Tópicos especiais em Estatística Aplicada - 2023/2

EPC3

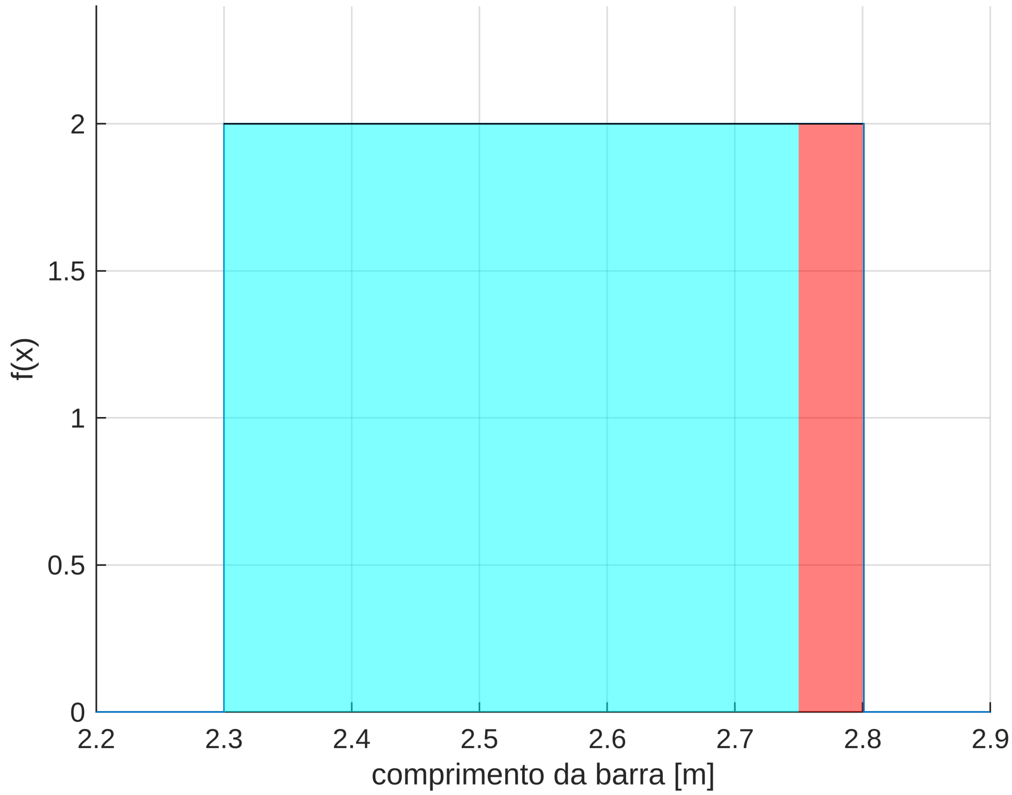
Aluno: João Gabriel Santos Custodio

1) A função densidade de probabilidade do comprimento de uma barra de metal é dada por f(x) = 2 para 2.3 < x < 2.8 m.

1.1. Se as especificações para produzir esta barra forem comprimento de 2.25 a 2.75 m, qual a proporção de barras que não atenderá as especificações?

Para o intervalo de 2.75<x<2.8 as barras não atenderão as especificações. A proporção de barras que não atenderão as especificações pode ser encontrada sendo a área sob a curva desse intervalo de probabilidade.

```
x = linspace(2.2, 2.9, 700);
prob = 2;
figure;
grid(), hold on;
y1 = cat(2, cat(2, zeros(1,100), ones(1,500)*prob), zeros(1,100));
stairs(x, y1)
ylim([0, 1.2*prob])
area([2.3, 2.75], [2, 2], 'FaceColor', 'cyan', 'FaceAlpha', 0.5);
area([2.75, 2.8], [2, 2], 'FaceColor', 'red', 'FaceAlpha', 0.5);
ylabel("f(x)")
xlabel("comprimento da barra [m]")
hold off
```



Portanto, evidencia-se que a área sob o gráfico onde não atende as especificações é a área destacada em vermelho no gráfico acima. Dessa forma, temos:

(2.8-2.75)*2 = 0.05*2 = 0.10 = 10%.

Logo 10% das barras produzidas não atenderão as espeficicações.

1.2. Suponha que a função densidade de probabilidade seja f(x) = 2 para o intervalo de 0.5 m de comprimento. Sobre qual valor deveria f(x) ser centrada para obter a máxima proporção de barras dentro da especificação?

