UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

DEPARTAMENTO DE ESTATÍSTICA

TRABALHO DE CONCLUSÃO DE CURSO

ALUNO: RICARDO FERREIRA VITELLI

PROFESSOR ORIENTADOR : SÉRGIO FISCHER

APLICAÇÃO DE MODELOS DE PREVISÃO EM SÉRIES TEMPORAIS

1 - INTRODUÇÃO

A Secretaria Municipal dos Transportes é o orgão municipal responsável pelo cálculo do preço das tarifas de ônibus no município de Porto Alegre.

A estrutura da base tarifária utilizada constitui-se de dois agrupamentos específicos constituidos pelas seguintes parcelas:

a) Custos diretos de operação do veículo (custo independente)

b) Custos indiretos, incorridos na operaçao do veículo (custos dependentes).

Esta definição utilizando os termos custo independente e custo dependente, não esta relacionada com a definição de dependência ou independência em Estatística, é apenas uma questão de nomenclatura da SMT.

As parcelas incluídas no custo independente foram as seguintes:

- Consumo de combustíveis;
- consumo de óleo e lubrificantes;
- consumo de pneumáticos e câmaras;
- consumo de peças e acessórios.

As parcelas incluídas no custo dependente foram as seguintes:

- Depreciação dos veículos;
- Remuneração do capital investido;
- despesas administrativas;
- pessoal de operação.

O primeiro agrupamento de parcelas, refere-se às despesas não rateáveis aos diversos veículos. O segundo é vinculado na medida da utilização dos veículos.

2 - EQUIPE DE CADASTRO E REGISTRO

A equipe de cadastro e registro é o orgão de coordenação e controle das atividades pertinentes a cadastro e registro de equipamento de sinalização e de pessoal e Estatística, relativas ao transporte urbano do município.

A equipe de cadastro e registro compreende:

a) Setor de Cadastro e Registro;

b) Setor de Estatística.

Segundo determinação da Secretaria Municipal dos Transportes, as atividades destes setores são as seguintes:

2.1. O setor de cadastro e registro é o orgão encarregado de registrar toda a sinalização vertical, horizontal e semafórica realizada, bem como suas alterações, e pessoal ocupado no transporte público.

2.2. O setor de Estatística é o orgão encarregado do conjunto de atividades que busca estudar o universo peculiar do objeto do transporte público e do tráfego urbano, utilizando as técnicas de amostragem probabilística com vistas à influência das estatísticas das populações analisadas.

Ao setor de Estatística compete:

a) estabelecer normas, coletar e coordenar o levantamento de dados estatísticos;

b) fornecer dados estatísticos para a supervisão de operações, a fim de servir de suporte à elaboração de seus programas de trabalho;

c) observar as normas e diretrizes emanadas da Secretaria de Planejamento do Município, na apresentação de dados estatísticos;

d) estudar, participar e estimular procedimentos que visem simplificar a reprodução sistemática de informações através de processamento eletrônico de dados;

e) elaborar planilhas diversas para levantamento de passageiros, fluxos de veículos e pedestres, acidentes de trânsito e outros, para posterior representação gráfica comparativa;

f) articular-se com a SSP-RS- Secretaria de Segurança Pública- RS- para coleta de dados estatísticos de acidentes de trânsito na capital;

g) exercer atividades pertinentes ou que lhe forem delegadas.

3 - PROBLEMA

Na Secretaria Municipal dos Transportes existiam duas séries temporais com dados mensais relativos a número de passageiros transportados no município de Porto Alegre e quilometragem mensal percorrida pelos ônibus no mesmo município.

É importante para a Secretaria Municipal dos Transportes, os valores futuros da série e portanto coube a mim, no meu período de estágio efetuar a previsão para os valores das séries para os 12 meses do ano de 1985.

4 - OBJETIVO

O objetivo desse trabalho de conclusão do curso de Bacharel em Estatística da Universidade Federal do Rio Grande do Sul, é o de obter um modelo estatístico capaz de melhor gerar previsões para as séries citadas anteriormente e que seja passível de utilização por parte do setor de Estatística da Secretaria Municipal dos Transportes do Município de Porto Alegre.

A escolha do modelo ficou condicionada a necessidade de não utilizar técnicas estatísticas muito complexas, pois é de conveniência que ele possa ser usado por parte da equipe técnica da Secretaria Municipal dos Transportes. Em vista destas condições, optou-se por utilizar o modelo clássico de decomposição, por este ser um modelo de fácil utilização e compreenção por parte de pessoas que não tenham conhecimentos estatísticos avançados.

5 - PROCEDIMENTO

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas.

Qualquer que seja a classificação que façamos para os modelos de séries temporais, podemos considerar um número muito grande de modelos diferentes para descrever o comportamento de uma série particular.

Quanto a previsão em séries temporais, esta não constitui um fim em si, mas apenas um meio de fornecer informações para uma consequente tomada de decisões, visando a determinados objetivos.

Os procedimentos de previsão utilizados na prática variam muito, podendo ser simples e intuitivos ou mais quantitativos e complexos.

Ossprocessos de previsão com modelos de séries de tempo são procedimentos que visam a extender a valores futuros o modelo descrito e ajustado aos valores passados e ao valor presente da variável. Portanto a previsão se torna o cálculo do valor esperado de uma futura observação, condicionado aos valores passados e ao valor presente da variável.

Ou seja, chamando de $\hat{X}_t(h)$ o valor previsto estimado para um horizonte de "h" períodos de tempo futuros e "t" o período de origem da previsão, então,

$$\hat{X}_t(h)=E(X_{t+h}/\ X_t,\ X_{t-1},\ \ldots)$$

5.1. Série do número de passageiros transportados.

Inicialmente vou enunciar o procedimento para a estimação da série relativa ao número de passageiros transportados no Município de Porto Alegre (ANEXO) relativos mês a mês.

A série primeiramente possuia 94 observações mas os dados utilizados foram relativos somente aos 49 últimos valores, isto devido ao fato de que os escolares não eram utilizados na contagem do número de passageiros nos valores da série anterior ao mês de Outubro de 1980.

Logo a série agora com 49 observações, foi plotada (GRÁFICO ANEXO) e pode-se observar que a mesma apresentava sazonalidade.

O teste feito para se verificar a existência da componente sazonal foi o teste não-paramétrico de Friedman.

