UNIVERSIDADE CATÓLICA PORTUGUESA

FACULDADE DE CIÊNCIAS ECONÓMICAS E EMPRESARIAIS

LICENCIATURAS EM ECONOMIA E GESTÃO

ESTATÍSTICA I

1ª Frequência – 27 de Outubro de 2008

APRESENTE TODOS OS CÁLCULOS EFECTUADOS
JUSTIFIQUE TODAS AS SUAS RESPOSTAS

I (6 valores)

Um estabelecimento comercial recebeu, num dia, 200 clientes que aí efectuaram compras. Os montantes dessas compras (em euros) distribuem-se de acordo com a tabela abaixo apresentada.

Valor das Compras (em euros)	N.º de Clientes
] 0 -20]	10
] 20 – 40]	50
] 40 – 80]	80
] 80 – 120]	40
] 120 – 160]	20
	200

a) Elabore o histograma e o polígono de frequências absolutas da distribuição da variável "montante das compras";

b) *i)* Calcule a média, a mediana e a moda desta variável;
ii) No contexto deste problema, interprete cada um dos valores calculados;

c) *i)* A partir das medidas de tendência central determinadas em b), estude a assimetria da distribuição da variável;
ii) No contexto concreto deste problema, interprete o resultado.

d) Calcule a variância, o desvio-padrão e o coeficiente de variação da variável "montante das compras".

e) Neste estabelecimento, os sacos de plástico são distribuídos gratuitamente aos clientes. Estudos anteriores sugerem que, em média, por cada 10 euros de compras, é utilizado um saco. Se o custo de cada saco para o estabelecimento for de 1 cêntimo, qual a despesa suportada com a distribuição dos sacos de plástico? Comente a exactidão deste valor.

II (6 valores)

Foi efectuado um inquérito junto de 1000 indivíduos, entre os 11 e os 35 anos, sobre a utilização do *e-mail* com a seguinte pergunta "*quantos dias por semana consultam o seu e-mail?*". As respostas, de acordo com a respectiva idade, foram as seguintes:

Variável Y (N.º de Dias)	Variável X (Idade, em anos)			
	[11 – 15]	] 15 – 19]	] 19 – 23]	] 23 – 35]
] 0-1]	2	10	24	12
] 1 – 2]	4	16	15	12
] 2 - 5]	14	24	21	16
] 5 – 7]	80	150	240	360

a) *i)* Compare a dispersão da variável X (idade) com a dispersão da variável Y (número de dias por semana).
ii) Justifique a escolha do indicador utilizado.

b) Considere os indivíduos que consultam o *e-mail* 2 ou menos vezes por semana.
i) Calcule a idade média destes indivíduos;
ii) Calcule a percentagem de indivíduos cuja idade está acima da média.

c) Considere a amostra dividida em 2 grupos: o grupo dos indivíduos até aos 19 anos e o grupo dos indivíduos mais de 19 anos.
i) Compare o número médio de dias por semana em que o *e-mail* é consultado pelos indivíduos de cada um destes grupos etários;
ii) Com base nos resultados de *i)*, o que pode concluir sobre a independência estatística entre as variáveis X e Y? Justifique.

d) Suponha que o coeficiente de correlação linear de Pearson entre as variáreis X e Y é igual a 0,35. Como interpreta, por palavras, o valor deste indicador?

III (5 valores)

Para estudar a evolução do preço dos bens de consumo numa determinada economia, estes foram agrupados em 3 grandes categorias (A, B e C).
O quadro abaixo sintetiza a informação disponível para cada uma das classes de bens:

- Preços entre 2003 e 2006;
- Montante total de despesa efectuada pelas famílias em 2003 e 2004;
- Estrutura da despesa das famílias em 2005 e 2006.

a) *i)* Analise a evolução dos **preços** durante este período através do cálculo dos índices de preços relativos, agregados e ponderados, para cada um dos anos, utilizando como base o ano de 2004;
ii) Como se designa o índice que calculou?

b) Discuta a **reversibilidade em relação ao tempo** do índice calculado em a) ilustrando para os anos de 2004 e 2005.

c) A partir de 2006 procedeu-se a uma **mudança de base do índice**, passando 2006 a ser o novo ano base. Em 2007, o índice, calculado na nova base tem o valor de 105. Usando esta informação e os valores dos índices que calculou em a), determine a taxa de crescimento média anual dos preços entre 2003 e 2007. Apresente e justifique todos os cálculos que efectuar.

