155	145	168	176
42	144	143	158	145	167	160
43	144	148	157	147	161	169
44	132	131	155	150	169	172
45	144	145	142	154	163	163
46	145	138	147	149	160	156
47	133	136	154	146	162	174

№	dev10	mal10	dev12	mal12	dev15	mal15
48	142	145	143	149	157	164
49	131	139	157	146	155	168
50	148	147	155	145	161	175

Статистические графики - Условные обозначения статистических данных посредством линий, геометрических фигур, рисунков или картосхем.

Виды графиков в пакете STSTGRAPHICS Plus, отображающие результаты проведенного анализа.

- Frequency Histogram (гистограмма);
- Density Trace (графики плотности вероятности);
- Box-and-Whisker Plot («ящик-с-усами»);
- Quantile Plot;
- Quantile-Quantile Plot.

Frequency Histogram.

Frequency Histogram - графическое изображение зависимости частоты попадания элементов выборки от соответствующего интервала группировки (рис.39).

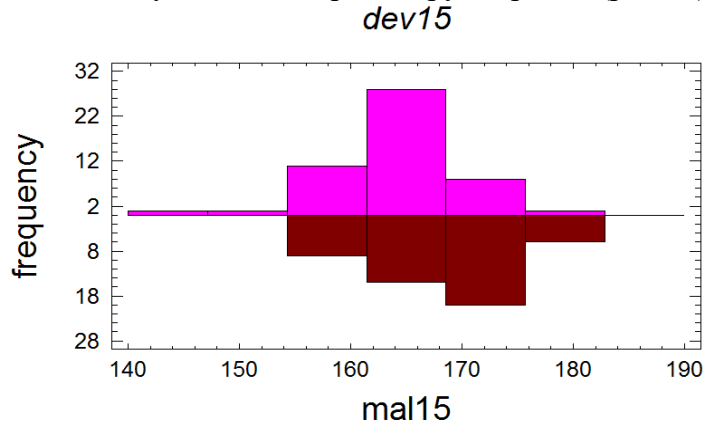


Рис.96. График Frequency Histogram.

На этом графике представлены гистограммы распределения двух выборок. При этом одна из них представлена в инверсном виде. Это представление итоговых результатов удобно использовать при необходимости продемонстрировать формы распределений.

Density Trace.

Density Trace - плотность вероятности распределения двух выборок (рис.40).

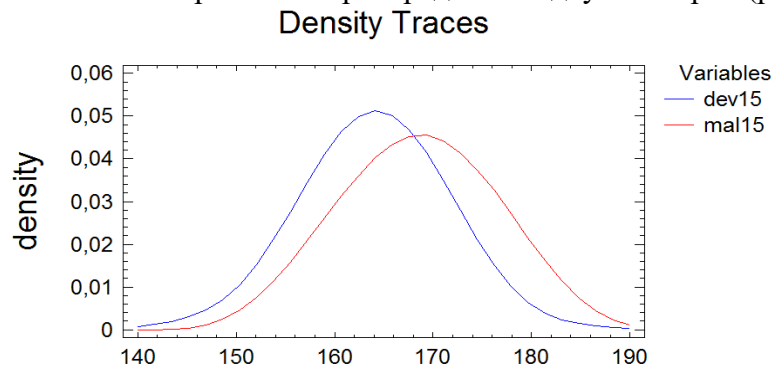


Рис.40. График Density Trace.

Этот график также демонстрирует плотность вероятности распределений двух выборок. Чем ближе графики друг к другу, тем меньше отличий между выборками и наоборот.

Box-and-Whisker Plot.

Box-and-Whisker Plot - график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей (рис.41).

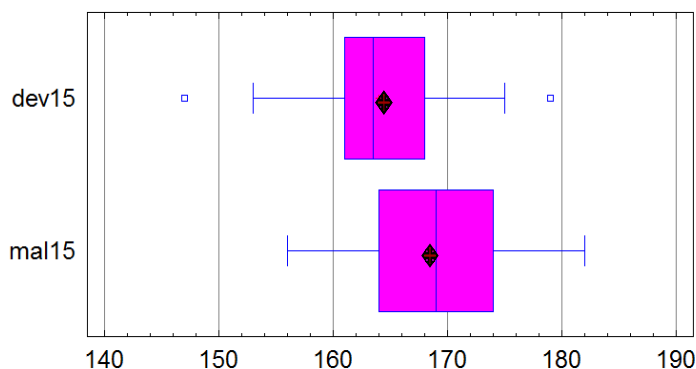


Рис.41. График Box-and-Whisker Plot.

Этот график представляет собой прямоугольник. Внутри этого прямоугольника попадает 50% всех результатов. Линия внутри этого прямоугольника соответствует медиане. Значение среднего арифметического помечено крестиком. «Усы» соответствуют максимальному и минимальному значению. Исключения составляют случаи отклонения значений от границ прямоугольника на полторы длины (помечается на диаграмме квадратиками). Эти значения наносятся на график отдельно. На рис.38 видно, что в группе девочек экстремальные значения выходят за границы, соответствующие 1,5 длины прямоугольника. В группе мальчиков таких значений нет. Этот график наглядно демонстрирует отличия выборок: чем меньше пересечение проекций графиков на ось абсцисс, тем больше различий между выборками.

Quantile Plot.

Quantile Plot - кумуляты, построенные на основе данных двух выборок (рис.42).

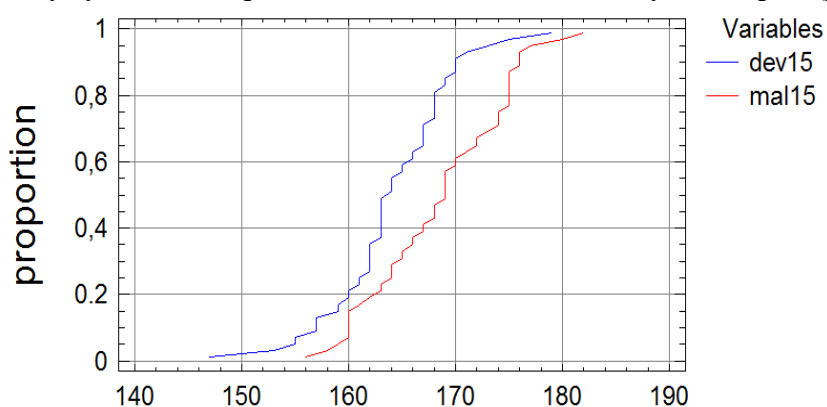


Рис.42. График Quantile Plot.

Чем больше отличий между выборками, тем дальше друг от друга графики.

Quantile-Quantile Plot.

Для построения графика ранжируем значения двух выборок, например, mal15 и dev15. И для каждой пары значений изображаем точку на координатной плоскости. На графике (рис.43) отчетливо видно превышение результатов одной выборки над другой, точки располагаются над биссектрисой первого координатного угла.

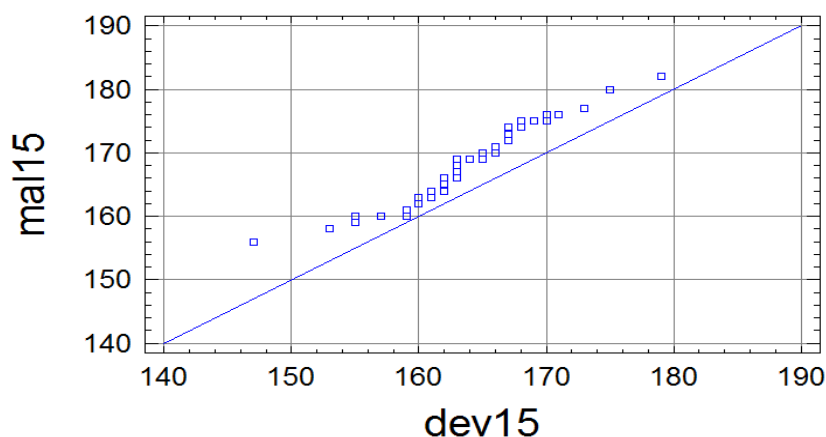


Рис.43. График Quantile-Quantile Plot.

В нашем случае отчетливо видно, что *рост мальчиков 15 лет больше роста их сверстниц.*

Ход выполнения практического занятия:

Математическое решение задачи (в тетради):

1. Объединяем 2 выборки X и Y в одну и ранжируем их по возрастанию;
2. Присваиваем ранги объединенной выборки;
3. После этого суммируем по отдельности ранги, относящиеся к каждой из выборок.

Меньшая из сумм принимается в качестве критерия $W_{набл.}$

4. Его значение сравнивается с критическим $W_{кр.}$ и формулируем вывод;
5. Рассчитываем критерий Манна-Уитни U рассчитывается по формуле 12;
6. Полученное значение $U_{набл.}$ сравнивается с $U_{кр.}$ И формулируем вывод.

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. Запустить программу STATGRAPHICS Plus и в электронной таблице создать столбцы, имеющие следующие имена:

dev10 – рост 10-летних девочек Санкт-Петербурга
mal10 – рост 10-летних мальчиков Санкт-Петербурга
dev12 – рост 12-летних девочек Санкт-Петербурга
mal12 – рост 12-летних мальчиков Санкт-Петербурга
dev15 – рост 15-летних девочек Санкт-Петербурга
mal15 – рост 15-летних мальчиков Санкт-Петербурга

2. Занести в столбцы результаты, представленные в таблице 13.

3. Сохранить результат в файл **ROST**.

Так как распределение роста девочек и мальчиков 15 лет соответствует нормальному закону, нет никаких препятствий для применения F-критерия Фишера и t-критерия Стьюдента. В то же время использование критерия Манна-Уитни не выдвигает никаких требований к характеру распределения результатов, поэтому он может быть использован в любых случаях, если необходимо сравнить две выборки.

4. В меню выбрать **Compare/Two-Samples/Two-Sample Comparison** (*Сравнить/Две выборки/Сравнение двух выборок*). Появится диалоговое окно **Two-Sample Comparison**. Перенести в поле Sample 1 (первая выборка) переменную *dev15*, а в поле **Sample 2** (вторая выборка) переменную *mal15*, рис. 31.

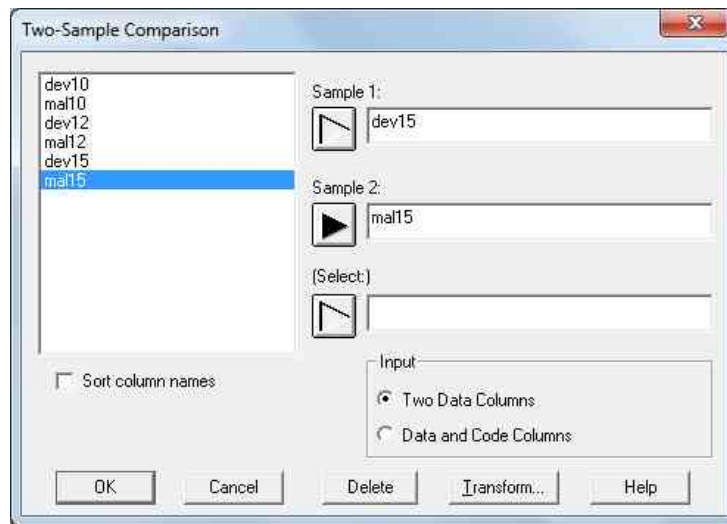


Рис. 31. Диалоговое окно Two-Sample Comparison.

