

Conceitos Iniciais de Estatística –Módulo 6 :
PROBABILIDADE – VARIÁVEL ALEATÓRIA CONTÍNUA
Prof. Rogério Rodrigues



CONCEITOS INICIAIS DE ESTATÍSTICA:

► **PROBABILIDADE / VARIÁVEL ALEATÓRIA CONTÍNUA**

CURSO : ADMINISTRAÇÃO

PERÍODO : 4º

1) INTRODUÇÃO :

Como sabemos, uma variável aleatória é contínua se seus valores são dados em intervalos. Por isso, o cálculo de probabilidades relativas a essas variáveis implicam em funções contínuas e, na maioria das vezes, funções desconhecidas ou analiticamente imprevisíveis. Entretanto, muitas das variáveis analisadas na maioria das pesquisas socioeconômicas correspondem à funções conhecidas ou se aproximam razoavelmente delas. No caso das análises com amostras processadas e expressas convenientemente por distribuições e histogramas, é elementar o cálculo de probabilidades que se referem aos limites das classes de variáveis, porém as probabilidades relacionadas à variáveis entre os limites das classes requerem o emprego de modelos auxiliares.

Exemplo ilustrativo 1 : A distribuição a seguir, registra o preço de 100 ações no mercado.

i	PREÇOS (US\$)	f_i	f_r	F_b	F_r
1	[2 , 4[	3	3%	3	3%
2	[4 , 6[	8	8%	11	11%
3	[6 , 8[	15	15%	26	26%
4	[8 , 10[	20	20%	46	46%
5	[10 , 12[	24	24%	70	70%
6	[12 , 14[	14	14%	84	84%
7	[14 , 16[	10	10%	94	94%
8	[16 , 18[	6	6%	100	100%
TOTAIS ►		100	100%		

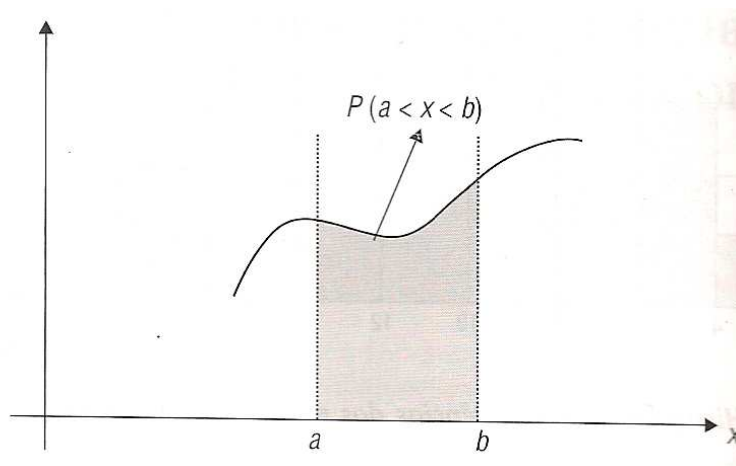
Consultando a distribuição, é possível calcular algumas probabilidades como:

→ Probabilidade do preço de uma ação ser inferior a 6 dólares: $\frac{F_{b_2}}{100} = \frac{11}{100} = F_{r_2} = 11\%$;

→ Probabilidade do preço de uma ação ser no mínimo 12 dólares: $\frac{f_6 + f_7 + f_8}{100} = \frac{30}{100} = 30\%$;

→ Probabilidade do preço de uma ação ser de 10 a 16 dólares: $\frac{f_5 + f_6 + f_7}{100} = \frac{44}{100} = 44\%$;

No caso mais geral do cálculo de probabilidade com variáveis aleatórias contínuas, emprega-se modelos expressos por funções matemáticas denominadas *Funções Densidade de Probabilidade*. Cada probabilidade de um intervalo $a < X < b$ corresponde à área sob a curva representativa da função nesse intervalo. A área total sob a curva é 1.