Preços

Bens	2003	2004	2005	2006
A	12	15	15	25
B	110	149	155	200
C	370	450	600	800

Despesa das Famílias

	Despesa Total		Estrutura da Despesa (em %)	
Bens	2003	2004	2005	2006
A	3600	4200	16%	18%
B	15950	20860	61%	58%
C	7400	8100	23%	23%
Total	26950	33160	100%	100%

IV
~~III~~ (3 valores)

Atrás do México e da Turquia

Portugal no topo das desigualdades da OCDE

21.10.2008 - 09h00 Lusa

*Um estudo da **Organização para a Cooperação e o Desenvolvimento Económico (OCDE)** mostra que Portugal é um dos países com maiores desigualdades na distribuição de rendimentos pelos cidadãos, em paralelo com os Estados Unidos e apenas atrás da Turquia e do México. No seu relatório **"Crescimento e Desigualdades"** divulgado hoje, citado pela Lusa, os autores deste estudo colocaram a Dinamarca e a Suécia à frente dos países mais justos, com um **coeficiente** de 0,23, e o México a liderar a tabela dos mais injustos (0,47), seguido da Turquia (0,43) e de Portugal e dos Estados Unidos (ambos com 0,38).*

(Nota: O coeficiente a que o estudo faz referência é o índice de Gini)

O quadro abaixo ilustra a distribuição de rendimento numa economia hipotética.

i) Calcule o valor do Índice de Gini para esta economia;

ii) Utilizando como referência a informação contida no excerto da notícia publicada no jornal Público como avalia a desigualdade na distribuição do rendimento da economia ilustrada no quadro?

Classes de Rendimento	N.º de Indivíduos
0- 100	100
100 – 500	150
500 - 1000	200
Total	450

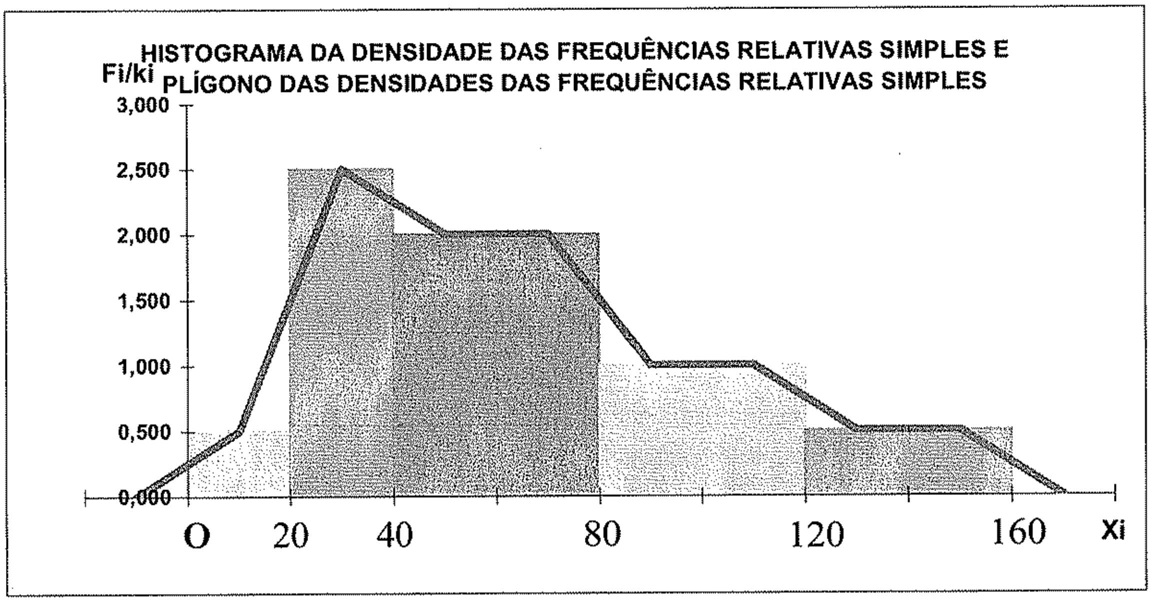
ESTATÍSTICA I - 1º TESTE - 27 DE OUTUBRO DE 2008

É apresentada a resolução de referência. Outras formas de responder correctas foram consideradas correctas. Quando são apresentadas resoluções alternativas, só uma é que foi exigida.