Neste teste os meses são considerados tratamentos e os anos blocos . Os tratamentos são ordenados dentro dos blocos associando se a cada elemento do bloco o valor de sua ordem (em sentido crescente).

Testa-se então as seguintes hipóteses:

H_o: $a_2=a_3=\ldots=a_{13}=0$

H_1: Todos os parâmetros são diferentes de zero.

Agora calcula-se a seguinte estatística:

$$T_2 = \frac{12}{pk(k+1)} \sum_{j=1}^{k} R^2_{.j} \; -3p\,(k+1) \, ,$$

onde: p= número de blocos (para se usar blocos completos utilizou-se 3 anos)

k = número de tratamentos

$R_{.j}$ = soma dos postos (ordens) no j-ésimo bloco

Assim a estatística encontrada foi:

$T_2 = 27,87$

Para se verificar a significância do teste compara-se o valor de T_2 com o valor tabelado para o teste a 95% de confiança. Como não há valor tabelado para k=12, então aproxima-se o valor da estatística do teste a uma distribuição χ^2 com k-1 graus de liberdade.

$\chi^2_{11(0,05)} = 19,70 < T_2 = 27,87$ então rejeita-se H_0, isto é, aceita-se a existência da componente sazonal na série com 95% de confiança.

Como era de interesse da Secretaria Municipal dos Transportes utilizar o modelo deforma que esse pudesse ser alimentado e de fácil utilização, o modelo escolhido para se prever os valores futuros da série foi o modelo clássico de decomposição.

Agora vou espor algumas considerações sobre este modelo.

Este modelo é decomposto em quatro componentes descritas a seguir.

a) Componente tendencial T(t)

A tendência, também chamada por alguns autores de tendência secular, é caracterizada como aquele movimento regular e contínuo de longo prazo, refletindo um movimento ascendente ou descendente em longo período de tempo. De certa forma, pode ser vista como aquela componente que descreve as variações graduais que se mantêm em um longo período de observação da variável no tempo. Estatisticamente, pode-se caracterizá-la pelo fato de que a expectância de Y, E(Y), varia no tempo.

b) Componente Sazonal S(t)

As variações sazonais são aquelas variações periódicas (cíclicas) que ocorrem com certa regularidade dentro de um curto período de tempo. Embora o próprio nome dê a entender que esses movimentos ocorrem por um período anual, de acordo com as estações climáticas, esses movimentos podem ser estendidos a qualquer intervalo de curto prazo, como diário, horário, semanal, mensal, trimestral, etc.

c) Componente Cíclica C(t)

As variações cíclicas são aquelas variações que se referem as oscilações de longo prazo que caracterizam, em geral, os ciclos econômicos. São as flutuaçõesde longo prazo em torno da curva de tendência.

d) Componente Aleatória a_t

As variações aleatórias, também chamadas residuais, referem-se não só aqueles movimentos esporádicos ocasionados por eventos aleatórios imprevisíveis, tais como as calamidades da natureza , mas também ao conjunto de todos aqueles movimentos da série que não forem passíveis de identificação em seus demais componentes , uma vez que não obedecem a nenhuma lei comportamental capaz de ser descrita de forma determinística, através de relações funcionais exclusivamente matemáticas.

Para a composição do modelo clássico a componente cíclica foi desconsiderada, por abranger um perído intervalar de tempo muito curto para que se pudesse perceber a presença dessa componente.

A componente tendencial foi obtida através da escolha da função de regressão entre quatro seguintes:

a) Função reta do tipo: $Y = a_o + a_1X + e$

b) Função Parábola do tipo: $Y = a_o + a_1X + a_2X^2 + e$

c) Função exponencial do tipo: $Y = a_o a_1^X e$

d) Função Logarítmica do tipo: $Y = a_o X^{a_1} e$

Onde:

Y é a variável número de passageiros

X é a variável tempo

e "e" é a componente aleatória, tal que E N(0,1)

A escolha do modelo baseou-se no critério de menor variância para as seguintes séries:

a) Reta

$$VAR\left[\frac{\Delta Y}{\Delta X}\right] = 7.10^{12}$$

b) Parábola

$$VAR\left[\frac{\Delta^2 Y}{\Delta X^2}\right] = 2.10^{13}$$

c) Exponencial

$$VAR\left[\frac{\Delta \text{ Log } Y}{\Delta X}\right] = 0,0025$$

d) Logarítmica

$$VAR\left[\frac{\Delta \text{ Log } Y}{\Delta \text{ Log } X}\right] = 9.2343$$

Onde:
$$\Delta W = W_t - W_{t+1}$$
$$\Delta^2 W = \Delta(\Delta W)$$
$$(\Delta W)^2 = (W_t - W_{t+1})^2$$

No processo de cálculo, obteve-se a seguinte função:

$\hat{Y} = 26\ 925\ 584{,}99(\ 0{,}9966012529\ )^{X}$,

oriunda da solução do seguinte sistema de equações:

$$n \text{ Log } a_0 + \text{Log } a_1 \sum_{i=1}^{m} X_i = \sum_{i=1}^{m} \text{Log } Y_i$$

$$\text{Log } a_0 \sum_{i=1}^{m} X_i + \text{Log } a_1 \sum_{i=1}^{m} X_i^2 = \sum_{i=1}^{m} (X_i \text{ Log } Y_i)$$

Onde:

n é o número de termos da série

X é a variável tempo

Y é a variável número de passageiros

Para a escolha do modelo deve-se considerar que o modelo aditivo é adequado quando a componente sazonal não depende das outras componentes, como a componente tendencial.

Primeiramente ajustou-se a série a um modelo multiplicativo.

Para a obtenção dos índices de sazonalidade no modelo utilizou-se uma média móvel de tamanho 12. Como os valores não ficaram centrados fez-se uma média móvel de tamanho dois. A partir desse processo, obteve-se os seguintes índices sazonais:

Mês	Índice sazonal
Jan.	0,943011979
Fev.	0.806614738
Mar.	1,075964979

Mês	Índice sazonal
Abr.	0,987195928
Maio	1,038357888
Jun.	0,981146057
Jul.	1,009715980
Ago.	1,062703443
Set.	0,989872095
Out.	1,027212735
Nov.	1,003510538
Dez.	1,074693643

Obs: Estes ídices sazonais foram ajustados de forma que eles somem 12.

Estimando-se os valores da série encontrou-se um erro quadrático médio da ordem de $3{,}74185.10^{11}$.

