**Excel. Биржевая диаграмма, она же блочная, она же ящичная**

В одном из [комментариев](http://baguzin.ru/wp/?page_id=718#comments) на сайте меня попросили рассказать про ящичные диаграммы. На мой взгляд, эти диаграммы в повседневной офисной практике используются незаслуженно редко. И тому я вижу несколько объяснений:

* в Excel эти диаграммы носят название *биржевые* (рис. 1), что, казалось бы, сужает область их применения до специальных сфер деятельности;
* большинство менеджеров видят мир детерминированным, и потому использование диаграмм, отражающих вероятностный подход, даже не приходит им в голову;
* возможности Excel в построении таких диаграмм ограничены, а установка надстроек может вызывать затруднения.

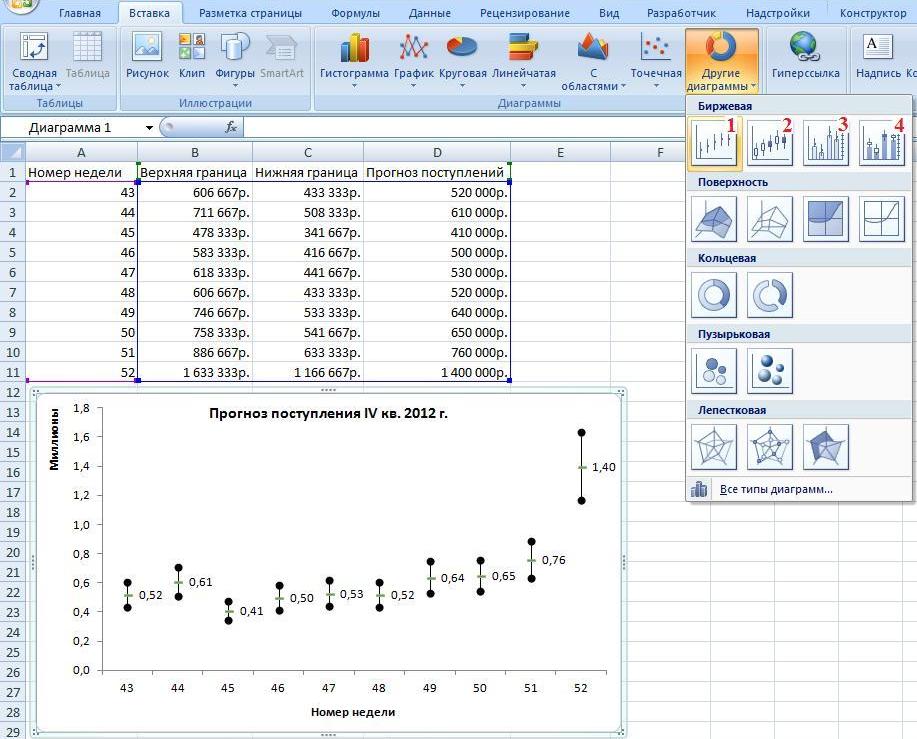


Рис. 1. Меню выбора биржевой диаграммы

Можно выделить следующие области применения ящичных диаграмм[[1]](#footnote-1):

* изменение цен акций и объемов торгов;
* анализ статистических данных; например, в метеорологии или менеджменте качества;
* представление результатов маркетинговых исследований, социологических опросов;
* составление прогнозов.

В Excel доступны четыре типа биржевых диаграмм (см. нумерацию на рис. 1), содержащих от 3 до 5 набора данных:

|  |  |
| --- | --- |
| № на рис. 1 | Ряды данных в порядке их расположения |
| 1 | Максимальное значение – минимальное – закрытие |
| 2 | Открытие – максимальное – минимальное – закрытие |
| 3 | Объем – максимальное – минимальное – закрытие |
| 4 | Объем – открытие – максимальное – минимальное – закрытие |

В качестве категорий (ось Х) можно использовать даты или названия (например, акций). Так на рис. 1 использован первый тип диаграммы, и данные расположены в порядке: верхняя граница – нижняя граница – прогноз поступлений. А в качестве категории используется номер недели.