2) DISTRIBUIÇÃO NORMAL – CURVA NORMAL :

Dos modelos de distribuição usados para variáveis aleatórias contínuas, o mais adotado é o da *Distribuição Normal*, deduzida em 1753 por De Moivre, redescoberta por Laplace em 1774 e por Gauss em 1809. Por isso, sua curva, em forma de sino, é conhecida como *Curva de Gauss*. Essa função é dada por

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

em que:

$$-\infty < x < \infty$$

μ = média da distribuição

σ = desvio padrão da distribuição

$$\pi = 3,1416\dots$$

$$e = 2,7182\dots$$

O fato é que a área sob o gráfico da função acima implica em calcular a integral dessa função no intervalo do qual se quer a probabilidade, ou seja,

$$P(a < x < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Como essa integral não pode ser analiticamente calculada, resolve-se o problema com uma transformação de variáveis que conduz à chamada *Distribuição Normal Padronizada*. A transformação citada considera

$\frac{X_i - \mu}{\sigma} = Z_i$, em que a média de Z é zero e sua variância é igual a 1. Então, a função densidade de Z é

dada por

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

em que:

$$-\infty < x < \infty$$

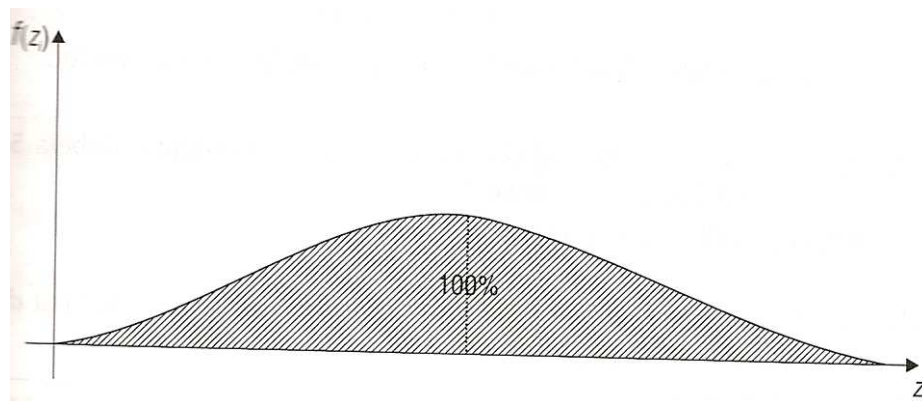
$$\mu(z) = 0$$

$$\sigma(z) = 1$$

$$\pi = 3,1416\dots$$

$$e = 2,7182\dots$$

e seu gráfico é:



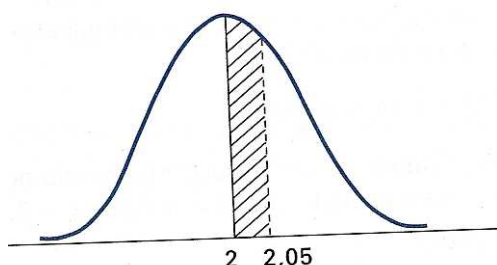
Para se calcular a probabilidade de um intervalo $a < X < b$, faz-se $P(a < X < b) = P(z_1 < Z < z_2)$ em que $Z_1 = \frac{a - \mu}{\sigma}$ e $Z_2 = \frac{b - \mu}{\sigma}$. Com esses valores, usa-se uma tabela de distribuição padronizada como a que se segue:

[illegible]

Exemplo ilustrativo 2 : Uma máquina produz parafusos com diâmetros cuja média é 2 cm e o desvio padrão é 0,04 cm. Qual é a probabilidade de essa máquina produzir um parafuso com diâmetro entre 2 e 2,05 cm ?

Resolução :

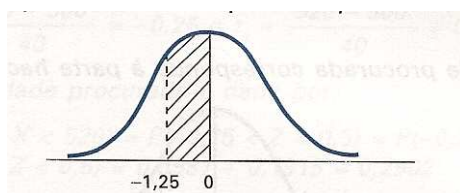
A probabilidade pedida é $P(2 < X < 2,05)$. No gráfico essa probabilidade corresponde à área hachurada:



Então, temos que $z_1 = \frac{2 - 2}{0,04} = 0$ e $z_2 = \frac{2,05 - 2}{0,04} = 1,25 \Rightarrow P(2 < X < 2,05) = P(0 < Z < 1,25)$.

