

Tractor casa

no futebol

Paulo Alexandre Moura Rebello

Porto, Outubro 2005

O Factor casa no Futebol

Introdução

O futebol é, actualmente, uma importante indústria. Os clubes de topo gerem orçamentos avultados.

Sendo que os jogadores são profissionais e recebem muito acima da média da população e as regras são uniformes para todos os jogos, seria de esperar que as equipas atingissem iguais performances no seu estádio ou noutra campo qualquer.

Mas a análise estatística prova que isto não acontece. Existe um fenómeno chamado “factor casa” que faz com que as equipas obtenham melhores resultados quando jogam no seu estádio.

Este trabalho pretende medir a importância dessa variável no futebol. As causas, essas, terão que ser dadas provavelmente pela Psicologia (em concreto a Psicologia do desporto)...

Índice

Relatório _____ **[3-6]**

Análise dos dados _____ **[7-24]**

1ª Parte _____ [8-15]

2ª Parte _____ [16-19]

3ª Parte _____ [20-24]

Dados _____ **[25-38]**

Relatório

O estudo

O factor casa não afecta de igual modo todas as equipas. Por isso o estudo está dividido em três partes.

1ª parte

É analisado o rácio entre os pontos ganhos em casa e o total de pontos obtidos para cada uma das equipas.

Pretende-se provar que o factor casa existe e demonstrar a sua importância na performance das equipas.

2ª parte

É analisado a forma como a variável %Pontos em Casa se relaciona com uma segunda variável: o total de pontos obtidos.

Pretende-se demonstrar que as melhores equipas estão menos “dependentes” do factor casa para pontuar.

3ª parte

São comparadas duas amostras com o objectivo de analisar a forma como os diferentes instrumentos estatísticos utilizados reagem a variações e reforçar as teorias defendidas.

Amostra

Neste estudo serão analisadas as 5 últimas edições dos quatro campeonatos nacionais de futebol mais importantes* da Europa.

*Sendo subjectivo, considera-se como sendo as equipas portuguesas, inglesas, espanholas e italianas aquelas que melhores resultados têm obtido nas provas Europeias.

Organização do estudo

O estudo está assim organizado:

“*Texto livre*” para enunciar formulas ou simplesmente os valores encontrados.

“ <i>Texto in-box</i> ” para expor interpretações em relação aos valores encontrados.

Método utilizado

Recolhas da amostra

nos seguintes sítios da Internet:

- Federação Portuguesa de Futebol
<http://www.fpf.pt/>
- Liga Italiana de Futebol Profissional
<http://www.lega-calcio.it>
- FIFA
<http://www.fifa.com/en>
- Liga Inglesa de Futebol Profissional
<http://www.premierleague.com>
- ZeroZero – Base de dados
<http://www.zerozero.pt>
- Jornal A bola
<http://www.abola.pt>
- Liga de Futebol Profissional Espanhola
<http://www.lfp.es/>
- Acção local de estatística aplicada
<http://alea-estp.ine.pt/>
- A foto utilizada na primeira página é do estádio do Manchester United

Análise dos dados

através do Excel com o auxílio do Dossier IV – Estatística com Excel do ALEA.

Folhas utilizadas:

- Folha principal com a análise de 384 observações (20 campeonatos de 4 países).
- Folha só com campeões com a análise de 71 observações de todos os campeões em Portugal

Variáveis utilizadas:

- Rácio entre pontos em casa e pontos no total
- Total de Pontos
- Rácio entre pontos fora e pontos no total
- Diferença entre pontos em casa e pontos fora (como cálculo auxiliar)

Análise dos dados

1ª Parte

A importância do factor casa

Provar que o “Factor casa” existe através da análise de medidas de localização

2004/2005	Jogos	1) Pontos C	2) Pontos F	3) Total	4) %PontosC	5) C-F
Juventus	38	47	39	86	55%	8
Milan	38	38	41	79	48%	-3

Começando então por analisar a amostra, para cada classificação final foram calculados os Pontos obtidos em casa 1), os Pontos obtidos fora 2) e o Total de pontos 3) (1+2). A variável analisada é a %Pontos em Casa 4) que se obtém pelo rácio dos pontos em casa por o total de pontos (1/3).

$$\left(\frac{\text{Pontos.em.Casa}}{\text{Total.de.Pontos}} \right)$$

Por simplificação a variável irá designar-se simplesmente por “%PC”.

A coluna 5) é a diferença entre pontos em casa e pontos fora (1-2) e foi criada como cálculo auxiliar.

O estudo incidirá sobre a coluna 4) pois somente o “rácio entre os pontos em casa e o Total de Pontos” traduz verdadeiramente o factor casa, ou seja, 1 ponto a mais conseguido em casa por uma equipa com 101 pontos vale menos que um ponto a mais para uma equipa que tem apenas 11 pontos. O Factor casa é muito maior na segunda hipótese e só a coluna 4) reflecte isso.

Pontos C	Pontos F	Total	%PontosC	C-F
51	50	101	50%	1
6	5	11	55%	1

dimensão da amostra	382	6)
minimo da amostra	0,41025641	7)
máximo da amostra	0,923076923	8)
número de classes	10	
precisão	2	
precisão dos limites das classes	0,01	
amplitude média das classes	0,05	
número de classes utilizadas	10	

para esta diferença uma vez que não tem qualquer influência para as conclusões que se pretendem tirar.

Ao todo temos 382 classificações finais de equipas analisadas 6), cujo o valor mais baixo da “%PC” encontrado é de 41% 7) enquanto o mais alto é de 92% 8).

É possível tirar as primeiras conclusões:

A medida que se pode obter a partir do máximo e do mínimo é a amplitude da amostra.

A amplitude da “%PC” é de 51 pp. o que não se pode considerar elevado tendo em conta que o valor poderia no limite variar 101 pp. (de [0 a 100])

Quando o valor “%PC” é inferior a 50% significa que a equipa conquistou mais pontos fora, quando é superior a 50% a equipa somou mais pontos em casa. Quanto mais se afastar do 50% maior é a diferença de pontos ganhos em casa e fora.

Caso não existisse factor casa a variável calculada tenderia para os 50%, ou seja metade dos pontos conquistados foram em casa, a outra metade fora.

Pela análise dos extremos é possível observar que o mínimo apenas se afasta 9 pp. dos 50% enquanto o máximo se afasta 42 pp., o que indicia (sem certezas) que a média andará longe (para valores maiores) dos 50%.

Através da função “CONTAR” do Excel, é possível saber quantos valores na amostra são maiores que 50%. Os resultados obtidos são os seguintes: 358 valores são >50% e somente 18 são inferiores que 50% (só 6 são efectivamente 50%)

Não seria de esperar que os valores coincidissem todos com 50% mas a distribuição andaria a volta de 50% e os valores superiores seriam ≈ aos valores inferiores.

O facto dos valores superiores a 50% serem 20 vezes em maior número que os valores inferiores só pode ser explicado pelo factor casa.

	aproximada	exacta
média	0,63	0,631

9)

Através desta tabela podemos saber o valor da média 63% 9). Vamos atribuir à média o valor representativo da amostra pois os valores distribuem-se em seu redor, ou seja, a 63% é o valor que anula a soma de todos os desvios sobre ele mesmo (entenda-se por desvio a diferença entre um valor da amostra e 63%).

Seria de esperar que a média fosse 50%, ou seja, que, em média, as equipas ganhassem tantos pontos em casa como fora. Mas a análise estatística diz-nos que, em média, 63% dos pontos são conseguidos em casa. Mais uma vez, só o factor casa pode explicar esta variação.

[De referir que os 63% representam, defacto, um factor casa considerável. Aos 63% de “%PC” corresponde 37% de %P.Fora (1-63%), o que torna possível afirmar que, em média, cada equipa conquista 1,71 vezes mais pontos em casa do que fora (63%/37%), ou se se preferir, conquista mais 71% de pontos em casa do que fora.]

Para a mediana a interpretação é a mesma. De referir que a mediana é, nesta amostra, igual à média o que indicia que não temos outliers significativos pois a mediana é uma medida mais resistente que a média. De notar que mesmo o primeiro quartil apresenta um valor acima de 50% (factor casa!). Significa que 25% das observações da amostra se encontram abaixo dos 57%.

11)

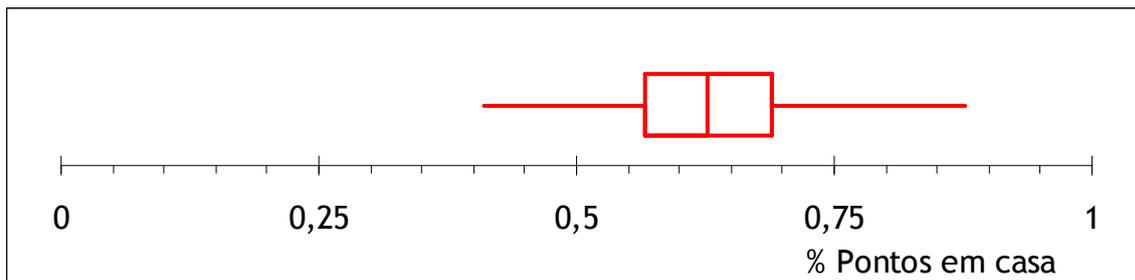
10)

11)

F*(me)	0,5	F*(q1)	0,25	F*(q3)	0,75
0,61	0,43194	0,56	0,22775	0,66	0,65445
0,635	0,53665	0,585	0,32461	0,71	0,79058
me=	0,626	q1=	0,566	q3=	0,695

A Mediana 10) diz-nos qual o valor central da amostra (estando os dados agrupados). Como a amostra tem número par de observações, a mediana é a média aritmética dos dois valores centrais.

Os quartis dividem a amostra (estando os dados agrupados) em quatro partes iguais 11).



(O gráfico exclui 2 outliers à direita, de 89,7% e de 92,3%.)

Unindo os extremos e os quartis obtemos o diagrama de extremos e quartis.

Através da análise ao diagrama é possível ver nitidamente o factor casa uma vez que os valores se encontram distribuídos à direita de 50%. Caso o factor casa não existisse a “caixa” estaria centrada sobre os 50% que coincidiriam com a mediana. Seria aceitável a existência de extremos que “afastassem” o primeiro e o terceiro quartil do centro mas a mediana estaria sempre muito próximo dos 50%.