5. В появившемся окне анализа (**Two-Sample Comparison**) появится текстовое окно общего анализа (**Analysis Summary**), в котором сообщается, что сравниваются переменные dev15 и mal15. В первой выборке 50 значений (от 147 и 179), во второй выборке также 50 значений (от 156 до 182).

6. Щелкнуть на кнопке табличных опций (**Tabular options**). Поставить флажки только на следующих полях: **Comparison of means** (сравнение средних арифметических – t-критерий Стьюдента), **Comparison of Standard Deviation** (сравнение стандартных отклонений – F-критерий Фишера) и **Comparison of Medians** (сравнение медиан – критерий Манна-Уитни).

7. В результате получим три окна:

Comparison of means

Comparison of Standard Deviation

Comparison of Medians

8. Т.к. использование t-критерия Стьюдента требует предварительной проверки данных с применением F-критерия Фишера, то начинаем анализ со второго окна.

9. Второе окно имеет вид (рис.32):

Comparison of Standard Deviations		
	dev15	mal15
Standard deviation	5,72499	6,28649
Variance	32,7755	39,52
Df	49	49
Ratio of Variances = 0,82934		
95,0% Confidence Intervals		
Standard deviation of dev15: [4,78228;7,1341]		
Standard deviation of mal15: [5,25132;7,83381]		
Ratio of Variances: [0,470631;1,46145]		
F-test to Compare Standard Deviations		
Null hypothesis: sigma1 = sigma2		
Alt. hypothesis: sigma1 NE sigma2		
F = 0,82934 P-value = 0,51496		

Рис. 32. Окно Comparison of Standard Deviation

Для того, чтобы увидеть все альтернативные гипотезы необходимо щелчком правой кнопки мыши вызвать контекстное меню и выбрать команду **Pane Options**. Появится диалоговое окно **Comparison of Standard Deviation Options**.

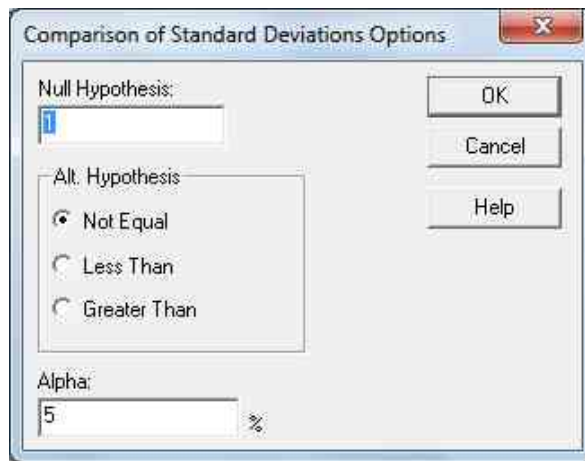


Рис. 33. Диалоговое окно *Comparison of Standard Deviation Options*

Здесь можно выбрать одну из 3 альтернативных гипотез: Not Equal – не равны, Less

Than – меньше чем, Greater Than – больше чем.

```
F-test to Compare Standard Deviations
Null hypothesis: sigma1 = sigma2
Alt. hypothesis: sigma1 > sigma2
F = 0,82934 P-value = 0,74252
```

```
F-test to Compare Standard Deviations
Null hypothesis: sigma1 = sigma2
Alt. hypothesis: sigma1 < sigma2
F = 0,82934 P-value = 0,25748
```

Рис. 34. Альтернативные гипотезы:

Less Than ($\sigma_1 < \sigma_2$) и *Grater Then* ($\sigma_1 > \sigma_2$)

10. В окне *Comparison of Standard Deviation* приводятся:

a. Значения стандартного отклонения (Standard deviation), дисперсии (Variance) и числа степеней свободы (Df) для изучаемых переменных.

b. Доверительные интервалы (Confidence Intervals) для стандартного отклонения.

c. Значения F-критерия Фишера, при различных альтернативных гипотезах. В нашем случае нет оснований предполагать, что значения дисперсий отличаются, поэтому выбираем альтернативу Not Equal (не равны). Значение критерия Фишера $F=0,82934$ и $P\text{-value}=0,51496$ (экспериментально рассчитанной вероятности ошибки) свидетельствует о том, что между дисперсиями нет доверительных отличий, $P > 0,05$. Следовательно, вариативность роста 15-летних мальчиков такая же, как и 15-летних девочек.

Примечание: в пакете *STATGRAPHICS Plus* при расчете величины критерия в числитель заносится значение меньшей дисперсии, а в знаменатель – большей, поэтому значение критерия меньше единицы.

11. Переходим к анализу окна **Comparison of Means** (рис. 35):

```
Comparison of Means
-----
95,0% confidence interval for mean of dev15: 164,0 +/- 1,62703 [162,373,165,627]
95,0% confidence interval for mean of mal15: 168,52 +/- 1,78661 [166,733,170,307]
95,0% confidence interval for the difference between the means
not assuming equal variances: -4,52 +/- 2,38651 [-6,90651,-2,13349]

t test to compare means

Null hypothesis: mean1 = mean2
Alt. hypothesis: mean1 NE mean2
not assuming equal variances: t = -3,75896 P-value = 0,000291839
```

Рис. 35. Окно *Comparison of Means*.

В окне *Comparison of Means* приводятся доверительные интервалы для математического ожидания роста девочек и мальчиков 15 лет, и их разности. После этого приводятся значения t-теста ($t=-3,75896$) и P-value для альтернативной гипотезы **mean1 NE mean2** (генеральные средние не равны). Две другие альтернативные гипотезы (**mean1 > mean2** и **mean1 < mean2**) можно выбрать при необходимости (см.п.9)

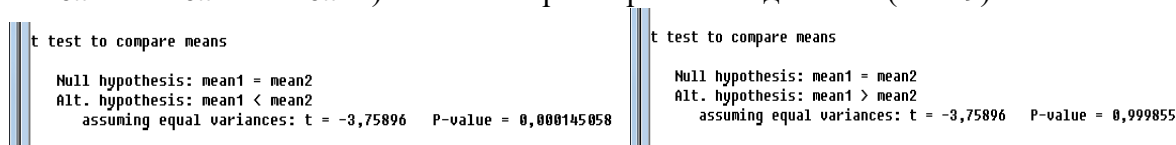


Рис. 36. Альтернативные гипотезы ($mean1 > mean2$ и $mean1 < mean2$) окна анализа *Comparison of Means*

Так как мы предполагаем, что генеральное среднее, характеризующее рост девочек меньше генерального среднего, характеризующего рост мальчиков, то выбираем альтернативной гипотезой гипотезу $mean1 < mean2$.

После этого делаем **вывод**: так как P-value очень мал ($P\text{-value}=0,000145058$), гипотеза о равенстве средних отвергается. Следовательно, рост 15 летних девочек достоверно меньше роста 15 летних мальчиков, $P<0,001$.

Примечание: в каждой из альтернативных гипотез приводятся значения t-критерия Стьюдента и P-value для двух случаев:

- a. Если дисперсии равны (*assuming equal variances*)
- b. Если дисперсии не равны (*not assuming equal variances*)

12. Данные значения можно получить при установке или снятии соответствующей галочки (*Assume Equal Sigmas*) в окне *Comparison of Means Options* (рис.37):

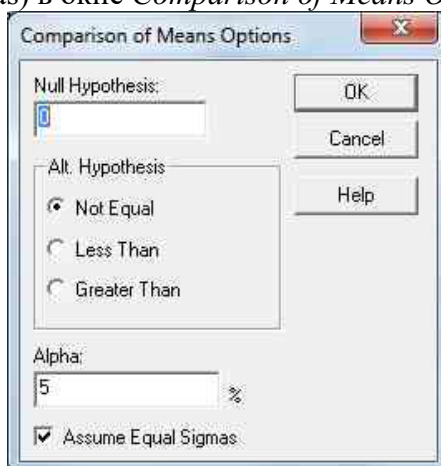


Рис.37. Окно *Comparison of Means Options*

Так как установлено, что группы мальчиков и девочек 15 лет не отличаются друг от друга по вариативности роста (между дисперсиями, характеризующими вариативность роста мальчиков и девочек нет достоверных отличий), выбираем первый вариант (*assuming equal variances*).

13. Третье окно имеет вид:

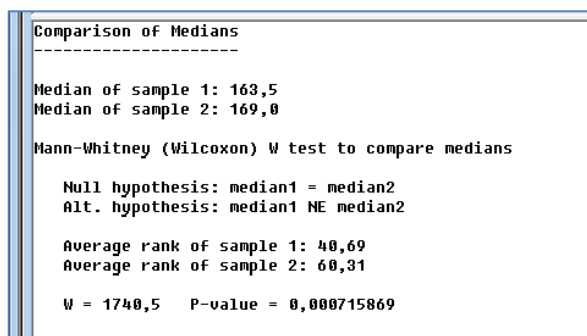


Рис. 38. Окно *Comparison of Medians*

В окне **Comparison of Medians** приводятся значения медианы первой выборки (dev15) и второй выборки (mal15). Медианы соответственно равны: 163,5 см и 169,0 см. После этого формируется нулевая гипотеза. Проверяется равенство медиан генеральных совокупностей ($median1=median2$). После формулировки нулевой гипотезы приводятся величины средних рангов для первой выборки (40,69) и второй выборки (60,31). После этого приводятся значения критерия Манна-Уитни $W=1740,5$ и P-value для альтернативной гипотезы: $median1 \neq median2$ (медианы не равны). Так как P-value очень мал ($P\text{-value}=0,000715869$), то гипотеза о равенстве медиан отвергается, рост 15 летних девочек достоверно меньше роста 15 летних мальчиков ($P<0,001$).

14. Выводы записать в окно StatReporter.

15. Самостоятельно проверить гипотезу о том, что вариативность роста мальчиков 10 лет такая же, как и девочек этого же возраста.

16. Проверить гипотезу о том, что рост 10-летних девочек меньше роста 10-летних мальчиков.

Примечание: распределение роста девочек и мальчиков 10 лет считать соответствующим нормальному закону.

17. Проверить гипотезу о том, что рост 12-летних девочек меньше роста 12-летних мальчиков.

Примечание: распределение роста девочек и мальчиков 12 лет считать соответствующим нормальному закону.

18. Для построения графиков результата анализа используем файл ROST.

19. В окне проведения предварительного анализа (Analysis Summary) щелкнуть на кнопке графических опций. Появится диалоговое окно (рис.44). Выбрать необходимые графики.

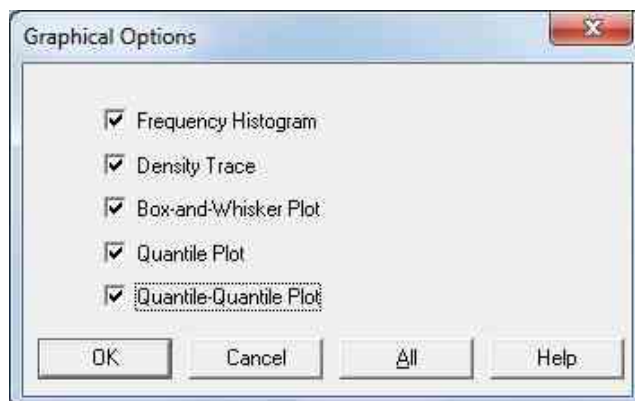


Рис.44. Диалоговое окно *Graphical Options*.