Procuremos na tabela anterior: Na primeira coluna, encontramos o valor 1,2 e na primeira linha, encontramos o 5, que é o último algarismo do número 1,25. Na interseção da linha com a coluna encontrada, achamos o valor 0,3944 e $P(2 < X < 2,05) = P(0 < Z < 1,25) = 0,3944$ ou 39,44%.

Exemplo ilustrativo 3 : Suponhamos que a transformação na variável z gerasse a necessidade de calcular $P(-1,25 < Z < 0)$. Essa probabilidade equivale à área hachurada no gráfico abaixo:



Resolução :

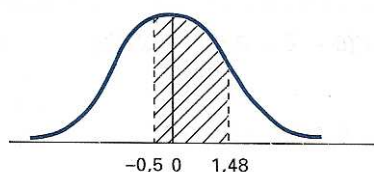
Já vimos que $P(0 < Z < 1,25) = 0,3944$. Pela simetria da curva temos: $P(-1,25 < Z < 0) = P(0 < Z < 1,25) = 0,3944$ ou 39,44%.

Exemplo ilustrativo 4 : Uma micro perfuratriz cirúrgica faz furos tais que a média dos diâmetros é 2,01 mm e o desvio padrão é 0,02 mm. Qual é a probabilidade de essa máquina fazer um furo entre 2,000 mm e 2,0396 mm ?

Resolução :

Temos que $P(2,000 < X < 2,0396) = P(z_1 < z < z_2)$. Então, $z_1 = \frac{2,00 - 2,01}{0,02} = -0,5$. Do mesmo modo,

temos $z_2 = \frac{2,0396 - 2,01}{0,02} = 1,48$, de acordo com o gráfico abaixo:

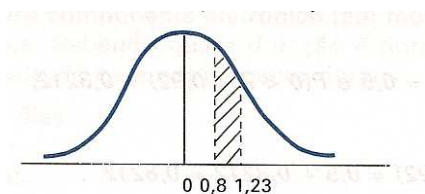


$P(2,000 < X < 2,0396) = P(-0,5 < z < 1,48) = P(-0,5 < z < 0) + P(0 < z < 1,48)$. Consultando a tabela anterior, temos $P(2,000 < X < 2,0396) = 0,1915 - 0,4306 = 0,6221$ ou 62,21%.

Exemplo ilustrativo 5 : Considere o tempo, em minutos, que um mergulhador consegue ficar submerso sem equipamento de respiração. Um levantamento feito com um grupo de mergulhadores calculou a média desses tempos : achou 4,4 minutos, com um desvio padrão de 3 minutos. Qual é, nesse grupo, a probabilidade de um mergulhador ficar submerso entre 6,8 e 8,09 minutos?

Resolução :

Temos que $P(6,8 < X < 8,09) = P(z_1 < z < z_2)$. Então, $z_1 = \frac{6,8 - 4,4}{3} = 0,8 \text{ min}$. Do mesmo modo, temos $z_2 = \frac{8,09 - 4,4}{3} = 1,23$, de acordo com o gráfico abaixo:



$P(6,8 < X < 8,09) = P(0,8 < z < 1,23) = P(0 < z < 1,23) - P(0 < z < 0,8) = 0,3907 - 0,2881 = 0,1026$ ou 10,26%.

Exemplo ilustrativo 6 : Uma clínica de exames laboratoriais entrevistou um grupo de 42 pacientes sob jejum, colheu amostra de sangue desses pacientes e mediu a concentração de glicose em todos eles. O resultado é dado pela distribuição abaixo:

i	Glicose (mg/dl)	f_i	x_i	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
1	[70 , 75[	6	72,5	435,0	162,31	973,85
2	[75 , 80[	7	77,5	542,5	59,91	419,35
3	[80 , 85[	7	82,5	577,5	0,018	0,12
4	[85 , 90[	8	87,5	700,0	5,11	40,86
5	[90 , 95[	8	92,5	740,0	52,71	421,66
6	[95 , 100[	6	97,5	585,0	150,31	901,85
TOTAIS ►		42		3.580,0		2.757,69

