

APLICAÇÃO DA METODOLOGIA DE BOX-JENKINS À SÉRIE TEMPORAL DE
TURISMO: DORMIDAS MENSAIS NA REGIÃO NORTE DE PORTUGAL

Paula Odete Fernandes^{*}
e-mail: pof@ipb.pt

Francisco José Terroso Cepeda^{**}
e-mail: cepeda@ipb.pt

RESUMO

Com o objectivo de aprofundar o conhecimento do sector turístico do Norte de Portugal e contribuir para o enriquecimento de estudos econométricos sobre este tema, efectuou-se este trabalho em que se aplicou a metodologia de *Box-Jenkins* ao tratamento e análise da série temporal de turismo: Dormidas Mensais na Região Norte de Portugal, correspondente ao período de Janeiro de 1987 a Dezembro de 1997, tendo sido a mesma construída com base na informação publicada pelo Instituto Nacional de Estatística (INE).

O modelo ajustado ARIMA, resultante da metodologia aplicada, apresentou qualidades estatística e de ajustamento aceitáveis, pelo que se procedeu à sua utilização para efeitos previsionais. Procedeu-se ainda, neste contexto, a uma comparação dos resultados obtidos com os realmente apurados pelo INE, para o período de 1998, concluindo-se que o modelo ajustado modeliza convenientemente a série de referência, dado que os valores observados são próximos dos resultados obtidos através do modelo estimado.

Palavras Chave: Estacionaridade, Sazonalidade, Estimação, Função de Autocorrelação, Função de Autocorrelação Parcial, Modelos ARIMA e Previsão.

* Prof.^a Adjunta do Departamento de Economia e Gestão da Escola Superior de Tecnologia e de Gestão/Instituto Politécnico de Bragança.

** Professor Coordenador com Agregação do Departamento de Economia e Gestão da Escola Superior de Tecnologia e de Gestão/Instituto Politécnico de Bragança.

1. INTRODUÇÃO

A actividade turística devido ao elevado número de bens e serviços que engloba, tem um peso significativo na economia de qualquer país, não só pelo elevado número de empregos que envolve, mas também pela capacidade de captação de divisas nos vários sectores da actividade económica a que está ligada.

Consequentemente, é fundamental a compreensão do fenómeno turístico em Portugal, para que as entidades envolvidas em cada sector de actividade possam efectivar medidas estratégicas capazes de tirar o máximo partido de situações emergentes. Daí o interesse de conhecer previsões da procura turística do Norte de Portugal. Para tal, aplicou-se a metodologia de *Box-Jenkins* tendo-se elegido a série temporal: Dormidas¹ Mensais da Região Norte de Portugal, dada que é uma das variáveis que melhor traduz a procura efectiva. Através do modelo *Box-Jenkins* univariado será feita a análise (com respectivas previsões) da série de referência com recurso ao conhecimento dos seus valores passados. Ao longo do presente trabalho esta série de referência passará a designar-se por DN_t .

2. MODELAÇÃO DE BOX-JENKINS PARA A SÉRIE DN_t

A metodologia de Box-Jenkins, permite efectuar uma análise ao comportamento das séries temporais, com base num duplo estudo e conjunto: por um lado uma componente autorregressiva que se estabelece em função da história estatística passada das variáveis consideradas e, por outro lado, um tratamento dos factores aleatórios ou estocásticos, especificado por processos de médias móveis. Estes modelos, devido ao seu esquema de delineamento e resolução operativa, permitem a incorporação de análises de carácter sazonal, o isolamento da componente tendência e, inclusivé aprofundar as inter-relações entre estas componentes que se integram na evolução das séries em estudo (Parra & Domingo, 1987; Chu, 1998).

Talaya e Palomo (1984), referem que a flexibilidade desta técnica permite introduzir, dentro dos modelos, variáveis artificiais que representem movimentos atípicos dentro das

¹ Entende-se por dormidas a permanência num estabelecimento que fornece alojamento, considerada em relação a cada indivíduo, e por um período compreendido entre as 12 horas de um dia e as 12 horas do dia seguinte.

séries produzidos por elementos ou factores externos, podendo-se optar por utilizar funções de transferências que incluam como “*inputs*” diversas variáveis explicativas reais ou fictícias.

