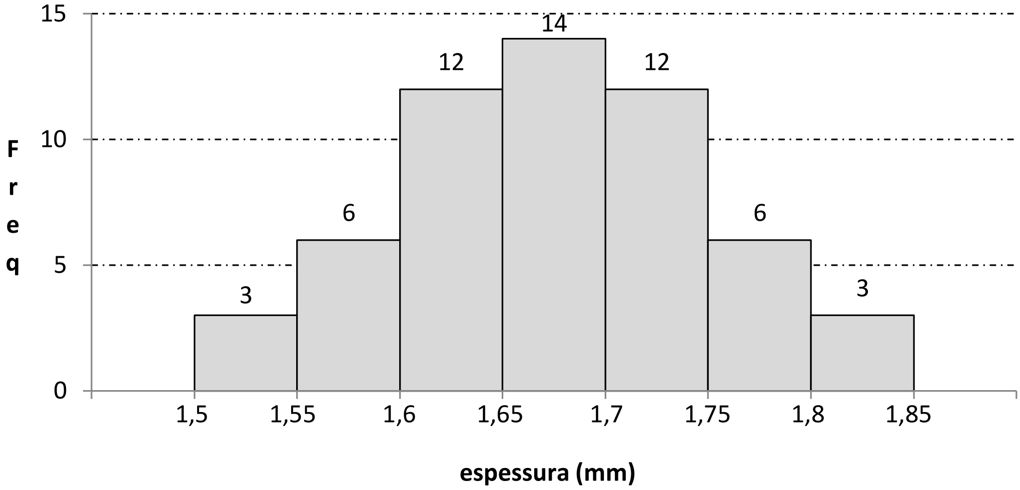
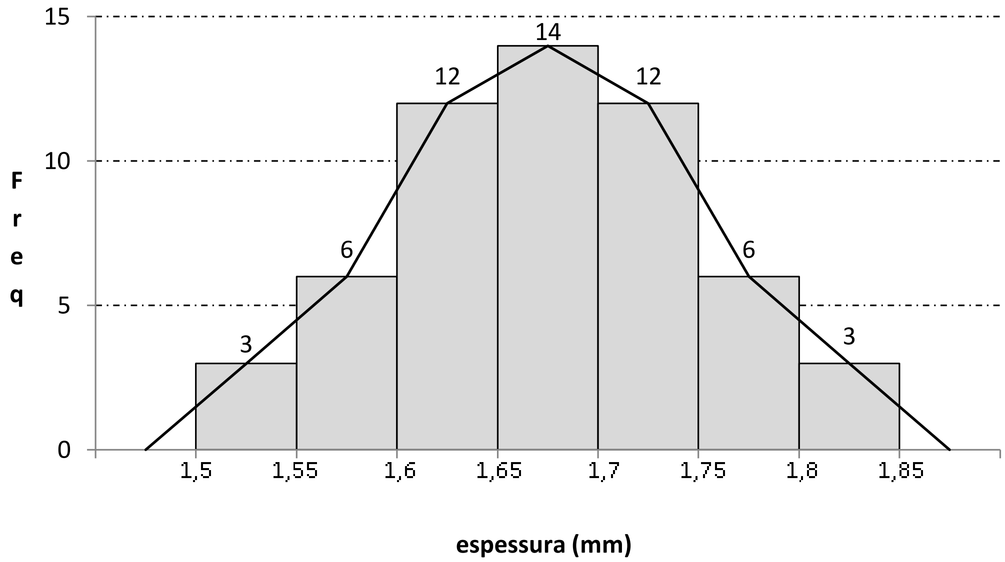
Variável Aleatória Contínua

Até o momento, nosso estudo focou a modelagem de experimentos aleatórios, em que as características de interesse (variáveis aleatórias) eram provenientes de contagem e, por esse motivo, foi denominada discreta. Nosso estudo agora se refere às variáveis aleatórias contínuas, ou seja, aquelas provenientes de contagem. Considere a distribuição das espessuras (mm) de placas de metal de determinada empresa, representada no histograma abaixo. Note que a variável contínua é a medida da espessura de tais placas.



