

CE301 - Estatística Básica - Prova 1

1o. Semestre 2025

Nome: _____

Data: ____ / ____ / ____ GRR: _____ Assinatura: _____

Um questionário foi aplicado a uma amostra de 10 alunos de um curso de MBA de uma universidade americana, fornecendo as seguintes informações: **love** (1: sem relacionamento estável, 2: relacionamento estável, 3: sentimento profundo de pertencimento e cuidado), **sex** (1: atividade sexual satisfatória, 0: não), **work** (escala de 5 pontos em que 1: sem trabalho, 3: trabalho ok, 5: amo meu trabalho), **money** (renda familiar em milhares de dólares) e **happy** (felicidade numa escala de 10 pontos em que 10 é o mais feliz). Os dados da amostra estão descritos na Tabela 1.

Tabela 1: Dados da amostra

happy	money	sex	love	work
7	90	1	2	2
8	45	1	3	4
2	0	0	2	2
8	35	1	3	3
8	62	0	3	4
6	45	0	2	3
5	70	0	2	3
4	88	1	1	2
7	40	0	2	3
5	56	1	2	3

Com base nos dados, responda as questões de 1 a 11.

Nas respostas use pelo menos 2 casas decimais.

Nos gráficos atente-se para a legenda e escalas.

-
- 1) Quais são os tipos de variáveis coletadas? Classifique-as em qualitativa nominal, qualitativa ordinal, quantitativa discreta e quantitativa contínua. (0,5 ponto)
 - happy: quantitativa contínua
 - money: quantitativa contínua
 - sex: qualitativa nominal
 - love: qualitativa ordinal
 - work: quantitativa continua/qualitativa ordinal
 - 2) Considere que existe um cadastro de alunos alocados em 5 turmas. Primeiramente foi selecionada uma turma por meio de um sorteio em que todas as turmas tinham a mesma probabilidade de fazer parte da amostra.

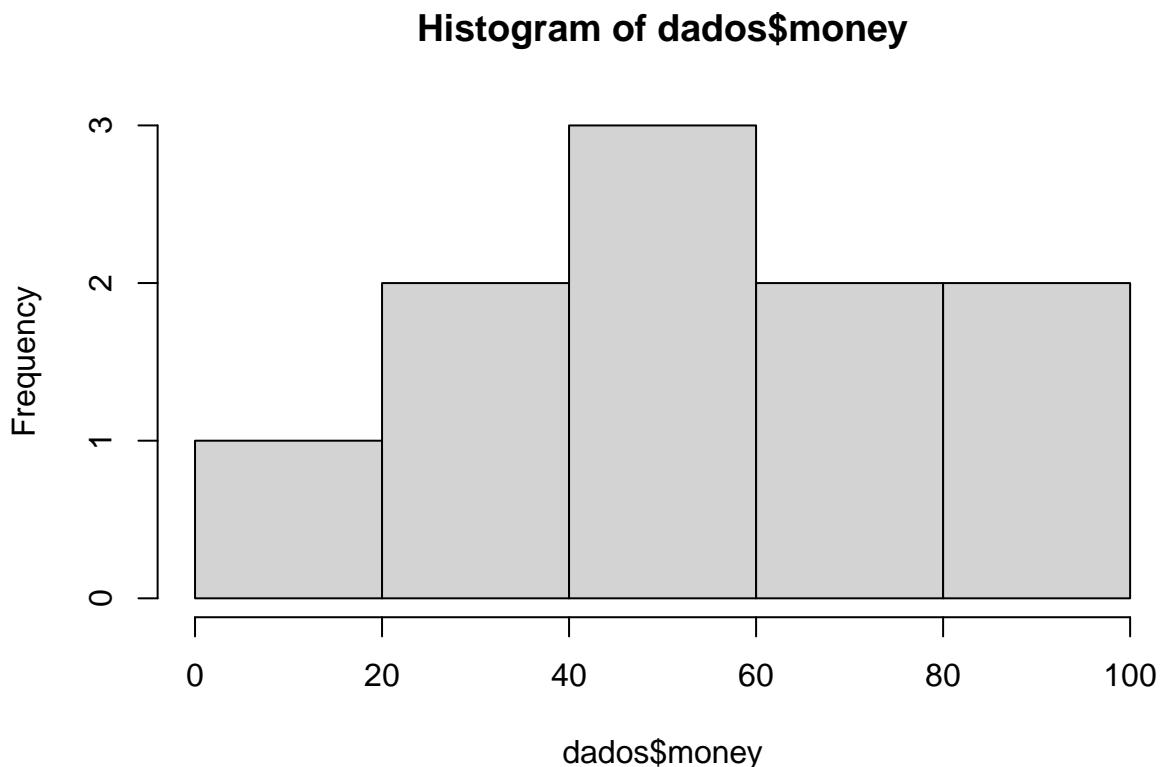
Após o sorteio da turma, uma amostra aleatória de alunos da turma sorteada foi selecionada em que todos os elementos tinham a mesma probabilidade de serem selecionados. Qual o tipo de amostragem utilizada? Este plano de amostragem corresponde a um método probabilístico ou não probabilístico? Justifique sua resposta. (0,5 ponto)

