

Практическая работа 7

для зарубежных студентов
по направлению подготовки «Лечебное дело»

Компьютерный анализ медицинских данных

Вариационные ряды. Вычисление и оценка показателей вариационного ряда. Стандартизация

Цель работы: научиться вычислять средние величины, оценивать показатели вариационного ряда, изучить метод стандартизации.

Вариационные ряды. Средние величины.

В прикладной статистике объектом исследования являются полученные в результате наблюдений статистические данные. Статистические данные – это совокупность объектов (наблюдений, случаев) и признаков (переменных), их характеризующих.

Переменная (признак) – некоторое общее для всех изучаемых объектов, например людей, свойство, конкретные проявления которого могут меняться от объекта к объекту. Это может быть или физиологический признак (рост, вес, пол, острота зрения и т.д.), или психическое свойство (темперамент, характер, степень интеллектуальности и др.), которые характеризуют состояние рассматриваемого объекта. Переменные могут быть измерены в различных шкалах.

Различные проявления признака для разных объектов называют значениями. С точки зрения теории вероятностей и математической статистики изучаемая переменная также может трактоваться как случайная величина.

На сегодняшний день общепринятого согласия в вопросе «правильной» классификации типов переменных не достигнуто, поэтому для простоты мы примем за основу три типа переменных: *непрерывные*, *дискретные* и *категориальные* (номинальные):

Непрерывные переменные (*continuous variables*) могут принимать любые численные значения, которые естественным образом упорядочены на числовой оси (рост, вес, АД, острота зрения, РОЭ).

Дискретные переменные (*discrete variables*) могут принимать счётное множество упорядоченных значений, которые могут просто обозначать целочисленные данные или ранжировать данные по степени проявления на упорядоченной ранговой шкале (клиническая стадия опухоли, тяжесть состояния пациента).

Категориальные переменные (*categorical variables*) являются неупорядоченными и используются для качественной классификации (пол, цвет глаз, место жительства, тип темперамента); в частности, они могут быть бинарными (дихотомическими) и иметь категорические значения: 1/0, да/нет, имеется/отсутствует.

Тип переменных определяет набор статистических методов анализа данных.

Полученный в результате статистических наблюдений первичный материал необходимо сгруппировать для дальнейшего использования. При этом группировке подлежат только качественно однородные элементы.

Вариационный ряд – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Количественные (числовые) данные предполагают, что переменная принимает некоторое числовое значение. Из них выделяют дискретные данные и непрерывные.

Численное значение каждого отдельного признака или явления, входящего в вариационный ряд, называется вариантой и обозначается символом V или x , или y .

Вариационный ряд, в котором каждая варианта встречается только один раз, называют **простым**.

При увеличении числа наблюдений, как правило, встречаются повторяющиеся значения вариант. В этом случае создается сгруппированный вариационный ряд, где указывается число повторений (частота, обозначается буквой «р»).

Ранжированный вариационный ряд состоит из вариант, расположенных в порядке возрастания или убывания.

Интервальный вариационный ряд составляют при большом числе единиц наблюдения (более нескольких десятков) с целью упрощения последующих вычислений, выполняемых, как правило, без использования компьютера.

Также вариационные ряды делятся на **дискретные** и **непрерывные**.

Непрерывный вариационный ряд включает значения вариант, которые могут выражаться любыми значениями.

Если в вариационном ряде значения признака (варианты) заданы в виде отдельных конкретных чисел, то такой ряд называют **дискретным**.

Особое место в статистическом анализе принадлежит определению среднего уровня изучаемого признака или явления. Средний уровень признака измеряют средними величинами.

Средняя величина характеризует общий количественный уровень изучаемого признака и является групповым свойством статистической совокупности. Она нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и выдвигает на первый план основное, типичное свойство изучаемого признака.

В медицине средние величины широко используются:

- **Для оценки состояния здоровья населения: характеристики физического развития** (рост, вес, окружность грудной клетки и пр.), **выявления распространенности и длительности различных заболеваний, анализа демографических показателей** (естественного движения населения, средней продолжительности предстоящей жизни, воспроизводства населения, средней численности населения и др.).
- **Для изучения деятельности ЛПУ, медицинских кадров и оценки качества их работы, планирования и определения потребности населения в различных видах медицинской помощи** (среднее число обращений или посещений на одного жителя в год, средняя длительность пребывания больного в стационаре, средняя продолжительность обследования больного, средняя обеспеченность врачами, койками и пр.).
- **Для характеристики санитарно-эпидемиологического состояния** (средняя запыленность воздуха в цехе, средняя площадь на одного человека, средние нормы потребления белков, жиров и углеводов и т. д.).
- **Для определения медико-физиологических показателей в норме и патологии, при обработке лабораторных данных, для установления достоверности результатов выборочного исследования в социально-гигиенических, клинических, экспериментальных исследованиях.**

Вычисление средних величин выполняется на основе вариационных рядов. Общими характеристиками значений признака, отражаемого в вариационном ряду, являются средние величины. К ним относятся:

Мода (Mo) - значение наиболее часто встречающейся варианты.

Медиана (Me) - значение варианты, делящей **ранжированный** вариационный ряд пополам (с каждой стороны медианы находится половина вариант).

Средняя арифметическая величина (M или \bar{x}) – это общая количественная характеристика определенного признака изучаемых явлений, составляющих

качественно однородную статистическую совокупность. Различают *среднюю арифметическую простую* и *взвешенную*.

Простая средняя арифметическая вычисляется по формуле:

$$M = \frac{\sum V}{n},$$

где $\sum V$ - сумма вариантов;

n - число наблюдений.

В сгруппированном вариационном ряду определяют **взвешенную среднюю арифметическую**:

$$M = \frac{\sum Vp}{n},$$

$\sum Vp$ - сумма произведений вариант на их частоты;

n - число наблюдений.

В редких случаях, когда имеется симметричный вариационный ряд, значения моды и медианы совпадают со значением средней арифметической.

Средние арифметические величины могут не в полной мере отражать свойства вариационного ряда, в особенности, когда необходимо сопоставление с другими средними. Близкие по значению средние могут быть получены из рядов с различной степенью рассеяния. Чем ближе друг к другу отдельные варианты по своей количественной характеристике, тем меньше **рассеяние (вариабельность) ряда**, тем типичнее его средняя.

Основными параметрами, позволяющими оценить вариабельность признака:

- ✓ Размах;
- ✓ Амплитуда;
- ✓ Среднее квадратическое отклонение;
- ✓ Коэффициент вариации.

