MAE116 – Noções de Estatística
Grupo A – 1º semestre de 2020
Aula de revisão I

Medidas-resumo

Medidas de posição

- Média: soma das observações dividida pelo número delas.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- **Moda**: valor mais *frequente*.

- **Mediana**: é a realização que ocupa a *posição central* da série de observações, quando estão ordenadas em ordem crescente.

Medidas de dispersão

- **Variância**:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$
$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n \times \overline{x}^2\right).$$

- **Desvio padrão**: é definido como a raiz quadrada da variância, isto é, $s = \sqrt{s^2}$

- **Coeficiente de variação**: exprime a *variabilidade em relação à média*.

$$CV = \frac{s}{\overline{x}} \times 100\%$$

Exercício 1

A Maratona de Boston, realizada nos Estados Unidos, é a segunda mais antiga das maratonas, atrás apenas da maratona olímpica de Atenas. No entanto, as mulheres só começaram a correr oficialmente a partir de 1972. Os tempos (em minutos, arredondado à primeira decimal) para a mulher vencedora de 2000 a 2018 aparecem na seguinte tabela. Em 2014, a etíope Buzunesh Deba estabeleceu o novo recorde para mulheres, 2 horas 19 minutos e 59 segundos.

Ano	*Tempo*	*Ano*	*Tempo*	*Ano*	*Tempo*
2000	*146,2*	*2006*	*143,6*	*2012*	*151,8*
2001	*143,9*	*2007*	*149,3*	*2013*	*146,4*
2002	*140,7*	*2008*	*145,4*	*2014*	*140,0*
2003	*145,3*	*2009*	*152,3*	*2015*	*144,9*
2004	*144,5*	*2010*	*146,2*	*2016*	*149,3*
2005	*145,2*	*2011*	*142,6*	*2017*	*141,9*
				2018	*159,9*

a) *Calcule o tempo médio registrado pelas mulheres vencedoras na Maratona de Boston no período de 2000 a 2018.*

Vamos denotar por $x_1, x_2, \ldots, x_{19}$ os valores dos tempos (em minutos) das 19 mulheres vencedoras na Maratona de Boston no período de 2000 e 2018. Assim, por exemplo, $x_1 = 146,2$ é o valor do tempo da mulher vencedora no ano 2000 e $x_{19} = 159,9$ é o valor do tempo da mulher vencedora no ano 2018.
Assim, o tempo médio é dado por

$$
\begin{aligned}
\overline{x} &= \frac{1}{19}\sum_{i=1}^{30} x_i = \frac{x_1 + x_2 + \cdots + x_{19}}{19} \\
&= \frac{146,2 + 143,9 + \cdots + 159,9}{19} = \frac{2.779,4}{19} \cong 146,28.
\end{aligned}
$$

Portanto, em média, as mulheres vencedoras na Maratona de Boston no período de 2000 a 2018 gastam 146,28 minutos para concluir a prova.

b) *Calcule o coeficiente de variação dos períodos 2000–2009 e 2010–2018. Interprete.*

Vamos considerar os tempos nos períodos 2000–2009 e 2010–2018 e denotar por $\overline{x}_1$ e $\overline{x}_2$ as respectivas médias. Analogamente, denotamos por s_1 e s_2 os respectivos desvios padrão e por CV_1 e CV_2 os respectivos coeficientes de variação. Assim, temos que

-
$$
\begin{aligned}
\overline{x}_1 &= \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{x_1 + x_2 + \cdots + x_{10}}{10} \\
&= \frac{146,2 + 143,9 + \cdots + 152,3}{10} = \frac{1.456,4}{10} = 145,64.
\end{aligned}
$$

-
$$
\begin{aligned}
\overline{x}_2 &= \frac{1}{9}\sum_{i=11}^{19} x_i = \frac{x_{11} + x_{12} + \cdots + x_{19}}{9} \\
&= \frac{146,2 + 146,6 + \cdots + 159,9}{9} = \frac{1.323}{9} = 147.
\end{aligned}
$$

Vamos obter agora o desvio padrão para o período 2000–2009.

- Para obter s_1, temos que

$$
\begin{aligned}
\sum_{i=1}^{10} x_i^2 &= x_1^2 + x_2^2 + \cdots + x_{10}^2 \\
&= 146,2^2 + 143,9^2 + \cdots + 152,3^2 = 212.201,4.
\end{aligned}
$$

De modo que

$$
\begin{aligned}
s_1^2 &= \frac{1}{10-1}\left(\sum_{i=1}^{10} x_i^2 - 10 \times \overline{x}_1^2\right) \\
&= \frac{1}{9}\left[212.201,4 - 10 \times 145,64^2\right] = \frac{91,304}{9} \cong 10,1449.
\end{aligned}
$$

Logo,

$$
s_1 = \sqrt{10,1449} \cong 3,1851.
$$

- Para obter s_2, temos que

$$\sum_{i=11}^{19} x_i^2 = x_{11}^2 + x_{12}^2 + \cdots + x_{19}^2$$
$$= 146,2^2 + 146,6^2 + \cdots + 159,9^2 = 194.775,5.$$