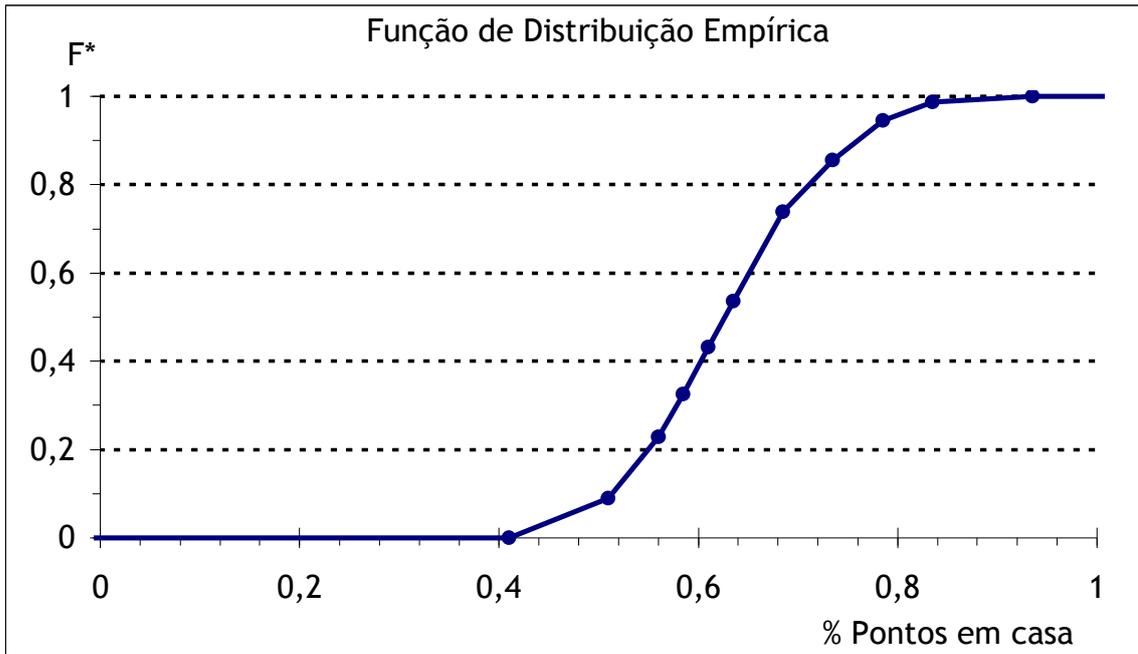
Obtemos também a informação que as observações estão mais concentradas no intervalo compreendido entre 57% e 62% do que no resto da amostra, ou seja, 25% das equipas têm entre 57% e 62% de “%PC” intervalo esse que corresponde apenas a 10% dos valores possíveis na amostra.

Adicionalmente podemos observar que a distribuição parece ser simétrica.

Os quartis foram calculados com base na função distribuição empírica. A função de distribuição tem como equação as últimas 3 colunas da **Tabela 1**), isto admitindo um crescimento uniforme dentro de cada classe. Representa a frequência relativa acumulada (F^* , coluna seleccionada). Graficamente apresenta o seguinte aspecto:

A escala do eixo da variável “%PC” é colocada entre 0% e 100% (valores possíveis). A leitura do gráfico permite confirmar que o valor mínimo surge apenas aos 41% (altura em que o gráfico começa a tomar valores diferentes de zero) enquanto que o máximo é atingido já depois dos 90% altura em que o total de frequências acumuladas atinge 1.

É possível ainda observar que em 41% a recta começa a crescer de forma lenta e que acelera a partir dos 51% (começam a registar-se mais observações) e mantém o crescimento constante (o número de observações registado é semelhante) até atingir 64% onde continua a crescer mas a ritmos decrescentes (o número de observações diminui).



$$F^*(x) = \begin{cases} 0, & x < 0,41 \\ -0,36 + 0,89 x, & 0,41 \leq x < 0,51 \\ -1,32 + 2,77 x, & 0,51 \leq x < 0,56 \\ -1,94 + 3,87 x, & 0,56 \leq x < 0,59 \\ -2,18 + 4,29 x, & 0,59 \leq x < 0,61 \\ -2,12 + 4,19 x, & 0,61 \leq x < 0,64 \\ -2,02 + 4,03 x, & 0,64 \leq x < 0,69 \\ -0,88 + 2,36 x, & 0,69 \leq x < 0,74 \\ -0,45 + 1,78 x, & 0,74 \leq x < 0,79 \\ 0,29 + 0,84 x, & 0,79 \leq x < 0,84 \\ 0,88 + 0,13 x, & 0,84 \leq x < 0,94 \\ 1, & 0,94 < x \end{cases}$$

A função de distribuição empírica tem esta expressão matemática e foi obtida através dos valores da **Tabela 1**.

(Em cima temos a sua representação gráfica).

Tabela 1)

amplitude (h _i)	Classes ([l _{i-1} ; l _i [c _i	n _i	f _i	h _i	f _i /h _i	F*(l _i)	m _i	b _i	F*(x) = b _i + m _i x, l _{i-1} ≤ x < l _i
0,1	[0,41 ; 0,51[0,46	34	0,089	0,1	0,89005	0,08901	0,89	-0,36	-0,36 + 0,89 x
0,05	[0,51 ; 0,56[0,535	53	0,139	0,05	2,77487	0,22775	2,77	-1,32	-1,32 + 2,77 x
0,025	[0,56 ; 0,59[0,5725	37	0,097	0,025	3,87435	0,32461	3,87	-1,94	-1,94 + 3,87 x
0,025	[0,59 ; 0,61[0,5975	41	0,107	0,025	4,29319	0,43194	4,29	-2,18	-2,18 + 4,29 x
0,025	[0,61 ; 0,64[0,6225	40	0,105	0,025	4,18848	0,53665	4,19	-2,12	-2,12 + 4,19 x
0,05	[0,64 ; 0,69[0,66	77	0,202	0,05	4,03141	0,73822	4,03	-2,02	-2,02 + 4,03 x
0,05	[0,69 ; 0,74[0,71	45	0,118	0,05	2,35602	0,85602	2,36	-0,88	-0,88 + 2,36 x
0,05	[0,74 ; 0,79[0,76	34	0,089	0,05	1,7801	0,94503	1,78	-0,45	-0,45 + 1,78 x
0,05	[0,79 ; 0,84[0,81	16	0,042	0,05	0,8377	0,98691	0,84	0,29	0,29 + 0,84 x
0,1	[0,84 ; 0,94[0,885	5	0,013	0,1	0,13089	1	0,13	0,88	0,88 + 0,13 x

A moda não é imediata no caso da variável ser contínua.

É feita com base no histograma.

O histograma é feito com base nos valores da **tabela 2)**

Ao organizar os dados em classes há alguma informação que é perdida. A amplitude das classes da **Tabela2)** foi assim dividida pelas respectivas classes de forma a minimizar essa perda. Para classes com poucos elementos é dada uma maior amplitude para classes com maior concentração, como é o caso da classe modal, a amplitude é mais reduzida para ser possível observar como é que os valores se encontram distribuídos.

Podemos obter o valor da moda (60,3%) através da fórmula de King:

$$0,59 + \frac{4,19}{3,87 + 4,19} * 0,025$$

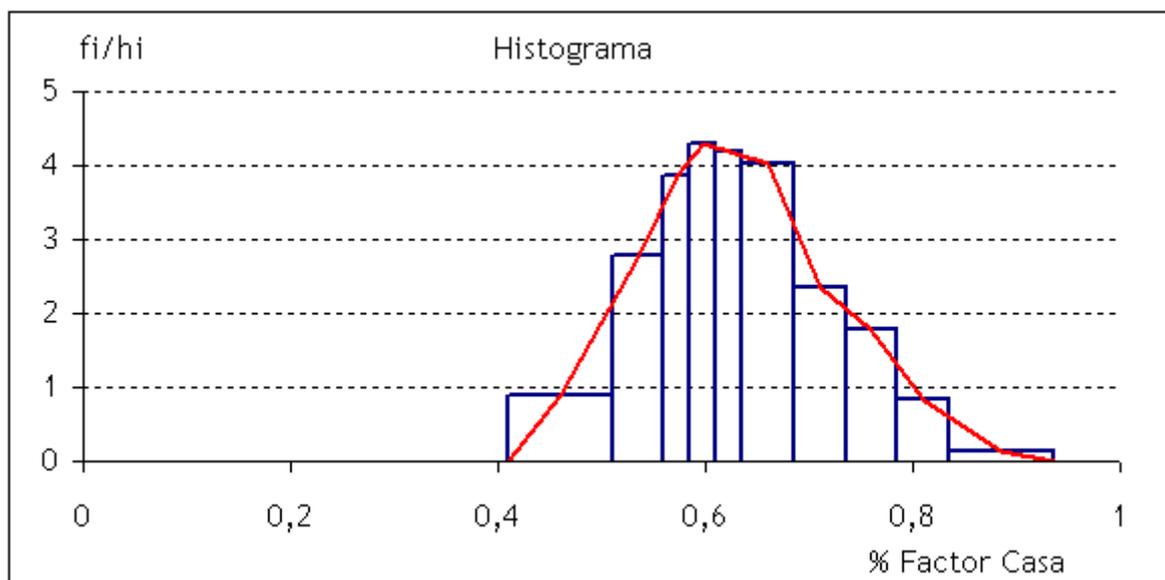
Podemos afirmar que a classe que regista maior número de observações é a que compreende o intervalo [59%;61%]. Através da fórmula de King podemos afirmar que o valor que se observou mais vezes estará muito próximo de 60,3%, (apesar de, por se tratar de uma variável contínua, ser difícil que o valor se repita, ou mesmo que se verifique!).

O valor encontrado para a moda, 60,3% vem mais uma vez confirmar que o factor casa existe. Caso assim não fosse, a moda da amostra seria ≈ 50%.

Podemos tirar outra conclusão desta observação. Pelo facto da moda ser diferente da média e da mediana a amostra não é (perfeitamente) simétrica.

Tabela 2)

l_0	limites classes (l_{i-1})	amplitude (h_i)	Classes ($[l_{i-1}; l_i[$)	c_i	n_i	f_i	h_i	f_i/h_i
0,41	0,41	0,1	[0,41 ; 0,51[0,46	34	0,089	0,1	0,89005
	0,51	0,05	[0,51 ; 0,56[0,535	53	0,139	0,05	2,77487
	0,56	0,025	[0,56 ; 0,59[0,5725	37	0,097	0,025	3,87435
	0,585	0,025	[0,59 ; 0,61[0,5975	41	0,107	0,025	4,29319
	0,61	0,025	[0,61 ; 0,64[0,6225	40	0,105	0,025	4,18848
	0,635	0,05	[0,64 ; 0,69[0,66	77	0,202	0,05	4,03141
	0,685	0,05	[0,69 ; 0,74[0,71	45	0,118	0,05	2,35602
	0,735	0,05	[0,74 ; 0,79[0,76	34	0,089	0,05	1,7801
	0,785	0,05	[0,79 ; 0,84[0,81	16	0,042	0,05	0,8377
l_{11}	0,835	0,1	[0,84 ; 0,94]	0,885	5	0,013	0,1	0,13089
0,93	0,935		Totais		382	1		



Análise da dispersão

Podemos obter informações acerca de como os valores se encontram distribuídos pela amostra. Para tal usamos medidas de dispersão.

A primeira medida calculada foi a amplitude da amostra (51 pp.) que é uma medida de dispersão, mas é a menos resistente. Podemos calcular a diferença entre o 3º e o 1º quartis para obter a amplitude inter quartil. o valor obtido é 12,4 pp.