Приблизительно о вариабельности признака можно судить по размаху и амплитуде вариационного ряда. **Размах** указывает на максимальную (V_{max}) и минимальную (V_{min}) варианты в ряду. **Амплитуда (A_m)** является разностью этих вариант:

$$A_m = V_{max} - V_{min}$$

Еще одной важнейшей характеристикой вариабельности ряда является дисперсия (D). Она учитывает величину отклонения (d) каждой варианты вариационного ряда от его средней арифметической, т.е. ($d=V-M$). Поскольку отклонения вариант от средней могут быть положительными и отрицательными, то их сумма может оказаться равной нулю ($\sum d=0$). Чтобы избежать этого, величины отклонения (d) возводятся во вторую степень и усредняются. Таким образом, дисперсия вариационного ряда является средним квадратом отклонений вариант от средней арифметической и вычисляется по формуле:

$$D = \frac{\sum d^2}{n}.$$

Но наиболее часто применяется более удобный параметр, вычисляемый на основе дисперсии - среднее квадратическое отклонение (σ).

Поскольку дисперсия выражается квадратом отклонений, ее величина не может использоваться в сопоставлении со средней арифметической. Для этих целей применяется **среднее квадратическое отклонение (σ)**.

$$\sigma = \sqrt{\frac{\sum d^2}{n}}. \quad (1)$$

Оно характеризует среднее отклонение всех вариантов вариационного ряда от средней арифметической величины в тех же единицах, что и сама средняя величина, поэтому они могут использоваться совместно.

Формула (1) применяется при числе наблюдений (n) больше 30. В противном случае значение среднего квадратического отклонения будет иметь погрешность, связанную с математическим смещением ($n-1$). В связи с этим, более точный результат может быть получен с помощью учета такого смещения в формуле расчета *стандартного отклонения*:

$$s = \sqrt{\frac{\sum d^2}{n-1}}.$$

Стандартное отклонение (s) – это оценка среднеквадратического отклонения случайной величины X относительно её математического ожидания на основе несмещённой оценки её дисперсии.

При значениях $n > 30$ среднее квадратическое отклонение (σ) и стандартное отклонение (s) будут одинаковыми ($\sigma = s$). Поэтому в большинстве практических пособий эти критерии рассматриваются как равнозначные. В программе Excel вычисление стандартного отклонения может быть выполнено функцией =СТАНДОТКЛОН(диапазон). А с целью расчета среднего квадратического отклонения требуется создать соответствующую формулу.

Величина среднего квадратического отклонения позволяет судить о характере однородности вариационного ряда и исследуемой группы:

- Если величина σ небольшая, то это свидетельствует о достаточно высокой однородности изучаемого явления. В этом случае среднюю арифметическую следует признать вполне характерной для данного вариационного ряда.
- Слишком малая величина σ заставляет думать об искусственном подборе наблюдений.
- При очень большой величине σ средняя арифметическая в меньшей степени характеризует вариационный ряд, что говорит о значительной вариабельности изучаемого признака или явления или о неоднородности исследуемой группы.

Заметим, что сопоставление величины среднего квадратического отклонения возможно только для признаков одинаковой размерности. Действительно, при сравнении разнообразия веса новорожденных детей и взрослых, мы всегда получим более высокие значения σ у взрослых.

Сравнение вариабельности признаков различной размерности может быть выполнено с помощью **коэффициента вариации**. Он выражает разнообразие в процентах от средней величины, что позволяет производить сравнение различных признаков. Коэффициент вариации в медицинской литературе обозначается символом C .

$$C = \frac{\sigma}{M} \cdot 100\%.$$

Значения коэффициента вариации менее 10% свидетельствует о малом рассеянии, от 10 до 20% – о среднем, более 20% – о сильном рассеянии вариант вокруг средней арифметической.

Для формулировки общего вывода об изучаемом явлении, результаты, полученные на основе выборочной совокупности, должны быть, перенесены на

генеральную совокупность статистическими методами. При этом следует помнить, что найденные на основе данных выборочной совокупности величины, например, среднее арифметическое, как правило, при повторных исследованиях под влиянием случайных явлений может меняться.

Чтобы определить степень совпадения выборочного исследования и генеральной совокупности, необходимо оценить величину ошибки, которая неизбежно возникает при выборочном наблюдении – **ошибки репрезентативности (m)** (или **средней ошибки средней арифметической**). Она является разностью между средними, полученными при выборочном статистическом наблюдении, и аналогичными величинами, которые были бы получены при изучении генеральной совокупности.

Ошибку репрезентативности нельзя смешивать с ошибками регистрации или ошибками внимания (описки, просчеты, опечатки и др.). Величина ошибки репрезентативности зависит от объема выборки и от варибельности признака. Чем больше число наблюдений, тем ближе выборка к генеральной совокупности и тем меньше ошибка. Чем более изменчив признак, тем больше величина статистической ошибки.

На практике для вычисления ошибки репрезентативности в вариационных рядах пользуются следующей формулой:

$$m = \frac{\sigma}{\sqrt{n-1}},$$

При выполнении статистического анализа на основе вычисления относительных величин построение вариационного ряда не является обязательным. При этом определение средней ошибки для относительных показателей может выполняться по упрощенной формуле:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя, выраженного в процентах, промилле и т.д.;

q – величина, обратная P и выраженная как $(1-P)$, $(100-P)$, $(1000-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;

n – число наблюдений в выборочной совокупности.

Эта формула может применяться только в том случае, когда значение показателя меньше его основания. В ряде случаев расчета интенсивных показателей такое условие не соблюдается, и показатель может выражаться числом более 100% или 1000‰. В подобной ситуации выполняется построение вариационного ряда и вычисление ошибки репрезентативности по формуле для средних величин на основе среднего квадратического отклонения.

Прогнозирование величины средней арифметической в генеральной совокупности выполняется с указанием **доверительных границ** – минимального и максимального значений. Поскольку выборочная средняя является случайной величиной, такой прогноз выполняется с приемлемым для исследователя уровнем вероятности. В медицинских исследованиях он составляет не менее 95%.

Согласно теории вероятности, в явлениях, подчиняющихся нормальному закону распределения (рис. 1), между значениями средней арифметической, среднего квадратического отклонения и вариантами существует строгая зависимость - выполняется **правило трех сигм: 99,7% значений варьирующего признака находятся в пределах $M \pm 3\sigma$** .

Кроме того 68,3% варьирующего признака находятся в пределах $M \pm 1\sigma$, 95,5% — в пределах $M \pm 2\sigma$.

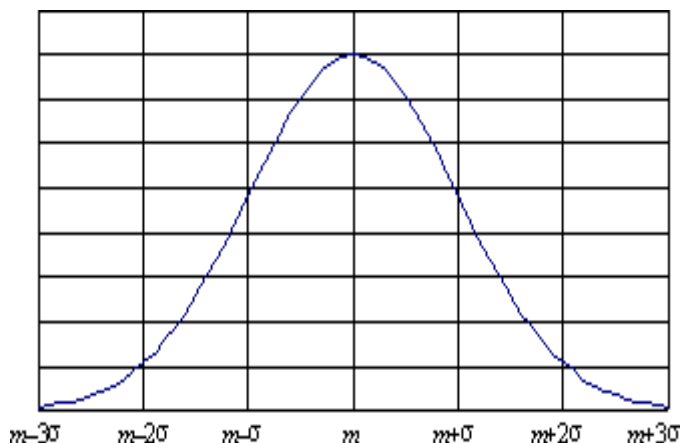


Рис. 1. Плотность вероятностей нормального распределения.