20. В окне StatGallery изменить расположение панелей. Для этого щелчком правой кнопки мыши вызвать контекстное меню и выбрать команду Arrange Panes. В появившемся диалоговом окне StatGallery Options включить переключатель By Columns. В колонках 1 и 2 установить по 2 строки (Rows), а в 3-ем столбце – 1 строку.

21. Аналогично создать три страницы (кнопка Next Page) в окне StatGallery.
22. Результаты графического анализа данных для разных возрастных групп поместить на отдельные страницы.
23. Выводы по результатам анализа записать в окне StatReporter.
24. Сохранить файл с графическим представлением статистических данных в папку со своей фамилией и номером группы на рабочем столе с именем «ROST».

Практическое задание № 8 **«Сравнение связанных выборок»**

по теме «Проверка правдоподобия статистических гипотез».

Время, отведенное на выполнение практического занятия: 4 часа.

Цель: сформировать умение проводить сравнение связанных выборок и формулировать выводы на основе полученных результатов анализа.

Теоретические сведения к практическому занятию:

При педагогических исследованиях используются сравнения парных выборок, полученных до и после проведения эксперимента. Такие выборки считаются связанными. Для оценки достоверности различий между такими выборками используется модифицированный t-критерий Стьюдента и непараметрический критерий знаковых рангов Вилкоксона.

t-критерий Стьюдента (проверка гипотезы о равенстве нулю среднего арифметического генеральной совокупности).

При использовании этого критерия гипотеза формулируется в отношении разностей сопряженных пар наблюдений.

Условия применения: образуют $d_i = x_i - y_i$ – разности связанных пар результатов измерений и предполагает нормальное распределение этих разностей в генеральной совокупности с параметрами μ_d, σ_d .

Гипотеза H_0 : $\mu_d = 0$

Альтернативная гипотеза: $\mu_d \neq 0$

Рассчитываем t-критерий Стьюдента по формуле:

$$t_{\text{набл}} = \frac{\bar{d}}{S_d} \sqrt{n} \quad , \quad (18)$$

где \bar{d} – среднее арифметическое разностей, S_d – стандартное отклонение, n – объем выборки.

После расчёта значение $t_{\text{набл}}$ сравнивается с $t_{\text{кр}}$ (см. Приложение 2). Если $t_{\text{набл}} \geq t_{\text{кр}}$, то гипотеза о равенстве генерального среднего нулю отвергается на уровне значимости α

Критерий знаковых рангов Вилкоксона (проверка гипотезы о равенстве нулю медианы генеральной совокупности).

Этот критерий применяется при проверке гипотезы о различии двух связанных выборок (например, исследование одних и тех же спортсменов до и после эксперимента).

Условия применения: этот критерий используется для любого распределения связанных выборок.

Гипотеза H_0 : $Me_d = 0$

Альтернативная гипотеза: $Me_d \neq 0$

Пример: С помощью критерия знаковых рангов Вилкоксона, сравнить вес до (vesdo) и после эксперимента (vespo) и сформулировать вывод на основе полученного результата анализа. Данные взять из таблицы 14.

Таблица 14. Экспериментальные данные веса женщин до и после эксперимента.

№	vesdo	vespo	d_i	Ранги $ d_i $
1	59,1	58,0	1,1	9 (+)
2	62,3	62,3	0	
3	58,6	58,0	0,6	4 (+)
4	60,2	59,1	1,1	9 (+)
5	63,4	60,2	3,2	21 (+)
6	78,6	68,3	10,3	29 (+)
7	55,4	57,0	-1,6	11 (-)
8	64,9	60,8	4,1	23 (+)
9	65,0	62,0	3,0	18,5 (+)
10	63,2	60,7	2,5	15,5 (+)
11	72,9	65,4	7,5	28 (+)
12	56,1	55,0	1,1	9 (+)
13	69,2	62,9	6,3	27 (+)
14	68,8	64,0	4,8	26 (+)
15	58,8	58,0	0,8	5 (+)
16	66,2	64,2	2,0	13 (+)
17	62,5	62,8	-0,3	3 (-)
18	68,7	65,6	3,1	20 (+)
19	62,2	60,0	2,2	14 (+)
20	66,5	62,0	4,5	25 (+)
21	62,0	60,3	1,7	12 (+)
22	60,1	60,0	0,1	1 (+)
23	63,2	60,7	2,5	15,5 (+)
24	61,3	60,3	1,0	7 (+)
25	60,3	59,4	0,9	6 (+)
26	65,1	62,4	2,7	17 (+)
27	70,2	66,0	4,2	24 (+)
28	65,3	62,3	3,0	18,5 (+)
29	58,2	58,0	0,2	2 (+)
30	64,5	60,9	3,6	22 (+)
			R (-)	14
			R (+)	415

Решение:

1. Вычислить разности $d_i = x_i - y_i$ (таблица 14, столбец 4); пары с одинаковыми значениями отбрасываются;

2. Находим ранги абсолютных значений разностей (таблица 14, столбец 5) и отмечаем ранги, относящиеся к положительным и отрицательным значениям разностей;

3. Находим суммы рангов отрицательных R(-) и положительных R(+) разностей, меньшая из сумм рангов принимается в качестве значения критерия $W_{набл.}$ в нашем примере $W_{набл.}=14$;

4. Из таблицы критических значений W -критерия Вилкоксона для сопряженных пар (приложение 5) находим $W_{кр.}$. И сравниваем с $W_{набл.}$.

5. Если $W_{набл.} \leq W_{кр.}$, то нулевая гипотеза отбрасывается – наблюдаемое различие связанных выборок является статистически значимым на уровне значимости α . В нашем примере $W_{кр.}=69$ при $\alpha=0,01$. Так как $W_{набл.} < W_{кр.}$, то различие в весе женщин до и после эксперимента достоверны, $P < 0,01$.

Форма выполнения практического занятия: решение профессиональной задачи:

1. письменно в тетради сравнить связанные выборки с помощью t-критерия Стьюдента и критерия знаковых рангов Вилкоксона.

2. на компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0. сравнить связанные выборки с помощью t-критерия Стьюдента, критерия знаков и критерия знаковых рангов Вилкоксона.

Задача: Для выявления влияния большого объема скоростно-силовой нагрузки на уровень скоростно-силовых способностей спринтеров, имеющих квалификацию I разряда и мастера спорта, был проведен педагогический эксперимент в течение 5 недель, когда спортсмены выполняли ежедневную нагрузку большего объема. В начале и конце этого периода у них были измерены результаты в тройном прыжке с места, по которым оценивался уровень скоростно-силовых способностей. Ниже приведены экспериментальные данные 10 спринтеров, участвующих в эксперименте.

Таблица 15. Результаты в тройном прыжке с места до и после эксперимента.

До эксперимента (x_i , см)									
920	911	887	928	943	932	905	921	932	917
После эксперимента (y_i , см)									
903	891	863	891	930	924	891	898	892	890

Ход выполнения практического занятия:

Математическое решение задачи (в тетради):

1. t-критерий Стьюдента:
 - a. Проверить близость распределения к нормальному закону;
 - b. Вычислить разности $d_i = x_i - y_i$;
 - c. Найти среднее арифметическое разностей \bar{d} ;
 - d. По формул 18 найти значение $t_{набл.}$;
 - e. По таблице критических значений найти $t_{кр.}$;
 - f. Сравнить $t_{набл.}$ с $t_{кр.}$ и сформулировать вывод.
2. Критерий знаковых рангов Вилкоксона:
 - a. Вычислить разности $d_i = x_i - y_i$;
 - b. Найти ранги абсолютных значений разностей и отметить ранги, относящиеся к положительным и отрицательным значениям разностей;
 - c. Найти суммы рангов отрицательных и положительных;
 - d. Меньшую из сумм рангов принять в качестве $W_{набл.}$;
 - e. По таблице критических значений найти $t_{кр.}$;
 - f. Сравнить $t_{набл.}$ с $t_{кр.}$ и сформулировать вывод.

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. В статистическом пакете **STATGRAPHICS Plus 5.0** создать файл JUMP.
2. В электронной таблице создать два столбца с именами jumpdo и jumpro.
3. Проверить закон распределения экспериментальных данных для того, чтобы определить с помощью какого критерия будут сравниваться связанные выборки.
4. Выбрать пункт меню Compare / Two Samples / Paired Samples Comparison. Появится окно Paired Samples Comparison, в котором нужно в поле Sample 1 занести результаты в тройном прыжке с места до эксперимента, а в поле Sample 2 - после эксперимента, нажать кнопку ОК. Появится окно общего анализа (Analysis Summary), в котором сообщается, что 10 разностей варьируются от 8,0 до 40,0.
5. Щелкнуть на кнопке табличных опций (Tabular Options), и выбрать Hypothesis Tests (критерии для проверки гипотез). Появится окно результатов анализа (рис.42)

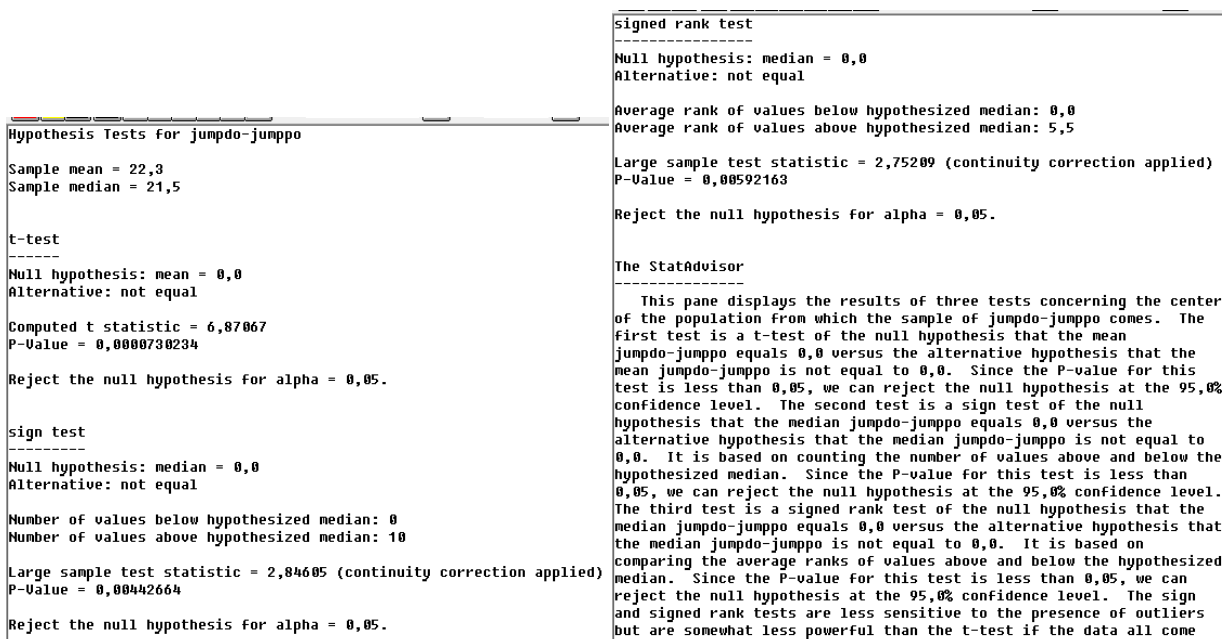


Рис.42. Окно результата анализа сравнения двух выборок.