Para ser possível a comparação deste valor com o de outras amostras é necessário relativiza-lo. Para isso vamos dividi-lo pela mediana (62.6%) o que resulta em 19,8.

Vamos utilizar ainda o coeficiente de variação para fazemos comparações futuras. O coeficiente de variação é obtido pelo rácio do desvio padrão sobre a média multiplicado por 100 o que resulta em 14,4(4).

$$\frac{\text{Desvio.Padrão}}{\text{Média}} \times 100 = 14.4(4)$$

F*(me)	0,5	F*(q1)	0,25	F*(q3)	0,75
0,61	0,43194	0,56	0,22775	0,66	0,65445
0,635	0,53665	0,585	0,32461	0,71	0,79058
me=	0,626	q1=	0,566	q3=	0,695

Podemos afirmar que 50% das equipas tem uma “%PC” compreendida entre 57% e 69% e que correspondem as observações mais centrais.

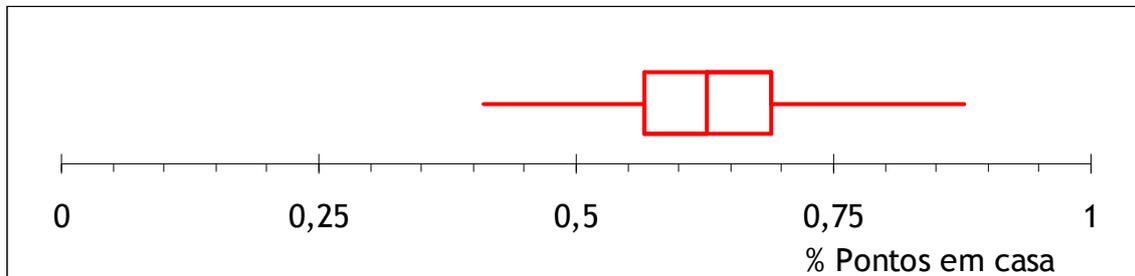
Temos a informação que o desvio padrão é de 9,1%. O desvio padrão (14) é calculado fazendo a raiz quadrada da variância (12). A vantagem do desvio padrão em relação à variância é que ao contrário desta, que vem em unidades ao quadrado, o desvio padrão é dado na mesma unidade da amostra. O desvio médio (14) obtém-se calculando a média dos desvios em módulo para a média. É de 7,3%.

	aproximada	exacta
12) variância	0,009780743	0,008264175
13) desvio padrão	0,099	0,091
14) desvio médio	0,077	0,073

Analisando o desvio médio podemos afirmar que, em média, os valores se afastam de 63% 7,3pp. (para valores maiores ou menores).

Mas para calcular a dispersão o desvio padrão costuma ser o mais utilizado (9,1 pp.).

Análise da Assimetria



Seria de esperar que a distribuição fosse simétrica, ou seja, que os pontos ganhos em casa fossem, em média, tantos quanto os pontos ganhos fora e que essa e que se encontra-se harmoniosamente distribuídos em torno dos 50% (%PC).

Uma vez que a existe o factor casa, a média, a mediana e a moda deslocam-se para valores superiores a 50% o que não implica que a distribuição deixe de ser simétrica (pode manter-se simétrica mas agora em relação ao valor 63%)

A primeira indicação que temos é que a (média = mediana) \approx moda.

O Grau de assimetria de Pearson obtém-se pelo seguinte cálculo $\frac{63,1 - 60,3}{9,1} = 0,3$

Quadro 1)

Grau de assimetria de Pearson 0,3	g)
coeficiente de assimetria de Fisher 0,179	g1)
grau de assimetria de Bowley 0,032	g')

momento centrado de ordem 3 0,000144227

O facto de a média ser igual à mediana e \approx à moda constitui um forte indicio de simetria.

Analisando o **Quadro1)** concluímos que os valores são todos positivos pelo que a amostra é assimétrica positiva. Porém, analisando com pormenor verificamos que os valores alcançados são baixos. A moda, provavelmente influenciada pela organização dos dados em classes, apresenta um valor diferente da média e da mediana, condicionando assim o **g)** que se afasta do **g1)** e do **g')**.

Podemos ainda referir que p momento centrado de ordem 3 ≈ 0 . Quando os momentos centrados de ordem impar são iguais a 0 a distribuição é perfeitamente simétrica.

Analisando o diagrama de extremos e quartis verificamos que os dados se encontram distribuídos simetricamente.

Por estes motivos, podemos considerar que a distribuição é simétrica.

Análise da concentração

2004/2005	Jogos	1) Pontos C	2) Pontos F	3) Total	4) %PontosC	5) C-F
Juventus	38	47	39	86	55%	8
Milan	38	38	41	79	48%	-3

Não é possível fazer a análise da concentração para o Factor casa.

A coluna 4), que tem sido estudada, não permite a análise da concentração uma vez que é condição para que se possa efectuar essa análise que os valores possam ser somados. Ora a soma de percentagens não faz sentido.

A diferença entre pontos em casa e pontos no fora 5) poderia servir como alternativa, uma vez que as conclusões em relação ao factor casa são idênticas a coluna 4) e tem como vantagem trabalhar com valores absolutos, o que possibilitaria a sua soma, mas também não é possível porque esta assume valores negativos.

2ª PARTE

Estatística descritiva bidimensional

Vimos que a média da percentagem de pontos em casa sobre os pontos no total para o conjunto das equipas é de 63%. Mas tal como já concluímos existe dispersão, ou seja este rácio não é igual para todas as equipas. Será que essa dispersão é aleatória ou resulta de outra variável?

2004/2005	Jogos	1) Pontos C	2) Pontos F	3) Total	4) %PontosC	5) C-F	Extracto da amostra	
Juventus	38	47	39	86	55%	8	0,546511628	86
Milan	38	38	41	79	48%	-3	0,481012658	79

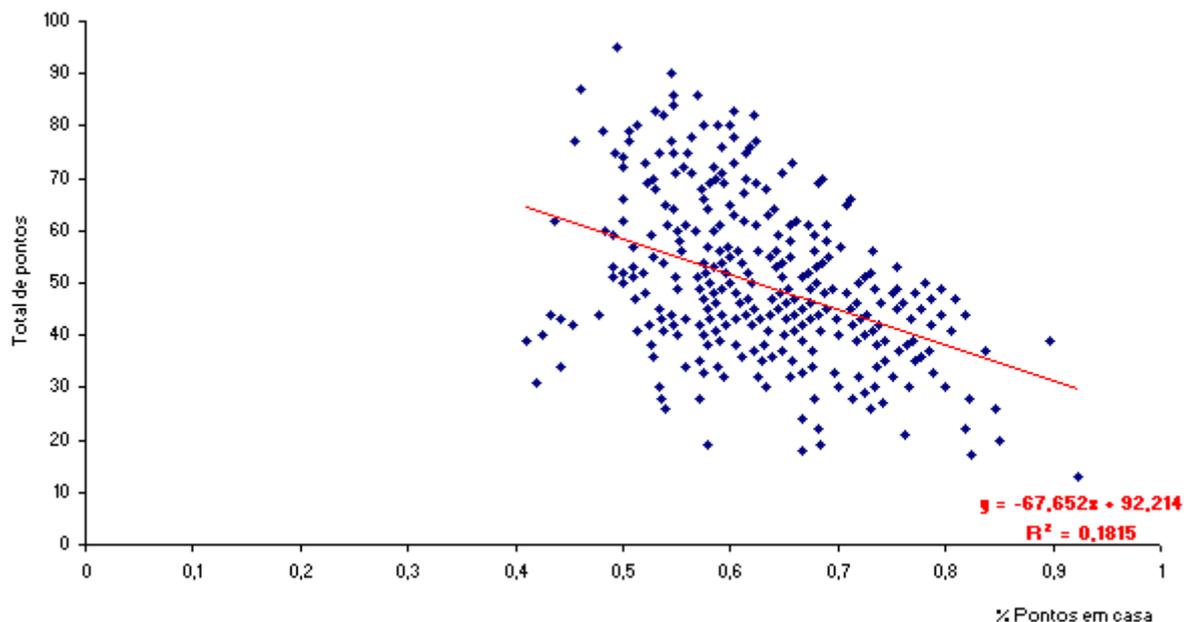
Vamos acrescentar agora a cada elemento da amostra (rácio P.casa/P.total *100) o respectivo valor de pontos totais. Por outras palavras, vamos criar pares ordenados com os valores das colunas 4) e 3) conforme demonstra este **extracto da amostra**.

Facilmente se consegue fazer a representação gráfica das duas variáveis, colocando no eixo dos “y” os valores do total de pontos e o correspondente valor em % de pontos em casa no eixo dos “x”.

Cada ponto azul corresponde a um par ordenado. A disposição dos pares ordenados sugere uma recta em torno do qual de encontram os pontos. Essa recta está representada a vermelho, designa-se por recta de regressão do total de pontos sobre %pontos casa e (como todas as rectas) pode-se escrever na forma $y=mx+b$.

A qualidade do ajustamento da recta ao conjunto de pontos é medida pelo R^2 (coeficiente de determinação).

Diagrama de dispersão



Seria de esperar que o factor casa afectasse todas as equipas de idêntica forma, dito por outras palavras, seria de esperar que a equipa campeã tivesse proporcionalmente mais pontos em casa do que a equipa que fica em último. Traduzindo em números, admitindo um dado campeonato

2xx0/2xx1		Pontos C	Pontos F	Total	%PontosC	C-F
Juventus	38	40	20	60	67%	20
Milan	38	20	10	30	67%	10
Inter	38	10	5	15	67%	5

em que a “%PontosC” é 67%, seria de esperar que o último conquistasse o dobro dos pontos em casa assim como os clubes do meio, assim como o campeão, tal como está exemplificado nesta tabela com classificações hipotéticas (feita com estes valores para tornar claro o exemplo de proporcionalidade).

É lógico supor que um campeonato pudesse afastar-se desta lógica, mas a média de vários campeonatos (a nossa amostra de 20 campeonatos) deveria corroborar este princípio.

Seria de esperar que a amostra real que temos obtivesse a mesma relação de proporcionalidade entre os primeiros e os últimos classificados que a tabela supra referida.

Média da variável pontos no total)

	exacta
média	49,5

A recta de regressão do total de pontos sobre %pontos casa deveria ter declive 0 estar ao nível 49,5 que é a **média da variável pontos no total**. Os pontos correspondentes aos pares ordenados estariam em torno da recta.

Mas...

Pela análise ao **diagrama de dispersão** conclui-se que assim não acontece. As equipas com mais pontos tendem a ter um factor casa mais reduzido enquanto que as equipas com menos pontos dependem mais do factor casa para pontuarem.

