

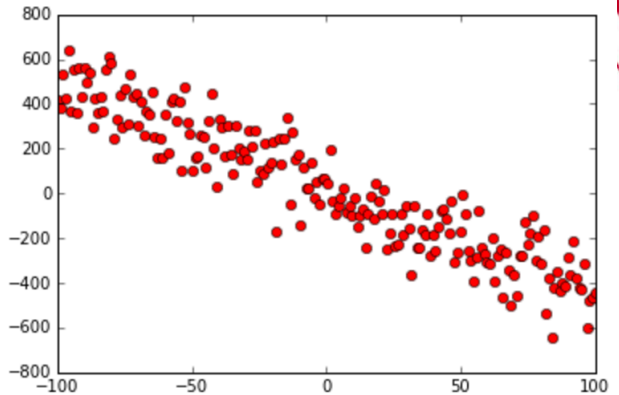
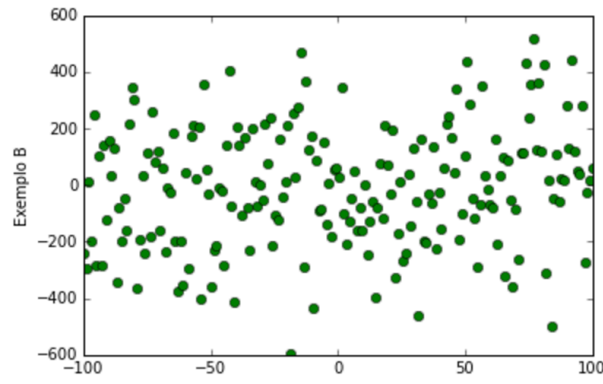
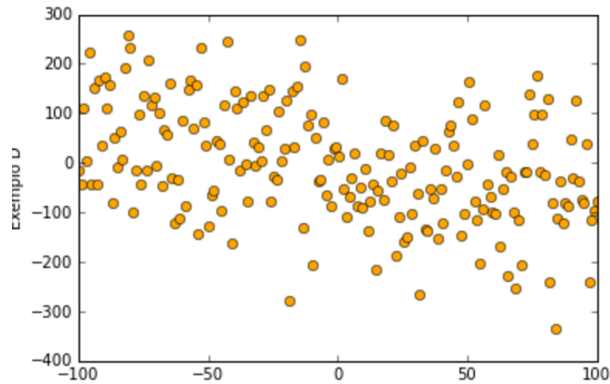
# **Ciência dos Dados**

## **Aula 08**

**Questões para peer learning**

**Assuntos: Revisão e Probabilidade**

# Questão 0



O coeficiente de correlação  $\rho$  para os três gráficos é:

A. 0.25, 0.10, 0.85

B. -0.39, 0.15, -0.93

C. -680.0, 59.0, -1400.0

D. 0.10, 0.05, -0.60

# Questão 1 – Interpretação Professora

Uma professora leciona sua disciplina em três turmas e calculou algumas medidas resumo expressas na Tabela 1.

**Tabela 1.** Algumas medidas resumo para as variáveis *Grau de atenção (nota de 0 a 10)* e *Número de faltas*, segundo cada *Turma*.

Grau de atenção	Medidas	A	B	C
	Média	7,54	7,52	7,53
	Desvio padrão	1,84	1,85	1,86
	Assimetria	Simetria	Direita	Esquerda
Número de faltas	Média	3,26	3,27	3,25
	Desvio padrão	0,51	0,53	0,52
	Assimetria	Direita	Esquerda	Simetria

Com os resultados da Tabela 1 desta questão, avalie se a seguinte interpretação é verdadeira ou falsa:

“Para a amostra em questão, há indícios de que possivelmente é mais interessante dar aulas na turma A se for pelo *Grau de atenção*; entretanto, a turma C tem mais comprometimento com a disciplina por faltarem menos nas aulas”.

## Questão 1

Grau de atenção	Medidas	A	B	C
	Média	7,54	7,52	7,53
	Desvio padrão	1,84	1,85	1,86
	Assimetria	Simetria	Direita	Esquerda

Número de faltas	Média	3,26	3,27	3,25
	Desvio padrão	0,51	0,53	0,52
	Assimetria	Direita	Esquerda	Simetria

Interpretação é verdadeira ou falsa:

“Para a amostra em questão, há indícios de que possivelmente é mais interessante dar aulas na turma A se for pelo *Grau de atenção*; entretanto, a turma C tem mais comprometimento com a disciplina por faltarem menos nas aulas”.

- A) Verdadeira, pois a média me convence.
- B) Falsa, pois a variabilidade não me convence.
- C) Verdadeira, pois a simetria me convence.
- D) Falsa, pois a simetria não me convence.