Большинство экспериментальных исследований, в том числе и в области медицины, связано с измерениями, результаты которых могут принимать практически любые значения в заданном интервале, поэтому, как правило, описываются моделью непрерывных случайных величин. В связи с этим в большинстве статистических методов рассматриваются непрерывные распределения. Одним из таких распределений, имеющим основополагающую роль в математической статистике, является **нормальное** или **Гауссово распределение**.

Это объясняется целым рядом причин.

1. Прежде всего, многие экспериментальные наблюдения можно успешно описать с помощью нормального распределения (измерения веса, роста и других физиологических параметров организма человека).

2. Многие распределения, связанные со случайной выборкой, при увеличении объема последней переходят в нормальное.

3. Нормальное распределение хорошо подходит в качестве приближенного описания других непрерывных распределений (например, асимметричных).

В то же время следует отметить, что в медицинских данных встречается много экспериментальных распределений, описание которых моделью нормального распределения невозможно. Для этого в статистике разработаны методы, которые принято называть «Непараметрическими».

Выбор статистического метода, который подходит для обработки данных конкретного эксперимента, должен производиться в зависимости от принадлежности полученных данных к нормальному закону распределения. Проверка гипотезы на подчинение признака нормальному закону распределения выполняется с помощью гистограммы распределения частот (графика), а также ряда статистических критериев. Среди них:

- Критерий асимметрии (β);
- Критерий проверки на эксцесс (γ);
- Критерий Шапиро – Уилкса (W).

Анализ характера распределения данных (его еще называют проверкой на нормальность распределения) осуществляется по каждому параметру. Чтобы уверенно судить о соответствии распределения параметра нормальному закону, необходимо достаточно большое число единиц наблюдения (не менее 30 значений).

Для нормального распределения критерии асимметрии и эксцесса принимают значение 0. Если распределение смещено вправо $\beta > 0$ (положительная асимметрия), при $\beta < 0$ - график распределения смещен влево (отрицательная

асимметрия). Критерий асимметрии проверяет форму кривой распределения. В случае нормального закона $\gamma=0$. При $\gamma > 0$ кривая распределения острее, если $\gamma < 0$ пик более сглаженный, чем функция нормального распределения.

Пример построения вариационных рядов, вычисления средних величин, создания графика распределения признака и проверки на нормальность распределения.

Условие задачи: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №1 больницы выполнено 16 измерений пульса у детей:

Иванов Василий	– 120 уд.в мин.
Сидоров Костя	– 130 – “ -
...	- 115
...	- 120
...	- 120
...	- 125
...	- 110
...	- 125
...	- 115
...	- 120
...	- 125
...	- 135
...	- 115
...	- 130
...	- 125
...	- 120

Задание: а) составьте простой не ранжированный вариационный ряд, вычислите среднюю арифметическую величину и определите степень рассеяния вариант в вариационном ряду;

б) выполните упорядочение (ранжирование) ряда по возрастанию и определите медиану;

в) составьте сгруппированный ранжированный вариационный ряд, постройте график распределения признака, определите моду и среднюю взвешенную величину;

г) определите вид распределения признака и статистические критерии нормальности распределения;

д) вычислите ошибку репрезентативности и доверительные границы колебания средней арифметической в генеральной совокупности.

Решение: запустите программу Excel, откройте файл в папке своей учебной группы под именем «Статистика–Фамилии студентов». Создайте НОВЫЙ лист, переименуйте его, обозначив названием «Сред_вел». На этом листе введите данные и решение задачи, как показано ниже, сохраните изменения и покажите результат работы преподавателю.

а) Построение простого вариационного ряда частоты пульса детей, поступивших в отделение №1 больницы. Простой вариационный ряд представляет собой таблицу, в которой подлежащим является изучаемый признак, обозначаемый знаком V (варианта). Все полученные в эксперименте данные вносят в таблицу в

порядке их записи в журнал регистрации (условия задачи). Если значения вариант повторяются, то каждое из них заносится в отдельную ячейку таблицы. Затем вычисляют сумму вариант, среднюю арифметическую, отклонения (d) каждой варианты от средней величины и квадрат отклонения (d^2) соответствующими формулами (таблица 1).

Таблица 1

Простой не ранжированный вариационный ряд

	V	$d=V-M$	d^2
1 Иванов Василий	120	-1,88	3,52
2 Сидоров Костя	130	8,13	66,02
3 ...	115	-6,88	47,27
4 ...	120	-1,88	3,52
5 ...	120	-1,88	3,52
6 ...	125	3,13	9,77
7 ...	110	-11,88	141,02
8 ...	125	3,13	9,77
9 ...	115	-6,88	47,27
10 ...	120	-1,88	3,52
11 ...	125	3,13	9,77
12 ...	135	13,13	172,27
13 ...	115	-6,88	47,27
14 ...	130	8,13	66,02
15 ...	125	3,13	9,77
16 ...	120	-1,88	3,52
Сумма (Σ) =	1950	0,00	643,75

$n=16$

Вычисление средней арифметической (M) в простом вариационном ряду выполняется по формуле:

$$\bar{M} = 1950/16 = \underline{121,9} \text{ уд/мин.}$$

В программе Excel для вычисления средней арифметической может применяться функция =СРЗНАЧ(Диапазон ячеек). Использование этой функции даст такой же результат.

Определение степени рассеяния вариант в вариационном ряду с помощью среднего квадратического отклонения (σ - Сигма) и коэффициента вариации (C). Среднее квадратическое отклонения вычисляется по формуле:

$$\sqrt{\frac{\Sigma}{n}} = \text{КОРЕНЬ}(643,75 / 16) = \underline{\pm 6,343}.$$

Для упрощения расчета среднего квадратического отклонения при $n > 30$ в программе Excel может использоваться функция =СТАНДОТКЛОН(Диапазон). Если выполнить вычисления этой функцией в нашем вариационном ряду, то будет получено значение стандартного отклонения = 6,55. Это значение отличается от величины среднего квадратического отклонения, что обусловлено малым числом наблюдений ($n = 16$).

Коэффициент вариации (C) вычисляется по формуле:

— $= 6,343/121,9 * 100 = 5,2\%$, что является малым рассеянием (<10%).