Depois de analisarmos o factor casa, chegamos à conclusão que existe um segundo factor que explica a dispersão não aleatória do primeiro → O factor *Total de pontos*

É possível tirar esta conclusão pois existe claramente uma relação estatística entre as duas variáveis, ou seja, o aumento de uma unidade na variável “%PC” implica uma diminuição em média da outra variável, “pontos no total”, no entanto a variável “pontos no total” não pode ser determinada pela “%PC”. Podemos então afirmar que existe uma correlação negativa entre as duas variáveis.

A recta de regressão do total de pontos sobre %pontos casa (calculada pelo método dos mínimos quadrados) apresenta a equação que está na tabela de regressão. A interpretação que se pode tirar destes valores é que por cada unidade aumentada na variável “%PC” a variável “Pontos no total” decresce, em média, 67,652 unidades. Como a variável “%PC” está em percentagem será mais lógico entender este valor dividido por 100 e interpretar da seguinte forma: Por cada unidade (em pontos percentuais) acrescida à variável “%PC” a variável “Pontos no Total” sofre, em média, uma variação de -0,67652. É esta a interpretação subjacente à coluna “variação” da tabela de regressão. Esta interpretação é, no entanto, bastante curiosa: O que nos é dito, por outras palavras, é

que através da regressão, sabendo a forma como os pontos estão distribuídos (em casa e fora) é possível prever o total de pontos de uma equipa, e que por cada acréscimo de um ponto percentual conquistado fora, a equipa tende, em média, a obter mais 0,67652 pontos no total, o que não é de compreensão imediata porque estamos a liderar apenas com uma variação na percentagem dos pontos conquistados fora em relação ao total, e consequentemente a uma variação no sentido inverso aos pontos conquistados em casa, não estamos a falar de valores absolutos (!!!).

A tabela de regressão foi feita com base nesta equação e apresenta alguns valores possíveis para a variável “%PC” com intervalos de 1pp. e desde o mínimo até ao máximo que a variável toma na amostra. É possível verificar como as médias das duas variáveis se intersectam (na coluna seleccionada). Esta seria a situação geral caso não existisse o “*factor Total de pontos*” (a recta de regressão ao nível da média de y com declive nulo).

O valor 92,214 teoricamente é o total de pontos que, em média, uma equipa obtém caso tenha um factor casa nulo (carece de sentido).

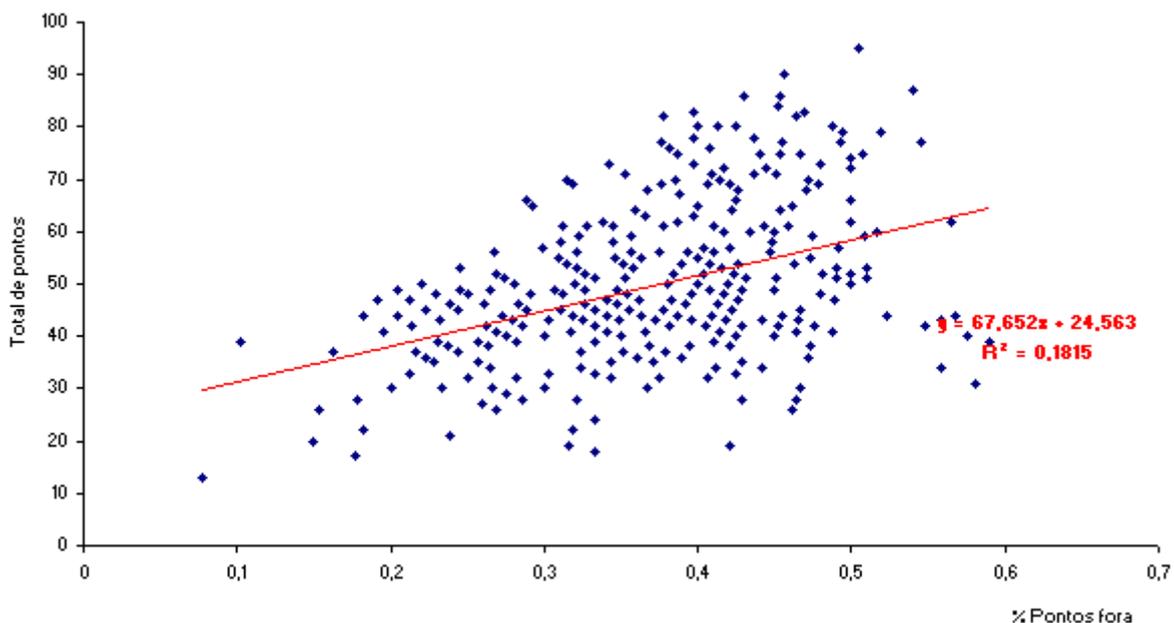
Podemos ainda medir a qualidade do ajustamento através do valor de R^2 (0,1815). Sendo que quando R^2 é nulo não existe correlação e que quando este é =1 a correlação é máxima podemos interpretar este valor como sendo baixo. Existe de facto correlação entre a “%PC” e “P. Total” mas não se pode considerar uma correlação muito forte. ou seja. existem muitas excepções.

Tabela de Regressão						
Pontos no total (y) = -67,652 "% P.Casa/Total"(x) + 92,214						
% P.Casa	P.Total	Varição		% P.Casa	P.Total	Varição
41%	64,477	-	a	68%	46,206	-0,67652
42%	63,8	-0,67652	-67,652	69%	45,529	-0,67652
43%	63,124	-0,67652		70%	44,852	-0,67652
44%	62,447	-0,67652	a/100	71%	44,175	-0,67652
45%	61,771	-0,67652	-0,67652	72%	43,498	-0,67652
46%	61,094	-0,67652		73%	42,821	-0,67652
47%	60,418	-0,67652	b	74%	42,144	-0,67652
48%	59,741	-0,67652	92,214	75%	41,467	-0,67652
49%	59,065	-0,67652		76%	40,79	-0,67652
50%	58,388	-0,67652		77%	40,113	-0,67652
51%	57,711	-0,67652		78%	39,436	-0,67652
52%	57,035	-0,67652		79%	38,759	-0,67652
53%	56,358	-0,67652		80%	38,082	-0,67652
54%	55,682	-0,67652		81%	37,405	-0,67652
55%	55,005	-0,67652		82%	36,728	-0,67652
56%	54,329	-0,67652		83%	36,051	-0,67652
57%	53,652	-0,67652		84%	35,374	-0,67652
58%	52,976	-0,67652		85%	34,697	-0,67652
59%	52,299	-0,67652		86%	34,02	-0,67652
60%	51,622	-0,67652		87%	33,343	-0,67652
61%	50,945	-0,67652		88%	32,666	-0,67652
62%	50,268	-0,67652		89%	31,989	-0,67652
63%	49,591	-0,67652		90%	31,312	-0,67652
64%	48,914	-0,67652		91%	30,635	-0,67652
65%	48,237	-0,67652		92%	29,958	-0,67652
66%	47,56	-0,67652		93%	29,281	-0,67652
67%	46,883	-0,67652				

A título de curiosidade foram correlacionadas as variáveis “% pontos fora” e “Pontos total” e tal como se impunha o diagrama de dispersão, bem como a recta de regressão, são precisamente o inverso do que tínhamos calculado até aqui.

No entanto torna-se mais fácil de perceber o 2º factor calculado. Podemos afirmar que ao aumentar 1pp. a “% de pontos ganhos fora”, em média, obtemos um acréscimo na classificação final de 0,67652.

A explicação (extra-estatística) pode ter haver com o facto de o campeonato ser uma prova de regularidade. As equipas que dependem mais do factor casa são mais irregulares desde logo porque os jogos em casa são alternados com jogos fora. As que não dependem tanto conseguem ter melhores sequências ao longo do campeonato do que as outras. Mas tal como já foi visto anteriormente, a correlação não se pode considerar muito forte. De facto existe, pode ser vantajoso saber tirar partido da informação que fornece, mas não determina, há muitas excepções.



3ª Parte

(Todo o texto é interpretativo, não sendo enunciadas fórmulas)

Foi criada uma nova amostra desta vez com a classificação obtida por todos os campeões em Portugal desde o início da competição.

Tabela de todos os campeões em Portugal											
Ano	J	P.Casa	P.For	P.Total	%P.Casa	Ano	J	P.Casa	P.For	P.Total	%P.Casa
2004 / 2005	34	39	26	65	60%	1968 / 1969	26	23	16	39	59%
2003 / 2004	34	51	31	82	62%	1967 / 1968	26	26	15	41	63%
2002 / 2003	34	49	37	86	57%	1966 / 1967	26	26	17	43	60%
2001 / 2002	34	42	33	75	56%	1965 / 1966	26	22	17	39	56%
2000 / 2001	34	48	29	77	62%	1964 / 1965	26	25	18	43	58%
1999 / 2000	34	42	35	77	55%	1963 / 1964	26	24	22	46	52%
1998 / 1999	34	48	31	79	61%	1962 / 1963	26	23	25	48	48%
1997 / 1998	38	38	28	66	58%	1961 / 1962	26	23	20	43	53%
1996 / 1997	34	39	46	85	46%	1960 / 1961	26	26	20	46	57%
1995 / 1996	34	48	36	84	57%	1959 / 1960	26	23	22	45	51%
1994 / 1995	34	32	30	62	52%	1958 / 1959	26	23	18	41	56%
1993 / 1994	34	30	24	54	56%	1957 / 1958	26	24	19	43	56%
1992 / 1993	34	29	25	54	54%	1956 / 1957	26	23	18	41	56%
1991 / 1992	34	32	24	56	57%	1955 / 1956	26	25	18	43	58%
1990 / 1991	38	35	34	69	51%	1954 / 1955	26	21	18	39	54%
1989 / 1990	34	32	27	59	54%	1953 / 1954	26	26	17	43	60%
1988 / 1989	34	48	31	79	61%	1952 / 1953	26	24	19	43	56%
1987 / 1988	38	38	28	66	58%	1951 / 1952	26	23	18	41	56%
1986 / 1987	30	28	21	49	57%	1950 / 1951	26	23	22	45	51%
1985 / 1986	30	29	20	49	59%	1949 / 1950	26	22	23	45	49%
1984 / 1985	30	29	26	55	53%	1948 / 1949	26	24	18	42	57%
1983 / 1984	30	28	24	52	54%	1947 / 1948	26	23	18	41	56%
1982 / 1983	30	28	23	51	55%	1946 / 1947	26	26	21	47	55%
1981 / 1982	30	25	21	46	54%	1945 / 1946	22	21	17	38	55%
1980 / 1981	30	29	21	50	58%	1944 / 1945	18	15	15	30	50%
1979 / 1980	30	30	22	52	58%	1943 / 1944	18	17	14	31	55%
1978 / 1979	30	27	23	50	54%	1942 / 1943	18	18	12	30	60%
1977 / 1978	30	28	23	51	55%	1941 / 1942	22	22	16	38	58%
1976 / 1977	30	28	23	51	55%	1940 / 1941	14	12	11	23	52%
1975 / 1976	30	25	25	50	50%	1939 / 1940	18	18	16	34	53%
1974 / 1975	30	26	23	49	53%	1938 / 1939	14	13	10	23	57%
1973 / 1974	30	28	21	49	57%	1937 / 1938	14	13	10	23	57%
1972 / 1973	30	30	28	58	52%	1936 / 1937	14	14	10	24	58%
1971 / 1972	30	28	27	55	51%	1935 / 1936	14	13	8	21	62%
1970 / 1971	26	25	16	41	61%	1934 / 1935	14	14	8	22	64%
1969 / 1970	26	25	21	46	54%						

A competição foi sofrendo alterações ao longo do tempo, por exemplo, no princípio competiam somente 8 equipas. O valor do primeiro campeonato deve entrar para os cálculos estatísticos com uma ponderação diferente do valor do último campeonato onde há mais equipas e consequentemente mais jogos. Porém isso representa muitas dificuldades de cálculo.