При использовании четырех наборов данных (тип 2) диаграмма оправдывает свое альтернативное название, так как на ней изображаются ящички/блоки. В качестве примера я взял [многолетние данные](http://ru.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B8%D0%BC%D0%B0%D1%82_%D0%9C%D0%BE%D1%81%D0%BA%D0%B2%D1%8B) наблюдений за температурой в Москве (рис. 2). Каждый «бочонок» соответствует совокупности наблюдений за один месяц в течение многих лет (если быть точным, 133 лет).

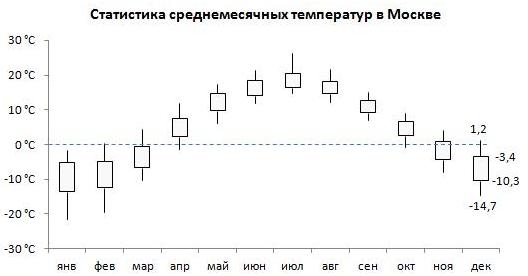


Рис. 2. Биржевая диаграмма второго типа; статистика среднемесячных температур в Москве за период наблюдений (1879–2012 гг.); на примере декабря: низ «усика» соответствует минимальной среднемесячной температуре, низ блока соответствует значению на одно стандартное отклонение ниже среднего (μ – σ), верх блока = μ + σ, верх «усика» – максимальная среднемесячная температура; в интервал μ ± σ, то есть внутрь блока, попадает 68,3% всех наблюдений

*Внимание!* Excel настроен на построение именно *биржевых* диаграмм. Из-за этого мастер диаграмм не всегда справляется с построением диаграммы, на основе данных, содержащих отрицательные значения. Например, если на рис. 2 вы возьмете только данные за январь, февраль и март, то Excel «ругнется», что размещение данных неверное. Если ваши данные содержат отрицательные числа, а строк меньше, чем параметров (которых, напомню, используется от 3 до 5), просто продублируйте строки, чтобы их стало больше, чем параметров, постройте диаграмму, а затем уменьшите ее область построения и удалите лишние строки.

Если значение «открытие» больше значения «закрытие», на биржевой диаграмме второго типа блок будет закрашен (рис. 3).

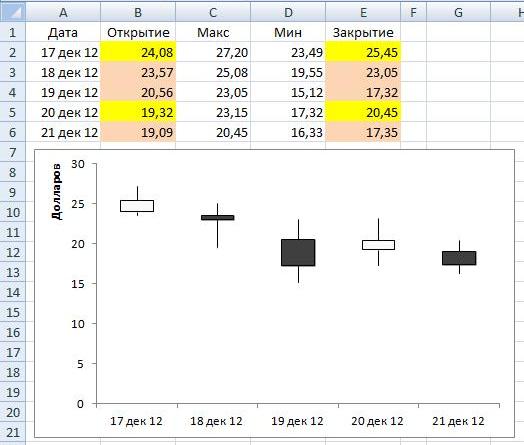


Рис. 3. Биржевая диаграмма второго типа; динамика стоимости акций; 17 и 20 декабря цена закрытия больше цены открытия – ящички светлые, 18, 19 и 21 декабря цена закрытия меньше цены открытия – ящички темные

Третий и четвертый тип биржевой диаграммы в Excel аналогичен первому и второму типу плюс еще один параметр – объем торгов (размер выборки и т.п.). Поместите столбец с объемом сразу за столбцом категорий. Соответствующая диаграмма имеет вторую ось ординат, для отражения объема торгов или его аналога (рис. 4). Левая ось ординат для объема торгов, правая – для цены акций.