Assim, a metodologia proposta por G. Box e G. Jenkins, em 1970, para o estudo das séries temporais, pressupõe as fases (Murteira, Müller & Turkman; 1993):

- Identificação do modelo, através de um número suficientemente grande de dados relativamente recentes;
- Estimação dos seus parâmetros após o modelo estar identificado;
- Avaliação do ajustamento atrás referido e, caso necessário, tentativa de encontrar um novo modelo que melhor descreva o fenómeno em causa;

como etapas primordiais precedentes à fase de previsão.

Este modelo permite escolher, de entre uma vasta classe de modelos algébricos, aquele que melhor se ajusta à série de valores que se possui. Trata-se dos chamados modelos ARIMA (*Autoregressive Integrated Moving Average*), bastante adequados para previsões de curto prazo e no caso de séries que contenham variações sazonais (Witt & Witt; 1992). No entanto, os modelos introduzidos por *Box* e *Jenkins* descrevem exclusivamente séries estacionárias, ou seja, com média e variância constantes no tempo e autocovariância dependente apenas do grau de defasamento entre as variáveis, devendo começar-se por verificar ou provocar a estacionaridade da série (Pulido; 1989).

3. MODELAÇÃO UNIVARIANTE DA SÉRIE DN_t

3.1. Apresentação da Série

A série DN_t , cujos dados disponíveis, recolhidos pelo INE, respeitam ao período compreendido entre Janeiro de 1987 e Dezembro de 1997, correspondendo a 132 dados mensais ao longo de 11 anos (ver Anexo I - Quadro 1). Esta série é considerada como significativa da actividade turística, devido às suas características, apenas contempla visitantes que pretendam usufruir das capacidades turísticas da região Norte de Portugal.

3.2. Estacionarização da Série

A análise da estacionaridade da série começa por ser intuitiva, resultante do conhecimento que se tem sobre a natureza do fenómeno em análise e da observação do cronograma da série.

Observando as figuras 1 e 2, verifica-se, facilmente, a existência de oscilações pouco irregulares, mas sugerem a não estabilização da média e a presença de sazonalidade (“picos” regularmente espaçados no tempo, máximos nos meses de Verão e mínimos nos meses de Inverno), ou seja, os valores das dormidas são em função da altura do ano.

Figura 1 – Cronograma da série DN_t , para o período de 1987:01 a 1997:12.

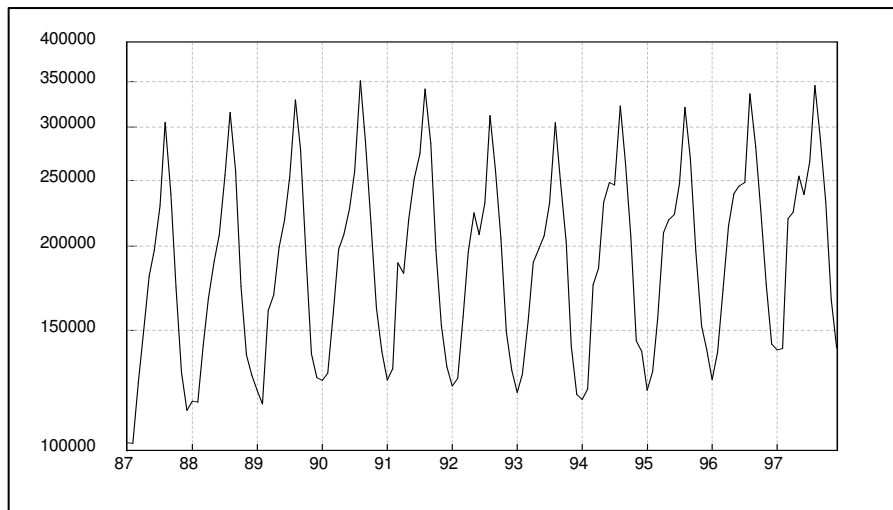
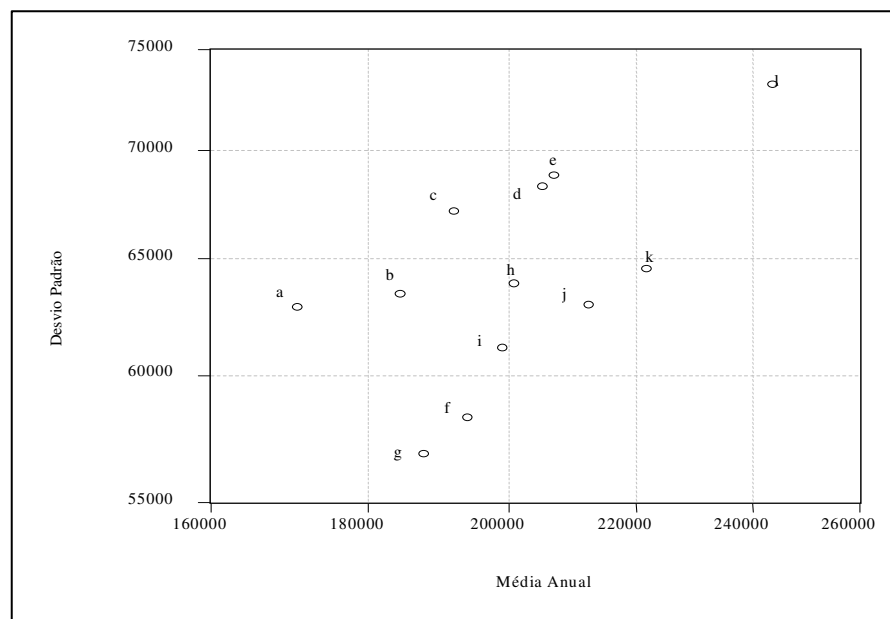