Вариабельность признака (рассеяние) оценивается как малая при $C < 10\%$, средняя при $10\% < C < 20\%$, высокая при $C > 20\%$

Вывод: средняя частота пульса пациентов изучаемой группы составляет 121,9 ударов в минуту, вариабельность пульса малая.

б) Преобразование не ранжированного вариационного ряда в ранжированный (таблица 2) производится в имеющей таблице простого ряда с помощью команды сортировки по столбцу **V**. С этой целью необходимо предварительно выделить диапазон ячеек таблицы, включающий данные из сортируемого столбца, а затем на вкладке «Главная» нажать кнопку «Сортировка и фильтр». При выполнении сортировки требуется автоматически расширить диапазон сортируемых ячеек.

Таблица 2

Ранжированный вариационный ряд

	V	d=V-M	d²
1	110	-11,88	141,02
2	115	-6,88	47,27
3	115	-6,88	47,27
4	115	-6,88	47,27
5 Иванов Василий	120	-1,88	3,52
6	120	-1,88	3,52
7	120	-1,88	3,52
8	120	-1,88	3,52
9	120	-1,88	3,52
10	125	3,13	9,77
11	125	3,13	9,77
12	125	3,13	9,77
13	125	3,13	9,77
14 Сидоров Костя	130	8,13	66,02
15	130	8,13	66,02
16	135	13,13	172,27
Сумма (Σ) = =СУММ(Диапазон ячеек)	1950	0	643,75
Средняя арифметическая = =СРЗНАЧ(Диапазон ячеек)	121,875		

Для определения медианы (**Me**) в ранжированном вариационном ряду находят значение варианты, расположенной в центре ряда.

Me = варианта, занимающая срединное положение (если ряд состоит из четного числа вариантов, медианой является полусумма двух центральных вариантов) = **120уд/мин.**

В программе Excel для определения медианы применяется функция =МЕДИАНА(Диапазон ячеек).

в) Построение сгруппированного ранжированного ряда выполняется в отдельной таблице (таблица 3), которая должна включать столбец частоты повторений вариант (p). Данные из сгруппированного вариационного ряда используются для построения точечной диаграммы распределения частот. Он демонстрирует частоту встречаемости вариант (p), указываемую по вертикальной оси точечной диаграммы (рис. 2).

	V	p
1	110	1
2	115	3
3	120	5
4	125	4
5	130	2
6	135	1
$\Sigma=$	735	16

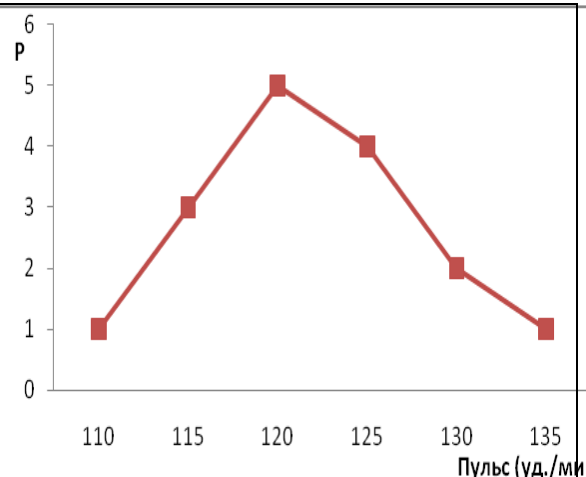


Таблица 3
Сгруппированный ранжированный вариационный ряд

Рис. 2. График распределения признака

Определение моды может выполняться в сгруппированном вариационном ряду выбором значения варианты, которая повторяется наибольшее число раз.

M_o = наиболее часто повторяющаяся варианта = **120 уд/мин** (встречается 5 раз).

В программе Excel для определения моды может применяться функция =МОДА(Диапазон данных), которая используется только в простом вариационном ряду.

Для сгруппированного вариационного ряда средняя арифметическая вычисляется по модифицированной формуле и называется «Средняя взвешенная»:

$$M = \frac{\sum Vp}{n} = \frac{110*1 + 115*3 + 120*5 + 125*4 + 130*2 + 135*1}{1 + 3 + 5 + 4 + 2 + 1} = \frac{735}{16} = 121,9$$

Независимо от способа построения вариационного ряда и метода вычисления средней арифметической, для одинаковых данных значения их средней величины должны совпадать.

г) Оценка нормальности распределения выполняется с помощью графика распределения (рис. 2) по форме линии, которая должна приближаться к виду кривой Гаусса (функции нормального распределения). Статистические критерии нормальности распределения вычисляются с помощью функций или в модуле «Описательная статистика», который может быть вызван командой «Данные» - «Анализ данных» - «Описательная статистика».

Если указанная команда отсутствует разделе «Данные», необходимо выполнить установку надстройки программы Excel «Пакет анализа». Для этого требуется открыть меню настройки программы кнопкой «Office», выбрать «Параметры Excel» -> «Надстройки» -> «Перейти» -> «Пакет анализа» и «Пакет анализа - VBA» -> «ОК».

Модуль «Описательная статистика» производит вычисления только в простых вариационных рядах. В связи с этим необходимо использовать диапазон ячеек таблицы, в котором составлен ПРОСТОЙ вариационный ряд созданный в пункте б) настоящей задачи. Результат вычислений, выполненный модулем «Описательная статистика», показан в таблице 4.

Таблица 4

Простой не ранжированный вариационный ряд					Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
	V	d=V-M	d ²			
1 Иванов Василий	120	-1,88	3,52			
2 Сидоров Костя	130	8,13	66,02			
3 ...	115	-6,88	47,27		Среднее 121,875	
4 ...	120	-1,88	3,52		Стандартная ошибка 1,63777	
5 ...	120	-1,88	3,52		Медиана 120	
6 ...	125	3,13	9,77		Мода 120	
7 ...	110	-11,88	141,02		Стандартное отклонение 6,551081	
8 ...	125	3,13	9,77		Дисперсия выборки 42,91667	
9 ...	115	-6,88	47,27		Экцесс -0,16979	
10 ...	120	-1,88	3,52		Асимметричность 0,209598	
11 ...	125	3,13	9,77		Интервал 25	
12 ...	135	13,13	172,27		Минимум 110	
13 ...	115	-6,88	47,27		Максимум 135	
14 ...	130	8,13	66,02		Сумма 1950	
15 ...	125	3,13	9,77		Счет 16	
16 ...	120	-1,88	3,52		Наибольший(1) 135	
□□	1950	0,00	643,75		Наименьший(1) 110	
n=16					Уровень надежности(95,0%) 3,490827	

Экцесс = -0,17. Для нормального распределения эксцесс = 0. **Распределение близко к нормальному.**

Асимметричность = 0,2. Для нормального распределения асимметричность = 0. **Распределение близко к симметричному.**

Применяя стандартные функции =ЭКССЕСС(Диапазон) и =СКОС(Диапазон) мы получим эти же значения.

