Variáveis Aleatórias Contínuas

Departamento de Estatística

Recapitulação

- Já vimos que existem dois tipos de Variáveis Aleatórias:
  - Discretas e
  - Contínuas.
- Sabemos como trabalhar com variáveis aleatórias Discretas;
- Iremos ver nesse slide como lidar com variáveis aleatórias Contínuas;
- Para tal, iniciaremos com um exemplo:

Exemplo – V. A. Contínua

- Estudos anteriores revelam a existência de um grande lençol freático no subsolo de uma região. No entanto, sua profundidade ainda não foi determinada, sabendo-se apenas que o lençol pode estar situado em qualquer ponto entre 20 e 100 metros.
- Vamos supor que escolhemos, ao acaso, um ponto nessa região e dispomos de uma sonda que, ao fazer a perfuração, detecta a profundidade do reservatório de água.
- Denotamos por X a variável aleatória representando a profundidade.

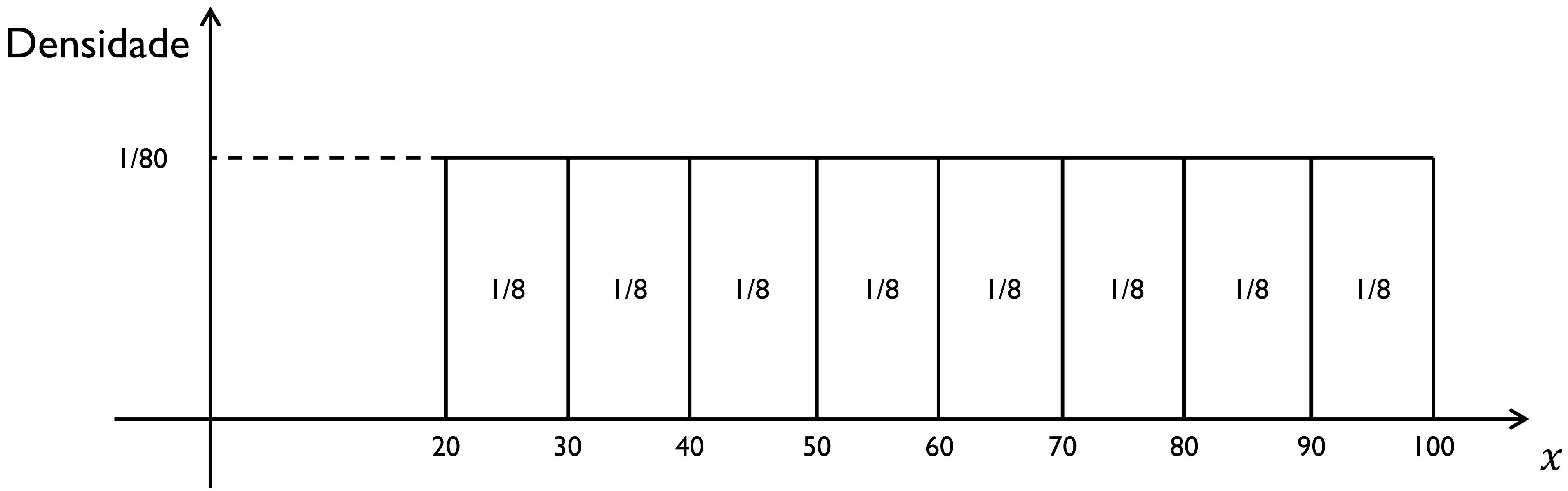
Exemplo – V. A. Contínua

- É razoável assumirmos que a sonda pode parar em qualquer ponto entre 20 e 100 metros, sem que tenhamos motivos para privilegiar essa ou aquela profundidade, ou seja, consideramos todos os pontos como equiprováveis.
- Ao pensar em atribuir uma probabilidade para cada ponto, chegamos a uma dificuldade, como temos infinitos pontos e todos são equiprováveis, teríamos infinitas probabilidades o que causaria uma probabilidade total de valor infinito, e não 1 como deve ser.