Para modelarmos essa distribuição, podemos considerar os pontos médios das bases superiores dos retângulos e ligarmos os mesmos por uma curva. Teremos, então:

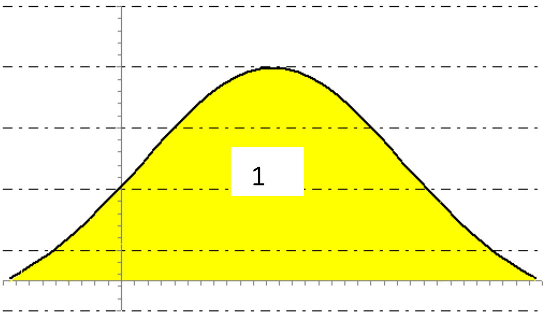


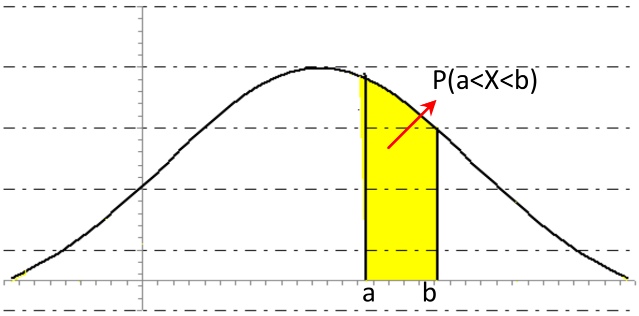
Essa curva é, portanto, uma função que representa aproximadamente o comportamento da produção das tais placas de metal em relação à espessura. A determinação da lei que define essa função não será tratada nesta disciplina. Entretanto, as características para que essa função permita que se calculem probabilidades, em relação à variável aleatória contínua (v.a.c.) X: espessura das tais placas de metal (mm) é o foco de nosso estudo.

A probabilidade **será definida como a medida da área sob a curva dessa função**. Assim, a medida da **área total** sob a curva deve ser **um**. Observe, ainda, que definindo probabilidade como medida da área sob a curva, ao calcular a **probabilidade de uma placa ter espessura igual a exatamente um valor é zero** (não tem área quando não se tem duas dimensões). Esse aspecto pode ser relacionado com o fato de não se conhecer o valor exato de qualquer medida. Essa função é denominada ***função densidade de probabilidades da v.a.c. X*** (f.d.p. de X).

Podemos, agora, apresentar a definição:

Dada uma v.a.c. X, dizemos que uma função $f: \mathbb{R} \to \mathbb{R}$ é **função densidade de probabilidades de X**, quando: $\boxed{\boldsymbol{f(x) \geq 0 \,, \forall\, x \in \mathbb{R}}}$ (ela não assume valores negativos) e, ainda, $\boxed{\int_{-\infty}^{+\infty} \boldsymbol{f(x)dx = 1}}$ (lembrando que podemos calcular a medida da área sob uma curva utilizando integral, essa condição significa que a medida da área total sob a curva deve ser igual a um). Assim, teremos que $\boxed{\boldsymbol{P(a} < X < b) = \int_{\boldsymbol{a}}^{\boldsymbol{b}} \boldsymbol{f(x)dx}}$.





A seguir apresentamos um exemplo de aplicação desse conceito.

Exemplo *Em um teste educacional com crianças, o tempo de uma bateria de questões de raciocínio verbal e lógico é medido e anotado para ser comparado com um modelo teórico. Este teste é utilizado para identificar o desenvolvimento das crianças e auxiliar a aplicação de medidas corretivas. O modelo teórico considera a variável aleatória T: tempo de teste em minutos como uma função* $f: \mathbb{R} \to \mathbb{R}$ *definida por:*

$$f(t) = \begin{cases} k(t-4) & se\ \ 8 \leq t < 10 \\ 6\,k & se\ \ 10 \leq t \leq 15 \\ 0 & outros\ valores \end{cases} \qquad k \in \mathbb{R}$$