Assim,

$$s_2^2 = \frac{1}{9-1}\left(\sum_{i=11}^{19} x_i^2 - 9 \times \overline{x}_2^2\right)$$
$$= \frac{1}{8}\left[194.775,5 - 9 \times 147^2\right] = \frac{294,5}{8} \cong 36,8125.$$

Portanto,

$$s_2 = \sqrt{36,8125} \cong 6,0673.$$

Dessa forma, temos que

- $CV_1 = \dfrac{s_1}{\overline{x}_1} \times 100\% = \dfrac{3,1851}{145,64} \times 100\% = 2,19\%.$
- $CV_2 = \dfrac{s_2}{\overline{x}_2} \times 100\% = \dfrac{6,0673}{147} \times 100\% = 4,13\%.$

Assim, considerando a dispersão relativa à média, medida pelo coeficiente de variação (CV), nota-se que os tempos registrados das ganhadoras na maratona no período 2010–2018 são mais dispersos do que os tempos registrados no período 2000–2009.

c) *Construa o boxplot dos tempos. A americana Desi Linden foi a vencedora em 2018, o que podemos dizer de seu tempo?*

Para construirmos o boxplot, precisaremos relembrar alguns conceitos.

Um pouco mais sobre medidas-resumo

Considere o conjunto de n observações *ordenadas* em *ordem crescente*.

- **Mínimo**: menor valor do conjunto de valores.
- **Máximo**: maior valor do conjunto de valores.
- **Percentil de ordem** $p \times 100$ $(0 < p < 1)$: é o valor que ocupa a posição $p \times (n+1)$ do conjunto de dados ordenado.
  - Percentil 25 : 1º Quartil (Q_1); Posição $0,25 \times (n+1)$;
  - Percentil 50 : Mediana ou 2º Quartil (Md); Posição $0,5 \times (n+1)$;
  - Percentil 75 : 3º Quartil (Q_3); Posição $0,75 \times (n+1)$.
- **Distância interquartil**: denotada por d_q, é a diferença entre o terceiro e primeiro quartis, isto é,

$$d_q = Q_3 - Q_1.$$

Quando a posição do percentil procurado resulta em número não inteiro, tomamos a média entre os dois vizinhos mais próximos. Por exemplo, se o conjunto de dados ordenados for

$$1, 3, 5, 7, 9, 10, 12, 13$$

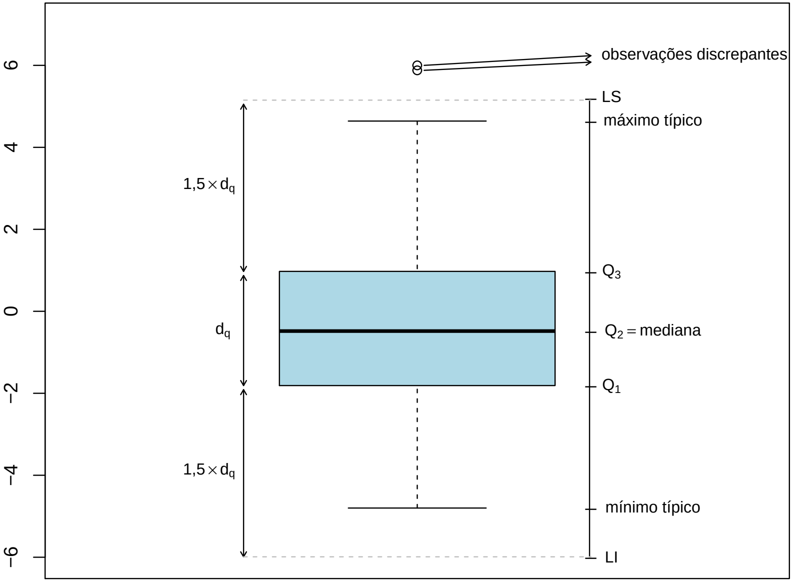
temos: posição da mediana $= \dfrac{8+1}{2} = 4,5.$

A mediana será $\dfrac{7+9}{2} = 8.$

O boxplot utiliza essas medidas na sua construção, conforme figura a seguir. Além dessas medidas, precisamos encontrar o LS, LI e o máximo e o mínimo típicos.

- $LS = Q_3 + 1,5 \times d_q$ e $LI = Q_1 - 1,5 \times d_q$ são os limites superior e inferior, respectivamente.
- Máximo típico: maior valor menor que o LS.
- Mínimo típico: menor valor maior que o LI.

Veja a figura a seguir.


observações discrepantes
LS
máximo típico
1,5 × d_q
Q3
d_q
Q2 = mediana
Q1
1,5 × d_q
mínimo típico
LI
6
4
2
0
−2
−4
−6

- Os dois pontos acima do LS são chamadas de *observações discrepantes*. Qualquer observação que ficar acima do LS ou abaixo do LI são chamadas dessa forma.

Voltando para os dados em estudo, a tabela a seguir apresenta as observações da variável Tempo ordenadas em ordem crescente.