Exemplo – V. A. Contínua

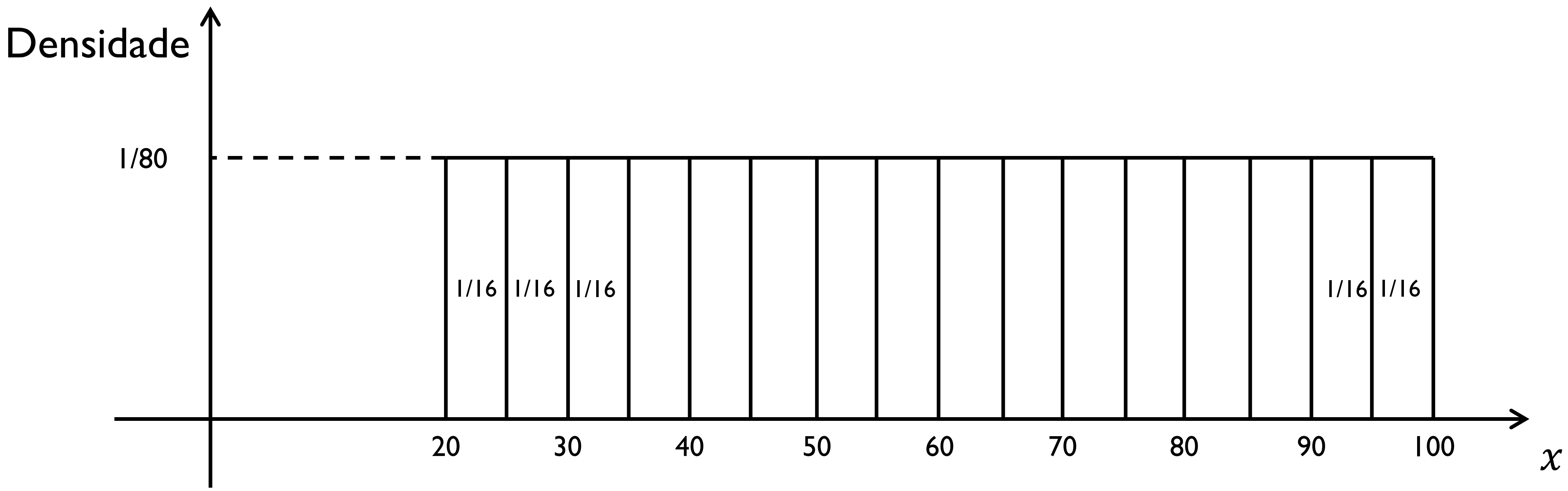
- A solução é não considerar valores únicos no cálculo da probabilidade, e sim intervalos de valores.
- Nesse caso o espaço amostral corresponde ao intervalo $[20,100]$ e as profundidades são igualmente prováveis.
- Suponha que dividimos o espaço amostral em 8 intervalos de comprimento 10. É razoável atribuir aos intervalos a probabilidade 1/8, correspondendo à relação entre o comprimento de cada um deles e o comprimento do espaço amostral.

Exemplo – V. A. Contínua



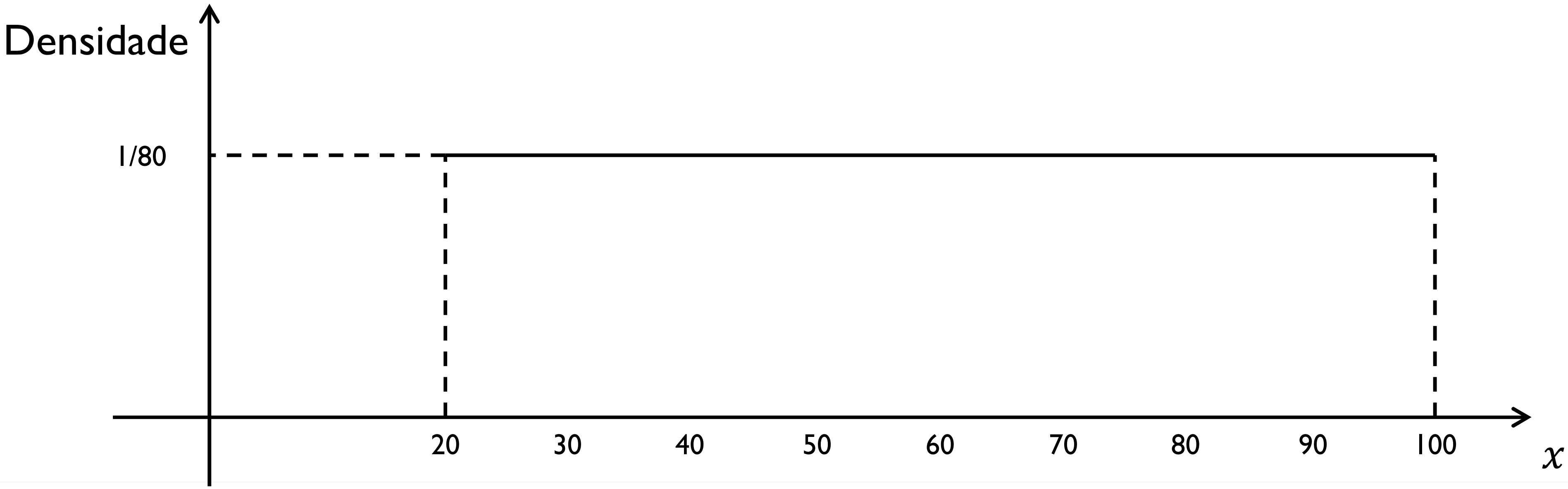
Exemplo – V. A. Contínua

- E se dividíssemos o espaço amostral em 16 intervalos?



