

# 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: representa solidão e isolamento, 2: representa um conjunto de relacionamentos seguros, 3: sentimento profundo de pertencimento e cuidado no contexto de alguma família ou comunidade), **sex** (1: atividade sexual satisfatória, 0: atividade sexual não satisfatória), **work** (escala de 5 pontos em que 1: procurando trabalho, 3: trabalho é “ok”, 5: adora o trabalho), **money** (renda familiar em milhares de dólares) e **happy** (felicidade medida em uma escala de 10 pontos, sendo 1: estado suicida, 5: sensação de “apenas levando a vida” e 10: estado eufórico). 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

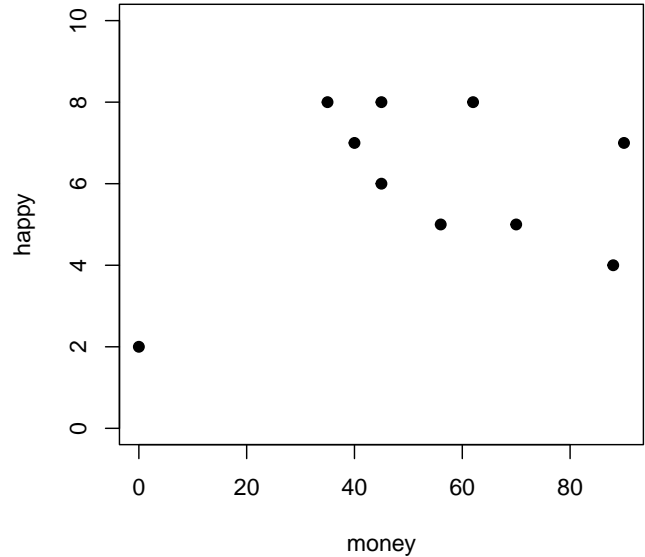
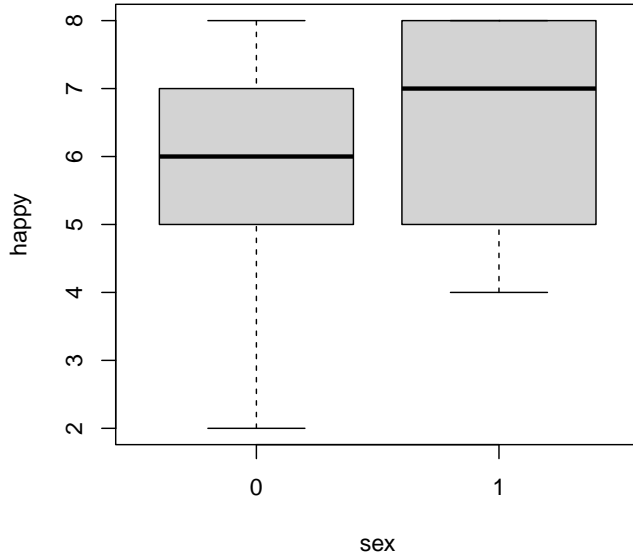
Com base nos dados, responda as questões de 1 a 11.

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)
- 2) Considere que exista 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)
- 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)
- 4) Monte uma tabela de frequências para a variável **money**. Use faixas de tamanho 20, partindo de 40 até 180. Qual é a faixa modal? (0,5 ponto)
- 5) Obtenha média e desvio-padrão das variáveis **money** e **happy**. (1 ponto)

- 6) As variáveis **happy** e **money** estão em diferentes escalas, qual delas apresenta maior variabilidade? Utiliza uma medida de comparação adequada. (1 ponto)
- 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)
- 9) Monte uma tabela de dupla entrada usando frequências absolutas para **love** e **sex**. O que você conclui? (1 ponto)
- 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)
- 12) Responda de forma sucinta: (1 ponto)
  - a) Qual a diferença entre amostragem casual simples e amostragem sistemática?
  - b) Por que é melhor evitar o gráfico de setores?
  - c) Uma medida de tendência central é suficiente para representar uma variável? Explique.
  - d) O que é um ponto atípico ou outlier?
  - e) O que o coeficiente de correlação expressa? Comente suas características.

---


$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \bar{y} = \frac{\sum_{i=1}^k f_i \cdot y_i}{\sum_{i=1}^k f_i} \quad A = \max(y) - \min(y)$$

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

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

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

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


---



1) Quais são os tipos de variáveis coletadas? Classifique-as em qualitativa nominal, qualitativa ordinal, quantitativa discreta e quantitativa contínua.

love  $\rightarrow$  qualitativa ordinal  
 sex  $\rightarrow$  qualitativa nominal  
 work  $\rightarrow$  qualitativa ordinal  
 money  $\rightarrow$  quantitativa contínua  
 happy  $\rightarrow$  qualitativa ordinal

2) Considere que exista 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.