6. В этом окне приводятся результаты анализа данных с использованием трех тестов: одного параметрического (t-test) и двух непараметрических: критерия знаков (sign test) и критерия знаковых рангов Вилкоксона (signed rank test). После выводов результатов анализа данных с использованием всех трех критериев СтатКонсультант сообщает. Что в этом окне приводятся результаты трех тестов. Описываются условия применения каждого из тестов. Кроме того, дается краткая методика расчета с использованием каждого из критериев.

Примечание: если предположение о соответствии распределения результатов нормальному закону не подтверждается, то нужно использовать два других критерия, которые обладают меньшей мощностью, однако, и требования к исходным данным у них ниже.

7. Сделать выводы самостоятельно и записать в окно StatReporter.

8. Сохранить файлы с данными, результатами статистической обработки и статистическими выводами в папку со своей фамилией и номером группы на рабочем столе с именем «JUMP».

Практическое задание № 9

«Решение регрессионных задач в спорте»

по теме «Регрессионный и корреляционный анализ».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение выявлять и анализировать зависимости случайных величин между ними и представлять результаты в виде статистического вывода.

Теоретические сведения к практическому занятию:

Регрессионный анализ - это метод установления аналитического выражения стохастической зависимости между исследуемыми признаками. Уравнение регрессии показывает, как в среднем изменяется у при изменении любого из x_i .

Если независимая переменная одна - это простой регрессионный анализ. Если же их несколько (п 2), то такой анализ называется многофакторным.

В ходе регрессионного анализа решаются две основные задачи:

1. Построение уравнения регрессии, т.е. нахождение вида зависимости между результатным показателем и независимыми факторами x_1, x_2, \dots, x_p .

2. Оценка значимости полученного уравнения, т.е. определение того, насколько выбранные факторные признаки объясняют вариацию признака у.

Применяется регрессионный анализ главным образом для планирования, а также для разработки нормативной базы.

В отличие от корреляционного анализа, который только отвечает на вопрос, существует ли связь между анализируемыми признаками, регрессионный анализ дает и ее формализованное выражение. Кроме того, если корреляционный анализ изучает любую взаимосвязь факторов, то регрессионный - одностороннюю зависимость, т.е. связь, показывающую, каким образом изменение факторных признаков влияет на признак результативный.

Регрессионный анализ - один из наиболее разработанных методов математической статистики. Строго говоря, для реализации регрессионного анализа необходимо выполнение ряда специальных требований (в частности, x_1, x_2, \dots, x_n ; y должны быть независимыми, нормально распределенными случайными величинами с постоянными дисперсиями). В реальной жизни строгое соответствие требованиям регрессионного и корреляционного анализа встречается очень редко, однако оба эти метода весьма распространены в педагогических исследованиях. Зависимости в физической культуре и спорте могут быть не только прямыми, но и обратными и нелинейными. Регрессионная модель может быть построена при наличии любой зависимости, однако в многофакторном анализе используют только линейные модели.

Построение уравнения регрессии осуществляется, как правило, методом наименьших квадратов, суть которого состоит в минимизации суммы квадратов отклонений фактических значений результатного признака от его расчетных значений, т.е.:

$$j = a + b_1x_1 + b_2x_2 + \dots + b_nx_n - \text{расчетное значение результатного фактора.}$$

Коэффициенты регрессии рекомендуется определять с помощью аналитических пакетов для персонального компьютера или специального калькулятора.

Необходимо отметить, что в педагогических исследованиях корреляционный и регрессионный анализы нередко объединяются в один - корреляционно-регрессионный анализ. Подразумевается, что в результате такого анализа будет построена регрессионная зависимость (т.е. проведен регрессионный анализ) и рассчитаны коэффициенты ее тесноты и значимости (т.е. проведен корреляционный анализ). В известном смысле корреляционная связь носит более общий характер, поскольку она не предполагает наличия зависимости «причина - следствие».

Ход выполнения практического занятия:

1. Подготовка данных для использования пакета *Statgraphics Plus*

Данные могут быть введены непосредственно в *Statgraphics Plus* или могут быть импортированы из текстовых файлов.

1) ввести исходные данные или открыть существующий текстовый файл с анализируемыми данными:

№	Y	X1	X2	X3
1	74	33	56	10
2	84	34	58	8
3	73	36	67	7
4	93	35	70	9
5	56	33	73	8
6	71	37	77	7
7	117	39	78	12
8	111	42	99	9
9	135	43	93	13
10	125	44	96	12

Для импортирования файла в главное меню необходимо найти и выполнить команду чтения файла с исходными данными: ***File\ Open\ Open Data File***

В появившемся окне указать имя папки, тип файла *All Files* \ имя файла *Вариант X\Открыть*. В новом появившемся окне *Read ASCII File* нажать кнопку *tab delimited* и клавишу **OK**.

В нижней левой части экрана появится командное меню с надписью *Вари...* Нажав клеточную клавишу вызвать команду *Развернуть*. При этом на экране появится таблица с данными варианта задания:

Таблица 16. Исходные данные для регрессионного анализа

	Col 1	Col 2	Col 3	Col 4	Col 5
1	№	Y	X1	X2	X3
2	1	74	33	56	10
3	2	84	34	58	8
4	3	73	36	67	7
5	4	93	35	70	9
6	5	56	33	73	8
7	6	71	37	77	7
8	7	117	39	78	12
9	8	111	42	99	9
10	9	135	43	93	13
11	10	125	44	96	12

2) Редактировать данные

Убрать с экрана лишние надписи путем удаления соответствующих строк и колонок таблицы данных. Для удаления необходимо мышью пометить удаляемые данные, нажатием правой клавиши мышки вызвать контекстное меню и выполнить команду *Delete*. В рассматриваемом примере необходимо убрать первую колонку с порядковыми номерами и первую строку с идентификаторами переменных. При этом необходимо запомнить размещение данных во вновь полученной таблице:

Таблица 17. Регрессионный анализ

	Col_1[1]	Col_2	Col_3	Col_4
1	507,20	19,50	359,90	448,10
2	506,60	19,80	187,10	451,90
3	487,80	21,10	375,20	447,90
4	496,00	18,60	287,90	444,30
5	493,60	19,60	444,00	443,20
6	458,90	11,70	462,40	411,70
7	429,30	10,50	459,50	328,60
8	386,90	13,60	511,30	314,70
9	311,50	10,80	328,60	259,40
10	302,20	10,90	350,00	187,70
11	262,00	10,30	298,70	238,50
12	242,20	10,60	529,30	269,40
13	231,90	8,50	320,00	284,00
	214,30	6,70	502,00	172,30
	208,40	8,30	194,90	166,40

В первой колонке содержатся данные зависимой переменной Y, а во второй, третьей и четвертой колонках – соответственно X1, X2 и X3.

3) Сохранить данные.

Каждой переменной в файле данных *Statgraphics Plus* будет соответствовать отдельная колонка. Используя меню *File/SAVE/Save Date File As...*, сохраните данный файл с расширением, <имя файла>.sf3 или с расширением .txt.

2. Решение задач в *Statgraphics Plus*

2.1. Выбор факторных признаков для построения регрессионной модели на основе анализа матрицы коэффициентов корреляции.

Для построения матрицы парной корреляции всех переменных с помощью пакета *Statgraphics Plus* необходимо выполнить следующие действия:

1) в главном меню выбрать и нажать клавишу *Summary stats*. При этом на экране появится диалоговое окно *Multiple-Variable Analysis*;

2) заполнить диалоговое окно ввода данных идентификаторами колонок, в которых содержатся данные для анализа. В рассматриваемом примере необходимо ввести: *Col_1, Col_2, Col_3*. Нажать кнопку *OK*.

3) в появившемся подменю *Multiple-Variable Analysis* выбрать и нажать клавишу *Tabular options*. В окне табличных настроек поставить флажок напротив *Correlations*, нажать клавишу *OK*. При этом на экране появиться матрица коэффициентов парной корреляции.

4) Для записи матрицы корреляции в файл с исходными данными вызвать пункт подменю *Save results*, в окне *Correlations* установить флажок и нажать *OK*. Файлу будет присвоен идентификатор *CMAT*. Матрица коэффициентов парной корреляции будет продолжением таблицы с исходными данными с колонками *CMAT_1, CMAT_2* и т.д. Матрица коэффициентов парной корреляции для рассматриваемого примера имеет вид:

Таблица 18. Матрица коэффициентов парной корреляции

	CMAT_1	CMAT_2	CMAT_3	CMAT_4
1	1,0	0,883818	0,0288061	0,937679
2	0,883818	1,0	0,181598	0,888499
3	0,0288061	0,181598	1,0	0,0333198
4	0,937679	0,888499	0,0333198	1,0

Значения коэффициентов парной корреляции указывают на весьма тесную связь зависимой переменной y с факторными переменными x_1 , x_2 и x_3 . В то же время межфакторная связь $r_{x_1x_2} = 0,921$ весьма тесная и превышает тесноту связи x_2 с y , что свидетельствует о наличии мультиколлинеарности. В связи с этим для улучшения разрабатываемой модели необходимо исключить из ее структуры фактор x_2 как малоинформативный и недостаточно статистически надежный.

Коэффициенты частной корреляции дают более точную характеристику тесноты связи двух переменных, чем коэффициенты парной корреляции, т.к. очищают парную зависимость от воздействия других переменных.

5) Для отображения на экране результатов вычисления линейных коэффициентов частной корреляции необходимо в окне табличных настроек *Tabular options* поставить флажок напротив *Partial Correlations*, а в подменю *Save results* в окне *Partial Correlations* установить флажок и нажать *OK*. Файлу будет присвоен идентификатор *PMAT*. Матрица коэффициентов частной корреляции будет продолжением матрицы коэффициентов парной корреляции с колонками *PMAT_1, PMAT_2, PMAT_3* и *PMAT_4*. Матрица коэффициентов частной корреляции для рассматриваемого примера имеет вид:

Таблица 19. Матрица коэффициентов частной корреляции

	PMAT_1	PMAT_2	PMAT_3	PMAT_4
1		0,3394	0,125615	0,684299
2	0,3394		-0,351905	0,383612
3	0,125615	-0,351905		0,119295
4	0,684299	0,383612	0,119295	

Как следует из полученной матрицы, наиболее тесно связан y с факторами x_1 , x_3 , коэффициенты частной корреляции равны $r_{yx1.x2}=0,712$, $r_{yx3.x2}=0,737$. Связь y с фактором x_2 гораздо слабее и равна $r_{yx2.x1}=-0,371$. Межфакторная зависимость x_1 и x_2 , равная $r_{x1x2,y}=0,855$ выше, чем парная y и x_2 : $r_{yx2.x1}=-0,371$.