Calcule:

- A média da distribuição, sua variância e desvio padrão;
- A probabilidade de encontrar nesse grupo de pacientes alguém com uma concentração de glicose entre 96,4 e 98,5 mg/dl.
- A probabilidade de encontrar nesse grupo de pacientes alguém com uma concentração de glicose superior a 96 mg/dl.
- A probabilidade de encontrar nesse grupo de pacientes alguém com uma concentração de glicose inferior a 87 mg/dl.

Resolução :

a) Pela distribuição, temos que $\bar{x} = \frac{3,580}{42} = 85,24 \text{ mg / dl}$, $s^2 = \frac{2.757,69}{41} = 67,26$ e $s = 8,2 \text{ mg/dl}$;

b) $P(96,4 < X < 98,5) = P\left(\frac{96,4 - 85,24}{8,2} < z < \frac{98,5 - 85,24}{8,2}\right) = P(1,36 < z < 1,62) = P(0 < z < 1,62) -$

$- P(0 < z < 1,36) \Rightarrow \text{Tabela} \Rightarrow P(96,4 < X < 98,5) = 0,4474 - 0,4131 = 0,0343$ ou 3,43%. (Veja o gráfico na página seguinte)

c) $P(X > 96) = P\left(z > \frac{96 - 85,24}{8,2}\right) = P(z > 1,31) = P(z > 0) - P(0 < z < 1,31) = 0,5 - 0,4049 = 0,0951$
ou 9,51%. (Veja o gráfico abaixo)

Gráfico do item b

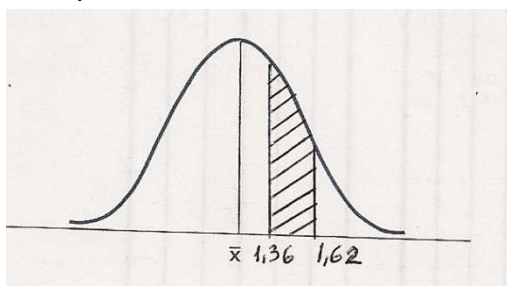
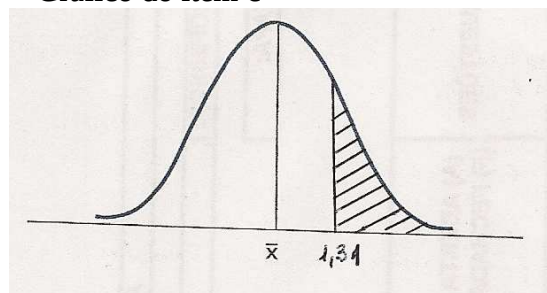
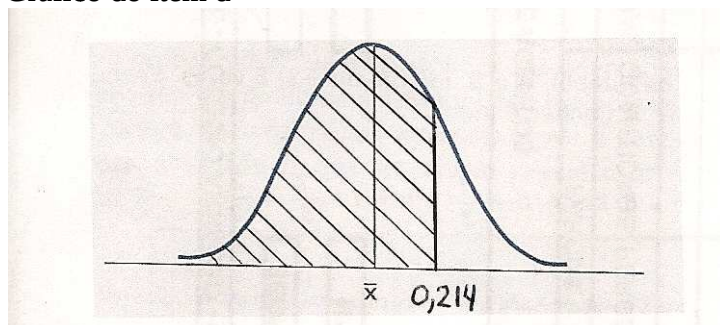


Gráfico do item c



d) $P(X < 75) = P\left(z < \frac{87 - 85,24}{8,2}\right) = P(z < 0,214) = P(z < 0) + P(0 < z < 0,214) =$
 $= 0,5 + 0,0832 = 0,5832$ ou 58,32%. (Veja o gráfico abaixo)

Gráfico do item d



Exercícios propostos :

1) Os salários semanais dos operários industriais são distribuídos normalmente em torno da média de R\$ 500,00, com desvio padrão de R\$ 40,00. Calcule a probabilidade de um operário ter um salário semanal situado entre R\$ 490,00 e R\$ 520,00.