## Questão 2 – Variabilidade

Se tivermos os desvios padrão (dp) de duas variáveis quantitativas X e Y iguais a:

$$\text{dp}(X) = 0,05 \text{ litros} \quad \text{dp}(Y) = 0,05 \text{ litros}$$

Podemos dizer que a variabilidade nessas duas variáveis são:

- A) Iguais já que ambas variáveis são em litros.
- B) Diferentes apenas quando as médias são diferentes.
- C) Iguais mesmo quando as unidades de medidas são diferentes.
- D) Diferentes mesmo que as médias sejam iguais.

## Questão 3 – Reta ajustada

Filtrando os aposentados na PNAD 2011 e, ainda, restringindo ser aposentados do sexo masculino, com 53 a 90 anos de idade, aposentados pelo sistema público de previdência e residentes em áreas urbanas.

Para os aposentados que **não** se declararam economicamente ativos, a reta ajustada considerando Idade ( $w$ ) e Renda familiar ( $y$ ) é dada por:

$$\hat{y} = 4.668,00 - 25,19 w.$$

E, para os aposentados economicamente ativos, esta mesma relação é representada pela reta:

$$\hat{y} = 6.741,30 - 55,13 w.$$

- A) Aposentados economicamente ativos ganharam, em média, mais do que os aposentados não economicamente ativos, não importando a idade.
- B) Aposentados economicamente ativos ganharam, em média, mais do que os aposentados não economicamente ativos, para idades altas.
- C) Aposentados economicamente ativos ganharam, em média, mais do que os aposentados não economicamente ativos, para idades baixas.
- D) Nada se pode dizer sobre a associação entre as variáveis Renda familiar e Idade.

## Questão 4 – Probabilidade...

A Netflix considera diversas variáveis antes de sugerir uma lista de filmes a um usuário. Entre elas, considere o período que usuário acessa a conta (Diurno e Noturno) e, por consequência, o gênero sugerido ao usuário (Romance e Ação).

A tabela abaixo apresenta as **frequências absolutas** que relacionam essas duas variáveis.

Período	Gênero	
	Romance	Ação
Diurno	35	15
Noturno	45	105

## Questão 4 – Probabilidade

A tabela abaixo apresenta as **frequências absolutas** que relacionam essas duas variáveis.

Período	Gênero	
	Romance	Ação
Diurno	35	15
Noturno	45	105

Assumindo D: período diurno e R: gênero romance, é correto afirmar:

- A)  $P(D \cap R) = 22,5\%$
- B)  $P(D^c \cap R^c) = 52,5\%$
- C)  $P(R^c) = 40\%$
- D)  $P(D) = 2 * P(D^c)$



## Questão 5 – Probabilidade

A tabela abaixo apresenta as **frequências absolutas** que relacionam essas duas variáveis.

Período	Gênero	
	Romance	Ação
Diurno	35	15
Noturno	45	105

É correto afirmar:

- A) No período noturno, a probabilidade de sugerir Romance é 30%.
- B) Se filme sugerido é Ação, a probabilidade de ter sido acessado a Noite é 10,5%
- C) A probabilidade de filme sugerido ser Romance e período acessado ser Diurno é 35%
- D) A probabilidade do filme não ser acessado no período Noturno é 75%

## Questão 6 – Probabilidade

A tabela abaixo apresenta as **probabilidade conjuntas** e as **probabilidades marginais**.

Período	Gênero		Total
	Romance	Ação	
Diurno	0,175	0,075	0,250
Noturno	0,225	0,525	0,750
Total	0,400	0,600	1,000

Assumindo D: período diurno e R: gênero romance, é incorreto afirmar:

- A)  $P((D \cup R)^c) = 47,5\%$
- B)  $1 - P(R^c) = 40\%$
- C)  $P(R|D) = 70\%$
- D)  $P(D \cup R) = 17,5\%$

# Questão 7 – Probabilidade

Considerando uma amostra de estudantes de engenharia, a tabela abaixo apresenta as **probabilidade conjuntas** e as **probabilidades marginais**, por gênero e se assiste filmes?

	Gênero		Total
	Masculino	Feminino	
Assiste Filmes?			
Sim	42%	18%	60%
Não	28%	12%	40%
Total	70%	30%	100%

É correto afirmar:

- A) As mulheres são minoria no Brasil.
- B) É fácil encontrar uma estudante de engenharia mulher e que não assiste filmes.
- C) Assistir filme é independente do gênero.
- D) Entre os homens, eles gostam mais de assistir filmes do que entre as mulheres.