И. ВЫСОЦКИЙ,
г. Москва

ТЕМА: «ПРЕДСТАВЛЕНИЕ ДАННЫХ»

СЦЕНАРИИ УРОКОВ

Продолжение. Начало см. № 4, 5 за 2023 г.

Урок: «Диаграммы рассеивания»

Часть 1

Материалы представляют собой примерный сценарий урока по теме «Диаграммы рассеивания». Учитель может на свое усмотрение использовать сценарий целиком или частично наряду с собственными разработками и материалами учебника. Некоторые материалы сценария подготовлены учителем С.Н. Полункиной[1]. Авторы будут благодарны за замечания и предложения по структуре и содержанию сценариев.

Оборудование: линейка или угольник, цветные карандаши, миллиметровая бумага, калькулятор.

Цель урока: ознакомление учащихся с представлением информации на диаграммах рассеивания. Учащиеся должны приобрести навык построения диаграмм рассеивания, а также должно появиться представление о том, как связь между двумя изменчивыми числовыми величинами отражается графически на диаграмме рассеивания.

Диаграмма рассеивания помогает определить примерный характер связи между двумя числовыми величинами. Каждому наблюдению соответствует точка, координаты которой на плоскости равны значениям двух наблюдаемых величин.

Цель построения: визуальное восприятие двух величин, наблюдаемых совместно в одном опыте.

Задание 1. Рассмотрите с учащимися примеры диаграмм из учебника (с. 39–42, примеры 1, 3) и постройте с ними диаграммы рассеивания из упражнения № 6 (с. 43).

1. Есть ли связь между ростом и весом человека? Для наглядного ответа на этот вопрос построим диаграмму рассеивания. Данными для этой диаграммы служит набор пар чисел.

Каждая пара — это рост и вес одного человека. В таблице приведены значения роста и веса 15 юношей.

Рост (см)	167	169	179	178	177	175	171	181
Вес (кг)	62	67	70	72	70	69	63	80

Рост (см)	174	175	180	174	172	178	171
Вес (кг)	73	66	75	70	67	74	66

[1] Текст статьи С.Н. Полункиной «Диаграммы рассеивания». — Режим доступа: ptlab.mccme.ru/sites/ptlab.mccme.ru/files/diagrammy_rasseivaniya.pdf.

3. Самочувствие человека во многом определяется температурой тела и артериальным давлением. По данным обследования 25 человек в больнице построена диаграмма рассеивания для температуры и давления (рис. 1).

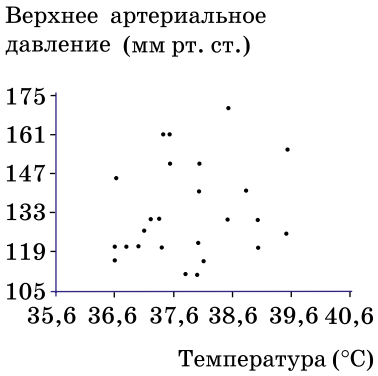


Рис. 1

6. Мальчики на соревнованиях прыгали в длину с места и бежали 60 м. Их результаты приведены в таблице.

Прыжок (см)	180	194	190	215	210
Бег (с)	10,8	10,2	10,6	9,6	10,2

Прыжок (см)	170	175	202	205
Бег (с)	11,0	11,6	10,4	10,0

Прыжок (см)	195	205	200	190	186
Бег (с)	11,0	9,8	10,6	10,8	10,7

Постройте диаграмму рассеивания. Можно ли утверждать, что результаты прыжков с места связаны со скоростью бега на 60 м?

В результате разбора примеров и построения диаграмм у учащихся получаются три диаграммы рассеивания (рис. 2).

Совокупность точек на каждой диаграмме образует скопление (облако, «созвездие»). По форме скопления можно предположить наличие или отсутствие связи между этими величинами.

Желательный результат обсуждения. Удобно обвести облако фломастером или карандашом так, чтобы все (или почти все) точки попали внутрь овала. Если облако круглое или вытянутое по горизонтали или вертикали (пример 3: «Температура — давление»), то можно предположить, *что связи между величинами нет* — значения одной из них не влияют на значения другой.

Если облако имеет явный наклон, то важно, в какую сторону наклонено облако. Если облако вытянуто вправо-вверх (пример 1: «Рост — вес), это говорит о том, что чем больше одна величина, тем больше вторая. Тогда говорят, что между величинами *наблюдается положительная связь*.

Если же область точек вытянута вправо-вниз (диаграмма упражнения 6), то чем больше первая величина, тем меньше вторая, и наоборот. В этом случае *связь называют отрицательной*.

Предложите учащимся построить на листе миллиметровой бумаги диаграммы рассеивания по таблицам в заданиях 2 и 3.

Задание 2. В таблице представлены данные о протяженности и количестве станций линий Московского метрополитена по состоянию на сентябрь 2018 года. Постройте диаграмму рассеивания по данным о линиях московского метро, обведите скопление точек. Как вы думаете, связаны ли длина линии метро и количество станций на ней?