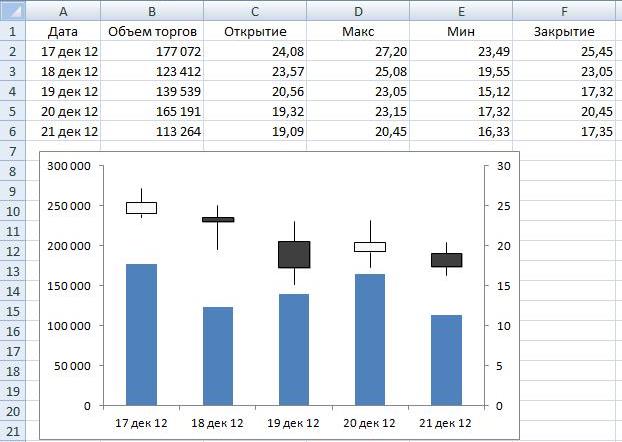


Рис. 4. Биржевая диаграмма четвертого типа; динамика торгов и стоимости акций

Использование блочных диаграмм в статистике было известно задолго до появления Excel. Английский термин – box-and-whisker diagram (диаграмма *ящик с усами*). Блочная диаграмма представляет собой удобное средство для изображения пяти базовых статистических показателей. Например, на рис. 5 изображена блочная диаграмма, иллюстрирующая показатели среднегодовой доходности 15 высокорисковых фондов.[[2]](#footnote-2)

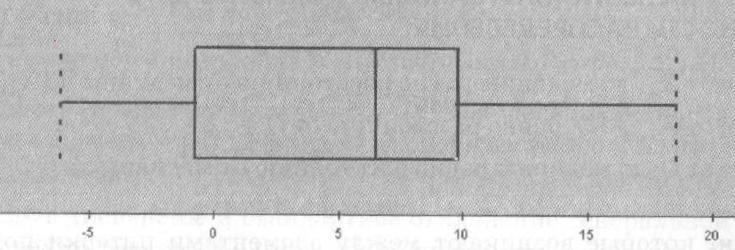


Рис. 5. Блочная диаграмма, иллюстрирующая показатели доходности 15 фондов с очень высоким уровнем риска; по оси Х среднегодовая доходность в процентах

Вертикальная линия, проведенная внутри прямоугольника, отмечает медиану. Левая сторона прямоугольника соответствует первому квартилю, Q1, а правая сторона — третьему квартилю, Q3. Таким образом, прямоугольник содержит средние 50% элементов выборки. Младшие 25% данных изображаются в виде линии (так называемый *ус*), соединяющей левую сторону прямоугольника с наименьшим выборочным значением Xmin. Следовательно, старшим 25% данных соответствует линия, соединяющая правую сторону прямоугольника с наибольшим выборочным значением Xmax.

Блочная диаграмма, представленная на рис. 5, демонстрирует, что показатели среднегодовой доходности 15 фондов с очень высоким уровнем риска имеют практически симметричное распределение, поскольку расстояние между медианой и наибольшим значением приблизительно равно расстоянию между наименьшим значением и медианой. Однако другие характеристики распределения указывают на несимметричность. Правый ус диаграммы длиннее левого, поскольку выборка содержит выброс, равный 18,5% (правая граница), а медиана расположена ближе к правой стороне диаграммы, чем к левой.

На рис. 6 изображены четыре типа распределений, а также соответствующие им блочные диаграммы.

