UNIVERSIDADE DE SÃO PAULO
Faculdade de Zootecnia e Engenharia de Alimentos

ZAB1111 – Estatística Básica

Aula 10 – PROBABILIDADE (4)

MODELOS PARA VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Material preparado pelo Prof. Dr. César Gonçalves de Lima – FZEA/USP

7. VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Definição: Uma variável aleatória contínua é uma função que pode assumir infinitos valores num intervalo de números reais.

Associaremos a cada subintervalo do seu domínio uma probabilidade usando uma função densidade de probabilidade ($f.d.p.$).

Definição 1. Uma função $f(x)$ definida para todo x $[a, b]$, é chamada **função densidade de probabilidade** (f.d.p.) se satisfaz as seguintes condições:

a) $f(x)$ é positiva, para todo $x \in [a, b]$;

b) $\int_a^b f(x)dx = 1$, ou seja, a área sob a curva representativa de $f(x)$, entre as abscissas a e b, é igual a um.

Observe que:

a) A função $f(x)$ não define uma probabilidade.

b) O que define uma probabilidade é <u>o valor da integral</u> definida de $f(x)$ no intervalo $[x_1, x_2]$, por exemplo, que coincide com a <u>área da região sob a curva</u> de $f(x)$, o eixo das abscissas e os limites de integração.

c) Para **calcular a probabilidade** da v.a. X assumir valores entre x_1 e x_2, com $x_1 < x_2$, precisamos resolver:

$$P(x_1 < \mathrm{X} < x_2) = \int_{x_1}^{x_2} f(x)dx$$

d) $P(X = k) = 0$ porque $\int_k^k f(x)dx = [F(x)]_k^k = F(k) - F(k) = 0$.

e) Definimos a **função distribuição acumulada** ($f.d.a.$) da variável contínua X, com f.d.p. $f(x)$, por:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)dz$$

Importante: Conhecendo a $f.d.a.$ da v.a. X

$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

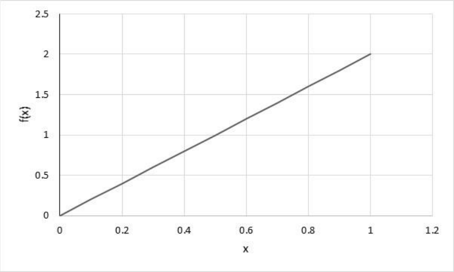
f) Se X é uma v.a. <u>contínua</u> definida no intervalo $[a; b]$ e $f(x)$ é sua f.d.p. definimos:

Média ou Esperança Matemática: $\mu = E(\mathrm{X}) = \int_a^b x\, f(x)\, dx$

Variância: $\sigma^2 = var(\mathrm{X}) = \int_a^b [x - E(X)]^2 f(x)\, dx$

Exemplo 1. Dada a função $f(x) = 2x$, $x \in [0, 1]$ pede-se:

i) Verificar se $f(x)$ é uma função densidade de probabilidade.



$f(x)$ é positiva para $x \in [0, 1]$

$$\int_0^1 2x\, dx = x^2 \Big|_0^1 = 1$$

ii) Calcular $P(0 < X < 0{,}5)$ e $P(0{,}2 < X < 0{,}7)$.

- $P(0 < X < 0{,}5) = x^2|_0^{0{,}5} = 0{,}25$
- $P(0{,}2 < X < 0{,}7) = x^2|_{0{,}2}^{0{,}7} = 0{,}49 – 0{,}04 = 0{,}45$

iii) Calcular a média dessa distribuição

- $\mu = E(\mathrm{X}) = \int_0^1 x\ (2x)dx = \int_0^1 2x^2\ dx = \frac{2}{3} x^3 \mid_0^1 = \frac{2}{3}$

A seguir conheceremos três modelos para $v.a.$ contínuas: os modelos Uniforme, Exponencial e Normal, que são úteis e muito usados em diversas áreas de pesquisa.

Para cada modelo vamos conhecer sua f.d.p., seu gráfico, sua média, variância e alguma forma simples de calcular probabilidades.

8. ALGUNS MODELOS PROBABILÍSTICOS PARA V. A. CONTÍNUA

8.1. DISTRIBUIÇÃO UNIFORME: é a distribuição de probabilidades mais simples. Sua f.d.p. é uma constante e $P(x_1 < X < x_2)$ é proporcional ao tamanho do intervalo.