Utilizou-se também um modelo com ouso de variáveis enumerativas (DUMMY VARIABLES) , que servem como fatores de ajustamento para a sazonalidade. No caso, as variáveis enumerativas assumem somente dois valores 0 ou 1.

Dessa forma o modelo passa a ser um modelo múltiplo da forma:

$$\hat{Y} = a_1 a_2^X \prod_{i=3}^{13} a_i^{b_i}$$

Onde os b_i são as variáveis enumerativas e os a_i para i maior que dois são os índices de sazonalidade.

Para melhor se entender o modelo deve-se observar através da tabela 1 na página seguinte, de que forma foram utilizadas as variáveis enumerativas.

Tabela com os valores da série e as variáveis enumerativas.

TABELA 1

t	Y(t)	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
1	28 049 167	0	0	0	0	0	0	0	0	0	0	0
2	26 444 380	1	0	0	0	0	0	0	0	0	0	0
3	28 199 158	0	1	0	0	0	0	0	0	0	0	0
4	25 784 804	0	0	1	0	0	0	0	0	0	0	0
5	21 872 035	0	0	0	1	0	0	0	0	0	0	0
6	27 581 215	0	0	0	0	1	0	0	0	0	0	0
7	26 358 300	0	0	0	0	0	1	0	0	0	0	0
8	27 405 289	0	0	0	0	0	0	1	0	0	0	0
9	26 466 381	0	0	0	0	0	0	0	1	0	0	0
10	27 296 266	0	0	0	0	0	0	0	0	1	0	0
11	27 118 171	0	0	0	0	0	0	0	0	0	1	0
12	25 436 365	0	0	0	0	0	0	0	0	0	0	1
13	27 183 723	0	0	0	0	0	0	0	0	0	0	0
14	26 019 715	1	0	0	0	0	0	0	0	0	0	0
15	27 193 754	0	1	0	0	0	0	0	0	0	0	0
16	24 319 670	0	0	1	0	0	0	0	0	0	0	0
17	20 529 113	0	0	0	1	0	0	0	0	0	0	0
18	28 310 551	0	0	0	0	1	0	0	0	0	0	0
19	25 211 255	0	0	0	0	0	1	0	0	0	0	0
20	26 003 741	0	0	0	0	0	0	1	0	0	0	0
21	24 182 466	0	0	0	0	0	0	0	1	0	0	0
22	25 819 330	0	0	0	0	0	0	0	0	1	0	0
23	26 910 746	0	0	0	0	0	0	0	0	0	1	0
24	24 623 300	0	0	0	0	0	0	0	0	0	0	1
25	25 199 993	0	0	0	0	0	0	0	0	0	0	0
26	24 888 636	1	0	0	0	0	0	0	0	0	0	0
27	26 712 649	0	1	0	0	0	0	0	0	0	0	0
28	22 554 710	0	0	1	0	0	0	0	0	0	0	0
29	18 798 948	0	0	0	1	0	0	0	0	0	0	0
30	26 878 884	0	0	0	0	1	0	0	0	0	0	0
31	23 462 178	0	0	0	0	0	1	0	0	0	0	0
32	25 022 266	0	0	0	0	0	0	1	0	0	0	0
33	23 248 568	0	0	0	0	0	0	0	1	0	0	0
34	22 727 771	0	0	0	0	0	0	0	0	1	0	0
35	25 397 800	0	0	0	0	0	0	0	0	0	1	0

CONT:

t	Y(t)	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
36	23 678 409	0	0	0	0	0	0	0	0	0	0	1
37	23 934 030	0	0	0	0	0	0	0	0	0	0	0
38	23 374 180	1	0	0	0	0	0	0	0	0	0	0
39	25 335 663	0	1	0	0	0	0	0	0	0	0	0
40	22 431 878	0	0	1	0	0	0	0	0	0	0	0
41	19 719 271	0	0	0	1	0	0	0	0	0	0	0
42	23 463 118	0	0	0	0	1	0	0	0	0	0	0
43	22 798 049	0	0	0	0	0	1	0	0	0	0	0
44	23 853 141	0	0	0	0	0	0	1	0	0	0	0
45	22 144 440	0	0	0	0	0	0	0	1	0	0	0
46	23 659 199	0	0	0	0	0	0	0	0	1	0	0
47	24 915 349	0	0	0	0	0	0	0	0	0	1	0
48	23 117 017	0	0	0	0	0	0	0	0	0	0	1
49	24 893 536	0	0	0	0	0	0	0	0	0	0	0

FONTE: Dados brutos - SMT.

A obtenção dos parâmetros do modelo foi feita através de uma máquina HP 85 que utiliza regressão linear múltipla, portanto foi necessário linearizar-se o modelo da seguinte forma:

$$Y = a_1 a_2^x \prod_{i=3}^{13} a_i^{b_i}$$

$$\text{Log } Y = \text{Log}\left[a_1 a_2^x \prod_{i=3}^{13} a_i^{b_i}\right]$$

$$\text{Log } Y = \text{Log } a_1 + \text{Log } a_2^x + \sum_{i=3}^{13} \text{Log } a_i^{b_i}$$

$$\text{Log } Y = \text{Log } a_1 + x \text{ Log } a_2 + \sum_{i=3}^{13} (b_i \log a_i)$$

Desta maneira pode-se usar o modelo como regressão linear.

A partir desta regressão, obteve-se os seguintes resultados:

- Coeficiente de regressão múltipla (R^2) da ordem de 92,93 por cento, isto é, as variações das variáveis enumerativas e da variável tempo explicam 92,93% das variações da variável número de passageiros transportados.

