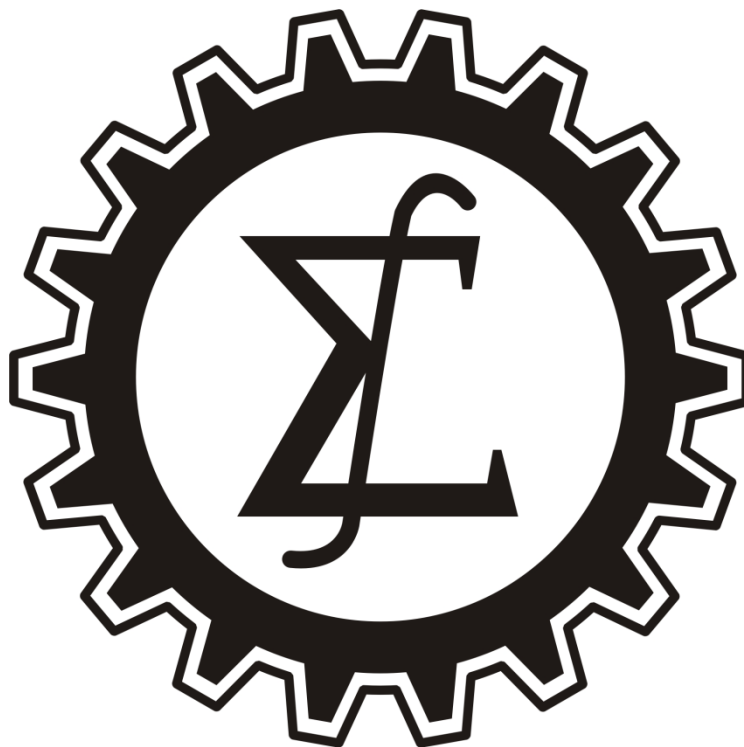


Trabalho prático de Introdução à Probabilidade e Estatística



Realizado:

Sérgio Cláudio Allibhai Nº31500

João Pedro Calhau Nº31621

André Filipe Figueira Nº31626

Luís Filipe Zurrapa Nº32330

Curso:

Engenharia Informática

Data:

20/06/2014

Índice

	Pág.
0.Introdução.....	3
1. Materiais e métodos.....	3
2. Analise dos resultados.....	4
3. Conclusão.....	18

Introdução

No âmbito da disciplina de Introdução a Probabilidade e Estatística foi nos apresentados a opção de realizar um trabalho prático. Sendo escolhida esta opção, foi nos fornecidos uma série de dados pela docente e um enunciado.

Este trabalho prático consiste numa análise de dados sobre a percentagem de pessoas com acesso à internet de pessoas entre os 16 e 74 anos em 3 regiões, sendo essa a região da zona Euro, e num contexto mais detalhado, a região de Portugal e a região da Grécia, isto entre 2002 e 2013, a excepção do ano de 2012, que não nos foi apresentado dados relativos a esse ano.

Para a realização deste trabalho recorreremos aos apoios dos slides das aulas teóricas, apontamentos das aulas práticas e utilização do software IBM SPSS, para conseguir responder as perguntas apresentadas pelo enunciado fornecido.

Materiais e métodos

Materiais

Para realização deste trabalho contamos com o suporte do software IBM SPSS 21, com o apoio dos slides das aulas teóricas bem como os apontamentos das mesmas e aulas praticas, recorreremos também ao auxílio do livro “ Estatística e Probabilidades” de Anabela Afonso e Carla Nunes

Métodos

Para a resolução deste trabalho utilizamos vários métodos, desde o cálculo de medidas de tendência central e não central a assimetrias e achatamentos, incluindo entre estes medidas de dispersão e coeficientes de variação. Foram utilizados testes de hipótese e intervalos de confiança. Foi também necessário comparar médias e fazer regressões lineares, bem como interpretar p-values e testes de normalidade.

Por fim determinaram-se os coeficientes de correlação e construíram-se as retas de regressão dos mínimos quadrados.