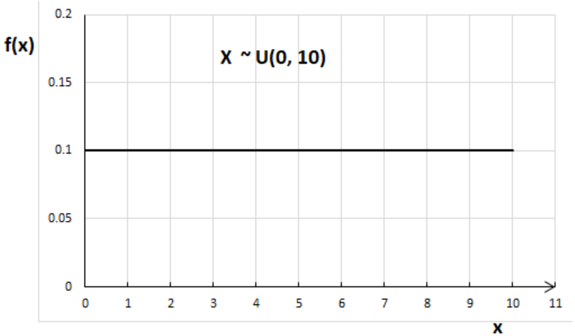
Definição 8.1. A v.a. contínua X tem distribuição uniforme no intervalo real $[a, b]$ se a sua f.d.p. for dada por:

$$f(x) = \frac{1}{b-a}$$

Pode-se provar que: $E(X) = \frac{a+b}{2}$, $var(X) = \frac{(b-a)^2}{12}$ e $F(x) = \frac{x-a}{b-a}$ para todo $x \in [a, b]$,

Nota: As probabilidades calculadas em intervalos de mesma amplitude são iguais.

Exemplo 1. Se $X \sim U(0{,}10)$ calcular $P(1 < X < 3)$ e $P(2 < X < 4)$



Para calcular probabilidades vamos usar a função de distribuição acumulada:

$$F(x) = \frac{x-a}{b-a} = \frac{x-0}{10-0} = \frac{x}{10}$$

- $P(1 < X < 3) = F(3) - F(1) = \frac{3}{10} - \frac{1}{10} = 0{,}3 - 0{,}1 = 0{,}2$
- $P(2 < X < 4) = F(4) - F(2) = \frac{4}{10} - \frac{2}{10} = 0{,}4 - 0{,}2 = 0{,}2$

8.2. A DISTRIBUIÇÃO EXPONENCIAL é usada para descrever o tempo até a ocorrência de um evento.

Exemplos: tempo de vida de uma bateria de celular; tempo exigido para um técnico executar certa tarefa; tempo de chegada de um carro a um posto de pedágio *etc*.

Definição 8.2. Dizemos que a v.a. contínua X, definida para valores positivos, tem distribuição *exponencial* de parâmetro $\lambda > 0$, se a sua f.d.p. é dada por:

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} = \frac{1}{\lambda} exp(-x/\lambda)$$

Pode-se provar que:

$E(X) = \lambda$, $var(X) = \lambda^2$ e $F(x) = P(X \leq x) = 1 - e^{-x/\lambda}$, para $x > 0$.

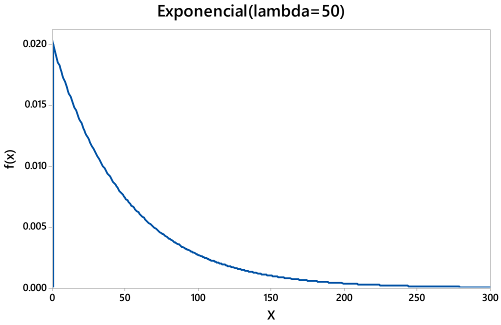


Figura 8.1. Função densidade de probabilidade de $X \sim exp(\lambda = 50)$

Note que: $P(0<X<50) > P(50<X<100) > P(200<X<250)$, ou seja, as probabilidades calculadas em intervalos de mesma amplitude serão tão menores, quanto mais distantes da origem estiverem os seus limites.

Para calcular probabilidades de uma variável com distribuição exponencial é mais comum usar a sua <u>função distribuição acumulada</u>:

$$F(x) = P(X \leq x) = 1 - e^{-x/\lambda}$$

Exemplo: Calcular a probabilidade da variável $X \sim exp(\lambda)$ assumir um valor entre x_1 e x_2.

$$\begin{aligned} P(x_1 < X < x_2) &= F(x_2) - F(x_1) \\ &= \left[1 - e^{-x_2/\lambda}\right] - \left[1 - e^{-x_1/\lambda}\right] \\ &= e^{-x_1/\lambda} - e^{-x_2/\lambda} \end{aligned}$$

Exemplo 2. A duração da carga (em horas) de um determinado tipo de bateria de celular é uma v.a. T, contínua, com distribuição exponencial de média $\lambda = 50$h. Calcular a probabilidade de que a carga desta bateria dure entre 50 e 60 horas de uso.