Grupo I

li	*Li*	*xi*	*Ki*	*Fi*	*fi*	*Fi/Ki*	*fi/Ki*	*Si*	*si*	*xiFi*
0	20	10	20	10	0,05	0,500	0,00250	10	0,050	100
20	40	30	20	50	0,25	2,500	0,01250	60	0,300	1500
40	80	60	40	80	0,40	2,000	0,01000	140	0,700	4800
80	120	100	40	40	0,20	1,000	0,00500	180	0,900	4000
120	160	140	40	20	0,10	0,500	0,00250	200	1,000	2800
Soma				200	1,00					13200
Divisão po n										66

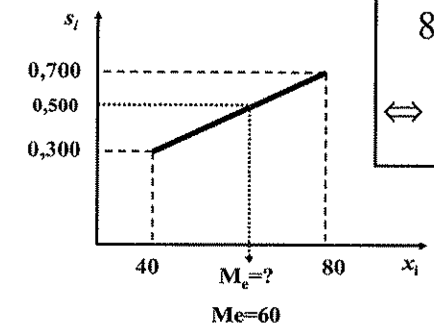
a)



b) i)

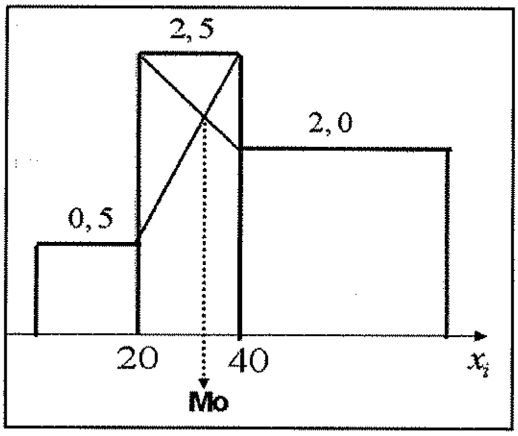
$$\bar{X}=\frac{\sum_{i=1}^{m} x_i F_i}{n}=\frac{13200}{200}=66 \quad euros$$

Classe onde se acumulam 50% das observações:



$$\frac{0,70-0,30}{80-40}=\frac{0,50-0,30}{M_e-40}\Leftrightarrow$$

$$\Leftrightarrow M_e=40+\frac{0,50-0,30}{0,40}\times 40=60 \quad euros$$



Fórmula de King:

$$\frac{2,5-2,0}{40-M_o}=\frac{2,5-0,5}{M_o-20}\Leftrightarrow$$

$$\Leftrightarrow M_o=20+\frac{(2,5-0,5)}{(2,5-0,5)+(2,5-2,0)}\times 20=$$

$$=36 \quad euros$$

b) ii) Média: Há um total de 13200 euros em compras. Se igualmente pelos 200 clientesa dava 66 euros por cliente.

Mediana: Metade dos clientes fez 60 euros ou menos de compras e metade fez 60 euros ou mais de compras.

Moda: O valor de compras com maior ocorrência é 36 euros.

c) i) $M_o \leq M_e \leq \bar{X}$ Trata-se de uma assimetria positiva ou enviesamento à esquerda.

c) ii) A maior parte das ocorrências é de compras de valor mais baixo.

d)

li	*Li*	*xi*	*Fi*	$\left(x_i - \overline{X}\right)^2 Fi$	$x_i^{\ 2} Fi$
0	20	10	10	31360	1000
20	40	30	50	64800	45000
40	80	60	80	2880	288000
80	120	100	40	46240	400000
120	160	140	20	109520	392000
Soma			200	254800	1126000
Divisão por *n*				1274	5630

Variância:

$$s_x^{\ 2} = \frac{1}{n}\sum_i F_i\left(x_i - \overline{x}\right)^2 = \frac{254800}{200} = 1274$$

Desvio-Padrão:

$$s_x = +\sqrt{s_x^{\ 2}} = +\sqrt{1274} = 35,693$$

Coeficiente de Variação:

$$C_v = \frac{s_x}{\overline{x}} = \frac{35,693}{66} = 0,5408 \quad ou \quad 54,08\%$$

e)

$$\sum_{i=1}^{m} x_i F_i = 13200 \qquad euros$$

Total de receitas de compras para a empresa.

$$\frac{13200}{10} = 1320 \quad \text{sacos}$$

$$1320\,\text{sacos} \times 0,01 = 13,2 \qquad \text{euros}$$

Não é um valor exacto uma vez que estamos a trabalhar com o centro da classe.