Figura 2 – Dispersão dos valores de DN_t .



Após a análise das FAC e FACP (funções autocorrelação e autocorrelação parcial) estimadas da série, chega-se a conclusões semelhantes à anterior. A FAC (figura 3) toma a forma de uma sinusóide amortecida, não decrescendo para zero à medida que k aumenta, o que confirma estar-se perante um processo não estacionário. A sazonalidade também é evidente da observação da FAC, que apresenta autocorrelações positivas elevadas para valores de k múltiplos de 12. Os picos ocorrem todos os anos nos meses de Agosto, sugerindo uma sazonalidade de período 12, dado estar-se a trabalhar com valores mensais. A FACP (figura 4) apresenta valores significativos para os dois primeiros *lags*, caindo em seguida para zero, evidenciando mais tarde alguns picos.

Figura 3 – FAC estimada da série DN_t .

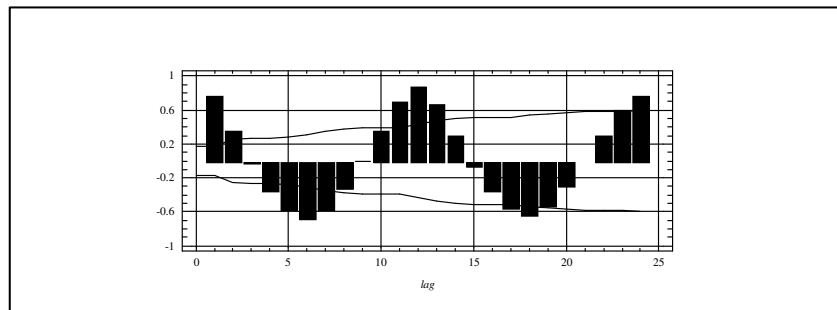
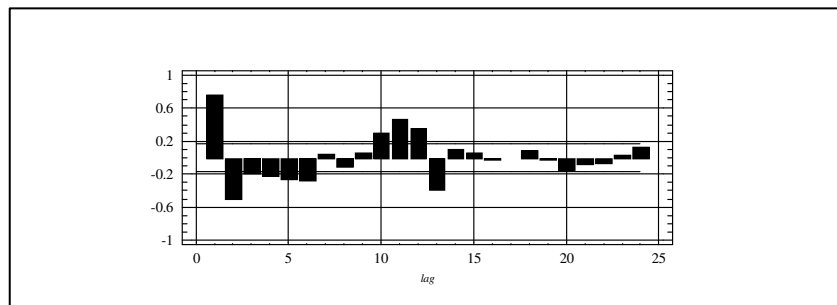


Figura 4 – FACP estimada da série DN_t .