a) *determine o valor de k para que a função do modelo teórico seja função densidade de probabilidades da v.a.c. T e esboce o seu gráfico;*
b) *calcule a probabilidade de que o tempo de teste seja de 9 a 12 minutos;*
c) *calcule a probabilidade de que o tempo de teste seja exatamente igual a 10 minutos;*
d) *calcule a probabilidade de que o tempo de teste seja de pelo menos 9minutos e meio.*

a) Inicialmente, dentre todas as funções definidas pelo modelo teórico (para cada valor real de k temos uma função), deveremos determinar aquela que possua as características necessárias para ser função densidade de probabilidades de uma v.a.c. Para tal, devemos ter: $\int_{-\infty}^{+\infty} \boldsymbol{f(t)dt = 1}$. Como a função, dependendo do intervalo da reta, é definida por uma lei diferente, para determinar essa integral, precisamos decompor na soma de integrais, conforme muda a lei que define a função. Teremos: $\int_{-\infty}^{+\infty} \boldsymbol{f(t)dt = 1} \Rightarrow \int_{-\infty}^{\boldsymbol{8}} \boldsymbol{0\ dt} + \int_{\boldsymbol{8}}^{\boldsymbol{10}} k(t-4)\ \boldsymbol{dt} + \int_{\boldsymbol{10}}^{\boldsymbol{15}} 6\ k\ \boldsymbol{dt} + \int_{\boldsymbol{15}}^{+\infty} \boldsymbol{0\ dt = 1}$

Como a primeira e a última integrais são iguais a zero e a segunda integral pode ser decomposta na soma de duas outras, obtemos:

$$\int_{8}^{10} kt\ \boldsymbol{dt} + \int_{8}^{10} (-4k)\ \boldsymbol{dt} + \int_{10}^{15} 6\ k\ \boldsymbol{dt} = \boldsymbol{1}$$

Utilizando propriedade de integral, podemos colocar as constantes fora da integral, obtendo:

$$\boldsymbol{k}\int_{8}^{10} t\ \boldsymbol{dt} - \boldsymbol{4k}\int_{8}^{10} \boldsymbol{dt} + \boldsymbol{6k}\int_{10}^{15} \boldsymbol{dt = 1}$$

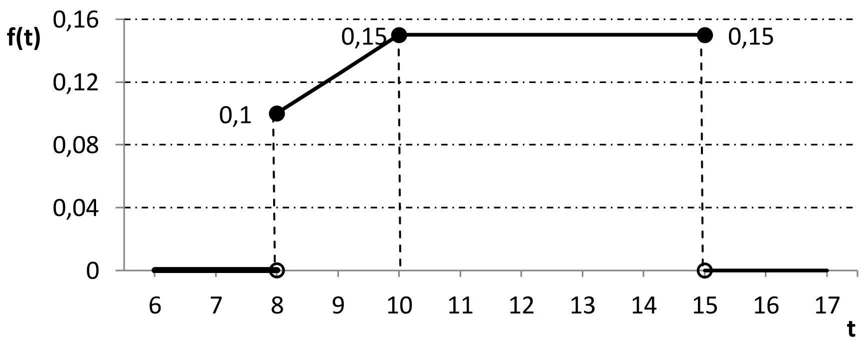
Resolvendo as integrais, temos:

$$k\left(\frac{t^2}{2}\right)_8^{10} - 4k(t)_8^{10} + 6k(t)_{10}^{15} = 1 \Rightarrow k\left(\frac{10^2-8^2}{2}\right) - 4k(10-8) + 6k(15-10) = 1$$

$$18k - 8k + 30k = 1 \Rightarrow 40k = 1 \Rightarrow k = \frac{1}{40} = 0{,}025$$

Logo, a f.d.p. da v.a.c. T é definida por: $f(t) = \begin{cases} 0{,}025t - 0{,}1 & se\ \ 8 \le t < 10 \\ 0{,}15 & se\ \ 10 \le t \le 15 \\ 0 & outros\ valores \end{cases}$

A representação gráfica dessa função é:



Observe que a medida da área sob a curva é um. Como nesse caso as representações são lineares, podemos constatar esse fato, calculando as áreas, valendo-se das figuras geométricas que elas representam. Assim, no intervalo $]-\infty\,,0]$ temos que a medida da área é igual a zero; no intervalo $[8\,,10[$ a medida da área do trapézio que é igual a $\frac{0{,}15+0{,}1}{2} \times 2 = 0{,}25$; no intervalo $[10\,,15]$ a medida da área do retângulo é igual a $5 \times 0{,}15 = 0{,}75$ e, finalmente, no intervalo

$]15,+\infty[$ temos que a medida da área é igual a zero. Somando-se as medidas das áreas, temos que a medida da área total sob a curva da f.d.p. é $0 + 0{,}25 + 0{,}75 + 0 = 1$.

b) $P(9 \leq T \leq 12) = \int_9^{12} f(t)dt = \int_9^{10}(0{,}025t - 0{,}1)dt + \int_{10}^{12} 0{,}15dt =$

$$= 0{,}025\int_9^{10} tdt - 0{,}1\int_9^{10} dt + 0{,}15\int_{10}^{12} dt = 0{,}2375 - 0{,}1 + 0{,}3 = 0{,}4375$$

Note que, embora tenhamos utilizado integral para o cálculo da medida da área, podemos utilizar a área das figuras geométricas correspondentes que chegaremos à mesma resposta. Valemo-nos do cálculo integral, por ser uma disciplina que você já estudou e porque nos casos em que a lei não é linear, ela é a única possibilidade.

c) $P(T = 10) = 0$, porque a probabilidade em um ponto, no caso das variáveis contínuas, é sempre zero. Note que podemos fazer também: $P(10 \leq T \leq 10) = \int_{10}^{10} f(t)dt = 0$.

d) $P(T \geq 9{,}5) = \int_{9,5}^{+\infty} f(t)dt = \int_{9,5}^{10}(0{,}025t - 0{,}1)dt + \int_{10}^{15} 0{,}15dt + \int_{15}^{+\infty} 0dt =$

$$= 0{,}025\int_{9,5}^{10} tdt - 0{,}1\int_{9,5}^{10} dt + 0{,}15\int_{10}^{15} dt = 0{,}121875 - 0{,}05 + 0{,}75 = 0{,}821875$$

Note que essa probabilidade pode ser calculada por meio da medida da área de figuras geométricas correspondentes, como também, valendo-se do complementar, isto é, basta calcularmos a medida da área total (1) menos a medida da área no intervalo de 8 a 9,5, que obteremos o mesmo resultado (por meio do cálculo integral, ou, por meio da medida da área da figura geométrica correspondente).

A seguir, propomos algumas situações para você resolver.

Exercício 1. *A quantia gasta anualmente, em milhões de reais, na manutenção do asfalto em uma cidade do interior é representada pela função:* $\boldsymbol{f(y)} = \begin{cases} \boldsymbol{k(2y-1), se\ 0{,}5 \leq y} < 2 \\ \boldsymbol{0, caso\ contr}\text{á}\boldsymbol{rio} \end{cases}$.

Determine:

(a) o valor de k para que a função seja f.d.p. e esboce o seu gráfico;

(b) a probabilidade da quantia gasta em um dado ano ser inferior a 0,8 milhões;

(c) a probabilidade da quantia gasta em um dado ano não ser inferior a 1,5 milhões.

Exercício 2. *Em certa região, fósseis de pequenos animais são frequentemente encontrados e um arqueólogo estabeleceu o seguinte modelo de probabilidades para o comprimento (em centímetros):* $\boldsymbol{f(x)} = \begin{cases} \boldsymbol{k\,x} \ , \ \boldsymbol{se} \ \boldsymbol{4 \leq x \leq 8} \\ -\boldsymbol{2\,k\,x + 0,6} \ , \ \boldsymbol{se} \ \boldsymbol{8} < x \leq 10 \\ \boldsymbol{4\,k} \ , \ \boldsymbol{se} \ \boldsymbol{10} < x \leq 11 \\ \boldsymbol{0} \ , \ \boldsymbol{outros\ valores} \end{cases}$

a) determine o valor de k para que a função f seja f.d.p. e esboce o seu gráfico;

b) para um fóssil, encontrado nessa região, determine a probabilidade de o comprimento do fóssil ser superior a 5cm, mas, inferior a 10,5cm.