Exemplo – V. A. Contínua

- O histograma mostra que, apesar de termos intervalos diferentes, a densidade permanece a mesma.
- Podemos continuar esse procedimento, aumentando cada vez mais o número de faixas, diminuindo suas amplitudes, até, obtermos (teoricamente) infinitos intervalos e o seguinte histograma:



Exemplo – V. A. Contínua

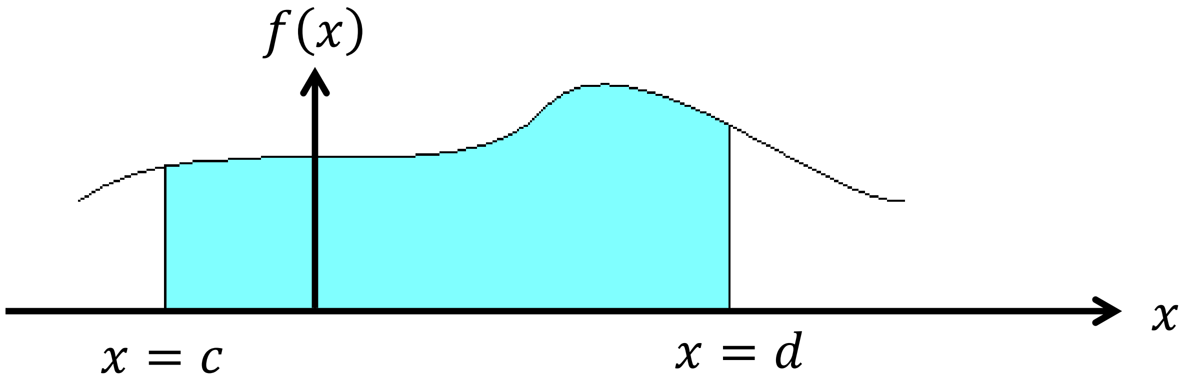
- Sendo assim, a probabilidade de uma variável aleatória contínua é definida pela área abaixo de uma função positiva, denominada *densidade de probabilidade*.
- A densidade em si, não é uma probabilidade, mas uma função matemática que auxilia na atribuição de probabilidades.
- Para a variável aleatória contínua X representando a profundidade do lençol freático, a função densidade f é dada por:
- $f(x) = \begin{cases} \frac{1}{80}, & 20 \leq x \leq 100 \\ 0, & x < 20 \text{ ou } x > 100 \end{cases}$

Definição

- Diz-se que X é uma **variável aleatória contínua**, se existir uma função f, denominada **função densidade de probabilidade** (fdp) de X que satisfaça às seguintes condições:
  - $f(x) \geq 0$ para todo x;
  - A área definida por $f(x)$ é igual a 1;
  - Para quaisquer a, b, com $-\infty < a < b < +\infty$, teremos $P(a \leq X \leq b) = \int_a^b f(x)\,dx$.

Observações

- $P(c < X < d)$ representa a área sob a curva no gráfico abaixo:



- $P(X = x_0) = 0;$
  - $P(c \leq X \leq d) = P(c \leq X < d) = P(c < X \leq d) = P(c < X < d).$
- $f(x)$ não representa a probabilidade de coisa alguma!
  - Somente quando a função for integrada entre dois limites, ela produzirá uma probabilidade.

Exercício 1

- Voltando ao caso do lençol freático cuja variável aleatória X representando sua profundidade possui a seguinte função densidade:
- $f(x) = \begin{cases} 1/80; & 20 \leq x \leq 80 \\ 0; & x < 20 \text{ } ou \text{ } x > 80 \end{cases}$
- Qual a probabilidade de a profundidade estar entre 30 e 50 metros? E abaixo de 50 metros?

O Valor Esperado de uma Variável Aleatória Contínua

- Seja X uma variável aleatória contínua com f.d.p. f. O **valor esperado** de X é definido como

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx,$$

- se, e somente se,

$$\int_{-\infty}^{+\infty} |x| f(x) dx$$

- for finita.

A Variância de uma Variável Aleatória

- Seja X uma variável aleatória. Definimos a **variância** de X, denotada por $V(X)$ ou ${\sigma_X}^2$, da seguinte maneira:

$$V(X) = E\left[\left(X - E(X)\right)^2\right] = E(X^2) - [E(X)]^2$$

- A raiz quadrada positiva de $V(X)$ é denominada o **desvio-padrão** de X, e é denotado por σ_X.

Exercício 2

- Existe uma variável aleatória X com $E(X) = 3$ e $E(X^2) = 8$?

Distribuição Normal

- Uma das distribuições contínua mais utilizadas é a distribuição normal ou Gaussiana;
- A distribuição normal pode ser entendida como uma aproximação da distribuição binomial com probabilidade de sucesso constante e tamanho da amostra tendendo ao infinito;
- Podendo, no entanto, representar qualquer intervalo pertencente ao conjunto dos números reais, diferentemente da Binomial.

