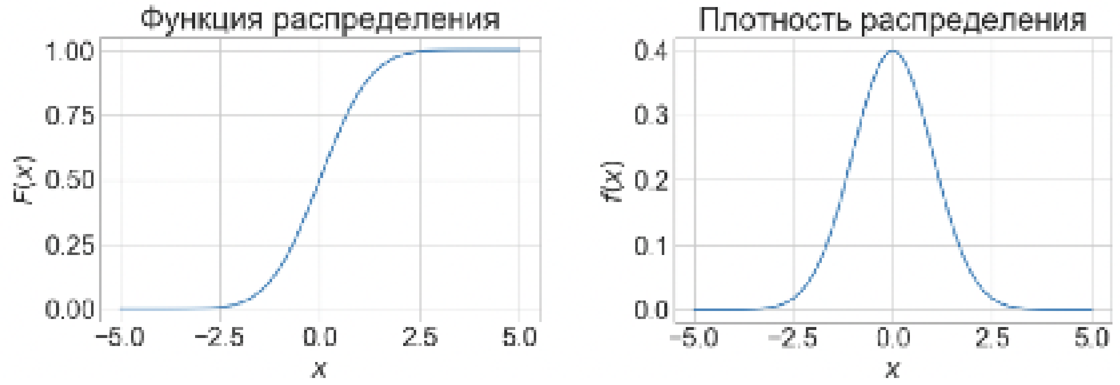
Квиз-1 (DA6)

1

Выберите в списке ниже все верные утверждения: 1 балл

- [ ] Если случайная величина Z распределена на отрезке [0;3], функция f_Z(z) = 3 z^2 может быть её плотностью распределения
- [x] Если случайная величина Z распределена на отрезке [0;1], функция F_Z(z) = z может быть её функцией распределения
- [ ] Распределение дискретной случайной величины можно описать с помощью плотности распределения
- [x] Любую случайную величину можно однозначно задать с помощью функции распределения (не бывает двух разных случайных величин с одинаковой функцией распределения)
- [ ] Любую случайную величину можно однозначно задать с помощью математического ожидания и дисперсии (не бывает двух разных случайных величин с одинаковыми математическими ожиданиями и дисперсиями)
- [x] Медиана и квантиль уровня 50% это одно и то же
- [x] Площадь под плотностью распределения должна быть равна единице
- [ ] Мода и квантиль уровня 50% это одно и то же

Устройство сундука



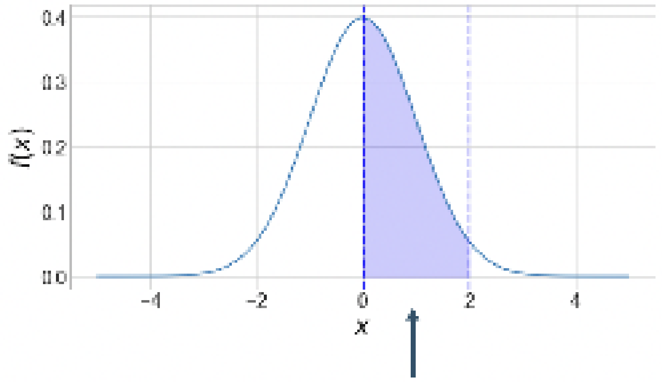
Медиана - срединоное значение, которое делит данные пополам

Квантиль 50% - значение ниже которого лежит 50%

Мода - самое часто встречающееся значение

Непрерывные случайные величины

Распределение непрерывной случайной величины описывается **плотностью распределения вероятностей.**



Пример: нормальное распределение

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Площадь равна вероятности попасть на отрезок от нуля до двух $= \int_0^2 f(x)\ dx$

Площадь под всей плотностью должна быть равна 1

2

Случайная величина X имеет распределение как на картинке. Найдите дисперсию случайной величины X. 1 балл

X	-1	0	1	2
$P(X=k)$	0.1	0.2	0.3	0.4

Var(X) = E(X^2)-E^2(X)