Utilizando o intervalo de 0.5m de comprimento, basta centralizar f(x) no centro do intervalo especificado no item 1.1 e encontrar a mediana. Nesse caso, como já mencionado, o intervalo é de 2.3<x<2.75.

```
x1_2 = linspace(2.3, 2.75, 1000);
median(x1_2)
```

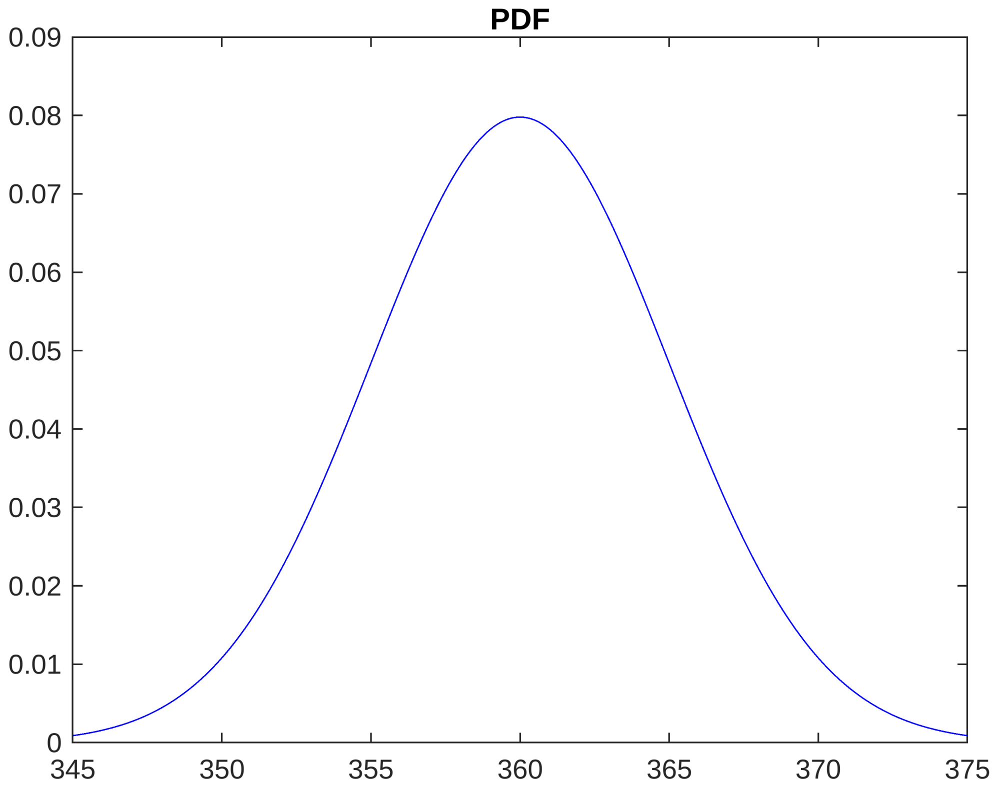
```
ans = 2.5250
```

Conclui-se, portanto, que para nenhuma barra ser produzida fora do intervalo especificado, f(x) há de ser centrada no valor 2.53.

2. O volume de enchimento de uma máquina de enchimento automatizada usada para encher refrigerantes é normalmente distribuído com uma média de 360ml e um desvio padrão de 5ml.

Abaixo temos o gráfico que mostra a PDF do problema:

```
media = 360; % 360ml de média de enchimento normalmente distribuidos
sigma = 5; % 5ml de desvio padrão
x = media - 3*sigma: 0.1: media+3*sigma;
pdf = normpdf(x, media, sigma);
figure
plot(x, pdf, 'blue')
ylim([0, 0.09])
title("PDF")
```



2.1 Qual a probabilidade de que o volume seja menor que 355ml?

```
volume_prob2_1 = find(x<355);
probabilidade2_1 = trapz(x(volume_prob2_1), pdf(volume_prob2_1))
```

```
probabilidade2_1 = 0.1525
```

Fazendo uma análise estatística, é necessário calcular a integral do gráfico de forma que os limites de integração sejam limitados para valores até x<355 como calculado pelo código acima.