Линия	Длина (км)	Количество станций
Сокольническая	32,5	22
Замоскворецкая	42,8	23
Арбатско-Покровская	45,1	22
Филевская	14,9	13
Кольцевая	19,3	12
Калужско-Рижская	37,6	24
Таганско-Краснопресненская	42,2	23
Калининская	16,5	8
Солнцевская	24,9	12
Серпуховско-Тимирязевская	41,2	25
Люблинско-Дмитровская	38,3	23
Большая кольцевая	10,5	5
Каховская	3,3	3
Бутовская	10	7
Монорельс	4,7	6
МЦК	54	31

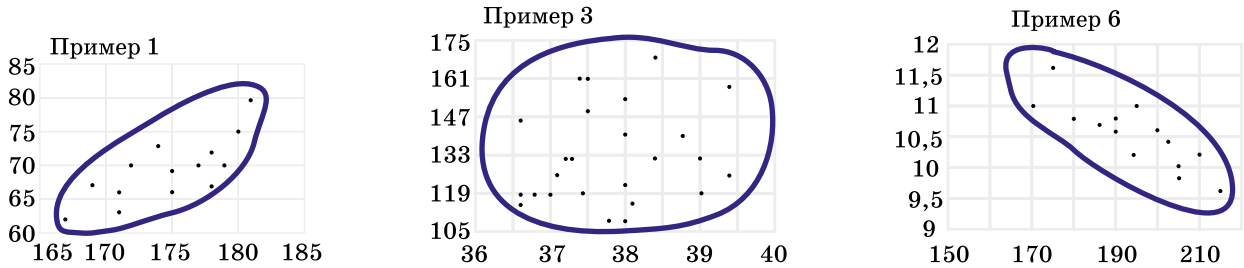


Рис. 2. Положительная связь «Рост — вес» (слева), отсутствие связи «Температура — давление» (в центре) и отрицательная связь «Длина прыжка — время забега» (справа)

Желательный результат обсуждения. Получившееся облако имеет явный наклон вправо-вверх — наблюдается положительная связь (рис. 3).

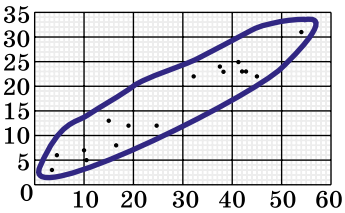


Рис. 3. Диаграмма «Протяженность линии — число станций»

Предложите учащимся сравнить эту диаграмму с диаграммой «Рост — вес» (задание 1). В обоих случаях величины положительно связаны, но в задаче о метрополитене точки плотнее расположены внутри облака, поэтому овал сильнее вытянут. Можно предположить, что связь между длиной линии метро и количеством станций сильнее (теснее), чем связь между ростом и весом человека. Покажите, как выглядят диаграммы с разной силой связи (рис. 4).

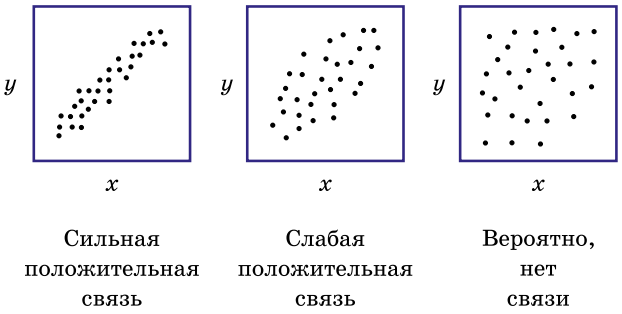


Рис. 4. Разный характер связи: от сильной до отсутствия

Иногда случается так, что связь есть, а на диаграмме ее не видно. И наоборот — может показаться, что связь есть, а на самом деле она отсутствует. Причины могут быть разными; чаще всего так бывает, если данных мало. Чем больше имеется наблюдений, тем более обоснованным будет предположение.

Иногда облако на диаграмме имеет не овальную, а более сложную форму. Это означает, что связь между величинами тоже сложная.

Задание 3. Завод выпускает стальные газовые трубы. В таблице приведены данные контрольных измерений изготовленных труб: диаметр и длина трубы массой 1 т. По этим данным постройте диаграмму рассеивания. Как вы думаете, есть ли связь между диаметром трубы и ее длиной?

Диаметр (мм)	Метров в 1 тонне	Диаметр (мм)	Метров в 1 тонне	Диаметр (мм)	Метров в 1 тонне
21,3	810	42,3	279	88,5	120
21,3	780	48	300	88,5	105
21,3	722	48	260	114	92
26,8	655	48	230	114	82
26,8	580	60	236	114	75
26,8	545	60	205	140	74
33,5	459	60	165	140	66
33,5	421	75,5	175	140	55
33,5	380	75,5	140	165	63
42,3	360	75,5	125	165	56
42,3	325	88,5	135	165	46

Желательный результат обсуждения. Посоветуйте учащимся для удобства округлять данные: диаметр — до целого числа миллиметров, длину — до целого числа метров, кратного 5 м. Невысокая точность не помешает главной цели — визуальному анализу.

На диаграмме (рис. 5) видно, что диаметр и длина трубы связаны (связь отрицательная), но облако данных в данном случае не овальное, а имеет изогнутую форму.

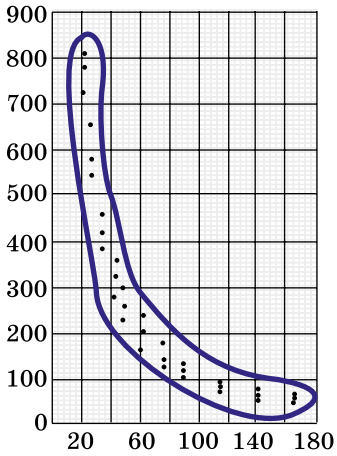


Рис. 5

На рисунке 6 схематично показано, как могут выглядеть диаграммы рассеивания при сложных связях: точки группируются не внутри овала, а внутри более сложной, часто невыпуклой фигуры.