Amostragem por conglomerados: método probabilístico.

- 3) Monte uma tabela de frequências para a variável **work**. Use frequências absolutas e relativas. Qual seria o gráfico mais adequado para representar esta tabela? (0,5 ponto)

work	f	fr
2	3	0.3
3	5	0.5
4	2	0.2

- 4) Monte uma tabela de frequências para a variável **money**. Use faixas de tamanho 20, partindo de 0 até 100. Qual é a faixa modal? (0,5 ponto)



- 5) Obtenha média e desvio-padrão das variáveis **money** e **happy**. (1 ponto)

```
mean(dados$money)
```

```
## [1] 53.1
```

```
mean(dados$happy)
```

```
## [1] 6
```

- 6) As variáveis **happy** e **money** estão em diferentes escalas, qual delas apresenta maior variabilidade? Utiliza uma medida de comparação adequada. (1 ponto)

```
sd(dados$happy)/mean(dados$happy)
```

```
## [1] 0.3333333
```

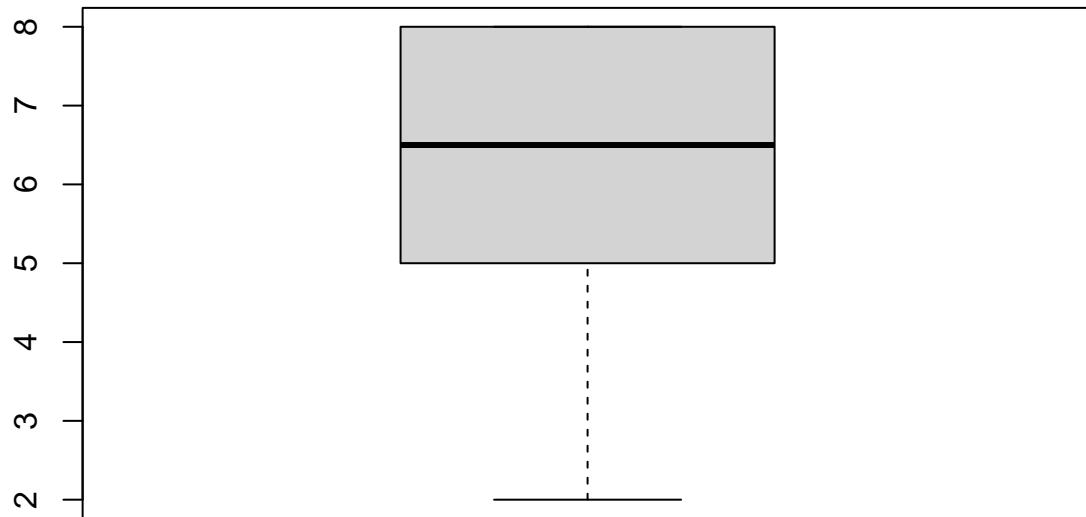
```
sd(dados$money)/mean(dados$money)
```

```
## [1] 0.5023109
```

- 7) Com base na tabela do item (4), esboce o histograma da variável **money**. O que você conclui a respeito da simetria? (1 ponto)

- 8) Obtenha as quantidades necessárias e esboce o box-plot da variável **happy**. Coloque nos eixos os valores utilizados para o esboço. O que você conclui a respeito da simetria e da presença de valores atípicos? (1 ponto)

```
boxplot(dados$happy)
```



```
summary(dados$happy)
```