Dessa forma, foi encontrado o resultado sendo uma probablidade de 15,25%

2.2 Se os refrigerantes com menos de 350ml e mais de 370ml forem eliminados, qual a proporção de refrigerantes eliminados?

```
volume_prob2_2_less = find(x<350);
volume_prob2_2_bigger = find(x>370);
probabilidade2_2 = (trapz(x(volume_prob2_2_less), pdf(volume_prob2_2_less))
+ trapz(x(volume_prob2_2_bigger), pdf(volume_prob2_2_bigger)))
```

```
probabilidade2_2 = 0.0407
```

Da mesma forma que calcular o resultado concatenando as áreas que correspondem aos refrigerentes eliminados para os volumes menores que 350ml e maiores que 370ml. Sendo assim, a proporção encontrada é de 4 refrigerantes eliminados para 100 produzidos.

2.3 Determine as especificações de volume de modo que 99% dos refrigerantes sejam aceitos.

```
for xaxis = 360:1:380
    volume_prob2_3 = find(x<xaxis);
    probabilidade2_3 = trapz(x(volume_prob2_3), pdf(volume_prob2_3),2);
    if probabilidade2_3 > 0.99
        X = ['Para uma especificacao de ', num2str(xaxis), ' temos uma prob
de ', num2str(probabilidade2_3), ' serem aceitos'];
        disp(X)
    end
end
```

```
Para uma especificacao de 373 temos uma prob de 0.99371 serem aceitos
Para uma especificacao de 374 temos uma prob de 0.99593 serem aceitos
Para uma especificacao de 375 temos uma prob de 0.99721 serem aceitos
Para uma especificacao de 376 temos uma prob de 0.9973 serem aceitos
Para uma especificacao de 377 temos uma prob de 0.9973 serem aceitos
Para uma especificacao de 378 temos uma prob de 0.9973 serem aceitos
Para uma especificacao de 379 temos uma prob de 0.9973 serem aceitos
Para uma especificacao de 380 temos uma prob de 0.9973 serem aceitos
```

Com o código acima, foi possível encontrar a porcentagem de refrigerantes aceitos para um intervalo de especificações que foi arbritariamente escolhido de 360ml até 380, em seguida foi mostrado apenas os valores que satisfazem a especificação com 99% de refrigerantes aceitos na produção. Com isso, foi possível determinar que para uma linha de produção para 373ml, temos que 99% dos refrigerantes serão aceitos.

3. Seja uma variável aleatória X com distribuição binomial. Escolha um valor de n > 10 e p conforme abaixo.

3.1 Escolha p tal que X possa ser aproximada por uma variável aleatória Y com distribuição normal, plotando os gráficos da distribuição de X e de Y.

Para este problema, será utilizado n de tal forma que $\mathrm{np} > 5\, e\, n(1-p) > 5$.

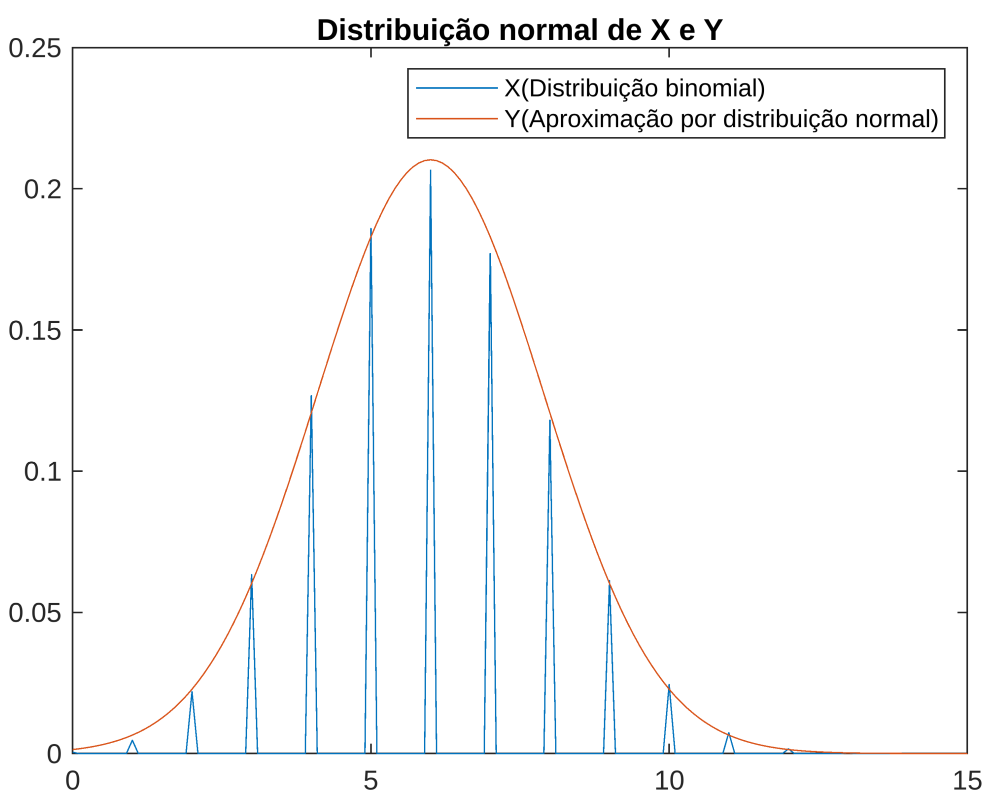
Escolhendo n = 15, então $p > \frac{5}{15}$=0.33, assim podemos escolher p = 0.4 e todas inequações são atendidas.