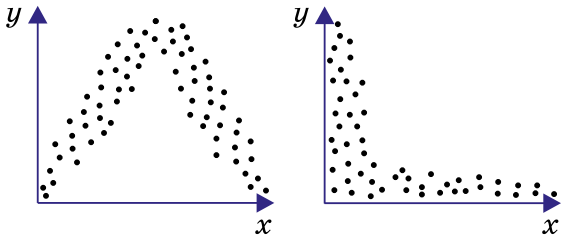


Рис. 6

№ п/п	Год выпуска	Пробег (тыс. км)	Цена (тыс. руб)	№ п/п	Год выпуска	Пробег (тыс. км)	Цена (тыс. руб)	№ п/п	Год выпуска	Пробег (тыс. км)	Цена (тыс. руб)
1	2015	26	1170	11	2015	95	707	21	2011	150	427
2	2010	40	410	12	2012	100	519	22	2012	156	495
3	2016	50	828	13	2014	100	580	23	2007	160	315
4	2014	70	698	14	2012	101	440	24	2008	170	395
5	2014	52	750	15	2013	105	515	25	2008	176	350
6	2013	60	615	16	2010	110	470	24	2009	210	355
7	2015	66	610	17	2011	120	584	27	2009	210	290
8	2016	70	865	18	2013	130	550	28	2007	285	230
9	2016	75	684	19	2009	135	329	29	2008	280	130
10	2016	93	735	20	2011	141	479				

Выводы. Чтобы увидеть, связаны ли между собой две изменчивые величины, и понять характер связи, зачастую удобно строить диаграммы рассеивания. Связь бывает положительной или отрицательной, сильной или слабой. Эти свойства обычно хорошо видны на диаграммах рассеивания.

Важно!

1. Наличие связи не означает, что одна величина обусловлена другой. Наблюдаемая связь далеко не всегда причинно-следственная. Механизм связи может быть очень сложным. Часто величины связаны косвенно. Например, рассматривая пару «Длина прыжка — время забега», нельзя сказать, что школьник дальше прыгает потому, что быстрее бегает, или наоборот. Общее физическое развитие, рост, тренированность обуславливают обе величины.

2. Диаграммы рассеивания не позволяют сделать абсолютно точный вывод, даже если наличие связи (или ее отсутствие) кажется очевидным. Ни один статистический метод не позволяет делать доказанные заключения. Диаграммы рассеивания так же, как и другие методы представления данных, позволяют лишь выдвинуть гипотезу о том, что связь есть, или дают основания усомниться в этом, наблюдая форму облака.

Задачи, рекомендуемые для решения на уроке: с. 42 № 1, № 5.

Рекомендуемое домашнее задание: с. 42 № 2–4.

Урок: «Диаграммы рассеивания»

Часть 2

Урок предполагает наличие у учащихся знаний о линейной функции, ее графике и угловом коэффициенте прямой. Если эти темы еще не изучены, то целесообразно перенести этот урок на более позднее время.

Оборудование: линейка или угольник, цветные карандаши или фломастеры, миллиметровая бумага, калькулятор.

Цель урока: формирование представления об оценке параметров линейных зависимостей с помощью диаграмм рассеивания. Учащиеся должны научиться проводить ось диаграммы и выражать линейную связь между величинами формулой.

Задание 4. На одном из популярных сайтов, посвященных продаже подержанных автомобилей, есть данные о ценах на автомобили «Шкода Октавия». В таблице (вверху страницы) приведены данные о 29 предложениях в октябре 2018 года.

Можно рассмотреть разные величины и их соответствующие значения. Рассмотрим, например, пару «Возраст машины — цена». Ясно, что чем старше автомобиль, тем он дешевле (ведь это не раритетные коллекционные машины). Цена каждый год снижается примерно на 10%[2]. Поэтому связь между ценой и возрастом нелинейная. Давайте рассмотрим связь между возрастом автомобиля на момент продажи и его пробегом. Очевидно, пробег автомобиля пропорционален его сроку службы. Если машина прослужила год, то она могла наездить, предположим, 20 000 км. Еще год — еще примерно столько же, если образ жизни водителя не поменялся, и так далее. Логично считать, что между величинами «возраст авто» и «пробег авто» должна быть положительная линейная связь. Построим диаграмму рассеивания и «на глаз» проверим, правдоподобно ли это предположение.

[2] Данные Российского союза страховщиков.

Желательный результат обсуждения. Сначала таблицу нужно сделать удобнее: добавим столбец «Возраст» вместо ненужного уже столбца «Цена». На построенной диаграмме рассеивания для пары «Возраст — пробег» (рис. 7) почти все точки образуют хорошо оформленное наклонное и вытянутое облако.

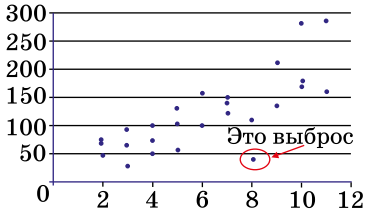


Рис. 7. Диаграмма «Возраст (г) – пробег (тыс. км)»

Исключение составляет точка (8; 40). Эта выбивающаяся точка — *выброс*, то есть нетипичное и даже немного удивительное явление. Как мог получиться выброс? Возможны разные причины. Предложите школьникам подумать, как могло получиться, что у этой машины оказался такой небольшой пробег за восемь лет. Возможны разные варианты. Например:

– Хозяин заболел и не смог ездить. Машина стояла в гараже несколько лет, а теперь он или его родственники продают машину.

– Автомобиль ездил только летом, вывозил семью на дачу.

– Может быть, хозяин пытается «набить цену» и «скрутил пробег»? Такое тоже бывает…

– Может быть, человек, внося данные на сайт, просто ошибся, а 40 тыс. км на самом деле 140 тыс. км?

Определить по диаграмме точную причину невозможно, да и не нужно. Важно, что эта точка выбивается из общей картины, и мы ее не будем учитывать, а рассмотрим облако без нее. Выбросы нужно удалять; эта процедура чистки данных позволяет получить более точную картину явления.

Важно! Не существует единого правила, какие точки считать выбросами, а какие нет. Есть общепринятые методы, но в данном случае мы просто видим, что точка далеко от облака.

Как расположены точки в облаке? Они сгруппированы около некоторой прямой. Это подтверждает предположение о линейном характере связи. Но что это за прямая? Если мы ее проведем, то сможем написать ее уравнение. Тогда мы будем знать, как связь между пробегом и возрастом приближенно выражается формулой.

Провести прямую можно разными способами. Мы будем использовать самый простой — «глазомерный». Найдем прямую, которую можем назвать *осью диаграммы*[3]. Искомая прямая будет описывать линейную модель связи величин в нашей задаче.