Distribuição Normal

- Uma variável aleatória X tem distribuição normal com parâmetros μ e σ^2 [$X \sim N(\mu, \sigma^2)$]se sua densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}$$

- Em que $-\infty < x < \infty; -\infty < \mu < \infty$ e $\sigma^2 > 0$.
- Observações:
  - π representa uma constante, aproximadamente 3,1415;
  - e representa uma outra constante, aproximadamente 2,7182;
  - μ representa a média da distribuição;
  - σ^2 representa a variância da distribuição;
  - Juntos, os parâmetros μ e σ^2 definem uma função densidade de probabilidade normal.

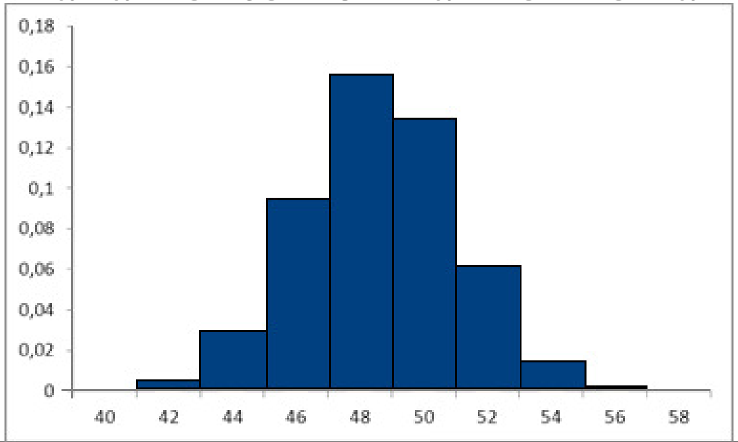
Propriedades da Distribuição Normal

- A distribuição normal é unimodal e simétrica em torno de sua média μ;
  - $P(X < \mu) = P(X > \mu)$;

- O desvio padrão σ é uma medida da dispersão dos dados ao redor da média μ:
  - $P(\mu - \sigma < X < \mu + \sigma) = 0{,}6826$;
  - $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0{,}9546$;
  - $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0{,}9974$.

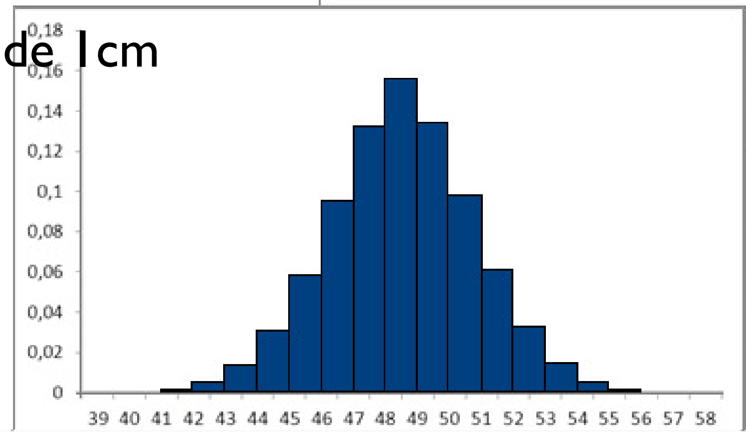
Exemplo

- Suponha que o comprimento de recém-nascidos do sexo feminino não portadores de anomalias congênitas seja uma variável aleatória com distribuição aproximadamente normal de média 48,54cm e desvio-padrão 2,5cm.