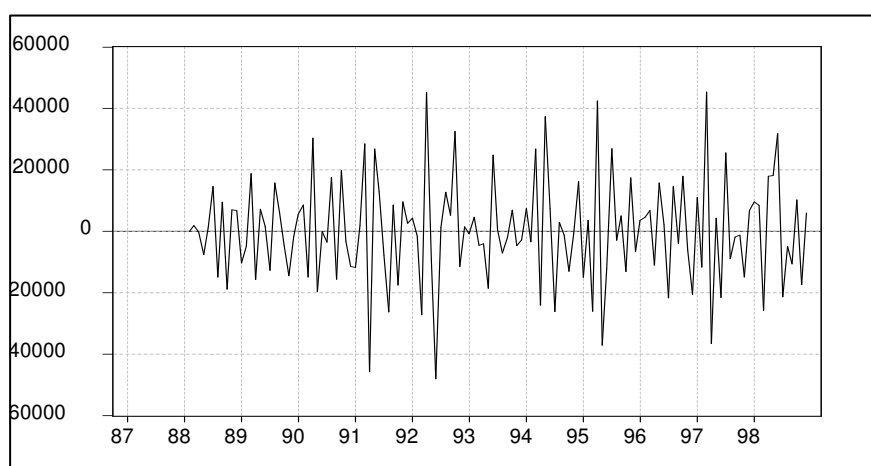


Para se aplicar a metodologia de *Box-Jenkins* é necessário em primeiro lugar identificar a série e remover a não estacionaridade. Assim, deve aplicar-se uma ou mais transformações sobre os valores da série de modo a obter uma outra série (transformada dos valores originais) estacionária. Tais transformações, apesar de conservarem a estrutura geral da série, exercem efeitos consideráveis sobre o conjunto de dados que podem tornar o seu estudo mais cómodo, alterando a sua escala (eventualmente diminuindo a sua amplitude), atenuando assimetrias, eliminando possíveis *outliers*, diminuindo resíduos e atingindo

finalmente os objectivos em causa: estabilizando variâncias e linearizando tendências (Otero; 1993).

A estacionarização (figura 5) da série em estudo, foi conseguida após a aplicação de uma diferenciação simples [$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t$] e uma diferença sazonal [$\nabla_s X_t = X_t - X_{t-s} = (1 - B^s) X_t$], à série inicial. É conveniente minimizar as diferenciações dos dados (para evitar a sobre-diferenciação ou *overdifferencing*), pois diferenciar origina aumento da variância do erro da previsão (Murteira, Müller & Turkman, 1993; González, 1999).

Figura 5 – Cronograma da série, após diferenciações simples e sazonal.



Assim, sendo DN_t a série inicial, a série que servirá de base ao estudo da identificação tem a forma: $[(1-B) (1-B^{12})DN_t]$, e o seus correlogramas (FAC e FACP) (figuras 6 e7).

3.2. Identificação do Modelo

Como instrumento básico de identificação do processo ARMA em causa, deverão ser consideradas as FAC e FACP estimadas da série estacionarizada, que podem ser observadas nas figuras que se seguem (considerando as bandas de significância $\pm 1.96 \sigma$).

Figura 6 – FAC estimada da série após diferenciações simples e sazonal.

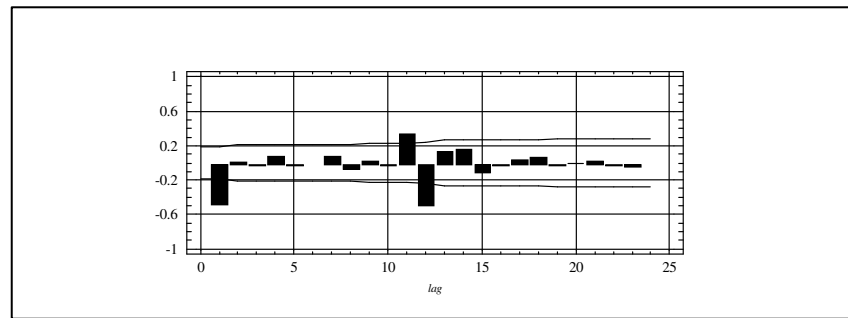
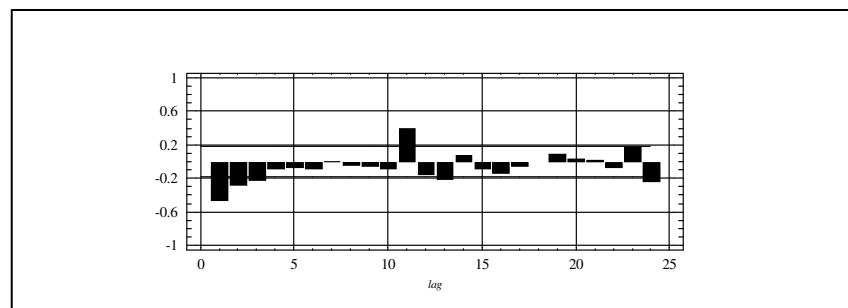


Figura 7 – FACP estimada da série após diferenciações simples e sazonal.