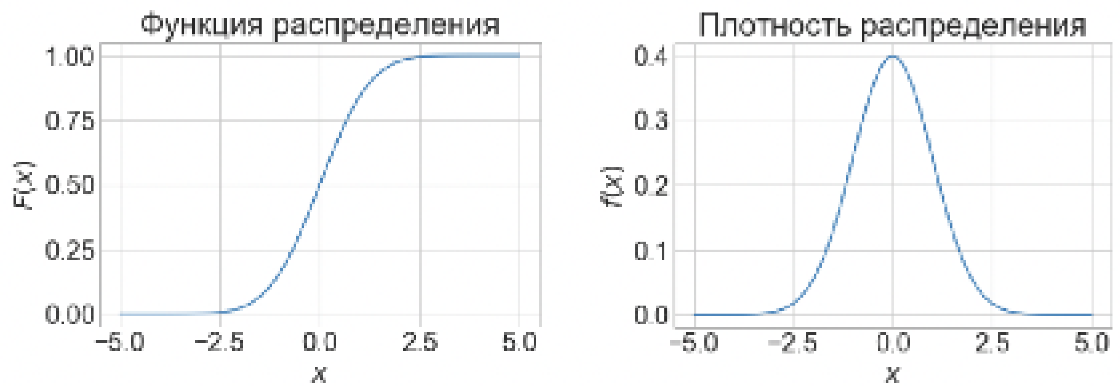
1. E(X) = -1*0.1+0*0.2+1*0.3+2*0.4=1
2. E^2(X) = 1^2 = 1
3. E(X^2) = ((-1)^2)*0.1+(0^2)*0.2+(1^2)*0.3+(2^2)*0.4=2
4. Var(X) = 2-1=1

3

Выберите все утверждения, которые верны для любой функции распределения F(x) 1 балл

- [x] Она не убывает
- [ ] Она возрастает
- [ ] Она не возрастает
- [ ] F(x) принимает любые значения на [0;+∞)
- [ ] F(x)<0
- [x] F(x)≤1
- [ ] Если взять производную от плотности распределения, можно получить функцию распределения

Устройство сундука



Важные свойства

1. Плотность определена только для непрерывных случайных величин
2. $f(x) = F'(x)$
3. $\int_{-\infty}^{+\infty} f(t)\ dt = 1, \quad f(t) \geq 0 \ \ \forall\, t$
4. $F(x)$ не убывает, лежит между 0 и 1
5. $\mathbb{P}(a \leq X \leq b) = \int_a^b f(t)\ dt = F(b) - F(a)$
6. Вероятность того, что непрерывная случайная величина попадёт в точку, равна нулю

4

Случайные величины X и Y независимы. Найдите мат. ожидание случайной величины Z = 2X - Y + 1, если известно следующее:

1 балл

$$\mathbb{E}(X) = -1, \mathbb{E}(X^2) = 5, \mathbb{E}(Y) = 2, \mathbb{E}(Y^2) = 13.$$

Мой ответ

Z = 2X-Y+1

E(Z) = E(2X-Y+1) = E(2X)-E(Y)+E(1)=

= 2E(X)-E(Y)+1 = 2*(-1) - 2 + 1 =

=-2 - 2 + 1 = -3

Свойства математического ожидания

X, Y — случайные величины $\quad a$ — константа

1. $\mathbb{E}(a) = a$
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
3. $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$
4. $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, если независимы
5. Математическое ожидание случайной величины – не случайно
6. $\mathbb{E}\big(X - \mathbb{E}(X)\big) = \mathbb{E}(X) - \mathbb{E}\big(\mathbb{E}(X)\big) = \mathbb{E}(X) - \mathbb{E}(X) = 0$

5

Случайные величины X и Y независимы. Найдите дисперсию случайной величины Z = 2X - Y + 1 если известно следующее: 1 балл

$$\mathbb{E}(X) = -1, \mathbb{E}(X^2) = 5, \mathbb{E}(Y) = 2, \mathbb{E}(Y^2) = 13.$$

Мой ответ

Z = 2X-Y+1

Var(Z) = Var(2X-Y+1) =

= Var(2X)+Var(Y)+Var(1)=

= 4*Var(X)+Var(Y) =

= 4*(E(X^2)-E^2(X))+E(Y^2)-E^2(Y) =

=4*(5-(-1)^2)+13-2^2=25

Свойства дисперсии

X, Y – случайные величины a – константа

1. $Var(a) = 0$
2. $Var(X + Y) = Var(X) + Var(Y)$, если независимы
3. $Var(X - Y) = Var(X) + Var(Y)$, если независимы
4. $Var(a \cdot X) = a^2 \cdot Var(X)$
5. Дисперсия случайной величины – не случайна

6

Ваня любит пить чай. Иногда он пьёт его с сахаром, иногда без. На этой неделе он помечал цифрой 1 дни, когда пил чай с сахаром. Получилось 1,1,0,0,1,0. Найдите *среднее значение* сахарных дней в жизни Вани 1 балл

$$\bar{x} = \frac{X_1 + X_2 + \cdots X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