Analise dos resultados

1. Neste trabalho está em estudo 3 variáveis sendo elas o Ano: Variável quantitativa discreta (que vai 2002 a 2013 excluindo 2012), a Região: Variável qualitativa nominal (Euro Área, Portugal e Grécia) e o AcessoNet: Variável quantitativa discreta (% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa).

2. De seguida iremos analisar as variáveis, para mais fácil iremos fragmentar o problema em 3 partes regidas pela região:

Euro Área:

Statistics

% de indivíduos, com idades entre os 16
e os 74 anos, com acesso à internet em
casa

N	Valid	11
	Missing	0
Mean		49,27
Std. Error of Mean		4,559
Median		49,00
Mode		26 ^a
Std. Deviation		15,120
Variance		228,618
Skewness		,039
Std. Error of Skewness		,661
Kurtosis		-1,267
Std. Error of Kurtosis		1,279
Range		46
Minimum		26
Maximum		72
Sum		542
	10	27,40
	25	36,00
Percentiles	50	49,00
	75	64,00
	90	71,00

a. Multiple modes exist. The smallest
value is shown

Medidas de tendência central:

Média: A percentagem média da população da Euro Área que teve acesso à Internet em casa entre 2002/2013 com exceção de 2012 foi de 49,27%.

Mediana: Em 50% dos dados observados notou-se que, na região da Euro Área, 49% da população, com idades entre os 16 e 74 anos, teve acesso à internet em casa.

Moda: A moda é o valor que se regista com maior frequência, mas neste caso existem múltiplas modas, sendo o menor valor entre elas 26%.

Medidas de tendência não central:

Quartis: Em 10% dos dados observados notou-se que, na região da Euro Área, 27,40% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 25% dos dados observados notou-se que, na região da Euro Área, 36,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 50% dos dados observados notou-se que, na região da Euro Área, 49,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 75% dos dados observados notou-se que, na região da Euro Área, 64,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 90% dos dados observados notou-se que, na região da Euro Área, 71,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Amplitude inter-quartilica: $Q = Q_{0,75} - Q_{0,25} = 64 - 36 = 28\%$

Intervalo de Variação: $Q' = Q_{0,9} - Q_{0,1} = 71 - 27,40 = 43,60\%$

Medidas de Dispersão:

Variância: Os valores obtidos desviam-se dos valores esperados por 228,618%.

Desvio-Padrão: Os valores obtidos desviam-se da média por 15,120%.

Coeficiente de dispersão: $S/X = 15,120/49,27 = 0,3079$

Coeficiente de variação: $(S/X) * 100\% = 0,307 * 100 = 30,79\%$

Da análise do coeficiente de variação temos que a média é representativa, pois é inferior a 50%.

Medidas de Achatamento:

Achatamento (Kurtosis) = Kurtosis / Std. Error of Kurtosis = -0,9906

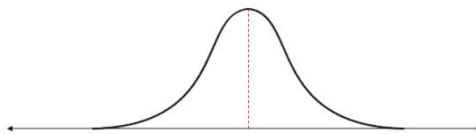
Quanto da medida de achatamento visto que $|-0,9906| < 1.96$ assumimos uma distribuição mesocúrtica.



Medidas de Assimetria:

Assimetria (Skewness) = Skewness / Std. Error of Skewness = 0,0590

Quanto a medida de assimetria visto que $|0.0590| < 1.96$ assumimos uma distribuição simétrica.



Portugal:

Statistics

% de indivíduos, com idades entre os 16
e os 74 anos, com acesso à internet em
casa

N	Valid	11
	Missing	0
Mean		30,82
Std. Error of Mean		4,592
Median		27,00
Mode		12 ^a
Std. Deviation		15,230
Variance		231,964
Skewness		,463
Std. Error of Skewness		,661
Kurtosis		-1,126
Std. Error of Kurtosis		1,279
Range		45
Minimum		12
Maximum		57
Sum		339
	10	12,60
	25	17,00
Percentiles	50	27,00
	75	45,00
	90	55,60

a. Multiple modes exist. The smallest
value is shown

Medidas de tendência central:

Média: A percentagem média da população de Portugal que teve acesso à Internet em casa entre 2002/2013 com exceção de 2012 foi de 30,82%

Mediana: Em 50% dos dados observados notou-se que, na região de Portugal, 27% da população, com idades entre os 16 e 74 anos, teve acesso à internet em casa.