Utilizando a equação abaixo, podemos calcular $\mu\, e\, \sigma$ para o problema:

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)$$

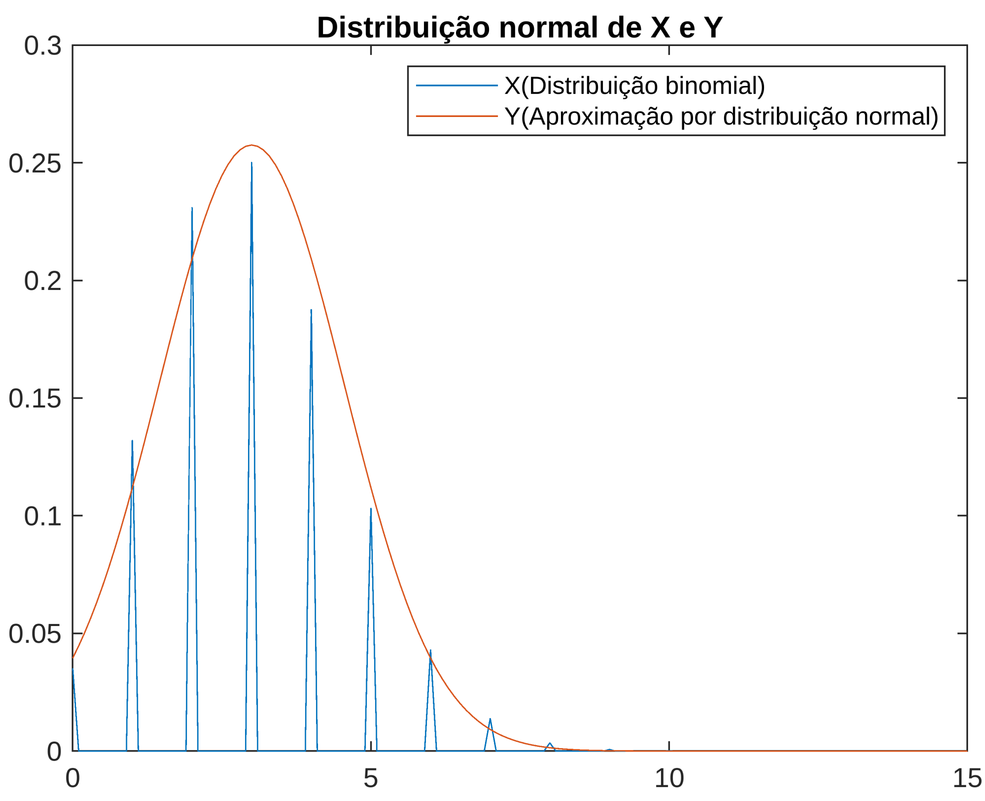
```
n = 15;
p = 0.4;

axisX = 0:0.1:n;
X = binopdf(axisX,n,p);

mu2 = n*p;
sigma2 = sqrt(n*p*(1-p));
Y = normpdf(axisX,mu2,sigma2);

figure;
plot(axisX,X)
hold on
plot(axisX,Y)
title("Distribuição normal de X e Y")
legend("X(Distribuição binomial)","Y(Aproximação por distribuição normal)")
```



3.2 Repita para p tal que esta aproximação não seja possível, mostrando isso através dos gráficos de distribuição correspondentes.