Exercício 3. *O tempo de corrosão, em anos, de uma peça metálica pode ser considerado uma v.a.c. X com f.d.p. dada por:* $\boldsymbol{f(x)} = \begin{cases} \boldsymbol{a\,x} \ , \ \boldsymbol{se} \ \boldsymbol{0 \leq x \leq 1} \\ \boldsymbol{a} \ , \ \boldsymbol{se} \ \boldsymbol{1} < x \leq 2 \\ -\boldsymbol{ax + 3a} \ , \ \boldsymbol{se} \ \boldsymbol{2} < x \leq 3 \\ \boldsymbol{0} \ , \ \boldsymbol{outros\ valores} \end{cases}$ *Determine:*

a) o valor da constante a e esboce o gráfico de f(x);

b) a porcentagem de peças em que o tempo de corrosão seja maior que 2,3 anos;

c) entre 500 peças, o número de peças que se espera ter o tempo de corrosão inferior a 2 anos.

Depois de definirmos a função densidade de probabilidades que permite determinar as probabilidades para o caso das variáveis aleatórias contínuas, o próximo passo é a transposição do conceito de **Função distribuição de probabilidades acumulada**, ou, **função repartição**.

Função Distribuição Acumulada de Probabilidades (Função Repartição) se define como no caso das v.a.'s discretas, ou seja:

$$F : \Re \to [0, 1]$$
$$t \mapsto F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)dx \, .$$

Considere o exemplo em que a f.d.p. da v.a. contínua X é dada por: $f : \Re \to \Re$, tal que:

$$f(x) = \begin{cases} 0, \ se \ x < 0 \\ \dfrac{x}{10} + \dfrac{1}{10}, \ se \ 0 \leq x \leq 2 \\ -\dfrac{3x}{40} + \dfrac{9}{20}, \ se \ 2 < x \leq 6 \\ 0, \ se \ x > 6 \end{cases}$$

Assim, nesse exemplo, a Função Distribuição Acumulada de Probabilidades será: $F : \Re \to [0,1]$, tal que:

se $t < 0$, então, $F(t) = \int_{-\infty}^{t} 0\,dx = 0$

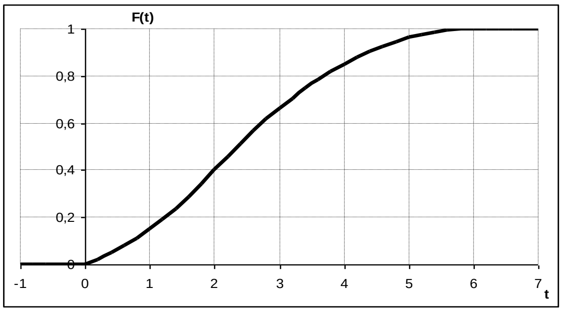
se $0 \leq t \leq 2$, então, $F(t) = \int_{-\infty}^{0} 0 \ dx + \int_{0}^{t} \left(\dfrac{x}{10} + \dfrac{1}{10} \right) dx = \dfrac{t^2}{20} + \dfrac{t}{10}$

se $2 < t \le 6$, então, $F(t) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{2}\left(\frac{x}{10}+\frac{1}{10}\right)dx + \int_{2}^{t}\left(-\frac{3x}{40}+\frac{9}{20}\right)dx = -\frac{3t^2}{80}+\frac{9t}{20}-\frac{7}{20}$

se $t > 6$, então, $F(t) = \int_{-\infty}^{0} 0\,dx + \int_{0}^{2}\left(\frac{x}{10}+\frac{1}{10}\right)dx + \int_{2}^{6}\left(-\frac{3x}{40}+\frac{9}{20}\right)dx + \int_{6}^{+\infty} 0\,dx = 1$

Temos, portanto:

$$F(t) = \begin{cases} 0, \; se \; t < 0 \\ \frac{t^2}{20}+\frac{t}{10}, \; se \; 0 \le t \le 2 \\ -\frac{3t^2}{80}+\frac{9t}{20}-\frac{7}{20}, \; se \; 2 < t \le 6 \\ 1, \; se \; t > 6 \end{cases}$$ ou