Все это приводит к выводу о необходимости исключить фактор x_2 из правой части уравнения множественной регрессии.

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи между анализируемыми переменными.

На основе анализа матрицы коэффициентов парной корреляции делаем вывод о целесообразности построения двухфакторного регрессионного уравнения $y=f(x_2, y)$

2.2. Построение линейного уравнения регрессии.

По заданию необходимо построить регрессионную модель со значимыми факторами в линейной форме.

Оценку параметров уравнения регрессии необходимо осуществить по методу наименьших квадратов. Для проведения регрессионного анализа с помощью пакета **Statgraphics Plus** выполнить следующие действия:

1) В главном меню вызвать программу **Multiple regression**. Заполнить диалоговое окно ввода данных. В окне **Dependent Variable** занести имя колонки с результирующей переменной – **Col_3** В окне **Independent Variables** занести имена колонок с отобранными факторными переменными – **Col_3 Col_2** Нажать клавишу **OK**. Результат вычисления функции появится в отдельном окне **Multiple regression Analysis**. Для рассматриваемого примера результаты вычислений будут следующими:

Multiple Regression Analysis

Dependent variable: Col_3

Parameter	Standard Estimate	T Error	Statistic	P-Value
CONSTANT	427,592	85,3512	5,00979	0,0002
Col_2	-4,00536	6,01557	-0,665832	0,5172

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	5448,7	1	5448,7	0,44	0,5172
Residual	159775,0	13	12290,4		

Total (Corr.) 165223,0 14
R-squared = 3,29778 percent
R-squared (adjusted for d.f.) = 0,0 percent
Standard Error of Est. = 110,862
Mean absolute error = 90,2232
Durbin-Watson statistic = 2,3116

Рис.43. Окно результата регрессионного анализа.

В первой колонке *Parameter* перечисляются используемые в модели независимые переменные, а во второй колонке *Estimate* содержатся коэффициенты уравнения регрессии.

По результатам вычислений составим уравнение множественной регрессии вида:

$$y = 427,592 - 4,00536 * x_2$$

Значения случайных ошибок параметров уравнения регрессии записаны в колонке *Standard Error*:

$$m_{b0}=85,3512; m_{b1}= 6,01557;$$

Они показывают, какое значение данной характеристики сформировалось под влиянием случайных факторов. Эти значения использованы при расчетном определении *t*-критерия Стьюдента, приведенные в колонке *T Statistic*:

$$t_{b0}=5,00979 \quad t_{b1}=-0,665832$$

Если расчетное значение критерия больше табличного значения, то можно делать вывод о существенности данного параметра, который формируется под воздействием неслучайных величин x_1 и x_3 . В противном случае принимается гипотеза о случайной природе значения коэффициента уравнения, и соответствующая переменная исключается из структуры модели как неинформативный фактор.

Табличное значение *t*-критерия Стьюдента для уровня значимости $\alpha=0,05$ и числа степеней свободы $f=13$ равно $t_m=2,16$

В рассматриваемом примере расчетные значения *t*-критерия Стьюдента больше табличного значения $t_m=2,16$, что подтверждает статистическую значимость параметров b_0 , b_1 .

Величина b_0 оценивает агрегированное влияние неучтенных факторов на результирующую переменную y . Величины b_1 , b_3 указывают, что с увеличением x_1 и x_3 на единицу их значений результирующая переменная увеличится соответственно на 4,00536. Сравнивать эти значения между собой не следует, т.к. они зависят от единицы измерения каждого признака и поэтому не сопоставимы между собой.

2) Оценить надежность уравнения регрессии в целом и коэффициента детерминации.

Во второй таблице *Analysis of Variance* приводятся результаты дисперсионного анализа. В первой строке таблицы *Model* приводится дисперсия регрессии, во второй *Residual* – дисперсия остатка, а в третьей - *Total (Corr.)* – общая дисперсия. Колонки слева на право содержат: сумму квадратов – *Sum of Squares*, число степеней свободы – *Df*, оценку дисперсии – *Mean Square*, расчетный критерий Фишера – *F-Ratio*.

Оценку надежности уравнения регрессии в целом и коэффициента детерминации дает расчетное значение *F* - критерия Фишера, равно $F_p=0,44$. Сравнение расчетного значения критерия с табличным позволяет проверить гипотезу об адекватности модели и значимости коэффициента детерминации. Табличное значение *F* – критерия Фишера для уровня значимости $\alpha=0,05$ и числа степеней свободы $f_1=1$, $f_2=13$ равно $F_T=4,67$. Расчетное значение *F* - критерия Фишера больше табличного значения $F_p=0,44$ больше $F_T=4,74$. Значит модель неадекватна.

Уравнение регрессии следует признать неадекватным, модель считается незначимой.

В нижней части таблицы приводятся:

Коэффициент детерминации *R-squared* в процентах;

R-squared = 3,29778 percent

Нормированный коэффициент *R*-квадрат *R-squared(adjusted for d.f.)*;

R-squared (adjusted for d.f.) = 0,0 percent

Стандартная ошибка *Standard Error of Est.*;

Standard Error of Est. = 110,862

Среднее значение абсолютной ошибки *Mean absolute error*;

Mean absolute error = 90,2232

Статистика Дурбина – Ватсона *Durbin-Watson statistic*- 2,3116

Коэффициент детерминации R-squared $R^2= 3,29778 \%$ показывает долю вариации результативного признака под воздействием изучаемых факторов. Следовательно, около 3,29% вариации зависимой переменной учтено в модели и обусловлено вариацией включенных факторов.

Коэффициент множественной корреляции $R=6,01557$ показывает тесноту связи зависимой переменной y со всеми включенными в модель объясняющими факторами x_1, x_3 .

Скорректированный коэффициент детерминации ***R-squared(adjusted for d.f.)*** равен 0,0 %. Он определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсии и позволяет сравнивать модели с разным числом факторов. Оба коэффициента детерминации указывают на высокую детерминированность результатов y факторами x_1 и x_3 .

Статистика Дурбина – Ватсона ***Durbin-Watson statistic=2,3116*** используется при проверке коррелированности остатков модели.

2.3. Расчет частных коэффициентов эластичности

1) Для получения статистических оценок по исходным данным и результатам вычислений необходимо вызвать из главного меню ***Statgraphics Plus*** программу ***Summary ststis***. При этом на экране появится диалоговое окно ***Multiple-Variable Analysis***. Заполнить диалоговое окно ввода данных идентификаторами колонок, в которых содержатся данные для анализа. В рассматриваемом примере необходимо ввести: ***Col_2, Col_3***. Нажать кнопку ***OK***.

2) В появившемся подменю ***Multiple-Variable Analysis*** выбрать и нажать клавишу ***Tabular options***. В окне табличных настроек поставить флажок напротив ***Summary ststistics***, нажать клавишу ***OK***. При этом на экране появится таблица со статистическими данными переменных, записанных в колонках ***Col_2, Col_3***:

Summary Statistics

	<i>Col_2</i>	<i>Col_3</i>
<i>Count</i>	15	15
<i>Average</i>	13,3667	374,053
<i>Variance</i>	24,2595	11801,7
<i>Standard deviation</i>	4,9254	108,635
<i>Minimum</i>	6,7	187,1
<i>Maximum</i>	21,1	529,3
<i>Std. skewness</i>	0,756068	-0,36014
<i>Std. kurtosis</i>	-1,14019	-0,675618
<i>Sum</i>	200,5	5610,8

Рис.44. Окно числовых характеристик.

Строки таблицы в порядке следования сверху вниз содержат следующие статистические данные по каждой переменной: количество наблюдений, среднее значение, дисперсия, стандартное отклонение, минимальное значение, максимальное значение, коэффициент асимметрии, коэффициент крутизны, сумма.

Средние значения ***Average*** использовать для расчета коэффициентов эластичности регрессионной модели.

3) Рассчитать частные коэффициенты эластичности по формулам:

$$\overline{\varepsilon}_{y x_i} = b_i \frac{\overline{x_i}}{\overline{y}}, \quad (19)$$

где b_i – коэффициент регрессии при x_i в уравнении множественной регрессии;

\bar{x}_i – среднее значение i -го фактора;

\bar{y} - среднее значение результирующей переменной.

Для решаемой задачи получаем:

$$\varepsilon_{yx1} = (6,01557 \cdot -0,4) / 374 = -169,167,$$

По значениям частных коэффициентов эластичности можно сделать вывод о более сильном влиянии на результирующий признак y фактора x_1 . При изменении на 1% факторной переменной x_1 результирующая переменная y изменится на 1,4%.

2.4. Построение прогноза

1) Для построения прогноза вначале необходимо ввести прогнозные значения выбранных факторов $x_1=38$ и $x_3=10$ в таблицу с исходными данными:

Таблица 20. Исходные данные для построения прогноза

	Col_1[1]	Col_2	Col_3	Col_4
1	74	33	56	10
2	84	34	58	8
3	73	36	67	7
4	93	35	70	9
5	56	33	73	8
6	71	37	77	7
7	117	39	78	12
8	111	42	99	9
9	135	43	93	13
10	125	44	96	12
11		38		10

2) Рассчитать прогнозируемое значение результирующей переменной (точечный прогноз), для чего в меню программы **Multiple regression** нажать клавишу **Save results..** В окне сохранения результатов **Save results Options** поставить флажок напротив прогнозного значения - **Predicted Values**, нажать клавишу **OK**. В окне табличных данных появится колонка прогноза - **Predicted** с рассчитанными по регрессионной модели данными:

Predicted

80,1464

72,0052

73,3791

81,5203

68,3756

77,0087

113,695

106,928

134,099

131,843

98,2946

Последняя строка таблицы содержит прогнозируемое значение результирующей переменной:

$$y_{np} = 98,2946.$$

3) Определить интервальные значения прогнозируемой переменной для уровня значимости $\alpha=0,05$.

В окне сохранения результатов **Save results Options** поставить два флажка напротив нижний предел - **Lower Limits for Predictions** и верхний предел **Upper Limits for Predictions**,

нажать клавишу **OK**. В окне табличных данных появятся колонки прогноза - **Predicted** с интервальными значениями: нижним пределом - **Lower Plims** и верхним пределом - **Upper Plims** рассчитанными для уровня значимости $\alpha=0,05$.

Predicted	Lower Plims	Upper Plims
80,1464	54,65	105,643
72,0052	48,6046	95,4058
73,3791	49,1401	97,618
81,5203	58,4868	104,554
68,3756	44,4929	92,2583
77,0087	52,3479	101,67
113,695	89,3879	138,002
106,928	81,6375	132,218
134,099	108,674	159,523
131,843	106,727	156,959
98,2946	75,6931	120,896

Рис.45. Окно табличных данных с прогнозом

Для рассматриваемого примера интервальный прогноз находится в нижней строке таблицы:

$$y_{np} \in [75,6931; 120,896]$$

2.5. Запись результатов вычислений в рабочую папку и завершение работы с программой **Statgraphics Plus**.

1) Результаты вычислений, отображаемые в окне пакета программ **Statgraphics Plus** записывать в свою папку в виде отдельных файлов. Для записи содержимого экрана необходимо выполнить следующую последовательность команд:

File | **Save As** | **Save StatFollio As...** | далее укажите <имя папки> для записи, тип файла **StatFollios** | ***.sgp** | < имя файла >.