X_ср = (1+1+0+0+1+0)/6 = 0.5

7

Ваня любит пить чай. Иногда он пьёт его с сахаром, иногда без. На этой неделе он помечал цифрой 1 дни, когда пил чай с сахаром. Получилось 1,1,0,0,1,0. Найдите *дисперсию* сахарных дней в жизни Вани

$$\hat{\sigma}^2 = \frac{(X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + \cdots + (X_n - \bar{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})$$

$$\text{Var(X)} = \frac{(1-0.5)^2+(1-0.5)^2+(0-0.5)^2+(0-0.5)^2+(1-0.5)^2+(0-0.5)^2}{6} = 0.25$$

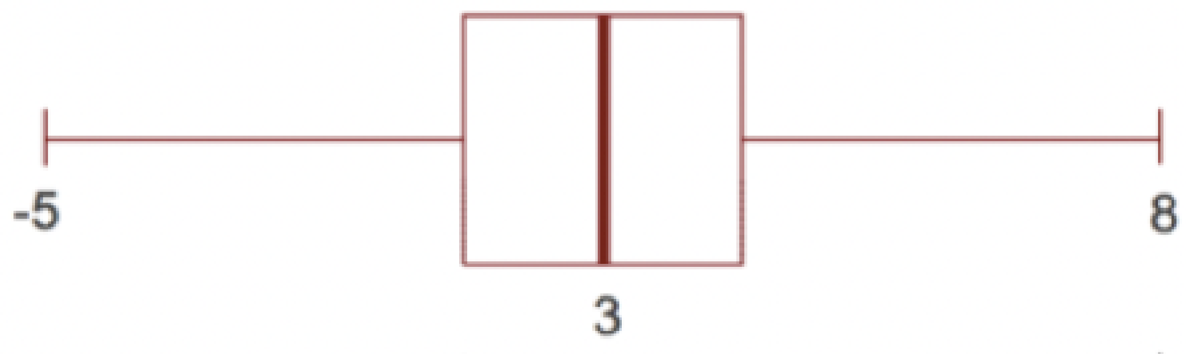
8

Измерен рост 25 человек. Средний рост оказался равным 160 см. Медиана оказалась равной 155 см. Машин рост в 163 см был ошибочно внесен как 173 см. Выберите все верные утверждения из списка ниже. 1 балл

- [x] Когда ошибку исправят, медиана никак не изменится
- [x] Когда ошибку исправят, среднее уменьшится
- [ ] Когда ошибку исправят, медиана увеличится
- [ ] Когда ошибку исправят, среднее увеличится
- [ ] Когда ошибку исправят, среднее никак не изменится
- [ ] Медиана и среднее нечувствительны к выбросам в отличие от моды

9

Вася по данным построил следующий ящик с усами. Найдите значение интерквантильного размаха. Ответ округлите до двух знаков после точки. 1 балл



Дано:
$q1-1.5IQR = -5$
$q2 = 3$
$q3+1.5IQR = 8$

Найти:
$IQR = q3-q1$

$$\begin{cases} q1-1.5(q3-q1) = -5 \\ q3+1.5(q3-q1) = 8 \end{cases} \quad \begin{cases} q1 - 1.5q3 + 1.5q1 = -5 \\ q3 + 1.5q3 - 1.5q1 = 8 \end{cases} \quad \begin{cases} 2.5q1 - 1.5q3 = -5 \\ 2.5q3 - 1.5q1 = 8 \end{cases} \quad \begin{matrix} q1 - 0.6q3 = -2 \\ 2.5q3 - 1.5q1 = 8 \end{matrix}$$