Essa função pode ser usada para calcular probabilidades, como fizemos no caso das variáveis aleatórias discretas. A diferença, neste caso das variáveis contínuas, é que a variável contínua tem probabilidade igual a zero, em qualquer ponto. Como consequência, a probabilidade de uma v.a.c. assumir valores em um intervalo real $[a\,,b]$ sempre será dada pela diferença $F(b) - F(a)$, mesmo que o intervalo seja aberto, ou aberto em um dos extremos do intervalo e fechado no outro. Por exemplo:

$$P(1 \le X \le 5) = F(5) - F(1) = \frac{65}{80} = 0{,}8125 = P(1 < X \le 5) = P(1 \le X < 5) = P(1 < X < 5)$$

$$P(X \le 2) = F(2) = 0{,}4 \qquad\qquad P(X < 2) = F(2) = 0{,}4$$

$$P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 0{,}15$$

Da mesma forma, podemos fazer a transposição do conceito de média (valor esperado, ou, esperança matemática), propriedades da média, variância, propriedades de variância e o conceito de desvio padrão. A priori, pode-se afirmar que os conceitos e propriedades são exatamente os mesmos, a menos da ferramenta utilizada para calculá-los.

Comparando o cálculo da média para os dois tipos de variáveis, temos que:

Se X é v.a.d.:

$E(X) = \sum_i x_i \, p_X(x_i)$

Se X é v.a.c.:

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

Observe que, nas variáveis contínuas, em lugar de somatória temos a integral; em lugar de valores x_i temos $x \in \mathbb{R}$ e em lugar de $p_X(x_i)$ temos $f(x)$, que fornece probabilidades, quando integrada.

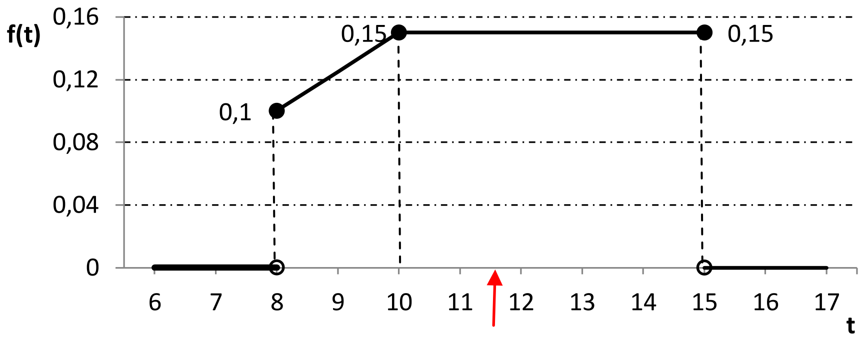
Por exemplo, no caso da situação, anteriormente apresentada, do teste educacional com crianças em que a variável aleatória contínua T: tempo do teste em minutos tem a seguinte função distribuição de probabilidades: $f(t) = \begin{cases} 0{,}025t - 0{,}1 & se \;\; 8 \le t < 10 \\ 0{,}15 & se \;\; 10 \le t \le 15 \\ 0 & outros \; valores \end{cases}$, calculemos $E(T)$.

$$E(T)=\int_{-\infty}^{+\infty} t\,f(t)dt=\int_{-\infty}^{8} t\,0dt+\int_{8}^{10} t\,(0{,}025t-0{,}1)dt+\int_{10}^{15} t\,0{,}15dt+\int_{15}^{+\infty} t\,0dt=$$

$$=0{,}025\int_{8}^{10} t^2dt-0{,}1\int_{8}^{10} t\,dt+0{,}15\int_{10}^{15} t\,dt==0{,}025\left(\frac{t^3}{3}\right)_8^{10}-0{,}1\left(\frac{t^2}{2}\right)_8^{10}+0{,}15\left(\frac{t^2}{2}\right)_{10}^{15}=\frac{1397}{120}\cong$$

$$\cong 11{,}641667$$

Assim, o tempo esperado do teste é de aproximadamente 11,641667 minutos.