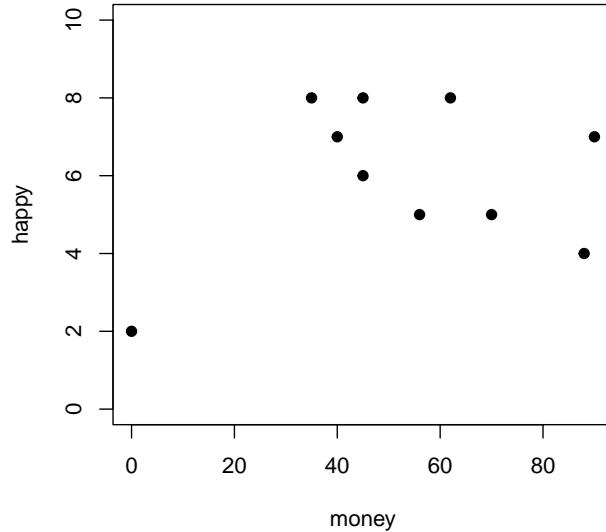
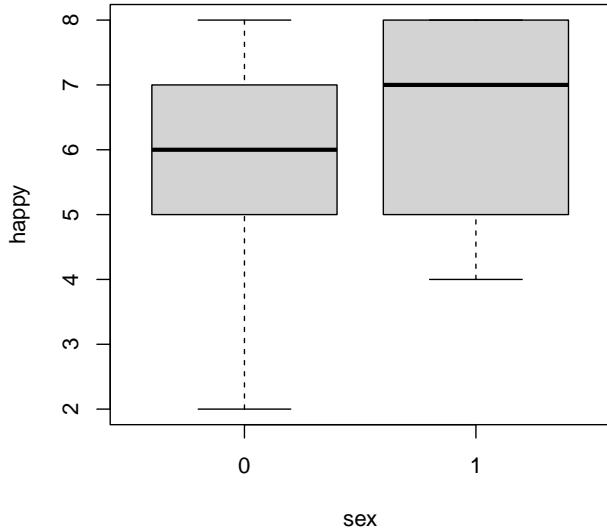
```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      2.00    5.00   6.50    6.00   7.75    8.00
```

- 9) Monte uma tabela de dupla entrada usando frequências absolutas para **love** e **sex**. O que você conclui? (1 ponto)

```
table(dados$love,dados$sex)
```

```
##  
##      0 1  
##  1 0 1  
##  2 4 2  
##  3 1 2
```

- 10) Avalie os gráficos abaixo. O que você conclui a respeito da relação entre as variáveis **happy** e **money**? (1 ponto)



- 11) Obtenha uma medida de associação entre **sex** e **love**. O que você conclui? (1 ponto)

```
chisq.test(table(dados$love,dados$sex))
```

```
## Warning in chisq.test(table(dados$love, dados$sex)): Aproximação do  
## qui-quadrado pode estar incorreta
```

```
##  
## Pearson's Chi-squared test  
##  
## data: table(dados$love, dados$sex)  
## X-squared = 2, df = 2, p-value = 0.3679
```

- 12) Responda de forma sucinta: (1 ponto)

- Qual a diferença entre amostragem casual simples e amostragem sistemática?
- Por que é melhor evitar o gráfico de setores?
- Uma medida de tendência central é suficiente para representar uma variável? Explique.
- O que é um ponto atípico ou outlier?
- O que o coeficiente de correlação expressa? Comente suas características.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^k f_i \cdot y_i}{\sum_{i=1}^k f_i}$$

$$A = \max(y) - \min(y)$$

$$DAM_{média} = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}|$$

$$DAM_{mediana} = \frac{1}{n} \sum_{i=1}^n |y_i - md|$$

$$s^2 = Var(y) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right) \quad s = \sqrt{s^2}$$

$$CV = 100 \cdot \frac{s}{\bar{y}} \quad z = \frac{y_i - \bar{y}}{s} \quad H = - \sum_{i=1}^S f_i \cdot ln(f_i) \quad Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$Cov(y_1, y_2) = \frac{\sum_{i=1}^n (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{n-1} \quad r = \frac{\sum_{i=1}^n (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (y_{1i} - \bar{y}_1)^2} \cdot \sqrt{\sum_{i=1}^n (y_{2i} - \bar{y}_2)^2}} = \frac{Cov(y_1, y_2)}{\sqrt{Var(y_1)} \cdot \sqrt{Var(y_2)}}$$