Para satisfazer a não possibilidade da aproximação, basta escohermos p de tal forma que p<0.33, ou seja, se p=0.2 veremos que a aproximação não será possível.

```
p2 = 0.2;
% Criando X
inter = 0:0.1:n;
X2 = binopdf(inter,n,p2);
% Criando Y
mu3 = n*p2;
sigma3 = sqrt(n*p2*(1-p2));
Y2 = normpdf(inter,mu3,sigma3);
figure;
plot(inter,X2)
hold on
plot(inter,Y2)
title("Distribuição normal de X e Y")
legend("X(Distribuição binomial)","Y(Aproximação por distribuição normal)")
```



Com o gráfico acima, fica evidente que a aproximação falha em representar perfeitamente a distribuição binomial pois há valores que não possuem intersecção com a Distribuição binomial.

4. Gere uma variável aleatória X com distribuição normal padronizada contendo 500 amostras.

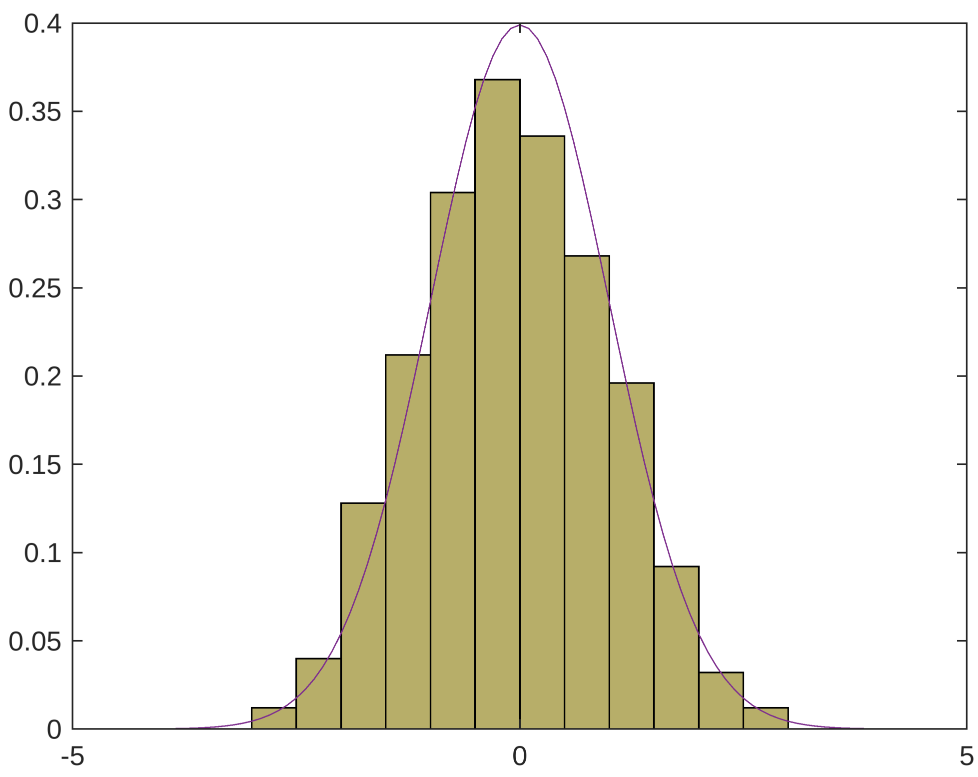
```
x = randn(500,1);
```

4.1 Faça um histograma de X e sobre ele plote a pdf (0,1)

```
figure()
histogram(x,'Normalization', 'pdf');
hold on
mu = 0;
sigma = 1;
x_axis = -5:0.1:5;
f = exp(-(x_axis-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
plot(x_axis, f)
```

4.2 Faça um histograma de $Y = X^2$ e sobre ele plote a pdf

```
Y = x.^2;
histogram(x,'Normalization', 'pdf');
hold on
mu = 0;
sigma = 1;
x_axis = -5:0.1:5;
f = exp(-(x_axis-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
plot(x_axis, f)
```



4.3 Obtenha $P(-2 \leq X \leq 2)$ e y tal que $P(Y \leq y) = P(-2 \leq X \leq 2)$

4.4 Plote a pdf χn2 para n = 1,2, . . ,10, explique o efeito de n sobre a pdf, e para n = 10 compare χn 2 com (n, √2n).

Gere uma variável aleatória X com distribuição uniforme com 1000 amostras e intervalo [a,b] de sua escolha.

```
x_rand = 50.*rand(1,1000);
```

5. Selecione aleatoriamente 20 amostras de X e construa o intervalo de confiança para a média com α = 5%. Considere a variância conhecida, obtida da população de 1000 amostras. Repita este procedimento e construa 50 intervalos para a média. Plote os intervalos das médias junto com a média de X e comente os resultados que obteve, comparando os intervalos estimados e verificando se eles contêm a média conhecida da população.