Ano	2014	2002	2017	2011	2006	2001	2004	2015	2005
Tempo	140,0	140,7	141,9	142,6	143,6	143,9	144,5	144,9	145,2

2003	2008	2000	2010	2013	2007	2016	2012	2009	2018
145,3	145,4	146,2	146,2	146,4	149,3	149,3	151,8	152,3	159,9

Dessa forma,

- A posição do $Q_1 = 0,25 \times (19+1) = 5 \Rightarrow Q_1 = 143,6$.
- A posição do $Q_2 = 0,5 \times (19+1) = 10 \Rightarrow Q_2 = 145,3$.
- A posição do $Q_3 = 0,75 \times (19+1) = 15 \Rightarrow Q_3 = 149,3$.

Assim,

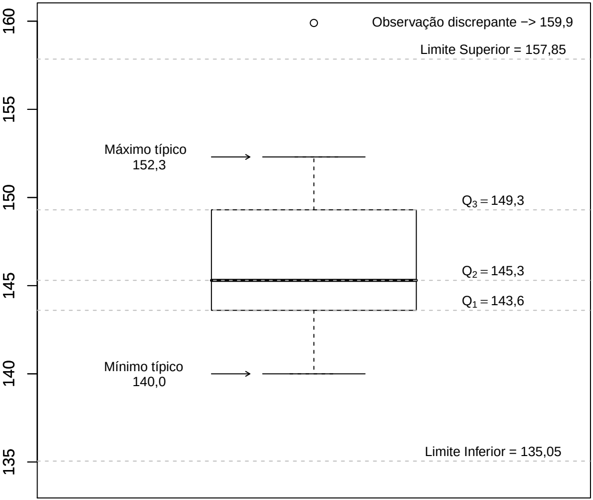
-

$$
\begin{aligned}
LI &= Q_1 - 1,5 \times d_q = Q_1 - 1,5 \times (Q_3 - Q_1) \\
&= 143,6 - 1,5 \times (149,3 - 143,6) = 135,05.
\end{aligned}
$$

-

$$
\begin{aligned}
LS &= Q_3 + 1,5 \times d_q = Q_3 + 1,5 \times (Q_3 - Q_1) \\
&= 149,3 + 1,5 \times (149,3 - 143,6) = 157,85.
\end{aligned}
$$

Dessa forma, o boxplot para a variável tempo é apresentado na figura a seguir.



160
155
150
145
140
135
Observação discrepante –> 159,9
Limite Superior = 157,85
Máximo típico
152,3
$Q_3 = 149{,}3$
$Q_2 = 145{,}3$
$Q_1 = 143{,}6$
Mínimo típico
140,0
Limite Inferior = 135,05

Finalmente, o boxplot mostra que o tempo registrado pela Americana Desi Linden na maratona de 2018, 159,9 min, pode ser considerado como ponto discrepante.

Exercício 2

Num estudo realizado na Universidade de Califórnia em Los Angeles registraram-se a idade, em meses, em que uma criança falou sua primeira palavra e o resultado de um teste de aptidão (Gesell Adaptive Score) aplicado muito depois. A seguir estão apresentados os valores para 21 crianças.

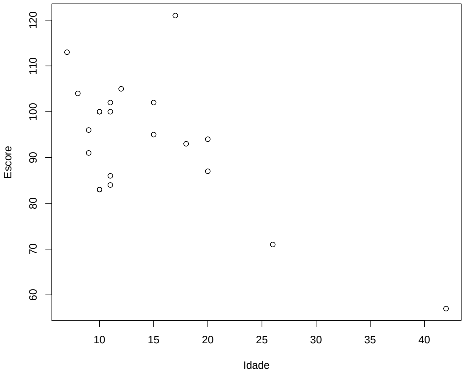
Criança	*Idade*	*Escore*	*Criança*	*Idade*	*Escore*	*Criança*	*Idade*	*Escore*
1	15	95	8	11	100	15	11	102
2	26	71	9	8	104	16	10	100
3	10	83	10	20	94	17	12	105
4	9	91	11	7	113	18	42	57
5	15	102	12	9	96	19	17	121
6	20	87	13	10	83	20	11	86
7	18	93	14	11	84	21	10	100

Seja X a variável idade na primeira palavra e Y escore do teste Gesell Adaptive.

Use que $\sum_{i=1}^{21} X_i = 302$ e $\sum_{i=1}^{21} Y_i = 1967$; e $\sum_{i=1}^{21} X_i^2 = 5.606$ e $\sum_{i=1}^{21} Y_i^2 = 188.155$

a) *Construa o diagrama de dispersão do escore do teste Gesell Adaptive versus a idade na primeira palavra. Comente o gráfico.*

Figura: Diagrama de dispersão entre escore e idade.



