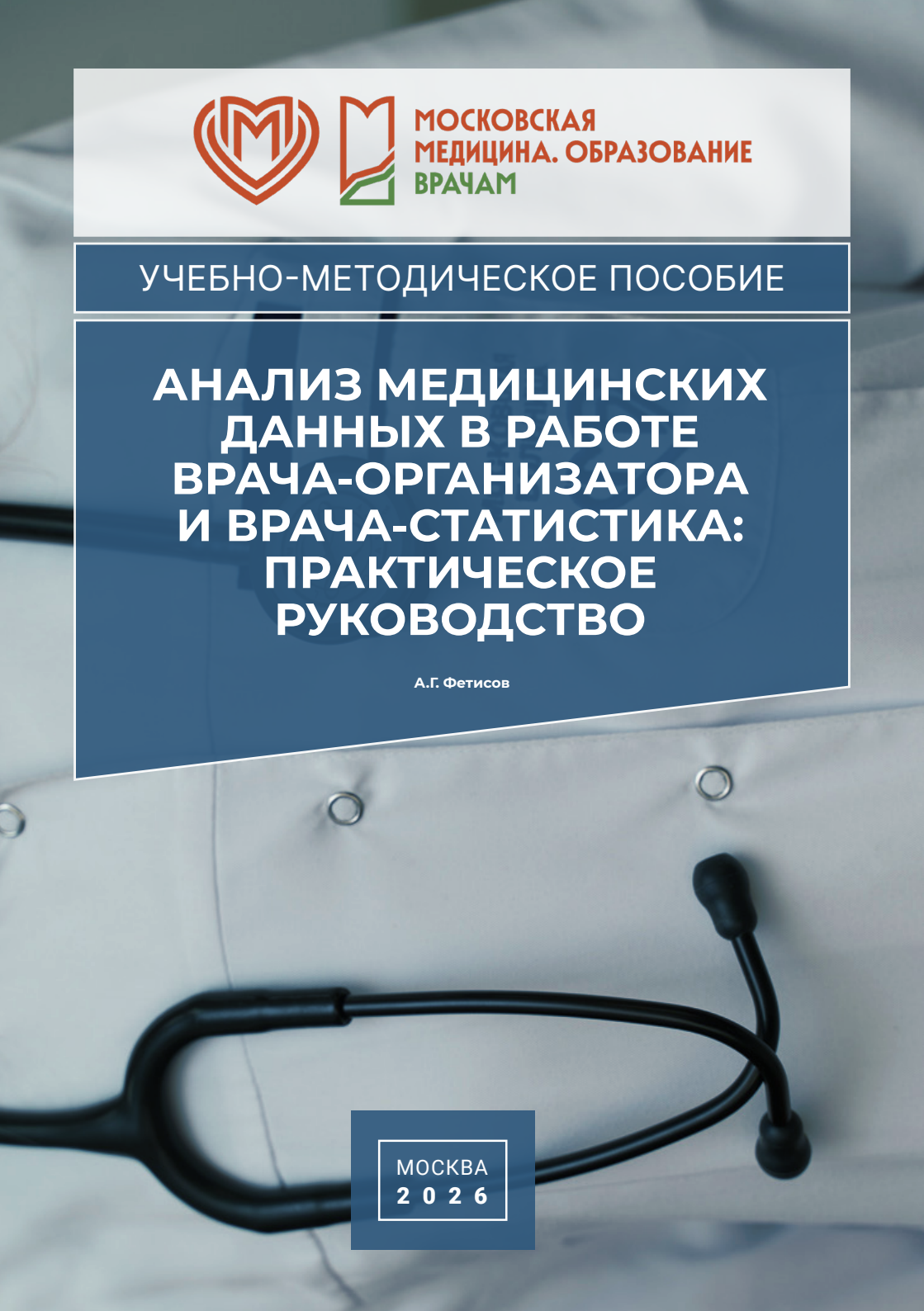


МОСКОВСКАЯ МЕДИЦИНА. ОБРАЗОВАНИЕ ВРАЧАМ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ В РАБОТЕ ВРАЧА-ОРГАНИЗАТОРА И ВРАЧА-СТАТИСТИКА: ПРАКТИЧЕСКОЕ РУКОВОДСТВО

А.Г. Фетисов

МОСКВА
2026

Государственное бюджетное учреждение города Москвы
«Научно-исследовательский институт организации
здравоохранения и медицинского менеджмента
Департамента здравоохранения города Москвы»

АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ В РАБОТЕ ВРАЧА-ОРГАНИЗАТОРА И ВРАЧА-СТАТИСТИКА: ПРАКТИЧЕСКОЕ РУКОВОДСТВО

Учебно-методическое пособие

А.Г. Фетисов

Москва
2026

УДК 614.2
ББК 51.1

Рецензенты:

Дурнев Роман Александрович – д-р техн. наук, первый вице-президент ФГБУ «Российская академия ракетных и артиллерийских наук»;
Гажева Анастасия Викторовна – канд. мед. наук, доцент, начальник отдела координации организационно-методической работы в здравоохранении ГБУ «НИИОЗММ ДЗМ».

Фетисов А.Г.
Анализ медицинских данных в работе врача-организатора и врача-статистика: практическое руководство: учебно-методическое пособие [Электронный ресурс] / А.Г. Фетисов. – Электрон. текстовые дан. – М.: ГБУ «НИИОЗММ ДЗМ», 2026. – URL: https://niioz.ru/moskovskaya-meditsina/izdaniya-nii/metodicheskie-posobiya/ – Загл. с экрана. – 178 с.

ISBN 978-5-907952-99-7

Практическое руководство посвящено эффективной работе с медицинскими данными в среде табличного процессора Microsoft Excel. В отличие от классических учебников по статистике, пособие решает прикладную задачу: помочь врачу-статистику, организатору здравоохранения и другим специалистам самостоятельно превращать сырые выгрузки из медицинских информационных систем в структурированные, наглядные и обоснованные отчеты.

Издание адресовано практикующим врачам-статистикам, заместителям главных врачей по организационно-методической работе, заведующим отделениями, а также всем медицинским специалистам, кто сталкивается с необходимостью обработки и анализа ведомственных данных в Excel. Руководство может использоваться как для самостоятельного изучения, так и в системе дополнительного профессионального образования.

УДК 614.2
ББК 51.1

Утверждено и рекомендовано к печати
Научно-методическим советом ГБУ «НИИОЗММ ДЗМ»
(Протокол № 4 от 21 апреля 2026 г.).

Самостоятельное электронное издание сетевого распространения.

Минимальные системные требования: браузер Internet Explorer/Safari и др.;
скорость подключения к Сети 1 МБ/с и выше.

ISBN 978-5-907952-99-7
9 785907 952997 >

© Фетисов А.Г., 2026
© ГБУ «НИИОЗММ ДЗМ», 2026

ОГЛАВЛЕНИЕ

Введение 5

Часть 1.
От хаоса к структуре: гигиена данных как основа анализа 8

Глава 1.1. Две философии таблиц: «плохая» и «хорошая» 8

Глава 1.2. Первичная очистка данных: базовые инструменты Excel 10

Глава 1.3. Фильтрация и сортировка: искусство отбора данных 15

Глава 1.4. Контроль качества данных: условное форматирование как страж порядка 22

Часть 2.
Математика здравого смысла: описательная статистика 33

Глава 2.1. Считаем итоги: от «сколько всего» к структуре 33

Глава 2.2. Средние величины: когда среднее арифметическое обманывает 39

Глава 2.3. Структура и доли: искусство отвечать на вопрос «Какую часть?» 47

Глава 2.4. Вариабельность данных: искусство видеть разброс 55

Часть 3.
Сравнение и группировка: главные инструменты аналитика 66

Глава 3.1. Сводные таблицы – король анализа 66

Глава 3.2. Формулы с условиями: когда одной функции недостаточно 74

Глава 3.3. Вертикальный поиск: искусство связывать таблицы 82

Глава 3.4. Динамические диапазоны и структурированные ссылки: как сделать формулы умными 92

ОГЛАВЛЕНИЕ

Часть 4.
Графики для главного врача: визуализация данных....98

Глава 4.1. Как выбрать правильный тип графика....98

Глава 4.2. Создание и настройка диаграмм:
от цифр к убедительной презентации....108

Глава 4.3. Прогнозирование: линии тренда как взгляд в будущее....119

Глава 4.4. Интерактивные дашборды: объединяем все в единую картину...127

Часть 5.
Power Query....138

Глава 5.1. Знакомство с Power Query: ваш личный ассистент
по подготовке данных....138

Глава 5.2. Очистка данных за 10 минут: превращаем хаос в порядок....147

Глава 5.3. Склеивание и объединение: из множества
файлов единый отчет....152

Часть 6.
Бонус: быстрая статистика (пакет анализа)....158

Глава 6.1. Что такое «пакет анализа» и как его включить....158

Глава 6.2. Описательная статистика одним кликом....162

Глава 6.3. Простые сравнения для отчетов....167

Заключение....173

Введение



Здравствуйте, уважаемые коллеги!

Эта книга адресована вам, врачам-статистикам, врачам-методистам, заведующим отделениями, заместителям главных врачей по организационно-методической работе и всем, кто ежедневно работает с медицинской статистикой. Мы хорошо понимаем: вы не математики и не программисты. Вы врачи, которые отвечают за качество и эффективность медицинской помощи, ваша задача – превращать горы цифр в понятные и обоснованные управленческие решения.

В повседневной практике перед каждым из нас постоянно возникают одни и те же вопросы. Сколько пациентов реально пролечено в отделении за месяц? Как отличается нагрузка на персонал в понедельник и пятницу? Есть ли устойчивая динамика смертности, средней длительности лечения или числа послеоперационных осложнений? Где находятся узкие места процесса? На что именно нужно обратить внимание главному врачу? Традиционные бумажные отчеты и примитивные подсчеты на калькуляторе уже давно не справляются с современным объемом данных. При этом, как показывает практика, большинство врачей-организаторов до сих пор тратят часы и дни на ручную обработку таблиц в Excel, совершая одни и те же типичные ошибки.

Цель настоящего пособия – дать вам практическое руководство, которое позволит за считаные минуты превращать сырые выгрузки из медицинских информационных систем в чистые надежные данные, проводить полноценный статистический анализ без привлечения внешних специалистов, а также представлять результаты в виде понятных сводных таблиц и наглядных графиков, готовых для включения в доклад главному врачу или в отчет для контролирующих органов. Особое внимание уделено автоматизации рутинных ежемесячных и ежеквартальных отчетов, чтобы освободить ваше время для настоящей аналитической работы.

Как устроена книга?

Пособие построено по принципу «от простого к сложному» и ориентировано исключительно на медицинскую практику. Первая часть посвящена гигиене данных. Мы разберем, почему красивые таблицы с объединенными ячейками и цветными заливками – это главный враг достоверного анализа; научимся превращать хаос в умные таблицы. Во второй части речь пойдет об описательной статистике, о том, как считать средние, медианы, доли и разброс показателей так, чтобы цифры не вводили в заблуждение. Третья часть знакомит с главными рабочими инструментами аналитика: сводными таблицами, условными функциями и вертикальным поиском. Четвертая часть посвящена визуализации данных: выбору правильного типа графика и подготовке его для презентации руководству. Пятая часть раскрывает возможности Power Query как средства настоящей автоматизации ежемесячной рутины. И наконец, бонусная шестая часть рассказывает о пакете анализа для тех, кому нужно быстро получить расширенную статистику, включая разбор практического кейса с парным t-тестом на медицинских данных.

Книга максимально практическая. Каждый раздел содержит пошаговые инструкции, реальные медицинские кейсы из работы врача-статистика и готовые шаблоны. В конце разделов предусмотрены упражнения с учебными датасетами, которые доступны для скачивания по ссылкам в приложении.

Технические требования и выбор инструмента

Прежде чем мы перейдем к практическим занятиям, важно ответить на вопрос, который неизбежно встает перед каждым организатором здравоохранения: какой инструмент выбрать для повседневной работы? Рынок предлагает несколько вариантов, и у каждого есть свои сильные стороны. Наша задача – сделать выбор осознанно, исходя из реальных задач медицинской статистики.

Google Sheets – это облачный инструмент, который выигрывает там, где на первом месте стоит оперативность обмена информацией и совместная работа. Несколько специалистов – заведующий отделением, заместитель главного врача, методист – могут одновременно работать над одной сводкой, видеть изменения в реальном времени и оставлять комментарии. Это идеальный вариант для небольших оперативных отчетов, контроля дневных показателей или быстрого согласования данных. Однако при переходе к полноценному статистическому анализу его возможности начинают буксовать. Sheets заметно теряет производительность уже при 50–100 тыс. строк, что критично для годовых реестров пациентов. В нем нет мощного инструмента для автоматизации, подобного Power Query. Кроме того, работа с конфиденциальными персональными данными в облаке Google

может противоречить внутренним политикам безопасности и требованиям Росздравнадзора по защите персональных данных (Федеральный закон № 152-ФЗ). Поэтому Sheets остается отличным дополнением для финальной презентации данных, но не базовым инструментом для их глубокой обработки.

Apache OpenOffice Calc – это бесплатный открытый табличный процессор, который может рассматриваться как вариант для учреждений с нулевым бюджетом на ПО. Он позволяет решать базовые задачи: использовать функции ЕСЛИ, СУММЕСЛИ, строить несложные сводные таблицы (DataPilot) и графики. Но для серьезной работы врача-статистика его ограничения становятся критическими: отсутствует Power Query, сводные таблицы имеют жесткие ограничения, производительность падает при работе с большими массивами, нет встроенного статистического пакета. Calc – это возможный выбор для очень простых задач, но он не подходит для автоматизации и глубокого анализа.

Microsoft Excel (версии 2016 и выше) остается безусловным лидером в работе врача-статистика и врача-организатора, где цена ошибки – это искаженные показатели качества медицинской помощи и эффективности использования ресурсов. Наше пособие полностью построено на его возможностях, потому что Excel уверенно работает с миллионами строк, предоставляет мощный инструмент автоматизации Power Query, позволяет проводить глубокий анализ с помощью сводных таблиц, срезов, временных шкал и встроенного пакета анализа (t-тесты, корреляция, описательная статистика), а также обеспечивает безопасность и контроль, позволяя работать полностью офлайн и хранить данные на локальных серверах учреждения.

Практический итог нашей работы прост: для ежедневной рутины и оперативной совместной правки можно использовать Google Sheets, для простых расчетов при минимальных бюджетах базово подойдет OpenOffice Calc, но для полноценного статистического анализа, автоматизации ежемесячной и годовой отчетности, научной работы и уверенного доклада главному врачу ваш выбор – Microsoft Excel. Именно поэтому все наши дальнейшие главы об умных таблицах, Power Query, сводных таблицах и статистических тестах построены на Excel как на самом надежном и функциональном инструменте, отвечающем самым высоким требованиям здравоохранения.

Мы начинаем с главного: учимся делать так, чтобы цифры не врали. Добро пожаловать в мир точной и удобной медицинской статистики в Excel!



Часть 1. От хаоса к структуре: гигиена данных как основа анализа

Глава 1.1. Две философии таблиц: «плохая» и «хорошая»

Представьте, что вы получаете очередную выгрузку из медицинской информационной системы. Открыв файл, вы с первых секунд понимаете: анализ превратится в испытание. Заголовки для красоты разбросаны по нескольким строкам, ячейки объединены, между записями зияют пустые строки, а в столбцах с цифрами прячутся текстовые пометки. Это классический пример плохой таблицы, которая, к сожалению, слишком хорошо знакома каждому врачу-статистику.

Почему такой формат – главный враг достоверности? Потому что он закладывает мины под все дальнейшие расчеты. Объединенные ячейки искажают работу сводных таблиц, текстовые значения делают невозможным подсчет средних, а ручные цветовые заливки, которые кажутся такими удобными для визуального выделения, Excel попросту не видит. Потратив часы на анализ такой таблицы, вы рискуете в итоге представить руководству искаженные показатели смертности, средней длительности лечения или структуры коечного фонда.

Хорошая новость в том, что привести данные в порядок можно быстро и эффективно, если усвоить несколько простых, но железных правил. Настоящая хорошая таблица, готовая к анализу, выглядит на первый взгляд скучно, но именно эта строгость – залог надежности. Вот ее ключевые признаки.

Строгая структура: первая строка содержит только уникальные и понятные заголовки столбцов (например, «Дата_госпитализации», «Ф.И.О. пациента», «Отделение», «Койко-дни»).

Отсутствие «украшательств»: внутри области данных нет ни одной объединенной ячейки, пустых строк или столбцов. Любое форматирование (цвет,

шрифт) должно быть строго функциональным, а лучше отсутствовать вовсе.

Чистота типов данных: каждый столбец содержит информацию только одного типа: только даты, только числа или только текст. Смешивание форматов недопустимо.

Ручные подписи и цветные ячейки – это зло не только потому, что Excel их не видит. Они нарушают главный принцип работы с данными: автоматизацию. Когда вы будете добавлять новые строки за следующий месяц, все эти красивости придется расставлять вручную заново, что неизбежно приведет к ошибкам и потере времени.

Первая привычка профессионала: умная таблица (Smart Table) Самый мощный и простой прием, который должен стать вашим рефлексом, – это мгновенное превращение обычного диапазона в умную таблицу. Делается это в два клика.

1. Выделите любую ячейку внутри ваших данных.

2. Нажмите сочетание клавиш Ctrl + T (или Ctrl + L). В появившемся окне убедитесь, что Excel правильно определил диапазон, и нажмите OK.

Что вы получаете в результате? Во-первых, автоматические фильтры в заголовках. Во-вторых, что важнее, таблица становится динамической: при добавлении новых строк она расширяется автоматически, и все формулы, сводные таблицы и графики, созданные на ее основе, будут подхватывать эти изменения без вашего участия. Вы можете сразу дать таблице понятное имя (например, «Пациенты_2026») через вкладку «Конструктор».

Практический кейс. В одном из крупных стационаров мы столкнулись с проблемой: ежемесячная выгрузка из МИС приходила с объединенными ячейками в столбце «Отделение». На разбор этой красоты уходило 30–40 мин. После того как в отделении ввели железное правило «сначала умная таблица», время первичной подготовки данных сократилось до 2–3 мин.

Запомните главное правило, которое стоит написать на стикере и приклеить к монитору: «Прежде чем считать хоть одну цифру – преврати данные в умную таблицу». Это первый и самый важный шаг к тому, чтобы ваши цифры перестали врать и начали работать на вас.

Глава 1.2. Первичная очистка данных: базовые инструменты Excel

Даже после того как мы превратили обычный диапазон в умную таблицу, данные из медицинских информационных систем редко бывают готовы к анализу. Выгрузки из 1С-Медицина, ЕГИСЗ или региональных МИС почти всегда приносят с собой привычный багаж проблем: повторяющиеся записи одного и того же пациента, разнобой в обозначении пола, лишние пробелы, Ф.И.О. в одной ячейке и даты, которые Excel упорно считает текстом. На первый взгляд эти мелочи кажутся несущественными, но именно они приводят к серьезным искажениям итоговых показателей: завышенному числу пролеченных больных, неверной структуре заболеваемости, ошибочному расчету средней длительности лечения и койко-дней.

Первичная очистка данных – это не скучная формальность, а самый эффективный способ сэкономить время. В реальной практике именно здесь, на этапе подготовки, экономится до 80% тех часов, которые раньше уходили на мучительное исправление ошибок в уже готовых отчетах. Освоив несколько простых инструментов, вы перестанете тратить дни на рутину и сможете сосредоточиться на настоящей аналитической работе.

Удаление дубликатов: когда повтор не лечит.

Одна из самых распространенных проблем, с которой мы сталкиваемся, – это повторяющиеся строки. В медицинской практике дубликаты возникают постоянно и по разным причинам. Один и тот же пациент может быть дважды выгружен за сутки как вызов скорой помощи. Техническая ошибка системы способна создать две записи о повторной госпитализации в приемное отделение с интервалом в несколько минут. Даже при проведении диспансеризации нередко появляются дублирующие строки с небольшой разницей во времени. Каждый такой повтор искусственно завышает показатели объема медицинской помощи, искажая все последующие расчеты от нагрузки на отделения до структуры коечного фонда.

Удаление дубликатов в Excel выполняется буквально в несколько кликов. На вкладке «Данные» находим кнопку «Удалить дубликаты» и перед нами открывается окно, где Excel предлагает выбрать столбцы для поиска совпадений. В медицинской практике обычно отмечают ключевые поля: «Номер истории болезни», «Ф.И.О. пациента», «Дата рождения» и «Дата обращения». После нажатия ОК программа выдает сообщение «Удалено X дубликатов». В крупном стационаре за месяц может удаляться от 50 до 300 повторных записей, каждая из которых, останься она в таблице, искажала бы реальную картину (рис. 1.1).

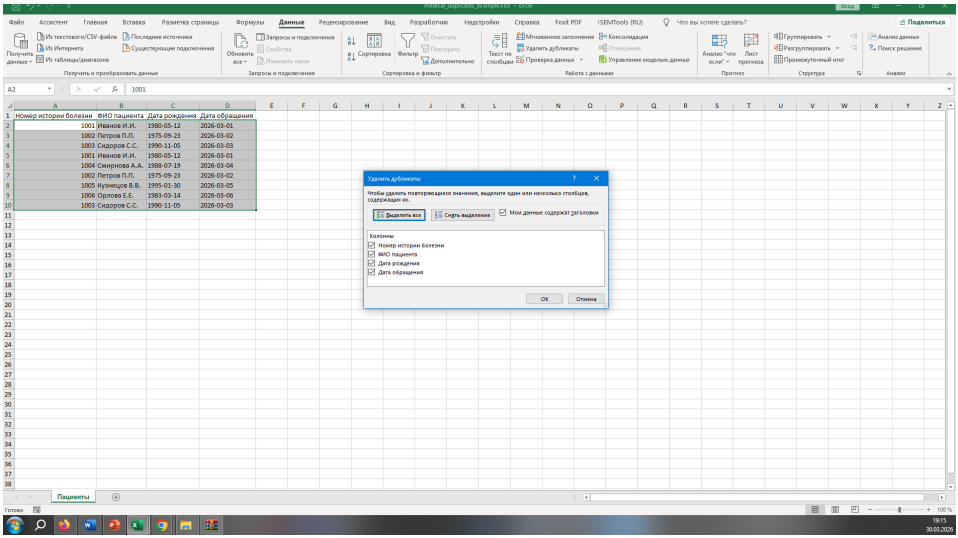


Рисунок 1.1. Удалить дубликаты

Однако здесь необходим профессиональный подход. Прежде чем нажать кнопку удаления, всегда стоит сделать копию исходного листа. Иногда повтор – это не ошибка, а реальное клиническое событие: пациент мог поступить повторно с ухудшением состояния в тот же день. В таких случаях разумнее добавить вспомогательный столбец с формулой для выявления потенциальных дублей и проверить их вручную. Это займет несколько минут, но убережет от потери важной информации.

Приведение текста к единому виду: укрощение хаоса.

Вторая головная боль врача-статистика – разнобой в написании одних и тех же сущностей. Пол пациента может обозначаться десятком способов: «м», «М», «муж», «Мужской», «мужской», «1», «Муж.». Диагнозы приходят в виде «Гипертоническая болезнь», «ГБ», «гипертония», «Артериальная гипертензия». Фамилии врачей могут содержать лишние пробелы или быть записаны разным регистром. Если оставить все как есть, сводная таблица покажет не два отделения, а десять потому, что название одного и того же отделения введено с разными вариантами написания.

Справиться с этим хаосом помогает простой, но мощный инструмент «Найти и заменить» (Ctrl + H). Выделяем столбец, открываем окно и последовательно заменяем варианты написания на единый стандарт. Например, находим «муж» – заменяем на «Мужской», находим «м» – заменяем на «Мужской», находим «1» –

заменяем на «Мужской». После нескольких таких проходов столбец становится пригодным для анализа (рис. 1.2).

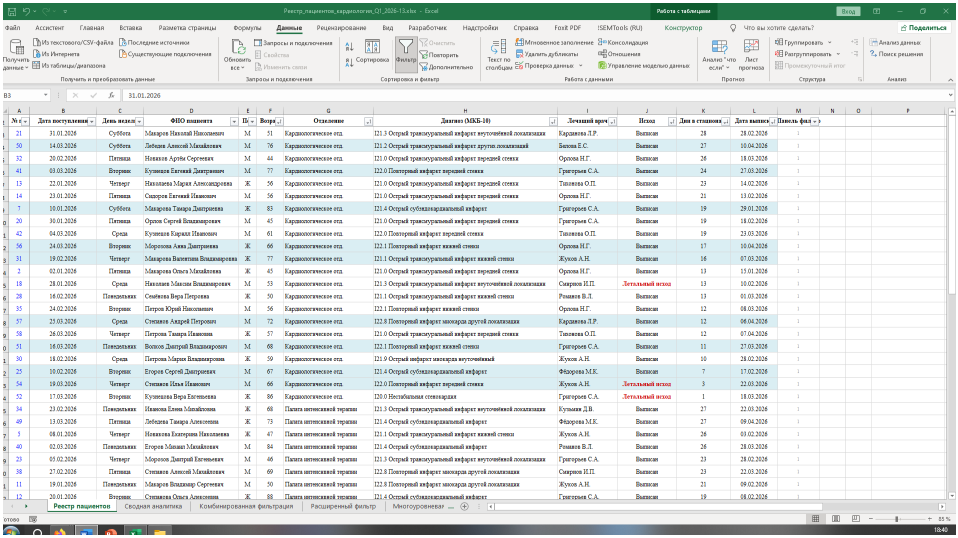


Рисунок 1.2. «Найти и заменить»

Однако даже после замены слов в ячейках часто остаются лишние пробелы в начале или конце невидимые глазу, но смертельные для точности сопоставления данных. Здесь на помощь приходит функция СЖПРОБЕЛЫ. В соседнем столбце вводим формулу =СЖПРОБЕЛЫ(A2), протягиваем вниз, затем копируем результат и вставляем обратно специальным способом («Специальная вставка» → «Значения»). Теперь текст чист, пробелы удалены, Excel больше не путает «Отделение кардиологии» с «Отделение кардиологии».

Разделение данных по столбцам: из одной ячейки – в структуру.

Еще одна частая проблема – данные, склеенные в одной ячейке. Классический пример: Ф.И.О. пациента и дата рождения записаны вместе «Иванов Иван Иванович 12.05.1965». Для анализа структуры по возрасту, полу или частотности фамилий такой формат совершенно непригоден. Нам нужны отдельные столбцы: фамилия, имя, отчество, дата рождения.

Инструмент «Текст по столбцам» решает эту задачу за минуту. Выделяем столбец, на вкладке «Данные» выбираем «Текст по столбцам», указываем «С разделителями» и ставим галочку «Пробел». На следующем шаге можно за-

дать формат данных: для Ф.И.О. выбираем «Текст», для даты рождения – «Дата». Excel аккуратно разносит каждую часть по отдельным столбцам, превращая хаотичную строку в аккуратную структуру (рис. 1.3).

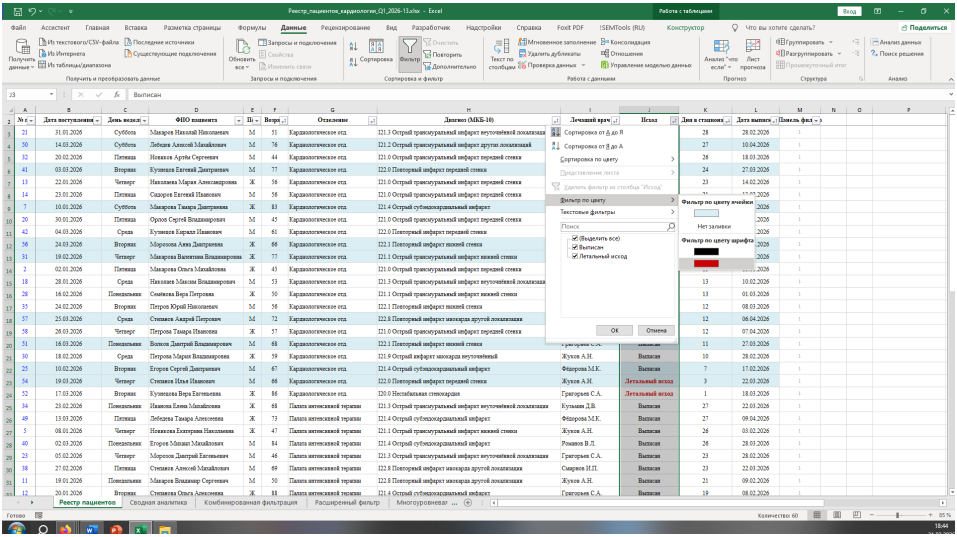


Рисунок 1.3. Инструмент «Текст по столбцам»

Проверка и исправление типов данных: заставляем Excel видеть числа.

Самый коварный враг достоверного анализа – неверно распознанные типы данных. Вы можете часами строить сводные таблицы, но, если Excel считает числа текстом, средние значения не будут считаться, а суммы окажутся нулевыми. Распознать проблему просто: в стандартной настройке числа выравниваются по правому краю, а текст – по левому. Если в столбце с койко-днями или возрастом пациентов цифры прижаты влево, значит, Excel их не видит.

Особенно часто это происходит с датами. Они могут выглядеть привычно: «15.03.2025», но функция СРЗНАЧ или группировка по месяцам в сводной таблице не работают. Еще одна распространенная ситуация: при выгрузке из зарубежных систем десятичная запятая заменена на точку, Excel отказывается воспринимать такие числа.

Способы исправления есть, и они просты. Самый быстрый лайфхак: выделить столбец, перейти на вкладку «Данные», выбрать «Текст по столбцам» и сразу нажать «Готово» без каких-либо дополнительных настроек. Excel просыпается и переопределяет тип данных автоматически. Для дат можно также использо-

вать формат ячеек (Ctrl + 1 → вкладка «Число» → «Дата»). Для чисел с неверным разделителем помогает функция ПОДСТАВИТЬ или временная замена точки на запятую через «Найти и заменить» (рис. 1.4).

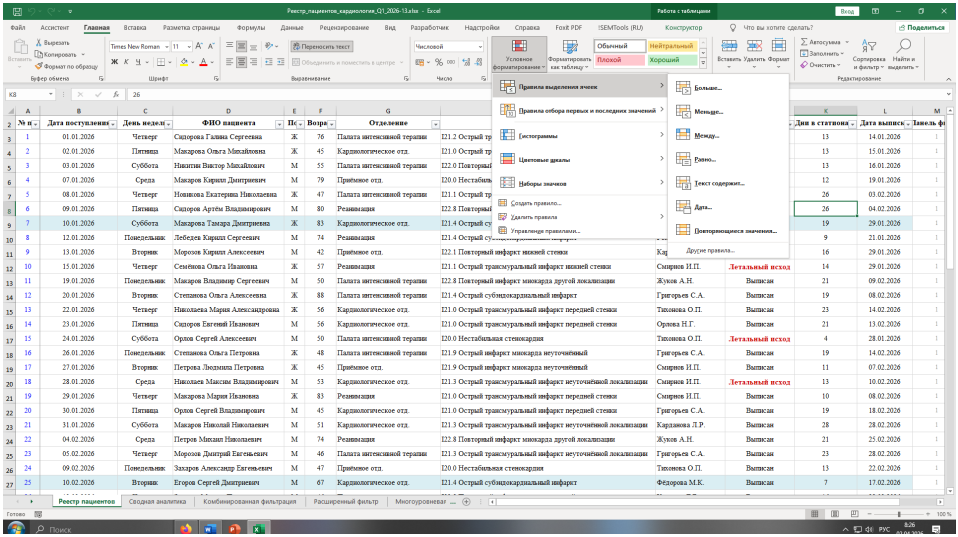


Рисунок 1.4. Столбец «Койко-день» до и после исправления типа данных — числа выровнены по правому краю

Практический кейс: правило пятиминутной очистки.

В одном многопрофильном стационаре на 800 коек мы столкнулись с типичной ситуацией: ежемесячная подготовка данных для отчета отнимала 2–3 ч, каждый раз находились нестыковки, которые приходилось перепроверять. После внедрения простого алгоритма – «пятиминутки очистки» – время подготовки сократилось до 15–20 мин., а жалобы главного врача на странные цифры исчезли.

Вот этот алгоритм, который мы рекомендуем взять за основу (рис. 1.5).

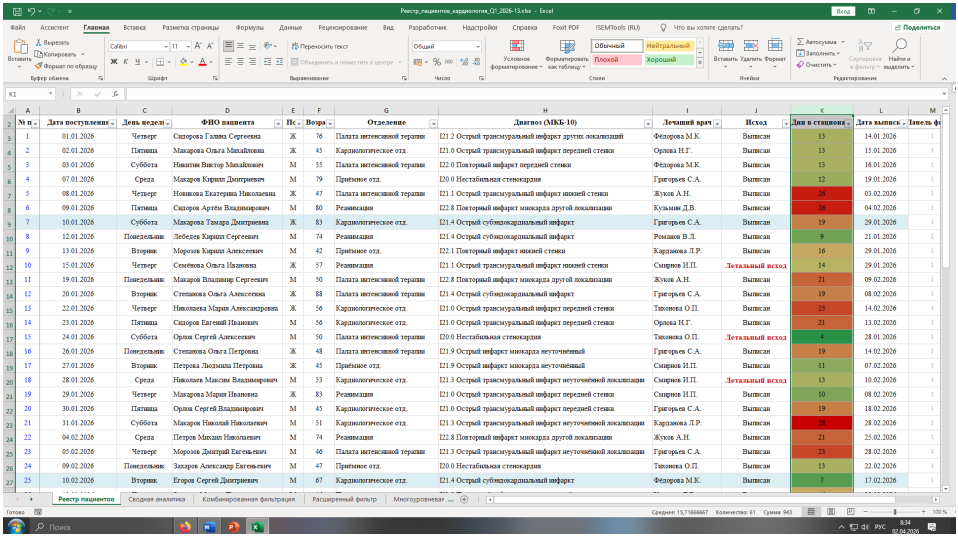
1. **Умная таблица: Ctrl + T сразу после открытия файла.**

2. **Удаление дубликатов: по номеру истории болезни и дате обращения.**

3. **Приведение текста к единому виду: замена вариантов написания пола,**

отделений, диагнозов, затем СЖПРОБЕЛЫ.

4. **Разделение Ф.И.О. и исправление типов данных: «Текст по столбцам» и проверка числовых столбцов.**



№ п	Дата поступления	День недели	ФИО пациента	Пс	Возр	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дня в стациона	Дата выпис	Панель ф
1	01.01.2026	Четверг	Сидорова Галина Сергеевна	Ж	76	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Фёдорова М.К.	Выписан	13	14.01.2026	1
2	02.01.2026	Пятница	Макарова Ольга Михайловна	Ж	45	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Орлова Н.Г.	Выписан	13	15.01.2026	1
3	03.01.2026	Суббота	Никитин Виктор Михайлович	М	55	Палата интенсивной терапии	I22.0 Повторный инфаркт передней стенки	Фёдорова М.К.	Выписан	13	16.01.2026	1
4	07.01.2026	Среда	Макаров Кирилл Дмитриевич	М	79	Приёмное отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Выписан	12	19.01.2026	1
5	08.01.2026	Четверг	Новикова Екатерина Николаевна	Ж	47	Палата интенсивной терапии	I21.1 Острый трансмуральный инфаркт нижней стенки	Жуков А.Н.	Выписан	26	03.02.2026	1
6	09.01.2026	Пятница	Сидоров Артём Владимирович	М	80	Реанимация	I22.8 Повторный инфаркт миокарда другой локализации	Кузьмин Д.В.	Выписан	26	04.02.2026	1
7	10.01.2026	Суббота	Макарова Тамара Дмитриевна	Ж	83	Кардиологическое отд.	I21.4 Острый субэпикардиальный инфаркт	Григорьев С.А.	Выписан	19	29.01.2026	1
8	12.01.2026	Понедельник	Лебедев Кирилл Сергеевич	М	74	Реанимация	I21.4 Острый субэпикардиальный инфаркт	Романов В.Л.	Выписан	9	21.01.2026	1
9	13.01.2026	Вторник	Морозов Кирилл Алексеевич	М	42	Приёмное отд.	I22.1 Повторный инфаркт нижней стенки	Кардакова Л.Р.	Выписан	16	29.01.2026	1
10	15.01.2026	Четверг	Семёнова Ольга Ивановна	Ж	57	Реанимация	I21.1 Острый трансмуральный инфаркт нижней стенки	Смирнов И.П.	Летальный исход	14	29.01.2026	1
11	19.01.2026	Понедельник	Макаров Владимир Сергеевич	М	50	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Жуков А.Н.	Выписан	21	09.02.2026	1
12	20.01.2026	Вторник	Степанова Ольга Алексеевна	Ж	88	Палата интенсивной терапии	I21.4 Острый субэпикардиальный инфаркт	Григорьев С.А.	Выписан	19	08.02.2026	1
13	22.01.2026	Четверг	Николаева Мария Александровна	Ж	56	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	23	14.02.2026	1
14	23.01.2026	Пятница	Сидоров Евгений Иванович	М	56	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Орлова Н.Г.	Выписан	21	13.02.2026	1
15	24.01.2026	Суббота	Орлов Сергей Алексеевич	М	50	Палата интенсивной терапии	I20.0 Нестабильная стенокардия	Тихонова О.П.	Летальный исход	4	28.01.2026	1
16	26.01.2026	Понедельник	Степанова Ольга Петровна	Ж	48	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Григорьев С.А.	Выписан	19	14.02.2026	1
17	27.01.2026	Вторник	Петрова Людмила Петровна	Ж	45	Приёмное отд.	I21.9 Острый инфаркт миокарда неуточнённый	Смирнов И.П.	Выписан	11	07.02.2026	1
18	28.01.2026	Среда	Николаев Максим Владимирович	М	53	Кардиологическое отд.	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Смирнов И.П.	Летальный исход	13	10.02.2026	1
19	29.01.2026	Четверг	Макарова Мария Ивановна	Ж	83	Реанимация	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	10	08.02.2026	1
20	30.01.2026	Пятница	Орлов Сергей Владимирович	М	45	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Григорьев С.А.	Выписан	19	18.02.2026	1
21	31.01.2026	Суббота	Макаров Николай Николаевич	М	51	Кардиологическое отд.	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Кардакова Л.Р.	Выписан	28	28.02.2026	1
22	04.02.2026	Среда	Петров Михаил Николаевич	М	74	Реанимация	I22.8 Повторный инфаркт миокарда другой локализации	Жуков А.Н.	Выписан	21	25.02.2026	1
23	05.02.2026	Четверг	Морозов Дмитрий Евгеньевич	М	46	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Григорьев С.А.	Выписан	23	28.02.2026	1
24	09.02.2026	Понедельник	Захаров Александр Евгеньевич	М	47	Приёмное отд.	I20.0 Нестабильная стенокардия	Тихонова О.П.	Выписан	13	22.02.2026	1
25	10.02.2026	Вторник	Егоров Сергей Дмитриевич	М	67	Кардиологическое отд.	I21.4 Острый субэпикардиальный инфаркт	Фёдорова М.К.	Выписан	7	17.02.2026	1

Рисунок 1.5. Схема алгоритма «пятиминутной очистки» в виде последовательных шагов

Главное правило этой главы можно сформулировать коротко и жестко: чистые данные – фундамент достоверной медицинской статистики. Прежде чем считать средние величины, доли, строить сводные таблицы или графики, убедитесь, что в таблице нет мусора. Только тогда ваши цифры будут точно отражать реальную работу отделения и всего учреждения и помогут принимать правильные управленческие решения.

Глава 1.3. Фильтрация и сортировка: искусство отбора данных

После того как мы очистили данные от дубликатов, привели текст к единому виду и исправили типы данных, наступает момент, когда таблица становится по-настоящему послушной. Теперь она готова к главному: к тому, чтобы мы могли задавать ей точные вопросы и получать быстрые ответы. Инструменты фильтрации и сортировки в Excel позволяют делать это буквально за секунды,

избавляя от мучительного пролистывания тысяч строк в поисках нужной информации.

Представьте себе типичную задачу врача-статистика. Вам нужно ответить на вопросы: сколько пациентов с острым инфарктом миокарда поступило в кардиологическое отделение за последний квартал? Как распределились случаи летальных исходов по дням недели? Какие пять врачей имеют наибольшую нагрузку в приемном отделении? Без инструментов фильтрации и сортировки ответы на эти вопросы потребовали бы часов ручной работы, с ними – нескольких кликов.

Обычный фильтр: первый шаг к точным ответам.

Когда мы превратили данные в умную таблицу (напомню, Ctrl + T), в заголовках каждого столбца автоматически появились маленькие кнопки-стрелочки. Это и есть обычный фильтр – самый простой и при этом невероятно мощный инструмент отбора данных. Нажав на такую стрелочку, вы видите полный список уникальных значений в столбце и можете одним кликом оставить только нужные, сняв галочки с остальных (рис. 1.6).

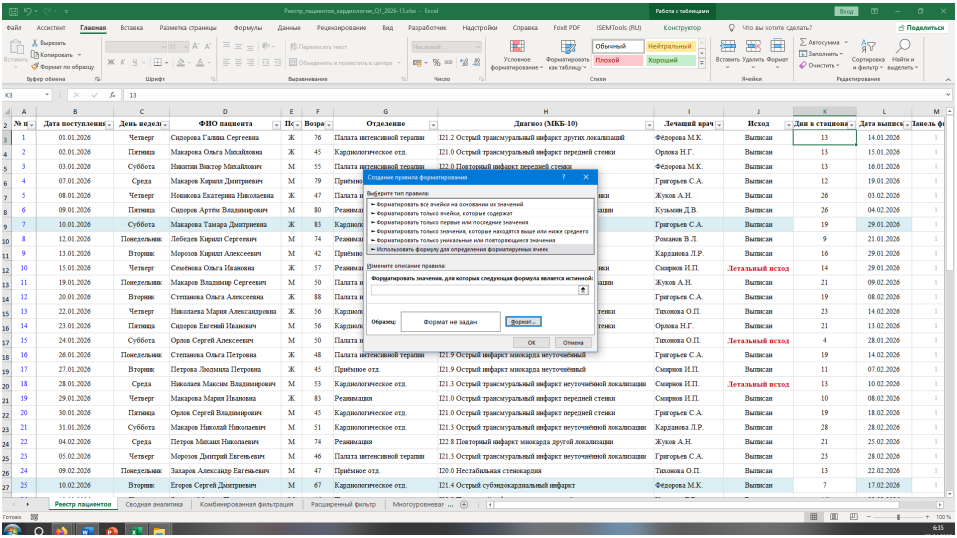


Рисунок 1.6. Кнопка фильтра в заголовке столбца «Отделение» с раскрытым списком уникальных значений

В медицинской практике фильтр становится незаменимым помощником в самых разных ситуациях. Вы можете отобрать только пациентов конкретного отделения, чтобы проанализировать структуру заболеваемости именно в нем, или оставить только вызовы скорой помощи за прошедшие сутки, чтобы оценить ночную нагрузку, или отфильтровать только те случаи, где длительность лечения превысила 30 дней, – это позволит быстро найти выпадающие случаи для клинико-экспертной работы.

Особенно ценна возможность комбинировать фильтры по нескольким столбцам одновременно. Например, вы можете отфильтровать пациентов кардиологического отделения, у которых в диагнозе присутствует «инфаркт», а возраст превышает 65 лет. Excel показывает только те строки, которые удовлетворяют всем условиям одновременно. Внизу строки состояния при этом отображается количество отфильтрованных записей, и вы сразу видите, сколько случаев попало в выборку (рис. 1.7).

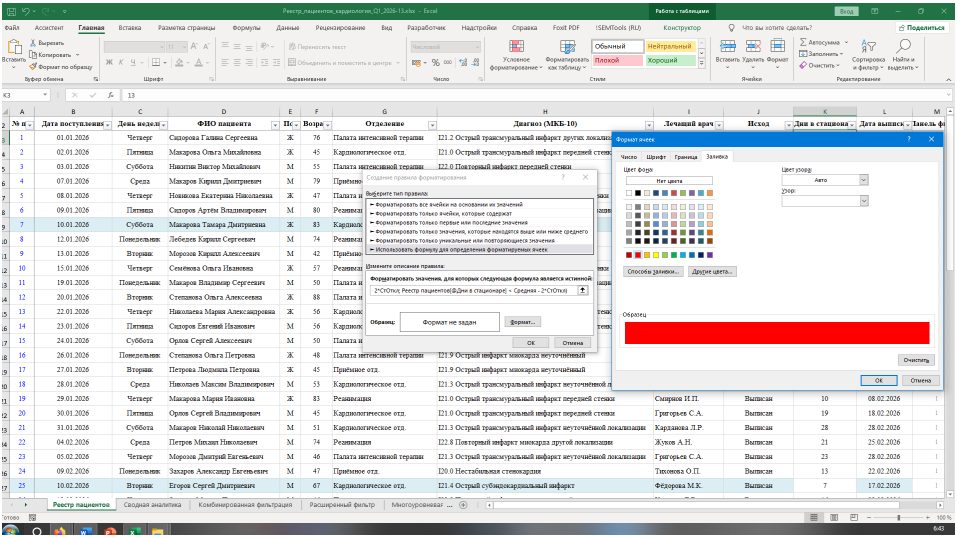


Рисунок 1.7. Строка состояния Excel с отображением количества отфильтрованных записей

Расширенный фильтр: когда простого фильтра недостаточно.

Обычный фильтр удобен для быстрых отборов, но у него есть ограничения. Что делать, если нужно отобрать пациентов с длительностью лечения от 7 до 14 дней? Или выбрать всех пациентов, у которых дата выписки приходится на выходные дни? Или отфильтровать записи, где фамилия врача начинается на

определенную букву? Для таких задач существует расширенный фильтр.