Resolução: Se $\mathrm{T} \sim exp(50) \Rightarrow f(t) = \frac{1}{50} e^{-t/50} \Rightarrow F(t) = 1 - e^{-t/50}$, para qualquer $t > 0$. Então:

$$
\begin{aligned}
P(50 \leq \mathrm{T} \leq 60) &= \int_{50}^{60} \frac{1}{50} e^{-t/50}\, dt = F(60) - F(50) \\
&= e^{-50/50} - e^{-60/50} = e^{-1} - e^{-1,2} \\
&= 0,3679 - 0,3012 = 0,0667
\end{aligned}
$$

$\therefore$ A chance de a carga desta bateria de celular durar entre 50 e 60 horas é de apenas 0,0667.

8.3. O MODELO NORMAL (ou de Gauss)

A distribuição normal foi introduzida pelo matemático Abraham de Moivre (1733) e é uma das distribuições probabilísticas mais importantes da Estatística, pois é usada para descrever inúmeros fenômenos físicos, biológicos e financeiros.

Definição 8.3. Dizemos que a variável contínua X tem distribuição normal, com parâmetros μ e σ^2 se a sua f.d.p. é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad \text{para } -\infty < x < \infty$$

Pode-se provar que: $\mu = E(\text{X})$, $\sigma^2 = var(\text{X})$ e $F(x) = P(X \leq x)$ é difícil de ser calculada.

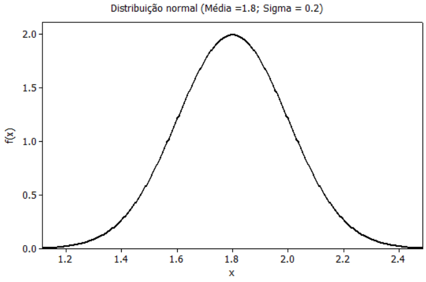
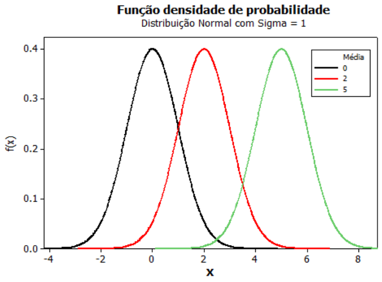


Figura 8.3. Distribuição normal com média $\mu = 1{,}8$ e desvio padrão $\sigma = 0{,}2$.

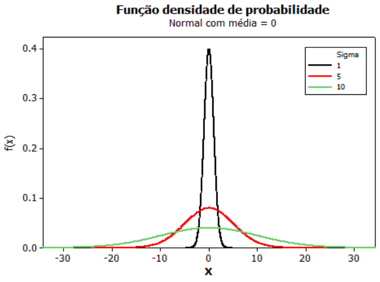
Características interessantes do gráfico da distribuição normal:

- Tem a forma de um sino.
- Tem uma assíntota horizontal: $f(x) = 0$.
- É simétrico em relação ao ponto de abscissa $x = \mu$.
- Como a curva é simétrica: $P(X > \mu) = P(X < \mu) = 0{,}5$

- $x = \mu$ é a abscissa do ponto de máximo absoluto da função e coincide com a mediana da distribuição.
- Os pontos de abscissas $x_1 = \mu - \sigma$ e $x_2 = \mu + \sigma$ são pontos de inflexão da função.



a) Médias diferentes e mesmo desvio padrão



b) Desvios padrões diferentes e mesma média

Figura 8.4 Distribuição normal com diferentes médias e variâncias

Cálculo de probabilidades:

- Calcular $P(x_1 < \mathrm{X} < x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1)$ é difícil!
- Para facilitar o cálculo de probabilidades usamos uma tabela baseada de uma variável **normal padronizada** ou **reduzida**.

  Para padronizar uma variável $X \sim N(\mu, \sigma^2)$ usamos

$$Z = \frac{X-\mu}{\sigma} \text{ e } Z \sim N(\mu{=}0; \sigma^2{=}1)$$

  As probabilidades $P(0 < Z < z_c)$ ou $F(x_c) = P(Z < z_c)$ estão tabeladas.

- Para calcular $P(x_1 < X < x_2)$ **padronizamos** os limites de integração e usamos a Tábua 1 para calcular a probabilidade:

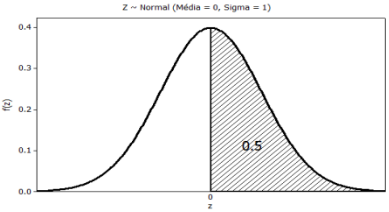
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2)$$

Exemplo 4. Assumindo que o peso de frangos ao abate tem distribuição normal de média 1,80 kg e desvio padrão igual a 0,14 kg, calcular a probabilidade de encontrar um frango com peso

(a) superior a 1,80 kg
(b) inferior a 1,90 kg
(c) inferior a 1,70 kg
(d) entre 1,80 e 2,00 kg
(e) superior a 2,10 kg
(f) entre 1,60 e 1,70 kg.