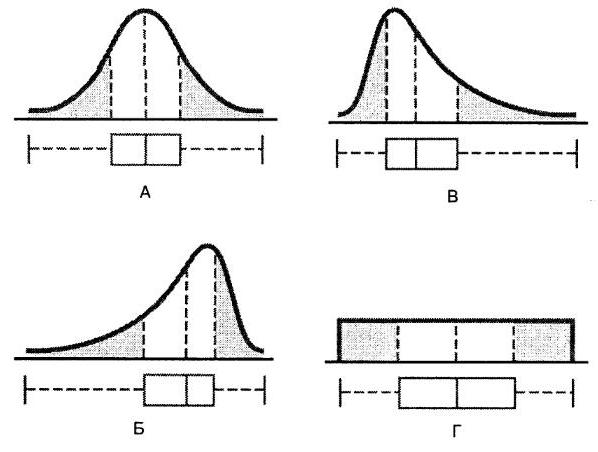


Рис. 6. Четыре гипотетических распределения, исследованных с помощью блочной диаграммы. Область, расположенная под каждым ящиком, разбита квартилями, входящими в пятерку базовых показателей. А – колоколообразное распределение, Б – распределение с отрицательной асимметрией, В – распределение с положительной асимметрией, Г – прямоугольное распределение.

Если данные распределены совершенно симметрично, как на рис. 6А и 6Г, среднее выборочное значение и медиана совпадают. Кроме того, длина левого уса равна длине правого, а линия медианы проходит через середину прямоугольника.

Если распределение данных имеет отрицательную асимметрию, как на рис. 6Б, среднее выборочное значение смещается вдоль левого хвоста. Отрицательная асимметрия проявляется в виде высокой концентрации данных в правой половине шкалы. При этом 75% всех данных расположены между левой стороной прямоугольника (первый квартиль, Q1) и концом правого уса (наибольшее выборочное значение, Xmax). Следовательно, вдоль длинного левого уса распределены всего 25% данных. Это свидетельствует о сильной асимметрии распределения.

Если распределение данных имеет положительную асимметрию, как на рис. 6В, пик распределения смещается влево. Теперь 75% всех данных расположены между началом левого уса (наименьшее выборочное значение, Xmin) и правой стороной прямоугольника (третий квартиль, Q3). Остальные 25% данных распределены вдоль длинного правого уса.

К сожалению, стандартные средства Excel не позволяют построить блочную диаграмму с пятью базовыми показателями статистики: Xmin, Q1, медиана, Q3, Xmax. Можно воспользоваться надстройкой [PHStat2](http://www.pearsonhighered.com/phstat/) или аналогичными. Я «исхитрился» и построил такую диаграмму, правда, не идеально красивую.

В качестве исходной возьмите диаграмму, как на рис. 2. Добавьте к данным еще один ряд – средние значениия μ. Чтобы сделать это, для начала выделите диаграмму и правой кнопкой мыши вызовите контекстное меню (рис. 7А).

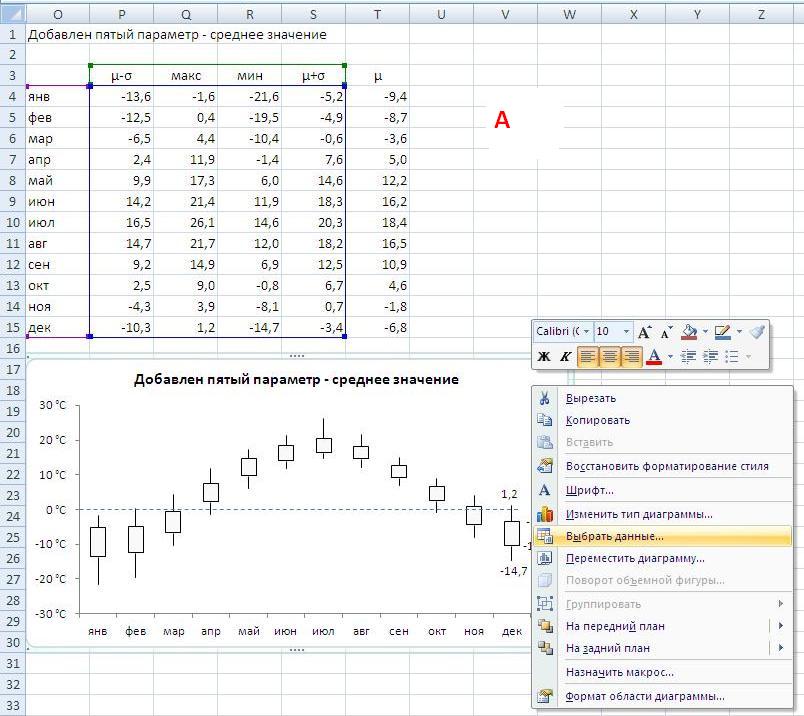


Рис. 7А. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; А – выбрать данные

В открывшемся окне «Выбор источника данных» нажмите «Добавить» (рис. 7Б).