Расширенный фильтр работает по принципу «критерии отдельно, данные отдельно». Вы создаете небольшую таблицу с условиями отбора где-нибудь на свободном месте листа, а затем указываете Excel, что нужно отфильтровать данные по этим критериям. Верхняя строка таблицы критериев должна содержать точные заголовки столбцов, к которым применяются условия. В строках ниже вы записываете сами условия (рис. 1.8).

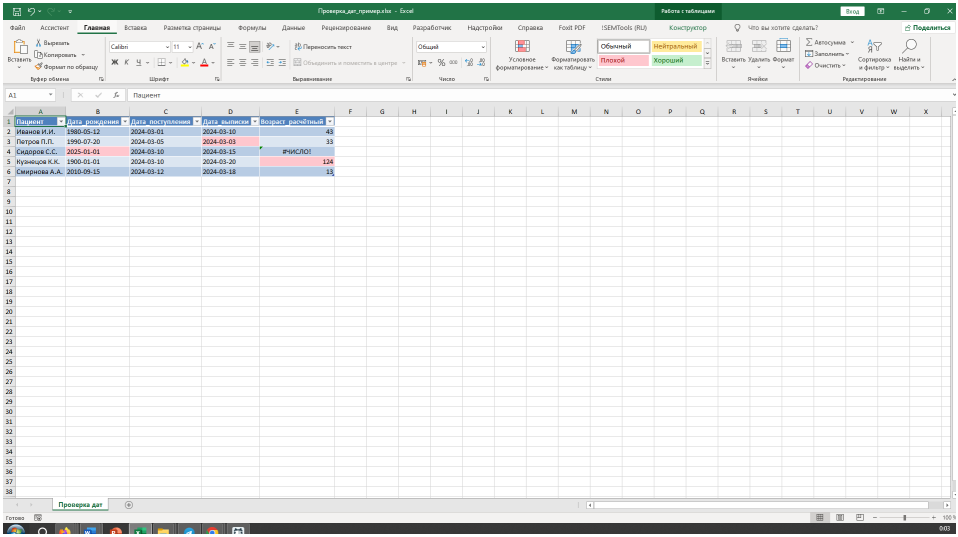


Рисунок 1.8. Пример таблицы критериев для расширенного фильтра с условиями по койко-дням и дате выписки

Главная сила расширенного фильтра в гибкости условий. Вы можете использовать операторы сравнения: >7 и <14 для диапазона длительности лечения. Можно применять логические связки: если записать условия в одной строке, они работают как «И» (должны выполняться одновременно); если в разных строках – как «ИЛИ» (достаточно выполнения хотя бы одного). Можно использовать подстановочные символы: звездочка (*) заменяет любую последовательность символов, вопросительный знак (?) – любой одиночный символ. Например, условие Иванов* отберет всех пациентов, чья фамилия начинается на «Иванов», независимо от инициалов.

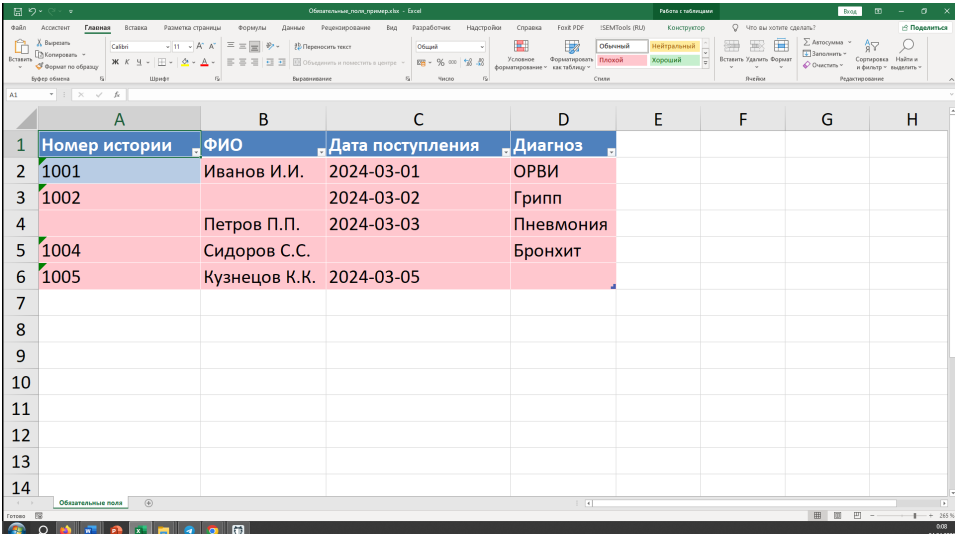
В работе врача-статистика расширенный фильтр особенно полезен при подготовке сложных выборок для клинико-статистических исследований. Напри-

мер, вы можете создать критерии для отбора пациентов с сахарным диабетом 2-го типа, у которых зафиксированы осложнения в определенный период, и при этом возраст на момент госпитализации был в заданном диапазоне. Такую выборку можно сохранить, экспортировать на отдельный лист и использовать для дальнейшего углубленного анализа.

Многоуровневая сортировка: наводим порядок в хаосе.

Если фильтр помогает отобрать нужные данные, то сортировка позволяет организовать их в удобном для анализа порядке. Но редко когда бывает достаточно отсортировать данные по одному столбцу. Гораздо чаще требуется многоуровневая сортировка, когда порядок записей определяется несколькими условиями.

Представьте себе задачу: нужно подготовить список пациентов для обхода заведующего отделением. Логично отсортировать сначала по отделениям, чтобы не переключаться между корпусами. Внутри каждого отделения по дате поступления, чтобы видеть самых «свежих» пациентов первыми. А внутри одной даты – по фамилии, чтобы быстро найти нужную карту. Многоуровневая сортировка решает эту задачу за несколько секунд (рис. 1.9).



	A	B	C	D	E	F	G	H
1	Номер истории	ФИО	Дата поступления	Диагноз				
2	1001	Иванов И.И.	2024-03-01	ОРВИ				
3	1002		2024-03-02	Грипп				
4		Петров П.П.	2024-03-03	Пневмония				
5	1004	Сидоров С.С.		Бронхит				
6	1005	Кузнецов К.К.	2024-03-05					
7								
8								
9								
10								
11								
12								
13								
14								

Рисунок 1.9. Окно «Сортировка» с настроенными уровнями: сначала по отделению, затем по дате поступления, затем по фамилии

Выполняется она через окно «Сортировка» (вкладка «Данные» → «Сортировка»). В этом окне вы добавляете уровни один за другим, указывая для каждого столбец, порядок (по возрастанию или убыванию) и при необходимости учитывать ли регистр или сортировать по цвету ячейки. Excel последовательно применяет уровни: сначала группирует все строки по первому столбцу, затем внутри каждой группы сортирует по второму и т.д.

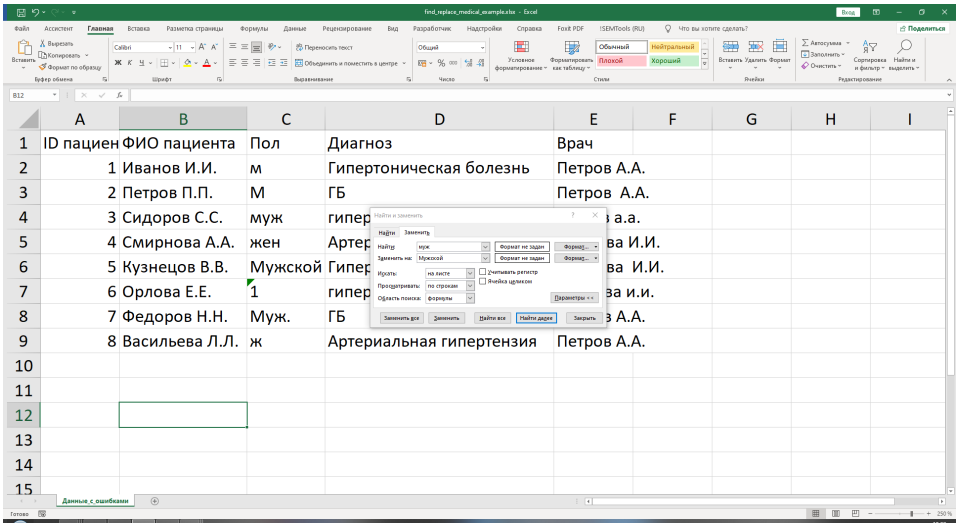
В медицинской статистике многоуровневая сортировка применяется постоянно (рис. 1.10).

- **По отделению, затем по длительности лечения (убывание), чтобы быстро выявить пациентов с наибольшим сроком госпитализации в каждом отделении.**

- **По дате выписки, затем по исходу, чтобы проанализировать, как распределялись летальные исходы во времени.**

- **По врачу для оценки активности разных специалистов.**

- **По нозологии, затем по возрасту для изучения возрастной структуры внутри групп заболеваний.**



	A	B	C	D	E	F	G	H	I
1	ID пациен	ФИО пациента	Пол	Диагноз	Врач				
2	1	Иванов И.И.	м	Гипертоническая болезнь	Петров А.А.				
3	2	Петров П.П.	М	ГБ	Петров А.А.				
4	3	Сидоров С.С.	муж	гипер	а.а.				
5	4	Смирнова А.А.	жен	Артер	ва И.И.				
6	5	Кузнецов В.В.	Мужской	Гипер	ва И.И.				
7	6	Орлова Е.Е.	1	гипер	ва и.и.				
8	7	Федоров Н.Н.	Муж.	ГБ	в А.А.				
9	8	Васильева Л.Л.	ж	Артериальная гипертензия	Петров А.А.				
10									
11									
12									
13									
14									
15									

Рисунок 1.10. Пример отсортированной таблицы с выделенными уровнями группировки

Важный нюанс: если вы работаете с умной таблицей, сортировка сохраняется как часть ее состояния. Вы можете отсортировать данные одним способом сегодня, сохранить файл, открыть завтра, и порядок сохранится. При этом фильтры и сортировка не нарушают целостность данных: все строки остаются на месте, меняется только порядок их отображения.

Поиск по цвету и условному значку: визуальная навигация.

Иногда данные уже содержат визуальные пометки – цветовые заливки ячеек или значки условного форматирования. Возможно, это ручные пометки, которые вы унаследовали от предыдущего отчетного периода, а возможно, результат автоматического условного форматирования, настроенного для выделения критических значений. Excel позволяет фильтровать и сортировать именно по этим визуальным признакам.

Эта возможность особенно полезна, когда вы работаете с данными, где уже есть некоторая предварительная разметка. Например, в ежемесячной выгрузке красным цветом выделены пациенты с летальным исходом, а желтым – случаи с осложнениями. С помощью фильтра по цвету вы можете одним кликом оставить только красные строки и проанализировать все летальные случаи, не отвлекаясь на остальные записи (рис. 1.11).

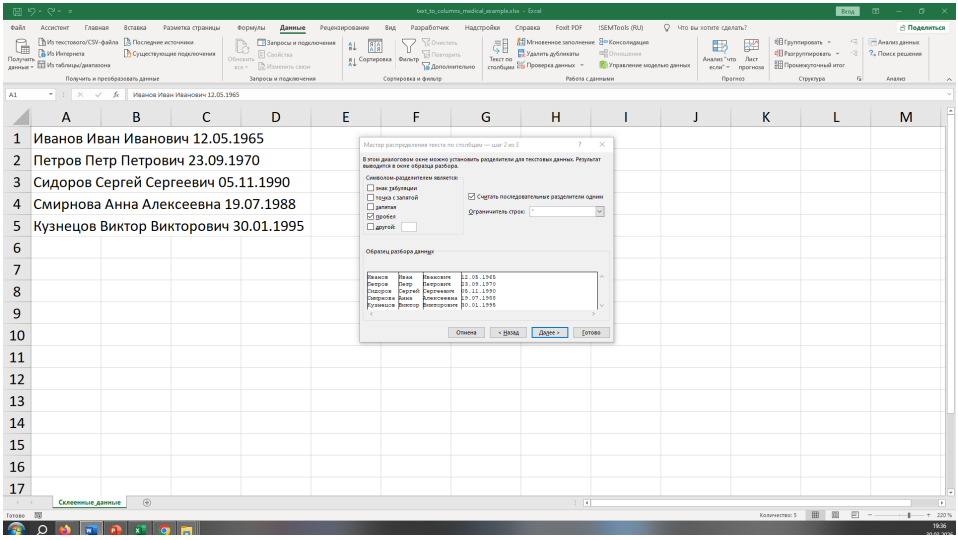


Рисунок 1.11. Меню фильтра по цвету ячейки с выбором доступных цветов

Однако здесь важно сделать оговорку. Поиск по цвету – это удобный инструмент навигации, но он не должен подменять собой чистую структуру данных. Если цветовая разметка становится основным способом выделения значимых записей, это признак того, что в таблицу стоит добавить отдельный служебный столбец с логическим признаком (например, «Летальный_исход»: Да/Нет). Такой подход надежнее, потому что цвет – это форматирование, которое может случайно сбиться, а значение в ячейке – это данные, которые всегда можно проанализировать формулами.

Практический совет: создаем панель быстрого анализа.

Главный вывод этой главы можно сформулировать так: фильтрация и сортировка превращают гору данных в структурированную информацию, готовую для осмысленного анализа. Освоив эти инструменты, вы перестанете тонуть в тысячах строк выгрузок и начнете видеть в них ответы на управленческие вопросы: кто, когда, сколько и с каким результатом. Это и есть первый шаг от простого подсчета цифр к настоящей аналитической работе врача-организатора.

Глава 1.4. Контроль качества данных: условное форматирование как страж порядка

Мы научились превращать хаотичные выгрузки в умные таблицы, удалять дубликаты, приводить текст к единому виду, разделять склеенные данные и исправлять типы. Казалось бы, данные готовы к анализу. Но есть еще один, финальный, шаг, который отделяет надежный анализ от потенциальной катастрофы. Этот шаг – визуальный контроль качества данных, или, проще говоря, умение заставить Excel самому подсвечивать подозрительные места.

Представьте себе: в таблице с 10 000 строк есть одна ячейка, где вместо числа 5 случайно введено 55. Или в столбце с датами поступления скрывается

29 февраля невисокосного года. Или в поле «Возраст» у пациента указано 150 лет. Такие ошибки не видны невооруженным глазом, но они способны исказить средние значения, процентили и выводы отчета. Именно здесь на помощь приходит условное форматирование – инструмент, который автоматически раскрашивает ячейки, нарушающие заданные правила, и позволяет увидеть аномалии за секунды.

В этой главе мы разберемся, как настроить условное форматирование для выявления самых распространенных проблем в медицинских данных: выбросов в числовых показателях, нелогичных дат, текстовых нестыковок и незаполненных критических полей.

Что такое условное форматирование и зачем оно врачу-статистику

Условное форматирование – это функция Excel, которая автоматически применяет заданное оформление (цвет заливки, цвет шрифта, значки) к ячейкам, удовлетворяющим определенному условию. Вы не раскрашиваете каждую ячейку вручную – вы задаете правило, Excel сам находит и помечает те записи, которые требуют внимания.

В работе врача-статистика условное форматирование решает три ключевые задачи:

- обнаружение ошибок ввода. Автоматическая подсветка значений, выходящих за физиологически возможные пределы (возраст 150 лет, длительность лечения 0 дней, отрицательное число койко-дней);

- выявление выбросов. Быстрое нахождение аномально высоких или низких показателей, которые могут искажать средние значения;

- контроль полноты данных. Подсветка пустых ячеек в критических полях (номер истории болезни, дата поступления, диагноз).

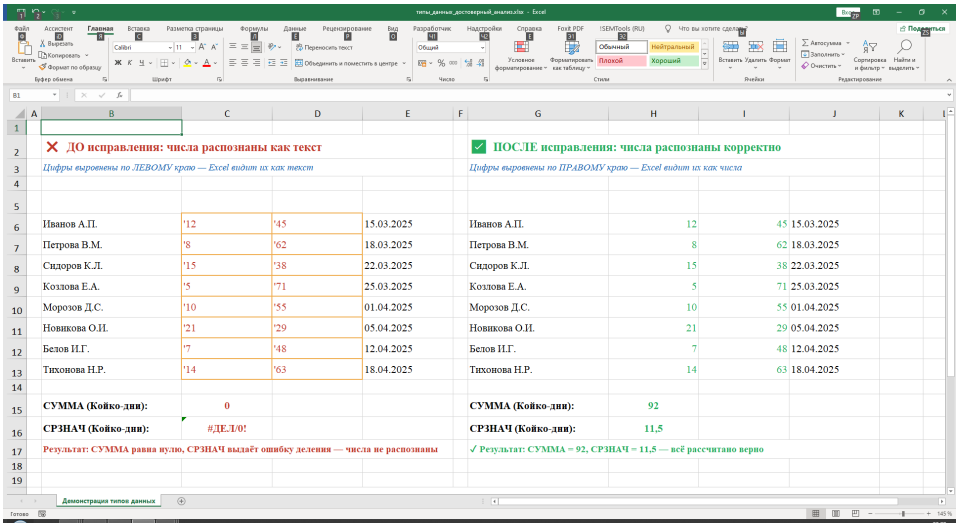
Найти и исправить одну ошибку вручную несложно. Но, когда ошибок десятки, а строк тысячи, только автоматический контроль позволяет не пропустить ни одной.

Основные типы правил условного форматирования

Excel предлагает несколько встроенных типов правил, которые покрывают большинство повседневных задач. Рассмотрим их на медицинских примерах.

1. Правила выделения ячеек

Это самый простой и интуитивно понятный тип. Вы задаете условие (больше, меньше, равно, между, содержит текст, дата) и формат выделения. Excel подсвечивает все ячейки, удовлетворяющие условию (рис. 1.12).



	B	C	D	E	G	H	I	J
2	✕ ДО исправления: числа распознаны как текст				✓ ПОСЛЕ исправления: числа распознаны корректно			
3	*Цифры выровнены по ЛЕВОМУ краю — Excel видит их как текст*				*Цифры выровнены по ПРАВОМУ краю — Excel видит их как числа*			
6	Иванов А.П.	'12	'45	15.03.2025	Иванов А.П.	12	45	15.03.2025
7	Петрова В.М.	'8	'62	18.03.2025	Петрова В.М.	8	62	18.03.2025
8	Сидоров К.Л.	'15	'38	22.03.2025	Сидоров К.Л.	15	38	22.03.2025
9	Козлова Е.А.	'5	'71	25.03.2025	Козлова Е.А.	5	71	25.03.2025
10	Морозов Д.С.	'10	'55	01.04.2025	Морозов Д.С.	10	55	01.04.2025
11	Новикова О.И.	'21	'29	05.04.2025	Новикова О.И.	21	29	05.04.2025
12	Белов И.Г.	'7	'48	12.04.2025	Белов И.Г.	7	48	12.04.2025
13	Тихонова Н.Р.	'14	'63	18.04.2025	Тихонова Н.Р.	14	63	18.04.2025
15	**СУММА (Койко-дни):**	0			**СУММА (Койко-дни):**	92		
16	**СРЗНАЧ (Койко-дни):**	#ДЕЛ/0!			**СРЗНАЧ (Койко-дни):**	11,5		
17	**Результат: СУММА равна нулю, СРЗНАЧ выдаёт ошибку деления — числа не распознаны**				✓ **Результат: СУММА = 92, СРЗНАЧ = 11,5 — всё рассчитано верно**			

Рисунок 1.12. Меню «Правила выделения ячеек» с раскрытыми вариантами

Медицинские примеры

- Выделить красным все случаи, где длительность лечения превышает 30 дней (аномально долгая госпитализация).
- Выделить желтым пациентов с возрастом более 100 лет (потенциальная ошибка ввода).
- Выделить синим все строки, где диагноз содержит слово «острый» (для быстрого поиска экстренных случаев).
- Выделить серым ячейки, где поле «Дата выписки» пусто, а пациент уже должен быть выписан.

2. Правила первых и последних значений

Эти правила выделяют определенный процент или количество самых больших или самых маленьких значений. Удобно для быстрого поиска экстремумов.

Медицинские примеры

- Выделить 10% пациентов с самой большой длительностью лечения для

разбора клинико-экспертной комиссией.

- Выделить 5% самых коротких сроков госпитализации, возможно, это пациенты, выписанные «по требованию» или переведенные.

- Выделить три максимальных значения койко-дней по каждому отделению для выявления рекордсменов.

3. Гистограммы, цветовые шкалы и наборы значков

Это визуальные способы показать распределение значений прямо в ячейках. Они не заменяют графики, но позволяют быстро оценить картину, не отрываясь от таблицы (рис. 1.13).

- **Гистограммы** – длина цветной полоски в ячейке пропорциональна значению. Позволяет мгновенно увидеть, где значения маленькие, а где большие.

- **Цветовые шкалы** – градиент от зеленого (маленькие значения) через желтый к красному (большие). Хорошо показывает распределение длительности лечения по отделениям.

- **Наборы значков** – стрелки, кружочки, флажки. Например, зеленый кружок для значений в норме, желтый – на границе, красный – за пределами.

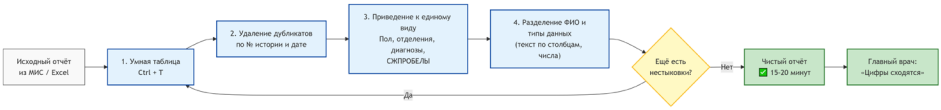


Рисунок 1.13. Таблица с цветовой шкалой в столбце «Длительность лечения»

4. Использование формул для сложных условий

Встроенных правил часто бывает недостаточно. Например, нужно выделить пациентов, у которых дата выписки раньше даты поступления. Или выделить строки, где возраст пациента и дата рождения противоречат друг другу. Для таких случаев Excel позволяет создавать правила на основе формул, и это самый мощный, хотя и требующий некоторого навыка инструмент (рис. 1.14).

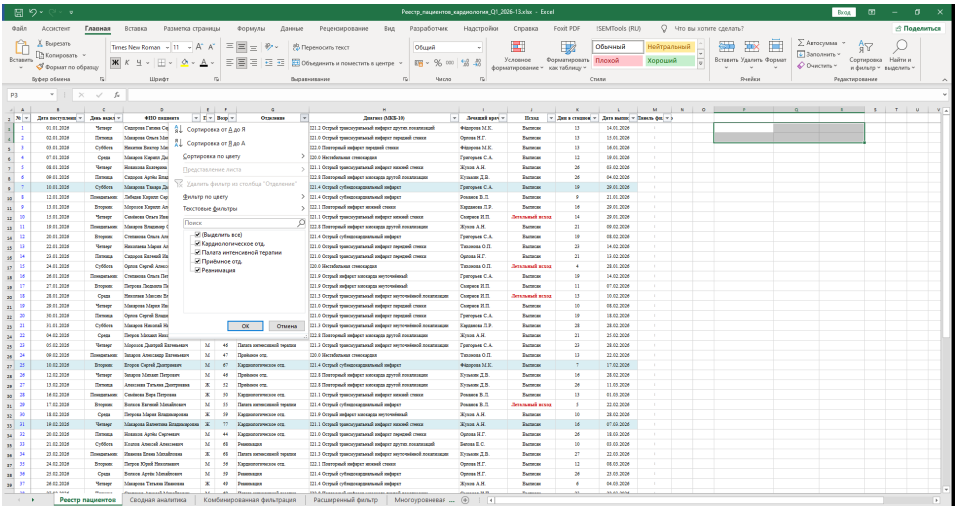


Рисунок 1.14. Окно создания правила с формулой, поле для ввода формулы

КЕЙС 1: выявление аномальных значений длительности лечения

Вернемся к нашему реестру пациентов. В столбце «Дни в стационаре» собрана информация о длительности госпитализации. Нам нужно быстро найти все случаи, где этот показатель выглядит подозрительно. Что считать аномалией? В медицине нет единого ответа: все зависит от профиля отделения. Для реанимации 30 дней это много, для отделения сестринского ухода – норма. Поэтому мы будем использовать подход, основанный на статистике: выделим значения, которые значительно отличаются от основной массы.

Шаг 1. Определяем критерии аномалий

Самый простой способ – использовать правило «среднее ± 2 стандартных отклонения», о котором мы говорили в главе 2.4. Но для условного форматирования нам нужно создать правило на основе формулы.

Вычисляем в отдельных ячейках.

- Средняя длительность: =СРЗНАЧ(Реестр пациентов[@Дни в стационаре]) → поместим в ячейку, назовем ее Средняя.

- Стандартное отклонение: =СТАНДОТКЛОН.В(Реестр пациентов[@Дни в стационаре]) → поместим в ячейку, назовем ее СтОткл.

Шаг 2. Создаем правило для выбросов

Выделяем столбец с койко-днями. Выбираем «Условное форматирование» → «Создать правило» → «Использовать формулу для определения форматируемых ячеек». Вводим формулу:

=ИЛИ(Реестр пациентов[@Дни в стационаре] > Средняя + 2*СтОткл; Реестр пациентов[@Дни в стационаре] < Средняя – 2*СтОткл).

Задаем формат, например, красную заливку и белый шрифт (рис. 1.15).

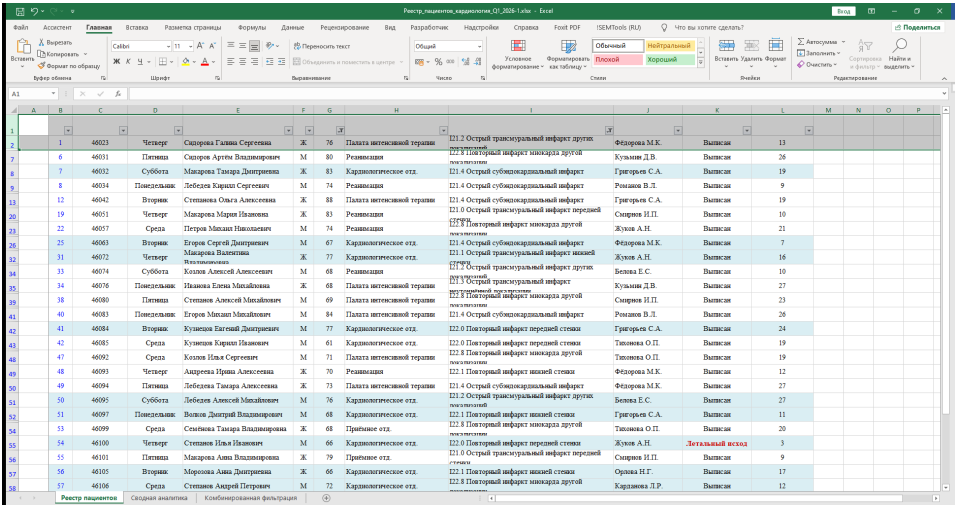


Рисунок 1.15. Окно создания правила с формулой и выбранным форматом

Шаг 3. Анализируем результат

Теперь все аномально длинные и аномально короткие случаи госпитализации подсвечены красным. Мы можем отфильтровать их по цвету и быстро проверить:

- не являются ли они ошибками ввода (например, 999 койко-дней вместо 9);
- не связаны ли они с особыми клиническими ситуациями, которые требуют внимания;
- не нарушены ли стандарты лечения.

Важное замечание: правило «среднее ± 2 сигмы» работает хорошо для симметричных распределений. Если в ваших данных есть явные выбросы, они искажают и среднее, и стандартное отклонение. В таких случаях лучше использовать правило на основе квартилей (межквартильный размах), но это уже более сложный подход, который мы можем освоить позже, когда перейдем к продвинутым инструментам.

КЕЙС 2: проверка логической согласованности дат

В медицинских данных очень часто встречаются ошибки в датах. Пациент не может быть выписан раньше, чем поступил. Дата рождения не может быть позже даты поступления. Сроки лечения не могут быть отрицательными. Условное форматирование помогает найти такие нестыковки за секунды.

Создаем несколько правил на основе формул (рис. 1.16).

Правило 1: дата выписки раньше даты поступления

= Реестр пациентов[@Дата_выписки] < Реестр пациентов[@Дата_поступления]

Формат: красная заливка.

Правило 2: дата рождения позже даты поступления

= Реестр пациентов[@Дата_рождения] > Реестр пациентов[@Дата_поступления]

Формат: красная заливка.

Правило 3: возраст пациента не соответствует дате рождения.

Здесь нужно более сложное правило, но для начала достаточно проверить, что возраст, рассчитанный по дате рождения, попадает в разумные пределы (например, от 0 до 120 лет). Добавляем вспомогательный столбец «Возраст_расчетный» с формулой =РАЗНДАТ([@Дата_рождения];[@Дата_поступления];“y”), а затем правилом подсвечиваем ячейки, где возраст меньше 0 или больше 120.

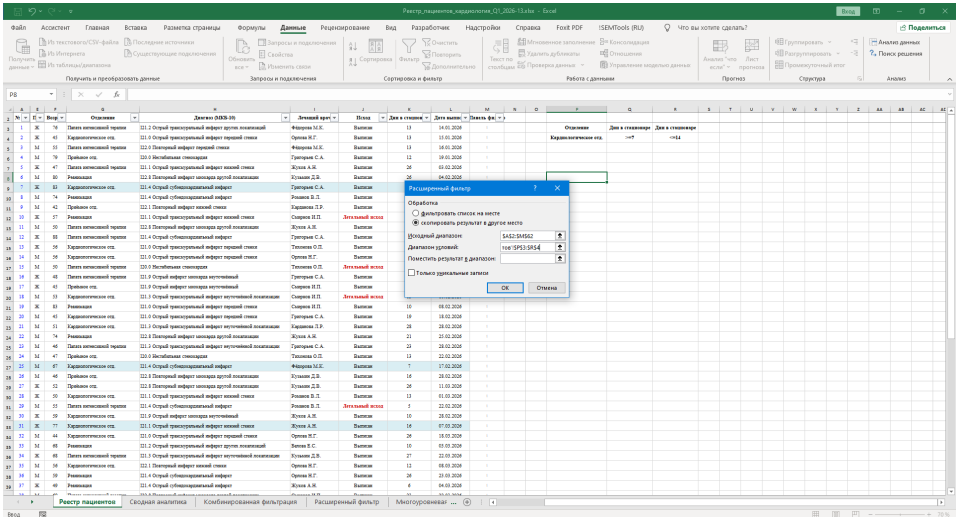


Рисунок 1.16. Таблица с подсвеченными ошибками в датах

КЕЙС 3: контроль полноты заполнения критических полей

Некоторые поля в медицинских базах данных должны быть заполнены всегда. Номер истории болезни, дата поступления, Ф.И.О. пациента, диагноз. Если хотя бы одно из этих полей пусто, запись не может считаться полноценной и должна быть исправлена.

Создаем правило, которое подсвечивает строку целиком, если в любом из критических полей есть пустота. Для этого используем формулу с функцией ИЛИ и ЕПУСТО:

=ИЛИ(ЕПУСТО(Реестр пациентов [@[Номер истории]]); ЕПУСТО(Реестр пациентов [@[ФИО]]); ЕПУСТО(Реестр пациентов [@[Дата поступления]]))

Применяем это правило ко всей таблице. Теперь строки с незаполненными критическими полями подсвечены. Их нужно либо заполнить, либо удалить, либо разобраться, почему данные отсутствуют (рис. 1.17).

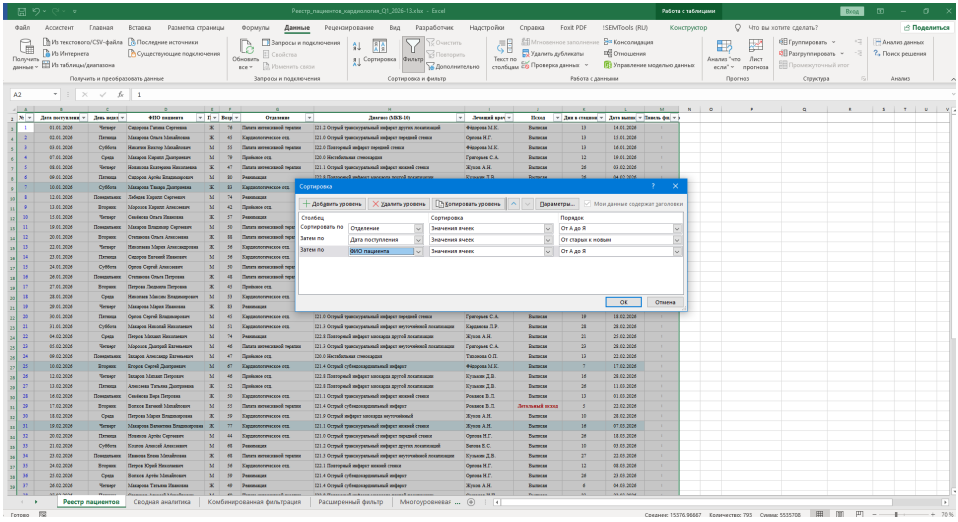


Рисунок 1.17. Таблица с подсвеченными строками, где пропущены обязательные поля

Условное форматирование – это не только инструмент разовой проверки. Его можно превратить в постоянный элемент вашего отчетного шаблона. Создайте лист с набором правил, которые проверяют:

- наличие дубликатов (с помощью формулы =СЧЕТЕСЛИ(диапазон; ячейка)>1);
- диапазоны допустимых значений для каждого числового показателя;
- логическую согласованность дат;
- заполненность критических полей.

Каждый раз, когда вы загружаете новую выгрузку, эти правила автоматически подсвечивают проблемные места. Вы тратите 2–3 мин. на просмотр подсвеченных ячеек, исправляете ошибки – и можете быть уверены, что данные готовы к анализу.

Что делать с найденными ошибками?

Условное форматирование нашло проблемные ячейки. Но что дальше? Алгоритм действий.

1. **Анализируем характер ошибки.** Это системная проблема (например, все даты в одном формате не распознаются) или единичная опечатка?

2. **Если ошибка системная,** возвращаемся к инструментам очистки из главы 1.2 («Найти и заменить», «Текст по столбцам», исправление типов данных).

3. **Если ошибка единичная,** исправляем вручную или, если ошибок много, используем фильтр по цвету, чтобы отобрать все проблемные строки и править их группами.

4. **После исправления** обновляем условное форматирование (правила остаются, но подсветка должна исчезнуть, если ошибки устранены). Если подсветка осталась, значит, проблема не решена, ищем дальше.

Главное правило: условное форматирование – это инструмент обнаружения, а не исправления. Оно показывает, где копать, но не заменяет вашего профессионального суждения о том, как правильно исправить данные.

Завершение части 1: контрольный список перед анализом.

Поздравляю! Мы завершили первую часть книги, посвященную гигиене данных. Теперь, прежде чем приступать к любому анализу (будь то расчет средних, построение сводных таблиц или визуализация), у вас есть четкий алгоритм подготовки данных.

- **Умная таблица** – диапазон превращен в умную таблицу (Ctrl + T).

- **Удаление дубликатов** – повторяющиеся записи удалены или помечены.

- **Единообразие текста** – пол, диагнозы, отделения приведены к единому виду, лишние пробелы удалены (СЖПРОБЕЛЫ).

- **Разделение данных** – Ф.И.О. и другие составные поля разделены по столбцам («Текст по столбцам»).

- **Правильные типы данных** – числа выровнены вправо, даты распознаны Excel.

- **Контроль качества** – условное форматирование не показывает критических ошибок (аномальных значений, противоречивых дат, пустых обязательных полей).

Только после того как все пункты этого списка выполнены, вы можете быть уверены: ваши цифры не будут врать. Данные готовы, и можно переходить к главному – анализу и интерпретации.

Главный вывод этой главы и всей первой части книги можно сформулировать так: качество анализа определяется качеством данных. Условное форматирование – это ваш верный страж, который не даст пропустить ошибки входа. Научившись видеть аномалии до того, как они исказят отчет, вы сэкономите часы работы и сохраните репутацию специалиста, чьим цифрам можно доверять.

Добро пожаловать в часть 2, где мы, наконец, начнем считать, анализировать и принимать обоснованные управленческие решения.



Часть 2. Математика здравого смысла: описательная статистика

Глава 2.1. Считаем итоги: от «сколько всего» к структуре

Прежде чем переходить к сложным методам анализа, давайте честно признаемся: бóльшая часть вопросов, которые ежедневно задают врачу-статистику и врачу-организатору, требует не магистральных регрессий и не многофакторного анализа, а умения правильно посчитать самые простые вещи.

Начнем с самого простого и одновременно самого важного вопроса: сколько? Сколько пациентов пролечено в отделении за месяц? Сколько койко-дней проведено в стационаре? Сколько вызовов скорой помощи обслужено? Эти цифры – основа любого отчета, точка отсчета для всех последующих расчетов.

И здесь, казалось бы, все очевидно: выделил столбец, посмотрел в строке состояния – и готово. Но медицинская статистика редко бывает такой прямолинейной. Что считать пролеченным больным: каждый случай госпитализации или каждого уникального пациента? Как быть с выбывшими и переведенными? Как посчитать общее число, если в таблице есть дубликаты, которые мы не заметили? Ответы на эти вопросы требуют не только технических навыков, но и понимания того, что именно мы считаем и зачем.

- Автосумма: первая линия обороны от ошибок

Самый быстрый способ получить итог по столбцу с числами – кнопка «Автосумма» на вкладке «Главная» или сочетание клавиш Alt + =. Выделяете ячейку под столбцом с данными, нажимаете Автосумму, и Excel автоматически предлагает формулу =СУММ(диапазон). Если данные находятся в умной таблице, формула будет выглядеть как =СУММ(Реестр пациентов[Дни в стационаре]) – это читаемо, понятно и автоматически расширяется при добавлении новых строк.

Но Автосумма умеет больше, чем просто суммировать. Если нажать на стрелочку рядом с кнопкой, открывается меню с другими функциями: среднее зна-

чение, количество чисел, максимум, минимум. Это удобно, когда нужно быстро получить базовые характеристики без написания формул вручную.

Однако здесь кроется первый подводный камень. Автосумма по умолчанию считает только ячейки с числами. Если в столбце есть текстовые значения или пустые ячейки, они просто игнорируются. В медицинских данных это может сыграть злую шутку. Например, если в столбце «Койко-дни» вместо числа стоит прочерк или буква О (вместо нуля), этот пациент не попадет в подсчет. Итоговая сумма окажется заниженной, а средняя длительность лечения – искаженной. Поэтому перед любым подсчетом итогов мы еще раз проверяем типы данных, так, как учили в главе 1.2.

- Функции СУММ, СЧЕТ и СЧЕТЗ: выбираем правильный инструмент

Для более контролируемого подсчета итогов лучше использовать функции напрямую. В Excel есть три ключевые функции, которые на первый взгляд похожи, но решают разные задачи.

СУММ складывает только числовые значения. Если в диапазоне есть текст или логические значения (ИСТИНА/ЛОЖЬ), они игнорируются. Это главный инструмент для подсчета койко-дней, суммы расходов, объема медицинской помощи в числовом выражении.

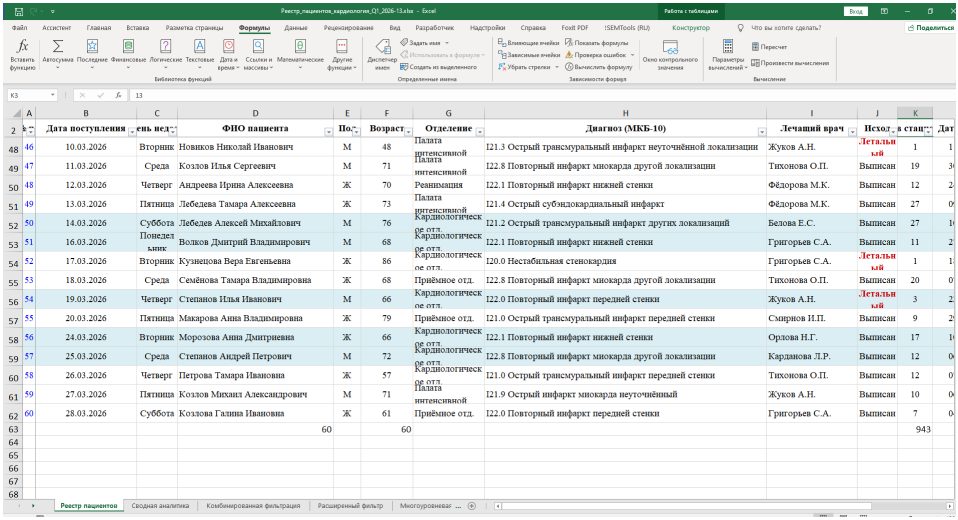
Пример: =СУММ(Таблица1[Койко-дни]) – общее число койко-дней, проведенных всеми пациентами за период.

СЧЕТ подсчитывает количество ячеек, содержащих числа. Текстовые ячейки, пустые ячейки и логические значения не учитываются. Эта функция незаменима, когда нужно посчитать, сколько пациентов имеют заполненный числовой показатель. Например, сколько пациентов получили определенный лабораторный тест (если результат записан числом) или сколько дней в отчетном периоде были непустыми.

Пример: =СЧЕТ(Таблица1[Возраст]) – количество пациентов, у которых в столбце «Возраст» указано число. Если возраст не указан или указан текстом, пациент не учитывается.

СЧЕТЗ подсчитывает количество любых непустых ячеек. Текст, числа, даты, ошибки – все, что не является пустотой, попадает в подсчет. Это главная функция для подсчета количества записей в таблице независимо от того, какие данные в них содержатся.

Пример: =СЧЕТЗ(Таблица1[Ф.И.О. пациента]) – общее количество пациентов в таблице (или, точнее, количество строк, где заполнена фамилия) (рис. 2.1).



№	Дата поступления	День нед.	ФИО пациента	Пол	Возраст	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стац.	Дата
46	10.03.2026	Вторник	Новиков Николай Иванович	М	48	Палата интенсивной	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный	1	1
47	11.03.2026	Среда	Козлов Илья Сергеевич	М	71	Палата интенсивной	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	3
48	12.03.2026	Четверг	Андреева Ирина Алексеевна	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	2
49	13.03.2026	Пятница	Лебедева Тамара Алексеевна	Ж	73	Палата интенсивной	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27	0
50	14.03.2026	Суббота	Лебедев Алексей Михайлович	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	1
51	16.03.2026	Понедельник	Волков Дмитрий Владимирович	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	2
52	17.03.2026	Вторник	Кузнецова Вера Евгеньевна	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный	1	1
53	18.03.2026	Среда	Семёнова Тамара Владимировна	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	0
54	19.03.2026	Четверг	Степанов Илья Иванович	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный	3	2
55	20.03.2026	Пятница	Макарова Анна Владимировна	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	2
56	24.03.2026	Вторник	Морозова Анна Дмитриевна	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	1
57	25.03.2026	Среда	Степанов Андрей Петрович	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Кардашова Л.Р.	Выписан	12	0
58	26.03.2026	Четверг	Петрова Тамара Ивановна	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12	0
59	27.03.2026	Пятница	Козлов Михаил Александрович	М	71	Палата интенсивной	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	0
60	28.03.2026	Суббота	Козлова Галина Ивановна	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7	0
				60		60				943	

Рисунок 2.1. Таблица с тремя разными результатами для СУММ, СЧЕТ и СЧЕТЗ, наглядно показывающая разницу

Почему это важно для врача-статистика? Потому что выбор неправильной функции может привести к серьезным ошибкам в отчетности. Представьте, что в столбце «Номер истории болезни» некоторые ячейки пусты (технический сбой), а некоторые содержат текст «Не введен». Функция СЧЕТ не увидит ни пустых ячеек, ни текста и покажет заниженное число. Функция СЧЕТЗ покажет все непустые ячейки, включая «Не введен», что тоже неверно. В такой ситуации правильным подходом будет сначала очистить данные, заменив «Не введен» на пустоту, а затем использовать СЧЕТЗ или (еще лучше) добавить столбец-флаг, где четко обозначено, является ли запись полноценным случаем лечения.

Кейс: считаем общее число пролеченных больных за месяц

Давайте рассмотрим реальную задачу, которая встает перед врачом-статистиком каждый месяц. Из медицинской информационной системы выгружен реестр пациентов стационара за январь. В таблице есть следующие столбцы:

- Номер истории болезни;
- Ф.И.О. пациента;

- Дата поступления;
- Дата выписки;
- Отделение;
- Исход (выписан, переведен, умер).

Задача: посчитать общее число пролеченных больных за месяц. Казалось бы, что может быть проще? Но на практике возникает несколько вопросов.

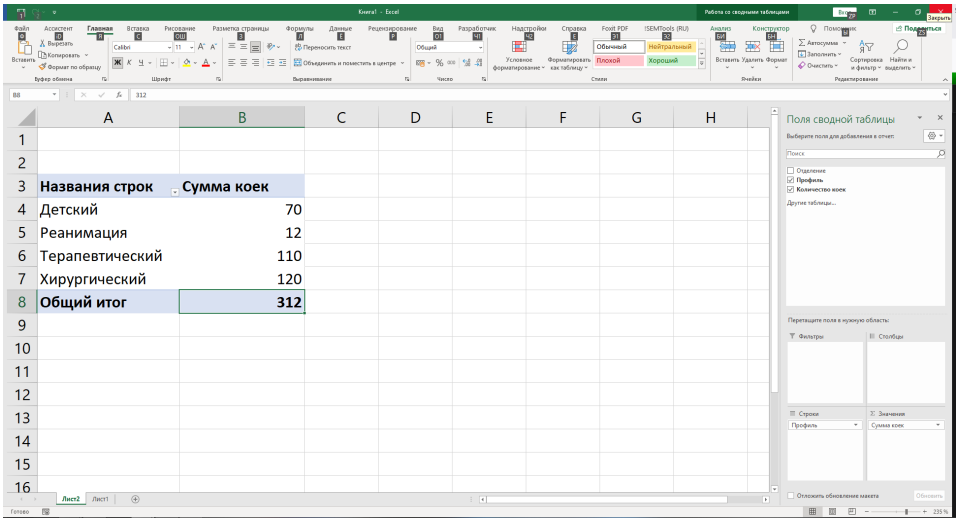
Первый вопрос: что считать пролеченным больным? Каждый случай госпитализации или каждого уникального пациента? Если пациент был госпитализирован дважды за месяц (например, поступил, выписался, через неделю поступил снова), он должен считаться как один или как два пролеченных?