Сначала найдем центральную точку. Например, можно найти средний возраст (6,1 года) и средний пробег (125 тыс. км). Напомните учащимся исключить выброс из расчетов. Поставим эту центральную точку (рис. 8).

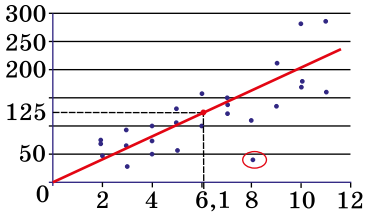


Рис. 8

Через эту точку на глазок (но по линейке) проводим прямую так, чтобы она прошла примерно посередине облака. У нас получилась прямая, которая проходит очень близко к началу координат. И это разумно: ведь новая машина имеет нулевой пробег[4].

Чем плотнее точки группируются вокруг оси диаграммы, тем сильнее связь и тем лучше построенная прямая описывает эту связь.

Уравнение прямой составить несложно. Напишем уравнение прямой

$$y = kx + b.$$

Теперь возьмем на построенной прямой две точки. Например, в нашем случае удобно взять начало координат и центр:

x	0	6,1
y	0	125

Подставим эти числа в уравнение. Получаем систему двух уравнений с неизвестными k и b:

$$\begin{cases} 0 = k \cdot 0 + b, \\ 125 = k \cdot 6{,}1 + b, \end{cases}$$

откуда $b = 0$, $k = \frac{125}{6{,}1} \approx 20{,}49$. Получаем уравнение оси: $y = 20{,}49x$. Это значит, что машины в рассмотренной совокупности пробегают в среднем по 20,49 тыс. км в год.

[3] Иногда используется термин «линия тренда». Этот термин оправдан тогда, когда речь идет о процессе во времени: изменение какой-либо величины во времени. В нашем случае ось диаграммы с натяжкой можно назвать трендом. А вот в следующем примере «Протяженность линии метро — количество станций на ней» зависимости от времени нет. Здесь упоминание тренда неуместно. Поэтому мы везде используем нейтральное название «ось диаграммы рассеивания».

[4] Машина, возраст которой 0 полных лет, не старше года. Будем считать, что ей полгода. Если за год машина проходит k км, то за полгода она проходит в два раза меньше: $\frac{k}{2}$ км. Это пробег машины возрастом 0 лет. Поэтому $b = 0{,}5k$. Получаем уравнение $y = kx + 0{,}5k$. Теперь достаточно одной центральной точки (6,1; 125), чтобы определить k: $125 = k(6{,}1 + 0{,}5)$, откуда $k \approx 18{,}94$. Получается уравнение $y = 18{,}94x + 9{,}47$. Разные предположения дают разные методы и приводят к разным, но близким моделям.

С помощью полученной формулы можно ответить на вопрос о приблизительном (типичном, ожидаемом) пробеге легкового автомобиля известного возраста. Например, машина, которой 5,5 лет, согласно нашей линейной модели имеет пробег $20{,}49 \cdot 5{,}5 \approx 112{,}7$ тыс. км.

Предложите учащимся самостоятельно провести ось диаграммы, построенной в задаче о Московском метрополитене (см. предыдущий урок) и написать ее уравнение.

Задание 5. Найдите координаты центра облака данных и проведите ось диаграммы. Найдите уравнение оси. При помощи полученной формулы связи между протяженностью линии метро и числом станций на ней в Московском метрополитене оцените ожидаемое количество станций на линии длиной 30 км.

Линия	Длина (км)	Количество станций
Сокольническая	32,5	22
Замоскворецкая	42,8	23
Арбатско-Покровская	45,1	22
Филевская	14,9	13
Кольцевая	19,3	12
Калужско-Рижская	37,6	24
Таганско-Краснопресненская	42,2	23
Калининская	16,5	8
Солнцевская	24,9	12
Серпуховско-Тимирязевская	41,2	25
Люблинско-Дмитровская	38,3	23
Большая кольцевая	10,5	5
Каховская	3,3	3
Бутовская	10	7
Монорельс	4,7	6
МЦК	54	31

Желательный результат обсуждения. Выбросов на первый взгляд нет (рис. 9). Координаты центра облака (27,4; 16,2). Через эту точку учащиеся на глаз строят прямую. В этом случае (из некоторых соображений) разумно взять точку (0; 1) (*почему*?) и провести прямую через центр облака и через центр. Не исключено, что эта прямая окажется подходящей осью.

Оси у каждого ученика получатся разные, поэтому и уравнения будут немного отличаться. Теперь предложите им найти, какое число станций должно быть на линии протяженностью 30 км (такой линии на самом деле нет).

В нашем примере получилось: $y = 0{,}53x + 1{,}7$ (см. рис. 9). Это значит, что (согласно нашей модели) в Московском метрополитене на каждый километр линий приходится примерно 0,53 станции или 53 станции на 100 км. На линии протяженностью 30 км окажется примерно 17,6 станции.

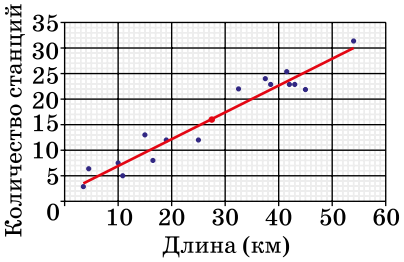


Рис. 9. Диаграмма «Протяженность линии метро (км) — число станций»

Важно! Об округлении: в жизни количество станций, безусловно, целое число. Многие школьники захотят округлить полученный результат до целого. Округление возможно, но необязательно: метод может дать дробное число.

На прошлом уроке учащиеся построили диаграмму рассеивания «Дальность прыжка — время забега». Можно ли считать, что если человек вдвое дальше прыгает, то он вдвое быстрее бегает? Или что есть вообще какая-то линейная связь между дальностью прыжка и временем забега на 60 м? Обсудите это. Если есть мнение, что диаграмма являет подобие линейной зависимости, предложите построить ось этой диаграммы (рис. 10).