Resposta:

Com a variavel x_rand previamente instanciada, podemos calcular o valor da variância e de sua média.

```
sigma_xrand = std(x_rand)
```

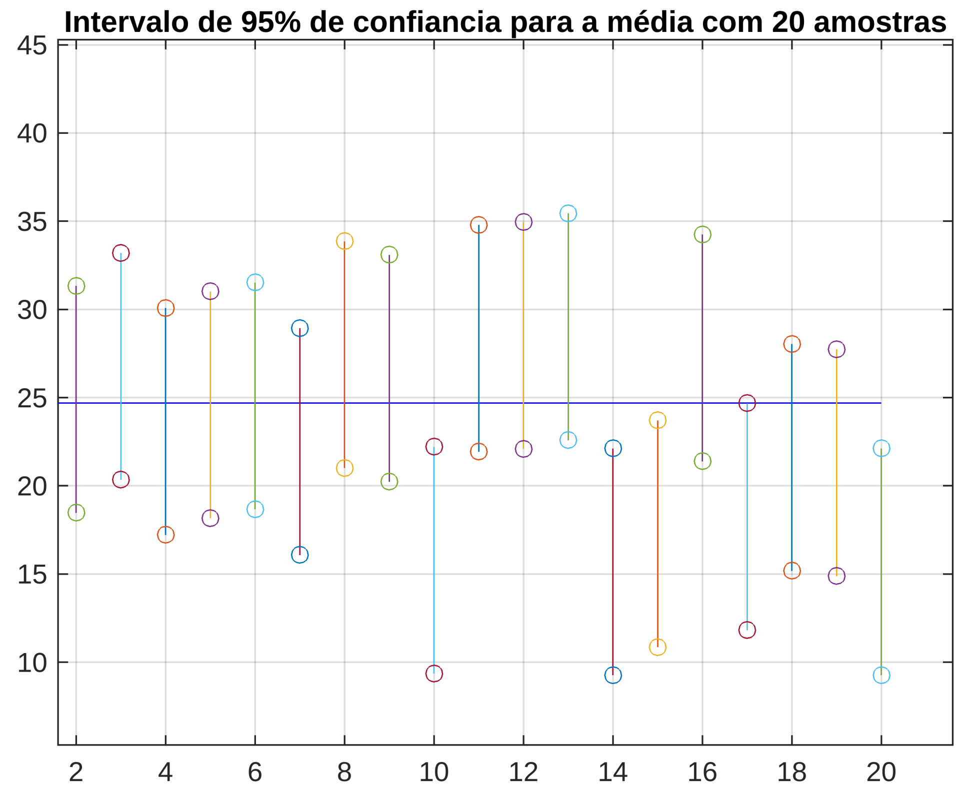
```
sigma_xrand = 14.6708
```

```
mu_xrand = mean(x_rand)
```