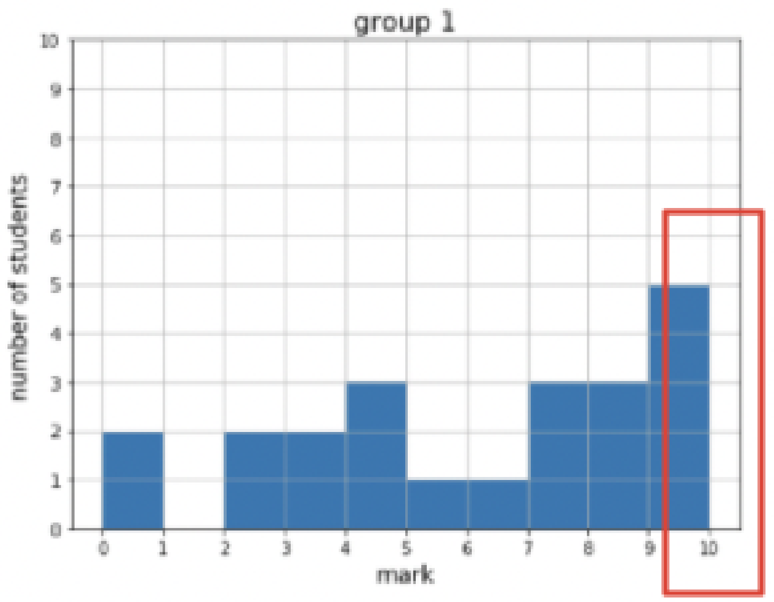
$q1 = -0.125$

$q3 = 3.125$

$q3 - q1 = 3.125-(-0.125) = 3.25$

10

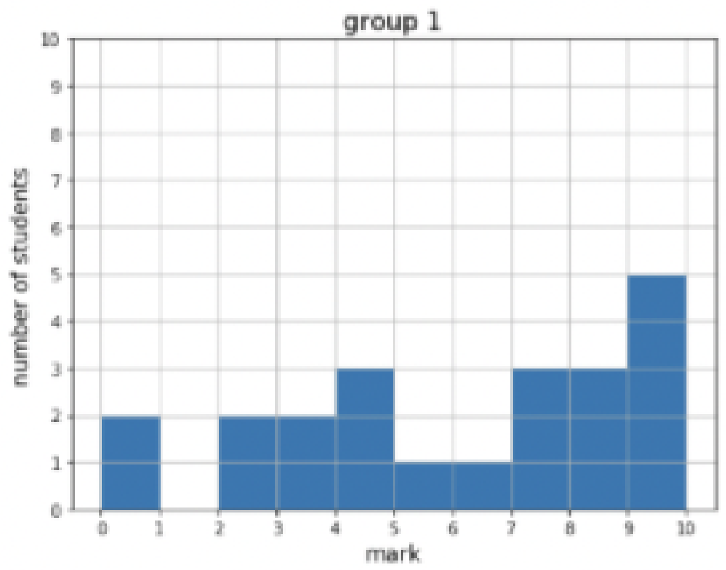
Группа на факультете "Э" написала контрольную работу по статистике. Ниже приведена гистограмма оценок за эту работу с шагом один (в столбец включается значение правой границы). Найдите значение *моды*.



Мода - самое часто встречающееся значение

11

Группа на факультете "Э" написала контрольную работу по статистике. Ниже приведена гистограмма оценок за эту работу с шагом один (в столбец включается значение правой границы). Найдите значение *медианы*.



Оценка	1	3	4	5	6	7	8	9	10
Частот	2	2	2	3	1	1	3	3	5

1 1 3 3 4 4 5 5 5 6 $\boxed{7\ 8}$ 8 8 9 9 9 10 10 10 10 10

(7+8)/2=7.5

12

Какие из нижеперечисленных статистик нечувствительны к выбросам? 1 балл

- [ ] Выборочное среднее
- [ ] Выборочная дисперсия
- [ ] Среднеквадратичное отклонение
- [x] Среднее абсолютное отклонение от медианы

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - med(x_1,\dots,x_n)|$$

- [x] Медиана
- [x] Интерквантильный размах: IQR=Q3-Q1
- [x] Квантиль уровня 35%
- [ ] Размах выборки max(X1,...Xn) min(X1....Xn)

13

У нас есть данные с результатами того, как студенты написали контрольную работу. В оценках есть пропуски. Мы не хотели бы выбрасывать эти наблюдения. Как лучше всего заполнить их для дальнейшей работы?

- (•) Средними
- ( ) Везде поставить цифру 42
- ( ) Нулями
- ( ) Дисперсиями

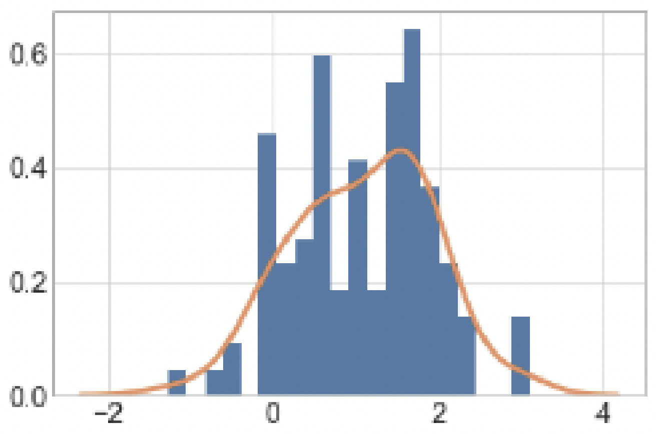
14

Выберите все верные утверждения 1

- [x] Ядерная оценка плотности позволяет получить график плотности в виде непрерывной линии
- [ ] Наивный прогноз на основе среднего устойчив к выбросам
- [x] Когда данные измерены в разном масштабе, это искажает подсчёт расстояний между объектами
- [x] Стандартизация по правилу z=(x-x.mean())/x.std() позволяет перейти к безразмерной величине и уравнять в правах разные переменные

Гистограмма

- Ядерные оценки плотности (kernel density estimation, KDE) позволяют получить график плотности в виде непрерывной кривой



Похожесть

Добрыня:	Ярополк:
90 кг	60 кг
1.9 м	1.7 м

- В Анализе данных часто ищут похожие объекты на основе расстояния между ними
- Какое расстояние между Добрыней и Ярополком?

