Segunda avaliação

Prof. Thiago Gentil Ramires - Probabilidade e Estatística

Nome: Eduardo Yuji Yoshida Yamada

RA: 2320606

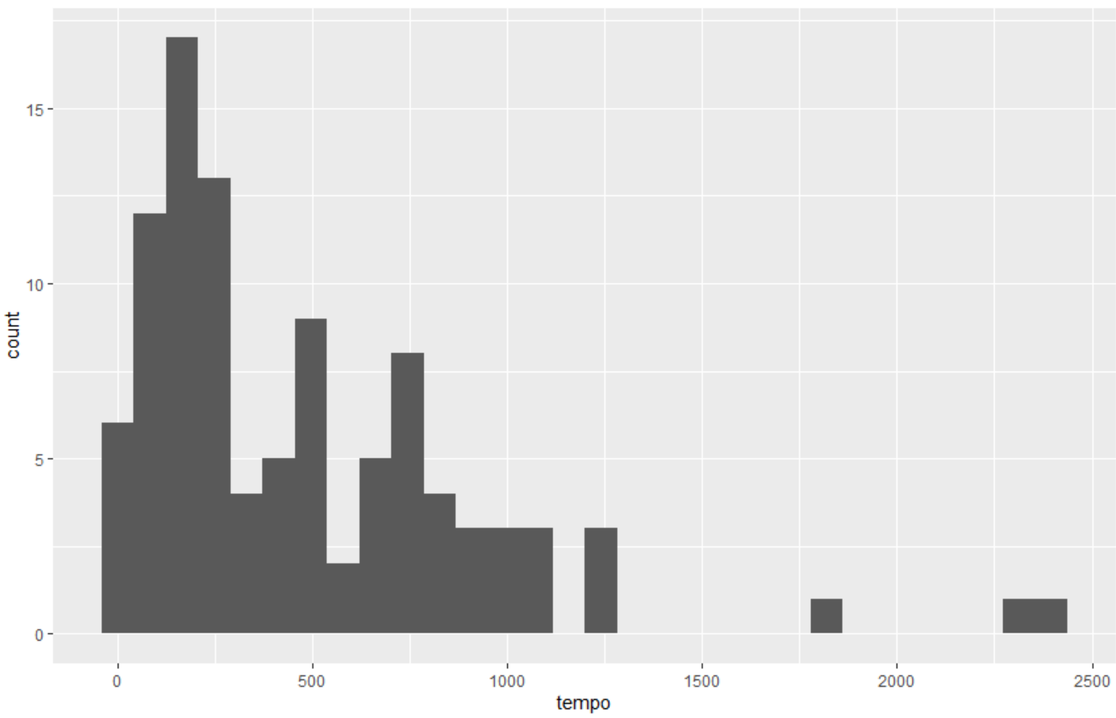
Disciplina matriculado(a): Probabilidade e Estatística – Eng. Comp.

Para todas as questões abaixo, **interprete os resultados e apresente os códigos e gráficos, quando necessário**. As bases de dados estão em anexo do Google Classroom, já salvas em CSV, com o separador decimal em Inglês, ou seja, as decimais estão separadas por ponto. **Cada questão vale 0,5 pontos.**

Questão 1) O tempo de duração (em horas) de 100 vigas metálicas, após teste de força, está apresentado no arquivo ex1.csv. Determine:

a) Qual o melhor modelo de probabilidade que representa tais tempos?

Resposta: Exponencial, o gráfico gerado se comporta como uma função exponencial



Códigos:

```
install.packages('ggplot2')

library(ggplot2)

dados = read.csv('ex1.csv',sep = ',',dec = '.',header = T)

ggplot(dados,aes(tempo))+geom_histogram(bins=30)
```

b) Determine a probabilidade de uma viga durar menos de 300 horas.

Resposta: `0.464611`

Códigos:

```
media = mean(dados$tempo)
pexp(300, 1/media)
```

c) Determine a probabilidade de uma viga durar entre 200 e 400 horas.

Resposta: `0.2246091`

Códigos:

```
pexp(400, 1/media)-pexp(199, 1/media)
```

d) Serão descartadas 70% das vigas com menor tempo de duração. Determine o tempo ideal de descarte, ou seja, o tempo que limita o descarte das vigas.