2) Завершить работу с пакетом программ **Statgraphics Plus**.

Практическое задание № 10

«Расчёт коэффициентов корреляции»

по теме «Начала корреляционного анализа».

Время, отведенное на выполнение практического занятия: 2 часа.

Цель: сформировать умение выявлять и анализировать зависимости случайных величин между ними и представлять результаты в графическом виде.

Теоретические сведения к практическому занятию:

Корреляция - это зависимость между двумя переменными величинами.

Корреляционное поле - это графическое представление статистических данных в декартовой (прямоугольной ХОУ) системе координат, где каждой паре (X,Y) на плоскости соответствует точка.

Виды связей между двумя случайными величинами:

1. **Функциональная зависимость** - взаимосвязь, при которой каждому значению одного показателя соответствует строго определенное значение другого (рис.46).

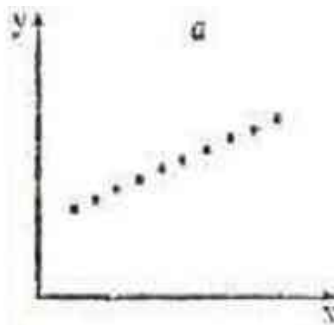


Рис.46. Графическое представление функциональной зависимости.

2. *Статистическая зависимость* - взаимосвязь, при которой одному значению первого показателя может соответствовать несколько значений второго показателя.

3. *Корреляционная зависимость* - зависимость заключается в том, что средняя величина одного показателя изменяется в зависимости от значения другого (рис.47).

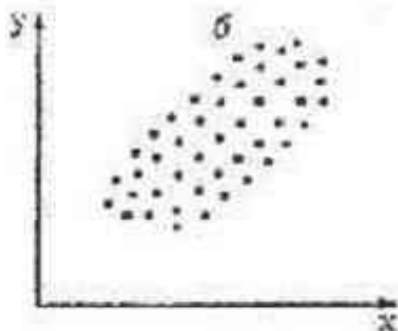


Рис.47. Графическое представление корреляционной зависимости.

4. *Независимые случайные величины* – связь отсутствует (рис.48).

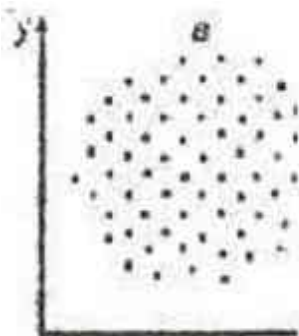


Рис.48. Графическое представление независимых величин.

Корреляционный анализ - раздел математической статистики, исследующий зависимость между случайными величинами на основе различных выборочных оценок генеральных коэффициентов корреляции.

Регрессионные зависимости (регрессия) - вид математической модели, описывающий зависимость одной случайной величины от ряда других, случайных или неслучайных величин. Записывается в виде функции, линейной или не линейной, которая описывает эту зависимость приближённо.

В качестве меры зависимости случайных величин в математической статистике наиболее распространены две оценки *коэффициента корреляции*:

- Брава-Пирсона (r);
- Спирмена (r_s).

Коэффициент корреляции Браве-Пирсона.

Для определения коэффициента корреляции Браве-Пирсона принимается предположение о двумерном нормальном распределении генеральной совокупности, из которой получены экспериментальные данные.

Коэффициент корреляции k Браве-Пирсона вычисляются по формуле:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2\right) \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2\right)}} \quad (20)$$

Коэффициент корреляции Спирмена.

Если требуется установить связь между двумя признаками, значения которых в генеральной совокупности распределены не по нормальному закону, то можно воспользоваться коэффициентом ранговой корреляции Спирмена r_s :

$$r_s = 1 - 6 \frac{\sum d_i^2}{n(n^2 - 1)} \quad (21)$$

где d_i – разность рангов i -той пары значений x_i и y_i .

Основные свойства оценок корреляции:

1. Обе оценки r и r_s принимают значения из сегмента $[-1; 1]$;
2. $r = +1$ или $r = -1$ тогда и только тогда, когда случайные величины X и Y связаны жёсткой линейной зависимостью, т.е. $Y = aX + b$ (на корреляционном поле изображающие точки ложатся на какую-либо прямую);
3. В случае нормального распределения системы $(X; Y)$ равенство $r = 0$ равносильно независимости случайных величин X и Y ;
4. Если же система $(X; Y)$ распределена не по нормальному закону, равенство $r = 0$ может выполняться и при наличии нелинейной зависимости случайных величин X и Y ;
5. Если $r \neq 0$, то чем ближе $|r|$ к единице, тем сильнее линейная зависимость случайных величин X и Y .

Значения принимаемые величиной r :

- $|r| = 1$ – линейная зависимость
- $0,7 \leq |r| \leq 0,99$ – сильная статистическая зависимость
- $0,5 \leq |r| \leq 0,7$ – средняя статистическая зависимость
- $0,2 \leq |r| \leq 0,5$ – слабая статистическая зависимость
- $0,09 \leq |r| \leq 0,2$ – очень слабая статистическая зависимость

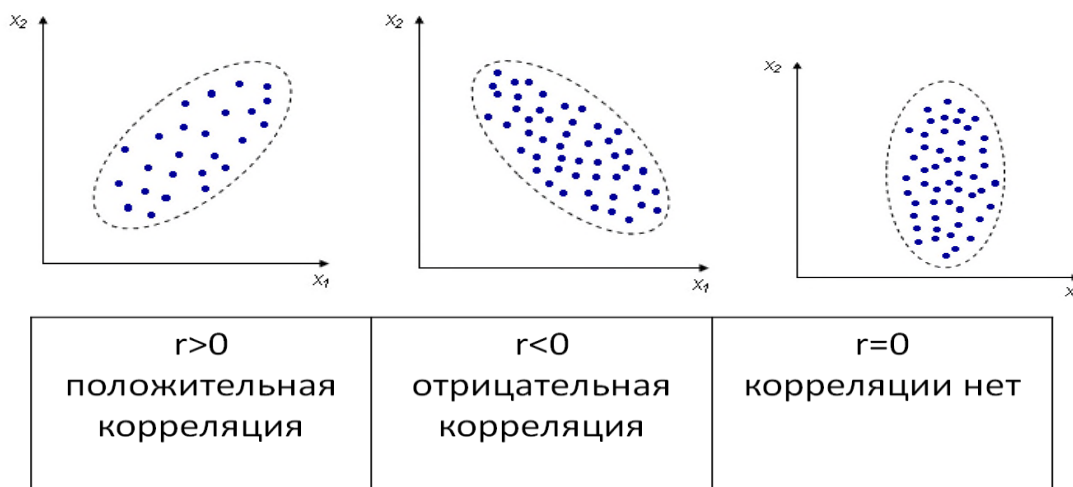


Рис.49. Значения корреляции (r).

Значимость коэффициента корреляции.

Для оценки значимости коэффициента корреляции необходимо сравнить его с r -критическим, величина которого зависит от объема выборки и уровня значимости. Если $r \geq r_{кр}$, то коэффициент корреляции статистически достоверен.

В научных публикациях необходимо указать объем выборки, чтобы читатель мог оценить достоверность вычисленных коэффициентов корреляции. Иногда в публикациях приводятся только значимые коэффициенты корреляции, а вместо незначимых ставятся прочерки. В таблице 17 авторы указали, что объем выборки $n=32$ и $\alpha=0,05$ составляет $r_{кр.}=0,349$ (см. Приложение 6). Следовательно, все коэффициенты корреляции достоверны.

Таблица 21. Коэффициенты корреляции между результатами в скоростно-силовых тестах и результатом в толкании ядра с разгоном ($n=32$, по Я.Е. Ланка, Ан.А. Шалманов, 1982).

Упражнение	1	2	3	4	5	6
1 Толкание ядра с разгона	1	0,97	0,84	0,83	0,73	0,73
2 Толкание ядра с места		1	0,84	0,82	0,74	0,76
3 Бросок ядра назад			1	0,85	0,71	0,66
4 Бросок ядра вперед				1	0,66	0,62
5 Приседание со штангой					1	0,58
6 Жим штанги лежа						1

Форма выполнения практического занятия: решение профессиональной задачи:

1. письменно в тетради вычислить коэффициент корреляции Спирмена и оценить значимость найденных значений;

2. на компьютере с помощью программы статистической обработки данных STATGRAPHICS Plus 5.0 вычислить коэффициенты корреляции Браве-Пирсона и Спирмена, оценить значимость результатов и представить результаты в графическом виде.

Задача: Для совершенствования методики педагогического контроля физической подготовленности юных тяжелоатлетов весовой категории до 60 кг, имеющих спортивную квалификацию, близкую к I разряду, изучалась взаимосвязь отдельных упражнений для оценки физической подготовленности со спортивным результатом. Ниже приведены результаты в одном из упражнений (прыжок в высоту с места) и спортивный результат в толчке для 12 спортсменов данной категории (таблица 16):

Таблица 22. Результаты упражнения (прыжок в высоту с места) и спортивный результат в толчке

Прыжок в высоту с места, см:											
57	60	58	61	63	58	55	64	65	64	66	61
Результат в толчке, кг:											
107,5	110	110	115	115	107,5	107,5	120	122,5	112,5	120	110

Вычислить коэффициенты корреляции Браве-Пирсона и Спирмена, оцените значимость найденных коэффициентов и постройте корреляционное поле.

Ход выполнения практического занятия:

Математическое решение задачи (в тетради):

1. Вычислить коэффициент корреляции Спирмена по формуле 20;
2. Сравнить результаты с $r_{кр.}$;
3. Оценить значимость найденного коэффициента корреляции и сделать вывод.

Решение задачи в пакете STATGRAPHICS Plus 5.0 (на компьютере):

1. В статистическом пакете **STATGRAPHICS Plus 5.0** создать файл CORREL.
2. Внести данные из таблицы 16.
3. Для расчета коэффициентов корреляции Браве-Пирсона и Спирмена необходимо выбрать Describe / Numeric Data / Multiple Variable Analysis. В появившемся диалоговом окне Multiple Variable Analysis (рис.50) перенести переменные в поле Data. Нажать ОК.

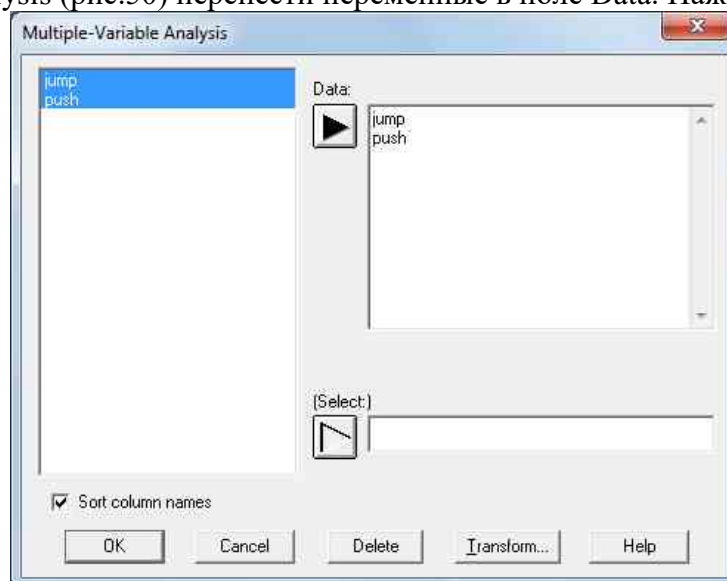


Рис.50. Диалоговом окне Multiple Variable Analysis

4. Появится окно общего анализа Analysis Summary (рис.51). в этом окне сообщается название переменных, а также количество случаев, содержащихся в переменной.