Segundo a Figura acima, o diagrama de dispersão sugere a existência de relação linear decrescente entre o escore do teste Gesell Adaptive e a idade na primeira palavra. A medida que a idade (em que criança falou sua primeira palavra) aumenta o resultado do teste tende a diminuir.

b) *Calcule o coeficiente de correlação linear de Pearson e interprete.*

O coeficiente de correlação de Pearson

É uma medida que avalia o quanto a "nuvem de pontos" no diagrama de dispersão aproxima-se de uma reta. Se temos um conjunto de n observações de duas variáveis X e Y, então, esse coeficiente é calculado por

$$r = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{yx}}{(n-1) s_y s_x}.$$

- O coeficiente de correlação está entre -1 e 1, isto é

$$-1 \leq r \leq 1.$$

- $r = -1$ indica associação linear *negativa e perfeita*;
- $r = 1$ indica associação linear *positiva e perfeita*;
- $r = 0$ indica *ausência* de associação linear.

Assim, temos que calcular algumas quantidades. São elas:

-
$$\overline{y} = \frac{1}{21}\sum_{i=1}^{21} y_i = \frac{1967}{21} = 93,6667.$$

-
$$\overline{x} = \frac{1}{21}\sum_{i=1}^{21} x_i = \frac{302}{21} = 14,3809.$$

-
$$\begin{aligned} s_y^2 &= \frac{1}{21-1}\left(\sum_{i=1}^{21} y_i^2 - 21 \times \overline{y}^2\right) \\ &= \frac{1}{20}\left[188155 - 21 \times 93,6667^2\right] \\ &= \frac{3912,5355}{20} = 195,6268. \end{aligned}$$

Portanto, o desvio padrão é $s_y = \sqrt{195,6268} = 13,9867.$

-

$$
\begin{aligned}
s_x^2 &= \frac{1}{21-1}\left(\sum_{i=1}^{21} x_i^2 - 21 \times \overline{x}^2\right) \\
&= \frac{1}{20}\left[5606 - 21 \times 14,3809^2\right] \\
&= \frac{1262,9840}{20} = 63,1492,
\end{aligned}
$$

resultanto o desvio padrão $s_x = \sqrt{63,1492} = 7,9465$.

-

$$
\begin{aligned}
\sum_{i=1}^{21} x_i y_i &= x_1 y_1 + x_2 y_2 + \cdots + x_{21} y_{21} \\
&= 15 \times 95 + 26 \times 71 + \cdots + 10 \times 100 \\
&= 26864.
\end{aligned}
$$

Assim,

$$r = \frac{\displaystyle\sum_{i=1}^{21} y_i x_i - 21 \times \overline{yx}}{(21-1) s_y s_x}$$
$$= \frac{26864 - 21 \times 93,6667 \times 14,3809}{20 \times 13,9867 \times 7,9465} = \frac{-1423,2404}{2222,9062} = -0,6403.$$

Note que observamos uma correlação linear negativa entre as variáveis escore e idade, o que já era esperado, com base no diagrama de dispersão do item a).

c) *Obtenha a reta de regressão ajustada $\hat{Y} = a + bX$. Interprete o valor do coeficiente b. Considerando a reta ajustada encontre o escore médio do teste Gesell Adaptive para crianças que falaram sua primeira palavra aos 2 anos.*

A reta de regressão

A reta ajustada da forma $\hat{Y} = a + bX$ é chamada reta de regressão ajustada.

- a é o intercepto da reta e
- b é o coeficiente angular ou a inclinação.

Se temos um conjunto de n observações, esses coeficientes podem ser calculados da seguinte maneira

$$b = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\overline{x}}{(n-1)s_x^2} \quad \text{e} \quad a = \overline{y} - b \times \overline{x}$$

Note que o numerador de b é igual ao numerador de r (coeficiente de correlação linear de Pearson).

Inicalmente, encontremos os valores de b e a.

$$b = \frac{\displaystyle\sum_{i=1}^{21} y_i x_i - 21 \times \overline{yx}}{(21-1)s_x^2}$$
$$= \frac{-1.423,2404}{20 \times 63,1492} = \frac{-1.423,2404}{1.262,984} = -1,1269,$$

e $a = \overline{y} - (-1,1269) \times \overline{x} = 93,6667 + 1,1269 \times 14,3809 = 109,8725.$

Assim, a reta ajustada é dada por

$$\hat{Y} = 109,8725 - 1,1269X.$$

Note que, como $b = -1,1269$, estima-se que, para um aumento de 1 mês na idade na qual a criança falou sua primeira palavra (X), o escore do teste (Y) diminui, **em média**, 1,1269 pontos.
Finalmente, temos que o escore médio do teste Gesell Adaptive para crianças que falaram sua primeira palavra aos 2 anos (24 meses) é

$$\hat{Y} = 109,8725 - 1,1269 \times 24 \cong 83 \text{ pontos.}$$

Probabilidade

- Experimento aleatório: fenômeno que, quando repetido nas mesmas condições, pode levar a resultados diferentes.
- Espaço amostral (Ω): conjunto de todos os resultados possíveis de um experimento aleatório.
- Eventos: subconjuntos do espaço amostral.

Composição de eventos

Sejam A e B dois eventos de um espaço amostral.