Amostragem por conglomerados, pois as turmas são os conglomerados. Inicialmente, sortamos os conglomerados e dentro da turma sorteada, sortamos os elementos. É probabilístico, pois todos os elementos têm chance de compor a amostra.

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?

work	Frequência absoluta	Frequência relativa	Frequência absoluta acumulada	Frequência relativa acumulada
1	0	$0/10 = 0$	0	0
2	3	$3/10 = 0,30$	3	0,3
3	5	$5/10 = 0,50$	8	0,8
4	2	$2/10 = 0,20$	10	1
5	0	$0/10 = 0$	10	1
TOTAL	10	1		

Gráfico de barras

4) Monte uma tabela de frequências para a variável money. Use faixas de tamanho 20, partindo de 40 até 180. Qual é a faixa modal?

money	Frequência absoluta	Frequência relativa	Frequência abs. acumulada	Frequência rel. acumulada
$[0, 20)$	1	$1/10 = 0,1$	1	0,1
$[20, 40)$	1	$1/10 = 0,1$	2	0,2
$[40, 60)$	4	$4/10 = 0,4$	6	0,6
$[60, 80)$	2	$2/10 = 0,2$	8	0,8
$[80, 100]$	2	$2/10 = 0,2$	10	1
Total	10	1,00		

A faixa modal é a  $[40, 60)$ .

5) Obtenha média e desvio-padrão das variáveis money e happy ②

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

$$\frac{1}{10-1} \left( \sum_{i=1}^{10} y_i^2 - \frac{(\sum_{i=1}^{10} y_i)^2}{10} \right) = \frac{1}{9} \left[ 90^2 + 45^2 + 0^2 + 35^2 + 62^2 + 45^2 + 70^2 + 88^2 + 40^2 + 56^2 - \frac{(90+45+0+35+62+45+70+88+40+56)^2}{10} \right] = \frac{1}{9} (34599 - 28196,1)$$

$$= \frac{1}{9} (6402,9) = \boxed{711,43} \rightarrow \sqrt{s^2} = \sqrt{711,43} = \boxed{26,672}$$

↳ variância ↳ desvio-padrão

$$\bar{x} = \frac{\sum_{i=1}^n y_i}{n} = \frac{(90+45+0+35+62+45+70+88+40+56)}{10} = \boxed{53,1}$$

↳ média

happy

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

$$\frac{1}{10-1} \left( \sum_{i=1}^{10} y_i^2 - \frac{(\sum_{i=1}^{10} y_i)^2}{10} \right) = \frac{1}{9} \left[ (7^2 + 8^2 + 2^2 + 8^2 + 8^2 + 6^2 + 5^2 + 4^2 + 7^2 + 5^2) - \frac{(7+8+2+8+8+6+5+4+7+5)^2}{10} \right] = \frac{1}{9} (396 - 360) = \frac{1}{9} (36) = \boxed{4}$$

↳ variância

$$\rightarrow \sqrt{s^2} = \sqrt{4} = \boxed{2}$$

↳ desvio padrão

$$\bar{x} = \frac{\sum_{i=1}^n y_i}{n} = \frac{(7+8+2+8+8+6+5+4+7+5)}{10} = \boxed{6}$$

↳ média

6) As variáveis happy e money estão em diferentes escalas, qual delas apresenta maior variabilidade?

money

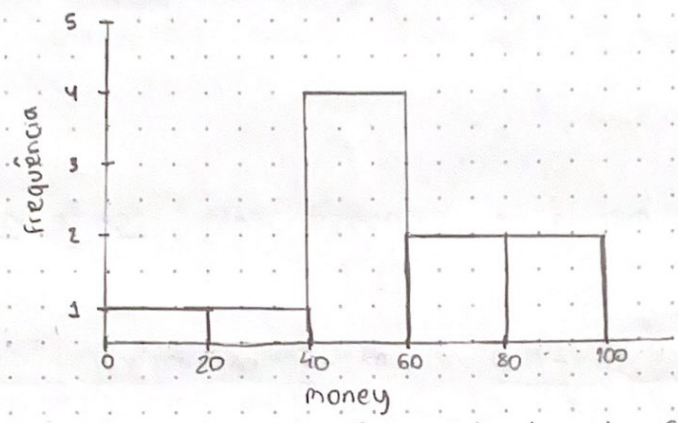
$$CV = 100 \cdot \frac{s}{\bar{y}} = 100 \cdot \frac{26,67}{53,1} = 50,22\%$$

happy

$$CV = 100 \cdot \frac{s}{\bar{y}} = 100 \cdot \frac{2}{6} = 33,33\%$$

A variável money apresenta maior variabilidade, pois seu coeficiente de variação é maior que da variável happy.