Utilizando a Tábua 1:

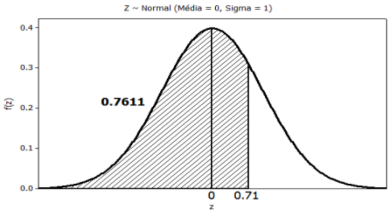
$$\text{(a) } P(X > 1{,}80) = P\left(\frac{X - 1{,}80}{0{,}14} \geq \frac{1{,}80 - 1{,}80}{0{,}14}\right)$$
$$= P(Z > 0) = 0{,}50$$



(b) $P(X < 1{,}90) = P\left(Z < \frac{1{,}90-1{,}80}{0{,}14}\right) =$

$= P(Z < 0{,}71)$

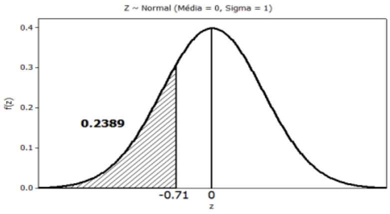
$= 0{,}50 + 0{,}2611 = 0{,}7611$



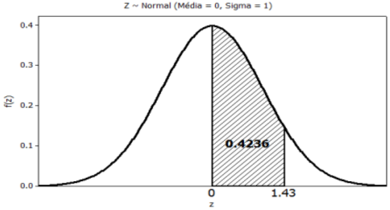
(c) $P(X < 1{,}70) = P\left(Z < \frac{1{,}70-1{,}80}{0{,}14}\right)$

$= P(Z < -0{,}71)$

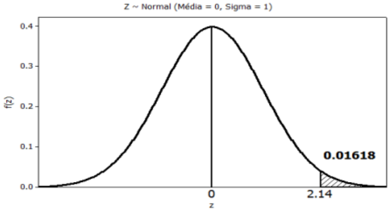
$= 0{,}50 - 0{,}2611 = 0{,}2389$



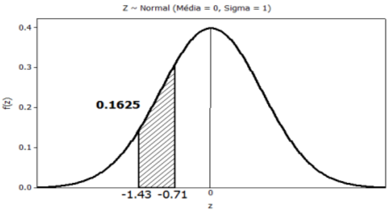
(d) $P(1{,}80 < X < 2{,}00)$
$= P(0 < Z < 1{,}43) = 0{,}4236$



(e) $P(X > 2{,}10) = P(Z > 2{,}14)$
$= 0{,}50 – 0{,}4838 = 0{,}0162$



(f) $P(1{,}60 < X < 1{,}70)$
$= P(-1{,}43 < Z < -0{,}71)$
$= 0{,}4236 - 0{,}2611 = 0{,}1625$



Também podemos calcular probabilidades usando a distribuição acumulada de probabilidades da normal padrão, que é dada por:

$$F(z_c) = P(Z \leq z_c) = \int_{-\infty}^{z_c} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt$$

Essas probabilidades acumuladas também estão tabeladas e para calcular, por exemplo, $P(x_1 < X < x_2)$ fazemos:

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = F(z_2) - F(z_1)$$

Exemplo 5. Admitindo que a altura dos alunos de Estatística tem distribuição normal com média $\mu = 1{,}67m$ e desvio padrão $\sigma = 0{,}08m$ calcule a probabilidade de um aluno escolhido ao acaso ter altura:

$a)\ P(X > 1{,}67)$ $\quad b)\ P(X > 1{,}80)$ $\quad c)\ P(1{,}60 < X < 1{,}75)$

$d)\ P(X < 1{,}50)$ $\quad e)\ P(X > 1{,}85)$

Usando a tabela com as probabilidades $F(z_c) = P(Z \leq z_c)$ temos:

$a)\ P(X > 1{,}67) = P(Z > 0) = 0{,}5$

$$b)\ P(X > 1{,}80) = P(Z > 1{,}63) = 1 - F(1{,}63)$$

$$= 1 - 0{,}9484 = 0{,}0516$$

$$c)\ P(1{,}60 < X < 1{,}75) = P(-0{,}88 < Z < 1{,}00)$$

$$= F(1{,}00) - F(-0{,}88) = 0{,}8413 - 0{,}1894 = 0{,}6519$$

$$d)\ P(X < 1{,}50) = P(Z < -2{,}13) = F(-2{,}13) = 0{,}0166$$

$$e)\ P(X > 1{,}85) = P(Z > 2{,}25) = 1 - F(2{,}25) = 1 - 0{,}9878$$
$$= 0{,}0122$$

8.4. APROXIMAÇÃO DA DISTRIBUIÇÃO BINOMIAL PELA NORMAL

Podemos usar a distribuição normal, que é associada a variáveis contínuas, para calcular valores $aproximados$ para as probabilidades de uma distribuição binomial, que está associada a variáveis aleatórias discretas.