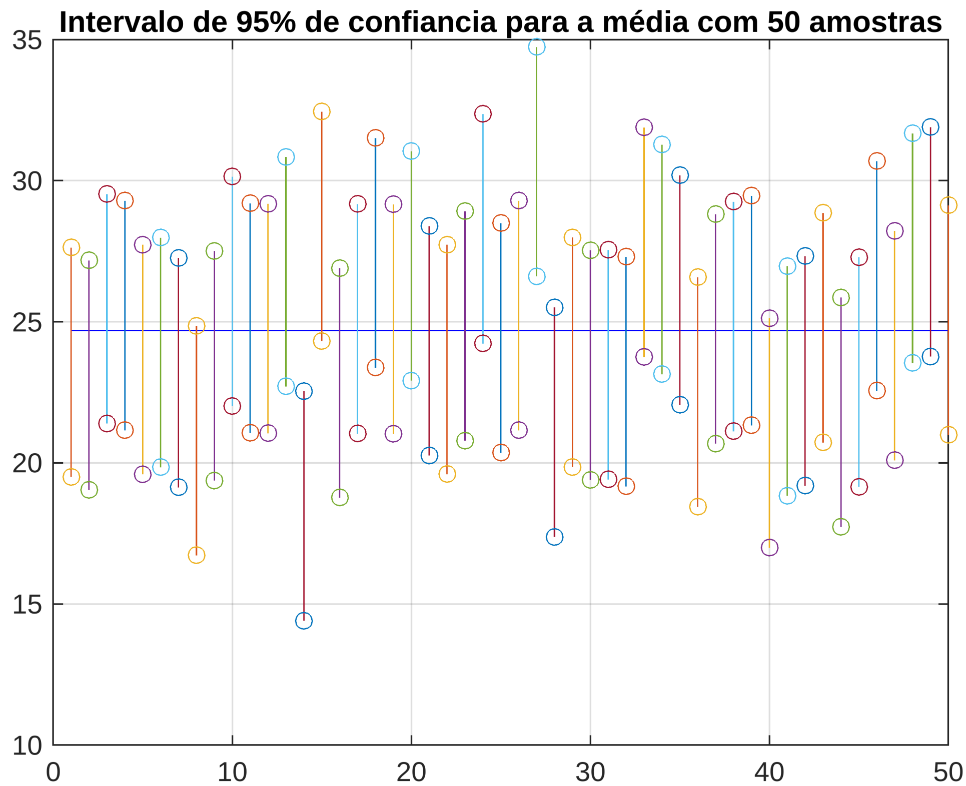
```
mu_xrand = 24.6967
```

Com esses valores obtidos, pode-se obter o intervalo de confiança de x_rand para 20 amostras e logo em seguida para 50 amostras e posteriormente discutir o resultado obtido.

```
%% 20 amostras
n_20 = 20;
z_a2 = 1.96;
figure(1)
plot(1:1:n_20,mu_xrand*ones(1,n_20),"b")
hold on
for i = 1:1:n_20
i_20 = randperm(length(x_rand),n_20);
vet_50 = x_rand(i_20);
x_bar = mean(vet_50);
inter_low = x_bar-z_a2*sigma_xrand/sqrt(n_20);
inter_high = x_bar+z_a2*sigma_xrand/sqrt(n_20);
plot([i,i],[inter_low, inter_high])
scatter([i,i],[inter_low, inter_high])
end
grid("on")
title("Intervalo de 95% de confiancia para a média com 20 amostras")
xlim([1.6 21.6])
ylim([5.3 45.3])
hold off
```



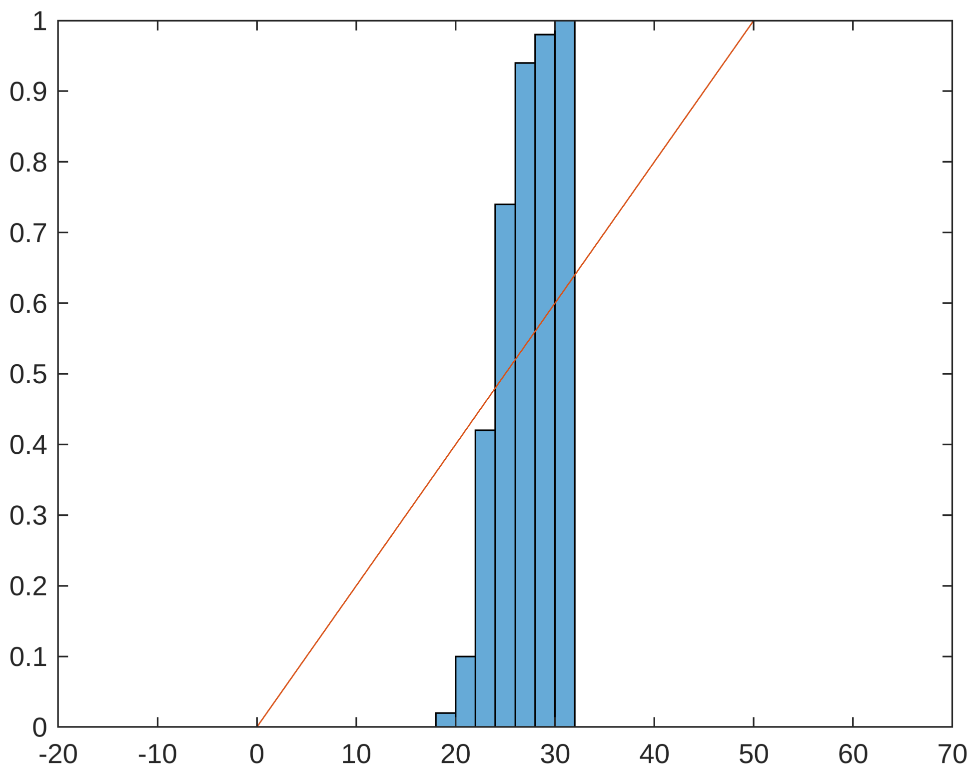

```
%% Agora com 50 amostras

n_50 = 50;
z_a2 = 1.96;
media_50 = zeros(1,50);
figure
plot(1:1:n_50,mu_xrand*ones(1,n_50),"b")
hold on
for i = 1:1:n_50
    i_50 = randperm(length(x_rand),n_50);
    vet_50 = x_rand(i_50);
    x_bar = mean(vet_50);
    media_50(1, i) = x_bar;
    inter_low = x_bar-z_a2*sigma_xrand/sqrt(n_50);
    inter_high = x_bar+z_a2*sigma_xrand/sqrt(n_50);
    plot([i,i],[inter_low, inter_high])
    scatter([i,i],[inter_low, inter_high])
end
grid("on")
title("Intervalo de 95% de confiancia para a média com 50 amostras")
hold off
```