2) Um teste de seleção para administradores candidatos a vagas numa empresa tem distribuição normal com média 100 e desvio padrão 10. Determine a probabilidade de um administrador submetido a esse teste ter nota

- a) maior do que 120;
- b) maior do que 80;
- c) entre 85 e 15;
- d) maior do que 100.

3) Os pesos de 600 estudantes são normalmente distribuídos com média 65,3 kg e desvio padrão 5,5 kg. Determine o número de estudantes que pesam

- a) entre 60 e 70 kg;
- b) mais que 63,2 kg;
- c) menos que 68 kg.

4) A duração de um certo componente eletrônico tem média de 850 dias e desvio padrão de 40 dias. Sabendo que a duração é normalmente distribuída, calcule a probabilidade de esse componente durar

- entre 700 e 1.000 dias;
- mais de 800 dias;
- menos de 750 dias.

5) Uma fábrica de pneus fez um teste para medir o desgaste de seus pneus e verificou que ele obedecia a uma distribuição normal de média 48.000 km e desvio padrão de 2.000 km. Calcular a probabilidade de um pneu escolhido ao acaso

- durar mais que 46.000 km;
- durar entre 45.000 e 50.000 km.

6) Uma variável aleatória contínua X pode ser descrita assim: $X = N(\bar{x}; \sigma^2)$, em que $\bar{x}$ é a média e σ^2 é a variância. Seja X a variável contínua, tal que $X = (12; 25)$. Qual é a probabilidade de uma observação ao acaso

- ser menor do que -3?
- cair entre -1 e 15?

7) O número $\frac{1+\sqrt{5}}{2}$, aproximadamente igual a 1,618, é conhecido como *Número de ouro*, pois, segundo a concepção do classicismo grego, tem sua origem na divisão de um todo em duas partes, tais que a menor caiba na maior o mesmo número de vezes que esta caiba no todo. Constata-se a presença deste número na morfologia dos seres vivos. Um exemplo disso é a razão entre o comprimento total das duas primeiras falanges de um dedo humano e o comprimento da terceira falange do mesmo dedo. Suponha que um pesquisador, depois de examinar uma amostra de uma população, concluiu que essa razão oscilava em torno da média 1,618, com desvio padrão igual 0,4, numa distribuição normal. Determine a probabilidade de um indivíduo dessa população apresentar uma razão

- compreendida entre 1,5 e 2,3;
- inferior a 1,5;
- superior a 1,62;

8) Sobre a proporção de ouro nas falanges do dedo, um professor resolveu usar a sua turma de 42 alunos como amostra. Para fazer a coleta de dados, ele mediu os dedos de todos da turma e calculou a razão entre as falanges. Os resultados das proporções constatadas estão ordenadas na tabela abaixo:

1,25	1,27	1,28	1,28	1,29	1,30	1,32
1,33	1,33	1,35	1,36	1,36	1,37	1,38
1,38	1,38	1,39	1,40	1,42	1,44	1,44
1,47	1,52	1,54	1,57	1,62	1,63	1,63
1,63	1,64	1,64	1,64	1,65	1,65	1,66
1,66	1,67	1,68	1,68	1,69	1,71	1,72

a) Construa uma distribuição com classes de mesma amplitude no formato abaixo:

i	RAZÃO	f_i	x_M	$f_i \cdot x_M$	$(x_M - \bar{x})^2$	$f_i \cdot (x_M - \bar{x})^2$
TOTAIS →						