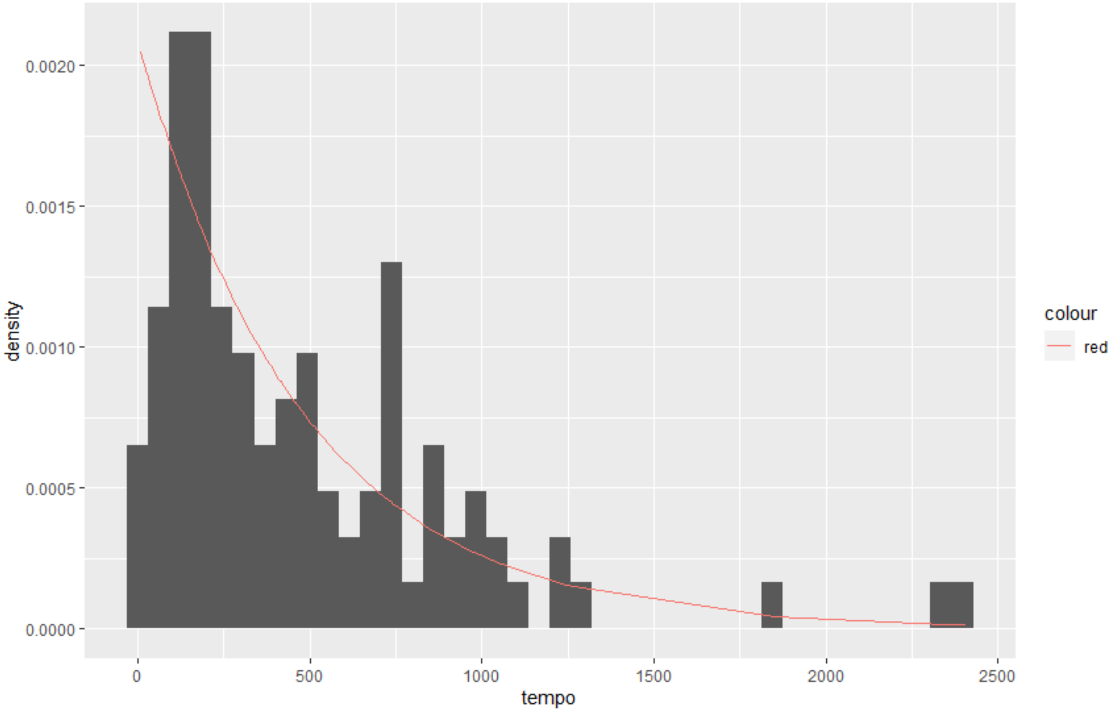
Resposta: `578.1273 horas`

Códigos:

```
qexp(0.7, 1/media)
```

e) Apresente graficamente o histograma dos dados juntamente com o modelo escolhido (ajustado) na resposta A).

Resposta:



Códigos:

dados$ptempo=dexp(dados$tempo,1/media)

ggplot(dados,aes(tempo))+geom_histogram(aes(y=..density..),bins=40)+geom_line(aes(tempo,ptempo,col='red'))

Questão 2) Uma empresa está interessada em estudar o comportamento dos produtos eletrônicos produzidos por ela. Em um teste com 20 desses produtos produzidos, 6 apresentaram defeitos. Em uma semana foram produzidos 100 novos produtos. Determine:

a) A probabilidade de mais de 75 ou mais não apresentarem defeito?

Resposta: `0.1631301 Soma das probabilidades de 75 até 100`

Códigos:

sum(dbinom(75:100,100,14/20))

b) A probabilidade de exatamente 70 não apresentarem defeito?

Resposta: `0.08678386`

Códigos:

dbinom(70,100,14/20)

c) A probabilidade de menos de 20 produtos apresentarem defeito?

Resposta: 0.008887208

Códigos:

pbinom(19,100,6/20)

d) Se no próximo mês forem produzidos 4000 produtos, quantos irão falhar em média?