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$\sqrt{(1.9 - 1.7)^2 + (90 - 60)^2} = \sqrt{0.04 + 900}$$

Вес вносит в расстояние более весомый вклад из-за того, что он измерен в кг

Способы масштабирования

i - номер наблюдения

Нормализация (Standard Scaler):

$$x_i^* = \frac{x_i - \bar{x}}{\hat{\sigma}}$$

Масштабирование на отрезок [0; 1] (Minmax Scaler):

$$x_i^{\wedge} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Устойчивая к выбросам нормализация (Robust Scaler):

$$x_i^{**} = \frac{x_i - med(X)}{Q_3 - Q_1}$$

Квиз-2 (DA6)

1

Выберите все верные утверждения

1 балл

- [ ] Если X — это длина питона в метрах, а Y — его вес в килограммах, то корреляция Пирсона между этими величинами измеряется в метрах, умноженных на килограммы
- [ ] Если корреляция между двумя случайными величинами равна нулю, это означает, что они независимы между собой
- [ ] Случайные величины X и Y зависимы. Случайные величины Y и Z зависимы. Тогда случайные величины X и Z зависимы
- [ ] Если коэффициент корреляции равен 0, значит между величинами нет никакой взаимосвязи
- [ ] Корреляция между количеством сна в часах и избыточным весом равная -0.7 означает, что избыточный вес — причина недосыпа
- [x] Умножение случайной величины на положительную константу не изменяет корреляции
- [ ] Умножение случайной величины на положительную константу не изменяет ковариации
- [x] Умножение случайной величины на отрицательную константу меняет знак корреляции
- [x] Корреляция величины с самой собой равна 1

2

Известно, что E(X)=−1, E(Y)=−1, Var(X)=9, Var(Y)=4, Corr(X, Y) = 1. 1 балл
Найдите: E(Y−2X−3).

E(Y-2X-3) = E(Y)-2E(X)-3 = -1 - 2 *(-1) -3 = -2

Свойства математического ожидания

X, Y — случайные величины a — константа

1. $\mathbb{E}(a) = a$
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
3. $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$
4. $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, если независимы
5. Математическое ожидание случайной величины – не случайно
6. $\mathbb{E}\big(X - \mathbb{E}(X)\big) = \mathbb{E}(X) - \mathbb{E}\big(\mathbb{E}(X)\big) = \mathbb{E}(X) - \mathbb{E}(X) = 0$

3

Известно, что E(X)=−1, E(Y)=−1, Var(X)=9, Var(Y)=4, Corr(X, Y) = 1.
Найдите: Var(Y−2X−3).

Corr(X, Y) = 1 -> зависимые

Var (Y-2X-3) = 1*Var(Y) + 4*Var(X) - 2*2*Cov(Y,X) =

= Var(Y) + 4*Var(X) - 4*Corr(X,Y)*std(X)*std(Y) =

= 4 + 4*9 -4*1*3*2 = 16

9. Если X и Y зависимы, тогда

$$\mathbb{E}\,(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) + Cov(X,Y)$$

$$Var\,(aX + bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var(Y)} + 2 \cdot a \cdot b \cdot Cov(X,Y)$$
$$Var\,(aX - bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var(Y)} - 2 \cdot a \cdot b \cdot Cov(X,Y)$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X) \cdot \sigma(Y)}$$

$$\sigma(X) = \sqrt{Var(X)}$$

Свойства дисперсии

X, Y — случайные величины $\quad a$ — константа

1. $Var(a) = 0$
2. $Var(X+Y) = Var(X) + Var(Y)$, если независимы
3. $Var(X-Y) = Var(X) + Var(Y)$, если независимы
4. $Var(a \cdot X) = a^2 \cdot Var(X)$
5. Дисперсия случайной величины – не случайна

4

Известно, что E(X)=−1, E(Y)=−1, Var(X)=9, Var(Y)=4, Corr(X, Y) = 1.
Найдите: E[2Y·(X−1)].