Прыжок (см)	180	194	190	215	210	170	175
Бег (с)	10,8	10,2	10,6	9,6	10,2	11,0	11,6

Прыжок (см)	202	205	195	205	200	190	186
Бег (с)	10,4	10,0	11,0	9,8	10,6	10,8	10,7

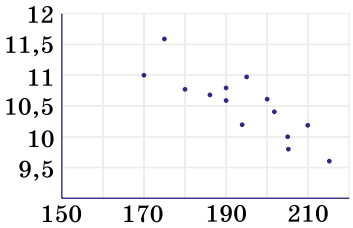


Рис. 10. Таблица и диаграмма «Дальность прыжка (см) — время забега на 60 м (с)»

Желательный результат обсуждения. Чем быстрее человек бегает, тем дальше он прыгает. Но многое еще зависит от силы и направления толчка при прыжке. Взаимосвязь между данными сложная. Однако форма диаграммы позволяет надеяться, что связь похожа на линейную. Можно провести ось и написать ее уравнение, но нужно помнить, что эта линейная формула опишет модель неточно и вряд ли позволит нам сделать хорошую оценку значений.

Обсудите еще одну задачу прошлого урока: про рост и вес. Можно ли считать, что вес и рост связаны линейно? Имеет ли смысл искать линейную зависимость в этом случае?

Рост (см)	167	169	179	178	177	175	171	181
Вес (кг)	62	67	70	72	70	69	63	80

Рост (см)	174	175	180	174	172	178	171	
Вес (кг)	73	66	75	70	67	74	66	

Желательный результат обсуждения. Вспомним формулу объема параллелепипеда: $V = abc$. Конечно, человек не параллелепипед, но между ростом и объемом тела (следовательно, весом) зависимость довольно сложная — третьей степени (не зря же объем измеряется в кубических метрах). Линейной связи нет. Построить ось и найти формулу можно, но для очень высоких и очень низких людей эта формула работать, скорее всего, не будет.

Выводы. Диаграммы рассеивания позволяют визуально выявить положительную или отрицательную связь. Но выявить характер связи более определенно — линейная она или более сложная — часто не удается. В двух первых примерах мы понимали, что линейная связь есть: во сколько раз больше возраст машины, во столько же раз больше пробег; во сколько раз длиннее линия метро, во столько же раз на ней больше станций. Конечно, все изменчиво, разброс данных велик, но облака точек в этих примерах вытянуты вдоль прямых, и мы понимаем, что так должно быть, исходя из здравого смысла. В таких случаях ось диаграммы позволяет лучше понять характер связи и даже количественно оценить ее параметры, написав уравнение.

Важно! Характер линейной связи — положительная она или отрицательная — определяет знак углового коэффициента k в уравнении оси диаграммы.

Часто связь двух величин не описывается линейной функцией, хотя может казаться, что большая часть облака (или даже все облако) вытянута вдоль прямой. В этом случае прямолинейная ось диаграммы описывает истинную связь приближенно только на том участке, где облако прямолинейно.

Рекомендуемое домашнее задание. Таблица содержит данные об атмосферном давлении в полдень 31 октября 2018 года в 34 городах России. В третьей колонке дана средняя высота центра города над уровнем моря.

1. Постройте диаграмму рассеивания по данным таблицы (внизу страницы) на миллиметровой бумаге.

2. Имеется ли связь между данными? Если да, охарактеризуйте ее (положительная она или отрицательная).

3. Что вызывает изменчивость атмосферного давления? Приведите один-два фактора, объясняющих, почему если связь между высотой и давлением существует, то она не очень сильная.

4. По вашему мнению, есть ли основания считать, что эта связь хотя бы приближенно описывается линейной функцией?

Город	Атмосферное давление (мм рт. ст.)	Высота города над уровнем моря (м)	Город	Атмосферное давление (мм рт. ст.)	Высота города над уровнем моря (м)
Абакан	738	245	Невинномысск	742	320
Барнаул	749	180	Новосибирск	752	150
Владикавказ	712	692	Омск	760	90
Грозный	756	130	Пенза	770	200
Екатеринбург	753	270	Пермь	764	149
Ессентуки	701	640	Санкт-Петербург	770	3
Иркутск	721	440	Ставрополь	728	620
Казань	774	60	Тверь	764	146
Калининград	762	5	Тюмень	767	102
Кемерово	749	140	Улан-Удэ	716	500
Кизляр	778	–5	Уфа	771	150
Кисловодск	692	825	Хабаровск	753	72
Красноярск	740	287	Ханты-Мансийск	766	50
Магас	720	600	Челябинск	756	225
Москва	765	187	Черкесск	723	530
Нальчик	733	512	Чита	708	650

5. Найдите координаты центра облака данных и проведите ось диаграммы.

6. Напишите уравнение оси диаграммы.

7. Согласно полученному уравнению, какое атмосферное давление можно ждать 31 октября в месте, которое находится на высоте 400 м над уровнем моря?

Урок-практикум: «Диаграммы рассеивания»

Урок и домашнее задание предполагает выполнение заданий на компьютере.

Оборудование: ПК с установленным процессором электронных таблиц, файл Les11.xlsx.

Цель урока: учащиеся должны научиться строить диаграмму рассеивания, ось диаграммы и выражать линейную связь между величинами средствами ПК.

На предыдущих двух уроках учащиеся работали с диаграммой рассеивания, построенной по данным о протяженности линий Московского метрополитена и количестве станций на них. Познакомьте учащихся с возможностями и функциями MS Excel, предназначенными для построения диаграмм рассеивания.

Задание 1. В таблицах файла Les11.xlsx представлены данные (по информации на осень 2018 года) о протяженности линий и количестве станций двух самых загруженных метрополитенов Европы — Московского и Парижского.