Moda: A moda é o valor que se regista com maior frequência, mas neste caso existem múltiplas modas, sendo o menor valor entre elas 12%.

Medidas de tendência não central:

Quartis: Em 10% dos dados observados notou-se que, na região de Portugal, 12,60% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 25% dos dados observados notou-se que, na região de Portugal, 17,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 50% dos dados observados notou-se que, na região de Portugal, 27,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 75% dos dados observados notou-se que, na região de Portugal, 45,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 90% dos dados observados notou-se que, na região de Portugal, 55,60% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Amplitude inter-quartilica: $Q = Q_{0,75} - Q_{0,25} = 45 - 17 = 28\%$

Intervalo de Variação: $Q' = Q_{0,9} - Q_{0,1} = 55,60 - 12,60 = 43 \%$

Medidas de Dispersão:

Variância: Os valores obtidos desviam-se dos valores esperados por 231,964%.

Desvio-Padrão: Os valores obtidos desviam-se da média por 15,230%.

Coeficiente de dispersão: $S/X = 15,230/30,82 = 0,4942$

Coeficiente de variação: $(S/X) * 100\% = 0,4942 * 100 = 49,42\%$

Da análise do coeficiente de variação temos que a média é representativa, pois é inferior a 50%.

Medidas de Achatamento:

Achatamento (Kurtosis) = Kurtosis / Std. Error of Kurtosis = -0,880

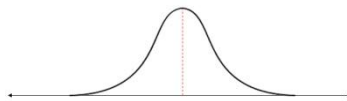
Quanto da medida de achatamento visto que $|-0,880| < 1.96$ assumimos uma distribuição mesocúrtica.



Medidas de Assimetria:

Assimetria (Skewness) = Skewness / Std. Error of Skewness = 0,700

Quanto a medida de assimetria visto que $|0.700| < 1.96$ assumimos uma distribuição simétrica.



Grécia:

Statistics

% de indivíduos, com idades entre os 16
e os 74 anos, com acesso à internet em
casa

N	Valid	11
	Missing	1
Mean		25,73
Std. Error of Mean		4,746
Median		21,00
Mode		8 ^a
Std. Deviation		15,742
Variance		247,818
Skewness		,723
Std. Error of Skewness		,661
Kurtosis		-,514
Std. Error of Kurtosis		1,279
Range		48
Minimum		8
Maximum		56
Sum		283
	10	8,40
	25	12,00
Percentiles	50	21,00
	75	38,00
	90	53,80

a. Multiple modes exist. The smallest
value is shown

Medidas de tendência central:

Média: A percentagem média da população da Grécia que teve acesso à Internet em casa entre 2002/2013 com exceção de 2012 foi de 25,73%

Mediana: Em 50% dos dados observados notou-se que, na região da Grécia, 21% da população, com idades entre os 16 e 74 anos, teve acesso à internet em casa.

Moda: A moda é o valor que se regista com maior frequência, mas neste caso existem múltiplas modas, sendo o menor valor entre elas 8%.

Medidas de tendência não central:

Quartis: Em 10% dos dados observados notou-se que, na região da Grécia, 8,40% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 25% dos dados observados notou-se que, na região da Grécia, 12,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 50% dos dados observados notou-se que, na região da Grécia, 21,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 75% dos dados observados notou-se que, na região da Grécia, 38,00% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Em 90% dos dados observados notou-se que, na região da Grécia, 53,80% da população, com idades entre os 16 e os 74 anos, teve acesso à internet em casa.

Amplitude inter-quartilica: $Q = Q_{0,75} - Q_{0,25} = 38 - 12 = 26\%$

Intervalo de Variação: $Q' = Q_{0,9} - Q_{0,1} = 53,80 - 8,40 = 45.4 \%$

Medidas de Dispersão:

Variância: Os valores obtidos desviam-se dos valores esperados por 247,818%.