д) Определение ошибки средней арифметической (**m**, ошибка репрезентативности или стандартная ошибка) выполняется по формуле:

$$\frac{\sigma}{\sqrt{n}} = 6,343 / \text{КОРЕНЬ}(16-1) = \underline{\underline{1,64}}$$

Программа Excel позволяет вычислять этот критерий с помощью модуля «Описательная статистика». При этом ошибка средней арифметической (**m**) называется «Стандартная ошибка», так как в качестве среднего квадратического отклонения используется «Стандартное отклонение» (таблица 16). Поэтому в вариационных рядах с числом единиц наблюдения менее 30 возможно несущественное отличие значений полученных расчетной формулой и модулем описательной статистики.

Таким образом, с вероятностью P>95% доверительные границы колебания средней арифметической в генеральной совокупности $M \pm 2m = 121,9 \pm 2*1,64 =$

121,9 ± 3,28 уд/мин.

Для вычисления доверительных границ в программе Excel может применяться статистическая функция =ДОВЕРИТ(0,05; Значение стандартного отклонения; Число единиц наблюдения). При ее использовании с вычисленными ранее значениями задачи получим:

$$=ДОВЕРИТ(0,05; 6,55; 16) = 3,21.$$

Вывод: Средняя частота пульса пациентов 1-го отделения с вероятностью 95,5% составляет от 118,6 до 125,2 ударов в минуту.

Пример сравнения рассеяния вариационных рядов.

Условие задачи: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №2 больницы выполнено 17 измерений пульса у детей: 1. Казаков Саша – 130 уд.в мин; 2. Литвинов Сережа – 135 уд.в мин.; 3...– 125; 4...– 115; 5 ...– 125; 6 ... – 125; 7 ... – 120; 8 ... – 125; 9 ... – 130; 10 ... – 120; 11 ... – 140; 12 ... – 145; 13 ... – 115; 14 ...– 130; 15 ... – 125; 16 ... – 120; 17 ... – 125.

Задание: а) создать простой, ранжированный вариационный ряды, определить средние величины и доверительные границы средней арифметической;

б) составить сгруппированный, ранжированный вариационный ряд и построить график распределения признака, выполнить проверку его на нормальность;

в) определить параметры варибельности признака: амплитуду, размах, среднее квадратическое отклонение, коэффициент вариации и ошибку репрезентативности;

г) сравнить характеристики рассеяния вариационных рядов, используемых в примерах данного раздела.

Решение: запустите программу Excel, откройте требуемый файл в папке своей учебной группы под именем «Статистика–Фамилии студентов». На листе «Сред_вел» этого файла введите данные в соответствии с таблицей 19, выполните вычисления с помощью формул и модуля «Описательная статистика». Покажите результат работы преподавателю.

а) Построение простого ранжированного вариационного ряда (таблица 5) и вычисление их основных характеристик модулем «Описательная статистика».

Таблица 5

Простой ранжированный вариационный ряд и его характеристики

	V	d=V-M	d ²
1	115	11,18	124,91
2	115	11,18	124,91
3	115	11,18	124,91
4	120	6,18	38,15
5	120	6,18	38,15
6	125	1,18	1,38
7	125	1,18	1,38
8	125	1,18	1,38
9	125	1,18	1,38
10	125	1,18	1,38
11	125	1,18	1,38
12 Казаков	130	-3,82	14,62
13	130	-3,82	14,62
14	130	-3,82	14,62
15 Литвинов	135	-8,82	77,85
16	140	-13,82	191,09
17	145	-18,82	354,33
Σ=	2145	0,00	1126,47

n=17

Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
Столбец1	
Среднее	126,1764706
Стандартная ошибка	2,035051839
Медиана	125
Мода	125
Стандартное отклонение	8,390733685
Дисперсия выборки	70,40441176
Эксцесс	0,3573165
Асимметричность	0,659063476
Интервал	30
Минимум	115
Максимум	145
Сумма	2145
Счет	17
Наибольший(1)	145
Наименьший(1)	115
Уровень надежности(95,0%)	4,314116203

$$M_2 = \underline{126,2}$$

$$m_2 = \underline{2,04}$$

Сигма (σ) = 8,390734

$$C = 8,39 / 126,2 * 100 = 6,6\%$$

Вывод: Средняя частота пульса пациентов 2-го отделения с вероятностью 95,5% составляет $126,2 \pm 2,04$ ударов в минуту, варибельность малая.

б) Создание сгруппированного вариационного ряда (таблица 6) и построение графика распределения (рис. 3).

Таблица 6		ранжированный	
Сгруппированный, вариационный ряд			
	V	p	
1	115	3	
2	120	2	
3	<u>125</u>	<u>6</u>	
4	130	3	
5	135	1	
6	140	1	
7	145	1	
	$\Sigma p =$	17	
$n = 17$			



Рис. 3. График распределения признака.

Me = варианта, занимающая срединное положение = МЕДИАНА(Диапазон) = 125 уд/мин.

Mo = наиболее часто повторяющаяся варианта = МОДА(Диапазон) = 125 уд/мин.

в) Параметры варибельности определяются по результатам обработки данных модулем «Описательная статистика» (таблица 17). Амплитуда (интервал) = 30 уд/мин, размах от 115 до 145 уд/мин, среднее квадратическое отклонение = 8,4 уд/мин, коэффициент вариации = 6,6%, ошибка репрезентативности = $\pm 2,04$ уд/мин.

г) Сравнение рассеяния вариационных рядов, используемых в примерах данного раздела выполняется на основе коэффициента вариации первого ($C_1 = 5,4\%$) и второго вариационных рядов ($C_2 = 6,6\%$).

Вывод: варибельность пульса пациентов 1-го отделения ниже, чем пациентов 2-го отделения больницы и в обоих случаях малая (<10%).

Метод стандартизации

Основным принципом подбора групп статистического наблюдения в случае необходимости их сравнения является однородность сравниваемых совокупностей по характеризующим признакам (признакам, по которым выполняется группировка совокупностей). Подбор групп с одинаковым уровнем характеризующих признаков позволяет исключить влияние на результативные признаки и акцентировать внимание на изучаемых особенностях явления.

Например, общие интенсивные коэффициенты (рождаемости, смертности, младенческой смертности, заболеваемости и т.д.) правильно отражают частоту явлений при их сопоставлении лишь в том случае, если состав сравниваемых совокупностей однороден. Если же они имеют различный возрастно-половой или профессиональный состав, разное соотношение по нозологическим формам или другим характеризующим признакам, то, ориентируясь на общие показатели, сопоставляя их, можно сделать неправильный вывод об истинных причинах разницы показателей сравниваемых совокупностей.