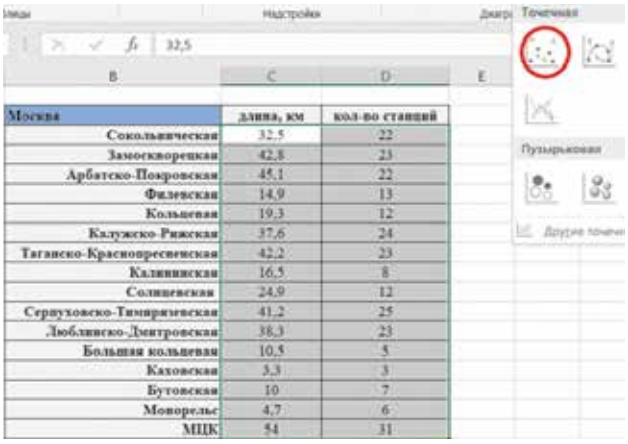
Предложите учащимся сравнить структуру каждого метрополитена. Станции в Париже расположены плотнее: общая длина линий меньше, чем в Москве, а количество станций — больше. Это заметно на схемах метро (рис. 1).

Рис. 1

Москва	Длина (км)	Количество станций
Сокольническая	32,5	22
Замоскворецкая	42,8	23
Арбатско-Покровская	45,1	22
Филевская	14,9	13
Кольцевая	19,3	12
Калужско-Рижская	37,6	24
Таганско-Краснопресненская	42,2	23
Калининская	16,5	8
Солнцевская	24,9	12
Серпуховско-Тимирязевская	41,2	25
Люблинско-Дмитровская	38,3	23
Большая кольцевая	10,5	5
Каховская	3,3	3
Бутовская	10	7
Монорельс	4,7	6
МЦК	54	31

Париж	Длина (км)	Количество станций
Линия 1	16,6	25
Линия 2	12,3	25
Линия 3	11,7	25
Линия 3bis	1,3	4
Линия 4	12,1	27
Линия 5	14,6	22
Линия 6	13,6	28
Линия 7	22,4	38
Линия 7bis	3,1	8
Линия 8	23,4	38
Линия 9	19,6	37
Линия 10	11,7	23
Линия 11	6,3	13
Линия 12	15,3	29
Линия 13	24,3	32
Линия14	9	9

Для построения диаграммы рассеивания нужно выделить данные (ячейки C3 : D18) и в меню «Вставка» выбрать «Построение точечной диаграммы» (рис. 2).



Москва	длина, км	кол-во станций
Сокольническая	32,5	22
Замоскворецкая	42,8	23
Арбатско-Покровская	45,1	22
Филевская	14,9	13
Кольцевая	19,3	12
Калужско-Рижская	37,6	24
Таганско-Краснопресненская	42,2	23
Калининская	16,5	8
Солнцевская	24,9	12
Серпуховско-Тимирязевская	41,2	25
Люблинско-Дмитровская	38,3	23
Большая кольцевая	10,5	5
Каховская	3,3	3
Бутовская	10	7
Монорельс	4,7	6
МЦК	54	31

Рис. 2

В результате на листе появится диаграмма рассеивания (рис. 3).

Рис. 3

Оформите диаграмму, переместите ее на удобное место на листе, впишите название диаграммы, наименования осей (рис. 4), укажите форму и размер точек, размер шрифта и т.д.

На предыдущем уроке учащиеся строили диаграмму по информации о линиях Московского метрополитена вручную. Предложите сравнить ее с диаграммой, построенной в MS Excel. Предложите школьникам самостоятельно построить и оформить такую же диаграмму рассеивания для Парижского метрополитена. В результате у учащихся должны получиться две диаграммы. Для удобства сравнения поместите их рядом, например, как на рисунке 5.

Когда на предыдущем уроке школьники проводили ось диаграммы, они делали это на глаз, стараясь провести ее через центр облака. Для этой цели мы советовали искать центр облака отдельно. Табличный процессор MS Excel имеет встроенную функцию для построения оси диаграммы. Для этого он использует специальный метод — подбирает прямую таким образом, чтобы сумма квадратов ошибок (отклонений точек от прямой) была наименьшей. При этом прямая обязательно пройдет через центр диаграммы. Доказательство этого факта можно будет обсудить со школьниками позже, изучая тему «Свойства дисперсии».