- Os parâmetros encontrados foram os seguintes:

$\text{Log } a_1 = 7{,}451040583$	$a_1 = 28\ 251\ 439{,}6$
$\text{Log } a_2 = -1{,}571767308\ 10^{-3}$	$a_2 = 0{,}9963874132$
$\text{Log } a_3 = -1{,}901198653\ 10^{-2}$	$a_3 = 0{,}9571676533$
$\text{Log } a_4 = 1{,}075503078\ 10^{-2}$	$a_4 = 1{,}0250735580$
$\text{Log } a_5 = -4{,}110545191\ 10^{-2}$	$a_5 = 0{,}9096923617$
$\text{Log } a_6 = -1{,}095689346\ 10^{-1}$	$a_6 = 0{,}7770179748$
$\text{Log } a_7 = 9{,}771332704\ 10^{-3}$	$a_7 = 1{,}0227543440$
$\text{Log } a_8 = -2{,}405314999\ 10^{-2}$	$a_8 = 0{,}9461235444$
$\text{Log } a_9 = -2{,}988882681\ 10^{-3}$	$a_9 = 0{,}9931414712$
$\text{Log } a_{10} = -2{,}913861537\ 10^{-2}$	$a_{10} = 0{,}9351071651$

$\text{Log } a_{11} = -1{,}237984807 \; 10^{-2}$ $a_{11} = 0{,}9718967995$

$\text{Log } a_{12} = 1{,}065291924 \; 10^{-2}$ $a_{12} = 1{,}0248325701$

$\text{Log } a_{13} = -2{,}011606345 \; 10^{-2}$ $a_{13} = 0{,}9547374023$

O erro quadrático médio obtido foi $3{,}38963.10^{11}$

Assim sendo o modelo utilizando-se das variáveis enumerativas para se estimar os índices sazonais apresentou um menor erro quadrático médio em relação ao modelo clássico de decomposição, onde os índices sazonais foram obtidos por médias móveis.

Testou-se também a eficiência de um modelo de regressão linear do tipo:

$$Y = a_1 + a_2 X + \sum_{i=3}^{13} a_i b_i$$

Utilizando-se novamente a máquina HP 85 obteve-se as seguintes soluções para:

- Coeficiente de regressão múltipla igual a 92,69%, isto é, a partir deste modelo, as variações das variáveis enumerativas e da variável tempo, explica-se 92,69% das variações da variável número de passageiros transportados.

- Os parâmetros encontrados foram os seguintes:

a_1 = 28 100 933,84

a_2 = -89 953,76152

a_3 = -1 120 130,858

a_4 = 648 411,4039

a_5 = -2 349 185,584

a_6 = -5 802 155,573

a_7 = 616 398,4385

a_8 = -1 394 644,3

a_9 = -191 026,7785

$a_{10} = -1661\ 718{,}527$

$a_{11} = -706\ 587{,}0154$

$a_{12} = 593\ 241{,}7461$

$a_{13} = -1\ 188\ 548{,}242$

- O erro quadrático médio obtido foi da ordem de $3{,}41756.10^{11}$.

Comparando-se estes dois modelos verifica-se que não há diferença significativa do valor de R^2. Assim sendo, optou-se pelo de regressão linear múltipla por ser de melhor acesso o seu uso por parte do corpo técnico da Secretaria Municipal dos Transportes.

Assim, segue-se que o modelo finalmente proposto para se prever valores futuros do número de passageiros transportados fica sendo o seguinte:

$$Y = a_1 + a_2 X + \sum_{i=3}^{13} a_i b_i + e$$

Onde os valores dos parâmetros já foram expostos anteriormente.

5.2. Série da quilometragem percorrida pelos ônibus no Município de Porto Alegre.

Esta série é composto de 49 observações (ANEXO) e o procedimento inicial foi o de plotar os valores da série (GRÁFICO ANEXO) e observar o seu comportamento.

Inicialmente foi feito um teste de Friedman para se verificar a existência ou não da componente sazonal na série.

Neste teste os meses são considerados tratamentos e os anos blocos. Os tratamentos são ordenados dentro dos blocos.

Testa-se as seguintes hipóteses:

H_0: $a_2=a_3=\ldots=a_{13}=0$

H_1: Todos os parâmetros são diferentes de zero.

Agora calcula-se a seguinte estatística:

$$T_2 = \frac{12}{pk(k+1)} \sum_{j=1}^{k} R_{.j}^2 - 3p(k+1),$$

onde: p = número de blocos(para usar blocos completos se considerou 3 anos)

k = número de tratamentos

$R_{.j}$ = soma dos postos (ordens) no j-ésimo bloco

Assim a Estatística encontrada foi:

$T_2 = 25{,}03$

Para se verificar a significância do teste compara-se o valor de T_2 com o valor tabelado para o teste a 95% de confiança. Como não há valor tabelado para k=12, então aproxima-se o valor do teste a uma distribuição χ^2 com k-1 graus de liberdade.

$\chi^2_{11(0,05)} = 19{,}70 < T_2 = 25{,}03$ então rejeita-se H_0, isto é, aceita-se a existência da sazonalidade na série com 95% de confiança.

Através do gráfico dos valores da série, pode-se observar a sazonalidade nos dados da série e particularmente que os valores do ano de 1982 apresentaram um comportamento bem atípico apesar de não se encontrar uma razão aparente para tanto.

O primeiro modelo testado foi feito nos mesmos moldes da série anterior, isto é, a função que melhor se ajustava para se encontrar a componente tendencial foi a função exponencial.

A funçao encontrada foi a seguinte:

$$\hat{Y} = 8\,365\,777{,}49(\,0{,}9981821571)^{X}$$

Para a parte sazonal da série fez-se médias móveis de tamanho 12 e novamente média móvel de tamanho 2 para centrar os valores.

Assim os valores obtidos para os índices sazonais foram os seguintes:

Mês	Índice sazonal
Jan.	0,982856771
Fev.	0,876685620
Mar.	1,082892563
Abr.	0,982191966
Maio	0,958483360
Jun.	0,995358403
Jul.	1,026341148
Ago.	1,037217981
Set.	1,001317303
Out.	1,010546408
Nov.	1,018214042
Dez.	1,027894436

Para a componente cíclica da série optou-se pela não inclusão dela no modelo, já que o período intervalar de tempo da série é muito pequeno para se detectar esta componente.

O erro quadrático médio encontrado foi de $1{,}62303.10^{11}$.

Apesar deste erro quadrático médio ser menor que o encontrado na série anterior, os valores desta série são bem menores e este erro é bem mais significativo que o anterior. Isto se observou quando da estimativa dos valores da série, estes não apresentaram resultados satisfatórios.

Testou-se novamente, pelas mesmas razões da série anterior, um modelo linear múltiplo para a série. Neste modelo utilizou-se novamente as variáveis enumerativas que para a série anterior se mostraram eficientes.

Este modelo apresentou um erro quadrático médio da ordem de $1{,}35674.10^{11}$ que é melhor que o obtido anteriormente mas este modelo apresentou também um $R^2 = 56{,}65\%$, o que representa um coeficiente muito pouco significativo, isto é, as variações das variáveis enumerativas e a variação da variável tempo explicam 56,65% das variações da variável quilometragem percorrida pelos ônibus.

