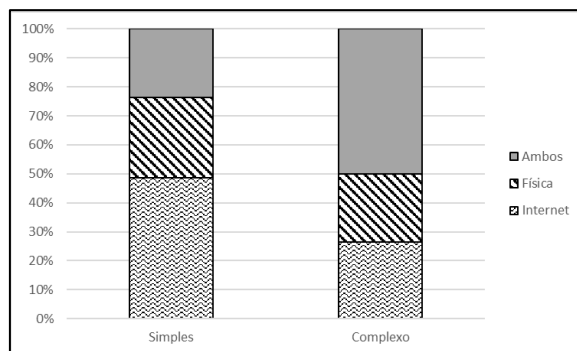


**Questão 1 (2 pontos)**

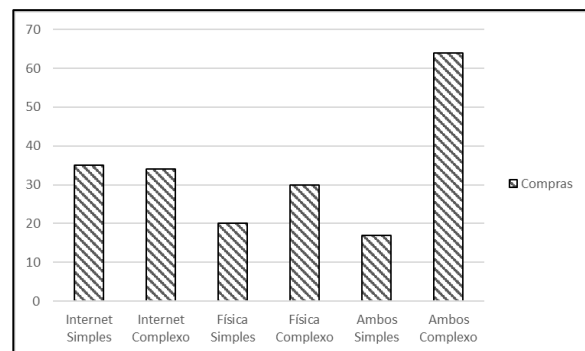
Uma pesquisa com consumidores sobre hábitos de compra buscou relacionar o tipo de produto/serviço comprado (simples ou complexo, na visão do consumidor) ao meio utilizado (exclusivamente pela Internet, exclusivamente em loja física, ou em ambos os meios). Para isso, entrevistaram-se consumidores sobre suas últimas compras. Os resultados obtidos estão disponíveis na planilha q1 do arquivo dadosA1.xlsx, e reproduzidos na tabela abaixo:

		Meio de compra		
		Internet	Física	Ambos
Tipo de produto	Simples	35	20	17
	Complexo	34	30	64

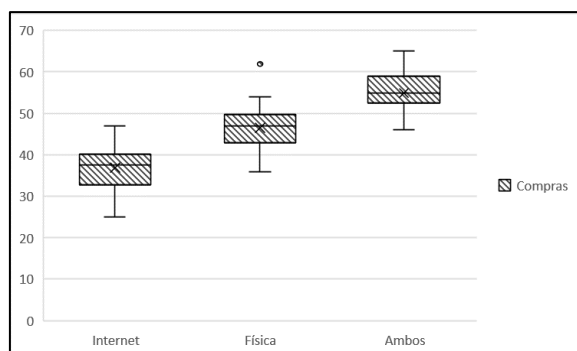
- a) (**½ ponto**) Qual a proporção de pessoas que compraram produtos simples exclusivamente pela Internet?
- b) (**½ ponto**) Das pessoas que usaram a Internet (exclusivamente ou não), qual a proporção que comprou produtos complexos?
- c) (**1 ponto**) Há evidência de associação entre o tipo de produto e o meio de compra? Justifique com cálculos e conceitos.
- d) (**½ ponto adicional/opcional**) Qual dos gráficos abaixo melhor expressa visualmente os resultados apresentados? Justifique.



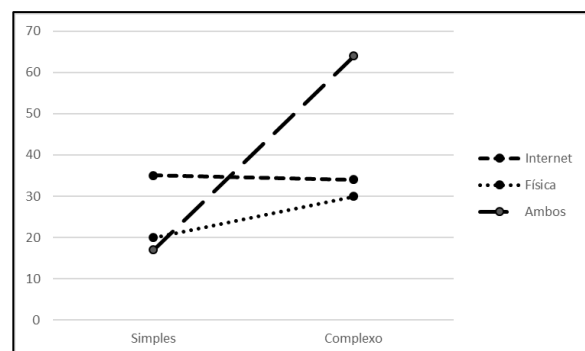
(a)



(b)



(c)



(d)