Рис. 7Б. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; Б – добавить ряд

В открывшемся окне «Изменение ряда» выберите имя ряда и значения (рис. 7В). Нажмите 2 раза Ok.

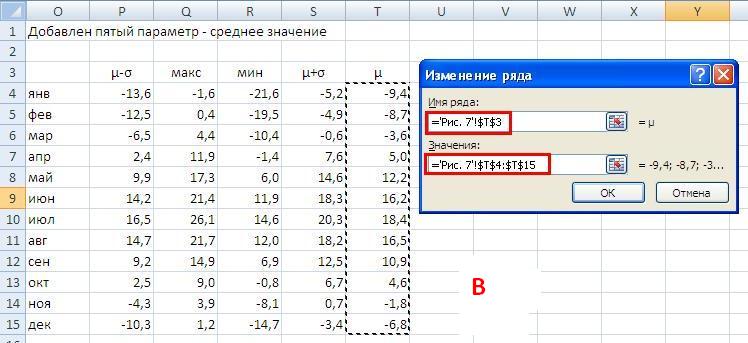


Рис. 7В. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; В – выбрать имя ряда и значения

Ряд на диаграмме добавился, но он не виден. Это связано с тем, что по умолчанию на биржевой диаграмме ряды не имеют ни цвета линий, ни маркеров. Все видимые элементы диаграммы выполнены с помощью полос повышения и планок погрешности; см. вкладку Excel «Работа с диаграммами» – «Макет», область Анализ (выделено в правой верхней части рис. 7Г). Обратите также внимание, что после добавления ряда μ ящички уменьшились по высоте (сравните размер ящичков на рисунках 7А и 7Г). Это связано с еще одной особенностью биржевых диаграмм – низ ящичков всегда соответствует значениям первого ряда. Первого – в окне «Выбор источника данных». В нашем случае (см. рис. 7Б) это ряд «μ–σ». Верх ящичков соответствует значениям последнего ряда. В нашем случае сначала это был ряд «μ+σ» (см. рис. 7Б), а после добавления ряда «μ», именно он стал последним. Чтобы исправить ситуацию, надо просто изменить порядок рядов в окне «Выбор источника данных». Откройте это окно еще раз, выделите ряд «μ», и передвиньте его вверх с помощью стрелки ↑. Ящички вернуться к исходному размеру.

Теперь, чтобы отформатировать вновь созданный ряд средних, выделите диаграмму, и пройдите по меню Работа с диаграммами – Макет. В области «Текущий фрагмент» щелкните на «Область диаграммы» и выберите ряд «μ» (рис. 7Г).