Desvio-Padrão: Os valores obtidos desviam-se da média por 15,742%.

Coeficiente de dispersão: $S/X = 15,742/25,73 = 0,6118$

Coeficiente de variação: $(S/X) * 100\% = 0,4942 * 100 = 61,18\%$

Da análise do coeficiente de variação temos que a média não é representativa, pois é superior a 50%.

Medidas de Achatamento:

Achatamento (Kurtosis) = Kurtosis / Std. Error of Kurtosis = -0,402

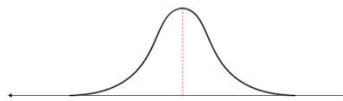
Quanto da medida de achatamento visto que $|-0,402| < 1.96$ assumimos uma distribuição mesocúrtica.



Medidas de Assimetria:

Assimetria (Skewness) = Skewness / Std. Error of Skewness = 1,09

Quanto a medida de assimetria visto que $|1,09| < 1.96$ assumimos uma distribuição simétrica.



3. a) A média e o desvio-padrão antes calculados assumem os mesmos valores que as estimativas pontuais da média e do desvio-padrão.

b)

Descriptives^a

Região		Statistic	Std. Error
% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa	Grecia Mean	25,73	4,746
	95% Confidence Interval for Mean	Lower Bound	15,15
		Upper Bound	36,30
	5% Trimmed Mean	25,03	
	Median	21,00	
	Variance	247,818	
	Std. Deviation	15,742	
	Minimum	8	
	Maximum	56	
	Range	48	
	Interquartile Range	26	
	Skewness	,723	,661
	Kurtosis	-,514	1,279

a. There are no valid cases for % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa when Região = ,000. Statistics cannot be computed for this level.

Descriptives

Região				Statistic	Std. Error
% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa	PORTUGAL	Mean		30,82	4,592
		95% Confidence Interval for Mean	Lower Bound	20,59	
			Upper Bound	41,05	
		5% Trimmed Mean		30,41	
		Median		27,00	
		Variance		231,964	
		Std. Deviation		15,230	
		Minimum		12	
		Maximum		57	
		Range		45	
		Interquartile Range		28	
		Skewness		,463	,661
		Kurtosis		-1,126	1,279

Descriptives

Região				Statistic	Std. Error
% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa	EURO area	Mean		49,27	4,559
		95% Confidence Interval for Mean	Lower Bound	39,11	
			Upper Bound	59,43	
		5% Trimmed Mean		49,30	
		Median		49,00	
		Variance		228,618	
		Std. Deviation		15,120	
		Minimum		26	
		Maximum		72	
		Range		46	
		Interquartile Range		28	
		Skewness		,039	,661
		Kurtosis		-1,267	1,279

Para um intervalo de confiança de 95% para a média tiramos, destas tabelas, que:

Para a Euro área, temos um limite superior de 59,43 e um limite inferior de 39,11

Para Portugal, temos um limite superior de 41,05 e um limite inferior de 20,59.

Para a Grécia, temos um limite superior de 36,30 e um limite inferior de 15,15.

c)

Tests of Normality

Região		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa	EURO area	,139	11	,200 [*]	,963	11	,808
	Grecia	,163	11	,200 [*]	,924	11	,354
	PORTUGAL	,151	11	,200 [*]	,937	11	,486

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H_0 : "As amostras provêm de uma população gaussiana" vs H_1 : "As amostras não provêm de uma população gaussiana"

Rejeitamos H_0 se $\alpha \geq p\text{-value}$

Como as amostras têm um tamanho inferior a 15, iremos interpretar os valores do teste de normalidade segundo Shapiro-Wilk. Segundo o output, podemos reparar que todos os p-value são superiores aos níveis de significância usuais (1%, 5% e 10%) logo podemos admitir que as amostras provem de uma população gaussiana.

d) Euro Área vs Portugal:

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
% de indivíduos com idades entre os 16 e os 74 anos, com acesso à internet em casa	Equal variances assumed	,006	,940	2,852	20	,010	18,455	6,471	4,957	31,952
	Equal variances not assumed			2,852	19,999	,010	18,455	6,471	4,957	31,952

Euro Área vs Grécia:

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
% de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa	Equal variances assumed	,009	,924	3,578	20	,002	23,545	6,581	9,817	37,274
	Equal variances not assumed			3,578	19,968	,002	23,545	6,581	9,816	37,275

Portugal vs Grécia:

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
% de indivíduos, com idades	Equal variances assumed	,001	,980	,771	20	,450	5,091	6,604	- 8,685	18,867

entre os 16			,771	19,978	,450	5,091	6,604	-	18,868
e os 74	Equal							8,686	
anos, com	variances								
acesso à	not								
internet em	assumed								
casa									

Ao analisar as tabelas podemos dizer que, as Variâncias nunca são iguais nos 3 casos, logo analisa mos a segunda facha das tabelas. Assim sendo no Intervalo de confiança a 95% para a igualdade de medias, as tabelas Euro Área vs Grécia e Euro Área vs Portugal não contem o valor 0, logo não pode mos admitir a igualdade das medias, visto que não satisfazem a condição $\mu_1 - \mu_2 = 0$; Por outro lado a tabela Portugal vs Grécia mostra que a 95% de confiança podere mos admitir a igualdade de μ , segundo $\mu_1 - \mu_2 = 0$.

4 -

- Para ajustar 2 variáveis a um modelo linear tem de ser necessariamente o Ano como variável independente (X) e o AcessoNet terá de ser a variável dependente (Y).
- Estimar e explicar os parâmetros

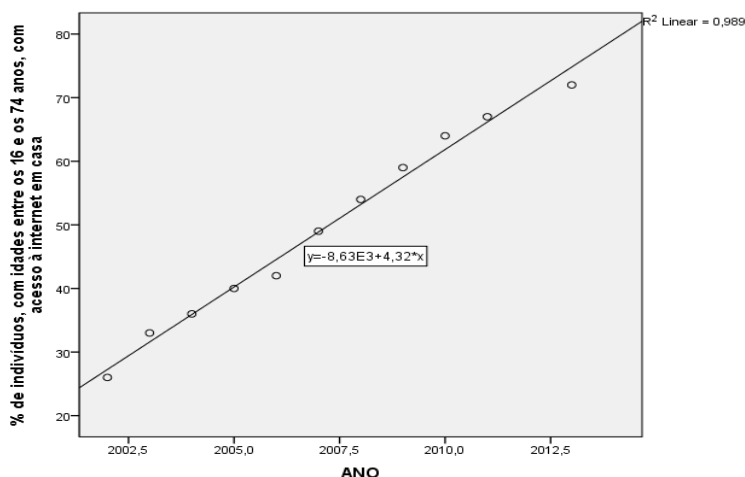
EU

Coefficients^a

Model	Unstandardized Coefficients		Standardize d Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-8628,000	311,855		-27,667	,000
ANO	4,323	,155	,994	27,825	,000

a. Dependent Variable: % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$$Y = 4,323 \cdot X - 8628$$



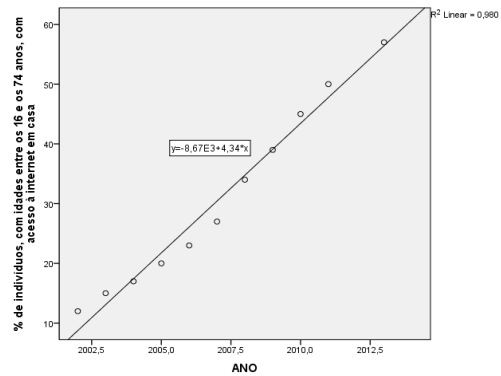
PT

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-8670,600	417,705		-20,758	,000
ANO	4,335	,208	,990	20,832	,000

a. Dependent Variable: % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$$\hat{Y} = 4,335X - 8628$$