Grupo II

$F(x_i;y_j)$			Variável *X* [Idade em anos]				$F(y_j)$
			[11-15[	[15-19[	[19-23[	[23-35]	
			$x_1=13,0$	$x_2=17$	$x_3=21,0$	$x_4=29,0$	
Variável Y [dias por semana]	] 0 - 1]	$y_1=0,5$	2	10	24	12	48
	] 1 - 2]	$y_2=1,5$	4	16	15	12	47
	] 2 - 5]	$y_3=3,5$	14	24	21	16	75
	] 5 - 7]	$y_4=6,0$	80	150	240	360	830
$F(x_i)$			100	200	300	400	1000

a) i)

Variável X

x_i	F_i	f_i	x_iF_i	x_if_i	$x_i-\overline{X}$	$(x_i-\overline{X})^2F_i$	$x_i^2F_i$
13	100	0, 100	1300	1, 300	-9, 600	9216, 000	16900
17	200	0, 200	3400	3, 400	-5, 600	6272, 000	57800
21	300	0, 300	6300	6, 300	-1, 600	768, 000	132300
29	400	0, 400	11600	11, 600	6, 400	16384, 000	336400
Total	1000	1, 000	22600	22, 600	--	32640, 000	543400
Divisão por *n*			22,600			32, 640	543,400
Raiz quadrada						5,713	

$$Coeficiente\ de\ \text{var}\,iação\ de\ X=\frac{S_x}{\overline{X}}=\frac{5,713}{22,6}=\quad 0,2527939$$

Variável Y

y_j	F_j	f_j	y_jF_j	y_jf_j	$y_j-\overline{Y}$	$(y_j-\overline{Y})^2F_j$	$y_j^2F_j$
0,5	48	0, 048	24	0, 024	-4, 837	1123, 035	12
1,5	47	0, 047	70,5	0, 071	-3, 837	691, 961	105,75
3,5	75	0, 075	262,5	0, 263	-1, 837	253, 093	918,75
6,0	830	0, 830	4980	4, 980	0, 663	364, 842	29880
Total	1000	1, 000	5337	5, 337	--	2432, 931	30916,5
Divisão por *n*			5,337			2, 433	30,917
Raiz quadrada						1,560	

$$Coeficiente\ de\ \text{var}\,iação\ de\ Y=\frac{S_Y}{\overline{Y}}=\frac{1,560}{5,337}=\quad 0,2922589$$

A dispersão relativa da variável Y é superior à dispersão relativa da variável X, sendo a dispersão medida em ambos os casos a partir do coeficiente de variação.

a) ii)

A comparação das dispersões tem de ser feita com base no coeficiente de variação porque este é uma medida relativa de dispersão. Deve ser utilizado pois trata-se de variáveis relativas a fenómenos diferentes, além do mais expressos em unidades diferentes.

Nota: Não se pode usar o desvio-padrão para fazer a comparação porque é uma medida absoluta de dispersão, expressa nas mesmas unidades do fenómeno em estudo.

b) i)

Variável X condicionada a Y pertencer ao intervalo]0-2]

Classe	x_i	F_i	f_i	x_iF_i	$S_i(\leq)$	$s_i(\leq)$
[11-15[	13	6	0, 063	78	6	0, 063
[15-19[	17	26	0, 274	442	32	0, 337
[19-23[	21	39	0, 411	819	71	0, 747
[23-35]	29	24	0, 253	696	95	1, 000
	Total	95	1, 000	2035		
Divisão por *n*				21,421		

$$\bar{X} = \frac{x_iF_i}{n} = \frac{2035}{95} = 21,421$$

b) ii)