E(2Y(X-1)) = E(2Y)*E(X-1) + Cov(2Y, X-1) =

= 2*E(Y)*(E(X)-1) + 2*Cov(Y, X) =

= 2*E(Y)*E(X) - 2*E(Y) + 2*Corr(Y, X)*std(X)*std(Y) =

= 2*(-1)*(-1) - 2 * (-1) + 2*1*3*2 = 16

9. Если X и Y зависимы, тогда

$$\mathbb{E}\,(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) + Cov(X,Y)$$

$$Var\,(aX + bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) + 2 \cdot a \cdot b \cdot Cov(X,Y)$$
$$Var\,(aX - bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) - 2 \cdot a \cdot b \cdot Cov(X,Y)$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X) \cdot \sigma(Y)}$$

$$\sigma(X) = \sqrt{Var(X)}$$

Свойства математического ожидания

X, Y — случайные величины $\quad a$ — константа

1. $\mathbb{E}(a) = a$
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
3. $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$
4. $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, если независимы
5. Математическое ожидание случайной величины – не случайно
6. $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$

Свойства ковариации

X, Y, Z — случайные величины $\quad a$ — константа

1. $Cov(X,Y) = Cov(Y,X)$
2. $Cov(a\,,b) = 0$
3. $Cov(a \cdot X,\ Y) = a \cdot Cov(X,Y)$
4. $Cov(X + a,\ Y) = Cov(X,Y)$
5. $Cov(X + Z,\ Y) = Cov(X,Y) + Cov(Z,Y)$
6. $Cov(X,X) = Var(X)$

5

Известно, что E(X)=−1, E(Y)=−1, Var(X)=9, Var(Y)=4, Corr(X, Y) = 1. Найдите Corr(Y−2X−3,X).

$$\mathrm{Corr(Y\text{-}2X\text{-}3,X)} = \frac{\mathrm{Cov(Y\text{-}2X\text{-}3,X)}}{\mathrm{(Var(Y\text{-}2X\text{-}3)*Var(X))\hat{}0.5}} =$$

$$= \frac{\mathrm{Cov(Y\text{-}2X,X)}}{\mathrm{(16*9)\hat{}0.5}} = \frac{\mathrm{Cov(Y,X)+Cov(\text{-}2X,X)}}{4*3} =$$

$$= \frac{\mathrm{Cov(Y, X)\text{-}2*Cov(X, X)}}{12} = \frac{\sigma(X)\sigma(Y)\mathrm{corr(X,Y)\text{-}2Var(X)}}{12} =$$

$$= \frac{3*2*1 - 2*9}{12} = \frac{-12}{12} = -1$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

$$\sigma(X) = \sqrt{Var(X)}$$

Свойства ковариации

X, Y, Z — случайные величины a — константа

1. $Cov(X,Y) = Cov(Y,X)$
2. $Cov(a, b) = 0$
3. $Cov(a \cdot X,\ Y) = a \cdot Cov(X,Y)$
4. $Cov(X + a,\ Y) = Cov(X,Y)$
5. $Cov(X + Z,\ Y) = Cov(X,Y) + Cov(Z,Y)$
6. $Cov(X,X) = Var(X)$

Свойства дисперсии

X, Y — случайные величины a — константа

1. $Var(a) = 0$
2. $Var(X+Y) = Var(X) + Var(Y)$, если независимы
3. $Var(X-Y) = Var(X) + Var(Y)$, если независимы
4. $Var(a \cdot X) = a^2 \cdot Var(X)$
5. Дисперсия случайной величины – не случайна

9. Если X и Y зависимы, тогда

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) + Cov(X,Y)$$

$$Var(aX + bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) + 2 \cdot a \cdot b \cdot Cov(X,Y)$$
$$Var(aX - bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) - 2 \cdot a \cdot b \cdot Cov(X,Y)$$

6

Пусть X_i — длина i-го питона в сантиметрах, а Y_i — в дециметрах. Чему равен выборочный коэффициент корреляции между этими наборами данных?

$$\hat{\rho}(X,Y) = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\hat{\sigma}_x \cdot \hat{\sigma}_y}$$

Y = 10X

Corr(10X, Y) = Corr(Y,Y) = 1

Свойства ковариации

X, Y, Z – случайные величины a – константа

1. $Cov(X,Y) = Cov(Y,X)$
2. $Cov(a, b) = 0$
3. $Cov(a \cdot X, Y) = a \cdot Cov(X,Y)$
4. $Cov(X + a, Y) = Cov(X,Y)$
5. $Cov(X + Z, Y) = Cov(X,Y) + Cov(Z,Y)$
6. $Cov(X,X) = Var(X)$

7

Выберите все верные ответы. Дисперсию случайной величины X можно найти, зная:

- [ ] E(X^2), E(X)
- [ ] E(XY), E(Y)
- [x] E^2(X), E(X)
- [x] Var(X+Y) , Cov(X,Y) и Var(Y)

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

$$Var\ (aX + bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(\mathrm{Y}) + 2 \cdot a \cdot b \cdot Cov(X, Y)$$

Var(X+Y) = Var(X) + Var(Y) + 2*Cov(X,Y)