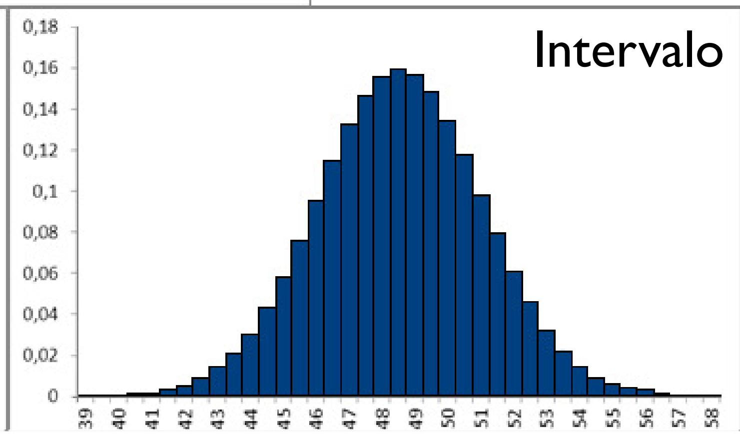


Intervalo de 2cm

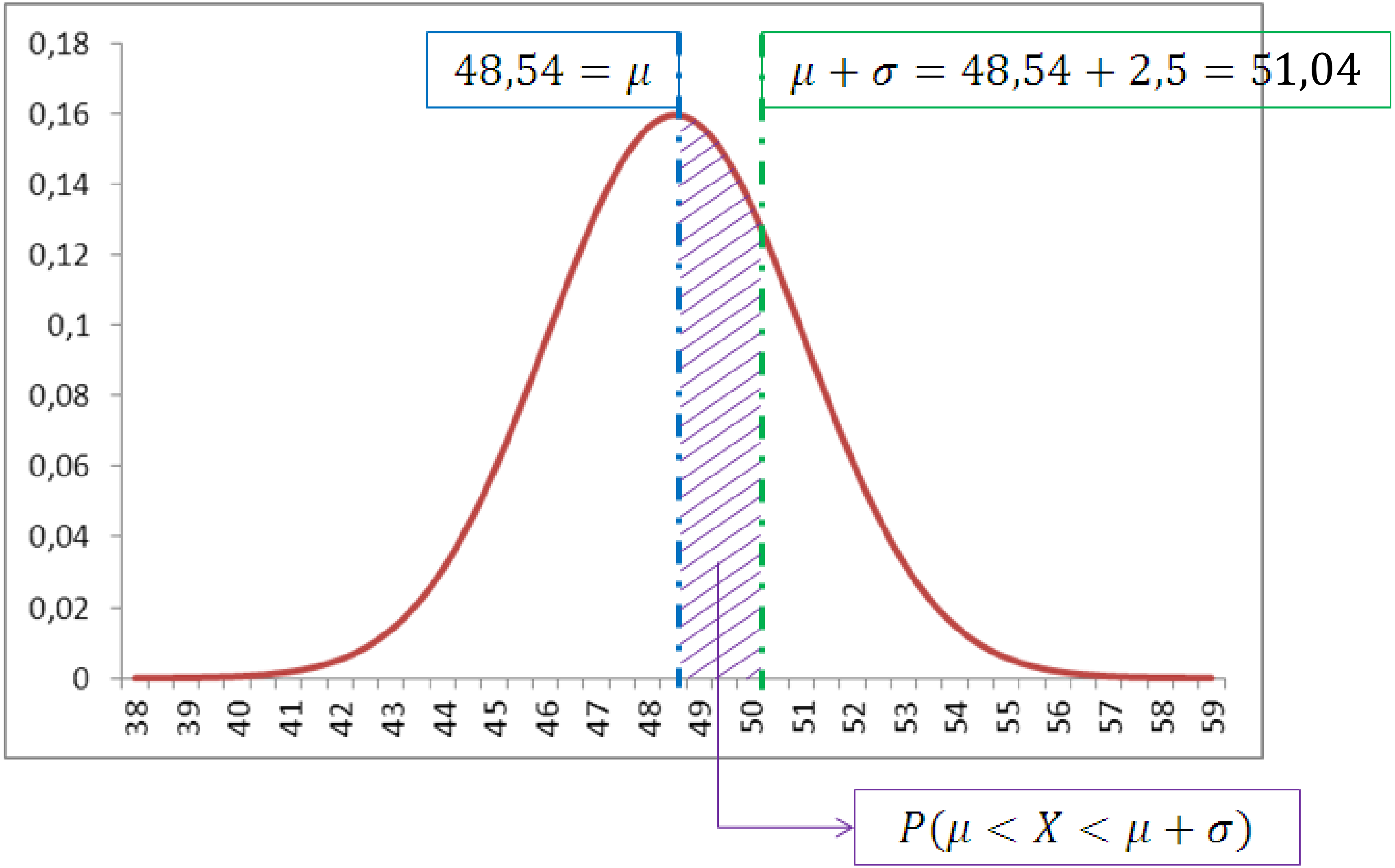
Intervalo de 1cm



Intervalo de 0,5cm



Exemplo



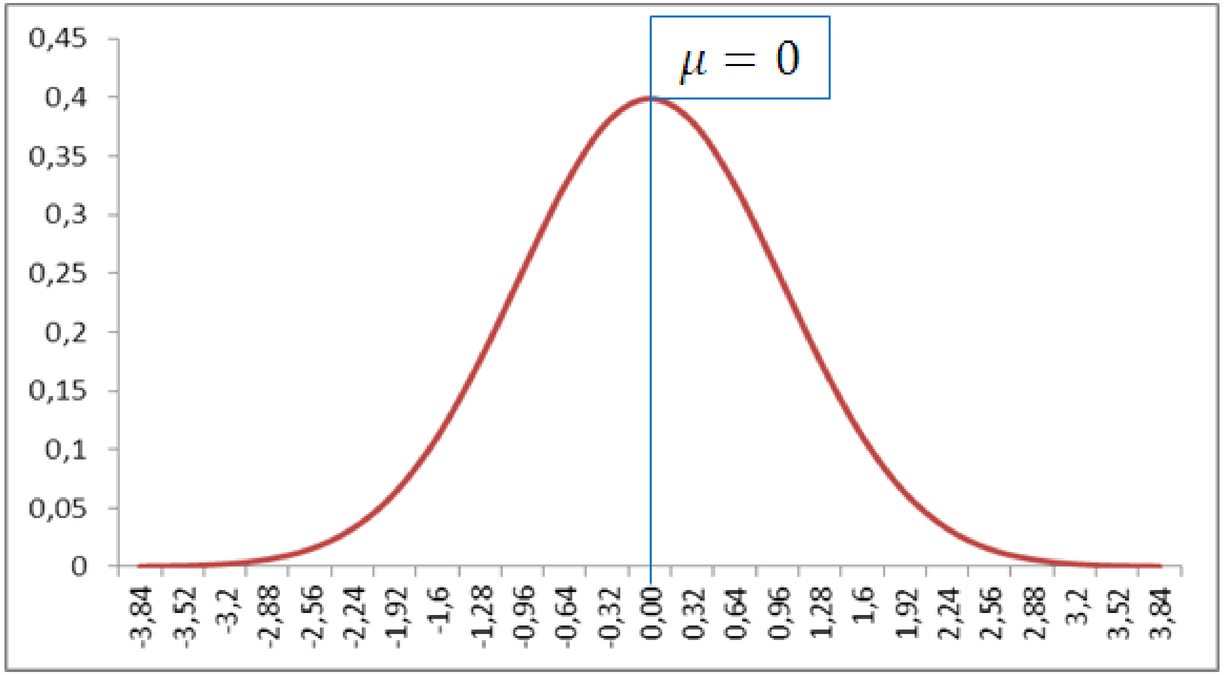
$48{,}54 = \mu$
$\mu + \sigma = 48{,}54 + 2{,}5 = 51{,}04$
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
$P(\mu < X < \mu + \sigma)$

Como Calcular a Probabilidade de X Pertencer a Determinado Intervalo

- Basta calcular a área sob a curva normal relativa a f.d.p. da variável aleatória X;
- Para calcular a área sob um gráfico, é necessário resolver uma integral, nem sempre trivial;
- Como fugir do cálculo de uma integral cada vez que quiser calcular uma probabilidade?
- Como a construção de tabelas para todas as possíveis variáveis aleatórias pertencentes a uma distribuição normal é impossível (existem infinitas combinações de médias e desvios padrão), utiliza-se a tabela da distribuição normal padrão.

Distribuição Normal Padrão

Diz-se que uma variável aleatória Z segue uma distribuição normal padrão se ela segue uma distribuição normal com média 0 e desvio padrão 1 ($\mu = 0$ e $\sigma = 1$).



Distribuição Normal Padrão

Tabela 42: Distribuição Normal Padrão (Z~N(0,1)). Corpo da tabela dá a probabilidade p, tal que p=P(Z≥z_c).

$P(Z \geq z_c)$	**Segunda decimal de z_c**									
Parte inteira e primeira decimal de z_c	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010

Como usar a tabela?