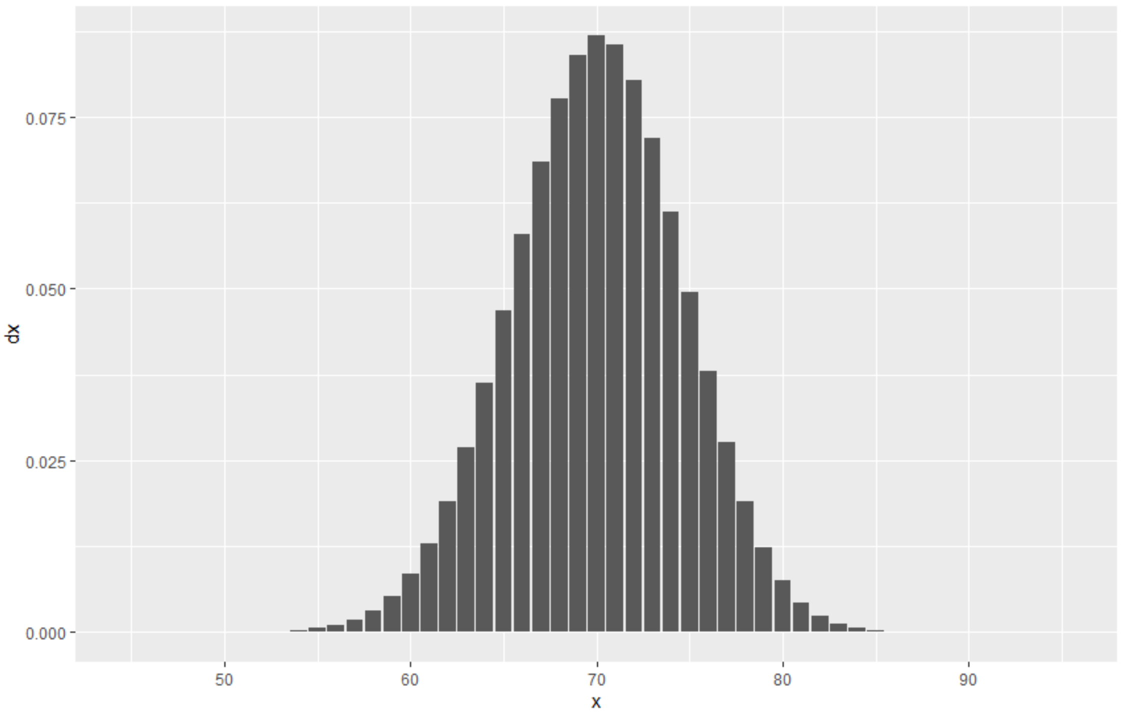
Resposta: 1200

Códigos:

*4000*6/20*

e) Apresente graficamente todas as probabilidades de não apresentar falha, considerando os n=100 produtos.

Resposta: O valor de x foi selecionado de 45 até 95, pois o intervalo antes de 50 e após 90 não apresenta gráfico. É possível ver que em 70, possui a maior probabilidade de não apresentar falhas



Códigos:

x=45:95

```
dx=dbinom(x,100,14/20)

dados1 = data.frame(x, dx)

ggplot(dados1, aes(x, dx))+geom_col()
```

Questão 3) Um algoritmo de detecção de anomalias capta, em média, 20 erros por hora. Determine:

a) A probabilidade de detectar 15 erros em uma hora?

Resposta: 0.05164885

Códigos:

```
dpois(15, 20)
```

b) A probabilidade de detectar entre 20 e 30 erros em uma hora?

Resposta: 0.5162681

Códigos:

```
sum(dpois(20:30,20))
```

c) A probabilidade de detectar mais de 449 erros em um dia?

Resposta: 0.919181

Códigos:

```
1-ppois(449,20*24)
```

d) Um novo algoritmo foi testado, sendo que a média de detecção em uma hora foi de 30 erros. Esse algoritmo será adquirido pela empresa caso detecte na próxima hora mais de 34 erros. Determine a probabilidade da compra ser realizada.