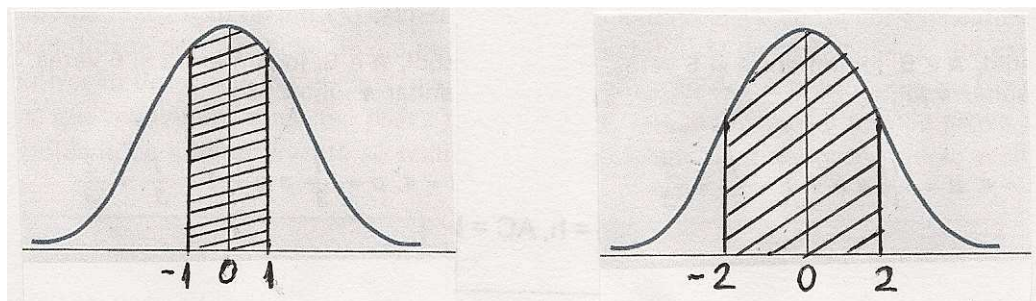
- b) Calcule a média, a variância e o desvio padrão da distribuição.
- c) Supondo a distribuição normal, calcule a probabilidade de uma razão ser
- 1º) compreendida entre 1,59 e 1,62;
 - 2º) menor do que 1,60;
 - 3º) maior do que 1,55.

3) DISTRIBUIÇÃO NORMAL – Efeitos do Desvio Padrão :

É desejável que a probabilidade de uma variável X pertença ao intervalo limitado por $\mu \pm \sigma$, de preferência, ou aos intervalos $\mu \pm 2\sigma$ e $\mu \pm 3\sigma$. Verifiquemos essas possibilidades:

a) Para $(\mu - \sigma) < X < (\mu + \sigma)$, temos, pela distribuição normal padronizada:

$\frac{(\mu - \sigma) - \mu}{\sigma} < Z < \frac{(\mu + \sigma) - \mu}{\sigma} \Rightarrow -1 < Z < 1$, cuja representação gráfica seria como na primeira figura abaixo.



Consultando a tabela de Distribuição normal padronizada, temos para $0 < z < 1$, $P(Z) = 0,3413$. Como a área correspondente ao intervalo $-1 < Z < 1$ é o dobro de $P(Z)$, temos $P(-1 < Z < 1) = 0,6826 = 68,26\%$.

b) Para $(\mu - 2\sigma) < X < (\mu + 2\sigma)$, temos, pela distribuição normal padronizada:

$\frac{(\mu - 2\sigma) - \mu}{\sigma} < Z < \frac{(\mu + 2\sigma) - \mu}{\sigma} \Rightarrow -2 < Z < 2$, cuja representação gráfica seria como na segunda figura acima.

Consultando a tabela de Distribuição normal padronizada, temos para $0 < z < 2$, $P(Z) = 0,4772$. Como a área correspondente ao intervalo $-2 < Z < 2$ é o dobro de $P(Z)$, temos $P(-2 < Z < 2) = 0,9544 = 95,44\%$.

b) Para $(\mu - 3\sigma) < X < (\mu + 3\sigma)$, temos, pela distribuição normal padronizada:

$\frac{(\mu - 3\sigma) - \mu}{\sigma} < Z < \frac{(\mu + 3\sigma) - \mu}{\sigma} \Rightarrow -3 < Z < 3$, cuja representação gráfica seria como na segunda figura acima.

Consultando a tabela de Distribuição normal padronizada, temos para $0 < z < 3$, $P(Z) = 0,4987$. Como a área correspondente ao intervalo $-3 < Z < 3$ é o dobro de $P(Z)$, temos $P(-3 < Z < 3) = 0,9974 = 99,74\%$.

Então, as probabilidades obtidas não dependem dos valores da média(μ) e do desvio padrão(σ). Esses valores apenas determinam se a curva é *platicúrtica* (se a dispersão medida por σ for grande) ou *leptocúrtica* (se a dispersão medida por σ for pequena).

4) DISTRIBUIÇÃO NORMAL – Combinação Linear de Distribuições Normais Independentes:

A combinação linear de duas distribuições normais independentes é também uma distribuição normal, ou seja, se X , Y e Z são variáveis aleatórias contínuas com distribuição normal, então $W = aX + bY + c$, com a , b e c constantes, será uma distribuição normal tal que

$$\mu_W = a.\mu_X + b.\mu_Y + c \quad \text{e} \quad \sigma_W = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Exemplo ilustrativo 7 :

Os gastos mensais de um restaurante com bebidas, segundo uma distribuição normal, são expressos por $N_B(12.400, 4.440.000)$ e os gastos com comida, segundo uma distribuição normal, são expressos por $N_C(32.200, 2.110.000)$. Qual é a probabilidade de a despesa total com esses itens ser

- menor do que 37.000 reais?
- compreendida entre 35.000 e 38.000 reais?
- superior a 42.000 reais?