Foi calculada a média à variável “%PC” com a ponderação dos jogos efectuados e o valor exacto é 55,76%. A média da mesma variável sem ponderação dá 55,83%. Como o erro é muito pequeno não vamos considerar qualquer ponderação (com a nota que os valores a que chegar-mos serão sempre aproximações).

Amostra dos 20 campeonatos

dimensão da amostra	382
mínimo da amostra	0,41025641
máximo da amostra	0,923076923

Amostra dos campeões

dimensão da amostra	71
mínimo da amostra	0,458823529
máximo da amostra	0,636363636

A amplitude da amostra dos campeões é menor. É preciso fazer a ressalva que sendo a sua dimensão menor há desde logo uma maior probabilidade de a amplitude ser menor também uma vez que esta é a medida de localização mais sensível, mas ainda assim é possível tirar conclusões, desde logo porque a variação para o mínimo da amostra é de apenas 4pp. enquanto que a variação para o máximo é de 29 pp. Curioso observar como o máximo da amostra dos campeões é a média da amostra dos 20 campeonatos o que por si só comprova que os clubes campeões têm, em média, menos “%PC” que os restantes.

A moda é, na amostra dos campeões e calculada pela fórmula de King, de 56,7% $0,558 + \frac{5,07}{9,01 + 5,07}(0,583 - 0,558)$. É mais baixa do que a moda da amostra dos 20 campeonatos pelo que irá influenciar a média.

Amostra dos 20 campeonatos

F*(me) 0,5	F*(q1) 0,25	F*(q3) 0,75
0,61 0,43194	0,56 0,22775	0,66 0,65445
0,635 0,53665	0,585 0,32461	0,71 0,79058
me= 0,626	q1= 0,566	q3= 0,695

Amostra dos campeões

F*(me) 0,5	F*(q1) 0,25	F*(q3) 0,75
0,558 0,4507	0,533 0,22535	0,558 0,4507
0,583 0,78873	0,558 0,4507	0,583 0,78873
me= 0,5616	q1= 0,5357	q3= 0,5801

exacta
média 0,631

exacta
média 0,5583

Tal como já se esperava a média dos campeões é bem mais baixa do que a dos 20 campeonatos.

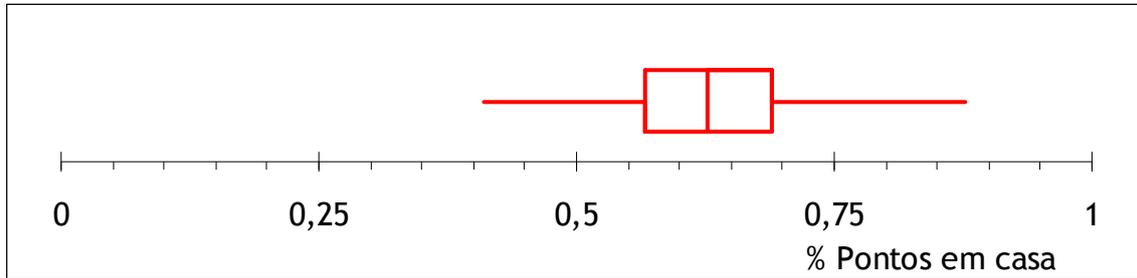
Através destes valores podemos afirmar que, em média, os campeões têm menos “%PC” que os restantes clubes.

Aos 63% corresponde 37% de %P.Forá (1-63%), o que torna possível afirmar que, em média, cada equipa conquista 1,71 vezes mais pontos em casa do que fora (63%/37%). Por cálculos análogos, em média, os campeões conquistam apenas 1,22 vezes mais pontos em casa do que fora.

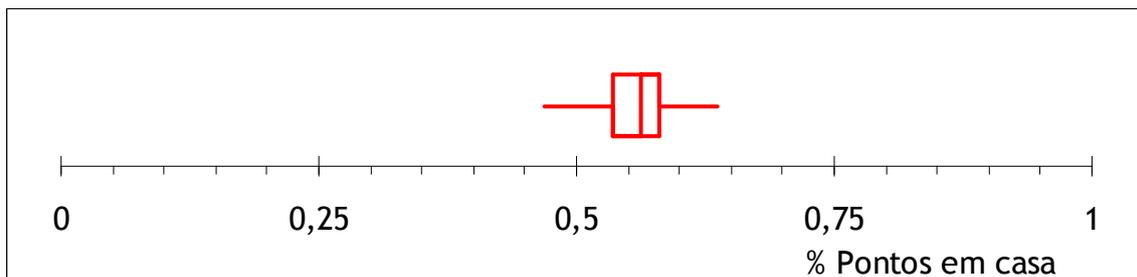
Através da amplitude inter quartis retiramos semelhantes conclusões à amplitude da amostra e ainda é possível observar como os valores centrais da amostra dos campeões estão mais concentrados do que os da amostra dos 20 campeonatos (mesmo feita a ressalva da dimensão das amostras).

Podemos ter em conta uma medida relativa para compararmos as amplitudes inter quartis dividindo pela mediana. Assim, para a amostra dos 20 campeonatos o valor encontrado é de 19,8 pp. enquanto que para a amostra dos campeões é de apenas 7,9 pp. o que indica uma amplitude relativamente menor.

Amostra dos 20 campeonatos



Amostra dos campeões

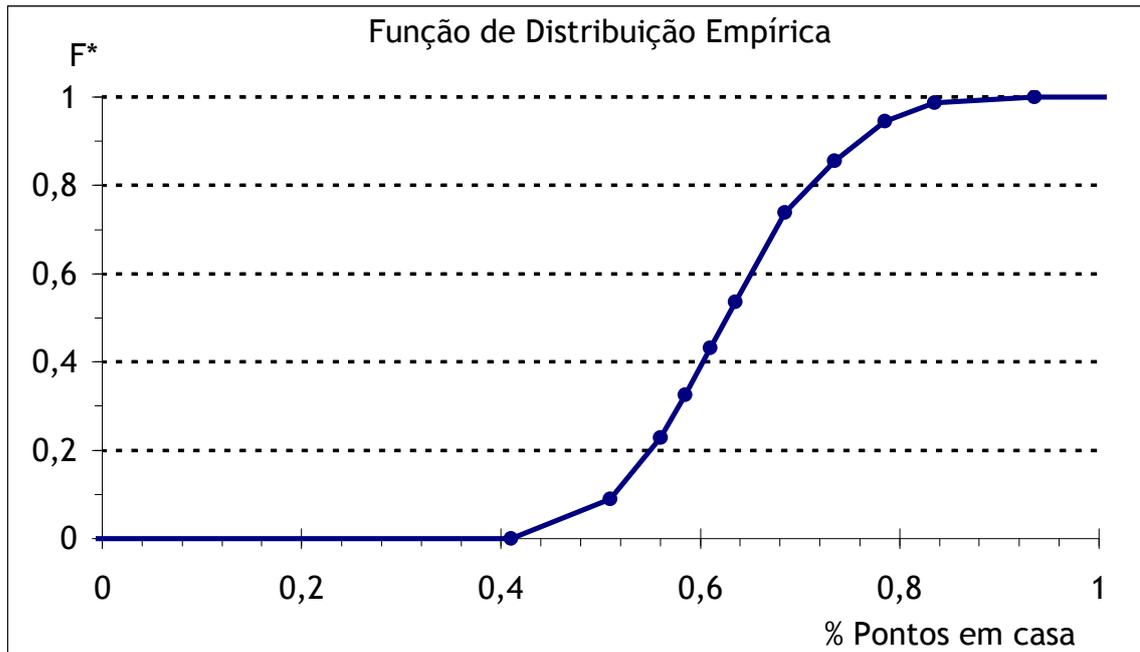


Através dos gráficos de extremos e quartis tornam-se mais nítidas as conclusões agora tiradas pois é possível a sua visualização gráfica. Mais uma vez, mesmo atendendo ao facto da amostra dos campeões ser bem mais pequena, é possível verificar como esta se situa à esquerda da amostra dos 20 campeonatos, mas ainda assim acima dos 50%.

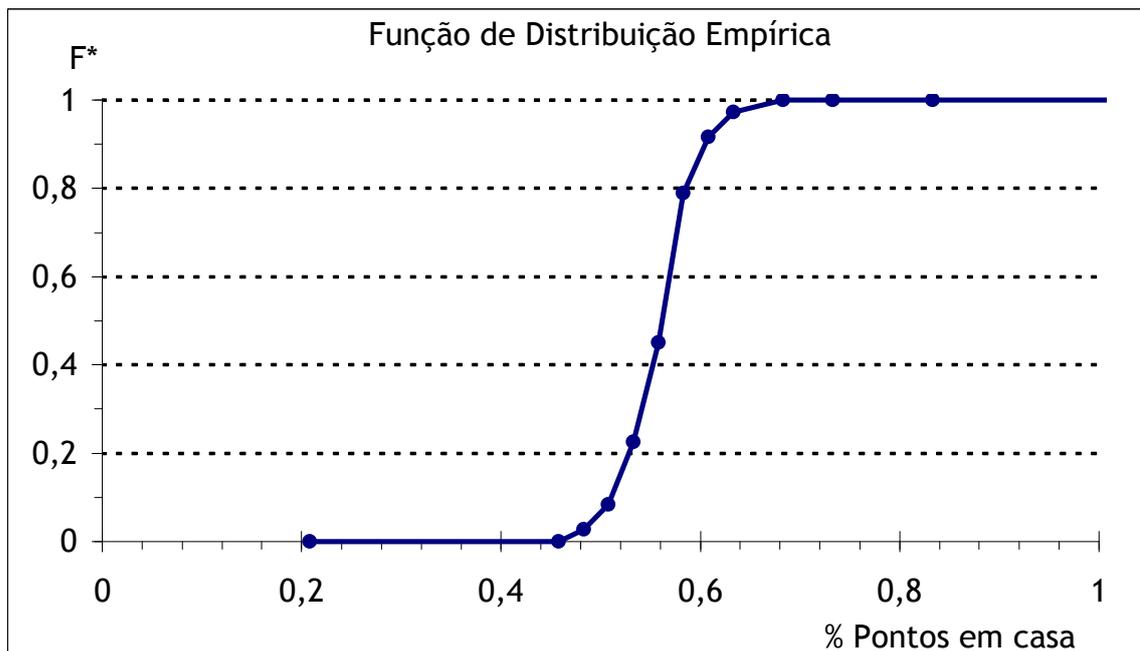
A amostra dos 20 campeonatos tem uma amplitude maior sendo por isso o gráfico de maior dimensão também.