В соответствии с правилами медицинской статистики, пролеченный больной – это каждый законченный случай госпитализации. То есть если пациент поступил дважды, это два пролеченных. Но, чтобы ответить на этот вопрос, нужно не только техническое умение, но и понимание правил учета, принятых в вашем учреждении.

Второй вопрос: как быть с переведенными? Если пациент поступил в кардиологию, затем переведен в реанимацию, а затем выписан, сколько у нас пролеченных? Один – это один законченный случай. Но в реестре могут быть две записи: перевод из отделения в отделение часто регистрируется как отдельное событие. Значит, прежде чем считать, нужно либо очистить таблицу от промежуточных переводов, либо четко понимать, какую именно строку считать основной.

Третий вопрос: есть ли в таблице дубликаты? Технические сбои могли создать повторные записи о поступлении одного и того же пациента в один день.

Четвертый вопрос: все ли записи действительно относятся к отчетному месяцу? Иногда в выгрузку попадают пациенты, поступившие в предыдущем месяце, но выписанные в текущем. В зависимости от требований отчета они могут учитываться либо в месяце поступления, либо в месяце выписки, либо не учитываться вовсе (рис. 2.2).



Названия строк	Сумма коек
Детский	70
Реанимация	12
Терапевтический	110
Хирургический	120
Общий итог	**312**

Рисунок 2.2. Фрагмент реестра пациентов с выделенными сложными случаями

Теперь посмотрим, как эти вопросы решаются с помощью инструментов, которые мы уже освоили, и функций подсчета.

Шаг 1. Подготовка данных. Превращаем диапазон в умную таблицу

(Ctrl + T). Удаляем дубликаты по номеру истории и дате поступления, если они действительно дубликаты. Если сомневаемся, создаем вспомогательный столбец для проверки.

Шаг 2. Определяем период. Добавляем вспомогательный столбец «Месяц_поступления» с формулой = МЕСЯЦ([@Дата поступления]). Или, если отчет считает по выписке, используем дату выписки.

Шаг 3. Фильтруем нужный месяц. С помощью фильтра оставляем только записи за январь.

Шаг 4. Решаем проблему переводов. Если в учреждении принято считать по законченным случаям, нам нужно оставить только одну строку на каждый случай. Обычно это строка с датой выписки или последняя запись о пациенте в периоде. Можно отсортировать таблицу по дате выписки (убывание) и затем с помощью функции СЧЕТЕСЛИ пометить уникальные номера историй, оставляя только последнюю запись.

Шаг 5. Считаем. Теперь, когда таблица очищена и содержит только те случаи, которые должны войти в отчет, считаем количество записей функцией СЧЕТЗ по столбцу с номером истории болезни (или любому другому, где гарантированно нет пустот). Формула будет выглядеть примерно так:

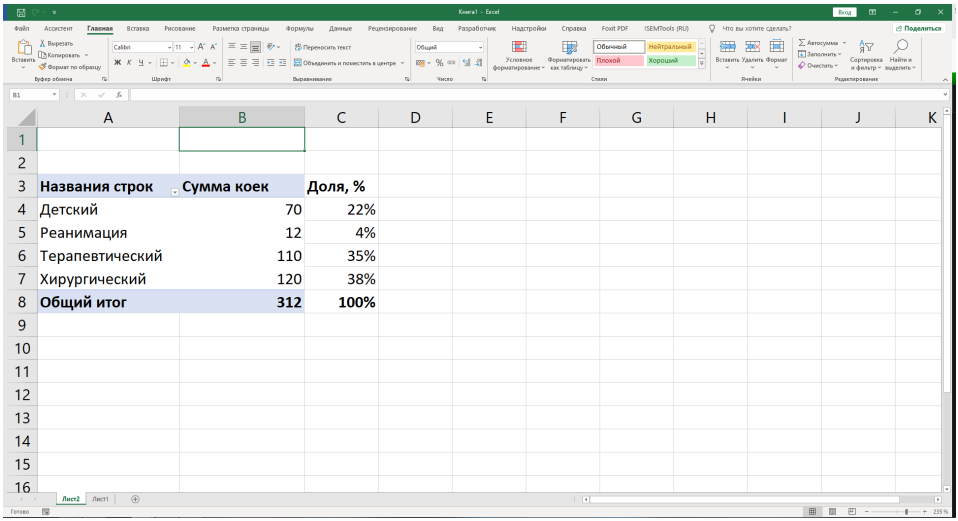
=СЧЕТЗ(Реестр пациентов[Номер истории болезни]).

Если же мы хотим посчитать общее количество дней лечения (койко-дни) за месяц, используем СУММ по столбцу «Койко-дни»:

=СУММ(Реестр пациентов[Дни в стационаре]).

- Практический совет: создаем шаблон ежемесячного отчета
- Чтобы не повторять одни и те же действия каждый месяц, мы рекомендуем создать шаблон с уже настроенными формулами. На отдельном листе размещаем ключевые показатели:
- общее число пролеченных (по законченным случаям);
- общее количество койко-дней;
- средняя длительность лечения (об этом поговорим в следующей главе);
- количество пациентов по отделениям;
- количество исходов (выписан, переведен, умер).

Все эти показатели рассчитываются формулами, ссылающимися на умную таблицу с данными. Каждый месяц вы просто заменяете старую выгрузку новой, обновляете фильтр по месяцу, и все итоги пересчитываются автоматически. Это превращает ежемесячную рутину в пятиминутную задачу (рис. 2.3).



	A	B	C
1			
2			
3	**Названия строк**	**Сумма коек**	**Доля, %**
4	Детский	70	22%
5	Реанимация	12	4%
6	Терапевтический	110	35%
7	Хирургический	120	38%
8	**Общий итог**	**312**	**100%**
9			
10			
11			
12			
13			
14			
15			
16			

Рисунок 2.3. Пример листа-шаблона с отчетом и формулами, ссылающимися на умную таблицу

Главный вывод этой главы можно сформулировать так: правильный подсчет итогов начинается не с формулы, а с понимания того, что именно мы считаем. Количество записей, количество чисел, количество заполненных ячеек – это три разных показателя, и каждый имеет свое место в медицинской статистике. Освоив эти базовые функции, вы закладываете фундамент для всех последующих расчетов: средних величин, долей, показателей разброса. И, как всегда, в работе с данными главное правило остается неизменным: прежде чем считать, убедитесь, что данные чистые, а вы точно понимаете, какой вопрос задаете.

Глава 2.2. Средние величины: когда среднее арифметическое обманывает

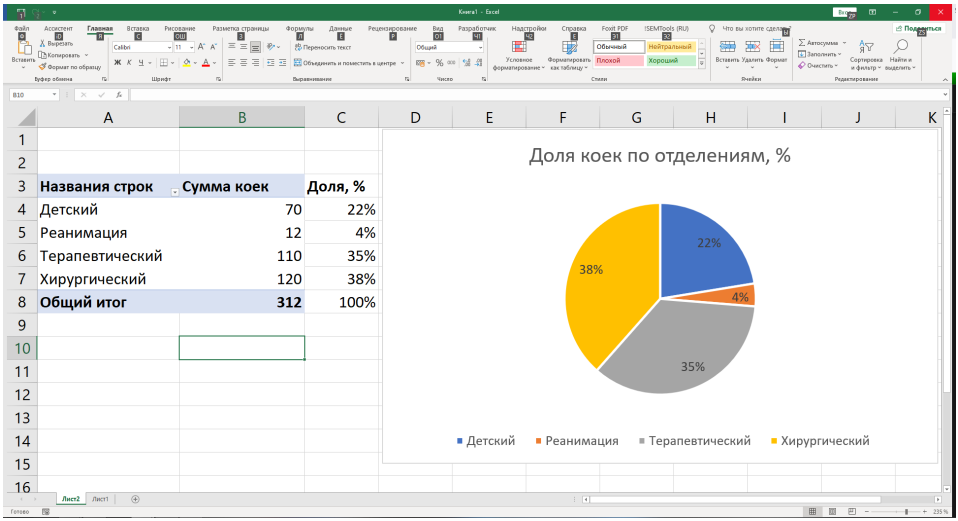
Представьте себе ситуацию. В отделении работают 10 врачей. У девяти из них зарплата составляет 50 000 руб., а у заведующего отделением – 200 000 руб. Вычисляем среднее арифметическое: (9 × 50 000 + 200 000) / 10 = 65 000 руб. Цифра получилась ровная, красивая, но... отражает ли она реальность? Для девяти рядовых врачей эта средняя зарплата выглядит завышенной, а для заведующего – заниженной. По сути, она не описывает типичную зарплату ни для одной группы.

Это классический пример того, как среднее арифметическое может вводить в заблуждение. В медицинской статистике таких ситуаций не меньше, а цена ошибки гораздо выше. Средняя длительность лечения в отделении может быть

искажена одним пациентом, который провел в стационаре полгода. Средний возраст пациентов с редким заболеванием может быть смещен единичным пожилым пациентом. Среднее время ожидания скорой помощи может не отражать реальную ситуацию из-за одного вызова, где бригада ехала три часа в пробке.

В этой главе мы разберемся, когда можно доверять среднему арифметическому, а когда нужно обращаться к другим инструментам – медиане и моде. Эти показатели не отменяют друг друга, а дополняют, позволяя увидеть данные с разных сторон и избежать ловушки средней температуры по больнице.

- Среднее арифметическое (СРЗНАЧ): сила и слабость
- Среднее арифметическое – это, пожалуй, самый известный статистический показатель. В Excel он вычисляется функцией СРЗНАЧ или кнопкой
- Автосумма → «Среднее». Формула суммирует все числовые значения в диапазоне и делит на их количество. Все просто, понятно и в большинстве случаев достаточно.
- В работе врача-статистика среднее арифметическое применяется повсеместно (рис. 2.4).
- Средняя длительность лечения – общее число койко-дней, деленное на количество пролеченных больных.
- Средний возраст пациента – сумма возрастов, деленная на количество пациентов.
- Средняя нагрузка на врача – общее количество посещений, деленное на число врачей.



Названия строк	Сумма коек	Доля, %
Детский	70	22%
Реанимация	12	4%
Терапевтический	110	35%
Хирургический	120	38%
Общий итог	**312**	100%

Рисунок 2.4. Формула СРЗНАЧ, диапазон с длительностью лечения

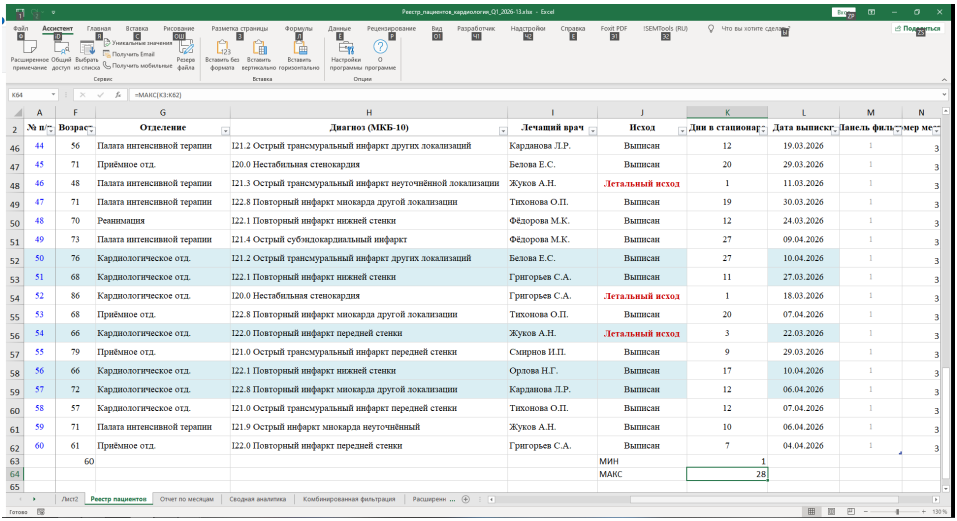
Однако у среднего арифметического есть серьезное ограничение: оно чрезвычайно чувствительно к выбросам – экстремальным значениям, которые сильно отличаются от основной массы данных. Один аномально длинный случай госпитализации может поднять среднюю длительность лечения по отделению на несколько дней, создавая ложное впечатление, что все пациенты лежат дольше нормы. Один пациент с редким заболеванием, доживший до 100 лет, может сдвинуть средний возраст в группе, где основная масса – люди 40–50 лет.

Вот почему в медицинской статистике среднее арифметическое никогда не используется изолированно. Оно обязательно сопровождается показателями разброса (о них в следующей главе) и в случаях с потенциальными выбросами дополняется медианой.

- Медиана (МЕДИАНА): защита от выбросов

Медиана – это золотая середина в самом прямом смысле слова. Если упорядочить все значения по возрастанию, медиана будет тем значением, которое окажется ровно посередине. Половина наблюдений находится ниже медианы, половина – выше. В отличие от среднего арифметического, медиана практически не чувствительна к выбросам.

В Excel медиана вычисляется функцией МЕДИАНА. Синтаксис прост: =МЕДИАНА(диапазон). Функция игнорирует текстовые и пустые ячейки, работая только с числами (рис. 2.5).



№ п/п	Возраст	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стационаре	Дата выписки	Панель фильтров	Номер ме
44	56	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Кардашова Л.Р.	Выписан	12	19.03.2026	1	3
45	71	Приёмное отд.	I20.0 Нестабильная стенокардия	Белова Е.С.	Выписан	20	29.03.2026	1	3
46	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1	11.03.2026	1	3
47	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	30.03.2026	1	3
48	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	24.03.2026	1	3
49	73	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27	09.04.2026	1	3
50	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	10.04.2026	1	3
51	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	27.03.2026	1	3
52	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1	18.03.2026	1	3
53	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	07.04.2026	1	3
54	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3	22.03.2026	1	3
55	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	29.03.2026	1	3
56	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	10.04.2026	1	3
57	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Кардашова Л.Р.	Выписан	12	06.04.2026	1	3
58	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12	07.04.2026	1	3
59	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	06.04.2026	1	3
60	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7	04.04.2026	1	3
	60				МИН	1			
					МАКС	28			

Рисунок 2.5. Формула МЕДИАНА рядом с формулой СРЗНАЧ

Кейс: типичная зарплата врача в отделении

Вернемся к примеру с зарплатой, который мы привели в начале главы. У нас есть 10 наблюдений: 50 000 (9 раз) и 200 000 (1 раз). Вычисляем в Excel:

- =СРЗНАЧ(диапазон) → 65 000 руб.;
- =МЕДИАНА(диапазон) → 50 000 руб.

Какая цифра лучше описывает типичную зарплату врача в этом отделении? Очевидно, медиана. Она показывает: половина врачей получает 50 000 руб. или меньше, половина – 50 000 руб. или больше. Единичный высокий оклад заведующего на нее практически не повлиял.

В медицинской практике аналогичных ситуаций множество. Представьте, что вы анализируете длительность лечения в отделении реанимации. Десять пациентов провели в отделении 2–3 дня, а один пациент с тяжелой сочетанной травмой – 45 дней. Среднее арифметическое покажет около 6–7 дней, что не отражает реальность ни для кого. Медиана покажет 2–3 дня – именно столько лежало пациентов (большинство). И это гораздо более полезная информация для планирования работы отделения.

Когда использовать медиану в медицинской статистике

Медиана становится предпочтительным выбором в следующих ситуациях.

- Данные имеют асимметричное распределение (например, длительность лечения: большинство пациентов выписываются быстро, но есть небольшое число долгожителей).

- Есть явные выбросы, которые не являются ошибкой ввода, но отражают реальные редкие случаи.

- Показатель не имеет нормального распределения (о том, как это проверить, поговорим в части 6).

- Вы хотите описать типичного пациента, а не среднего по больнице.

В отчетах для главного врача часто полезно приводить оба показателя рядом: среднюю длительность лечения (чувствительную к изменениям в тяжелых случаях) и медиану (характеризующую основную массу пациентов). Их сравнение дает ценную информацию о структуре пациентов: если медиана заметно ниже среднего, значит, в отделении есть группа пациентов с аномально долгим лечением, требующая внимания.

- Мода (МОДА.НСК и МОДА.ОДН): самое частое значение

Мода – это значение, которое встречается в наборе данных чаще всего.

В отличие от среднего и медианы, которые всегда существуют для любого набора чисел, мода может отсутствовать (если все значения уникальны) или быть множественной (если несколько значений встречаются одинаково часто).

В Excel для работы с модой есть две основные функции (рис. 2.6).

- МОДА.ОДН возвращает одно значение, которое встречается чаще всего. Если таких значений несколько, возвращается первое из них.

- МОДА.НСК возвращает массив всех значений, встречающихся с максимальной частотой. В более новых версиях Excel эта функция может выводить несколько значений автоматически.

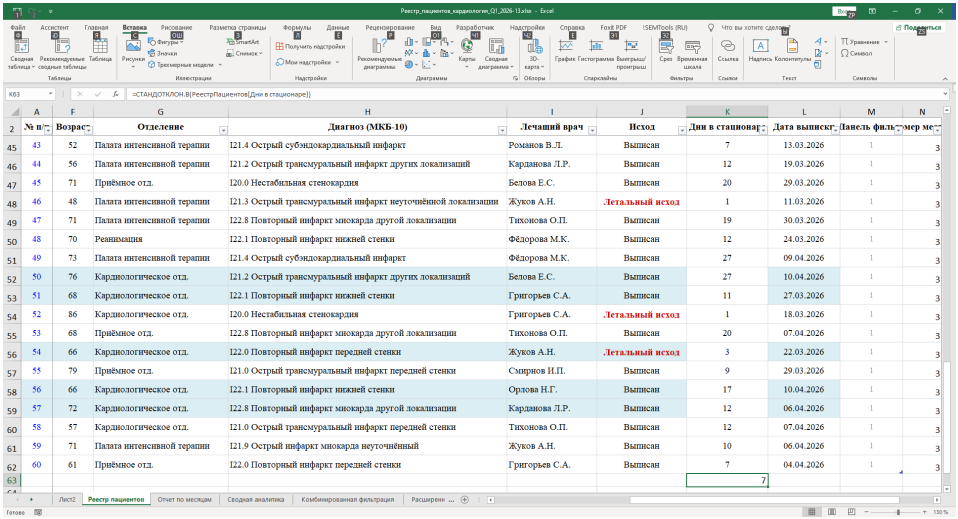


Рисунок 2.6. Формула МОДА.ОДН под столбцом «Дни в стационаре», результат — самое часто встречающееся значение

Кейс: самый частый диагноз в приемном покое

Представьте, что мы анализируем 500 обращений в приемный покой за месяц. В столбце «Диагноз при поступлении» содержатся записи «Гипертоническая болезнь», «Острый инфаркт миокарда», «Бронхит», «Гипертоническая болезнь», «Острый инфаркт миокарда» и т.д. Нам нужно быстро понять, какое состояние встречается чаще всего, чтобы оценить нагрузку на профильные отделения.

Функция =МОДА.ОДН в данном случае не сработает, потому что она работает только с числами. Для текстовых данных нам нужен другой подход. Самый простой способ найти моду для текста – использовать вспомогательную сводную таблицу или комбинацию функций НАИБОЛЬШИЙ и СЧЕТЕСЛИ. Но есть и более элегантное решение: создать сводную таблицу с диагнозами в строках и количеством в значениях, а затем отсортировать по убыванию. Самый частый диагноз окажется на первой строке (рис. 2.7).

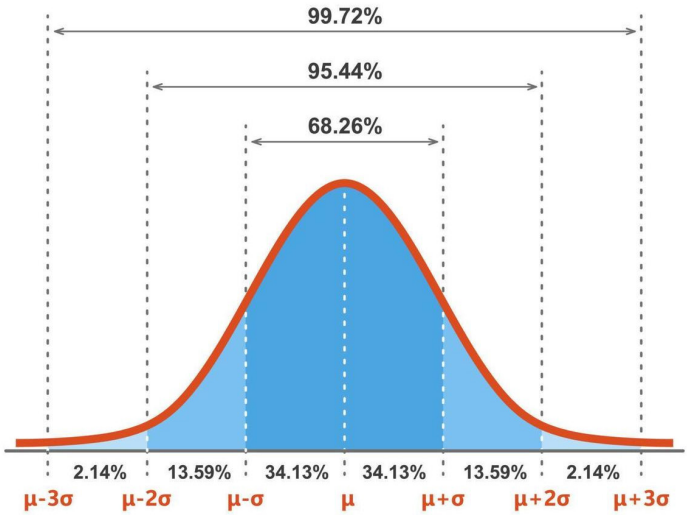


Рисунок 2.7. Сводная таблица с диагнозами, отсортированная по убыванию количества

В этом кейсе мода дает руководству важную информацию: если чаще всего в приемном покое регистрируются сердечно-сосудистые заболевания, значит, нужно усиливать кардиологическую службу, увеличивать количество кардиологических коек или пересматривать маршрутизацию пациентов.

Когда мода полезна в медицинской статистике

Мода особенно ценна при работе с категориальными данными, т.е. с текстовыми значениями, которые можно сгруппировать:

- наиболее частый диагноз в стационаре или поликлинике;
- наиболее распространенный исход лечения (выписан, переведен, умер);
- самая частая возрастная группа пациентов;
- наиболее популярный день недели для госпитализаций (помогает планировать работу приемного отделения);
- самый частый врач, выписывающий направление на госпитализацию.

Для числовых данных мода применяется реже, но тоже может быть полезна. Например, если мы анализируем возраст пациентов, мода покажет наиболее

частый возраст в выборке. Это может быть важно для планирования профилактических программ.

Практический совет: три кита описательной статистики.

В профессиональной работе врача-статистика среднее, медиана и мода не противопоставляются, а дополняют друг друга. В хорошем отчете вы часто встретите все три показателя рядом (табл. 2.1).

Таблица 2.1. Средняя, медиана, мода

Показатель	Значение, дни	Интерпретация
Среднее	7,2	Чувствительно к тяжелым случаям
Медиана	4,0	Половина пациентов лежит 4 дня или меньше
Мода	3	Чаще всего пациенты выписываются на 3-й день

Сравнение этих трех цифр дает быстрый диагноз структуре данных. Если среднее заметно больше медианы, значит, есть группа пациентов с аномально долгим лечением. Если среднее и медиана близки, а мода рядом, распределение симметричное, среднее можно использовать уверенно. Если мода сильно отличается от медианы, возможно, данные имеют несколько пиков (например, две основные группы пациентов с разной длительностью лечения).

Главный вывод этой главы можно сформулировать так: выбор средней величины – это не технический, а содержательный вопрос. Прежде чем нажать кнопку «СРЗНАЧ», задайте себе вопрос: что я хочу узнать? Типичное значение для большинства? Общую сумму, разделенную на количество? Самое частое явление? Ответ на этот вопрос определит, какой инструмент выбрать. И помните: в медицинской статистике нет единственно правильной средней – есть правильная средняя для правильного вопроса.

Глава 2.3. Структура и доли: искусство отвечать на вопрос «Какую часть?»

Средние величины показывают типичные значения, но они не дают ответа на другой важный вопрос: как распределяется целое по своим составным частям? Какая доля пациентов проходит лечение в терапевтическом отделении, а какая – в хирургическом? Какой процент коечного фонда приходится на реанимацию? Насколько изменилась структура госпитализаций по сравнению с прошлым годом? Эти вопросы требуют умения работать с долями и процентами, теми самыми относительными величинами, которые превращают абстрактные цифры в понятные управленческие решения.

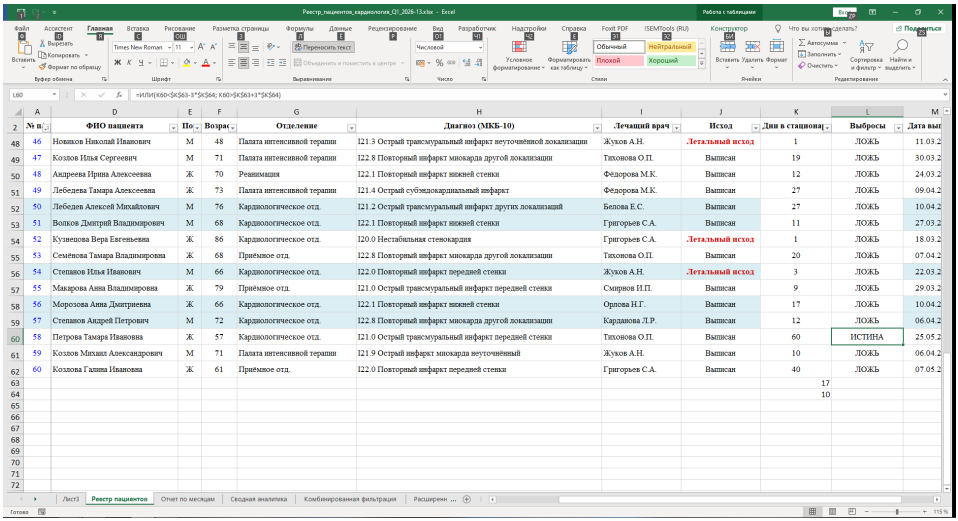
Врачи-статистики часто шутят: «Абсолютные цифры для отчета в Минздрав, а проценты для разговора с главным врачом». В этой шутке большая доля правды. Руководителю важно не только знать, что в отделении пролечены 500 пациентов, но и понимать, какую долю от общей нагрузки учреждения это составляет, как изменилась эта доля по сравнению с прошлым месяцем и на какие профили приходится основной объем работы.

В этой главе мы разберемся, как правильно рассчитывать доли и проценты, как отличать абсолютные величины от относительных и почему не всякий процент заслуживает доверия. И, конечно, рассмотрим все это на реальных медицинских кейсах.

Простое правило: часть делим на целое.

Формула расчета доли в Excel выглядит настолько просто, что кажется, будто здесь нечему учиться: = значение / общая_сумма. Но за этой простотой скрывается множество подводных камней. Главный из них: правильно определить, что в каждом конкретном случае является частью, а что – целым.

Представьте, что вы анализируете структуру пациентов по отделениям. Целое – это общее количество пролеченных больных за период. Часть – количество пациентов в каждом отделении. Формула, например, для кардиологического отделения: = количество_пациентов_в_кардиологии / общее_количество_пациентов. Умножаем на 100%, получаем долю в процентах (рис. 2.8).



№ п/п	ФИО пациента	Пол	Возраст	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стационаре	Выбросы	Дата выписки
46	Новиков Николай Иванович	М	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1	ЛОЖЬ	11.03.2
47	Козлов Илья Сергеевич	М	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	ЛОЖЬ	30.03.2
48	Андреева Ирина Алексеевна	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	ЛОЖЬ	24.03.2
49	Лебедева Тамара Алексеевна	Ж	73	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27	ЛОЖЬ	09.04.2
50	Лебедев Алексей Михайлович	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	ЛОЖЬ	10.04.2
51	Волков Дмитрий Владимирович	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	ЛОЖЬ	27.03.2
52	Кузнецова Вера Евгеньевна	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1	ЛОЖЬ	18.03.2
53	Семёнова Тамара Владимировна	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	ЛОЖЬ	07.04.2
54	Степанов Илья Иванович	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3	ЛОЖЬ	22.03.2
55	Макарова Анна Владимировна	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	ЛОЖЬ	29.03.2
56	Морозова Анна Дмитриевна	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	ЛОЖЬ	10.04.2
57	Степанов Андрей Петрович	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Кардашова Л.Р.	Выписан	12	ЛОЖЬ	06.04.2
58	Петрова Тамара Ивановна	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	60	ИСТИНА	25.05.2
59	Козлов Михаил Александрович	М	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	ЛОЖЬ	06.04.2
60	Козлова Галина Ивановна	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	40	ЛОЖЬ	07.05.2

Рисунок 2.8. Таблица с абсолютными числами пациентов по отделениям и рассчитанными долями в процентах

Однако уже на этом простом шаге могут возникнуть искажения. Что если пациент проходил лечение последовательно в двух отделениях? В зависимости от правил учета, он может быть учтен в обоих отделениях (тогда сумма долей превысит 100%) или только в одном (тогда нужно решить, в каком именно). Что если некоторые пациенты не были распределены по отделениям из-за ошибки ввода? Тогда знаменатель (целое) и сумма частей могут не сойтись. Именно поэтому перед любым расчетом структуры мы проверяем, что данные приведены к единому принципу учета.

В Excel проще всего рассчитывать доли с помощью формул, ссылающихся на итоговую ячейку. Чтобы при копировании формулы вниз ссылка на итог оставалась неизменной, используется абсолютная ссылка (со знаком $). Например, если итог находится в ячейке B8, формула для первой строки будет выглядеть так: =B4/$B$8. Затем ее можно протянуть вниз на все строки (рис. 2.9).

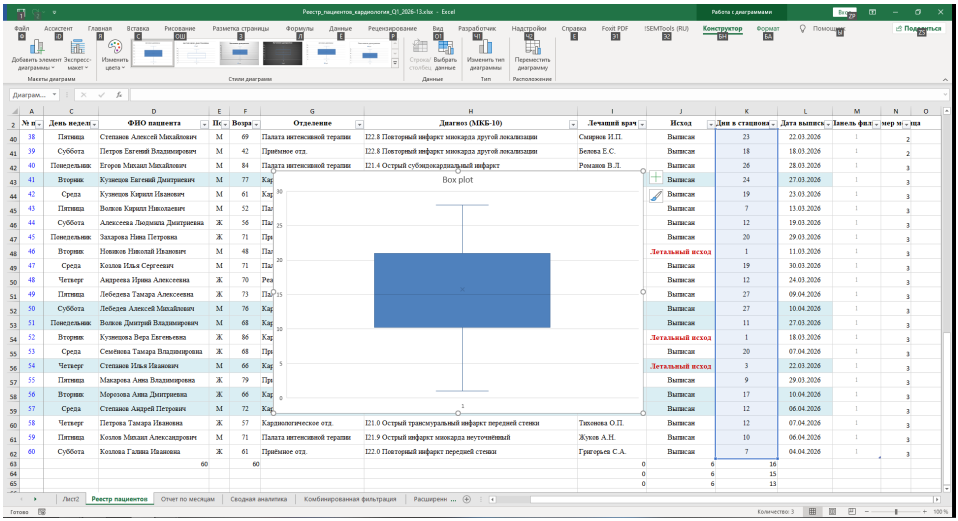


Рисунок 2.9. Строка формул

Абсолютные и относительные величины: два взгляда на одни и те же данные.

В медицинской статистике абсолютные и относительные величины не противопоставляются, а работают в паре. Каждая из них дает информацию, которую невозможно получить из другой.

Абсолютные величины – это просто количество: 500 пациентов, 1200 койко-дней, 15 летальных исходов. Они показывают масштаб явления, объем работы, нагрузку на систему. Без абсолютных цифр невозможно планировать ресурсы: сколько нужно коек, лекарств, персонала.

Относительные величины – это доли, проценты, соотношения:

30% пациентов проходят лечение в кардиологии, 5% случаев заканчиваются летально, 3,2 койко-дня в среднем на одного пациента. Они показывают структуру, интенсивность, сравнительную значимость явлений.

Опасность возникает, когда используют только один тип величин. Только абсолютные цифры могут скрыть изменение структуры: например, количество пациентов в кардиологии выросло с 200 до 220, но если общее число пациентов по стационару выросло с 1000 до 1500, то доля кардиологии на самом деле снизилась. И наоборот, только проценты могут скрыть масштаб: доля летальных исходов осталась на уровне 5%, но если общее число пациентов выросло вдвое, то доля кардиологии на самом деле снизилась.

В хорошем отчете всегда присутствуют оба взгляда. Для ключевых показателей приводятся и абсолютные значения, и относительные, и это позволяет руководству видеть полную картину.

Кейс: рассчитываем структуру коечного фонда

Давайте рассмотрим реальную задачу, которая регулярно встает перед врачом-статистиком любого стационара. Нужно проанализировать структуру коечного фонда: сколько процентов коек приходится на терапевтические профили, сколько – на хирургические, сколько – на детские, сколько – на реанимацию и т.д. Эта информация необходима для планирования развития учреждения, отчетности перед контролирующими органами и обоснования заявок на финансирование.

Исходные данные могут выглядеть по-разному. В простейшем случае у нас есть таблица с перечнем отделений и количеством коек в каждом (табл. 2.2).

Таблица 2.2. Перечень отделений и количество коек

Отделение	Количество коек
Кардиологическое	60
Неврологическое	50
Хирургическое	80
Травматологическое	40
Реанимация	12
Детское соматическое	45
Детское хирургическое	25

Задача: рассчитать долю терапевтических коек, хирургических коек и детских коек в общей структуре. Для этого нам нужно сгруппировать отделения по профилям.

Шаг 1. Определяем профильную принадлежность каждого отделения

Создаем вспомогательный столбец «Профиль», где вручную (или с помощью функции ВПР из справочника) присваиваем каждому отделению категорию (табл. 2.3):

- кардиологическое, неврологическое → «Терапевтический»;
- хирургическое, травматологическое → «Хирургический»;
- детское соматическое, детское хирургическое → «Детский»;
- реанимация → «Реанимация» (часто выделяется отдельно).

Таблица 2.3. Профиль

Отделение	Профиль	Количество коек
Кардиологическое	Терапевтический	60
Неврологическое	Терапевтический	50
Хирургическое	Хирургический	80
Травматологическое	Хирургический	40
Реанимация	Реанимация	12
Детское соматическое	Детский	45
Детское хирургическое	Детский	25

Шаг 2. Суммируем койки по профилям

Самый удобный способ – сводная таблица. Выделяем исходную таблицу (превращенную в умную), вставляем сводную таблицу, в строки помещаем поле «Профиль», в значения – поле «Количество коек» (сумма). Получаем аккуратный итог (рис. 2.10).

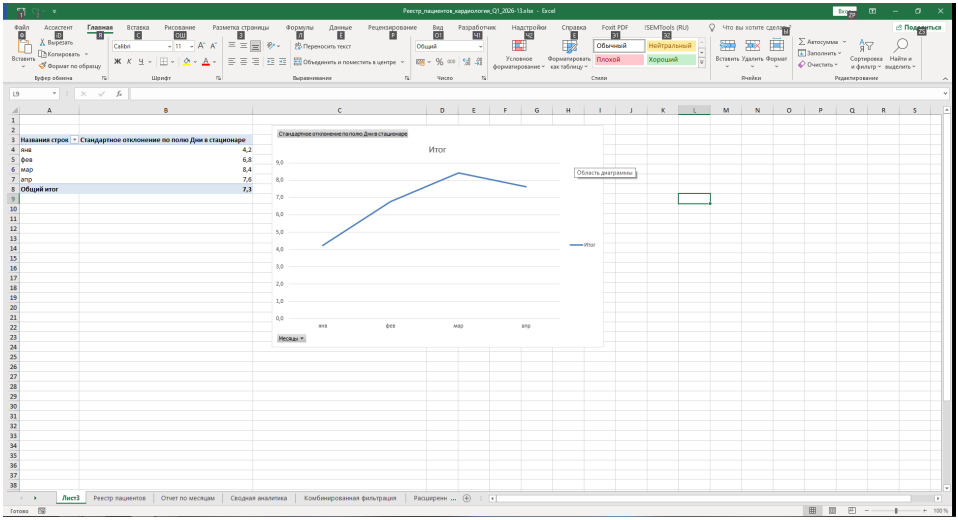


Рисунок 2.10. Сводная таблица с суммами коек по профилям

Шаг 3. Рассчитываем доли

Добавляем столбец «Доля, %» рядом со сводной таблицей. Формула: = значение / общий_итог × 100. Общий итог берем из ячейки с суммой (312). Получаем (рис. 2.11).

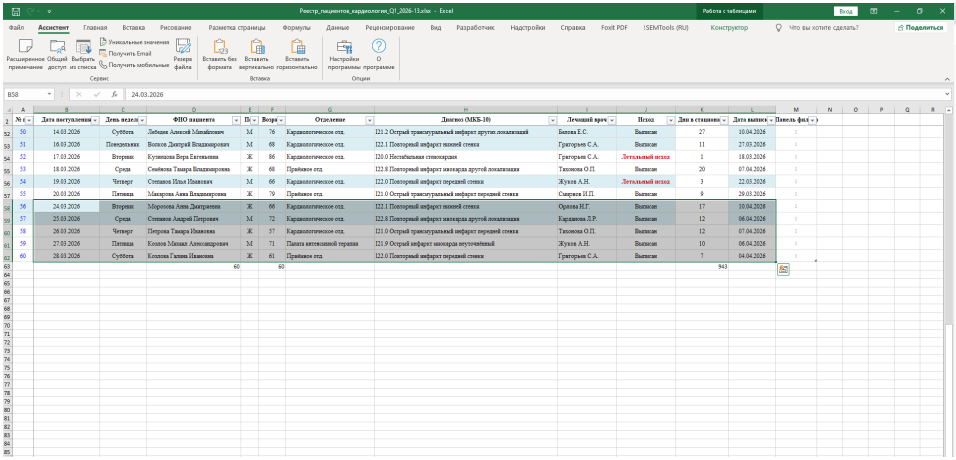


Рисунок 2.11. Итоговая таблица с рассчитанными долями, выделенными условным форматированием

Шаг 4. Анализируем и интерпретируем

Теперь у нас есть не просто абсолютные цифры, а понимание структуры. Мы видим, что хирургический профиль занимает почти 40% коечного фонда, терапевтический – чуть более 35%, детский – около 22%. Это важная информация для главного врача: если в регионе растет потребность в детской хирургии, а доля детских коек остается неизменной, возможно, нужно пересматривать структуру.

Но структура коечного фонда – это только статический срез. В динамике она становится еще более ценной. Сравнивая доли за несколько лет, можно увидеть тренды: увеличивается ли доля реанимационных коек (что может говорить о росте тяжести пациентов), снижается ли доля терапевтических коек (возможно, происходит перепрофилирование), как меняется соотношение между взрослыми и детскими отделениями.

Особые случаи: когда проценты могут ввести в заблуждение.

В работе с долями и процентами важно помнить о нескольких ловушках, в которые попадают даже опытные аналитики.

Маленькие знаменатели. Если общее количество наблюдений мало, даже один случай может дать большой процент, который будет неустойчив и плохо интерпретируем. Например, «доля летальных исходов в отделении за месяц составила 33%» звучит страшно, но если в отделении за месяц пролечено всего 3 пациента, то 33% – это один случай. В таких ситуациях разумнее приводить абсолютные цифры или указывать доверительные интервалы (о них поговорим в части 6).

Процент от процента. В сложных отчетах иногда пытаются складывать или вычитать проценты, полученные из разных знаменателей. Это грубая ошибка. Например, если 50% пациентов отделения – мужчины, а среди мужчин 20% – курильщики, нельзя сказать, что курильщики составляют 30% от всех пациентов. Нужно умножать: 0,5 × 0,2 = 0,1 = 10%.

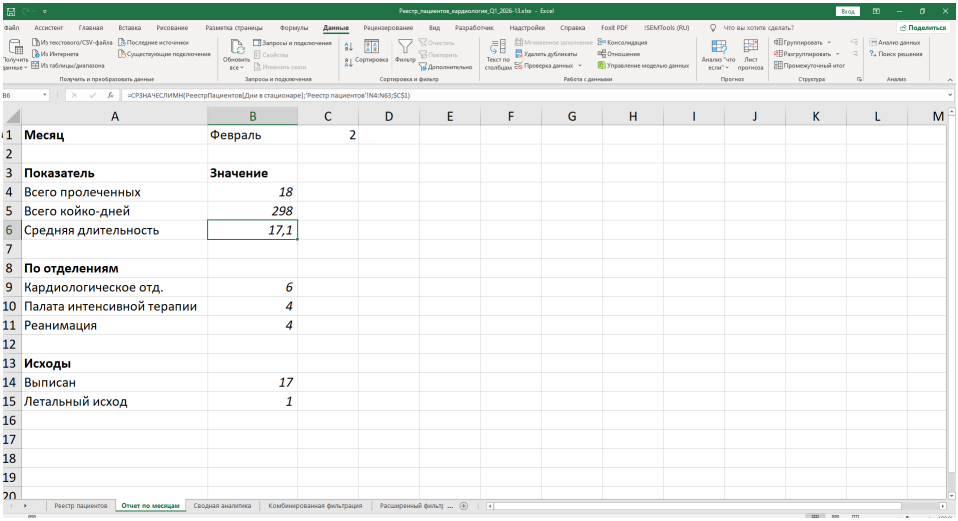
Эффект низкой базы. Когда показатель вырастает с 1 до 2%, говорят о росте на 100% (в относительном выражении). Это впечатляет, но может скрывать, что абсолютный прирост составил всего 1 случай на 100 пациентов.

В медицинской статистике принято указывать и абсолютные, и относительные изменения, чтобы избежать манипуляций.

Практический совет: визуализация структуры.

Доли и проценты проще всего воспринимаются в визуальном формате. Для отображения структуры идеально подходят круговые диаграммы (когда нужно показать соотношение частей целого) и столбчатые диаграммы с накоплением (когда нужно сравнить структуру нескольких периодов или групп).

В Excel создать такую диаграмму можно за несколько кликов: выделяем данные (категории и значения), на вкладке «Вставка» выбираем нужный тип. Для круговой диаграммы важно отсортировать данные по убыванию, чтобы самые крупные сегменты были на первом плане. Если сегментов много (более 5–6), мелкие доли лучше объединить в категорию «Прочие» – так диаграмма останется читаемой (рис. 2.12).



	A	B	C
1	**Месяц**	Февраль	2
2			
3	**Показатель**	**Значение**	
4	Всего пролеченных	*18*	
5	Всего койко-дней	*298*	
6	Средняя длительность	*17,1*	
7			
8	**По отделениям**		
9	Кардиологическое отд.	*6*	
10	Палата интенсивной терапии	*4*	
11	Реанимация	*4*	
12			
13	**Исходы**		
14	Выписан	*17*	
15	Летальный исход	*1*	
16			
17			
18			
19			
20			

Рисунок 2.12. Круговая диаграмма структуры коечного фонда с подписями долей

Главный вывод этой главы можно сформулировать так: доли и проценты – это язык, на котором руководство говорит с аналитикой. Научившись правильно рассчитывать структуру и интерпретировать относительные величины, вы сможете не просто предоставлять цифры, а рассказывать историю о том, как устроено ваше учреждение, куда движутся потоки пациентов и на какие точки нужно обратить внимание в первую очередь. И помните: абсолютные цифры показывают масштаб, проценты – значимость. Только вместе они дают полную картину.

Глава 2.4. Вариабельность данных: искусство видеть разброс

До сих пор мы говорили о показателях, которые описывают центр распределения: среднее, медиана, мода. Но данные – это не только «где середина». Не менее важно понимать, как данные разбросаны вокруг этой середины. Одно и то же среднее значение может скрывать принципиально разные реальности.

Представьте два отделения. В первом средняя длительность лечения составляет 7 дней, и все пациенты лежат от 6 до 8 дней. Во втором средняя тоже

7 дней, но одни пациенты выписываются на следующий день после поступления, а другие проводят в стационаре по три недели. Средняя одинакова, но реальность разная. В первом отделении все стабильно, процессы предсказуемы. Во втором – хаос, нет единых стандартов, а средняя цифра не описывает ни одного реального пациента.

Это и есть вариабельность – разброс данных. В медицинской статистике понимание разброса критически важно для оценки качества медицинской помощи, выявления аномалий и принятия управленческих решений. В этой главе мы разберемся, как измерять разброс и что делать с полученными цифрами.

Зачем нужно знать разброс: от среднего к реальности

Когда мы слышим фразу «средняя длительность лечения в отделении 7 дней», у нас возникает естественный вопрос: а насколько можно доверять этой цифре? Если все пациенты лежат от 6 до 8 дней – доверие высокое. Если разброс от 1 до 30 дней – средняя практически бесполезна для планирования.

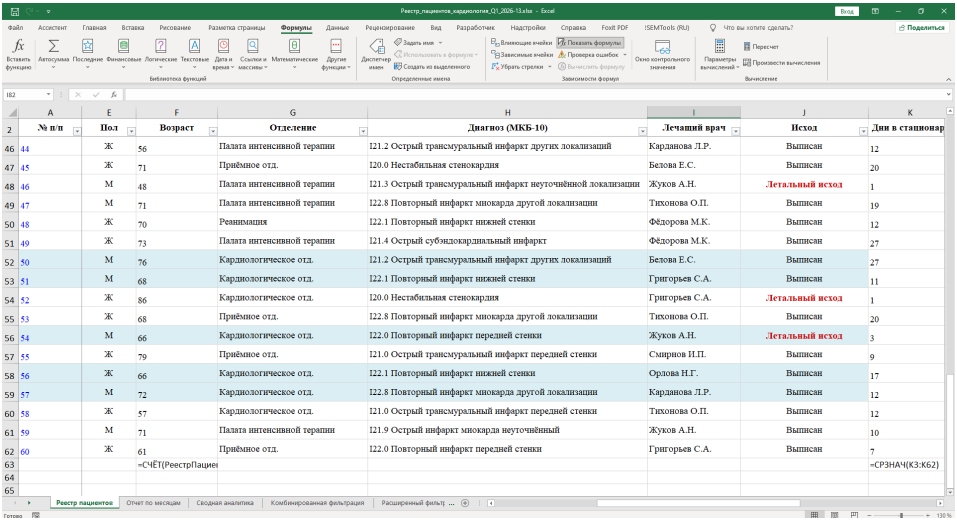
В работе врача-организатора показатели разброса дают ответы на важные вопросы.