- $P(Z \geq 2{,}65) = 0{,}0040;$

- $P(Z \leq 0{,}5) = 1 - P(Z \geq 0{,}5) = 1 - 0{,}3085 = 0{,}6915;$

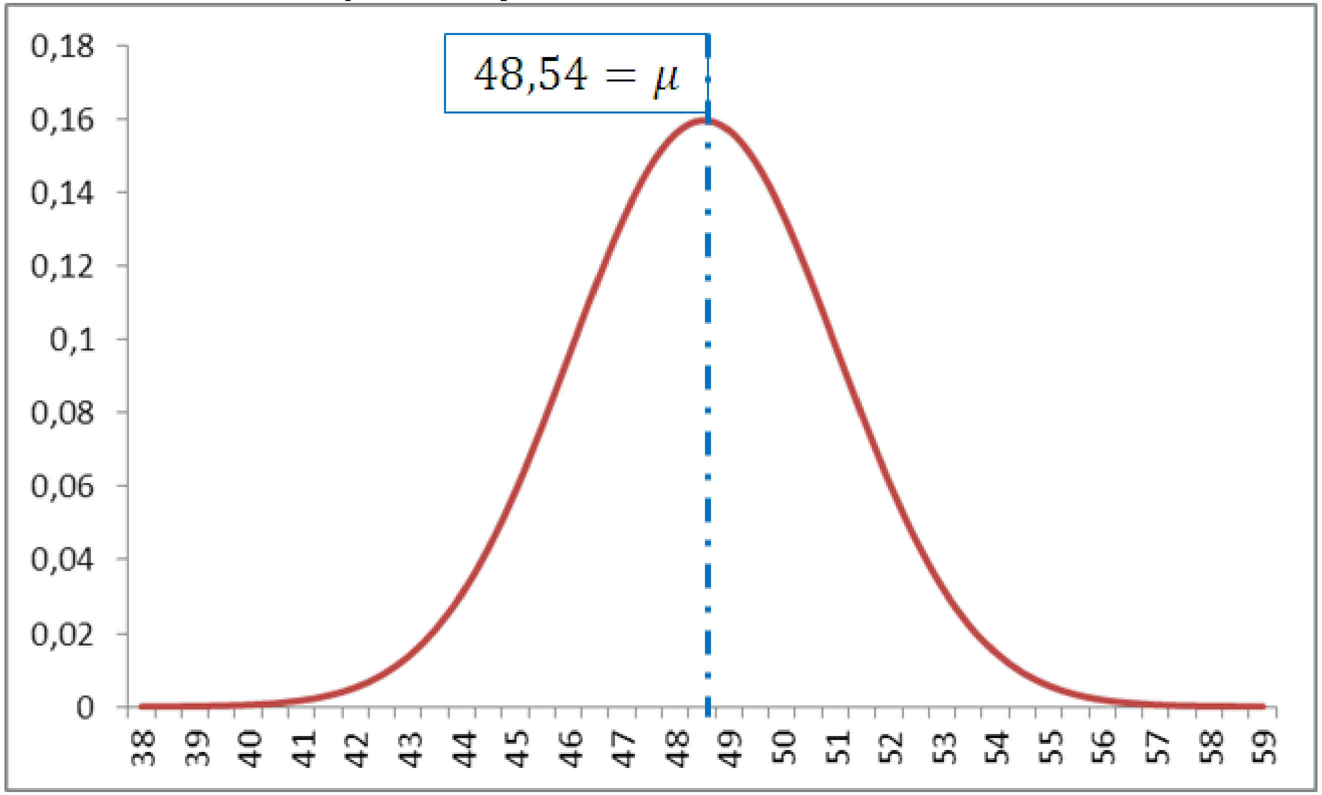
- $P(Z \leq -1{,}85) = P(Z \geq 1{,}85) = 0{,}0322;$

- $P(Z \geq -2{,}46) = 1 - P(Z \leq -2{,}46) = 1 - P(X \geq 2{,}46) = 0{,}9931;$

- $P(0{,}71 \leq Z \leq 1{,}93) = 1 - P(Z \geq 1{,}93) - P(Z \leq 0{,}71) = 1 - P(Z \geq 1{,}93) - \big(1 - P(Z \geq 0{,}71)\big) = P(Z \geq 0{,}71) - P(Z \geq 1{,}93) = 0{,}2389 - 0{,}0268 = 0{,}2121;$

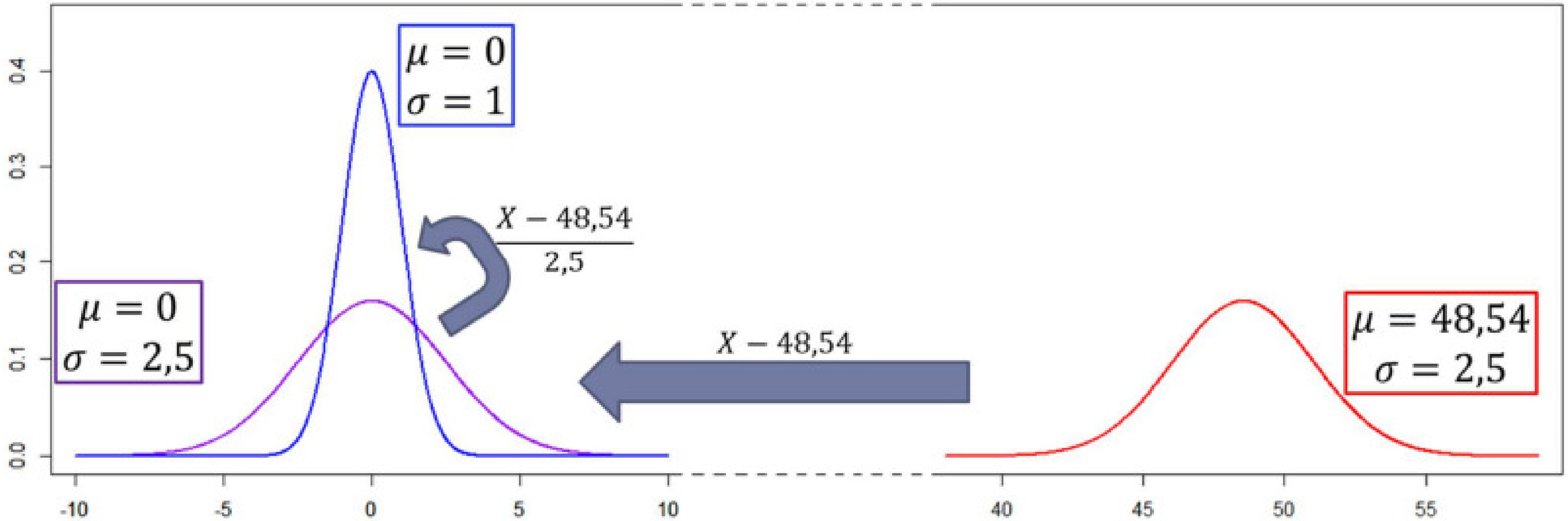
Exemplo

- Voltemos para o exemplo dado, em que se pretende estudar o comprimento de recém nascidos ($\mu = 48{,}54$cm e $\sigma = 2{,}5$cm).