Diante dos resultados obtidos nos gráficos para 20 amostras e 50 amostras, fica evidente que com o aumento das amostras o intervalo de confiança intercede, para esse caso do gráfico em especial, quase todas as retas das médias de cada variavel pseudo-aleatória que foi gerada. Vale salientar que se for gerando outra amostragem aleatória a intersecção com o intervalo de confiança não necessariamente terá 100% de acerto, mas é perfeitamente possível.

6. Plote o histograma cumulativo das 50 médias calculadas e compare com a cdf teórica da média amostral de X.

```
x_acc = mu_xrand-3*sigma_xrand:0.1:mu_xrand+3*sigma_xrand;
cdf_X = cdf("Uniform", x_acc, 0, 50);
figure
histogram(media_50, "Normalization", "cdf")
hold on
plot(x_acc, cdf_X)
hold off
```



Observando o gráfico acima, vale comentar que o comportamento crescente e linear com tendência ao valor 1 da CDF era esperado visto que a maior parte dos valores se concentra na média, e os valores extremos são cada vez mais improváveis à medida que nos afastamos da média.

7. Repita a atividade 5 fazendo um intervalo de confiança para a variância da variável X gerada com distribuição normal, média μ e variância σ escolhidas. Faça o histograma das 50 variâncias e analise.

8. Um fabricante de equipamentos seleciona aleatoriamente 1000 unidades e verifica que 5 têm defeito.

8.1. Construa um intervalo de confiança de 95% para a proporção de unidades com defeito.

De acodo com o enunciado:

$n = 1000$

$x = 5$

Nível de confiança = 0,95

Primeiramente, devemos calcular a Proporção Amostral do problema, e essa pode ser obtida com a equação abaixo

$$\text{PA} = \frac{x}{n}$$

$$\text{PA} = \frac{x}{n} = \frac{5}{1000} = 0,005$$

$$\text{Erro}_{\text{padrão}} = \sqrt{\frac{\text{PA}(1-\text{PA})}{n}} = \sqrt{0.005\frac{(0.995)}{10}} = 0.0022$$

Tendo PA, e sabendo que Z é valor crítico da distribuição normal padrão correspondente ao nível de confiança desejado (95%) então Z = 1.96. Dessa forma podemos calcular os limites das extremidades do intervalo de confiança:

$$\text{Limite}_{\text{inferior}} = \text{PA} - Z * \text{Erro}_{\text{padrão}} = 0.005 - 1.96 * 0.0022 = 0.0006$$

$$\text{Limite}_{\text{isuperior}} = \text{PA} + Z * \text{Erro}_{\text{padrão}} = 0.005 + 1.96 * 0.0022 = 0.0094$$

Potanto, obtem-se que o intervalo de confiança para o problema está entre 0.0006 e 0.0094.

8.2. Há evidências para suportar a afirmação de que a fração de peças com defeito é menor que 1%?

Para concluirmos que essa afirmação está correta, precisamos calcular quanto que vale Z para PO = 0,01.

$$Z = \frac{\text{PA} - \text{PO}}{\sqrt{\frac{(\text{PO}(1-\text{PA}))}{n}}} = \frac{0.005 - 0.01}{\sqrt{\frac{0.01(0.995)}{1000}}} = -1.5851$$

$$Z > -1.96$$

Uma vez que encontramos um valor de Z maior que o Valor Crítico, não podemos afirmar que a fração de peças com defeito é menor que 1%, isto para um nível de confiança de 95%.