## Questão 2 (8 pontos)<sup>1</sup>

A empresa ACME Ltda. é um grande fornecedor mundial de equipamentos elétricos e eletrônicos para clientes industriais. Recentemente, contratou uma pesquisa com seus clientes, cujos dados estão disponíveis na planilha q2 do arquivo `dadosA1.xlsx`. O conjunto de dados consiste de 100 observações sobre 14 variáveis separadas em três tipos de informações:

- Atributos que influenciam as escolhas dos clientes (**x1 a x7**), constituindo notas dadas pelos clientes entrevistados em avaliação sobre a ACME;
- Características das compras atuais dos clientes com a ACME (**x8 e x9**);
- Características das empresas clientes que responderam à pesquisa (**x10 a x14**).

Os dados disponíveis devem proporcionar à ACME uma melhor compreensão acerca das características de seus clientes assim como da relação entre suas percepções sobre a ACME e suas ações comerciais (satisfação e volume de compra). O significado de cada atributo está descrito na planilha `q2-dicionario` do arquivo `dadosA1.xlsx`.

Responda as perguntas abaixo, deixando explícitos os cálculos no arquivo `dadosA1.xlsx` (sugestão: **crie uma planilha para cada item**).

- a) **(2 pontos)** Faça uma análise exploratória da variável **x9** (**Satisfaction Level**). Isso inclui o cálculo de:
- i. (1 ponto) histograma (gráfico) de frequências relativas com 8 intervalos de amplitude 0,5 iniciando no valor 3,0;
  - ii. (½ ponto) medidas resumo (média, desvio padrão, mínimo, Q1, mediana, Q3 e máximo);
  - iii. (½ ponto) *box-plot* (explicitando o cálculo dos limites superior e inferior).
- b) **(½ ponto)** Calcule a média (**m**) e o desvio padrão (**dp**) da variável **x8** (**Usage Level**). Usando esses valores, crie uma variável chamada **x8c** (**Usage Level Class**), correspondente à classificação da variável **x8** (**Usage Level**) em três níveis (“alto”, “médio” e “baixo”) definidos pelos valores **m+0,5dp** e **m-0,5dp**.
- c) **(1,5 pontos)** Faça uma análise exploratória conjunta das variáveis **x9** (**Satisfaction Level**) e **x8c** (**Usage Level class**). Isso inclui o cálculo de:
- i. **(1 ponto)** medidas resumo (as mesmas do item a-ii) por categoria;
  - ii. **(½ ponto)** *box-plot* por categoria (um único gráfico com as três categorias).
- d) **(2 pontos)** Calcule o **R<sup>2</sup>** e avalie o grau de associação entre as variáveis **x8c** (**Usage Level class**) e **x9** (**Satisfaction Level**).
- e) **(1,5 pontos)** A diretoria da ACME quer avaliar a associação entre a variável **x9** (**Satisfaction Level**) e as variáveis **x1** (**Speed Delivery**) e **x2** (**Price Level**) separadamente. Para isso, calcule o coeficiente de correlação dos dois pares de variáveis (**x9-x1** e **x9-x2**) e conclua sobre a associação entre as variáveis em cada caso. NÃO UTILIZE as funções **COVARIACÃO.P**, **COVARIACÃO.S**, e **CORREL** do Excel (calcule “na marra”, pelas fórmulas apropriadas).

$$CORR(x, y) = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{dp(x)} \right) \left( \frac{y_i - \bar{y}}{dp(y)} \right) = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}};$$

$$COV(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- f) **(½ ponto)** Utilizando os resultados do item anterior, há evidência de que a satisfação do cliente (medida pela variável **x9**) seja causada preponderantemente por alguma das duas variáveis analisadas (**x1** ou **x2**)? Explique.

<sup>1</sup> Adaptado para uso do conjunto de dados disponível para download em [https://mvstats.com/wp-content/uploads/2022/01/Multivariate\\_Data\\_Analysis\\_6e\\_Datasets.zip](https://mvstats.com/wp-content/uploads/2022/01/Multivariate_Data_Analysis_6e_Datasets.zip)