- $A \cup B$ representa a *união* dos eventos A e B - ocorrência de pelo menos um dos eventos, A **ou** B.
- $A \cap B$ representa a *interseção* dos eventos A e B - ocorrência simultânea dos eventos, A **e** B.

Calculando probabilidades

Considere o espaço amostral $\Omega = \{\omega_1, \omega_2, \ldots\}$ e um evento $A \subset \Omega$. A probabilidade do evento A é

$$\mathbb{P}(A) = \sum_{\omega_j \in A} \mathbb{P}(\omega_j).$$

Se as probabilidades de todos os resultados são iguais, temos

$$\mathbb{P}(A) = \frac{\text{número de elementos de A}}{\text{número de elementos de } \Omega}.$$

- Evento complementar: A^c é o evento complementar de A; isto é, o evento em Ω em que A não ocorre.

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

- Regra da adição:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Probabilidade condicional e independência

A probabilidade condicional de A dado que B ocorreu é

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{se } \mathbb{P}(B) > 0.$$

- A e B são independentes se:

$$\mathbb{P}(A \mid B) = \mathbb{P}(A), \quad \mathbb{P}(B) > 0.$$

- Regra do produto de probabilidades:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) \times \mathbb{P}(A \mid B),$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B \mid A).$$

- Se A e B são independentes:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

Exercício 3

Uma empresa que oferece cursos de preparação para uma prova de proficiência de língua estrangeira sabe que 2/3 de seus clientes têm diploma universitário e que 1/3 não têm este título. Após completar o curso, a proporção de clientes sem diploma universitário que foram reprovados no teste de proficiência é 1/5 e 1/10 para aqueles com título universitário. Qual é a probabilidade de que

a) *Um cliente escolhido ao acaso seja reprovado no teste de proficiência.*

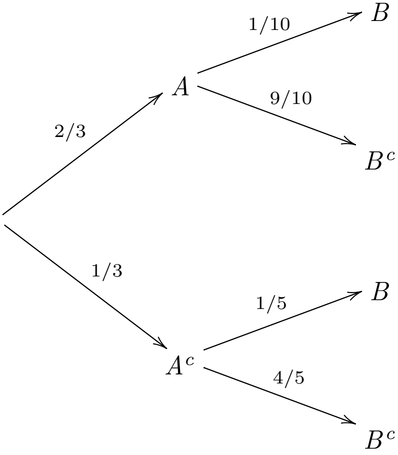
Vamos considerar os seguintes eventos.

- A : "O cliente, selecionado ao acaso da empresa, tem diploma universitário"
- B : "O cliente, selecionado ao acaso da empresa, é reprovado no teste de proficiência"

Pelo enunciado, temos as seguintes probabilidades

$$\mathbb{P}(A) = \frac{2}{3}, \quad \mathbb{P}(A^c) = 1 - \frac{2}{3} = \frac{1}{3},$$

$$\mathbb{P}(B \mid A) = \frac{1}{10} \quad \text{e} \quad \mathbb{P}(B \mid A^c) = \frac{1}{5},$$



$$\mathbb{P}(A \cap B) = \frac{2}{3} \times \frac{1}{10} = \frac{1}{15}$$

$$\mathbb{P}(A \cap B^c) = \frac{2}{3} \times \frac{9}{10} = \frac{3}{5}$$

$$\mathbb{P}(A^c \cap B) = \frac{1}{3} \times \frac{1}{5} = \frac{1}{15}$$

$$\mathbb{P}(A^c \cap B^c) = \frac{1}{3} \times \frac{4}{5} = \frac{4}{15}$$

Figura: Diagrama de árvore.

Relembre que os últimos “galhos” da árvore são probabilidades condicionais. Assim, por exemplo,

$$\frac{9}{10} = \mathbb{P}(B^c \mid A) = 1 - \mathbb{P}(B \mid A) = 1 - \frac{1}{10}.$$

Finalmente, estamos interessados na probabilidade de que um cliente escolhido ao acaso seja reprovado no teste de proficiência, isto é, $\mathbb{P}(B)$. Note que temos

$$\begin{aligned}
\mathbb{P}(B) &= \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) \\
&= \frac{2}{3} \times \frac{1}{10} + \frac{1}{3} \times \frac{1}{5} \\
&= \frac{1}{15} + \frac{1}{15} = \frac{2}{15} \cong 0,13.
\end{aligned}$$

b) *Um cliente escolhido ao acaso tenha diploma universitário, sabendo-se que ele foi reprovado no teste.*

Estamos interessados em $\mathbb{P}(A \mid B)$. Assim, temos

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/15}{2/15} = \frac{1}{2} = 0,5.$$

Variável aleatória discreta

Uma variável aleatória é *discreta* quando o conjunto de valores possíveis que ela assume for *finito* ou *infinito enumerável*.

