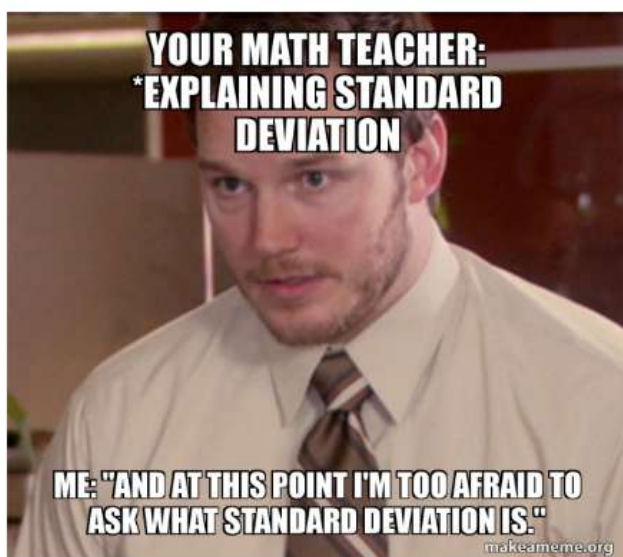


Биномиальное распределение и стандартное отклонение

На этом небольшом уроке мы познакомимся с довольно естественным распределением — биномиальным — и порешаем задачки, в которых это распределение возникает. Кроме того мы узнаем, что такое *стандартное отклонение*.



Биномиальное распределение

Напомним, по определению случайная величина X имеет распределение Бернулли, если $X = 1$ с вероятностью p и $X = 0$ с вероятностью $1 - p$.

Определение. Пусть X_1, \dots, X_n независимы и имеют распределение Бернулли с вероятностью успеха p . Тогда распределение случайной величины $S := X_1 + \dots + X_n$ называется *биномиальным распределением с n степенями свободы*.

Обозначение. Биномиальное распределение обозначается $Bin(n, p)$. Фразу " S имеет биномиальное распределение с n степенями свободы" записывают так: $S \sim Bin(n, p)$.

В DS биномиальное распределение встречается часто. Например, когда

- n пользователей независимо друг от друга могут совершать или не совершать какое-то действие (зайти на сайт, совершить покупку, кликнуть на рекламу).
- n одинаковых механизмов могут выходить или не выходить из строя.
- Вообще, каждый раз когда есть n похожих и независимых ситуаций с двумя возможными исходами, мы приходим к биномиальному распределению.

Комментарий. Биномиальное распределение хорошо приближается *нормальным распределением*, которое мы узнаем на следующих неделях, когда будем проходить непрерывные распределения. Поэтому на практике часто работают не с биномиальным, а с нормальным распределением. Тем не менее, сначала лучше понять биномиальное распределение.

Задача с проверкой. Биномиальное распределение и стандартное отклонение 1

Задача. В вашей фирме работают n разработчиков. Каждый рабочий день каждый разработчик приходит в офис с вероятностью p , и с вероятностью $(1 - p)$ остаётся работать из дома. Разработчики приходят работать независимо друг от друга.

1. Какова вероятность, что сегодня конкретные k разработчиков будут работать из офиса, а все остальные – из дома?
2. Какова вероятность, что сегодня в офис придут ровно k разработчиков? Имена пришедших не уточняются – важно только чтобы пришло ровно k .
3. Найдите распределение случайной величины S , где S – число разработчиков, которые пришли сегодня в офис.

Распределение из Пункта 3 и будет биномиальным распределением.

Задача для проверки. Пусть разработчиков зовут Аня, Гриша, Вика, Петя и Илья, и каждый из них приходит в офис с вероятностью 0.75.

Заполните пропуски

1. Какова вероятность того, что сегодня в офис придут только Аня и Петя? Ответ округлите до четвёртого знака после запятой

2. Какова вероятность того, что в офис придут ровно три разработчика? Ответ округлите до четвёртого знака после запятой

Задача с проверкой. Биномиальное распределение и стандартное отклонение 2

Задача. У вас та же фирма, что и в задаче с предыдущего шага: есть n разработчиков, каждый приходит в офис с вероятностью p (независимо от остальных разработчиков).

Пусть $X_i = 1$ если i -ый разработчик пришёл, и $X_i = 0$ в противном случае. Тогда $S := X_1 + \dots + X_n$ это количество пришедших разработчиков.

1. Найдите $E[S]$.
2. Найдите $Var(S)$.

Задача для проверки. Пусть в фирме 7 разработчиков, и каждый приходит в офис с вероятностью 0.8.

Заполните пропуски

1. Чему равно матожидание количества пришедших работников?

2. Чему равна дисперсия количества пришедших разработчиков?

Задача с проверкой. Биномиальное распределение и стандартное отклонение 3

Задача. Теперь пусть разработчики либо все одновременно приходят (с вероятностью p), либо все одновременно не приходят (с вероятностью $1 - p$). Обозначим число пришедших разработчиков за T .

1. Найдите $E[T]$.
2. Найдите $Var(T)$.