Considerando que, de certa forma, integral é soma, as mesmas propriedades que estudamos para o caso das variáveis aleatórias discretas (de soma), continuam valendo para o caso de X ser uma v.a.c. (de integral). Sendo k uma constante real, X e Y v.a's.c.:

$E(k)=k$ $\qquad E(kX)=kE(X)$ $\qquad E(X+k)=E(X)+k$

$E(X+Y)=E(X)+E(Y)$ $\qquad E(X-Y)=E(X)-E(Y)$

Além de que: $E(X^k)=\int_{-\infty}^{+\infty} x^k\,f(x)dx$, para qualquer constante real k.

Lembrando que o cálculo da variância de uma variável aleatória só depende de se saber calcular valor esperado, temos que a variância é dada por:

$$Var(X)=E(X^2)-[E(X)]^2=\int_{-\infty}^{+\infty} x^2\,f(x)dx-\left[\int_{-\infty}^{+\infty} x\,f(x)dx\right]^2$$

Sendo k uma constante real e X uma v.a.c., valem as propriedades:

$Var(K)=0$ $\qquad Var(X+K)=Var(X)$ $\qquad Var(KX)=k^2Var(X)$

O desvio padrão, como já definido por: $DP(X)=\sqrt{Var(X)}$.

No exemplo do teste com crianças, em que a v.a.c. T: tempo dispensado pela criança para realizar o teste, em que a f.d.p. é dada por:

$f(t)=\begin{cases}0{,}025t-0{,}1 & se\ \ 8\le t<10\\ 0{,}15 & se\ \ 10\le t\le 15\\ 0 & outros\ valores\end{cases}$ e em relação à qual temos que $E(T)=\frac{1397}{120}\cong$ $11{,}641667$ e que $E(T^2)\cong 139{,}383333$, teremos que $Var(T)=E(T^2)-[E(T)]^2\cong 139{,}383333-(11{,}641667\,)^2\cong 3{,}854923$, ou seja, a variância do tempo dispensado pela criança para o teste é $3{,}854923min^2$. Dessa forma, $DP(X)=\sqrt{Var(X)}\cong\sqrt{3{,}854923}\cong 1{,}963396$, ou seja, o desvio padrão do tempo dispensado pela criança na realização do teste é aproximadamente $1{,}963396min$.

Com essa afirmação, com uma aproximação não precisa, pode-se dizer que aproximadamente 68% das crianças dispensam de 9,7 a 13,6 minutos na realização do tal teste (observe que essa porcentagem é aproximada e que o intervalo é: $[E(T) - DP(T)\,;\ E(T) + DP(T)]$).

A seguir, apresentamos algumas situações para aplicação desses conceitos.

Exercício 4. *A quantia gasta anualmente, em milhões de reais, na manutenção do asfalto em uma cidade do interior é representada pela função:* $\boldsymbol{f(y)} = \begin{cases} \frac{\mathbf{8}}{\mathbf{9}}\boldsymbol{y} - \frac{\mathbf{4}}{\mathbf{9}}, \boldsymbol{se}\ \mathbf{0,5} \leq \boldsymbol{y} < 2 \\ \mathbf{0}, \boldsymbol{caso\ contr}\text{á}\boldsymbol{rio} \end{cases}$*. Determine:*

(a) o valor esperado de Y; *(b) a variância de Y;* *(c) o desvio padrão de Y.*

(d) determine a função distribuição de probabilidades acumulada de Y (ou, repartição);

(e) a probabilidade da quantia gasta ser menor que um milhão de reais, usando a repartição;

(f) $\boldsymbol{P}(\mathbf{0,8} \leq \boldsymbol{Y} < 1{,}7)$*, usando a função acumulada (ou, repartição).*