Resolução :

Aqui, temos $N = N_B + N_C$ ou $N = N_B(12.400, 4.440.000) + N_C(32.200, 2.110.000) = N(44.600, 6.550.000) \Rightarrow \mu = 44.600 \text{ reais e } \sigma = \sqrt{6.550.000} = 2.559,30 \text{ reais}$

- $P(X < 37.000) = P(Z < \frac{37.000 - 44.600}{2.559,30}) = P(Z < -2,97) = 0,5 - P(0 < Z < 2,97) = 0,5 - 0,4985 = 0,15\%.$
- $P(35.000 < X < 38.000) = P(\frac{35.000 - 44.600}{2.559,30} < Z < \frac{38.000 - 44.600}{2.559,30}) = P(-3,75 < Z < -2,58) = P(0 < Z < 3,75) - P(0 < Z < 2,58) = 0,4999 - 0,4951 = 0,48\%.$
- $P(X > 42.000) = P(Z > \frac{42.000 - 44.600}{2.559,30}) = P(Z > -1,02) = 0,5 + P(0 < z < 1,02) = 0,5 + 0,3461 = 84,61\%.$

Exercícios propostos :

9) Em uma empresa, a montagem de uma peça é feita em duas etapas. Os tempos requeridos para essas etapas são independentes e têm as seguintes distribuições: $N_1(75 \text{ seg} ; 16,81 \text{ seg}^2)$ e $N_2(129 \text{ seg} ; 106,09 \text{ seg}^2)$. Qual é a probabilidade de montar a peça em menos de 200 segundos?

10) Certo produto tem peso médio de 10 g e desvio padrão de 0,5 g. Esse produto é embalado em caixas de 120 unidades que pesam, em média 150 g e têm desvio padrão de 8g. Qual é a probabilidade de que uma caixa cheia pese mais do que 1.370 g?

11) Um avião de turismo de quatro lugares pode levar uma carga útil de 350 kg. Supondo que os passageiros tenham peso de 70 kg, com distribuição normal de peso, e desvio padrão de 20 kg, e que a bagagem de cada passageiro pese em média 12 kg, com desvio padrão de 5 kg e distribuição normal de peso, calcule a probabilidade de

- haver sobrecarga se o piloto não pesar os quatro passageiros e as respectivas bagagens;
- que o piloto tenha que tirar pelo menos 50kg de gasolina para evitar sobrecarga.

12) Em uma distribuição normal, 28% dos elementos são superiores a 34 e 12% são inferiores a 19. Encontrar a média e a variância da distribuição.

- 13)** As variáveis $X_1 \cong N(10 ; 9)$, $X_2 \cong N(-2 ; 4)$ e $X_3 \cong N(5 ; 25)$ são independentes. Determinar a distribuição de $Y = X_1 + X_2 + X_3$.
- 14)** Suponha que os diâmetros médios dos parafusos produzidos por uma fábrica seja de 0,25 polegadas e o desvio padrão de 0,02 polegadas. Um parafuso é considerado defeituoso se seu diâmetro é maior do que 0,28 polegadas ou menor do que 0,20 polegadas. Encontrar a porcentagem de parafusos defeituosos.
- 15)** Suponha que a duração de vida de dois equipamentos, E_1 e E_2 , tenham respectivamente distribuições $N(45 ; 9)$ e $N(40 ; 36)$. Se o equipamento tiver que ser usado por um período de 45 horas, qual deles deve ser preferido?
- 16)** Certa máquina de empacotar determinado produto oferece variações de peso com desvio padrão de 20 g. Em quanto deve ser regulado o peso médio para que apenas 10% dos pacotes tenham menos que 400g?