Однако выполнить подбор групп с одинаковым уровнем характеризующих признаков сложно. Если не удастся провести формирование однородных выборок, или уже после проведения исследования оказывается, что сравниваемые группы не одинаковы по характеризующим признакам, может использоваться МЕТОД СТАНДАРТИЗАЦИИ.

Метод стандартизации - это метод расчета условных (стандартизованных) показателей, заменяющих общие интенсивные (или средние) величины в тех случаях, когда их сравнение затруднено из-за несопоставимости групп. Он позволяет определить, какие показатели были бы получены в случае однородности групп, т.е. устранить (элиминировать) возможное влияние различий в составе совокупностей по какому-либо признаку на величину сравниваемых интенсивных показателей. Следовательно, метод стандартизации применяется тогда, когда имеющиеся различия в составе сравниваемых совокупностей могут повлиять на размеры общих коэффициентов.

Стандартизованные показатели – это условные, величины, они не отражают истинных размеров явлений, но свидетельствуют о том, каковы были бы значения сравниваемых интенсивных показателей, если бы были исключены различия в составах совокупностей.

Для того, чтобы устранить влияние неоднородности составов сравниваемых совокупностей на величину получаемых коэффициентов, их приводят к единому стандарту, то есть условно допускается, что состав совокупностей одинаков.

В качестве стандарта можно принять величину среды одной или всех сравниваемых групп, уровень явления в этих группах или какой-либо близкой по существу третьей совокупности. От выбора стандарта зависит способ вычисления, т.е. модификация метода. Принято выделять *прямой*, *обратный* и *косвенный* варианты соответственно. Варианты используемого метода стандартизации не влияют на конечный вывод и определяются исследователем.

Методика выполнения расчетов прямого метода стандартизации состоит из 5 этапов:

- ПЕРВЫЙ ЭТАП. Вычисление общих показателей (относительных или средних величин) отдельно для каждой группы.
- ВТОРОЙ ЭТАП. Выбор стандарта осуществляется произвольно.
- ТРЕТИЙ ЭТАП. Вычисление ожидаемых величин, демонстрирующих уровень явления, который мог бы быть зафиксирован при отсутствии неоднородности в группах.
- ЧЕТВЕРТЫЙ ЭТАП. Определение стандартизованных показателей.
- ПЯТЫЙ ЭТАП. Сравнение групп по стандартизованным показателям и формулирование вывода.

Пример вычисления стандартизованных показателей прямым методом стандартизации

Условия задачи: проведено углубленное обследование жителей 2-х микрорайонов города N, при котором получены следующие данные. В 1-м районе обследовано 795 человек, из них 156 мужчины, 639 женщины. Выявлена стенокардия у 3 мужчин и 30 женщин. Во 2-м районе обследовано 720 человек (585 мужчин, 135 женщин). Заболевание выявлено у 15 мужчин и 12 женщин.

Задание: сравнить патологическую пораженность стенокардией в 2-х районах города.

Решение: необходимо запустить программу Excel, ввести данные условия задачи и выполнить вычисления по этапам, как показано ниже, сохранить файл.

1-й этап: создание статистической таблицы и вычисление общих относительных величин ($P_{общ.}$), как показано в таблице 1.

Районы:	Абсолютные данные				1. Патологическая пораженность на 100 обсл. ($P_{общ.}$)		2. Стандарт (1-й + 2-й районы)	3. Ожидаемые величины по районам	
	1-й		2-й		1-й	2-й		1-й	2-
Пол	Обсл.	Заб.	Обсл.	аб.					
м	156	3	585	15	1,9	2,6	741	14,3	19
ж	639	30	135	12	4,7	8,9	774	36,3	68
Оба пола	795	33	720	27	4,2	3,8	1515	50,6	87
								4. Стандартизованные величины ($P_{ст.}$)	
								3,3	5

Таблиц

а 1 Вычисление относительных и стандартизованных величин

2-й этап: выбор стандарта. Для прямого метода стандартизации за стандарт выбирается число обследованных в обоих районах (уровень среды) среди мужчин и женщин отдельно, а затем их сумма, т.е. $741 = 156 + 585$ и т.д.

3-й этап: вычисление ожидаемых величин. Они вычисляются с помощью пропорций, которыми определяют возможное число заболевших, если бы в изучаемых районах был одинаков состав среды:

Из 100 болевших	-	$P_{общ.}$
Из числа стандарта	-	$X_{Ожид.число болевших}$

$$x = \frac{P_{общ.} * \text{Стандарт}}{100}$$

Ожидаемая величина болевших мужчин района 1:

Из 100 муж. заболело	-	$P_{общ.}$
Из числа станд. муж.	-	$X_{Ожид.число болевших}$

$$x = \frac{1,9 * 741}{100} = \mathbf{36,3}$$

Ожидаемая величина болевших женщин района 1:

Из 100 жен. заболело	-	$P_{общ.}$
Из числа станд. жен.	-	$X_{Ожид.число болевших}$

$$x = \frac{4,7 * 774}{100} = \mathbf{14,3}$$

Аналогично составляются пропорции для района 2.

Ожидаемые величины по каждому району суммируются.

Для района 1: $36,3 + 14,3 = \mathbf{50,6}$.

Для района 2: $68,8 + 19,0 = \mathbf{87,8}$.

4-й этап: вычисление стандартизованных величин. Расчет выполняется по принципам вычисления относительных величин, но на основе ожидаемых значений с помощью составления пропорции:

Из общего числа обслед.	-	болело число ожидаемых
Из 100 чел.	-	$P_{см.}$

$$P_{см.} = \frac{\text{Ожид.вел.} * 1}{\text{Общее число об}}$$

Для района 1:

Из 1515 чел	-	болело 50,6
Из 100 чел.	-	$P_{см.}$

$$P_{см1.} = \frac{50,6 * 100}{1515} = \mathbf{3,3}$$

Для района 2:

Из 1515 чел	-	болело 87,8
Из 100 чел.	-	$P_{см.}$

$$P_{см2.} = \frac{87,8 * 100}{1515} = \mathbf{5,8}$$

5-й этап: оценка достоверности стандартизованных величин, их сопоставление между собой и с относительными величинами и формулирование вывода.

Вычисление критерия достоверности Стьюдента выполняется с целью подтверждения гипотезы о закономерности выявленных различий.

Критерий достоверности Стьюдента для относительных величин рассчитывается по формуле:

$$t = \left| \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}} \right|,$$

где: P_1 – относительная величина (показатель) 1-й группы;

P_2 – относительная величина (показатель) 2-й группы;

m_1 – ошибка репрезентативности 1-го показателя;

m_2 – ошибка репрезентативности 2-го показателя.

Для вычисления ошибки репрезентативности относительной величины используется формула:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где: P – величина относительного показателя;

q – величина, обратная P и вычисленная как $(1-P)$, $(100-P)$, $(100-P)$ и т. д., в зависимости от основания, на которое рассчитан показатель;

n – число наблюдений.