Рис. 5

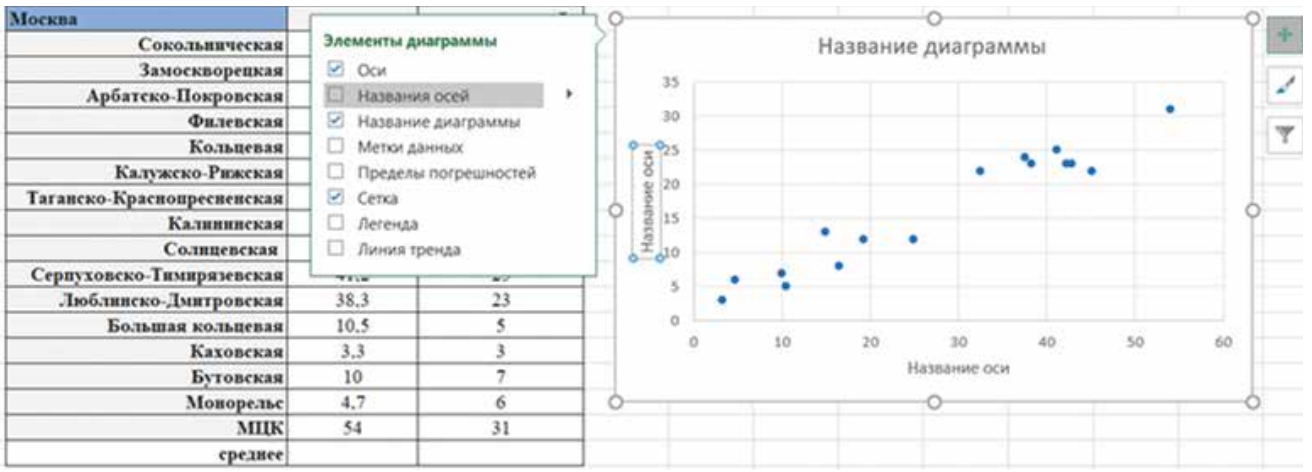


Рис. 4

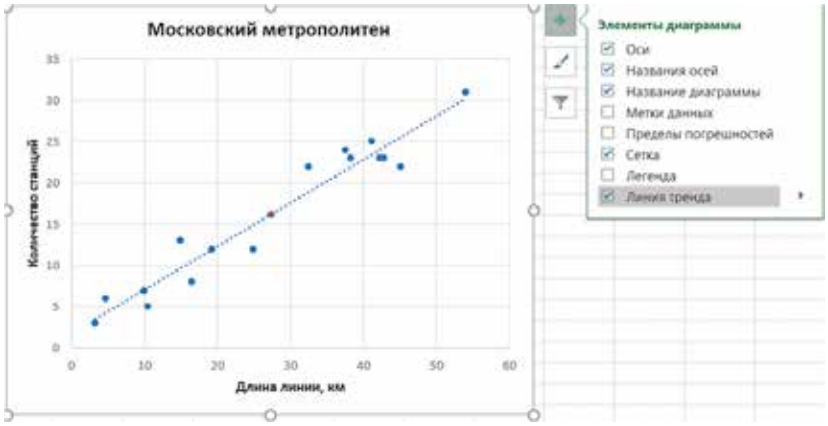


Рис. 6

Воспользуемся этой встроенной функцией. Для этого поставим галочку слева от элемента «Линия тренда» в меню «Элементы диаграммы» (рис. 6).

Если у вас более старая версия MS Excel, то эту функцию вы найдете в диалоговом окне, которое появится, если навести указатель на какую-нибудь точку данных и щелкнуть правой кнопкой мыши (рис. 7).

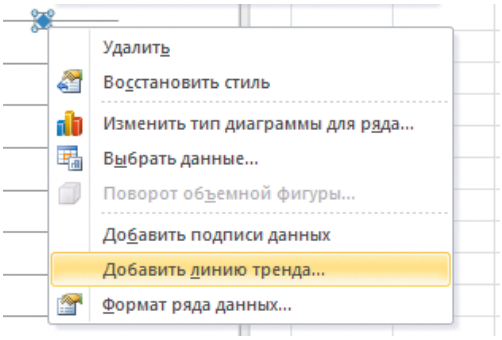


Рис. 7

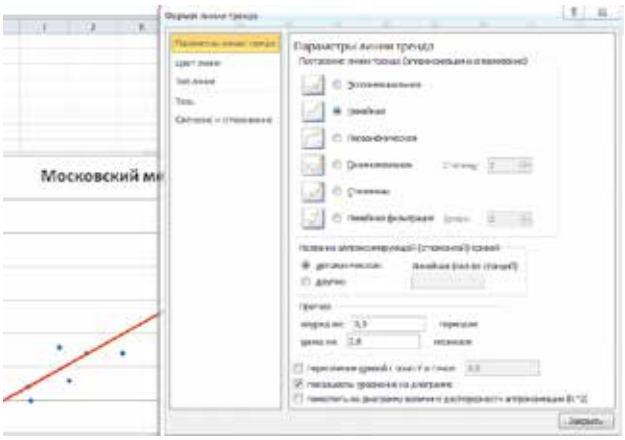


Рис. 8

Дважды щелкнув мышью, наведенной на получившуюся ось, учащиеся могут открыть диалоговое окно, где можно указать свойства оси. Посоветуйте сделать ось толстой красной сплошной линией (диалоги «Цвет линии» и «Тип линии»).

В диалоге «Параметры линии тренда» укажите «Показывать уравнение на диаграмме» и введите подходящее значение в поле «назад на» так, чтобы дотянуть линию влево до оси ординат (рис. 8). Это позволит продемонстрировать школьникам смысл свободного коэффициента в уравнении прямой: ордината точки пересечения прямой с вертикальной координатной осью.

Если все сделано правильно, на диаграмме появится уравнение, оно подвижно. Его можно переместить так, чтобы оно не закрывало данные, указать подходящий размер шрифта (правая кнопка мыши на уравнении вызывает контекстное меню). Предложите школьникам заняться итоговым оформлением. Можно поменять цвет или способ заливки диаграмм, добавить эмблемы метрополитенов в Москве и Париже, рисунки и т.п. Результат может выглядеть так, как на рисунке 9, или иначе — в зависимости от вкусов и дизайнерских талантов школьников.

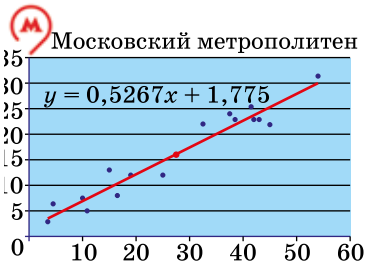


Рис. 9

Важный элемент уравнения прямой — угловой коэффициент. Обсудите со школьниками его смысл в данной задаче. Угловой коэффициент виден на диаграмме, но его нельзя скопировать — это картинка. Чтобы получить угловой коэффициент оси диаграммы рассеивания отдельно, в MS Excel предусмотрена функция

Москва	Длина (км)	Количество станций
Сокольническая	32,5	22
Замоскворецкая	42,8	23
Арбатско-Покровская	45,1	22
Филевская	14,9	13
Кольцевая	19,3	12
Калужско-Рижская	37,6	24
Таганско-Краснопресненская	42,2	23
Калининская	16,5	8
Солнцевская	24,9	12
Серпуховско-Тимирязевская	41,2	25
Люблинско-Дмитровская	38,3	23
Большая кольцевая	10,5	5
Каховская	3,3	3
Бутовская	10	7
Монорельс	4,7	6
МЦК	54	31
Коэффициент наклона		

«НАКЛОН». Синтаксис функции простой:

=НАКЛОН(значения y; значения x).