A amostra dos campeões, ao contrário da amostra dos 20 campeonatos, não é simétrica, sendo perceptível um enviesamento à direita.

Amostra dos 20 campeonatos



Amostra dos campeões



Os gráficos de distribuição empírica mostram-nos como as frequências acumuladas vão evoluindo. É possível observar a maior amplitude e dispersão da amostra dos 20 campeonatos (começa a crescer primeiro e só atinge os 100% depois da amostra dos campeões e nunca cresce com um declive tão acentuado como o gráfico dos campeões).

Amostra dos 20 campeonatos

	exacta
variância	0,008264175
desvio padrão	0,091
desvio médio	0,073

Amostra dos campeões

	exacta
variância	0,001371052
desvio padrão	0,037
desvio médio	0,0286

Os desvios confirmam analiticamente aquilo que já foi inferido graficamente: a dispersão é menor na amostra dos campeões.

Isto explica-se precisamente pelo factor *pontos no total*. Os campeões, que são quem têm mais pontos no total, têm um factor casa mais reduzido, e a amostra dos campeões só inclui estas equipas, ao contrário da amostra dos 20 campeonatos que inclui os campeões e as “outras”, obtendo assim uma dispersão maior.

O coeficiente de variação para a amostra dos 20 campeonatos é 14,4 enquanto que para a amostra dos campeões é de apenas 6,6, ou seja, as observações variam relativamente menos em relação à média na amostra dos campeões.

Conclusão

O remate do trabalho não será extenso uma vez que as conclusões foram tiradas ao longo do trabalho e resumem-se a 2 ideias principais:

1º Ficou demonstrado, através dos instrumentos estatísticos utilizados, que a equipa que joga em casa tem uma vantagem extra em relação à equipa visitante.

2º Ficou também demonstrado que essa vantagem varia conforme a equipa. As equipas de topo tendem a depender menos do factor casa para pontuar.

Dados

Portugal

Época 2004/2005

Equipa	J	P.Casa	P.Forá	P.Total	%PontosC	C-F
<u>Benfica</u>	34	39	26	65	60%	13
<u>F.C. Porto</u>	34	27	35	62	44%	-8
<u>Sporting</u>	34	34	27	61	56%	7
<u>Sp. Braga</u>	34	32	26	58	55%	6
<u>V. Guimarães</u>	34	32	22	54	59%	10
<u>Boavista</u>	34	30	20	50	60%	10
<u>Marítimo</u>	34	34	15	49	69%	19
<u>Rio Ave</u>	34	31	16	47	66%	15
<u>Belenenses</u>	34	35	11	46	76%	24
<u>V. Setúbal</u>	34	26	18	44	59%	8
<u>Penafiel</u>	34	24	19	43	56%	5
<u>Nacional</u>	34	28	13	41	68%	15
<u>Gil Vicente</u>	34	26	14	40	65%	12
<u>Académica</u>	34	24	14	38	63%	10
<u>U. Leiria</u>	34	20	18	38	53%	2
<u>Moreirense</u>	34	23	11	34	68%	12
<u>Estoril</u>	34	24	6	30	80%	18
<u>Beira Mar</u>	34	16	14	30	53%	2
					62%	10

Época 2003/2004

Equipa	J	P.Casa	P.Forá	P.Total	%PontosC	C-F
<u>F.C. Porto</u>	34	51	31	82	62%	20
<u>Benfica</u>	34	37	37	74	50%	0
<u>Sporting</u>	34	44	29	73	60%	15
<u>Nacional</u>	34	41	15	56	73%	26
<u>Sp. Braga</u>	34	29	25	54	54%	4
<u>Marítimo</u>	34	34	14	48	71%	20
<u>Rio Ave</u>	34	33	15	48	69%	18
<u>Boavista</u>	34	30	17	47	64%	13
<u>Moreirense</u>	34	33	13	46	72%	20
<u>U. Leiria</u>	34	28	17	45	62%	11
<u>Beira Mar</u>	34	27	14	41	66%	13
<u>Gil Vicente</u>	34	28	12	40	70%	16
<u>Académica</u>	34	20	18	38	53%	2
<u>V. Guimarães</u>	34	25	12	37	68%	13
<u>Alverca</u>	34	20	15	35	57%	5
<u>Belenenses</u>	34	22	13	35	63%	9
<u>P. Ferreira</u>	34	20	8	28	71%	12
<u>E. Amadora</u>	34	14	3	17	82%	11
					65%	12,66666667

Epoca 2002/2003

Equipa	J	P.Casa	P.Forá	P.Total	%PontosC	C-F
<u>F.C. Porto</u>	34	49	37	86	57%	12
<u>Benfica</u>	34	37	38	75	49%	-1
<u>Sporting</u>	34	29	30	59	49%	-1
<u>V. Guimarães</u>	34	25	25	50	50%	0
<u>U. Leiria</u>	34	29	20	49	59%	9
<u>P. Ferreira</u>	34	31	14	45	69%	17
<u>Boavista</u>	34	29	15	44	66%	14
<u>Gil Vicente</u>	34	29	15	44	66%	14
<u>Marítimo</u>	34	35	9	44	80%	26
<u>Belenenses</u>	34	30	13	43	70%	17
<u>Moreirense</u>	34	26	14	40	65%	12
<u>Nacional</u>	34	26	14	40	65%	12
<u>Beira Mar</u>	34	29	10	39	74%	19
<u>Sp. Braga</u>	34	28	10	38	74%	18
<u>Académica</u>	34	28	9	37	76%	19
<u>Varzim</u>	34	28	8	36	78%	20
<u>Santa Clara</u>	34	26	9	35	74%	17
<u>V. Setúbal</u>	34	13	18	31	42%	-5
				65%		12,16666667

Epoca 2001/2002

Equipa	J	P.Casa	P.Forá	P.Total	%PontosC	C-F
<u>Sporting</u>	34	42	33	75	56%	9
<u>Boavista</u>	34	37	33	70	53%	4
<u>F.C. Porto</u>	34	39	29	68	57%	10
<u>Benfica</u>	34	38	25	63	60%	13
<u>Belenenses</u>	34	29	28	57	51%	1
<u>Marítimo</u>	34	35	21	56	63%	14
<u>U. Leiria</u>	34	35	20	55	64%	15
<u>P. Ferreira</u>	34	27	19	46	59%	8
<u>Sp. Braga</u>	34	31	11	42	74%	20
<u>V. Guimarães</u>	34	30	12	42	71%	18
<u>Beira Mar</u>	34	28	11	39	72%	17
<u>Gil Vicente</u>	34	29	9	38	76%	20
<u>V. Setúbal</u>	34	29	9	38	76%	20
<u>Santa Clara</u>	34	28	9	37	76%	19
<u>Varzim</u>	34	23	9	32	72%	14
<u>Salgueiros</u>	34	23	7	30	77%	16
<u>Farense</u>	34	23	5	28	82%	18
<u>Alverca</u>	34	20	7	27	74%	13
				67%		13,83333333

Epoca 2000/2001

Equipa	J	P.Casa	P.Forá	P.Total	%PontosC	C-F
<u>Boavista</u>	34	48	29	77	62%	19
<u>F.C. Porto</u>	34	47	29	76	62%	18
<u>Sporting</u>	34	38	24	62	61%	14
<u>Sp. Braga</u>	34	40	17	57	70%	23
<u>U. Leiria</u>	34	36	20	56	64%	16

<u>Benfica</u>	34	37	17	54	69%	20
<u>Belenenses</u>	34	32	20	52	62%	12
<u>Beira Mar</u>	34	28	21	49	57%	7
<u>P. Ferreira</u>	34	28	20	48	58%	8
<u>Alverca</u>	34	25	18	43	58%	7
<u>Marítimo</u>	34	33	10	43	77%	23
<u>Salgueiros</u>	34	23	20	43	53%	3
<u>Farense</u>	34	26	13	39	67%	13
<u>Gil Vicente</u>	34	29	8	37	78%	21
<u>V. Guimarães</u>	34	19	17	36	53%	2
<u>Campomaiorense</u>	34	21	11	32	66%	10
<u>D. Aves</u>	34	18	4	22	82%	14
<u>E. Amadora</u>	34	13	6	19	68%	7
					65%	13,16666667

Média do Campeonato Português 65%

Tabela de todos os campeões em Portugal

Ano	J	P.Casa	P.Fora	P.Total	%P.Casa	Ano	J	P.Casa	P.Fora	P.Total	%P.Casa
2004 / 2005	34	39	26	65	60%	1968 / 1969	26	23	16	39	59%
2003 / 2004	34	51	31	82	62%	1967 / 1968	26	26	15	41	63%
2002 / 2003	34	49	37	86	57%	1966 / 1967	26	26	17	43	60%
2001 / 2002	34	42	33	75	56%	1965 / 1966	26	22	17	39	56%
2000 / 2001	34	48	29	77	62%	1964 / 1965	26	25	18	43	58%
1999 / 2000	34	42	35	77	55%	1963 / 1964	26	24	22	46	52%
1998 / 1999	34	48	31	79	61%	1962 / 1963	26	23	25	48	48%
1997 / 1998	38	38	28	66	58%	1961 / 1962	26	23	20	43	53%
1996 / 1997	34	39	46	85	46%	1960 / 1961	26	26	20	46	57%
1995 / 1996	34	48	36	84	57%	1959 / 1960	26	23	22	45	51%
1994 / 1995	34	32	30	62	52%	1958 / 1959	26	23	18	41	56%
1993 / 1994	34	30	24	54	56%	1957 / 1958	26	24	19	43	56%
1992 / 1993	34	29	25	54	54%	1956 / 1957	26	23	18	41	56%
1991 / 1992	34	32	24	56	57%	1955 / 1956	26	25	18	43	58%
1990 / 1991	38	35	34	69	51%	1954 / 1955	26	21	18	39	54%
1989 / 1990	34	32	27	59	54%	1953 / 1954	26	26	17	43	60%
1988 / 1989	34	48	31	79	61%	1952 / 1953	26	24	19	43	56%
1987 / 1988	38	38	28	66	58%	1951 / 1952	26	23	18	41	56%
1986 / 1987	30	28	21	49	57%	1950 / 1951	26	23	22	45	51%
1985 / 1986	30	29	20	49	59%	1949 / 1950	26	22	23	45	49%
1984 / 1985	30	29	26	55	53%	1948 / 1949	26	24	18	42	57%
1983 / 1984	30	28	24	52	54%	1947 / 1948	26	23	18	41	56%
1982 / 1983	30	28	23	51	55%	1946 / 1947	26	26	21	47	55%
1981 / 1982	30	25	21	46	54%	1945 / 1946	22	21	17	38	55%
1980 / 1981	30	29	21	50	58%	1944 / 1945	18	15	15	30	50%
1979 / 1980	30	30	22	52	58%	1943 / 1944	18	17	14	31	55%
1978 / 1979	30	27	23	50	54%	1942 / 1943	18	18	12	30	60%
1977 / 1978	30	28	23	51	55%	1941 / 1942	22	22	16	38	58%
1976 / 1977	30	28	23	51	55%	1940 / 1941	14	12	11	23	52%
1975 / 1976	30	25	25	50	50%	1939 / 1940	18	18	16	34	53%
1974 / 1975	30	26	23	49	53%	1938 / 1939	14	13	10	23	57%
1973 / 1974	30	28	21	49	57%	1937 / 1938	14	13	10	23	57%
1972 / 1973	30	30	28	58	52%	1936 / 1937	14	14	10	24	58%
1971 / 1972	30	28	27	55	51%	1935 / 1936	14	13	8	21	62%
1970 / 1971	26	25	16	41	61%	1934 / 1935	14	14	8	22	64%
1969 / 1970	26	25	21	46	54%						