В программе Excel значения этих критериев может быть вычислено следующими формулами:

m_1 : =КОРЕНЬ(3,3*(100-3,3)/(795-1)), результат 0,63;

m_2 : =КОРЕНЬ(5,8*(100-5,8)/(720-1)), результат 0,87;

t : =ABS((3,3-5,8)/КОРЕНЬ(0,63^2+0,87^2)), результат 2,3 > 2, т.е.

различия достоверны.

Для удобства сопоставления полученных значений составляется таблица, в которой сравнивают интенсивные и стандартизованные показатели (таблица 2).

Таблица 2

Сопоставление интенсивных и стандартизованных величин

Показатели	Район 1	Район 2	Результаты сравнения
Интенсивные ($P_{\text{общ.}}$)	4,2	3,8	Заболеваемость в 1-м районе > чем во 2-м районе
Стандартизованные ($P_{\text{ст.}}$)	$3,3 \pm 0,63$	$5,8 \pm 0,87$	Заболеваемость в 1-м районе < чем во 2-м районе

Вывод: сравнение стандартизованных показателей выявления стенокардии в районах 1 и 2 позволяет сделать заключение, что, если бы состав обследованных по полу в этих районах был одинаков, то показатель заболеваемости стенокардией в 1-ом районе был бы значительно ниже, чем во 2-ом (уровень значимости $p < 0,05$).

Из анализа общих интенсивных показателей такой вывод сделать нельзя, так как на общие показатели оказал влияние разный состав осмотренных по полу в этих районах. Из литературных данных известно, что стенокардия приблизительно в 2 раза чаще встречается среди мужчин, чем среди женщин. Поэтому в 1-ом районе, где были обследованы преимущественно мужчины, общий интенсивный показатель был завышен (4,2 на 100 обследованных), а во 2-ом районе – занижен (3,8 на 100 обследованных), так как были обследованы в основном женщины. Метод стандартизации позволил увидеть истинную картину заболеваемости, устранив влияние фактора, связанного с полом обследованных.

Практические задания

Задание 1.

Запустите программу Excel, откройте файл в папке своей учебной группы под именем «Статистика–Фамилии студентов», создайте НОВЫЙ лист, переименуйте его, обозначив названием «Сред_вел», решите требуемый вариант задания, сохраните, покажите файл преподавателю.

Вариант 1

а) В районе N, где расположена тепловая электростанция, в одной из точек жилого поселка было взято 100 проб атмосферного воздуха. Количество пыли в пробах воздуха измерялось следующими цифрами: 0,09 мг/м³ в 2-х пробах, 0,08 мг/м³ – 2 раза, 0,15 мг/м³ – 16 раз, 0,12 мг/м³ – 14 раз, 0,14 мг/м³ – 30 раз, 0,16 мг/м³ – 4 раза, 0,13 мг/м³ – 16 раз, 0,11 мг/м³ – 9 раз, 0,10 мг/м³ – 5 раз, 0,17 мг/м³ – 2 раза.

1) Составьте простой, ранжированный вариационный ряд и определите среднесуточную концентрацию пыли, ее вариабельность, доверительные границы колебаний средней величины, **не пользуясь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п. 1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия массы тела у новорожденных, детей первого года жизни и семилетних, если известны следующие параметры:

Возраст	Средняя масса (M), кг	σ , кг
Новорожденные	3,4	$\pm 0,5$

1 год	10,5	$\pm 0,8$
7 лет	22,9	$\pm 2,7$

Вариант 2

а) В питьевой воде, которой снабжаются дома жителей района N, определяли концентрацию соединений фтора. В 2-х пробах было обнаружено 0,5 мг/л этих соединений, в 4-х – 0,6 мг/л, в 8-и – 0,9 мг/л, в 9-и – 0,4 мг/л, в 16-и – 0,8 мг/л, в 15-ти – 0,9 мг/л, в 20-и – 1,2 мг/л, в 24-х – 1,1 мг/л, в 42-х – 1,3 мг/л, в 50-и – 1,0 мг/л, в 24-х – 1,5 мг/л, в 23-х – 1,6 мг/л, в 10-ти – 0,7 мг/л, в 8-и – 1,4 мг/л, в 4-х – 0,3 мг/л.

1) Составьте простой, ранжированный вариационный ряд, Определите среднюю концентрацию соединений фтора в питьевой воде и доверительные границы колебаний средней величины, **не пользуясь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п. 1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия лабораторных анализов с различной размерностью:

Наименование теста	Средний показатель	σ
Общий белок крови, мг%	6,8	$\pm 0,4$
СОЭ, мм/ч	9	± 2
Лейкоциты	8000	± 800

Вариант 3

а) При стоматологическом обследовании группы школьников 4-х классов сельского района были получены следующие результаты: 2 школьника имели по 5 кариозных зубов, 28 школьников – по 1 кариозному зубу, 8 школьников – по 4 кариозных зубов, 1 школьник – 8 кариозных зубов, 20 школьников – по 3 кариозных зуба, 16 школьников – по 2 кариозных зуба и 6 школьников не имели пораженных кариесом зубов.

1) Составьте простой, ранжированный вариационный ряд, определите среднее число зубов пораженных кариесом у школьников района, степень варибельности этого признака и доверительные границы колебаний средней величины, **не пользуясь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п. 1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия антропометрических данных у мальчиков 7-и летнего возраста:

Показатель	M	σ
Рост, см	123,4	$\pm 4,9$
Масса тела, кг	24,2	$\pm 3,1$

Окружность грудной клетки, см	60,1	±2,5
-------------------------------	------	------

Вариант 4

а) Перед сдачей экзамена у студентов определялась частота пульса. Были получены следующие данные: у 2 студентов — 76 ударов в минуту, у 3 студентов – 80 ударов в минуту, у 4 студентов – 108 ударов в минуту, у 2 студентов – 116 ударов в минуту, у 20 студентов – 88 ударов в минуту, у 6 студентов – 98 ударов в минуту, у 17 студентов – 86 ударов в минуту, у 11 студентов – 92 удара в минуту.

1) Составьте простой, ранжированный вариационный ряд и определите среднюю частоту пульса у студентов перед экзаменом, степень вариабельности признака и доверительные границы колебаний средней величины, **не пользуясь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п.1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия антропометрических данных у девушек 17-и летнего возраста:

Показатель	М	σ
Рост, см	161,2	±5,1
Масса тела, кг	55,8	±7,2
Жизненная емкость легких, см ³	3400	±250

Вариант 5

а) Исследовалась длительность лечения больных пневмонией в стационаре центральной районной больницы N-ского района. Были получены следующие результаты: 10 дней лечилось 2 больных, 9 дней – 1 больной, 11 дней – 1 больной, 12 дней – 1 больной, 8 дней – 3 больных, 13 дней – 1 больной, 21 день – 3 больных, 7 дней – 1 больной, 22 дня – 3 больных, 14 дней – 2 больных, 20 дней – 5 больных, 15 дней – 2 больных, 16 дней – 3 больных, 17 дней – 4 больных, 19 дней – 8 больных, 18 дней – 7 больных.