- Насколько процессы стабильны? Маленький разброс говорит о стандартизации, большой – о вариативности практик.
- Есть ли выбросы, требующие внимания? Аномально долгие случаи лечения могут скрывать врачебные ошибки или системные проблемы.
- Однородна ли группа пациентов? Если разброс велик, возможно, мы сравниваем несравнимое, например, пациентов с разной тяжестью состояния.
- Как изменилась ситуация? Рост разброса во времени может сигнализировать о проблемах.

Показатели разброса – это не просто технические цифры. Это инструмент диагностики организационных проблем, который часто дает больше информации, чем само среднее значение.

Минимум и максимум: границы показателя

Самый простой способ оценить разброс – посмотреть на минимальное и максимальное значения. В Excel для этого используются функции МИН и МАКС. Их синтаксис прост: =МИН(диапазон), =МАКС(диапазон) (рис. 2.13).



№ п/п	Пол	Возраст	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стационаре
44	Ж	56	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Карданова Л.Р.	Выписан	12
45	Ж	71	Приёмное отд.	I20.0 Нестабильная стенокардия	Белова Е.С.	Выписан	20
46	М	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1
47	М	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19
48	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12
49	Ж	73	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27
50	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27
51	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11
52	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1
53	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20
54	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3
55	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9
56	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17
57	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Карданова Л.Р.	Выписан	12
58	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12
59	М	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10
60	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7
		=СЧЁТ(РеестрПациен					=СРЗНАЧ(K3:K62)

Рисунок 2.13. Таблица с данными по длительности лечения, рядом МИН и МАКС, показывающие границы

В медицинской статистике минимум и максимум дают быструю картину экстремумов. Например, анализируя длительность лечения в отделении, вы видите минимум 1 день и максимум 28 дней. Уже эта информация полезна: кто-то выписался на следующий день после поступления (возможно, это был пациент, поступивший планово для обследования), а кто-то провел в стационаре полтора месяца (возможно, тяжелый пациент, требующий особого внимания).

Однако у минимума и максимума есть серьезное ограничение: они зависят от единичных наблюдений. Один аномально длинный случай – и максимум уже не отражает реальность основной массы пациентов. Поэтому минимум и максимум всегда интерпретируются в контексте: вместе с другими показателями разброса.

Практический совет: в отчетах для руководства часто полезно указывать не только минимум и максимум, но и, например, 10-й и 90-й процентили, т.е. значения, ниже которых лежит 10% наблюдений и выше которых 10% наблюдений. Это позволяет отсечь экстремумы и показать диапазон, в котором находится основная масса пациентов. О том, как считать процентили, мы поговорим в конце этой главы.

Стандартное отклонение: как далеко мы ушли от нормы

Минимум и максимум показывают границы, но не дают представления о том, как данные распределены между этими границами. Для этого существует стандартное отклонение – один из самых важных показателей вариабельности.

Стандартное отклонение отвечает на вопрос: в среднем насколько каждое наблюдение отличается от среднего значения? Если стандартное отклонение маленькое, данные тесно сгруппированы вокруг среднего. Если большое, данные разбросаны широко, и среднее плохо описывает отдельные наблюдения.

В Excel для расчета стандартного отклонения по выборке используется функция СТАНДОТКЛОН.В (или ее англоязычный аналог STDEV.S). Формула: =СТАНДОТКЛОН.В(диапазон). Если у вас есть данные по всей генеральной совокупности (например, все пациенты отделения за год), используется СТАНДОТКЛОН.Г (STDEV.P). В большинстве медицинских задач мы работаем с выборками, поэтому применяем В (рис. 2.14).

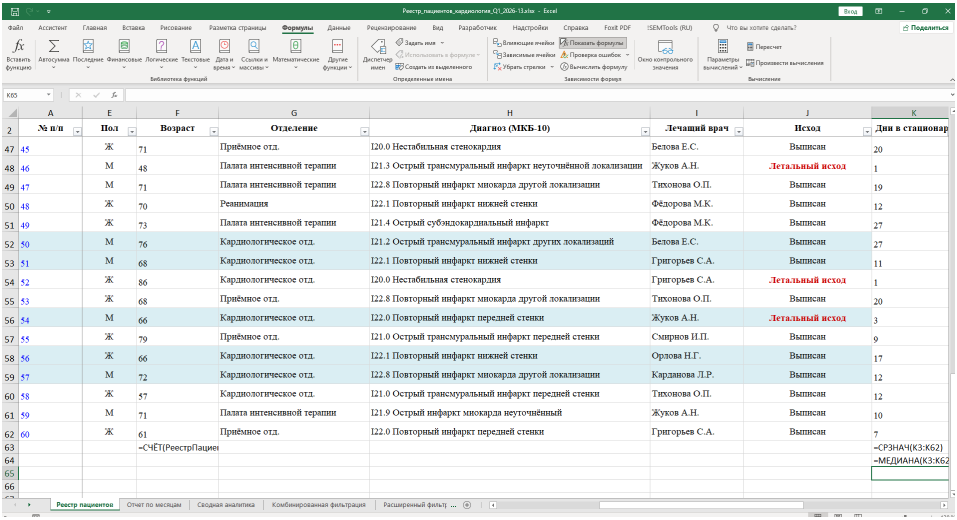


Рисунок 2.14. Формула СТАНДОТКЛОН.В в строке формул

Кейс: все ли врачи лечат одинаково?

Представьте, что мы анализируем работу хирургического отделения. Нас интересует вопрос: насколько единообразна практика лечения среди врачей отделения? Если одни врачи выписывают пациентов через 5–6 дней, а другие держат их по 12–14 дней при схожих диагнозах, это повод задуматься о стандартизации подходов.

У нас есть данные по 100 пациентам, пролеченным в отделении за месяц. Для каждого пациента известны лечащий врач, диагноз, длительность лечения. Мы хотим оценить разброс длительности лечения в целом по отделению, а затем сравнить разброс внутри группы каждого врача.

Шаг 1. Рассчитываем общие показатели по отделению

Вычисляем:

- средняя длительность лечения: =СРЗНАЧ(диапазон) → 8,2 дня;
- стандартное отклонение: =СТАНДОТКЛОН.В(диапазон) → 4,5 дня.

Стандартное отклонение 4,5 дня при средней 8,2 дня – это довольно большой разброс. Это значит, что типичное отклонение от среднего составляет более половины самого среднего. Уже на этом этапе мы понимаем: практика лечения в отделении неоднородна.

Шаг 2. Разбиваем по врачам

С помощью сводной таблицы группируем данные по лечащим врачам и для каждого вычисляем среднюю длительность лечения и стандартное отклонение (табл. 2.4).

Таблица 2.4. Средняя длительность лечения и стандартное отклонение

Врач	Врач	Врач	Врач
Иванов	25	6,5	1,8
Петрова	20	7,2	2,1
Сидоров	18	10,5	4,9
Кузнецов	22	8,8	4,2
Васильева	15	7,8	2,0

Шаг 3. Анализируем результаты

Мы видим интересную картину. У врачей Иванова, Петровой и Васильевой стандартное отклонение невелико (1,8–2,1 дня): их пациенты лечатся достаточно предсказуемо, практика стабильна. А вот у врачей Сидорова и Кузнецова стандартное отклонение значительно выше (4,2–4,9 дня): в их практике встречаются и очень короткие, и очень длинные случаи лечения.

Что это может означать? Вариантов несколько.

- Возможно, Сидоров и Кузнецов берут более тяжелых пациентов, тогда высокий разброс оправдан, и нужно смотреть на распределение диагнозов.
- Возможно, у них нет четких стандартов лечения, тогда это сигнал для клинико-экспертной комиссии.
- Возможно, в их практике есть отдельные выбросы – аномально длинные случаи, которые нужно разобрать индивидуально.

Без показателя разброса мы бы этого не увидели. Ограничившись средними значениями (6,5 против 10,5 дня), мы могли бы сделать вывод, что Сидоров просто дольше лечит пациентов. Но высокое стандартное отклонение подсказывает: проблема не в том, что он лечит дольше, а в том, что он лечит нестабильно.

Шаг 4. Формируем рекомендации

На основе этого анализа врач-статистик может подготовить для заведующего отделением рекомендации:

- провести разбор клинических случаев у врачей с высоким разбросом;
- сравнить диагнозы и тяжесть пациентов у разных врачей;
- рассмотреть возможность внедрения клинических рекомендаций для стандартизации подходов.

Это пример того, как простой статистический показатель становится инструментом управления качеством медицинской помощи.

Правило выявления выбросов: среднее ± 2–3 сигмы

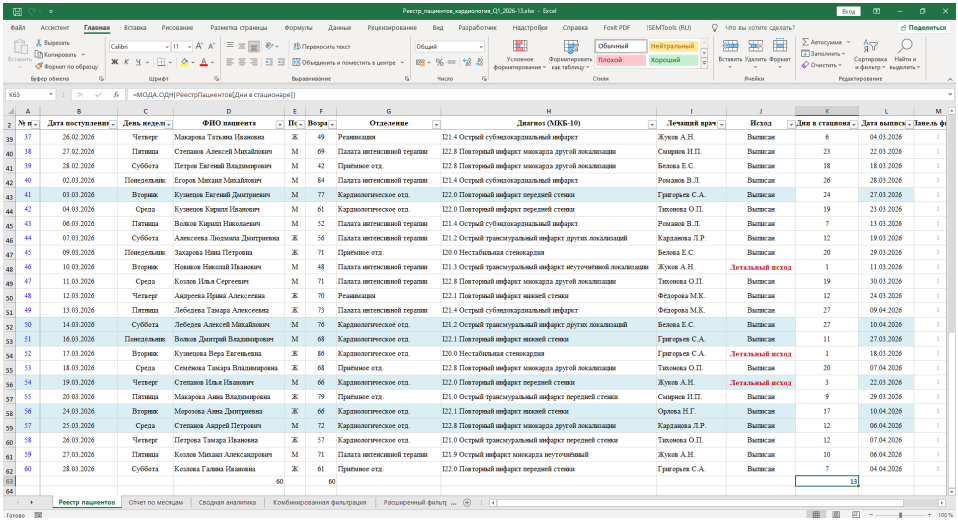
Одна из важнейших задач анализа разброса – выявление аномалий, или выбросов. Это наблюдения, которые значительно отличаются от основной массы данных. В медицинской статистике выбросы могут быть:

- ошибками ввода (например, вместо 3 дней указано 30 из-за опечатки);
- особыми клиническими случаями (редкое осложнение, уникальный пациент);
- сигналами системных проблем (врачебные ошибки, сбои в маршрутизации).

Классический способ выявления выбросов основан на свойствах нормального распределения. Если данные распределены нормально (по закону Гаусса), то:

- в интервале среднее ± 1 × стандартное отклонение находится около 68% всех наблюдений;
- в интервале среднее ± 2 × стандартное отклонение находится около 95% всех наблюдений;
- в интервале среднее ± 3 × стандартное отклонение находится около 99,7% всех наблюдений.

Соответственно наблюдения, которые выходят за пределы среднее ± 3 × стандартное отклонение, с высокой вероятностью являются выбросами – аномалиями, требующими внимания (рис. 2.15).



№ п	Дата поступления	День недел	ФИО пациента	Пс	Возра	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стациона	Дата выпис	Панель ф
37	26.02.2026	Четверг	Макарова Татьяна Ивановна	Ж	49	Реанимация	I21.4 Острый субэпикардиальный инфаркт	Жуков А.Н.	Выписан	6	04.03.2026	
38	27.02.2026	Пятница	Степанов Алексей Михайлович	М	69	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Смирнов И.П.	Выписан	23	22.03.2026	
39	28.02.2026	Суббота	Петров Евгений Владимирович	М	42	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Белова Е.С.	Выписан	18	18.03.2026	
40	02.03.2026	Понедельник	Егоров Михаил Михайлович	М	84	Палата интенсивной терапии	I21.4 Острый субэпикардиальный инфаркт	Романов В.Л.	Выписан	26	28.03.2026	
41	03.03.2026	Вторник	Кузнецов Евгений Дмитриевич	М	77	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	24	27.03.2026	
42	04.03.2026	Среда	Кузнецов Кирилл Иванович	М	61	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Тихонова О.П.	Выписан	19	23.03.2026	
43	06.03.2026	Пятница	Волков Кирилл Николаевич	М	52	Палата интенсивной терапии	I21.4 Острый субэпикардиальный инфаркт	Романов В.Л.	Выписан	7	13.03.2026	
44	07.03.2026	Суббота	Алексеева Людмила Дмитриевна	Ж	56	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Карлакова Л.Р.	Выписан	12	19.03.2026	
45	09.03.2026	Понедельник	Захарова Нина Петровна	Ж	71	Приёмное отд.	I20.0 Нестабильная стенокардия	Белова Е.С.	Выписан	20	29.03.2026	
46	10.03.2026	Вторник	Новиков Николай Иванович	М	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1	11.03.2026	
47	11.03.2026	Среда	Козлов Илья Сергеевич	М	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	30.03.2026	
48	12.03.2026	Четверг	Андреева Ирина Алексеевна	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	24.03.2026	
49	13.03.2026	Пятница	Лебедева Тамара Алексеевна	Ж	73	Палата интенсивной терапии	I21.4 Острый субэпикардиальный инфаркт	Фёдорова М.К.	Выписан	27	09.04.2026	
50	14.03.2026	Суббота	Лебедев Алексей Михайлович	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	10.04.2026	
51	16.03.2026	Понедельник	Волков Дмитрий Владимирович	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	27.03.2026	
52	17.03.2026	Вторник	Кузнецова Вера Евгеньевна	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1	18.03.2026	
53	18.03.2026	Среда	Семёнова Тамара Владимировна	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	07.04.2026	
54	19.03.2026	Четверг	Степанов Илья Иванович	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3	22.03.2026	
55	20.03.2026	Пятница	Макарова Анна Владимировна	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	29.03.2026	
56	24.03.2026	Вторник	Морозова Анна Дмитриевна	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	10.04.2026	
57	25.03.2026	Среда	Степанов Андрей Петрович	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Карлакова Л.Р.	Выписан	12	06.04.2026	
58	26.03.2026	Четверг	Петрова Тамара Ивановна	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12	07.04.2026	
59	27.03.2026	Пятница	Козлов Михаил Александрович	М	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	06.04.2026	
60	28.03.2026	Суббота	Козлова Галина Ивановна	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7	04.04.2026	
				60	60					13		

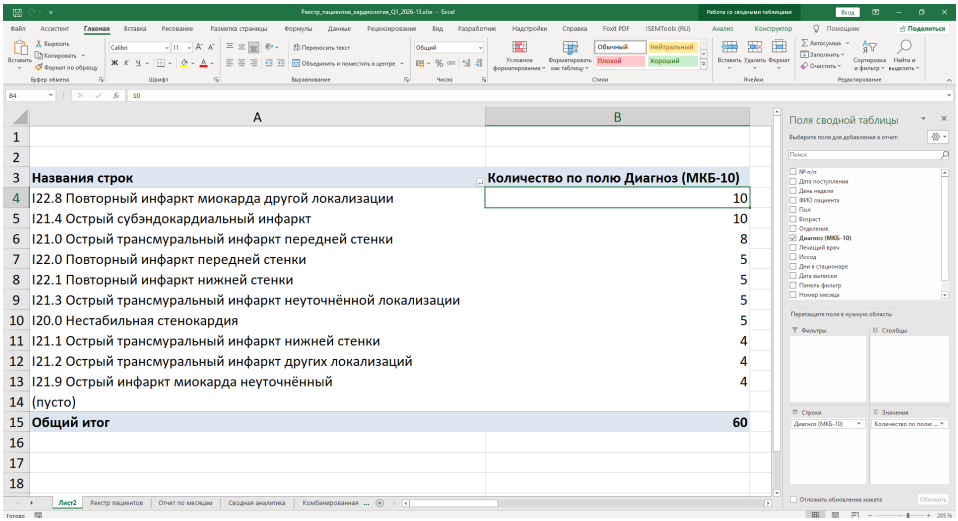
Рисунок 2.15. Нормальное распределение

В Excel это правило реализуется просто (рис. 2.16).

1. Вычисляем среднее (=СРЗНАЧ) и стандартное отклонение (=СТАНДОТКЛОН.В).

2. В соседнем столбце создаем формулу для проверки: =ИЛИ(значение < среднее – 3*ст.откл; значение > среднее + 3*ст.откл).

3. Фильтруем строки, где формула вернула ИСТИНА – это потенциальные выбросы.



	A	B
1		
2		
3	**Названия строк**	**Количество по полю Диагноз (МКБ-10)**
4	I22.8 Повторный инфаркт миокарда другой локализации	10
5	I21.4 Острый субэндокардиальный инфаркт	10
6	I21.0 Острый трансмуральный инфаркт передней стенки	8
7	I22.0 Повторный инфаркт передней стенки	5
8	I22.1 Повторный инфаркт нижней стенки	5
9	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	5
10	I20.0 Нестабильная стенокардия	5
11	I21.1 Острый трансмуральный инфаркт нижней стенки	4
12	I21.2 Острый трансмуральный инфаркт других локализаций	4
13	I21.9 Острый инфаркт миокарда неуточнённый	4
14	(пусто)	
15	**Общий итог**	**60**
16		
17		
18		

Рисунок 2.16. Таблица с выделенным выбросам, формула проверки в строке формул

Важное предостережение: правило «среднее ± 3 сигмы» работает хорошо, когда данные распределены нормально. В медицинских данных это бывает не всегда: длительность лечения, например, часто имеет асимметричное распределение (большинство пациентов лечатся недолго, некоторые долго).

В таких случаях лучше использовать альтернативные методы, например, межквартильный размах (о нем ниже). Но как первая линия обороны, особенно в сочетании с визуальным контролем, это правило очень полезно.

Расширенные показатели разброса: квартили и процентили

Для более глубокого анализа разброса, особенно при асимметричных распределениях, используются квартили и процентили. В Excel они вычисляются функциями:

- КВАРТИЛЬ.ВКЛ (или КВАРТИЛЬ.ИСКЛ) – для расчета квартилей.
- ПРОЦЕНТИЛЬ.ВКЛ (или ПРОЦЕНТИЛЬ.ИСКЛ) – для расчета любого процентиля.

Квартили делят данные на четыре равные части.

- Первый квартиль (Q1) – 25% наблюдений ниже этого значения.
- Второй квартиль (Q2) – это медиана, 50% наблюдений ниже.
- Третий квартиль (Q3) – 75% наблюдений ниже.

Разница между третьим и первым квартилями называется межквартильным размахом (IQR). Это устойчивая к выбросам мера разброса. Для выявления выбросов по квартилям используется правило: выбросами считаются значения, выходящие за пределы Q1 – 1,5 × IQR и Q3 + 1,5 × IQR (рис. 2.17).

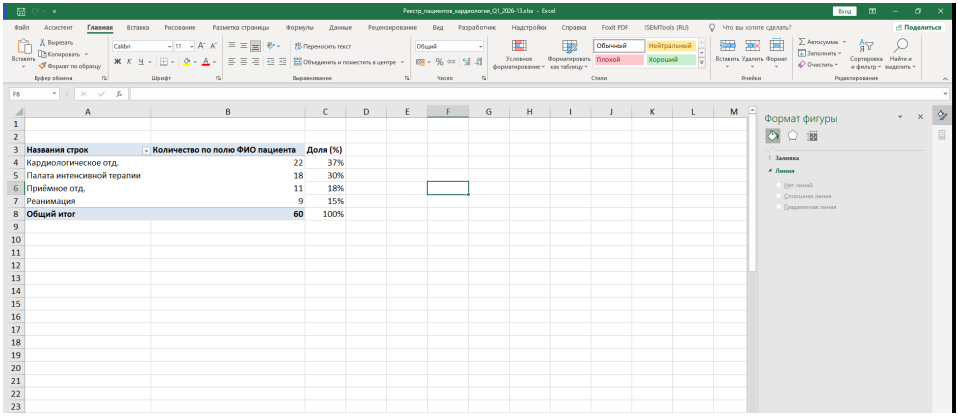


Рисунок 2.17. Ящик с усами (box plot), показывающий медиану, квартили, границы усов

В Excel можно построить диаграмму «ящик с усами» (box plot), которая наглядно показывает все эти показатели. Такая диаграмма – отличный способ визуализировать разброс и выбросы одним графиком.

Практический совет: создаем панель контроля качества.

Для регулярного мониторинга вариабельности можно создать простую панель контроля. Каждый месяц вы обновляете данные и смотрите на динамику стандартного отклонения по ключевым показателям (длительность лечения, сроки ожидания, нагрузка на врача). Если стандартное отклонение начинает расти – это ранний сигнал о том, что процессы становятся менее стабильными.

Возможно, появились новые врачи, изменилась структура пациентов или нарушились стандарты. Вы успеваете заметить проблему до того, как она отразится на средних показателях (рис. 2.18).

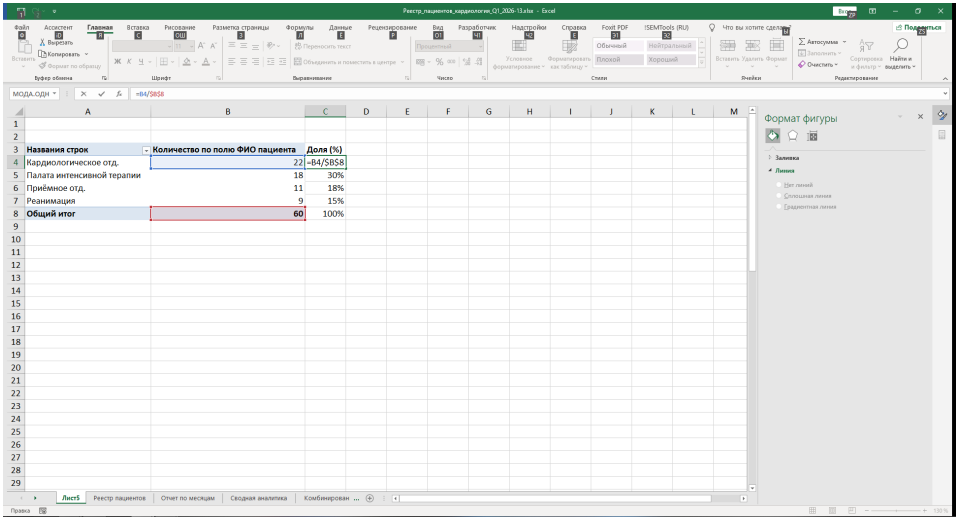
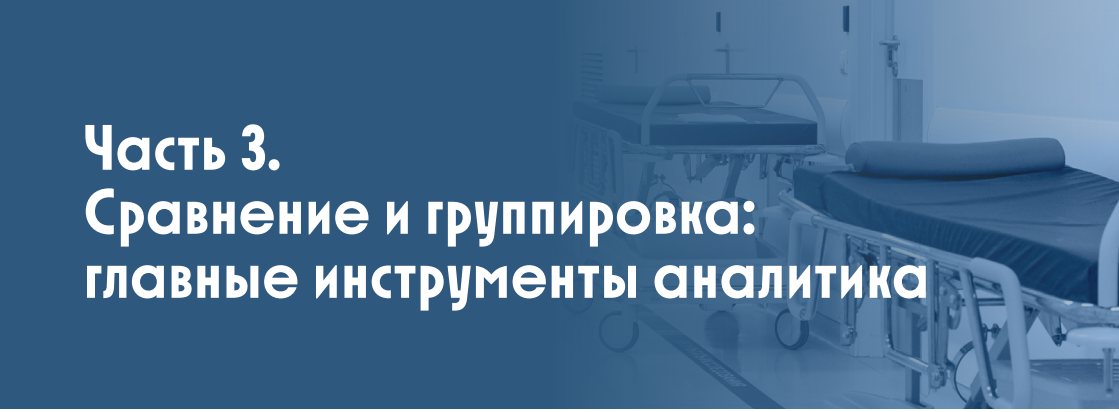


Рисунок 2.18. Линейный график, показывающий динамику стандартного отклонения длительности лечения по месяцам

Главный вывод этой главы можно сформулировать так: среднее без разброса – это половина правды. Показатели вариабельности превращают абстрактную «среднюю температуру по больнице» в инструмент диагностики реальных процессов. Минимум и максимум показывают границы, стандартное отклонение – меру хаоса, а правило выбросов помогает находить аномалии, требующие внимания. Освоив эти инструменты, вы сможете не только отвечать на вопрос «Сколько в среднем?», но и понимать, насколько этой средней можно доверять и где скрываются настоящие проблемы.



Часть 3. Сравнение и группировка: главные инструменты аналитика

Глава 3.1. Сводные таблицы – король анализа

До сих пор мы учились готовить данные к анализу: очищать, проверять, приводить к единому виду. Но настоящая аналитическая работа начинается там, где мы начинаем сравнивать, группировать, искать связи и отвечать на вопросы, которые ставит руководство. Сколько пациентов прошло через каждое отделение? Как распределяется нагрузка по дням недели? Какая средняя длительность лечения в кардиологии и хирургии? Как подтянуть фамилии врачей из одного справочника в другой?

Эти задачи решаются тремя мощными инструментами Excel, которые мы разберем в этой части: сводными таблицами, формулами с условиями и функциями вертикального поиска. Каждый из них – самостоятельный мир, но вместе они образуют фундамент, на котором строится аналитическая работа врача-статистика. Освоив их, вы перестанете тратить часы на ручной подсчет и начнете получать ответы на управленческие вопросы за минуты.

Если бы мне предложили назвать один инструмент Excel, который превращает врача-статистика из человека с калькулятором в настоящего аналитика, я бы без колебаний назвал сводные таблицы. Этот инструмент настолько мощен, что его часто называют «король анализа», и это не преувеличение.

Что такое сводная таблица? Это интерактивный отчет, который позволяет за считаные секунды группировать, суммировать, сравнивать и фильтровать тысячи строк данных. Вы задаете вопрос («Покажи мне среднюю длительность лечения по каждому отделению»), и сводная таблица дает ответ. Вы меняете вопрос

(«А теперь покажи по врачам»), и ответ меняется мгновенно. Никаких формул, никаких ручных подсчетов, только перетаскивание полей мышью.

В этой главе мы разберемся с устройством сводных таблиц, научимся создавать их с нуля и решим три реальных медицинских кейса, которые возникают в работе каждого врача-статистика.

Что такое сводная таблица и зачем она нужна?

Представьте, что у вас есть таблица с 10 000 записей о пациентах, поступивших в стационар за год. Вам нужно ответить на вопросы:

- сколько пациентов прошло через каждое отделение;
- какова средняя длительность лечения в кардиологии;
- как изменилось число госпитализаций по месяцам;
- какая нозология встречается чаще всего?

Вручную считать все это – неделя работы. С помощью формул – несколько часов. С помощью сводной таблицы – 2 мин.

Сводная таблица – это не просто суммирование, это динамический инструмент, который позволяет:

- группировать данные по любому признаку (отделение, врач, диагноз, месяц, день недели);
- агрегировать значения разными способами (сумма, среднее, количество, максимум, минимум, процент от итога);
- фильтровать данные прямо внутри отчета, не трогая исходную таблицу;
- менять структуру отчета в реальном времени, перетаскивая поля мышью.

И все это без единой формулы, без риска ошибиться в ссылках, без необходимости перестраивать отчет каждый месяц.

Структура сводной таблицы: четыре области власти.

Любая сводная таблица состоит из четырех основных областей, которые определяют, как будет выглядеть отчет. Понимание этих областей – ключ к управлению сводной таблицей (рис. 3.1).

Строки – поля, которые определяют, по каким категориям будут сгруппированы данные по вертикали. Если поместить сюда «Отделение», каждое отделение получит свою строку в отчете.

Столбцы – поля, которые группируют данные по горизонтали. Если поместить сюда «Месяц», месяцы распределятся по столбцам, и вы увидите динамику по месяцам.

Значения – поля, которые будут подсчитаны (суммированы, усреднены, посчитаны). Сюда помещаются числовые показатели: койко-дни, возраст, количество записей.

Фильтры – поля, которые позволяют ограничить отчет определенными значениями. Например, можно поставить фильтр по году и смотреть данные только за 2025 г.

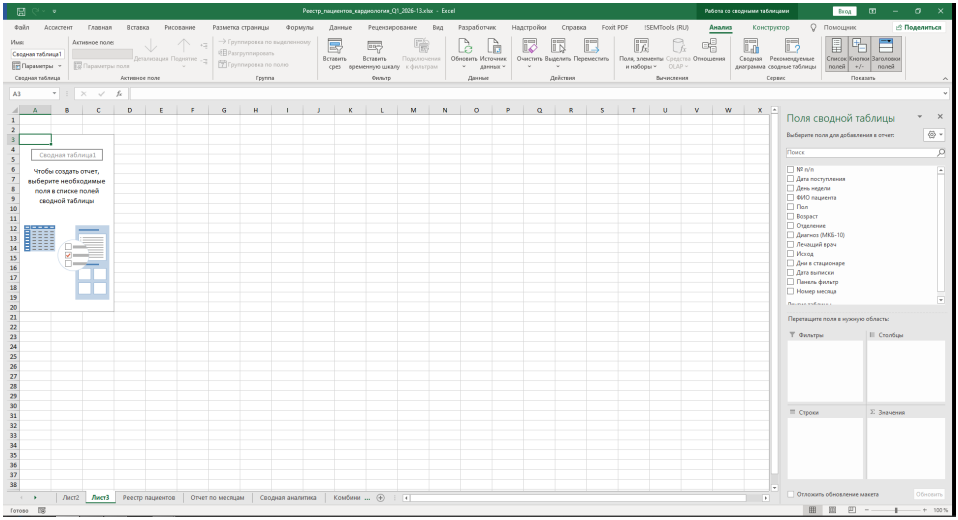


Рисунок 3.1. Схема сводной таблицы с выделенными четырьмя областями и пояснениями

Визуально это напоминает конструктор: вы берете поля из списка и перетаскиваете их в нужную область. Excel мгновенно перестраивает отчет.

Создание первой сводной таблицы: пошаговая инструкция

Давайте создадим первую сводную таблицу на реальных данных. У нас есть умная таблица «Пациенты» со следующими столбцами:

- Дата поступления;
- Отделение;
- Лечащий врач;
- Диагноз;
- Койко-дни;
- Возраст;
- Исход.

Шаг 1. Выделяем данные

Кликните в любую ячейку внутри умной таблицы. Excel автоматически определит диапазон.

Шаг 2. Вставляем сводную таблицу

Перейдите на вкладку «Вставка» → «Сводная таблица». Появится диалоговое окно. Excel уже подхватил диапазон (в случае умной таблицы – имя таблицы). Выберите, куда поместить сводную таблицу: на новый лист (рекомендуется) или на существующий (рис. 3.2).

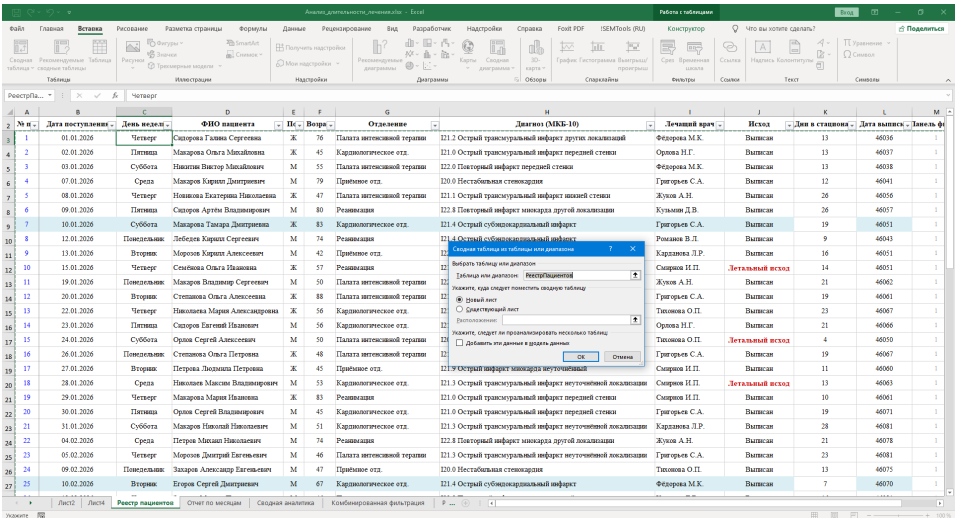


Рисунок 3.2. Окно создания сводной таблицы с выделенным диапазоном и выбором места размещения

Шаг 3. Работаем с полями

Справа появится панель «Поля сводной таблицы». Здесь перечислены все столбцы исходной таблицы. Наша задача – распределить их по областям (рис. 3.3).

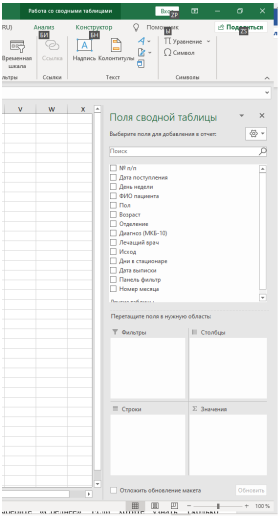


Рисунок 3.3. Панель «Поля сводной таблицы» с перечнем полей

Шаг 4. Выбираем функцию в значениях

Когда вы перетаскиваете числовое поле в область «Значения», Excel по умолчанию использует функцию «Сумма» для чисел и «Количество» для текста. Но это можно изменить. Нажмите на поле в области «Значения» → «Параметры полей значений» и выберите нужную функцию: сумма, среднее, количество, максимум, минимум, произведение и др. (рис. 3.4).

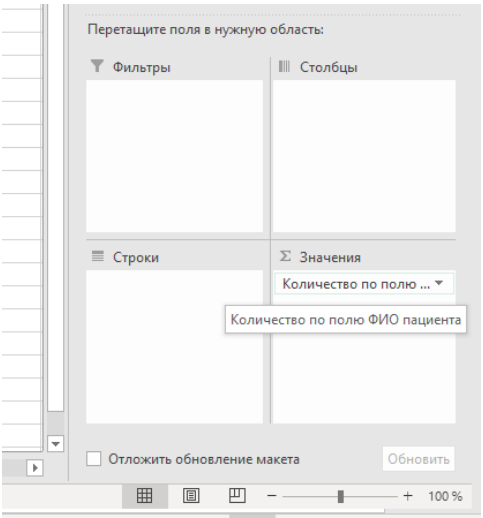


Рисунок 3.4. Окно «Параметры полей значений» с выбором функции

Это один из самых важных моментов. Если вы хотите посчитать среднюю длительность лечения, выберите «Среднее». Если хотите узнать, сколько пациентов было в каждом отделении, выберите «Количество» (не «Сумма»!). Если случайно оставить «Сумму» для текстового поля, результат будет 0 – распространенная ошибка новичков.

Кейс 1: Средняя длительность лечения по каждому отделению

Задача: сравнить эффективность работы отделений стационара. Нужно рассчитать среднюю длительность лечения для каждого отделения за месяц.

Шаг 1. Создаем сводную таблицу

Помещаем поле «Отделение» в область «Строки», поле «Койко-дни» – в область «Значения».

Шаг 2. Меняем функцию в значениях

По умолчанию Excel покажет сумму койко-дней. Нам нужно среднее. Кликаем на поле в области «Значения» → «Параметры полей значений» → выбираем «Среднее» (рис. 3.5).

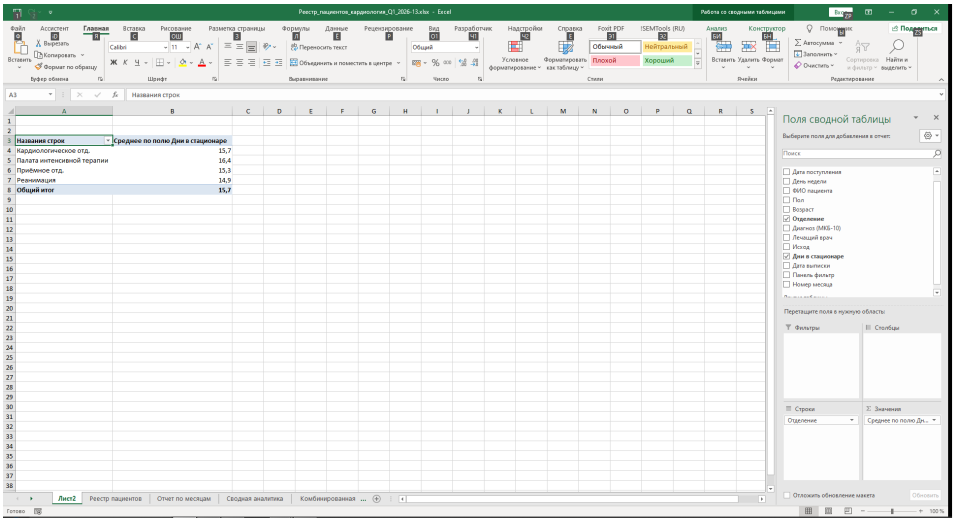


Рисунок 3.5. Сводная таблица с отделениями в строках и средней длительностью лечения в значениях

Шаг 3. Форматируем результат

Средняя длительность лечения обычно отображается с одним десятичным знаком. Кликаем правой кнопкой по ячейкам с результатами → «Формат ячеек» → «Числовой» → устанавливаем один знак после запятой.

Шаг 4. Анализируем результат

Мы видим, что средняя длительность лечения колеблется от 14,9 дня в реанимации до 16,4 дня в палате интенсивной терапии.

Практические советы по работе со сводными таблицами

Всегда используйте умные таблицы в качестве источника. Если вы добавите новые строки в умную таблицу, сводная таблица увидит их после обновления. Если источник – обычный диапазон, придется каждый раз менять диапазон вручную.

Обновление сводной таблицы. Сводная таблица не обновляется автоматически при изменении данных. После добавления или изменения записей нажмите правой кнопкой по сводной таблице → «Обновить» (или используйте сочетание Alt + F5).

Срезы (slicers) – удобная замена фильтрам. Вкладка «Анализ сводной таблицы» → «Вставить срез». Срезы – это красивые кнопки, которые позволяют фильтровать отчет одним кликом. Их можно использовать для создания интерактивных панелей управления.

Группировка дат. В сводной таблице даты можно группировать по дням, месяцам, кварталам, годам. Кликните правой кнопкой по любой дате → «Группировать» и выберите нужный уровень. Это избавляет от необходимости создавать вспомогательные столбцы (рис. 3.6).

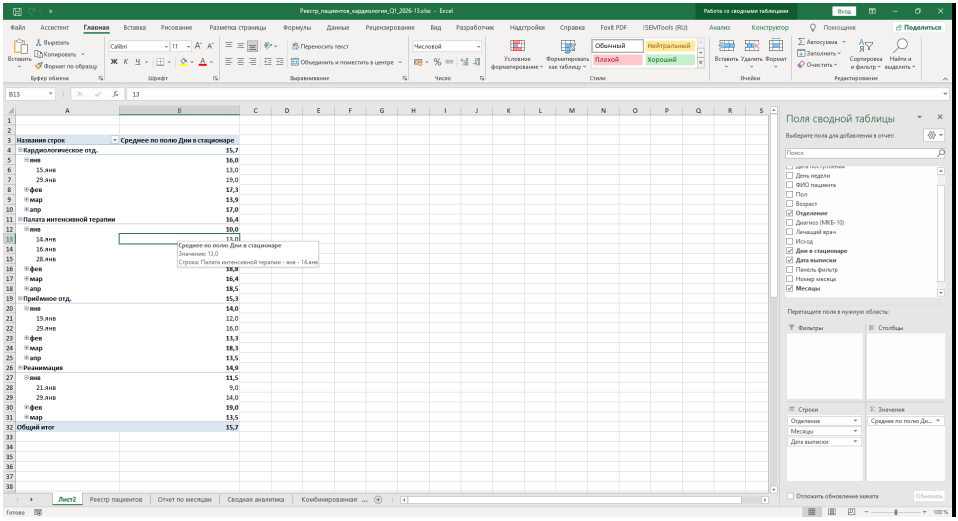


Рисунок 3.6. Окно группировки дат с выбором «Месяцы» и «Годы»

Главный вывод этой главы можно сформулировать так: сводные таблицы превращают хаос данных в структурированную информацию за секунды. Они позволяют врачу-статистику не тратить время на рутинные подсчеты, а сосредоточиться на главном – интерпретации результатов и подготовке управленческих решений. Освоив сводные таблицы, вы сделаете огромный шаг от «человека, который считает» к «аналитику, который управляет».

Глава 3.2. Формулы с условиями: когда одной функции недостаточно

Сводные таблицы – мощный инструмент, но не всегда удобный. Иногда нужно получить быстрый ответ на конкретный вопрос, не создавая целый отчет. Иногда результат должен быть встроен в сложную формулу. Иногда нужно посчитать что-то, что сводная таблица делает неочевидным образом. Именно здесь на помощь приходят формулы с условиями – функции, которые считают только ячейки, соответствующие заданным критериям.

Представьте себе типичные вопросы врача-статистика: сколько пациентов старше 60 лет с гипертонической болезнью состоят на диспансерном учете? Какова общая сумма койко-дней по кардиологическому отделению за последний квартал? Какая средняя температура у пациентов с подтвержденным гриппом? На каждый из этих вопросов можно ответить с помощью одной-единственной функции, если знать, какую именно выбрать.

В этой главе мы разберем три семейства функций, которые должны быть в арсенале каждого аналитика: СЧЕТЕСЛИ и СЧЕТЕСЛИМН (подсчет количества по условиям), СУММЕСЛИ и СУММЕСЛИМН (суммирование по условиям), СРЗНАЧЕСЛИ (среднее по условиям). Каждую функцию рассмотрим на реальных медицинских кейсах.

Зачем нужны функции с условиями

Функции с условиями решают одну простую задачу: выполнить действие (посчитать, сложить, усреднить) не над всеми ячейками диапазона, а только над теми, которые соответствуют определенным критериям. Это похоже на фильтрацию данных внутри формулы, без создания вспомогательных таблиц и без изменения исходных данных.

Главное преимущество этих функций – гибкость. Вы можете встроить их в любую формулу, комбинировать между собой, использовать в сложных отчетах. Они работают на лету: изменились данные – изменился результат. И, что особенно важно, они не требуют перестройки структуры листа, как сводные таблицы.

В Excel есть функции для одного условия (СЧЕТЕСЛИ, СУММЕСЛИ, СРЗНАЧЕСЛИ) и для нескольких условий одновременно (СЧЕТЕСЛИМН, СУММЕСЛИМН, СРЗНАЧЕСЛИМН). Вторые появились в более новых версиях Excel и значительно расширили возможности анализа.

СЧЕТЕСЛИ и СЧЕТЕСЛИМН: считаем по правилам

Начнем с самого понятного: подсчета количества. Функция СЧЕТЕСЛИ подсчитывает количество ячеек в диапазоне, которые удовлетворяют одному условию. Синтаксис прост: =СЧЕТЕСЛИ(диапазон; критерий) (рис. 3.7).

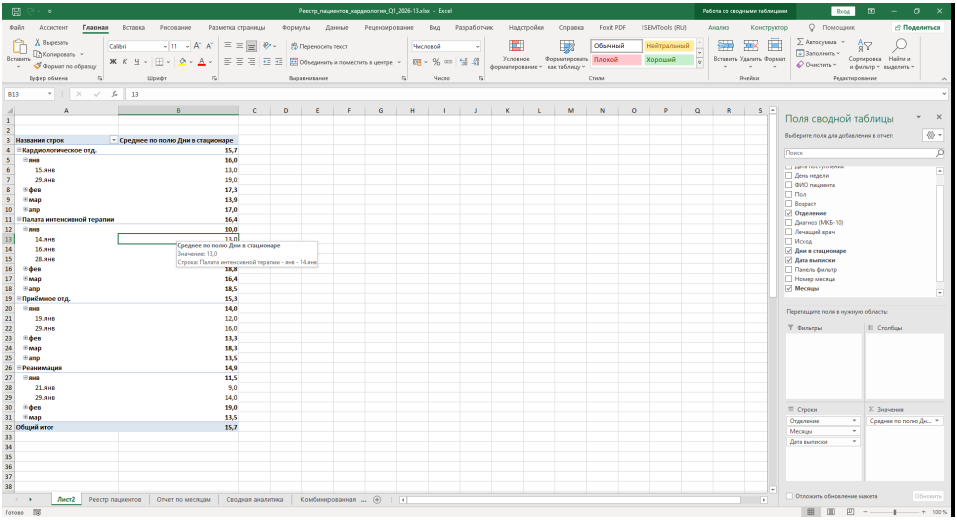


Рисунок 3.7. Формула СЧЕТЕСЛИ в строке формул с пояснением аргументов (подсчет количества пациентов старше 60 лет)

Кейс: пациенты старше 60 лет с гипертонией

Представьте, что у нас есть таблица диспансерного учета пациентов поликлиники. В столбцах: Ф.И.О., возраст, диагноз (основной), дата последнего визита. Нам нужно быстро посчитать, сколько пациентов старше 60 лет состоят на учете с диагнозом «гипертоническая болезнь». Задача требует двух условий: возраст > 60 и диагноз = «гипертоническая болезнь». Здесь одной СЧЕТЕСЛИ недостаточно – нужна СЧЕТЕСЛИМН.

Формула будет выглядеть так:

==СЧЕТЕСЛИМН(РеестрПациентов[Возраст];“>60”) РеестрПациентов[Диагноз (МКБ-10)]; “I20.0 Нестабильная стенокардия”) (рис. 3.8).