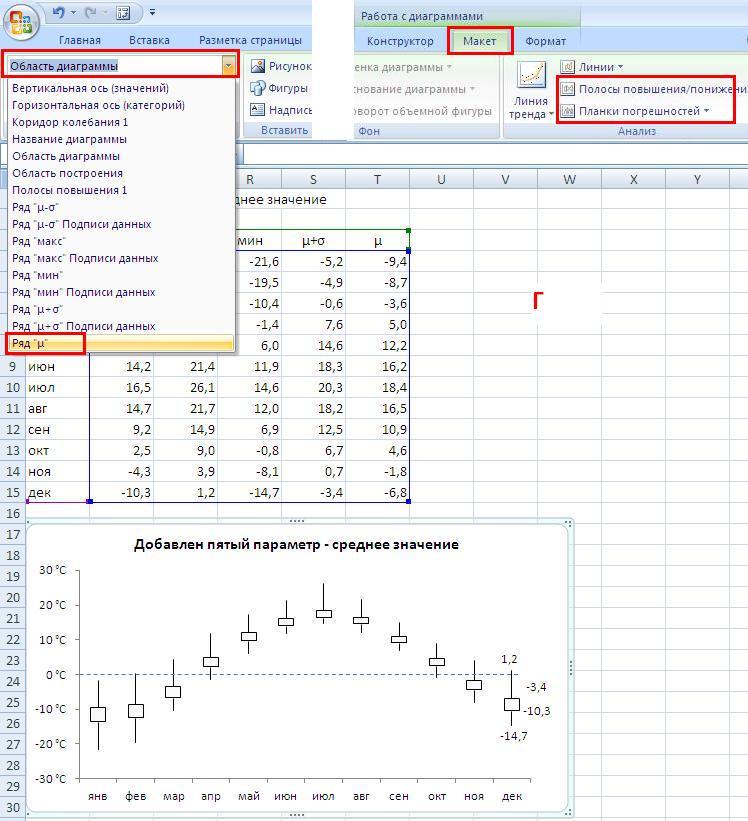


Рис. 7Г. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; Г – активировать новый ряд на диаграмме

Ряд «μ» выделился (рис. 7Д). Это видно, во-первых, по тому, что на диаграмме появились точечки вокруг ряда (пока невидимого), а на листе выделен диапазон Т3:Т15. Щелкните кнопку «Формат выделенного фрагмента».

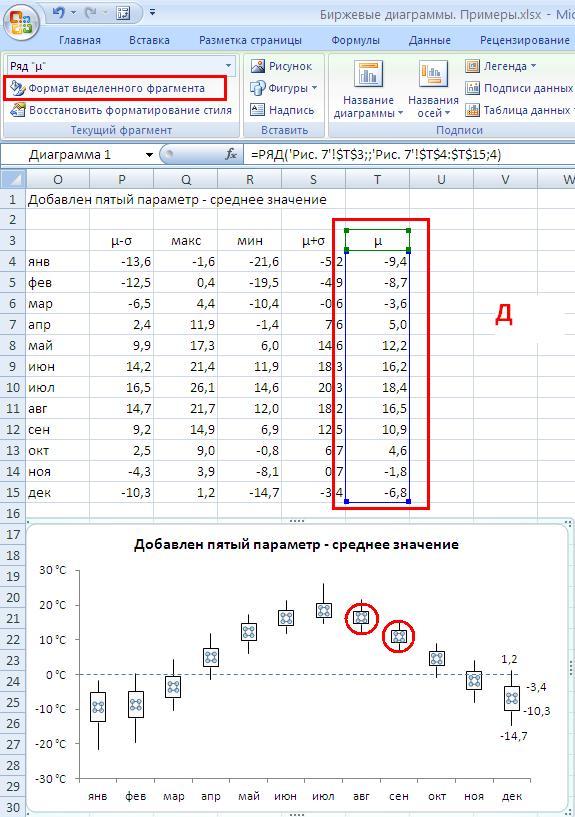


Рис. 7Д. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; Д – формат выделенного фрагмента

В открывшемся окне «Формат ряда данных», установите Цвет линии маркера – Нет линий, Заливка маркера – Сплошная, цвет – черный, Параметры маркера – как на рис. 7Е

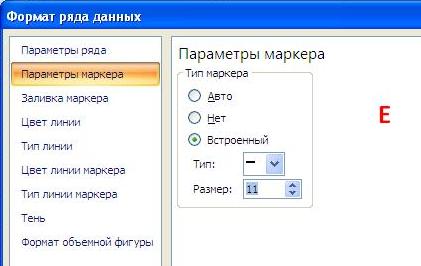


Рис. 7Е. Добавляем на стандартную биржевую диаграмму (тип 2) ряд средних значений; Е – параметры маркера

Итог усилий представлен на рис. 7. Стоит ли она затраченных усилий, решать вам.