7) Com base do item (4), esboce o histograma da variável money. O que você conclui a respeito da simetria? (3)



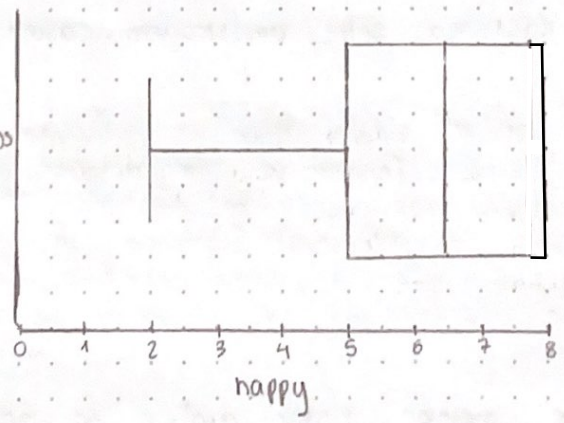
Pelo gráfico, percebe-se que ela é aproximadamente simétrica. Mas de acordo com a regra de medidas de tendência central, a distribuição de money é assimétrica à direita, pois:

$$45 < 50,5 < 53,1$$

(moda)      (mediana)      (média)

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?

Não há valores atípicos e os dados são assimétricos à esquerda, ou seja, estão concentrados à direita.



9) Monte uma tabela de dupla entrada usando frequências absolutas para as variáveis love e sex. O que você conclui?

	sex		
love	①	②	Total
①	1	0	1
②	2	4	6
③	2	1	3
Total	5	5	10

Conclusão  
 Não parece ter associação clara entre as variáveis, pois nota-se que a maioria das pessoas que não têm atividade sexual satisfatória estão no nível 2 da variável love. E pessoas com nível 1 da variável love, têm atividade sexual satisfatória.

- 10) Observa-se uma leve tendência de quem ganha mais dinheiro, é mais feliz. (4)  
 No Boxplot observa-se que a mediana da variável happy é maior em pessoas que têm atividade sexual satisfatória do que em quem não tem.

11) Obtenha uma medida de associação entre sex e love. O que você conclui?

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \quad e_{ij} = \frac{\text{Total Linha} \times \text{Total Coluna}}{\text{Total Geral}}$$

Célula (Love, sex)	Frequência ( $O_{ij}$ )	valor esperado ( $e_{ij}$ )	$(O_{ij} - e_{ij})^2 / e_{ij}$
1, 0	0	$(1 \times 5) / 10 = 0,5$	$(0 - 0,5)^2 / 0,5 = 0,5$
1, 1	1	$(1 \times 5) / 10 = 0,5$	$(1 - 0,5)^2 / 0,5 = 0,5$
2, 0	4	$(6 \times 5) / 10 = 3$	$(4 - 3)^2 / 3 = 0,33$
2, 1	2	$(6 \times 5) / 10 = 3$	$(2 - 3)^2 / 3 = 0,33$
3, 0	1	$(3 \times 5) / 10 = 1,5$	$(1 - 1,5)^2 / 1,5 = 0,17$
3, 1	2	$(3 \times 5) / 10 = 1,5$	$(2 - 1,5)^2 / 1,5 = 0,17$
Total	10	10	$\chi^2 \approx 2$

Há uma associação moderada entre as variáveis love e sex.

12) a) Qual a diferença entre amostragem casual simples e amostragem sistemática?

Na amostragem casual simples, todos os elementos da população têm a mesma chance de compor a amostra. Escolhe-se um determinado número de elementos de forma independente da população. Já na amostragem sistemática, todos os elementos dentro do intervalo apresentam a mesma probabilidade de serem selecionados. Inicialmente, define-se um número  $a$  sorteia-se um elemento a cada número de unidades escolhido.

b) Por que é melhor evitar gráfico de setores?

Porque se for uma variável com muitos níveis, o gráfico ficará visualmente poluído e pouco informativo e níveis com frequências iguais a 0 deixam de aparecer no gráfico, diferente de um gráfico de barras.

c) Uma medida de tendência central é suficiente para representar uma variável? Explique

Não. Pois conjuntos de dados diferentes podem ter a mesma medida de tendência central. Mas um conjunto de dados pode ser mais disperso e outro mais concentrado. Por isso, usamos as medidas de posição relativa e medidas de dispersão para descrever a variabilidade.

d) O que é um ponto atípico ou outlier? (5)

Um outlier é um elemento que não se comporta como o restante dos dados. Pode indicar um erro ou um caso excepcional na amostra.

e) O que o coeficiente de correlação expressa? Comente suas características.

O coeficiente de correlação é usado para identificar se há relação linear entre as variáveis quantitativas. Ele assume valores entre  $-1$  e  $1$ .

Se o valor for maior que  $0$ , existe uma relação linear positiva.

Se o valor for menor que  $0$ , existe uma relação linear negativa.

Se o valor for igual a  $0$ , não existe associação linear. É o

fato de existir correlação entre duas variáveis não significa que uma tem influência sobre a outra.