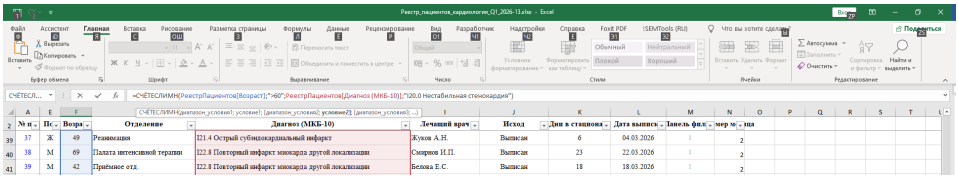


Рисунок 3.8. Формула СЧЕТЕСЛИМН с выделенными парами «диапазон–критерий»

В этой формуле:

первая пара (диапазон возрастов и условие «больше 60»);

вторая пара (диапазон диагнозов и условие «равно гипертоническая болезнь»).

Excel считает только те строки, где оба условия выполнены одновременно. Результат – количество пациентов в этой группе.

Важные нюансы работы с критериями

Текстовые критерии заключаются в кавычки: «Гипертоническая болезнь».

Числовые критерии тоже в кавычках, если используются операторы сравнения: «>60», «<=30», «<>0» (не равно нулю).

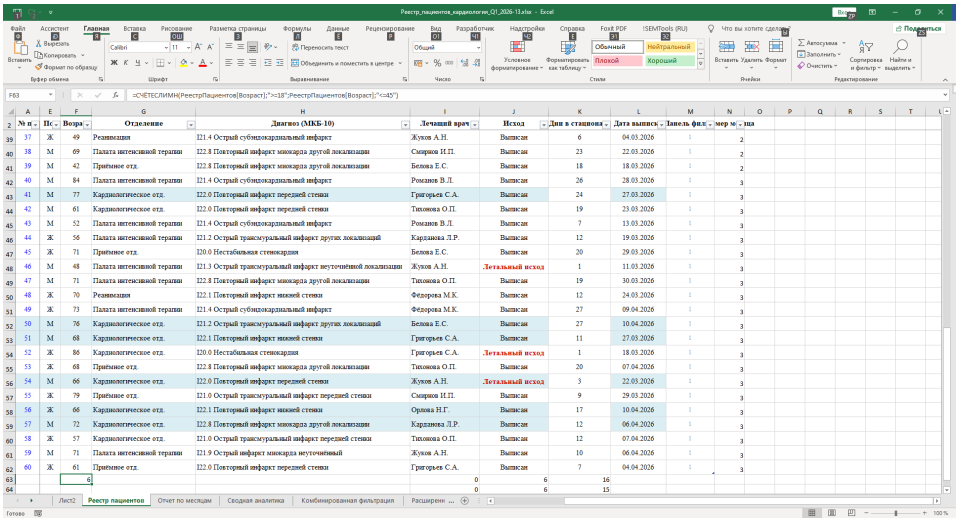
Можно использовать подстановочные символы: звездочка * заменяет любую последовательность символов, вопросительный знак ? – один символ. Например, "грипп*" найдет «грипп», «гриппозная инфекция», «грипп А».

Если критерий ссылается на ячейку, кавычки не нужны, а оператор сравнения можно объединить через &: «>»&A1.

Практический пример: возрастные группы

Часто требуется разбить пациентов на возрастные группы. С помощью СЧЕТЕСЛИМН это делается просто. Чтобы посчитать пациентов в возрасте от 18 до 45 лет (включительно), используем:

=СЧЕТЕСЛИМН(РеестрПациентов[Возраст];«>=18»;РеестрПациентов[Возраст];«<=45») (рис. 3.9).



№ п	Пс	Возра	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стациона	Дата выпис	Панель фил	мер м	ица
37	Ж	49	Реанимация	I21.4 Острый субэндокардиальный инфаркт	Жуков А.Н.	Выписан	6	04.03.2026	1		2
38	М	69	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Смирнов И.П.	Выписан	23	22.03.2026	1		2
39	М	42	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Белова Е.С.	Выписан	18	18.03.2026	1		2
40	М	84	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Романов В.Л.	Выписан	26	28.03.2026	1		3
41	М	77	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	24	27.03.2026	1		3
42	М	61	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Тихонова О.П.	Выписан	19	23.03.2026	1		3
43	М	52	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Романов В.Л.	Выписан	7	13.03.2026	1		3
44	Ж	56	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Карданова Л.Р.	Выписан	12	19.03.2026	1		3
45	Ж	71	Приёмное отд.	I20.0 Нестабильная стенокардия	Белова Е.С.	Выписан	20	29.03.2026	1		3
46	М	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1	11.03.2026	1		3
47	М	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	30.03.2026	1		3
48	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	24.03.2026	1		3
49	Ж	73	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27	09.04.2026	1		3
50	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	10.04.2026	1		3
51	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	27.03.2026	1		3
52	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1	18.03.2026	1		3
53	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	07.04.2026	1		3
54	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3	22.03.2026	1		3
55	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	29.03.2026	1		3
56	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	10.04.2026	1		3
57	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Карданова Л.Р.	Выписан	12	06.04.2026	1		3
58	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12	07.04.2026	1		3
59	М	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	06.04.2026	1		3
60	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7	04.04.2026	1		3
		0					0	6	16		
							0	6	15		

Рисунок 3.9. Таблица с результатами подсчета по возрастным группам, формулы рядом

СУММЕСЛИ и СУММЕСЛИМН: складываем с умом

Функции суммирования с условиями работают по тому же принципу, но добавляют третий аргумент – диапазон суммирования. СУММЕСЛИ суммирует ячейки, соответствующие одному условию. СУММЕСЛИМН – нескольким условиям одновременно.

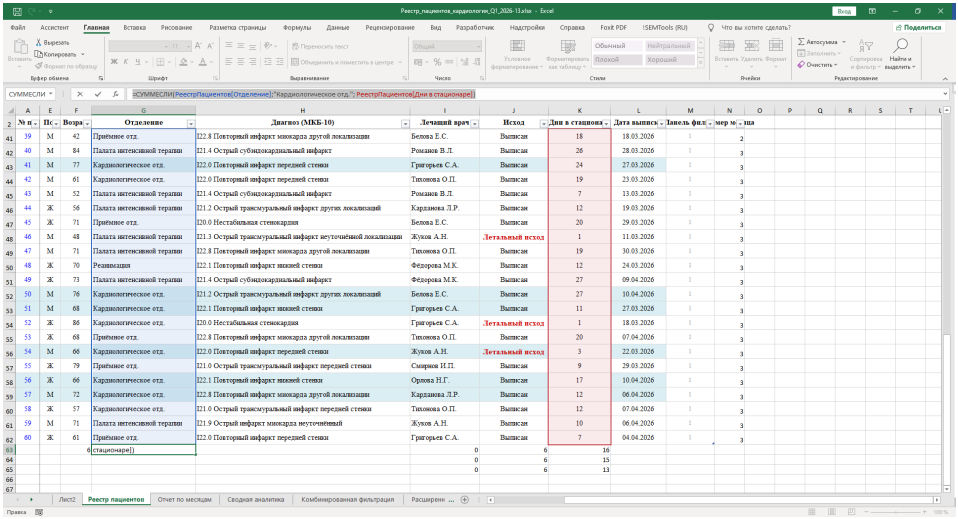
Синтаксис СУММЕСЛИМН имеет особенность: сначала указывается диапазон суммирования, а затем пары «диапазон условий – критерий»:

=СУММЕСЛИМН(диапазон_суммирования; диапазон_условия1; критерий1; диапазон_условия2; критерий2; ...)

Кейс: общий койко-день по отделениям

Вернемся к реестру пациентов стационара. Нам нужно быстро посчитать общее количество койко-дней по каждому отделению, не строя сводную таблицу. Для одного отделения (например, кардиологического) используем СУММЕСЛИ:

=СУММЕСЛИ(РеестрПациентов[Отделение];«Кардиологическое отд.»; РеестрПациентов[Дни в стационаре]) (рис. 3.10).



№ п	Пс	Возра	Отделение	Диагноз (МКБ-10)	Лечащий врач	Исход	Дни в стациона	Дата выписк	Панель фил	мер м	ица
39	М	42	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Белова Е.С.	Выписан	18	18.03.2026	1	2	
40	М	84	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Романов В.Л.	Выписан	26	28.03.2026	1	3	
41	М	77	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	24	27.03.2026	1	3	
42	М	61	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Тихонова О.П.	Выписан	19	23.03.2026	1	3	
43	М	52	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Романов В.Л.	Выписан	7	13.03.2026	1	3	
44	Ж	56	Палата интенсивной терапии	I21.2 Острый трансмуральный инфаркт других локализаций	Кардакова Л.Р.	Выписан	12	19.03.2026	1	3	
45	Ж	71	Приёмное отд.	I20.0 Нестабильная стенокардия	Белова Е.С.	Выписан	20	29.03.2026	1	3	
46	М	48	Палата интенсивной терапии	I21.3 Острый трансмуральный инфаркт неуточнённой локализации	Жуков А.Н.	Летальный исход	1	11.03.2026	1	3	
47	М	71	Палата интенсивной терапии	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	19	30.03.2026	1	3	
48	Ж	70	Реанимация	I22.1 Повторный инфаркт нижней стенки	Фёдорова М.К.	Выписан	12	24.03.2026	1	3	
49	Ж	73	Палата интенсивной терапии	I21.4 Острый субэндокардиальный инфаркт	Фёдорова М.К.	Выписан	27	09.04.2026	1	3	
50	М	76	Кардиологическое отд.	I21.2 Острый трансмуральный инфаркт других локализаций	Белова Е.С.	Выписан	27	10.04.2026	1	3	
51	М	68	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Григорьев С.А.	Выписан	11	27.03.2026	1	3	
52	Ж	86	Кардиологическое отд.	I20.0 Нестабильная стенокардия	Григорьев С.А.	Летальный исход	1	18.03.2026	1	3	
53	Ж	68	Приёмное отд.	I22.8 Повторный инфаркт миокарда другой локализации	Тихонова О.П.	Выписан	20	07.04.2026	1	3	
54	М	66	Кардиологическое отд.	I22.0 Повторный инфаркт передней стенки	Жуков А.Н.	Летальный исход	3	22.03.2026	1	3	
55	Ж	79	Приёмное отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Смирнов И.П.	Выписан	9	29.03.2026	1	3	
56	Ж	66	Кардиологическое отд.	I22.1 Повторный инфаркт нижней стенки	Орлова Н.Г.	Выписан	17	10.04.2026	1	3	
57	М	72	Кардиологическое отд.	I22.8 Повторный инфаркт миокарда другой локализации	Кардакова Л.Р.	Выписан	12	06.04.2026	1	3	
58	Ж	57	Кардиологическое отд.	I21.0 Острый трансмуральный инфаркт передней стенки	Тихонова О.П.	Выписан	12	07.04.2026	1	3	
59	М	71	Палата интенсивной терапии	I21.9 Острый инфаркт миокарда неуточнённый	Жуков А.Н.	Выписан	10	06.04.2026	1	3	
60	Ж	61	Приёмное отд.	I22.0 Повторный инфаркт передней стенки	Григорьев С.А.	Выписан	7	04.04.2026	1	3	
61		6	стационаре)]				0	6	16		
62							0	6	15		
63							0	6	13		

Рисунок 3.10. Формула СУММЕСЛИ с пояснением: проверяем отделение, суммируем койко-дни

Если нужно посчитать койко-дни по нескольким условиям, например по кардиологическому отделению и только за март, добавляем второе условие через СУММЕСЛИМН:

=СУММЕСЛИМН(РеестрПациентов[Дни в стационаре];РеестрПациентов[Отделение];«Кардиологическое отд.»;РеестрПациентов[Номер месяца];3)

СРЗНАЧЕСЛИ и СРЗНАЧЕСЛИМН: усредняем целенаправленно

Средние величины с условиями работают по тому же принципу. СРЗНАЧЕСЛИ вычисляет среднее арифметическое для ячеек, соответствующих одному условию. СРЗНАЧЕСЛИМН – нескольким условиям.

Синтаксис СРЗНАЧЕСЛИМН аналогичен СУММЕСЛИМН:

=СРЗНАЧЕСЛИМН(диапазон_усреднения; диапазон_условия1; критерий1; диапазон_условия2; критерий2; ...)

Кейс: средняя температура у пациентов с гриппом

Представьте, что мы анализируем данные по пациентам с гриппом, поступив-

шим в инфекционное отделение. В таблице есть столбцы: диагноз, температура при поступлении, возраст, длительность лихорадки. Нам нужно рассчитать среднюю температуру у пациентов с подтвержденным гриппом.

Формула с одним условием:

=СРЗНАЧЕСЛИ(Таблица1[Диагноз]; «грипп*»; Таблица1[Температура])

Обратите внимание на звездочку в критерии – она позволяет учесть все варианты написания: «Грипп», «Грипп А», «Гриппозная инфекция».

Если нужно уточнить: средняя температура только у взрослых пациентов (возраст ≥ 18) с гриппом, добавляем второе условие:

=СРЗНАЧЕСЛИМН(Таблица1[Температура]; Таблица1[Диагноз]; «грипп*»; Таблица1[Возраст]; «>=18») (рис. 3.11).

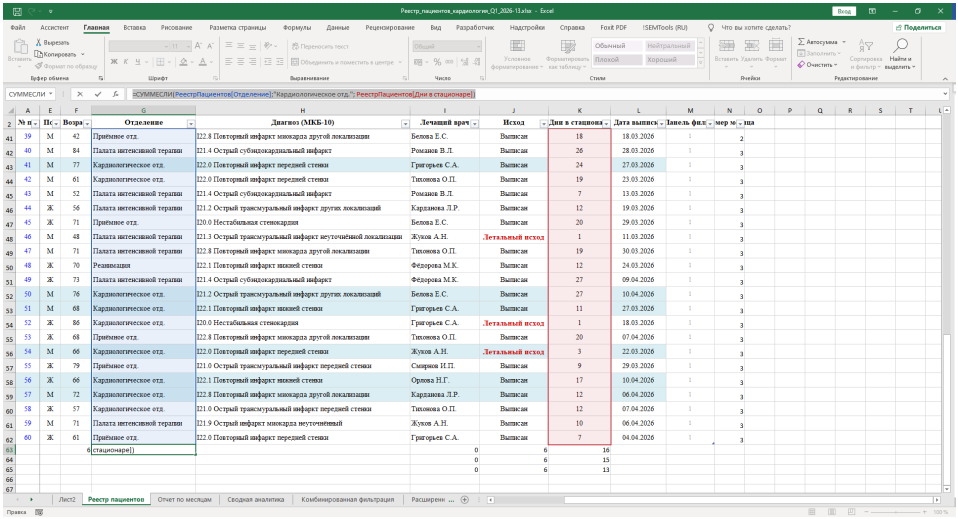


Рисунок 3.11. Формула СРЗНАЧЕСЛИМН с двумя условиями, результат 38,5 °C

Практический пример: сравнительный анализ

С помощью СРЗНАЧЕСЛИМН можно сравнивать показатели между группа-

ми. Например, средняя длительность лихорадки у пациентов с гриппом, получавших противовирусную терапию, и у тех, кто ее не получал. Это требует создания отдельного столбца «Терапия» (Да/Нет) и двух формул:

С терапией: =СРЗНАЧЕСЛИМН(Таблица1[Длительность_лихорадки]; Таблица1[Диагноз]; «грипп*»; Таблица1[Терапия]; «Да»)

Без терапии: =СРЗНАЧЕСЛИМН(Таблица1[Длительность_лихорадки]; Таблица1[Диагноз]; «грипп*»; Таблица1[Терапия]; «Нет»)

Сравнение результатов позволяет оценить эффективность терапии.

Комбинирование функций с условиями: расширяем возможности

Функции с условиями можно комбинировать между собой и с другими функциями Excel, создавая сложные аналитические конструкции.

Пример: доля пациентов в возрастной группе

Чтобы посчитать долю пациентов старше 60 лет среди всех пациентов с гипертонией, используем комбинацию:

=СЧЕТЕСЛИМН(Таблица1[Диагноз]; «Гипертоническая болезнь»; Таблица1[Возраст]; «>60») / СЧЕТЕСЛИ(Таблица1[Диагноз]; «Гипертоническая болезнь»)

Умножаем на 100%, получаем процентное соотношение (рис. 3.12).

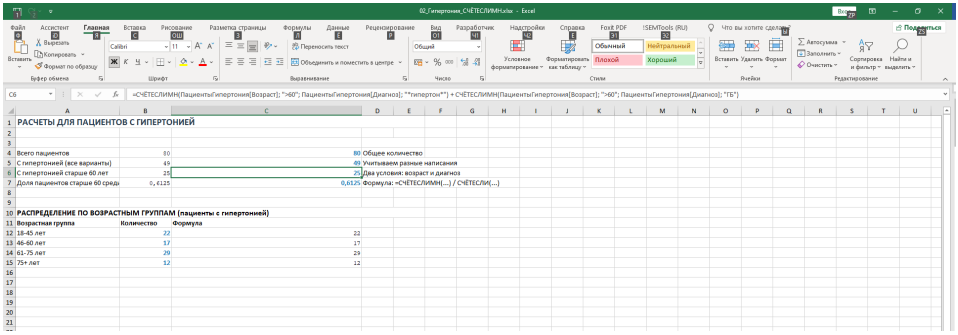


Рисунок 3.12. Доля пациентов в возрастной группе

Эта матрица обновляется автоматически при добавлении новых данных и позволяет отслеживать изменения средних показателей по месяцам.

Когда использовать сводные таблицы, а когда формулы с условиями

У начинающих аналитиков часто возникает вопрос: «Что выбрать: сводную таблицу или функции с условиями?» Ответ зависит от задачи.

Сводные таблицы лучше, когда:

- нужно быстро исследовать данные, перебирая разные варианты группировки;
- требуется интерактивность (срезы, возможность менять структуру отчета);
- нужно представить результат в виде наглядного графика;
- данные обновляются редко, отчет создается один раз.
- Формулы с условиями лучше, когда:
- результат нужно встроить в сложную формулу или в итоговую панель показателей;
- требуется автоматическое обновление без дополнительных действий (обновление сводной таблицы – лишнее движение);
- данные обновляются часто, нужен мгновенный пересчет;
- нужно получить единичное значение, а не целый отчет.

В идеале нужно владеть обоими инструментами и выбирать тот, который лучше подходит для конкретной задачи.

Практический совет: создаем панель ключевых показателей

Одно из лучших применений формул с условиями – создание панели ключевых показателей (KPI Dashboard). На отдельном листе размещаются ячейки с формулами, которые считают самые важные метрики:

- общее количество пролеченных пациентов (с учетом фильтра по периоду);

- количество пациентов в критических возрастных группах;
- общий койко-день по отделениям;
- среднюю длительность лечения;
- долю летальных исходов;
- количество пациентов с определенными диагнозами.

Все эти показатели считаются через СЧЕТЕСЛИМН, СУММЕСЛИМН и СРЗНАЧЕСЛИМН. Панель обновляется мгновенно при изменении исходных данных и дает руководству быстрый ответ на самые частые вопросы.

Главный вывод этой главы можно сформулировать так: формулы с условиями – это скальпель аналитика. Они позволяют делать точные, целенаправленные расчеты, не отвлекаясь на построение громоздких отчетов. Освоив СЧЕТЕСЛИМН, СУММЕСЛИМН и СРЗНАЧЕСЛИМН, вы сможете отвечать на сложные вопросы руководства одной формулой, экономя время и исключая ошибки ручного подсчета.

Глава 3.3. Вертикальный поиск: искусство связывать таблицы

До сих пор мы работали с одной таблицей. Но в реальной жизни врача-статистика данные редко лежат в одном месте. Выгрузка из МИС содержит номера историй болезни, но не фамилии врачей – они в другом справочнике. Табель учета рабочего времени содержит табельные номера сотрудников, а их Ф.И.О. – в отдельной таблице кадров. Реестр пациентов содержит коды диагнозов по МКБ-10, а их расшифровка – в классификаторе. Как соединить эти данные? Как заставить Excel найти нужную информацию в одной таблице на основе ключа из другой?

Ответ – функции вертикального поиска. Это семейство инструментов, которые позволяют подтягивать данные из одной таблицы в другую по общему ключу. Без них аналитик был бы обречен на бесконечное копирование и вставку, ручной поиск и неизбежные ошибки. С ними связывание таблиц становится делом нескольких секунд.

В этой главе мы разберем три подхода к вертикальному поиску: классиче-

ский ВПР (VLOOKUP), современный ПРОСМОТРХ (XLOOKUP) и универсальную связку ИНДЕКС + ПОИСКПОЗ. Каждый имеет свои сильные стороны, и, освоив все, вы сможете решать любые задачи по объединению данных.

1. Зачем связывать таблицы: жизнь без ВПР – ад

Представьте себе типичную ситуацию. Вы получили из бухгалтерии табель учета рабочего времени врачей. В нем есть даты, табельные номера, количество отработанных часов. Но для отчета главному врачу нужны Ф.И.О. и названия отделений. Эти данные есть в кадровой справке – отдельной таблице, где каждому табельному номеру соответствуют Ф.И.О. и отделение. Как их соединить?

Вариант первый (варварский): распечатать обе таблицы, вооружиться лупой и вручную вписать фамилии в табель. 100 строк – час работы. 1000 строк – день работы. А если данные обновляются каждый месяц?

Вариант второй (цивилизованный): использовать ВПР. Одна формула – и фамилии подтягиваются автоматически. Изменяются исходные данные – обновляется результат. Никакой рутины, никаких ошибок.

Именно ради таких задач и существуют функции вертикального поиска. Они экономят часы ручной работы и позволяют работать с данными так, как это задумано: таблицы хранятся отдельно, а связываются на лету по ключевым полям.

2. ВПР (VLOOKUP): классика жанра

ВПР (вертикальный просмотр) – это самая известная функция поиска в Excel. Она ищет значение в первом столбце указанного диапазона и возвращает значение из того же строки в указанном столбце.

Синтаксис ВПР:

=ВПР(искомое_значение; таблица_для_поиска; номер_столбца; [интервальный_просмотр])

- Искомое_значение – то, что мы ищем (например, табельный номер).
- Таблица_для_поиска – диапазон или имя таблицы, где мы ищем. Важно: искомое значение должно находиться в первом столбце этого диапазона.
- Номер_столбца – номер столбца в таблице_для_поиска (начиная с 1), из которого нужно вернуть значение.

- Интервальный_просмотр – ЛОЖЬ (0) для точного поиска, ИСТИНА (1) для приблизительного. В 99% медицинских задач используем ЛОЖЬ.

Кейс: подтягиваем Ф.И.О. врачей в табель учета времени

У нас есть две таблицы.

- Табель (умная таблица «Табель»): столбцы: Дата, Табельный_номер, Часы.

- Справочник_сотрудников (умная таблица «Сотрудники»): столбцы: Табельный_номер, Ф.И.О., Отделение, Должность.

Задача: в табеле добавить столбец с Ф.И.О. врача, подтянутыми из справочника по табельному номеру.

Формула в столбце Ф.И.О таблицы «Табель»:

=ВПР([@Табельный_номер]; Сотрудники; 2; ЛОЖЬ) (рис. 3.13).

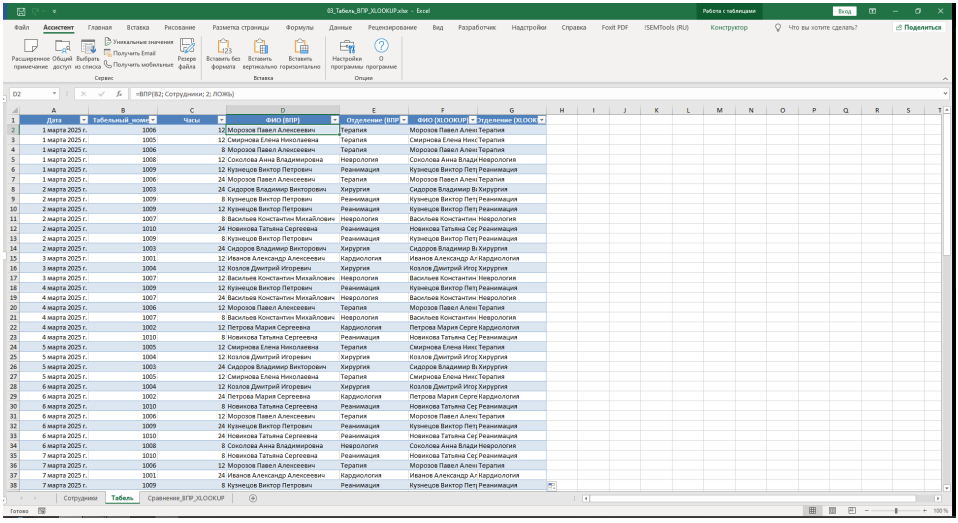


Рисунок 3.13. Формула ВПР в строке формул с пояснением каждого аргумента

Разберем

- [@Табельный_номер] – ищем значение из текущей строки столбца «Табельный_номер».

- Сотрудники – ищем в таблице «Сотрудники». Excel автоматически подставляет имя таблицы, если она преобразована в умную.

- 2 – возвращаем значение из второго столбца таблицы «Сотрудники» (это столбец Ф.И.О.).

- ЛОЖЬ – ищем точное совпадение.

Формула протягивается вниз автоматически, если таблица умная. Каждому табельному номеру сопоставляется фамилия врача.

Подводные камни ВПР, которые нужно знать

1. Искомое значение должно быть в первом столбце таблицы_для_поиска. Если ваши ключи находятся не в первом столбце, ВПР не сработает. Это самое частое ограничение, из-за которого аналитики переходят на другие функции.

2. Типы данных должны совпадать. Если в одной таблице табельный номер – число, а в другой – текст (например, «00123» вместо 123), ВПР не найдет совпадения. Перед поиском убедитесь, что типы данных одинаковы.

3. ВПР ищет только первое совпадение. Если в таблице несколько строк с одинаковым ключом, ВПР вернет значение из первой найденной. В справочниках это обычно не проблема (ключи уникальны), но в других сценариях нужно учитывать.

4. Ошибка #Н/Д означает, что значение не найдено. Это может быть реальным отсутствием данных или несовпадением типов. Не игнорируйте такие ошибки, разбирайтесь в причине.

5. При копировании формулы не забывайте про абсолютные ссылки. Если таблица_для_поиска задана как диапазон (например, $A$2:$B$100), используйте $, чтобы диапазон не съезжал при копировании. В случае умных таблиц эта проблема не возникает: имя таблицы всегда абсолютно.

[Место для рисунка: пример ошибки #Н/Д с пояснением возможных причин]

3. ПРОСМОТРХ (XLOOKUP): современная замена ВПР

В новых версиях Excel (Microsoft 365, Excel 2021 и выше) появилась функция ПРОСМОТРХ (XLOOKUP), которая решает почти все недостатки ВПР. Она проще, гибче и мощнее.

Синтаксис ПРОСМОТРХ:

=ПРОСМОТРХ(искомое_значение; массив_поиска; массив_возврата; [значение_при_отсутствии]; [режим_сопоставления]; [режим_поиска])

- Искомое_значение – то, что ищем.
- Массив_поиска – столбец или строка, где ищем.
- Массив_возврата – столбец или строка, откуда возвращаем значение.
- Значение_при_отсутствии – опционально, что показывать, если значение не найдено (например, «Не найден»).
- Остальные параметры для продвинутых сценариев, в простых случаях не нужны.

Преимущества XLOOKUP перед ВПР

1. Не требует, чтобы искомое значение было в первом столбце. Вы можете искать в любом столбце и возвращать из любого другого.
2. Может искать слева направо и справа налево. ВПР ищет только вправо, XLOOKUP – в любом направлении.
3. Простой синтаксис. Не нужно считать номера столбцов, просто указываете диапазон для поиска и диапазон для возврата.
4. Встроенная обработка ошибок. Можно задать текст или значение, которое будет показано, если совпадение не найдено.
5. Может искать не только первое, но и последнее совпадение. Через параметр режим_сопоставления.

Тот же кейс с табелем с помощью ПРОСМОТРХ:

=ПРОСМОТРХ([@Табельный_номер]; Сотрудники[Табельный_номер]; Сотрудники[ФИО]; «Не найден») (рис. 3.14).

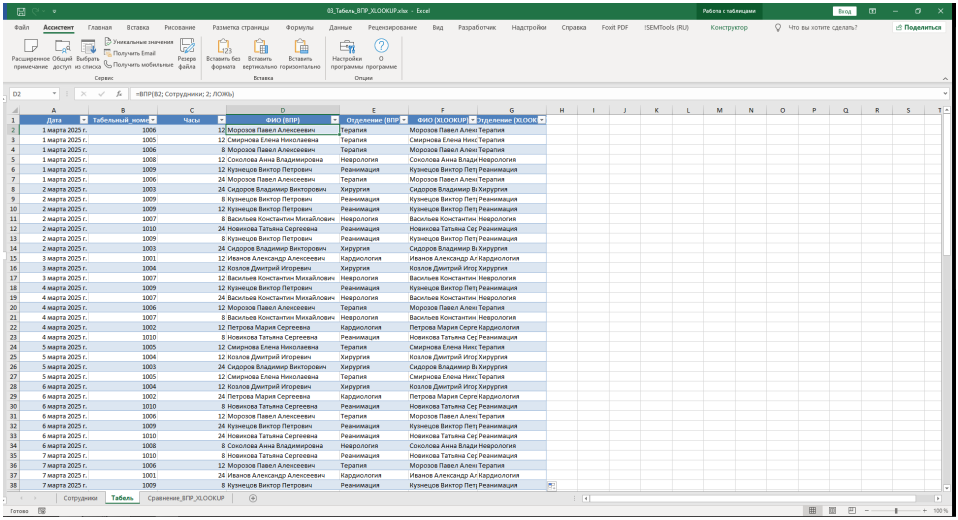


Рисунок 3.14. Формула ПРОСМОТРХ в строке формул

Формула читается почти как обычная фраза: «Найди табельный номер в столбце “Табельный_номер” таблицы “Сотрудники” и верни соответствующее значение из столбца “Ф.И.О.”. Если не найдешь – покажи “Не найден”».

Если в вашей организации используется Microsoft 365 или Excel 2021, я настоятельно рекомендую использовать ПРОСМОТРХ вместо ВПР. Она проще, надежнее и меньше подвержена ошибкам.

4. ИНДЕКС + ПОИСКПОЗ: для сложных случаев

До появления ПРОСМОТРХ связка ИНДЕКС + ПОИСКПОЗ была «золотым стандартом» для сложного поиска. Она работает в любых версиях Excel и остается актуальной, особенно когда нужно решить задачу, с которой не справляются ни ВПР, ни ПРОСМОТРХ (например, поиск по двум критериям одновременно).

Как это работает

- ПОИСКПОЗ находит позицию (номер строки) искомого значения в массиве.

- ИНДЕКС возвращает значение из указанного диапазона по номеру строки (и столбца).

Синтаксис:

=ИНДЕКС(массив_возврата; ПОИСКПОЗ(искомое_значение; массив_поиска; 0)) (рис. 3.15).

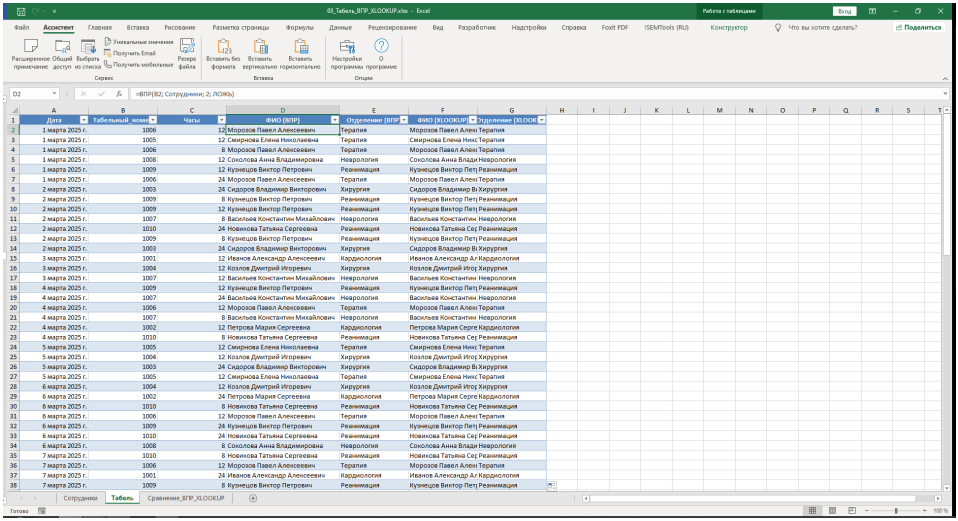


Рисунок 3.15. Схема работы связки ИНДЕКС+ПОИСКПОЗ с пояснением этапов

Кейс: поиск с двумя условиями

Представьте более сложную задачу. У нас есть таблица с результатами лабораторных исследований: Ф.И.О. пациента, дата исследования, показатель (например, уровень глюкозы). Нужно найти уровень глюкозы для конкретного пациента на конкретную дату. Здесь одного ключа недостаточно, нужно искать по двум условиям: Ф.И.О. и дата.

Связка ИНДЕКС + ПОИСКПОЗ с использованием массива условий решает эту задачу:

=ИНДЕКС(ЛабИсследования[Значение]; ПОИСКПОЗ(1; (ЛабИсследования[Ф.И.О._пациента]=B3)×(ЛабИсследования[Дата_исследования] =B4)×(ЛабИсследования[Показатель]=B5); 0)) (рис. 3.16).

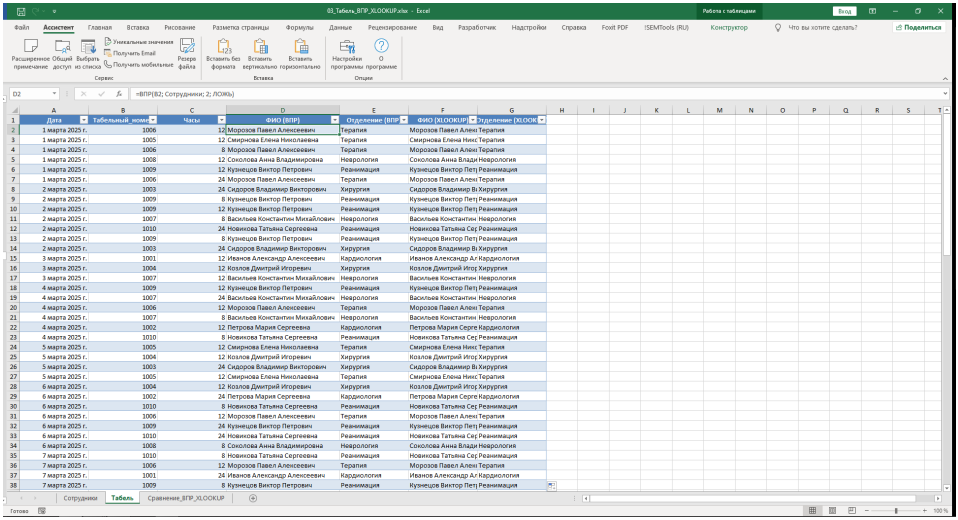


Рисунок 3.16. Формула ИНДЕКС+ПОИСКПОЗ с двумя условиями

В новых версиях Excel эту задачу можно решить и через ПРОСМОТРХ с конкатенацией, но связка ИНДЕКС+ПОИСКПОЗ остается надежным классическим решением, работающим во всех версиях.

Другой сценарий: поиск влево

Классический ВПР не умеет искать значения слева от ключа. Связка ИНДЕКС+ПОИСКПОЗ умеет. Если ключ находится не в первом столбце, а возвращаемое значение слева от него, используем:

=ИНДЕКС(Таблица[Столбец_слева]; ПОИСКПОЗ(искомое; Таблица[Ключевой_столбец]; 0)).

5. Сравнение трех подходов: что и когда использовать (табл. 3.1).

Таблица 3.1. Сравнение трех подходов

Функция	Когда использовать	Преимущества	Недостатки
ВПР	Классические задачи, совместимость со старыми версиями Excel	Работает везде, все знают	Ищет только вправо, нужно считать номера столбцов, нет встроенной обработки ошибок
ПРОСМОТРХ	Excel 2021, Microsoft 365 – основной выбор для новых проектов	Простой синтаксис, ищет в любом направлении, встроенная обработка ошибок, ищет по нескольким критериям	Не работает в старых версиях Excel
ИНДЕКС+ ПОИСКПОЗ	Сложные сценарии, поиск по двум условиям, работа в старых версиях	Максимальная гибкость, работает везде	Более сложный синтаксис, два вложенных вызова

6. Практический совет: создаем единую систему связанных таблиц

В медицинской организации данные часто хранятся в виде набора взаимосвязанных таблиц.

- Справочник сотрудников (табельный номер, Ф.И.О., отделение, должность).
- Справочник отделений (код отделения, название, профиль, заведующий).
- Справочник диагнозов (код МКБ-10, название, группа заболеваний).
- Реестр пациентов (номер истории, Ф.И.О., дата поступления, отделение, лечащий врач, диагноз).

Вместо того чтобы хранить в реестре пациентов полные названия отделений и фамилии врачей (что приводит к избыточности и ошибкам при переименованиях), мы храним только коды (идентификаторы). А в момент анализа или формирования отчета используем ПРОСМОТРХ или ВПР, чтобы подтянуть нужные названия из справочников (рис. 3.17).

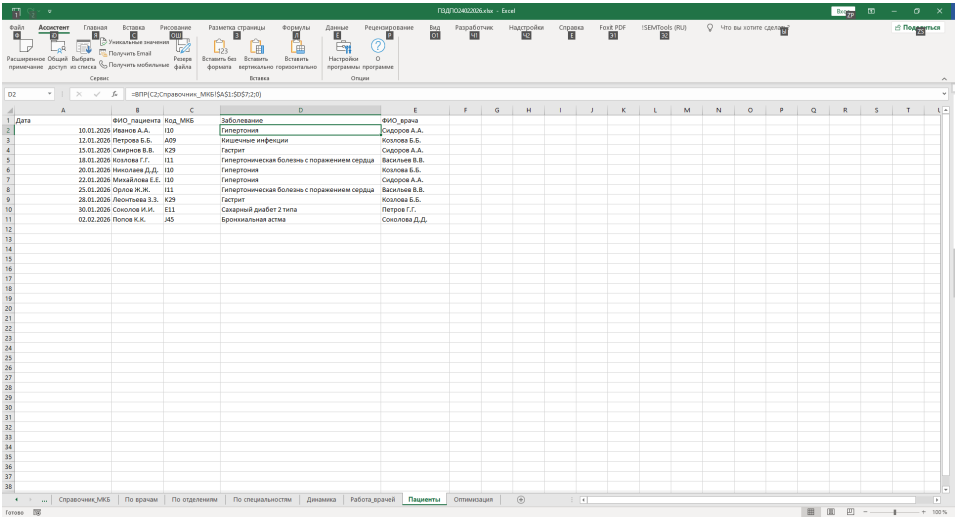


Рисунок 3.17. Схема взаимосвязи таблиц: реестр пациентов связан со справочником сотрудников, отделений и диагнозов

Такой подход называется нормализацией данных. Он обеспечивает:

- целостность: данные хранятся в одном месте, исправлять ошибки нужно только один раз;
- экономию места: вместо длинных названий хранятся короткие коды;
- гибкость: при изменении названия отделения достаточно обновить одну строку в справочнике, а не тысячи записей в реестре.

Главный вывод этой главы можно сформулировать так: умение связывать таблицы превращает разрозненные справочники и выгрузки в единую информационную систему. Освоив ВПР, ПРОСМОТРХ и ИНДЕКС+ПОИСКПОЗ, вы перестанете тратить время на ручное копирование данных и сможете работать

с медицинской информацией так, как это задумано в связанной, согласованной, легко обновляемой структуре.

Глава 3.4. Динамические диапазоны и структурированные ссылки: как сделать формулы умными

Мы прошли большой путь. Научились превращать хаотичные выгрузки в умные таблицы, удалять дубликаты, исправлять типы данных, контролировать качество с помощью условного форматирования. Освоили сводные таблицы, которые позволяют группировать и анализировать данные за секунды. Научились использовать СЧЕТЕСЛИМН, СУММЕСЛИМН и СРЗНАЧЕСЛИМН для точечных расчетов. Овладели искусством связывать таблицы через ВПР, ПРОСМОТРХ и ИНДЕКС+ПОИСКПОЗ.

Но есть один важный момент, который превращает хорошую аналитическую работу в отличную. Это умение делать формулы динамическими, такими, которые автоматически адаптируются к изменению объема данных, не требуя ручной корректировки каждый месяц.

Представьте себе ситуацию. Вы создали сложную панель показателей: средняя длительность лечения по отделениям, доля летальных исходов, количество пациентов в возрастных группах. Все формулы настроены, все работает. Но проходит месяц, вы загружаете новые данные, и оказывается, что часть формул съехала, потому что строк стало больше, а диапазоны остались старыми. Знакомо?

В этой главе мы разберемся, как сделать так, чтобы этого больше не происходило. Познакомимся со структурированными ссылками умных таблиц, научимся создавать динамические именованные диапазоны и поймем, почему Таблица1[Койко-дни] – это гораздо лучше, чем B2:B1000.

7. Проблема статических диапазонов: почему формулы ломаются

Когда мы только начинаем работать в Excel, нас учат записывать формулы так: =СРЗНАЧ(B2:B100). Все просто и понятно. Но у этого подхода есть три серьезные проблемы:

Проблема 1: новые строки не учитываются. Если в следующем месяце у вас будет не 99 записей, а 150, формула =СРЗНАЧ(B2:B100) продолжит считать

только первые 99. Новые пациенты не попадут в расчет. Вы этого можете даже не заметить: цифры изменятся незначительно, но будут уже неверными.

Проблема 2: пустые строки внизу искажают результат. Если вы предусмотрительно расширили диапазон до B2:B2000, но реальных данных только 500 строк, формула будет учитывать 1500 пустых ячеек (которые Excel считает как нули), искажая среднее значение.

Проблема 3: ошибки при вставке и удалении строк. Если вы вставите новую строку в середину диапазона, Excel в большинстве случаев корректно обновит ссылку. Но, если удалите строку, может возникнуть ошибка #ССЫЛКА! А если скопируете формулу на другой лист, ссылки могут съехать непредсказуемо.

Выход из этой ситуации: использовать динамические диапазоны, которые автоматически расширяются и сужаются вместе с данными.

8. Структурированные ссылки умных таблиц: элегантное решение

Самый простой и эффективный способ сделать формулы динамическими – использовать умные таблицы (Smart Tables), о которых мы говорили в главе 1.1. Когда вы превращаете диапазон в умную таблицу (Ctrl + T), Excel присваивает ей имя (по умолчанию «Таблица1», «Таблица2» и т.д.) и начинает работать со структурированными ссылками.

Вместо B2:B100 вы пишете Таблица1[Койко-дни]. Что это дает? (табл. 3.2)

- Автоматическое расширение. Когда вы добавляете новую строку в конец таблицы, Excel включает ее в диапазон. Формула =СРЗНАЧ(Таблица1[Койко-дни]) автоматически начнет учитывать новые данные.

- Читаемость. Таблица1[Койко-дни] гораздо понятнее, чем B2:B100. Глядя на формулу, вы сразу понимаете, какие данные обрабатываются.

- Устойчивость к изменениям. При вставке и удалении строк структурированные ссылки не ломаются. Они всегда указывают на нужный столбец в нужной таблице.

- Автозаполнение. Когда вы начинаете вводить формулу, Excel предлагает имена таблиц и столбцов — не нужно запоминать адреса.

Таблица 3.2. Как это выглядит на практике

Вместо	Пишем
=СУММ(B2:B100)	=СУММ(Таблица1[Койко-дни])
=СРЗНАЧ(C2:C100)	=СРЗНАЧ(Таблица1[Возраст])
=СЧЕТЕСЛИ(D2:D100; «Кардиология»)	=СЧЕТЕСЛИ(Таблица1[Отделение]; «Кардиология»)
=ВПР(A2; Справочник!$A$2:$B$100; 2; 0)	=ПРОСМОТРХ([@Табельный_номер]; Сотрудники[Табельный_номер]; Сотрудники[Ф.И.О.])

Обратите внимание на последний пример: [@Табельный_номер] – это ссылка на значение в текущей строке умной таблицы. Она работает как относительная ссылка, но выглядит понятнее и не съезжает при копировании.

Специальные обозначения в структурированных ссылках

Таблица1[Столбец] – весь столбец таблицы (все строки).

Таблица1[@Столбец] – значение из текущей строки.

Таблица1[[#Все];[Столбец]] – явное указание на весь столбец (редко требуется).

Таблица1[[#Данные];[Столбец]] – только строки с данными (без строки итогов, если она есть).

9. Кейс: автоматизация ежемесячного отчета

Давайте соберем все вместе и посмотрим, как использование умных таблиц и структурированных ссылок превращает ежемесячный отчет из рутинной работы в автоматический процесс.

У нас есть:

- ежемесячная выгрузка из МИС – файл с реестром пациентов;
- справочники: отделения, врачи, диагнозы (в отдельных файлах или на отдельных листах).

Настройка один раз

1. Превращаем выгрузку в умную таблицу (назовем ее «Реестр»).
2. Создаем лист «Отчет» с ключевыми показателями.

- Общее количество пациентов: =СЧЕТЗ(Реестр[Номер истории]).
- Общий койко-день: =СУММ(Реестр[Койко-дни]).
- Средняя длительность лечения: =СРЗНАЧ(Реестр[Койко-дни]).
- Количество по отделениям: сводная таблица на основе Реестр.
- Динамика по месяцам: сводная таблица с группировкой дат.

3. Настраиваем связи со справочниками через ПРОСМОТРХ.

Каждый месяц

4. Копируем новую выгрузку в таблицу «Реестр» (удаляем старые данные, вставляем новые).
5. Нажимаем «Обновить все» (вкладка «Данные» → «Обновить все»).
6. Готово. Все формулы и сводные таблицы пересчитаны автоматически.