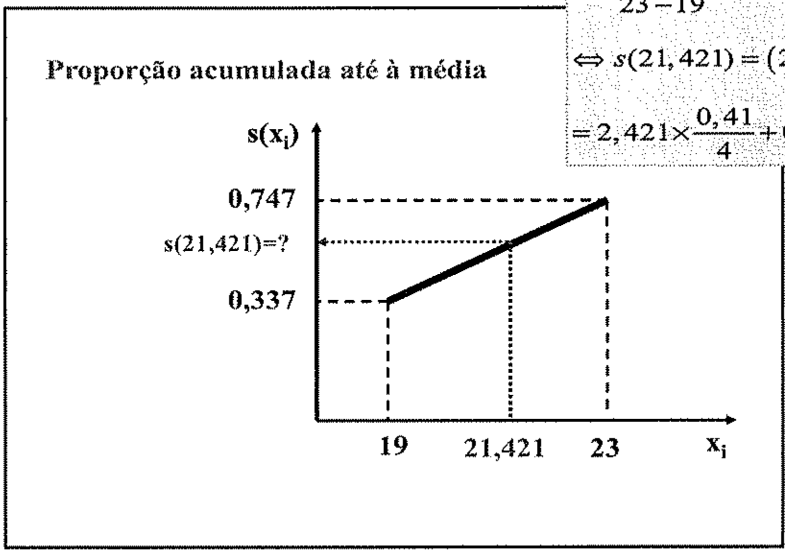
PERCENTAGEM DE OBSERVAÇÕES SITUADAS ACIMA DA MÉDIA:

$$1 - s(\bar{X}) = 1 - s(21,421)$$

$$\frac{0,747-0,337}{23-19} = \frac{s(21,421)-0,337}{21,421-19} \Leftrightarrow$$

$$\Leftrightarrow s(21,421) = (21,421-19)\times\frac{(0,747-0,337)}{(23-19)} + 0,337 =$$

$$= 2,421\times\frac{0,41}{4} + 0,337 = 0,585$$



$$1 - s(\bar{X}) = 1 - s(21,421) = 1 - 0,585 = 0,415 \quad ou \quad 41,5\%$$

c) i)

Variável Y condicionada a X pertencer ao intervalo [11-19]

Classe	y_j	F_j	f_j	y_jF_j	y_jf_j
]0 - 1]	0,5	12	0, 040	6	0, 020
]1 - 2]	1,5	20	0, 067	30	0, 100
]2 - 5]	3,5	38	0, 127	133	0, 443
]5 - 7]	6,0	230	0, 767	1380	4, 600
Total		300	1, 000	1549	5, 163
Divisão por *n*				5,163	

$$\bar{Y} = \frac{y_jF_j}{n} = \frac{1549}{300} = 5{,}163$$

Variável Y condicionada a X pertencer ao intervalo]19-35]

Classe	y_j	F_j	f_j	y_jF_j	y_jf_j
]0 - 1]	0,5	36	0, 051	18	0, 026
]1 - 2]	1,5	27	0, 039	40,5	0, 058
]2 - 5]	3,5	37	0, 053	129,5	0, 185
]5 - 7]	6,0	600	0, 857	3600	5, 143
Total		700	1, 000	3788	5, 411
Divisão por *n*				5,411	

$$\bar{Y} = \frac{y_jF_j}{n} = \frac{3788}{700} = 5{,}411$$

O grupo etário mais velho consulta o email mais dias por semana.

c) ii)

Quando há independência entre X e Y, então,

$$f(y_j \backslash x_1) = f(y_j \backslash x_2) = f(y_j \backslash x_3) = \ldots = f(y_j)$$

pelo que $\bar{Y} = \sum y_j f_j$ é também o valor que se obtém igual para todas as classes de x_i.

Portanto, quando há independência entre X e Y, então $\bar{Y} \backslash x_1 = \bar{Y} \backslash x_2 = \bar{Y} \backslash x_3 = \bar{Y}$, o mesmo acontecendo para qualquer subconjunto em que as frequências relativas ou a média sejam calculadas.

Este resultado não se verifica pelo que X e Y não são independentes.

d)

$r_{X,Y}$=0,35.

* A correlação é positiva, o que significa que as variáveis evoluem no mesmo sentido (quando uma aumenta a outra aumenta e quando uma diminui a outra diminui).

* A correlação é fraca a moderada (o valor está mais próximo de 0 do que de 1)

ESTATÍSTICA I - 1º TESTE - 27 DE OUTUBRO DE 2008

É apresentada a resolução de referência. Outras formas de responder correctas foram consideradas correctas. Quando são apresentadas resoluções alternativas, só uma é que foi exigida.