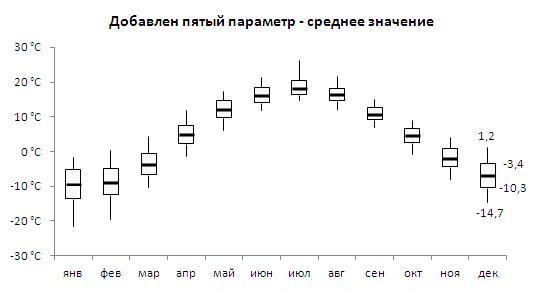


Рис. 7. Ящичная диаграмма с пятью базовыми показателями статистики: Xmin, Q1, медиана, Q3, Xmax.

\* \* \*

На закуску рассмотрим пример использования блочных диаграмма в практике «шесть сигм» (это одна из концепций менеджмента качества; для первого знакомства см., например, [Пит Панде, Ларри Холп. Что такое «шесть сигм»?](http://baguzin.ru/wp/?p=2405)).

Допустим, ваша задача выявить из «моря» факторов (X1, X2… Xn) тот, который оказывает наибольшее влияние на объем производства (Y). Для начала соберите все данные, относящиеся к делу (рис. 8).

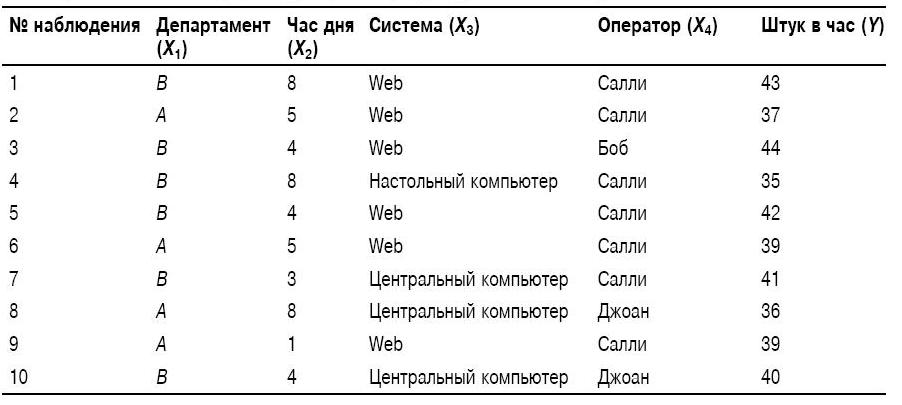


Рис. 8. Форма для регистрации данных наблюдений

Чтобы узнать, влияет ли наблюдаемый входной фактор (Xm) на выходной (Y), следует создать несколько диаграмм «ящик с усами» для выходного фактора таким образом, чтобы каждая отдельная диаграмма соответствовала одному условию входной переменной.

Данные на рис. 8 — это часть большого списка данных, собранных для одного процесса. Ключевой результат (Y) показывает часовую производительность в штуках. Возможные входные факторы влияния: департамент, выполняющий транзакцию (Х1), час дня, когда транзакция была выполнена (Х2), вид использованной рабочей системы (Х3) и имя работника, выполнившего транзакцию (Х4).

Какой эффект оказывает конкретный работник (Х4) на производительность (Y)? На рис. 9 представлен набор ящичных диаграмм Y для каждого значения входного фактора Х4.