Exemplo

- Se subtrairmos 48,54cm de todas as observações teremos uma distribuição normal com média 0cm e desvio padrão 2,5cm.
- Se, após subtrairmos 48,54cm, dividirmos todas as observações por 2,5cm teremos uma distribuição normal com média 0cm e desvio padrão 1cm.



Distribuição Normal Padronizada

- Uma variável aleatória X que siga um distribuição normal com média $\mu \neq 0$ ou desvio padrão $\sigma \neq 1$ pode ser padronizada pela seguinte expressão:
- $Z = \frac{X-\mu}{\sigma}$
- Conhecendo a relação entre uma variável aleatória X seguindo uma distribuição normal diferente da padrão e a variável aleatória Z que segue uma distribuição normal padrão, é possível calcular as probabilidades relativas à variável X utilizando a tabela de probabilidades de Z.

Exemplo

- Tomando X a variável aleatória relativa ao comprimento de recém nascidos ($\mu =$ 48,54cm e $\sigma =$ 2,5cm).
- $\mathrm{P}(X \geq 48{,}54) =?$
  - Primeiro Passo: Relacionar X com Z
    - Para $X = 48{,}54 \Rightarrow Z = \frac{X-\mu}{\sigma} = \frac{48{,}54-48{,}54}{2{,}5} = 0$
  - $\mathrm{P}(X \geq 48{,}54) = \mathrm{P}(Z \geq 0) = 0{,}5$
- $\mathrm{P}(X \leq 44{,}79) =?$
    - Para $X = 44{,}79 \Rightarrow Z = \frac{X-\mu}{\sigma} = \frac{44{,}79-48{,}54}{2{,}5} = \frac{-3{,}75}{2{,}5} = -1{,}5$
  - $\mathrm{P}(X \leq 44{,}79) = \mathrm{P}(Z \leq -1{,}5) = \mathrm{P}(Z \geq 1{,}5) = 0{,}0668$
- $\mathrm{P}(45{,}965 \leq X \leq 50{,}465) =?$
    - Para $X = 45{,}965 \Rightarrow Z = \frac{45{,}965-48{,}54}{2{,}5} = \frac{-2{,}575}{2{,}5} = -1{,}03$
    - Para $X = 50{,}465 \Longrightarrow Z = \frac{50{,}465-48{,}54}{2{,}5} = \frac{1{,}925}{2{,}5} = 0{,}77$
  - $\mathrm{P}(45{,}965 \leq X \leq 50{,}465) = \mathrm{P}(-1{,}03 \leq Z \leq 0{,}77) = 1 - \mathrm{P}(Z \leq -1{,}03) - \mathrm{P}(Z \geq 0{,}77) = 1 - 0{,}1515 - 0{,}2206 = 0{,}6279$

Exercício 3

- Doentes, sofrendo de certa moléstia, são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal, de média 15 e desvio padrão 2 (em dias).
- Seja X o *tempo de cura*, calcule:
  - Que proporção desses pacientes demora mais de 17 dias para se recuperar?
  - Qual o tempo máximo necessário para a recuperação de 25% dos pacientes?
  - Considere que 100 pacientes são escolhidos ao acaso, qual seria o número esperado de doentes curados em menos de 11 dias?

Propriedade da Distribuição Normal

- Tem-se que qualquer combinação linear de variáveis Normais independentes, também terá uma distribuição Normal;
- Ou seja, se $X_1, X_2, \ldots, X_n$ formam uma sequência de variáveis aleatórias $N(\mu_i; \sigma_i{}^2)$ independentes e $a_1, a_2, \ldots, a_n$, são constantes quaisquer, então:

$$W = \sum_{i=1}^{n} a_i X_i$$

- terá distribuição normal com parâmetros:
  - $\mu_W = \sum_{i=1}^{n} a_i \mu_i;$
  - $\sigma_w{}^2 = \sum_{i=1}^{n} a_i{}^2 \sigma_i{}^2.$

Exercício 4

- 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):
- $L = 2L_A + 5L_I + 3L_C$
- Com L_A, L_I e L_C representando, respectivamente, os lucros diários nos setores de agricultura, indústria e comércio.
- As distribuições de probabilidade dessas variáveis aleatórias são: $L_A \sim N(3{,}4), L_I \sim N(6{,}9)$ e $L_C \sim N(4{,}16)$.
- Supondo independência entre os três setores, qual será a probabilidade de um lucro diário acima de 50 mil?