Portanto testou-se também o modelo de regressão exponencial múltiplo, este apresentou um $R^2 = 56{,}91\%$; pouco distinto do modelo anterior.

Assim resolveu-se testar o modelo de regressão Logarítmica múltiplo, do tipo:

$$Y = a_0 X^{a_1} \sum_{i=3}^{13} a_i^{b_i} \quad ,$$

obtendo-se então um $R^2 = 51{,}80\%$, índice inferior aos obtidos anteriormente.

Desta forma voltou-se a observar o gráfico da série e nota-se que para os meses de Janeiro de 1981 e Março de 1982, os valores são, muito diversos dos valores obtidos para os mesmos meses em outros anos.

Assim sendo pode-se pensar que deva ter acontecido erro na transferência dos dados. Pode também ter acontecido alguma coisa não aleatória que influenciou nos resultados obtidos.

O procedimento inicialmente adotado foi o seguinte:

A média do processo (da série) foi de 8 034 059,934 e o seu desvio padrão 585 926,1331, assim construiu-se um intervalo para os valores da série da seguinte maneira:

A partir da média, adicionou-se e diminuiu-se dois desvios-padrões, construindo-se assim um intervalo para os valores da série com 95,45% de confiança.

$(\mu \pm 2\sigma) = (8\,034\,059{,}934 \pm 2 \times 585\,926{,}1331)$

$(6\,842\,781{,}182 \; ; \; 9\,186\,485{,}715)$

Desta forma pode-se observar que os valores dos meses de Janeiro de 1981 e Março de 1982 estavam fora deste intervalo, assim também como um terceiro ponto que era o mês de Fevereiro de 1983.

Portanto estes valores foram substituidos pelas médias dos meses correspondentes em anos posterior e anterior.

Com a nova série, calculou-se novamente a regressão, obtendo-se os seguintes resultados:

- Para a regressão linear múltipla R^2 = 64,09%
- Para a regressão exponencial múltipla R^2 = 64,37%

Com estes resultados observa-se que está ocorrendo uma sensível melhora nos valores de R^2 para a série, apesar destes não serem ainda significativos.

Para melhorar os resultados optou-se por eliminar da série o ano de 1982, por este apresentar-se de maneira atípica com relação aos outros anos da série. Assim sendo os novos resultados obtidos foram os seguintes:

- Para a regressão linear mútipla R^2 = 81,04%
- Para a regressão exponencial múltipla R^2 = 81,17%

Agora sim observa-se uma sensível melhora nos valores de R^2, onde para ambos os casos pode-se verificar que a retirada dos valores da série para 1982 se mostrou satisfatória.

Para melhor se prever os valores da série e obter resultados mais confiáveis assumindo que o modelo será incrementado com os valores futuros da série, resolveu-se finalmente testar o R^2 para o caso de se considerar os valores da série a partir de Janeiro de 1983. Apesar de não se obter dois períodos completos da série que seriam dois anos.

Assim finalmente os resultados obtidos foram:

- Para a regressão linear múltipla R^2 = 88,13%
- Para a regressão exponencial múltipla R^2 = 90,39%

Em vista destes resultados optou-se finalmente em utilizar o modelo de regressão exponencial múltiplo, por este apresentar um R^2 de 90,39% que foi o melhor resultado obtido entre os modelos apresentados.

Agora ilustro através da tabela 2 a maneira pela qual estimou se os valores do modelo:

Tabela da quilometragem percorrida pelos ônibus e var. enumerativas
TABELA 2

t	Y(t)	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
1	7 619 079	0	0	0	0	0	0	0	0	0	0	0
2	7 157 611	1	0	0	0	0	0	0	0	0	0	0
3	8 110 750	0	1	0	0	0	0	0	0	0	0	0
4	7 465 941	0	0	1	0	0	0	0	0	0	0	0
5	8 045 882	0	0	0	1	0	0	0	0	0	0	0
6	7 717 844	0	0	0	0	1	0	0	0	0	0	0
7	7 744 964	0	0	0	0	0	1	0	0	0	0	0
8	8 120 721	0	0	0	0	0	0	1	0	0	0	0
9	7 714 779	0	0	0	0	0	0	0	1	0	0	0
10	7 789 594	0	0	0	0	0	0	0	0	1	0	0
11	7 570 045	0	0	0	0	0	0	0	0	0	1	0
12	8 023 469	0	0	0	0	0	0	0	0	0	0	1
13	7 828 908	0	0	0	0	0	0	0	0	0	0	0
14	7 164 236	1	0	0	0	0	0	0	0	0	0	0
15	7 806 781	0	1	0	0	0	0	0	0	0	0	0
16	7 525 458	0	0	1	0	0	0	0	0	0	0	0
17	7 973 065	0	0	0	1	0	0	0	0	0	0	0
18	7 616 850	0	0	0	0	1	0	0	0	0	0	0

CONT:

t	Y(t)	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}
19	7 850 552	0	0	0	0	0	1	0	0	0	0	0
20	7 944 548	0	0	0	0	0	0	1	0	0	0	0
21	7 335 790	0	0	0	0	0	0	0	1	0	0	0
22	7 869 944	0	0	0	0	0	0	0	0	1	0	0

FONTE: Dados brutos - SMT

Os resultados obtidos para os parâmetros foram os seguintes:

$\text{Log } a_1 = 6\ 88964945$ $\qquad a_1 = 7\ 756\ 208{,}049$

$\text{Log } a_2 = -2{,}639452131.10^{-4}$ $\qquad a_2 = 0{,}9993924284$

$\text{Log } a_3 = -3{,}256892019.10^{-2}$ $\qquad a_3 = 0{,}9277502495$

$\text{Log } a_4 = 1{,}349256303.10^{-2}$ $\qquad a_4 = 1{,}0315554150$

$\text{Log } a_5 = -1{,}220125276.10^{-2}$ $\qquad a_5 = 0{,}9722965557$

$\text{Log } a_6 = 1{,}685342945.10^{-2}$ $\qquad a_6 = 1{,}0395692610$

$\text{Log } a_7 = -1{,}846407835.10^{-3}$ $\qquad a_7 = 0{,}9957575138$

$\text{Log } a_8 = 5{,}741637378.10^{-3}$ $\qquad a_8 = 1{,}0133083870$

$\text{Log } a_9 = 1{,}887767459.10^{-2}$ $\qquad a_9 = 1{,}0444260000$

$\text{Log } a_{10} = -9{,}305081696.10^{-3}$ $\qquad a_{10} = 0{,}9788021583$

$\text{Log } a_{11} = 8{,}316906017.10^{-3}$ $\qquad a_{11} = 1{,}0193349290$

$\text{Log } a_{12} = -7{,}647591548.10^{-3}$ $\qquad a_{12} = 0{,}9825449062$

$\text{Log } a_{13} = 1{,}788007167.10^{-2}$ $\qquad a_{13} = 1{,}0420296380$

Que são os parâmetros do modelo escolhido :