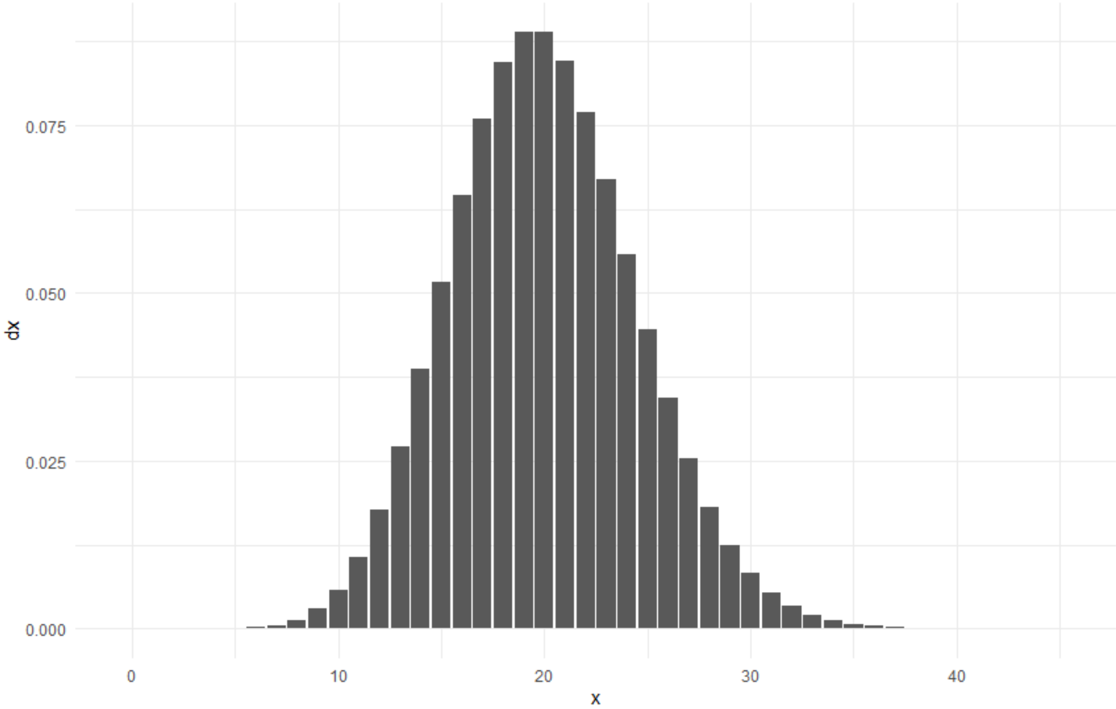
Resposta: 0.2026917

Códigos:

```
1-ppois(34,30)
```

e) Apresente o gráfico das probabilidades de detecção de erros do algoritmo, considerando o algoritmo com média de 20 erros por hora.

Resposta: O valor de x foi selecionado de 0 até 45, pois o intervalo após 40 não apresenta gráfico. É possível ver que perto de 20, possui a maior probabilidade de ocorrer um erro



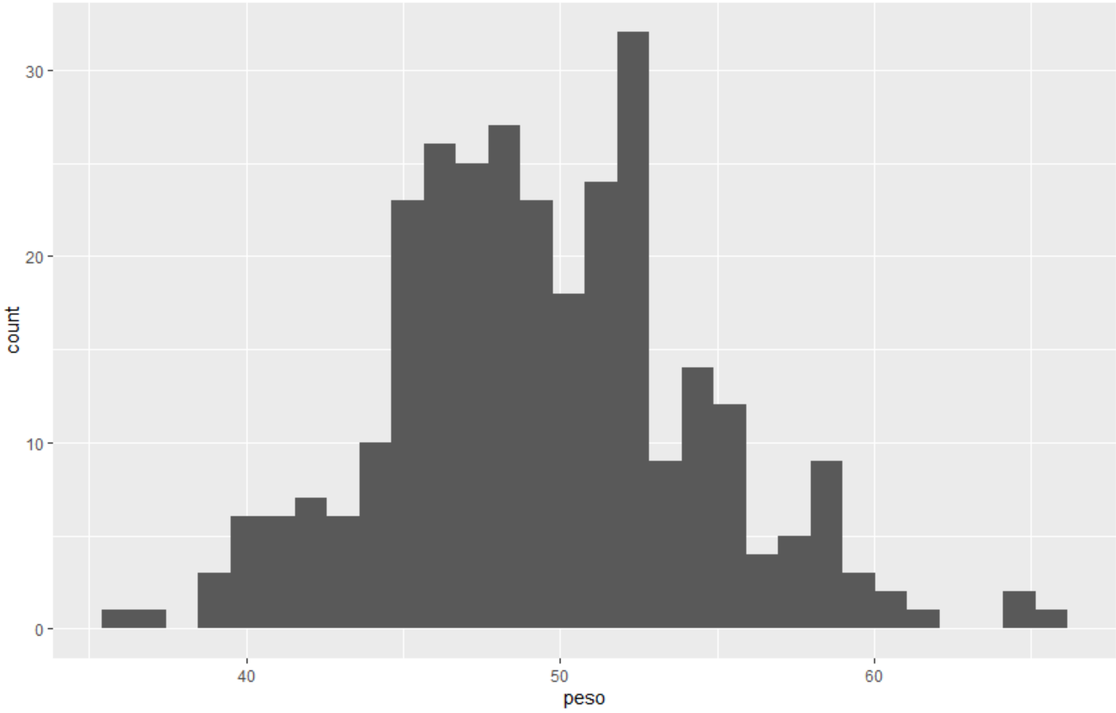
Códigos:

```
x=0:45
dx=dpois(x,20)
dados2=data.frame(x,dx)
ggplot(dados2,aes(x,dx))+geom_col()+theme_minimal()
```

Questão 4) O peso de um produto, em Kg, foi determinado em uma amostra de tamanho 300 (Ver anexo ex4.csv). Determine:

a) Qual o melhor modelo de probabilidade que representa tais pesos?