Grupo III

a) i) **Índice de preços relativos agregados**:

Face aos dados do enunciado, há 2 métodos alternativos para calcular os índices pretendidos.

1º método)

Utilizando a definição geral do índice: $I_{(t/0)}^{(p)} = \left(\sum_{i=1}^{n} \frac{p_{ti}}{p_{0i}} \times w_i \right) \times 100$ em que os ponderadores são a % da despesa em cada um dos bens, no ano base, 2004:

$$w_i = \frac{p_{i04} q_{i04}}{\sum_{i=A,B,C} \left(p_{i04} q_{i04} \right)}$$

w_A = 4200 / (4200+20860+8100) = 0,13
w_B = 20860 / (4200+20860+8100) = 0,63
w_C = 8100 / (4200+20860+8100) = 0,24

$I_{03/04}$ = ((12/15) * 0,13 + (110/149) * 0,63 + (370/450)* 0,24) * 100 = 77
$I_{04/04}$ = 100
$I_{05/04}$ = ((15/15) * 0,13 + (155/149) * 0,63 + (600/450)* 0,24) * 100 = 110,1
$I_{06/04}$ = ((25 / 15) * 0,13 + (200/149) * 0,63 + (800/450)* 0,24) * 100 = 148,7

Ano	$I_{t/04}$
2003	77,0
2004	100,0
2005	110,1
2006	148,7

2º método)

O índice pretendido é o índice de preços de Laspeyers, para o qual conhecemos a fórmula de cálculo simplificada:

$$I_{t/0}^{(p)}(L)=\frac{\sum_{i=1}^{n}p_{ti}q_{0i}}{\sum_{i=1}^{n}p_{0i}q_{0i}}\times 100$$

Para aplicar a fórmula, apenas temos de determinar as <u>quantidades dos bens em 2004</u>, a partir da informação sobre a despesa e os preços:

$Q_A = 4200 / 15 = 280$
$Q_B = 20860 / 149 = 140$
$Q_C = 8100 / 450 = 18$

A despesa total em 2004 é igual a 33160 (= 4200+20860+8100) (é o denominador da fórmula do índice)

$I_{03/04} = ((12*280+110*140+370*18) / 33160) * 100 = 77$
$I_{04/04} = 100$
$I_{05/04} = ((15*280+155*140+600*18) / 33160) * 100 = 110{,}1$
$I_{06/04} = ((25*280+200*140+800*18) / 33160) * 100 = 148{,}7$

b) Reversibilidade em relação ao tempo:

Um índice goza da propriedade da reversibilidade em relação ao tempo se se verificar a seguinte condição:

$$\frac{I_{t/0}}{100}=\frac{100}{I_{0/t}}$$

Utilizando os anos de 2004 e 2005, podemos escrever:

$$I_{05/04}^{(p)}=\frac{\sum_{i=1}^{n}p_{05i}q_{04i}}{\sum_{i=1}^{n}p_{04i}q_{04i}}\times 100 \qquad I_{04/05}^{(p)}=\frac{\sum_{i=1}^{n}p_{04i}q_{05i}}{\sum_{i=1}^{n}p_{05i}q_{05i}}\times 100$$

Então, é fácil de ver que, neste caso, $\frac{I_{05/04}^{(p)}}{100}\neq\frac{100}{I_{04/05}^{(p)}}$, pois

$$\frac{\sum_{i=1}^{n}\left(p_{05i}q_{04i}\right)}{\sum_{i=1}^{n}\left(p_{04i}q_{04i}\right)} \neq \frac{\sum_{i=1}^{n}\left(p_{05i}q_{05i}\right)}{\sum_{i=1}^{n}\left(p_{04i}q_{05i}\right)}$$

Conclusão: O índice não é reversível em relação ao tempo.

Nota: A resposta podia, em alternativa, ser dada calculando valor do índice $I_{04/05}^{(p)}$ e comparando-o com o valor já conhecido de $I_{05/04}^{(p)} = 110,1$.
Para calcular $I_{04/05}^{(p)}$ temos de utilizar directamente a definição do índice:

$I_{(04/05)}^{(p)} = \left(\sum_{i=1}^{n}\frac{p_{04i}}{p_{05i}} \times w_i\right) \times 100$ em que as ponderações são a estrutura da despesa em 2005, disponíveis no enunciado.