Время на подготовку отчета – 5–10 мин. вместо 2–3 ч (рис. 3.18).

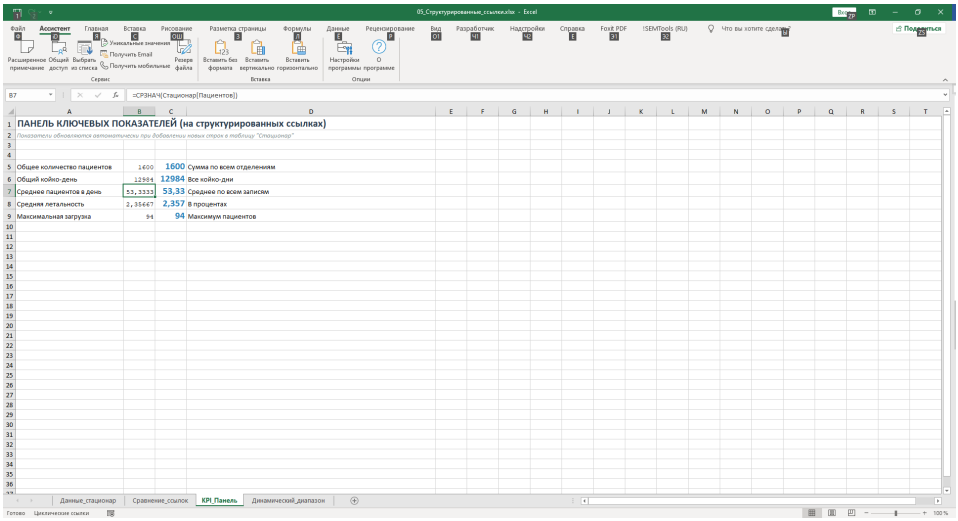


Рисунок 3.18. Схема автоматизированного отчета с указанием источников данных и направлений обновления

13. Практические советы по работе со структурированными ссылками

1. Переименовывайте таблицы. Имя «Таблица1» ничего не говорит о содержании. В свойствах таблицы (вкладка «Конструктор») задайте понятное имя: Реестр_пациентов, Справочник_отделений, Вызовы_2025.

2. Используйте автозаполнение. Когда вы начинаете вводить формулу и набираете имя таблицы, Excel предлагает список доступных столбцов. Это ускоряет работу и исключает опечатки.

3. Будьте осторожны с копированием формул. Если вы копируете формулу со структурированными ссылками на другой лист или в другую книгу, ссылки могут сломаться, если там нет таблицы с таким же именем. Используйте диспетчер имен, чтобы проверить все ссылки.

4. Проверяйте формулы с [@. Ссылки вида [@Столбец] работают только внутри умной таблицы. Если вы используете такую формулу вне таблицы, она вернет ошибку.

5. Превращайте старые файлы в умные таблицы. Если у вас есть исторические данные в формате обычных диапазонов, преобразуйте их в умные таблицы. Это займет минуту, но сэкономит часы в будущем.

14. Завершение части 3: что мы освоили

Поздравляю! Мы завершили третью часть книги, посвященную главным инструментам аналитика. Теперь в вашем арсенале:

- сводные таблицы для быстрой группировки, агрегации и анализа данных. Умеете группировать по дням недели, считать средние по отделениям, строить динамику по месяцам;
- формулы с условиями (СЧЕТЕСЛИМН, СУММЕСЛИМН, СРЗНАЧЕСЛИМН) для точечных расчетов, панелей показателей и сложной аналитики;
- вертикальный поиск (ВПР, ПРОСМОТРХ, ИНДЕКС+ПОИСКПОЗ) для связывания таблиц, подтягивания справочных данных, создания единой информационной системы;
- структурированные ссылки и умные таблицы для автоматического расширения диапазонов, устойчивости к изменениям и полной автоматизации ежемесячных отчетов.

Эти четыре инструмента – фундамент, на котором строится профессиональная работа с медицинскими данными в Excel. Освоив их, вы перестанете быть человеком с калькулятором и станете настоящим аналитиком, способным не только считать, но и видеть скрытые закономерности, формулировать выводы и обосновывать управленческие решения.

В следующей части мы перейдем к визуализации – научимся превращать цифры в понятные графики, которые можно сразу вставлять в доклад главному врачу. Но прежде чем двигаться дальше, убедитесь, что вы уверенно владеете инструментами из этой части. Они понадобятся нам снова и снова.



Часть 4. Графики для главного врача: визуализация данных

Глава 4.1. Как выбрать правильный тип графика

Представьте себе, что вы подготовили отчет по работе стационара за год. В вашем распоряжении десятки показателей: динамика госпитализаций по месяцам, структура коечного фонда, распределение пациентов по возрастам, средняя длительность лечения по отделениям, сравнение показателей с прошлым годом. Как все это представить так, чтобы главный врач увидел главное за три минуты?

Ответ начинается с выбора правильного типа графика. В Excel десятки типов диаграмм, но в реальной работе врача-статистика достаточно освоить четыре основных. Каждый из них решает свою задачу, и ошибка в выборе может сделать визуализацию бесполезной или даже вводящей в заблуждение.

В этой главе мы разберем эти четыре типа: линейчатую диаграмму (для динамики), гистограмму (для сравнения категорий), круговую диаграмму (для структуры) и точечную диаграмму (для связей). Поймем, для каких задач каждый из них предназначен, и научимся делать выбор осознанно, не потому что так красивее, а потому что так понятнее.

Принцип один: график должен отвечать на вопрос

Прежде чем открывать вкладку «Вставка» и выбирать тип диаграммы, задайте себе вопрос: какую историю я хочу рассказать с помощью этого графика? Или, еще проще: на какой вопрос руководства я хочу ответить?

Вот четыре основных типа вопросов и соответствующие им типы графиков (табл. 4.1).

Таблица 4.1. Типы графиков

Какой вопрос?	Какой график?
Как меняется показатель во времени?	Линейчатая диаграмма (график)
Как сравнить значения по категориям?	Гистограмма (столбцы)
Как распределено целое по частям?	Круговая диаграмма
Есть ли связь между двумя показателями?	Точечная диаграмма

Запомните это соответствие, и вы никогда не ошибетесь с выбором.

Линейчатая диаграмма (график): для динамики и трендов

Линейчатая диаграмма – это главный инструмент для отображения изменений во времени. Если у вас есть показатель, который измеряется в последовательные моменты времени (дни, недели, месяцы, годы), линейчатый график покажет, как он меняется, и позволит увидеть тренды, сезонность и аномалии.

Когда использовать

- Динамика числа госпитализаций по месяцам.
- Изменение средней длительности лечения в течение года.
- Динамика летальности по кварталам.
- Сравнение нескольких показателей во времени (например, число вызовов скорой и число госпитализаций).

Когда НЕ использовать

- Если категории не упорядочены во времени (например, отделения или диагнозы), для этого есть гистограмма.
- Если точек во времени слишком мало (2–3), столбцы будут нагляднее.

Как построить в Excel

Выделите данные (даты в одном столбце или строке, значения – в соседних). Вкладка «Вставка» → «График» → «С маркерами» или «Сглаженные линии».

Для временных рядов лучше использовать вариант с маркерами – так легче увидеть фактические точки.

Кейс: динамика вызовов скорой помощи по месяцам

У нас есть данные по вызовам скорой помощи за два года. Нужно показать главному врачу сезонность и сравнить с предыдущим годом. Строим линейный график: ось X – месяцы, ось Y – количество вызовов, две линии – 2024 и 2025 гг. (рис. 4.1).

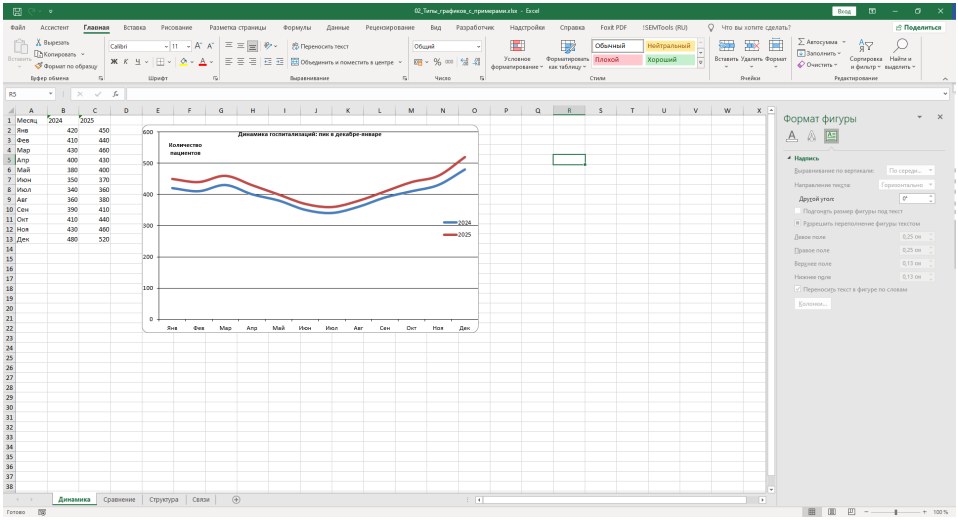


Рисунок 4.1. Линейный график с двумя линиями, показывающий рост вызовов в зимние месяцы

График наглядно демонстрирует, что пик вызовов приходится на декабрь–январь (сезон ОРВИ и гриппа), а также что в 2025 г. общее число вызовов выросло по сравнению с 2024 г. Это позволяет обосновать необходимость увеличения числа бригад в зимние месяцы.

Практические советы по настройке

- Ось X должна быть временной. Убедитесь, что Excel распознал даты именно как даты, а не как текст. Если на оси подписан каждый день, а данных много, используйте формат оси «Месяц» или «Квартал», чтобы избежать нагромождения.

- Не перегружайте. Если на одном графике больше 3–4 линий, он становится трудночитаемым. В таком случае лучше разбить на несколько графиков или использовать сводную диаграмму со срезами.

- Добавьте линию тренда. Если нужно показать общее направление (рост или снижение) на фоне колебаний, добавьте линию тренда (правой кнопкой по линии → «Добавить линию тренда» → «Линейная»).

- Гистограмма (столбцы): для сравнения категорий

Гистограмма – это универсальный инструмент для сравнения значений по категориям. Столбцы легко воспринимаются глазом, их высота позволяет мгновенно оценить, какая категория лидирует, а какая отстает.

Когда использовать

- Сравнение количества пациентов по отделениям.

- Сравнение средней длительности лечения по отделениям или врачам.

- Распределение вызовов по дням недели.

- Сравнение показателей текущего и прошлого периода (с группировкой).

Когда НЕ использовать

- Если нужно показать динамику во времени, то лучше график.

- Если категорий очень много (более 10–12), столбцы сольются, лучше использовать горизонтальную гистограмму или таблицу (табл. 4.2).

Таблица 4.2. Варианты гистограмм

Тип	Когда использовать
Обычная (кластерная)	Сравнение значений по одной категории или нескольких категорий без наложения
С накоплением	Показать не только общий объем, но и вклад каждой составляющей
Нормированная с накоплением	Показать долю каждой составляющей в процентах (100% на столбец)
Горизонтальная	Когда названия категорий длинные или их много

Кейс: сравнение средней длительности лечения по отделениям

Нужно показать заведующему стационаром, в каких отделениях средняя длительность лечения превышает нормативы. Строим обычную гистограмму: ось X – отделения, ось Y – средняя длительность лечения. Добавляем горизонтальную линию на уровне норматива (например, 10 дней) – это можно сделать через добавление ряда данных с константой (рис. 4.2).

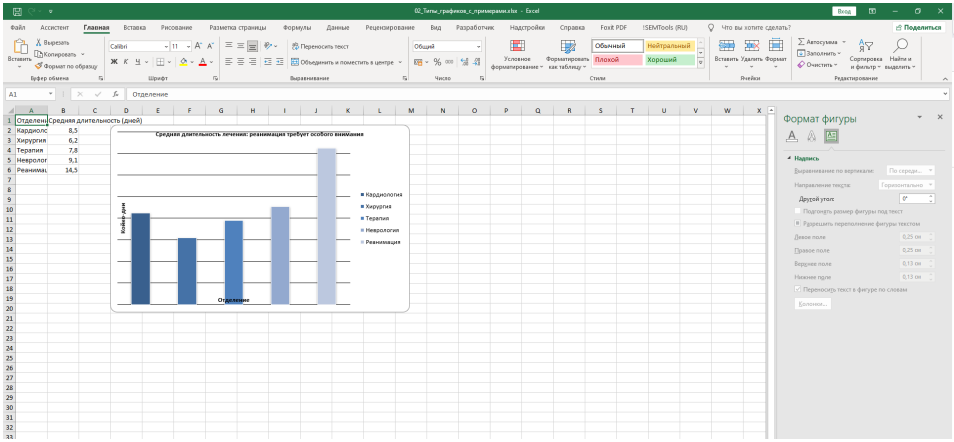


Рисунок 4.2. Гистограмма с выделенными отделениями, где средняя длительность превышает норматив

График сразу показывает, что в хирургическом и травматологическом отделениях показатели в норме, а в кардиологическом и неврологическом – превышение. Это сигнал для разбора причин.

Практические советы по настройке

- Сортируйте категории. Если порядок категорий не задан логикой (например, дни недели или месяцы), отсортируйте их по убыванию или возрастанию значения. Это делает график понятнее.

- Используйте подписи данных. Если значений немного, добавьте числовые подписи над столбцами: так не придется прицеливаться глазами к оси Y.

- Контролируйте масштаб оси Y. Ось Y должна начинаться с нуля. Если начать не с нуля, разница между столбцами визуально усилится, что может ввести в заблуждение. Excel по умолчанию делает правильно, но после ручной настройки проверьте.

- Круговая диаграмма: только для структуры и только 2–5 секторов

Круговая диаграмма – самый популярный и одновременно самый неправильно используемый тип графиков. Она хороша ровно для одной задачи: показать, как целое делится на части. Но у нее есть жесткие ограничения, которые нужно знать и соблюдать.

Когда использовать

- Структура коечного фонда по профилям.

- Распределение пациентов по возрастным группам.

- Доля летальных исходов по отделениям.

- Структура вызовов скорой по нозологиям.

Когда НЕ использовать

- Если секторов больше 5–6, диаграмма превращается в пиццу, которую невозможно прочитать.

- Если сектора примерно равны по размеру, сложно оценить, какой больше.

- Если нужно сравнить структуры за разные периоды, лучше использовать нормированную гистограмму с накоплением.

Правила работы с круговой диаграммой

1. Не больше 5 секторов. Если категорий больше, объедините мелкие доли в одну категорию «Прочие». Лучше показать 5 крупных кусков и один «Прочие», чем 10–12 тонких ломтиков, которые неразличимы.

2. Сортируйте сектора по убыванию. Самый крупный сектор начинайте с 12 ч и располагайте по часовой стрелке в порядке убывания. Так глаз сразу видит главное.

3. Используйте подписи долей. Подписывайте не только названия категорий, но и проценты. Лучше всего вынести подписи за пределы круга с линиями-выносками, чтобы не загромождать диаграмму.

4. Избегайте 3D-эффектов. Объемные круговые диаграммы искажают восприятие размеров секторов. Используйте плоский вариант.

Кейс: структура коечного фонда

Нужно показать руководству, как распределены койки по профилям.

В стационаре 7 профилей, но три из них составляют менее 5% каждый. Объединяем их в категорию «Прочие», получаем 5 секторов: терапевтический (35%), хирургический (30%), детский (18%), реанимация (10%), прочие (7%) (рис. 4.3).

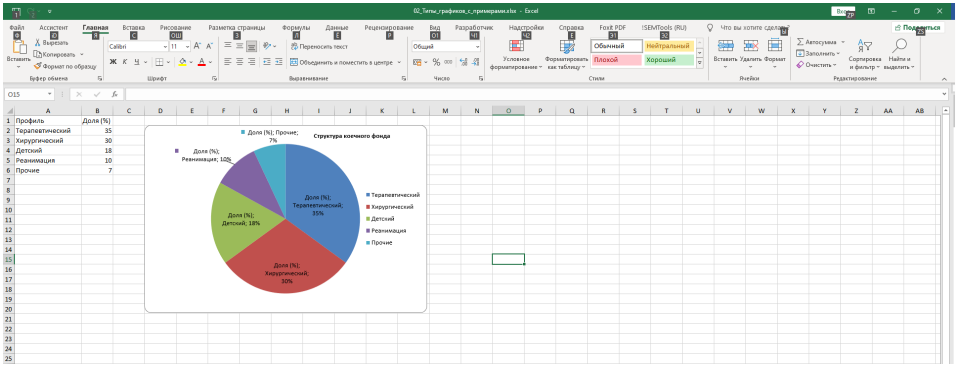


Рисунок 4.3. Круговая диаграмма структуры коечного фонда с выделенными долями

График наглядно показывает, что основа коечного фонда – терапевтический и хирургический профили. Это может быть аргументом при обсуждении перераспределения ресурсов.

Альтернатива: кольцевая диаграмма

Кольцевая диаграмма – разновидность круговой, которая позволяет разместить в центре итоговый показатель (например, общее число коек). Это удобно, когда нужно одновременно показать структуру и общий объем.

Точечная диаграмма (диаграмма рассеяния): для поиска связей

Точечная диаграмма – это инструмент для выявления взаимосвязей между двумя числовыми показателями. Каждая точка на графике соответствует одному наблюдению, ее положение определяется значениями двух показателей. Если точки выстраиваются вдоль линии, значит, есть корреляция.

Когда использовать

- Связь между возрастом пациента и длительностью лечения.
- Связь между нагрузкой на врача (число пациентов) и частотой осложнений.
- Связь между временем ожидания в приемном покое и удовлетворенностью пациентов.
- Выявление выбросов (аномальных сочетаний показателей).

Когда НЕ использовать

- Если нужно показать динамику или сравнение категорий, есть другие инструменты.
- Если оба показателя категориальные (текстовые), точечная диаграмма работает только с числами.

Как построить в Excel

Выделите два столбца с числами. Вкладка «Вставка» → «Точечная» → «Точечная с маркерами». Excel построит график, где по оси X – первый выделенный столбец, по оси Y – второй (рис. 4.4).

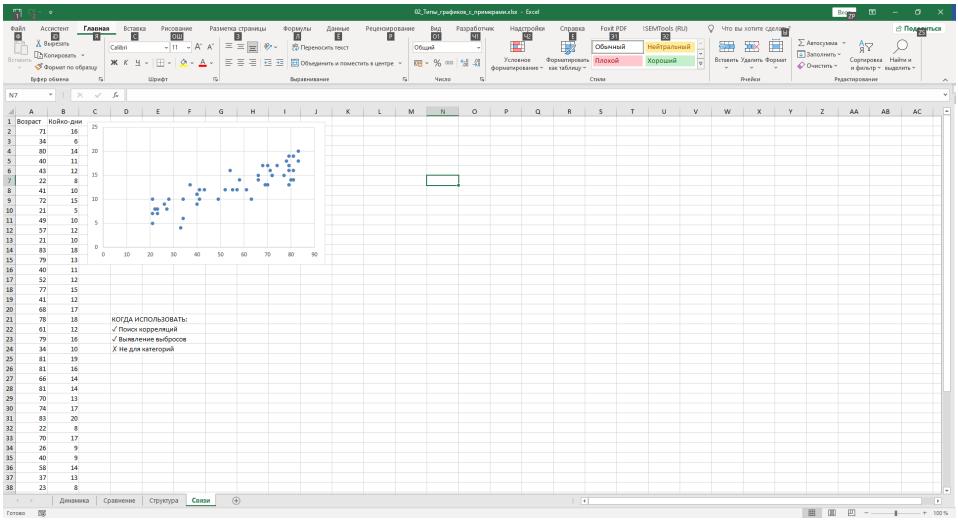


Рисунок 4.4. Точечная диаграмма с пояснением: каждая точка один пациент, ось X – возраст, ось Y – длительность лечения

Практические советы по настройке

- Добавьте линию тренда и коэффициент R^2. Линия тренда показывает общее направление, а R^2 (коэффициент детерминации) – насколько сильна связь. Чем ближе R^2 к 1, тем сильнее связь. В Excel его можно отобразить прямо на графике.

- Используйте разные цвета для групп. Если нужно сравнить связь в разных группах (например, мужчины и женщины, разные отделения), раскрасьте точки по-разному. Это делается через добавление дополнительных рядов данных.

- Масштабируйте оси. Если точки сконцентрированы в одной области, а есть выбросы, график может стать нечитаемым. Попробуйте использовать логарифмический масштаб или исключить выбросы для основного графика, показав их отдельно.

Какой график выбрать: краткий алгоритм

Вот простой алгоритм, который поможет вам быстро определиться с выбором.

1. Что я хочу показать?

- Изменение во времени → Линейчатая диаграмма (график).
- Сравнение категорий → Гистограмма (столбцы).
- Структуру целого → Круговая диаграмма (только 2–5 секторов).
- Связь между показателями → Точечная диаграмма.

2. Если выбрал круговую: сколько секторов?

- 2–5 → можно использовать.
- 6 и более → объединить мелкие в «Прочие» или использовать нормированную гистограмму с накоплением.

3. Если выбрал гистограмму: сколько категорий?

- 3–8 → обычная вертикальная.
- 8–15 → горизонтальная (чтобы подписи поместились).
- Более 15 → лучше таблица или группировка.

4. Если выбрал линейчатую: сколько линий?

- 1–3 → хорошо читается.
- 4–6 → можно, но осторожно, используйте разные типы линий и маркеры.
- Более 6 → разбить на несколько графиков.

Практический совет: не перегружайте график

Одна из самых распространенных ошибок – попытка уместить в один график все и сразу. Сложный график с множеством элементов не делает вас продвинутым аналитиком. Он делает вас непонятным.

Лучше сделать два простых графика, чем один сложный. Лучше добавить подпись или пояснение, чем заставлять зрителя догадываться. Лучше удалить лишние элементы (сетку, легенду, если она не нужна, рамки), чем загромождать пространство.

Вот чек-лист хорошего графика.

- Тип графика соответствует задаче.
- Заголовок понятен и отвечает на вопрос «Что показывает график?».
- Оси подписаны (название показателя и единицы измерения).
- Масштаб оси Y начинается с нуля (кроме точечных диаграмм).
- Цвета не мешают восприятию (нет радуги без необходимости).
- Легенда есть, если на графике больше одного ряда данных.
- Подписи данных есть, если точность важнее чистоты визуала.
- График понятен без дополнительных пояснений.

Главный вывод этой главы можно сформулировать так: выбор типа графика – это не вопрос вкуса, а вопрос точности передачи информации. Линейчатая диаграмма показывает динамику, гистограмма – сравнение, круговая – структуру, точечная – связи. Освоив эти четыре типа и научившись применять их по назначению, вы превратите свои отчеты из простыни цифр в понятные, убедительные истории, которые руководство сможет воспринять за минуты, а не за часы.

В следующей главе мы разберем практические кейсы визуализации медицинских данных, от настройки осей до подготовки графиков для презентации главному врачу.

Глава 4.2. Создание и настройка диаграмм: от цифр к убедительной презентации

Построить график в Excel – дело нескольких секунд. Сделать так, чтобы этот график был понятен, читаем и готов к вставке в доклад главному врачу или в отчет для Министерства здравоохранения, – задача, требующая внимания к деталям. Разница между сырым графиком, который Excel построил по умолчанию, и настроенной диаграммой, которая работает на вашу аргументацию, – это разница между сообщением «Вот какие-то цифры» и «Вот вывод, который нужно принять к сведению».

В этой главе мы разберемся, как настраивать диаграммы для максимальной наглядности. Научимся добавлять и форматировать подписи данных, легенду, названия осей. Поймем, как выбирать цвета и шрифты, чтобы график органично вписывался в презентацию или текстовый документ. Закрепим все это на практическом кейсе – построении графика динамики смертности за 5 лет, который можно сразу вставлять в годовой отчет.

Из чего состоит диаграмма: знакомство с элементами

Прежде чем настраивать диаграмму, важно понимать, из каких элементов она состоит и как они называются. Это поможет вам быстро находить нужные настройки (рис. 4.5).

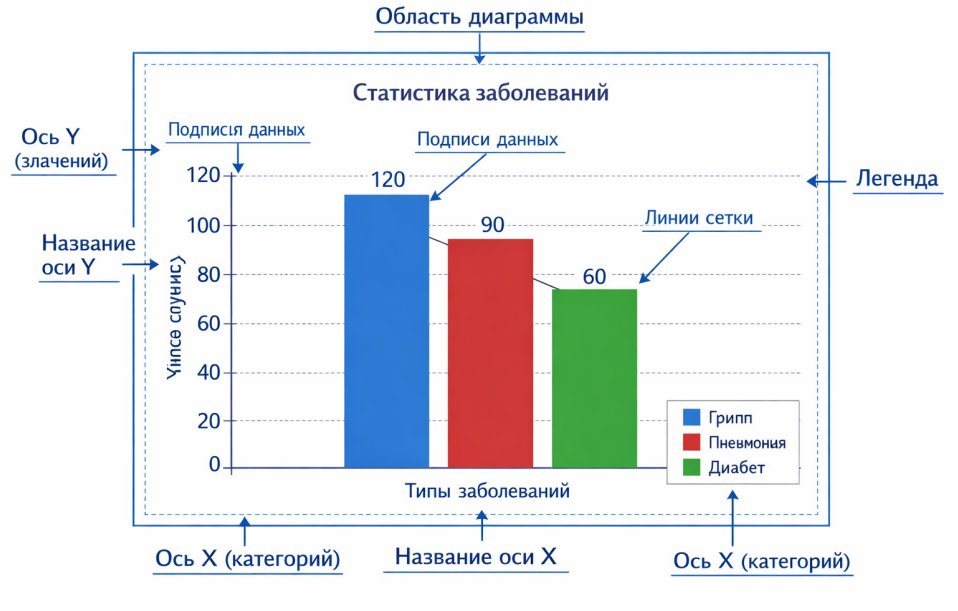


Рисунок 4.5. Схема диаграммы с подписями всех элементов: область диаграммы, область построения, заголовок диаграммы, ось X (категорий), ось Y (значений), названия осей, легенда, подписи данных, линии сетки

Основные элементы диаграммы

- Область диаграммы: весь прямоугольник, в котором находится диаграмма.
- Область построения: внутренняя часть, где отображаются столбцы, линии, точки.
- Заголовок диаграммы: название, которое объясняет, что показывает график.
- Ось X (ось категорий): обычно время или категории сравнения.
- Ось Y (ось значений): числовые значения показателя.
- Названия осей: пояснения, что отложено по осям (единицы измерения).
- Легенда: расшифровка цветов, если на графике несколько рядов данных.
- Подписи данных: числовые значения непосредственно над столбцами или линиями.
- Линии сетки: вспомогательные линии для оценки значений.

Excel добавляет не все эти элементы по умолчанию. Некоторые нужно включать вручную, некоторые настраивать. И это нормально: стандартная диаграмма – только заготовка, а ваша задача – довести ее до ума.

Добавление и настройка элементов диаграммы

Самый удобный способ управлять элементами диаграммы – использовать кнопку «Элементы диаграммы» (зеленый плюсик), которая появляется справа от диаграммы при ее выделении.

Заголовок диаграммы

По умолчанию Excel добавляет заголовок «Название диаграммы». Это текст-заготовка, который нужно заменить на содержательный. Хороший заголовок отвечает на вопрос «Что показывает график?» и формулируется как вывод, а не просто констатация факта (табл. 4.3).

Таблица 4.3. Название диаграммы

Плохо	Хорошо
«Динамика смертности»	«Смертность в стационаре снизилась на 15% за 5 лет»
«Госпитализации по месяцам»	«Пик госпитализаций приходится на декабрь–январь»
«Средняя длительность лечения»	«Средняя длительность лечения стабильна на уровне 7,2 дня»

Чтобы изменить заголовок, кликните по нему дважды или выделите и начните печатать.

Названия осей

Названия осей – это подсказка, что именно отложено по осям. Для оси Y (значений) обязательно указывайте единицы измерения: «Количество пациентов», «Койко-дни», «Смертность, %», «Длительность лечения, дни».

Добавляются названия осей через кнопку «Элементы диаграммы» → «Названия осей» → выбрать нужные (горизонтальная, вертикальная).

Легенда

Легенда нужна только тогда, когда на графике больше одного ряда данных (например, две линии – 2024 и 2025 гг., или столбцы по разным отделениям). Если ряд один, легенда занимает место и не несет полезной информации, ее лучше отключить.

Легенду можно перемещать: выделите ее, затем перетащите в удобное место (например, в верхний правый угол или под заголовок). Также можно изменить размер шрифта и цвет через контекстное меню.

Подписи данных

Подписи данных – это числовые значения прямо на элементах диаграммы (над столбцами, рядом с точками на линии). Они делают график самодостаточным: зрителю не нужно прицеливаться глазами к оси Y, чтобы оценить значение.

Добавляются через «Элементы диаграммы» → «Подписи данных». Для столбцов лучше выбрать положение «Снаружи у края», для линий — «Сверху» или «Слева».

Важное предостережение: если на графике много точек (например, ежемесячные данные за 5 лет – 60 точек), подписи данных сделают его нечитаемым. В таких случаях подписи добавляют только для ключевых точек (максимумы, минимумы, начало и конец периода) или используют подписи-выноски для отдельных значений.

Линии сетки

Линии сетки помогают глазу двигаться от точки графика к оси значений. По умолчанию Excel добавляет основные линии сетки по оси Y. Их можно убрать или добавить дополнительные (промежуточные) через «Элементы диаграммы» → «Линии сетки».

Совет: для презентаций часто имеет смысл убрать линии сетки, оставив только оси и подписи данных, – график становится чище. Для отчетов в Word линии сетки могут быть полезны для точной оценки значений.

Форматирование цветов и шрифтов: готовим график к вставке в презентацию

График, который выглядит хорошо на экране монитора, может потеряться в презентации, если не учесть особенности проектора, цветовой гаммы корпоративного шаблона и читаемости шрифтов.

Цвета: меньше – лучше

Стандартная палитра Excel часто бывает слишком яркой и пестрой. Для презентаций и официальных отчетов лучше использовать спокойную, профессиональную гамму.

- Для линейчатых графиков: один цвет (например, синий) для линии, другой (серый) – для области под линией, если она есть.
- Для гистограмм с группировкой: 2–3 цвета, которые хорошо сочетаются. Избегайте радуги из 5–6 цветов.
- Для круговых диаграмм: используйте оттенки одного цвета или корпоративную палитру учреждения.

Как изменить цвет

1. Выделите элемент диаграммы (столбец, линию, сектор).
2. На вкладке «Формат» или через контекстное меню выберите «Заливка»

или «Цвет контура».

3. Выберите подходящий цвет. Если в организации есть корпоративные цвета (например, в логотипе), используйте их.

Шрифты: читаемость прежде всего

В презентациях и отчетах шрифт должен быть крупным и четким. Основные правила

- Заголовок диаграммы: 14–16 pt, полужирный.
- Названия осей и легенда: 10–12 pt.
- Подписи данных: 9–10 pt, не полужирный, чтобы не перегружать.
- Тип шрифта: используйте тот же, что и в основном документе (чаще всего Calibri, Arial или Times New Roman).

Удаление лишнего: принцип минимализма

Стандартная диаграмма Excel содержит много элементов, которые не всегда нужны.

- Рамка вокруг области диаграммы: часто не нужна, ее можно убрать (Формат → Контур фигуры → Нет контура).
- Фон области построения: по умолчанию серый с белой заливкой. Для презентаций лучше сделать его полностью белым или прозрачным.
- Горизонтальная линия на уровне нуля: иногда лишняя.
- Маркеры на линии: если точек много, маркеры могут мешать; их можно убрать или оставить только для ключевых точек.

Правило: уберите все, что не помогает понять данные. Каждый элемент графика должен иметь функцию.

Экспорт графика для вставки в Word или PowerPoint

Самый простой способ перенести график из Excel в другой документ: скопировать и вставить. Но есть нюансы (табл. 4.4).

Таблица 4.4. Способы вставки

Способ вставки	Что происходит	Когда использовать
Вставить как изображение	График становится картинкой, не обновляется при изменении данных	Финальная версия отчета, когда данные больше не меняются
Вставить с сохранением связи	График обновляется при изменении исходных данных в Excel	Черновики, интерактивные документы
Вставить как объект Excel	График можно редактировать двойным кликом в Word/PowerPoint	Когда нужно поправить форматирование уже после вставки

Для финального отчета главному врачу рекомендуется вставлять график как изображение (PNG или просто как картинку). Так вы гарантируете, что график будет выглядеть одинаково на любом компьютере и не поедет при открытии документа.

Чтобы сохранить график как изображение

- Правой кнопкой по диаграмме → «Сохранить как рисунок».
- Выберите формат PNG (лучше для презентаций) или JPEG.
- **Кейс: построить график динамики смертности за 5 лет.**

Теперь применим все, что мы узнали, к реальной медицинской задаче. У нас есть данные по летальности в стационаре за 5 лет (2021–2025). Нужно построить график, который наглядно покажет динамику, и подготовить его для вставки в годовой отчет (табл. 4.5).

Таблица 4.5. Исходные данные

Год	Летальность
2021	2,8
2022	2,5
2023	2,4
2024	2,1
2025	1,9

Шаг 1. Строим базовый график

Выделяем данные (годы и проценты). Вкладка «Вставка» → «График» → «С маркерами». Получаем базовую диаграмму.

Шаг 2. Добавляем заголовок

Клик по заголовку «Название диаграммы». Вводим содержательный заголовок, который сразу сообщает вывод: «Летальность в стационаре снизилась на 32% за 5 лет». (2,8% → 1,9% – это снижение на 0,9 процентного пункта, что в относительном выражении составляет около 32%.)

Шаг 3. Добавляем названия осей

Кнопка «Элементы диаграммы» (зеленый плюсик) → «Названия осей» → выбираем «Основная горизонтальная» и «Основная вертикальная». Подписываем:

- Ось X: «Год».
- Ось Y: «Летальность, %».

Шаг 4. Настраиваем ось Y

По умолчанию Excel мог начать ось Y не с нуля. Для графика динамики это допустимо, но нужно быть внимательным. Если начать ось с 1,5%, снижение будет выглядеть более драматичным, чем есть на самом деле. В нашем случае лучше оставить с нуля: это честнее.

Шаг 5. Добавляем подписи данных

Кнопка «Элементы диаграммы» → «Подписи данных» → выбираем положение «Сверху» (для линии лучше над точками). Теперь над каждой точкой подписано значение летальности.

Шаг 6. Настраиваем легенду

На графике один ряд данных, легенда не нужна. Убираем ее: «Элементы диаграммы» → убираем галочку с «Легенда».

Шаг 7. Настраиваем цвет и стиль линии

Выделяем линию. На вкладке «Формат» выбираем:

- Цвет линии: темно-синий (или корпоративный цвет учреждения).
- Толщина: 2,25 pt.
- Маркеры: круглые, заливка белая, контур синий.

Шаг 8. Оформляем область диаграммы.

Убираем рамку вокруг диаграммы: выделяем область диаграммы → «Формат» → «Контур фигуры» → «Нет контура». Заливку оставляем белую или прозрачную.

Шаг 9. Добавляем линию тренда (по желанию)

Чтобы подчеркнуть направление динамики, можно добавить линейную линию тренда. Правой кнопкой по линии данных → «Добавить линию тренда» → «Линейная». В параметрах линии тренда можно вывести на график уравнение и коэффициент R^2, но для презентации руководству это часто избыточно, достаточно самой линии.

Шаг 10. Финальная проверка

Проверяем график по чек-листу (рис. 4.6.).

- Заголовок содержит вывод, а не просто название.
- Оси подписаны, есть единицы измерения.
- Подписи данных присутствуют и читаются.

- Легенда убрана (один ряд данных).
- Цвета профессиональные, не кричащие.
- Рамка убрана, фон белый.
- График понятен без дополнительных пояснений.

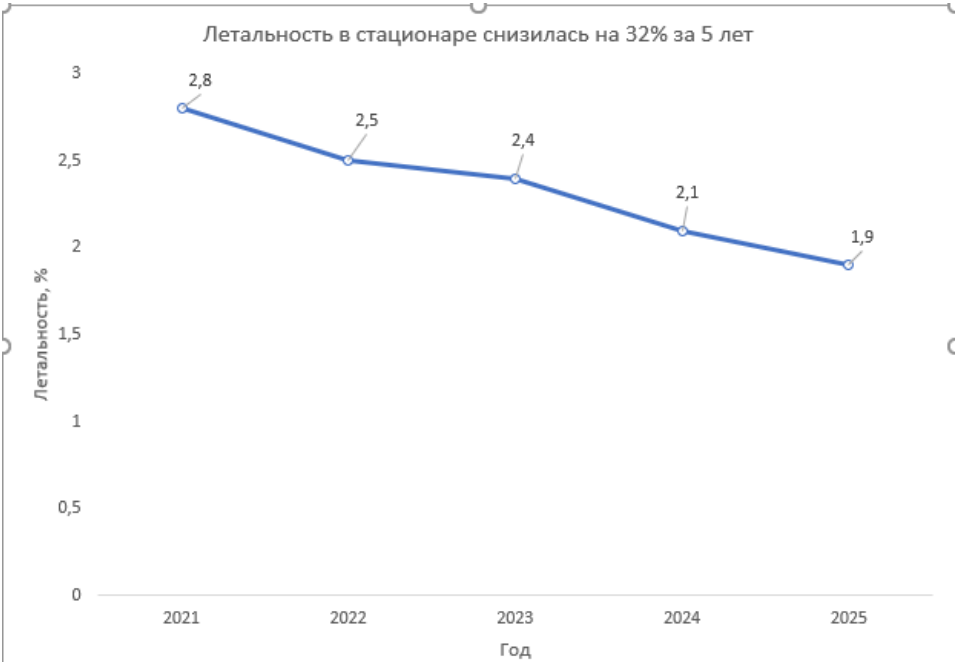


Рисунок 4.6. Финальный, полностью настроенный график, готовый к вставке в отчет

Шаг 11. Вставляем в отчет

Копируем график, переходим в Word или PowerPoint. Вставляем через «Специальная вставка» → «Рисунок (метафайл)» или «PNG» для финальной версии. Если нужно будет обновлять график позже, выбираем вариант с сохранением связи.

- Практический совет: создаем шаблон корпоративного стиля

Если вы регулярно готовите отчеты и презентации, имеет смысл создать шаблон диаграммы с уже настроенными цветами, шрифтами и стилями. Excel позволяет сохранить диаграмму как шаблон (.crtx) и затем применять его к любым новым графикам.

Как сохранить шаблон

1. Настройте одну диаграмму так, как вам нужно.
2. Правой кнопкой по диаграмме → «Сохранить как шаблон».
3. Задайте имя, например «Медицинский_отчет».

Как применить шаблон

1. Постройте новую диаграмму.
2. Правой кнопкой по диаграмме → «Изменить тип диаграммы» → вкладка «Шаблоны» → выберите свой шаблон.

Теперь каждый новый график будет сразу соответствовать корпоративному стилю, экономя вам время на форматирование.

Главный вывод этой главы можно сформулировать так: настройка диаграммы – это не второстепенная работа, а важная часть аналитического процесса. Хорошо настроенный график говорит сам за себя, экономит время руководства и делает ваши выводы убедительными. Плохо настроенный, даже с правильными цифрами, может остаться непонятым и неуслышанным. Уделите несколько минут форматированию, и ваши отчеты начнут работать на вас.

В следующей главе мы разберем продвинутые приемы визуализации: комбинированные диаграммы, использование срезов для интерактивности и создание дашбордов для оперативного мониторинга показателей.

Глава 4.3. Прогнозирование: линии тренда как взгляд в будущее

Анализ прошлого – это важно. Но руководителю нужен не только отчет о том,

что уже произошло. Ему нужно понимание того, что будет дальше. Сколько пациентов поступит в следующем году? Как изменится нагрузка на отделения? Нужно ли увеличивать финансирование или, наоборот, можно оптимизировать ресурсы? Ответы на эти вопросы лежат в области прогнозирования – умения экстраполировать прошлые тенденции в будущее.

Excel не является профессиональным инструментом для сложного прогнозирования (для этого есть SPSS, R, Python), но для большинства задач врача-статистика его возможностей вполне достаточно. Линии тренда позволяют быстро увидеть направление динамики, оценить ее силу и сделать обоснованный прогноз на ближайшую перспективу.

В этой главе мы разберемся, как добавлять линии тренда на графики, как выбирать правильный тип тренда, что означают загадочные R^2 и уравнение на графике и, самое главное, как использовать эти инструменты для обоснования управленческих решений. Все это на реальном медицинском кейсе: прогнозирование числа госпитализаций на следующий год для обоснования бюджета.

Что такое линия тренда и зачем она нужна

Линия тренда – это сглаженная линия, которая показывает общее направление изменения данных, отфильтровывая случайные колебания. Если на графике точки прыгают вверх-вниз, линия тренда помогает увидеть: а в целом-то они растут, снижаются или остаются на месте?

В работе врача-статистика линии тренда решают три задачи.

1. Визуализация динамики: показать, куда движется показатель (рост, снижение, стабильность).

2. Оценка силы связи: насколько данные соответствуют выбранной модели (через коэффициент R^2).

3. Прогнозирование: экстраполяция тренда на будущие периоды.

Важно понимать: линия тренда – это не магический кристалл. Это математическая модель, которая предполагает, что прошлые закономерности сохранятся в будущем. Чем стабильнее процесс, тем надежнее прогноз. Чем больше данных, тем точнее оценка.

Добавление линии тренда: пошаговая инструкция

Добавить линию тренда в Excel проще простого. Вот как это делается.

1. Постройте график (линейчатый, точечный или гистограмму – линия тренда доступна для большинства типов).

2. Кликните правой кнопкой мыши по линии данных (по точкам или по линии, соединяющей точки).

3. В контекстном меню выберите «Добавить линию тренда».

Справа появится панель «Формат линии тренда», где можно выбрать тип, настроить параметры и включить отображение уравнения и R^2.

Выбор типа тренда: линейный, экспоненциальный и др.

Excel предлагает несколько типов линий тренда. Выбор правильного типа – ключевой момент, от которого зависит качество прогноза.

Линейный тренд

Используется, когда данные изменяются равномерно: каждый год прирост примерно одинаковый (например, +100 пациентов в год). Линейный тренд – самый простой и наиболее часто используемый в медицинской статистике.

Формула: y = a + bx, где x – время, y – прогнозируемое значение.

Когда использовать: госпитализации, которые растут с постоянной скоростью; стабильный прирост диспансерной группы; равномерное снижение смертности.

Экспоненциальный тренд

Используется, когда данные изменяются с ускорением: каждый год прирост становится все больше (или снижение становится все медленнее). Например, если число пациентов растет не линейно, а с каждым годом все быстрее.

Формула: y = a * e^(bx).

Когда использовать: быстрорастущие эпидемии; внедрение новых методов лечения, дающее ускоряющийся эффект; рост обращений по новым профилям.

Логарифмический тренд

Используется, когда данные быстро растут в начале, а затем рост замедляется и выходит на плато. Например, внедрение скрининговой программы: в первые годы рост выявляемости высокий, затем стабилизируется.

Когда использовать: насыщение спроса; освоение новых мощностей; эпидемии, идущие на спад.

Полиномиальный тренд

Используется, когда данные имеют несколько волн (подъемов и спадов). Например, сезонные колебания с общей тенденцией. Степень полинома (2, 3, 4) показывает количество изгибов.

Когда использовать: сложные циклические процессы; данные с четко выраженной сезонностью.

Степенной тренд

Используется реже, в основном для процессов с постоянно меняющейся скоростью.

Для большинства медицинских задач врача-статистика достаточно двух типов: линейного (для равномерных изменений) и экспоненциального (для ускоряющихся процессов). Если вы не уверены, какой тип выбрать, начните с линейного: он наиболее консервативен и понятен руководству.

Отображение уравнения и величины достоверности (R^2): что это значит простыми словами

Когда вы добавляете линию тренда, Excel может показать на графике две загадочные вещи: уравнение и коэффициент R^2. Разберемся, что они означают и зачем нужны.

Уравнение линии тренда

Это математическая формула, которая описывает линию тренда. Для линейного тренда она выглядит так: $y = a + bx$. Где:

- y – прогнозируемое значение (то, что мы хотим предсказать);
- x – период времени (например, номер года);
- a – точка пересечения с осью Y (значение в нулевой точке);

- b – коэффициент наклона (показывает, на сколько в среднем изменяется показатель за один период).

Пример: если уравнение y = 1200 + 50x, это значит, что базовый уровень – 1200, а каждый следующий год приносит прирост в 50 единиц.

Коэффициент достоверности R^2 (R-квадрат)

Это показатель того, насколько точно линия тренда описывает реальные данные. R^2 принимает значения от 0 до 1.

- $R^2 = 1$ – идеальное совпадение: все точки лежат точно на линии тренда. В реальной медицинской статистике такого практически не бывает.

- $R^2 = 0{,}8–0{,}99$ – очень сильная связь: линия тренда хорошо описывает данные, прогноз будет достаточно точным.

- $R^2 = 0{,}5–0{,}79$ – умеренная связь: общая тенденция видна, но есть заметные колебания. Прогноз следует использовать с осторожностью.

- $R^2 < 0{,}5$ – слабая связь: линия тренда плохо описывает данные. Возможно, вы выбрали неправильный тип тренда или данные слишком хаотичны для прогнозирования.

Важно: R^2 – это не достоверность в строго научном смысле, а, скорее, мера соответствия. В отчетах для руководства не обязательно углубляться в математику, но полезно знать: чем ближе R^2 к 1, тем надежнее прогноз.