- A função de probabilidade atribui a cada valor x_i da v.a. discreta X sua probabilidade de ocorrência.

x	x_1	$\dots$	x_n
$\mathbb{P}(X = x)$	$\mathbb{P}(X = x_1)$	$\dots$	$\mathbb{P}(X = x_n)$

Uma função de probabilidade deve satisfazer

$$0 \leq \mathbb{P}(X = x_i) \leq 1, \forall i = 1, \dots, n$$

e

$$\sum_{i=1}^{n} \mathbb{P}(X = x_i) = 1.$$

- O valor médio ou esperança da distribuição de X é

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i).$$

- A variância da distribuição de X pode ser escrita como

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

- O desvio padrão definido como a raiz quadrada positiva da variância, isto é,

$$DP(X) = \sqrt{Var(X)}.$$

Modelos Probabilísticos Discretos

Distribuição de Bernoulli

- Experimentos que admitem apenas dois resultados: sucesso ou fracasso (por exemplo).
- A variável aleatória X, que assume apenas os valores 0 (fracasso) e 1 (sucesso), com função de probabilidade

x	1	0
$\mathbb{P}(X=x)$	p	$1-p$

é chamada variável aleatória de Bernoulli e denotamos $X \sim Bernoulli(p)$.

- $$\mathbb{E}(X) = p \quad \text{e} \quad Var(X) = p(1-p).$$

Distribuição Binomial

- Repetições independentes de um ensaio de Bernoulli com a mesma probabilidade de ocorrência de "sucesso".
- A v.a. X que representa o *número de sucessos* em n ensaios de Bernoulli *independentes* e com a *mesma probabilidade* p de sucesso, tem distribuição binomial com parâmetros n e p. Escrevemos $X \sim b(n;p)$
- $$\mathbb{E}(X) = np$$
  e
  $$Var(X) = np(1-p).$$

Exercício 4

No processo de empacotamento de uma empresa, que vende produtos em embalagens de 1 kg, considera-se como limite inferior o peso de 995 g, sendo que os pacotes devem ter peso superior a este limite para venda. No entanto, no processo de embalagem produz-se 4% de pacotes abaixo do limite. Nas inspeções costuma-se recolher uma amostra aleatória de 20 pacotes do produto finalizado e pesar cada um deles.

a) *Desta forma, qual é a probabilidade de que na amostra:*
   i. *apenas um pacote esteja abaixo do limite de peso?*
   ii. *no máximo dois pacotes estejam abaixo do limite de peso?*

Inicialmente, vamos considerar o experimento de selecionar *ao acaso* um pacote do produto finalizado e pesá-lo. Note que, se considerarmos como *sucesso* (S) o pacote selecionado estar com peso abaixo do limite, então

$$p = \mathbb{P}(S) = \frac{4}{100} = \frac{1}{25} = 0,04.$$

Note que este experimento é repetido independentemente **20** vezes, pois 20 pacotes são recolhidos. Além disso, a probabilidade do peso do pacote ser inferior ao limite é a mesma para todos os 20 pacotes.

Assim, temos um experimento do tipo *sucesso* (peso do pacote abaixo do limite) e *fracasso* (peso do pacote acima do limite) repetido 20 vezes de forma *independente* e com probabilidade de sucesso p *constante*.

Vamos definir a variável aleatória X como sendo o número de pacotes com peso abaixo do limite dentre os 20 pacotes selecionados aleatoriamente para a inspeção.

Dessa forma, X tem distribuição binomial com $n = 20$ e $p = \mathbb{P}(S) = 0,04$, isto é, $X \sim b(20; 0,04)$.

Vamos obter as probabilidades dessa distribuição através do `Rcmdr`, utilizando os seguintes comandos:

`Distribuições > Distribuições discretas > Distribuição binomial > Probabilidades da binomial...`

Em seguida, insira o número de experimentos de Bernoulli (n) e a

probabilidade de sucesso (p); neste caso, $n = 20$ e $p = 0,04$. A saída correspondente deve ser similar à da figura a seguir.

	Probability
0	4.420024e-01
1	3.683354e-01
2	1.457994e-01
3	3.644985e-02
4	6.454662e-03
5	8.606215e-04
6	8.964808e-05
7	7.470673e-06
8	5.058268e-07
9	2.810149e-08
10	1.287985e-09
11	4.878731e-11
12	1.524603e-12
13	3.909240e-14
14	8.144249e-16
15	1.357375e-17
16	1.767415e-19
17	1.732760e-21
18	1.203306e-23
19	5.277656e-26
20	1.099512e-28

No item i., estamos interessados em saber a probabilidade de apenas um pacote estar abaixo do limite de peso, isto é, $\mathbb{P}(X = 1)$, pela tabela, temos

$$\mathbb{P}(X = 1) = 0,3683354 \cong 0,3683.$$

No item ii., queremos saber qual é a probabilidade de que no máximo dois pacotes estejam abaixo do limite de peso, isto é, $\mathbb{P}(X \leq 2)$, assim

$$\begin{aligned} \mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &\cong 0,4420 + 0,3683 + 0,1458 = 0,9561. \end{aligned}$$

Uma outra forma de calcular $\mathbb{P}(X \leq 2)$ é pelo `Rcmdr`.