Var(X) = Var(Y)+2*Cov(X,Y)-Var(X+Y)

8

Нормальное распределение

Случайные величины X и Y имеют совместное нормальное распределение с нулевым математическим ожиданием и единичной ковариационной матрицей. Выберите все верные утверждения: 1 балл

- [ ] Var(X−Y)<Var(X)+Var(Y)
- [ ] Corr(X,Y)>0
- [x] Cov(X,Y)=Corr(X,Y)=Var(X)−Var(Y)
- [x] E(XY)=E(X)E(Y)
- [ ] Corr(X,Y)<0
- [x] X,Y независимы
- [ ] Распределение X может быть дискретным
- [x] Var(X)=1

$$N \sim \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Cov(X,Y) = 0 -> независимые

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right]$$

Математическое ожидание

$\mathbb{E}(X_1) = \mu_1$

$\mathbb{E}(X_2) = \mu_2$

Ковариационная матрица

$Var(X_1) = \sigma_1^2$

$Var(X_2) = \sigma_2^2$

$Cov(X_1, X_2) = \rho$

Свойства дисперсии

X, Y — случайные величины $\quad a$ — константа

1. $Var(a) = 0$
2. $Var(X + Y) = Var(X) + Var(Y)$, если независимы
3. $Var(X - Y) = Var(X) + Var(Y)$, если независимы
4. $Var(a \cdot X) = a^2 \cdot Var(X)$
5. Дисперсия случайной величины – не случайна

Свойства математического ожидания

X, Y — случайные величины $\quad a$ — константа

1. $\mathbb{E}(a) = a$
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
3. $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}(X)$
4. $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$, если независимы
5. Математическое ожидание случайной величины – не случайно
6. $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$

9

Выберите все верные утверждения: 1 балл

- [x] Коэффициент асимметрии используются для анализа скошенности. Он помогает сравнивать различные распределения, например, с нормальным
- [ ] Произведение нормальных случайных величин будет нормальной случайной величиной, если эти величины независимы
- [x] Сумма нормальных случайных величин будет нормальной случайной величиной
- [x] Функцию распределения для нормальной случайной величины нельзя найти в аналитическом виде
- [x] Большое отрицательное значение коэффициента эксцесса говорит о том, что в выборке вероятны выбросы из-за более тяжёлых, чем у нормального распределения хвостов
- [ ] В диапазоне от -3σ до 3σ лежит 95% всех значений нормально распределённой случайной величины
- [ ] Эксцесс позволяет сравнивать различные распределения с нормальным на предмет скошенности

Эксцесс и куртосис

Эксцессом случайной величины X называют величину

$$\beta_X = \underbrace{\frac{\mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^4\right]}{\sigma^4}}_{\text{Куртосис}} - 3$$

- Число 3 вычитается из куртосиса, чтобы эксцесс нормального распределения был равен нулю
- Если хвосты распределения легче, а пик острее, чем у нормального распределения, тогда $\beta_X > 0$
- Если хвосты распределения тяжелее, а пик более приплюснутый, тогда $\beta_X < 0$

Коэффициент асимметрии (skewness)

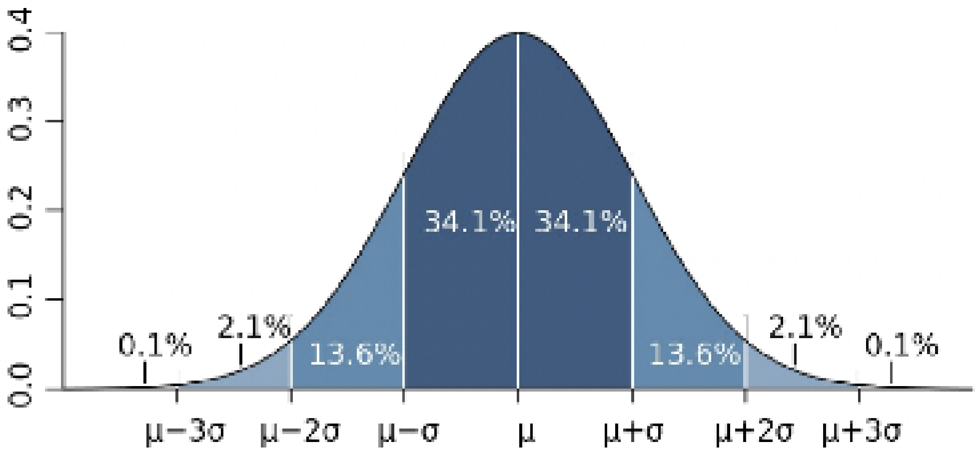
Коэффициентом асимметрии случайной величины X называют величину

$$A_X = \frac{\mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^3\right]}{\sigma^3}$$