Как включить отображение уравнения и R^2

В панели «Формат линии тренда» внизу поставьте галочки.

- «Показывать уравнение на диаграмме».

- «Показывать величину достоверности аппроксимации (R^2) на диаграмме».

Excel поместит эти данные в текстовое поле на графике. Его можно переместить в удобное место.

Кейс: спрогнозировать число госпитализаций на следующий год, обосновать бюджет

Теперь применим все знания на практике. У нас есть данные по числу госпитализаций в стационаре за последние 6 лет (2020–2025). Нужно спрогнозировать количество госпитализаций на 2026 г. и использовать этот прогноз для обоснования бюджетной заявки (табл. 4.6).

Таблица 4.6. Исходные данные

Год	Госпитализации
2020	4200
2021	4350
2022	4500
2023	4800
2024	5100
2025	5500

Шаг 1. Строим график и добавляем линию тренда

Выделяем данные, строим линейчатый график. Добавляем линию тренда (правой кнопкой по линии → «Добавить линию тренда»).

Шаг 2. Выбираем тип тренда

Смотрим на данные: прирост ускоряется (2020→2021: +150, 2023→2024: +300, 2024→2025: +400). Это не линейный, а ускоряющийся рост. Пробуем экспоненциальный тренд. Включаем отображение уравнения и R^2.

R^2 = 0,98 – очень высокая точность. Экспоненциальная модель хорошо описывает данные.

Шаг 3. Прогнозируем на 2026 г.

В панели «Формат линии тренда» находим раздел «Прогноз». Устанавливаем значение «Вперед» на 1 период (или на количество периодов, на которые хотим спрогнозировать). Excel автоматически продолжает линию тренда и показывает прогнозное значение.

В нашем случае прогноз на 2026 г. составляет 6050 госпитализаций.

Шаг 4. Проверяем прогноз с помощью линейного тренда

(для сравнения)

Для полноты картины добавляем линейный тренд (можно добавить второй ряд данных или временно переключиться). Линейный тренд дает прогноз

5900 госпитализаций. Какой выбрать? Экспоненциальный лучше описывает ускоряющуюся динамику, но он более оптимистичен. Для бюджетного планирования можно взять среднее между двумя моделями или выбрать более консервативный вариант.

[Место для рисунка: график с двумя линиями тренда – экспоненциальной и линейной – и прогнозными значениями]

Шаг 5. Формируем аргументацию для бюджета

Теперь у нас есть прогноз. Но просто сказать: «Мы ожидаем 6050 госпитализаций», – недостаточно. Нужно обосновать, почему руководство должно выделить дополнительные средства.

Структура обоснования

1. Факты прошлого. За последние 5 лет число госпитализаций выросло на 31% (с 4200 до 5500). Прирост ускоряется: в 2024–2025 гг. он составил 400 пациентов.

2. Тренд. Анализ показывает устойчивую тенденцию к росту. Модель прогнозирования (экспоненциальная, $R^2 = 0,98$) дает высокую точность.

3. Прогноз. Ожидаемое число госпитализаций в 2026 г. – 6050 пациентов. Это на 550 больше, чем в 2025 г. (+10%).

4. Потребность в ресурсах. Рост на 550 госпитализаций потребует:

- о дополнительных койко-дней: при средней длительности лечения 7,2 дня – около 4000 дополнительных койко-дней;

- о дополнительного финансирования питания, медикаментов, коммунальных услуг;

- о возможно, расширения штата (расчет на основе норматива нагрузки на одну ставку).

Шаг 6. Представляем график

Готовим финальный график с прогнозом (рис. 4.7).

- Основные данные (2020–2025): синяя линия с маркерами.
- Прогноз на 2026 г.: пунктирная линия или столбец другого цвета.
- Линия тренда: тонкая серая линия.
- Заголовок «Прогноз госпитализаций на 2026 год: ожидаемый рост 10%».
- Подпись: *«Модель экспоненциального тренда, R^2 = 0,98 (высокая точность)»*.

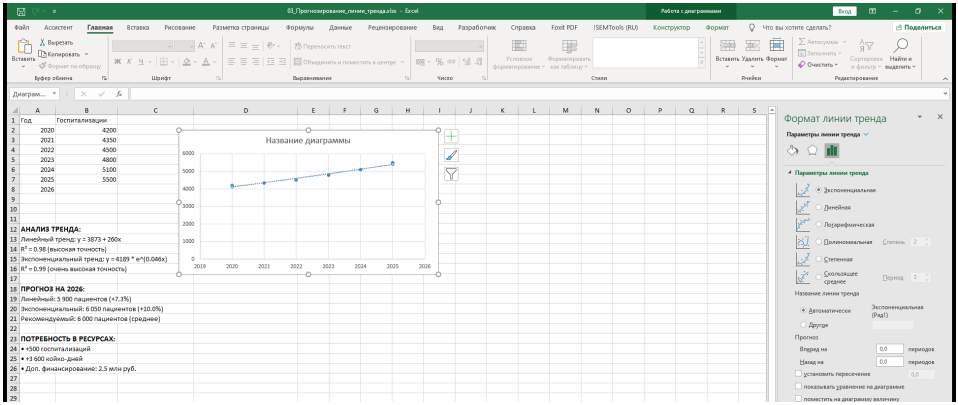


Рисунок 4.7. Финальный график с прогнозом, готовый к вставке в презентацию

Когда прогноз может не сработать: ограничения линий тренда

Линии тренда – мощный инструмент, но у них есть ограничения, о которых нужно помнить и предупреждать руководство.

1. Тренд не учитывает внешние факторы.

Если в 2026 г. произойдет эпидемия, изменится законодательство, откроется новое отделение или, наоборот, закроется больница в соседнем районе, прогноз сломается. Линия тренда знает только прошлые цифры, не причины.

2. Тренд не работает при резких изменениях.

Если в данных был единичный аномальный год (например, 2020 г. с пандемией), тренд может быть искажен. В таких случаях лучше исключить аномальный период из расчетов или использовать скользящее среднее.

3. Экстраполяция на слишком далекое будущее ненадежна.

Прогноз на 1–2 года вперед может быть достаточно точным. На 5–10 лет – это уже не прогноз, а предположение при неизменных условиях, которое редко сбывается.

4. R^2 не гарантирует, что модель правильная.

Высокий R^2 может получиться даже при неправильном выборе типа тренда, если данных мало. Всегда смотрите на график: если линия тренда идет в обход очевидных точек, возможно, тип выбран неверно.

Практический совет. Презентации для руководства всегда сопровождайте прогноз оговоркой: «Прогноз основан на данных прошлых лет и предполагает сохранение текущих условий. При изменении внешних факторов прогноз может быть скорректирован». Это профессионально и честно.

Глава 4.4. Интерактивные дашборды: объединяем всё в единую картину

Мы изучили разные типы графиков и научились их настраивать.

По отдельности каждый из них решает свою задачу: линейчатый показывает динамику, гистограмма – сравнение, круговая – структуру, точечная – связи.

Но в реальной работе врача-статистика редко бывает достаточно одного графика. Руководству нужна целостная картина: как меняются показатели во времени, как они соотносятся между собой, что происходит в разных отделениях, какие тренды требуют внимания.

Здесь на помощь приходят дашборды (панели управления) – интерактивные сборники визуализаций, которые позволяют видеть ключевые показатели в одном месте, фильтровать их по разным срезам и оперативно получать ответы на вопросы. Дашборд превращает разрозненные графики в единую историю, которую руководитель может прочитать за несколько минут.

В этой главе мы разберемся, как создавать дашборды в Excel, объединяя сводные таблицы, сводные диаграммы, срезы и временные шкалы. Научимся настраивать интерактивность, чтобы графики менялись при выборе отделения, периода или врача. И, конечно, применим все это на медицинском кейсе – создании дашборда для мониторинга работы стационара.

1. Что такое дашборд и зачем он врачу-статистику

Дашборд (от англ. dashboard – приборная панель) – это компактный, обычно одностраничный, сборник визуализаций, который позволяет отслеживать ключевые показатели в реальном времени или в динамике. Хороший дашборд отвечает на три вопроса

- Что происходит сейчас? (Текущие значения показателей.)
- Как это меняется? (Динамика, тренды.)
- Где проблемы? (Отклонения, выбросы, аномалии.)

В работе врача-статистика дашборд становится незаменимым инструментом по нескольким причинам.

Для руководства: возможность за 5 мин. оценить ситуацию в стационаре, не погружаясь в кипу бумажных отчетов. Кликнул по отделению, увидел динамику. Выбрал месяц – сравнил с предыдущим.

Для вас: автоматизация рутины. Вместо того чтобы каждый раз заново строить графики, вы обновляете исходные данные, и дашборд пересчитывается сам. Время подготовки отчета сокращается с часов до минут.

Для аналитики: возможность быстро исследовать данные. Покрутил срезы – увидел, что в кардиологии рост госпитализаций, а в неврологии снижение. Переключил на врачей – нашел специалиста с аномально высокой длительностью лечения.

2. Основные элементы дашборда

Любой дашборд состоит из трех типов элементов.

1. Ключевые показатели (KPI)

Крупные цифры, которые показывают самое главное: общее число пациентов, среднюю длительность лечения, летальность, занятость коек. Часто сопровождаются динамикой (стрелка вверх/вниз, процент изменения).

2. Визуализации

Графики и диаграммы, которые показывают динамику, структуру, сравнения. Обычно 3–5 графиков на одном листе.

3. Элементы управления (срезы, фильтры)

Интерактивные кнопки, которые позволяют фильтровать все графики одновременно. Выбрали отделение – все показатели и графики перестроились под него.

3. Создаем дашборд: пошаговая инструкция

Давайте создадим дашборд для мониторинга работы стационара. У нас есть умная таблица «Реестр» с данными о пациентах за последние 3 года. Нужно показать ключевые показатели, динамику госпитализаций, структуру по отделениям, распределение по возрастам и дать возможность фильтровать по году, отделению и врачу.

Шаг 1. Подготавливаем основу

Создаем новый лист и называем его «Дашборд». Здесь мы будем размещать все элементы. Исходные данные остаются на другом листе (или в другой книге).

Шаг 2. Создаем ключевые показатели (KPI)

Для каждого ключевого показателя создаем отдельную ячейку с формулой, которая считает нужное значение. Чтобы KPI были интерактивными, они должны ссылаться на сводные таблицы или использовать функции с условиями, которые учитывают выбранные срезы.

Самый удобный способ – создать несколько сводных таблиц на отдельном листе (лист «Данные_для_дашборда»), а на дашборде вывести их значения через простые ссылки.

Пример KPI

- Общее число пациентов: =СЧЕТЗ(Реестр[Номер истории]) или ссылка на ячейку со сводной таблицей.
- Общий койко-день: =СУММ(Реестр[Койко-дни]).
- Средняя длительность лечения: =СРЗНАЧ(Реестр[Койко-дни]).
- Летальность, %: =СЧЕТЕСЛИ(Реестр[Исход]; «умер») / СЧЕТЗ(Реестр[-Номер истории]) × 100.

Оформляем KPI крупным шрифтом (20–24 pt), полужирным, с подписями мелким шрифтом. Добавляем динамику – изменение по сравнению с предыдущим периодом (рис. 4.8).

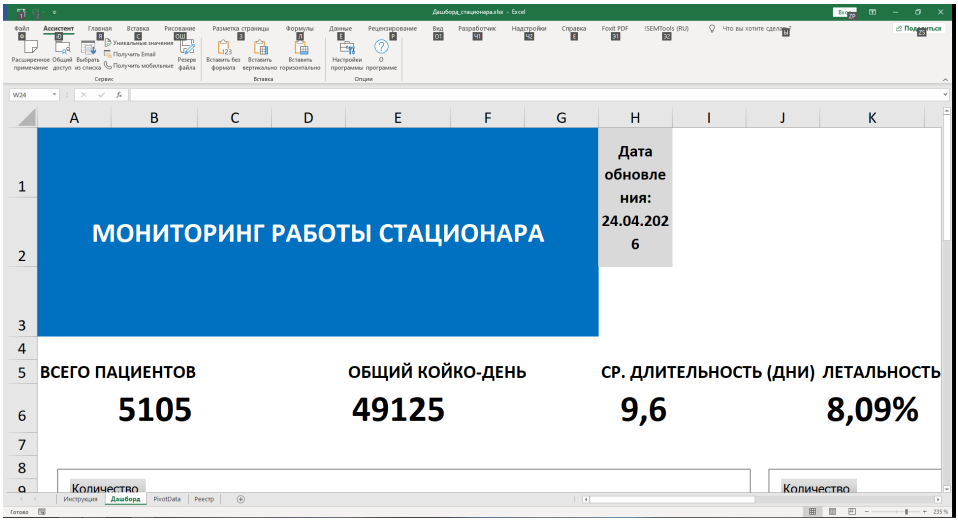


Рисунок 4.8. Место для рисунка: блок KPI на дашборде

Шаг 3. Создаем сводные таблицы и сводные диаграммы

Для каждого графика, который мы хотим разместить на дашборде, создаем сводную таблицу и на ее основе сводную диаграмму (рис. 4.9).

График 1. Динамика госпитализаций по месяцам (линейчатый)

- Сводная таблица: строки – дата (сгруппирована по месяцам), значения – количество номеров историй.
- Сводная диаграмма: линейчатая с маркерами.

График 2. Структура госпитализаций по отделениям (круговая)

- Сводная таблица: строки – отделение, значения – количество пациентов.
- Сводная диаграмма: круговая (с объединением мелких секторов в «Прочие»).

График 3. Средняя длительность лечения по отделениям (гистограмма)

- Сводная таблица: строки – отделение, значения – среднее койко-дней.
- Сводная диаграмма: гистограмма.

График 4. Распределение пациентов по возрастам (гистограмма)

- Добавляем в исходные данные вспомогательный столбец «Возрастная группа» (0–17, 18–44, 45–64, 65+).
- Сводная таблица: строки – возрастная группа, значения – количество пациентов.
- Сводная диаграмма: гистограмма.

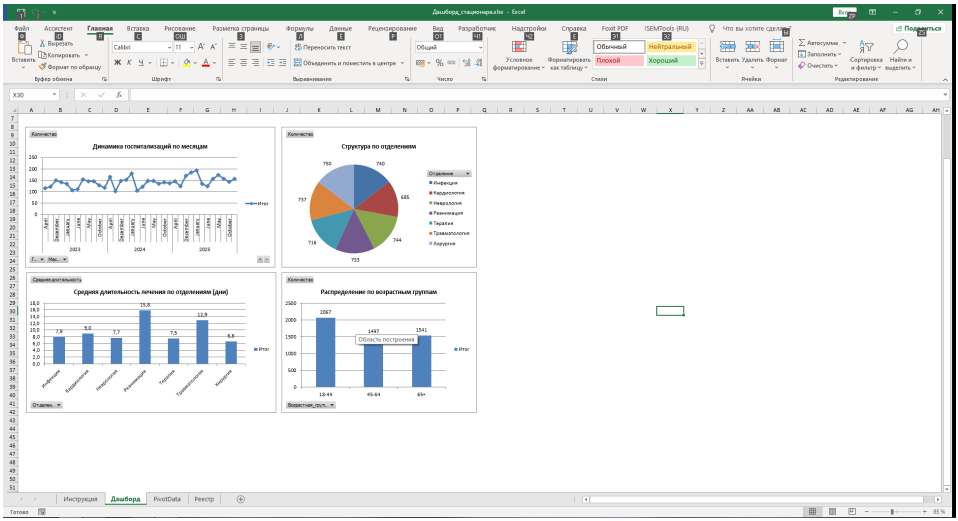


Рисунок 4.9. Четыре сводные диаграммы, размещенные на листе дашборда

Шаг 4. Добавляем срезы (slicers) для интерактивности

Срезы – это кнопки, которые позволяют фильтровать все сводные таблицы и диаграммы одновременно.

Как добавить срез (рис. 4.10)

1. Выделите любую сводную таблицу.

2. Вкладка «Анализ сводной таблицы» → «Вставить срез».

3. Выберите поля, по которым хотите фильтровать: «Год», «Отделение», «Врач».

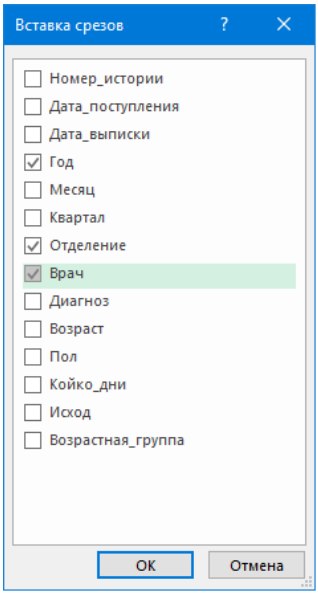


Рисунок 4.10. Окно выбора полей для срезов

После создания срезов их нужно связать со всеми сводными таблицами на дашборде. Для этого (рис. 4.11):

1. правой кнопкой по срезу → «Связи отчета»;
2. в появившемся окне отметить все сводные таблицы, которые должны фильтроваться этим срезом.

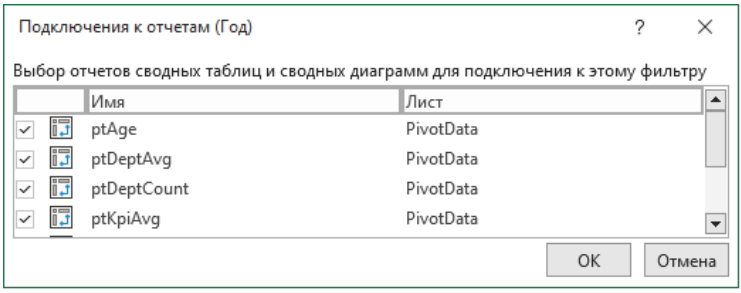


Рисунок 4.11. Окно «Связи отчета» с отмеченными сводными таблицами

Теперь, когда вы нажимаете на кнопку «Кардиология» в срезе отделений, все графики перестраиваются, показывая данные только по кардиологическому отделению. Это и есть интерактивность.

Шаг 5. Настраиваем внешний вид срезов

Срезы по умолчанию выглядят как серые кнопки. Их можно настроить под стиль дашборда:

- изменить количество столбцов в срезе (чтобы кнопки располагались в несколько рядов);
- выбрать стиль оформления (на вкладке «Параметры среза»);
- задать размер и положение.

Для удобства можно добавить временную шкалу (timeline) для фильтрации по датам – это специальный срез для полей с датами (рис. 4.12).

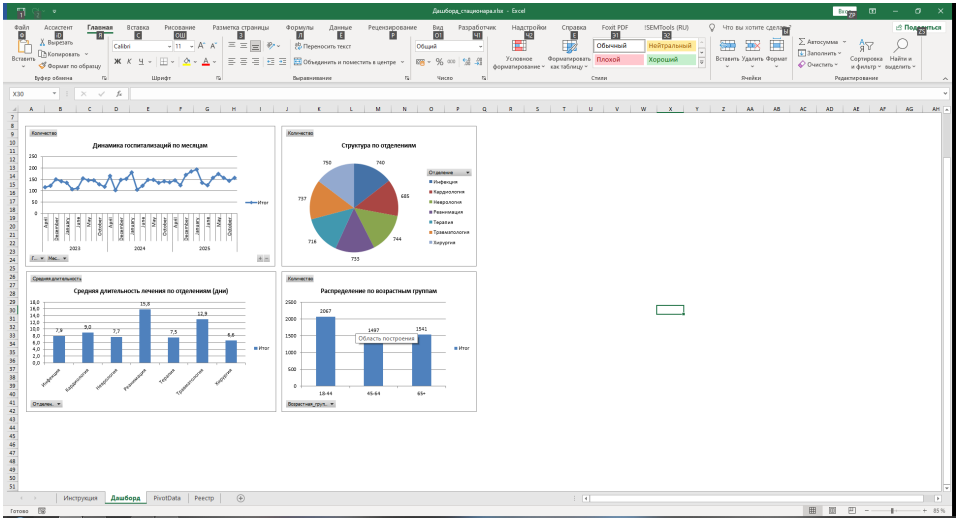


Рисунок 4.12. Настроенные срезы (год, отделение, врач)

Шаг 6. Добавляем элементы управления для KPI

KPI, которые мы создали через прямые формулы, не будут автоматически фильтроваться срезами. Чтобы они тоже реагировали на выбор, нужно:

- либо использовать функции СЧЕТЕСЛИМН, СУММЕСЛИМН, которые учитывают значения в срезах (сложно),
- либо создать отдельные сводные таблицы для KPI и вывести их значения на дашборд через ссылки.

Шаг 7. Оформляем дашборд

Теперь, когда все работает, приводим внешний вид к единому стандарту (рис. 4.13).

- Убираем рамки у сводных диаграмм, делаем фон белым.
- Убираем лишние элементы (линии сетки, легенду, если она дублируется).
- Выравниваем графики по сетке.
- Добавляем заголовок дашборда «Мониторинг работы стационара» и дату последнего обновления (формула =СЕГОДНЯ()).
- Используем единую цветовую гамму (например, синий для основного, серый для второстепенного).

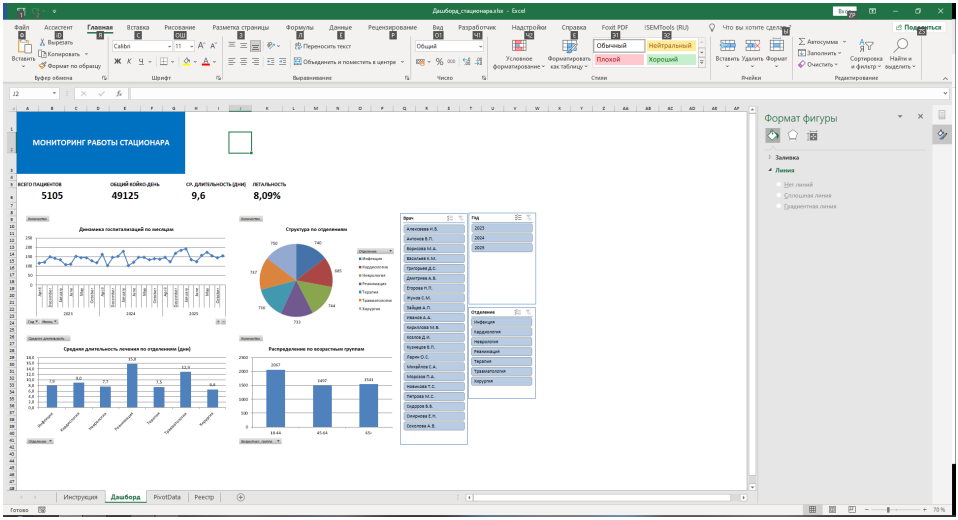


Рисунок 4.13. Финальный дашборд с настроенным оформлением

7. Кейс: дашборд для главного врача

Теперь, когда мы знаем, как все устроено, давайте представим, как этот дашборд работает в реальной жизни.

Ситуация. Главный врач заходит в кабинет, открывает файл дашборда. На экране панель управления стационаром. Он видит:

- KPI: 5105 пациентов, 49 125 койко-дней, средняя длительность 9,6 дня, летальность 8,1%;
- динамика: пик госпитализаций в марте, спад летом;
- структура: отделения распределены относительно равномерно (от 685 до 750 пациентов);
- срезы: выбраны все данные.

Главный врач задает вопрос: «А что у нас в кардиологии происходит с длительностью лечения?»

Вы кликаете в срезе «Отделение» на «Кардиологическое». Все графики перестраиваются. KPI показывают данные только по кардиологии. Гистограмма длительности по врачам кардиологии показывает, что у одного врача средняя длительность 11 дней, а у остальных – 7–8. Это сигнал для разбора.

Главный врач: «А по месяцам как?»

Вы смотрите на линейчатый график. В кардиологии рост госпитализаций в осенне-зимний период, но длительность лечения остается стабильной.

Все это занимает 2–3 мин. Никаких дополнительных отчетов, никакой ручной подготовки. Дашборд дал ответы на все вопросы (рис. 4.14).

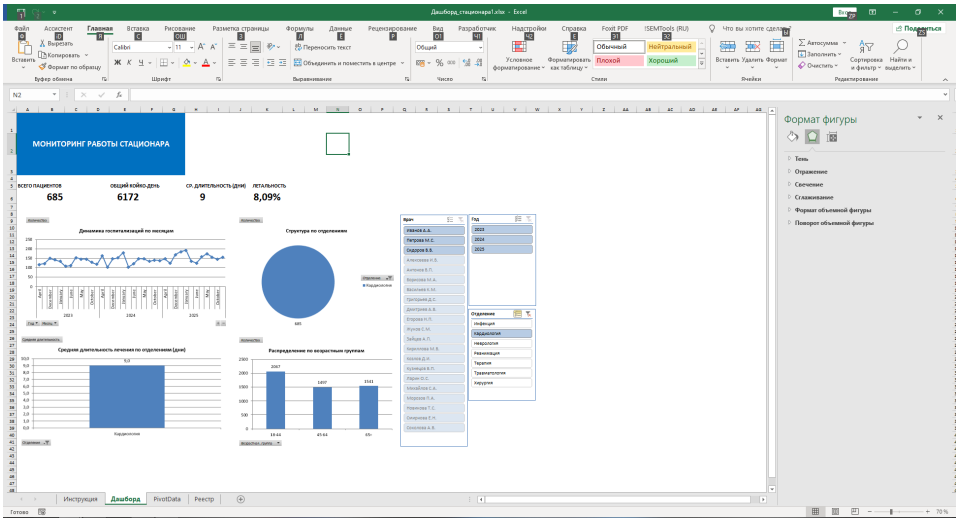


Рисунок 4.14. Дашборд с выбранным отделением «Кардиологическое» – все графики перестроены

8. Обновление дашборда: как сделать его живым

Главное преимущество дашборда – возможность быстрого обновления.

Алгоритм

1. Получаем новую выгрузку из МИС (например, данные за следующий месяц).

2. Заменяем старые данные в умной таблице «Реестр» новыми (удаляем старые, вставляем новые).

3. Нажимаем «Обновить все» (вкладка «Данные» → «Обновить все»).

4. Все сводные таблицы, графики, KPI пересчитываются автоматически. Срезы работают с новыми данными.

Время обновления: 1–2 мин. Никакой ручной настройки.

9. Практический совет: начните с малого

Если вы никогда не создавали дашборды, не пытайтесь сразу сделать сложную панель на 10 графиков. Начните с малого.

1. Создайте один-два графика на основе сводных таблиц.
2. Добавьте срез по отделениям.
3. Убедитесь, что графики фильтруются.
4. Постепенно добавляйте новые элементы.

Даже простой дашборд с 2–3 графиками и срезами сэкономит вам часы ручной работы и сделает ваши отчеты более профессиональными.

Главный вывод этой главы и всей части 4 можно сформулировать так: визуализация – это не украшение, а способ мышления. Отдельные графики отвечают на конкретные вопросы, а дашборд собирает эти ответы в единую картину, позволяя руководству видеть ситуацию целиком и принимать обоснованные решения. Освоив инструменты визуализации от выбора типа графика до создания интерактивной панели управления, вы переходите на новый уровень: от поставщика цифр к стратегическому партнеру руководителя.

В следующей части мы перейдем к Power Query – инструменту, который полностью автоматизирует подготовку данных и сделает ваш дашборд по-настоящему безотходным.



Часть 5: Знакомство с Power Query: ваш личный ассистент по подготовке данных

Глава 5.1. Знакомство с Power Query: ваш личный ассистент по подготовке данных

Мы научились готовить данные, строить сводные таблицы, писать формулы с условиями, связывать таблицы, создавать наглядные графики и даже интерактивные дашборды. Но есть одна проблема, которая остается с нами на всем этом пути: данные нужно готовить каждый месяц заново.

Каждый месяц вы получаете новую выгрузку из медицинской информационной системы. Каждый месяц вы повторяете одни и те же действия: удаляете дубликаты, приводите текст к единому виду, разделяете Ф.И.О., исправляете типы данных, подтягиваете справочники. Это занимает часы.

И каждый месяц вы думаете: «Неужели нельзя сделать это один раз, а потом просто нажимать кнопку?»

Можно. И этот инструмент называется Power Query.

Power Query – это технология, которая превращает рутинную подготовку данных в автоматический процесс. Вы настраиваете последовательность действий один раз, а затем каждый месяц просто обновляете запрос. Все очистки, преобразования, объединения таблиц выполняются автоматически. Это не просто экономия времени, это переход на новый уровень профессионализма.

В этой части книги мы познакомимся с Power Query, научимся загружать данные, очищать их, объединять несколько файлов в один и создавать полностью автоматизированную систему отчетности. Начнем с главного: с понимания, что такое Power Query и где он находится.

Представьте себе: у вас есть личный ассистент, которому можно один раз показать, как нужно обрабатывать ежемесячную выгрузку, а потом просто говорить: «Сделай, как в прошлый раз». Ассистент помнит все ваши действия: удалить такие-то столбцы, переименовать вот эти, заменить «м» на «Мужской», разделить Ф.И.О., подтянуть названия отделений из справочника. И делает все это за минуты, без ошибок, независимо от того, сколько строк в выгрузке (тысяча или сто тысяч).

Power Query – это и есть такой ассистент. Это инструмент, который позволяет записать последовательность действий по преобразованию данных, а затем автоматически применять ее к новым данным. В мире профессиональной аналитики этот процесс называется ETL – Extract, Transform, Load (извлечь, преобразовать, загрузить). Звучит сложно, но на деле все гораздо проще.

Что такое ETL и зачем это врачу-статистику

ETL – это трехэтапный процесс работы с данными, который Power Query выполняет за вас.

Extract (Извлечь) – загрузить данные из источника. Источником может быть файл Excel, папка с файлами, база данных, веб-страница или даже папка с выгрузками за несколько месяцев. Power Query умеет забирать данные откуда угодно.

Transform (Преобразовать) – очистить и привести данные к нужному виду. Удалить лишние столбцы, переименовать, изменить тип данных, заменить значения, разделить текст по столбцам, объединить несколько таблиц, отфильтровать строки, удалить дубликаты – все это Power Query делает автоматически по заданному вами сценарию.

Load (Загрузить) – поместить результат в Excel. После преобразования данные загружаются в виде умной таблицы или сводной таблицы, готовой к дальнейшему анализу.

Почему это важно для врача-статистика? Потому что 80% времени в работе с данными тратится именно на подготовку, на то, что в ETL называется Transform. Power Query забирает на себя эту рутину. Вы один раз настраиваете процесс преобразования, а затем просто обновляете запрос.

Реальный пример из жизни врача-статистика

Каждый месяц вы получаете из МИС выгрузку в формате Excel. В ней:

- лишние столбцы (служебная информация, которая не нужна в отчете);
- объединенные ячейки в заголовках (их нужно разобрать);
- Ф.И.О. в одной ячейке (нужно разделить на фамилию, имя, отчество);
- пол обозначен то «м», то «М», то «мужской» (нужно привести к единому виду);
- дата поступления в текстовом формате (нужно превратить в дату);4
- код отделения, а нужно название (подтянуть из справочника).

Раньше вы делали это вручную каждый месяц, 2–3 ч работы. С Power Query вы делаете это один раз. В следующем месяце вы просто загружаете новую выгрузку в ту же папку, нажимаете «Обновить» и через минуту получаете готовую очищенную таблицу.

Где находится Power Query в Excel

Power Query встроен в Excel начиная с версии 2016. В более ранних версиях (Excel 2010, 2013) он был отдельной надстройкой, но мы ориентируемся на современные версии.

Excel 2016, 2019, 2021, Microsoft 365 (Windows)

Power Query находится на вкладке «Данные» в группе «Получение и преобразование данных» (рис. 5.1).

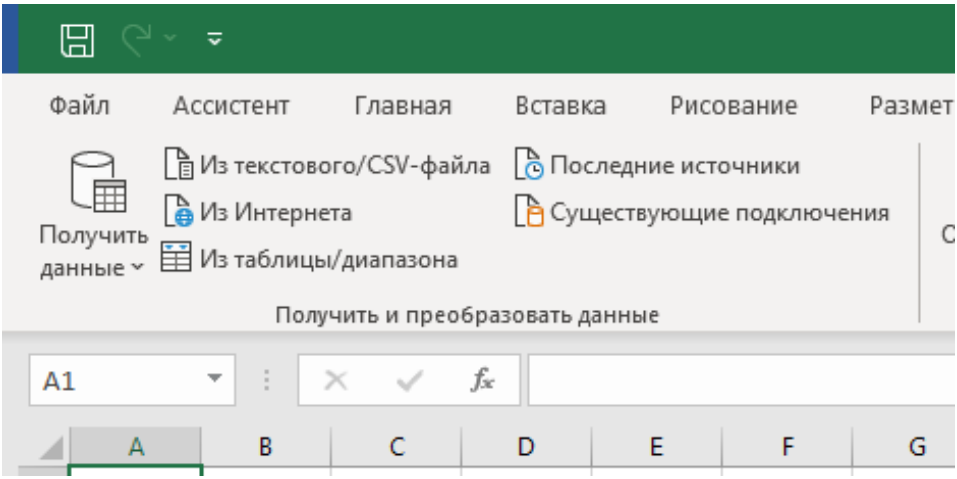


Рисунок 5.1. Вкладка «Данные» в Excel с выделенной группой «Получение и преобразование данных»

Вот основные кнопки, которые вам понадобятся.

- Из файла: загрузить данные из файла Excel, CSV, XML, JSON, папки и т.д.
- Из таблицы/диапазона: загрузить данные из текущего листа в Power Query.
- Последние использованные источники: быстро открыть недавние запросы.
- Обновить всё: обновить все запросы и загрузить свежие данные.

Excel для Mac (Microsoft 365)

В Mac-версии Power Query появился позже, но начиная с Microsoft 365 для Mac он доступен. Расположение аналогичное: вкладка «Данные» → группа «Получение данных». Однако функциональность может быть немного ограничена по сравнению с Windows-версией (например, некоторые расширенные преобразования могут отсутствовать). В тексте мы будем отмечать такие нюансы.

Главное, что нужно запомнить: Power Query – это не формула, не макрос и не надстройка. Это встроенный инструмент Excel, который находится прямо на вкладке «Данные». Он всегда под рукой.

Как выглядит Power Query: знакомство с интерфейсом

Когда вы загружаете данные в Power Query, открывается отдельное окно Редактор Power Query. Это отдельная среда внутри Excel, специально предназначенная для преобразования данных (рис. 5.2).

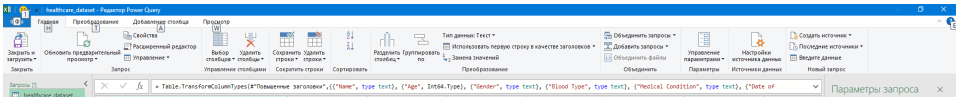


Рисунок 5.2. Окно Редактора Power Query с подписями основных элементов

Редактор Power Query состоит из нескольких зон.

1. Лента (в верхней части)

Здесь собраны все инструменты для преобразования данных. Они сгруппированы по вкладкам.

- Главная – основные действия: загрузить, обновить, удалить строки, сохранить запрос.
- Преобразование – самое важное для нас: изменить тип данных, разделить столбцы, заменить значения, отфильтровать строки.
- Добавление столбца – создать новые столбцы на основе существующих (например, выделить возраст из даты рождения).
- Просмотр – настройки отображения.

2. Панель запросов (слева)

Список всех запросов в текущей книге. Здесь можно переключаться между разными таблицами, которые вы загружаете и преобразуете.

3. Предварительный просмотр данных (в центре)

Здесь отображается результат преобразований на данный момент. Вы видите, как меняются данные после каждого шага. Это позволяет сразу проверять, правильно ли вы все делаете.

4. Панель настроек (справа)

Здесь отображаются шаги (steps) – последовательность всех действий, которые вы применили к данным. Каждый шаг можно удалить, отредактировать или изменить порядок. Это сердце Power Query: вы не просто применяете действия, вы записываете сценарий (рис. 5.3).

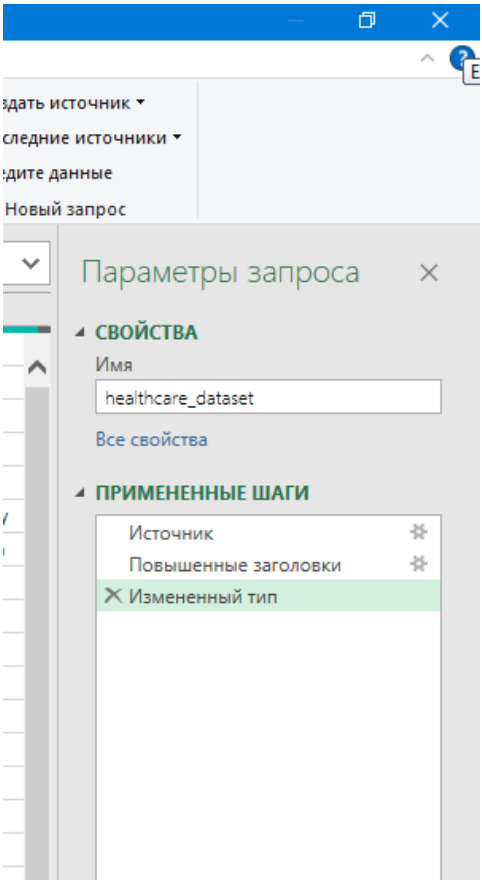


Рисунок 5.3. Панель «Шаги» с последовательностью действий

Первое знакомство: загружаем данные в Power Query

Давайте сделаем первый шаг: загрузим данные в Power Query. Это проще, чем кажется.

Вариант 1. Загрузить данные из текущей таблицы Excel

1. Выделите любую ячейку внутри вашей умной таблицы (или обычного диапазона).

2. На вкладке «Данные» нажмите «Из таблицы/диапазона».

3. Если данные еще не были умной таблицей, Excel спросит, нужно ли ее создать. Нажмите ОК.

Откроется Редактор Power Query с вашими данными.

Вариант 2. Загрузить данные из внешнего файла Excel

1. На вкладке «Данные» → «Из файла» → «Из книги Excel».
2. Найдите нужный файл на компьютере или в сети.
3. В открывшемся окне «Навигатор» выберите нужный лист или именованный диапазон.
4. Нажмите «Преобразовать данные» (чтобы открыть Редактор) или «Загрузить» (чтобы сразу загрузить в Excel) (рис. 5.4).

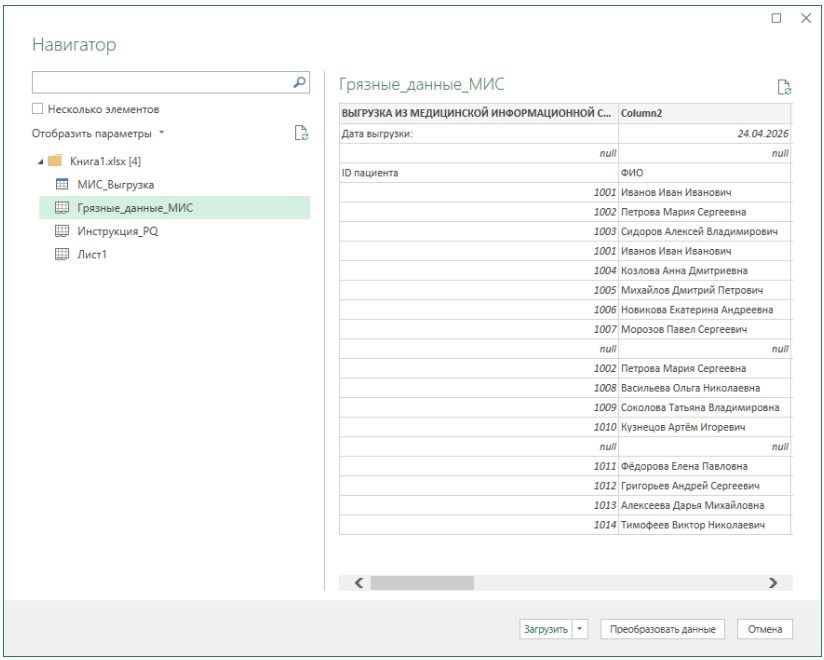


Рисунок 5.4. Окно «Навигатор» с выбором листа и кнопкой «Преобразовать данные»

Вариант 3. Загрузить данные из папки с несколькими файлами

Это одна из самых мощных функций Power Query. Если вы получаете ежемесячные выгрузки в отдельных файлах, Power Query может объединить их все в одну таблицу.

1. На вкладке «Данные» → «Из файла» → «Из папки».

2. Укажите путь к папке, где лежат файлы (например, «C:\Отчеты\2025\Выгрузки»).

3. Power Query покажет список всех файлов в папке.

4. Нажмите «Объединить и преобразовать данные» или «Преобразовать данные» для более тонкой настройки.

Первые шаги в Редакторе Power Query

Когда вы впервые попадаете в Редактор Power Query, он может показаться непривычным. Но на самом деле большинство действий здесь повторяют то, что вы уже умеете делать в Excel, просто интерфейс другой.

Что вы можете делать в Редакторе (и это будет записано в шаги)

- Удалить столбцы: правой кнопкой по заголовку столбца → «Удалить».

- Переименовать столбцы: двойной клик по заголовку.

- Изменить тип данных: иконка слева от заголовка столбца (например, «123» для чисел, «календарь» для дат).

- Удалить дубликаты: правой кнопкой по столбцу → «Удалить дубликаты».

- Заменить значения: правой кнопкой по столбцу → «Заменить значения» (аналог «Найти и заменить»).

- Разделить столбец: правой кнопкой → «Разделить столбец» (аналог «Текст по столбцам»).

- Отфильтровать строки: нажать на стрелочку в заголовке столбца и снять галочки с ненужных значений.

- Добавить условный столбец: вкладка «Добавление столбца» → «Условный столбец».

Важнейшее отличие от Excel: все эти действия не применяются к данным мгновенно, навсегда. Они записываются в панели «Шаги» справа. Вы можете вернуться к любому шагу, изменить его, удалить, вставить новый между существующими. Данные каждый раз пересчитываются заново по этому сценарию.

Зачем врачу-статистику осваивать Power Query

Может возникнуть вопрос: «Я и так неплохо справляюсь с формулами и сводными таблицами. Зачем мне еще и Power Query?»

Ответ прост. Формулы и сводные таблицы работают с данными, которые уже находятся в Excel. Power Query работает с данными до того, как они попали в Excel. Он берет на себя всю грязную работу: загрузку из разных источников, очистку, преобразование, объединение. В результате в Excel попадают уже готовые, идеально подготовленные данные.

Что дает Power Query врачу-статистику

- Экономия времени. То, что раньше занимало 2–3 ч в месяц, теперь занимает 5–10 мин.
- Исключение ошибок. Человек, выполняя рутинные действия, рано или поздно ошибается. Power Query делает это одинаково каждый раз.
- Воспроизводимость. Вы можете передать свой запрос коллеге, и он получит тот же результат. Не нужно объяснять «Сначала удали вот эти столбцы, потом замени...» – все уже в запросе.
- Масштабируемость. Формулы в Excel начинают тормозить при
- 100 000 строк. Power Query легко обрабатывает миллионы строк.
- Объединение файлов. Power Query может объединить все файлы из папки в одну таблицу, даже если файлов 50 и в каждом по 50 000 строк.

Главный вывод этой главы можно сформулировать так: Power Query – это инструмент, который превращает рутинную подготовку данных из неизбежного зла в автоматический процесс, выполняемый за минуты. Он находится прямо на вкладке «Данные», не требует программирования и запоминает все ваши действия, чтобы повторить их в следующий раз. Освоив Power Query, вы освободите себе часы времени каждый месяц и сможете сосредоточиться на том, что действительно важно: на анализе и интерпретации данных.

Глава 5.2. Очистка данных за 10 минут: превращаем хаос в порядок

В предыдущей главе мы познакомились с Power Query, узнали, где он находится и как устроен. Теперь переходим к самому главному: к практической очистке данных. Именно здесь Power Query раскрывает свою мощь, превращая то, что раньше занимало часы ручной работы, в 10-минутную настройку, которая затем работает автоматически.

Представьте себе типичную выгрузку из медицинской информационной системы. В ней есть все: и нужные столбцы, и лишние, и пустые строки, и дубликаты, и разнобой в написании, и даты, которые Excel упорно считает текстом. Раньше вы вооружались инструментами из первой части книги и тратили час-два на приведение всего этого в порядок. Теперь вы сделаете это один раз в Power Query, а затем будете просто обновлять запрос.

В этой главе мы разберем четыре ключевых действия, которые составляют основу очистки любых медицинских данных: удаление дубликатов, замена значений, удаление пустых строк и столбцов, изменение типов данных. Каждое действие мы рассмотрим на реальных примерах и покажем, как оно выполняется в Power Query.

Удаление дубликатов: избавляемся от повторов

Дубликаты – одна из самых частых проблем в медицинских выгрузках. Один и тот же пациент может быть заведен дважды, повторная госпитализация может быть зарегистрирована как два отдельных случая с одинаковыми данными, технический сбой может создать дублирующие строки.

Как это делается в Power Query

1. В редакторе Power Query выделите столбец (или несколько столбцов), по которым нужно искать дубликаты. Обычно это комбинация «Номер истории болезни», «Дата поступления», «Ф.И.О.».

2. На вкладке «Главная» → «Удалить строки» → «Удалить дубликаты» (рис. 5.5).

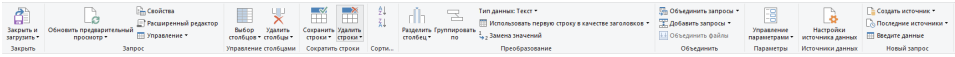


Рисунок 5.5. Меню «Удалить строки»

Важный нюанс. Если вы выделите один столбец, Power Query удалит строки, где значения в этом столбце повторяются. Если выделите несколько столбцов (удерживая Ctrl), дубликатами будут считаться строки, где все выделенные столбцы совпадают. В медицинской практике чаще используется второй вариант – удаление полных дубликатов случая госпитализации.