Важно. Сравните с ответами в задаче с предыдущего шага. Обратите внимание, что $E[S] = E[T]$, а $Var(S)$ меньше, чем $Var(T)$. Почему так получилось?

Задача для проверки. Пусть разработчиков 7, и все либо одновременно приходят в офис с вероятностью 0.8, либо все не приходят.

Заполните пропуски

1. Чему равно матожидание числа пришедших разработчиков?

2. Чему равна дисперсия числа пришедших разработчиков?

Стандартное отклонение

Представьте, что вы наблюдаете за стаей пингвинов. Пусть случайная величина X – возраст пингвина. Обычно на сколько лет возраст пингвина отличается от $E[X]$? Примерно на год, на два, на три?

Ясно, что число $Var(X)$ в качестве ответа не подойдёт, потому что $Var(X)$ это средний квадрат отклонения. Про то, почему $Var(X)$ не подойдёт, можно думать ещё и так. Если X измеряется в годах, то $E[X]$ тоже измеряется в годах. И отклонение от $E[X]$ должно измеряться в годах. Но $Var(X) = E[(X - E[X])^2]$ измеряется в годах в квадрате. В квадратных годах ;)

Напоминание. В [этой](#) задаче урока про дисперсию ответ был в квадратных рублях.

В качестве ответа на наш вопрос разумно взять среднее значение модуля разности между X и $E[X]$, то есть число $E[|X - E[X]|]$. К сожалению, такое определение не удобно для вычислений, потому что вычисления с участием модуля обычно требуют разбора случаев.

В результате для оценки отклонения X от $E[X]$ используют *стандартное отклонение*.

Определение. *Стандартным отклонением* случайной величины X называется $\sqrt{Var(X)}$.

В вычислениях обычно используют дисперсию. А когда нужно отвечать на какие-то реальные вопросы, используют стандартное отклонение.

Пример. Вы узнали, что средний рост взрослого пингвина равен 90 см. Дисперсия равна 100, то есть стандартное отклонение равно $\sqrt{100} = 10$ см. Пусть вы встретили пингвина ростом 130 см – то есть отклоняющегося на 4 стандартных отклонения от среднего роста. Ясно, что это необычный, аномальный пингвин. А может, и не пингвин вовсе, а кто-то другой.



Преподаватель составил историю оценок своих студентов и на их основе вывел, что оценка его студентов представляет собой следующую случайную величину:

Значение	3	4	5	6	7	8	9	10
Вероятность	0.06	0.08	0.09	0.3	0.2	0.12	0.1	0.05

(Преподаватель никому не ставит меньше 3, потому что это незачёт, а преподаёт он историю на математическом факультете.)

Комментарий. Да, в этой задаче придётся посчитать. Но зато она похожа на задачи, которые могут встретиться в реальной жизни.

Заполните пропуски

1. Найдите дисперсию этой случайной величины. Ответ округлите до 3 знаков после запятой:

2. Найдите стандартное отклонение этой случайной величины. Ответ округлите до 3 знаков после запятой:

Честный кубик бросают 10 раз, найдите значение стандартного отклонения для суммы выпавших значений.

Ответ округлите до 3 знаков после запятой.



Введите численный ответ

Дополнительная задача. Биномиальное распределение и стандартное отклонение

Давайте ещё раз убедимся, что для измерения отклонений сложно придумать что-то лучше, чем наше определение дисперсии и стандартного отклонения.

1. Зачем вообще искать средний квадрат отклонения? Давайте лучше искать само среднее отклонение, а не его квадрат. Найдите $E[X - E[X]]$ для любой случайной величины X .
2. Хорошо, со средним отклонением не получилось. Тогда давайте искать средний модуль отклонения: $H[X] := E[|X - E[X]|]$. Пусть теперь мы хотим использовать H в наших вычислениях, а не дисперсию. Докажите, что даже для независимых X и Y не всегда выполнено $H[X + Y] = H[X] + H[Y]$.

Как мы помним, для независимых X и Y выполнено $Var(X + Y) = Var(X) + Var(Y)$. Так что дисперсия ещё и этим удобнее для вычислений, чем H (а не только тем, что H использует модуль, а Var нет).

Проверка. $E[X - E[X]] =$

Введите численный ответ

Введите число

Дополнительная задача. Биномиальное распределение и стандартное отклонение

Пусть случайная величина принимает значение $a \in [0, 10]$ с вероятностью p и значение $b \in [0, 10]$ с вероятностью $(1 - p)$.

Мы рассматриваем всевозможные числа a, b, p . Каким может быть максимальное значение стандартного отклонения такой случайной величины?

Введите численный ответ

Что мы прошли на этом уроке

На этом уроке мы

- поработали с биномиальным распределением $Bin(n, p)$, нашли его дисперсию и математическое ожидание
- узнали, что стандартное отклонение это корень из дисперсии, то есть $\sqrt{Var[X]}$

Что нас ждёт на следующем уроке

На следующем уроке мы

- начнём обсуждать вероятностные пространства со счётным пространством исходов
- пройдем понятие из матана: *ряды*
- изучим, как ряды можно исследовать на *сходимость*