- Если плотность распределения симметрична, то $A_X = 0$
- Если левый хвост тяжелее, то $A_X > 0$
- Если правый хвост тяжелее, то $A_X < 0$

Свойства нормального распределения

$$\mathbb{P}(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.997$$



$X \sim N(\mu_x, \sigma_x^2)$ $\quad Y \sim N(\mu_y, \sigma_y^2)$ $\quad a$ – константа

3. $X + Y \sim N(\mu_x + \mu_y,\ \sigma_x^2 + \sigma_y^2)$
4. $X + a \sim N(\mu_x + a,\ \sigma_x^2)$
5. $a \cdot X \sim N(a \cdot \mu_x,\ a^2 \cdot \sigma_x^2)$

Нормальная случайная величина устойчива
к суммированию и линейным преобразованиям

10-12

У случайно выбираемого взрослого мужчины рост в сантиметрах (X) и вес в килограммах (Y) являются нормальным случайным вектором Z=(X,Y) с математическим ожиданием E(Z)=(180,90) и ковариационной матрицей как на картинке ниже.

Рассмотрим величину U=X−Y.

Считается, что человек страдает избыточным весом, если U≤80.

Ковариационная матрица Z

$$Var(Z) = \begin{pmatrix} 100 & 35 \\ 35 & 25 \end{pmatrix}$$

Свойства математического ожидания

X, Y — случайные величины $\qquad a$ — константа

1. $\mathrm{E}(a) = a$
2. $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
3. $\mathrm{E}(a \cdot X) = a \cdot \mathrm{E}(X)$
4. $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$, если независимы
5. Математическое ожидание случайной величины – не случайно
6. $\mathrm{E}(X - \mathrm{E}(X)) = \mathrm{E}(X) - \mathrm{E}(\mathrm{E}(X)) = \mathrm{E}(X) - \mathrm{E}(X) = 0$

9. Если X и Y зависимы, тогда

$$\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y) + Cov(X, Y)$$

$$Var(aX + bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) + 2 \cdot a \cdot b \cdot Cov(X, Y)$$
$$Var(aX - bY) = a^2 \cdot Var(X) + b^2 \cdot \mathrm{Var}(Y) - 2 \cdot a \cdot b \cdot Cov(X, Y)$$

Таблица 3.2 Нормативная функция нормального распределения

z	0,0	−0,1	−0,2	−0,3	−0,4	−0,5	−0,6	−0,7	−0,8	−0,9
Φ(z)	0,500	0,460	0,421	0,382	0,345	0,309	0,274	0,242	0,212	0,184
z	−1,0	−1,1	−1,2	−1,3	−1,4	−1,5	−1,6	−1,7	−1,8	−1,9
Φ(z)	0,159	0,136	0,115	0,097	0,081	0,067	0,055	0,045	0,036	0,029
z	−2,0	−2,1	−2,2	−2,3	−2,4	−2,5	−2,6	−2,7	−2,8	−2,9
Φ(z)	0,023	0,018	0,014	0,011	0,008	0,006	0,005	0,004	0,003	0,002
z	−3,0	−3,1	−3,2	−3,3	−3,4	−3,5	−3,6	−3,7	−3,8	−3,9
Φ(z)	0,013	0,0011	0,0007	0,0005	0,0003	0,0002	0,0002	0,0001	0,0001	0,000
z	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Φ(z)	0,500	0,540	0,579	0,618	0,655	0,691	0,726	0,758	0,788	0,816
z	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
Φ(z)	0,841	0,864	0,885	0,903	0,919	0,933	0,945	0,955	0,964	0,971
z	2,0	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9
Φ(z)	0,977	0,982	0,986	0,989	0,992	0,994	0,995	0,996	0,997	0,998
z	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	38	3,9
Φ(z)	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,000

U = X−Y

E(X) = 180
E(Y) = 90

Из ковариационной матрицы:

Var(X) = 100
Var(Y) = 25
Cov(X, Y) = 35

1. E(U)?

E(X-Y) = E(X) - E(Y) = 180 - 90 = 90

2. Var(U)?

Var(X-Y) = Var(X) + Var(Y) - 2*Cov(X, Y) = 100 + 25 - 2*35 = 55

3. Считается, что человек страдает избыточным весом, если U≤80. Найдите вероятность того, что случайно выбранный мужчина страдает избыточным весом. Ответ округлите до сотых.

U ~ N(90, 55)

$$P(U \le 80) = P\left(\frac{U-90}{7.4} \le \frac{80-90}{7.4}\right) = P\left(\frac{U-90}{7.4} \le -1.35\right) =$$

$$= F(-1.35) = 0.089$$