Espanha

2000/2001

		Pontos C	Pontos F	Total	%PontosC	C-F
Real Madrid	38	48	32	80	60%	16
Deportivo	38	48	25	73	66%	23
Mallorca	38	46	25	71	65%	21
Barcelona	38	40	23	63	63%	17
Valencia	38	40	23	63	63%	17
Celta	38	40	19	59	68%	21
Villarreal	38	33	24	57	58%	9
Málaga	38	35	21	56	63%	14
Espanyol	38	30	20	50	60%	10
Alavés	38	32	17	49	65%	15
Las Palmas	38	33	13	46	72%	20
Rayo	38	27	16	43	63%	11
R.Sociedad	38	24	19	43	56%	5
Athletic	38	31	12	43	72%	19
Zaragoza	38	33	9	42	79%	24
Valladolid	38	33	9	42	79%	24
Osasuna	38	26	16	42	62%	10
Oviedo	38	33	8	41	80%	25
Racing	38	35	4	39	90%	31
Numancia	38	29	10	39	74%	19
					68%	17,55

2001/2002

		Pontos C	Pontos F	Total	%PontosC	C-F
Valencia	38	46	32	75	61%	14
Deportivo	38	43	25	68	63%	18
Real Madrid	38	47	25	66	71%	22
Barcelona	38	41	23	64	64%	18
Celta	38	35	23	60	58%	12
Real Betis	38	38	19	59	64%	19
Alavés	38	31	24	54	57%	7
Sevilla	38	27	21	53	51%	6
Málaga	38	31	20	53	58%	11
Athletic	38	25	17	50	50%	8
Rayo	38	36	13	49	73%	23
Valladolid	38	31	16	48	65%	15
R.Sociedad	38	30	19	47	64%	11
Espanyol	38	38	12	47	81%	26
Villarreal	38	33	9	43	77%	24
Mallorca	38	26	9	43	60%	17
Osasuna	38	24	16	42	57%	8
Tenerife	38	26	8	41	63%	18

Las Palmas	38	29	4	40	73%	25
Zaragoza	38	25	10	37	68%	15
					64%	15,85

2002/2003

	Pontos C	Pontos F	Total	%PontosC	C-F	
Real Madrid	38	44	29	78	56%	15
R.Sociedad	38	45	25	76	59%	20
Deportivo	38	42	19	72	58%	23
Celta	38	36	23	61	59%	13
Valencia	38	33	25	60	55%	8
Barcelona	38	35	21	56	63%	14
Athletic	38	36	23	55	65%	13
Real Betis	38	33	26	54	61%	7
Mallorca	38	26	22	52	50%	4
Sevilla	38	31	25	50	62%	6
Osasuna	38	27	13	47	57%	14
Atlético	38	29	17	47	62%	12
Málaga	38	31	17	46	67%	14
Valladolid	38	31	9	46	67%	22
Villarreal	38	29	10	45	64%	19
Racing	38	29	17	44	66%	12
Espanyol	38	30	18	43	70%	12
Recreativo	38	23	15	36	64%	8
Alavés	38	23	11	35	66%	12
Rayo	38	20	12	32	63%	8
				62%	12,8	

2003/2004

	Pontos C	Pontos F	Total	%PontosC	C-F	
Valencia	38	39	34	77	51%	5
Barcelona	38	36	31	72	50%	5
Deportivo	38	42	30	71	59%	12
Real Madrid	38	41	25	70	59%	16
Athletic	38	35	27	56	63%	8
Sevilla	38	38	21	55	69%	17
Atlético	38	38	19	55	69%	19
Villarreal	38	35	21	54	65%	14
Real Betis	38	27	26	52	52%	1
Málaga	38	34	19	51	67%	15
Mallorca	38	26	20	51	51%	6
Zaragoza	38	28	18	48	58%	10
Osasuna	38	25	15	48	52%	10
Albacete	38	29	15	47	62%	14
R.Sociedad	38	28	16	46	61%	12
Racing	38	19	15	43	44%	4
Espanyol	38	26	13	43	60%	13
Valladolid	38	27	13	41	66%	14
Celta	38	16	12	39	41%	4
Murcia	38	22	12	26	85%	10
				59%	10,45	

2004/2005

		Pontos C	Pontos F	Total	%PontosC	C-F
Barcelona	38	46	38	84	55%	8
Real Madrid	38	46	36	80	58%	10
Villarreal	38	46	29	65	71%	17
Real Betis	38	41	29	62	66%	12
Espanyol	38	41	21	61	67%	20
Sevilla	38	35	17	60	58%	18
Valencia	38	38	17	58	66%	21
Athletic	38	37	19	51	73%	18
Deportivo	38	25	25	51	49%	0
Málaga	38	28	17	51	55%	11
Zaragoza	38	36	25	50	72%	11
Atlético	38	39	20	50	78%	19
Getafe	38	37	23	47	79%	14
R.Sociedad	38	31	18	47	66%	13
Osasuna	38	33	18	46	72%	15
Racing	38	30	24	44	68%	6
Mallorca	38	23	17	39	59%	6
Levante	38	24	14	37	65%	10
Numancia	38	21	23	29	72%	-2
Albacete	38	19	4	28	68%	15
					66%	12,1

Média do campeonato Espanhol

64%

Inglaterra

2004/2005		Pontos C	Pontos F	Total	%PontosC	C-F
Chelsea	38	47	48	95	49%	-1
Arsenal	38	44	39	83	53%	5
Man Utd	38	42	35	77	55%	7
Everton	38	38	23	61	62%	15
Bolton	38	32	26	58	55%	6
Liverpool	38	40	18	58	69%	22
Middlesbrough	38	33	22	55	60%	11
Man City	38	30	22	52	58%	8
Tottenham	38	32	20	52	62%	12
Aston Villa	38	30	17	47	64%	13
Charlton	38	28	18	46	61%	10
Birmingham	38	30	15	45	67%	15
Fulham	38	28	16	44	64%	12
Newcastle	38	28	16	44	64%	12
Blackburn	38	23	19	42	55%	4
Portsmouth	38	28	11	39	72%	17
West Brom	38	23	11	34	68%	12
Crystal Palace	38	23	10	33	70%	13
Norwich	38	26	7	33	79%	19
Southampton	38	24	8	32	75%	16
					63%	11,4

2003/2004		Pontos C	Pontos F	Total	%PontosC	C-F
Arsenal	38	49	41	90	54%	8
Chelsea	38	40	39	79	51%	1
Man Utd	38	40	35	75	53%	5
Liverpool	38	34	26	60	57%	8
Aston Villa	38	33	23	56	59%	10
Newcastle	38	38	18	56	68%	20
Bolton	38	26	27	53	49%	-1
Charlton	38	27	26	53	51%	1
Fulham	38	31	21	52	60%	10
Birmingham	38	29	21	50	58%	8
Middlesbrough	38	28	20	48	58%	8
Southampton	38	30	17	47	64%	13
Portsmouth	38	34	11	45	76%	23
Tottenham	38	31	14	45	69%	17
Blackburn	38	19	25	44	43%	-6
Man City	38	24	17	41	59%	7
Everton	38	29	10	39	74%	19
Leeds	38	22	11	33	67%	11
Leicester	38	19	14	33	58%	5
Wolverhampton	38	26	7	33	79%	19
					60%	9,3

2002/2003		Pontos C	Pontos F	Total	%PontosC	C-F
Man Utd	38	50	33	83	60%	17
Arsenal	38	47	31	78	60%	16
Newcastle	38	47	22	69	68%	25
Chelsea	38	41	26	67	61%	15
Liverpool	38	35	29	64	55%	6
Blackburn	38	34	26	60	57%	8
Everton	38	38	21	59	64%	17
Southampton	38	35	17	52	67%	18
Man City	38	29	22	51	57%	7
Tottenham	38	31	19	50	62%	12
Charlton	38	27	22	49	55%	5
Middlesbrough	38	37	12	49	76%	25
Birmingham	38	29	19	48	60%	10
Fulham	38	36	12	48	75%	24
Leeds	38	24	23	47	51%	1
Aston Villa	38	35	10	45	78%	25
Bolton	38	29	15	44	66%	14
West Ham	38	22	20	42	52%	2
West Brom	38	14	12	26	54%	2
Sunderland	38	11	8	19	58%	3
					62%	12,6

2001/2002		Pontos C	Pontos F	Total	%PontosC	C-F
Arsenal	38	40	47	87	46%	-7
Liverpool	38	41	39	80	51%	2
Man Utd	38	35	42	77	45%	-7
Newcastle	38	39	32	71	55%	7
Leeds	38	33	33	66	50%	0
Chelsea	38	37	27	64	58%	10
West Ham	38	40	13	53	75%	27
Aston Villa	38	31	19	50	62%	12
Tottenham	38	34	16	50	68%	18
Blackburn	38	30	16	46	65%	14
Middlesbrough	38	26	19	45	58%	7
Southampton	38	26	19	45	58%	7
Charlton	38	21	23	44	48%	-2
Fulham	38	28	16	44	64%	12
Everton	38	28	15	43	65%	13
Bolton	38	22	18	40	55%	4
Sunderland	38	28	12	40	70%	16
Ipswich	38	22	14	36	61%	8
Derby	38	19	11	30	63%	8
Leicester	38	16	12	28	57%	4
					59%	7,65

2000/2001		Pontos C	Pontos F	Total	%PontosC	C-F
Man Utd	38	47	33	80	59%	14
Arsenal	38	48	22	70	69%	26
Liverpool	38	43	26	69	62%	17
Leeds	38	36	32	68	53%	4
Ipswich	38	38	28	66	58%	10
Chelsea	38	42	19	61	69%	23
Sunderland	38	34	23	57	60%	11