Exercício 5. *Em certa região, fósseis de pequenos animais são frequentemente encontrados e um arqueólogo estabeleceu o seguinte modelo de probabilidades para o comprimento (em centímetros):* $\boldsymbol{f(x)} = \begin{cases} \mathbf{0,025}\,\boldsymbol{x}\ ,\ \boldsymbol{se}\ \mathbf{4} \leq \boldsymbol{x} \leq \mathbf{8} \\ -\,\mathbf{0,05}\,\boldsymbol{x} + \mathbf{0,6}\ ,\ \boldsymbol{se}\ \mathbf{8} < x \leq 10 \\ \mathbf{0,1}\ ,\ \boldsymbol{se}\ \mathbf{10} < x \leq 11 \\ \mathbf{0}\ ,\ \boldsymbol{outros\ valores} \end{cases}$ *a) Determine o comprimento médio e o desvio padrão dos fósseis, segundo o modelo;* *b) Se forem encontrados 20 fósseis, qual a probabilidade de que pelo menos três fósseis tenham comprimento inferior a 7cm?*

c) Determine a função distribuição de probabilidades acumulada de X.

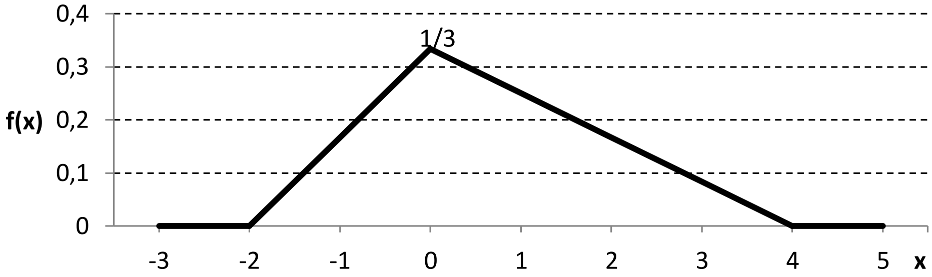
d) Calcule $\boldsymbol{P}(\mathbf{5} < X < 10{,}2)$*, usando a função acumulada.*

Exercício 6. *O tempo de corrosão, em anos, de uma peça metálica pode ser considerado uma v.a.c. X com f.d.p. dada por:* $\boldsymbol{f(x)} = \begin{cases} \mathbf{0,5}\,\boldsymbol{x}\ ,\ \boldsymbol{se}\ \mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1} \\ \mathbf{0,5}\ ,\ \boldsymbol{se}\ \mathbf{1} < x \leq 2 \\ -\mathbf{0,5}\boldsymbol{x} + \mathbf{1,5}\ ,\ \boldsymbol{se}\ \mathbf{2} < x \leq 3 \\ \mathbf{0}\ ,\ \boldsymbol{outros\ valores} \end{cases}$. *Determine:*

a) o tempo médio e o desvio padrão do tempo de corrosão de uma peça; *b) a porcentagem de peças em que o tempo de corrosão seja maior que o tempo médio mais 2,3 desvios padrão;*

c) a probabilidade do tempo de corrosão de a peça não diferir da média em mais de dois desvios padrão.

Exercício 7. *Considere que uma v.a.c. X tem a seguinte função densidade de probabilidades:*



Determine: (a) $\boldsymbol{E}(\boldsymbol{X})$; *(b)* $\boldsymbol{DP}(\boldsymbol{X})$; *(c)* $\boldsymbol{E}(-\mathbf{4}\boldsymbol{X}^{\mathbf{2}}+\mathbf{8}\boldsymbol{X}-\mathbf{9})$;

(d) $\boldsymbol{DP}(-\mathbf{9}\boldsymbol{X}+\mathbf{5})$; *(e)* $\boldsymbol{DP}(-\mathbf{2}\boldsymbol{X}^{\mathbf{2}}++\mathbf{5}\boldsymbol{X}+\mathbf{1})$; *(f) função repartição;*

(g) usando a função repartição, determine:

(i) $\boldsymbol{P}(\boldsymbol{X}>1)$ *(ii)* $\boldsymbol{P}(\boldsymbol{X}<2)$ *(iii)* $\boldsymbol{P}(-\mathbf{1}<X<3)$ *(iv)* $\boldsymbol{P}(\boldsymbol{X}>1 \mid X<2)$ (essa é condicional!)