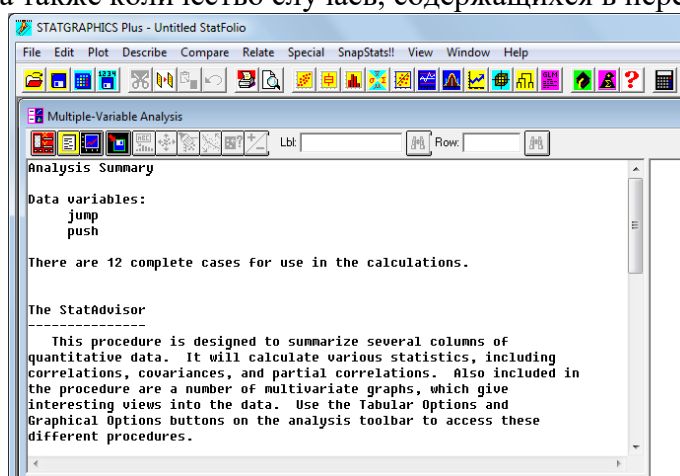


Рис.51. Окно общего анализа Analysis Summary.

5. Для расчета коэффициентов корреляции необходимо щелкнуть на кнопке табличных опций (Tabular Options). В появившемся диалоговом окне выбрать Correlation (коэффициент корреляции Браве-Пирсона) и Rank Correlation (коэффициент корреляции Спирмена) (рис.52). Щелкнуть ОК.

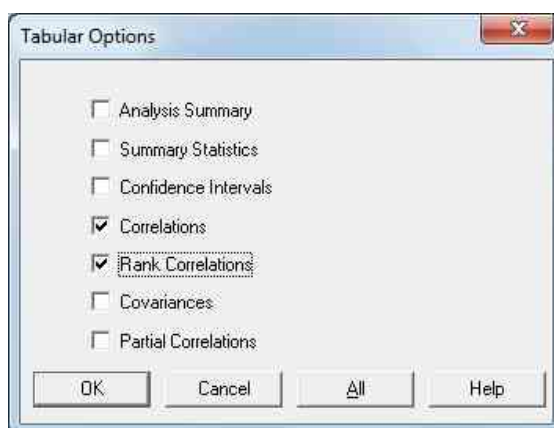


Рис.52. Диалоговое окно *Tabular Options*.

6. Появятся два окна: *Correlations* и *Spearman Rank Correlations* (рис.53).

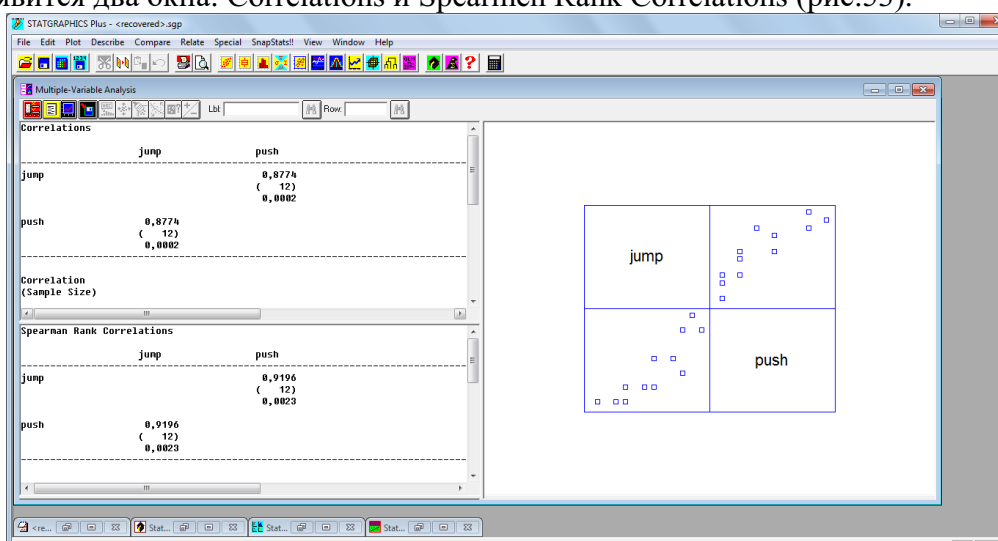


Рис.53. Окна *Correlations* и *Spearman Rank Correlations* и корреляционное поле.

Рассмотрим окно *Correlations*: в нем представлен коэффициент корреляции Браве-Пирсона, равный 0,8774, число парных измерений 12, *P-value* равен 0,0002.

Следует обратить внимание на сообщение СтатКонсультанта (StatAdvisor) в окне *Correlations*. Последнее предложение имеет следующий вид: *The following pairs of variables have P-value below 0,05: jump and push.*

СтатКонсультант сообщает, что *P-value* меньше 0,05 имеют следующие пары переменных: *jump and push*. Анализ этого сообщения позволяет быстро выявить все достоверные коэффициенты корреляции.

Таким образом, взаимосвязь результатов прыжков в высоту с места и результатов в толчке у тяжелоатлетов первого разряда высокая $r=0,8774$, коэффициент корреляции достоверен, $P<0,001$.

7. Самостоятельно проанализировать окно *Spearman Rank Correlation* и выводы по двум коэффициентам корреляции записать в окне *StatReporter*.

8. Чтобы наглядно представить взаимосвязь между результатами нужно построить корреляционное поле. Для этого необходимо выбрать в меню команду *Plot / Scatterplots / X-Y Plot*. В появившемся диалоговом окне переменную *jump* занести в поле *X*, а переменную *push* – в поле *Y*. Построенное корреляционное поле скопировать в окно *StatGallery*.

9. Сохранить файлы с данными, результатами статистической обработки, статистическими выводами и графическими данными в папку со своей фамилией и номером группы на рабочем столе с именем «CORREL».

ПРИЛОЖЕНИЯ

Приложение 1

Критические значения F-критерия Фишера

(верхние числа в строке соответствуют уровню значимости 0,05; средние — 0,01; нижние — 0,001)

		$v_1 = n_1 - 1$ — число степеней свободы для большей дисперсии											
		4	5	6	7	8	9	10	12	16	20	40	100
$v_2 = n_2 - 1$ — число степеней свободы для меньшей дисперсии	4	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,8	5,8	5,7	5,7
		16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,1	14,0	13,7	13,6
		53,4	51,7	50,5	49,8	49,0	48,6	48,2	47,4	46,6	46,2	45,4	44,7
	5	5,2	5,0	5,0	4,9	4,8	4,8	4,7	4,7	4,6	4,6	4,5	4,4
		11,4	11,1	10,7	10,5	10,3	10,3	10,1	9,9	9,7	9,6	9,3	9,1
		31,1	29,8	28,8	28,2	27,6	27,3	27,0	26,4	25,8	25,4	24,8	24,3
	6	4,5	4,4	4,3	4,2	4,2	4,1	4,0	4,0	3,9	3,9	3,8	3,7
		9,2	8,8	8,5	8,3	8,1	8,0	7,9	7,7	7,5	7,4	7,1	7,0
		21,9	20,8	20,0	19,5	19,0	18,8	18,5	18,0	17,5	17,2	16,6	16,2
	7	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,5	3,4	3,3	3,3
		7,9	7,5	7,2	7,0	6,8	6,7	6,6	6,5	6,3	6,2	5,9	5,8
		17,2	16,2	15,5	15,1	14,6	14,4	14,2	13,7	13,2	13,0	12,5	12,1
8	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,2	3,2	3,1	3,0	
	7,0	6,6	6,4	6,2	6,0	5,9	5,8	5,7	5,5	5,4	5,1	5,0	
	14,4	13,5	12,9	12,5	12,0	11,8	11,6	11,2	10,8	10,5	10,1	9,7	
9	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	3,0	2,9	2,8	2,8	
	6,4	5,1	5,8	5,6	5,4	5,5	5,3	5,1	4,9	4,8	4,6	4,4	
	12,6	11,7	11,1	10,8	10,4	10,2	10,0	9,6	9,2	8,9	8,5	8,1	
10	3,5	3,3	3,2	3,1	3,1	3,0	3,0	3,0	2,9	2,8	2,7	2,6	
	6,0	5,6	5,4	5,2	5,1	5,0	4,9	4,7	4,5	4,4	4,1	4,0	
	11,3	10,5	9,9	9,6	9,2	9,0	8,9	8,5	8,1	7,8	7,4	7,1	
12	3,3	3,1	3,0	2,9	2,9	2,8	2,8	2,7	2,6	2,5	2,4	2,4	
	5,4	5,1	4,8	4,7	4,5	4,4	4,3	4,2	4,0	3,7	3,6	3,5	
	9,6	8,9	8,4	8,1	7,7	7,5	7,4	7,0	6,7	6,5	6,1	5,7	
14	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4	2,4	2,3	2,2	
	5,0	4,7	4,5	4,3	4,1	4,0	3,9	3,8	3,6	3,5	3,3	3,1	
	8,6	7,9	7,4	7,1	6,8	6,6	6,5	6,1	5,8	5,6	5,2	4,9	
16	3,0	2,9	2,7	2,7	2,6	2,5	2,5	2,4	2,3	2,3	2,2	2,1	
	4,8	4,4	4,2	4,0	3,9	3,8	3,7	3,5	3,4	3,3	3,0	2,9	
	7,9	7,3	6,8	6,5	6,2	6,1	5,9	5,6	5,3	5,1	4,7	4,4	
18	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3	2,2	2,2	2,1	2,0	
	4,6	4,2	4,0	3,8	3,7	3,6	3,5	3,4	3,2	3,1	2,8	2,7	
	7,5	6,8	6,4	6,1	5,8	5,6	5,5	5,1	4,8	4,7	4,3	4,0	
20	2,9	2,7	2,6	2,5	2,5	2,4	2,4	2,3	2,2	2,1	2,0	1,9	
	4,4	4,1	3,9	3,7	3,6	3,5	3,4	3,2	3,0	3,0	2,7	2,5	
	7,1	6,5	6,0	5,7	5,4	5,3	5,1	4,8	4,5	4,4	4,0	3,7	
40	2,6	2,4	2,3	2,2	2,2	2,1	2,1	2,0	1,9	1,8	1,7	1,6	
	3,8	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,5	2,4	2,1	1,9	
	5,8	5,2	4,8	4,6	4,3	4,2	4,0	3,7	3,5	3,3	3,0	2,6	
100	2,5	2,3	2,2	2,1	2,0	2,0	1,9	1,9	1,7	1,7	1,5	1,4	
	3,5	3,2	3,0	2,8	2,7	2,6	2,5	2,4	2,2	2,1	1,8	1,6	
	5,0	4,5	4,1	3,9	3,7	3,4	3,4	3,1	2,8	2,7	2,3	1,9	

Примечание. Таблица составлен по Г.Ф. Лакину, 1980.