Exemplo 6. Uma moeda é lançada 10 vezes. Seja X o "número de caras em 10 lançamentos", então a probabilidade exata de ocorrerem 7 caras ou mais é:

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$
$$= 0,117 + 0,044 + 0,010 + 0,001 = 0,172.$$

Podemos calcular um valor aproximado de $P(X \geq 7)$ usando a distribuição distribuição normal de média

$$\mu = np = (10)(0,5) = 5$$

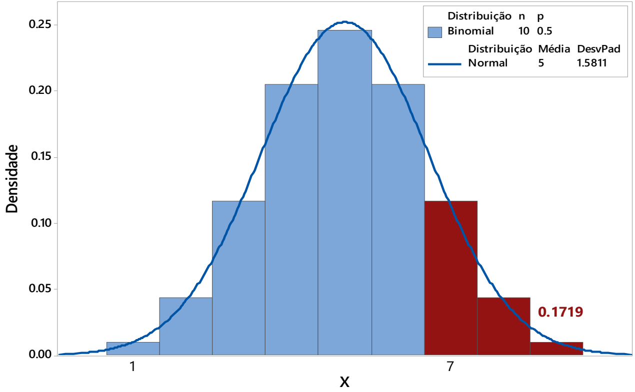
e variância

$$\sigma^2 = np(1-p) = (10)(0,5)(0,5) = 2,5$$

ou seja, vamos usar a <u>variável contínua</u> $W \sim N(\mu = 5; \sigma^2 = 2,5)$ para calcular um <u>valor aproximado</u> para $P(X \geq 7)$, em que X é uma variavel discreta:

$$P(X \geq 7) \cong P(W \geq 6,5) = P(Z \geq 0,949) = 0,1711$$

Perceba que este valor obtido é bem próximo da probabilidade exata calculada com a distribuição binomial.



k	$P(X = k)$
0	0,001
1	0,010
2	0,044
3	0,117
4	0,205
5	0,246
6	0,205
7	0,117
8	0,044
9	0,010
10	0,001

Figura 8.5. Aproximação da $Binomial(10; 0{,}5)$ pela $N(0{,}5; 2{,}5)$

Observações importantes:

- A aproximação de probabilidades da distribuição binomial pela distribuição normal será tanto melhor quanto maior for o valor de n e mais próximo de 0,5 for o valor de $p = P(sucesso)$.

- Quando o valor de n for grande ($n \longrightarrow \infty$) e o valor de p for muito pequeno ($p \rightarrow 0$), podemos obter melhores aproximações para as probabilidades de uma $Binomial\ (n, p)$ utilizando uma distribuição de $Poisson\ (\lambda = np)$.

EXERCÍCIOS PROPOSTOS

1) Suponha que o tempo de vida, T, de um vírus exposto ao meio ambiente segue uma distribuição exponencial com parâmetro $\lambda = 5$ segundos. Calcule a probabilidade do tempo de vida do vírus:

a) estar entre 3 e 5 segundos, $P(3 < T < 5)$.

b) ser superior a 4,5 segundos, $P(T > 4{,}5)$.

c) ser superior a 4 segundos, dado que é superior a 3 segundos, ou seja, $P(T > 4 \mid T > 3)$

2) Quando a v.a. X tem distribuição normal de média μ e variância σ^2, escrevemos $X \sim N(\mu, \sigma^2)$. Com base nesta distribuição calcule:

a) $P(X > 2\sigma)$ b) $P(|X - \mu| \leq 1{,}5\sigma)$ c) $P(|X - \mu| > \sigma)$

d) O número a tal que $P(\mu - a\sigma < X < \mu + a\sigma) = 0{,}89$

3) As notas finais de Cálculo II dos alunos de pós-graduação em Engenharia de Biossistemas têm distribuição normal, com média 6,4 e desvio padrão 1,4.

a) Um professor de Bioquímica decide atribuir conceito A para alunos com notas superiores a 8,5; conceito B para alunos com notas entre 5 e 8,5 e conceito C para alunos com notas inferiores a 5. Em uma classe de 50 alunos, qual é o número esperado de alunos com conceito A? E com conceito B? E com conceito C?

b) Outro professor deseja atribuir o conceito A para 20% dos melhores alunos z(maiores notas), o conceito B para os próximos 40% e o conceito C para 40% dos piores alunos (menores notas). Encontre as notas limites para cada um dos conceitos.

c) Qual método você acha mais interessante? Por quê?