Медицинский кейс. В выгрузке из приемного покоя обнаружились дубликаты: один и тот же пациент (Иванов И.И.) с одинаковой датой и временем поступления записан дважды. Выделяем столбцы «Ф.И.О.», «Дата поступления», «Время поступления», нажимаем «Удалить дубликаты». Остается одна запись.

В панели «Шаги» появляется запись «Удалены дубликаты». В следующий раз это действие повторится автоматически.

Замена значений: приводим к единому виду

Разнобой в написании – вечная головная боль врача-статистика. Пол может быть обозначен как «м», «М», «муж», «Мужской», «мужской », «1». Отделения могут называться «Кардиология», «кардио», «Кардиологическое отделение». Диагнозы – «ГБ», «гипертония», «Гипертоническая болезнь».

Power Query позволяет заменить все эти варианты на единый стандарт за несколько кликов.

Как это делается

1. Выделите столбец, в котором нужно заменить значения.

2. Нажмите правой кнопкой по заголовку столбца → «Заменить значения».

3. В открывшемся окне введите «Найти» (то, что нужно заменить) и «Заменить на» (единый стандарт).

4. Нажмите ОК (рис. 5.6).

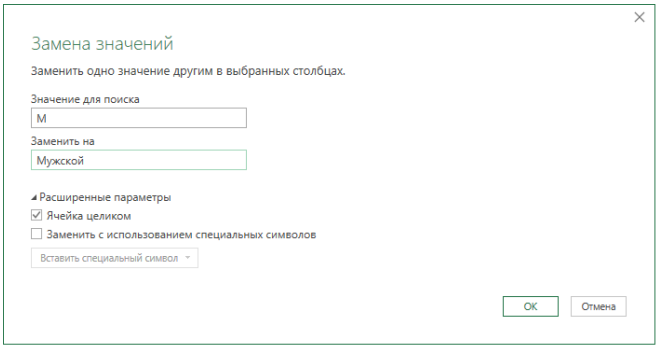


Рисунок 5.6. Окно «Заменить значения» с примером: найти «м», заменить на «Мужской»

Важная особенность Power Query: замены можно делать последовательно, и все они запишутся в шаги. Вы можете сначала заменить «м» на «Мужской», затем «М» на «Мужской», затем «муж» на «Мужской». Каждая замена – отдельный шаг, который можно редактировать или удалить.

Медицинский кейс

В столбце «Пол» встречаются варианты: «м», «М», «муж», «Мужской», «Муж», «1», пустые ячейки. Создаем последовательные замены:

- «м» → «Мужской»;
- «М» → «Мужской»;
- «муж» → «Мужской»;
- «Муж» → «Мужской»;
- «1» → «Мужской»;
- пустые ячейки оставляем пустыми или заменяем на «Не указан».

После всех замен все мужчины будут обозначены единообразно. То же самое делаем для женщин.

Удаление пустых строк и столбцов: убираем мусор

Выгрузки из МИС часто содержат пустые строки (служебные разделители) и пустые столбцы (поля, которые не используются в данной организации). Power

Query умеет удалять и то, и другое одним кликом.

Удаление пустых строк

1. Выделите один или несколько столбцов (или всю таблицу).

2. На вкладке «Главная» → «Удалить строки» → «Удалить пустые строки».

Power Query удалит все строки, где во всех столбцах (или в выделенных) нет данных.

Удаление пустых столбцов

1. На вкладке «Главная» → «Удалить столбцы» → «Удалить пустые столбцы».

Power Query найдет и удалит все столбцы, в которых все ячейки пусты.

Медицинский кейс. В выгрузке из стационара есть 20 столбцов. 5 из них служебные, всегда пустые. Еще 3 столбца заполнены только в 5% случаев и не используются в анализе. В конце файла 2 пустые строки-разделителя. Power Query удаляет все это за секунды. В Excel пришлось бы искать пустые столбцы вручную или писать сложные формулы.

Изменение типа данных: заставляем Excel правильно понимать цифры и даты

Самый коварный враг достоверного анализа – неверно распознанные типы данных. Числа, которые Excel считает текстом, не суммируются. Даты, которые считаются текстом, не группируются в сводных таблицах по месяцам. Power Query позволяет исправить это на этапе загрузки.

Как это делается

1. В редакторе Power Query обратите внимание на иконки слева от заголовков столбцов:

- «123» – число;
- «Abc» – текст;
- «Календарь» – дата;

- «Часы» – время;
- «123 с дробью» – десятичное число.

2. Чтобы изменить тип, кликните по иконке и выберите нужный тип из списка.

Медицинский кейс. В выгрузке столбец «Дата поступления» имеет иконку «Abc» (текст), хотя данные выглядят как даты. Кликаем по иконке, выбираем «Дата». Power Query преобразует текст в даты. В столбце «Койко-дни» числа выровнены влево (признак текста). Меняем тип на «Целое число». Все числа становятся корректными.

Важный нюанс. Если в столбце есть нечисловые значения (например, прочерк вместо нуля), Power Query может выдать ошибку при попытке изменить тип на число. В таких случаях сначала нужно заменить прочерки на нули (или на пустоту) с помощью «Заменить значения», а затем менять тип.

- Практический сценарий: очистка выгрузки из стационара за 10 мин.

Давайте соберем все вместе и представим реальный сценарий работы врача-статистика. У нас есть выгрузка из МИС со следующими проблемами (табл. 5.1).

- Лишние столбцы (A, B, C, D, E, F, G – нужны только A, C, D, F).
- Пустые строки в конце файла.
- Пустые столбцы (K, M, N).
- Дубликаты по Ф.И.О. и дате поступления.
- Пол обозначен разными вариантами.
- Дата поступления – текст.
- Койко-дни – текст.

Таблица 5.1. Настройка запроса в Power Query (один раз)

Шаг	Действие	Результат
1.	Загружаем файл в Power Query	Данные в редакторе
2.	Удаляем лишние столбцы (правой кнопкой → «Удалить»)	Остались только нужные
3.	Удаляем пустые столбцы («Главная» → «Удалить столбцы» → «Удалить пустые столбцы»)	Пустые столбцы исчезли
4.	Удаляем пустые строки («Главная» → «Удалить строки» → «Удалить пустые строки»)	Пустые строки удалены
5.	Выделяем столбцы «Ф.И.О.» и «Дата поступления», удаляем дубликаты	Повторные записи удалены
6.	Заменяем значения в столбце «Пол»: «м» → «Мужской», «М» → «Мужской», «ж» → «Женский», «Ж» → «Женский»	Пол приведен к единому виду
7.	Меняем тип столбца «Дата поступления» на «Дата»	Даты стали корректными
8.	Меняем тип столбца «Койко-дни» на «Целое число»	Числа стали корректными
9.	Нажимаем «Закрыть и загрузить»	Очищенная таблица загружена в Excel

Глава 5.3. Склеивание и объединение: из множества файлов единый отчет

Мы научились очищать один файл. Но в реальной работе врача-статистика данные почти всегда приходят в виде множества файлов: выгрузки за январь, февраль, март… или отчеты по отделениям, которые нужно собрать воедино. Раньше вы открывали каждый файл, копировали данные, вставляли в общую таблицу. И так 12 раз подряд. Это занимало часы, риск ошибки был огромен.

Power Query решает эту задачу кардинально. Он умеет объединять файлы из папки в одну таблицу (это называется добавление запросов, Append) и связывать разные таблицы по ключевым полям (это называется объединение запросов, Merge). В этой главе мы разберем оба подхода на медицинских кейсах.

Добавление запросов (Append): склеиваем файлы за разные месяцы

Представьте, что у вас есть 12 файлов – выгрузки из стационара за каждый месяц года. Все файлы имеют одинаковую структуру (одинаковые столбцы). Вам нужно собрать их в одну таблицу, чтобы проанализировать годовые данные.

Вариант 1. Склеиваем несколько файлов, выбранных вручную

Если файлов немного и их список меняется от раза к разу, можно загрузить каждый файл отдельным запросом, а затем склеить их.

1. Загрузите первый файл в Power Query (Данные → Из файла → Из книги Excel). Нажмите «Преобразовать данные». Затем «Файл» → «Закрыть и загрузить».

2. Повторите для второго файла, третьего и т.д. У вас появится несколько запросов в панели слева.

3. На вкладке «Главная» → «Добавить запросы» → стрелка вниз → «Добавить запросы».

4. В открывшемся окне выберите запросы, которые нужно склеить, и нажмите ОК (рис. 5.7).

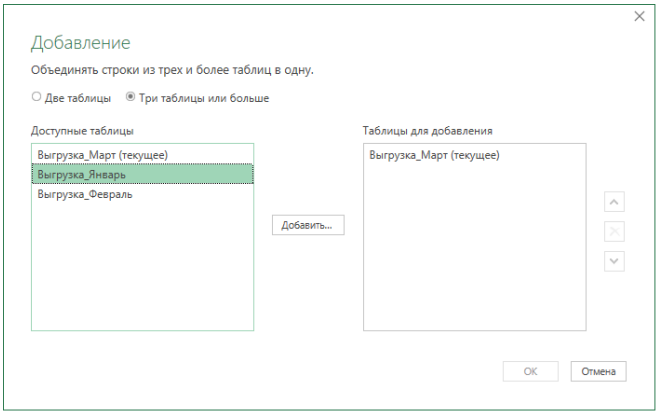


Рисунок 5.7. Окно «Добавление запросов» с выбранными запросами «Январь», «Февраль», «Март»

Power Query создаст новый запрос, который объединяет все выбранные таблицы одну под другой.

Вариант 2. Склеиваем все файлы из папки (автоматически)

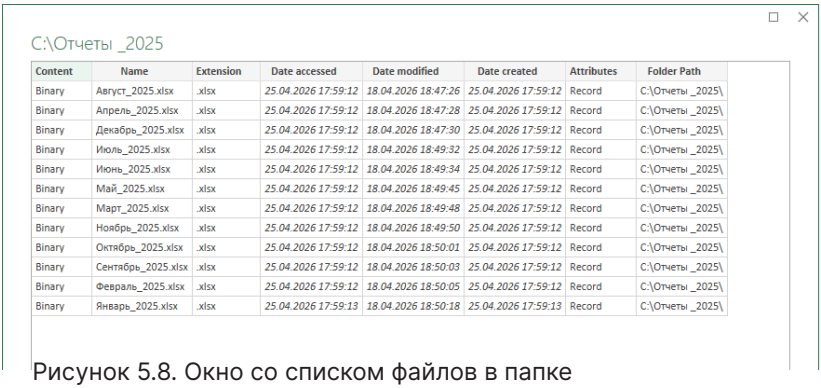
Это самый мощный сценарий. Если вы каждый месяц кладете файлы в одну и ту же папку с одинаковой структурой, Power Query может объединить их все автоматически, без необходимости каждый раз указывать, какие файлы добавлять (рис. 5.8).

1. На вкладке «Данные» → «Из файла» → «Из папки».

2. Укажите путь к папке, где лежат ваши ежемесячные файлы (например, «C:\Отчеты_2025»).

3. Power Query покажет список всех файлов в папке.

4. Нажмите кнопку «Объединить» и преобразовать данные» (или «Преобразовать данные» для более тонкой настройки).


C:\Отчеты _2025

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	Август_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:47:26	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Апрель_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:47:28	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Декабрь_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:47:30	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Июль_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:49:32	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Июнь_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:49:34	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Май_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:49:45	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Март_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:49:48	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Ноябрь_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:49:50	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Октябрь_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:50:01	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Сентябрь_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:50:03	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Февраль_2025.xlsx	.xlsx	25.04.2026 17:59:12	18.04.2026 18:50:05	25.04.2026 17:59:12	Record	C:\Отчеты _2025\
Binary	Январь_2025.xlsx	.xlsx	25.04.2026 17:59:13	18.04.2026 18:50:18	25.04.2026 17:59:13	Record	C:\Отчеты _2025\

Рисунок 5.8. Окно со списком файлов в папке

Power Query автоматически:

- загрузит все файлы из папки;
- распознает структуру (по первому файлу);
- объединит их в одну таблицу;
- добавит служебный столбец с именем файла-источника (полезно, чтобы знать, из какого месяца данные).

Результат: один запрос, который всегда смотрит на папку. Когда вы кладете в папку новый файл (например, выгрузку за следующий месяц), просто нажимаете «Обновить», и новый файл автоматически добавляется в общую таблицу.

Медицинский кейс: годовой отчет из 12 месячных выгрузок

В стационаре каждый месяц формируется файл «Стационар_месяц.xlsx». Все файлы хранятся в папке «C:\Отчеты_2025». Врач-статистик настраивает запрос на папку один раз в январе. В течение года он просто кладет новые файлы в эту папку. В декабре, чтобы получить годовой отчет, он нажимает «Обновить»,

и все 12 файлов объединены в одну таблицу. Никакого копирования, никакой ручной склейки.

Объединение запросов (Merge): аналог ВПР внутри Power Query

Добавление запросов (Append) склеивает таблицы по вертикали (одну под другой). Объединение запросов (Merge) связывает таблицы

по горизонтали (подтягивает данные из одной таблицы в другую по ключу). Это аналог функций ВПР или ПРОСМОТРХ, но внутри Power Query, с теми же преимуществами: настройка один раз, автоматическое обновление.

Когда это нужно

- У вас есть реестр пациентов с кодом отделения, а нужно подтянуть название отделения из справочника.
- У вас есть табель учета времени с табельными номерами, а нужно подтянуть Ф.И.О. из кадрового справочника.
- У вас есть выгрузка диагнозов с кодами МКБ, а нужно подтянуть расшифровку из классификатора.

Как это делается

1. У вас должно быть два запроса (или два источника): основной и справочный. Например, «Справочник врачей» и «Справочник_отделений».
2. Выделите основной запрос («Справочник врачей»). На вкладке «Главная» → «Объединить запросы» → стрелка вниз → «Объединить запросы как новые» (или «Объединить запросы» для добавления в существующий).
3. В открывшемся окне выберите:

- о второй запрос («Справочник_отделений»);
- о столбцы, по которым будете связывать (например, «Код_отделения» в основном и «Код_отделения» в справочнике);
- о тип объединения: чаще всего «Левое внешнее» (оставить все строки из основного запроса, добавить данные из справочника, если есть совпадение) (рис. 5.9).

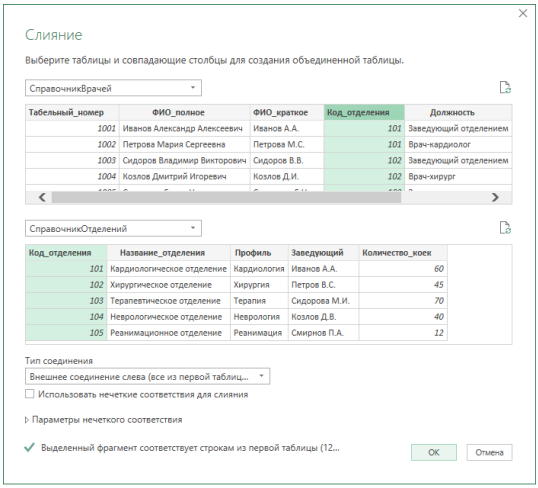


Рисунок 5.9. Окно «Объединение» с выбранными таблицами и столбцами

4. Нажмите OK. Power Query создаст новый столбец с названием «Справочник_отделений» (или как вы назвали), в котором будут вложенные таблицы.

5. Нажмите на значок расширения (две стрелки) в заголовке нового столбца и выберите, какие поля из справочника нужно добавить (например, «Название отделения», «Профиль», «Заведующий»). Снимите галочку «Использовать исходное имя столбца как префикс» (рис. 5.10).

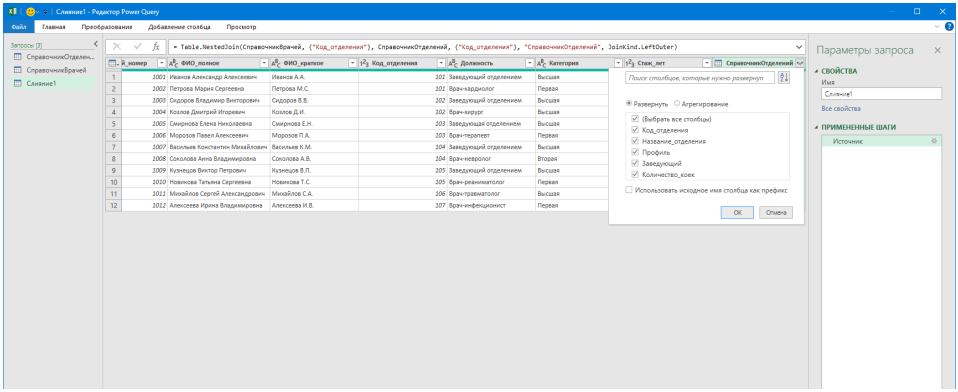


Рисунок 5.10. Окно расширения столбца с выбором полей

6. Power Query подтянет выбранные поля в основной запрос. Лишние столбцы из справочника можно удалить.

Важное отличие от ВПР. ВПР в Excel подтягивает данные статически: если справочник изменился, формула пересчитается, но, если вы добавляете новые строки в основной таблице, их нужно обрабатывать отдельно. Power Query делает это динамически: при обновлении запроса все связи перестраиваются заново.

Главный вывод этой главы: Power Query превращает сборку годовых отчетов из утомительной рутины в автоматический процесс. Добавление запросов (Append) склеивает файлы из папки, объединение запросов (Merge) подтягивает справочные данные, и все это настраивается один раз. Освоив эти инструменты, вы сможете тратить минуты на то, на что раньше уходили часы, и быть уверенными, что данные не содержат ошибок ручной обработки.



Часть 6. Бонус: быстрая статистика (пакет анализа)

Глава 6.1. Что такое «пакет анализа» и как его включить

В большинстве повседневных задач врача-статистика достаточно описательной статистики: средних, процентов, графиков. Но иногда руководство задает вопросы, на которые нельзя ответить без статистических тестов: «Случайно ли снизилась летальность в этом году или это закономерность?», «Действительно ли в нашем отделении пациенты лечатся дольше, чем в соседнем?», «Есть ли связь между возрастом пациента и длительностью лечения?».

Для таких случаев в Excel есть Пакет анализа (Analysis ToolPak) – набор статистических инструментов, которые работают в несколько кликов. Он не требует знания формул или программирования. Вы выбираете нужный инструмент, указываете диапазоны данных и получаете готовый результат с интерпретацией.

В этой части книги мы познакомимся с пакетом анализа, научимся его включать, разберем самые полезные инструменты: описательную статистику,

t-тест для сравнения групп и корреляцию для оценки связей. Это бонусная часть, но именно она превращает вас из «человека, который считает» в «аналитика, который доказывает».

Что такое надстройка и зачем она нужна

Надстройки Excel – это дополнительные модули, которые расширяют функциональность программы. Некоторые из них поставляются вместе с Excel, но не включены по умолчанию, чтобы не перегружать интерфейс для тех, кому они не нужны. Пакет анализа как раз такая надстройка.

В пакет анализа входят инструменты для:

- описательной статистики: расчет средних, медиан, стандартных отклонений, асимметрии, эксцесса и других показателей одним кликом;

- сравнения групп: t-тесты для независимых и зависимых выборок, дисперсионный анализ (ANOVA);
- анализа связей: корреляция, ковариация;
- прогнозирования: регрессионный анализ, скользящее среднее, экспоненциальное сглаживание;
- Проверки гипотез: критерии согласия, ранговые тесты.

[Место для рисунка: список инструментов Пакета анализа в окне выбора]

Для врача-статистика, который не занимается научными исследованиями на уровне диссертации, достаточно освоить три-четыре инструмента. Но даже этого набора хватит, чтобы отвечать на большинство вопросов руководства, требующих статистического обоснования.

1. Как включить пакет анализа

Активация надстройки занимает меньше минуты. В разных версиях Excel интерфейс может незначительно отличаться, но принцип один.

Excel 2016, 2019, 2021, Microsoft 365 (Windows)

1. Откройте Excel. Нажмите на вкладку «Файл» (в левом верхнем углу).
2. В открывшемся меню выберите «Параметры» (в самом низу).
3. В окне «Параметры Excel» выберите раздел «Надстройки».
4. В нижней части окна в выпадающем списке «Управление» выберите «Надстройки Excel» и нажмите «Перейти».
5. В открывшемся окне «Надстройки» поставьте галочку напротив «Пакет анализа» (Analysis ToolPak).
6. Нажмите ОК.

После этого на вкладке «Данные» справа появится новая группа «Анализ», а в ней кнопка «Анализ данных».

2. Что делать, если пакета анализа нет в списке

В редких случаях пакет анализа может быть не установлен на компьютере. Это происходит, если при установке Office был выбран вариант «без дополнительных компонентов». В такой ситуации:

1. закройте Excel;
2. откройте «Панель управления» → «Программы и компоненты»;
3. найдите Microsoft Office в списке, нажмите «Изменить»;
4. выберите «Добавить или удалить компоненты»;
5. найдите «Средства анализа Excel» (или «Analysis ToolPak») в разделе «Надстройки Excel» и установите его.

После завершения установки повторите шаги по включению надстройки.

3. Как устроен пакет анализа: общий принцип работы

Все инструменты пакета анализа работают по единому принципу.

- Вы нажимаете кнопку «Анализ данных» на вкладке «Данные».
- В открывшемся окне выбираете нужный инструмент из списка (рис. 6.1).

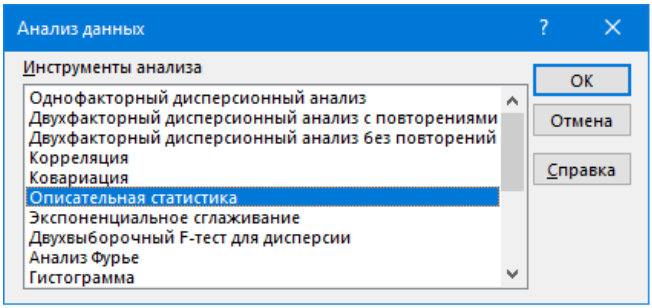


Рисунок 6.1. Анализ данных. Выбор инструмента

6. **В окне настройки инструмента указываете:**

- входной интервал: диапазон данных, которые нужно проанализировать;
- группировка: по столбцам или по строкам (обычно «по столбцам»);
- метки в первой строке: если первая строка содержит заголовки;
- выходной интервал: куда вывести результат (можно на новый лист,
- в новую книгу или в указанный диапазон на текущем листе).

7. **Нажимаете ОК (рис. 6.2).**

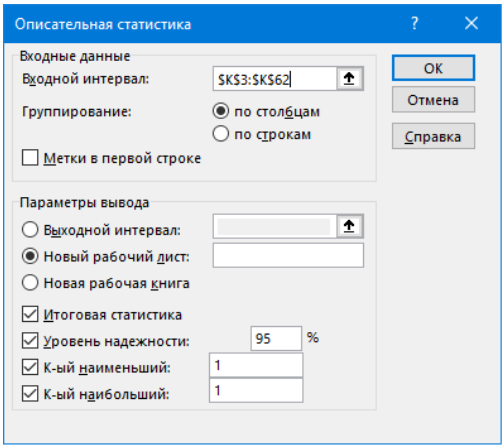


Рисунок 6.2. Окно выбора инструмента анализа и окно настройки для одного из инструментов

Excel выполняет расчеты и выводит результат в указанное место.

Результат – это статический текст и числа. Если исходные данные изменятся, результат не обновится автоматически. Для повторного расчета нужно запустить инструмент заново.

Это важно понимать: пакет анализа – это инструмент для разового углубленного анализа, а не для автоматических отчетов. Для ежедневного мониторинга мы используем сводные таблицы и формулы. А для проверки гипотез, сравнения групп, оценки связей запускаем пакет анализа.

8. Практический совет: когда использовать пакет анализа

Пакет анализа не нужен для рутинных отчетов. Его сила в ответах на вопросы, которые требуют статистического обоснования. Вот несколько примеров из практики врача-статистика.

- «Случайно ли снизилась летальность?»: t-тест покажет, значимо ли различие между показателями этого и прошлого года.

- «Действительно ли в нашем отделении лечатся дольше?»: t-тест сравнит среднюю длительность лечения в вашем отделении и в соседнем.

- «Есть ли связь между возрастом пациента и длительностью лечения?»: корреляционный анализ покажет силу и направление связи. «Какой фактор сильнее влияет на длительность лечения: возраст или тяжесть состояния?»: регрессионный анализ позволит оценить вклад каждого фактора.

В следующих главах мы разберем самые полезные инструменты пакета анализа: описательную статистику, t-тест и корреляцию. Этого набора достаточно для 90% статистических задач, возникающих в работе врача-статистика.

Главный вывод этой главы: пакет анализа – это скрытый ресурс Excel, который превращает программу из табличного процессора в полноценный инструмент статистического анализа. Его легко включить, он не требует знания формул, а результаты выводятся в удобном виде. Освоив его, вы сможете не только описывать данные, но и доказывать свои выводы статистически.

Глава 6.2. Описательная статистика одним кликом

В первой части книги мы подробно разбирали, как считать средние, медианы, стандартные отклонения с помощью формул. Это полезно, когда нужно встроить эти показатели в автоматический отчет. Но иногда возникает задача: быстро получить полный набор описательных статистик по одному или нескольким показателям. Среднее, медиана, мода, стандартное отклонение, дисперсия, асимметрия, эксцесс, минимум, максимум, сумма, количество – всё это одним запуском.

Именно для этого в пакете анализа есть инструмент «Описательная статистика». Он выдает исчерпывающую информацию о выбранных данных за секунды.

9. Когда нужна описательная статистика

Описательная статистика – это первое, что делают аналитики, когда получают новые данные. Она позволяет:

- быстро оценить распределение: нормальное оно или нет (по асимметрии и эксцессу);
- увидеть разброс: стандартное отклонение, минимум, максимум;
- понять центральную тенденцию: среднее, медиана, мода;
- проверить качество данных: если количество наблюдений меньше ожидаемого, возможно, есть пропуски.

В работе врача-статистика описательная статистика полезна при:

- анализе новых выгрузок перед построением отчетов;
- подготовке данных для научной статьи или диссертации;
- сравнении показателей в динамике (например, средняя длительность лечения в этом и прошлом году).

10. Запуск инструмента «Описательная статистика»

Предположим, у нас есть столбец с длительностью лечения пациентов за год (1000 наблюдений). Мы хотим получить полную статистическую характеристику этого показателя.

Шаг 1. Подготовка данных

Убедитесь, что данные находятся в одном столбце (или нескольких столбцах, если нужно проанализировать несколько показателей). В первой строке можно указать заголовок (он будет использован в результатах).

Шаг 2. Запуск инструмента

1. На вкладке «Данные» нажмите «Анализ данных».
2. В списке инструментов выберите «Описательная статистика» и нажмите ОК.

Шаг 3. Настройка параметров

В открывшемся окне «Описательная статистика» укажите:

- входной интервал: диапазон ячеек с данными (например, B2:B1001). Если первая строка содержит заголовок, поставьте галочку «Метки в первой строке»;
- группировка: оставьте «по столбцам» (стандартный вариант);
- выходной интервал: укажите, куда вывести результат (например, на новый лист, в новую книгу или в ячейку на текущем листе);
- в разделе «Параметры вывода» поставьте галочку «Итоговая статистика». Это основной набор показателей;
- дополнительно можно включить:
    - «уровень надежности»: доверительный интервал для среднего (по умолчанию 95%);
    - «K-й наибольший»: вывести k-е по величине значение;
    - «K-й наименьший»: вывести k-е по величине наименьшее значение (рис. 6.3).

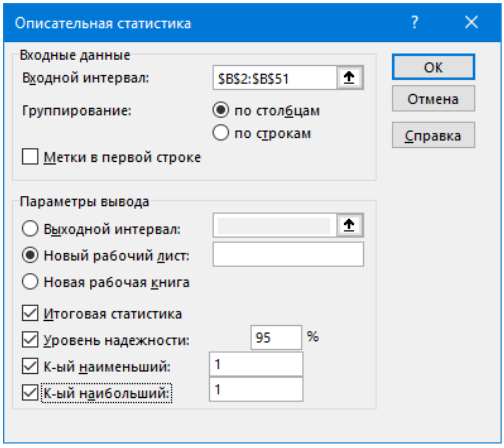


Рисунок 6.3. Окно «Описательная статистика» с заполненными полями

Шаг 4. Получение результата

После нажатия ОК Excel выводит таблицу, например, с такими результатами (табл. 6.1).

Таблица 6.1. Показатели анализа длительности лечения

Показатель	Что означает	Для длительности лечения, дни
Среднее	Среднее арифметическое	7,2
Стандартная ошибка	Стандартное отклонение среднего	0,12
Медиана	Серединное значение	6,8
Мода	Наиболее часто встре-чающееся значение	5,0
Стандартное отклонение	Мера разброса	3,8
Дисперсия выборки	Квадрат стандартного отклонения	14,44
Эксцесс	Острота пика распределения	1,2 (положительный — пик острее нормального)
Асимметричность	Смещение распределения	0,8 (положительная — «хвост» справа)
Минимум	Наименьшее значение	1
Максимум	Наибольшее значение	28
Сумма	Общая сумма	7200
Счет	Количество наблюдений	1000
Уровень надежности (95,0%)	Доверительный интер-вал для среднего	7,2 ± 0,24

11. Интерпретация результатов: что означают эти цифры

Для врача-статистика важно не просто получить цифры, но и понять, что они говорят о данных.

Среднее и медиана. В нашем примере среднее (7,2) больше медианы (6,8). Это признак положительной асимметрии: есть небольшое число пациентов с очень долгим лечением, которые тянут среднее вверх. Если бы среднее было меньше медианы, это указывало бы на отрицательную асимметрию (несколько пациентов с очень коротким лечением).

Асимметричность (Skewness). Положительное значение (0,8) подтверждает, что распределение смещено вправо. Если значение асимметрии находится в пределах от -0,5 до 0,5, распределение считается приблизительно симметричным. Если больше 1 или меньше -1, то асимметрия выраженная.

Эксцесс (Kurtosis). Положительное значение (1,2) говорит о том, что распределение более острое (с более высокой концентрацией вокруг центра) по сравнению с нормальным. Это часто встречается в медицинских данных.

Стандартное отклонение. 3,8 дня при среднем 7,2 дня – довольно большой разброс. Это говорит о том, что длительность лечения сильно варьируется от пациента к пациенту.

Доверительный интервал. Уровень надежности 95% означает, что мы можем быть на 95% уверены, что истинное среднее значение в генеральной совокупности находится в интервале от 6,96 до 7,44 дня.

12. Сравнение нескольких показателей

Описательная статистика может анализировать несколько столбцов одновременно. Например, у нас есть данные по длительности лечения в трех отделениях. Указываем входной интервал, охватывающий все три столбца. Excel выведет отдельную таблицу для каждого столбца.

Это позволяет быстро сравнить распределения показателей в разных группах.

13. Практический совет: используйте описательную статистику для проверки данных

Описательная статистика – это еще и мощный инструмент контроля качества данных. Вот на что обращать внимание.

- Минимум и максимум. Если минимум отрицательный для показателя, который не может быть отрицательным (возраст, койко-дни), или максимум аномально высок (возраст 150 лет), это сигнал об ошибках ввода.

- Количество наблюдений (Счет). Если оно меньше ожидаемого, возможно, в данных есть пропуски, которые не были обработаны.

- Асимметрия и эксцесс. Аномально высокие значения могут указывать на наличие выбросов, которые искажают распределение.

Запускайте описательную статистику после первичной очистки данных: это поможет убедиться, что все в порядке, перед тем как строить отчеты и графики.

Главный вывод этой главы: описательная статистика в пакете анализа – это самый быстрый способ получить полную картину распределения данных. Вместо того чтобы вводить десятки формул, вы запускаете один инструмент и получаете все сразу. Это незаменимый помощник при первичном анализе новых данных и при подготовке к более сложным статистическим тестам.

Глава 6.3. Простые сравнения для отчетов

Описательная статистика показывает, как устроены данные. Но часто перед врачом-статистиком встают вопросы, требующие сравнения: «Значимо ли различается средняя длительность лечения в двух отделениях?», «Есть ли связь между возрастом пациента и длительностью лечения?», «Случайно ли снизилась летальность или это закономерность?».

Для ответа на такие вопросы в пакете анализа есть два ключевых инструмента: двухвыборочный t-тест (для сравнения средних в двух группах) и корреляция (для оценки связи между двумя показателями).

14. Двухвыборочный t-тест: сравниваем два отделения

t-тест (или t-критерий Стьюдента) – это статистический метод, который позволяет определить, значимо ли различаются средние значения в двух группах. В медицинской статистике он применяется постоянно:

- сравнение средней длительности лечения в двух отделениях;

- сравнение среднего возраста пациентов в двух группах (например, с осложнениями и без);

- сравнение показателей до и после внедрения нового стандарта лечения.

В пакете анализа есть несколько вариантов t-теста. Для большинства задач врача-статистика подходит «Двухвыборочный t-тест с одинаковыми дисперсиями» (если разброс в группах примерно одинаков) или «Двухвыборочный t-тест с различными дисперсиями» (если разброс разный).

Кейс: сравнение средней длительности лечения в кардиологии и хирургии

У нас есть данные по длительности лечения пациентов в двух отделениях: кардиологии (120 пациентов) и хирургии (95 пациентов). Нужно проверить, значимо ли различается средняя длительность лечения.

Шаг 1. Подготовка данных

Данные двух групп должны находиться в разных столбцах, например:

- столбце A: длительность лечения в кардиологии;
- столбце B: длительность лечения в хирургии.

Шаг 2. Запуск t-теста

1. На вкладке «Данные» нажмите «Анализ данных».
2. Выберите «Двухвыборочный t-тест с одинаковыми дисперсиями» (или с различными в зависимости от предварительной оценки разброса) и нажмите ОК.

[Место для рисунка: окно выбора инструментов с выделенным t-тестом]

Шаг 3. Настройка параметров

В открывшемся окне укажите:

- интервал переменной 1: диапазон данных кардиологии (например, A2:A121);
- интервал переменной 2: диапазон данных хирургии (например, B2:B96);

- поставьте галочку «Метки», если в первой строке есть заголовки;
- альфа: уровень значимости (0,05 – стандартный для медицинских исследований);
- выходной интервал: укажите, куда вывести результат (рис. 6.4).

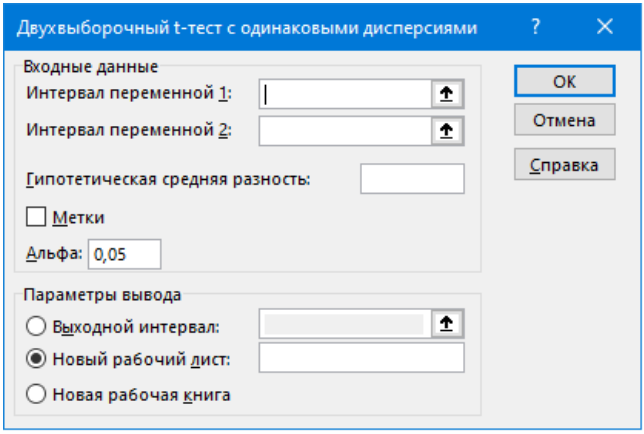


Рисунок 6.4. Окно настройки t-теста с заполненными полями

Шаг 4. Интерпретация результатов

Excel выводит таблицу с результатами (табл. 6.2).

Таблица 6.2. Ключевые значения

Показатель	Кардиология, дни	Хирургия, дни
Среднее	8,5	6,2
Дисперсия	4,2	3,8
Наблюдения	120	95
t-статистика	4,78	
t-критическое двухстороннее	1,97	

Как интерпретировать

- Если |t-статистика| > t-критическое, различие между группами статистически значимо (p <0,05).

- Если |t-статистика| < t-критическое, различие незначимо (p >0,05).

В нашем примере 4,78 > 1,97. Значит, средняя длительность лечения в кардиологии (8,5 дня) значимо выше, чем в хирургии (6,2 дня). Это не случайное колебание, а статистически подтвержденный факт.

Что дает этот результат для управленческого решения. Можно обоснованно говорить, что кардиологические пациенты требуют более длительного лечения. Это может влиять на планирование коечного фонда, расчет нагрузки на персонал и финансовое планирование.

Когда использовать t-тест

- Сравнение показателей до и после внедрения стандарта (здесь используется парный t-тест – другой вариант).

- Сравнение двух отделений, двух групп пациентов, двух периодов.

- Проверка, не случайны ли различия, которые вы видите в сводных таблицах.

15. Корреляция: связь между возрастом и длительностью лечения

Корреляционный анализ показывает, есть ли связь между двумя числовыми показателями, и, если есть, насколько она сильна и в каком направлении. Коэффициент корреляции (r) может принимать значения от -1 до +1:

- r от 0 до +1: положительная связь (чем больше один показатель, тем больше другой);

- r от 0 до -1: отрицательная связь (чем больше один, тем меньше другой);

- r близкий к 0: связи нет.

В медицинской статистике корреляция используется для поиска факторов, влияющих на исходы, и для проверки гипотез о взаимосвязях (табл. 6.3).

Таблица 6.3. Как интерпретировать силу корреляции

r	Интерпретация
0,0–0,3	Слабая связь
0,3–0,7	Умеренная связь
0,7–1,0	Сильная связь

Для медицинских данных коэффициент 0,65 – это умеренная, но важная связь. Она может быть аргументом для создания отдельных протоколов лечения для пожилых пациентов.

Когда использовать корреляцию

- Поиск факторов, влияющих на длительность лечения, летальность, частоту осложнений.
- Проверка гипотез о связи между лабораторными показателями и исходами.
- Подготовка данных для более сложного регрессионного анализа.

16. Практический совет: что означают p-значения (и почему они важны)

В результатах t-теста вы можете увидеть не только t-критическое, но

и p-значение (например, p = 0,0002). Это более точный показатель статистической значимости.

- p <0,05: различие статистически значимо (вероятность ошибки менее 5%).
- p >0,05: различие незначимо (может быть случайным).

Чем меньше p, тем убедительнее результат. В медицинских исследованиях обычно принимают порог 0,05. Если p = 0,0002, это очень сильное доказательство того, что различия не случайны.

В отчетах для руководства можно формулировать так: «Различие средней длительности лечения между кардиологическим и хирургическим отделениями статистически значимо ($p < 0,05$)». Это звучит профессионально и убедительно.

Главный вывод этой главы: t-тест и корреляция – это простые, но мощные инструменты, которые позволяют перейти от описания данных к проверке гипотез и обоснованию выводов. Освоив их, вы сможете отвечать на вопросы руководства не предположениями, а статистически подтвержденными фактами.

Заключение



Уважаемые коллеги!

Мы прошли долгий путь, от хаотичных выгрузок из медицинских информационных систем до полностью автоматизированных отчетов, наглядных дашбордов и статистически обоснованных выводов. Каждая глава этой книги была посвящена инструменту, который делает вашу работу точнее, быстрее и профессиональнее.

Давайте еще раз кратко пройдем по главным вехам этого пути.

Часть 1. Гигиена данных. Мы поняли главное: качество анализа определяется качеством данных. Научились превращать плохие таблицы в умные, удалять дубликаты, приводить текст к единому виду, исправлять типы данных и контролировать качество с помощью условного форматирования. Эти навыки – фундамент, без которого все дальнейшие расчеты теряют смысл.

Часть 2. Описательная статистика. Освоили «математику здравого смысла»: считаем итоги, средние величины, медианы, структуру и доли, оцениваем разброс данных. Поняли, когда среднее арифметическое обманывает, а когда на помощь приходит медиана. Научились видеть за цифрами реальные процессы.

Часть 3. Главные инструменты аналитика. Сводные таблицы, формулы с условиями, вертикальный поиск и структурированные ссылки – этот набор превращает вас из человека с калькулятором в аналитика, способного отвечать на сложные вопросы за секунды. Вы научились группировать данные, связывать таблицы и создавать динамические формулы, которые не ломаются при добавлении новых строк.

Часть 4. Визуализация данных. Цифры становятся аргументами, когда они правильно представлены. Вы научились выбирать правильный тип графика для каждой задачи, настраивать его для презентации, использовать линии тренда

для прогнозирования и создавать интерактивные дашборды, которые позволяют руководству видеть картину целиком.

Часть 5. Power Query. Это инструмент, который выводит вашу работу на новый уровень. Вы настроили автоматическую очистку данных, научились объединять десятки файлов в один отчет, подтягивать справочники и настраивать обновление по расписанию. Теперь то, что раньше занимало часы, делается за минуты.

Часть 6. Пакет анализа. Бонус, который открывает дверь в мир углубленной статистики. Вы узнали, как включить этот мощный инструмент, как одним кликом получить полную описательную статистику, как сравнивать группы с помощью t-теста и оценивать связи с помощью корреляции. Теперь вы можете не только описывать данные, но и доказывать свои выводы.

Как поддерживать навык

Освоить инструменты – это полдела. Главное – сохранить и развить навык, сделать его частью повседневной работы. Вот несколько советов, которые помогут вам не растерять приобретенное.

1. Используйте новые инструменты в рутине

Начните с малого. В следующий раз, когда вам понадобится посчитать количество пациентов по отделениям, не делайте это вручную – постройте сводную таблицу. Когда придет новая выгрузка, не чистите ее вручную, настройте запрос в Power Query. Каждая рутинная задача – это возможность отточить навык.

2. Создайте свою библиотеку шаблонов

Сохраняйте файлы с настроенными запросами, сводными таблицами, дашбордами. В следующий раз, когда понадобится похожий отчет, вы не начнете с нуля: вы возьмете шаблон и адаптируете его под новые данные. Это экономит время и гарантирует качество.

3. Документируйте свои настройки

Power Query запоминает шаги, но не всегда понятно, зачем был сделан тот или иной шаг. Добавляйте комментарии в шаги (правой кнопкой → «Свойства» → «Описание»). Через месяц вы сами будете благодарны себе за это.

4. Учитесь у коллег и учите коллег

Обменивайтесь опытом с коллегами-статистиками. Покажите, как вы автоматизировали ежемесячный отчет. Посмотрите, как это сделали они. Совместное обсуждение решений часто приводит к новым идеям и оптимизациям.

5. Не бойтесь экспериментировать

Excel и Power Query прощают ошибки. Вы всегда можете отменить действие, вернуться к предыдущему шагу, попробовать другой подход. Самый надежный способ освоить инструмент – пробовать решать с его помощью реальные задачи, даже если сначала не все получается.

6. Следите за обновлениями

Microsoft постоянно развивает Excel и Power Query. В новых версиях появляются функции, которые могут упростить вашу работу. Например, функция ПРОСМОТРX (XLOOKUP) значительно удобнее классического ВПР. Будьте в курсе новинок.

7. Главное не инструмент, а цель

Помните: вы не осваиваете Excel ради Excel. Вы осваиваете инструменты, чтобы быстрее и точнее отвечать на вопросы руководства, обосновывать управленческие решения, улучшать качество медицинской помощи. Каждый настроенный отчет, каждый автоматизированный процесс – это вклад в вашу профессиональную репутацию и в итоге в здоровье пациентов.

Вместо напутствия



Уважаемые коллеги!

Медицинская статистика – это не скучная обязанность, не бумажная работа, которую приходится делать, потому что так надо. Это мощный инструмент управления, который позволяет видеть ситуацию объективно, находить точки роста, обосновывать решения и доказывать эффективность.

Вы врачи-организаторы, врачи-статистики, заведующие отделениями, заместители главных врачей. Вы те, кто превращает горы цифр в понятные управленческие решения. И от того, насколько точно и быстро вы это делаете, зависит качество медицинской помощи в вашем учреждении.

Книга, которую вы держите в руках (или читаете на экране), задумывалась как практическое руководство, а не как теоретический трактат. Мы сознательно ушли от сложных формул и абстрактных рассуждений в сторону пошаговых инструкций, реальных медицинских кейсов и готовых шаблонов. Надеюсь, это помогло вам не просто узнать о возможностях Excel, но и начать применять их в своей повседневной работе.

Теперь слово за вами. Открывайте свои выгрузки, применяйте новые инструменты, экспериментируйте, автоматизируйте. И помните: цифры не врут, когда их считает тот, кто знает, как это делать правильно.

Спасибо, что были с нами. Успехов в работе!

Автор выражает благодарность коллегам из медицинских организаций, чьи реальные кейсы легли в основу примеров этой книги. Ваша ежедневная работа – лучший источник вдохновения для таких пособий.

Информация об авторе

Фетисов Александр Георгиевич, канд. тех. наук, доцент кафедры информационных систем и технологий в здравоохранении ГБУ города Москвы «Научно-исследовательский институт организации здравоохранения и медицинского менеджмента Департамента здравоохранения города Москвы», ORCID: 0000-0003-2106-9122.

Научное электронное издание

АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ В РАБОТЕ ВРАЧА-ОРГАНИЗАТОРА И ВРАЧА-СТАТИСТИКА: ПРАКТИЧЕСКОЕ РУКОВОДСТВО

Фетисов Александр Георгиевич

Корректор: Л.И. Базылевич
Дизайнер-верстальщик: А. И. Кораблева

Объем данных: 2,9 МБ
Дата подписания к использованию: 13.11.2026
URL: https://niioz.ru/moskovskaya-meditsina/izdaniya-nii/metodicheskie-posobiya/

ГБУ «НИИОЗММ ДЗМ»,
115088, г. Москва, ул. Шарикоподшипниковская, д. 9
Тел.: +7 (495) 530-12-89
Электронная почта: niiozmm@zdrav.mos.ru

НИИ
ОРГАНИЗАЦИИ
ЗДРАВООХРАНЕНИЯ
И МЕДИЦИНСКОГО
МЕНЕДЖМЕНТА

МОСКВА
2026