$$Y = a_1 a_2^X \prod_{i=3}^{13} a_i b_i$$

6 - CONCLUSÕES

Dos modelos inicialmente utilizados para se estimar a série de número de passageiros, o que apresentou melhores resultados foi o que se utilizou das variáveis enumerativas. Estas se mostraram mais eficientes do que o modelo clássico de decomposição utilizando um processo com médias móveis para a obtenção dos índices sazonais. A eficiência no caso foi estabelecida pela comparação dos erros quadráticos médios das séries estimadas por estes modelos.

No primeiro caso, isto é, modelo clássico de decomposição usando-se médias móveis de tamanho 12, para a obtenção dos índices sazonais e uma função exponencial para a tendência, verificou-se um erro quadrático médio de $3{,}74185.10^{11}$, enquanto que para o caso do modelo exponencial com o uso de variáveis enumerativas, este erro diminuiu para $3{,}38963.10^{11}$.

No entanto o modelo utilizado foi o de regressão linear, visto que o seu R^2 não apresentou diferença significativa com o da regressão exponencial e é de mais fácil manejo. O R^2 no primeiro caso foi de 92,69% e no segundo caso 92,93%.

Portanto, a seguir apresentamos a série de número de passageiros transportados, tabela 3, com os valores da série de Outubro de 1980 até Outubro de 1984 e as respectivas estimativas dos pontos da série, juntamente com as previsões desta a partir de Novembro de 1984 até Outubro de 1985.

Tabela comparativa dos valores reais e estimados para a série Nº pas.

TABELA 3

t	Y(t)	$\hat{Y}(t)$	$\lvert Y - \hat{Y} \rvert$	Erro Percentual
1	28 049 167	28 010 980	38 187	0,14
2	26 444 380	26 800 895	356 515	1,35
3	28 199 158	28 479 484	280 326	0,99
4	25 784 804	25 391 933	392 871	1,52
5	21 872 035	21 849 009	23 026	0,11
6	27 581 215	28 177 610	596 395	2,16
7	26 358 300	26 076 613	281 687	1,07
8	27 405 289	27 190 277	215 012	0,78
9	26 466 381	25 629 631	836 750	3,16
10	27 296 266	26 494 809	801 457	2,94
11	27 118 171	27 704 684	586 513	2,16
12	25 436 365	25 832 940	396 575	1,56
13	27 183 723	26 931 535	252 188	0,93
14	26 019 715	25 721 450	298 265	1,15
15	27 193 754	27 400 039	206 285	0,76
16	24 319 670	24 312 488	7 182	0,03
17	20 529 113	20 769 564	240 451	1,17
18	28 310 551	27 098 165	1 212 386	4,28
19	25 211 255	24 997 168	214 087	0,85
20	26 003 741	26 110 832	107 091	0,41
21	24 182 466	24 550 186	367 720	1,52
22	25 819 330	25 415 364	403 966	1,56
23	26 910 746	26 625 239	285 507	1,06
24	24 623 300	24 753 495	130 195	0,53
25	25 199 993	25 852 090	652 097	2,59
26	24 888 636	24 642 005	246 631	0,99
27	26 712 649	26 320 594	392 055	1,47
28	22 554 710	23 233 043	678 333	3,01
29	18 798 948	19 690 119	891 171	4,74
30	26 878 884	26 018 719	860 165	3,20

t	Y(t)	Ŷ(t)	$\|Y - \hat{Y}\|$	Erro Percentual
31	23 462 178	23 917 723	455 545	1,94
32	25 022 266	25 031 387	9 121	0,04
33	23 248 568	23 470 741	222 173	0,96
34	22 727 771	24 335 919	1 608 148	6,92
35	25 397 800	25 545 794	147 994	0,58
36	23 678 409	23 674 050	4 359	0,02
37	23 934 030	24 772 645	838 615	3,50
38	23 374 180	23 562 560	188 380	0,81
39	25 335 663	25 241 149	94 514	0,37
40	22 431 878	22 153 598	278 280	1,24
41	19 719 271	18 610 674	1 108 597	5,62
42	23 463 118	24 939 274	1 476 156	6,29
43	22 798 049	22 838 278	40 229	0,18
44	23 853 141	23 951 942	98 801	0,41
45	22 144 440	22 391 296	246 856	1,11
46	23 659 199	23 256 474	402 725	1,70
47	24 915 349	24 466 349	449 000	1,80
48	23 117 017	22 594 605	522 412	2,26
49	24 893 536	23 693 200	1 200 336	4,82
50		22 483 115		
51		24 161 703		
52		21 074 153		
53		17 531 229		
54		23 859 829		
55		21 758 833		
56		22 872 496		
57		21 311 851		
58		22 177 029		
59		23 386 904		
60		21 515 160		
61		22 613 754		

FONTE: SMT

NOTA: Os valores da série para os tempos 50 a 61 são os valores previstos para os períodos de nov/84 a out/85.

Quanto a série quilometragem percorrida pelos ônibus, optou-se finalmente por utilizar o modelo de regressão exponencial múltiplo, com um R^2 de 90,39%. A seguir, tabela 4, apresentamos os valores da série juntamente com as estimativas dos pontos da série e as previsões de Novembro de 1984 até Outubro de 1985.

Tabela comparativa dos valores reais e estimados para Quil. Perc.