4) Para se ajustar a uma máquina a correia deve ter entre 60 e 62 cm de comprimento. Tendo em vista o processo de fabricação, o comprimento dessas correias pode ser considerado uma variável aleatória com distribuição Normal de média 60,7 cm e desvio padrão 0,8 cm. Pergunta-se:

a) Qual a probabilidade de uma correia, escolhida ao acaso, não se ajustar à máquina?

b) Um grande revendedor dessas correias estabelece um controle de qualidade nos lotes que compra da fábrica: ele sorteia 4 correias do lote e só aceita o lote se o comprimento médio estiver dentro do tamanho aceito pela máquina. Calcule a probabilidade de aceitação do lote.

   Dicas: Ver Apêndice

5) As notas da parte de Biologia do exame de admissão ao Medical College (EUA) têm distribuição normal com média 8,0 e desvio padrão 2,6. Dentre os 600 candidatos que fizeram o exame, quantos deverão ter nota entre 6,0 e 7,0?

APÊNDICE

Pode-se provar que qualquer <u>combinação linear de variáveis normais</u> independentes também terá distribuição normal, ou seja, se as variáveis $X_1, X_2, \ldots, X_k$ formam uma sequência de variáveis independentes com distribuição $N(\mu_i, \sigma_i^2)$ e $c_1, c_2, \ldots, c_k$ são constantes quaisquer então a variável

$$W = \sum_{i=1}^{k} c_i X_i = c_1 X_1 + c_2 X_2 + \cdots + c_k X_k$$

também terá distribuição normal com média

$$\mu_W = E\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} E(c_i X_i) = \sum_{i=1}^{k} c_i E(X_i) = \sum_{i=1}^{k} c_{i\,i} \mu_i$$

e variância

$$\sigma_W^2 = var\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} var(c_i X_i) = \sum_{i=1}^{k} c_i^2 \, var(X_i) = \sum_{i=1}^{k} c_i^2 \sigma_i^2$$

Exemplo[i] Uma corretora negocia títulos na Bolsa de Valores e utiliza um modelo probabilístico para avaliar seus lucros. Suas aplicações financeiras de compra e venda atingem três áreas: agricultura, indústria e comércio. Admita que o seguinte modelo representa o comportamento do lucro diário da corretora (em milhares de reais):

$$L = 2L_A + 5L_I + 3L_C$$

Em que L_A, L_I e L_C representam os lucros diários nas áreas de agricultura, indústria e comércio. As distribuições de probabilidades associadas a essas variáveis $L_A \sim N(3;4)$, $L_I \sim N(6;9)$ e $L_C \sim N(4;16)$. Supondo independência entre os resultados das três áreas, qual é a probabilidade de a corretora ter um lucro diário superior a R$ 50 mil?

Como o lucro diário da corretora, L, é uma combinação linear de três normais independentes, L terá distribuição normal com parâmetros:

$$\mu_L = 2\mu_A + 5\mu_I + 3\mu_C = 2(3) + 5(6) + 3(4) = 48$$

e

$$\sigma_L^2 = 2^2(4) + 5^2(9) + 3^2(16) = 385$$

Então $L \sim N(48; 385)$ e, portanto,

$$P(L > 50) = P\left(\frac{L-48}{\sqrt{385}} > \frac{50-48}{\sqrt{385}}\right) = P(Z > 0{,}10) = 0{,}4602$$

Indicando que num determinado dia, a chance de a corretora ter um lucro superior a R$50 mil é de aproximadamente 46%.

Exercício. Se as n variáveis X_i, $i = 1,2,\dots,n$, têm distribuição $N(\mu,\sigma^2)$ e são independentes, mostre que a nova variável

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

tem distribuição normal com média μ e variância $\frac{\sigma^2}{n}$.

---

i Magalhães, M.N.; Pedroso de Lima, A.C. Noções de Probabilidade e Estatística, EDUSP, 2008