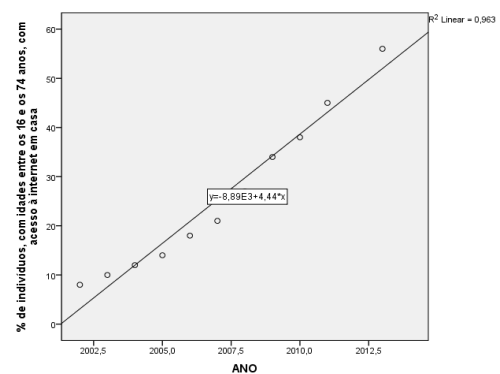
GR

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-8893,000	579,622		-15,343	,000
ANO	4,444	,289	,982	15,387	,000

a. Dependent Variable: % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$$\hat{Y} = 4,444X - 8893$$



c)

os testes de hipóteses consistem em verificar se a recta passa pela origem e se seu declive é nulo ou não, ou seja se existe relação linear,.

H0 = a recta passa pelo declive

H1 = a recta não passa pelo declive

H0: $\beta_0 = 0$ vs H1: $\beta_0 \neq 0$;

Visto que na Euro área, na PT e na GR os p-values do valor de β_0 se aproximam do valor 0, então assim rejeitamos a hipótese nula para qualquer dos níveis de significância usuais (1%, 5% e 10%). Podemos então assim concluir que a recta de regressão linear, para todas as regiões não passa pela origem.

H0 = o avançar do ano não influencia linearmente a percentagem de pessoas com acesso a internet.

H1 = o avançar do ano influencia linearmente a percentagem de pessoas com acesso a internet.

H0: $\beta_1 = 0$ vs H1: $\beta_1 \neq 0$;

Visto que na Euro área, na PT e na GR os p-values dos valores da amostra Ano são aproximadamente zero, então assim rejeitamos a hipótese nula para qualquer dos níveis de significância usuais (1%,5% e 10%). Podemos então assim concluir que existe evidência estatística suficiente para afirma que o avançar do ano influencia linearmente a percentagem de pessoas com acesso a internet.

d)

Euro área

R = 0.994

Significa que tem uma forte associação linear positiva entre o ano e a % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$R^2 = 0.989$

Significa que 98.9 % da variabilidade do avanço do tempo é explicada pela relação linear que possui com a % da população entre os 16 e 74 anos com acesso a internet em casa.

PT

R=0,990

Significa que tem uma forte associação linear positiva entre o ano e a % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$$R^2 = 0.980$$

Significa que 98.0 % da variabilidade do avanço do tempo é explicada pela relação linear que possui com a % da população entre os 16 e 74 anos com acesso a internet em casa.

GR

$$R = 0,982$$

Significa que tem uma forte associação linear positiva entre o ano e a % de indivíduos, com idades entre os 16 e os 74 anos, com acesso à internet em casa

$$R^2 = 0.963$$

Significa que 96.3 % da variabilidade do avanço do tempo é explicada pela relação linear que possui com a % da população entre os 16 e 74 anos com acesso a internet em casa.

e)

Euro área

$$\hat{Y} = 4.323 \cdot 2015 - 8628 \Leftrightarrow Y = 82.845\%$$

PT

$$\hat{Y} = 4.335 \cdot 2015 - 8670.600 \Leftrightarrow Y = 64.425\%$$

GR

$$\hat{Y} = 4.444 \cdot 2015 - 8893.000 \Leftrightarrow Y = 61.66\%$$

Conclusão

Com a realização deste trabalho pretendíamos como aspeto inicial começar a trabalhar com o software SPSS. Para utilizar de maneira mais eficaz este recurso utilizamos os conhecimentos aprendidos ao longo do ano. Tendo em conta isto podemos dizer que os objetivos que nos foram propostos para este trabalho foram cumpridos e que neste momento temos pelo menos as noções básicas deste software preparando-nos para qualquer necessidade no futuro.

Na resolução deste trabalho notou-se preocupação por parte das variâncias das nossas amostras em estudo, pois este valor era valores muito altos, significando que existe grande dispersão nos dados recolhidos.