TABELA 4

t	Y(t)	$\hat{Y}(t)$	$\lvert Y - \hat{Y} \rvert$	Erro Percentual
1	7 619 079	7 751 496	123 417	1,74
2	7 157 611	7 187 083	29 472	0,41
3	8 110 750	7 986 384	124 366	1,53
4	7 465 941	7 523 023	57 082	0,76
5	8 045 882	8 038 651	7 231	0,09
6	7 717 844	7 695 190	22 654	0,29
7	7 744 964	7 826 065	81 101	1,05
8	8 120 721	8 061 495	59 226	0,73
9	7 714 779	7 550 381	164 398	2,13
10	7 789 594	7 858 269	68 675	0,88
11	7 570 045	7 570 045	0	0
12	8 023 469	8 023 469	0	0
13	7 828 908	7 695 169	133 739	1,71
14	7 164 236	7 134 857	29 379	0,41
15	7 806 781	7 928 350	121 569	1,56
16	7 525 458	7 468 357	57 101	0,76
17	7 973 065	7 980 237	7 172	0,09
18	7 616 850	7 639 273	22 423	0,29

CONT:

t	Y(t)	$\hat{Y}(t)$	$\lvert Y - \hat{Y} \rvert$	Erro Percentual
19	7 850 552	7 769 197	81 355	1,04
20	7 944 548	8 002 915	58 367	0,73
21	7 335 790	7 495 516	159 726	2,18
22	7 869 944	7 801 167	68 777	0,87
23		7 515 037		
24		7 965 166		
25		7 639 252		
26		7 083 012		
27		7 870 739		
28		7 414 088		
29		7 922 249		
30		7 583 762		
31		7 712 742		
32		7 944 762		
33		7 441 049		
34		7 744 479		

FONTE: SMT

NOTA: Os valores da série para os tempos 23 a 34 são os previstos para nov/84 a out/85.

ANEXO

TADELA 5

Série mensal do número de passageiros transportados pelos ônibus no município de Porto Alegre, no período de outubro de 1980 até outubro de 1984.

t	Y(t)
1	28 049 167
2	26 444 380
3	28 199 158
4	25 784 804
5	21 872 035
6	27 581 215
7	26 358 300
8	27 405 289
9	26 466 381
10	27 296 266
11	27 118 171
12	25 436 365
13	27 183 723
14	26 019 715
15	27 193 754
16	24 319 670
17	20 529 113
18	28 310 551
19	25 211 355
20	26 003 741
21	24 182 466
22	25 819 330

CONT:

t	Y(t)
23	26 910 746
24	24 623 300
25	25 199 993
26	24 888 636
27	26 712 649
28	22 554 710
29	18 798 948
30	26 878 884
31	23 463 178
32	25 022 266
33	23 248 668
34	22 727 771
35	25 397 800
36	23 678 409
37	23 934 030
38	23 374 180
39	25 335 663
40	22 431 878
41	19 719 271
42	23 463 118
43	22 798 049
44	23 853 141
45	22 144 440
46	23 659 199
47	24 915 349
48	23 117 017
49	24 893 536

FONTE: SMT

TABELA 6

Série mensal do número de quilômetros percorridos pelos ônibus no município de Porto Alegre, desde outubro de 1980 até outubro de 1984.

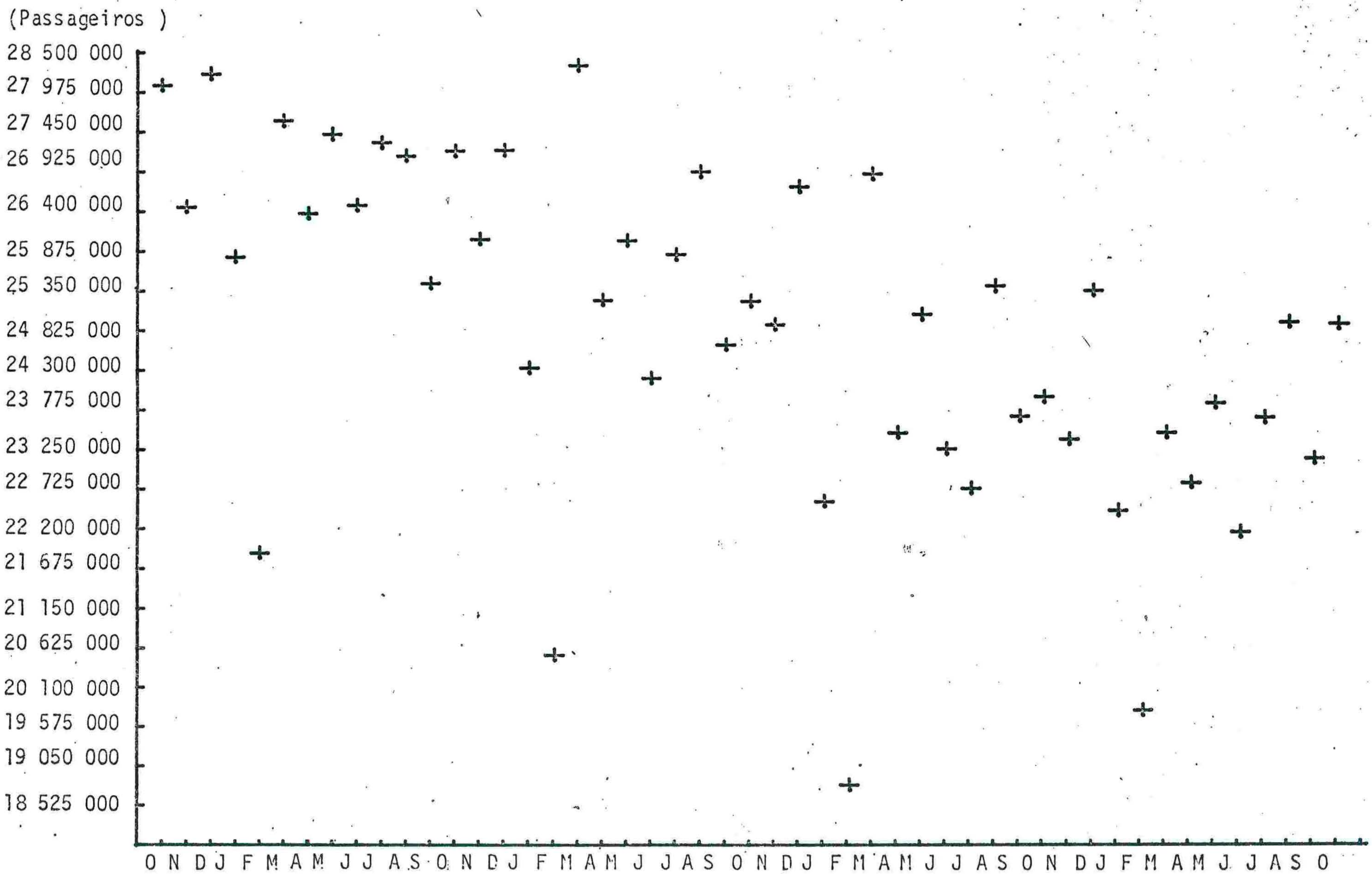
t	Y(t)
1	8 907 859
2	8 183 570
3	8 812 219
4	6 669 342
5	7 732 029
6	8 816 358
7	8 426 461
8	8 036 564
9	8 134 705
10	8 453 521
11	8 372 241
12	8 112 276
13	8 323 043
14	7 973 410
15	8 441 120
16	8 096 474
17	7 150 986
18	9 969 821
19	8 149 983
20	7 208 743
21	8 300 245
22	8 666 977
23	8 580 184
24	8 337 649
25	8 233 324
26	8 939 891
27	8 206 649
28	7 619 079
29	6 642 195
30	8 110 750