Впишите эту функцию в поле D20 (рис. 10) и в качестве аргументов укажите соответствующие массивы (сначала массив значений, то есть число станций, а потом массив протяженности).

В поле D20 появится значение углового коэффициента (рис. 11).

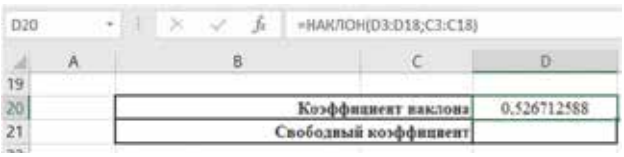


Рис. 11

Чтобы получить свободный член оси диаграммы рассеивания, в MS Excel предусмотрена функция «ОТРЕЗОК»:

=ОТРЕЗОК (значения y; значения x).

Впишите эту функцию в поле D21. Аргументы этой функции будут совпадать с ранее введенной функцией «НАКЛОН».

Обсудите с учащимися некоторые вопросы.

– В чем смысл свободного члена в уравнении? [Для московского метро он приближенно равен 1,78, а для парижского — 4,26].

– Подумайте вместе, может ли свободный член быть очень большим?

– Чему он должен равняться в «идеальном метро», где все станции на всех линиях расположены строго на одинаковом расстоянии друг от друга?

Желательный результат обсуждения. Свободный член в уравнении показывает количество станций на линии «нулевой длины». Воображаемая линия метро нулевой длины все равно должна иметь одну станцию — начало. Кто-то из школьников скажет, что даже очень короткая линия должна иметь две конечные станции. Поэтому разумно считать, что свободный член должен быть где-то между 1 и 2. В парижском метро свободный коэффициент не удовлетворяет этому предположению. Это, вероятно, говорит о неравномерности распределения станций по линиям.

Если останется время, предложите учащимся построить диаграмму для Пекинского метрополитена (вверху страницы).

Выводы. Диаграммы рассеивания — весьма распространенный способ представления и изучения данных. Настолько распространенный, что разработчики электронных таблиц потрудились сделать специальные готовые функции. Построить диаграмму, ее ось и вывести уравне-

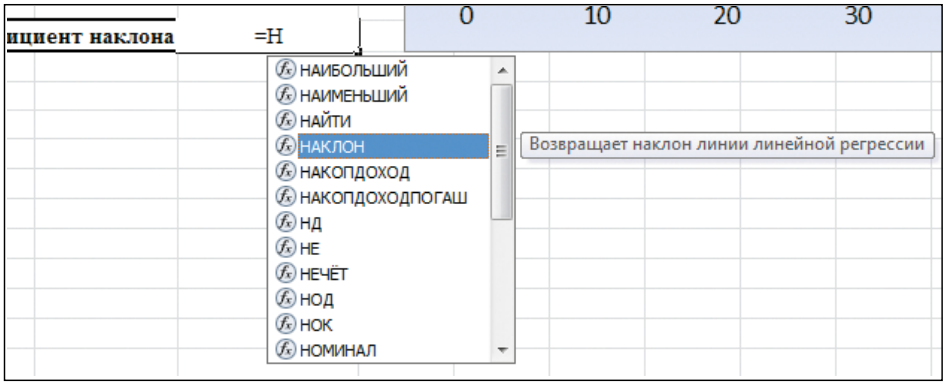


Рис. 10

Линия метро	Протяженность (км)	Количество станций	Линия метро	Протяженность (км)	Количество станций
1	30,4	23	14 запад	12,7	7
2	23,1	18	14 восток	28,6	19
4	28,2	24	15	45,7	18
5	27,6	23	Батун	18,9	13
6	42,8	26	Чанпин	32	12
7	23,7	19	Дасин	21,7	12
8	26,6	17	Фаншань	24,8	11
9	16,5	12	Ичжуан	23,3	13
10	57,1	45	Аэропорт-экспресс	28,5	4
13	40,9	16			

ние — очень легко. Но по-прежнему нужно помнить, что все определяет характер данных. Если зависимость линейная и мы это понимаем, то прямая ось и ее уравнение имеют смысл. А если данные носят нелинейный характер, то прямолинейная ось и линейное уравнение превращаются в «арифметическую забаву», как говорил Борис Гнеденко о подобного рода упражнениях.

Рекомендуемое домашнее задание. Таблица содержит данные (по информации на осень 2018 года) о количестве линий и общем количестве станций метрополитенов 16 крупных городов мира (таблица внизу страницы).

Выполните задание средствами MS Excel.

1. Постройте диаграмму рассеивания по данным таблицы.

2. Имеется ли связь между данными? Если да, охарактеризуйте ее (положительная она или отрицательная).

3. По вашему мнению, есть ли основания считать, что эта связь хотя бы приближенно описывается линейной функцией?

4. Проведите ось диаграммы и найдите ее формулу.

5. На основе данных таблицы определите, сколько в среднем приходится станций метро на одну линию.

Город	Число линий	Число станций	Город	Число линий	Число станций
Пекин	19	332	Шэньчжэнь	8	199
Лондон	11	270	Марсель	2	24
Токио	13	285	Медельин	2	28
Киото	1	32	Ренн	1	15
Гуанчжоу	13	235	Брюссель	4	59
Мехико	12	195	Лилль	2	62
Нанкин	9	164	Милан	4	113
Гонконг	10	104	Стокгольм	3	100
Чунцин	6	155	Шэньчжэнь	8	199
Сингапур	5	69	Валенсия	9	138
Дели	6	146	Барселона	12	187
Вашингтон	6	91			