A identificação das componentes sazonal e não sazonal é feita separadamente recorrendo aos figurinos teóricos (Otero; 1993).

A análise das FAC e FACP, estimadas após diferenciações simples e sazonal, parecem sugerir apenas um processo, para a componente não sazonal, ARMA (0,1) visto que a primeira estimativa da FAC é significativa decaindo bruscamente para zero a partir do primeiro *lag* e os valores iniciais da FACP são significativos e a decair de uma forma exponencial. Relativamente, à componente sazonal, as FAC e FACP estimadas sugerem igualmente um processo ARMA (0,1) atendendo aos valores da FAC estimada no *lag* 12 e 24 (o primeiro é significativo, enquanto o segundo é sem qualquer expressão) e aos valores da FACP para os mesmos *lags*, ambos significativos.

Identificados todos os parâmetros do modelo, poder-se-á concluir que se trata de um modelo ARIMA (0,1,1) \times (0,1,1)₁₂ sobre a série,

$$\text{ou seja, } (1-B)(1-B^{12})DN_t = (1-\theta B)(1-\Theta B^{12})\epsilon_t.$$

3.3. Estimaco e validao do Modelo

Através do programa *Statgraphics Plus*, foram obtidos os seguintes resultados:

Quadro 1 – Estimativas obtidas para os coeficientes.

Factor	Lag	Coeficiente	Rácio t	P-Value
Médias Móveis	1	0.692241	10.1299	0.000000
Médias Móveis	12	0.744521	12.2922	0.000000

Assim, o modelo estimado vem dado pela expresso:

$$(1-B)(1-B^{12})DN_t = (1-0.692241B)(1-0.744521B^{12})\epsilon_t.$$

A anlise da significncia estatística dos coeficientes estimados do modelo, revela que os mesmos so, significativamente, diferentes de zero, ao nvel de significncia de 5%, ou seja, os rcios t para os parmetros estimados, levam a concluir que ambos os coeficientes so estatisticamente significativos.

Quanto à invertibilidade, a das duas componentes (no sazonal e sazonal), verificam as condies de invertibilidade, uma vez que as estimativas dos parmetros das componentes de mdias mveis so, em mdulo, inferiores à unidade.

A anlise da matriz das correlaes pode dar indicaes sobre a necessidade da rejeio do mesmo. Com efeito, ainda que os parmetros sejam significativos, eles sero inadequados se existir forte correlao entre eles, o que afecta a estabilidade do modelo, tornando-o inadequado para a previso (Murteira, Mller & Turkman; 1993).

Observando a matriz de correlaes (quadro 2) entre os parmetros conclui-se que existe muita fraca correlao entre eles, j que apenas correlaes superiores, em valor absoluto, a 0.7 assinalam parmetros fortemente correlacionados.

Quadro 2 – Matriz de correlaes entre os parmetros.

	MA (1)	MA (12)
MA (1)	1	-0.0648947
MA (12)	-0.0648947	1

A hiptese da existncia de redundncia, no se coloca neste modelo. Pela anlise ao teste de correlao (afasta a hiptese de os parmetros assumirem o valor 1) e anlise dos

coeficientes estimados (valores não muito próximos de 1), afastam, simultaneamente, a hipótese da existência de redundâncias. Modelos com redundâncias ou quase redundâncias são instáveis, dando origem a estimativas com pouca qualidade (Almargo; 1979).

Uma vez avaliada a qualidade estatística do modelo, interessa avaliar a qualidade do ajustamento, o que é feito com base na análise dos respectivos resíduos. Com efeito, se este explica correctamente a série em causa, os resíduos estimados deverão ter um comportamento análogo ao de um ruído branco (*white noise*), nomeadamente, em termos de FAC e FACP estimadas, isto é, sem valores significativos.