Приложение 2
Критические значения *t*-критерия Стьюдента
 (ν — число степеней свободы)

Уровни значимости									
ν	0,1	0,05	0,01	0,001	ν	0,1	0,05	0,01	0,001
1	6,314	12,706	63,657	636,619	21	1,721	2,080	2,831	3,819
2	2,920	4,308	9,925	31,599	22	1,717	2,074	2,819	3,792
3	2,353	3,182	5,841	12,924	23	1,714	2,069	2,807	3,768
4	2,132	2,776	4,604	8,610	24	1,711	2,064	2,797	3,745
5	2,015	2,571	4,032	6,869	25	1,708	2,060	2,787	3,725
6	1,943	2,447	3,707	5,959	26	1,706	2,056	2,779	3,707
7	1,895	2,365	3,499	5,408	27	1,703	2,052	2,771	3,690
8	1,860	2,306	3,355	5,041	28	1,701	2,048	2,763	3,674
9	1,833	2,262	3,250	4,781	29	1,699	2,045	2,756	3,659
10	1,812	2,228	3,169	4,587	30	1,697	2,042	2,750	3,646
11	1,796	2,201	3,106	4,437	40	1,684	2,021	2,704	3,551
12	1,782	2,179	3,055	4,318	50	1,676	2,009	2,678	3,505
13	1,771	2,160	3,012	4,221	60	1,664	2,000	2,660	3,505
14	1,761	2,145	2,977	4,140	80	1,664	1,990	2,639	3,416
15	1,753	2,131	2,947	4,073	100	1,660	1,984	2,626	3,391
16	1,746	2,120	2,921	4,015	120	1,658	1,980	2,617	3,373
17	1,740	2,110	2,898	3,965	200	1,653	1,972	2,601	3,340
18	1,734	2,101	2,878	3,922	500	1,648	1,965	2,586	3,310
19	1,729	2,093	2,861	3,883	∞	1,645	1,960	2,580	3,291
20	1,725	2,086	2,845	3,850					

0,9 0,95 0,99 0,999 0,9 0,95 0,99 0,999
 Доверительные уровни

Примечание. Таблица составлена по Л.Н. Большеву и Н.В. Смирнову, 1968; М.Дж. Кендаллу и А.М. Стьюарту, 1973

Приложение 3
Критические значения W-критерия Вилкоксона
для независимых выборок

$n_2 \backslash N_1$	3	4	5	6	7	8	9	10	11	12	13	14
5	6	11	17									
6	7	12	18	26								
7	7	13	20	27	36							
8	8	14	21	29	38	49						
9	8	15	22	31	40	51	63					
10	9	15	23	32	42	53	65	78				
11	9	16	24	34	44	55	68	81	96			
12	10	17	26	35	46	58	71	85		115		
13	10	18	27	37	48	60	73	88	103	119	137	
14	11	19	28	38	50	63	76	91	106	123	141	160
15	11	20	29	40	52	65	79	94	110	127	145	164
16	12	21	31	42	54	67	82	97	114	131	150	169
Уровень значимости 0,05												
5			15									
6		10	16	23								
7		10	17	24	32							
8		11	17	25	34	43						
9	6	11	18	26	35	45	56					
10	6	12	19	27	37	47	58	71				
11	6	12	20	28	38	49	61	74	87			
12	7	13	21	30	40	51	63	76	90	106		
13	7	14	22	31	41	53	65	79	93	109	125	
14	7	14	22	32	43	54	67	81	96	112	129	147
15	8	15	23	33	44	56	70	84	99	115	133	151
16	8	15	24	34	46	58	72	86	102	119	137	155
Уровень значимости 0,01												

Примечание. Таблица составлена по Г.Ф. Лакину, 1980.

Приложение 4
Критические значения U-критерия Манна-Уитни
для независимых выборок

		Size of the largest sample (n_2)																												
		5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30			
Size of the smallest sample (n_1)	3	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	13	13			
	4	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14	15	16	17	17	18	19	20	21	22	23			
	5	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20	22	23	24	25	27	28	29	30	32	33			
	6		5	6	8	10	11	13	14	16	17	19	21	22	24	25	27	29	30	32	33	35	37	38	40	42	43			
	7			8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54			
	8				13	15	17	19	22	24	26	29	31	34	36	38	41	43	45	48	50	53	55	57	60	62	65			
	9					17	20	23	26	28	31	34	37	39	42	45	48	50	53	56	59	62	64	67	70	73	76			
	10						23	26	29	33	36	39	42	45	48	52	55	58	61	64	67	71	74	77	80	83	87			
	11							30	33	37	40	44	47	51	55	58	62	65	69	73	76	80	83	87	90	94	98			
	12								37	41	45	49	53	57	61	65	69	73	77	81	85	89	93	97	101	105	109			
	13									45	50	54	59	63	67	72	76	80	85	89	94	98	102	107	111	116	120			
	14										55	59	64	67	74	78	83	88	93	98	102	107	112	118	122	127	131			
	15											64	70	75	80	85	90	96	101	106	111	117	122	125	132	138	143			
	16												75	81	86	92	98	103	109	115	120	126	132	138	143	149	154			
	17													87	93	99	105	111	117	123	129	135	141	147	154	160	166			
	18														99	106	112	119	125	132	138	145	151	158	164	171	177			
	19															113	119	126	133	140	147	154	161	168	175	182	189			
	20																127	134	141	149	156	163	171	178	186	193	200			
	21																	142	150	157	165	173	181	188	196	204	212			
	22																		158	166	174	182	191	199	207	215	223			
	23																			175	183	192	200	209	218	226	235			
	24																				192	201	210	219	228	238	247			
	25																					211	220	230	239	249	258			
	26																						230	240	250	260	270			
	27																							250	261	271	282			

Приложение 5
Критические значения W-критерия Вилкоксона для сопряженных пар

<i>n</i>	α		<i>n</i>	α		<i>n</i>	α	
	0,05	0,01		0,05	0,01		0,05	0,01
6	1		13	18	11	20	53	39
7	3		14	22	14	21	60	44
8	5	1	15	26	17	22	67	50
9	7	3	16	31	21	23	74	56
10	9	4	17	36	24	24	82	62
11	12	6	18	41	29	25	90	69
12	15	8	19	47	33			

Примечание. Таблица составлена по В. Ю. Урбаху, 1964.

Приложение 6

Критические значения коэффициентов корреляции r-Пирсона (r-Спирмена)
(для проверки ненаправленных альтернатив, n — объем выборки)

n	P				n	P				n	P			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	33	0,291	0,344	0,442	0,547	61	0,213	0,252	0,327	0,411
6	0,729	0,811	0,917	0,974	34	0,287	0,339	0,436	0,539	62	0,211	0,250	0,325	0,408
7	0,669	0,754	0,875	0,951	35	0,283	0,334	0,430	0,532	63	0,209	0,248	0,322	0,405
8	0,621	0,707	0,834	0,925	36	0,279	0,329	0,424	0,525	64	0,207	0,246	0,320	0,402
9	0,582	0,666	0,798	0,898	37	0,275	0,325	0,418	0,519	65	0,206	0,244	0,317	0,399
10	0,549	0,632	0,765	0,872	38	0,271	0,320	0,413	0,513	66	0,204	0,242	0,315	0,396
11	0,521	0,602	0,735	0,847	39	0,267	0,316	0,408	0,507	67	0,203	0,240	0,313	0,393
12	0,497	0,576	0,708	0,823	40	0,264	0,312	0,403	0,501	68	0,201	0,239	0,310	0,390
13	0,476	0,553	0,684	0,801	41	0,260	0,308	0,398	0,495	69	0,200	0,237	0,308	0,388
14	0,458	0,532	0,661	0,780	42	0,257	0,304	0,393	0,490	70	0,198	0,235	0,306	0,385
15	0,441	0,514	0,641	0,760	43	0,254	0,301	0,389	0,484	80	0,185	0,220	0,286	0,361
16	0,426	0,497	0,623	0,742	44	0,251	0,297	0,384	0,479	90	0,174	0,207	0,270	0,341
17	0,412	0,482	0,606	0,725	45	0,248	0,294	0,380	0,474	100	0,165	0,197	0,256	0,324
18	0,400	0,468	0,590	0,708	46	0,246	0,291	0,376	0,469	110	0,158	0,187	0,245	0,310
19	0,389	0,456	0,575	0,693	47	0,243	0,288	0,372	0,465	120	0,151	0,179	0,234	0,297
20	0,378	0,444	0,561	0,679	48	0,240	0,285	0,368	0,460	130	0,145	0,172	0,225	0,285
21	0,369	0,433	0,549	0,665	49	0,238	0,282	0,365	0,456	140	0,140	0,166	0,217	0,275
22	0,360	0,423	0,537	0,652	50	0,235	0,279	0,361	0,451	150	0,135	0,160	0,210	0,266
23	0,352	0,413	0,526	0,640	51	0,233	0,276	0,358	0,447	200	0,117	0,139	0,182	0,231
24	0,344	0,404	0,515	0,629	52	0,231	0,273	0,354	0,443	250	0,104	0,124	0,163	0,207
25	0,337	0,396	0,505	0,618	53	0,228	0,271	0,351	0,439	300	0,095	0,113	0,149	0,189
26	0,330	0,388	0,496	0,607	54	0,226	0,268	0,348	0,435	350	0,088	0,105	0,138	0,175
27	0,323	0,381	0,487	0,597	55	0,224	0,266	0,345	0,432	400	0,082	0,098	0,129	0,164
28	0,317	0,374	0,479	0,588	56	0,222	0,263	0,341	0,428	450	0,078	0,092	0,121	0,155
29	0,311	0,367	0,471	0,579	57	0,220	0,261	0,339	0,424	500	0,074	0,088	0,115	0,147
30	0,306	0,361	0,463	0,570	58	0,218	0,259	0,336	0,421	600	0,067	0,080	0,105	0,134
31	0,301	0,355	0,456	0,562	59	0,216	0,256	0,333	0,418					
32	0,296	0,349	0,449	0,554	60	0,214	0,254	0,330	0,414					

Примечание. Таблица составлена по Л.Н. Большеву и Н.В. Смирнову, 1968; E. Tiit, 1972.

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ УЧЕБНЫХ ИЗДАНИЙ, ИНТЕРНЕТ-РЕСУРСОВ, ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ

Основные источники:

1. Статистика : учебник и практикум для среднего профессионального образования / под редакцией И. И. Елисейевой. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 361 с.

Дополнительные источники:

1. Яковлев, В. Б. Статистика. Расчеты в Microsoft Excel : учебное пособие для среднего профессионального образования / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 353 с.

2. Основы математической статистики: Учебное пособие для институтов физ.культ. / Под ред. В.С. Иванова. – М.: Физкультура и спорт, 1990. – 176 с., ил. (электронная версия)

3. Измерения и методы математической статистики в физическом воспитании и спорте: Учебное пособие для ВУЗов Л.В. Денисова, И.В. Хмельницка, Л.А. Харченко. – К.: Олимп. лит., 2008. – 127 с.