CONT:

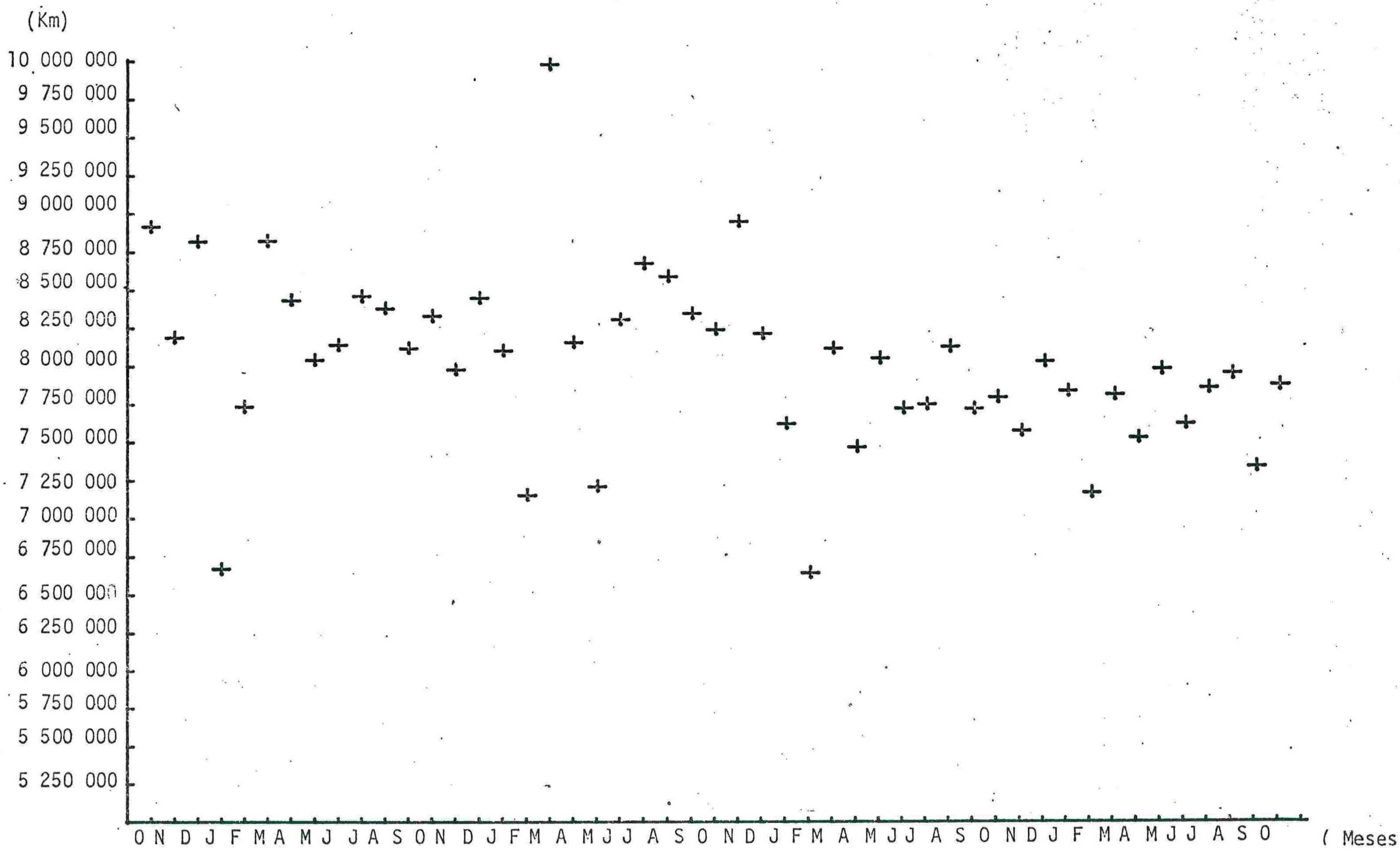
t	Y(t)
31	7 465 941
32	8 045 882
33	7 717 844
34	7 744 964
35	8 120 721
36	7 714 779
37	7 789 594
38	7 570 045
39	8 023 469
40	7 828 908
41	7 164 236
42	7 806 781
43	7 525 458
44	7 973 065
45	7 616 850
46	7 850 552
47	7 944 548
48	7 335 790
49	7 869 944

FONTE: SMT

GRÁFICO DO NÚMERO DE PASSAGEIROS TRANSPORTADOS EM PORTO ALEGRE (OUT. 80 - OUT. 84)



GRÃFICO DA QUILOMETRAGEM PERCORRIDA PELOS ÔNIBUS EM PORTO ALEGRE (OUT. 80 - OUT. 84)



BIBLIOGRAFIA:

HOFFMANN, Rodolfo & VIEIRA, Sônia. Análise de Regressão. São Paulo, Ed. da Universidade de São Paulo, 1977. 339 p.

WONNACOTT, Ronald J. & WONNACOTT, Thomas H. Econometria. Rio de Janeiro, 2 ed. , 1978. 424 p.

MORETTIN, Pedro Alberto & TOLOI, Clélia Maria de Castro. 13º Colóquio Brasileiro de Matemática. Rio de Janeiro, Vol. 1, 1981, 356 p.

PEREIRA, Basílio de B. Séries temporais multivariadas. 6º SINAPE, 1983.

SOUZA, Reinaldo. Metodologias para a análise e previsão de séries temporais univariadas e multivariadas. Revista de econometria, ano 1, nº 2, novembro de 1981.

CHAO, Lincoln L. Estadistica para las ciências administrativas.Bogotá, McGraw Hill, 1975.