Observando as FAC e FACP estimadas residuais do modelo em causa (figuras 8 e 9), verifica-se a ausência de qualquer *lag* com valor significativo ao nível dos 5%.

Figura 8 – FAC estimada residual do modelo.

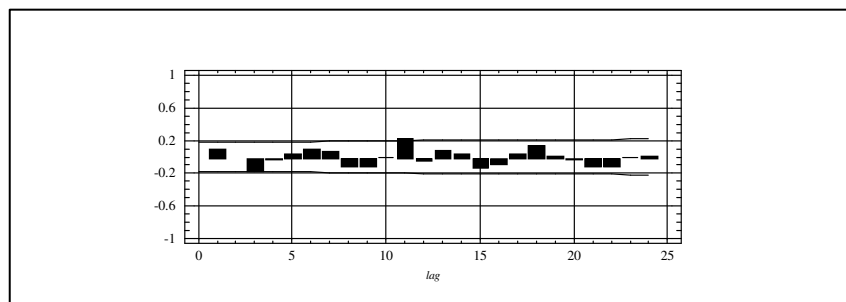
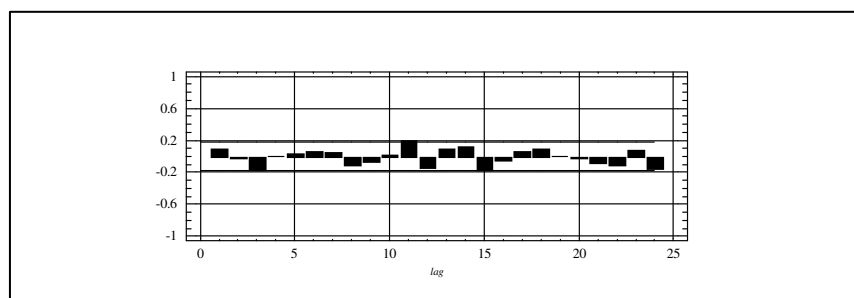


Figura 9 – FACP estimada residual do modelo.



A estatística de *Ljung-Box*, apresenta, para o *lag* 12, o valor de $Q=12.459$ e um *P-value* de 0.44, para o *lag* 24, o valor de $Q=21.066$ e *P-value*=0.635 e para o *lag* 36, o valor de $Q=27.642$ e um *P-value* de 0.84, pelo que se aceita a ideia de que os resíduos do modelo estimado se comportam como um ruído branco.

3.4. Previsão

A previsão de valores futuros de uma série, representa um dos objectivos mais importantes da sua análise. A mesma pode realizar-se através de diferentes metodologias, tendo em conta o tipo de utilização, a extensão (longo, médio e curto termo) e, principalmente a disponibilidade de dados (Harvey & Todd, 1983; Talaya, 1993; Talaya, 1996).

Encontrado o modelo que, aparentemente, melhor adere à evolução das Dormidas mensais no Norte de Portugal, registadas no período de 1987 a 1997, procura-se avaliar a sua capacidade preditiva, efectuando previsões para 1998 e 1999, dado que a metodologia de *Box-Jenkins* presta-se a previsões de curto prazo.

A função de previsão, que para cada *lag* $t+m$ introduz o preditor X_{t+m} , é dada por:

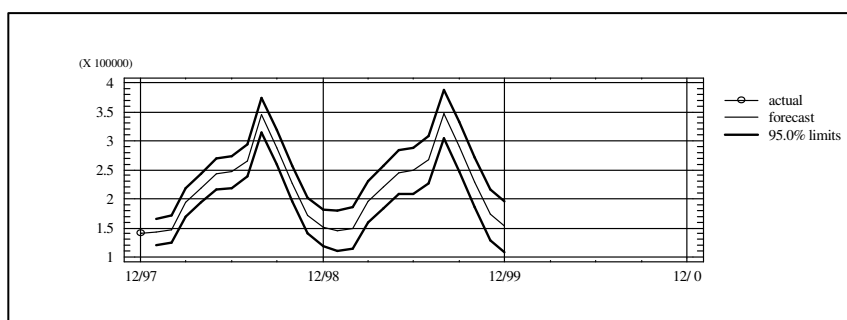
$$X_t(m) = E\{ X_{t+m} \mid X_t, X_{t-1}, X_{t-2}, \dots \}.$$

No quadro 3, apresentam-se as previsões e os respectivos limites inferior e superior do intervalo de confiança a 95%. Na figura 10 apresenta-se a evolução gráfica das previsões e dos limites a partir das observações mais recentes da série.