$I_{04/05}$ = (15/15*0,16+149/155*0,61+450/600*0,23) * 100 = 91,8

Assim, verifica-se que $\frac{110,1}{100} \neq \frac{100}{91,8}$, logo não há reversibilidade no tempo.

c) $I_{07/06} = 105$

Qual o crescimento médio **anual** dos preços entre 2003 e 2007?
A taxa de crescimento média anual entre 2003 e 2007 (período de 4 anos), que designo por y_{03-07}, pode calcular-se a partir da taxa de crescimento total entre 2003 e 2007, que designo por x_{03-07}, utilizando a média geométrica:

$$1 + y_{03-07} = \sqrt[4]{\left(1 + x_{03-07}\right)}$$

- Sabe-se que a taxa de crescimento **total** entre 2003 e 2007 , representada por x_{03-07}, pode ser obtida a partir do índice de preços de 2007 com base em 2003:

$$x_{03-07} = \frac{I_{07/03}}{100} - 1$$

- Como obter $I_{07/03}$?
Fazendo o encadeamento dos índices, podemos escrever:

$$\frac{I_{07/03}}{100} = \frac{I_{07/06}}{100} \times \frac{I_{06/04}}{100} \times \frac{I_{04/03}}{100}$$

Nesta expressão, conhecemos os valores de:
$I_{07/06} = 105$ (enunciado) e $I_{06/04} = 149$ (alínea a)).
Conhecemos também o valor de $I_{03/04} = 77$ com o qual vamos calcular $I_{04/03}$:

$$\frac{I_{04/03}}{100} = \frac{100}{77} \rightarrow I_{04/03} = 130.$$

(Nota: este é um cálculo aproximado pois já sabemos que não há reversibilidade no tempo para este índice)

Podemos então obter o valor pretendido:

$$\frac{I_{07/03}}{100} = \frac{105}{100} \times \frac{149}{100} \times \frac{130}{100} \rightarrow I_{07/03} = 203$$

$$x_{03-07} = \frac{I_{07/03}}{100} - 1 = \frac{203}{100} - 1 = 1{,}03 \rightarrow \text{Taxa crescimento total entre 2003 e 2007} = 103\%$$

Taxa de crescimento média anual entre 2003 e 2007 = y

$$1 + y = \sqrt[4]{(1 + x_{03-07})} = \sqrt[4]{(1 + 1{,}03)} = 1{,}19$$

Resposta: A taxa de crescimento média anual entre 2003 e 2007 é 19 % ao ano.

Grupo IV

Para calcular o valor do índice de Gini utilizando a fórmula conhecida (e disponível no formulário) começamos por calcular as variáveis *pi* e *qi* em que:
pi representa a frequência relativa acumulada até ao fim da classe *i*
qi representa a % do rendimento total detida pelos indivíduos até ao fim da classe *i*.

$$IG = 1 - \sum_{j=1}^{m} (q_j + q_{j-1}) \times (p_j - p_{j-1})$$

Índice de Gini

Classes de Rendimento	Fi	Xi	pi	pi-pi-1	XiFi	XiFI%	qi	qi+qi-1	
			(= fi)	(A)			XiFi%acum	(B)	(A)*(B)
0- 100	100	50	0,222	0,222	5000	0,025	0,025	0,025	0,006
100 – 500	150	300	0,556	0,333	45000	0,225	0,250	0,275	0,092
500 - 1000	200	750	1,000	0,444	150000	0,750	1,000	1,250	0,556
Total	450			1,000	200000	1,000			0,653

IG=1-0,653=0,347
IG= 0,347

b) O valor do índice de Gini situa-se sempre entre 0 e 1. Quanto mais próximo de 0 estiver o valor do índice mais igualitária é a distribuição do atributo (neste caso rendimento) entre os indivíduos. Um valor muito elevado do índice, em contrapartida, corresponde a uma situação de distribuição pouco igualitária, muito assimétrica do atributo.
No caso da economia hipotética do enunciado, o valor do índice é ligeiramente mais baixo que o valor para Portugal e USA, o que pode ser interpretado como representando uma distribuição do rendimento um pouco menos assimétrica, longe contudo ainda dos valores Dinamarca e da Suécia.