1) Составьте простой, ранжированный вариационный ряд и определите среднюю длительность лечения пневмонии, степень вариабельности признака и доверительные границы колебаний средней величины, **не пользуясь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п.1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия антропометрических данных у 12-летних мальчиков:

Показатель	М, см	σ
Рост	142,0	±8,5

Окружность грудной клетки	66,0	$\pm 4,0$
Окружность головы	50,0	$\pm 2,0$

Вариант 6

а) Исследовалась длина тела новорожденных девочек по данным родильного дома. Были получены следующие данные: у 8 девочек рост составил 48 см, у 6 девочек – 51 см, у 7 девочек – 53 см, у 1 девочки – 49 см, у 9 девочек – 52 см, у 8 девочек – 50 см, у 1 девочки – 47 см, у 3 девочек – 46 см, у 2 девочек – 54 см, у 1 девочки – 55 см, у 1 девочки – 56 см.

1) Составьте простой, ранжированный вариационный ряд, определите среднюю длину тела новорожденных девочек, степень вариабельности признака и доверительные границы колебаний средней величины, **не пользуйсь надстройкой Пакет анализа (Анализ данных)**.

2) С помощью пакета анализа найдите основные точечные характеристики полученного вами вариационного ряда. Сравните полученные результаты с результатами, полученными в п. 1).

3) Составьте сгруппированный, ранжированный вариационный ряд и выполните построение графика распределения признака.

4) Сравните полученный график с кривой нормального распределения. Какой вывод вы можете сделать?

б) Сравните характер разнообразия антропометрических данных у 12-летних девочек:

Показатель	М	σ
Рост, см	140	$\pm 9,5$
Масса тела, кг	40	± 6
Жизненная емкость легких, см ³	2300	± 360

Задание 2. Вычисление стандартных показателей в программе Excel.

Запустите программу Excel, сохраните новую книгу под именем «Стандартизация», переименуйте первый лист этой книги в «Стандарт», выполните на нем вычисление стандартизованных показателей и сделайте вывод

Вариант 1

Проведено изучение заболеваемости населения в городах N и M. В городе N проживало 30000 человек, из них 20000 мужчин и 10000 женщин. В течение года болели 5000 мужчин и 4000 женщин. В городе M проживало 65000 человек (25000 мужчин, 40000 женщин). Болели в течение года 7000 мужчин и 14000 женщин. Требуется рассчитать интенсивные и стандартизованные показатели заболеваемости (на 1000 жителей), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вариант 2

Требуется сравнить физическое развитие учеников 8-10 классов в 2-х школах. В 1-й школе обследовано 195 человек, из них 31 мальчик и 164 девочки, выявлено, что отставание физического развития есть у 1-го мальчика и 5-ти девочек.

Во 2-й школе обследовано 182 человека, из них 125 мальчиков и 57 девочек, имели отставание 5 мальчиков и 4 девочки. Вычислите интенсивные и стандартизованные показатели частоты отклонений физического развития детей в этих школах, (на 100 детей) сравните их и оцените достоверность различий между ними, сделайте вывод.

Вариант 3

Необходимо сравнить качество протезирования зубов в 2-х стоматологических поликлиниках. В 1-й поликлинике изготовлено 400 протезов, из них 35 мостовидных и 365 единичных; плохая фиксация протезов зафиксирована у 1-го пациента с мостовидным и 5-ти пациентов с единичными протезами. Во 2-й поликлинике установлено 350 протезов, из них 250 мостовидных и 100 единичных; нарушение фиксации обнаружено 12-и пациентов с мостовидными и 5-и с единичными протезами. Вычислите интенсивные и стандартизованные показатели качества протезирования в поликлиниках (на 100 пациентов), проанализируйте и оцените достоверность различий между ними, сделайте вывод.

Вариант 4

Сравните загрязненность воздуха по количеству проб, не отвечающих санитарным требованиям, в 2-х районах города N. В 1-м районе взято 300 проб, из них 30 в жилых кварталах и 270 вблизи автомобильных дорог; обнаружено несоответствие требованиям в 2-х квартальных пробах и 5 у автодороги. Во 2-м районе взято 200 проб, из них 150 квартальных и 50 у дороги; имели отклонение от норматива загрязненности 12 квартальных и 5 проб рядом с автодорогой. Требуется рассчитать интенсивные и стандартизованные показатели загрязненности воздуха в 2-х районах (на 100 проб), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вариант 5

Требуется сравнить качество работы 2-х лечебно-профилактических учреждений (ЛПУ) по числу послеоперационных осложнений. В 1-й клинике проведено 600 хирургических операций, из них 55 полостных в общем отделении и 545 малоинвазивных в эндоскопическом отделении; послеоперационные осложнения зафиксированы у 20-и пациентов, перенесших полостную операцию, и у 5-ти пациентов из эндоскопического отделения. Во 2-й клинике выполнено 350 операций, из них 250 полостных и 100 малоинвазивных; осложнения были зафиксированы у 5-и пациентов общего отделения и у 4-х - эндоскопического. Вычислите интенсивные и стандартизованные показатели послеоперационных осложнений в 2-х ЛПУ (на 100 операций), проанализируйте и оцените достоверность различий между ними, сделайте вывод.

Вариант 6

Сравните загрязненность 2-х водоемов по количеству проб, не отвечающих санитарным требованиям. В 1-м водоеме взято 300 проб, из них 30 поверхностных и 270 с глубины 3 метра; обнаружено несоответствие требованиям в 5-ти поверхностных пробах и 7-ми глубинных. Во 2-м водоеме взято 320 проб, из них 270 поверхностных и 50 глубинных; имели отклонение от норматива загрязненности 8 поверхностных и 3 глубинные пробы. Необходимо рассчитать интенсивные и стандартизованные показатели загрязненности водоемов (на 100 проб), проанализировать и оценить достоверность различий между ними, сделать вывод.

Вопросы для самоконтроля

1. В каких случаях применяется метод стандартизации?
2. В чем заключается суть метода стандартизации?
3. Какие варианты (модификации) метода стандартизации вам известны?
4. Этапы прямого метода стандартизации.
5. О чем свидетельствуют стандартизованные показатели?
6. Как оцениваются стандартизованные показатели?



По окончании работы необходимо:

- 1) Файлы с выполненными заданиями сохранить на рабочем столе и закрыть программу.
- 2) Переместить свои файлы в папку **Документы \Папка с номером группы**.
- 3) Сообщить преподавателю о выполнении заданий.