Quadro 3 – Previsões da série DN_t .

Período	Limite Inferior	Previsões	Limite Superior	Período	Limite Inferior	Previsões	Limite Superior
Jan-98	119,274	141,899	164,524	Jan-99	106,797	144,726	182,656
Fev-98	123,420	147,198	170,975	Fev-99	110,499	149,974	189,450
Mar-98	172,059	196,962	221,864	Mar-99	158,695	199,688	240,681
Abr-98	195,490	221,495	247,500	Abr-99	181,686	224,171	266,656
Mai-98	219,912	246,998	274,084	Mai-99	205,669	249,623	293,577
Jun-98	218,316	246,465	274,614	Jun-99	203,637	249,039	294,441
Jul-98	236,117	265,313	294,509	Jul-99	221,006	267,837	314,668
Ago-98	314,112	344,340	374,568	Ago-99	298,570	346,813	395,056
Set-98	257,247	288,495	319,742	Set-99	241,277	290,917	340,557
Out-98	196,295	228,549	260,804	Out-99	179,899	230,921	281,943
Nov-98	138,767	172,018	205,269	Nov-99	121,949	174,339	226,730
Dez-98	115,710	149,948	184,186	Dez-99	107,471	152,219	205,966

Figura 9 – Evolução das Previsões da série DN_t , para o período de 1998:01 a 1999:12.



Apreciando, agora em termos comparativos (Quadro 4) as previsões e as observações mensais para o ano de 1998, verifica-se que as previsões obtidas são próximas dos valores reais.

Quadro 4 – Erro da Previsão efectuada para a série DN_t .

Período	Observações^{a)}	Previsões	Erro da Previsão	% Erro
Jan-98	148,218	141,899	6,319	4.26
Fev-98	157,415	147,198	10,217	6.49
Mar-98	209,929	196,962	12,967	6.18
Abr-98	232,767	221,495	11,272	4.84
Mai-98	280,326	246,998	33,328	11.89
Jun-98	296,612	246,465	50,147	16.91
Jul-98	303,866	265,313	38,553	12.69
Ago-98	377,645	344,340	33,305	8.82
Set-98	309,700	288,495	21,205	6.85
Out-98	263,522	228,549	34,973	13.27
Nov-98	180,796	172,018	8,778	4.86
Dez-98	161,273	149,948	11,325	7.02

^{a)} dados fornecidos pelo INE.

Repare-se que os valores previstos para o ano de 1998, revelam-se “subavaliados” quando comparados com os valores das dormidas realmente apuradas no mesmo período, sendo de destacar o período de Maio a Outubro, o que sugerem que, na realidade, o ano de 1998 tenha ultrapassado as expectativas, revelando-se assim um bom ano no domínio do turismo. A plausibilidade desta hipótese poderá ser sustentada pelo facto de, em 1998, Portugal ter sido palco da exposição mundial, EXPO’98, o que sem dúvida, terá constituído um pólo adicional de atracção e promoção turística.

4. CONSIDERAÇÕES FINAIS

Este trabalho teve como objectivo aprofundar o conhecimento do sector turístico do Norte de Portugal e contribuir para o enriquecimento de estudos econométricos sobre este tema. Utilizou-se para isso a modelização de *Box-Jenkins* à série Dormidas mensais na Região Norte de Portugal, com a pretensão de comparar as suas previsões obtidas com as observadas.

O modelo que se estimou tem validade para o conjunto de dados que lhe serviu de suporte. Descreve uma realidade, transmitida pela série anteriormente referida, no período de Janeiro de 1987 a Dezembro de 1997, fornecida pelo INE, tendo permitido efectuar previsões das dormidas mensais para os anos de 1998 e 1999. Os valores observados, para o período de 1998, são próximos dos obtidos pelo modelo estimado, para o mesmo período. Terá todo interesse verificar se acontece o mesmo para o período de 1999, aquando da publicação oficial dos dados pelo INE.

