

**Atenção:** 1. Apresentar sempre a hipótese nula e alternativa, a estatística teste e respectiva distribuição, a região de rejeição.

2. Quem tiver usado o EXCEL para os cálculos tem, no final de gravar o ficheiro no Desktop com nome (1º e último) e nº aluno

Questão aberta:

1. Uma análise histórica à duração de filmes de longa metragem indica que 9% têm uma duração inferior a 90 minutos, 36% têm uma duração entre 90 e 120 minutos, 52% têm uma duração entre 120 e 180 minutos e 3% duram mais 180 minutos. Uma amostra de 100 filmes de longa metragem produzidos em 2019 forneceu os dados apresentados na tabela abaixo:[35 pontos]

	A1	A2	A3	A4
Duração	≤ 90	(90, 120]	(120, 180]	> 180
Nº filmes	6	51	36	7

Teste se a duração dos filmes produzidos em 2019 segue o padrão histórico?

$$H_0: p_{01} = 0.09, p_{02} = 0.36, p_{03} = 0.52, p_{04} = 0.03$$

	A1	A2	A3	A4
Duração	≤ 90	(90, 120]	(120, 180]	> 180
Nº filmes ( $f_{obs}$ )	6	51	36	7
$p_{0j}$	0.09	0.36	0.52	0.03
Nº filmes ( $f_{esper.}$ )	9	36	52	3
$q_{j=}$	1	6.25	2.618	

Estatística teste:  $Q = \sum_{j=1}^4 \frac{(n_j - n \cdot p_{0j})^2}{n \cdot p_{0j}} \sim \chi^2_{(4-1)}$  como  $3 < 5$  para que o teste seja

válido tem de se agregar as duas últimas classes pelo que se tem:

$$Q = \sum_{j=1}^3 \frac{(n_j - n \cdot p_{0j})^2}{n \cdot p_{0j}} \sim \chi^2_{(3-1)}$$

$$q_{obs} = \frac{(6 - 9)^2}{9} + \frac{(51 - 36)^2}{36} + \frac{[(36 + 7) - 55]^2}{55} = 9.868$$

valor.  $p = P(Q > q_{obs}) = P(\chi^2_{(2)} > 9.868) = 0.0072 < 0.01$  Rejeita-se  $H_0$  e conclui-se que a duração dos filmes produzidos em 2019 não segue o padrão histórico.

2. O dono de um ginásio pretende saber se existe relação entre as modalidades escolhidas pelos seus clientes e as suas idades. Para isso seleccionou aleatoriamente alguns clientes e registou a idade e modalidade praticada de cada um deles na tabela que se segue:[20 pontos]

Idade	Modalidade			$N_i$
	Musculação	Pilates	Dança	
35 ou menos	150	150	100	400
Mais de 35	150	250	200	600
$N_j$	300	400	300	1000

O que pode concluir sobre a independência entre as preferências e a idade a um nível de significância de 1%

$$H_0: p_{ij} = p_i \cdot p_j \quad \forall (i, j) \quad i, = 1, 2; j = 1, 2, 3 \quad \text{contra} \quad H_0: \exists p_{ij} \neq p_i \cdot p_j$$

$$\text{Estatística t este: } Q = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(n_{ij} - n \cdot p_i \cdot p_j)^2}{n \cdot p_i \cdot p_j} \sim \chi^2_{\frac{(2-1) \cdot (3-1)}{2}}$$

Tabela dos  $\hat{p}_{ij}$ :

Idade	Modalidade			$\hat{p}_i$
	Musculação	Pilates	Dança	
35 ou menos	0.12	0.16	0.12	0.4
Mais de 35	0.18	0.24	0.18	0.6
$\hat{p}_j$	0.3	0.4	0.3	

Data	
Level of Significance	0.01
Number of Rows	2
Number of Columns	3
Degrees of Freedom	2

Results	
Critical Value	9.21034
Chi-Square Test Statistic	19.09722
p-Value	7.13E-05
<b>Reject the null hypothesis</b>	

Observed Frequencies				
Row variable	Column variable			Total
	C1	C2	C3	
R1	150	150	100	400
R2	150	250	200	600

Expected Frequencies				
Row variable	Column variable			Total
	C1	C2	C3	
R1	120	160	120	400
R2	180	240	180	600
Total	300	400	300	1000

Pelo que se conclui que a modalidade escolhida não é independente da idade