Resposta: Pela análise gráfica é possível definir que o melhor modelo é o modelo Normal



Códigos:

dados3 = read.csv('ex4.csv', dec = ".", header = T)

ggplot(dados3,aes(peso))+geom_histogram()

b) Determine a probabilidade de um produto não pesar entre 45 e 55kg.

Resposta: 0.3137118

Códigos:

pnorm(45, mean(dados3$peso),sd(dados3$peso))+(1-pnorm(55,

mean(dados3$peso),sd(dados3$peso)))

c) Determine a probabilidade de um peso ter exatamente 50Kg.

Resposta: 0, pois não existe probabilidade de um ponto para modelos contínuos

d) Os produtos serão classificados entre: leves (30% mais baixos), médios (pesos entre 30 a 70%) e pesados (os 30% maiores). Determine os limites dos pesos para realizar essa classificação.

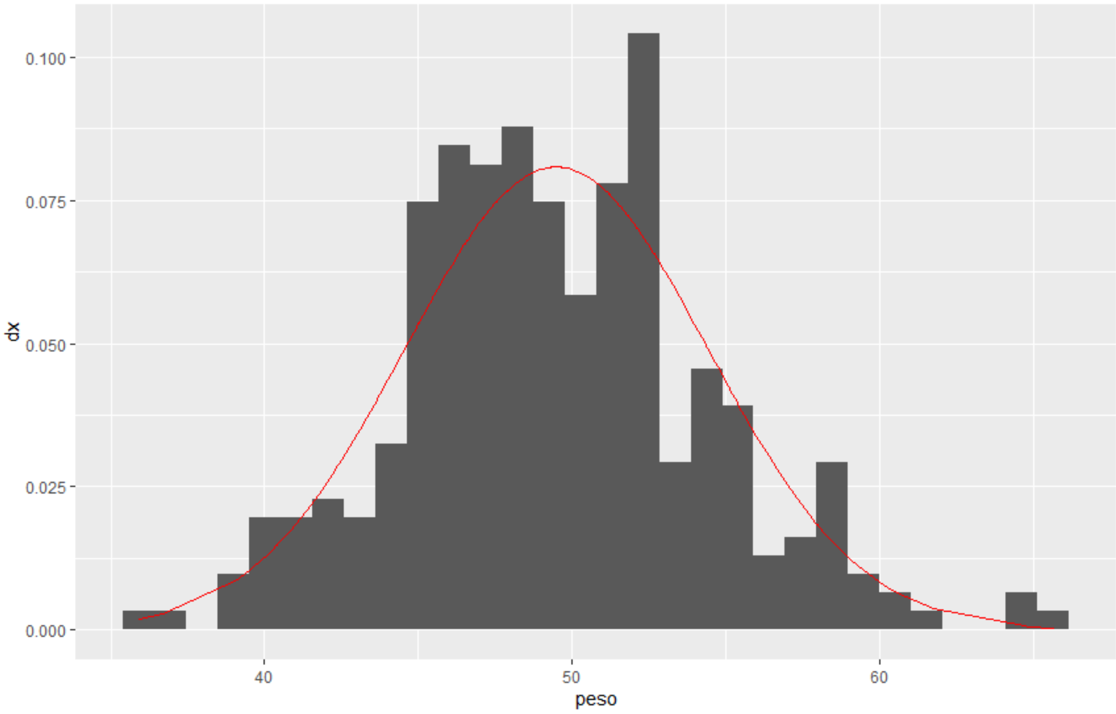
Resposta: Produtos com peso abaixo de 46.90295kg são classificados leves, produtos acima de 52.08065kg são classificados como pesados, e os produtos com peso nesse intervalo são classificados como médios

<u>Códigos:</u>

```
qnorm(0.3, mean(dados3$peso), sd(dados3$peso))
qnorm(0.7, mean(dados3$peso), sd(dados3$peso))
```

e) Apresente graficamente o histograma dos dados juntamente com o modelo escolhido (ajustado) na resposta A).

<u>Resposta:</u>



<u>Códigos:</u>

```
dados3$dx=dnorm(dados3$peso, mean(dados3$peso), sd(dados3$peso))
ggplot(dados3, aes(peso, dx))+geom_histogram(aes(y=..density..))
+geom_line(col='red')
```