Pelo `Rcmdr`, utilizamos os seguintes comandos:
`Distribuições > Distribuições discretas > Distribuição binomial > Probabilidades das caudas da binomial...`

Em seguida, insira o número de experimentos de Bernoulli (n) e a probabilidade de sucesso (p) e o valor que você quer calcular a probabilidade; neste caso, $n = 20$ e $p = 0,04$. O valor a ser colocado para calcular a probabilidade é 2 e marque a opção `Cauda inferior`.
A saída deverá ser `0.9561372`.

b) *Qual é o número esperado de pacotes abaixo do limite? E qual é o desvio padrão?*

O número esperado de pacotes abaixo do limite é

$$\mathbb{E}(X) = n \times p = 20 \times 0,04 = 0,8.$$

O desvio padrão do número de pacotes abaixo do limite é

$$DP(X) = \sqrt{Var(X)} = \sqrt{np(1-p)} = \sqrt{20 \times 0,04 \times 0,96} \cong 0,8764.$$

Variável aleatória contínua

Uma variável aleatória é *contínua* quando o conjunto de valores possíveis que ela assume for *não enumerável*.

- Uma v.a. contínua é caracterizada por sua função densidade de probabilidade $f(x)$.

A distribuição normal

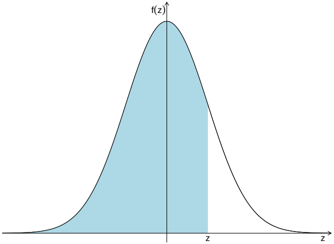
- Seja X uma variável aleatória com distribuição normal com parâmetros μ e σ^2, isto é, $X \sim N(\mu, \sigma^2)$. Então,
  - μ é o valor esperado (média) de X, com $-\infty < \mu < \infty$;
  - σ^2 é a variância de X, com $\sigma^2 > 0$.
- Se $X \sim N(\mu, \sigma^2)$, então $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.
- A variável aleatória $Z \sim N(0, 1)$ denomina-se *normal padrão* ou *reduzida*.
- $$\begin{aligned}\mathbb{P}(a < X < b) \\ &= \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right).\end{aligned}$$

Exemplos

A tabela da normal padrão que está sendo utilizada apresenta os valores para o cálculo de

$$A(z) = \mathbb{P}(Z \leq z),$$

para $z \geq 0$. Veja figura abaixo.



A seguir, apresentamos alguns exemplos de utilização da tabela.

- $\mathbb{P}(Z < 0,32) = A(0,32) = 0,6255.$
- $\mathbb{P}(0 < Z < 0,32) = A(0,32) - A(0) = 0,6255 - 0,5 = 0,1255.$
- $\mathbb{P}(-1,32 < Z < 0) = \mathbb{P}(0 < Z < 1,32) = A(1,32) - A(0) = 0,9066 - 0,5 = 0,4066.$
- $\mathbb{P}(Z < z) = 0,975$. Note que z é tal que $A(z) = 0,975$. Pela tabela, $z = 1,96$.
- $\mathbb{P}(-z \leq Z \leq z) = 0,80$. Temos $\mathbb{P}(Z > z) = 0,10$. Logo, z é tal que $\mathbb{P}(Z > z) = 0,10$ $\Rightarrow 1 - \mathbb{P}(Z < z) = 0,10 \Rightarrow A(z) = 0,9$. Assim, $z = 1,28$.

	Segunda parte decimal de z									
Parte inteira e primeira decimal de z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Exercício 5

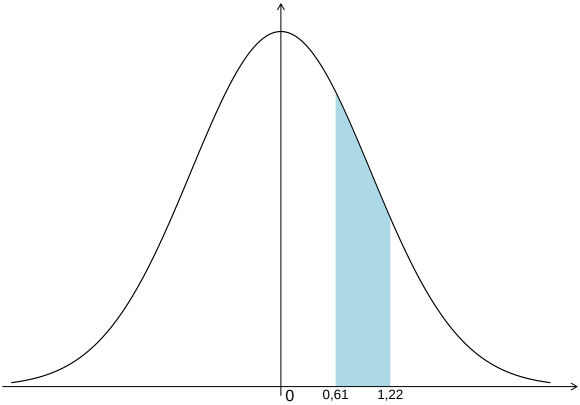
Um comerciante pretende classificar galinhas criadas numa granja de acordo com o peso do seguinte modo: 15% das mais leves como pequenas, os 50% seguintes como médias e os 35% seguintes como grandes. Sabe-se que a distribuição dos pesos de galinhas de granja pode ser representado por uma distribuição Normal, com média 4 kg e desvio padrão 0,82 kg.

a) *Sorteia-se uma galinha da produção. Qual é a probabilidade de que tenha peso entre 4,5 kg e 5 kg?*

Vamos definir a variável aleatória X como sendo o peso de uma galinha selecionada ao acaso nesta granja. Temos que $X \sim N(4; 0,82^2)$.