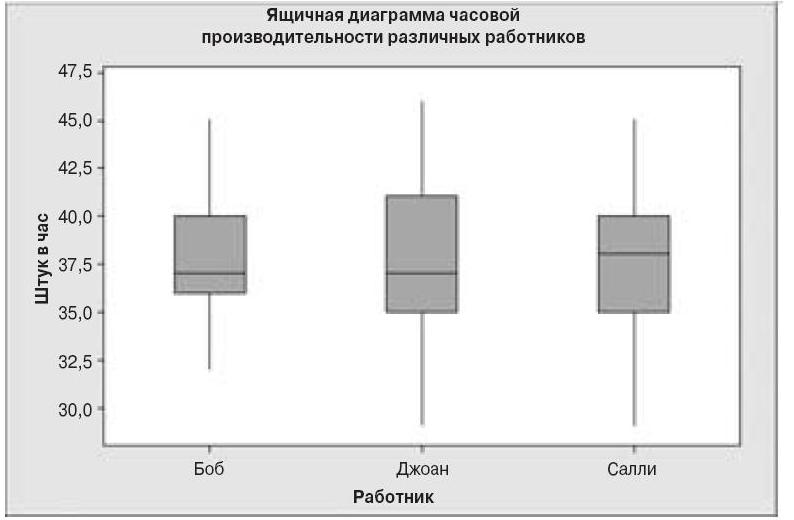


Рис. 9. Влияние каждого отдельного работника на часовую производительность

Зависит ли количество выполненных за час транзакций от того, кто делан эту работу – Боб, Джоан или Салли? Из рис. 9 ясно видно, что операторы почти не разнятся между собой по производительности. У них приблизительно один и тот же средний уровень и вариация одной и той же величины. Из этого следует вывод, что переменная Х4 (работник) не является ключевым фактором влияния на вариацию результата.

Что можно сказать о зависимости результата от того, какой департамент (Х3) выполняет транзакции? На рис. 10 изображена еще одна группа диаграмм «ящик с усами» для двух департаментов.

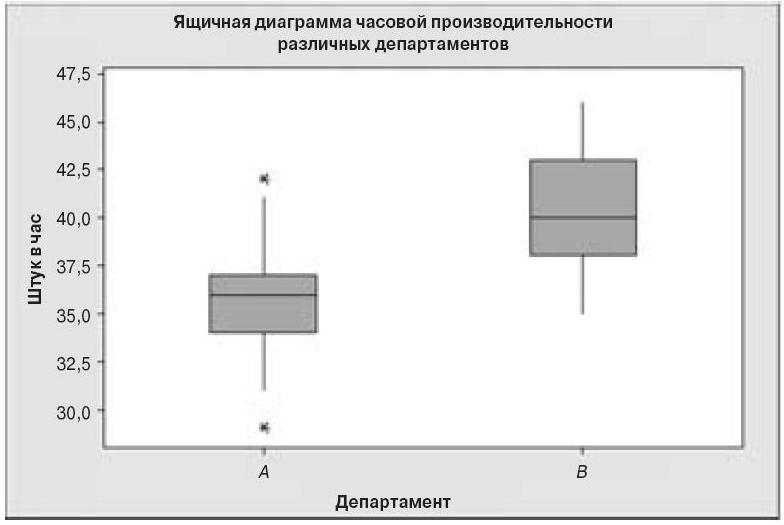


Рис. 10. Влияние каждого департамента на часовую производительность

Видно, что разница между центрами вариации для департаментов А и В существенна по сравнению со средней шириной вариации внутри департаментов. Вывод: от того, какой департамент выполняет транзакции, действительно зависит часовое количество последних.

1. Чтобы разнообразить изложение, я буду употреблять все три названия диаграмм в качестве синонимов. При написании заметки использованы материалы книг: Джон Уокенбах Microsoft Office Excel 2007. Библия пользователя, стр. 391, 392, Дэвид Левин и др. Статистика для менеджеров с использованием Microsoft Excel, стр. 214–217, Крейг Джиджи и др. Шесть сигм для «чайников», стр. 158–160, а также учебные материалы Microsoft: [Представление данных в биржевой диаграмме](http://office.microsoft.com/ru-ru/excel-help/HA001233750.aspx), [Создание биржевой диаграммы](http://office.microsoft.com/ru-ru/excel-help/HA001117942.aspx). [↑](#footnote-ref-1)
2. Из книги Д. Левин и др. Статистика для менеджеров с использованием Microsoft Excel, стр. 214. [↑](#footnote-ref-2)