Aston Villa	38	32	22	54	59%	10
Charlton	38	38	14	52	73%	24
Southampton	38	35	17	52	67%	18
Newcastle	38	34	17	51	67%	17
Tottenham	38	39	10	49	80%	29
Leicester	38	34	14	48	71%	20
Derby	38	31	11	42	74%	20
Everton	38	26	16	42	62%	10
Middlesbrough	38	19	23	42	45%	-4
West Ham	38	24	18	42	57%	6
Coventry	38	19	15	34	56%	4
Man City	38	15	19	34	44%	-4
Bradford	38	19	7	26	73%	12
					63%	13,35
Média do campeonato Inglês					61%	

Itália

2004/2005		Pontos C	Pontos F	Total	%PontosC	C-F
Juventus	38	47	39	86	55%	8
Milan	38	38	41	79	48%	-3
Inter	38	40	32	72	56%	8
Udinese	38	31	31	62	50%	0
Sampdoria	38	33	28	61	54%	5
Palermo	38	34	19	53	64%	15
Messina	38	37	11	48	77%	26
Roma	38	26	19	45	58%	7
Livorno	38	32	13	45	71%	19
Reggina	38	27	17	44	61%	10
Lecce	38	32	12	44	73%	20
Cagliari	38	36	8	44	82%	28
Lazio	38	24	20	44	55%	4
Siena	38	23	20	43	53%	3
Chievo	38	29	14	43	67%	15
Fiorentina	38	28	14	42	67%	14
Parma	38	33	9	42	79%	24
Bologna	38	25	17	42	60%	8
Brescia	38	21	20	41	51%	1
Atalanta	38	27	8	35	77%	19
					63%	11,55

2003/2004		Pontos C	Pontos F	Total	%PontosC	C-F
Milan	34	44	38	82	54%	6
Roma	34	40	31	71	56%	9
Juventus	34	40	29	69	58%	11
Inter	34	31	28	59	53%	3
Parma	34	32	26	58	55%	6
Lazio	34	34	22	56	61%	12
Udinese	34	25	25	50	50%	0
Sampdoria	34	28	18	46	61%	10
Chievo	34	21	23	44	48%	-2
Lecce	34	22	19	41	54%	3
Brescia	34	22	18	40	55%	4
Bologna	34	26	13	39	67%	13
Reggina	34	20	14	34	59%	6
Siena	34	25	9	34	74%	16
Perugia*	34	19	13	32	59%	6
Modena	34	21	9	30	70%	12
Empoli	34	24	6	30	80%	18
Ancona	34	12	1	13	92%	11
					61%	8

2002/2003		Pontos C	Pontos F	Total	%PontosC	C-F
JUVENTUS	34	40	32	72	56%	8
INTER	34	35	30	65	54%	5
MILAN	34	40	21	61	66%	19
LAZIO	34	29	31	60	48%	-2
PARMA	34	35	21	56	63%	14
UDINESE	34	38	18	56	68%	20
CHIEVO	34	36	19	55	65%	17
ROMA	34	32	17	49	65%	15
PERUGIA	34	31	11	42	74%	20
BRESCIA	34	25	17	42	60%	8
BOLOGNA	34	30	11	41	73%	19
EMPOLI	34	20	18	38	53%	2
MODENA	34	24	14	38	63%	10
REGGINA	34	29	9	38	76%	20
ATALANTA	34	23	15	38	61%	8
PIACENZA	34	22	8	30	73%	14
COMO	34	16	8	24	67%	8
TORINO	34	16	5	21	76%	11
					64%	12

2001/2002		Pontos C	Pontos F	Total	%PontosC	C-F
JUVENTUS	34	42	29	71	59%	13
ROMA	34	43	27	70	61%	16
INTER	34	36	33	69	52%	3
MILAN	34	29	26	55	53%	3
CHIEVO	34	32	22	54	59%	10
LAZIO	34	36	17	53	68%	19
BOLOGNA	34	38	14	52	73%	24
PERUGIA	34	34	12	46	74%	22
ATALANTA	34	24	21	45	53%	3
PARMA	34	32	12	44	73%	20
TORINO	34	26	17	43	60%	9
PIACENZA	34	25	17	42	60%	8
UDINESE	34	17	23	40	43%	-6
BRESCIA	34	23	17	40	58%	6
VERONA	34	30	9	39	77%	21
LECCE	34	15	13	28	54%	2
FIorentina	34	15	7	22	68%	8
VENEZIA	34	12	6	18	67%	6
					62%	10,38889

2000/2001		Pontos C	Pontos F	Total	%PontosC	C-F
ROMA	34	41	34	75	55%	7
JUVENTUS	34	38	35	73	52%	3
LAZIO	34	41	28	69	59%	13
PARMA	34	31	25	56	55%	6
INTERNAZIONALE	34	33	18	51	65%	15
MILAN	34	33	16	49	67%	17

ATALANTA	34	21	23	44	48%	-2
BRESCIA	34	28	16	44	64%	12
FIorentina	34	28	15	43	65%	13
BOLOGNA	34	31	12	43	72%	19
PERUGIA	34	26	16	42	62%	10
UDINESE	34	22	16	38	58%	6
LECCE	34	23	14	37	62%	9
VERONA	34	31	6	37	84%	25
REGGINA	34	28	9	37	76%	19
VICENZA	34	23	13	36	64%	10
NAPOLI	34	22	14	36	61%	8
BARI	34	17	3	20	85%	14
					64%	11,33333
				Média do campeonato de Itália	63%	

Edição Outubro 2010:
Inclusão das notas finais
dos trabalhos:

Este trabalho obteve 18
valores sendo o melhor.

U. PORTO**FEP** FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

Licenciatura em GESTÃO
LGE206 - Probabilidades e Estatística
Ano Lectivo 2005/2006
Exame de Recurso - 10/02/2006

CÓDIGO	NOME	EXAME	Trabalho	NOTA
		9,6	-	10
		11,4	14,0	12
		15,0	-	15
		15,6	-	16
		12,8	-	13
		9,6	13,0	11
		10,9	-	11
		11,7	13,0	12
		16,9	14,0	16
		10,2	-	10
		13,7	17,0	15
		7,7	-	8
		9,5	-	10
		0,3	11,0	4
		15,8	-	16
		14,5	8,0	13
		12,0	-	12
		2,7	15,0	6
		7,6	-	8
		3,1	-	3
		10,9	-	11
		18,7	-	DEF. NOTA
		5,4	-	5
		6,7	-	7
		11,5	-	12
		9,5	-	10

24-02-2006 19:47

Licenciatura em GESTÃO
 LGE206 - Probabilidades e Estatística
 Ano Lectivo 2005/2006
 Exame de Recurso - 10/02/2006

CÓDIGO	NOME	EXAME	Trabalho	NOTA
		11,1	-	11
		7,8	-	8
		8,8	7,0	8
		12,8	-	13
		5,7	10,0	7
		6,4	13,0	8
		13,3	11,0	13
		10,6	-	11
		desistiu	-	0
		5,6	-	6
		18,7	-	DEF. NOTA
		7,6	14,0	10
		5,8	-	6
		12,5	-	13
		13,5	-	14
		4,1	-	4
		11,8	-	12
		6,2	-	6
		3,2	-	3
		14,5	-	15
		15,1	-	15
		14,0	17,0	15
		11,5	-	12
		13,6	-	14
		15,5	-	16
		12,8	16,0	14

24-02-2006 19:47

U. PORTO**FEP** FACULDADE DE ECONOMIA
UNIVERSIDADE DO PORTO

Licenciatura em GESTÃO
LGE206 - Probabilidades e Estatística
Ano Lectivo 2005/2006
Exame de Recurso - 10/02/2006

CÓDIGO	NOME	EXAME	Trabalho	NOTA
		12,7	11,0	12
		9,1	-	9
		9,7	-	10
		16,7	-	17
		11,6	-	12
		9,5	-	10
		10,1	-	10
		11,0	-	11
		8,3	-	8
		13,2	-	13
		7,6	-	8
		2,4	8,0	4
	Paulo Rebelo	17,5	18,0	18
		15,7	10,0	14
		6,4	-	6
		14,4	12,0	14
		12,2	-	12
		6,7	-	7
		5,3	-	5
		8,4	-	8
		9,9	-	10
		9,7	11,0	10
		6,5	-	7
		5,9	-	6
		8,0	-	8
		6,8	-	7

24-02-2006 19:47

CÓDIGO	NOME	EXAME	Trabalho	NOTA
		9,6	10,0	10
		9,5	-	10
		11,7	-	12
		10,1	14,0	11
		18,3	15,0	17
		8,0	13,0	10
		11,7	-	12
		19,2	-	DEF. NOTA
		desistiu	-	

**Consulta da correcção dos pontos pelos alunos
 dia 2/03/2006, 5ª feira, 12:00-13:00, sala 221**

**A prova (escrita) para os alunos com nota "DEF. NOTA" terá lugar no dia
 3/03/2006, 6ª feira. Devem comparecer às 10 horas na sala 221.**

24-02-2006 19:47

Comentário dos Professores de Estatística, avaliadores do trabalho:

O trabalho prático proposto aos alunos da disciplina de Probabilidades e Estatística do 2º ano da Licenciatura em Gestão da Faculdade de Economia do Porto (do plano de estudos que vigorou entre 2004 e 2008) refere-se ao tópico de Análise Exploratória de Dados, e tinha como objectivo confrontar o aluno com uma aplicação a dados reais das técnicas aprendidas nas aulas da disciplina.

Os dados a analisar deveriam ser recolhidos pelos alunos, sendo utilizado software estatístico e/ou folhas de cálculo, para a sua análise. A correcta interpretação dos resultados obtidos era considerada primordial na avaliação do trabalho prático.

A experiência mostrou que os alunos aderiam com entusiasmo ao trabalho prático, tendo sido referidas apreciações positivas pela ligação que este lhes proporcionava entre a matéria leccionada e a vida real, permitindo nalguns casos a exploração de dados referentes a assuntos de interesse pessoal.

O aluno Paulo Rebelo manifestou um elevado interesse pela disciplina, encontrando nela um conjunto de ferramentas de grande utilidade prática, em particular para aplicação no seu hobby de apostas desportivas. O trabalho que apresentou “O Factor casa no Futebol” consistiu num estudo bastante completo dos resultados de vários campeonatos europeus de futebol, permitindo-lhe chegar a algumas conclusões, confirmando empiricamente as hipóteses avançadas. Este trabalho prático cumpriu sem margem para dúvida os objectivos pedagógicos propostos, demonstrando que o Paulo Rebelo dominou bem os conceitos teóricos pertinentes, sabendo aplicá-los de forma correcta e eficiente em situações reais, permitindo-lhe de seguida desenvolver uma actividade profissional em que utiliza a Estatística com sucesso.

Porto e FEP, 2 de Março de 2011

A equipa docente

Prof. Doutora Ana Cristina Freitas

Prof. Doutor Jorge Pereira

Prof. Doutora Paula Brito