Como apreciação final, resta salientar que sempre que se trate de efectuar previsões, torna-se necessário dispor de informações estatísticas, atempadamente, ou corre-se o risco de não se poder utilizar os resultados das previsões antes da ocorrência dos fenómenos. Esta é uma questão que se coloca com particular aquidade na gestão ou na economia, pois quem investiga está sempre dependente do sistema oficial de recolha e publicação dos dados estatísticos.

5. BIBLIOGRAFIA

- ALMARGO, José; (1979); “Aplicaciones del enfoque Box-Jenkins a series del Turismo Español”; *Cuadernos Economicos de Información Comercial Española*; nº 11/12, pp. 53/107.
- CHU, Fong-Lin; (1998); “Forecasting Tourist Arrivals: nonlinear sine wave or ARIMA?”; *Journal of Travel Research*; Vol. 36; pp. 79/84.
- GONZÁLEZ, Carlos Sánchez; (1999); “*Métodos Econometricos*”; Editorial Ariel; Barcelona.
- HARVEY, A.C. and Todd, P.H.J.; (1983); “Forecasting Economic Time Series with structural and Box-Jenkins Models: a case study”; *Journal of Business y Economic Statistics*; Vol. 1, nº 4 ; pp. 299/307.
- MURTEIRA, Bento J.F., Müller, Daniel A. e Turkman, K. Feridun; (1993); “*Análise de sucessões cronológicas*”; McGraw-Hill; Lisboa.
- OTERO, José M^a; (1993); “*Econometría – series temporales y predicción*”; Editorial AC; Madrid.
- PARRA, S. B. y Domingo, J. U.; (1987); “Analise de series temporales de turismo de la Comunidad Valenciana”; *Estadística Española*; nº 114; pp. 111/132.
- PULIDO, Antonio; (1989); “*Predicción Económica y Empresarial*”; Ediciones Pirámide; Madrid.
- TALAYA, A.E. y Palomo, M.F.; (1984); “Técnicas de previsión y análisis de comportamiento de la demanda”; *Estudios Turisticos*; nº84; pp.3/16.
- TALAYA, Agueda Esteban; (1993); “Previsiones Turisticas: Metodologia y Resultados”; *Estudios Turisticos*; nº118; pp.57/71.
- TALAYA, Agueda Esteban; (1996); “Previsiones de la demanda turistica”; *Revista de Economía-Información Comercial Española*; nº749; pp.89/97.
- WITT, Stephen F. and Witt, Christine A.; (1995); “Forecasting tourism demand: a review of empirical research”; *International Journal of Forecasting*; nº 11; pp. 447/475.

ANEXO I

Quadro 1 – Dormidas mensais na Região Norte de Portugal, para o período de 1987:01 a 1997:12.

Mês	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Jan	102,447	118,011	122,217	126,671	126,826	124,194	121,469	118,606	122,480	126,910	140,430
Fev	102,123	117,547	116,837	129,802	131,653	127,474	129,284	122,988	130,393	139,403	141,183
Mar	125,401	142,687	160,658	158,701	188,999	157,536	154,734	175,261	156,645	172,393	219,465
Abr	150,042	167,118	169,326	197,757	182,290	196,087	189,142	185,525	209,263	213,973	224,382
Mai	180,430	189,823	199,158	207,876	219,187	223,918	198,402	232,075	218,666	239,142	253,833
Jun	197,113	207,729	218,595	227,159	251,295	207,907	207,216	248,237	222,720	245,264	238,334
Jul	229,293	254,523	252,634	257,633	273,927	231,801	231,453	246,274	247,589	248,398	266,993
Ago	304,847	315,113	329,014	351,500	341,490	312,026	304,576	322,366	320,750	336,086	345,672
Set	238,542	258,287	278,074	284,867	283,378	259,023	249,583	266,094	269,433	280,769	288,409
Out	173,503	174,359	189,664	216,286	197,241	205,400	202,792	206,256	196,466	225,734	232,052
Nov	130,187	137,933	138,683	162,062	152,554	149,289	141,976	144,803	152,340	175,438	166,835
Dez	114,229	128,774	127,730	139,683	132,802	130,963	120,748	139,706	140,643	143,163	141,349
Total	2,048,157	2,211,904	2,302,590	2,459,997	2,481,642	2,325,618	2,251,375	2,408,191	2,387,388	2,546,673	2,658,937

Fonte: Elaboração própria, com base nos dados fornecidos pelo INE.