Estamos interessados em $\mathbb{P}(4,5 < X < 5)$.

$$\begin{aligned}
\mathbb{P}(4,5 < X < 5) &= \mathbb{P}\left(\frac{4,5-4}{0,82} < \frac{X-4}{0,82} < \frac{5-4}{0,82}\right) \\
&= \mathbb{P}\left(\frac{0,5}{0,82} < Z < \frac{1}{0,82}\right) \\
&= \mathbb{P}\left(0,61 < Z < 1,22\right)
\end{aligned}$$


0
0,61
1,22

Assim,

$$\begin{aligned}
\mathbb{P}(4,5 < X < 5) &= \mathbb{P}\left(0,61 < Z < 1,22\right) \\
&= \mathbb{P}(Z < 1,22) - \mathbb{P}(Z < 0,61) \\
&= A(1,22) - A(0,61) \\
&= 0,8888 - 0,7291 \\
&= 0,1597.
\end{aligned}$$

b) *Quais são os limites de peso para cada classificação?*

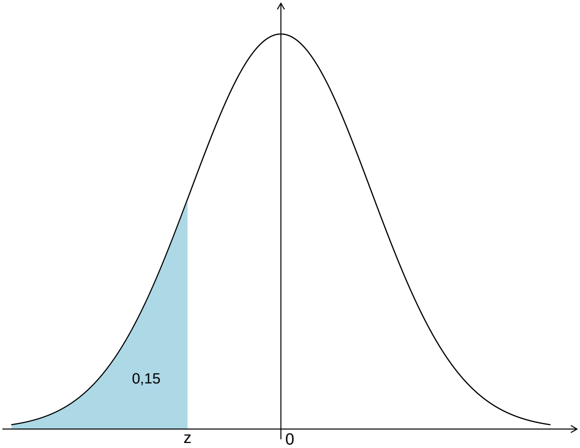
Nosso objetivo é encontrar os valores de x_1 que separa as primeiras 15% galinhas mais leves e o valor de x_2 que separa as 15%+50% = 65% mais leves; deixando as galinhas acima de x_2 como as grandes (35%).

$$\overset{15\%}{\text{———————}}\ x_1\ \overset{50\%}{\text{———————}}\ x_2\ \overset{35\%}{\text{———————}}$$

Assim, x_1 é tal que $\mathbb{P}(X < x_1) = 0,15$. Dessa forma, temos que

$$\begin{aligned}
\mathbb{P}(X < x_1) &= 0,15 \\
\Rightarrow \mathbb{P}\left(\frac{X-4}{0,82} < \frac{x_1-4}{0,82}\right) &= 0,15 \\
\Rightarrow \mathbb{P}\left(Z < \frac{x_1-4}{0,82}\right) &= 0,15.
\end{aligned}$$

Assim, precisamos encontrar, na tabela da normal padrão, qual é o valor que deixa uma probabilidade acumulada de 0,15. Isto é, qual o valor z tal que $\mathbb{P}(Z < z) = 0,15$. Este valor é $-1,04$.

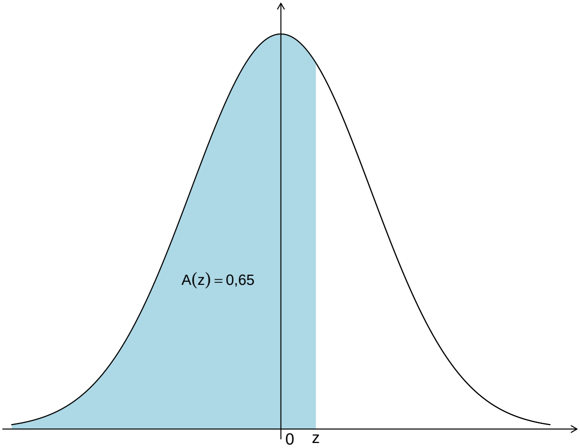

0,15
z
0

Assim, temos

$$\frac{x_1 - 4}{0,82} = -1,04 \Rightarrow x_1 = -1,04 \times 0,82 + 4 = 3,15.$$

Analogamente, precisamos encontrar o valor de x_2 tal que $\mathbb{P}(X < x_2) = 0,65$. Dessa forma, temos que

$$\begin{aligned} \mathbb{P}(X < x_2) &= 0,65 \\ \Rightarrow \mathbb{P}\left(\frac{X - 4}{0,82} < \frac{x_2 - 4}{0,82}\right) &= 0,65 \\ \Rightarrow \mathbb{P}\left(Z < \frac{x_2 - 4}{0,82}\right) &= 0,65. \end{aligned}$$


A(z) = 0,65
0
z

Pela tabela, esse valor é $0,39$.

$$\frac{x_2 - 4}{0,82} = 0,39 \Rightarrow x_2 = 0,39 \times 0,82 + 4 = 4,32.$$

Portanto, temos os seguintes limites dos pesos para cada classificação:

- $x \leq 3,15 \Rightarrow$ Galinhas leves;
- $x \in (3,15; 4,32) \Rightarrow$ Galinhas médias;
- $x \geq 4,32 \Rightarrow$ Galinhas pesadas.